Layered Schemes for Large-Alphabet Secret Key Distribution

Hongchao Zhou
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: hongchao@mit.edu

Ligong Wang
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: wlg@mit.edu

Gregory Wornell
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: gww@mit.edu

***Abstract*—We discuss the design of practical codes for large-alphabet secret key distribution, motivated by the application of high-dimensional quantum key distribution. We introduce and study a simple scheme called layered scheme, which can be treated as a variant of coded modulation.The idea of layered schemes is to first split the observed large-alphabet symbols into bit layers, and to then encode all the bit layers either independently or jointly using binary codes. The channels that we are interested in are more general than the AWGN channels or Rayleigh fading channels studied in coded modulation. We present and compare different implementations of layered schemes, i.e., based on independent parallel encoding or joint encoding, and we investigate different approaches in how to map large-alphabet symbols into bit layers. Both theoretical analyses and simulation results show that layered schemes have good performances on q-ary channels such as uniform-error channels and limited-magnitude-error channels.**

I. Introduction

We consider the problem of secret key distribution (SKD), which aims at establishing a secret key between two terminals based on their correlated observations [1], [16]. Specifically, we try to design efficient codes for SKD when the initial observations of the two parties are memoryless sequences, with each symbol in the sequences being drawn from a relatively large alphabet. Demands for such codes come from, e.g., practical high-dimensional quantum key distribution (QKD) [2]–[4], [7], [10]. To meet these demands, we introduce a type of efficient coding schemes which we call *layered schemes*.

A. The Problem

Consider a scenario where two terminals, Alice and Bob, initially observe sequences $\mathbf{X} \in \mathcal{X}^n$ and $\mathbf{Y} \in \mathcal{Y}^n$, respectively. We assume that the eavesdropper, Eve, has no initial knowledge about $\mathbf{X}$ or $\mathbf{Y}$. The goal of SKD is to extract a secret key $\mathbf{S} \in \{0,1\}^s$ between Alice and Bob based on these observations. The secret key $\mathbf{S}$ should be almost uniformly distributed on $\{0,1\}^s$, and should be almost completely unknown to Eve, in the sense of [1], [16].

We assume that the sequences $\mathbf{X}$ and $\mathbf{Y}$ are both memoryless, with each pair (X,Y) drawn from the same joint distribution P_{XY}. Without loss of generality, let $\mathcal{X} = \{0,1,...,q-1\}$. Fixing P_X, We can treat $Y \in \mathcal{Y}$ as the output after transmitting X over a channel characterized by the transition law $P_{\mathcal{C}}(Y|X)$.

As an example of the above setting, think of the optical SKD problem discussed in [2], [14]. A source generates random entangled-photon pairs which travel to Alice and Bob separately, where some photons may be lost in transmission. Alice and Bob record the detection times of the photons, with precision up to time-slots of a certain length. They divide all the time-slots into frames where each frame contains q slots. Due to technical constraints, q is typically on the order of 10 to 1000. Through public discussion Alice and Bob can locate all the frames in which they each observed exactly one photon. They can use their relative detection positions in these frames (i.e., X and Y) to distill the secret key. There can be two types of errors in this example. First, due to transmission loss, Alice's and Bob's detections may come from different photon pairs, which result in their detection positions being independent. Second, due to detection jitters, when they detect photons from the same source pair, their detection positions can differ by one or two slots. The channel from X to Y is then the result of combining these two types of effects. Because of the high loss-rate and low detection-efficiency in today's optical systems, the error probability of $X \neq Y$ can be rather high, e.g, 50%, raising challenges in code design.

A typical SKD protocol consists of two steps. In the first step, often called *information reconciliation* in cryptography, Alice and Bob communicate over a public channel (which is authentic but public to Eve). Based on the messages transmitted and on $\mathbf{X}$ and $\mathbf{Y}$, respectively, Alice generates a sequence $\mathbf{W} \in \{0,1\}^w$ and Bob generates a sequence $\mathbf{W}' \in \{0,1\}^w$. In this step they try to make $\mathbf{W} = \mathbf{W}'$ with high probability, but Eve can have some information about $\mathbf{W}$. In the second step, *privacy amplification* [5] is applied to the sequence $\mathbf{W}$ and $\mathbf{W}'$ to extract the secret key $\mathbf{S}$ and $\mathbf{S}'$. If $\mathbf{W} = \mathbf{W}'$, then $\mathbf{S} = \mathbf{S}'$. Furthermore, privacy amplification can ensure that Eve has virtually no information about $\mathbf{S}$. Since there exist standard techniques for privacy amplification, in this paper we focus on the first, information-reconciliation step. We try to minimize the amount of information that is leaked to Eve while ensuring $\mathbf{W} = \mathbf{W}'$ with high probability.

We define the key rate of this step as

$$r = P(\mathbf{W} = \mathbf{W}')\frac{H(\mathbf{W}|\mathbf{W} = \mathbf{W}') - t}{n}, \qquad (1)$$

where t is the number of bits (correlated with $\mathbf{W}$) communicated between Alice and Bob. This definition of the key rate is in the nonasymptotic regime, and is hence slightly different from the existing definition in, e.g., [16], which focuses on the limit where n tends to infinity and where the probability $\mathbf{W} = \mathbf{W}'$ tends to one. In this limit, it can be shown using results of [1] that r tends to $I(X;Y)$. For finite n, we show in Appendix A that it always holds that $r \leq I(X;Y) + \frac{1}{n}$. We henceforth call $I(X;Y)$ the *maximal key rate* between $\mathbf{X}$ and $\mathbf{Y}$.

B. Slepian-Wolf Coding

A simple one-way information-reconciliation scheme directly applies a Slepian-Wolf [18] code. In this scheme, Alice sends a message $\mathbf{R} \in \{0,1\}^t$ that is a deterministic function of $\mathbf{X}$ to Bob; Bob then tries to recover the sequence $\mathbf{X}$ based on $\mathbf{R}$ and $\mathbf{Y}$. They use $\mathbf{X}$ as the common sequence $\mathbf{W}$. In [18], Slepian and Wolf showed that the shortest length of the binary message $\mathbf{R}$ that can guarantee Bob's successful decoding of $\mathbf{X}$ is asymptotically equal to $H(\mathbf{X}|\mathbf{Y})$.

In practice, Slepian-Wolf codes are often constructed from linear channel codes. Specifically, let $\mathcal{C}$ be a linear code with a parity-check matrix $\mathbf{H}$. In a corresponding Slepian-Wolf code, the message $\mathbf{R}$ sent from Alice to Bob is the fully compressed version of the syndrome of $\mathbf{X}$, namely, of $\mathbf{HX}$. It is easy to show that, if the code $\mathcal{C}$ can correct the error $\mathbf{Y} - \mathbf{X}$, then Bob can retrieve $\mathbf{X}$ from $\mathbf{Y}$ and $\mathbf{R}$ successfully.

To implement the above scheme requires efficient linear q-ary codes, which are hard to find or construct for moderately large q. One candidate is a Reed-Solomon code, but its blocklength is limited by q and it is inefficient when error probability is high. Another candidate is a large-alphabet LDPC code, which is more efficient than Reed-Solomon codes. However, traditional belief-propagation decoding is not scalable to large alphabet size q for practical use [8], while verification-based decoding for LDPC codes, as of today, only works for extremely large q [15].

C. Layered Schemes

The main idea of layered schemes for large-alphabet SKD is to convert the problem into binary problems by mapping each symbol $X \in \mathcal{X} = \{0,1,...,q-1\}$ into $k = \lceil \log_2 q \rceil$ bits. Doing this, Alice splits the sequence $\mathbf{X}$ into k bit layers. She then applies Slepian-Wolf coding to the k bit layers, either independently or jointly, and sends the encoded bits to Bob.

Layered schemes can be seen as the reverse of *coded modulation* (see Fig. 1), which was proposed to achieve both power and bandwidth efficiencies in communication. In coded modulation, the transmitter first encodes the message into a binary codeword using an error-correction code, and then maps the codeword to a sequence $\mathbf{X}$ with alphabet size $|\mathcal{X}| = 2^k$.

Well-known coded modulation schemes include multilevel coding (MLC) [12], [19] and bit-interleaved coded modulation (BICM) [9]. These schemes have been extensively studied for Gaussian channels or Rayleigh fading channels, and they can be converted to Slepian-Wolf codes for sources with corresponding joint distributions. However, existing work on coded modulation has not considered general q-ary channels or the specific optical setting discussed in Section I-A, and the alphabet size q was constrained to be a power of 2 (not necessary in layered schemes).

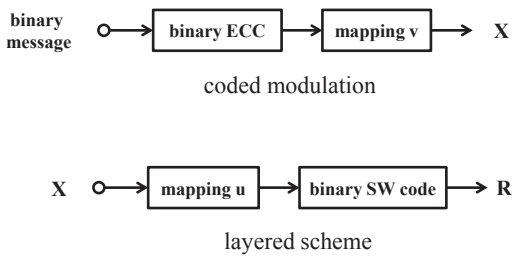


Fig. 1. Layered scheme and coded modulation.

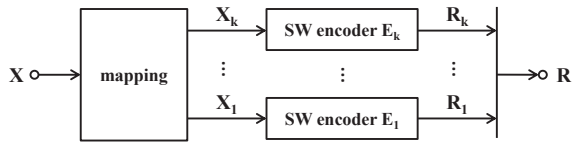
One drawback of layered schemes (or of coded modulation) is the high latency. However, in contrast to communication, latency is less important in SKD. In SKD, Alice and Bob usually generate a secret-bit stream in a block-by-block way. These secret bits are stored for further use following the well-known one-time-pad scheme [17]. Thanks to this cashing mechanism, even if there is a loss or delay in some block in the key-distribution process, it will not introduce any delay in the real-time data communication unless the cashe is empty. Hence, while in communication the decoding error rate of each block should be made extremely small to avoid retransmission (which will introduce delay), in SKD we are more interested in the statistical performance of different blocks, i.e., the key rate that we defined in (1).

In some cases (Sections III-C and IV-C) our layered schemes involve interactive communication, i.e., they involve communication from Bob to Alice. This can simplify and improve the performances of the schemes. Interactive communication is often allowed and widely used in SKD problems. But the kind of interactive communication we propose cannot be used for channel codes, even in the presence of feedback.

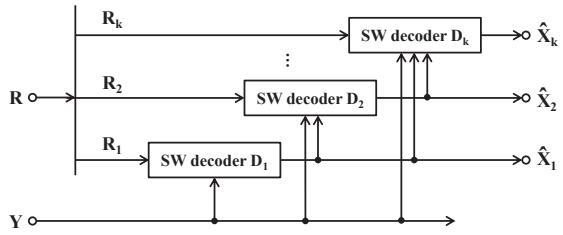
The rest of this paper is organized as follows. Section II presents layered schemes and their implementations based on independent encoding and joint encoding. Section III compares two different encoding approaches of layered schemes. Section IV studies properties of layered schemes for certain classes of $P_{\mathcal{C}}(Y|X)$ and investigates the role of interactive communication. Section V shows simulation results and demonstrates that performances of layered schemes can be near optimal.

II. Layered Schemes

As depicted in Fig. 1, a layered scheme has two steps. In the first step, an injective mapping $u\colon \mathcal{X} \to \{0,1\}^k$ with $k = \lceil \log_2 |\mathcal{X}| \rceil$ is applied to map each symbol X in $\mathbf{X}$ to k bits $(X_1, X_2, ..., X_k)$. The sequence $\mathbf{X}$ is hence split into k bit layers, henceforth denoted by $\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}$, where $\mathbf{X_i}$ with $1 \leq i \leq k$ contains the ith bit of $u(X)$ for every X in $\mathbf{X}$. In the second step, Alice generates a message $\mathbf{R}$ by applying a binary Slepian-Wolf code to $[\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}]$ and sends $\mathbf{R}$



(a) Encoder



(b) Decoder

Fig. 2. Independently-Encoded Layered Scheme.

to Bob. After receiving $\mathbf{R}$, Bob tries to recover every layer $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X_k}$, and hence also the original sequence $\mathbf{X}$, based on $\mathbf{Y}$ and $\mathbf{R}$. The sequence $\mathbf{X}$ is used as the common sequence shared between Alice and Bob (i.e., the sequence denoted $\mathbf{W}$ in the Introduction), upon which privacy-amplification will be applied.

In the second step above, we consider two different approaches for Alice to apply a Slepian-Wolf code to $[\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}]$. The first approach is to apply a (possibly different) Slepain-Wolf code to each bit layer independently. Doing this will produce k output bit-strings, the concatenation of which is the message $\mathbf{R}$. This approach is similar in spirit to MLC in channel coding. The second approach is to apply a single binary Slepian-Wolf code to the whole vector $[\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}]$, generating the message $\mathbf{R}$ directly. This approach is similar to BCIM in channel coding. We call the first approach an *independently-encoded* scheme, and the second approach a *jointly-encoded* scheme.

A. Independently-Encoded Scheme

The diagram of the independently-encoded layered scheme is sketched in Fig. 2. In this scheme, Alice encodes each bit layer $\mathbf{X_i}$ with $1 \leq i \leq k$ independently based on binary Slepian-Wolf coding. As a result, she gets k messages denoted by $\mathbf{R_1}, \mathbf{R_2}, ..., \mathbf{R_k}$, whose concatenation $\mathbf{R}$ is sent to Bob.

Bob decodes each bit layer $\mathbf{X_i}$ with $1 \leq i \leq k$ based on the received message $\mathbf{R_i}$, his observation $\mathbf{Y}$, and the decoding results of previous layers, i.e., $\widehat{\mathbf{X}}_\mathbf{1}^{\mathbf{i-1}} = [\widehat{\mathbf{X}}_\mathbf{1}, \widehat{\mathbf{X}}_\mathbf{2}, ..., \widehat{\mathbf{X}}_\mathbf{i-1}]$. This is the multistage decoding for Slepian-Wolf codes.

The idea behind this scheme is that the errors in different bit layers (for the same symbol) are correlated. For instance, consider the model where X is uniformly distributed on $\{0, 1, ..., q-1\}$ with $q = 2^k$ for some integer k, and where $P_\mathcal{C}(Y|X)$ is a uniform-error channel (i.e., given $Y \neq X$, Y takes value in the remaining $2^k - 1$ symbols with equal probabilities). If a certain symbol erred in one bit layer, then (irrespective of the mapping u) the probability that it errs in the next bit layer is 0.5. Hence the next bit can be treated as erased [20].

For the ith layer, it is convenient to think about the equivalent channel which takes input $\mathbf{X_i}$ and outputs $\mathbf{Y}, \mathbf{X_1^{i-1}}$. The transition law of this channel can be easily obtained as

$$P(Y = y, X_1^{i-1} = x_1^{i-1} | X_i = x_i) = \frac{\sum_{x \in A(x_1, x_2, ..., x_i)} P_\mathcal{C}(y|x) P(X = x)}{\sum_{x \in A(x_i)} P(X = x)},$$

where $A(x_1, x_2, ..., x_i)$ denotes the set of all $x \in \mathcal{X}$ such that the jth bit of $u(x)$ is x_j with $1 \leq j \leq i$; similarly, $A(x_i)$ denotes the set of all $x \in \mathcal{X}$ such that the ith bit of $u(x)$ is x_i.

Theorem 1. *Provided that an asymptotically optimal Slepian-Wolf code for each bit layer can be found, the maximum achievable asymptotic rate (i.e., the maximum rate in the limit where blocklength n tends to infinity and where the probability that $\mathbf{W} = \mathbf{W}'$ is required to tend to one) of independently-encoded layered schemes is*

$$r^*_{independent} = I(X; Y).$$

Proof: Let t_i be the length of the message $\mathbf{R_i}$. From our definition (1) we have

$$r = P(\mathbf{X} = \hat{\mathbf{X}}) \frac{H(\mathbf{X}|\mathbf{X} = \hat{\mathbf{X}}) - \sum_{i=1}^k t_i}{n}.$$

Assume that the first $i-1$ bit layers are successfully decoded. If the Slepian-Wolf code is asymptotically optimal for the equivalent channel i, then [18]

$$\lim_{n \to \infty} \frac{t_i}{H(\mathbf{X_i}|\mathbf{Y}, \mathbf{X_1^{i-1}})} = 1$$

and, for any positive ϵ, for large enough n, $P(\mathbf{X_i} \neq \widehat{\mathbf{X}}_\mathbf{i}) < \epsilon$. Using the chain rule for conditional entropies we obtain

$$\begin{aligned} \lim_{n \to \infty} \sum_{i=1}^k \frac{t_i}{n} &= \lim_{n \to \infty} \sum_{i=1}^k \frac{H(\mathbf{X_i}|\mathbf{Y}, \mathbf{X_1^{i-1}})}{n} \\ &= \lim_{n \to \infty} \frac{H(\mathbf{X}|\mathbf{Y})}{n} \\ &= H(X|Y), \end{aligned}$$

where the last step follows from our assumption that the sequences $\mathbf{X}$ and $\mathbf{Y}$ are memoryless.

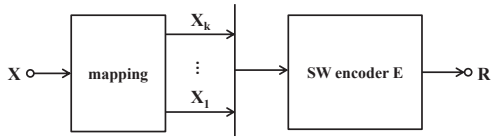
We also have

$$\begin{aligned} H(\mathbf{X}) &\leq H(\mathbf{X}|\mathbf{1}_{(\mathbf{X_1^k} = \hat{\mathbf{X}}_\mathbf{1}^\mathbf{k})}) + 1 \\ &\leq P(\mathbf{X_1^k} = \hat{\mathbf{X}}_\mathbf{1}^\mathbf{k}) H(\mathbf{X}|\mathbf{X_1^k} = \hat{\mathbf{X}}_\mathbf{1}^\mathbf{k}) + \epsilon k n + 1 \\ &\leq H(\mathbf{X}|\mathbf{X_1^k} = \hat{\mathbf{X}}_\mathbf{1}^\mathbf{k}) + \epsilon k n + 1, \end{aligned}$$

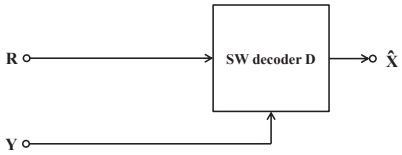
so $H(\mathbf{X}|\mathbf{X_1^k} = \hat{\mathbf{X}}_\mathbf{1}^\mathbf{k}) \geq H(\mathbf{X}) - \epsilon k n - 1$.

Hence, as $n \to \infty$, we have

$$\begin{aligned} r &\geq (1 - \epsilon k) \frac{H(\mathbf{X}) - \epsilon k n - 1 - \sum_{i=1}^k t_i}{n} \\ &= (1 - \epsilon k) H(X) - \epsilon k - H(X|Y) \\ &= I(X; Y) - \epsilon k (H(X) + 1). \end{aligned}$$



(a) Encoder



(b) Decoder

Fig. 3. Jointly-Encoded Layered Scheme.

Setting ϵ arbitrarily small proves the theorem. ∎

Theorem 1 shows that the independently-encoded scheme is theoretically optimal, irrespective of the chosen mapping u.

B. Jointly-Encoded Scheme

In the jointly-encoded layered scheme, as sketched in Fig. 3, Alice treats $[\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}]$ as a whole binary sequence and applies a single Slepian-Wolf code to this sequence. As a result, she gets the message $\mathbf{R}$, which she sends to Bob.

To decode $[\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}]$ based on the received message $\mathbf{R}$ and the observation $\mathbf{Y}$, Bob ignores the dependence between the bits in $[\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_k}]$. For instance, for the Maximum Likelihood Decoder, Bob computes the likelihood of X_i only based on Y and the index i. For the ith layer, we can think of the equivalent channel which takes input $\mathbf{X_i}$ and outputs $\mathbf{Y}$. The transition law of this equivalent channel is

$$P(Y = y|X_i = x_i) = \frac{\sum_{x \in A(x_i)} P_{\mathcal{C}}(y|x) P(X = x)}{\sum_{x \in A(x_i)} P(X = x)}.$$

Assume that the binary Slepian-Wolf code (as well as the decoder for each layer) is optimal, then the message length t satisfies

$$\lim_{n \to \infty} \frac{t}{\sum_{i=1}^{k} H(\mathbf{X_i}|\mathbf{Y})} = 1.$$

This implies the following:

Theorem 2. *Provided that an asymptotically optimal binary Slepian-Wolf code can be found, the maximum asymptotic rate of the jointly-encoded layered scheme is*

$$r^*_{joint} = H(X) - \sum_{i=1}^{k} H(X_i|Y).$$

Note that there is generally a gap between r^*_{joint} and $r^*_{\text{independent}} = I(X;Y)$. This gap depends on the distribution P_{XY} and on the mapping $u \colon \mathcal{X} \to \{0,1\}^k$.

C. Slepian-Wolf Codes based on LDPC codes

We next demonstrate how to construct a Slepian-Wolf code from a binary LDPC code. For a binary-input channel with input $\mathbf{X}$ and output $\mathbf{Y}$, we assume that C is an LDPC code with parity-check matrix $\mathbf{H}$ such that for all $\mathbf{X} \in C$, $\mathbf{X}$ can be successfully decoded from $\mathbf{Y}$ with probability close to one.

The encoding of the Slepian-Wolf code based on C is very simple: the message $\mathbf{R}$ sent from $\mathbf{X}$ to $\mathbf{Y}$ is the compressed version of $\mathbf{HX}$. Note that if we define a coset code $C_{\mathbf{R}}$ as

$$C_{\mathbf{R}} = \{\mathbf{X}' \in \{0,1\}^n \colon \mathbf{HX}' = \mathbf{R}\},$$

then $\mathbf{X}$ is a codeword in $C_{\mathbf{R}}$. Decoding $\mathbf{Y}$ and $\mathbf{R}$ jointly to recover $\mathbf{X}$ is now equivalent to decoding the coset code $C_{\mathbf{R}}$. One such decoder based on belief-propagation is described in [11], [13]. Specifically, we label the bits in $\mathbf{R}$ to the check nodes, then the belief passed from a check node c to a variable node v is

$$\mathbf{m}_{cv} = (-1)^{R_c} 2 \tanh^{-1} \left(\prod_{v' \in N(c)/v} \tanh\left(\frac{\mathbf{m}_{v'c}}{2}\right) \right), \quad (2)$$

where $\mathbf{m}_{vc}$ is the message passed from a variable node $\mathbf{v}$ to a check node $\mathbf{c}$, $N(c)$ is the set of variable nodes that connect to check node $\mathbf{c}$, and R_c is the bit labeled on the check node c. Compared to the belief-propagation decoder for the original LDPC code, it only changes the signs of the beliefs from the check nodes with $R_c = 1$.

D. Channel Adapters

Slepian-Wolf codes based on LDPC codes have near-optimal performances for binary-input symmetric channels with equiprobable input distribution. Here, a binary-input channel $\mathcal{C} \colon \{0,1\} \to \mathcal{Y}$ is said to be symmetric if and only if there exists a bijective function $\sigma \colon \mathcal{Y} \to \mathcal{Y}$ such that $\sigma^2(y) = y$ for all $y \in \mathcal{Y}$ and

$$P_{\mathcal{C}}(y|x = 0) = P_{\mathcal{C}}(\sigma(y)|x = 1)$$

for all $y \in \mathcal{Y}$.

Unfortunately, the equivalent channels yielded by layered schemes are not always symmetric. In [11], a tool called channel adapter was introduced to force the symmetry of the binary-input channels for communication. The same idea can be used for Slepian-Wolf coding. Let $\mathbf{X} \in \{0,1\}^n$ be the binary sequence observed by Alice and $\mathbf{Y} \in \mathcal{Y}^n$ be the sequence observed by Bob. Alice draws a random sequence $\mathbf{Z}$ uniformly from $\{0,1\}^n$, computes $\mathbf{X}' = \mathbf{X} \oplus \mathbf{Z}$, and sends $\mathbf{Z}$ to Bob. We thus obtain a new channel whose input is $X' \in \{0,1\}$ and whose output is $(Y, Z) \in \mathcal{Y} \times \{0,1\}$. It is easy to see that $P(X' = 1) = P(X' = 0)$. This new channel is symmetric, because

$$P_{\mathcal{C}}((Y,Z)|X = 0) = P_{\mathcal{C}}(\sigma(Y,Z)|X = 1)$$

with $\sigma(Y,Z) = (Y, Z \oplus 1)$ and $\sigma^2(Y,Z) = (Y,Z)$. If we construct a Slepian-Wolf code based on LDPC codes for the sequences $\mathbf{X}'$ and $(\mathbf{Y}, \mathbf{Z})$, then it has near-optimal performance due to the symmetry of the new channel. Hence

the length of the message $\mathbf{R}$ sent by Alice is close to $nH(X'|Y,Z)$. After successfully decoding $\mathbf{X}'$, Bob can further retrieve $\mathbf{X} = \mathbf{X}' \oplus \mathbf{Z}$.

Observing that $H(X'|Y,Z) = H(X \oplus Z|Y,Z) = H(X|Y,Z) = H(X|Y)$, we conclude that: A Slepian-Wolf code for an arbitrary binary-input channel that is based on channel adapters is asymptotically optimal if the underlying Slepian-Wolf code is asymptotically optimal for binary-input symmetric channels with equiprobable input distribution.

III. Comparison of Layered Schemes

In this section we introduce different mappings u and compare the two classes of layered schemes.

A. Mappings

Let $w\colon \{0,1,...,2^k-1\} \to \{0,1\}^k$ be a bijective mapping. We consider u which is w with the input constrained on the first $|\mathcal{X}|$ values.

The simplest mapping is the binary representation, i.e., the unique function $w_{\text{bin}}\colon \{0,1,...,2^k-1\} \to \{0,1\}^k$ such that $w_{\text{bin}}(x) = [x_1, x_2, ..., x_k]$ with $x = \sum_{i=1}^{k} x_i 2^{i-1}$.

Gray mapping, where two successive values in $\{0,1,...,2^k-1\}$ differ in only one bit [6], is widely used in BICM [9]. We denote it as w_{Gray}.

Both the binary representation and the Gray mapping are easy to construct but generally suboptimal in terms of r^*_{joint}. It is often computationally difficult to find the best mapping w by brute-force searching. One idea is to first randomly generate a bijective function w, and then to optimize w based on a heuristic approach. Specifically, we switch the outputs of w for two distinct input values $a, b \in \{0,1,...,2^k-1\}$ if this operation leads to a better mapping. We do this until no such two distinct input values can be found. We use w_{search} to denote the mapping constructed in this way. As we shall see, this heuristic approach can often lead to reasonably good mappings.

B. Maximal Key Rates

We next compare $r^*_{\text{independent}}$ and r^*_{joint}, which are given in Theorems 1 and 2, respectively, where the latter depends on the chosen mapping u.

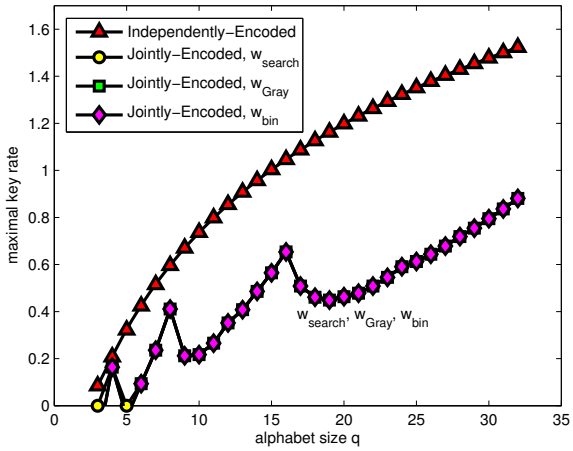
An important class of q-ary channel is uniform-error channels. Such a channel's transition law $P_{\mathcal{C}}(y|x)$ with $x, y \in \{0,1,...,q-1\}$ is given by

$$P_{\mathcal{C}}(y|x) = \begin{cases} 1-\delta & \text{if } y = x \\ \frac{\delta}{q-1} & \text{otherwise,} \end{cases} \tag{3}$$

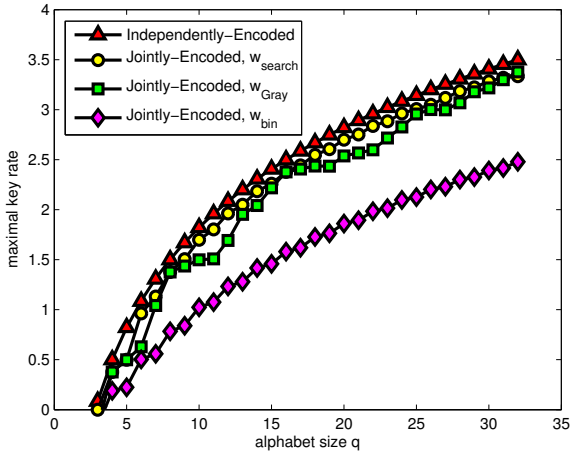
where δ is the symbol error rate. In high-dimensional QKD, this type of errors is usually caused by the dark current, photon transmission and detection losses [3].

Another interesting class of channels is local-error channels. Local errors have also been observed in QKD systems, and are often caused by jitters of electronics [3]. A simple local-error channel is one with transition law

$$P_{\mathcal{C}}(y|x) = \begin{cases} \frac{\delta}{2} & \text{if } y - x = \pm 1, \\ 1-\delta & \text{if } x = y, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$



(a) Uniform-Error Channel



(b) Local-Error Channel

Fig. 4. Maximal key rates of the layered schemes for different channels.

for all $x, y \in \{0,1,...,q-1\}$. Here, $-1 = q-1$ in the field of $\{0,1,2,...,q-1\}$.

Fig. 4 compares the maximal rates of the two classes of (i.e., independently-encoded and jointly-encoded) layered schemes, with different us, for the two channel models specified by (3) and (4), where we set the symbol error rate $\delta = 0.5$ and let X be uniformly distributed. It shows that the independently-encoded scheme is much more efficient than the jointly-encoded scheme for channels with a big fraction of uniform errors. We also observe that the Gray mapping w_{Gray} is much better than the binary representation w_{bin} for the jointly-encoded scheme for local-error channels.

C. Error Propagation

In practice, our independently-encoded schemes can suffer from error propagation: a decoding error in a bit layer will result in decoding errors in the following bit layers. To see this effect, let t_i denote the length of the message sent from Alice to Bob for the ith bit layer, and let e_i denote the decoding error

probability of the ith bit layer when the first $i-1$ bit layers are successfully decoded. The key rate of the independently-encoded scheme can then be written as

$$r_{\text{independent}} = \left(\prod_{i=1}^{k}(1-e_i)\right)\frac{nH(X)-\sum_{i=1}^{k}t_i}{n}.$$

Using interactive communication between Alice and Bob, which is often allowed in information reconciliation, can help to largely eliminate error-propagation effects. To this end, after decoding the ith layer, Bob checks whether decoding was successful or not. He can do this either by observing the decoding process, i.e., whether the beliefs of all the variables converge to certainty, or by Alice's adding extra redundant bits for error detection. If there is a decoding error, then Bob asks Alice to transmit the whole sequence $\mathbf{X_i}$ directly. As a result, the key rate of the independently-encoded scheme is improved to

$$r_{\text{independent}} = \sum_{i=1}^{k}\left[(1-e_i)\frac{nH(X_i|X_1^{i-1})-t_i}{n} + e_i\frac{H(X_i|X_1^{i-1})-H(X_i)}{n}\right].$$

D. Using Suboptimal Binary Slepian-Wolf Codes

Realistic Slepian-Wolf codes cannot achieve the theoretical limit. We next discuss how suboptimality of binary Slepian-Wolf codes affects the performance of the layered schemes. To simplify the discussion, we make a simple assumption (which may not be realistic): given an arbitrary binary-input channel, the binary Slepian-Wolf code for $\mathbf{X}\in\{0,1\}^n$ and $\mathbf{Y}\in\mathcal{Y}^n$ requires $\alpha H(\mathbf{X}|\mathbf{Y})$ bits where $\alpha>1$ for successful decoding. If the message length is shorter than $\alpha H(\mathbf{X}|\mathbf{Y})$, then $\mathbf{X}$ cannot be recovered; otherwise, $\mathbf{X}$ can be recovered surely.

Based on this assumption, we get the key rate of the jointly-encoded scheme:

$$r_{\text{joint}} = H(X)-\alpha\sum_{i=1}^{k}H(X_i|Y) \le H(X)-\alpha H(X|Y).$$

We also get the key rate of the independently-decoded scheme:

$$\begin{aligned} r_{\text{independent}} &= H(X)-\sum_{i=1}^{k}\min(H(X_i),\alpha H(X_i|X_1^{i-1},Y)) \\ &= H(X)-\alpha H(X|Y) \\ &\quad +\sum_{i=1}^{k}(\alpha H(X_i|X_1^{i-1},Y)-H(X_i))^+, \end{aligned}$$

where $x^+\triangleq\max\{x,0\}$.

We see that $r_{\text{joint}}\le r_{\text{independent}}$, i.e., even when we use binary Slepian-Wolf codes that are suboptimal (but with the same performance for the two schemes), the independently-encoded scheme is always better than the jointly-encoded scheme, irrespective of the mapping u.

E. Discussions: Independent Encoding vs. Joint Encoding

From the above analyses, we have the following simple observations, which we shall further verify by simulation in Section V.

(1) If the magnitude of errors is large, it is prone to apply the independently-encoded layered scheme rather than the jointly-encoded layered scheme, for much higher maximal key rate.

(2) Error propagation among different bit layers is a problem for the independently-encoded scheme. However, this effect can be eliminated by interactive communication between Alice and Bob.

(3) In high-speed SKD applications, hardware implementation of the underlying binary Slepian-Wolf decoders is required. Limited by hardware, the input length of the Slepian-Wolf code cannot be too large. If we assume that the binary Slepian-Wolf codes used for the two layered schemes have the same input length and approximately the same performance, then the overall performance of the jointly-encoded scheme should not be better than the independently-encoded scheme. However, in this case, the block length of the jointly-encoded scheme is actually k times shorter than that of the independently-encoded scheme. In other words, the independently-encoded scheme introduces more latency which, as we argued in the Introduction, is less important in SKD than in communications.

(4) We will demonstrate that in the independently-encoded scheme, interactive communication between Alice and Bob can further improve the practical performance of the scheme. In particular, Bob may send some useful information to Alice after decoding each bit layer, and based on this information Alice can better encode the next layer.

(5) As we have discussed many advantages of the independently-encoded scheme, it has an obvious disadvantage: it requires k binary Slepian-Wolf decoders, while the jointly-encoded scheme requires only one decoder, which is less complex on hardware.

IV. Channel Properties

In this section, we study some properties of certain channels that are useful in layered schemes.

A. Reflection-Symmetric Channels

A channel $\mathcal{C}\colon\mathcal{X}\to\mathcal{Y}$ with $\mathcal{X}=\{0,1,2,..,q-1\}$ for some even q is said to be *reflection-symmetric* if there exists a bijective function $\sigma\colon\mathcal{Y}\to\mathcal{Y}$ with $\sigma^2=I$ such that

$$P_{\mathcal{C}}(y|x)=P_{\mathcal{C}}(\sigma(y)|q-1-x)$$

for all $x\in\mathcal{X}, y\in\mathcal{Y}$. Examples of reflection-symmetric channels include the uniform-error channels (3) and the local-error channels (4).

We show that if a q-ary channel is reflection-symmetric with equiprobable input distribution, then all the binary equivalent channels yielded by the layered schemes are also symmetric (without channel adapters). Here, we consider mappings $u\colon\mathcal{X}\to\{0,1\}^k$ satisfying

$$u(x)=\overline{u(q-1-x)} \tag{5}$$

for all $x \in \mathcal{X}$, where $\overline{a} \triangleq [1, 1, ..., 1] - a$. For instance, when $q = 2^k$, binary representation satisfies (5).

Theorem 3. *For a reflection-symmetric channel with equiprobable input distribution, if the mapping u satisfies* (5)*, then the layered scheme based on u yield binary-input symmetric channels with equiprobable input distribution for all layers.*

Proof: Let us consider the equivalent channel i of the independently-encoded scheme. The channel's input is X_i and its output is (X_1^{i-1}, Y).

It is easy to see that the input has equiprobable distribution, i.e., $P(X_i = 0) = P(X_i = 1)$. We can see that this channel is also symmetric. To this end, observe

$$\begin{aligned} &P((X_1^{i-1}, Y)|X_i = 0) \\ &= 2 \sum_{x \in A(X_1^{i-1}, X_i = 0)} P(Y|x)P(x) \\ &= 2 \sum_{q-1-x \in A(\overline{X_1^{i-1}}, X_i = 1)} P(\sigma(Y)|q-1-x)P(q-1-x) \\ &= P(\overline{X_1^{i-1}}, \sigma(Y)|X_i = 1). \end{aligned}$$

Hence we obtain a bijective function σ' with $\sigma'(X_1^{i-1}, Y) = (\overline{X_1^{i-1}}, \sigma(Y))$ and $\sigma'^2 = I$. So the equivalent channels of the independently-encoded scheme are indeed symmetric.

The proof for jointly-encoded layered schemes are similar and are omitted. ∎

B. Limited-Magnitude-Error Channels

Given a channel $\mathcal{C}\colon \mathcal{X} \to \mathcal{Y}$ with $\mathcal{Y} = \mathcal{X} = \{0, 1, 2, ..., q-1\}$, we define its error magnitude as the minimal integer m such that $m = m_+ + m_-$ with

$$P_{\mathcal{C}}(y|x) = 0, \forall y - x < -m_- \text{ or } y - x > m_+. \tag{6}$$

For some channels with limited-magnitude errors, it is not necessary to split the sequence $\mathbf{X}$ into $k = \lceil \log_2 q \rceil$ bit layers. Instead, we can generate a new sequence $\mathbf{X}'$ such that $\mathbf{X}' = \mathbf{X} \mod \mathbf{m} + \mathbf{1}$. Then we apply a layered scheme to $\mathbf{X}'$ and $\mathbf{Y}$. In this case, the number of bit layers can be reduced to $\lceil \log_2(m+1) \rceil$, and the encoding/decoding is simplified. The following theorem shows that this simplification does not degrade performance.

Theorem 4. *In the above approach, $\mathbf{X}$ can be uniquely determined by $\mathbf{X}'$ and $\mathbf{Y}$, and $H(\mathbf{X}'|\mathbf{Y}) = H(\mathbf{X}|\mathbf{Y})$.*

Proof: Let $Z = Y - X' \mod m + 1$.

If $Z = 0$, then $X = Y$.

If $0 < Z \le m_+$, then $X = Y - Z$.

If $m_+ + 1 \le Z \le m$, then $X = Y + Z - m - 1$.

So $\mathbf{X}$ can be uniquely determined by $\mathbf{X}'$ and $\mathbf{Y}$, and $H(\mathbf{X}'|\mathbf{Y}) = H(\mathbf{X}', \mathbf{Y}|\mathbf{Y}) = H(\mathbf{X}|\mathbf{Y})$. ∎

C. Cyclic-Symmetric Channels

A channel $\mathcal{C}\colon \mathcal{X} \to \mathcal{Y}$ with $\mathcal{Y} = \mathcal{X} = \{0, 1, 2, .., q-1\}$ is said to be *cyclic-symmetric* if for all $x, y \in \{0, 1, ..., q-1\}$, we have $P_{\mathcal{C}}(y|x) = P_{\mathcal{C}}(y - x|0)$. Here, we define the operation '$-$' in the field of $\{0, 1, ..., q-1\}$, e.g., $0 - 1 = q - 1$. Both the uniform-error channels (3) and local-error channels (4) are cyclic-symmetric channels.

Theorem 5. *For a cyclic-symmetric channel with $q = 2^k$ and equiprobable input distribution, let u be the binary representation in the lowest-bit-first order. If $C = \min(\{i\colon X_i \neq Y_i\} \bigcup \{k+1\})$, then*

$$I(X; C) = 0.$$

Proof: We show that $P(X = x, C = i)$ is independent of x as follows:

$$\begin{aligned} P(X = x, C = i) &= \sum_{y\colon (y-x) = 2^{i-1} \bmod 2^i} P_{\mathcal{C}}(y|x)P(x) \\ &= \sum_{y\colon (y-x) = 2^{i-1} \bmod 2^i} P_{\mathcal{C}}(y - x|0)\frac{1}{2^k} \\ &= \sum_{e \in \mathcal{X}\colon e = 2^{i-1} \bmod 2^i} P_{\mathcal{C}}(e|0)\frac{1}{2^k}. \end{aligned}$$

The claim follows. ∎

The above theorem implies that for a cyclic-symmetric channel with $q = 2^k$, if we apply a layered scheme with the binary-representation mapping, then Bob can send C to Alice without disclosing any information about X. We can also write C as $[C_1, C_2, ..., C_k]$ with

$$C_i = \mathbf{1}_{(X_1^i \neq Y_1^i)}.$$

So, instead of transmitting C, Bob can transmit $C_1, C_2, ..., C_k$ to Alice. We can implement this process in the independently-encoded scheme, i.e, after decoding the ith bit layer, Bob sends C_i to Alice. The following example demonstrates how such interactive communication can be used to improve the practical performance of the independently-encoded scheme.

Example 6. *For a uniform-error channel specified by (3) with $q = 2^k$, we have*

$$P(X_i|X_1^{i-1}, Y, C_{i-1} = 1) = \frac{1}{2}.$$

It implies that if $C_{i-1} = 1$, then Bob knows nothing about X_i until receiving the message $\mathbf{R_i}$. In this case, if Alice knows $C_{i-1} = 1$, there is no need for her to encode X_i into the message $\mathbf{R_i}$. A simple approach is that Alice directly sends X_i to Bob without encoding it; she only needs to encode those bits with $C_{i-1} = 0$. In each bit layer, the bits to encode is a random variable upper-bounded by the block length n, which is not convenient for hardware decoding. Our idea of solving this problem is to encode n bits from the same bit layer but different blocks jointly.

In the above example, assume that $q = 32$ and that the binary Slepain-Wolf coding is not perfect: it requires to transmit

$1.4H(\mathbf{X}|\mathbf{Y})$ bits to recover $\mathbf{X}$. Then, without applying the interactive-communication mechanism, the maximal symbol error rate that allows non-zero key rate is $\delta = 0.620$. By applying the interactive-communication mechanism described in the example, the maximal symbol error rate that allows non-zero key rate is 0.883.

V. Simulation

In this section, we evaluate the performance of the layered schemes for large-alphabet secret key distribution by simulation. We first introduce the setup and some practical implementation issues. Then we provide some simulation results for different types of channels.

A. Setup

In the simulation, we set $q = 32$ with $k = 5$ and let X be uniformly distributed on $\{0, 1, ..., 31\}$. The blocklength for the independently-encoded scheme is $n = 4000$ and the the block length for the jointly-encoded scheme is 800 (for the same length of the underlying binary Slepian-Wolf codes). We assume that each binary Slepian-Wolf code associates with extra $d = 20$ parity-check bits so that the decoding error can be detected by Bob with high probability.

We use simple regular LDPC codes for the binary Slepian-Wolf coding. Specifically, given the block length n and message length r, the parity-check matrix $\mathbf{H}$ is randomly constructed such that each column has exactly 3 ones. (Note that by using irregular LDPC codes based on density evolution, we can get better performances of the layered schemes.)

In the independently-encoded scheme, we use the binary-representation with the lowest-bit-first order as the mapping u, and we let t_i be the message length of the binary Slepian-Wolf code for the ith bit layer. Then the rate of the scheme (eliminating the error-propagation effect based on interactive communication) is

$$r_{\text{independent}} = \sum_{i=1}^{5} \frac{(1 - e_i(t_i)) \max(n - t_i - d, 0)}{n}$$

with $n = 4000$, where $e_i(t_i)$ is the decoding error probability for the ith bit layer and it is a fixed function of t_i with given channel and Slepian-wolf code.

In the jointly-encoded scheme, we use the Gray mapping as the mapping u, and we let t be the message length of the binary Slepian-Wolf code. Then the rate of the scheme is

$$r_{\text{joint}} = \frac{(1 - e(t)) \max(n - t - d, 0)}{n/k}$$

with $n = 4000$, where $e(r)$ is the decoding error probability.

B. Message Length

According to the expressions of the key rates, the message length t_i (or t) for each binary Slepian-Wolf code can be optimized independently. For instance, the optimal length t_i^* for the equivalent channel i in the independently-encoded scheme is

$$t_i^* = \arg\max_{t_i} (1 - e_i(t_i)) \max(n - t_i - d, 0).$$

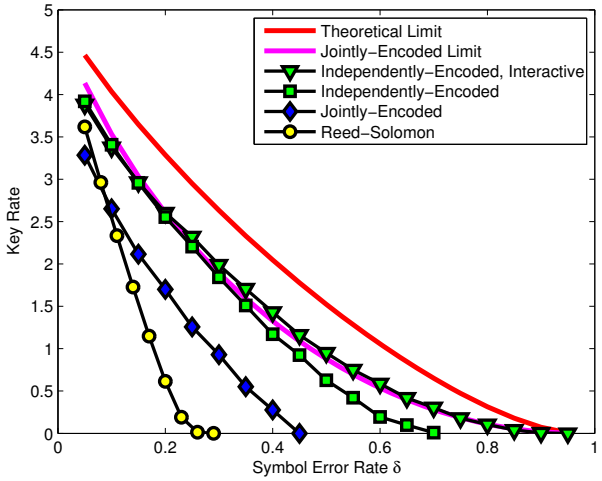


Fig. 5. Key rates of the layered schemes based on regular LDPC codes for uniform-error channels when $q = 32$ and $n = 4000$. The block length of the Reed-Solomon code is 1024.

Such a way of selecting t_i is very useful when symbol error rate is high. For instance, assume that it requires $t_i > n$ to make $e_i(t_i) \leq 0.01$, and it requires $t_i = 0.9n$ to make $e_i(t_i) = 0.5$. Obviously, the second t_i results in a higher key rate.

In the simulation, we ignore this edge effect of t_i (i.e., when symbol error rate is high, the optimal message length should be used). We use a simple empirical approach to determine $t_1, t_2, ..., t_k$ and t. Specifically, we let

$$t_i = \alpha_i H(\mathbf{X}|\mathbf{Y})$$

with $\alpha_i > 1$ when $\mathbf{X}$ and $\mathbf{Y}$ are the input sequence and output sequence of the equivalent channel i, respectively. At the beginning, we let $\alpha_i = 1$. We apply the binary Slepian-Wolf coding to many samples of $\mathbf{X}$ and $\mathbf{Y}$. If $\mathbf{X}$ cannot be successfully decoded, then we update $\alpha_i = \alpha_i + \alpha_c$ with a small constant α_c, e.g. 0.02. By running this procedure for enough samples, e.g., 1000 samples, we obtain a reasonably good message length t_i, based on which a binary Slepian-Wolf code is constructed as the basic component for the layered schemes.

C. Uniform-Error Channels

Fig. 5 shows the performances of the layered schemes for the uniform-error channels defined by (3). Due to the imperfectness of the regular LDPC codes that we used, there is a gap between the actual key rates of the layered schemes and their maximal key rates.

We compare the layered schemes with Slepian-Wolf codes based on Reed-Solomon codes. We consider a Reed-Solomon code over $F(q)$ with block length $n = q^m$. For such a code, in order to correct at most t symbol errors, it requires $r = 2mt$ redundant symbols. Given a symbol error rate δ, we can select the best t to maximize the key rate. In this case, the maximal

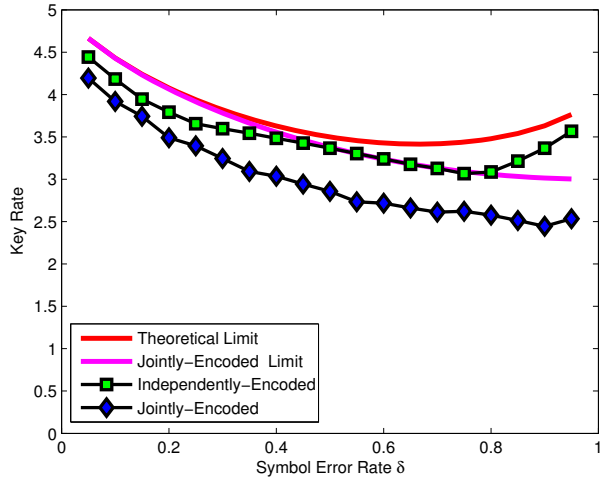


Fig. 6. Key rates of the layered schemes based on regular LDPC codes for local-error channels when $q = 32$ and $n = 4000$.

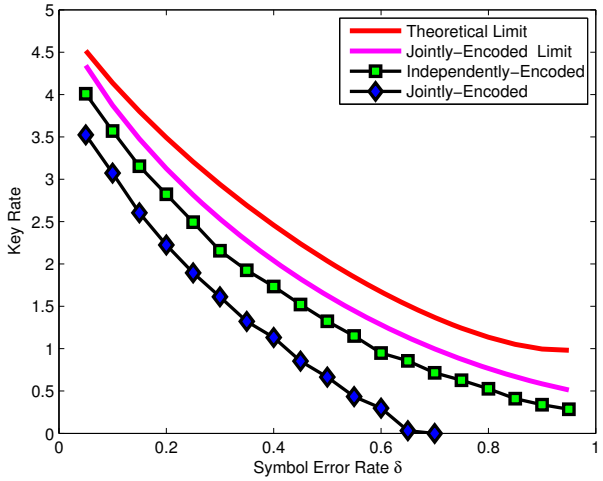


Fig. 7. Key rates of the layered schemes based on regular LDPC codes for hybrid-error channels when $q = 32$ and $n = 4000$.

key rate is

$$r_{\text{RS}} = \max_{t,m} \frac{P(|\mathbf{E}| \le t)[k(q^m - 2tm) - d]}{q^m},$$

where $P(|\mathbf{E}| \le t)$ is the probability that the number of erroneous symbols is at most t. In this simulation, we have $k = 5$ and we set $m = 2$. From Fig. 5, we see that there is a significant performance gain in the layered schemes over Reed-Solomon codes, especially when the symbol error rate is not small.

We also see that the independently-encoded scheme is much more efficient than the jointly-encoded scheme for correcting uniform errors. This possibly comes from the gap between the maximal key rates of the two schemes. Furthermore, interactive communication improves the performance of the independently-encoded scheme. Our intuition is that, as the symbol error rate increases, the improvement becomes stronger because there are more bits that can be treated as erased in the scheme.

D. Local-Error Channels

Fig. 6 shows the performances of the layered schemes for local-error channels described by (4). There is a smaller gap between the performances of the jointly-encoded scheme and of the independently-encoded scheme compared to on uniform-error channels. One observation is that, as the symbol error rate increases, the gap between their maximal key rates increases. Another observation is that the gap of the actual key rate and the maximal key rate of the jointly-encoded scheme is larger than that of the independently schemes.

We can explain the second observation as follows. The key rate of the independently-encoded scheme can be roughly written as

$$k - \sum_{i=1}^{k} \alpha(H(X_i|X_1^{i-1}, Y)) \tag{7}$$

for a concave function $\alpha(\cdot)$. It means that, in practice, if $H(X_i|Y)$ is small, the binary Slepian-Wolf code needs a large overhead. In the ideal case, $\alpha(x) = x$. Similarly, the key rate of the jointly-encoded scheme can be roughly written as

$$k\left(1 - \alpha\left(\frac{\sum_{i=1}^{k} H(X_i|Y)}{k}\right)\right). \tag{8}$$

The difference between (7) and (8) comes from two terms,

$$\sum_{i=1}^{k} \alpha(H(X_i|Y)) - \sum_{i=1}^{k} \alpha(H(X_i|X_1^{i-1}, Y)) \ge 0$$

introduced by the gap of the maximal key rates, and

$$k\alpha\left(\frac{\sum_{i=1}^{k} H(X_i|Y)}{k}\right) - \sum_{i=1}^{k} \alpha(H(X_i|Y)) \ge 0$$

introduced by the concavity of the function $\alpha(\cdot)$. This is an advantage of the independent-encoded scheme over the jointly-encoded scheme in practical use.

E. Hybrid-Error Channels

In the high-dimensional QKD systems described in [3], [14], a large number of uniform errors and local errors due to different mechanisms are observed. We call such a channel a *hybrid-error* channel.

Fig. 7 simulates a hybrid-error channel $\mathcal{C}\colon \{0, 1, ..., q-1\} \to \{0, 1, ..., q-1\}$ described by

$$P_{\mathcal{C}}(y|x) = \begin{cases} \frac{\delta}{4} & \text{if } y - x = \pm 1, \\ 1 - \delta & \text{if } x = y, \\ \frac{\delta}{2(q-3)} & \text{otherwise.} \end{cases}$$

We see that, even if all the symbols have errors, the independently-encoded scheme is still able to generate secret bits.

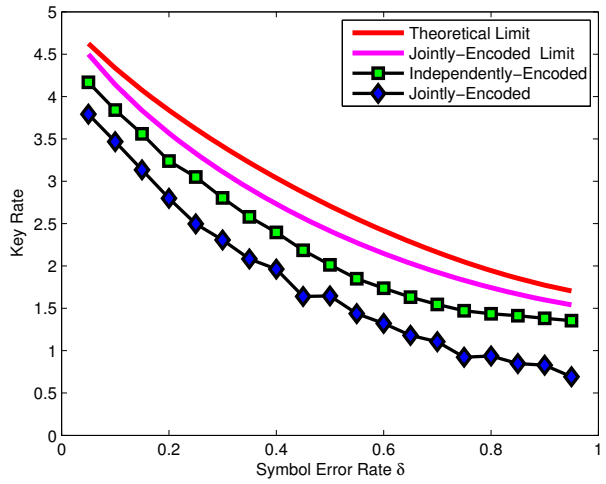


Fig. 8. Key rates of the layered schemes based on regular LDPC codes for discrete-Gaussian-error channels when $q = 32$ and $n = 4000$.

F. Discrete-Gaussian-Error Channels

Fig. 8 shows the performances of the layered schemes for a discrete channel with Gaussian noise, and we call it a discrete-Gaussian-error channel, defined by

$$P_{\mathcal{C}}(y|x) = \delta \frac{e^{-\frac{|y-x|^2}{2\sigma^2}}}{c}$$

for all $x, y \in \{0, 1, ..., q-1\}$, $x \neq y$. Here $c = \sum_{x=-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}$ is a normalization factor and δ is the symbol error rate. In this simulation, we choose $\sigma = 3$.

Appendix A
Upper Bound on the Key Rate

Theorem 7. *Let r be the key rate defined in* (1)*, then, for any $n > 0$,*

$$r \leq I(X;Y) + \frac{1}{n}.$$

Proof: Let $\mathbf{T} \in \{0,1\}^t$ be the set of bits that are communicated between Alice and Bob and are correlated with $\mathbf{W}$. Note that $\mathbf{W}$ can be uniquely determined by $\mathbf{X}$ and $\mathbf{T}$, and that $\mathbf{W}'$ can be uniquely determined by $\mathbf{Y}$ and $\mathbf{T}$.

First, we derive an upper bound on $H(\mathbf{W}|\mathbf{W} = \mathbf{W}') - t$ as follows:

$$\begin{aligned}
&H(\mathbf{W}|\mathbf{W} = \mathbf{W}') - t \\
&\quad \leq H(\mathbf{W}|\mathbf{W} = \mathbf{W}', \mathbf{T}) \\
&\quad = H(\mathbf{XW}|\mathbf{W} = \mathbf{W}', \mathbf{T}) - H(\mathbf{X}|\mathbf{W}, \mathbf{W} = \mathbf{W}', \mathbf{T}) \\
&\quad \leq H(\mathbf{X}|\mathbf{W} = \mathbf{W}', \mathbf{T}) - H(\mathbf{X}|\mathbf{Y}, \mathbf{W}, \mathbf{W} = \mathbf{W}', \mathbf{T}) \\
&\quad = H(\mathbf{X}|\mathbf{W} = \mathbf{W}', \mathbf{T}) - H(\mathbf{X}|\mathbf{Y}, \mathbf{W} = \mathbf{W}', \mathbf{T}) \\
&\quad = I(\mathbf{X}; \mathbf{Y}|\mathbf{W} = \mathbf{W}', \mathbf{T})
\end{aligned}$$

Following the proof of Theorem 1 in [16], specifically Eq. (13) therein, we have

$$I(\mathbf{X}; \mathbf{Y}|\mathbf{W} = \mathbf{W}', \mathbf{T}) \leq I(\mathbf{X}; \mathbf{Y}|\mathbf{W} = \mathbf{W}').$$

We then obtain

$$\begin{aligned}
r =& P(\mathbf{W} = \mathbf{W}') \frac{H(\mathbf{W}|\mathbf{W} = \mathbf{W}') - t}{n} \\
\leq& P(\mathbf{W} = \mathbf{W}') \frac{I(\mathbf{X}; \mathbf{Y}|\mathbf{W} = \mathbf{W}')}{n} \\
\leq& \frac{I(\mathbf{X}; \mathbf{Y}|\mathbf{1}_{(\mathbf{W}=\mathbf{W}')})}{n} \\
\leq& I(X;Y) + \frac{1}{n}.
\end{aligned}$$

This completes the proof. ■

References

[1] R. Ahlswede and I. Csiszár,"Common randomness in information theory and cryptography–Part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1121–1132, Jul. 1993.

[2] I. Ali-Khan and J. C. Howell, "Experimental demonstration of high two-photon time-energy entanglement," *Phys. Rev. A*, vol. 73, 031801(R) (2006).

[3] I. Ali-Khan, C. J. Broadbent, and J. C. Howell, "Large-alphabet quantum key distribution using energy-time entangled bipartite states," *Phys. Rev. Lett.* vol. 98, 060503 (2007).

[4] J. T. Barreiro, N. K. Langford, N. A. Peters, and P. G. Kwiat, "Generation of hyperentangled photon pairs," *Phys. Rev. Lett.* vol. 95, 260501 (2005).

[5] C. H. Bennett, G. Brassard, C. Crepeau, and U. M. Maurer, "Generalized privacy amplification," IEEE Trans. Inf. Theory, vol. 41, no. 6, pp. 1915–1923, 1995.

[6] J. R. Bitner, G. Ehrlich and E. M. Reingold, "Efficient generation of the binary reflected gray code and its applications," *Communications of the ACM*, vol. 19, pp. 517–521, 1976.

[7] N. J. Cerf, M. Bourennance, A. Karlsson, and N. Gisin, "Security of quantum key distribution using d-level systems", *Phys. Rev. Lett.*, vol. 88, 127902 (2002).

[8] M. C. Davey and D. MacKay, "Low density parity check codes over GF(q)," *IEEE Commun. Lett.*, vol. 2, pp. 165–167, 1998.

[9] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, 1998.

[10] M. N. O'Sullivan-Hale, I. Ali-Khan, R. W. Boyd, and J. C. Howell, "Pixel entanglement: experimental realization of optically entangled d=3 and d=6 qudits," *Phys. Rev. Lett.*, vo. 94, 220501 (2005).

[11] J. Hou, P. H. Siegel, L. B. Milstein and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2141–2155, 2003.

[12] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inform. Theory*, vol. 23, pp. 371–377, 1977.

[13] A. Kavčić, X. Ma, and M. Mitzenamcher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," IEEE Trans. Inform. Theory, vol. 49, pp. 1636–1652, 2003.

[14] Y. Kochman and G.W. Wornell, "On high-efficiency optical communication and key distribution" in *Proc. Information Theory and Application Workshop*, February 2012.

[15] M. G. Luby and M. Mitzenmacher, "Verification-based decoding for packet-based low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 120–127, 2005.

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

[17] C. E. Shannon, "Communication theory of secrecy systems," Bell Syst. Tech. J. vol. 28, pp. 656-715, 1949.

[18] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Theory*, vol. 19, pp. 471–480, 1973.

[19] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1361–1391, 1999.

[20] C. Weidmann, "Coding for the q-ary symmetric channel with moderate q," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, pp. 2156–2159, July 2008.