Communication Under Channel Uncertainty: An Algorithmic Perspective and Effective Construction

Holger Boche, *Fellow, IEEE*, Rafael F. Schaefer, *Senior Member, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—The availability and quality of channel state information heavily influences the performance of wireless communication systems. For perfect channel knowledge, optimal signal processing and coding schemes have been well studied and often closed-form solutions are known. On the other hand, the case of imperfect channel information is less understood and closed-form characterizations of optimal schemes remain unknown in many cases. This paper approaches this question from a fundamental, algorithmic point of view by studying whether or not such optimal schemes can be constructed algorithmically in principle (without putting any constraints on the computational complexity of such algorithms). To this end, the concepts of compound channels and averaged channels are considered as models for channel uncertainty and block fading and it is shown that, although the compound channel and averaged channel themselves are computable channels, the corresponding capacities are not computable in general, i.e., there exists no algorithm (or Turing machine) that takes the channel as an input and computes the corresponding capacity. As an implication of this, it is then shown that for such compound channels, there are no effectively constructible optimal (i.e., capacity-achieving) signal processing and coding schemes possible. This is particularly noteworthy as such schemes must exist (since the capacity is known), but they cannot be effectively, i.e., algorithmically, constructed. Thus, there is a crucial difference between the existence of optimal schemes and their algorithmic constructability. In addition, it is shown that there is no search algorithm that can find the maximal number of messages that can be reliably transmitted for a fixed blocklength. Finally, the case of partial channel knowledge is studied in which either the transmitter or the receiver have perfect channel knowledge while the other part remains uncertain. It is shown that also in the cases of an informed encoder and informed decoder, the capacity remains non-computable in general and, accordingly, optimal signal processing and coding schemes are not effectively constructible.**

***Index Terms*—Communication system, channel uncertainty, compound channel, block fading, Turing computability.**

I. Introduction

FUTURE communication systems beyond fifth generation (5 G) mobile networks will impose strict requirements on reliability, latency, and robustness. The Tactile Internet [2] with its sensitive applications including Industry 4.0 or Vehicle-to-Everything (V2X) communication imposes challenges on the robustness requirement; particularly when the system suffers from channel uncertainty. Such uncertainty easily arises due to the dynamic nature of the wireless medium, but also due to implementation issues such as estimation inaccuracy, finite transmit power for pilot signals, or limited feedback schemes.

The performance of wireless communication systems heavily depends on the availability and quality of channel state information (CSI). For the case of perfect CSI where transmitter and receiver perfectly know the channel, extensive studies have been performed and in certain cases even closed-form solutions have been derived for the optimal signal processing and coding. These studies have had significant impact on the development of wireless communication systems. For example, for multiple-input multiple-output (MIMO) systems with orthogonal frequency division multiplexing (OFDM) with low mobility of users, the optimal signal processing, i.e., beamforming along the channel singular vectors, and optimal coding are known resulting in Gaussian codebooks and the well-known waterfilling solution [3], [4]. For multi-user systems, solutions to the downlink beamforming problem have been derived in [5]. A general duality theory has been derived in [6] which allows to make optimal signal processing schemes for the uplink also applicable in the downlink. This has been subsequently extended to more general multi-user scenarios in [7]–[9]. Robust signal processing techniques for multi-antenna systems has been developed in [10]–[15] which have eventually led to the development of the interference calculus framework [16], [17] used for interference managed network utility optimization.

Due to the dynamic nature of the wireless channel, but also due to implementation issues such as estimation and feedback inaccuracy, practical systems always suffer from channel uncertainty. Accordingly, the provision of accurate CSI is a major challenge making the assumption of perfect CSI unrealistic and imperfect CSI must be taken into account in the system design. There are some specific uncertainty models that have been studied in greater detail, cf. for example [18]. However, for general

Manuscript received December 4, 2019; revised May 31, 2020 and September 8, 2020; accepted September 17, 2020. Date of publication October 6, 2020; date of current version November 6, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. V. Raghavan. The work of Holger Boche was supported in part by the German Federal Ministry of Education and Research (BMBF) within the national initiative for "*Molecular Communication (MAMOKO)*" under Grant 16KIS0914, and in part by the Gottfried Wilhelm Leibniz Prize of the German Research Foundation (DFG) under Grant BO 1734/20-1. The work of Rafael F. Schaefer was supported in part by the BMBF within the national initiative for "*Post Shannon Communication (NewCom)*" under Grant 16KIS1004, and in part by the DFG under Grant SCHA 1944/6-1. The work of H. Vincent Poor was supported by the U.S. National Science Foundation under Grants CCF-0939370 and CCF-1908308. This paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Barcelona, Spain, May 2020 [1]. *(Corresponding author: Rafael Schaefer.)*

Holger Boche is with the Institute of Theoretical Information Technology, Technische Universität München, 80290 Munich, Germany and the Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany (e-mail: boche@tum.de).

Rafael F. Schaefer is with the Information Theory and Applications Chair, Technische Universität Berlin, 10587 Berlin, Germany (e-mail: rafael.schaefer@tu-berlin.de).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Digital Object Identifier 10.1109/TSP.2020.3027902

channels with uncertainty, the situation is completely different. In this case, a general capacity formula has been established, but the optimal signal processing and coding schemes remain unknown in general. Such optimal schemes have been found only for very few specific cases and accordingly, common belief was that it is a hard problem to find the optimal signal processing and coding schemes. Surprisingly, to the best of our knowledge, the issue of why it is actually so hard to find such optimal schemes has not yet been studied.

As in the imperfect CSI case there is little to nothing known with respect to the optimal signal processing and coding and its closed-form solutions, the aim of this work is to study this question from a fundamental algorithmic point of view. We consider the general channel uncertainty models of *compound channels* [19], [20] and *averaged channels* [21], [22] which are introduced in Section II. For compound channels, the actual channel state is unknown. Rather, it is only known that it belongs to a known set of channels (uncertainty set) and that it remains constant for the whole duration of transmission. Averaged channels provide good models for block fading channels. Here, the unknown channel realization remains constant for a certain period of time and then changes according to a known underlying probability distribution that characterizes the fading process. In both cases, the capacity has been established [19], [20] and [21], respectively, but the optimal signal processing and coding schemes remain unknown in general.

To address this issue from a fundamental algorithmic point of view, we use the concept of a *Turing machine* [23]–[25] and the corresponding *computability framework* as introduced in Section III. The Turing machine is a mathematical model of an abstract machine that manipulates symbols on a strip of tape according to certain given rules. It can simulate any given algorithm and therewith provides a simple but very powerful model of computation. Turing machines have no limitations on computational complexity, unlimited computing capacity and storage, and execute programs completely error-free. They are further equivalent to the von Neumann-architecture without hardware limitations and the theory of recursive functions, cf. also [26]–[30]. Accordingly Turing machines provide fundamental performance limits for today's digital computers. Since base stations use digital hardware for signal processing, coding, and resource allocation, they are the ideal concept to study whether or not optimal signal processing and coding schemes can be found algorithmically in principle (without putting any constraints on the computational complexity of such an algorithm).

Communication from a computability or algorithmic point of view has attracted much less attention. In [31] the computability of the capacity functions of the wiretap channel under channel uncertainty and adversarial attacks is studied. The computability of the capacity of finite state channels is studied in [32] and of non-i.i.d. channels in [33]. These works have in common that they analyze the capacity function of various communication scenarios and analyze under which conditions the capacity function is non-computable.

Although the capacity of the compound channel is given in closed-form and single-letter entropic quantities, in Section IV it is then shown that the capacity itself is not computable in general even if the compound channel itself is computable, i.e., there exists no algorithm (or Turing machine) that takes the compound channel as an input and computes the capacity of the channel. As an implication of this, it is subsequently shown that for such compound channels, there is no optimal signal processing and coding scheme that achieves the capacity. This means that although such optimal schemes must exist (since the capacity is known), they cannot be effectively, i.e., algorithmically, constructed. This emphasizes that there is a crucial difference between the existence of such schemes and their algorithmic construction. This study is extended to the problem of maximal code construction. For the case of channel uncertainty it is shown that there is no Turing machine that can compute the maximal number of possible messages that can be reliably (tolerating a certain probability of error) transmitted for a given and fixed block length. This has significant consequences. It is further impossible to construct exhaustive search algorithms that determine the optimal pre- and post-processing; even if the set of all underlying parameters are finite and therewith also the space over which the search is performed. To the best of our knowledge, this is the first time that such a behavior has been observed.

These findings may provide an explanation for why there has been little progress for computer-aided automatic communication system design. This is in contrast to e.g. circuit design or system and control theory, whose development is driven by such computer-aided approaches. On the other hand for communication systems and their signal processing, progress and improvements are due to creative thinking and theoretical analysis.

II. Communication Under Channel Uncertainty

Practical systems always suffer from uncertainty in the channel state information (CSI). This uncertainty about the channel naturally arises due to the dynamic nature of the wireless medium, but also due to implementation issues such as estimation inaccuracy or limited feedback schemes. For example, practical systems always have to use pilots to estimate the communication channel. This estimation will always be imperfect as pilot signals are always limited in their energy and the corresponding estimation can only be performed in a certain signal-to-noise (SNR) regime. Interference, mobility of the users, and the time variance of the channel further decrease the quality of the channel estimate. Accordingly, it is impossible to precisely estimate the actual channel, rather the estimation will always result in a set of infinitely many possible channel realizations. As a result, the optimal signal processing and coding schemes must be robust according to the whole set of possible channels. We consider two models to capture these effects: *Compound channels* [19], [20] and *averaged channels* [21], [22]. These concepts capture the imperfectness of the channel knowledge caused by the nature of the wireless medium and the practical limitations as motivated above.

A. Compound and Averaged Channel Models

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite input and output alphabets and $\mathcal{S}$ an arbitrary state (uncertainty) set. Then for a fixed channel state $s \in \mathcal{S}$, the channel is given by a stochastic matrix $W_s : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ which we interchangeably also write as $W_s \in \mathcal{CH}(\mathcal{X};\mathcal{Y})$. The channel state $s \in \mathcal{S}$ is assumed to remain constant throughout the whole transmission so that the discrete memoryless channel

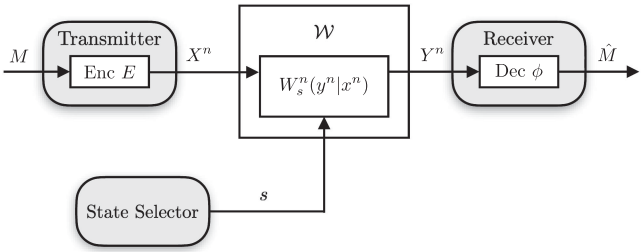


Fig. 1. Communication under channel uncertainty, where the channel state $s \in \mathcal{S}$ (which may even come from a malevolent adversary or jammer) is unknown to the transmitter and the receiver.

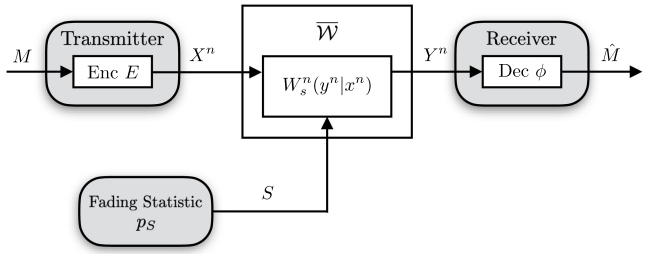


Fig. 2. Communication over block fading channels, where the fading state $s \in \mathcal{S}$ follows an underlying probability distribution $p_S \in \mathcal{P}(\mathcal{S})$ which is known to the transmitter and receiver. The unknown channel realization remains constant for n channel uses.

is given by $W_s(y^n|x^n) := \prod_{i=1}^n W_s(y_i|x_i)$ for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$.

Definition 1: The *compound channel*

$$\mathcal{W} := \{W_s \in \mathcal{CH}(\mathcal{X};\mathcal{Y}) : s \in \mathcal{S}\}$$

is given by the collection of all channels $W_s \in \mathcal{CH}(\mathcal{X};\mathcal{Y})$ for all possible channel states $s \in \mathcal{S}$. The set of all compound channels is denoted by $\mathcal{CC}(\mathcal{X},\mathcal{S};\mathcal{Y})$.

Communication under channel uncertainty (or imperfect CSI) is perfectly captured by the concept of compound channels. This is shown in Fig. 1.

Remark 1: The state set $\mathcal{S}$ is also known as uncertainty set and models how uncertain transmitter and receiver are about the channel. The state set $\mathcal{S}$ and therewith also the corresponding set $\mathcal{W}$ are known to the transmitter and receiver so that they always have some partial channel knowledge (as $\mathcal{W}$ is a strict subset of the set of all possible channels). The actual structure of the uncertainty set depends on several parameters such as pilot signal design, energy, mobility, interference, and many others. Note that in practical systems the channel cannot be known perfectly and there is always some uncertainty. It is important to understand how the uncertainty set affects the optimal signal processing and coding.

Remark 2: Note that the state set is in general not finite for many practical communication scenarios.

While a compound channel is solely defined by its state (uncertainty) set $\mathcal{S}$, for averaged channels, the transmitter and receiver know the uncertainty set $\mathcal{S}$ and the probability distribution $p_S \in \mathcal{P}(\mathcal{S})$, i.e., they know the probability $p_S(s)$ that channel realization W_s, $s \in \mathcal{S}$, governs the transmission. We will see later that the knowledge about p_S is used for the optimal encoder and decoder design, cf. (2) and (3).

Definition 2: The *averaged channel*

$$\overline{\mathcal{W}} := \{\mathcal{W} \in \mathcal{CC}(\mathcal{X},\mathcal{S};\mathcal{Y}), p_S \in \mathcal{P}(\mathcal{S})\}$$

is given by its corresponding compound channel $\mathcal{W} \in \mathcal{CC}(\mathcal{X},\mathcal{S};\mathcal{Y})$ and an additional probability distribution $p_S \in \mathcal{P}(\mathcal{S})$ on the state set $\mathcal{S}$. The set of all averaged channels is denoted by $\mathcal{AC}(\mathcal{X},\mathcal{S};\mathcal{Y})$.

This concept perfectly models block fading channels for which the unknown fading channel remains constant for a number of channel uses before changing again according to $p_S \in \mathcal{P}(\mathcal{S})$. Communication over averaged channels is shown in Fig. 2.

B. Codes and Performance for Compound Channels

Since the actual channel state is unknown to the transmitter and receiver, universal encoder and decoder are needed that are independent of the channel state.

Definition 3: An (n, M_n)-*code* for the compound channel consists of an encoder $E_n : \mathcal{M}_n \to \mathcal{X}^n$ at the transmitter with a set of messages $\mathcal{M}_n := \{1, \ldots, M_n\}$ and a decoder $\phi_n : \mathcal{Y}^n \to \mathcal{M}_n$ at the receiver.

As the receiver needs to decode the transmitted message for all possible channel realizations, we require the *maximum error criterion*

$$e_{\max,n} = \sup_{s\in\mathcal{S}} \max_{m\in\mathcal{M}_n} \sum_{y^n:\phi_n(y^n)\neq m} W_s^n(y^n|x_m^n) \leq \epsilon_n \qquad (1)$$

for $\epsilon_n > 0$ with $x_m^n = E_n(m)$ the codeword for message $m \in \mathcal{M}_n$.

Remark 3: The maximum error criterion is considered, but all results derived in this paper also hold for the average error criterion

$$\bar{e}_n = \sup_{s\in\mathcal{S}} \frac{1}{|\mathcal{M}_n|} \sum_{m\in\mathcal{M}_n} \sum_{y^n:\phi_n(y^n)\neq m} W_s^n(y^n|x_m^n) \leq \epsilon_n.$$

However, the maximum error criterion in (1) is stronger and particularly relevant for applications with very strict performance requirements such as Industry 4.0.

Definition 4: A rate $R > 0$ is called *achievable* for the compound channel $\mathcal{W}$ if there exists a sequence of (n, M_n)-codes such that we have $\frac{1}{n}\log M_n \geq R$ and $e_{\max,n} \leq \epsilon_n$ (or $\bar{e}_n \leq \epsilon_n$ respectively) with $\epsilon_n \to 0$ as $n \to \infty$. The *capacity* $C(\mathcal{W})$ is given by the supremum of all achievable rates R.

The capacity of the compound channel has been established in [19], [20] and displays a simple and heuristically expected structure.

Theorem 1 ([19], [20]): The capacity $C(\mathcal{W})$ of the compound channel $\mathcal{W}$ is

$$C(\mathcal{W}) = \sup_{p\in\mathcal{P}(\mathcal{X})} \inf_{s\in\mathcal{S}} I(p, W_s)$$

with $I(p, W_s)$ denoting the mutual information for input distribution $p \in \mathcal{P}(\mathcal{X})$ and channel $W_s \in \mathcal{CH}(\mathcal{X};\mathcal{Y})$.

C. Codes and Performance for Averaged Channels

Similar to compound channels, the state set $\mathcal{S}$ is known but the actual channel realization is unknown to the transmitter and

receiver in the case of averaged channels. However, they are further aware of the underlying probability distribution according to which the channel is changing, i.e., they know the probability $p_S \in \mathcal{P}(\mathcal{S})$. As the actual channel state is unknown, the encoder and decoder must be independent of it but can depend on the probability distribution p_S. In particular, all channel realizations $W_{\hat{s}}$ with

$$p_S(\hat{s}) = 0 \tag{2}$$

need not be considered in the design as these channel realizations do not appear and, accordingly, do not affect the error performance criterion (3).

Definition 5: An (n, M_n)-*code* for the averaged channel consists of an encoder $E_n : \mathcal{M}_n \to \mathcal{X}^n$ at the transmitter with a set of messages $\mathcal{M}_n := \{1, \ldots, M_n\}$ and a decoder $\phi_n : \mathcal{Y}^n \to \mathcal{M}_n$ at the receiver.

The maximum error criterion is then given as

$$e_{\max,n} = \max_{m \in \mathcal{M}_n} \sum_{s \in \mathcal{S}} \sum_{y^n : \phi_n(y^n) \neq m} W_s^n(y^n | x_m^n) p_S(s) \leq \epsilon_n \tag{3}$$

for $\epsilon_n > 0$ with $x_m^n = E_n(m)$ the codeword for message $m \in \mathcal{M}_n$. The definitions of an achievable rate and of the capacity $C(\overline{\mathcal{W}})$ of the averaged channel $\overline{\mathcal{W}}$ follow accordingly as in Definition 4.

The capacity of the averaged channel has been established in [21] and is given in the following theorem.

Theorem 2 ([21]): The capacity $C(\overline{\mathcal{W}})$ of the averaged channel $\overline{\mathcal{W}}$ is

$$C(\overline{\mathcal{W}}) = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{s \in \mathcal{S}} I(p, W_s).$$

We see that the capacity of the averaged channel $\overline{\mathcal{W}}$ equals the capacity of the corresponding compound channel $\mathcal{W}$, i.e., $C(\overline{\mathcal{W}}) = C(\mathcal{W})$. An interesting observation here is that the additional knowledge of a probability distribution for the channel state does increase the capacity compared to compound channels where the channel state is chosen arbitrarily. The reason behind this is that the encoder and decoder must be designed to work for all possible states regardless of their probability of occurrence.

Theorems 1 and 2 show that the capacity of the compound channel averaged channel are *analytically* well understood as they provide a closed-form single letter entropic expression for the capacity function. But surprisingly, it is not much known about the optimal signal processing and how this expression can be computed *algorithmically* on digital computers. Accordingly, in the remainder of this paper, we study the structure of the capacity and the algorithmic computability of optimal strategies.

III. Computability Framework

Here, we introduce the computability framework based on Turing machines which provides the needed background. For this we need some basic definitions and concepts of computability which are briefly reviewed. The concept of computability and computable real numbers was first introduced by Turing in [23] and [24].

A sequence of rational numbers $\{r_n\}_{n \in \mathbb{N}}$ is called a *computable sequence* if there exist recursive functions $a, b, s : \mathbb{N} \to \mathbb{N}$ with $b(n) \neq 0$ for all $n \in \mathbb{N}$ and

$$r_n = (-1)^{s(n)} \frac{a(n)}{b(n)}, \qquad n \in \mathbb{N}, \tag{4}$$

cf. [34, Def. 2.1 and 2.2] for a detailed treatment. A real number x is said to be computable if there exists a computable sequence of rational numbers $\{r_n\}_{n \in \mathbb{N}}$ such that

$$|x - r_n| < 2^{-n} \tag{5}$$

for all $n \in \mathbb{N}$. We denote the set of computable real numbers by $\mathbb{R}_c$. Based on this, we define the set of computable probability distributions $\mathcal{P}_c(\mathcal{X})$ as the set of all probability distributions $P_X \in \mathcal{P}(\mathcal{X})$ such that $P_X(x) \in \mathbb{R}_c$, $x \in \mathcal{X}$. Further, let $\mathcal{CH}_c(\mathcal{X}; \mathcal{Y})$ be the set of all computable channels, i.e., for a channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ we have $W(\cdot|x) \in \mathcal{P}_c(\mathcal{Y})$ for every $x \in \mathcal{X}$. Finally, computable compound channels and computable averaged channels are defined as follows.

Definition 6: A compound channel $\mathcal{W} = \{W_s \in \mathcal{CH}_c(\mathcal{X}; \mathcal{Y}) : s \in \mathcal{S}\}$ is said to be *computable* if there is a recursive function $\varphi : \mathcal{S} \to \mathcal{CH}_c(\mathcal{X}; \mathcal{Y})$ with $\varphi(s) = W_s$ for all $s \in \mathcal{S}$. The set of all computable compound channels is denoted by $\mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$.

Definition 7: An averaged channel $\overline{\mathcal{W}} = \{\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}; \mathcal{Y}), p_S \in \mathcal{P}(\mathcal{S})\}$ is said to be *computable* if the corresponding compound channel $\mathcal{W}$ is computable according to Definition 6 and p_S is a computable probability distribution. The set of all computable averaged channels is denoted by $\mathcal{AC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$.

In particular, this means we require that the compound set $\mathcal{W}$ is algorithmically constructible, i.e., for every state $s \in \mathcal{S}$ the channel W_s can be constructed by an algorithm with input s.

We further need the concepts of a recursive set and a recursively enumerable set as, for example, defined in [34].

Definition 8: A set $\mathcal{A} \subset \mathbb{N}$ is called *recursive* if there exists a computable function f such that $f(x) = 1$ if $x \in \mathcal{A}$ and $f(x) = 0$ if $x \notin \mathcal{A}$.

Computable functions $f : \mathbb{N} \to \mathbb{N}$ are recursive functions [35]. Thus, computable functions are exactly those functions that are computable by a Turing machine.

Definition 9: A set $\mathcal{A} \subset \mathbb{N}$ is *recursively enumerable* if there exists a recursive function whose domain is exactly $\mathcal{A}$.

We have the following properties which will be crucial later for proving the desired results; cf. also [34] for further details.

- $\mathcal{A}$ is recursive is equivalent to $\mathcal{A}$ is recursively enumerable and $\mathcal{A}^c$ is recursively enumerable.
- There exist recursively enumerable sets $\mathcal{A} \subset \mathbb{N}$ that are not recursive, i.e., $\mathcal{A}^c$ is not recursively enumerable. This means there are no computable, i.e., recursive, functions $f : \mathbb{N} \to \mathcal{A}^c$ with $[f(\mathbb{N})] = \mathcal{A}^c$.

Turing machines are extremely powerful compared to state-of-the-art digital signal processing (DSP) and field gate programmable array (FPGA) platforms and even current supercomputers. It is the most general computing model and is even capable of performing arbitrary exhaustive search tasks on arbitrary large but finite structures. The complexity can even grow faster than double-exponentially with the set of parameters of the underlying communication system (such as time, frequencies, transmit power, modulation scheme, number of transmit and receive antennas, etc.). We will discuss this in greater detail in Section IV-E, where we study search algorithms that compute the optimal pre- and post-processing for the cases of perfect CSI and imperfect CSI.

IV. Channel Uncertainty and Signal Processing Performance

In this section, we study the case, where the actual channel state is unknown to both transmitter and receiver. We first present a detailed analysis for compound channels and then discuss how these results extend to averaged channels as well.

A. Capacity Estimation for Compound Channels

We start with studying the algorithmic computability of the capacity of compound channels in greater detail. In particular, for a computable compound channel $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ we ask whether or not the capacity $C(\mathcal{W})$ can be algorithmically computed. The following result establishes a negative answer to this question. Subsequently, we discuss the implications and consequences for the optimal signal processing and coding schemes.

Theorem 3: Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary finite alphabets. Then there is a computable compound channel $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ such that

$$C(\mathcal{W}) \notin \mathbb{R}_c.$$

Proof: Before we prove the result, we present a brief outline of it. We start with an arbitrary recursively enumerable set $\mathcal{A} \subset \mathbb{N}$ that is not recursive. This set $\mathcal{A}$ is used to describe a computable compound channel which will consist of a computable sequence of binary symmetric channels. For this particular channel, we show that its capacity can be expressed by a binary entropy function $h_2(\cdot)$ which needs to be evaluated at a non-computable real number μ_*. With the help of Lemma 1 we then show the binary entropy $h_2(\mu_*)$ of a non-computable real number μ_* itself is a non-computable real number. This shows that the capacity of a computable compound channel is a non-computable real number.

Now, we turn to the actual proof which we carry out for $|\mathcal{X}| = |\mathcal{Y}| = 2$. The result for the general case $|\mathcal{X}| \geq 2, |\mathcal{Y}| \geq 2$ follows then immediately.

We consider the binary symmetric channel (BSC) given by

$$W(y|1,\mu) = \begin{pmatrix} 1-\mu \\ \mu \end{pmatrix}, \qquad W(y|2,\mu) = \begin{pmatrix} \mu \\ 1-\mu \end{pmatrix} \tag{6}$$

for some $\mu \in (0,1)$.

Let $\mathcal{A} \subset \mathbb{N}$ be a recursively enumerable set that is not recursive (i.e., $\mathcal{A}^c$ is not recursively enumerable) and $\varphi_{\mathcal{A}} : \mathbb{N} \to \mathcal{A}$ be a computable function with $\varphi_{\mathcal{A}}[\mathbb{N}] = \mathcal{A}$ and for every $k \in \mathcal{A}$ there exists one $s_k \in \mathbb{N}$ with $\varphi_{\mathcal{A}}(s_k) = k$ [36].

We choose

$$\mu_s = \sum_{l=1}^{s} \frac{1}{2^{\varphi_{\mathcal{A}}(l)+2}} \tag{7}$$

and observe that $\{\mu_s\}_{s=1}^{\infty}$ is a computable sequence of monotonically increasing rational numbers with

$$\begin{aligned} \mu_s &= \frac{1}{4}\sum_{l=1}^{s} \frac{1}{2^{\varphi_{\mathcal{A}}(l)}} < \frac{1}{4}\sum_{l=1}^{\infty} \frac{1}{2^{\varphi_{\mathcal{A}}(l)}} < \frac{1}{4}\sum_{l=1}^{\infty} \frac{1}{2^{l}} \\ &= \frac{1}{8}\sum_{l=0}^{\infty} \frac{1}{2^{l}} = \frac{1}{8}\frac{1}{1-\frac{1}{2}} = \frac{1}{4}. \end{aligned} \tag{8}$$

We now consider the computable compound channel $\mathcal{W} = \{W_s : s \in \mathbb{N}\}$ with

$$W_s(y|x) := W(y|x,\mu_s).$$

We have

$$\lim_{s\to\infty} \mu_s = \mu_*, \tag{9}$$

with $\mu_* \notin \mathbb{R}_c$ and $\mu_* < \frac{1}{4}$. Further, we have for every $s \in \mathbb{N}$

$$C(W_s) = \max_{p\in\mathcal{P}(\mathcal{X})} I(p, W_s) = I(p_*, W_s)$$

with $p_*(1) = p_*(2) = \frac{1}{2}$ since the channel W_s is symmetric. Accordingly, the capacity of the BSC W_s is

$$C(W_s) = 1 - h_2(\mu_s)$$

with $h_2(\cdot)$ the binary entropy function.

With this, we obtain for the compound channel $\mathcal{W} = \{W_n : n \in \mathbb{N}\}$

$$\begin{aligned} C(\mathcal{W}) &= \max_{p\in\mathcal{P}(\mathcal{X})} \inf_{s\in\mathbb{N}} I(p, W_s) \\ &= \inf_{s\in\mathbb{N}} I(p_*, W_s) = 1 - \sup_{s\in\mathbb{N}} h_2(\mu_s) = 1 - h_2(\mu_*) \end{aligned}$$

since $h_2(\mu_n) < h_2(\mu_{n+1})$. Note that $h_2(\cdot)$ is a continuous function.

Finally, it remains to show that, since $\mu_* \notin \mathbb{R}_c$, we also have $h_2(\mu_*) \notin \mathbb{R}_c$ so that $C(\mathcal{W}) \notin \mathbb{R}_c$ is shown. This is not straightforward due to the following observation: The binary entropy function $h_2(\mu_*) = -\mu_* \log \mu_* - (1-\mu_*)\log(1-\mu_*)$ is the sum of two terms and even if both of these are non-computable, we cannot easily conclude that the sum is non-computable as well. This implies that a Taylor expansion approach for h_2 does not allow to conclude that if $\mu \notin \mathbb{R}_c$ we also have $h_2(\mu) \notin \mathbb{R}_c$. Rather, we need to invoke a Newton fixed-point iteration argument to be able to conclude that for the above $\mu_* \notin \mathbb{R}_c$ we do have $h_2(\mu_*) \notin \mathbb{R}_c$. The details are delegated to the following lemma.

Lemma 1: For $\mu_* \notin \mathbb{R}_c$ from our construction above with $\lim_{s\to\infty} \mu_s = \mu_*$ and $\mu_s = \sum_{l=1}^{s} \frac{1}{2^{\varphi_{\mathcal{A}}(l)+2}}$ so that $\mu_* < \frac{1}{4}$, cf. (7) and (9), we have

$$h_2(\mu_*) \notin \mathbb{R}_c.$$

Proof: The proof is given in the Appendix A. ■

This completes the proof and shows that the capacity is a non-computable number. ■

Remark 4: The number μ_* is not a computable real number which can be shown as follows by contradiction: To do so, we assume that μ_* would be computable. Then, μ_* is of the form

$$\mu_* = \sum_{l=1}^{\infty} \frac{1}{2^{\varphi_{\mathcal{A}}(l)+2}}.$$

Based on an algorithm for the computation of μ_*, we would then be able to obtain an algorithm that decides for every natural number $k \in \mathbb{N}$ if $k \in \mathcal{A}$ or $k \in \mathcal{A}^c = \mathbb{N}\backslash\mathcal{A}$. Then, both $\mathcal{A}$ and $\mathcal{A}^c$ would be recursively enumerable and therewith $\mathcal{A}$ would be recursive. But this is a contradiction to the initial assumption since we assumed that the set $\mathcal{A}$ is recursively enumerable and but not recursive.

Some remarks are in order.

1) For $C(\mathcal{W}) \notin \mathbb{R}_c$, it is also not possible to algorithmically compute $C(\mathcal{W})$ although the compound channel itself is computable, i.e., $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$.

2) Theorem 3 does not directly imply that optimal (and almost optimal, respectively) signal processing strategies cannot be algorithmically computed.
3) In the definition of achievability, i.e., Definition 4, only the *existence* of an encoder and decoder is required, but not how these can actually be found and constructed.

In particular, the last question will be addressed in the next subsection.

B. Optimal Signal Processing and Coding for Compound Channels

In this section, we show that optimal signal processing and coding schemes exist for certain compound channels, but that these cannot be constructed algorithmically, i.e., by a computer-aided design. This emphasizes that there is a considerable difference between the existence of optimal schemes and their actual algorithmic construction. As a consequence, it is in general not possible to deduce algorithms for the construction of optimal schemes from the proof of their existence.

Before we start with the actual analysis, we present and discuss two examples to demonstrate that certain relevant objects can exist, but that these need not be algorithmically constructible. To show the existence of certain objects, in mathematics there are various non-constructive approaches which becomes clear particularly in the following example.

Example 1 (Specker Sequence): We start with an arbitrary computable sequence $\{a_k\}_{k\in\mathbb{N}}$ of rational numbers with $a_k < a_{k+1}$, $k \in \mathbb{N}$, and $0 \leq a_k \leq 1$. It is clear that there exists exactly one real number $a_* \in [0,1]$ such that $\lim_{k\to\infty} a_k = a_*$ holds and the sequence $\{a_k\}_{k\in\mathbb{N}}$ converges monotonically increasingly to a_*. This means the approximation error $a_* - a_k$, $k \in \mathbb{N}$, is monotonically decreasing with k and converges to zero.

Already in 1949, Specker constructed such a sequence $\{a_k\}_{k\in\mathbb{N}}$ for which the limit a_* is not a computable real number [37]. This number a_* exists but is a transcendental number, i.e., it has a unique binary representation

$$a_* = \sum_{n=1}^{\infty} b_n \frac{1}{2^n}$$

with $b_n \in \{0,1\}$. However, there exists no algorithm that can compute the coefficients b_n, $n \in \mathbb{N}$. The same result is also true for the decimal representation of the number a_*. This example has the following consequences. Whenever it is proved that every uniformly bounded, monotonically increasing sequence of real numbers has a limit, then this proof shows only the existence of this limit. Specker's example above demonstrates that it is possible that this limit cannot be algorithmically computed, i.e., there is no algorithm that takes the sequence $\{a_k\}_{k\in\mathbb{N}}$ as input and outputs for any given n of the binary representation of a_* the corresponding coefficients $b_n(a_*)$. Accordingly, such a Turing machine as shown in Fig. 3 cannot exist.

As the limit a_* cannot be constructed algorithmically, every proof for the existence of the limit a_* must have a non-constructive component as for example the following:

Since the computable sequence $\{a_k\}_{k\in\mathbb{N}}$ is strict monotonically increasing, for every $l \in \mathbb{N}$ there must exist an interval $(\frac{r}{2^l}, \frac{r+1}{2^l}]$, $0 \leq r \leq 2^l - 1$, such that $\{a_k\}_{k\in\mathbb{N}}$ and

$$\left(\frac{r_l}{2^l}, \frac{r_l+1}{2^l}\right] \tag{10}$$

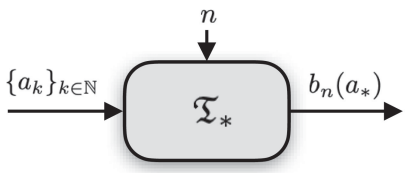


Fig. 3. Turing machine $\mathfrak{T}_*$ for computation of the binary representation of a_*. It receives the monotonically increasing sequence $\{a_k\}_{k\in\mathbb{N}}$ that converges to a_* as input and outputs the coefficient $b_n(a_*)$ for specified n.

for some r_l have an infinite intersection. This implies that $a_* \in (\frac{r_l}{2^l}, \frac{r_l+1}{2^l}]$ and therewith

$$\lim_{l\to\infty} \frac{r_l}{2^l} = a_* = \lim_{l\to\infty} \frac{r_l+1}{2^l}.$$

However, there is no Turing machine, i.e., algorithm, that takes $l \in \mathbb{N}$ as an input and outputs the number $r_l \in \mathbb{N}$ so that (10) is true. The sequence $\{\mu_s\}_{s\in\mathbb{N}}$ in (7) for example provides such a sequence that satisfies all conditions as imposed above on the Specker sequence.

The argument that every monotonically increasing but bounded sequence converges to a finite limit is frequently used in the area of signal processing. It appears for example in the iterative maximization of utility functions in wireless communication systems, cf. for example [38] and references therein. It is used to show that iterative approaches usually converge to (local) maxima. As argued above in Example 1 however, it is not clear in general whether or not such limits are computable real numbers. The same is also true for the iterative minimization of loss functions. The Specker sequence can be used here as well.

Example 2 (Fekete Lemma): The result of Fekete [39] plays an important role in information theory for characterizing the capacity; see for example [40]–[42].

Let $\{d_n\}_{n\in\mathbb{N}}$ be a monotonically increasing sequence of computable real numbers with

$$d_{m+n} \geq d_m + d_n, \tag{11}$$

$m, n \in \mathbb{N}$. Then, Fekete's Lemma ensures that the computable sequence $\{\frac{d_n}{n}\}_{n\in\mathbb{N}}$ converges to the limit C_* with

$$C_* = \lim_{n\to\infty} \frac{d_n}{n} = \sup_{n\in\mathbb{N}} \frac{d_n}{n}. \tag{12}$$

As we will see next, the limit C_* must not be a computable real number so that Fekete's Lemma shows only the existence of it. Similarly to the first example above, it is in general impossible to compute the binary representation of C_* based on the computable sequence $\{\frac{d_n}{n}\}_{n\in\mathbb{N}}$ and similarly there is no Turing machine as in Fig. 3 that computes the desired coefficients.

Before we prove this, we want to note that relation (11) immediately shows why Fekete's Lemma is relevant for determining the performance of optimal signal processing and coding schemes. If we have such optimal schemes for certain communication tasks for blocklength $m+n$, then these schemes should be superior to comparable schemes individually applied to blocklength m and blocklength n, i.e., for example first for the time duration $[0, m-1]$ and subsequently for $[m, m+n]$.

We can show that the sequences from the first example above immediately imply (12) as following: We take a Specker sequence $\{a_k\}_{k\in\mathbb{N}}$ and set $d_k = k \cdot a_k$, $k \in \mathbb{N}$. Then, $\{d_k\}_{k\in\mathbb{N}}$ is a computable sequence of computable real numbers and for

$m, n \in \mathbb{N}$ arbitrary, we have

$$d_{m+n} = (m+n)a_{m+n} = m \cdot a_{m+n} + n \cdot a_{m+n} \geq m \cdot a_m + n \cdot a_n.$$

Accordingly, $C_* = a_* \notin \mathbb{R}_c$ for the Specker sequence from the example above.

For the analysis of the constructability of optimal signal processing and coding schemes, we introduce some needed concepts.

Definition 10: An asymptotically optimal signal processing and coding scheme is said to be *effectively constructible* for the compound channel $\mathcal{W}$ if there is an algorithm $\mathcal{A}$ that computes for given error $\epsilon \in (0,1)$, $\epsilon \in \mathbb{R}_c$, and for every $m \in \mathbb{N}$ an $(n(\epsilon,m), M_n(\epsilon,m), \epsilon)$-code such that

$$\left|C(\mathcal{W}) - \frac{1}{n(\epsilon,m)} \log M_n(\epsilon,m)\right| < \frac{1}{m}, \tag{13}$$

i.e., the rate of the code is close to the capacity of the compound channel.

For communication systems with perfect channel state information at the transmitter and receiver, the optimal signal processing and coding schemes are, in general, effectively constructible (We will explicitly show this for the exhaustive search approach in Section IV-E). Often, there are even closed form solutions known. However, there will be always uncertainty in the channel state information and the optimal signal processing and coding is not fully explored and known. In this case, Definition 10 is crucial which means that if an algorithm $\mathcal{A}$ computes in the m-th step an $(n(\epsilon,m), M_n(\epsilon,m), \epsilon)$-code, then it computes an encoder $E_n : \mathcal{M}_n \to \mathcal{X}^n$ and decoder $\phi_n : \mathcal{Y}^n \to \mathcal{M}_n$ with $\mathcal{M}_n = \{1, \ldots, M_n(\epsilon,m)\}$ and $\bar{e}_n \leq \epsilon$. Thus, Definition 10 requires the algorithm to recursively depend on the channel and the given error. We obtain the following result.

Corollary 1: For the compound channel $\mathcal{W}$ of Theorem 3, the optimal signal processing and coding scheme are not effectively constructible.

Proof: Assume the optimal signal processing and coding scheme would be effectively constructible. Then (13) would be satisfied for all $m \in \mathbb{N}$. Then,

$$\zeta_n = \frac{1}{n(\epsilon,m)} \log M_n(\epsilon,m), \qquad m \in \mathbb{N}$$

is a computable sequence of computable reals that converges effectively to $C(\mathcal{W})$ due to (13). This implies that $C(\mathcal{W}) \in \mathbb{R}_c$. But this is a contradiction to Thorem 3 which proves the desired result. ■

We can further weaken the requirement of an effectively constructible code as follows.

Definition 11: An asymptotically optimal signal processing and coding scheme is said to be *weakly effectively constructible* for the compound channel $\mathcal{W}$ if there exists a $\epsilon \in (0,1)$, $\lambda \in \mathbb{R}_c$ and an algorithm $\mathcal{A}$ that computes for every $m \in \mathbb{N}$ an $(n(\epsilon,m), M_n(\epsilon,m), \lambda)$-code such that

$$\frac{1}{n(\epsilon,m)} \log M_n(\epsilon,m) \leq \frac{1}{n(\epsilon,m+1)} \log M_n(\epsilon,m+1) \tag{14}$$

and

$$\lim_{n\to\infty} \frac{1}{n(\epsilon,m)} \log M_n(\epsilon,m) = C(\mathcal{W}). \tag{15}$$

Some remarks are in order.

1) The concepts of effectively constructible in Definition 10 and weakly effectively constructible in Definition 11 differ in the following way: For the latter, the algorithm takes an error $\epsilon \in (0,1)$, $\epsilon \in \mathbb{R}_c$, as an input, i.e., the construction of the optimal signal processing and coding scheme is done for one specific error and not for any error $\epsilon \in (0,1), \epsilon \in \mathbb{R}_c$ as in the former case. Furthermore, effectively constructible requires that the rate of the code is guaranteed to be close to the capacity, cf. (13). On the other hand, weakly effectively constructible requires only that the rate of the code converges to the capacity.
2) We do not require that the algorithm of Definition 11 depends recursively on the channel.

Theorem 4: For the compound channel $\mathcal{W}$ of Theorem 3, the optimal signal processing and coding scheme is not weakly effectively constructible for any $\epsilon \in (0,1)$, $\epsilon \in \mathbb{R}_c$.

Proof: Assume that there exists a $\epsilon \in (0,1)$, $\epsilon \in \mathbb{R}_c$, for which such an algorithm $\mathcal{A}$ exists. Then

$$\zeta_m = \frac{1}{n(m)} \log M_n(m), \quad m \in \mathbb{N}$$

is a computable sequence of computable reals. We have $\zeta_m \leq \zeta_{m+1}$ and further

$$\overline{C}_N = \max_{p \in \mathcal{P}(\mathcal{X})} \min_{1 \leq s \leq N} I(p, W_s).$$

We observe that $\min_{1\leq s\leq N} I(p,W_s)$ is a computable function with respect to $p \in \mathcal{P}(\mathcal{X})$ since all W_n, $1 \leq n \leq N$, are computable channels. Therewith, $\overline{C}_N$ is a computable real and $\{\overline{C}_N\}_{N\in\mathbb{N}}$, a computable sequence of computable reals and we have have $\overline{C}_{N+1} \leq \overline{C}_N$.

This means that $C(\mathcal{W})$ is the limit of two computable sequences of computable reals, where one sequence is monotonically decreasing and the other is monotonically increasing. This implies that we must have $C(\mathcal{W}) \in \mathbb{R}_c$ which is a contradiction to Theorem 3. Accordingly, our initial assumption is wrong and the optimal signal processing and coding scheme is not weakly effectively constructible. ■

Remark 5: Recently, significant progress has been made in effective code constructions. For example for fixed and given computable binary symmetric channel, algorithms have been proposed that construct capacity-achieving polar codes. Now Theorem 4 shows that, in general, such algorithms cannot exist for computable compound channels.

It is interesting to compare Theorem 4 with the state-of-the-art design of optimal robust, i.e., for channel uncertainty or fading channels, signal processing schemes. To the best of our knowledge, there are no computer-aided automatic design approaches as, for example, in the area of circuit design, where this is a common approach. New approaches, in particular coding schemes, are obtained in a creative process and, accordingly, cannot be captured by Turing machines. Theorem 4 now shows that even for computable compound channels and computable averaged channels such an approach, i.e., a computer-aided automatic design approach, is not possible. In Section IV-E we will further show that even an exhaustive search is not possible to determine the optimal pre- and post-processing. This is in contrast to the case of perfect channel knowledge, where it is no problem to use Turing machines for an exhaustive search.

Theorem 4 further yields immediately the corresponding result for the averaged channel and therewith for block fading channels.

Corollary 2: For the compound channel $\mathcal{W}$ of Theorem 3 and for all fading statistics $p_S \in \mathcal{P}(\mathcal{S})$ with $p_S(s) > 0$, $s \in \mathcal{S}$, the optimal signal processing and coding scheme for the averaged channel $\overline{\mathcal{W}} = \{\mathcal{W}, p_S\}$ is not weakly effectively constructible for any $\epsilon \in (0,1)$, $\epsilon \in \mathbb{R}_c$.

It is interesting to observe that the result above is true for all possible fading statistics. This implies that the capacities have the same numeric value. Accordingly, the same conclusions and discussions for the compound channel also hold for the averaged channel.

We see that even for this minimal requirement of weakly effectively constructible, there is no algorithm that can compute such a scheme. This shows the following: Although we know from Theorem 1 that for all rates $R < C(\mathcal{W})$ for the compound channel $\mathcal{W}$ there exists suitable encoder and decoder, these cannot be constructed effectively, i.e., algorithmically.

Some further remarks are in order.

1) Every sequence of $(n(k), M_n(k), \epsilon)$-codes for the compound channel that satisfies both weakly effectively constructible conditions (14)-(15) must have the property that it is not computable. We know that the sequence of rates $\frac{1}{n(k)} \log M_n(k) = R(k)$, $k \in \mathbb{N}$ is not computable. Accordingly, there is no algorithm for computing $R(k)$, $k \in \mathbb{N}$. Moreover, any subsequence will not work either. We conclude that at least of the two sequences $\{n(k)\}_{k\in\mathbb{N}}$ or $\{M_n(k)\}_{k\in\mathbb{N}}$ is not computable.
2) Reversely, for a computable sequence of $(n(k), M_n(k), \epsilon)$-codes for the compound channel that satisfies the monotonicity condition (14), then there exists a $\gamma > 0$ such that $\frac{1}{n(k)} \log M_n(k) \leq C(\mathcal{W}) - \gamma$ is true for all $k \in \mathbb{N}$. This means for such an algorithm for the construction of (n, M_n, ϵ)-codes we have a loss in rate compared to the capacity.
3) We do not require that the algorithm of Definition 11 is computable, i.e., recursively depending on the channel. Therefore, for every channel and every ϵ one has to find an algorithm for the optimal signal processing and coding. Of course, from a practical point of view it would be advantageous to have a universal algorithm for the compound channel, i.e., for the whole state set and all ϵ. However, as there is already no algorithm for the weaker requirement of Definition 10, such a universal algorithm cannot exist.

Remark 6: Note that the uncertainty set $\mathcal{S}$ can be given with arbitrarily fine parameters in the sense that for any different states $s_1, s_2 \in \mathcal{S}$, and any tolerated error $\delta > 0$, the output distributions satisfy

$$\sum_{y\in\mathcal{Y}} \left|W_{s_1}(y|0) - W_{s_2}(y|0)\right| < \delta$$

and

$$\sum_{y\in\mathcal{Y}} \left|W_{s_1}(y|1) - W_{s_2}(y|1)\right| < \delta.$$

But this is no contradiction to the previous results and, accordingly, Theorem 3, Theorem 4, and Corollary 1 remain valid for the algorithmic computability of the capacity and the optimal signal processing and coding schemes.

This thinking about how the uncertainty set can be controlled in terms of tolerated uncertainty error $\delta > 0$ immediately applies to the case in which the channel parameters are estimated based on pilot signals. Here, the energy allocated to the pilot signals controls the uncertainty error of the channel estimates and therewith the uncertainty set. In the end, this uncertainty set will always be non-empty due to the noise and the finite energy that can be allocated in practical systems. Accordingly, there is an algorithmically constructed compound channel that exhibits the above discussed behavior.

C. Finite Blocklength Performance

Recently, the fundamental problem of analyzing the performance of coding schemes in the finite blocklength regime has attracted considerable interest. Since the seminal work [43], there has been considerable progress in understanding the finite blocklength performance, see e.g. also the recent monograph [44] and references therein. The main goal here is to study and characterize for a given fixed (and finite) blocklength n and a fixed (and non-vanishing) error ϵ the maximal achievable rate $R(n,\epsilon)$ and how large the gap is to the capacity C (which is the fundamental asymptotic limit, i.e., for $n \to \infty$ and $\epsilon_n \to 0$).

In [43] this question has been studied for a DMC W and it has been shown that in general we have $|C(W) - R(n,\epsilon)| \leq F(W,n,\epsilon)$ for some function $F(W,n,\epsilon)$. In particular, the dependency of the function F for a fixed W on the blocklength n is particularly interesting and it is desired to be computable, i.e., numerically evaluatable.

Our results in Corollary 1 and Theorem 4 immediately show that such a characterization for the compound channel $\mathcal{W}$, for which $R(n,\epsilon)$ and $F(\mathcal{W},n,\epsilon)$ in dependency on n are computable functions, is not possible.

Theorems 7-8 even show that for a fixed blocklength, it is impossible to construct search algorithms which find good pre- and post-processing that asymptotically achieve the capacity.

D. Averaged Channels

As the compound channel and averaged channel share the same capacity expression, cf. Theorems 1 and 2, the results established above for compound channels can be extended to averaged channels in a straightforward way.

As in Theorem 3 for compound channels, there are computable averaged channels $\overline{\mathcal{W}}$ such that the capacity $C(\overline{\mathcal{W}})$ cannot be algorithmically computed.

Theorem 5: Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary finite alphabets. Then there is a computable averaged channel $\overline{\mathcal{W}} \in \mathcal{AC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ such that

$$C(\overline{\mathcal{W}}) \notin \mathbb{R}_c.$$

We obtain the same consequences as for compound channels regarding the effective constructability.

Theorem 6: For the averaged channel $\overline{\mathcal{W}}$ of Theorem 5, the optimal signal processing and coding scheme is not effectively constructible and also not weakly effectively constructible for any $\epsilon \in (0,1)$, $\epsilon \in \mathbb{R}_c$.

E. Consequences for Finding the Optimal Signal Processing

Turing machines are extremely powerful and superior to current digital computers and can find optimal signal processing

strategies for many applications as already outlined in Section III. Note that in today's communication systems, the set of the underlying parameters that determines the signal processing are always discrete and finite. The reason for this is that all resources in these systems are addressed and digitally optimized in the baseband from which the analog signals are subsequently created. Even these analog signals are often optimized via digital processing [45]. As a consequence, all system parameters are discrete although the number of possible values can be very large. However, a Turing machine is very powerful and can perform exhaustive search even over very large sets, since complexity is not an issue here.

In the following we consider the problem of reliable transmission over a noisy channel, where we consider a fixed blocklength n and tolerated probability of error ϵ. We discuss the cases of perfect CSI and imperfect CSI to highlight how the available channel knowledge affects the optimal pre- and post-processing. We show that for perfect CSI, an exhaustive search for the optimal signal processing is always possible, while for imperfect CSI, this is no longer possible in general.

1) Optimal Pre- and Post-Processing for Perfect CSI: Let W be a discrete memoryless channel (DMC), for which all parameters are rational, and $\epsilon \in (0,1)$, $\epsilon \in \mathbb{Q}$, be the required error probability. We are now interested in finding an algorithm or Turing machine $\mathfrak{T}_{W,\epsilon}$ that computes for any input blocklength $n \in \mathbb{N}$ the optimal signal processing strategy for the DMC W and error probability ϵ. This means for input blocklength $n \in \mathbb{N}$, the Turing machine $\mathfrak{T}_{W,\epsilon}(n) = (E_n^*, \phi_n^*)$ outputs an optimal encoder E_n^* and a optimal decoder ϕ_n^* providing the maximal possible rate while guaranteeing the target error probability ϵ. This means the optimal encoder E_n^* is specified by the set of codewords

$$\{x^n(1), \ldots, x^n(M_n^*(\epsilon))\}$$

and the optimal decoder ϕ_n^* is specified by decoding sets

$$\{\mathcal{D}_n^*(1), \ldots, \mathcal{D}_n^*(M_n^*(\epsilon))\}$$

with $\mathcal{D}_n^*(m) = \bigcup_{y^n:\phi_n^*(y^n)=m} y^n$ such that the maximum probability of error satisfies

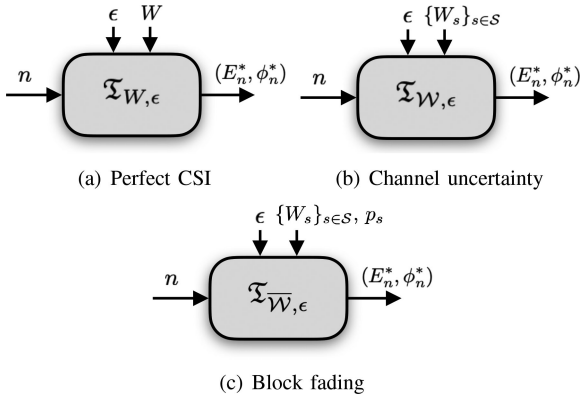
$$\max_m W^n\left(\mathcal{D}_n^*(m)|x^n(m)\right) < \epsilon, \tag{16}$$

and further, for all $M_n > M_n^*(\epsilon)$ it is impossible to find an encoder and decoder that satisfy (16).

The corresponding Turing machine $\mathfrak{T}_{W,\epsilon}$ is visualized in Fig. 4(a) and works as follows. We first fix the blocklength $n = 2$ and search for the maximum number of messages. We start with $\mathcal{M}_2 = \{1, 2\}$ and search over all possible pairs of codewords $x^2(1), x^2(2) \in \mathcal{X}^2$ if there are disjoint decoding sets $\mathcal{D}_2(1), \mathcal{D}_2(2) \subset \mathcal{Y}^2$ with $\mathcal{D}_2(1) \cup \mathcal{D}_2(2) = \mathcal{Y}^2$ such that the maximum probability of error

$$\max_{m=1,2} W^2\left(\mathcal{D}_2^c(m))|x^2(m)\right) < \epsilon, \tag{17}$$

is satisfied. Note that the left hand side and right hand side of (17) are rational numbers so that this comparison can be computed by $\mathfrak{T}_{W,\epsilon}$. Since the set of all possible pairs in $\mathcal{X}^2$ and the set of all possible decoding sets in $\mathcal{Y}^2$ are finite, $\mathfrak{T}_{W,\epsilon}$ can actually test all possible combinations. If there is no pair of codewords $(x^2(1), x^2(2))$ with corresponding decoding sets such that (17) is satisfied, the Turing machine stops and we set the maximum number of codewords as $M_2^*(\epsilon) = 1$. In the other case, if we are successful and there are pair of codewords



(a) Perfect CSI (b) Channel uncertainty

(c) Block fading

Fig. 4. Turing machine for computing the optimal pre- and post-processing. (a) For perfect CSI, the Turing machine $\mathfrak{T}_{W,\epsilon}$ gets W and ϵ as parameters and computes the optimal (E_n^*, ϕ_n^*) for any input blocklength n. (b) For channel uncertainty, the Turing machine $\mathfrak{T}_{\mathcal{W},\epsilon}$ gets the uncertainty set $\{W_s\}_{s\in\mathcal{S}}$ and ϵ as parameters and computes the optimal (E_n^*, ϕ_n^*) for any input blocklength n. (c) For block fading channels, the Turing machine $\mathfrak{T}_{\overline{\mathcal{W}},\epsilon}$ gets the uncertainty set $\{W_s\}_{s\in\mathcal{S}}$, the distribution p_S, and ϵ as parameters and computes the optimal (E_n^*, ϕ_n^*) for any input blocklength n.

and corresponding decoding sets that satisfy the error criterion, we increase the number of messages and consider the set of messages $\mathcal{M}_2 = \{1, 2, 3\}$. Next, we perform the same search for all possible codewords $x^2(1), x^2(2), x^2(3) \in \mathcal{X}^2$ and decoding sets $\mathcal{D}_2(1), \mathcal{D}_2(2), \mathcal{D}_2(3)$ if there are some that satisfy the corresponding error criterion. This procedure is continued inductively.

There exists an $\hat{M} \leq \min\{|\mathcal{X}|^2, |\mathcal{Y}|^2\}$ for which the search described above is not successful anymore. Then, the maximum number of codewords is $M_2^*(\epsilon) = \hat{M} - 1$ since $M_2^*(\epsilon) \geq \hat{M} - 1$ is obviously true. In addition, if $M_2^*(\epsilon) > \hat{M} - 1$ would be true, we must have $M_2^*(\epsilon) > \hat{M}$ since our search was not successful for $\hat{M}$. However, we now argue that this cannot be the case. Let an optimal encoder E_2^* be given by $\{x^2(1), \ldots, x^2(M_2^*(\epsilon))\}$ and an optimal decoder ϕ_n^* be specified by $\{\mathcal{D}_2^*(1), \ldots, \mathcal{D}_2^*(M_2^*(\epsilon))\}$ which determines an optimal pre- and post-processing. We now consider the encoder $\hat{E}_2$ given by

$$\left\{x^2(1), \ldots, x^2(\hat{M})\right\}$$

and the decoder $\hat{\phi}_2$ specified by

$$\left\{\hat{\mathcal{D}}_2(1), \ldots, \hat{\mathcal{D}}_2(\hat{M}-1), \hat{\mathcal{D}}_2(\hat{M})\right\}$$

with $\hat{\mathcal{D}}_2(\hat{M}) = \bigcup_{l=\hat{M}}^{M_2^*(\epsilon)} \mathcal{D}_2^*(l)$ as pre- and post-processing. This must satisfy

$$\max_{1\leq m\leq \hat{M}} W^2\left(\hat{\mathcal{D}}_2^c(m)|x^2(m)\right) < \epsilon,$$

i.e., our search algorithm would have already found $\hat{E}_2$ and $\hat{\phi}_2$ for $\hat{M}$ messages. This is a contradiction so that $M_2^*(\epsilon) = \hat{M} - 1$ must be true.

The algorithm above computed for a fixed blocklength the maximum number of messages such that the probability of error ϵ is satisfied (it is of course possible to further include additional resources and constraints such as beamforming weights or MIMO pre- and post-processing strategies into this search,

because all these signal processing steps are done in the baseband). Now, we inductively continue with the procedure above with increasing blocklength n. For a fixed n we compute $M_n^*(\epsilon)$ and then increase n by one and compute $M_{n+1}^*(\epsilon)$ with the algorithm above using $M_n^*(\epsilon)+1$ as the initial message size. With this procedure, the Turing machine $\mathfrak{T}_{W,\epsilon}$ computes the computable sequence $\{M_n^*(\epsilon)\}_{n\in\mathbb{N}}$ and with the strong converse to Shannon's coding theorem we have for all $\epsilon \in (0,1)$

$$\lim_{n\to\infty}\frac{1}{n}\log M_n^*(\epsilon) = \max_{p\in\mathcal{P}(\mathcal{X})} I(p,W) = C(W).$$

2) Optimal Pre- and Post-Processing for Imperfect CSI: Obviously, one would like to have a similar Turing machine that computes the optimal pre- and post-processing for the case of imperfect CSI as well. Such a Turing machine $\mathfrak{T}_{\mathcal{W},\epsilon}$ is depicted in Fig. 4(b) and we show next that such a Turing machine does not exist for computable compound channels $\mathcal{W}$ in general, even if all possible channel realizations in the uncertainty set and the required probability of error are rational. Note that we do not require that the Turing machine depends recursively on the channel; we only ask if it is possible to find such a search algorithm for a fixed and given channel and error.

To show this, we next prove the general result that for $\epsilon \in (0,\frac{1}{2})$, $\epsilon \in \mathbb{R}_c$, and the compound channel $\mathcal{W}$ from Theorem 3, the sequence of the maximum number of messages $\{M_n^*(\epsilon)\}_{n\in\mathbb{N}}$ is not a computable sequence of natural numbers, i.e., there is no Turing machine $\mathfrak{T}^*$ with

$$\mathfrak{T}^*(n) = M_n^*(\epsilon), \quad n\in\mathbb{N},$$

so that there is no algorithm that computes the sequence $\{M_n^*(\epsilon)\}_{n\in\mathbb{N}}$. As a consequence, there is no algorithm or Turing machine that can effectively compute the sequence of optimal pre- and post-processing.

Theorem 7: Let $\mathcal{W}$ be the compound channel from Theorem 3. Then for all $\epsilon \in (0,\frac{1}{2})$, $\epsilon \in \mathbb{R}_c$, the sequence of the maximum number of messages $\{M_n^*(\epsilon)\}_{n\in\mathbb{N}}$ is not a computable sequence of natural numbers. This holds then also for rational $\epsilon \in (0,\frac{1}{2}), \epsilon \in \mathbb{Q}$.

Proof: We prove the result by contradiction. Therefore, we assume that there exists an $\hat{\epsilon} \in (0,\frac{1}{2})$, $\hat{\epsilon} \in \mathbb{R}_c$, such that the sequence $\{M_n^*(\hat{\epsilon})\}_{n\in\mathbb{N}}$ is a computable sequence. Depending on $\hat{\epsilon}$ we can effectively, i.e., algorithmically, compute a natural number $\hat{n} = \hat{n}(\hat{\epsilon})$ (for a detailed proof we refer to Appendix B, cf. in particular (26)) such that for all $n \geq \hat{n}$ the rate

$$\hat{R}_n := \frac{1}{n}\log M_n^*(\hat{\epsilon}) \tag{18}$$

of this maximal code construction satisfies

$$\hat{R}_n \leq C(\mathcal{W}).$$

Since the rate is defined as in (18), the sequence $\{\hat{R}_n\}_{n\in\mathbb{N}}$ is a computable sequence. For $n \geq \hat{n}$ we define $U_n = \max_{l\in[\hat{n},n]} \hat{R}_l$ and for $n \geq \hat{n}$ we have $U_n \leq U_{n+1}$ and $U_{n+1} \geq \hat{R}_{n+1}$. Since the stronger converse holds for $\mathcal{W}$ (for a detailed discussion we refer to Appendix B), we have

$$\lim_{n\to\infty}\hat{R}_n = C(\mathcal{W})$$

and $C(\mathcal{W})$ is the limit of a computable monotonically increasing sequence of computable numbers. Since $C(\mathcal{W})$ is also the limit of a computable monotonically decreasing sequence of computable numbers, it must hold $C(\mathcal{W}) \in \mathbb{R}_c$, cf. [35]. Thus, the capacity value must be a computable real number which is a contradiction to Theorem 3, which shows that our initial assumption of $\{M_n^*(\hat{\epsilon})\}_{n\in\mathbb{N}}$ being a computable sequence is wrong proving the desired result. ■

This result allows us to study the behavior of our search algorithm (and also any other possible search algorithm) for the optimal pre- and post-processing for compound channels in more detail. Note that the proof of Theorem 3 has been presented for compound channels with $|\mathcal{X}| = |\mathcal{Y}| = 2$ and W_s, $s \in \mathcal{S}$, having rational entries. The maximum probability of error condition (16) becomes

$$\sup_{s\in\mathcal{S}}\max_m W_s^n\left(\mathcal{D}_n(m)|x^n(m)\right) < \epsilon.$$

We have seen that the sequence $\{M_n^*(\hat{\epsilon})\}_{n\in\mathbb{N}}$ is not computable. Here, we have only considered Turing machines that have to stop for every input. In the following, we weaken this assumption by requiring the Turing machine to stop only for those inputs that fulfill the desired properties. In the other case, the Turing machine runs forever and does not stop. We obtain the following result.

Theorem 8: For computable compound channels $\mathcal{W}$ from Theorem 3, there is no Turing machine $\mathfrak{T}^*$ that takes the blocklength n and the desired number of messages M as input and stops whenever $M \leq M_n^*(\epsilon)$ and otherwise runs forever.

Proof: We prove the result by contradiction. Therefore, we assume that there exists such a Turing machine $\mathfrak{T}^*$ that takes the blocklength n and the desired number of messages M as input and stops whenever $M \leq M_n^*(\epsilon)$ and otherwise runs forever. We use this Turing machine to construct a new Turing machine $\mathfrak{T}_{\text{opt}}$ as follows:

The Turing machine $\mathfrak{T}_{\text{opt}}$ provides for every $m \in \mathbb{N}$ a set $\mathcal{M}_m = \{M_m(2),\ldots,M_m(m)\}$ of $m-1$ natural numbers by the following procedure:

For $n = 2,\ldots,m$ we start the Turing machine $\mathfrak{T}^*$ in parallel for the inputs $M = 2,\ldots,m$. This means in total there are $(m-1)^2$ Turing machines $\{\mathfrak{T}^*(n,M)\}_{2\leq n\leq m, 2\leq M\leq m}$ and for every of these Turing machines there are m steps for the computation of $\mathfrak{T}^*(n,M)$. For fixed n some of these Turing machines stop. For the corresponding number M, for which the Turing machines $\mathfrak{T}^*(n,\cdot)$ for input M stop, we take the largest number and denote it by $M_m(n)$. We obtain the set $\mathcal{M}_m = \{M_m(2),\ldots,M_m(m)\}$ of these numbers. Further, we have that for every $n \in \mathbb{N}$ and $M \leq M_n^*(\epsilon)$ there must exist a number $\hat{l}$ such that the Turing machine $\mathfrak{T}^*$ for inputs n and M stops within $\hat{l}$ steps. Accordingly, we have $M_m(n) = M_n^*(\epsilon)$ for all sufficiently large natural numbers n. Thus, the sequence $\{M_n^*(\epsilon)\}_{n\in\mathbb{N}}$ must be a computable sequence of natural numbers. This is a contradiction to Theorem 7 and our initial assumption was wrong. This proves the desired result. ■

This result can be used to further study the behavior of our search algorithm (and also any other possible search algorithm).

Theorem 9: Every search algorithm for the maximal code construction for computable compound channels $\mathcal{W}$ from Theorem 3 has the following property. There exists a blocklength $\hat{n}$ such that one of the following statements is true:

1) If the algorithm stops for all $M \in \mathbb{N}$, then the algorithm provides wrong results for $M_n^*(\epsilon)$.
2) If the algorithm does not stop for all $M \in \mathbb{N}$, then the algorithm cannot stop for those M for which $M \leq M_n^*(\epsilon)$ is true.

Proof: Theorem 7 already shows that the search algorithm cannot always stop while always providing the correct answer. This justifies the first property. The second property follows immediately from Theorem 8. ■

We have seen that an exhaustive search over all resources is not possible on a Turing machine. This is insofar surprising as for a fixed blocklength n, all underlying parameters are finite and therewith the set of possibilities as well. In addition to this, all parameters itself are computable and therewith algorithmically tractable by a Turing machine.

To show that there is no effective implementation of an exhaustive search, there are two important things to highlight:

1) The capacity value $C(\mathcal{W})$ of the compound channel $\mathcal{W}$ is a non-computable number.
2) The sequence of the maximal number of messages converges to the capacity from below after a specific blocklength $\hat{n}$. It is possible to effectively compute this blocklength $\hat{n}$.

We note that we not only have shown that our presented search algorithm does not work under channel uncertainty, but further that there is no such exhaustive search algorithm possible at all. To the best of our knowledge, this is the first time that such a behavior has been observed.

Finally, we note that all these studies done for compound channels carry over to averaged channels as well.

V. Channel Knowledge At Transmitter or Receiver

Here, we discuss the cases in which either the encoder or decoder has knowledge about the actual channel realization to see whether such partial knowledge influences the computability of the corresponding capacity expression.

A. Informed Encoder

We start with the case of an informed encoder (IE) so that channel state information is available at the transmitter. As the actual channel state is known to the transmitter, it can adapt the encoder accordingly so that definition of a code changes a follows.

Definition 12: An (n, M_n)*-code* for the compound channel with informed encoder consists of set of encoders $E_{s,n} : \mathcal{M}_n \to \mathcal{X}^n$, one for each channel state $s \in \mathcal{S}$, at the transmitter with a set of messages $\mathcal{M}_n := \{1, \ldots, M_n\}$ and a decoder $\phi_n : \mathcal{Y}^n \to \mathcal{M}_n$ at the receiver.

The capacity is known in this case and is equal to the worst-case capacity.

Theorem 10 ([46]): The capacity $C_{\text{IE}}(\mathcal{W})$ of the compound channel $\mathcal{W}$ with informed encoder is

$$C_{\text{IE}}(\mathcal{W}) = \inf_{s\in\mathcal{S}} C(W_s) = \inf_{s\in\mathcal{S}} \sup_{p\in\mathcal{P}(\mathcal{X})} I(p, W_s).$$

Now we want to study how the available channel state information at the transmitter effects the computability of the capacity and the effective constructability of the corresponding codes. Let $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ be the computable compound channel from the proof of Theorem 1 whose capacity expression is not computable. The following result shows that also for an informed encoder, the capacity remains non-computable.

Theorem 11: For the computable compound channel $\mathcal{W}$ of Theorem 1 we have

$$C_{\text{IE}}(\mathcal{W}) \notin \mathbb{R}_c.$$

Proof: In the proof of Theorem 3 we have shown that for all $s \in \mathbb{N}$ we have

$$I(p_*, W_s) = \max_{p\in\mathcal{P}(\mathcal{X})} I(p, W_s)$$

with

$$p_* = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \in \mathcal{P}_c(\mathcal{X})$$

being the optimal input distribution. With this, we get

$$\begin{aligned} C(\mathcal{W}) &\geq \min_{s\in\mathbb{N}} \max_{p\in\mathcal{P}(\mathcal{X})} I(p, W_s) \\ &= \min_{s\in\mathbb{N}} C(W_s) = C_{\text{IE}}(\mathcal{W}). \end{aligned}$$

On the other hand, since channel state information at the transmitter can only increase the capacity, we also have

$$C(\mathcal{W}) \leq C_{\text{IE}}(\mathcal{W})$$

so that we actually have equality $C(\mathcal{W}) = C_{\text{IE}}(\mathcal{W})$. With this, the non-computability result of Theorem 3 immediately implies the non-computability for the informed encoder case as well. ■

With this, we immediately get the following result.

Theorem 12: For the compound channel $\mathcal{W}$ of Theorem 11, the optimal signal processing and coding scheme is not weakly effectively constructible for any $\epsilon \in (0, 1)$, $\epsilon \in \mathbb{R}_c$.

Proof: With the result of Theorem 11, the proof here is exactly as in Theorem 4 and omitted for brevity. ■

Theorems 11 and 12 show that in the case of an informed encoder, we have the same results and observations as in the case of no channel state information. Accordingly, channel state information at the transmitter does not effect the computability. Of course, in general, channel state information at the transmitter provides a gain in performance, but our construction above shows that this need not be the case in general.

B. Informed Decoder

We continue with the case of an informed decoder (ID) so that channel state information is available at the receiver but not at the transmitter. Accordingly, the receiver can adapt its decoder to the actual channel state and the definition of a code changes a follows.

Definition 13: An (n, M_n)*-code* for the compound channel with informed decoder consists of an encoder $E_n : \mathcal{M}_n \to \mathcal{X}^n$ at the transmitter with a set of messages $\mathcal{M}_n := \{1, \ldots, M_n\}$ and a set of decoders $\phi_{s,n} : \mathcal{Y}^n \to \mathcal{M}_n$, one for each channel state $s \in \mathcal{S}$, at the receiver.

It is known that channel state information at the receiver does not increase the capacity of a compound channel.

Theorem 13 ([46]): The capacity $C_{\text{ID}}(\mathcal{W})$ of the compound channel $\mathcal{W}$ with informed decoder is

$$C_{\text{ID}}(\mathcal{W}) = C(\mathcal{W}).$$

As the capacity in the case of an informed decoder is the same as without channel state information, all results from there immediately apply to the informed decoder case as well and the capacity is non-computable the corresponding optimal signal processing and coding schemes are not weakly effectively constructible.

VI. Conclusion

In this paper, the uncertainty models of compound and averaged channels have been studied in which the channel state is unknown to both encoder and decoder. We have also considered variations on this theme in which the channel state is either known to the encoder (compound channel with informed encoder) or known to the decoder (compound channel with informed decoder). For all these scenarios, it has been shown that the capacity of the compound channel and averaged channel are non-computable real numbers in general and that the corresponding optimal signal processing and coding schemes are not effectively constructible. This implies that the existence of certain schemes does not immediately imply their algorithmic constructability. For future work, it is of interest to further characterize the classes of such channels for which the capacity is a non-computable number. In addition, we have shown that it is impossible to algorithmically find good pre- and post-processing schemes in the case of channel uncertainty and block fading; even exhaustive search algorithms are impossible to implement algorithmically. To the best of our knowledge, this is the first time that such behavior has been observed.

It is interesting to observe that not all areas of signal processing and system theory allow an automatic design of optimal processing. For robust signal processing for communication systems under channel uncertainty and block fading, there is little known and we have shown that such a computer-aided approach is in general impossible.

Appendix

A. Proof of Lemma 1

Here we present the proof of Lemma 1. This is done by contradiction for which we assume that $\lambda = h_2(\mu_*)$ is a computable real number, i.e., $\lambda \in \mathbb{R}_c$.

Without loss of generality we can assume that there are values l_1, l_2, and l_3 for which the computable function $\varphi_{\mathcal{A}}$ takes on the values $\varphi_{\mathcal{A}}(l_1) = 1$, $\varphi_{\mathcal{A}}(l_2) = 2$, and $\varphi_{\mathcal{A}}(l_3) = 3$, since otherwise we would simply add these values to the set $\mathcal{A}$. With this, we have

$$\mu_* > \frac{1}{4} \cdot \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} \right) = \frac{1}{4} \cdot \frac{7}{8} = \frac{7}{32} = \delta,$$

cf. (8) and (9).

Note that the binary entropy function $h_2(\cdot)$ is a computable function on the interval $[\delta, \frac{1}{4}]$ and so is the first derivative $h_2'(\cdot)$. We have $h_2'(\mu_*) \geq c_1 > 0$, $\mu_* \in [\delta, \frac{1}{4}]$ and c_1 can actually be computed, i.e., $c_1 \in \mathbb{R}_c$. Furthermore, $h_2(\cdot)$ is a concave function and strictly monotonically increasing on $[0, \frac{1}{2}]$.

Next, we will use the following fact: Let $\mathbb{I} \subset \mathbb{R}$ be a closed interval and $f : \mathbb{I} \to \mathbb{R}$ be a strict monotonically increasing, twice continuously differentiable function. There exists an interior point $a \in \mathbb{I}$ with $f(a) = 0$. Let $x_0 \in \mathbb{I}$ a starting point and

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Further, let $C_1 = \max_{x \in \mathbb{I}} |f'(x)|$ and $C_2 = \max_{x \in \mathbb{I}} |f''(x)|$ with

$$\frac{C_2}{2C_1} |\mathbb{I}| = C_3 < 1$$

where $|\mathbb{I}|$ denotes the length of the interval $\mathbb{I}$. Then it holds that

$$|x_n - a| < \frac{1}{K} (C_3)^{2^n}$$

with $K = \frac{C_2}{2C_1}$. This means we have quadratic convergence.

Now, we use this result for the function

$$G(\lambda, p) := \lambda - h_2(p)$$

with fixed $\lambda = h_2(\mu_*)$ as initially assumed. By assumption we have $\lambda \in \mathbb{R}_c$ so that $G(\lambda, \cdot)$ is a computable continuous function in the interval $[\frac{1}{4} \cdot \frac{7}{8}, \frac{1}{4}]$. The first derivative of G is

$$G'(\lambda, p) = \frac{\partial}{\partial p} G(\lambda, p) = -h_2'(p)$$

and therewith $G'(\lambda, \cdot)$ is a computable continuous function on the interval $[\frac{1}{4} \cdot \frac{7}{8}, \frac{1}{4}]$.

For $p \in (0, 1)$ we have

$$\begin{aligned} h_2(p) &= -p \log p - (1-p) \log(1-p) \\ &= -(p \log p + (1-p) \log(1-p)) \\ h_2'(p) &= -\left(\log p + p \frac{1}{\ln 2} \frac{1}{p} - \log(1-p) \right. \\ &\qquad \left. - (1-p) \frac{1}{\ln 2} \frac{1}{1-p} \right) \\ &= -(\log p - \log(1-p)) \\ h_2''(p) &= -\left(\frac{1}{\ln 2 \cdot p} + \frac{1}{\ln 2 \cdot (1-p)} \right). \end{aligned}$$

We choose $p_0 = \frac{1}{4}$ as the starting point. With this,

$$p_{n+1} = p_n - \frac{G(\lambda, p_n)}{G'(\lambda, p_n)}, \quad n \in \mathbb{N},$$

is a computable sequence of computable real numbers, since $G(\lambda, \cdot)$ and $G'(\lambda, \cdot)$ are computable continuous functions. Now, for $p \in [\frac{1}{4} \cdot \frac{7}{8}, \frac{1}{4}]$ we have

$$\begin{aligned} |h_2'(p)| &= |-\log p - \log(1-p)| \\ &= \log \frac{1-p}{p} \\ &\leq \log \frac{\frac{3}{4}}{\frac{1}{4} \cdot \frac{7}{8}} \\ &= \log \left(\frac{3}{4} \cdot \frac{4 \cdot 8}{7} \right) \\ &= \log \frac{3 \cdot 8}{7}. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} |h_2''(p)| &< \frac{1}{\ln 2} \left(\frac{4 \cdot 8}{7} + \frac{1}{1 - \frac{1}{4}} \right) \\ &= \frac{1}{\ln 2} \left(\frac{4 \cdot 8}{7} + \frac{4}{3} \right) \end{aligned}$$

$$< \frac{2}{\ln 2}\frac{4\cdot 8}{7}.$$

For the interval $\mathbb{I} = [\frac{1}{4}\cdot\frac{7}{8}, \frac{1}{4}]$ we obtain

$$\frac{\max_{p\in\mathbb{I}}|h_2''(p)|}{2\cdot\max_{p\in\mathbb{I}}|h_2'(p)|} < \frac{\frac{2}{\ln}\frac{4\cdot 8}{7}}{2\cdot\log\frac{3\cdot 8}{7}} = \frac{4\cdot 8}{\ln 2\cdot\log\frac{3\cdot 8}{7}\cdot 7} = C_4.$$

The length of the interval is

$$|\mathbb{I}| = \frac{1}{4} - \frac{1}{4}\cdot\frac{7}{8} = \frac{1}{4}\cdot\left(1-\frac{7}{8}\right) = \frac{1}{32}$$

and we set

$$C_5 = C_4\cdot\frac{1}{32} = \frac{1}{\ln 2\cdot\log\frac{3\cdot 8}{7}\cdot 7} < 1$$

which is a computable real number. With this,

$$C_6 = \frac{\max_{p\in\mathbb{I}}|h_2''(p)|}{2\cdot\max_{p\in\mathbb{I}}|h_2'(p)|} \tag{19}$$

is obviously a computable real number as well and we have

$$|\mu_* - p_n| \leq \frac{1}{C_6}(C_5)^{2^n}.$$

The computable sequence $\{p_n\}_{n\in\mathbb{N}}$ of computable real numbers converges effectively to μ_*. Therefore, μ_* must be computable as well, i.e., $\mu_* \in \mathbb{R}_c$. But this is a contradiction since from our construction we have $\mu_* \notin \mathbb{R}_c$. This is implies that our initial assumption that $\lambda = h_2(\mu_*) \in \mathbb{R}_c$ was wrong. Accordingly, we have proved that $h_2(\mu_*) \notin \mathbb{R}_c$. ■

B. Stronger Converse

Here we show that the stronger converse according to [47] holds for the maximal code construction used in Theorem 7. The goal here is to compute a number $\hat{n}$ for the computable compound channel $\mathcal{W}$ from Theorem 3 for $\epsilon \in (0, \frac{1}{2})$, $\epsilon \in \mathbb{R}_c$ such that for all $n \geq \hat{n}$ the rate of the maximal code construction satisfies

$$\frac{1}{n}\log M_n^*(\epsilon) < C(\mathcal{W}).$$

It is important here that the number $\hat{n} = \hat{n}(\epsilon)$ is computable, i.e., it is not sufficient to only know the existence of this number, but it must actually be computable by a Turing machine. Then we are able to prove Theorem 7, since only then we can obtain a lower bound for $C(\mathcal{W})$. In the following, we follow [47], but emphasize that the main goal of [47] was to prove the existence of the necessary constants but not their explicit computability. For completeness, we present the whole proof instead of mentioning only the steps that needed to be adapted and changed.

For $p \in (0, \frac{1}{2})$ we consider a fixed binary symmetric channel W_p with cross-over probability p. For a fixed error $\epsilon \in (0, \frac{1}{2}), \epsilon \in \mathbb{R}_c$, and blocklength n, we want to establish an upper bound on the maximal number of messages $M_n^*(\epsilon)$ that can be transmitted over this channel. Let K be the largest natural number such that

$$\sum_{j=0}^{K}\binom{n}{j}p^j(1-p)^{n-j} \leq 1-\epsilon$$

holds. Then we have

$$M_n^*(\epsilon) < \frac{2^n}{\binom{n}{K}}$$

with

$$\binom{n}{K} = \frac{n!}{K!(n-K)!}.$$

For $\epsilon \in (0, \frac{1}{2})$ let β_ϵ be given by

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\beta_\epsilon}\exp(-\frac{1}{2}t^2)dt = 1-\epsilon. \tag{20}$$

Due to (20), the number β_ϵ is effectively computable. In [47] it is shown that $K = np + \beta(n,\epsilon)(np(1-p))^{\frac{1}{2}}$ is true and for all $\Delta > 0$ there exists an n_Δ such that for all $n \geq n_\Delta$ we have

$$\beta_\epsilon - \Delta \leq \beta(n,\epsilon) \leq \beta_\epsilon + \Delta.$$

In the following, we want to effectively compute n_Δ which is needed in the maximal code construction in the proof of Theorem 7. For this purpose, we want to understand the role of K by establishing lower and upper bounds. For this we need the Berry-Esseen theorem, cf. for example [48].

We first establish an upper bound on K as follows. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{P}[X_1 = 1] = p$ and $\mathbb{P}[X_1 = 0] = q = 1-p$. Then,

$$\mathbb{E}[X_1] = p,$$
$$\text{Var}(X_1) = p(1-p) = pq,$$
$$\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^3] = pq(p^2+q^2).$$

We have

$$S_n^* = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{npq}}$$

and therewith it follows from the Berry-Esseen theorem that

$$\mathbb{P}[X_1 + \cdots + X_n \leq K] = \mathbb{P}\left[S_n^* \leq \frac{K-np}{\sqrt{npq}}\right] \leq \phi\left(\frac{K-np}{\sqrt{npq}}\right) + \frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}$$

with $\phi(\cdot)$ denoting the probability density function of the normal distribution $\mathcal{N}(0,1)$ with zero mean and unit variance. To ensure that the left hand side is upper bounded by $1-\epsilon$, it is sufficient that the right hand side is upper bounded by $1-\epsilon$ as well. We have

$$\frac{K-np}{\sqrt{npq}} \leq \phi^{-1}\left(1-\epsilon-\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}\right)$$

which is equivalent to

$$K \leq np + \sqrt{npq}\,\phi^{-1}\left(1-\epsilon-\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}\right) \tag{21}$$

which provides the desired upper bound on K so that

$$\sum_{j=1}^{K}\binom{n}{j}p^j(1-p)^{n-(1-p)} \leq 1-\epsilon$$

holds.

Next, we establish in a similar fashion a lower bound on K. From the definition of K we have

$$\sum_{j=0}^{K+1}\binom{n}{j}p^j(1-p)^{n-j} > 1-\epsilon \tag{22}$$

so that

$$\mathbb{P}\left[S_n^* \le \frac{K+1-np}{\sqrt{npq}}\right] \ge \phi\left(\frac{K+1-np}{\sqrt{npq}}\right) - \frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}. \tag{23}$$

If the right hand side of (23) is larger than $1-\epsilon$, then this is also true for (22). Thus,

$$\phi\left(\frac{K+1-np}{\sqrt{npq}}\right) - \frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2} > 1-\epsilon$$

implies (22), i.e.,

$$\frac{K+1-np}{\sqrt{npq}} \ge \phi^{-1}\left(1-\epsilon+\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}\right)$$

so that

$$\begin{aligned} K &\ge np + \sqrt{npq}\,\phi^{-1}\left(1-\epsilon+\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}\right) - 1 \\ &= np + \sqrt{npq}\,\phi^{-1}\left(1-\epsilon+\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2} - \frac{1}{\sqrt{npq}}\right). \end{aligned} \tag{24}$$

For $\epsilon \in (0, \frac{1}{2})$, $\epsilon \in \mathbb{Q}$, we can now recursively compute a natural number n_Δ for every $p \in (0, \frac{1}{2})$, $q = 1-p$, $p \in \mathbb{Q}$, and every Δ such that

$$\begin{aligned} &\max\left\{\left|\phi^{-1}(1-\epsilon) - \phi^{-1}(1-\epsilon-\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2})\right|,\right. \\ &\left.\left|\phi^{-1}(1-\epsilon+\frac{0.8}{\sqrt{n}}\frac{p^2+q^2}{(pq)^2}) - \frac{1}{\sqrt{npq}} - \phi^{-1}(1-\epsilon)\right|\right\} \\ &\quad \le \Delta. \end{aligned}$$

Now, from [47, p. 211] we have

$$M_n^*(\epsilon) < \frac{2^n}{\binom{n}{j}}$$

or equivalently,

$$\log M_n^*(\epsilon) < n - \log\binom{n}{j}.$$

It remains to show that we can further bound $M_n^*(\epsilon)$ as follows:

$$\log M_n^*(\epsilon) < n\left(1 - p\log\frac{1}{p} - q\log\frac{1}{q}\right)$$

for all $n \ge n_1$. To achieve this, we make use of [47, Eq. (3)]. With

$$A = \frac{1}{2}\log(2\pi n) + \frac{1}{12[npq + Bn^{1/2}(q-p) - B^2]}$$

and $B = \beta(n,\epsilon)(pq)^{\frac{1}{2}}$, we obtain

$$\begin{aligned} -\log\binom{n}{K} &< A + \left(np + Bn^{\frac{1}{2}} + \frac{1}{2}\right)\log\left(p+\frac{B}{\sqrt{n}}\right) \\ &\quad + \left(nq - Bn^{\frac{1}{2}} + \frac{1}{2}\right)\log\left(q-\frac{B}{\sqrt{n}}\right) \\ &= np\log\left(p+\frac{B}{\sqrt{n}}\right) + nq\log\left(q-\frac{B}{\sqrt{n}}\right) \\ &\quad - Bn^{\frac{1}{2}}\left(\log(q-\frac{B}{\sqrt{n}}) - \log(p+\frac{B}{\sqrt{n}})\right) \\ &\quad + \frac{1}{2}\log\left(p+\frac{B}{\sqrt{n}}\right) + \frac{1}{2}\log\left(q-\frac{B}{\sqrt{n}}\right) + A \\ &< np\log p + nq\log q - Bn^{\frac{1}{2}}\left(\log\frac{q-\frac{B}{\sqrt{n}}}{p+\frac{B}{\sqrt{n}}}\right) \\ &\quad + \frac{1}{2}\log\left(p+\frac{B}{\sqrt{n}}\right) + \frac{1}{2}\log\left(q-\frac{B}{\sqrt{n}}\right) + A. \end{aligned} \tag{25}$$

Next, we will need the following elementary bound:

$$\log(1+x) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^k}{k}, \qquad |x| < 1$$

implies that

$$\log(1+x) < x, \qquad 0 < x < 1$$

and further

$$\log(1-x) = \sum_{k=1}^{\infty}(-1)^{k+1}(-1)^k\frac{x^k}{k} = -\sum_{k=1}^{\infty}\frac{x^k}{k} < -x.$$

We use this to obtain for $\frac{1}{2} < q < 1$ and $p > 1-q$

$$\begin{aligned} &np\log\left(p+\frac{B}{\sqrt{n}}\right) + nq\log\left(q-\frac{B}{\sqrt{n}}\right) \\ &\quad = np\log p + nq\log q + np\log\left(1+\frac{B}{p\sqrt{n}}\right) \\ &\qquad + nq\log\left(1-\frac{B}{q\sqrt{n}}\right) \\ &\quad < np\log p + nq\log q + np\frac{B}{p\sqrt{n}} - nq\frac{B}{q\sqrt{n}} \\ &\quad = np\log p + nq\log q + B\sqrt{n} - B\sqrt{n} \\ &\quad = np\log p + nq\log q. \end{aligned}$$

Inserting this into (25), we obtain

$$\begin{aligned} -\log\binom{n}{K} &< np\log p + nq\log q - Bn^{\frac{1}{2}}\left(\log\frac{q-\frac{B}{\sqrt{n}}}{p+\frac{B}{\sqrt{n}}}\right. \\ &\left. + \frac{1}{2\sqrt{n}}\log\left(p+\frac{B}{\sqrt{n}}\right) + \frac{1}{2\sqrt{n}}\log\left(q-\frac{B}{\sqrt{n}}\right) + \frac{1}{\sqrt{n}}A\right). \end{aligned}$$

We can recursively compute an $\hat{n} \ge n_\Delta$ such that for all $n \ge \hat{n}$ the part within the brackets is positive, i.e.,

$$\begin{aligned} &\log\frac{q-\frac{B}{\sqrt{n}}}{p+\frac{B}{\sqrt{n}}} + \frac{1}{2\sqrt{n}}\log\left(p+\frac{B}{\sqrt{n}}\right) \\ &\quad + \frac{1}{2\sqrt{n}}\log\left(q-\frac{B}{\sqrt{n}}\right) + \frac{1}{\sqrt{n}}A > 0. \end{aligned}$$

This implies then that for all $n \ge \hat{n}$ we have

$$-\log\binom{n}{K} < np\log p + nq\log q.$$

This shows that the stronger converse holds in our case. With this, we have established an effectively computable upper bound on $M_n^*(\epsilon)$ for the binary symmetric channel W_p.

Now, we want to determine the number $\hat{n}$ for the compound channel $\mathcal{W}$ from Theorem 3. We have to make sure that $\hat{n}$ can

be algorithmically computed. From the proof of Theorem 3 we know that our compound channel $\mathcal{W}$ satisfies

$$\mu_s = \sum_{l=1}^{s} \frac{1}{2^{\varphi_{\mathcal{A}}(l)+2}},$$

cf. also (7). This means that we get $\mu_1 = \frac{1}{8}$ and further

$$\mu_* = \sum_{l=1}^{\infty} \frac{1}{2^{\varphi_{\mathcal{A}}(l)+2}} \leq \left(\sum_{l=1}^{\infty} \frac{1}{2^l}\right) \cdot \frac{1}{4}$$

$$= \frac{1}{4}\left(\sum_{l=1}^{\infty} \frac{1}{2^l} - 1\right) = \frac{1}{4}\left(\frac{1}{1-\frac{1}{2}} - 1\right) = \frac{1}{4}.$$

For our compound channel, for $\epsilon \in (0, \frac{1}{2}), \epsilon \in \mathbb{Q}$, we can now compute the numbers n_Δ and $\hat{n}$, which depend on the parameter p (as in the calculation above), i.e., $n_\Delta = n_\Delta(p)$ and $\hat{n} = \hat{n}(p)$. We have seen that for our compound channel, we have

$$\frac{1}{8} \leq \mu_s \leq \frac{1}{4}$$

and both $n_\Delta(\frac{1}{4})$ and $\hat{n}(\frac{1}{4})$ are valid computable numbers so that we have

$$\forall n \geq \hat{n}\left(\frac{1}{4}\right) \Rightarrow \log M_n^*(\epsilon) < nC(\mathcal{W}). \tag{26}$$

Note that in the calculation above we have $\hat{n}(\frac{1}{4}) \geq n_\Delta(\frac{1}{4}) < month >$. Since $\hat{n}(\frac{1}{4})$ is recursively computable, the proof is complete. ■

ACKNOWLEDGMENT

The author would like to thank Arogyaswami Paulraj for valuable discussions and interesting comments on jamming and interference in 5 G mobile networks and sensible applications such as Industry 4.0 or Vehicle-to-Everything (V2X).

REFERENCES

[1] H. Boche, R. F. Schaefer, and H. V. Poor, "Robust transmission over channels with channel uncertainty: An algorithmic perspective," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Barcelona, Spain, May 2020, pp. 5230–5234.

[2] G. Fettweis *et al.*, "The tactile internet," ITU-T Tech. Watch Rep., Geneva, Switzerland, Tech. Rep., Aug. 2014. [Online]. Available: https://www.itu.int/oth/T2301000023/en

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[4] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K: Cambridge Univ. Press, 2005.

[5] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[6] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proc. IEEE 56th Veh. Technol. Conf.*, Sep. 2002, pp. 87–91.

[7] M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2324–2334, Jul. 2005.

[8] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: Duality and sum-MSE minimization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5436–5446, Nov. 2007.

[9] N. Vučić and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 714–725, Feb. 2009.

[10] J. W. Kim and C. K. Un, "An adaptive array robust to beam pointing error," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1582–1584, Jun. 1992.

[11] D. D. Feldman and L. J. Griffiths, "A projection approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 867–876, Apr. 1994.

[12] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[13] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[14] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.

[15] R. G. Lorenz and S. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[16] H. Boche and M. Schubert, "Concave and convex interference functions – general characterizations and applications," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4951–4965, Oct. 2008.

[17] H. Boche and M. Schubert, "A unifying approach to interference modeling for wireless networks," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3282–3297, Jun. 2010.

[18] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[19] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[20] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.*, vol. 4, no. 4, pp. 371–386, 1960.

[21] R. Ahlswede, "The weak capacity of averaged channels," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 11, pp. 61–73, Mar. 1968.

[22] R. Ahlswede, *Rudolf Ahlswede's Lectures on Information Theory 2: Transmitting and Gaining Data*, A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Berlin, Germany: Springer International Publishing, 2015.

[23] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proc. London Math. Soc.*, vol. 2, no. 42, pp. 230–265, 1936.

[24] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem. A correction," *Proc. London Math. Soc.*, vol. 2, no. 43, pp. 544–546, 1937.

[25] K. Weihrauch, *Computable Analysis - An Introduction*. Berlin, Germany: Springer-Verlag, 2000.

[26] J. Avigad and V. Brattka, "Computability and analysis: The legacy of Alan Turing," in *Turing's Legacy: Developments from Turing's Ideas in Logic*, R. Downey, Ed. Cambridge, UK: Cambridge Univ. Press, 2014.

[27] K. Gödel, "Die Vollständigkeit der Axiome des logischen Funktionenkalküls," *Monatshefte für Mathematik*, vol. 37, no. 1, pp. 349–360, 1930.

[28] K. Gödel, "On undecidable propositions of formal mathematical systems," *Notes by Stephen C. Kleene and Barkely Rosser on Lectures at the Institute for Advanced Study, Princeton, NJ*, 1934.

[29] S. C. Kleene, *Introduction to Metamathematics*. New York, NY, USA: Van Nostrand, Wolters-Noordhoffv, 1952.

[30] M. Minsky, "Recursive unsolvability of Post's problem of 'tag' and other topics in theory of Turing machines," *Ann. Math.*, vol. 74, no. 3, pp. 437–455, 1961.

[31] H. Boche, R. F. Schaefer, and H. V. Poor, "Secure communication and identification systems – Effective performance evaluation on Turing machines," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1013–1025, 2020.

[32] H. Boche, R. F. Schaefer, and H. V. Poor, "On the structure of the capacity formula for general finite state channels with applications," in *Proc. IEEE Inf. Theory Workshop*, Visby, Sweden, Aug. 2019, pp. 1–5.

[33] H. Boche, R. F. Schaefer, and H. V. Poor, "Coding for non-iid sources and channels: Entropic approximations and a question of Ahlswede," in *Proc. IEEE Inf. Theory Workshop*, Visby, Sweden, Aug. 2019, pp. 1–5.

[34] R. I. Soare, *Recursively Enumerable Sets and Degrees*. Berlin, Germany: Springer-Verlag, 1987.

[35] M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[36] T. Rado, "On non-computable functions," *Bell Syst. Tech. J.*, vol. 41, no. 3, pp. 877–884, May 1962.

[37] E. Specker, "Nicht konstruktiv beweisbare Sätze der Analysis," *J. Symbolic Logic*, vol. 14, no. 3, pp. 145–158, Sep. 1949.

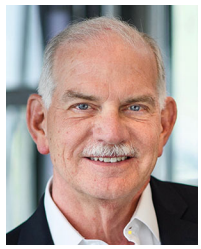
[38] S. Stanczak, M. Wiczanowski, and H. Boche, *Fundamentals of Resource Allocation in Wireless Networks*, 2nd ed. Berlin, Germany: Springer-Verlag, 2008.
[39] M. Fekete, "Über die Verteilung von Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Mathematische Zeitschrift*, vol. 17, no. 1, pp. 228–249, 1923.
[40] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
[41] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[42] T. M. Cover and J. A. Thomas, "Channel coding rate in the finite blocklength regime," *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[43] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
[44] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations Trends Commun. Inf. Theory*, vol. 11, no. 1-2, pp. 1–184, 2014.
[45] G. Wunder, R. F. H. Fischer, H. Boche, S. Litsyn, and J.-S. No, "The PAPR problem in OFDM transmission: New directions for a long-lasting problem," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 130–144, Nov. 2013.
[46] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin, Germany: Springer Verlag, 1978.
[47] L. Weiss, "On the strong converse of the coding theorem for symmetric channels without memory," *Quart Appl. Math.*, vol. 18, no. 3, pp. 209–214, Oct. 1960.
[48] R. Durrett, *Probability*, 2nd ed. London, U.K.: Duxbury Press, 1996.

Holger Boche (Fellow, IEEE) received the Dipl.-Ing. degree in electrical engineering, Graduate degree in mathematics, and the Dr.-Ing. degree in electrical engineering from the Technische Universität Dresden, Dresden, Germany, in 1990, 1992, and 1994, respectively. From 1994 to 1997, he did postgraduate studies at the Friedrich-Schiller Universität Jena, Jena, Germany. In 1998, he received the Dr. rer. nat. degree in pure mathematics from the Technische Universität Berlin, Berlin, Germany.

In 1997, he joined the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute (HHI), Berlin, Germany. From 2002 to 2010, he was Full Professor in mobile communication networks with the Institute for Communications Systems, Technische Universität Berlin. In 2003, he became the Director of the Fraunhofer German-Sino Laboratory for Mobile Communications, Berlin, and in 2004, he became the Director of the Fraunhofer Institute for Telecommunications, HHI. He was a Visiting Professor with ETH Zurich, Zurich, Switzerland, during 2004 and 2006 (Winter), and with KTH Stockholm, Stockholm, Sweden, in 2005 (Summer). Since October 2010, he has been a Full Professor with the Institute of Theoretical Information Technology, Technische Universität München, Munich, Germany. Among his publications is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Since 2014, Prof. Boche has been a member and Honorary Fellow of the TUM Institute for Advanced Study, Munich, Germany, and since 2018, a Founding Director of the Center for Quantum Engineering, Technische Universität München. He is a member of IEEE Signal Processing Society SPCOM and SPTM Technical Committees. He was Elected Member of the German Academy of Sciences (Leopoldina) in 2008 and of the Berlin Brandenburg Academy of Sciences and Humanities in 2009. He was a recipient of the Research Award "Technische Kommunikation" from the Alcatel SEL Foundation in October 2003, the "Innovation Award" from the Vodafone Foundation in June 2006, and the Gottfried Wilhelm Leibniz Prize from the Deutsche Forschungsgemeinschaft (German Research Foundation) in 2008. He was a co-recipient of the 2006 IEEE Signal Processing Society Best Paper Award and a recipient of the 2007 IEEE Signal Processing Society Best Paper Award. He was General Chair of the Symposium on Information Theoretic Approaches to Security and Privacy at IEEE GlobalSIP 2016.

Rafael F. Schaefer (Senior Member, IEEE) received the Dipl.-Ing. degree in electrical engineering and computer science from the Technische Universität Berlin, Germany, in 2007, and the Dr.-Ing. degree in electrical engineering from the Technische Universität München, Germany, in 2012. From 2013 to 2015, he was a Postdoctoral Research Fellow with Princeton University. Since 2015, he has been an Assistant Professor with the Technische Universität Berlin. Among his publications is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017). He was a recipient of the VDE Johann-Philipp-Reis Prize in 2013 and the best paper award of the German Information Technology Society (ITG-Preis) in 2016. He was one of the exemplary reviewers of the IEEE COMMUNICATION LETTERS in 2013. He is currently an Associate Editor for the IEEE TRANSACTIONS OF COMMUNICATIONS and a Consulting Editor of the IEEE OPEN JOURNAL OF SIGNAL PROCESSING. He is a member of the IEEE Information Forensics and Security Technical Committee.



H. Vincent Poor (Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer sciences from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he was the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, including most recently at Berkeley and Cambridge. His research interests include information theory, machine learning and network science, and their applications in wireless networks, energy systems and related fields. His publications in these areas includes the forthcoming book *Advanced Data Analytics for Power Systems* (Cambridge University Press, 2021).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a Foreign Member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, the 2019 ASEE Benjamin Garver Lamme Award, a D.Sc. *honoris causa* from Syracuse University, awarded in 2017, and a D.Eng. *honoris causa* from the University of Waterloo, awarded in 2019.