Reliable Deniable Communication: Hiding Messages in Noise

Pak Hou Che, Mayank Bakshi, Sidharth Jaggi, Mahdi Jafari Siavoshani
The Chinese University of Hong Kong

***Abstract*—A transmitter Alice may wish to *reliably* transmit a message to a receiver Bob over a binary symmetric channel (BSC), while simultaneously ensuring that her transmission is *deniable* from an eavesdropper Willie. That is, if Willie listening to Alice's transmissions over a "significantly noisier" BSC than the one to Bob, he should be unable to estimate even whether Alice is transmitting. Even when Alice's (potential) communication scheme is publicly known to Willie (with *no* common randomness between Alice and Bob), we prove that over n channel uses Alice can transmit a message of length $\mathcal{O}(\sqrt{n})$ bits to Bob, deniably from Willie. We also prove information-theoretically order-optimality of our results.**

I. Introduction

Alice is in jail, and may wish to communicate reliably with Bob in the neighboring cell, over n uses of a noisy BSC (if she stays silent, the input to the channel is all zeroes). Unfortunately, the warden Willie is monitoring Alice (though his observations are significantly noisier, since his CCTV camera is low-quality).[1] Willie only wishes to detect Alice's "transmission status" (*i.e.*, he only wants to know whether she's talking or not, and doesn't necessarily care *what* she's saying). Hence Alice wishes to use a communication scheme that is "deniable from Willie", *i.e.* Willie's *best* estimate of Alice's transmission status should be essentially statistically independent of his observations. In this work we show:

1) *Deniability – outer bound on codeword weight*: If the binary code Alice uses to encode her message contains many "high-weight codewords" (that is, have weight that is $\omega(\sqrt{n})$ over n channel uses), then her communication scheme cannot be deniable. In particular, Willie can simply count the number of non-zero symbols he observes to estimate, fairly accurately, Alice's transmission status. Hence, for deniability, Alice's code should have mostly "low-weight codewords".
2) *Reliability/deniability – outer bound on throughput*: Since the link to Bob is *also* noisy, Alice needs to code. We use information-theoretic inequalities to demonstrate that for any code is simultaneously highly deniable and reliable, a message of at most $\mathcal{O}(\sqrt{n})$ bits can be transmitted by Alice over n channel uses.[2]
3) *Reliability/deniability – achievable scheme*: If Willie's BSC is "sufficiently noisier" (in a precise sense that we quantify later) than Bob's BSC, then we design a communication scheme (publicly known to all parties – Alice, Bob and Willie) such that:
   - *Throughput*: It encodes a message with $\mathcal{O}(\sqrt{n})$ bits in n channel uses (or $\mathcal{O}(\sqrt{n})\log(n)$ bits when the channel to Bob is noiseless but the channel to Willie is still noisy).
   - *Reliability*: It enables Bob to correctly reconstruct Alice's message with high probability.
   - *Deniability*: Willie's best estimate of Alice's transmission status is essentially statistically independent of his observations of his channel.

The first two results above are analogues (for the scenario of the BSCs considered in this work) of theorems in recent work that motivated our work (the corresponding results for Additive White Gaussian Noise (AWGN) channels proved in [1, 2]). The last, corresponding to construction of "reliable and deniable public codes" is the main result of this work.

In particular, we stress again that in our model (unlike the models in most prior work) there is no common randomness that is hidden from Willie that Alice and Bob can leverage. The only asymmetry between Bob's and Willie's estimation abilities arises from the fact that Willie's observations of Alice's (possible) transmissions are noisier than Bob's. Hence the fact that we demonstrate the existence of *public* codes satisfying Result 3 above is a significant strengthening of the model in [1, 2], wherein in general common randomness is required, and consumed at a rate greater than the throughput of the reliable/deniable communication in the first place!

Also, in our model (and also in the model of [1, 2], but *not* in the vast majority of steganographic models), Alice's default transmission if she has nothing to say, is nothing. This default silence of Alice makes it challenging to hide the fact that she is *not* silent when she actually has something to say. The only reason we are able to achieve a non-zero throughput is due to the fact that Willie's observations of Alice's potential transmissions are noisy (and in particular, significantly noisier than Bob's). Hence the subtitle of this work – "hiding messages in noise".

II. Related work

We consider is a variant of the classical *steganography* problem, but we argue that our results are more "realistic" in some settings, and also technically more challenging to prove.

[1] If Willie's is as good as Bob's channel, then cno communication that is simultaneously reliable and deniable is possible, since Willie can use whatever decoding strategy Bob does.

[2] Note that this implies that Alice's rate decays to 0 asymptotically in n. Hence in this work we usually scale Alice's "throughput" (the number of bits in her message) with respect to $\sqrt{n}$, to obtain a quantity we call the "relative throughput".

A. Steganography

Most steganographic models make at least one of the following assumptions (*none of which we make*):

- **(A1) Non-zero covertext/stegotext:** In almost all works in the literature, Alice's *default* transmission (even if she has no hidden message to transmit to Bob) is usually non-zero, by this assumption. Many works characterize the capacity of various steganographic problems – see for instance [3–7]. An important exception to the non-zero covertext assumption occurs in the work of Bash, Goeckel and Towsley [1, 2] – we discuss this work in depth below.
- **(A2) Shared secret key/common randomness:** Many steganographic protocols require a key (that is often almost as large as the message being communicated) that is shared between Alice and Bob, and is kept secret from Willie, in advance of any communication. A variety of examples of such protocols can be found in, for example [8] or [3]. However, not all works make this assumption. Some exceptions to this assumption of a shared secret include works by [4, 7].
- **(A3) Noiseless communication:** Some works consider a model wherein the communication channel between Alice and Bob is noiseless. In some such scenarios, the optimal throughput can sometimes be boosted by a multiplicative factor of $\log n$ (for instance [8, Chapters 8 and 13]). Some model do consider noise – for instance due to an actively jamming warden (for instance [3]). In other models this may simply be random channel noise (for instance the work of [1, 2], and our work here).

B. The Square Root Law

The "Square Root Law" (often abbreviated as SRL in the literature [5, 9, 10]) can be perhaps characterized as an observation that in a variety of steganographic models, the throughput (the length of the message that Alice can communicate deniably and reliably with Bob) scales as $\mathcal{O}(\sqrt{n})$ (here n is the number of "channel uses" that Alice has access to).

We note that in our setting (and also that of [1, 2]), our throughput does indeed provably scale as the square-root of the number of channel uses. However, the critical reason underlying this scaling is that we consider the scenario wherein the covertext is all-zero – Alice must "whisper very softly", since she has no excuse if Willie hears something that cannot be explained by the noise on the channel to him.

C. The work of Bash, Goeckel and Towsley

The results and techniques closest to those in this work (and indeed the starting-point of our investigations) are those of [1, 2]. However, there are important differences in the models.

- **Public codes vs. shared secret keys:** The critical difference between our model and that of [1, 2] is that in our setting there is no shared secret key between Alice and Bob that is hidden from Willie. Hence our codes are "public". A setting wherein Alice's consumption of secret keys happens significantly faster ($\Omega(n)$) than her throughput ($\mathcal{O}(\sqrt(n))$) to Bob (as in [1, 2]) is not sustainable. Proving this required novel and powerful techniques that might be of independent interest.
- **Discrete vs. continuous channels:** In our work all channels are discrete (finite input and output alphabets) – in particular, for ease of presentation we focus here on the case wherein Alice's transmissions pass through independent BSCs. In contrast, the results of [1, 2] are for channels wherein the noise is AWGN.[3]

III. Model

A. Notation

Calligraphic symbols such as $\mathcal{C}$ denote sets. Boldface upper-case symbols such as $\mathbf{X}$ denote random variables, boldface lower-case symbols such as $\mathbf{x}$ denote particular instantiations of those random variables. Vectors are denoted by an arrow above a symbol, such as in $\vec{\mathbf{x}}$. For notational convenience, in this work, unless otherwise specified, all vectors are of length n, where n corresponds to the *block-length* (number of channel uses). All logarithms in this work are binary, unless otherwise stated. The *Hamming weight* (number of non-zero entries) of a vector $\vec{\mathbf{x}}$ is denoted by $wt(\vec{\mathbf{x}})$, and the *Hamming distance* between two vectors $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ of equal length (the number of corresponding entries in which $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$ differ) is denoted by $d(\vec{\mathbf{x}}, \vec{\mathbf{y}})$. For any two numbers a and b in the interval $[0, 1]$, we use $a * b$ to denote *binary convolution* of these two numbers, defined as $a(1-b)+b(1-a)$ – this corresponds to the noise parameter of the BSC comprising of a BSC(a) followed by a BSC(b). As is standard in an information-theoretic context, the notation $H(.)$ corresponds to the *(binary) entropy function*, $H(.|.)$ to *conditional entropy*, $I(.;.))$ to *mutual information*, and $D(.||.)$ to the *Kullback-Leibler divergence* between two distributions.

B. Communication System

The transmitter Alice is connected via a binary-input binary-output broadcast medium to the receiver Bob and the warden Willie. The channel from Alice to Bob is a Binary Symmetric Channel with crossover probability p_b (henceforth denoted BSC(p_b)). The channel from Alice to Willie is a BSC(p_w). By assumption, the noise on the two channels is independent, $p_b < p_w$, and all parties (Alice, Bob and Willie) know the channel parameters p_b and p_w.

Alice (potentially) wishes to communicate a *message* m uniformly at random from a set $\{1, \ldots, N\}$ to Bob – the symbol $\mathbf{M}$ denotes the random variable corresponding to Alice's message. For notational convenience we say that if Alice does not wish to communicate with Bob, her message is 0. Equivalently, if Alice does have a message she wishes to communicate to Bob, then a certain arbitrary binary variable $\mathbf{T}$ equals 1. Otherwise, $\mathbf{T}$ equals 0. Only Alice knows the value of $\mathbf{T}$ *a priori*.

[3] It is conceivable that our proof techniques also carries over to the AWGN model of [1, 2], but significant extensions would be required to translate our techniques from the discrete world over to the continuous version.

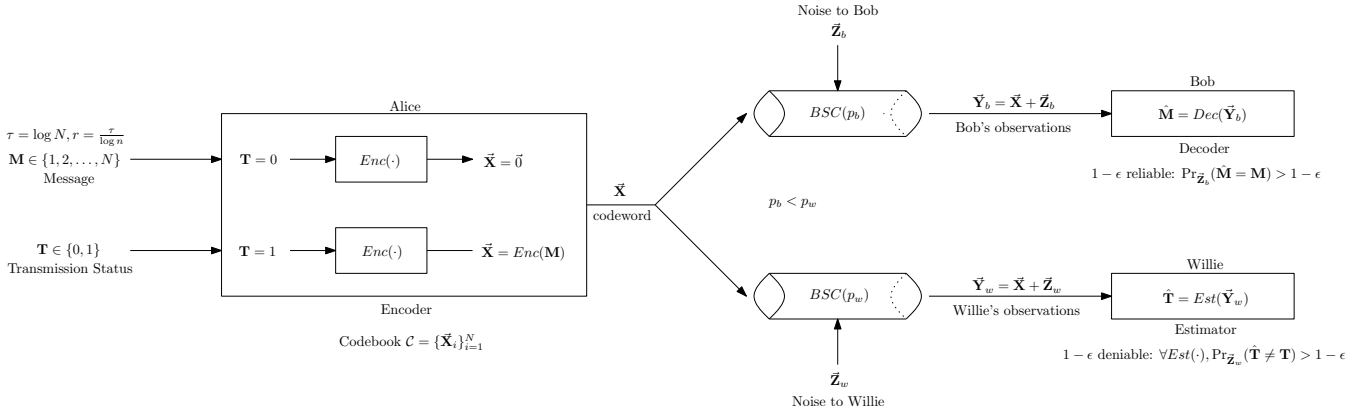

Fig. 1. Depending on her transmission status $\mathbf{T}$, Alice either broadcasts the all-zero dector $\vec{\mathbf{0}}$, or encodes her messages $\mathbf{M}$ into codewords $\vec{\mathbf{X}} = Enc(\mathbf{M})$ from her codebook $\mathcal{C}$. The codebook has $2^\tau = 2^{r\sqrt{n}}$ codewords, where the throughput τ scales as $\mathcal{O}(\sqrt{n})$. Bob receives $\vec{\mathbf{Y}}_b$ over a BSC(p_b), and Willie receives a noisier version $\vec{\mathbf{Y}}_w$ over a BSC(p_w), where $p_b < p_w$. It is desired that the codebook $\mathcal{C}$ be both reliable (*i.e.* Bob is able to decode $\mathbf{M}$ as $\hat{\mathbf{M}}$ with a "small" probability error) and deniable (*i.e.* Willie's observations give him essentially no information about Alice's transmission status.)

Alice encodes each message $m \in \{1, \ldots, N\}$ into a length-n binary *codeword* $\vec{\mathbf{x}}(m) = (\mathbf{x}(m, 1), \ldots, \mathbf{x}(m, n))$ using an *encoder* $Enc(.) : \{0\} \cup \{1, \ldots, N\} \rightarrow \{0, 1\}^n$. To simplify notation, we often denote $\vec{\mathbf{x}}(m)$ and $\vec{\mathbf{X}}(\mathbf{M})$ as $\vec{\mathbf{x}}$ and $\vec{\mathbf{X}}$ respectively, wherever it does not cause confusion. The encoder is required to satisfy the condition that the 0 message is always encoded to the length-n zero-vector $\vec{\mathbf{0}}$. The set $\{\vec{\mathbf{x}}(1), \ldots, \vec{\mathbf{x}}(N)\}$ of possible non-zero codewords that are the outputs of Alice's encoder is denoted by the *codebook* $\mathcal{C}$. The *throughput* τ of Alice's codebook is defined as $\log N$, and the *relative throughput* r of Alice's codebook is defined as $\log N/\sqrt{n}$. We note that in this work, the throughput of the codes we consider typically scale with the block-length n as $\mathcal{O}(\sqrt{n})$. This corresponds to a "rate" that decays to zero as n increases without bound, rather than converging to a constant as is common in many other communication settings. Hence we deliberately consider the relative throughput rather than the rate of our codes.

Bob receives the length-n binary vector $\vec{\mathbf{Y}}_b$. Here $\vec{\mathbf{Y}}_b = \vec{\mathbf{X}} \oplus \vec{\mathbf{Z}}_b$, where $\vec{\mathbf{Z}}_b$ denotes the noise added by the BSC(p_b) channel between Alice and Bob. Bob uses his *decoder* $Dec(.) : \{0, 1\}^n \rightarrow \{0\} \cup \{1, \ldots, N\}$ to generate his *estimate of Alice's message* as $\hat{\mathbf{M}} = Dec(\vec{\mathbf{Y}}_b)$. Bob's *probability of decoding error when Alice is transmitting*, $P_{e,T=1}$, is defined as $\Pr_{\mathbf{M}, \vec{\mathbf{Z}}_b}(\hat{\mathbf{M}} = \vec{\mathbf{0}} | T = 1) + \Pr_{\mathbf{M}, \vec{\mathbf{Z}}_b}(\hat{\mathbf{M}} \neq \mathbf{M} | T = 1)$. Bob's *probability of decoding error when Alice is not transmitting*, $P_{e,T=0}$, is defined as $\Pr_{\vec{\mathbf{Z}}_b}(\hat{\mathbf{M}} \neq \vec{\mathbf{0}} | T = 0)$. Bob's *overall probability of decoding error*, P_e, is defined as $P_{e,T=1} + P_{e,T=0}$.We say Alice's codebook $\mathcal{C}$ is $(1-\epsilon)$-*reliable* if Bob's probability of decoding error is less than ϵ.

Willie knows *a priori* both $Enc(.)$ and $Dec(.)$ (and hence also $\mathcal{C}$). Willie receives the length-n binary vector $\vec{\mathbf{Y}}_w$. Here $\vec{\mathbf{Y}}_w = \vec{\mathbf{X}} \oplus \vec{\mathbf{Z}}_w$, where $\vec{\mathbf{Z}}_w$ denotes the noise added by the BSC(p_w) channel between Alice and Willie. Willie uses his *estimator* $Est_{\mathcal{C}}(.) : \{0, 1\}^n \rightarrow \{0, 1\}$ to generate his *estimate of Alice's transmission status* as $\hat{\mathbf{T}} = Est_{\mathcal{C}}(\vec{\mathbf{Y}}_w)$.

We use a hypothesis-testing metric to quantify the *deniability of Alice's codebook* $\mathcal{C}$. Let *the probability of false alarm* $\Pr_{\vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 1 | \mathbf{T} = 0)$ be denoted by $\alpha(Est_{\mathcal{C}}(.))$. Analogously, let *the probability of missed detection* $\Pr_{\mathbf{M}, \vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 0 | \mathbf{T} = 1)$ be denoted by $\beta(Est_{\mathcal{C}}(.))$ (and for a specific transmitted codeword m, we denote $\Pr_{\vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 0 | \mathbf{M} = m)$ by $\beta^{(m)}(Est_{\mathcal{C}}(.))$). The quantities $\alpha(Est_{\mathcal{C}}(.))$ and $\beta(Est_{\mathcal{C}}(.))$ denote respectively the probability that Willie guesses Alice is transmitting even if she is not, and the probability that Willie guesses Alice is not transmitting even though she actually is. We say Alice's codebook $\mathcal{C}$ is $(1-\epsilon)$-*deniable* if there is no estimator $Est_{\mathcal{C}}(.)$ for Willie such that $\alpha(Est_{\mathcal{C}}(.)) + \beta(Est_{\mathcal{C}}(.)) < 1 - \epsilon$.

For any block-length n, we say a corresponding codebook $\mathcal{C}$ is *simultaneously* $(1-\epsilon)$-*reliable and* $(1-\epsilon)$-*deniable* if it simultaneously ensures that Bob's probability of decoding error is at most ϵ, and has deniability $1 - \epsilon$. A relative throughput r is said to be *reliably and deniably achievable* if for any $\epsilon > 0$ there exists a sufficiently large block-length n and a corresponding codebook $\mathcal{C}$ with relative throughput r over that block-length that is simultaneously $(1-\epsilon)$-reliable and $(1-\epsilon)$-deniable. The *optimal relative throughput* is the supremum of all reliably and deniably achievable relative throughputs.

IV. Main results/high-level intuition

We now present our main results and corresponding intuition. Unfortunately space paucity means we are unable to present our full proofs here – **we strongly recommend our full technical report HERE to the interested reader.**

Theorem 1 below proves an outer bound on the median of the Hamming weights of codewords of any codebook $\mathcal{C}$ that has high deniability. The intuition is that if many codewords have "high" Hamming weight, with non-negligible probability the Hamming weight of Willie's observed vector $\vec{\mathbf{y}}_w$ will be above a carefully chosen threshold.

Theorem 1. *If more than half of the codewords in* $\mathcal{C}$ *are of weight greater than* $c_1\sqrt{n}$*, then the deniability of Alice's codebook* $\mathcal{C}$ *is less than* $1/2 + \frac{4p_w(1-p_w)}{c_1^2(1-2p_w)^2}$.

Theorem 2 below proves an upper bound on the throughput τ of any code that simultaneously has high reliability and deniability. The proof technique follows standard information-theoretic converse arguments, but we critically need to use the fact (from Theorem 1) that codes that have high deniability do not have too many codewords of high Hamming weight.

Theorem 2. *If a codebook $\mathcal{C}$ is simultaneously $(1-\epsilon)$-deniable and $(1-\epsilon)$-reliable, the relative throughput is at most*

$$\sqrt{p_w(1-p_w)}\frac{1-2p_b}{1-2p_w}(1-2\epsilon)^{-3/2}\log\left[\frac{1-2p_b}{p_b}\right]$$

Next we state and prove one of the main results of this work – namely, that randomly chosen codes (chosen from a suitable ensemble) are with high probability simultaneously highly reliable and highly deniable.

To understand our proof, it is instructive to first understand those of [1, 2]. The code constructions in [1, 2] use common randomness between Alice and Bob to coordinate which of an ensemble of possible codebooks Alice actually uses to communicate with Bob. Since Bob knows the common randomness, he knows which codebook Alice actually used. However, since the common randomness is kept secret from Willie, from his perspective, if Alice transmits a non-zero codeword, the probability distribution on his received $\vec{\mathbf{y}}_w$ is a *code ensemble average* $\Pr_{\mathcal{C},\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$. Using some elegant statistical properties of this ensemble average distribution, and the distribution $\Pr_{\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=0)$ (corresponding to Willie's observations if Alice does not transmit anything), [1, 2] demonstrate that Willie is essentially unable to learn anything about the binary random variable $\mathbf{T}$. Their proof can be essentially summarized in the statement that "the ensemble average codebook is highly deniable". The challenge in extending their proof technique to a public codebook is that this proof says nothing about the existence of a single, public, highly deniable codebook.

Our key idea is to extend the analysis by proving that the *actual* distribution $\Pr_{\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$ on $\vec{\mathbf{y}}_w$ if Alice transmits a non-zero codeword is tightly concentrated about its ensemble average. However, our first "naïve" attempts in using standard concentration inequalities were unsuccessful, since for any $\vec{\mathbf{y}}_w$ the probability (averaged over Alice's choice of message, channel noise, and over all codebooks) that Willie actually observes $\vec{\mathbf{y}}_w$ is exceedingly small (decaying at least exponentially in n).

Hence we proceed indirectly. We first note that it suffices to prove that $\Pr_{\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$ converges point-wise to its ensemble average for "typical" $\vec{\mathbf{y}}_w$ (since the bulk of the probability mass of the ensemble average distribution falls in a certain range). For any $\vec{\mathbf{y}}_w$ in this range, we prove that expected number of codewords at a certain distance range (corresponding to the "typical" noise patterns $\vec{\mathbf{Z}}_w$) of each $\vec{\mathbf{y}}_w$ is super-polynomial. For random variables with such "large" (super-polynomial) expectations, standard arguments suffice to prove concentration with probability that is super-

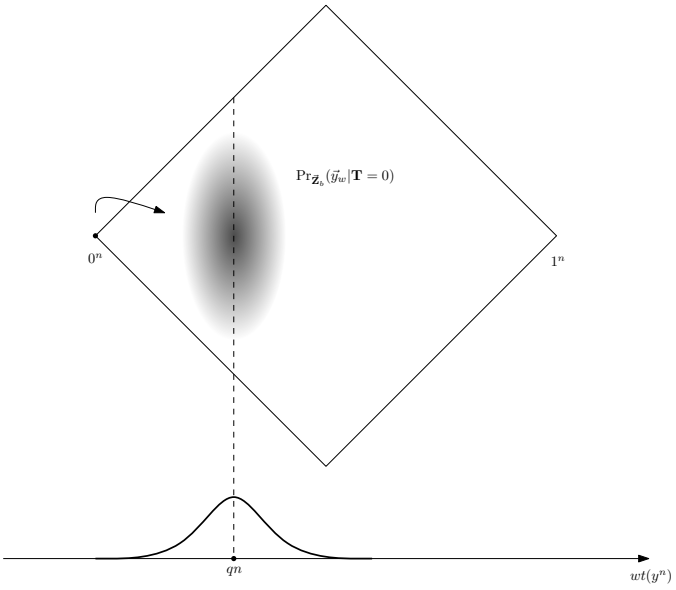

Fig. 2. The diamond at the top of this figure denotes the set (of size 2^n) of all possible $\vec{\mathbf{y}}_w$ that Willie may observe if Alice transmits nothing (her transmission status $\mathbf{T}=0$). In particular, the $\vec{\mathbf{y}}_w$ are arranged in a partial order, so that vectors with lower Hamming weight are to the left of vectors with higher Hamming weight. The curve at the bottom plots the probability distribution of observing $\vec{\mathbf{y}}_w$ of a particular weight. Since $\vec{\mathbf{X}}=0$, the "typical" $\vec{\mathbf{y}}_w$ that Willie observes are of weight approximately $p_w n$. Also, the probability distribution on $\vec{\mathbf{y}}_w$ equals $\Pr_{\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=0)$.

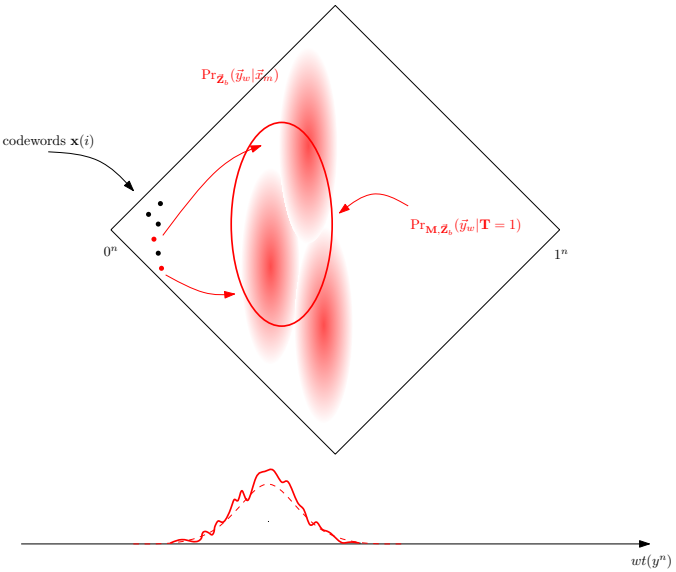

Fig. 3. The diamond at the top of this figure denotes the set (of size 2^n) of all possible $\vec{\mathbf{y}}_w$ that Willie may observe if Alice transmits a codeword uniformly at random from her codebook $\mathcal{C}$ (her transmission status $\mathbf{T}=1$). Since codewords in our codebook $\mathcal{C}$ have expected Hamming weight ρn, the "typical" $\vec{\mathbf{y}}_w$ that Willie observes are of weight approximately $(p_w * \rho)n$. The curve at the bottom plots the probability distribution of observing $\vec{\mathbf{y}}_w$ of a particular weight. In this case (unlike in Figure 2) the probability distribution on $\vec{\mathbf{y}}_w$ is somewhat "lumpy", since the probability that Willie observes a particular $\vec{\mathbf{y}}_w$ depends on the distribution of the Hamming distance between that particular $\vec{\mathbf{y}}_w$ and the set of codewords $\vec{\mathbf{x}} \in \mathcal{C}$. In particular, the probability distribution on $\vec{\mathbf{y}}_w$ equals $\Pr_{\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1) = \frac{1}{2^{rn}}\sum_{\vec{\mathbf{x}}\in\mathcal{C}}\Pr_{\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\vec{\mathbf{x}})$.

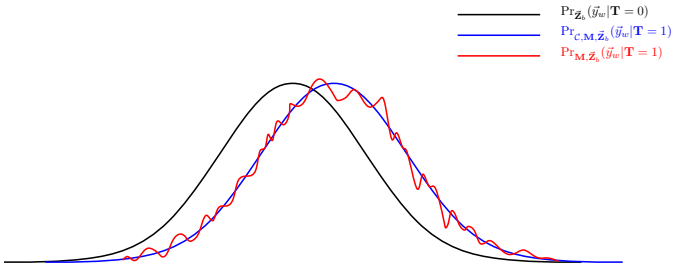

Fig. 4. Our proof that a random codebook $\mathcal{C}$ chosen with the "right" parameters (number of codewords, expected weight of codewords) proceeds as follows. We need to demonstrate that the probability distributions $\Pr_{\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=0)$ and $\Pr_{\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$ are "close" (in variational distance). However, since the latter distribution is complex (due to its dependence on the specific codebook $\mathcal{C}$), we do this comparison in two stages. We first demonstrate that the probability distribution $\Pr_{\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=0)$ and the *ensemble distribution* $\Pr_{\mathcal{C},\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$ (*i.e.* the weighted average over all possible codebooks $\mathcal{C}$ of the latter distribution) are "close". Next, we prove that with high probability over the choice of codebooks $\mathcal{C}$, the distribution of $\Pr_{\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$ is tightly concentrated around its expectation $\Pr_{\mathcal{C},\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$.

exponentially small in n. This allows us to show that with high probability over the ensemble average, a randomly chosen codebook satisfies the property that the number of codewords in "typical" Hamming shells around most "typical" $\vec{\mathbf{y}}_w$ are tightly concentrated around their expectations. Book-keeping calculations then enable us to show that this concentration in the distance-distribution of codewords translate to a pointwise concentration (with super-exponential probability) of $\Pr_{\mathbf{M},\vec{\mathbf{Z}}_b}(\vec{\mathbf{y}}_w|\mathbf{T}=1)$ to its ensemble average. This technique allows us to bypass the problem of the small expected values[4] of the random variables of primary interest (the probability of observing specific channel outputs if a specific codebook is used), by focusing instead on random variables with "large expected values"[5] (numbers of codewords of certain "types") that then enable us to recover the random variables of primary interest. One calculation that requires some care is that due to the low throughput of our codes (scaling as $\mathcal{O}(\sqrt{n})$) we need to define our typical sets carefully, to simultaneously ensure that they are high probability sets, but are also not "too large".

To complete the proof, we need to demonstrate that in fact a randomly chosen code (from the same ensemble as used to generate the highly deniable code above) is also highly reliable with sufficiently high probability (and hence a randomly chosen code is, with high probability, simultaneously high deniable and highly reliable). This follows from somewhat standard random coding arguments, if Bob decodes to the nearest codeword.

We generate the ensemble of all codebooks $\mathcal{C}$ containing $2^{r\sqrt{n}}$ codewords of block-length n, with each codeword generated by choosing each bit to be 1 with probability Bernoulli(ρ), and sample codebooks from this distribution (we call such codebooks *random public codebooks*). In what follows, we set ρ to equal $c_2/\sqrt{n}$, where c_2 is a code design parameter.

[4]Or, as George Walker Bush put it, "the soft bigotry of low expectations".
[5]Or, as Philip Pirrip might put it, we have "Great Expectations".

Let

$$c_8 = \sqrt{\frac{\left(\frac{33}{2\epsilon}\sqrt{2\ln\frac{11}{\epsilon}}+1\right)}{1-\delta}},$$

$$c_9 = \log\left(\frac{1-p_w}{p_w}\right)\left(2\ln\frac{11}{\epsilon}+\frac{2\sqrt{2}\epsilon}{11\sqrt{\frac{1}{p_w}+\frac{1}{1-p_w}}}\right),$$

and

$$c_{10} = \frac{2\sqrt{2}\epsilon}{11(1-2p_w)\sqrt{\frac{1}{p_w}+\frac{1}{1-p_w}}}.$$

Theorem 3. . *For any $p_w > \max\{\frac{1}{3}, \frac{1}{2}-\frac{1-2p_b}{2c_8}\}$, and any relative throughput $r \in (c_9, c_{10})$ with probability greater than $1-2^{-\theta(\sqrt{n})}$, a random public codebook $\mathcal{C}$ is simultaneously $(1-\epsilon)$-reliable and $(1-\epsilon)$-deniable.*

Note 1: One may notice that since we have not optimized all constants in our proof, we require not only that $p_w > p_b$, but in fact that p_w is rather close to $1/2$ (though it still works for values bounded away from $1/2$). It would be interesting to examine whether in fact any $p_w > p_w$ would also work.
Note 2: It can be verified that if p_b is zero while p_b is not, then in fact a throughput that scales as $\mathcal{O}(\sqrt{n}\log(n))$ is both reliably and deniably possible. (This does not contradict the outer bound in Theorem 2, since the higher order terms in the bound vanish, and the bound then becomes $\mathcal{O}(\sqrt{n}\log(n))$.) In that case, code construction and proof of it's properties proceeds as follows. Alice's codebook comprises of *all* codewords of weight at most $c_2\sqrt{n}$ (it can be verified that there are $2^{\sqrt{n}\log(n)}$) such codewords). Since there is no noise on Bob's channel, reliability is automatically guaranteed. As to deniability, it can be directly verified that the probability distribution for the specific code in this case matches the ensemble average, implying deniability.

As mentioned above, proofs of the above assertions may be found in the technical report at [11].

References

[1] B. Bash, D. Goeckel, and D. Towsley, "Square root law for communication with low probability of detection on awgn channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2012, pp. 448–452.
[2] ——, "Square root law for communication with low probability of detection on awgn channels," *arXiv preprint arXiv:1202.6423*, 2012.
[3] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2706–2722, 2008.
[4] B. Ryabko and D. Ryabko, "Asymptotically optimal perfect steganographic systems," *Problems of Information Transmission*, vol. 45, no. 2, pp. 184–190, 2009.
[5] A. D. Ker, "The square root law in stegosystems with imperfect information," in *Proceedings of the 12th Information Hiding Workshop*, 2010.
[6] ——, "The square root law requires a linear key," in *in Proc. 11th ACM Workshop on Multimedia and Security*, 2009, pp. 85–92.
[7] A. Ker, "The square root law does not require a linear key," in *Proceedings of the 12th ACM workshop on Multimedia and security*, 2010, pp. 213–224.
[8] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009.
[9] A. Ker, T. Pevnỳ, J. Kodovskỳ, and J. Fridrich, "The square root law of steganographic capacity," in *Proceedings of the 10th ACM workshop on Multimedia and security*. ACM, 2008, pp. 107–116.
[10] A. Ker, "A capacity result for batch steganography," *Signal Processing Letters, IEEE*, vol. 14, no. 8, pp. 525–528, 2007.
[11] P. H. Che, M. Bakshi, S. Jaggi, and M. J. Siavoshani, "Reliable deniable communication: Hiding messages in noise," *http://personal.ie.cuhk.edu.hk/∼sjaggi/arxiv_01.pdf*, 2013.