Ergodic Theory Meets Polarization I:
A Foundation of Polarization Theory

Rajai Nasser
Ecole Polytechnique Fédérale de Lausanne,
Lausanne, Switzerland
Email: rajai.nasser@epfl.ch

Abstract—An open problem in polarization theory is to determine the binary operations that always lead to polarization when they are used in Arıkan style constructions. This paper solves this problem by providing a necessary and sufficient condition for a binary operation to be polarizing. The characterization is given in terms of a new mathematical framework that we introduce. We show that a binary operation is polarizing if and only if its inverse is strongly ergodic.

I. INTRODUCTION

Polar codes are a class of codes invented by Arıkan [1] which achieves the capacity of symmetric binary-input memoryless channels with low encoding and decoding complexities. Arıkan’s construction is based on a basic transformation that is applied recursively. It was shown that by using this transformation, we can convert a set of identical and independent copies of a given single user binary-input channel, into a set of “almost perfect” and “almost useless” channels while preserving the total capacity. This phenomenon is called polarization and it is used to construct capacity-achieving polar codes.

Arıkan’s basic construction uses the XOR operation. Therefore, any attempt to generalize Arıkan’s technique to channels having a non-binary input alphabet $X$, has to replace the XOR operation by a binary operation $\ast$ on the input alphabet $X$. The first operation that was investigated is the addition modulo $q$, where $q = |X|$ and $X$ is endowed with the algebraic structure $\mathbb{Z}_q$. Şaşoğlu et al. [2] showed that if $q$ is prime, then the addition modulo $q$ leads to the same polarization phenomenon as in the binary input case.

Park and Barg [3] showed that if $q = 2^r$ with $r > 0$, then the addition modulo $q$ leads to a polarization phenomenon which is different from the polarization in the binary input case, but it can still be used to construct capacity-achieving polar codes. They showed that we have a multilevel polarization: while we don’t always have polarization to “almost perfect” or “almost useless” channels, we always have polarization to channels which are easy to use for communication. Sahebi and Pradhan [4] showed that multilevel polarization also happens if any Abelian group operation on the alphabet $X$ is used. This allows the construction of polar codes for arbitrary discrete memoryless channels (DMC) since any alphabet can be endowed with an Abelian group structure.

Polar codes for arbitrary DMCs were also constructed by Şaşoğlu [5] by using a special quasigroup operation that ensures two-level polarization. The author and Telatar showed in [6] that all quasigroup operations are polarizing (in the general multilevel sense) and can be used to construct capacity-achieving polar codes for arbitrary DMCs [7].

This paper determines all the polarizing operations by providing a necessary and sufficient condition for a binary operation to be polarizing.

II. PRELIMINARIES

All the sets that are considered in this paper are finite.

A. Easy channels

Notation 1. The symmetric capacity of a channel $W : X \rightarrow Y$, denoted $I(W)$, is the mutual information $I(X; Y)$, where $X$ is uniform in $X$ and $Y$ is the output of $W$ when $X$ is the input.

Definition 1. A channel $W : X \rightarrow Y$ is said to be $\delta$-easy if and only if there exists an integer $L \leq |X|$ and a random code $B$ of block length $L$ and rate $\log L$ (i.e., $B \in \mathcal{S} := \{C \subset X : |C| = L\}$), which satisfy the following:

- $|I(W) - \log L| < \delta$.
- If $C \in \mathcal{S}$ is chosen according to the distribution of $B$ and $X$ is chosen uniformly in $C$, then the marginal distribution of $X$ as a random variable in $X$ is uniform.
- If for each $C \in \mathcal{S}$ we fix a bijection $f_C : \{1, \ldots, L\} \rightarrow C$, then $I(W_B) > \log L - \delta$, where $W_B : \{1, \ldots, L\} \rightarrow Y \times S$ is the channel defined by:

$$W_B(y, C|a) = W(y|f_C(a)) \mathbb{P}_B(C).$$

Note that the value of $I(W_B)$ does not depend on the choice of the bijections $(f_C)_{C \in \mathcal{S}}$.

If $W$ is $\delta$-easy for a small $\delta$, then we can reliably transmit information near the symmetric capacity of $W$ using a code of blocklength $1$ (hence the easiness; there is no need to use codes of large blocklengths): we choose a random code according to $B$, we reveal this code to the receiver, and then we transmit information using this code. The rate of this code is equal to $\log L$ which is close to the symmetric capacity $I(W)$. On the other hand, the fact that $I(W_B) > \log L - \delta$ means that $W_B$ is almost perfect, which ensures that our simple coding scheme has a low probability of error.
B. Polarization process

**Definition 2.** Let $W : X \rightarrow Y$ be a channel and let $*$ be a binary operation on $X$. We define the two channels $W^- : X \rightarrow Y \times Y$ and $W^+ : X \rightarrow Y \times Y \times X$ as follows:

$$W^-(y_1, y_2, u_1) = \frac{1}{|X|} \sum_{u_2 \in X} W(y_1 | u_1 \ast u_2) W(y_2 | u_2),$$

$$W^+(y_1, y_2, u_1 | u_2) = \frac{1}{|X|} W(y_1 | u_1 \ast u_2) W(y_2 | u_2).$$

For every $s = (s_1, \ldots, s_n) \in \{-, +\}^n$, we define $W^s$ recursively as $W^s := ((W^{s_1})^{s_2} \ldots)^{s_n}$. 

**Definition 3.** Let $(B_n)_{n \geq 1}$ be i.i.d. uniform random variables in $\{-, +\}$. For each channel $W$ of input alphabet $X$, we define the channel-valued process $(W_n)_{n \geq 0}$ as follows:

$$W_0 := W,$$

$$W_{n+1} := W_{n} B_n \quad \forall n \geq 1.$$ 

**Definition 4.** A binary operation $*$ is said to be polarizing if and only if for every channel $W$ of input alphabet $X$, we have

- $I(W^-) + I(W^+) = 2I(W)$ (conservation property).
- For every $\delta > 0$, $W_n$ almost surely becomes $\delta$-easy, i.e.,

$$\lim_{n \to \infty} P(W_n \text{ is } \delta-\text{easy}) = 1.$$

### III. Ergodic Theory of Binary Operations

**A. Uniformity preserving and quasigroup operations**

**Definition 5.** A uniformity preserving operation $*$ on $X$ is a binary operation for which the mapping $f_* : X^2 \rightarrow X^2$ defined by $f_*(u_1, u_2) = (u_1 \ast u_2, u_1)$ is bijective. It is called uniformity preserving since $(U_1 \ast U_2, U_1)$ is uniform in $X^2$ if and only if $(U_1, U_2)$ is uniform in $X^2$.

**Remark 1.** A polarizing operation $*$ on $X$ must be uniformity preserving: if $*$ were not uniformity preserving, then $*$ would not have the conservation property of Definition 4. On the other hand, if $*$ is uniformity preserving, then $I(W^-) + I(W^+) = 2I(W)$ and $I(W^-) \leq I(W) \leq I(W^+)$. 

Because of Remark 1, we will only consider uniformity preserving operations in the rest of this paper.

**Remark 2.** It is easy to see that $*$ is uniformity preserving if and only if it satisfies the following condition:

- For any two elements $a, b \in X$, there exists a unique element $c \in X$ such that $a = c \ast b$. We denote this element $c$ by $a/\ast b$. We call $/\ast$ the inverse of $\ast$.

**Definition 6.** A uniformity preserving operation is said to be a quasigroup operation if it also satisfies the following:

- For any two elements $a, b \in X$, there exists a unique element $c \in X$ such that $a = b \ast c$.

**Notation 2.** Let $A$ and $B$ be two subsets of $X$. We define the set $A \ast B := \{a \ast b : a \in A, b \in B\}$. We simply denote $\{a\} \ast B$ and $A \ast \{b\}$ by $a \ast B$ and $A \ast b$, respectively.

If $*$ is uniformity preserving and $B \neq \emptyset$, then $|A \ast B| \geq |A|$.

**B. Irreducible and ergodic operations**

Remark 1 showed that being uniformity preserving is a necessary condition to be polarizing. On the other hand, [6] showed that being a quasigroup operation is a sufficient condition. Therefore, a necessary and sufficient condition must be a property that is stronger than uniformity preserving and weaker than quasigroup. A reasonable strategy to search for a necessary and sufficient condition is to relax the quasigroup property while keeping the uniformity preserving property.

The difference between a quasigroup operation and a uniformity preserving operation is that in the case of a quasigroup operation, any element is reachable from any other element by one multiplication on the right. This property does not always hold for a uniformity preserving operation.

One possible relaxation of the quasigroup property is to consider uniformity preserving operations where all the elements are reachable from each other by multiple multiplications on the right. The reason why we consider such binary operations is because of their good connectability properties: if the elements of $X$ are well connected under $*$, this will create strong correlations between the inputs of the synthetic channels $W^s$ which should ultimately lead to a polarization phenomenon.

**Definition 7.** Let $*$ be a uniformity preserving operation on a set $X$. We say that $a \in X$ is $*$-connectable to $b \in X$ in $l$ steps, and we write $a \xrightarrow{s}^l b$, if there exist $l$ elements $x_0, \ldots, x_{l-1}$ of $X$ such that $(\ldots ((a \ast x_0) \ast x_1) \ldots \ast x_{l-1}) = b$. We say that $a$ is $*$-connectable to $b$, and we write $a \xrightarrow{s}^l b$, if there exists $l > 0$ such that $a \xrightarrow{s}^l b$.

**Definition 8.** A uniformity preserving operation $*$ is said to be irreducible if all the elements of $X$ are $*$-connectable to each other. Define the period of an element $a \in X$ as $\per(a) := \gcd\{l > 0 : a \xrightarrow{s}^l a\}$. One can show that if $*$ is irreducible then $\per(a, a)$ is the same for all $a \in X$ (see [8]). We denote the common value of $\per(a, a)$ as $\per(*)$.

If there exists $l > 0$ such that all the elements of $X$ are $*$-connectable to each other in exactly $l$ steps, we say that the binary operation $*$ is ergodic. In this case, we call the minimum integer $l > 0$ which satisfies this property the connectability of the operation $*$, and we denote it by $\con(*)$.

**Remark 3.** In order to justify our choice of terminology in the previous definition, consider a sequence $(X_n)_{n \geq 0}$ of independent and uniformly distributed random variables in $X$. Define $(X_n)_{n \geq 0}$ recursively as follows: $X_0 = X_0'$ and $X_n = X_{n-1} \ast X_n'$ for $n > 1$. It is easy to see that $(X_n)_{n \geq 0}$ is a stationary Markov chain. We have the following:

- $*$ is irreducible if and only if $(X_n)_{n \geq 0}$ is irreducible.
- $*$ is ergodic if and only if $(X_n)_{n \geq 0}$ is ergodic.

**Proposition 1.** We have the following:

- Every quasigroup operation is ergodic, and every ergodic operation is irreducible.
- If $*$ is uniformity preserving but not irreducible, there exist two disjoint non-empty subsets $A_1$ and $A_2$ of $X$ such that $A_1 \cup A_2 = X$, $A_1 \ast X = A_1$ and $A_2 \ast X = A_2$. 
If $*$ is irreducible, there exists a partition $E$ of $X$ containing $n = \per(*)$ subsets $H_0, \ldots, H_{n-1}$ such that $H_i \ast X = H_{i+1 \mod n}$ for all $0 \leq i < n$. Moreover, we have $|H_0| = \ldots = |H_{n-1}|$.

If $*$ is irreducible, $\per(*) = 1$ if and only if $*$ is ergodic.

If $*$ is ergodic, all the elements of $X$ are $*$-connectable to each other in $s$-steps for all $s \geq \con(*)$.

If $*$ is ergodic, $\con(*) = 1$ if and only if $*$ is a quasigroup operation.

If $*$ is irreducible (resp. ergodic), then $/*$ is irreducible (resp. ergodic) as well.

**Proof:** See [8].

Although ergodic operations seem to have good connectability properties, this is not enough to ensure polarization as we will see in Section IV. It turns out that we need a stronger notion of ergodicity. But in order to define this stronger notion of ergodicity, we first need to define stable partitions.

**C. Balanced and stable partitions**

**Definition 9.** A partition $H$ of a set $X$ is said to be a balanced partition if and only if all the elements of $H$ have the same size. We denote the common size of its elements by $|H|$. The number of elements in $H$ is denoted by $|H|$. Clearly, $|X| = |H| \cdot |H|$ for a partition $H$.

**Notation 3.** Let $A$ and $B$ be two sets of subsets of $X$. We define $A \ast B := \{A \ast B : A \in A, B \in B\}$.

**Definition 10.** Let $H$ be a set of subsets of $X$, and let $*$ be a unimodular preserving operation on $X$. We define the set $H^* = H \ast H = \{A \ast B : A, B \in H\}$, and we define the sequence $(H^*)_{n \geq 0}$ recursively as follows:

- $H^0 = H$.
- $H^n = (H^{n-1})^* = H^{(n-1)*} \ast H^{(n-1)*}$ for all $n > 0$.

**Definition 11.** A partition $H$ of $X$ is said to be a periodic partition of $(X, *)$ if there exists $n > 0$ such that $H^n = H$. The minimum integer $n > 0$ satisfying $H^n = H$ is called the period of $H$, and it is denoted by $\con(H)$.

A partition $H$ of $X$ is said to be a stable partition of $(X, *)$ if and only if $H$ is both balanced and periodic.

**Example 1.** Let $Q = \mathbb{Z}_n \times \mathbb{Z}_n$ and define $(x_1, y_1) \ast (x_2, y_2) = (x_1 + y_1 + x_2 + y_2, y_1 + y_2)$ which is a quasigroup operation. For each $i, j \in \mathbb{Z}_n$, define $H_{i,j} = \{(j + ik, k) : k \in \mathbb{Z}_n\}$. Let $H_i = \{H_{i,j} : j \in \mathbb{Z}_n\}$ for $i \in \mathbb{Z}_n$. It is easy to see that $H_{i,0} = H_{i, \mod n}$ and so $H_0$ is a stable partition of period $n$.

**Proposition 2.** If $H$ is a stable partition of $(X, *)$, then for every $n > 0$, $H^n$ is a stable partition satisfying $\per(H^n) = \per(H)$ and $||H^n|| = ||H||$.

**Proof:** See [8].

The following proposition shows that the concept of stable partitions generalizes the concept of quotient groups:

**Proposition 3.** Let $(G, \ast)$ be a finite group, and let $H$ be a stable partition of $(G, \ast)$. There exists a normal subgroup $H$ of $G$ such that $H$ is the quotient group of $G$ by $H$.

**Definition 12.** A stable partition $H$ is said to be a sub-stable partition of another stable partition $H_2$ if for any $H_1 \in H_2$, there exists $H_2 \in H_2$ such that $H_1 \subset H_2$ (in such case, we clearly have that $||H_1||$ divides $||H_2||$). We denote this relation by $H_1 \preceq H_2$, and we say that $H_1$ is finer than $H_2$.

**D. Strong ergodicity**

Let $H$ be a stable partition. Let $H \in H$ and $x \in X$. For any sequence $(x_n)_{n \geq 0}$ satisfying $x_n \in H^n$ for all $n \geq 0$, define the sequence $(A_n)_{n \geq 0}$ and $(\con(n))_{n \geq 0}$ as follows:

- $A_0 = \{x\}$ and $H_0 = H$.
- $A_n = A_{n-1} \ast X_{n-1} = \{x \ast x_n \ast x_1 \ast \ldots \ast x_{n-1}\}$.
- $H_n = H_{n-1} \ast X_{n-1} = \{x \ast H_{n-1} \ast x_1 \ast \ldots \ast x_{n-1}\}$.

Since $x \in H$, we have $A_n \subset H \subset H^n$ and so $|A_n| \leq |H_n| = ||H^n|| = ||H||$ for all $n \geq 0$. Therefore, $|H_n|$ is constant. On the other hand, $|A_n| \geq |A_{n-1}|$ since $A_n = A_{n-1} \ast X_{n-1}$. Hence, $|A_n|$ is increasing and it is upper bounded by $||H||$.

Does $|A_n|$ reach $||H||$ or does $|A_n|$ remain strictly less than $||H||$ for all $n \geq 0$? In other words, do we have $A_n = H_n$ for some $n > 0$ or do $A_n$ remain a strict subset of $H_n$ for all $n \geq 0$? The answer depends on a stratified course on the sequence $(x_n)_{n \geq 0}$, so one can ask: is it possible to choose at least one sequence $(x_n)_{n \geq 0}$ for which $A_n = H_n$ for some $n > 0$?

What are the stable partitions $H$ for which it is always possible to reach a set in $H^n$ for some $n > 0$ starting from an arbitrary singleton in $X$ and then recursively multiplying on the right by sets chosen from $\mathcal{H}^k$ ($0 \leq k < n$)?

It is easy to see that for the trivial stable partition $H = \{X\}$, the above condition is equivalent to ergodicity. Therefore, satisfying the above condition for every stable partition is a stronger notion of ergodicity. Strong ergodicity turns out to be important for polarization theory as we will see in Section IV.

In this section, we formally define strong ergodicity.

**Notation 4.** Let $\mathcal{X} = (X_i)_{0 \leq i < k}$ be a sequence of subsets of $X$. For any $A \subset X$, we denote $(\cdots ((A \ast X_0) \ast X_1) \cdots \ast X_{k-1})$ by $A \ast \mathcal{X}$. If $A = \{a\}$, we write $a \ast \mathcal{X}$ to denote $\{a\} \ast \mathcal{X}$.

**Definition 13.** Let $H$ be a stable partition of $(X, \ast)$ where $\ast$ is uniformly preserving. A sequence $\mathcal{X} = (X_i)_{0 \leq i < k}$ is said to be $H$-sequence if $X_0 \in H$, $X_1 \in H^*$, ..., $X_{k-1} \in H^{(k-1)*}$. An $H$-sequence $\mathcal{X} = (X_i)_{0 \leq i < k}$ is said to be $H$-augmenting if $\per(H)$ divides $k$ and $A \subset A \ast \mathcal{X}$ for every $A \subset X$.

**Theorem 1.** Let $H$ be a stable partition of $(X, \ast)$ where $\ast$ is ergodic. There exists a unique sub-stable partition $K_H$ of $H$ such that:

- For every $K \in K_H$ and every $H$-sequence $\mathcal{X}$, we have $K \in K_H$.
- For every $K \in K_H$ and every $x \in K$, there exists an $H$-augmenting sequence $\mathcal{X}$ such that $x \in \mathcal{X} = K$.
- For every $K \in K_H$, every $x \in K$, and every $H$-augmenting sequence $\mathcal{X}$, $x \ast X \in K$.

$K_H$ is called the first residue of the stable partition $H$. We also have $K_H^\ast = K_{H^\ast}$, for all $n \geq 0$. 

Definition 14. A uniformity preserving operation $*$ is said to be strongly ergodic if for any stable partition $\mathcal{H}$ and for any $x \in X$, there exists an integer $n = n(x, \mathcal{H})$ such that for any $H \in \mathcal{H}^\ast$, there exists an $\mathcal{H}$-sequence $X_{x,H}$ of length $n$ such that $x \equiv x_{x,H} \equiv H$.

Theorem 2. We have the following:

1) Every strongly ergodic operation is ergodic.
2) If $*$ is strongly ergodic, there exists an integer $d \geq 0$ such that for any $s \geq d$, any stable partition $\mathcal{H}$, any $x \in X$ and any $H \in \mathcal{H}^\ast$, there exists an $\mathcal{H}$-sequence $X_{x,H}$ of length $s$ satisfying $x + X_{x,H} = H$. If $d$ is minimal with this property, we call it the strong connectability of $*$, and we denote it by $\text{socn}(*)$.
3) If $*$ is ergodic then $*$ is strongly ergodic if and only if $\kappa_{\mathcal{H}} = \mathcal{H}$ for every stable partition $\mathcal{H}$.
4) Every quasigroup operation is strongly ergodic.

Proof: See [8].

E. Generated stable partitions

Definition 15. Let $A$ and $B$ be two sets of subsets of $X$. We say that $A$ is finer than $B$ (or $B$ is coarser than $A$) if and only if for every $A \in A$ there exists $B \in B$ such that $A \subseteq B$. We write $A \trianglelefteq B$ to denote the relation “$A$ is finer than $B$.”

Let $A$ be a set of subsets of $X$. It is possible to find a stable partition of $(X, *)$ which is coarser than $A$ and finer than every other stable partition that is coarser than $A$. The following shows that the answer to this question is affirmative if $*$ is ergodic.

Proposition 4. Let $*$ be an ergodic operation on $X$, and let $A$ be a set of subsets of $X$. There exists a unique stable partition $\langle A \rangle$ which satisfies the following:

- $A \trianglelefteq \langle A \rangle$.
- For every stable partition $\mathcal{H}$, if $A \leq \mathcal{H}$ then $\langle A \rangle \leq \mathcal{H}$.

$\langle A \rangle$ is called the stable partition generated by $A$. We also have $\langle A \rangle^\ast = \langle A^\ast \rangle$ for every $n \geq 0$.

Proof: See [8].

Let $A$ be a set of subsets of $X$ which covers $X$ and does not contain the empty set. We have $\langle A \rangle$ which implies that $A^\ast \geq \langle A \rangle$ for every $n \geq 0$. Can we find $n > 0$ for which $A^\ast = \langle A \rangle^\ast$? The following shows that the answer to this question is affirmative if $*$ is strongly ergodic.

Definition 16. Let $A$ be a set of subsets of $X$. We say that $A$ is an $X$-cover if $\varnothing \notin A$ and $X = \bigcup_{A \in A} A$.

Theorem 3. Let $*$ be a strongly ergodic operation on $X$. For every $X$-cover $A$, there exists $n < 2^{|\mathcal{X}|}$ such that $A^\ast = \langle A \rangle$ and $\text{per}(\langle A \rangle)$ divides $n$, i.e., $A^\ast = \langle A \rangle = \langle A \rangle^\ast$.

Proof: See [8].
The rest of this section is dedicated to show that the strong ergodicity of */∗ is a sufficient condition for polarization.

**Definition 17.** Let ᨤ be a balanced partition of ᨧ and let Ṿ : ᨧ → ᨨ. We define the channel Ṿ|Ꮟ : ᨤ → ᨨ as:

\[ Ṿ|Ꮟ(y|H) = \frac{1}{|||H|||} \sum_{x \in H} Ṿ(y|x). \]

**Theorem 4.** Let ᨧ be an arbitrary set and let ∗ be a uniformity preserving operation on ᨧ such that */∗ is strongly ergodic. Let Ṿ : ᨧ → ᨨ be an arbitrary channel. Then for any δ > 0, we have:

\[ \lim_{n \to \infty} \frac{1}{2^n} \left\{ s \in \{-, +\}^n : \exists \mathcal{H}_s a stable partition of (XHR, */∗), \right. \]

\[ \left. |I(ℐ^{*}|\mathcal{H}_s)| - \log |\mathcal{H}_s| < \delta, \quad |I(ℐ∗) - \log |\mathcal{H}_s| < \delta \right\} = 1. \]

**Proof:** We only provide a sketch of the proof. Details can be found in [9].

Remark 1 shows that the process (ℐ(ℐn))n≥0 of Definition 3 is a martingale. Therefore, it converges almost surely and so \( |I(ℐ^{n+1}) - I(ℐn)| \) converges almost surely to zero, where \( k = 2^{2^n\epsilon} + \text{scon}(*) \). We can deduce from this that

\[ \forall \epsilon > 0, \lim_{n \to \infty} \mathbb{P}[|I(ℐ^{n}|) - I(ℐn)| < \epsilon] = 1, \]

where \( |k| = \{-, +\}^k \) is the sequence containing \( k \) minus signs (see [9]). Let ℐ∗ : ℬ → ℬ be a realization of ℐn satisfying \( |I(ℐ^{n}, |k|) - I(ℐn)| < \epsilon \). Let ℬ be a uniform random variable in ℬ and let ℬ be the output of the channel ℐ∗ when ℬ is the input. Fix γ > 0 and define:

- For each \( y \in ℬ \), let ℋy,γ = \{x ∈ ℬ : ℘X|Y(x|y) ≥ γ\}.
- For each \( D \subset ℬ \), let ℬD,γ = \{y ∈ ℬ : ℘D|Y(y|D) ≥ γ\}.

It can be shown that by choosing \( \epsilon = \epsilon(γ) \) to be small enough, \( |I(ℐ^{n}, |k|) - I(ℐn)| < \epsilon \) ensures that ℋs is an ℬ-cover (see [9] for details). Theorem 3 now implies the existence of \( m < 2^{2^n\epsilon} \) such that ℋm,γ = ℋs.

The fact that \( |I(ℐ^{n}, |k|) - I(ℐn)| < \epsilon(γ) \) is small and the fact that \( k - m > \text{scon}(∗) \) can be used to deduce that ℋs = ℋm,γ = ℋs (see [9] for details). This implies that ℋs must be a stable partition of (XHR, */∗).

Moreover, we can show that if \( \epsilon > 0 \) is small enough then

\[ \mathbb{P}_Y\left\{ y \in ℬ : \exists ℋ \in ℋs, |||X|Y=y - ℋ||| \leq \gamma \right\} > 1 - \gamma, \]

where \( ℋ \) is the probability distribution on ℬ that is uniform on ℬ and zero everywhere else. This essentially means that the output ℬ can reliably determine (i.e., with high probability) the unique ℬ ∈ ℋs containing ℬ. On the other hand, we cannot reliably determine from ℬ any other information about ℬ since, conditioned on ℬ, ℬ is almost uniform in ℬ. This means that both \( |I(ℐ^{n}, ℋs)| - \log |ℋs| \) and \( |I(ℐn)| - \log |ℋs| \) are small (see [9] for details).

**Lemma 1.** Let Ṿ : ℬ → ℬ be an arbitrary channel. If there exists a balanced partition ℋ of ℬ such that \( |I(ℐ) - \log |ℋ|| < δ \) and \( |I(ℐ|ℋ)| - \log |ℋ|| < \delta \), then Ṿ is δ-easy.

**Proof:** Let \( L = |ℋ| \) and let \( ℬ1, \ldots , ℬ_L \) be the L members of ℋ. For each \( 1 \leq i \leq L \), let \( ℬ_i \) be a random variable uniformly distributed in \( ℬ_i \). Define \( ℬ = \{ ℬ1, \ldots , ℬ_L \} \), which is a random set taking values in \( ℬ = \{ ℬ \subset ℬ : |邾| = L \} \). It can be shown that ℬ satisfies the conditions of Definition 1 (See [9] for details). Therefore, Ṿ is δ-easy.

**Theorem 5.** A uniformity preserving operation ∗ is polarizing if and only if */∗ is strongly ergodic.

**Proof:** The theorem follows from Proposition 5, Theorem 4 and Lemma 1.

**V. DISCUSSION AND CONCLUSION**

An ergodic theory for binary operations was developed and applied to provide a foundation for polarization theory. We showed that a uniformity preserving operation ∗ is polarizing if and only if */∗ is strongly ergodic. A natural question to ask is whether the strong ergodicity of */∗ implies the strong ergodicity of ∗. It is easy to see that ∗ is ergodic (resp. irreducible, quasigroup operation) if and only if */∗ is ergodic (resp. irreducible, quasigroup operation). We don’t know whether the same is true for strong ergodicity.

The potential applications of the ergodic theory of binary operations might extend beyond polarization theory: the mathematical framework that is developed here is fairly general and might be useful in areas outside polarization and information theory.

**ACKNOWLEDGMENT**

I would like to thank Emre Telatar for enlightening discussions and for his helpful feedback on the paper.

**REFERENCES**


