Explicit Design of Provably Covert Channel Codes

Shi-Yuan Wang and Matthieu R. Bloch
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
Email: shi-yuan.wang@gatech.edu, matthieu.bloch@ece.gatech.edu

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517930

***Abstract*—We design an explicit code ensuring provably covert communication over Binary Symmetric Channels (BSCs). This design complements an earlier work that provides a methodology for asymptotic optimal performance but falls short of offering explicit details for operation at finite block length. In particular, we show how to preserve covertness guarantees when facing the unavoidable compromises required by finite block length operation. Compared to the reference scheme without sophisticated coding, our scheme offers orders of magnitude savings in secret key bits. Key ingredients of our design include polar codes for source coding and invertible extractors.**

I. Introduction

While physical-layer security has extensively focused on confidentiality, there has been recent interest in understanding covertness, defined as the ability to ensure low-probability of detection against eavesdroppers. In many situations, covert communication is governed by a *square-root law* [1], by which the number of reliable and covert message bits scales at most as the square root of the block length. While the constant associated to the scaling plays the role of the covert capacity and has now been characterized for several channels and metrics [2]–[4], few actual codes achieving or approaching the limits with low-complexity have been proposed.

Such codes are known to exist, as shown with a concatenated scheme consisting of an outer Reed-Solomon and an inner random code of small length [5]. For asynchronous channels, polar codes provide a possible solution [6]. A non-linear channel code has also been proposed [7] but without theoretical guarantees. The most explicit coding scheme with provable guarantees, in the sense of offering not only a deterministic construction but also low-complexity encoding and decoding algorithms, was recently proposed by combining Pulse-Position Modulation (PPM), Multi-Level Coding (MLC), polar codes, and channel resolvability codes [8], [9].

Our objective here is to fill a missing gap in [8], [9], by which the construction of specific component codes was left to the reader. We offer the full details, including the practical considerations one faces when operating at finite length without compromising covertness. The rest of the paper is organized as follows. After briefly reviewing notation (Section II) and the MLC-PPM coding scheme of [9] (Section III), we show how to construct polar codes (Section IV-B) and invertible extractors (Section V) for MLC-PPM. We conclude with an explicit code design and simulation results (Section VI).

This work was supported by National Science Foundation award 1910859.

II. Notation

Calligraphic letters are reserved for sets, and $|\cdot|$ denotes the cardinality. For two distributions P and Q over the same set $\mathcal{X}$, the relative entropy is $\mathbb{D}(P\,\|\,Q) \triangleq \sum_x P(x)\log_2\frac{P(x)}{Q(x)}$, the variational distance is $\mathbb{V}(P,Q) \triangleq \frac{1}{2}\sum_x |P(x)-Q(x)|$, and the chi-squared distance is $\chi_2(P\,\|\,Q) \triangleq \sum_x \frac{(P(x)-Q(x))^2}{Q(x)}$. For $a,b\in\mathbb{R}$, if $a\leqslant b$, we define $[\![a,b]\!] \triangleq \{\lfloor a\rfloor, \lfloor a\rfloor+1,\cdots,\lceil b\rceil-1,\lceil b\rceil\}$. When used as a superscript or a subscript, $a:b$ also denotes $[\![a,b]\!]$. For $q\in\mathbb{N}^*$, any index set $\mathcal{S}\subseteq[\![1,q]\!]$, and a sequence $(X_i)_{i=1}^q$, $X_\mathcal{S}$ denotes $(X_i)_{i\in\mathcal{S}}$. Using the above notation, we also have $X_{1:q}=(X_i)_{i=1}^q$. For simplicity, we sometimes let X^q or the vector form $\mathbf{X}\triangleq(X_1,\cdots,X_q)^\intercal$ denote a sequence $X_{1:q}$ when it is clear from the context, as each component in $(X_i)_{i=1}^q$ is a scalar.

For $q\in\mathbb{N}^*$, $m\triangleq 2^q$, and $i\in[\![1,m]\!]$, we define a PPM symbol $\widetilde{x}_i$ of order m as a binary vector of length m such that the i-th component is one and all other components are zero. Let $d:\mathbb{F}_2^N\to\mathbb{N}$ be the binary-to-decimal converter, where the *leftmost* bit is the *least-significant* bit, and d^{-1} be its inverse function. For a set $\mathcal{S}\subseteq[\![1,q]\!]$, we define $\mathcal{A}^q(x_\mathcal{S})\triangleq\{j\in[\![1,q]\!]:(d^{-1}(j-1))_\mathcal{S}=x_\mathcal{S}\}$.

A Binary-Input Discrete Memoryless Channel (BI-DMC) with transition probability $W_{Y|X}$ is symmetric if there exists a permutation π of $\mathcal{Y}$ such that $\pi^{-1}=\pi$ and $W_{Y|X}(y|1)=W_{Y|X}(\pi(y)|0)$ for every $y\in\mathcal{Y}$. The use of PPM of order m over a BI-DMC defines a *super-channel* with transition probability $\widetilde{W}_{\widetilde{Y}|\widetilde{X}}=W_{Y|X}^{\otimes m}\triangleq\prod_{i=1}^m W_{Y|X}$ whose input alphabet is $\widetilde{\mathcal{X}}_q\triangleq\{\tilde{x}_i\}_{i=1}^m$ and output alphabet is $\widetilde{\mathcal{Y}}_q\triangleq\mathcal{Y}^{2^q}$.

III. MLC-PPM for Covert Communication

We consider a covert communication scheme over a BI-DMC $W_{Y|X}$ in which a transmitter, Alice, attempts to reliably communicate a uniformly distributed message $W\in[\![1,M]\!]$ with a legitimate receiver, Bob, in the presence of a passive adversary, Willie, who observes Alice's transmission through another BI-DMC $W_{Z|X}$. By convention, channel input "0" is the innocent symbol corresponding to the absence of communication. We also define, for $\theta\in\{0,1\}$, $P_\theta\triangleq W_{Y|X=\theta}$ and $Q_\theta\triangleq W_{Z|X=\theta}$. The encoding process is assisted by a uniformly distributed secret key $S\in[\![1,K]\!]$ shared only between Alice and Bob.

The objective of Alice is two-fold: 1) communicate with Bob reliably, measured by the average probability of error $\mathbb{P}(W\neq\widehat{W})$; 2) escape detection from the adversary, measured by the variational distance $\mathbb{V}(P_{\widetilde{\mathbf{Z}}},Q_0^{\otimes n})$ between the output distribution at Willie induced by the code $P_{\widetilde{\mathbf{Z}}}$ and innocent

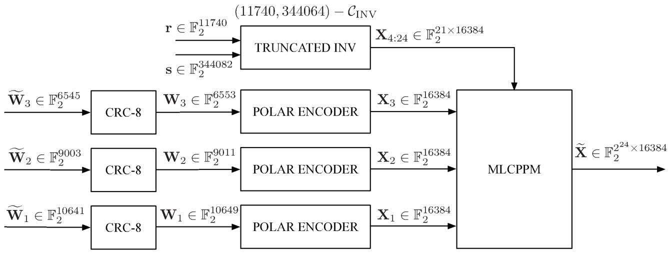
$(11740, 344064) - \mathcal{C}_{\text{INV}}$
$\mathbf{r} \in \mathbb{F}_2^{11740}$
$\mathbf{s} \in \mathbb{F}_2^{344082}$
TRUNCATED INV
$\mathbf{X}_{4:24} \in \mathbb{F}_2^{21\times 16384}$
$\widetilde{\mathbf{W}}_3 \in \mathbb{F}_2^{6545}$
CRC-8
$\mathbf{W}_3 \in \mathbb{F}_2^{6553}$
POLAR ENCODER
$\mathbf{X}_3 \in \mathbb{F}_2^{16384}$
$\widetilde{\mathbf{W}}_2 \in \mathbb{F}_2^{9003}$
CRC-8
$\mathbf{W}_2 \in \mathbb{F}_2^{9011}$
POLAR ENCODER
$\mathbf{X}_2 \in \mathbb{F}_2^{16384}$
MLCPPM
$\widetilde{\mathbf{X}} \in \mathbb{F}_2^{2^{24}\times 16384}$
$\widetilde{\mathbf{W}}_1 \in \mathbb{F}_2^{10641}$
CRC-8
$\mathbf{W}_1 \in \mathbb{F}_2^{10649}$
POLAR ENCODER
$\mathbf{X}_1 \in \mathbb{F}_2^{16384}$

Fig. 1. Illustration of the MLC-PPM encoding scheme with $\ell = 2^{14}$, $m = 2^{24}$. The encoding structure and numbers are those used in an example in Section VI.

distribution $Q_0^{\otimes n}$, where n is the number of channel uses. Because of the two-pronged objective, two coding mechanisms have to be simultaneously deployed: reliability, to control the probability of error, and resolvability [10], [11] to control the variational distance. It is shown in [9] that MLC-PPM with Multi-Stage Decoding (MSD) asymptotically achieves the covert capacity with polar codes. A transmission block of MLC-PPM is composed of ℓ PPM symbols, and each of the PPM symbol is of order $m \triangleq 2^q$, $q \in \mathbb{N}^*$. MLC allows us to decompose the PPM super-channel into q binary channels. Alice then divides her message $W \in [\![1, M]\!]$ into $(W_i)_{i=1}^q$ for each level, encodes every W_i into $\mathbf{X}_i \in \mathbb{F}_2^\ell$ with a component code at each level, and generates PPM symbols $\widetilde{\mathbf{X}} \in \widetilde{\mathcal{X}}_q^\ell$ from the encoded outputs at each level in parallel through a PPM mapper $\tilde{x} : x_{1:q} \mapsto \tilde{x}_{d(x_{1:q})+1}$. The component code of every level has block length ℓ, called the coding block length. The number of channel uses is therefore $n \triangleq m\ell$, called the overall block length. The encoding setup is illustrated in Fig. 1. In practice, as illustrated in Fig. 1 and detailed in the next sections, not all levels are actually used for reliable communication. The lower levels are jointly coded for reliability and resolvability using polar codes with Cyclic Redundancy Checks (CRCs) while the higher levels are coded for resolvability using only invertible extractors.

The choices of n and ℓ are not independent. Specifically, the PPM order m has to be chosen on the order of the coding block length ℓ to respect the square root law and ensure covertness [9], [12]. This creates a potential challenge by which the code rate at each level may change with the overall block length. Fortunately, under MSD starting from level q, the channel at every level $j \in [\![1, q]\!]$ is given by [9, (108)]

$$W_{Y|X}^{(j)}(y^{2^j}|x_j) = \frac{1}{2^j} \sum_{k \in \mathcal{A}^j(x_j)} P_0^{\otimes 2^j}(y^{2^j}) \frac{P_1(y_k)}{P_0(y_k)}, \quad (1)$$

which is invariant with respect to (w.r.t.) the number of levels q. The channel perceived by the eavesdropper is similar, with Q_θ in place of P_θ. This shows that increasing the coding block length only increases the number of levels while leaving each equivalent channel fixed. This also allows one to design reliability and resolvability codes of *fixed rate* at each level, for otherwise polar codes introduced in the following section would suffer from a slow polarization rate. Only the lower levels carry significant capacity, which justifies only coding for reliability and resolvability on the lower levels. For the higher levels of negligible capacity, one can sacrifice secret key bits S to only design a channel resolvability code without worrying about reliability.

Several challenges still remain. While polar codes can ensure joint reliability and resolvability, this requires the identification of "good" and "bad" bit channels. Since [9] relies on *source polarization* instead of channel polarization, one must adopt the approach of [13] to identify the bit channels as described in Section IV. Second, the use of invertible extractors for resolvability must account for constraints, such as choosing a finite field compatible with the target block length as shown in Section V. Third, the analysis in [9], which is based on the relative entropy, is not tight enough at finite length, since it incurs penalties associated with using Pinsker's and reverse Pinsker's inequalities repeatedly. We choose instead the variational distance metric to analyze the covertness. The variational distance is also the metric more operationally relevant to covert communications [4] and leads to a direct analysis in the following sections.

IV. Source Polarization Design

A. Preliminary: channel resolvability

Consider a Discrete Memoryless Channel (DMC) $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$. Let q_X denote the uniform distribution on $\mathcal{X}$ and let Q_Z denote the output distribution induced by q_X. Then, the distribution induced by j independent uses of q_X is

$$Q_{Z^j}(z^j) \triangleq \sum_{\mathbf{x}} W_{Z|X}^{\otimes j}(z^j|\mathbf{x}) q_X^{\otimes j}(\mathbf{x}). \quad (2)$$

The problem of channel resolvability investigates whether one can approximate Q_{Z^j} by using codewords chosen uniformly at random in a (j, K)-code $\mathcal{C}$. Note that the distribution induced by $\mathcal{C}$ is

$$P_{Z^j}(z^j) = \sum_{k=1}^{K} W_{Z|X}^{\otimes j}(z^j|\mathbf{x}_k) \frac{1}{K}. \quad (3)$$

Rates $R = \log_2 K_j / j$ are achievable if there exists a sequence of (j, K_j)-codes such that $\lim_{j\to\infty} \mathbb{V}(P_{Z^j}, Q_{Z^j}) = 0$ [14], [15].

B. Joint reliability and resolvability code

Polar codes with *binning* have been proved in [6], [11], [16], [17] to provide an explicit construction to achieve both reliability and resolvability asymptotically based on source polarization [18]. This requires identifying polarization sets associated to the channel $W_{Y|X}^{(j)}$ and $W_{Z|X}^{(j)}$ for each level of the MLC-PPM scheme. Specifically, we combine the channel input distribution P_X with $W_{Y|X}^{(j)}$ and $W_{Z|X}^{(j)}$, to obtain two Discrete Memoryless Sources (DMSs) $(\mathcal{X}, \mathcal{Y}, P_X W_{Y|X}^{(j)})$ and $(\mathcal{X}, \mathcal{Z}, P_X W_{Z|X}^{(j)})$, and define the polarization sets

$$\mathcal{V}_X \triangleq \{i \in [\![1, \ell]\!] : \mathbb{H}(U_i|U^{i-1}) \geqslant 1 - \eta_\ell\}, \quad (4)$$

$$\mathcal{H}_{X|Y} \triangleq \{i \in [\![1,\ell]\!] : \mathbb{H}\big(U_i|U^{i-1}Y^\ell\big) \geqslant \eta_\ell\}, \tag{5}$$
$$\mathcal{V}_{X|Y} \triangleq \{i \in [\![1,\ell]\!] : \mathbb{H}\big(U_i|U^{i-1}Y^\ell\big) \geqslant 1-\eta_\ell\}, \text{ and} \tag{6}$$
$$\mathcal{V}_{X|Z} \triangleq \{i \in [\![1,\ell]\!] : \mathbb{H}\big(U_i|U^{i-1}Z^\ell\big) \geqslant 1-\eta_\ell\}, \tag{7}$$

where X^ℓ is polarized as $U^\ell \triangleq X^\ell G_\ell$ with polar transform G_ℓ and $\eta_\ell \triangleq 2^{-\ell^\beta}$ with $\beta \in (0,\frac{1}{2})$. Based on (4)-(7), a joint reliability and resolvability code relies on the following sets.

- $\mathcal{V}_C \triangleq \mathcal{H}_{X|Y}\cap\mathcal{V}_{X|Z}$, which contains uniformly distributed bits C representing the code;
- $\mathcal{V}_{C'} \triangleq \mathcal{H}_{X|Y}\cap\mathcal{V}^c_X$, which contains non-uniformly (*almost deterministic*) distributed bits C', which will be secured by using one-time pad with secret key bits S';
- $\mathcal{V}_{W'} \triangleq \mathcal{H}^c_{X|Y}\cap\mathcal{V}_{X|Z}$, which contains uniformly distributed message bits W' that are almost independent from Willie's observation;
- $\mathcal{V}_{\overline{W}} \triangleq (\mathcal{H}^c_{X|Y}\cap\mathcal{V}_X)\setminus\mathcal{V}_{W'}$, which contains additional uniformly distributed message bits $\overline{W}$;
- $\mathcal{V}_S \triangleq \mathcal{H}_{X|Y}\cap\mathcal{V}^c_{X|Z}\cap\mathcal{V}_X$, which contains uniformly distributed secret key bits S.

The set $\mathcal{V}_W \triangleq \mathcal{V}_{W'}\cup\mathcal{V}_{\overline{W}}$ is the set of covert message bits, while $\mathcal{V}_{W'}$ is also held secret from Willie. We also need secret key bits to secure $\mathcal{V}_{C'}\cup\mathcal{V}_S$. Since each equivalent PPM super-channel $W^{(j)}$ in (1) is symmetric, the capacity-achieving input distribution P_X is uniform and therefore $\mathcal{V}^c_X = \varnothing$. Note that if Willie's channel is degraded w.r.t. to Bob's channel, then so is the equivalent channel defined in (1) corresponding to each level for Willie's channel [9, (104)].

C. Source upgrading and degrading

We now detail how to construct the polarization sets such as those in (4)-(7). Given a DMS $(\mathcal{X},\mathcal{Y},P_{XY})$ and a coding block length ℓ, let $(X^\ell,Y^\ell) \sim P^{\otimes\ell}_{XY}$ denote a sequence generated by the source. The source polarization applied on X^ℓ equivalently transforms ℓ independent and identical copies of P_{XY} into ℓ polar bit-sources, defined as

$$P^i_\ell(u_i,u^{i-1}y^\ell) \triangleq \sum_{u_{i+1:\ell}} P^{\otimes\ell}_{XY}(u^\ell G_\ell, y^\ell),$$

via a polar transform G_ℓ. Identifying $\mathcal{H}_{X|Y}$ and $\mathcal{V}_{X|Y}$ requires the knowledge of $P^i_\ell(u_i,u^{i-1}y^\ell)$ to obtain $\mathbb{H}\big(U_i|U^{i-1}Y^\ell\big)$. However, i-th bit source has the alphabet size $2^i\,|\mathcal{Y}|^\ell$, which grows exponentially fast in the block length, making exact calculations intractable.

In the context of channel polarization, [13], [19] suggest mitigating the intractable complexity through a quantization approach, by which the original channels are approximated through *upgrading* and *degrading* while minimizing the capacity difference. Since the channel polarization preserves the degradation and upgradation, we can use such an approximation before each polarization step as in [13, Algorithms A and B] to obtain upper and lower bounds on capacities or error probabilities for each bit channel as desired in the construction of polar codes. This approach can be adapted to source polarization as follows. Similar to the channel polarization, we define recursive transforms for the source distribution as

$$P^-(u_1,y^2) \triangleq \sum_{u_2} P(u_1+u_2,y_1)P(u_2,y_2), \tag{8}$$
$$P^+(u_2,u_1y^2) \triangleq P(u_1+u_2,y_1)P(u_2,y_2). \tag{9}$$

We then recursively construct the bit sources $P^i_\ell(u_i,u^{i-1}y^\ell)$. It is useful to view $(U_i,U^{i-1}Y^\ell) \sim P^i_\ell$ as a binary input process U_i interacting with a BI-DMC $W_{U^{i-1}Y^\ell|U_i}$, in which we call U_i the input and $U^{i-1}Y^\ell$ the output. One can then quantize U^{i-1},Y^ℓ to get upgraded and degraded source approximations, as defined next. For brevity, we denote $(U_i,U^{i-1}Y^\ell)$ by $(\widehat{X},\widehat{Y})$ in this subsection.

Definition 1. *Let $(\widehat{X},\widehat{Y}) \sim P_{\widehat{X}\widehat{Y}}$ and $(\widehat{X},\widehat{Z}) \sim Q_{\widehat{X}\widehat{Z}}$ be DMSs. The source $Q_{\widehat{X}\widehat{Z}}$ is (stochastically) degraded w.r.t. $P_{\widehat{X}\widehat{Y}}$ if there exists an intermediate channel $\Phi_{\widehat{Z}|\widehat{Y}} : \widehat{\mathcal{Y}} \to \widehat{\mathcal{Z}}$ such that*

$$Q_{\widehat{X}\widehat{Z}} = \sum_{y\in\widehat{\mathcal{Y}}} \Phi_{\widehat{Z}|\widehat{Y}}(z|y)P_{\widehat{X}\widehat{Y}}(x,y). \tag{10}$$

Similarly, $P_{\widehat{X}\widehat{Y}}$ is (stochastically) upgraded w.r.t. $Q_{\widehat{X}\widehat{Z}}$. One direct consequence is $\mathbb{H}(\widehat{X}|\widehat{Y}) \leqslant \mathbb{H}(\widehat{X}|\widehat{Z})$.

a) Source polarization construction: We modify the suboptimal channel upgrading and degrading algorithms [13], [19], apply them to the source distribution with polarization, and therefore obtain upper and lower bounds on the conditional entropy of each bit source while minimizing the conditional entropy difference introduced due to the quantization. This allows us to identify the sets in (4)-(7).

b) Encoding and decoding: The encoder and decoder follow from [18] and many decoding algorithms for channel polarization can be adopted, such as Successive Cancellation (SC) decoding [18], [20] and Successive Cancellation List (SCL) decoding [21].

Remark 1. *As the channel $W^{(j)}$ is symmetric and P_X is chosen to be uniform, the above upgrading and degrading could rely on the channel upgrading and degrading of [13], [19]. We have provided the source upgrading and degrading for completeness.*

D. Covertness analysis for polar coding scheme

We next investigate the covertness induced by the polar coding scheme in terms of the variational distance (See the discussion at end of Section III). Let $P_{\widetilde{\mathbf{Z}}}$ be the output distribution of the PPM super-channel when coding over ℓ PPM symbols of order m, and let Q^m_{PPM} be the output induced by a uniform distribution on PPM symbols of order m. For $j \in [\![1,q]\!]$, let $P_{\widetilde{\mathbf{Z}}^{(j)}}$ denote the output distribution induced by a code over the channel $W^{(j)}_{Z|X}$, and let $Q^{\otimes\ell}_{\widetilde{Z}^{(j)}}$ denote the output distribution when the input is uniform. The variational distance $\mathbb{V}(P_{\widetilde{\mathbf{Z}}},(Q^m_{\text{PPM}})^{\otimes\ell})$ satisfies [9, Lemma 8]

$$\mathbb{V}(P_{\widetilde{\mathbf{Z}}},(Q^m_{\text{PPM}})^{\otimes\ell}) \leqslant \sum_{j=1}^{q} \mathbb{V}(P_{\widetilde{\mathbf{Z}}^{(j)}},Q^{\otimes\ell}_{\widetilde{Z}^{(j)}}). \tag{11}$$

It remains to analyze $\mathbb{V}(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell})$ at each level. In fact, since the channel at each level is symmetric, when a *one-shot transmission* with ℓ super-channel uses occurs, we have $\mathbb{V}(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}) = 0$ [6, Lemma 9]. Therefore, the above polar coding scheme does not contribute to the covertness metric.

Remark 2. *Although we do not consider transmission over multiple blocks here, an efficient scheme would reuse part of secret and covert message bits as the key bits for the next transmission. This is referred to as the* chaining *technique [6], [11], [17], [22], [23] and ensures the key rate is asymptotically optimal. For brevity, we only consider the situation in which Willie's channel is degraded w.r.t. Bob's and therefore seek for the design that transmits more covert and secret message bits than the secret key bits it consumes. Handling the general case requires chaining the polar code construction over blocks as shown in [9] but does not present design difficulties. In addition, the randomness bits C are inherently secret from Willie's observation, and can therefore be shared publicly between Alice and Bob and also reused for multiple transmissions.*

However, the bits in $\mathcal{V}_{W'}$ and $\mathcal{V}_C$ are not perfectly *secret, that is, $\mathbb{I}(\widetilde{Z}; W'C)$ is not zero but vanishing in ℓ [6, Lemma 10]. Reusing randomness bits C and secret message bits W' for multiple transmissions must account for this leakage, and therefore $\mathbb{V}(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell})$ is not zero when we consider transmission over blocks [6, Lemma 12].*

V. Invertible Extractor and Channel Resolvability Code

A. Resolvability code design with invertible extractors

In the MLC-PPM scheme, we can design a single channel resolvability code for all the higher levels with a two-universal invertible extractors [9, Section V-D], [24]. Recall that the equivalent channel from level $u+1$ to q is given by [9, (143)]

$$W_{\widetilde{Z}|X}^{u+1:q}(\tilde{z}|x_{u+1:q}) = \frac{1}{2^u} \sum_{k \in \mathcal{A}^q(x_{u+1:q})} Q_0^{\otimes 2^q}(\tilde{z}) \frac{Q_1(z_k)}{Q_0(z_k)}. \quad (12)$$

We instantiate a pair of inverter-extractor for the above equivalent channel containing the following two components:

$$\begin{aligned} &\mathrm{Ext} : \mathbb{S} \times \mathbb{F}_2^{(q-u)\ell} \to \mathbb{F}_2^{(q-u)\ell-k} : (\mathbf{s}, \mathbf{x}_{u+1:q}) \mapsto \mathbf{b}, \\ &\mathrm{Inv} : \mathbb{S} \times \mathbb{F}_2^{(q-u)\ell-k} \times \mathbb{F}_2^k \to \mathbb{F}_2^{(q-u)\ell} : (\mathbf{s}, \mathbf{b}, \mathbf{r}) \mapsto \mathbf{x}_{u+1:q}, \end{aligned}$$

where Ext is a two-universal extractor with seed $\mathbf{s} \in \mathbb{S}$ and bin index $\mathbf{b} \in \mathbb{F}_2^{(q-u)\ell}$. Let $\mathcal{P}_{\mathbf{s},\mathbf{b}} \triangleq \{\mathbf{x}_{u+1:q} \in \mathbb{F}_2^{(q-u)\ell} : \mathrm{Ext}(\mathbf{s}, \mathbf{x}_{u+1:q}) = \mathbf{b}\}$ be the pre-image under $\mathbf{s}$ and $\mathbf{b}$. For any $\mathbf{s}$ and $\mathbf{b}$, we assume that Ext is regular, that is, all the elements in $\mathcal{P}_{\mathbf{s},\mathbf{b}}$ are *uniformly distributed* and $|\mathcal{P}_{\mathbf{s},\mathbf{b}}| = 2^k$. Consider the encoder ϕ defined as $\phi : \mathbb{F}_2^k \to \mathbb{F}_2^{(q-u)\ell} : \mathbf{r} \mapsto \mathrm{Inv}(\mathbf{s}, \mathbf{b}, \mathbf{r})$. Let $P_{\widetilde{\mathbf{Z}}|\mathbf{S}=\mathbf{s},\mathbf{B}=\mathbf{b}}$ be the distribution induced by ϕ with seed $\mathbf{s}$ and bin index $\mathbf{b}$, and let $Q_{\widetilde{\mathbf{Z}}}$ be the distribution induced by a uniform input.

Lemma 2. *The above encoder ϕ with secret key rate $R_{u+1:q} \triangleq \frac{k}{\ell}$ with $\mathbf{s}$ and $\mathbf{b}$ selected randomly according to uniform distributions $q_{\mathbf{S}}$ and $q_{\mathbf{B}}$, respectively, satisfies*

$$\begin{aligned} &\mathbb{E}_{\mathbf{S},\mathbf{B}}\{\mathbb{V}(P_{\widetilde{\mathbf{Z}}|\mathbf{S},\mathbf{B}}, Q_{\widetilde{\mathbf{Z}}})\} \\ &\quad \leqslant 2^{-\frac{\ell\varepsilon^2}{2\log_2^2(2^{q-u}+3)}} + \frac{1}{2}\sqrt{2^{-\ell(R_{u+1:q} - \mathbb{I}(X_{u+1:q};\widetilde{Z}) - \varepsilon)}}, \end{aligned} \quad (13)$$

if $R_{u+1:q} > \mathbb{I}(X_{u+1:q}; \widetilde{Z}) + \varepsilon$ for any $\varepsilon > 0$.

Proof: The proof follows by combining leftover hash lemma [25], [26] with [27, Lemma 3] and [28, Theorem 1]. ■

Similar to [9, (155)], encoding by ϕ achieves channel resolvability regardless of the choice of $\mathbf{b}$ because of the symmetry of the equivalent channel $W_{\widetilde{Z}|X}^{u+1:q}$. We implement an explicit invertible extractor with the finite-field multiplication [10], [29]. For $\mathbf{s} \in \mathbb{S} = \mathbb{F}_2^{(q-u)\ell} \setminus \{\mathbf{0}\}$, define $\mathbf{b} = \mathrm{Ext}(\mathbf{s}, \mathbf{x}_{u+1:q}) \triangleq \left(\mathbf{s}^{-1} \odot \mathbf{x}_{u+1:q}\right)|_{[\![1,(q-u)\ell-k]\!]}$, and $\mathrm{Inv}\,(\mathbf{s}, \mathbf{b}, \mathbf{r}) \triangleq \mathbf{s} \odot (\mathbf{b}\|\mathbf{r})$, where $\odot$ denote the multiplication in the field $\mathbb{F}_2^{(q-u)\ell}$, and $(\cdot)|_{[\![a,b]\!]}$ represents the bit positions in $[\![a, b]\!]$.

Remark 3. *There are efficient ways to implement the finite-field multiplication, e.g., the number-theoretic transform, while the technical difficulty is to find irreducible polynomials to construct large finite fields. For instance, [30, Section 7.3.1] suggests a family of trinomials in the form of $x^a + x^b + 1$, where a is some* Mersenne exponent*, but the available trinomials of this form are quite limited. Another way is the finite-field arithmetic using circulant matrices (FACM) [31], which provides an efficient construction if and only if the desired degree a of irreducible polynomial satisfies 1) $a+1$ is prime, and 2) 2 is a primitive root modulo $a+1$. Compared to the construction based on the Mersenne exponent, FACM provides many more available sizes of finite fields and is more suitable to our scenario.*

B. Truncated invertible extractor

We can efficiently implement the finite-field multiplication when $(q-u)\ell$ is either a *Mersenne exponent* or a size satisfying conditions to apply FACM, but the desired lengths of invertible extractors imposed by the overall block length might not match the available sizes. We present a simple remedy to this inflexibility.

We first generate a resolvability code at a slightly longer block length, which is an available field size closest to the desired length of the code. Then, we truncate some bits of the resulting codeword to fit in the desired coding scheme.

Formally, for $n' > n$, a given seed $\mathbf{s} \in \mathbb{S}$, a bin index $\mathbf{b} \in \mathbb{F}_2^{n'-k}$, and an inverter

$$\mathrm{Inv} : \mathbb{S} \times \mathbb{F}_2^{n'-k} \times \mathbb{F}_2^k \to \mathbb{F}_2^{n'} : (\mathbf{s}, \mathbf{b}, \mathbf{r}) \mapsto x^{n'} \triangleq \mathbf{s} \odot (\mathbf{b}\|\mathbf{r}),$$

the encoders ϕ' and ϕ'' are defined as $\phi' : \mathbb{F}_2^k \to \mathbb{F}_2^{n'} : \mathbf{r} \mapsto \mathrm{Inv}(\mathbf{s}, \mathbf{b}, \mathbf{r})$, and $\phi'' : \mathbb{F}_2^k \to \mathbb{F}_2^n : \mathbf{r} \mapsto \mathrm{Inv}(\mathbf{s}, \mathbf{b}, \mathbf{r})|_{[\![1,n]\!]}$, respectively, where we have used n' and n denoting the lengths of code before and after truncation. Now, let $X^{n'} \triangleq (X^n, X^{n_0})$. The truncated coding scheme induces

$$P_{Z^n}(z^n) \triangleq \sum_{z^{n_0}} \sum_{x^{n'} \in \mathcal{P}_{\mathbf{s},\mathbf{b}}} \frac{1}{|\mathcal{P}_{\mathbf{s},\mathbf{b}}|} W_{Z|X}^{\otimes n'}(z^{n'}|x^{n'}). \tag{14}$$

Similarly, the uniformly generated input induces

$$Q_{Z^n}(z^n) \triangleq \sum_{z^{n_0}} \sum_{x^{n'}} \frac{1}{2^{n'}} W_{Z|X}^{\otimes n'}(z^{n'}|x^{n'}). \tag{15}$$

The following lemma, which follows from the fact that marginalization does not increase variational distance [15], shows that the truncation does not increase the variational distance between the distribution induced by the resolvability code and target distribution.

Lemma 3. *If the encoder ϕ' defined above with $\mathbf{S}$ and $\mathbf{B}$ selected randomly according to uniform distributions satisfies $\lim_{n' \to \infty} \mathbb{E}_{\mathbf{S},\mathbf{B}}\{\mathbb{V}(P_{Z^{n'}}, Q_{Z^{n'}})\} = 0$, then the encoder ϕ'' also satisfies $\lim\limits_{n \to \infty} \mathbb{E}_{\mathbf{S},\mathbf{B}}\{\mathbb{V}(P_{Z^n}, Q_{Z^n})\} = 0$.*

Specifically, for any $\mathbf{s}$ and $\mathbf{b}$, $\mathbb{V}(P_{Z^n}, Q_{Z^n}) \leqslant \mathbb{V}(P_{Z^{n'}}, Q_{Z^{n'}})$.

VI. Numerical Results

The overall contribution of polar codes and invertible extractor in the MLC-PPM scheme to the variational distance is as follows: for any $\mathbf{b} \in \mathbb{F}_2^{(q-u)\ell-k}$ and $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E}_{\mathbf{S}}\{\mathbb{V}(P_{\widetilde{\mathbf{Z}}}, Q_0^{\otimes n})\} &\leqslant \sum_{j=1}^{u} \mathbb{V}(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}) \\ &\quad + \mathbb{E}_{\mathbf{S}}\{\mathbb{V}(P_{\widetilde{\mathbf{Z}}|\mathbf{S},\mathbf{B}=\mathbf{b}}, Q_{\widetilde{\mathbf{Z}}})\} + \mathbb{V}((Q_{\text{PPM}}^m)^{\otimes \ell}, Q_0^{\otimes n}) \\ &\leqslant 2^{-\frac{\ell\varepsilon^2}{2\log_2^2(2^{q-u}+3)}} + \frac{1}{2}\sqrt{2^{-\ell(R_{u+1:q} - \mathbb{I}(X_{u+1:q};\widetilde{Z}) - \varepsilon)}} \\ &\quad + \sqrt{\frac{\delta}{2} + \mathcal{O}\left(\frac{1}{m}\right)} = \delta_t, \end{aligned} \tag{16}$$

where u is the number of level used for reliability, and δ is the parameter that governs the covert stochastic process Q_{PPM}^m, defined in Section IV-D, in terms of $\mathbb{V}((Q_{\text{PPM}}^m)^{\otimes \ell}, Q_0^{\otimes n}) \leqslant \sqrt{\frac{\delta}{2} + \mathcal{O}\left(\frac{1}{m}\right)}$ such that $\ell = \left\lceil \frac{2\delta m}{\chi_2(Q_1 \,\|\, Q_0)} \right\rceil$ as in [12]. We ignore the $\mathcal{O}(\frac{1}{m})$ term in the following numerical evaluations, since it is on the order of 2^{-q}.

We finally conclude with a complete design example shown in Fig. 1. We consider the situation in which the main channel $W_{Y|X}$ and Willie's channel $W_{Z|X}$ are BSCs with crossover probabilities 0.1 and 0.42, respectively. By choosing $\delta = 2 \times 10^{-4}$ with coding block length of polar component codes $\ell = 2^{14}$, the minimum number of levels of MLC is $q = 24$, which means the order of PPM m is 2^{24}. The overall block length of the MLC-PPM $n = \ell m$ is 2^{38}. The capacities of the lowest 3 levels of the main channel are $0.742, 0.638,$ and 0.492 bits, respectively, while those of Willie's channel are $0.037, 0.019,$ and 0.0095 bits, respectively. We use the lowest 3 levels for reliability, and the code rates of each polar component code are chosen experimentally to be $0.65, 0.55,$ and 0.4, respectively, to back off enough from capacities and meet our reliability constraint of the error rate close to 10^{-5}. To further enhance the performance of SCL decoding, we also use a CRC-8 precoding for each polar component code, which introduces an additional rate penalty and requires extra 8 bits of secret key at each level to hide the dependency by using one-time pads with the appended bits. This amounts to 26189 covert message bits, with a Bit-Error-Rate (BER) performance 1.15×10^{-5} with list size 16 evaluated after 10000 iterations of simulation.

For the levels from 4 to 24, we design the invertible extractor described as follows. The length of required output size $(q-u)\ell$ is 344064, so the inverter operates in $\mathbb{F}_2^{344082}$ by using FACM, and then we truncate the result. We choose $\varepsilon = 0.7$ and the secret key rate $R_{u+1:q} \triangleq \mathbb{I}(X_{u+1:q}; \widetilde{Z}) + 1.01\varepsilon$ with $\mathbb{I}(X_{4:24}; \widetilde{Z}) = 0.0095$, which requires 11740 bits of secret key for the inverter, while the resolvability from the inverter $\mathbb{E}_{\mathbf{S}}\{\mathbb{V}(P_{\widetilde{\mathbf{Z}}|\mathbf{S},\mathbf{B}=\mathbf{b}}, Q_{\widetilde{\mathbf{Z}}})\}$ is upper-bounded by 1.82×10^{-3}. The total covertness metric (16) is upper bounded by $\delta_t = 1.182 \times 10^{-2}$. Note that the theoretical covert capacity of $W_{Y|X}$ derived in [4, (32)] is $\frac{2}{\sqrt{\chi_2(Q_1 \,\|\, Q_0)}} \mathbb{D}(P_1 \,\|\, P_0) = 15.65$, and the number of covert message bits of the present scheme is 26189, i.e., the covert throughput is $\frac{26189}{\sqrt{n} Q^{-1}\left(\frac{1-\delta_t}{2}\right)} \approx 3.372$.

With $\beta = 0.39$, we obtain the sizes of $\mathcal{V}_{W'}$ and $\mathcal{V}_S$ at each level based on upgrading, and we have $7751, 6817$ and 4942 bits of covert and secret message bits, while we do not need any secret key bits for $\mathcal{V}_S$ because $|\mathcal{V}_S|$ is zero at each level. This amounts to 19510 covert and secret message bits, which is more than the number of secret key bits consumed (11764 bits in total, including secret key bits used in inverter, $\mathcal{V}_S$ and CRC). Therefore, we have an efficient one-shot transmission scheme in terms of the secret key usage.

As a less favorable scenario, we consider the case where Willie's channel is a BSC with crossover probability 0.34, while the main channel and other parameters (i.e., $\ell, \delta, \varepsilon, \beta$) remain unchanged. The capacities of the lowest 3 levels of Willie's channel are $0.143, 0.08,$ and 0.042 bits, respectively. We need $m = 2^{26}$, $n = 2^{40}$, and $(q-u)\ell = 376832$. The inverter operates in $\mathbb{F}_2^{376836}$. In this case, $\mathbb{I}(X_{4:26}; \widetilde{Z}) = 0.0414$, which requires 12263 bits of secret key for the inverter, we have $\mathbb{E}_{\mathbf{S}}\{\mathbb{V}(P_{\widetilde{\mathbf{Z}}|\mathbf{S},\mathbf{B}=\mathbf{b}}, Q_{\widetilde{\mathbf{Z}}})\} \leqslant 5.2 \times 10^{-3}$ and $\delta_t = 1.52 \times 10^{-2}$. The covert throughtput is $\frac{26189}{\sqrt{n} Q^{-1}\left(\frac{1-\delta_t}{2}\right)} \approx 1.311$, while the covert capacity is 7.508. Also, the covert and secret bits at each level are $4603, 4447,$ and 3267, respectively, while the additional secret key bits required for $\mathcal{V}_S$ are $0, 4,$ and 23 bits, respectively. We still have an efficient scheme, as $12317 > 12263 + 24 + 27$.

In summary, for the one-shot transmission of the present MLC-PPM schemes, we require fewer secret key bits than the number of covert and secret message bits transmitted. Although secret key bits are required in our one-shot transmission, the scheme in [1] would require $26189 + 26189 \times 38 \approx 10^6$ secret key bits to enable the covert communication, which is 2 orders of magnitude more than ours.

References

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of Reliable Communication with Low Probability of Detection on AWGN Channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[2] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental Limits of Communication With Low Probability of Detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[3] M. R. Bloch, "Covert Communication over Noisy Channels: A Resolvability Perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[4] M. Tahmasbi and M. R. Bloch, "First- and Second-Order Asymptotics in Covert Communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[5] Q. Zhang, M. Bakshi, and S. Jaggi, "Covert Communication With Polynomial Computational Complexity," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1354–1384, Mar. 2020.

[6] G. Frèche, M. Bloch, and M. Barret, "Polar Codes for Covert Communications over Asynchronous Discrete Memoryless Channels," *Entropy*, vol. 20, no. 1, p. 3, Dec. 2017.

[7] M. Lamarca and D. Matas, "A non-linear channel code for covert communications," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, Morocco, Apr. 2019, pp. 1–7.

[8] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Codes for Covert Communication over Additive White Gaussian Noise Channels," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, Jul. 2019, pp. 977–981.

[9] ——, "Multilevel-Coded Pulse-Position Modulation for Covert Communications Over Binary-Input Discrete Memoryless Channels," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6001–6023, Oct. 2020.

[10] R. A. Chou, M. R. Bloch, and J. Kliewer, "Low-complexity channel resolvability codes for the symmetric multiple-access channel," in *Proc. of IEEE Information Theory Workshop*, Hobart, TAS, Australia, Nov. 2014, pp. 466–470.

[11] R. A. Chou and M. R. Bloch, "Polar Coding for the Broadcast Channel with Confidential Messages: A Random Binning Analogy," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2410–2429, May 2016.

[12] M. R. Bloch and S. Guha, "Optimal covert communications using pulse-position modulation," in *proc. of IEEE International Symposium on Information Theory*, Aachen, Germany, Jun. 2017, pp. 2825–2829.

[13] I. Tal and A. Vardy, "How to Construct Polar Codes," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[14] T. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[15] P. Cuff, "Distributed Channel Synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[16] M. Bloch, M. Hayashi, and A. Thangaraj, "Error-Control Coding for Physical-Layer Secrecy," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1725–1746, Oct. 2015.

[17] R. A. Chou, M. R. Bloch, and E. Abbe, "Polar Coding for Secret-Key Generation," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6213–6237, Nov. 2015.

[18] E. Arikan, "Source polarization," in *Proc. of IEEE International Symposium on Information Theory*, Austin, TX, Jun. 2010, pp. 899–903.

[19] A. Kartowsky and I. Tal, "Greedy-merge degrading has optimal power-law," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 917–934, Feb. 2019.

[20] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[21] I. Tal and A. Vardy, "List Decoding of Polar Codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[22] E. Sasoglu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *proc. of IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013, pp. 1117–1121.

[23] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving Marton's Region for Broadcast Channels Using Polar Codes," *IEEE Transactions on Information Theory*, vol. 61, no. 2, pp. 783–800, Feb. 2015.

[24] M. Tahmasbi and M. R. Bloch, "Toward Undetectable Quantum Key Distribution Over Bosonic Channels," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 2, pp. 585–598, Aug. 2020.

[25] R. Renner, "Security of quantum key destribution," *International Journal of Quantum Information*, vol. 6, no. 1, pp. 1–127, Feb. 2008.

[26] S. Watanabe and M. Hayashi, "Non-asymptotic analysis of privacy amplification via Rényi entropy and inf-spectral entropy," in *proc. of IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013, pp. 2715–2719.

[27] R. A. Chou, "Explicit Codes for the Wiretap Channel with Uncertainty on the Eavesdropper's Channel," in *Proc. of IEEE International Symposium on Information Theory*, Vail, CO, Aug. 2018, pp. 476–480.

[28] T. Holenstein and R. Renner, "On the Randomness of Independent Experiments," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1865–1871, Apr. 2011.

[29] M. Bellare and S. Tessaro, "Polynomial-Time, Semantically-Secure Encryption Achieving the Secrecy Capacity," *arXiv preprint*, vol. 1201.3160, Jan. 2012.

[30] G. Van Assche, *Quantum Cryptography and Secret-Key Distillation*. Cambridge: Cambridge University Press, 2006.

[31] M. Hayashi and T. Tsurumaru, "More Efficient Privacy Amplification With Less Random Seeds via Dual Universal Hash Function," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2213–2232, Apr. 2016.