# Countably Infinite Multilevel Source Polarization for Non-Stationary Erasure Distributions 

Yuta Sakai Department of ECE, National University of Singapore<br>Email: eleyuta@nus.edu.sg

Ken-ichi Iwata Hiroshi Fujisaki<br>Graduate School of Engineering, Graduate School of Natural Science and Technology, University of Fukui<br>Email: k-iwata@u-fukui.ac.jp<br>Email: fujisaki@ec.t.kanazawa-u.ac.jp


#### Abstract

Polar transforms are central operations in the study of polar codes. This paper examines polar transforms for nonstationary memoryless sources on possibly infinite source alphabets. This is the first attempt of source polarization analysis over infinite alphabets. The source alphabet is defined to be a Polish group, and we handle the Arikan-style two-by-two polar transform based on the group. Defining erasure distributions based on the normal subgroup structure, we give recursive formulas of the polar transform for our proposed erasure distributions. As a result, the recursive formulas lead to concrete examples of multilevel source polarization with countably infinite levels when the group is locally cyclic. We derive this result via elementary techniques in lattice theory.


## I. Introduction

Polar codes were invented by Arıkan [1] as a provably capacity-achieving channel coding technique for binary-input memoryless symmetric channels with low complexity encoding/decoding. The central operation in polar coding is socalled the polar transform, which creates worse and better channels than an original channel in a certain sense. It was shown that this coding technique can also be applied to binary source coding problems with side information [2]. Şaşoğlu [3] extended polar source coding from binary to non-binary alphabets. While Şaşoğlu's polar transform [3, Definition 4.1] is non-linear in a certain sence, Mori-Tanaka [4] established non-binary polar source coding with linear polar transforms over finite fields. In those studies, the polar transforms asymptotically create either deterministic or equiprobable conditional probability distributions, and these limiting proportions can be fully characterized by the conditional Shannon entropy of the original source. Such a two-level polarization phenomenon is sometimes referred to as strong polarization. ${ }^{1}$

## A. From Two-Level to Multilevel Polarization

On the other hand, there are studies of investigating polarization phenomena with three or more polarization levels

This work is supported in part by JSPS KAKENHI Grant Numbers 26420352 ; 17K06422; 17J11247; and 18K11465, and a Singapore NRF fellowship (R-263-000-D02-281).
${ }^{1}$ Note that the terminology "strong polarization" is somewhat ambiguous in the literature. Mori-Tanaka [4] said that a channel is polarized in a weaker sense if its polar transforms do not behave as the two-level polarization; Nasser [5] said that a channel is strongly polarizing if its polar transforms behave as the two-level polarization; and Błasiok et al. [6, Definition 1.4] defined a meaning of strong polarization in a different way.
[5], [7]-[16]. Such a phenomenon is referred to as multilevel polarization. Roughly speaking, multilevel polarization means that the polar transform asymptotically creates equiprobable conditional distributions on some cosets of normal subgroups, provided that the polar transform is based on a finite group (cf. [10, Theorem 6] and [15, Theorem V.1]). Unlike twolevel polarization, characterizing the limiting proportions of multilevel polarization still remains an open problem ${ }^{2}$ in general. In this paper, such limiting proportions are referred to as asymptotic distribution of multilevel polarization. In practice, the asymptotic distribution is an important indicator of constructing polar codes [17]. As a special case, the authors of this paper [13], [14] solved the asymptotic distribution for non-binary-input erasure-like channel models proposed in [12]. Particularly, if the input alphabet size is not a prime power, then our previous works [13], [14] give a non-trivial method for calculating the exact asymptotic distribution algorithmically. Recently, Nasser [5] characterized polarization levels, i.e., the support of the asymptotic distribution, for his proposed channels called automorphic-symmetric channels.

## B. Binary Operation defining Polar Transform

While the Arıkan-style two-by-two polar transform is always defined on a group of order two ${ }^{3}$ in the study of binary polar codes, note that it can be defined on several binary operations in the study of non-binary polar codes. Şaşoğlu [3, Definition 4.1] considered it based on certain finite quasigroups; Mori-Tanaka [4] on finite fields; Park-Barg [16] on cyclic groups in which those orders are a power of two; Sahebi-Pradhan [15] on finite abelian groups; Nasser-Telatar [10] on finite quasigroups; Nasser-Telatar [11] also on finite abelian groups; Nasser [7], [8] on more weaker finite algebras than quasigroups; Nasser [5] also on finite, not necessarily abelian, groups; and the authors [12]-[14] on finite cyclic groups. It is worth pointing out that by the structure theorem of finite abelian groups, the polar transform based on a finite abelian group can be considered as a polar coding for the multiple-access channel (MAC) [10], [11]. The neccesary and sufficient condition of MACs that the entire capacity region can be achieved by the MAC polarization has

[^0]been characterized by Nasser-Telatar [11] via discrete Fourier analysis over a finite abelian group.

## C. Related Problem: Sumset Inequality for Shannon Entropy

Beyond the study of polar codes, bounding some measures and/or entropy of a random variable (r.v.) generated by a group action between two independent r.v.'s are central issues [18][25]. This problem is information-theoretically analogous to additive combinatorics or Ruzsa calculus (cf. [26]). Ruzsa [18] and Tao [19] established the sumset and inverse sumset estimates on the Shannon entropy of abelian group-valued r.v.'s with countable order of the group (see also [20]). Kontoyiannis and Madiman [21] extended such estimates to the differential entropy of real-valued r.v.'s, where the group operation is addition of real numbers. Madiman and Kontoyiannis [24] further generalized such estimates to the differential entropy of Polish, locally compact, and abelian group-valued r.v.'s, where the probability density functions are defined with respect to a Haar measure. Applications of such sumset inequalities for the Shannon entropy to polar codes have been discussed in the literature [22], [23].

## D. Contributions of This Study

In this study, we considers the recursive application of Arıkan-style two-by-two polar transforms based on Polish groups with possibly infinite order. This is the first attempt of polarization analysis over infinite alphabets. The main aim of this study is to explore the asymptotic distribution of multilevel polarization. In the context of polar coding for a non-stationary source, we consider a mutually independent, but not necessarily identically distributed, sequence of group-valued r.v.'s with side information. This non-stationary setting is similar to the study of binary polar coding for non-stationary channels [27], [28]. Moreover, examining the asymptotic behavior of the conditional Shannon entropy with the recursive group actions among independent r.v.'s is somewhat related to the study of sumset inequalities for the Shannon entropy [18]-[25].

We tackle this problem by proposing erasure distributions based on the normal subgroup structure of a given group, see Definition 2 of Section III-A. Our group-based erasure distributions are a generalization of Şaşoğlu's one [3, Section 3.3.1], which is defined on the binary source alphabet. To simplify our analysis, we derive the recursive formulas of the polar transform for the group-based erasure distributions in Theorem 1 of Section III-B. When the group is locally cyclic, in Theorem 2 of Section IV-B, we give a method of computing the exact asymptotic distribution of multilevel polarization for groupbased erasure distributions. Theorem 2 is proved via the lattice structure of the normal subgroups [29], [30]. These results are more abstractly general than that of the authors' previous works [12]-[14], and the first instance of countably infinite polarization levels. We decsribe a relation to our previous works [12]-[14] in Section V-A1; we give the simplest case of countably infinite polarization levels with the Prüfer $p$-group in Section V-A2.

## II. Problem Formulation and Basic Lemmas

## A. Conditional Distribution and Conditional Shannon Entropy

Let $(\mathcal{X}, \mathcal{B})$ be a standard Borel space, $X$ an $(\mathcal{X}, \mathcal{B})$-valued r.v., and $Y$ a r.v. Denote by $P_{X \mid Y}$ a regular conditional distribution ${ }^{4}$ of $X$ relative to $Y$, i.e., it is a r.v. forming a probability measure on $(\mathcal{X}, \mathcal{B})$ almost surely (a.s.) and $P_{X \mid Y}(B)$ is a version of the conditional probability $\mathbb{P}\left(X^{-1}(B) \mid Y\right)$ for each $B \in \mathcal{B}$. Definition 1 introduces an equivalence relation between two r.v.'s relative to another r.v.

Definition 1. We say that two r.v.'s $Y$ and $Z$ are equivalent relative to an $(\mathcal{X}, \mathcal{B})$-valued r.v. $X$, denoted by $Y \equiv_{X} Z$, if $P_{X \mid Y}(B)=P_{X \mid Z}(B)$ a.s. for every $B \in \mathcal{B}$.

Definition 1 can be reduced to the equivalence relation $\stackrel{i}{\sim}$ introduced by Mori-Tanaka [4, p. 2722], provided that $\mathcal{X}$ is a finite alphabet. Now, we say that $X$ is conditionally discrete relative to $Y$ if there exists a $\mathcal{B}$-valued r.v. $D$ such that $D$ is countable a.s., and $P_{X \mid Y}(D)=1$ a.s. Then, the conditional entropy is defined by

$$
\begin{equation*}
H(X \mid Y):=\mathbb{E}\left[\sum_{x \in \mathcal{X}} P_{X \mid Y}(x) \log \frac{1}{P_{X \mid Y}(x)}\right] \tag{1}
\end{equation*}
$$

provided that $X$ is conditionally discrete relative to $Y$. Here, $\log$ stands for the natural logarithm satisfying $0 \log 0=0$.

The following lemma is trivial from the definitions.
Lemma 1. If $Y \equiv_{X} Z$, and if $X$ is conditionally discrete relative to $Y$ or $Z$, then it holds that

$$
\begin{equation*}
H(X \mid Y)=H(X \mid Z) \tag{2}
\end{equation*}
$$

Namely, the equivalence relation defined in Definition 1 classifies pairs of r.v.'s having the equal conditional Shannon entropy. Lemma 1 can be straightforwardly extended to a more general conditional quantity [33, Equation (4)].

## B. One-Step Polar Transform with a Polish Group

Let $G$ be a Polish group with group operation •, i.e., it is a topological group equipped with a complete separable metrizable topology. Here, the group $G$ is not necessarily abelian. Denote by $\mathcal{B}_{G}$ the Borel $\sigma$-algebra induced by the Polish topology of $G$. Clearly, the measurable space ( $G, \mathcal{B}_{G}$ ) is standard Borel. Assume that the mapping $(g, h) \mapsto g \bullet h$ is Borel-measurable; and consider two independent, but not necessarily identically distributed, $\left(G, \mathcal{B}_{G}\right)$-valued r.v.'s $X_{1}$ and $X_{2}$. The one-step polar transform generates two $\left(G, \mathcal{B}_{G}\right)$-valued r.v.'s $U_{1}$ and $U_{2}$ by

$$
\begin{align*}
& U_{1}=X_{1} \bullet X_{2}  \tag{3}\\
& U_{2}=X_{2} \tag{4}
\end{align*}
$$

The following lemma will be useful to simplify the subsequent analysis on the polar transform (3)-(4) for erasure distributions defined later in Definition 2.

[^1]Lemma 2. Given four r.v.'s $Y_{1}, Y_{2}, Z_{1}$, and $Z_{2}$, suppose that

$$
\begin{align*}
Y_{1} & \equiv X_{1} Z_{1},  \tag{5}\\
Y_{2} & \equiv X_{2} Z_{2},  \tag{6}\\
\left(X_{1}, Y_{1}, Z_{1}\right) & \Perp\left(X_{2}, Y_{2}, Z_{2}\right) . \tag{7}
\end{align*}
$$

Then, it holds that

$$
\begin{align*}
\left(Y_{1}, Y_{2}\right) & \equiv_{U_{1}}\left(Z_{1}, Z_{2}\right),  \tag{8}\\
\left(U_{1}, Y_{1}, Y_{2}\right) & \equiv_{U_{2}}\left(U_{1}, Z_{1}, Z_{2}\right) . \tag{9}
\end{align*}
$$

Proof of Lemma 2: We first prove $\left(U_{1}, Y_{1}, Y_{2}\right) \equiv_{U_{2}}$ $\left(U_{1}, Z_{1}, Z_{2}\right)$. For each $A_{1}, A_{2} \in \mathcal{B}_{G}$, it holds that

$$
\begin{align*}
& \mathbb{E}\left[\mathbf{1}_{\left\{U_{1} \in A_{1}\right\}} P_{U_{2} \mid U_{1}, Y_{1}, Y_{2}}\left(A_{2}\right) \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\text { a) }}{=} \mathbb{E}\left[\mathbf{1}_{\left\{U_{1} \in A_{1}\right\}} \mathbb{E}\left[\mathbf{1}_{\left\{U_{2} \in A_{2}\right\}} \mid U_{1}, Y_{1}, Y_{2}\right] \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\mathrm{b})}{=} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\left\{\left(U_{1}, U_{2}\right) \in A_{1} \times A_{2}\right\}} \mid U_{1}, Y_{1}, Y_{2}\right] \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\mathrm{c})}{=} \mathbb{E}\left[\mathbf{1}_{\left\{\left(U_{1}, U_{2}\right) \in A_{1} \times A_{2}\right\}} \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\mathrm{d})}{=} \mathbb{E}\left[\mathbf{1}_{\left\{U_{2} \in A_{2}\right\}} P_{U_{1} \mid U_{2}, Y_{1}, Y_{2}}\left(A_{1}\right) \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\mathrm{e})}{=} \mathbb{E}\left[\mathbf{1}_{\left\{X_{2} \in A_{2}\right\}} \mathbb{E}\left[\mathbf{1}_{\left\{X_{1} \bullet X_{2} \in A_{1}\right\}} \mid X_{2}, Y_{1}, Y_{2}\right] \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\mathrm{f})}{=} \mathbb{E}\left[\mathbf{1}_{\left\{X_{2} \in A_{2}\right\}} P_{X_{1} \mid Y_{1}, Y_{2}}\left(A_{1} \bullet X_{2}^{-1}\right) \mid Y_{1}, Y_{2}\right] \\
& \stackrel{(\mathrm{g})}{=} \mathbb{E}\left[\mathbf{1}_{\left\{X_{2} \in A_{2}\right\}} P_{X_{1} \mid Y_{1}}\left(A_{1} \bullet X_{2}^{-1}\right) \mid Y_{1}, Y_{2}\right] \\
& \stackrel{\text { h) }}{=} \int_{A_{2}} P_{X_{1} \mid Y_{1}}\left(A_{1} \bullet x_{2}^{-1}\right) P_{X_{2} \mid Y_{1}, Y_{2}}\left(\mathrm{~d} x_{2}\right) \\
& \stackrel{\text { (i) }}{=} \int_{A_{2}} P_{X_{1} \mid Y_{1}}\left(A_{1} \bullet x_{2}^{-1}\right) P_{X_{2} \mid Y_{2}}\left(\mathrm{~d} x_{2}\right) \\
& \stackrel{\text { (j) }}{=} \int_{A_{2}} P_{X_{1} \mid Z_{1}}\left(A_{1} \bullet x_{2}^{-1}\right) P_{X_{2} \mid Z_{2}}\left(\mathrm{~d} x_{2}\right) \\
& \stackrel{\text { (k) }}{=} P_{U_{1}, U_{2} \mid Z_{1}, Z_{2}}\left(A_{1} \times A_{2}\right) \\
& \stackrel{(1)}{=} \mathbb{E}\left[\mathbf{1}_{\left\{U_{1} \in A_{1}\right\}} P_{U_{2} \mid U_{1}, Z_{1}, Z_{2}}\left(A_{2}\right) \mid Z_{1}, Z_{2}\right] \tag{10}
\end{align*}
$$

a.s., where (a) follows by the definition of conditional probabilities; (b) follows from the fact that $\mathbf{1}_{\left\{U_{1} \in A_{1}\right\}}$ is $\sigma\left(U_{1}, Y_{1}, Y_{2}\right)$ measurable; (c) follows from the fact that

$$
\begin{equation*}
\mathcal{G} \subset \mathcal{H} \Longrightarrow \mathbb{E}[Z \mid \mathcal{G}]=\mathbb{E}[\mathbb{E}[Z \mid \mathcal{H}] \mid \mathcal{G}] \quad \text { (a.s.) } \tag{11}
\end{equation*}
$$

for a real-valued r.v. $Z$ (cf. [34, Theorem 9.1.5]); (d) follows as in (a)-(c); (e) follows by (3)-(4); (f) follows from $X_{1} \Perp X_{2}$ $\left(Y_{1}, Y_{2}\right)$ together with [34, Theorem 9.2.2]; (g) follows from [34, Theorem 9.2.1] and the fact that $\left(X_{1}, Y_{1}\right) \Perp\left(X_{2}, Y_{2}\right)$ implies $X_{1} \Perp Y_{2} \mid Y_{1}$; (h) follows since $P_{X_{2} \mid Y_{1}, Y_{2}}(\cdot)$ forms a probability measure on $\left(G, \mathcal{B}_{G}\right)$ a.s.; (i) follows from [34, Theorem 9.2.1] and the fact that $\left(X_{1}, Y_{1}\right) \Perp\left(X_{2}, Y_{2}\right)$ implies $X_{2} \Perp Y_{1} \mid Y_{2} ;(\mathrm{j})$ follows by $Y_{1} \equiv_{X_{1}} Z_{1}$ and $Y_{2} \equiv_{X_{2}} Z_{2}$; (k) follows as in (d)-(f); and (l) follows as in (a)-(c). Namely, Equation (10) is equivalent to $P_{U_{2} \mid U_{1}, Y_{1}, Y_{2}}\left(A_{2}\right)=P_{U_{2} \mid U_{1}, Z_{1}, Z_{2}}\left(A_{2}\right)$ a.s. for $A_{2} \in \mathcal{B}_{G}$; and thus, we have $\left(U_{1}, Y_{1}, Y_{2}\right) \equiv_{U_{2}}\left(U_{1}, Z_{1}, Z_{2}\right)$, as desired.

We next prove $\left(Y_{1}, Y_{2}\right) \equiv_{U_{1}}\left(Z_{1}, Z_{2}\right)$. By setting $A_{2}=G$, it can be verified from (10) that for each $A_{1} \in \mathcal{B}_{G}$,

$$
P_{U_{1} \mid Y_{1}, Y_{2}}\left(A_{1}\right)=\int_{G} P_{X_{1} \mid Y_{1}}\left(A_{1} \bullet x_{2}^{-1}\right) P_{X_{2} \mid Y_{2}}\left(\mathrm{~d} x_{2}\right)
$$



Fig. 1. Recursive construction of polar transforms up to $n \leq 3$ (see (13)-(14)).

$$
\begin{align*}
& =\int_{G} P_{X_{1} \mid Z_{1}}\left(A_{1} \bullet x_{2}^{-1}\right) P_{X_{2} \mid Z_{2}}\left(\mathrm{~d} x_{2}\right) \\
& =P_{U_{1} \mid Z_{1}, Z_{2}}\left(A_{1}\right) \tag{12}
\end{align*}
$$

a.s. Thus, we have $\left(Y_{1}, Y_{2}\right) \equiv_{U_{1}}\left(Z_{1}, Z_{2}\right)$, as desired.

Lemma 2 means that the polar transform (3)-(4) preserves the equivalence relation defined in Definition 1. In the context of polar channel codes, as an analoguos result to Lemma 2, the preserving property of a certain channel ordering was firstly shown by Korada [35, Lemma 4.7] (see also [17, Lemma 5]), and its generalization from binary to arbitrary input alphabets was given in [13, Lemma 2] under a quasigroup operation.

## C. Recursive Construction of Polar Transforms

This subsection introduces the recursive applications of the polar transform (3)-(4). For a sequence $\left\{X_{i}\right\}_{i=1}^{\infty}$ of $\left(G, \mathcal{B}_{G}\right)$ valued r.v.'s, we recursively construct the double sequence $\left\{U_{n}^{(i)}\right\}_{i=1, n=0}^{\infty}$ of r.v.'s by ${ }^{5}$

$$
\begin{align*}
U_{n}^{(2 i-1)} & :=U_{n-1}^{(i)} \bullet U_{n-1}^{\left(i+2^{n-1}\right)},  \tag{13}\\
U_{n}^{(2 i)} & :=U_{n-1}^{\left(i+2^{n-1}\right)} \tag{14}
\end{align*}
$$

for each $i \geq 1$ and $n \geq 1$, where $\left\{U_{0}^{(i)}\right\}_{i=1}^{\infty}:=\left\{X_{i}\right\}_{i=1}^{\infty}$. Figure 1 illustrates a diagram of (13) and (14). Note that $U_{n}^{(i)}$ is also $\left(G, \mathcal{B}_{G}\right)$-valued for all $n \geq 0$ and $i \geq 1$.

As a non-stationary source with side information, consider a sequence $\left\{Y_{i}\right\}_{i=1}^{\infty}$ of r.v.'s playing the role of side information for $\left\{X_{i}\right\}_{i=1}^{\infty}$. Suppose that the sequence $\left\{\left(X_{i}, Y_{i}\right)\right\}_{i=1}^{\infty}$ is mutually independent. Define the double sequence $\left\{\mathfrak{S}_{n}^{(i)}\right\}_{i=1, n=0}^{\infty}$ of conditional Shannon entropies by

$$
\begin{equation*}
\mathfrak{H}_{n}^{(i)}:=H\left(U_{n}^{(i)} \mid\left\{U_{n}^{(j)}\right\}_{j=1}^{i-1},\left\{Y_{j}\right\}_{j=1}^{\infty}\right) \tag{15}
\end{equation*}
$$

for each $n \geq 0$ and $i \geq 1$. Roughly speaking, a source polarization theorem explores how the sequence $\left\{\mathfrak{H}_{n}^{(i)}\right\}_{i=1}^{\infty}$ behaves for sufficiently large $n$. It can

[^2]be verified by (13)-(14) that $U_{n}^{(i)}$ is conditionally independent of $\left(\left\{U_{n}^{(j)}\right\}_{j=1}^{m 2^{n}},\left\{Y_{j}\right\}_{j=1}^{m 2^{n}},\left\{Y_{j}\right\}_{j=(m+1) 2^{n}+1}^{\infty}\right)$ relative to $\left(\left\{U_{n}^{(j)}\right\}_{j=m 2^{n}+1}^{i},\left\{Y_{j}\right\}_{j=m 2^{n}+1}^{(m+1) 2^{n}}\right)$, where $m:=\left\lfloor i / 2^{n}\right\rfloor$, and $\lfloor\cdot\rfloor$ stands for the floor function. Namely, it holds that
\[

$$
\begin{equation*}
\mathfrak{H}_{n}^{(i)}=H\left(U_{n}^{(i)} \mid\left\{U_{n}^{(j)}\right\}_{j=m 2^{n}+1}^{i-1},\left\{Y_{j}\right\}_{j=m 2^{n}+1}^{(m+1) 2^{n}}\right) . \tag{16}
\end{equation*}
$$

\]

The next section analyzes the conditional Shannon entropies $\left\{\mathfrak{H}_{n}^{(i)}\right\}_{i=1, n=0}^{\infty}$ for erasure distributions.

## III. Polar Transforms for Erasure Distributions

## A. Erasure Distribution Based on Group Structure

It is well-known that the polar transform for binary erasure channels (BECs) can be easily analyzed [1, Proposition 6] by a certain recursive formula. This convenient probabilistic model was translated from channel to source coding problems by Şaşoğlu [3, Section 3.3.1] by defining an erasure distribution as an analogy to the BEC. In this subsection, from the group theoretic perspective, we introduce a more general erasure distribution than Şaşoğlu's one.

Let $H \triangleleft G$ be a shorthand for a normal subgroup $H$ of $G$. The following definition gives an erasure distribution based on the normal subgroup structure of $G$.

Definition 2 (group-based erasure distribution). Let $X$ be an $\left(G, \mathcal{B}_{G}\right)$-valued r.v., and $Y$ a $\mathcal{Y}$-valued r.v., where

$$
y=\bigcup_{H \triangleleft G: H \text { is finite }} \frac{G}{H}=\{g \bullet H \mid g \in G \text { and finite } H \triangleleft G\} .
$$

Then, we say that $(X, Y)$ follows an erasure distribution if

$$
\begin{equation*}
P_{X \mid Y}(B)=\frac{|B \cap Y|}{|Y|} \quad \text { (a.s.), } \quad \text { for } B \in \mathcal{B}_{G} \tag{17}
\end{equation*}
$$

The alphabet $y$ given in Definition 2 is the collection of finite cosets of $G$, and note that $X \in Y$ a.s. if $(X, Y)$ follows an erasure distribution. Intuitively, this means that the side information $Y$ only tells us that $X$ belongs to a finite coset $Y$. It is clear that $X$ is conditionally discrete relative to $Y$ if $(X, Y)$ follows an erasure distribution, because $P_{X \mid Y}$ is a uniform distribution on a finite coset $Y$ a.s. (see (17)).

If the order of $G$ is two, then Definition 2 coincides with Şaşoğlu's one [3, Section 3.3.1]. Moreover, Definition 2 can be reduced to some know erasure-like channels as follows.

Remark 1. Let $(X, Y)$ a pair of r.v.'s following an erasure distribution in the sense of Definition 2. Suppose that

$$
\begin{equation*}
\mathbb{P}\{Y=g \bullet H \mid Y \in G / H\}=\mathbb{P}\{Y=H \mid Y \in G / H\} \tag{18}
\end{equation*}
$$

for every $g \in G$ and every finite $H \triangleleft G$, provided that $\mathbb{P}\{Y \in G / H\}>0$. If $G$ is a finite cyclic group, then the joint probability measure induced by $(X, Y)$ is equivalent to a modular arithmetic erasure channel [13, Definition 2] with uniform input distribution. As summarized in [13, Examples 14], modular arithmetic erasure channels can be further reduced to many other erasure-like channels given in [15], [36]. In addition, if $G$ is a finite elementary abelian group, then the joint probability measure induced by $(X, Y)$ is equivalent to a
combination of linear channels [13, Definition 23] with uniform input distribution, where the linear channels are constructed on the set of subspaces of a finite-dimensional vector space over the prime field.

Lemma 3 shows a simple formula for $H(X \mid Y)$ for an erasure distribution; it can be directly proven by (1) and Definition 2.
Lemma 3. If $(X, Y)$ follows an erasure distribution, then

$$
\begin{equation*}
H(X \mid Y)=\mathbb{E}[\log |Y|] \tag{19}
\end{equation*}
$$

It is worth pointing out that Lemma 3 is analogous to [10, Proposition 3] and [13, Proposition 2].

## B. Reduction of Polar Transforms to Erasure Distributions

Suppose that $\left(X_{i}, Y_{i}\right)$ follows an erasure distribution for $i \geq 1$. Let $\left\{Y_{n}^{(i)}\right\}_{i=1, n=0}^{\infty}$ be a double sequence of r.v.'s defined by ${ }^{6}$

$$
\begin{align*}
Y_{n}^{(2 i-1)} & :=Y_{n-1}^{(i)} \bullet Y_{n-1}^{\left(i+2^{n-1}\right)},  \tag{20}\\
Y_{n}^{(2 i)} & :=\phi\left(U_{n}^{(2 n-1)}, Y_{n-1}^{(i)}, Y_{n-1}^{\left(i+2^{n-1}\right)}\right) \tag{21}
\end{align*}
$$

for each $i \geq 1$ and $n \geq 1$, where $\left\{Y_{0}^{(i)}\right\}_{i=1}^{\infty}:=\left\{Y_{i}\right\}_{i=1}^{\infty}$, and the mapping $\phi: G \times \mathcal{Y} \times \mathcal{Y} \rightarrow \mathcal{Y} \cup\{\emptyset\}$ is defined by

$$
\begin{equation*}
\phi:(g, a \bullet H, b \bullet K) \mapsto\left(a^{-1} \bullet g \bullet H\right) \cap(b \bullet K) \tag{22}
\end{equation*}
$$

for $a, b, g \in G$ and finite $H, K \triangleleft G$. Theorem 1 shows that like the BEC [1, Proposition 6], the polar transform for erasure distributions generates other erasure distributions again.
Theorem 1. The pair $\left(U_{n}^{(i)}, Y_{n}^{(i)}\right)$ also follows an erasure distribution for each $i \geq 1$ and $n \geq 0$. In addition, for each $n \geq 0$ and $i \geq 1$, and with $m:=\left\lfloor i / 2^{n}\right\rfloor$, it holds that

$$
\begin{equation*}
Y_{n}^{(i)} \equiv_{U_{n}^{(i)}}\left(\left\{U_{n}^{(j)}\right\}_{j=m 2^{n}+1}^{i-1},\left\{Y_{j}\right\}_{j=m 2^{n}+1}^{(m+1) 2^{n}}\right) . \tag{23}
\end{equation*}
$$

Corollary 1. If $\left(X_{i}, Y_{i}\right)$ follows an erasure distribution, then

$$
\begin{equation*}
\mathfrak{H}_{n}^{(i)}=\mathbb{E}\left[\log \left|Y_{n}^{(i)}\right|\right] \quad \text { for } n \geq 0, i \geq 1 \tag{24}
\end{equation*}
$$

Proof of Theorem 1: By Lemma 2 and the recursivity of (13)-(14), it suffices to consider the one-step polar transform (3)-(4). Next, we prove Theorem 1 in two parts:

1) Proof for Minus Transform $\left(U_{1},\left(Y_{1}, Y_{2}\right)\right)$ : Since $X_{1}$ and $X_{2}$ are conditionally discrete relative to $Y_{1}$ and $Y_{2}$, respectively, it follows by the first equality of (12) that

$$
\begin{align*}
P_{U_{1} \mid Y_{1}, Y_{2}}\left(A_{1}\right) & =\sum_{u_{2} \in G} P_{X_{1} \mid Y_{1}}\left(A_{1} \bullet u_{2}^{-1}\right) P_{X_{2} \mid Y_{2}}\left(u_{2}\right) \\
& =\sum_{u_{1} \in A_{1}} \sum_{u_{2} \in G} P_{X_{1} \mid Y_{1}}\left(u_{1} \bullet u_{2}^{-1}\right) P_{X_{2} \mid Y_{2}}\left(u_{2}\right) \tag{25}
\end{align*}
$$

a.s. for every $A_{1} \in \mathcal{B}_{G}$. Thus, we see that $U_{1}$ is conditionally discrete relative to $\left(Y_{1}, Y_{2}\right)$ as well. It follows from (17) that

$$
\begin{equation*}
P_{X_{1} \mid Y_{1}}\left(u_{1} \bullet u_{2}^{-1}\right) P_{X_{2} \mid Y_{2}}\left(u_{2}\right)=\frac{\mathbf{1}_{\left\{u_{1} \bullet u_{2}^{-1} \in Y_{1}\right\} \cap\left\{u_{2} \in Y_{2}\right\}}}{\left|Y_{1}\right|\left|Y_{2}\right|} \tag{26}
\end{equation*}
$$

a.s. for every $\left(u_{1}, u_{2}\right) \in G^{2}$. Now, we readily see that for some $u_{1}, u_{2}, g_{1}, g_{2} \in G$ and $H, K \triangleleft G$, both $u_{1} \bullet u_{2}^{-1} \in g_{1} \bullet H$ and

[^3]$u_{2} \in g_{2} \bullet K$ hold if and only if the following system of two congruences holds:
\[

$$
\begin{cases}u_{2} \equiv g_{1}^{-1} \bullet u_{1} & (\bmod H),  \tag{27}\\ u_{2} \equiv g_{2} & (\bmod K)\end{cases}
$$
\]

By the Chinese Remainder Theorem in group theory, the system (27) has a unique solution $u_{2} \in G$ modulo $H \cap K$ if and only if $u_{1} \equiv g_{1} \bullet g_{2}(\bmod H \bullet K)$. Therefore, we have

$$
\begin{equation*}
\left\{u_{2} \in G \mid\left(u_{1} \bullet u_{2}^{-1}, u_{2}\right) \in Y_{1} \times Y_{2}\right\} \in \frac{G}{H \cap K} \tag{28}
\end{equation*}
$$

if and only if $Y_{1} \in G / H, Y_{2} \in G / K$, and $u_{1} \in Y_{1} \bullet Y_{2}$. Hence, it follows from (25), (26), and (28) that for each $A_{1} \in \mathcal{B}_{G}$,

$$
\begin{align*}
P_{U_{1} \mid Y_{1}, Y_{2}}\left(A_{1}\right) & =\sum_{u_{1} \in A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)} \sum_{u_{2} \in \phi\left(U_{1}, Y_{1}, Y_{2}\right)} \frac{1}{\left|Y_{1}\right|\left|Y_{2}\right|} \\
& =\sum_{u_{1} \in A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)} \frac{\left|\phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|}{\left|Y_{1}\right|\left|Y_{2}\right|} \\
& =\frac{\left|A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)\right|}{\left|Y_{1} \bullet Y_{2}\right|} \tag{29}
\end{align*}
$$

a.s., where the last equality follows by the identity ${ }^{7}$

$$
\begin{equation*}
|H \bullet K||H \cap K|=|H||K| \tag{30}
\end{equation*}
$$

for finite $H, K \triangleleft G$. Equation (29) implies that $\left(U_{1},\left(Y_{1}, Y_{2}\right)\right)$ follows an erasure distribution (see (17)). Furthermore, since $Y_{1} \bullet Y_{2}$ is a function of $\left(Y_{1}, Y_{2}\right)$, we see that $\sigma\left(Y_{1} \bullet Y_{2}\right) \subset \sigma\left(Y_{1}, Y_{2}\right)$. Hence, it can be verified by (11) and (29) that

$$
\begin{equation*}
Y_{1} \bullet Y_{2} \equiv_{U_{1}}\left(Y_{1}, Y_{2}\right) \tag{31}
\end{equation*}
$$

completing the proof for minus transform.
2) Proof for Plus Transform $\left(U_{2},\left(U_{1}, Y_{1}, Y_{2}\right)\right)$ : Since $X_{1}$ and $X_{2}$ are conditionally discrete relative to $Y_{1}$ and $Y_{2}$, respectively, it follows as in Steps (a)-(i) of (10) that for each $A_{1}, A_{2} \in \mathcal{B}_{G}$,

$$
\begin{align*}
& \mathbb{E}\left[\mathbf{1}_{\left\{U_{1} \in A_{1}\right\}} P_{U_{2} \mid U_{1}, Y_{1}, Y_{2}}\left(A_{2}\right) \mid Y_{1}, Y_{2}\right] \\
&=\sum_{u_{1} \in A_{1}} \sum_{u_{2} \in A_{2}} P_{X_{1} \mid Y_{1}}\left(u_{1} \bullet u_{2}^{-1}\right) P_{X_{2} \mid Y_{1}}\left(u_{2}\right) \\
& \stackrel{(\text { a })}{=} \sum_{u_{1} \in A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)} \sum_{u_{2} \in A_{2} \cap \phi\left(U_{1}, Y_{1}, Y_{2}\right)} \frac{1}{\left|Y_{1}\right|\left|Y_{2}\right|} \\
&=\sum_{u_{1} \in A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)} \frac{\left|A_{2} \cap \phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|}{\left|Y_{1}\right|\left|Y_{2}\right|} \\
& \stackrel{\text { (b) }}{=} \sum_{u_{1} \in A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)} \frac{P_{U_{1} \mid Y_{1}, Y_{2}\left(u_{1}\right)}^{P_{U_{1} \mid Y_{1}, Y_{2}}\left(u_{1}\right)} \frac{\left|A_{2} \cap \phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|}{\left|Y_{1}\right|\left|Y_{2}\right|}}{\stackrel{\text { (c) }}{=}} \sum_{u_{1} \in A_{1} \cap\left(Y_{1} \bullet Y_{2}\right)} \\
& \stackrel{P_{U_{1} \mid Y_{1}, Y_{2}}\left(u_{1}\right) \frac{\left|A_{2} \cap \phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|}{\mid=}}{\left|\phi\left[U_{1}, Y_{1}, Y_{2}\right)\right|} \\
&\left.\left.\mathbf{1}_{\left\{U_{1} \in A_{1}\right\}} \frac{\left|A_{2} \cap \phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|}{\left|\phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|} \right\rvert\, Y_{1}, Y_{2}\right] \tag{32}
\end{align*}
$$

a.s., where (a) follows from (26) and (28); (b) follows from the fact that $P_{U_{1} \mid Y_{1}, Y_{2}}\left(u_{1}\right)>0$ if and only if $u_{1} \in Y_{1} \bullet Y_{2}$ (see (29)); (c) follows from (29)-(30); and (d) follows from the

[^4]fact that $P_{U_{1} \mid Y_{1}, Y_{2}}(\cdot)$ forms a uniform distribution on $Y_{1} \bullet Y_{2}$ a.s. Therefore, it holds that for each $A_{2} \in \mathcal{B}_{G}$,
\[

$$
\begin{equation*}
P_{U_{2} \mid U_{1}, Y_{1}, Y_{2}}\left(A_{2}\right)=\frac{\left|A_{2} \cap \phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|}{\left|\phi\left(U_{1}, Y_{1}, Y_{2}\right)\right|} \tag{33}
\end{equation*}
$$

\]

a.s. Equation (33) implies that $\left(U_{2},\left(U_{1}, Y_{1}, Y_{2}\right)\right)$ follows an erasure distribution. Furthermore, since $\phi\left(U_{1}, Y_{1}, Y_{2}\right)$ is a function of $\left(U_{1}, Y_{1}, Y_{2}\right)$, it holds that $\sigma\left(\phi\left(U_{1}, Y_{1}, Y_{2}\right)\right) \subset \sigma\left(U_{1}, Y_{1}, Y_{2}\right)$. Hence, it can be verified by (11) and (33) that

$$
\begin{equation*}
\phi\left(U_{1}, Y_{1}, Y_{2}\right) \equiv_{U_{2}}\left(U_{1}, Y_{1}, Y_{2}\right), \tag{34}
\end{equation*}
$$

completing the proof for plus transform.
Proof of Corollary 1: Corollary 1 is now obvious from Lemmas $1-3$ and Theorem 1.

If the order of $G$ is two, then Theorem 1 can be reduced to the discussion in [3, Section 3.3.1], which is alanogous to the ease of analysing the polar transform for BECs [1, Proposition 6]. As shown in the following remark, Theorem 1 is indeed a generalization of the known formulas in easy case studies of Arrkan-style two-by-two polar transforms.

Remark 2. For a source $\left\{\left(X_{i}, Y_{i}\right)\right\}_{i=1}^{\infty}$, suppose the same hypothesis as Remark 1. If $G$ is a finite cyclic group, then Theorem 1 is a counterpart of the recursive formula for modular arithmetic erasure channels [13, Theorem 1]. In addition, if $G$ is a finite elementary abelian group, then Theorem 1 is a counterpart of the recursive formula for combinations of linear channels [10, Proposition 4] in a stationary setting.

By Corollary 1, the probability measures induced by $\left\{\left|Y_{n}^{(i)}\right|\right\}_{i=1}^{\infty}$ are important to analyze the asymptotic distribution of multilevel polarization for sufficiently large $n$. The next section explores the asymptotic distribution in a special case.

## IV. Multilevel Source Polarization Analysis

A group $G$ is said to be locally cyclic if every finitely generated subgroup of $G$ is cyclic. In this section, we investigate the multilevel polarization theorem for a non-stationary source $\left\{\left(X_{i}, Y_{i}\right)\right\}_{i=1}^{\infty}$, where $\left(X_{i}, Y_{i}\right)$ follows an erasure distribution with a locally cyclic group $G$ for each $i \geq 1$.

## A. Theorem 1 with a Locally Cyclic Group

It is known that every locally cyclic group is isomorphic to a subgroup of the additive rationals $\mathbb{Q}$ or of the additive quotient group $\mathbb{Q} / \mathbb{Z}$. Namely, the order of every locally cyclic group must be countable; and thus, it is worth mentioning that any subset of a locally cyclic Polish group $G$ is Borel.

Fix an index $i \geq 1$. Since every finite subgroup of a locally cyclic group is cyclic, and since every cyclic group of order $k$ is isomorphic to $\mathbb{Z} / k \mathbb{Z}$ under addition, we observe that (i) every locally cyclic group $G$ has at most a countable number of finite normal subgroups $N \triangleleft G$; and (ii) $\left|Y_{i}\right|=k$ if and only if $Y_{i} \in G / N$ with $N \cong \mathbb{Z} / k \mathbb{Z}$. Hence, it follows that $\mathbb{P}\left\{\left|Y_{i}\right|=k\right\}=\mathbb{P}\left\{Y_{i} \in G / N\right.$ for some $N$ isomorphic to $\left.\mathbb{Z} / k \mathbb{Z}\right\}$, and Lemma 3 can be rewritten by

$$
\begin{equation*}
H(X \mid Y)=\sum_{N \triangleleft G: N \text { is finite }}(\log |N|) \mathbb{P}\left\{Y_{i} \in G / N\right\} \tag{35}
\end{equation*}
$$

Namely, the probability $\mathbb{P}\left\{Y_{i} \in G / N\right\}$ is important in the subsequent analysis. Actually, if we define $\varepsilon_{n}^{(i)}(N):=\mathbb{P}\left\{Y_{n}^{(i)} \in G / N\right\}$ for each $i \geq 1, n \geq 0$, and finite $N \triangleleft G$, then it follows from Theorem 1 that for every $i \geq 1, n \geq 1$ and finite $N \triangleleft G$,

$$
\begin{align*}
\varepsilon_{n}^{(2 i-1)}(N) & =\sum_{H, K \triangleleft G: H \bullet K=N} \varepsilon_{n-1}^{(i)}(H) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(K),  \tag{36}\\
\varepsilon_{n}^{(2 i)}(N) & =\sum_{H, K \triangleleft G: H \cap K=N} \varepsilon_{n-1}^{(i)}(H) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(K) . \tag{37}
\end{align*}
$$

Based on the recursive formulas (36)-(37), we observe from Corollary 1 and (35) that for each $i \geq 1$ and $n \geq 0$,

$$
\begin{equation*}
\mathfrak{H}_{n}^{(i)}=\sum_{N \triangleleft G: N \text { is finite }}(\log |N|) \varepsilon_{n}^{(i)}(N) . \tag{38}
\end{equation*}
$$

## B. Asymptotic Distribution of Multilevel Polarization

In the previous subsection, the calculations of $\left\{\mathfrak{G}_{n}^{(i)}\right\}_{i=1, n=0}^{\infty}$ have been simplified to (38) via (36)-(37). Based on this, we now give a multilevel polarization theorem for erasure distributions with a locally cyclic group $G$. To formalize it, we now introduce some notations and definitions as follows:

For each finite $N \triangleleft G$, denote by $\mathcal{S}(N)$ the collection of overgroups $H \triangleright N$ satisfying the following two conditions: (i) there exists a $K \triangleleft G$ satisfying the proper subgroup chain $N \triangleleft K \triangleleft H$ and (ii) there are no distinct $K_{1}, K_{2} \triangleleft G$ satisfying the proper subgroup chain $N \triangleleft K_{1} \triangleleft K_{2} \triangleleft H$. Moreover, for each finite $N \triangleleft G$ and $H \in \mathcal{S}(N)$, denote by $\mathcal{M}(N, H)$ the collection of finite subgroups $K \triangleleft G$ satisfying the proper subgroup chain $N \triangleleft K \triangleleft H$. These notations will be used in Algorithm 1 later.

For each finite $N \triangleleft G$, define

$$
\begin{equation*}
Q(N):=\lim _{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^{m} \mathbb{P}\left\{Y_{i} \in G / N\right\} \tag{39}
\end{equation*}
$$

provided that the limit exists. Henceforth, assume that the limit $Q(N)$ always exists for each finite $N \triangleleft G$, and $Q(N):=0$ if $N$ is infinite. This existence of limits is a similar assumption to [27, Remark 1] in the study of polar codes for non-stationary channels, which can be considered as the existence of entropy rate for a sequence of independent r.v.'s (see also [28]). The following example shows some simple cases of (39).
Example 1. If $\left\{\left(X_{i}, Y_{i}\right)\right\}_{i=1}^{\infty}$ is a stationary source with generic distribution $P_{X, Y}$, i.e., if $P_{X_{i}, Y_{i}}=P_{X, Y}$ for all $i \geq 1$, then it is clear that $Q(H)=\mathbb{P}\{Y \in G / H\}$. Similarly, if $Y_{i}$ converges in distribution to $Y$ as $i \rightarrow \infty$, then it follows from the Cesáro mean that $Q(H)=\mathbb{P}\{Y \in G / H\}$.

It is clear that $Q(\cdot)$ forms a discrete probability distribution on the set of normal subgroups of $G$. As wil bel seen in Theorem 2 and Algorithm 1, the distribution $Q(\cdot)$ plays a significant role to characterize the asymptotic distribution of multilevel source polarization. For each $H, K, N \triangleleft G$, we define

$$
\begin{align*}
\chi(N, K, H) & :=\sum_{J \triangleleft G: J \bullet N=J \bullet K, J \cap H=J \cap K} Q(J),  \tag{40}\\
\beta(N, H) & :=\sum_{J \triangleleft G: J \cap H=J \cap N} Q(J) . \tag{41}
\end{align*}
$$

```
Algorithm 1: Solving \(\mu(N)\) used in (43) of Theorem 2
    Data: A locally cyclic group \(G\); a finite normal subgroup \(N \triangleleft G\);
            and a distribution \(Q(\cdot)\) defined in (39)
    Result: Probability masses \(\{\mu(H)\}_{H \triangleleft N}\)
    \(\alpha \longleftarrow 0 ; K \longleftarrow\{e\}\); and \(\mu(H) \longleftarrow 0\) for all \(H \triangleleft N\)
    while \(K \triangleleft N\) do
        if \(\mathcal{S}(K)\) is nonempty then
            \(\left(H_{1}, H_{2}\right) \longleftarrow \underset{\left(K_{1}, K_{2}\right): K_{1} \in \mathcal{S}(K), K_{2} \in \mathcal{M}\left(K, K_{1}\right)}{\arg \max } \chi\left(K, K_{2}, K_{1}\right)\)
            \(\mu(K) \longleftarrow \beta\left(K, H_{1}\right)-\alpha\)
                if there exists \(H_{3} \in \mathcal{M}\left(K, H_{1}\right)\) s.t. \(H_{3} \neq H_{2}\) then
                    \(\mu(K) \longleftarrow \mu(K)+\chi\left(K, H_{3}, H_{1}\right)\)
                \(K \longleftarrow H_{2}\)
        else if there exists an overgroup \(H \triangleright K\) then
            \(\mu(K) \longleftarrow \beta(K, H)-\alpha\)
                \(K \longleftarrow H\)
        else
            \(\mu(K) \longleftarrow 1-\alpha\)
        \(\alpha \longleftarrow \alpha+\mu(K)\)
```

The definitions (40)-(41) will be also used in Algorithm 1.
Finally, according to (15), we define

$$
\begin{equation*}
\mathfrak{H}_{n}^{(i)}[N]:=H\left(U_{n}^{(i)} \bullet N \mid\left\{U_{n}^{(j)}\right\}_{j=1}^{i-1},\left\{Y_{j}\right\}_{j=1}^{\infty}\right) \tag{42}
\end{equation*}
$$

for each $i \geq 1, n \geq 0$, and $N \triangleleft G$. Since $U_{n}^{(i)} \bullet N$ is a function of $U_{n}^{(i)}$, it is clear that $\mathfrak{H}_{n}^{(i)}[N] \leq \mathfrak{S}_{n}^{(i)}$ for each $N \triangleleft G$. In addition, we readily see that $\mathfrak{H}_{n}^{(i)}[G]=0$ and $\mathfrak{H}_{n}^{(i)}[\{e\}]=\mathfrak{H}_{n}^{(i)}$, where $e \in G$ stands for the identity element of $G$.

The following theorem characterizes the asymptotic distribution of multilevel polarization for a non-stationary source following erasure distributions.
Theorem 2. For any $\delta>0$ and finite $N \triangleleft G$, it holds that

$$
\begin{align*}
& \left.\lim _{n \rightarrow \infty} \lim _{m \rightarrow \infty} \frac{1}{m 2^{n}} \right\rvert\,\left\{1 \leq i \leq m 2^{n}: \mathfrak{H}_{n}^{(i)}[N]<\delta\right. \\
&\left|\mathfrak{G}_{n}^{(i)}-\log \right| N|\mid<\delta\} \mid=\mu(N) \tag{43}
\end{align*}
$$

where $\mu(N)$ can be exactly calculated by Algorithm 1 .
In (43), the limit with respect to $m$ is taken due to the non-stationarity of the source and the dependence of r.v.'s induced by the polar transforms (13)-(14), see Section III-B. Here, note that the number $m$ plays a similar role to that of Section III-B; however, it is not given as $m:=\left\lfloor i / 2^{n}\right\rfloor$ but given as independent of $n$ in Theorem 2.

We shall prove Theorem 2 by employing elementary techniques in lattice theory [30]. Basic notions and definitions in lattice theory can be found in Appendix A.

Proof of Theorem 2: Since the lattice of normal subgroups of a group is distributive if and only if the group is locally cyclic (cf. [29, Theorem 4 of Chapter 3]), it suffices to consider a distributive lattice $(L, \vee, \wedge, \leq)$. Define the double sequence $\left\{\boldsymbol{\varepsilon}_{n}^{(i)}\right\}_{i=1, n=0}^{\infty}$ of probability vectors $\boldsymbol{\varepsilon}_{n}^{(i)}:=\left\{\varepsilon_{n}^{(i)}(j)\right\}_{j \in L}$ by

$$
\begin{equation*}
\varepsilon_{n}^{(2 i-1)}(j):=\sum_{k, l \in L: k \vee l=j} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \tag{44}
\end{equation*}
$$

$$
\begin{equation*}
\varepsilon_{n}^{(2 i)}(j):=\sum_{k, l \in L: k \wedge l=j} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \tag{45}
\end{equation*}
$$

for each $i \geq 1$ and $n \geq 1$, where $\left\{\boldsymbol{\varepsilon}_{0}^{(i)}\right\}_{i=1}^{\infty}:=\left\{\boldsymbol{\varepsilon}_{i}\right\}_{i=1}^{\infty}$ is a given initial probability vector. Note that (44) and (45) correspond to (36) and (37), respectively. To analyze the probability vectors $\boldsymbol{\varepsilon}_{n}^{(i)}$, we further define partial sums of elements in $\varepsilon_{n}^{(i)}$ by

$$
\begin{align*}
\theta_{n}^{(i)}(a, b) & :=\sum_{j \in L: j \vee a=j \vee b} \varepsilon_{n}^{(i)}(j),  \tag{46}\\
\chi_{n}^{(i)}(a, c, b) & :=\sum_{j \in L: j \vee a=j \vee c, j \wedge b=j \wedge c} \varepsilon_{n}^{(i)}(j),  \tag{47}\\
\beta_{n}^{(i)}(a, b) & :=\sum_{j \in L: j \wedge a=j \wedge b} \varepsilon_{n}^{(i)}(j) \tag{48}
\end{align*}
$$

for each $a, b, c \in L$. When elements $a, b \in L$ are clear from the context, we simply write (46)-(48) as $\theta_{n}^{(i)}, \chi_{n}^{(i)}(c)$, and $\beta_{n}^{(i)}$. By defining

$$
\begin{equation*}
\mathcal{M}(a, b):=\{c \in L \mid a<c<b\} \tag{49}
\end{equation*}
$$

for each $a, b \in L$, we can show the following lemma.
Lemma 4. Let $a, b \in L$ be chosen so that $a<b$ and there is no pair $x, y \in L$ satisfying $a<x<y<b$. If $(L, \leq)$ is modular, then it holds that

$$
\begin{equation*}
\theta_{n}^{(i)}+\beta_{n}^{(i)}+\sum_{c \in \mathcal{M}(a, b)} \chi_{n}^{(i)}(c)=1 \tag{50}
\end{equation*}
$$

for every $a, b \in L, i \geq 1$, and $n \geq 0$.
Proof of Lemma 4: See Appendix B.
Note that every distributive lattice is modular. Lemma 4 means that the probability masses of $\boldsymbol{\varepsilon}_{n}^{(i)}$ are well-partitioned by (46)-(48). Let $a$ and $b$ be chosen so that $a<b$ and there is no pair $x, y \in L$ satisfying $a<x<y<b$. If $(L, \leq)$ is modular, then $|\mathcal{M}(a, b)|$ can be an arbitrary nonnegative integer. Particularly, if $(L, \leq)$ is distributive, then it can be verified that $0 \leq|M(a, b)| \leq 2$. Noting this fact, we can observe:

Lemma 5. Let $(L, \leq)$ be a distributive lattice, and let $a, b \in L$ be chosen so that $a<b$. Suppose that there is no pair $x, y \in L$ satisfying $a<x<y<b$. Then, it holds that

$$
\begin{align*}
\theta_{n}^{(2 i-1)} & =\theta_{n-1}^{(i)}+\theta_{n-1}^{\left(i+2^{n-1}\right)}-\theta_{n-1}^{(i)} \theta_{n-1}^{\left(i+2^{n-1}\right)}+C_{n}^{(i)}  \tag{51}\\
\beta_{n}^{(2 i-1)} & =\beta_{n-1}^{(i)} \beta_{n-1}^{\left(i+2^{n-1}\right)},  \tag{52}\\
\theta_{n}^{(2 i)} & =\theta_{n-1}^{(i)} \theta_{n-1}^{\left(i+2^{n-1}\right)},  \tag{53}\\
\beta_{n}^{(2 i)} & =\beta_{n-1}^{(i)}+\beta_{n-1}^{\left(i+2^{n-1}\right)}-\beta_{n-1}^{(i)} \beta_{n-1}^{\left(i+2^{n-1}\right)}+C_{n}^{(i)} \tag{54}
\end{align*}
$$

for every $n \geq 1$ and $i \geq 1$, where

$$
\begin{equation*}
C_{n}^{(i)}:=\sum_{c_{1}, c_{2} \in \mathcal{M}(a, b): c_{1} \neq c_{2}} \chi_{n-1}^{(i)}\left(c_{1}\right) \chi_{n-1}^{\left(i+2^{n-1}\right)}\left(c_{2}\right) \tag{55}
\end{equation*}
$$

Moreover, if there exists an $x \in L$ satisfying $a<x<b$, then it holds that

$$
\begin{align*}
\chi_{n}^{(2 i-1)}(c)= & \chi_{n-1}^{(i)}(c) \chi_{n-1}^{\left(i+2^{n-1}\right)}(c) \\
& +\chi_{n-1}^{(i)}(c) \beta_{n-1}^{\left(i+2^{n-1}\right)}+\beta_{n-1}^{(i)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c) \tag{56}
\end{align*}
$$

$$
\begin{align*}
\chi_{n}^{(2 i)}(c)= & \chi_{n-1}^{(i)}(c) \chi_{n-1}^{\left(i+2^{n-1}\right)}(c) \\
& +\chi_{n-1}^{(i)}(c) \theta_{n-1}^{\left(i+2^{n-1}\right)}+\theta_{n-1}^{(i)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c) \tag{57}
\end{align*}
$$

for every $n \geq 1$, every $i \geq 1$ and every $c \in \mathcal{M}(a, b)$.
Proof of Lemma 5: See Appendix C.
Since the positive divisors of a positive integer form a distributive lattice, Lemma 5 is a generalization of [13, Lemma 6] from a lattice of positive divisors with the stationary source setting to general distributive lattices with the nonstationary source setting. Therefore, as in [13, Lemma 7], we can obtain the following lemma.

Lemma 6. Let $(L, \leq)$ be a distributive lattice, and let $a, b \in L$ be chosen so that $a<b$. Suppose that there is no pair $x, y \in L$ satisfying $a<x<y<b$. Then, it holds that

$$
\begin{align*}
& \theta_{n}^{(2 i-1)}+\theta_{n}^{(2 i)}=\theta_{n-1}^{(i)}+\theta_{n-1}^{\left(i+2^{n-1}\right)}+C_{n}^{(i)}  \tag{58}\\
& \beta_{n}^{(2 i-1)}+\beta_{n}^{(2 i)}=\beta_{n-1}^{(i)}+\beta_{n-1}^{\left(i+2^{n-1}\right)}+C_{n}^{(i)} \tag{59}
\end{align*}
$$

for every $n \geq 1$ and $i \geq 1$. Moreover, if there exists an $x \in L$ satisfying $a<x<b$, then it holds that

$$
\begin{align*}
\chi_{n}^{(2 i-1)}(c)+\chi_{n}^{(2 i)}(c)=\chi_{n-1}^{(i)} & \left(1-\chi_{n-1}^{\left(i+2^{n-1}\right)}\right) \\
& +\chi_{n-1}^{\left(i+2^{n-1}\right)}\left(1-\chi_{n-1}^{(i)}\right) \tag{60}
\end{align*}
$$

Proof of Lemma 6: Lemma 6 is a direct consequence from Lemmas 4 and 5.

Based on this observation, we can prove Theorem 2 in a similar fashion to [13, Section IV].

## V. Concluding Remarks

We have explored the asymptotic distribution of multilevel source polarization over possibly infinite source alphabets by defining a convenient probabilistic model called an erasure distribution, which is defined in Definition 2. The analysis of Arıkan-style two-by-two polar transforms (13)-(14) based on a Polish group was simplified by Theorem 1 establishing recursive formulas of the polar transforms for erasure distributions. When the group is locally cyclic, Theorem 2 and Algorithm 1 give a method for calculating the exact asymptotic distribution of multilevel source polarization for erasure distributions. This is the first instance of multilevel source polarization with countably infinite levels, which is characterized by the structure of distributive lattices.

## A. Simple Instances of Theorem 2

In the following, we mention two examples of Theorem 2.

1) Modular Arithmetic Erasure Channels: As explained in Remarks 1 and 2, the erasure distribution defined in Definition 2 can be reduced to the modular arithmetic erasure channel [13, Definition 2], provided that $G$ is a finite cyclic group. Since every cyclic group is locally cyclic, Theorem 2 can also be reduced to the authors' previous results [13], [14], which is described in a stationary setting.
2) Prüfer p-group: Let $p \geq 2$ be a prime number. The Prüfer p-group $G$ can be defined by the Sylow $p$-subgroup of $\mathbb{Q} / \mathbb{Z}$ up to isomorphism, i.e., $G \simeq\left\{m / p^{n}+\mathbb{Z} \mid m \in \mathbb{Z}\right.$ and $\left.n \in \mathbb{Z}_{\geq 0}\right\}$. It is known that $H \triangleleft K$ or $K \triangleleft H$ for any $H, K \triangleleft G$, provided that $G$ is the Prüfer $p$-group. Thus, Corollary 2 can simplify Theorem 2 without Algorithm 1.
Corollary 2. Suppose that $G$ is the Prüfer p-group. For any $\delta>0$ and finite $N \triangleleft G$, and $Q(\cdot)$ as in (39), it holds that

$$
\begin{align*}
& \left.\lim _{n \rightarrow \infty} \lim _{m \rightarrow \infty} \frac{1}{m 2^{n}} \right\rvert\,\left\{1 \leq i \leq m 2^{n}: \mathfrak{G}_{n}^{(i)}[N]<\delta\right. \\
&\left|\mathfrak{S}_{n}^{(i)}-\log \right| N|\mid<\delta\} \mid=Q(N) \tag{61}
\end{align*}
$$

Therefore, the source polarization for erasure distributions can be simply characterized by the initial condition (39) without any other compulation method like Algorithm 1, provided that $G$ is the Prüfer $p$-group. This gives a simple and concrete instance of countably infinite polarization levels.

## B. Future Works

We have shown a possibility of multilevel polarization phenomena over infinite alphabets. Multilevel polarization analysis for more general source distributions than erasure distributions is an open problem; and inventing practical encoding/decoding schemes with the infinite polarization levels is also of interest in the study of source coding over an infinite source alphabet.

While Theorem 2 can be reduced to the authors' previous results [13], [14] as discussed in Section V-A1, it cannot be reduced to Nasser-Telatar's case study [10, Section VIII], because an elementary abelian group is not locally cyclic in general (cf. Remarks 1 and 2). Generalizing Theorem 2 from locally cyclic to abelian, or not necessarily abelian, groups is highly of interest in terms of the MAC polarization [10], [11].

## Acknowledgements

The authors would like to thank to Dr. Jun Muramatsu for his helpful discussions; Dr. Mine Alsan for her valuable comments greatly improving this paper; and the anonymous reviewers in ISIT'19 for carefully reading this paper and for giving many valuable advices.

## Appendix A <br> Brief Introduction to Lattice Theory

Definition 3 (partially ordered sets; posets). For a binary relation $\leq$ on a nonempty set $L$, the system $(L, \leq)$ is called a poset if it satisfies the following three properties: (i) $a \leq a$; (ii) $a \leq b$ and $b \leq a$ imply that $a=b$; and (iii) $a \leq b$ and $b \leq c$ imply that $a \leq c$, for all $a, b, c \in L$.

Let $(L, \leq)$ be a poset. As a strict relation, the binary relation $a<b$ is a shorthand for $a \leq b$ and $a \neq b$.

Definition 4 (predecessors and followers). For two elements $a, b \in L$ of a poset $(L, \leq)$, we say that $b$ covers $a$, or $a$ is covered by $b$, if $a<b$ and there is no $x \in L$ such that $a<x<b$. This relation is denoted by $a<b$.

For each $a, b \in L$, an upper bound of $a$ and $b$ is an element $u \in L$ satisfying $a \leq u$ and $b \leq u$; and a least upper bound $s$ of $a$ and $b$ is an upper bound of $a$ and $b$ satisfying $s \leq u$ for every upper bound $u$ of $a$ and $b$. If a least upper bound $s$ of $a$ and $b$ exists, then it is unique; and thus, it can be denoted by $a \vee b:=s$, provided that it exists. Analogously, for each $a, b \in L$, a lower bound of $a$ and $b$ is an element $l \in L$ satisfying $l \leq a$ and $l \leq b$; and a greatest lower bound $i$ of $a$ and $b$ is an upper bound of $a$ and $b$ satisfying $l \leq i$ for every lower bound $i$ of $a$ and $b$. If a greatest lower bound $i$ of $a$ and $b$ exists, then it is unique; and thus, it can be denoted by $a \wedge b:=i$, provided that it exists.

Definition 5 (lattices). A poset $(L, \leq)$ is called a lattice if every two elements $a, b \in L$ have the least upper bound $a \vee b$ and the greatest lower bound $a \wedge b$.

Given a lattice $(L, \leq)$, the binary operations $\vee$ and $\wedge$ are called a join and a meet, respectively, and the lattice is sometimes denoted by $(L, \vee, \wedge, \leq)$. These binary operations $\vee$ and $\wedge$ satisfy the following identities:
Lemma 7 ([30, Lemma 1 in page 8]). Let $(L, \leq)$ be a lattice. For every $a, b, c \in L$, it holds that
(i) $a \vee b=b \vee a$ and $a \wedge b=b \wedge a$;
(ii) $a \vee(b \vee c)=(a \vee b) \vee c$ and $a \wedge(b \wedge c)=(a \wedge b) \wedge c$; and
(iii) $a \vee(a \wedge b)=a \wedge(a \vee b)=a$.

We now give the notion of modularity as follows:
Definition 6 (modular lattices; [30, Section 7 of Chapter I]). A lattice $(L, \leq)$ is said to be modular if $a \leq c$ implies that $a \vee(b \wedge c)=(a \vee b) \wedge c$ for every $a, b, c \in L$.

We readily see that Definition 6 is equivalent to the following two identities:

$$
\begin{align*}
& {[(x \wedge y) \vee z] \wedge y=(x \wedge y) \vee(z \wedge y)}  \tag{62}\\
& {[(x \vee y) \wedge z] \vee y=(x \vee y) \wedge(z \vee y)} \tag{63}
\end{align*}
$$

We next give the notion of distributivity as follows:
Definition 7 (distributive lattices; [30, Section 6 of Chapter I]). A lattice $(L, \leq)$ is said to be distributive if

$$
\begin{align*}
& a \vee(b \wedge c)=(a \vee b) \wedge(a \vee c)  \tag{64}\\
& a \wedge(b \vee c)=(a \wedge b) \vee(a \wedge c) \tag{65}
\end{align*}
$$

for every $a, b, c \in L$.
Note that every distributive lattice is modular, but there is a modular lattice which is not distributive. Finally, we give the following lemma under the distributivity.
Lemma 8. Let $(L, \leq)$ be a distributive lattice, and let $a, b \in L$ be chosen so that $a \leq b$. Then, it holds that

$$
\begin{equation*}
j \wedge a=j \wedge b \text { and } k \wedge a=k \wedge b \tag{66}
\end{equation*}
$$

if and only if

$$
\begin{equation*}
(j \vee k) \wedge a=(j \vee k) \wedge b \tag{67}
\end{equation*}
$$

for every $j, k \in L$.
Proof of Lemma 8:

1) If part $\Leftarrow$ : We readily see that

$$
\begin{align*}
j \wedge a & =[(j \vee k) \wedge j] \wedge a \\
& =[(j \vee k) \wedge a] \wedge j \\
& =[(j \vee k) \wedge b] \wedge j \\
& =[(j \vee k) \wedge j] \wedge b \\
& =j \wedge b . \tag{68}
\end{align*}
$$

The identity $k \wedge a=k \wedge b$ can be shown similarly; and these imply the sufficiency, as desired.
2) Only if part $\Rightarrow$ : We readily see that

$$
\begin{align*}
(j \vee k) \wedge a & =(j \wedge a) \vee(k \wedge a) \\
& =(j \wedge b) \vee(k \wedge b) \\
& =(j \vee k) \wedge b, \tag{69}
\end{align*}
$$

which implies the necessity, as desired.
While Lemma 8 is elementary, this lemma is important in proving Theorem 2. Note that Lemma 8 does not hold, provided that a lattice $(L, \leq)$ is modular but not distributive.

## Appendix B

Proof of Lemma 4
Let $a, b \in L$ be chosen so that $a<b$, and let $i \in L$ be an arbitrary element. Since $a \wedge b=a<b=a \vee b$, note that $i \vee a \leq i \vee b$ and $i \wedge a \leq i \wedge b$. If $i \vee a=i \vee b$, then

$$
\begin{align*}
a & =(i \wedge a) \vee a \\
& \leq(i \wedge b) \vee a \\
& \stackrel{(\text { a) }}{=}(i \vee a) \wedge b \\
& =(i \vee b) \wedge b \\
& =b, \tag{70}
\end{align*}
$$

where (a) follows by the modularity (see Definition 6). As $a<b$, this implies that $i \wedge a \neq i \wedge b$; and therefore, since $i \wedge a \leq i \wedge b$, it follows that $i \wedge a<i \wedge b$ if $i \vee a=i \vee b$. Similarly, it can be dually verified that $i \vee a<i \vee b$ if $i \wedge a=i \wedge b$. Therefore, we have

$$
\begin{equation*}
\{j \in L \mid j \vee a=j \vee b\} \cap\{k \in L \mid k \wedge a=k \wedge b\}=\emptyset \tag{71}
\end{equation*}
$$

On the other hand, if $i \vee a<i \vee b$ and $i \wedge a<i \wedge b$, then it holds that $a<(i \wedge b) \vee a<b$, which implies the existence of $x \in L$ satisfying $a<x<b$.

We first consider the case where there is no $x \in L$ satisfying $a<x<b$, i.e., suppose that $a<b$. Then, the set $\mathcal{M}(a, b)$ defined in (49) is empty, and we observe that

$$
\begin{equation*}
\{\{j \in L \mid j \vee a=j \vee b\},\{k \in L \mid k \wedge a=k \wedge b\}\} \tag{72}
\end{equation*}
$$

forms a partition of $L$. Therefore, since $\varepsilon_{n}^{(i)}$ is a probability vector for each $n \geq 0$ and $i \geq 1$, it follows from (46)-(48) that Lemma 4 holds, provided that $a<b$.

We next consider the case where there is at least one $x \in$ $L$ such that $a<x \prec b$. Then, the set $\mathcal{M}(a, b)$ defined in
(49) is nonempty. Moreover, it follows by the modularity of Definition 6 that $a<y<b$ for every $y \in \mathcal{M}(a, b)$. Hence, we observe that neither $c_{1} \leq c_{2}$ nor $c_{2} \leq c_{1}$ for every distinct $c_{1}, c_{2} \in \mathcal{M}(a, b)$. If $i \vee a=i \vee b$, then it is clear that $i \vee c=i \vee b$ for every $c \in \mathcal{M}(a, b)$; and it follows that

$$
\begin{align*}
a & =(i \wedge a) \vee a \\
& \leq(i \wedge c) \vee a \\
& =(i \vee a) \wedge c \\
& =(i \vee c) \wedge c \\
& =c \tag{73}
\end{align*}
$$

for every $c \in \mathcal{M}(a, b)$, which implies that $i \wedge a<i \wedge c$ for every $c \in \mathcal{M}(a, b)$. Similarly, if $i \wedge a=i \wedge b$, then it can be dually verified that $i \vee c<i \vee b$ and $i \wedge c=i \wedge a$ for every $c \in \mathcal{M}(a, b)$. Now, suppose that $i \vee a<i \vee b$ and $i \wedge a<i \wedge b$. As shown in the two paragraphs back, there exists $c \in \mathcal{M}(a, b)$ such that

$$
\begin{equation*}
a<c=(i \wedge b) \vee a<b \tag{74}
\end{equation*}
$$

For such an element $c=(i \wedge b) \vee a=(i \vee a) \wedge b$, it holds that

$$
\begin{align*}
i \wedge c & =i \wedge[(i \wedge b) \vee a] \\
& \stackrel{(\text { a) }}{=}(i \wedge b) \vee(i \wedge a) \\
& =i \wedge b \tag{75}
\end{align*}
$$

and

$$
\begin{align*}
i \vee c & =i \vee[(i \vee a) \wedge b] \\
& \stackrel{(b)}{=}(i \vee a) \wedge(i \vee b) \\
& =i \vee a, \tag{76}
\end{align*}
$$

where (a) and (b) follow from (62) and (63), respectively. For each $c \in \mathcal{M}(a, b)$, define

$$
\begin{equation*}
S(c):=\{x \in L \mid x \vee a=x \vee c \text { and } x \wedge b=x \wedge c\} \tag{77}
\end{equation*}
$$

We now prove by contradiction that $S\left(c_{1}\right) \cap S\left(c_{2}\right)=\emptyset$ for every distinct $c_{1}, c_{2} \in \mathcal{M}(a, b)$. That is, suppose that there exists an element $d \in L$ satisfying $d \in S\left(c_{1}\right) \cap S\left(c_{2}\right)$. Then, we observe that

$$
\begin{align*}
& c_{1} \stackrel{(\mathrm{a})}{=} c_{1} \wedge b \\
& \stackrel{(\mathrm{~b})}{\leq}\left(d \vee c_{1}\right) \wedge b \\
& \stackrel{(\mathrm{c})}{=}\left(d \vee c_{2}\right) \wedge b \\
& \stackrel{(\mathrm{~d})}{=}(d \wedge b) \vee c_{2} \\
& \stackrel{(\mathrm{e})}{=}\left(d \wedge c_{2}\right) \vee c_{2} \\
& \stackrel{(\mathrm{f})}{=} c_{2}, \tag{78}
\end{align*}
$$

where (a) follows from $c_{1} \prec b$; (b) follows from $c_{1} \leq d \vee c_{1}$; (c) follows from $d \in S\left(c_{1}\right) \cap S\left(c_{2}\right)$, i.e., $d \vee c_{1}=d \vee a=d \vee c_{2}$; (d) follows from $c_{2} \leq b$ and the modular equality (6); (e) follows from $d \in S\left(c_{1}\right) \cap S\left(c_{2}\right)$, i.e., $d \wedge c_{1}=d \wedge b=d \wedge c_{2}$; and (f) follows from the absorption law: $(x \wedge y) \vee y=y$. This, however,
contradicts to $c_{1} \not \leq c_{2}$. Therefore, we have $S\left(c_{1}\right) \cap S\left(c_{2}\right)=\emptyset$. Concluding discussions of this paragraph, we observe that

$$
\begin{align*}
\emptyset & =\{j \in L \mid j \vee a=j \vee b\} \cap\{k \in L \mid k \wedge a=k \wedge b\}  \tag{79}\\
& =\{j \in L \mid j \vee a=j \vee b\} \cap S(c)  \tag{80}\\
& =\{k \in L \mid k \wedge a=k \wedge b\} \cap S(c)  \tag{81}\\
& =S\left(c_{1}\right) \cap S\left(c_{2}\right) \tag{82}
\end{align*}
$$

for every $c \in \mathcal{M}(a, b)$ and every distinct $c_{1}, c_{2} \in \mathcal{M}(a, b)$; and

$$
\begin{align*}
& \bigcup_{c \in \mathcal{M}(a, b)} S(c) \\
& =(\{j \in L \mid j \vee a=j \vee b\} \cup\{k \in L \mid k \wedge a=k \wedge b\})^{C}, \tag{83}
\end{align*}
$$

where $\mathcal{A}^{C}$ stands for the complement of a set $\mathcal{A}$. Thus, we have that

$$
\begin{align*}
&\{\{j \in L \mid j \vee a=j \vee b\},\{k \in L \mid k \wedge a=k \wedge b\}\} \\
& \cup\{S(c) \mid c \in \mathcal{M}(a, b)\} \tag{84}
\end{align*}
$$

forms a partition of $L$. Therefore, since $\varepsilon_{n}^{(i)}$ is a probability vector for each $n \geq 0$ and $i \geq 1$, it follows from (46)-(48) that Lemma 4 holds, provided that there exists an $x \in L$ satisfying $a<x<b$. This completes the proof of Lemma 4.

## Appendix C

## Proof of Lemma 5

By the duality between the join $\vee$ and the meet $\wedge$, it suffices to prove the identities of Lemma 5 only for the minus transforms (44). Let $a, b \in L$ be chosen so that $a<b$. A direct calculation shows

$$
\begin{align*}
\beta_{n}^{(2 i-1)} & \stackrel{(\mathrm{a})}{=} \sum_{\substack{j \in L: \\
j \wedge a=j \wedge b}} \varepsilon_{n}^{(2 i-1)}(j) \\
& \stackrel{(\mathrm{b})}{=} \sum_{\substack{j \in L: \\
j \wedge a=j \wedge b,}} \sum_{\substack{k, l \in L: \\
k \vee l=j}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& =\sum_{\substack{k, l \in L:}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& \stackrel{(\mathrm{c})}{=} \sum_{\substack{k \in L: \\
k \wedge a=k \wedge b}} \varepsilon_{n-1}^{(i)}(k) \sum_{\substack{l \in L: \\
l \wedge a=l \wedge b}} \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& =\beta_{n-1}^{(i)} \beta_{n-1}^{\left(i+2^{n-1}\right)} \tag{85}
\end{align*}
$$

for every $n \geq 1$ and $i \geq 1$, where (a) follows from (48); (b) follows from (44); and (c) follows from Lemma 8.

Suppose that there is at least one $x \in L$ such that $a<x<b$, i.e., $b$ covers $x$ and $x$ covers $a$. For each $c \in \mathcal{M}(a, b)$, we have

$$
\begin{aligned}
\chi_{n}^{(2 i-1)}(c) & \stackrel{(\text { a) }}{=} \sum_{\substack{j \in L: \\
j \vee a=j \vee c, j \wedge b=j \wedge c}} \varepsilon_{n}^{(2 i-1)}(j) \\
& \stackrel{(b)}{=} \sum_{\substack{i \in L: \\
i \vee c i \vee a, i \wedge c=i \wedge b}} \sum_{\substack{, l \in L \in L=j}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l)
\end{aligned}
$$

$$
\begin{aligned}
& =\sum_{k, l \in L:} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& (k \vee l) \vee c=(k \vee l) \vee a \text {, } \\
& (k \vee l) \wedge c=(k \vee l) \wedge b \\
& =\sum_{\substack{k, l \in L: \\
(k \vee l) \vee c=(k \vee l) \vee a,(k \vee) \wedge c(k \vee l) \wedge b,}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& k \vee a=k \vee b \text {, } \\
& l \vee a<l \vee b \\
& +\sum_{\substack{k, l \in L: \\
(k \vee) \vee c=(k \vee l) \vee a,(k \vee l) \wedge c=(k \vee l) \wedge b,}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& (k \vee l) \wedge c=(k \vee l) \wedge b, \\
& k \vee a<k \vee b \text {, } \\
& l \vee a=l \vee b \\
& +\sum_{\substack{k, l \in L: \\
(k \vee l) \vee c=(k \vee l) \vee a,(k \vee l) \wedge c=(k \vee l) \wedge b, k \vee a k \vee b, l \vee a<l \vee b}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& \left.\stackrel{(\mathrm{c})}{=} \sum_{\substack{k, l \in L: \\
(k \vee l) \wedge c=(k \vee l) \wedge b, k \vee a=k \vee b, l \vee a<l \vee b}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l)\right) \\
& \stackrel{(\mathrm{d})}{=} \sum_{\substack{k, l \in L: \\
k \wedge c=k \wedge b, l \wedge=l \wedge b, k \vee a=k \vee b,}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& l \vee a<l \vee b \\
& +\sum_{\substack{k, l \in L: \\
k \wedge=k \wedge b, l \wedge c=l \wedge b, k \vee a<k \vee b, l \vee a=l \vee b}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& l \vee a=l \vee b \\
& +\sum_{\substack{k, l \in L: \\
(k \vee l) \vee c=(k \vee l) \vee a, k \wedge c=k \wedge b, l \wedge c=l \wedge b, k \vee a<k \vee b, l \vee a<l \vee b}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
& \left.\stackrel{(\mathrm{e})}{=} \sum_{\substack{k, l \in L: \\
l \wedge a=l \wedge b, k \vee c k \vee b, k \vee c=k \vee a}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l)\right)
\end{aligned}
$$

$$
\begin{gather*}
+\sum_{\substack{k, l \in L: \\
k \wedge c=k \wedge b, l \wedge c=l \wedge b, k \vee c=k \vee a, l \vee c=l \vee a}} \varepsilon_{n-1}^{(i)}(k) \varepsilon_{n-1}^{\left(i+2^{n-1}\right)}(l) \\
=\beta_{n-1}^{(i)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c)+\chi_{n-1}^{(i)}(c) \beta_{n-1}^{\left(i+2^{n-1}\right)} \\
\\
\end{gather*}
$$

for every $n \geq 1$ and $i \geq 1$, where (a) follows from (47); (b) follows from (44); (c) follows from the fact that $j \vee a=j \vee b$ implies $j \vee c=j \vee a$, as shown in the proof of Lemma 4; (d) follows from Lemma 8; and (e) follows from the fact that (i) $k \wedge c=k \wedge b$ implies $k \vee a<k \vee b$ and (ii) $j \wedge c=j \wedge b$ and $j \vee a<j \vee b$ imply $j \vee c=j \vee a$, as shown in the proof of Lemma 4.

Finally, we observe that

$$
\begin{align*}
& \theta_{n}^{(2 i-1)} \stackrel{(\mathrm{a})}{=} 1-\left(\beta_{n}^{(2 i-1)}(a, b)+\sum_{c \in \mathcal{M}(a, b)} \chi_{n}^{(2 i-1)}(c)\right) \\
& \stackrel{(\mathrm{b})}{=} 1-\left(\beta_{n-1}^{(i)} \beta_{n-1}^{\left(i+2^{n-1}\right)}+\sum_{c \in \mathcal{M}(a, b)} \chi_{n}^{(2 i-1)}(c)\right) \\
& \stackrel{(\mathrm{c})}{=} 1-\left(\beta_{n-1}^{(i)} \beta_{n-1}^{\left(i+2^{n-1}\right)}+\sum_{c \in \mathcal{M}(a, b)}\left(\beta_{n-1}^{(i)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c)\right.\right. \\
&\left.\left.+\chi_{n-1}^{(i)}(c) \beta_{n-1}^{\left(i+2^{n-1}\right)}+\chi_{n-1}^{(i)}(c) \chi_{n-1}^{\left(i+2^{n-1}\right)}(c)\right)\right) \\
& \stackrel{(\mathrm{d})}{=} 1-\left(\beta_{n-1}^{(i)} \beta_{n-1}^{\left(i+2^{n-1}\right)}+\beta_{n-1}^{(i)} \sum_{c \in \mathcal{M}(a, b)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c)\right. \\
&+\beta_{n-1}^{\left(i+2^{n-1}\right)} \sum_{c \in \mathcal{M}(a, b)} \chi_{n-1}^{(i)}(c)-C_{n}^{(i)} \\
&\left.+\left(\sum_{c \in \mathcal{M}(a, b)} \chi_{n-1}^{(i)}(c)\right)\left(\sum_{c \in \mathcal{M}(a, b)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c)\right)\right) \\
&=1+C_{n}^{(i)}-\left(\beta_{n-1}^{(i)}+\sum_{c \in \mathcal{M}(a, b)} \chi_{n-1}^{(i)}(c)\right) \\
& \quad \times\left(\beta_{n-1}^{\left(i+2^{n-1}\right)}+\sum_{c \in \mathcal{M}(a, b)} \chi_{n-1}^{\left(i+2^{n-1}\right)}(c)\right) \\
& \stackrel{(\mathrm{e})}{=} 1+C_{n}^{(i)}-\left(1-\theta_{n-1}^{(i)}\right)\left(1-\theta_{n-1}^{\left(i+2^{n-1}\right)}\right) \\
&= \theta_{n-1}^{(i)}+\theta_{n-1}^{\left(i+2^{n-1}\right)}-\theta_{n-1}^{(i)} \theta_{n-1}^{\left(i+2^{n-1}\right)}+C_{n}^{(i)}, \tag{87}
\end{align*}
$$

where (a) follows from Lemma 4; (b) follows from (85); (c) follows from (86); (d) follows from (55); and (e) follows from Lemma 4. This completes the proof of Lemma 5.

## References

[1] E. Arıkan, "Channel polarization: A method for constructing capacityachieving codes for symmetric binary-input memoryless channels," IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051-3073, Jul. 2009.
[2] -_, "Source polarization," in Proc. IEEE Int. Symp. Inf. Theory, Austin, TX, USA, Jun. 2010, pp. 13-18.
[3] E. Şaşoğlu, "Polarization and polar codes," Found. Trends Commun. Inf. Theory, vol. 8, no. 4, pp. 259-381, Oct. 2012.
[4] R. Mori and T. Tanaka, "Source and channel polarization over finite fields and Reed-Solomon matrices," IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2720-2736, May 2014.
[5] R. Nasser, "On the polarization levels of automorphic-symmetric channels," in Proc. IEEE Int. Symp. Inf. Theory, Paris, France, Jul. 2019, [Online]. Available at https://arxiv.org/abs/1811.05203.
[6] J. Błasiok, V. Guruswami, P. Nakkiran, A. Rudra, and M. Sudan, "General strong polarization," Feb. 2018, [Online]. Avaibalbe at https://arxiv.org/ abs/1802.02718.
[7] R. Nasser, "An ergodic theory of binary operations-Part I: Key properties," IEEE Trans. Inf. Theory, vol. 62, no. 12, pp. 6931-6952, Dec. 2016.
[8] _—, "An ergodic theory of binary operations-Part II: Applications to polarization," IEEE Trans. Inf. Theory, vol. 63, no. 2, pp. 1063-1083, Feb. 2017.
[9] _, "Polarization and channel ordering: Characterizations and topological structures," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2017.
[10] R. Nasser and E. Telatar, "Polarization theorems for arbitrary DMCs and arbitrary MACs," IEEE Trans. Inf. Theory, vol. 63, no. 6, pp. 2917-2936, Jun. 2016.
[11] _, "Fourier analysis of MAC polarization," IEEE Trans. Inf. Theory, vol. 63, no. 6, pp. 3600-3620, Jun. 2017.
[12] Y. Sakai and K. Iwata, "A generalized erasure channel in the sense of polarization for binary erasure channels," in Proc. IEEE Inf. Theory Workshop, Cambridge, UK, Sept. 2016.
[13] Y. Sakai, K. Iwata, and H. Fujisaki, "Modular arithmetic erasure channels and their multilevel channel polarization," Apr. 2018, submitted to IEEE Trans. Inf. Theory, [Online]. Available at https://arxiv.org/abs/1804.09016.
[14] ——, "Asymptotic distribution of multilevel channel polarization for a certain class of erasure channels," in Proc. IEEE Int. Symp. Inf. Theory, Vail, CO, USA, Jun. 2018, pp. 856-860.
[15] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," IEEE Trans. Inf. Theory, vol. 59, no. 12, pp. 7839-7857, Dec. 2013.
[16] W. Park and A. Barg, "Polar codes for $q$-ary channels, $q=2^{r}$," IEEE Trans. Inf. Theory, vol. 59, no. 2, pp. 955-969, Feb. 2013.
[17] I. Tal and A. Vardy, "How to construct polar codes," IEEE Trans. Inf. Theory, vol. 59, no. 10, pp. 6562-6582, Oct. 2013.
[18] I. Z. Ruzsa, "Sumsets and entropy," Random Struct. Alg., vol. 34, no. 1, pp. 1-10, 2009.
[19] T. Tao, "Sumset and inverse sumset theory for Shannon entropy," Combin. Probab. Comput., vol. 19, no. 4, pp. 603-639, Jan. 2010.
[20] M. Madiman, A. W. Marcus, and P. Tetali, "Entropy and set cardinality inequalities for partition-determined functions," Random Struct. Alg., vol. 40, no. 4, pp. 399-424, Jul. 2012.
[21] I. Kontoyiannis and M. Madiman, "Sumset and inverse sumset inequalities for differential entropy and mutual information," IEEE Trans. Inf. Theory, vol. 60, no. 8, pp. 4503-4514, Aug. 2014.
[22] V. Guruswami and A. Velingker, "An entropy sumset inequality and polynomially fast convergence to Shannon capacity over all alphabets," in Proc. 30th Conf. Comput. Complex., Portland, Oregon, USA, Jun. 2015, pp. 42-57.
[23] E. Abbe, J. Li, and M. Madiman, "Entropies of weighted sums in cyclic groups and an application to polar codes," Entropy, vol. 19, no. 9, Sept. 2017.
[24] M. Madiman and I. Kontoyiannis, "Entropy bounds on abelian groups and the Ruzsa divergence," IEEE Trans. Inf. Theory, vol. 64, no. 1, pp. 77-92, Jan. 2018.
[25] J. Li and M. Madiman, "A combinatorial approach to small ball inequalities for sums and differences," Combin. Probab. Comput., vol. 28, no. 1, pp. 100-129, Jan. 2019.
[26] T. Tao and V. Vu, Additive Combinatorics. Cambridge, UK: Cambridge University Press, 2006.
[27] M. Alsan and E. Telatar, "A simple proof of polarization and polarization for non-stationary memoryless channels," IEEE Trans. Inf. Theory, vol. 62, no. 9, pp. 4873-4878, Sept. 2016.
[28] H. Mahdavifar, "Polar coding for non-stationary channels," Dec. 2018, [Online]. Available at https://arxiv.org/abs/1611.04203v3.
[29] O. Ore, "Structures and group theory II," Duke Math. J., vol. 4, no. 2, pp. 247-269, 1938.
[30] G. Birkhoff, Lattice Theory, 3rd ed. Providence: American Mathematical Society, 1967.
[31] R. M. Dudley, Real Analysis and Probability. Cambridge, UK: Cambridge University Press, 2002.
[32] R. Durrett, Probability: Theory and Examples, 5th ed. Cambridge, UK: Cambridge University Press, 2019.
[33] Y. Sakai, "Fano-type inequalities based on general conditional information measures over countably infinite alphabets with list decoding," Dec. 2018, submitted to IEEE Trans. Inf. Theory, [Online]. Available at https://arxiv.org/abs/1801.02876v6.
[34] K. L. Chung, A Course in Probability Theory, 3rd ed. San Diego: Academic Press, 2001.
[35] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.
[36] W. Park and A. Barg, "The ordered Hamming metric and ordered symmetric channels," in Proc. IEEE Int. Symp. Inf. Theory, St. Peterburg, Russia, Aug. 2011, pp. 2283-2287.


[^0]:    ${ }^{2}$ In [9, Section 9.2.1], Nasser raised such an open problem aiming to find a method for calculating the exact or approximated asymptotic distribution of multilevel polarizaiton.
    ${ }^{3}$ Groups of order two are unique up to isomorphism.

[^1]:    ${ }^{4}$ The regular conditional distribution $P_{X \mid Y}$ always exists, because $(\mathcal{X}, \mathcal{B})$ is standard Borel (see, e.g., [31, Theorem 10.2.2] or [32, Theorem 4.1.17]).

[^2]:    ${ }^{5}$ The recursive formulas (13)-(14) for binary-input non-starionary channels can be found in, e.g., [27, Section III] and [28, Equation (7)].

[^3]:    ${ }^{6}$ Assume w.l.o.g. that $Y_{n}^{(2 i)}:=\{e\}$ if $\phi\left(U_{n}^{(2 n-1)}, Y_{n-1}^{(i)}, Y_{n-1}^{\left(i+2^{n-1}\right)}\right)=\emptyset$.

[^4]:    ${ }^{7}$ Consider the second isomorphism theorem without the topological structure.

