A General Formula for Compound Channel Capacity

Sergey Loyka, Charalambos D. Charalambous

***Abstract*—A general formula for the capacity of arbitrary compound channels, which are not necessarily ergodic, stationary or information-stable, is obtained using the information density approach. A direct (constructive) proof is given. To prove achievability, we generalize Feinstein Lemma to the compound channel setting, and to prove converse, we generalize Verdu-Han Lemma to the same compound setting. This extends the general formula for channel capacity in [8] to arbitrary compound channels (not necessarily finite-state or countable).**

I. Introduction

CHANNEL state information (CSI) has a significant impact on channel performance as well as code design to achieve that performance. This effect is especially pronounced for wireless channels, due to their dynamic nature, limitations of a feedback link (if any), channel estimation errors etc. [1].

When only incomplete or inaccurate CSI is available, performance analysis and coding techniques have to be modified properly. The impact of channel uncertainty has been extensively studied since late 1950s [2]-[6]; see [7] for an extensive literature review up to late 1990s. Since channel estimation is done at the receiver (Rx) and then transmitted to the transmitter (Tx) via a limited (if any) feedback link, most studies concentrate on limited CSI available at the Tx end assuming full CSI at the Rx end, the assumption we adopt in this paper.

There are several typical approaches to model channel uncertainty. In the compound channel model, the channel is unknown to the Tx but is known to belong to a certain set of channels. A member of the channel uncertainty set (state set) is selected at the beginning and held constant during the entire transmission [3]-[5], thus modeling a scenario with little dynamics (channel coherence time significantly exceeds the codeword duration [1]). A more dynamic approach is that of the arbitrary-varying channel, where the channel is allowed to vary from symbol to symbol being unknown to the Tx (but also restricted to belong to a certain class of channels) [6]. A variation of the compound channel model is that of the composite channel where there is a probability assigned to each member of the compound channel set thus avoiding an over-pessimistic nature of the compound channel capacity when one channel is particularly bad but occurs with small probability [10]. Finally, incomplete CSI at the Tx end can be addressed by assuming that the channel is not known but its distribution is known to the Tx, the so-called channel distribution information (CDI) [1].

All the studies above of compound channels require members of the uncertainty (state) set to be information-stable. In this paper, we relax this assumption and obtain compound capacity of information-unstable channels using the compound channel approach [2]-[7] in combination with the information density approach [8]-[10]. This results in a general formula for the capacity of compound channels, which are not necessarily ergodic, stationary or information-stable, which extends the general capacity formula of regular (non-compound) channels in [8] to compound channels with arbitrary channel state sets.

The capacity of a class of compound channels was obtained earlier in [9] using the information density approach. However, (i) its proof is rather involved and indirect (first, a result is established for mixed channels; then, a certain equivalence is established between mixed and compound channels, which establishes the compound channel capacity in a rather elaborate and indirect way); and (ii) it holds for finite-state channels only[1]. In the present paper, we give a direct (simpler) proof by extending Feinstein and Verdu-Han Lemmas to compound channel setting (using an algorithmic code construction), which also holds for arbitrary state sets (not only countable, finite-state etc.). The main results are in Theorems 1, 2 in Section IV.

A formulation of channel uncertainty problem based on the information density approach was presented in [10] using the composite channel model. This, however, requires a probability measure associated with channel states, so that the channel input-output description is entirely probabilistic and the general formula in [8] applies to such setting. We consider the compound channel setting here, where there is no probability measure associated with channel states and a certain achievable performance has to be demonstrated for any member of the uncertainty class using a single code, for which the general formula in [8] is not applicable.

Section II introduces a (general) channel model and assumptions. The information density approach [8][9] is introduced and briefly reviewed in section III. In section IV, a general compound channel capacity formula is obtained using the

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ontario, Canada, e-mail: sergey.loyka@ieee.org

C.D. Charalambous is with the ECE Department, University of Cyprus, Nicosia, Cyprus, e-mail: chadcha@ucy.ac.cy

[1]While [9] claims its validity for countably-infinite-state channels, supremum over channel states is missing in error probability definition in [9] (Definition 3.3.1), so that arbitrary low error probability cannot be guaranteed for all channel states via large block length (see e.g. [1]-[7] for a proper definition of compound error probability). When re-instated, the upper bound at bottom of p. 199 becomes ∞ for infinite-state channels. Thus, Theorem 3.3.5 in [9] is proved for finite-state channels only. In fact, Example 1 in Section V shows that this Theorem is invalid for infinite-state channels in general.

information density approach, which holds for a wide class of channels including non-stationary, non-ergodic or information-unstable channels and arbitrary channel state sets (not only countable or finite-state).

II. Channel Model

Let us consider a generic discrete-time channel model shown in Fig. 1, where $X^n = \{X_1 ... X_n\}$ is a (random) sequence of n input symbols, $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ denotes all such sequences, and Y^n is the corresponding output sequence; $s \in \mathcal{S}$ denotes the channel state (which may also be a sequence) and $\mathcal{S}$ is the (arbitrary) uncertainty set; $p_s(y^n|x^n)$ is the channel transition probability; $p(x^n)$ and $p_s(y^n)$ are the input and output distributions under channel state s.

Let us assume that the full channel state information (CSI) is available at the receiver but not the transmitter (see e.g. [1] for a detailed motivation of this assumption; when the channel is quasi-static, this assumption is not necessary) and that the channel input $\boldsymbol{X}$ and state s are independent of each other. Following the standard approach (see e.g. [1]), we augment the channel output with the state: $Y^n \to (Y^n, s)$. The information density [11]-[13] between the input and output for a given channel state s and a given input distribution $p(x^n)$ is

$$i(x^n; y^n, s) = \ln \frac{p_s(x^n, y^n)}{p(x^n) p_s(y^n)} = i(x^n; y^n|s)$$

where we have used the fact that the input X^n and channel state s are independent of each other.

We will not assume any particular noise or channel distribution so that our results are general and apply to *any* such distribution; channels are allowed to be information-unstable.

III. Capacity of a Given Channel State

In this section, we will assume that a channel state s is given and known to both the Tx and Rx (alternatively, one may assume that the channel state set is a singleton) and review the corresponding results in [8][9] for this setting.

When the channel is information-stable, the normalized information density converges to the mutual information (per symbol) in probability as $n \to \infty$ (due to the law of large numbers) [11]-[13], whose operational meaning is the maximum achievable rate for a given input distribution $p(\boldsymbol{x})$, a channel state s and arbitrary small error probability. Maximizing it over $p(\boldsymbol{x})$ results in the channel capacity. In other cases (information-unstable channels), the normalized information density remains a random variable, even when $n \to \infty$, whose support set is in general an interval [8][9]. Following the analysis in [8], its infimum $\underline{I}(\boldsymbol{X}; \boldsymbol{Y}|s)$ is the largest achievable rate for a given channel state s, input distribution $p(\boldsymbol{x})$ and arbitrary-small error probability:

$$\underline{I}(\boldsymbol{X}; \boldsymbol{Y}|s) = \sup_R \left\{ R : \lim_{n\to\infty} \Pr\left\{ n^{-1} i(X^n; Y^n|s) \le R \right\} = 0 \right\} \tag{1}$$

Following Theorems 2 and 5 in [8], the channel capacity, for a given state s, is obtained by maximizing $\underline{I}(\boldsymbol{X}; \boldsymbol{Y}|s)$ over $p(\boldsymbol{x})$,

$$C(s) = \sup_{p(\boldsymbol{x})} \underline{I}(\boldsymbol{X}; \boldsymbol{Y}|s) \tag{2}$$

Note that this is a very general result, as the channel is not required to be information-stable (ergodic, stationary, etc.). The converse is proved via Verdu-Han Lemma (a lower bound to error probability, which is a dual of Feinstein bound) [8][9]. We definite (n, r_n, ε_{ns})-code in the standard way, where n is the block length, ε_{ns} is the error probability (for channel state s), $r_n = \ln M_n / n$ is the code rate and M_n is the number of codewords.

Lemma 1 (Verdu-Han Lemma [8][9])**.** *Every $(n, r_n, \varepsilon_{ns})$-code satisfies the following inequality,*

$$\varepsilon_{ns} \ge \Pr\left\{ n^{-1} i(X^n; Y^n|s) \le r_n - \gamma \right\} - e^{-\gamma n} \tag{3}$$

for any $\gamma > 0$, where X^n is uniformly distributed over all codewords and Y^n is the corresponding channel output under channel state s.

This is a slight re-wording of Lemma 3.2.2 in [9], where we explicitly indicate channel state s for future use.

On the other hand, the achievability of (2) for a given and known s (i.e. a single, known channel) was proved in [8] via Feinstein Lemma.

Lemma 2 (see e.g. [8][9])**.** *For arbitrary input X^n, any r_n and a given channel state s, there exists a code satisfying the following inequality,*

$$\varepsilon_{ns} \le \Pr\left\{ n^{-1} i(X^n; Y^n|s) \le r_n + \gamma \right\} + e^{-\gamma n} \tag{4}$$

for any $\gamma > 0$.

While this is sufficient to prove achievability for a given and known s, it does not work for the compound channel setting, since we need a code that works for the entire class of channels, not just a single channel as in (4).

IV. Compound Channel Capacity

In this section, we obtain a general formula for compound channel capacity of information-unstable channels by generalizing Lemmas 1 and 2 above to the compound channel setting. This result is more general than the corresponding result established in [9] (Theorem 3.3.5) for finite-state channels, since the former allows arbitrary uncertainty set $\mathcal{S}$. We define an (n, r_n, ε_n)-code for a compound channel in the same way as above, with the compound error probability

$$\varepsilon_n = \sup_s \varepsilon_{ns} \tag{5}$$

and require $\varepsilon_n \to 0$ as $n \to \infty$, which insures arbitrary low error probability for any channel in the uncertainty set for sufficiently large n [1]-[7] [2]. Codebooks are required to be independent of the actual channel state s while the decision regions are allowed to depend on s (since the receiver knows the channel).

It is immediate that the worst-case channel capacity is $\inf_{s\in\mathcal{S}} C(s)$, where $\mathcal{S}$ is the set of possible channel states (uncertainty set), but achieving this requires s to be known to the Tx. If this is not the case, it is far less trivial that the

[2] It is the missing $\sup_s$ in Definition 3.3.1 in [9] that makes Theorem 3.3.5 applicable to finite-state channels only.

compound channel capacity can be obtained by swapping $\sup$ and $\inf$ (see e.g. [7] for an extensive discussion of this issue; while the swapping works in many cases, there are examples when it does not [14]). This is the case for the general (possibly, information-unstable) compound channel considered here, whose capacity is established below.

Theorem 1. *Consider a general compound channel where the channel state $s \in \mathcal{S}$ is known to the receiver but not the transmitter and is independent of the channel input; the transmitter knows the (arbitrary) uncertainty set $\mathcal{S}$. Its compound channel capacity is given by*

$$C_c = \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \tag{6}$$

where $\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \sup_R \{R \in \Omega\}$,

$$\Omega = \left\{ R : \lim_{n\to\infty} \sup_{s\in\mathcal{S}} \Pr\left\{ \frac{1}{n} i(X^n;Y^n|s) \le R \right\} = 0 \right\} \tag{7}$$

Proof. To prove achievability and converse, we generalize Lemmas 1 and 2 above to the compound channel setting.

Lemma 3 (Feinstein Lemma for compound channels)**.** *For arbitrary input X^n and uncertainty set $\mathcal{S}$ and any r_n, there exists a (n, r_n, ε_n)-code (where the codewords are independent of channel state s), satisfying the following inequality,*

$$\varepsilon_n \le \sup_{s\in\mathcal{S}} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le r_n + \gamma \right\} + e^{-\gamma n} \tag{8}$$

for any $\gamma > 0$.

Proof. see Appendix. □

It is clear from the proof that the same inequality holds for both maximum and average error probability. Next, we generalize Verdu-Han Lemma to the compound channel setting.

Lemma 4 (Verdu-Han Lemma for compound channels)**.** *For any uncertainty set $\mathcal{S}$, every (n, r_n, ε_n)-code satisfies the following inequality,*

$$\varepsilon_n \ge \sup_{s\in\mathcal{S}} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le r_n - \gamma \right\} - e^{-\gamma n} \tag{9}$$

for any $\gamma > 0$, where X^n is uniformly distributed over all codewords and Y^n is the corresponding channel output under channel state s.

Proof. To prove this inequality, invoke (3) for a given channel state s and then maximize both sides over all possible channel states to obtain:

$$\begin{aligned} \varepsilon_n &= \sup_s \varepsilon_{ns} \\ &\ge \sup_s \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le r_n - \gamma \right\} - e^{-\gamma n} \end{aligned} \tag{10}$$

A subtle point here is that the original Verdu-Han Lemma allows codewords to depend on channel state while the compound codewords are independent of channel state. Since such a dependence can only decrease error probability, the desired inequality still holds. □

Now, to prove achievability in Theorem 1, fix $p(\boldsymbol{x})$ and set $r_n \le \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) - 2\gamma$ for any $\gamma > 0$. From Lemma 3,

$$\begin{aligned} \lim_{n\to\infty} \varepsilon_n &\le \lim_{n\to\infty} \sup_{s\in\mathcal{S}} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) - \gamma \right\} \\ &= 0 \end{aligned} \tag{11}$$

which shows that $\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) - 2\gamma$ is achievable $\forall \gamma > 0$, so that $C_c \ge \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y})$.

To prove the converse, set $r_n \ge \underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) + 2\gamma$ for any $\gamma > 0$, where $\boldsymbol{X}^*$ is the capacity-achieving input and $\boldsymbol{Y}^*$ is the corresponding output, and use Lemma 4 to obtain

$$\begin{aligned} \lim_{n\to\infty} \varepsilon_n &\ge \lim_{n\to\infty} \sup_{s\in\mathcal{S}} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le \underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) + \gamma \right\} \\ &\ge \varepsilon_0 > 0 \end{aligned} \tag{12}$$

for some fixed $\varepsilon_0 > 0$ (the last two inequalities follow from the definition of $\underline{I}$), so that no rate above $\underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*)$ is achievable.

It is clear from the proof that the same capacity holds under the maximum as well as average error probability. □

Remark 1. *$\underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y})$ is an extension of $\underline{I}(\boldsymbol{X},\boldsymbol{Y}|s)$ to the compound channel setting, not $\inf_s \underline{I}(\boldsymbol{X},\boldsymbol{Y}|s)$, in the general case.*

The relationship between $\underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y})$ and $\inf_s \underline{I}(\boldsymbol{X},\boldsymbol{Y}|s)$ is established below.

Proposition 1. *The following inequality holds for a general compound channel*

$$\underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y}) \le \underline{I}(\boldsymbol{X},\boldsymbol{Y}) = \inf_s \underline{I}(\boldsymbol{X},\boldsymbol{Y}|s) \tag{13}$$

Proof. The proof is by contradiction. Assume that $\underline{\underline{I}} > \underline{I}$, set $R = (\underline{\underline{I}} + \underline{I})/2 > \underline{I}$ and observe that $R < \underline{\underline{I}}$ and

$$\begin{aligned} &\lim_{n\to\infty} \sup_s \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le R \right\} \\ &\quad \ge \sup_s \lim_{n\to\infty} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \le R \right\} \ge \varepsilon_0 > 0 \end{aligned} \tag{14}$$

for some $\varepsilon_0 > 0$ - a contradiction, where the last two inequalities are from the definition of $\underline{I}$. Therefore, $\underline{\underline{I}} \le \underline{I}$. □

It can be demonstrated, via examples (see Example 1 below), that the inequality in (13) can be strict. To see when the equality is achieved, we need the following definition.

Definition 1. *A compound channel is uniform if*

$$\Pr\left\{ n^{-1} i(X^n;Y^n|s) \le \underline{I}(\boldsymbol{X},\boldsymbol{Y}) - \gamma \right\} \to 0 \ \forall \gamma > 0 \tag{15}$$

uniformly in $s \in \mathcal{S}$ as $n \to \infty$.

Note that while the point-wise convergence is insured for each s from the definition of $\underline{I}(\boldsymbol{X},\boldsymbol{Y})$, it does not have to be uniform and, indeed, examples can be constructed where it is not. In a sense, the uniform convergence here insures that the channel does not behave "too badly" as n increases.

For a uniform compound channel, one obtains the following result.

Proposition 2. *The following equality holds for a uniform compound channel*

$$\underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y}) = \underline{I}(\boldsymbol{X},\boldsymbol{Y}) = \inf_s \underline{I}(\boldsymbol{X},\boldsymbol{Y}|s) \tag{16}$$

Proof. We begin with the following Lemma.

Lemma 5. *Let the sequence $f_n(s) \to f(s)$ as $n \to \infty$ and the convergence is uniform. Then,*

$$\lim_{n\to\infty} \sup_s f_s(n) = \sup_s \lim_{n\to\infty} f_s(n) \tag{17}$$

We now show that (13) holds with equality for uniform compound channels. Indeed, set $R = \underline{I}(\boldsymbol{X},\boldsymbol{Y}) - \gamma$, and observe that, for any $\gamma > 0$,

$$\begin{aligned}\lim_{n\to\infty} \sup_s &\Pr\left\{n^{-1} i(X^n;Y^n|s) \le R\right\} \\ &= \sup_s \lim_{n\to\infty} \Pr\left\{n^{-1} i(X^n;Y^n|s) \le R\right\} = 0\end{aligned} \tag{18}$$

from which its follows that $\underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y}) \ge \underline{I}(\boldsymbol{X},\boldsymbol{Y})$. Combining this with (13), one obtains the desired result. □

We are now in a position to establish the capacity of uniform compound channels.

Theorem 2. *Consider a uniform compound channel where the channel state $s \in \mathcal{S}$ is known to the receiver but not the transmitter and is independent of the channel input; the transmitter knows the (arbitrary) uncertainty set $\mathcal{S}$. Its compound channel capacity is given by*

$$C_c = \sup_{p(\boldsymbol{x})} \inf_{s\in\mathcal{S}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) \tag{19}$$

Proof. Using Proposition 2 in Theorem 1 gives (19). □

Note that Theorems 1 and 2 hold for any alphabet and any uncertainty set. In many cases of practical interest (e.g. when the set of feasible input distributions $p(\boldsymbol{x})$ and/or the uncertainty set $\mathcal{S}$ are compact and $\underline{I}(x;y|s)$ is well-behaving), $\sup$ and/or $\inf$ can be substituted by $\max$ and/or $\min$. Unlike [9], the present result applies to arbitrary channel uncertainty sets (not just finite-state) and its proof is direct (i.e. not relying on mixed channels but directly constructing capacity-approaching codes for compound channel).

We remark that many well-known results (e.g. [5]) are special cases of Theorem 1 and 2. The latter is pleasantly similar to many know results for information-stable channels, which also include $\sup - \inf$ expression.

V. Example

To demonstrate the difference between Theorems 1 and 2 and the fact that inequality in (13) can be strict, consider the following binary non-stationary channel with memory:

$$p_s(y^n|x^n) = p_s(y^n) \text{ if } n \le s \tag{20}$$

i.e. the output is independent of the input. If $n > s$, then the channel is n-th extension of BSC with zero cross-over probability, and $\mathcal{S} = \{1, 2, ...\}$. This can model a channel with memory where the noise coherence time $\tau = s$ so that blocklength $n > \tau$ is required to achieve low error probability. Since $i(X^n;Y^n|s) = 0$ if $s \ge n$, it follows that $\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = 0$ while $\underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) = \ln 2 \ \forall s$ under equiprobable input, so that

$$\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = 0 < \underline{I}(\boldsymbol{X};\boldsymbol{Y}) = \inf_s \underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) = \ln 2 \tag{21}$$

and hence

$$C_c = \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = 0 < \ln 2 = \sup_{p(\boldsymbol{x})} \inf_{s\in\mathcal{S}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) \tag{22}$$

The compound capacity C_c is zero because for any blocklength, does not matter how large, there are always channel states with error probability close to 1 so that arbitrary low error probability is not attainable. The standard $\sup - \inf$ expression falls short of the channel capacity in this case because this compound channel is not uniform. It also demonstrates that Theorem 3.3.5 in [9] does not hold for infinite-state channels. Note that if the coherence time becomes bounded, i.e. $\tau = s \le S$, then $C_c = \sup_{p(\boldsymbol{x})} \inf_{s\le S} \underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) = \ln 2$ as one can use sufficiently-long codewords constructed for memoryless BSC (notice also that the channel becomes uniform in this case).

VI. Conclusion

A general formula for compound channel capacity has been established using the information density approach, which does not require the channel to be stationary, ergodic, or information-stable, and which applies to any channel uncertainty set (not only countable or finite-state). An example is provided, which show that finite and infinite-state compound channels can behave very differently.

VII. Acknowledgement

The authors are grateful to S. Verdu and E. Telatar for insightful discussions and suggestions, and A. Lapidoth for valuable comments.

VIII. Appendix: Proof of Lemma 3

Proof. Let us define

$$B_s(x^n) = \{y^n : i(x^n;y^n|s) \ge \ln\alpha\}, \ \alpha = M_n e^{n\gamma}, \tag{23}$$

$$\lambda_n = \sup_{s\in\mathcal{S}} \Pr\{i(X^n;Y^n|s) \le \ln\alpha\} + M_n/\alpha \tag{24}$$

and observe, for future use, that

$$\begin{aligned}1 \ge \Pr\{Y^n \in B_s(x^n)|x^n\} &= \sum_{y^n\in B_s(x^n)} p_s(y^n|x^n) \\ &\overset{(a)}{\ge} \alpha \sum_{y^n\in B_s(x^n)} p_s(y^n) = \alpha P_s(B_s(x^n))\end{aligned}$$

from which it follows that

$$P_s(B_s(x^n)) \le 1/\alpha \ \forall s, x^n, \tag{25}$$

where (a) is from $p_s(y^n|x^n) \ge \alpha p_s(y^n) \ \forall y^n \in B_s(x^n)$ [3].

We use an iterative codebook construction similar to that in [16] but properly extended to the compound channel setting here. First, fix the input distribution $p(\boldsymbol{x})$ and find x^n such that

$$x^n : \ \inf_s P_s(B_s(x^n)|x^n) \ge 1 - \lambda_n \tag{26}$$

[3]while we use summation which applies to discrete alphabets, it is clear that the same argument holds for continuous alphabets using integration/probability measures instead. This also applies to other arguments in the paper.

and use it as codeword 1, $\boldsymbol{u}_1 = x^n$ (note that this codeword is independent of channel state s); set the decision region $D_{1s} = B_s(\boldsymbol{u}_1)$ for this codeword, so that probability of correct decision for this codeword is at least $1 - \lambda_n$.

Next, find $x^n \neq \boldsymbol{u}_1$ such that

$$x^n : \inf_s P_s(B_s(x^n) - D_{1s}|x^n) \geq 1 - \lambda_n \tag{27}$$

and use it as codeword 2, $\boldsymbol{u}_2 = x^n$; set the decision region $D_{2s} = B_s(\boldsymbol{u}_2) - D_{1s}$.

For codeword K, find $x^n \neq \boldsymbol{u}_k, k = 1...K-1$, such that

$$x^n : \inf_s P_s\left(B_s(x^n) - \bigcup_{k=1}^{K-1} D_{ks}|x^n\right) \geq 1 - \lambda_n \tag{28}$$

and set $\boldsymbol{u}_K = x^n$, $D_{Ks} = B_s(\boldsymbol{u}_K) - \bigcup_{k=1}^{K-1} D_{ks}$.

Assume that the process stops at $k = K$, i.e. no further x^n can be found satisfying the required inequality, so that:

$$\inf_s P_s\left(B_s(x^n) - D_s|x^n\right) < 1 - \lambda_n \ \forall x^n \notin \{\boldsymbol{u}_k\}_{k=1}^K, \tag{29}$$

where $D_s = \bigcup_{k=1}^K D_{ks}$. The same inequality also holds for $x^n = \boldsymbol{u}_k$, since

$$B_s(\boldsymbol{u}_k) - D_s = B_s(\boldsymbol{u}_k) - \bigcup_{k=1}^{K} B_s(\boldsymbol{u}_k) = \emptyset \tag{30}$$

The following Lemma shows that a sufficiently large number of codewords can be constructed in this way.

Lemma 6. *The algorithm above generates $K > M_n$ codewords.*

Proof. To see this, observe that it follows from (29) and (30) that there exists such channel state s_0 that

$$P_s\left(B_s(x^n) - D_s|x^n\right) < 1 - \lambda_n \ \forall x^n, s = s_0 \tag{31}$$

For this channel state, one obtains:

$$\begin{aligned}\lambda_n &< 1 - \sum_{x^n} p(x^n) P_{s_0}\left(B_0 \cap D_{s_0}^c|x^n\right) \\ &= 1 - \sum_{x^n} p(x^n)(P_{s_0}\left(B_0|x^n\right) - P_{s_0}\left(B_0 \cap D_{s_0}|x^n\right)) \\ &= P_{s_0}\left(B_0^c(X^n)\right) + \sum_{x^n} p(x^n) P_{s_0}\left(B_0 \cap D_{s_0}|x^n\right) \end{aligned} \tag{32}$$

where $B_0 = B_{s_0}(x^n)$, D_s^c denotes the complement of D_s. Note that the 1st term in (32) is

$$P_{s_0}\left(B_0^c(X^n)\right) = \Pr\left\{i(X^n; Y^n|s_0) < \ln \alpha\right\} \tag{33}$$

and 2nd term t_2 can be upper bounded as follows:

$$\begin{aligned} t_2 &\leq \sum_{x^n} p(x^n) P_{s_0}\left(D_{s_0}|x^n\right) = \sum_{x^n} p(x^n) \sum_{k=1}^{K} P_{s_0}\left(D_{ks_0}|x^n\right) \\ &= \sum_{k=1}^{K} \Pr\left\{Y^n \in D_{ks_0}\right\} \leq \sum_{k=1}^{K} \Pr\left\{Y^n \in B_{s_0}(\boldsymbol{u}_k)\right\} \\ &\leq K/\alpha \end{aligned} \tag{34}$$

where we have used the facts that (i) the sets $\{D_{ks}\}_{k=1}^K$ are non-overlapping and (ii) $D_{ks} \in B_s(\boldsymbol{u}_k)$. The last inequality follows from $\Pr\left(Y^n \in B_s(\boldsymbol{u}_k)\right) \leq 1/\alpha$, which follows from (25). Combining (33) with (34) and using (24), one finally obtains:

$$\lambda_n < \Pr\left\{i(X^n; Y^n|s_0) \leq \ln \alpha\right\} + K/\alpha \tag{35}$$

$$\begin{aligned}\lambda_n &= \sup_{s \in \mathcal{S}} \Pr\left\{i(X^n; Y^n|s) \leq \ln \alpha\right\} + M_n/\alpha \\ &\geq \Pr\left\{i(X^n; Y^n|s_0) \leq \ln \alpha\right\} + M_n/\alpha \end{aligned} \tag{36}$$

from which it follows that $M_n < K$. □

Thus, one can always select M_n codewords using this iterative method. For this codebook, the maximum error probability $\varepsilon_{n,max}$ satisfies

$$\begin{aligned}\varepsilon_{n,max} &= \sup_s \max_k P_s(D_{ks}^c|\boldsymbol{u}_k) = \max_k \sup_s P_s(D_{ks}^c|\boldsymbol{u}_k) \\ &= \max_k (1 - \inf_s P_s(D_{ks}|\boldsymbol{u}_k)) \leq \lambda_n \end{aligned} \tag{37}$$

where $P_s(D_{ks}^c|\boldsymbol{u}_k)$ represents error probability when $\boldsymbol{u}_k$ is transmitted under channel state s and where $\inf_s P_s(D_{ks}|\boldsymbol{u}_k) \geq 1 - \lambda_n$ by code construction. Since $\varepsilon_{n,max} \leq \lambda_n$, so is the average error probability $\varepsilon_n \leq \lambda_n$, from which (8) follows. □

REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, No. 6, pp. 2619-2692, Oct. 1998.

[2] R.L. Dobrushin, "Optimal information Transmission through a channel with unknown parameters," *Radiotekhnika i Electronika*, vol. 4, pp. 1951-1956, 1959.

[3] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, pp. 1229-1241, December 1959.

[4] J. Wolfowitz, "Simultaneous channels," *Arch. Rat. Mech. Anal.*, vol. 4, pp. 371-386, 1960.

[5] W. L. Root, P. P. Varaya, "Capacity of Classes of Gaussian Channels", *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350-1393, Nov. 1968.

[6] I. Csiszar, "Arbitrary varying channels with general alphabets and states", *IEEE Trans. Inform. Theory*, vol. 38, pp. 1725-1742, Nov. 1992.

[7] A. Lapidoth and P. Narayan, "Reliable Communication Under Channel Uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, No. 6, Oct. 1998.

[8] S. Verdu, T.S. Han, "A General Formula for Channel Capacity", *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.

[9] T. S. Han, Information-Spectrum Method in Information Theory, New York: Springer, 2003.

[10] M. Effros, A. Goldsmith, Y. Liang, "Generalizing Capacity: New Definitions and Capacity Theorems for Composite Channels," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3069-3087, July 2010.

[11] R. L. Dobrushin, "A general formulation of the fundamental theorem of Shannon in information theory", Uspekhi Mat. Nauk, v. 14, no. 6(90), Nov.-Dec. 1959, pp.3–104.

[12] M. S. Pinsker, Information and Information Stability of Random Variables and Processes. San Francisco: Holden-Day, 1964.

[13] R.L. Stratonovich, Information Theory, Moscow: Sovetskoe Radio, 1974.

[14] A. Lapitoth and E. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. pp.973-983, May 1998.

[15] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, Wiley, New York, 2006.

[16] R.B. Ash, Information Theory, John Wiley & Sons, 1966.