Continuity of Channel Parameters and Operations under Various DMC Topologies

Rajai Nasser
EPFL, Lausanne, Switzerland
Email: rajai.nasser@epfl.ch

Abstract—We study the continuity of several channel parameters and operations under various topologies on the space of equivalent discrete memoryless channels (DMC). We show that mutual information, channel capacity, Bhattacharyya parameter, probability of error of a fixed code, and optimal probability of error for a given code rate and blocklength, are continuous under various DMC topologies. We also show that channel operations such as sums, products, interpolations, and Arikan-style transformations are continuous.

I. INTRODUCTION

Let \( \mathcal{X} \) and \( \mathcal{Y} \) be two finite sets and let \( W \) be a fixed channel with input alphabet \( \mathcal{X} \) and output alphabet \( \mathcal{Y} \). It is well known that the input-output mutual information is continuous on the simplex of input probability distributions. Many other parameters that depend on the input probability distribution were shown to be continuous on the simplex in [1].

If \( \mathcal{X} \) and \( \mathcal{Y} \) are finite sets, the space of channels with input alphabet \( \mathcal{X} \) and output alphabet \( \mathcal{Y} \) can be naturally endowed with the topology of the Euclidean metric, or any other equivalent metric. It is well known that the channel capacity is continuous in this topology. If \( \mathcal{X} \) and \( \mathcal{Y} \) are arbitrary, one can construct a topology on the space of channels using the weak-* topology on the output alphabet. It was shown in [2] that the capacity is lower semi-continuous in this topology.

The continuity results that are mentioned in the previous paragraph do not take into account “equivalence” between channels. Two channels are said to be equivalent if they are degraded from each other. This means that each channel can be simulated from the other by local operations at the receiver. Two channels that are degraded from each other are completely equivalent from an operational point of view: both channels can be simulated from the other by local operations at the receiver.

In [3], equivalent binary-input channels were identified with their \( L \)-density (i.e., the density of log-likelihood ratios). The space of equivalent binary-input channels was endowed with the topology of convergence in distribution of \( L \)-densities. Since the symmetric capacity\(^1\) and the Bhattacharyya parameter can be written as an integral of a continuous function with respect to the \( L \)-density [3], it immediately follows that these parameters are continuous in the \( L \)-density topology.

In [4], many topologies were constructed for the space of equivalent channels sharing a fixed input alphabet. In this paper, we study the continuity of several channel parameters and operations under these topologies.

The continuity of channel parameters is important both theoretically and practically. If a parameter (such as the optimal probability of error of a given code) is difficult to compute for a channel \( W \), one can approximate it by computing the same parameter for a sequence of channels \( (W_n)_{n \geq 0} \) that converges to \( W \) in some topology. Another application of the continuity of channel parameters and operations is the study of robustness of various system designs against the imperfect specification of the channel.

II. PRELIMINARIES

We assume that the reader is familiar with the basic concepts of general topology. If the reader finds any notation or concept unfamiliar, he can find its formal definition in the preliminaries section of [4]. Moreover, due to the space limitation, we are only able to explain the intuition behind the definitions and state the main results. The proofs can be found in [5].

A. Measure-theoretic notations

If \( (M, \Sigma) \) is a measurable space, we denote the set of probability measures on \( (M, \Sigma) \) as \( \mathcal{P}(M, \Sigma) \). Let \( P \in \mathcal{P}(M, \Sigma) \), and let \( f : M \to M' \) be a measurable mapping from \( (M, \Sigma) \) to another measurable space \( (M', \Sigma') \). The push-forward probability measure of \( P \) by \( f \) is the probability measure \( f_#P \) on \( (M', \Sigma') \) defined as \( (f_#P)(A') = P(f^{-1}(A')) \) for every \( A' \in \Sigma' \).

We denote the product of two probability measures \( P_1 \in \mathcal{P}(M_1, \Sigma_1) \) and \( P_2 \in \mathcal{P}(M_2, \Sigma_2) \) as \( P_1 \times P_2 \).

B. Random mappings

Let \( M \) and \( M' \) be two arbitrary sets and let \( \Sigma' \) be a \( \sigma \)-algebra on \( M' \). A random mapping from \( M \) to \( (M', \Sigma') \) is a mapping \( R \) from \( M \) to \( \mathcal{P}(M', \Sigma') \), i.e., for every \( x \in M \) we associate a probability distribution \( R(x) \) on \( M' \). \( R(x) \) can be

\(^1\)The symmetric capacity is the input-output mutual information with uniformly distributed input.
interpreted as the probability distribution of the random output given that the input is \( x \).

Let \( \Sigma \) be a \( \sigma \)-algebra on \( M \). We say that \( R \) is a measurable random mapping from \((M, \Sigma)\) to \((M', \Sigma')\) if the mapping \( R_B : M \to \mathbb{R} \) defined as \( R_B(x) = (R(x))(B) \) is measurable for every \( B \in \Sigma' \). Note that this definition of measurability is consistent with the measurability of ordinary mappings: let \( f \) be a mapping from \( M \) to \( M' \) and let \( D_f : M \to \mathcal{P}(M', \Sigma') \) be the random mapping defined as \( D_f(x) = \delta_{f(x)} \) for every \( x \in M \), where \( \delta_{f(x)} \in \mathcal{P}(M', \Sigma') \) is a Dirac measure centered at \( f(x) \). \( D_f \) is a measurable random mapping if and only if \( f \) is a measurable mapping.

Let \( P \) be a probability measure on \((M, \Sigma)\) and let \( R \) be a measurable random mapping from \((M, \Sigma)\) to \((M', \Sigma')\). The push-forward probability measure of \( P \) by \( R \) is the probability measure \( R_* P \) on \((M', \Sigma')\) defined as:

\[
(R_* P)(B) = \int_M R_B \cdot dP, \quad \forall B \in \Sigma'.
\]

Note that this definition is consistent with the push-forward of ordinary mappings: if \( f \) and \( D_f \) are as above, then \((D_f)_* P = f_* P\).

**C. Meta-probability measures**

Let \( X \) be a finite set. We denote the set of probability distributions on \( X \) as \( \Delta_X \). A meta-probability measure on \( X \) is a probability measure on the Borel sets of \( \Delta_X \). It is called a meta-probability measure because it is a probability measure on the space of probability distributions on \( X \). We denote the set of meta-probability measures on \( X \) as \( \mathcal{M}(\Delta_X) \).

Let \( f \) be a mapping from a finite set \( X \) to another finite set \( X' \). \( f \) induces a push-forward mapping \( f_* \) taking probability distributions in \( \Delta_X \) to probability distributions in \( \Delta_{X'} \). \( f_* \) in turn induces another push-forward mapping taking meta-probability measures in \( \mathcal{M}(\Delta_X) \) to meta-probability measures in \( \mathcal{M}(\Delta_{X'}) \). We denote this mapping as \( f_* \) and we call it the meta-push-forward mapping induced by \( f \). The mapping \( f_* \) is continuous under both the weak-* and the total variation topologies.

Let \( X_1 \) and \( X_2 \) be two finite sets. Let \( \text{Mul} : \Delta_{X_1} \times \Delta_{X_2} \to \Delta_{X_1 \times X_2} \) be defined as \( \text{Mul}(p_1, p_2) = p_1 \times p_2 \). For every \( \mathcal{M}_1 \in \mathcal{M}(\Delta_{X_1}) \) and \( \mathcal{M}_2 \in \mathcal{M}(\Delta_{X_2}) \), we define the tensor product of \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \) as \( \mathcal{M}_1 \otimes \mathcal{M}_2 = \text{Mul}_* (\mathcal{M}_1 \times \mathcal{M}_2) \in \mathcal{M}(\Delta_{X_1 \times X_2}) \). The tensor product is a continuous mapping from \( \mathcal{M}(\Delta_{X_1}) \times \mathcal{M}(\Delta_{X_2}) \) to \( \mathcal{M}(\Delta_{X_1 \times X_2}) \) under both the weak-* and the total variation topologies.

**D. DMC topologies**

In this subsection, we describe the topologies that we introduced in [4]. A more detailed summary can be found in Section III of [5].

We denote the set of all channels with input alphabet \( X \) and output alphabet \( Y \) as \( \text{DMC}_{X,Y} \). The space of channels with input alphabet \( X \) is defined as

\[
\text{DMC}_{X,*} = \prod_{n \geq 1} \text{DMC}_{X,[n]},
\]

where \( [n] = \{1, \ldots, n\} \) and \( \prod \) is the disjoint union symbol. Two channels with input alphabet \( X \) are said to be equivalent if they are degraded from each other. The equivalence relation on \( \text{DMC}_{X,Y} \) is denoted as \( \text{R}_{X,Y}^{(o)} \). The equivalence relation on \( \text{DMC}_{X,*} \) is denoted as \( \text{R}_{X,*}^{(o)} \).

We define the metric distance \( d_{X,Y} \) on \( \text{DMC}_{X,Y} \) as

\[
d_{X,Y}(W, W') = \frac{1}{2} \max_{x \in X} \sum_{y \in Y} |W(y|x) - W'(y|x)|.
\]

The topology on \( \text{DMC}_{X,Y} \) that is induced by \( d_{X,Y} \) is denoted as \( T_{X,Y} \). The quotient topological space obtained from \((\text{DMC}_{X,Y}, T_{X,Y})\) by \( R_{X,Y}^{(o)} \) is denoted as \((\text{DMC}_{X,Y}, T_{X,Y}^{(o)})\).

The quotient of \( \text{DMC}_{X,*} \) by \( R_{X,*}^{(o)} \) is denoted as \( \text{DMC}_{X,*}^{(o)} \).

The strong topology \( T_{X,*}^{(o)} \) on \( \text{DMC}_{X,*}^{(o)} \) is the finest topology that makes the inclusion mappings from \( \text{DMC}_{X,*}^{(o)} \rightarrow \text{DMC}_{X,[n]} \to \text{DMC}_{X,*}^{(o)} \) continuous for every \( n \geq 1 \).

In [4], we introduced the noisiness metric \( d_{X,*}^{(o)} \) on \( \text{DMC}_{X,*} \).

This metric compares the “noisiness levels” between channels. The noisiness topology \( T_{X,*}^{(o)} \) on \( \text{DMC}_{X,*} \) is the topology induced by \( d_{X,*}^{(o)} \).

Let \( W \) be a channel with input alphabet \( X \), and assume a uniform prior probability distribution on \( X \). The meta-probability measure that describes the possible posterior probability distributions is called the Blackwell measure\(^2\) of \( W \), and it is denoted as \( \text{MP}_W \). Two channels are equivalent if and only if they have the same Blackwell measures [6]. Moreover, there is a canonical bijection between \( \text{DMC}_{X,*}^{(o)} \) and the set of balanced and finitely supported meta-probability measures \( \mathcal{M}(\text{DMC}_{X,*}^{(o)}) \) [6]. This allows us to construct the weak-* topology and the total variation topology \( T_{X,Y} \) on \( \text{DMC}_{X,*} \) by transporting the corresponding topologies from \( \mathcal{M}(\text{DMC}_{X,*}^{(o)}) \) to \( \text{DMC}_{X,*} \) through the canonical bijection. We showed in [4] that the weak-* topology is exactly the same as the noisiness topology \( T_{X,*}^{(o)} \).

**III. CHANNEL PARAMETERS AND OPERATIONS**

**A. Useful parameters**

For every \( p \in \Delta_X \) and every \( W \in \text{DMC}_{X,Y} \), define \( I(p, W) \) as the mutual information \( I(X; Y) \) computed using the natural logarithm, where \( X \) is distributed as \( p \) and \( Y \) is the output of \( W \) when \( X \) is the input. The capacity of \( W \) is defined as \( C(W) = \sup_{p \in \Delta_X} I(p, W) \).

For every \( p \in \Delta_X \), the error probability of the MAP decoder of \( W \) under prior \( p \) is defined as:

\[
P_e(p, W) = 1 - \sum_{y \in Y} \max_{x \in X} p(x) W(y|x).
\]

Clearly, \( 0 \leq P_e(p, W) \leq 1 \).

\(^2\)In an earlier version of this work, I called \( \text{MP}_W \) the posterior meta-probability distribution of \( W \). Maxim Raginsky thankfully brought to my attention the fact that \( \text{MP}_W \) is called Blackwell measure.
For every $W \in \text{DMC}_{X,Y}$, define the Bhattacharyya parameter of $W$ as
\[ Z(W) = \frac{1}{|X|} \sum_{x_1,x_2 \in X, y \in Y} \sum_{x_1 \neq x_2} \sqrt{W(y|x_1)W(y|x_2)}, \]
if $|X| \geq 2$, and $Z(W) = 0$ if $|X| = 1$. It is easy to see that $0 \leq Z(W) \leq 1$. It was shown in [7] and [8] that
\[ \frac{1}{4} Z(W)^2 \leq P_e(\pi, W) \leq (|X| - 1) Z(W), \]
where $\pi$ is the uniform distribution on $X$.

An $(n,M)$-code $C$ on the alphabet $X$ is a subset of $X^n$ satisfying $|C| = M$. The blocklength of $C$ is $n$, and $M$ is the size of the code. The rate of $C$ is $\frac{n}{M} \log M$, and it is measured in nats. The error probability of the ML decoder for the code $C$ when it is used for a channel $W \in \text{DMC}_{X,Y}$ is given by:
\[ P_{e,C}(W) = 1 - \frac{1}{|C|} \sum_{C \ni y \in Y} \max_{x_1 \in X} \left\{ \sum_{i=1}^{n} W(y_i|x_i) \right\}. \]

The optimal error probability of $(n,M)$-codes for a channel $W$ is given by:
\[ P_{e,n,M}(W) = \min_{C \ni X'} \frac{1}{|C|} \sum_{C \ni y \in Y} \max_{x_1 \in X'} \left\{ \sum_{i=1}^{n} W(y_i|x_i) \right\}. \]

It is well known that all these parameters depend only on the $R_{X,Y}$-equivalence class of $W$. Therefore, we can define those parameters for every $W \in \text{DMC}_{X,Y}$. We can show that these parameters are continuous on $(\text{DMC}_{X,Y}^+, \mathcal{T}_{X,Y}^+)$ (see [5]).

B. Channel operations

For every two channels $W_1 \in \text{DMC}_{X_1,Y_1}$ and $W_2 \in \text{DMC}_{X_2,Y_2}$, define the channel sum $W_1 \oplus W_2 \in \text{DMC}_{X_1 \cup X_2, Y_1 \cup Y_2}$ of $W_1$ and $W_2$ as:
\[ (W_1 \oplus W_2)(y, i|x, j) = \begin{cases} W_i(y|x) & \text{if } i = j, \\ 0 & \text{otherwise}, \end{cases} \]
where $X_1 \cup X_2 = (X_1 \times \{1\}) \cup (X_2 \times \{2\})$ is the disjoint union of $X_1$ and $X_2$. $W_1 \oplus W_2$ arises when the transmitter has two channels $W_1$ and $W_2$ at his disposal and he can use exactly one of them at each channel use.

We define the channel product $W_1 \otimes W_2 \in \text{DMC}_{X_1 \times X_2, Y_1 \times Y_2}$ of $W_1$ and $W_2$ as:
\[ (W_1 \otimes W_2)(y_1, y_2|x_1, x_2) = W_1(y_1|x_1)W_2(y_2|x_2). \]
$W_1 \otimes W_2$ arises when the transmitter has two channels $W_1$ and $W_2$ at his disposal and he uses both of them at each channel use. Channel sums and products were first introduced by Shannon in [9].

For every $W_1 \in \text{DMC}_{X_1,Y_1}$, $W_2 \in \text{DMC}_{X_2,Y_2}$ and every $0 \leq \alpha \leq 1$, we define the $\alpha$-interpolation $[\alpha W_1, (1 - \alpha)W_2] \in \text{DMC}_{X_1 \times X_2, Y_1 \times Y_2}$ between $W_1$ and $W_2$ as:
\[ [\alpha W_1, (1 - \alpha)W_2](y, i|x) = \begin{cases} \alpha W_1(y|x) & \text{if } i = 1, \\ (1 - \alpha)W_2(y|x) & \text{if } i = 2. \end{cases} \]

Channel interpolation arises when a channel behaves as $W_1$ with probability $\alpha$ and as $W_2$ with probability $1 - \alpha$. The transmitter has no control on which behavior the channel chooses, but on the other hand, the receiver knows which behavior was chosen. Channel interpolations were used in [10] to construct interpolations between polar codes and Reed-Muller codes.

Fix a binary operation $\ast$ on $X$. For every $W \in \text{DMC}_{X,Y}$, define $W^- \in \text{DMC}_{X,Y}$ and $W^+ \in \text{DMC}_{X^2 \times X}$ as:
\[ W^-(y_1, y_2|u_1) = \frac{1}{|X|^2} \sum_{u_2 \in X} W(y_1|u_1 \ast u_2)W(y_2|u_2), \]
and
\[ W^+(y_1, y_2, u_1|u_2) = \frac{1}{|X|} W(y_1|u_1 \ast u_2)W(y_2|u_2). \]
These operations generalize Arkan’s polarization transformations [11].

All the channel operations that are defined here can be “quotiented” by the equivalence relations. We just need to realize that the equivalence class of the resulting channel depends only on the equivalent classes of the channels that were used in the operation (see [5]). For example, for every $W_1 \in \text{DMC}_{X_1,Y_1}$ and every $W_2 \in \text{DMC}_{X_2,Y_2}$, we can define the channel sum $W_1 \oplus W_2 \in \text{DMC}_{X_1 \cup X_2, Y_1 \cup Y_2}$ as the $R_{X_1 \cup X_2,Y_1 \cup Y_2}$-equivalence class of $W_1 \oplus W_2$ for any $W_i \in W_1$ and any $W_j \in W_2$. We can similarly define all the other operations on the quotient spaces similarly. All these operations are continuous on the quotient spaces (see [5]).

IV. CONTINUITY IN THE STRONG TOPOLOGY

Since the channel parameters $I, P_e, Z, P_{e,C}$ and $P_{e,n,M}$ are defined on $\text{DMC}_{X,n}$ for every $n \geq 1$ (see Section III-A), they are also defined on $\text{DMC}_{X,n}^+$, $\text{DMC}_{X,n}^+$ and $\text{DMC}_{X,n}^+$.

**Theorem 1.** Let $\mathcal{U}_X$ be the standard topology on $\Delta_X$. We have:

- $I : \Delta_X \times \text{DMC}_{X,n}^+ \to \mathbb{R}^+$ is continuous on $(\Delta_X \times \text{DMC}_{X,n}^+)(\mathcal{U}_X \otimes \mathcal{T}_{X,n}^+)$ and concave in $p$, where $\mathcal{U}_X \otimes \mathcal{T}_{X,n}^+ = \text{product topology of } \mathcal{U}_X \text{ and } \mathcal{T}_{X,n}^+$.
- $C : \text{DMC}_{X,n}^+ \to \mathbb{R}^+$ is continuous on $(\text{DMC}_{X,n}^+)(\mathcal{U}_X \otimes \mathcal{T}_{X,n}^+)$.
- $P_e : \Delta_X \times \text{DMC}_{X,n}^+ \to [0,1]$ is continuous on $(\Delta_X \times \text{DMC}_{X,n}^+)(\mathcal{U}_X \otimes \mathcal{T}_{X,n}^+)$ and concave in $p$.
- $Z : \text{DMC}_{X,n}^+ \to [0,1]$ is continuous on $(\text{DMC}_{X,n}^+)(\mathcal{U}_X \otimes \mathcal{T}_{X,n}^+)$.

For every code $C$ on $X$, $P_{e,C} : \text{DMC}_{X,n}^+ \to [0,1]$ is continuous on $(\text{DMC}_{X,n}^+)(\mathcal{T}_{X,n}^+)$.

For every $n > 0$ and every $1 \leq M \leq |X|^n$, the mapping $P_{e,n,M} : \text{DMC}_{X,n}^+ \to [0,1]$ is continuous on $(\text{DMC}_{X,n}^+)(\mathcal{T}_{X,n}^+)$. It is also possible to extend the definition of all the channel operations that were defined in section III-B to $\text{DMC}_{X,n}^+$.
Theorem 2. Assume that all spaces of equivalent channels are endowed with the strong topology. We have:

- The mapping \( \overline{W}_1, \overline{W}_2 : D M C^{(o)}_{X_1} \times D M C^{(o)}_{X_2} \to D M C^{(o)}_{X_1 \oplus X_2} \) is continuous.
- The mapping \( \overline{W}_1, \overline{W}_2 : D M C^{(o)}_{X_1} \to D M C^{(o)}_{X_1 \otimes X_2} \) is continuous.
- The mapping \( \overline{W}_1, \overline{W}_2, \alpha : D M C^{(o)}_{X_1} \times D M C^{(o)}_{X_1 \times X_2} \to D M C^{(o)}_{X_1 \times X_2} \) is continuous.
- For every binary operation \( \ast \) on \( X \), the mapping \( \bar{W} \mapsto \bar{W}^\ast \) from \( D M C^{(o)}_{X_1} \to D M C^{(o)}_{X_1} \) is continuous.
- For any binary operation \( \ast \) on \( X \), the mapping \( \bar{W} \mapsto \bar{W}^\ast \) from \( D M C^{(o)}_{X_1} \to D M C^{(o)}_{X_1} \) is continuous.

Corollary 1. \( (D M C^{(o)}_{X_1}, T^{(o)}_{s,s_{X_1}}) \) is strongly contractible\(^3\) to every point in \( D M C^{(o)}_{X_1} \).

5. Continuity in the Noisiness/Weak-* and the Total Variation Topologies

We need to express the channel parameters and operations in terms of the Blackwell measures.

A. Channel parameters

The following proposition shows that many channel parameters can be expressed as an integral of a continuous function with respect to the Blackwell measure:

Proposition 1. For every \( \bar{W} \in D M C^{(o)}_{X_1} \), we have:

\[
\forall p \in \Delta_X, \quad P_e(p, \bar{W}) = H(p) - |X| \cdot \int_{\Delta_X} \frac{p(x)p'(x) \log \frac{p(x)}{p'(x)}}{\sum_{y'} p(y')p'(y')} \cdot d M P_1(p').
\]

\[
\forall p \in \Delta_X, \quad P_e(p, \bar{W}) = 1 - |X| \cdot \int_{X} \max_{x \in X} \{ p(x) \times p'(x) \} \cdot d M P_1(p').
\]

\[
\text{If } |X| \geq 2, \quad Z(\bar{W}) = \frac{1}{|X| - 1} \sum_{x \neq x'} \int_{\Delta_X} \sqrt{p(x)p(x')} \cdot d M P_1(p).
\]

For every code \( C \subset X^n \), we have

\[
P_{e,C}(\bar{W}) = 1 - \frac{|X|^n}{|C|} \cdot \int_{\Delta_X} \max_{x_1 \in X^n} \left\{ \prod_{i=1}^n p_i(x_i) \right\} \cdot d M P_1^n(p^n).
\]

where \( H(p) \) is the entropy of \( p \), and \( M P^n_0(W) \) is the product measure on \( \Delta^n_X \) obtained by multiplying \( M P_0 \) with itself \( n \) times. Note that we adopt the standard convention that \( \log \frac{0}{0} = 0 \).

Theorem 3. Let \( U_X \) be the standard topology on \( \Delta_X \) and let \( T = T^{(o)}_{s,s_{X_1}} \) or \( T = T^{(o)}_{s,s_{TV,X_1}} \). We have:

- \( I : \Delta_X \times D M C^{(o)}_{X_1} \to \mathbb{R}^+ \) is continuous on \( (\Delta_X \times D M C^{(o)}_{X_1}, U_X \otimes T) \) and concave in \( p \).
- \( C : D M C^{(o)}_{X_1} \to \mathbb{R}^+ \) is continuous on \( (D M C^{(o)}_{X_1}, T) \).
- \( P_e : \Delta_X \times D M C^{(o)}_{X_1} \to [0, 1] \) is continuous on \( (\Delta_X \times D M C^{(o)}_{X_1}, U_X \otimes T) \) and concave in \( p \).
- \( Z : D M C^{(o)}_{X_1} \to [0, 1] \) is continuous on \( (D M C^{(o)}_{X_1}, T) \).
- For every code \( C \) on \( X \), \( P_{e,C} : D M C^{(o)}_{X_1} \to [0, 1] \) is continuous on \( (D M C^{(o)}_{X_1}, T) \).

B. Channel Operations

In the following, we show that we can express the channel operations in terms of Blackwell measures. We have all the tools to achieve this for the channel sum, channel product and channel interpolation. In order to express the channel polarization transformations in terms of the Blackwell measures, we need to introduce new definitions.

Let \( X' \) be a finite set and let \( \ast \) be a binary operation on \( X' \). We say that \( \ast \) is uniformity preserving if the mapping \( (a, b) \to (a \ast b, a) \) is a bijection from \( X' \) to itself [12]. For every \( a, b \in X \), we denote the unique element \( c \in X \) satisfying \( c \ast b = a \) as \( c = a/b \). It was shown in [8] that a binary operation \( \ast \) is polarizing if and only if it is uniformity preserving and \( / \ast \) is strongly ergodic. Binary operations that are not uniformity preserving are not interesting for polarization theory because they do not preserve the symmetric capacity [8]. Therefore, we will focus only on polarization transformations that are based on uniformity preserving binary operations.

Let \( \ast \) be a fixed uniformity preserving binary operation on \( X \). Define the mapping \( C^{-\ast} : \Delta_X \times \Delta_X \to \Delta_X \) as

\[
(C^{-\ast}(p_1, p_2))(u) = \sum_{u_2 \in X} p_1(u_1 \ast u_2)p_2(u_2).
\]

The probability distribution \( C^{-\ast}(p_1, p_2) \) can be interpreted as follows: let \( X_1 \) and \( X_2 \) be two independent random variables in \( X \) that are distributed as \( p_1 \) and \( p_2 \) respectively, and let \( (U_1, U_2) \) be the random pair in \( X^2 \) defined as \( (U_1, U_2) = (X_1/X_2, X_2) \), or equivalently \((X_1, X_2) = (U_1 + U_2, U_2)\). \( C^{-\ast}(p_1, p_2) \) is the probability distribution of \( U_1 \).

For every \( MP_1, MP_2 \in MP(X) \), we define the \((-\ast)\)-convolution of \( MP_1 \) and \( MP_2 \) as

\[
(MP_1, MP_2)^{-\ast} = C^{-\ast}_{\#}(MP_1 \times MP_2) \in MP(X).
\]

For every \( p_1, p_2 \in \Delta_X \) and every \( u_1 \in X \) satisfying \( (C^{-\ast}(p_1, p_2))(u_1) > 0 \), define \( C^{+,\#}(p_1, p_2) \in \Delta_X \) as

\[
(C^{+,\#}(p_1, p_2))(u_2) = \frac{p_1(u_1 \ast u_2)p_2(u_2)}{(C^{-\ast}(p_1, p_2))(u_1)}.
\]

The probability distribution \( C^{+,\#}(p_1, p_2) \) can be interpreted as follows: if \( X_1, X_2, U_1 \) and \( U_2 \) are as above, \( C^{+,\#}(p_1, p_2) \) is the conditional probability distribution of \( U_2 \) given \( U_1 = u_1 \).
Define the mapping $C^{+,\ast} : \Delta_X \times \Delta_X \rightarrow \mathcal{MP}(\mathcal{X})$ as follows:

$$C^{+,\ast}(p_1,p_2) = \sum_{u_1 \in \mathcal{X}} \frac{(C^{-\ast}(p_1,p_2))(u_1) \cdot \delta_{C^{+,\ast}(p_1,p_2)}(u_1)}{(C^{-\ast}(p_1,p_2))(u_1) \ast 0}$$

where $\delta_{C^{+,\ast}}(p_1,p_2)$ is a Dirac measure centered at $C^{+,\ast}(p_1,p_2)$. If $X_1, X_2, U_1$ and $U_2$ are as above, $C^{+,\ast}(p_1,p_2)$ is the meta-probability measure that describes the possible conditional probability distributions of $U_2$ that are seen by someone having knowledge of $U_1$. $C^{+,\ast}$ is a measurable random mapping from $\Delta_X \times \Delta_X$ to $\Delta_X$ (see [5]).

For every $\mathcal{MP}_1, \mathcal{MP}_2 \in \mathcal{MP}(\mathcal{X})$, we define the $(+\ast)$-convolution of $\mathcal{MP}_1$ and $\mathcal{MP}_2$ as:

$$(\mathcal{MP}_1, \mathcal{MP}_2)^{+,\ast} = C^{+,\ast}_{\ast}(\mathcal{MP}_1 \times \mathcal{MP}_2) \in \mathcal{MP}(\mathcal{X}) .$$

Proposition 2. We have:

- For every $\hat{W}_1 \in \mathcal{DMC}^{(o)}_{X_1,\ast}$ and $\overline{W}_2 \in \mathcal{DMC}^{(o)}_{X_2,\ast}$, we have:
  
  $$\mathcal{MP}_{\hat{W}_1 \otimes \overline{W}_2} = \mathcal{MP}_1 \mathcal{MP}_{\hat{W}_1} \ast \mathcal{MP}_2 \mathcal{MP}_{\overline{W}_2} ,$$

  where $\mathcal{MP}_W^{(o)}$ (respectively $\mathcal{MP}_{\overline{W}}^{(o)}$) is the meta-push-forward of $\mathcal{MP}_{\hat{W}_1}$ (respectively $\mathcal{MP}_{\overline{W}_2}$) by the canonical projection from $\mathcal{X}_1$ (respectively $\mathcal{X}_2$) to $\hat{X}_1 \coprod \hat{X}_2$.

- For every $\hat{W}_1 \in \mathcal{DMC}^{(o)}_{X_1,\ast}$ and $\overline{W}_2 \in \mathcal{DMC}^{(o)}_{X_2,\ast}$, we have:
  
  $$\mathcal{MP}_{\hat{W}_1 \otimes \overline{W}_2} = \mathcal{MP}_{\hat{W}_1} \otimes \mathcal{MP}_{\overline{W}_2} ,$$

  for every $\alpha \in [0, 1]$ and every $\hat{W}_1, \hat{W}_2 \in \mathcal{DMC}^{(o)}_{X,\ast}$, we have:

  $$\mathcal{MP}_{\hat{W}_1 \otimes (1-\alpha)\overline{W}_2} = \alpha \mathcal{MP}_{\hat{W}_1} + (1-\alpha) \mathcal{MP}_{\overline{W}_2} .$$

- For every uniformity preserving binary operation $\oplus$ on $\mathcal{X}$ and every $\hat{W} \in \mathcal{DMC}^{(o)}_{X,\ast}$, we have:

  $$\mathcal{MP}_{\hat{W}^{\oplus}} = (\mathcal{MP}_{\hat{W}}, \mathcal{MP}_{\hat{W}})^{+,\ast} .$$

- For every uniformity preserving binary operation $\otimes$ on $\mathcal{X}$ and every $\hat{W} \in \mathcal{DMC}^{(o)}_{X,\ast}$, we have:

  $$\mathcal{MP}_{\hat{W}^{\otimes}} = (\mathcal{MP}_{\hat{W}}, \mathcal{MP}_{\hat{W}})^{+,\ast} .$$

Note that the polarization transformation formulas in Proposition 2 generalize the formulas given by Raginsky in [13] for binary-input channels.

Theorem 4. Assume that all spaces of equivalent channels are endowed with the noisiness/weak-* or the total variation topology.

- The mapping $(\hat{W}_1, \overline{W}_2) \rightarrow \hat{W}_1 \oplus \overline{W}_2$ from $\mathcal{DMC}^{(o)}_{X_1,\ast} \times \mathcal{DMC}^{(o)}_{X_2,\ast}$ to $\mathcal{DMC}^{(o)}_{X_1 \coprod X_2,\ast}$ is continuous.

- The mapping $(\hat{W}_1, \overline{W}_2) \rightarrow \hat{W}_1 \otimes \overline{W}_2$ from $\mathcal{DMC}^{(o)}_{X_1,\ast} \times \mathcal{DMC}^{(o)}_{X_2,\ast}$ to $\mathcal{DMC}^{(o)}_{X_1 \coprod X_2,\ast}$ is continuous.

- The mapping $(\hat{W}_1, \overline{W}_2, \alpha) \rightarrow [\alpha \hat{W}_1, (1-\alpha)\overline{W}_2]$ from $\mathcal{DMC}^{(o)}_{X_1,\ast} \times \mathcal{DMC}^{(o)}_{X_2,\ast} \times [0, 1]$ to $\mathcal{DMC}^{(o)}_{X_1 \coprod X_2,\ast}$ is continuous.

- For every uniformity preserving binary operation $\ast$ on $\mathcal{X}$, the mapping $\hat{W} \rightarrow \hat{W}^{-}$ from $\mathcal{DMC}^{(o)}_{X,\ast}$ to $\mathcal{DMC}^{(o)}_{X,\ast}$ is continuous.

- For every uniformity preserving binary operation $\ast$ on $\mathcal{X}$, the mapping $\hat{W} \rightarrow \hat{W}^{+}$ from $\mathcal{DMC}^{(o)}_{X,\ast}$ to $\mathcal{DMC}^{(o)}_{X,\ast}$ is continuous.

Corollary 2. Both $(\mathcal{DMC}^{(o)}_{X,\ast}, T^{(o)}_{\mathcal{X}_1,\ast})$ and $(\mathcal{DMC}^{(o)}_{X,\ast}, T^{(o)}_{\mathcal{X}_2,\ast})$ are strongly contractible to every point in $\mathcal{DMC}^{(o)}_{X,\ast}$.

VI. DISCUSSION

It is worth mentioning that Theorem 1 can be proven from Theorem 3 because the noisiness topology is coarser than the strong topology. The main reason why we provided another proof for Theorem 1 in [5] is to show that the quotient formulation of the strong topology makes it easy to work with.

The continuity of the channel sum and the channel product on the whole product space $(\mathcal{DMC}^{(o)}_{X_1,\ast} \times \mathcal{DMC}^{(o)}_{X_2,\ast}, T^{(o)}_{\mathcal{X}_1,\ast} \otimes T^{(o)}_{\mathcal{X}_2,\ast})$ remains an open problem. As we explain in [5], it is sufficient to prove that the product topology $T^{(o)}_{\mathcal{X}_1,\ast} \otimes T^{(o)}_{\mathcal{X}_2,\ast}$ is compactly generated.

ACKNOWLEDGMENT

I would like to thank Emre Telatar and Mohammad Bazzi for helpful discussions. I am also grateful to Maxim Raginsky for his comments.

REFERENCES

[1] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
[2] H. Schwarte, “On weak convergence of probability measures, channel capacity and code error probabilities,” IEEE Transactions on Information Theory, vol. 42, no. 5, pp. 1549–1551, Sep 1996.
[3] T. Richardson and R. Urbanke, Modern Coding Theory. New York, NY, USA: Cambridge University Press, 2008.
[4] R. Nasser, “Topological structures on DMC spaces,” arXiv:1701.04467, Jan 2017.
[5] ——, “Continuity of channel parameters and operations under various DMC topologies,” arXiv:1701.04466, Jan 2017.
[6] E. Torgersen, Comparison of Statistical Experiments. ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1991.
[7] E. Şaşoğlu, E. Telatar, and E. Arıkın, “Polarization for arbitrary discrete memoryless channels,” in Information Theory Workshop, 2009. ITW 2009, IEEE, 2009, pp. 144–148.
[8] R. Nasser, “An ergodic theory of binary operations, part II: Applications to polarization,” IEEE Transactions on Information Theory, vol. 63, no. 2, pp. 1063–1083, Feb 2017.
[9] C. Shannon, “The zero error capacity of a noisy channel,” IRE Transactions on Information Theory, vol. 2, no. 3, pp. 8–19, September 1956.
[10] M. Mondelli, S. H. Hassani, and R. L. Urbanke, “From polar to reed-muller codes: A technique to achieve the finite-length performance,” IEEE Transactions on Communications, vol. 62, no. 9, pp. 3084–3091, Sept 2014.
[11] E. Arıkın, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” Information Theory, IEEE Transactions on, vol. 55, no. 7, pp. 3051–3073, 2009.
[12] R. Nasser, “An ergodic theory of binary operations, part I: Key properties,” IEEE Transactions on Information Theory, vol. 62, no. 12, pp. 6931–6952, Dec 2016.
[13] M. Raginsky, “Channel polarization and blackwell measures,” in 2016 IEEE International Symposium on Information Theory (ISIT), July 2016, pp. 56–60.