

Towards Secure Over-The-Air Computation

Matthias Frey,^{*} Igor Bjelaković,^{†‡} and Sławomir Stańczak^{†‡}

^{*}Department of Electrical and Electronic Engineering, The University of Melbourne, Australia, [†]Technische Universität Berlin, Germany, and [‡]Fraunhofer Heinrich Hertz Institute, Berlin, Germany

Abstract

We propose a new method to protect over-the-air (OTA) computation schemes against passive eavesdropping. Our method uses a friendly jammer whose signal is – contrary to common intuition – stronger at the legitimate receiver than it is at the eavesdropper. We focus on the computation of arithmetic averages over an additive white Gaussian noise (AWGN) channel. The derived secrecy guarantee translates to a lower bound on the eavesdropper’s mean square error while the question of how to provide operationally more significant guarantees such as semantic security remains open for future work. The key ingredients in proving the security guarantees are a known result on channel resolvability and a generalization of existing achievability results on coding for compound channels.

Index Terms

over-the-air computation, information-theoretic secrecy, compound channel, AWGN channel, friendly jamming, eavesdropper

I. INTRODUCTION

In many envisioned applications in wireless networks, the receiver requires only a function of values available at the distributed transmitters rather than the full information about the values themselves. Examples include distributed Federated Learning [1] and distributed anomaly detection in sensor networks [2]. In such cases, analog OTA computation schemes can deliver sizable performance gains over classical separation-based approaches especially when the number of transmitters is large [1]–[7].

In some of the foreseen scenarios, such as e-health or industrial applications, security and privacy are expected to be major concerns in addition to efficient resource usage. Information theoretic secrecy can complement classic cryptography in addressing these issues. A natural way to enhance security in an analog OTA computation setting such as [1]–[7] is to add a jammer to the system that deteriorates the signal-to-noise ratio (SNR) of the eavesdropper and thereby prevents it from reconstructing a low-noise estimate of the objective function. In this case, it is necessary to place the jammer so that its signal is significantly stronger at the eavesdropper than it is at the legitimate receiver. Since in general, the exact position of an eavesdropper is not known, jammers typically have to be placed at multiple locations. In this work, we propose to turn this situation around and place the jammer so that its signal is stronger at the legitimate receiver than it is at the eavesdropper. Such a setup is often easier to realize since the jammer can for instance be set up in proximity to the legitimate receiver and in certain settings, such as factory buildings, it may be feasible to assume that the attacker is located, e.g., outside of the building while the legitimate receiver and the transmitters are inside. Our proposed scheme operates under the assumption that the jamming signal is stronger at the legitimate receiver than it is at the eavesdropper. It is applicable to the special case of OTA computation of an arithmetic average over an AWGN channel. The main idea is to carefully construct a jamming signal in such a way that it can be fully reconstructed (and therefore canceled in post-processing) by the legitimate receiver, while the eavesdropper is impacted by the jamming signal as though it was white noise.

A. Prior Work

To the best of our knowledge, the OTA computation problem over a wiretap channel has not yet been considered in the literature. Therefore, in this subsection we briefly summarize the literature on the building blocks we use for the approach to the wiretap OTA computation channel that we propose in this work as well as for literature on concepts that are closely related to the ones presented in this paper.

a) OTA computation: The concept of analog OTA computation was originally introduced in [3] and further developed in [2], [4]. In [5], [7], we revisited this idea, adapted the existing scheme and provided an extension to a large class of functions and analysis of the estimation error for finite block length in a very general, fast-fading setting. There is also a digital version of OTA computation in which domain and range of the computed versions are finite, which was introduced in [8]. There are many more prior works in this area. For details, we refer the reader to the literature section in [7].

Part of this work was presented at the 2021 IEEE International Symposium on Information Theory, 12-20 July 2021, Melbourne, Victoria, Australia.

This work was supported by the German Research Foundation (DFG) within their priority program SPP 1798 “Compressed Sensing in Information Processing” and under grants STA 864/7-1 and STA 864/15-1. This work was also supported by the Federal Ministry of Education and Research of Germany in the program of “Souverän. Digital. Vernetzt.”. Joint project 6G-RIC, project identification numbers: 16KISK020K, 16KISK030.

b) Coding for compound channels: The compound channel problem was introduced independently in [9]–[11], while first independent results for the capacity expression can be found in [10], [11]. These works, however, explore mainly the case of finite input and output alphabets. The *semi-continuous* case in which only the input alphabet is assumed to be finite is briefly touched upon in [11] and studied in more detail in [12] which provides an example showing that the capacity expression from the finite case does not carry over to the semi-continuous case in general. The semi-continuous case was further explored in [13], [14]. In many cases of practical interest, the capacity expression from the finite case can be generalized to the *continuous* case in which neither input nor output alphabets are assumed to be finite, as was found in [15] for a class of Gaussian compound channels. Wiretap compound channels with finite alphabets are studied in [16]. Gaussian compound wiretap channels and related models have been investigated in [17]. However, the compound channel part in this work focuses on continuous-alphabet extensions of point-to-point compound channels.

c) Channel Resolvability and Semantic Security: The concept of channel resolvability was introduced in [18], [19]. Further results relevant in the context of this work appeared, e.g., in [20]–[23]. We use our generalization [24] for continuous channels as a basis for our proposed scheme. Although we cannot provide full semantic security guarantees in this work, we also heavily draw from the idea of obtaining semantic security by means of channel resolvability, which is developed in [25]–[28].

d) Friendly Jamming: The idea of friendly jamming has been used in [29] to aid a transmitter-receiver pair in protecting a point-to-point transmission from a passive eavesdropper. Distributed and centralized beamforming techniques are used so that the jamming signal impacts the signal-to-noise ratio at the eavesdropper but not at the legitimate receiver. Several more recent works (cf., e.g., [30]–[32]) have expanded upon this idea and refined the friendly jamming techniques. In the context of two-way wiretap channels, [33]–[35] use cooperative jamming, in which the transmitter/receiver nodes add artificial noise to their wiretap-encoded messages. In [34], channel resolvability is used to prove strong secrecy guarantees for such schemes. To the best of our knowledge, there are no prior works which use jamming to protect OTA computation against eavesdropping.

e) Physical Layer Security: The concept of information theoretic secrecy was introduced in [36] and the wiretap channel model together with a weaker, but more tractable notion of secrecy was introduced in [37]. Based on this, various stronger secrecy notions have been introduced and investigated (e.g., [27], [38], [39]). All of these existing works investigate how digitally coded transmissions can be protected against eavesdropping, while in the present work, we focus on uncoded analog transmissions over multiple-access channels.

f) Computational Wiretap Channels: [40], [41] study a system model in which a function computation is to be protected from an eavesdropper. Contrary to this work, there is only one transmitter, and the eavesdropper has the same channel output as the legitimate receiver. The security guarantee hinges upon the eavesdropper wanting to compute a function that is different from the receiver’s intended function, and one key application that is noted by the authors is therefore information-theoretic privacy.

B. Summary of the Main Contributions and Outline

The main contributions of this paper are as follows:

- 1) We propose a novel framework and result for incorporating security considerations into the OTA computation of an arithmetic average over an AWGN channel. In this framework, a friendly jammer is included in the system which deteriorates the eavesdropper’s SNR while not significantly impacting the legitimate receiver’s ability to obtain an approximation of the function value which is to be OTA computed.
- 2) In order to approach this problem, we observe a connection between the secure OTA computation problem and the problems of compound channel coding and channel resolvability for point-to-point channels. This connection is not dependent on the AWGN channel model and may therefore be useful also to establish results for more general channel models.
- 3) We prove a theorem on compound channel coding for continuous alphabets. It is a generalization of the result of the part of [15] which considers finite-dimensional Gaussian channels, and we can consequently recover this result as a special case.

In Section II, we state and prove our main result about OTA computation of an arithmetic mean over an AWGN channel. Part of the proof relies on technical results from later sections and is therefore deferred to Section IV. In Section III, we state and prove the point-to-point compound channel coding theorem that is required in the following section. In Section IV, we give the full details of the connection between the secure OTA computation problem, compound channel coding and channel resolvability that is used to establish the result of Section II. Section V concludes the paper and states open questions for future research.

Throughout the paper, we define notation where it is first used. For the reader’s convenience, a summary of notational symbols can be found in Fig. 1.

II. SYSTEM MODEL AND MAIN RESULT

In this section, we introduce the detailed system and channel model, and then proceed to state and discuss the main result of this paper. The part of the proof that requires the technical tools of later sections is only sketched here, while the full technical details are deferred to Section IV.

$f : \mathcal{S}_1 \times \dots \times \mathcal{S}_K \rightarrow \mathbb{R}$	Objective function to be approximated.
W	channel.
$\mathcal{X}_1, \dots, \mathcal{X}_K$	multiple-access channel input alphabets.
\mathcal{Y}	legitimate receiver's channel output alphabet; output alphabet of point-to-point channel.
$F^n = (F_1^n, \dots, F_K^n)$	Pre-processors for n channel uses at the transmitters.
$\mathfrak{A}_1, \dots, \mathfrak{A}_K$	transmitters.
D^n	Post-processor for n channel uses at the receiver.
\mathfrak{B}	legitimate receiver.
\tilde{f}	Estimator at the receiver for $f(s_1, \dots, s_K)$.
\mathfrak{E}	eavesdropper.
Z	eavesdropper's channel output.
\mathcal{Z}	eavesdropper's channel output alphabet.
\mathfrak{J}	jammer.
X	jammer's channel input; input of point-to-point channel.
\mathcal{X}	jammer's input alphabet; input alphabet of point-to-point channel.
Y	legitimate receiver's channel output; output of point-to-point channel.
$\ \cdot\ _{\text{TV}}$	total variation norm on the vector space of signed, finite measures.
T_1, \dots, T_K	transmitters' pre-processed channel inputs.
φ_N	probability density function of the standard normal distribution.
Φ_N	cumulative distribution function of the standard normal distribution.
$\exp(\cdot)$	exponentiation with Euler's number as basis.
$\log(\cdot)$	natural logarithm.
$\mu \ll \nu$	measure μ is absolutely continuous with respect to measure ν .
$\frac{d\mu}{d\nu}$	Radon-Nikodym derivative of μ with respect to ν .
$Q_{P,W}$	joint input-output distribution of channel W under input distribution P .
$R_{P,W}$	output distribution of channel W under input distribution P .
$\mathbf{i}_{P,W}(x^n; y^n)$	information density of the input-output pair (x^n, y^n) under a channel W with input distribution P .
$\mathbf{I}_{P,W}$	Mutual information between input and output of channel W under input distribution P .
$\mathbf{D}_\alpha(\cdot \cdot)$	Rényi divergence of order α .
$\mathbf{D}_1(\cdot \cdot)$	Kullback-Leibler divergence.
$(W_s)_{s \in \mathcal{S}}$	compound channel.
\mathcal{M}	randomness used in jamming strategy; transmitted message in compound channel.
$(\hat{W}_{\delta,j})_{j=1}^J$	sequence of channels that approximate a compound channel with error δ .
Sym_+^n	symmetric, positive semidefinite $n \times n$ matrices.
Sym_{++}^n	symmetric, positive definite $n \times n$ matrices.
N	additive channel noise.
(P, n, \mathcal{R}) -ensemble	random codebook ensemble with input distribution P , block length n and rate \mathcal{R} .
$\mathcal{C} = \mathcal{C}(m)_{m=1}^{\exp(n\mathcal{R})}$	codebook.
(c, C)	additive input cost constraint for a channel.
$W_{\mathfrak{B}}$	legitimate user's effective channel.
$W_{\mathfrak{E}}$	eavesdropper's effective channel.
$\hat{R}_{W^n, \mathcal{C}}$	distribution of the output of channel W^n if a uniformly random code word from \mathcal{C} is transmitted.

Fig. 1. Table of symbols.

A. Distributed Function Approximation with Jamming (DFA-J)

In the following, we introduce the system model for DFA-J which is an extension of the model used in [5].

Let $\mathcal{S}_1, \dots, \mathcal{S}_K$ be measurable spaces. The goal is to approximate functions $f : \mathcal{S}_1 \times \dots \times \mathcal{S}_K \rightarrow \mathbb{R}$ over a multiple-access channel W with measurable input alphabets $\mathcal{X}_1, \dots, \mathcal{X}_K$ and a measurable output alphabet \mathcal{Y} in a distributed setting. An admissible *distributed function approximation (DFA)* scheme for $f : \mathcal{S}_1 \times \dots \times \mathcal{S}_K \rightarrow \mathbb{R}$ for n channel uses is a pair (F^n, D^n) , consisting of:

- 1) A pre-processing function $F^n = (F_1^n, \dots, F_K^n)$ for the transmitters $\mathfrak{A}_1, \dots, \mathfrak{A}_K$, where each F_k^n is of the form

$$F_k^n(s_k) = (t_{k,i}(s_k, U_k(i)))_{i=1}^n \in \mathcal{X}_k^n$$

with i.i.d. random variables $U_k(1), \dots, U_k(n)$ and a measurable map

$$(s_k, u_1, \dots, u_n) \mapsto (t_{k,i}(s_k, u_i))_{i=1}^n \in \mathcal{X}_k^n.$$

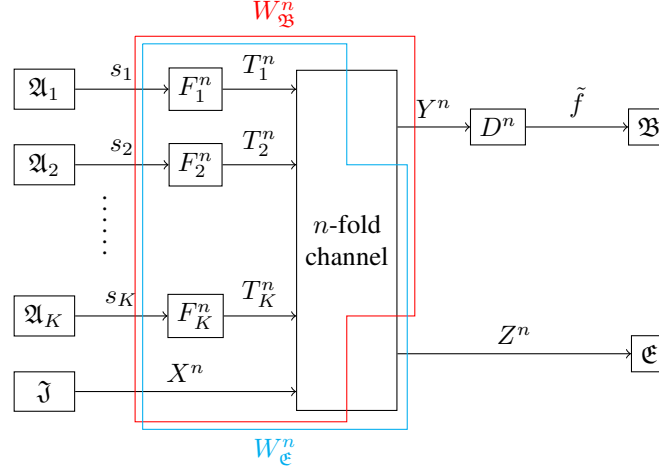


Fig. 2. System model for distributed function approximation with jamming (DFA-J) described in Section II-A.

- 2) A post-processing function D^n for the receiver \mathfrak{B} : The receiver is allowed to apply a measurable recovery function $D^n : \mathcal{X}^n \rightarrow \mathbb{R}$ upon observing the output of the channel.

Note that this, contrary to the system model in [5], imposes the restriction that the pre-processing is i.i.d. across channel uses, which will be crucial for the security extension to the approximation scheme. Although this definition of admissible schemes is slightly less general, the scheme proposed in [5] still is an admissible scheme even in this stricter sense.

So in order to approximate f , the transmitters apply their pre-processing maps to

$$(s_1, \dots, s_K) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_K$$

resulting in $F_1^n(s_1), \dots, F_K^n(s_K)$, which are sent over the channel. The receiver observes the output of the channel and applies the recovery map D^n . The whole process defines an estimate \tilde{f} of f .

Depending on the application at hand, there are multiple ways in which the quality of the estimate \tilde{f} can be quantified.

Definition 1. 1) Let $\delta, \varepsilon \in (0, 1)$ and f be given. We say that f is δ -approximated after n channel uses with confidence level ε if there is an approximation scheme (F^n, D^n) such that the resulting estimate \tilde{f} of f satisfies

$$\mathbb{P} \left(\left| \tilde{f} - f(s_1, \dots, s_K) \right| \geq \delta \right) \leq \varepsilon$$

for all $s^K := (s_1, \dots, s_K) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_K$.

- 2) We say that f is V -MSE-approximated if we have

$$\mathbb{E} \left(\left(\tilde{f} - f(s_1, \dots, s_K) \right)^2 \right) \leq V,$$

where the expectation is over the joint distribution of s_1, \dots, s_K and \tilde{f} which is induced by the distributed function approximation (DFA) scheme and the channel.

In this work, we extend the DFA system model adding an attacker \mathfrak{E} which attempts to eavesdrop on the transmission and wants to gain knowledge about s_1, \dots, s_K . At each channel use, \mathfrak{E} observes an output Z ranging over the eavesdropper's alphabet \mathcal{Z} . As a counter-measure, we add a friendly jammer \mathfrak{J} which transmits some jamming sequence X^n with the objective to prevent \mathfrak{E} from obtaining information while still allowing \mathfrak{B} to obtain a good estimate of $f(s_1, \dots, s_K)$. This extended model is depicted in Fig. 2.

Definition 2. A scheme for distributed function approximation with jamming (DFA-J) consists of:

- a DFA scheme; i.e., pre- and post-processing schemes, and
- a jamming strategy given by a probability distribution on \mathcal{X}^n .

We say that a DFA-J scheme allows reconstruction of the jamming signal with probability ϵ if there is a decoding function $\vartheta : \mathcal{Y}^n \rightarrow \mathcal{X}^n$ such that

$$\sup_{s_1 \in \mathcal{S}_1, \dots, s_K \in \mathcal{S}_K} \mathbb{P}_{s_1, \dots, s_K} (\vartheta(Y^n) \neq X^n) \leq \epsilon$$

and ϵ is the smallest number with this property.

The objective is to find admissible pre- and post-processing strategies as well as a jamming strategy such that \mathfrak{B} can obtain a good approximation \tilde{f} of $f(s_1, \dots, s_K)$ while bounding the usefulness of any information that \mathfrak{E} can obtain about s_1, \dots, s_K .

Together with the channel, a DFA-J scheme induces a probability distribution $\tilde{R}_{s_1, \dots, s_K}$ on \mathcal{Z}^n for each $(s_1, \dots, s_K) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_K$. How secure the scheme is depends on how strongly $\tilde{R}_{s_1, \dots, s_K}$ depends on s_1, \dots, s_K . In the following, we formalize this notion.

Any measurable function $g : \mathcal{S}_1 \times \dots \times \mathcal{S}_K \rightarrow \mathcal{T}$, where \mathcal{T} is a measurable space, is called an *eavesdropper's objective*.

Definition 3. 1) Given a real number $\delta \geq 0$, we say that a DFA-J scheme is δ -semantically secure if there is a probability measure μ on \mathcal{Z}^n such that for all $(s_1, \dots, s_K) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_K$,

$$\left\| \tilde{R}_{s_1, \dots, s_K} - \mu \right\|_{\text{TV}} \leq \delta, \quad (1)$$

where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm on finite signed measures. The probability measure μ can be arbitrary here except for the requirement that it is independent of s_1, \dots, s_K .

2) Let $g : \mathcal{S}_1 \times \dots \times \mathcal{S}_K \rightarrow \mathcal{T}$, where $\mathcal{T} \subseteq \mathbb{R}$ is measurable and bounded, be an eavesdropper's objective. Let $V \geq 0$ be a real number. We say that a DFA-J scheme is (g, V) -MSE-secure if under a uniform distribution of $g(s_1, \dots, s_K)$, for every estimator $d : \mathcal{Z}^n \rightarrow \mathcal{T}$, we have

$$\mathbb{E} \left((d(Z^n) - g(s_1, \dots, s_K))^2 \right) \geq V,$$

where the expectation is over the joint distribution of s_1, \dots, s_K and Z^n which results from the application of the DFA-J scheme and the channel.

If a scheme is (g, V) -MSE-secure, it means that any estimator used by the eavesdropper has a mean square error (MSE) of no less than V in the case of a uniformly distributed objective. This means that s_1, \dots, s_K are randomly distributed in such a way that $g(s_1, \dots, s_K)$ follows a uniform distribution which implies that s_1, \dots, s_K cannot be i.i.d. uniform in general. Our motivation for assuming a uniform objective instead of uniform i.i.d. s_1, \dots, s_K is the following: For many choices of g that we consider relevant (and in particular the computation of arithmetic mean on which we will focus in this paper), the function $g(s_1, \dots, s_K)$ tends to concentrate at its expectation for large values of K if s_1, \dots, s_K are independent. Since the statistical information is assumed to be known at the eavesdropper, it could therefore achieve low MSE even without intercepting any channel output. However, only the results for the AWGN channel in this section rely on such an assumption of uniformity while the technical results in Section IV assume that s_1, \dots, s_K are deterministic, but arbitrary. This means that the results of Section IV specialize to arbitrary stochastic models of s_1, \dots, s_K , and in particular to the uniform and non-independent case. Therefore, there is hope that this somewhat restrictive assumption could be lifted in future research.

In a sense made explicit by the following lemma, semantic security is the stronger of the two security notions from Definition 3.

Lemma 1. Let $\mathcal{T} := [a, b]$, let $g : \mathcal{S}_1 \times \dots \times \mathcal{S}_K \rightarrow \mathcal{T}$ be an eavesdropper's objective and $\delta \geq 0$ a real number. Assume that $g(s_1, \dots, s_K)$ is uniformly distributed on \mathcal{T} .

Then, any DFA-J scheme that is δ -semantically secure is also $(g, (1/12 - \delta)(b - a)^2)$ -MSE-secure.

Proof. Let $d : \mathcal{Z}^n \rightarrow \mathcal{T}$. Then, assuming the distribution of s_1, \dots, s_K corresponds to a uniform distribution on $[a, b]$ of $g(s_1, \dots, s_K)$, we have

$$\begin{aligned} \mathbb{E}_{s_1, \dots, s_K} \mathbb{E}_{\tilde{R}_{s_1, \dots, s_K}} \left((d(Z^n) - g(s_1, \dots, s_K))^2 \right) &= \mathbb{E}_{s_1, \dots, s_K} \int_0^{(b-a)^2} \tilde{R}_{s_1, \dots, s_K} \left((d(Z^n) - g(s_1, \dots, s_K))^2 > t \right) dt \\ &\stackrel{(1)}{\geq} \mathbb{E}_{s_1, \dots, s_K} \int_0^{(b-a)^2} \left(\mu \left((d(Z^n) - g(s_1, \dots, s_K))^2 > t \right) - \delta \right) dt \\ &= \mathbb{E}_{s_1, \dots, s_K} \mathbb{E}_{\mu} \left((d(Z^n) - g(s_1, \dots, s_K))^2 \right) - \delta(b-a)^2 \\ &\stackrel{(a)}{\geq} \mathbb{E}_{s_1, \dots, s_K} \mathbb{E}_{\mu} \left((\mathbb{E}_{\mu} g(s_1, \dots, s_K) - g(s_1, \dots, s_K))^2 \right) - \delta(b-a)^2 \\ &\stackrel{(b)}{=} \left(\frac{1}{12} - \delta \right) (b-a)^2, \end{aligned}$$

where the step (a) is because under μ , Z^n is independent of s_1, \dots, s_K , and therefore the MSE is minimized by the mean of $g(s_1, \dots, s_K)$. Finally, step (b) follows from the assumption that $g(s_1, \dots, s_K)$ is uniform on $[a, b]$, and hence its variance is known (see, e.g., [42, Example 3.4]). \square

B. AWGN Channel Model

In general, the approximation scheme even without an eavesdropper or jammer highly depends on the particular structure of the channel and f . In this work, we focus on the computation of arithmetic means over AWGN channels (although some

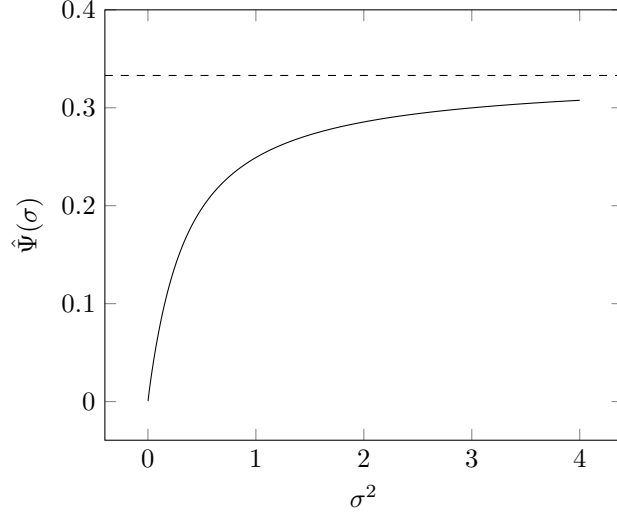


Fig. 3. Illustration of the MSE guarantees of Theorem 1. The dashed line is the MSE which an eavesdropper would have without any received signal (i.e., guessing the middle of the interval).

of the technical results we develop for this purpose hold under more general conditions). Specifically, the objective function is given as

$$f : (s_1, \dots, s_K) \mapsto \frac{1}{K} \sum_{k=1}^K s_k, \quad (2)$$

where for all k , $\mathcal{S}_k = [-1, 1]$. The channel is given by

$$Y = h_{\mathfrak{A}\mathfrak{B}} \sum_{k=1}^K T_k + h_{\mathfrak{J}\mathfrak{B}} X + N_{\mathfrak{B}} \quad (3)$$

$$Z = h_{\mathfrak{A}\mathfrak{E}} \sum_{k=1}^K T_k + h_{\mathfrak{J}\mathfrak{E}} X + N_{\mathfrak{E}}. \quad (4)$$

$N_{\mathfrak{B}}$ is centered normal with variance $\sigma_{\mathfrak{B}}^2$ and $N_{\mathfrak{E}}$ is centered normal with variance $\sigma_{\mathfrak{E}}^2$. The real channel coefficients $h_{\mathfrak{A}\mathfrak{B}}, h_{\mathfrak{J}\mathfrak{B}}, h_{\mathfrak{A}\mathfrak{E}}, h_{\mathfrak{J}\mathfrak{E}}$ are assumed deterministic and known everywhere. The channel is used n times with transmitter input sequences T_k^n for each $k \in \{1, \dots, K\}$ and X^n for the jammer. The input sequences are subject to the average power constraints

$$\frac{1}{n} \sum_{i=1}^n (T_{k,i})^2 \leq \mathfrak{P}_{\mathfrak{A}}, \quad \frac{1}{n} \sum_{i=1}^n (X_i)^2 \leq \mathfrak{P}_{\mathfrak{J}}.$$

C. Main Result

Theorem 1. Consider the wiretap channel given by (3) and (4) and the objective function f defined in (2). Assume that s_1, \dots, s_K are distributed in such a way that $f(s_1, \dots, s_K)$ is uniform in $[-1, 1]$. Define

$$\sigma_{\text{eff}, \mathfrak{B}}^2 := \frac{\sigma_{\mathfrak{B}}^2}{h_{\mathfrak{A}\mathfrak{B}}^2 K^2 \mathfrak{P}_{\mathfrak{A}}}, \quad \sigma_{\text{eff}, \mathfrak{E}}^2 := \frac{\sigma_{\mathfrak{E}}^2 + h_{\mathfrak{J}\mathfrak{E}}^2 \mathfrak{P}_{\mathfrak{J}}}{h_{\mathfrak{A}\mathfrak{E}}^2 K^2 \mathfrak{P}_{\mathfrak{A}}} \quad (5)$$

and

$$\Psi(t) := \int_0^t \int_{-\infty}^{\infty} \left(v + \frac{\varphi_N(-v) - \varphi_N(t-v)}{\Phi_N(t-v) - \Phi_N(-v)} - u \right)^2 \cdot \frac{1}{t} \varphi_N(u-v) dv du, \quad (6)$$

where φ_N denotes the probability density function and Φ_N the cumulative distribution function of the standard normal distribution, respectively. Assume that the channel from \mathfrak{J} to \mathfrak{B} is stronger than the channel from \mathfrak{J} to \mathfrak{E} , i.e., $h_{\mathfrak{J}\mathfrak{B}}/\sigma_{\mathfrak{B}} > h_{\mathfrak{J}\mathfrak{E}}/\sigma_{\mathfrak{E}}$. Then there is a DFA-J scheme and there are constants $\gamma_1, \gamma_2 > 0$ such that for sufficiently large n , the following hold:

- \mathfrak{B} can approximate the objective function $f(s_1, \dots, s_K)$ with a MSE not exceeding

$$\sigma_{\text{eff}, \mathfrak{B}}^2 \Psi\left(\frac{2}{\sigma_{\text{eff}, \mathfrak{B}}}\right) + \exp(-n\gamma_1) \quad (7)$$

- The scheme is (f, V) -MSE-secure, where

$$V := \sigma_{\text{eff}, \mathfrak{E}}^2 \Psi \left(\frac{2}{\sigma_{\text{eff}, \mathfrak{E}}} \right) - \exp(-n\gamma_2). \quad (8)$$

Here and for the remainder of the paper, the functions $\exp(\cdot)$ and $\log(\cdot)$ both use Euler's number as their basis. In order to understand the impact of the function Ψ that appears in the security and approximation guarantees, we refer the reader to the plot of the function $\hat{\Psi} : \sigma \mapsto \sigma^2 \Psi(2/\sigma)$ in Fig. 3.

The proof of Theorem 1 is divided into two main steps:

- In Lemma 2, we examine a scenario where the jammer employs white noise as a jamming signal. The instantaneous channel inputs of the jammer are known to \mathfrak{B} but not to \mathfrak{E} . We establish MSE-security and MSE-approximation guarantees for this case.
- We show that there is a jamming strategy induced by a suitable code book which guarantees that \mathfrak{B} can reconstruct the jammer's channel input with high probability, but for \mathfrak{E} , the jamming strategy resembles the case of white noise. The MSE-security and MSE-approximation guarantees then follow via a comparison to the case of Lemma 2. This part of the proof is deferred to the end of Section IV, after all necessary technical ingredients are introduced. The reconstruction of the jamming signal at \mathfrak{B} is based on the compound channel coding result in Section III, and the resemblance of white noise at \mathfrak{E} is based on a known channel resolvability result.

Lemma 2. Consider the wiretap channel given by (3) and (4) and the objective function f defined in (2). Assume that $f(s_1, \dots, s_K)$ is uniformly distributed on \mathcal{T} . Furthermore, suppose that the jamming sequence X^n is i.i.d. centered Gaussian with variance \mathfrak{P}_3 and that it is known at the legitimate receiver while the eavesdropper has only statistical information. Then, under the definitions (5) and (6), there is a DFA-J scheme which is $(f, \sigma_{\text{eff}, \mathfrak{E}}^2 \Psi(2/\sigma_{\text{eff}, \mathfrak{E}}))$ -MSE-secure and $(\sigma_{\text{eff}, \mathfrak{B}}^2 \Psi(2/\sigma_{\text{eff}, \mathfrak{B}}))$ -MSE-approximates f at the receiver.

The proof of Lemma 2 is based on a few facts from statistics. We only state the relevant lemmas here. Since they are straightforward consequences of elementary known facts about minimum MSE estimators, we expect that they are folklore in the field of statistics, however, we are not aware of any reference that states these facts in the form in which we need them for our proof. Therefore, we include the proofs of the following two lemmas in the appendix for the sake of completeness.

Lemma 3. If U is distributed uniformly on $[a, b]$ and, conditioned on U , V_1, \dots, V_n are i.i.d. normally distributed with mean U and variance σ^2 , then the minimum MSE estimator for estimating U from the observations V_1, \dots, V_n is

$$\hat{U} := \bar{V} + \frac{\sigma}{\sqrt{n}} \cdot \frac{\varphi_N \left(\frac{a - \bar{V}}{\sigma/\sqrt{n}} \right) - \varphi_N \left(\frac{b - \bar{V}}{\sigma/\sqrt{n}} \right)}{\Phi_N \left(\frac{b - \bar{V}}{\sigma/\sqrt{n}} \right) - \Phi_N \left(\frac{a - \bar{V}}{\sigma/\sqrt{n}} \right)}, \quad (9)$$

where $\bar{V} := \frac{1}{n} \sum_{i=1}^n V_i$.

Lemma 4. Under the assumptions of Lemma 3, the estimator \hat{U} satisfies

$$\mathbb{E} \left((U - \hat{U})^2 \right) = \frac{\sigma^2}{n} \Psi \left(\frac{b - a}{\sigma/\sqrt{n}} \right),$$

with Ψ as defined in (6).

Proof of Lemma 2. We use the following transmission strategy:

$$X_i : \text{Gaussian with mean 0 and variance } \mathfrak{P}_3, \quad (10)$$

$$F_k^n : s_k \mapsto (1, \dots, 1) \cdot s_k \sqrt{\frac{\mathfrak{P}_2}{n}} \quad (11)$$

The receiver can obtain

$$\begin{aligned} Y'_i &:= \frac{Y_i - h_{3\mathfrak{B}} X_i}{h_{2\mathfrak{B}} K \sqrt{\mathfrak{P}_2/n}} \\ &= \frac{h_{2\mathfrak{B}} \sum_{k=1}^K T_k + N_{\mathfrak{B},i}}{h_{2\mathfrak{B}} K \sqrt{\mathfrak{P}_2/n}} \\ &= \frac{h_{2\mathfrak{B}} \sum_{k=1}^K s_k \sqrt{\mathfrak{P}_2/n} + N_{\mathfrak{B},i}}{h_{2\mathfrak{B}} K \sqrt{\mathfrak{P}_2/n}} \\ &= f(s_1, \dots, s_K) + \frac{N_{\mathfrak{B},i}}{h_{2\mathfrak{B}} K \sqrt{\mathfrak{P}_2/n}}. \end{aligned}$$

We define the post-processing operation D^n at the receiver as first obtaining Y'_1, \dots, Y'_n and then computing the minimum MSE estimator from Lemma 3. With this choice, Lemma 4 yields the claimed reconstruction error guarantee.

On the other hand, the output at \mathfrak{E} is given by

$$\begin{aligned} Z_i &\stackrel{(4)}{=} h_{\mathfrak{A}\mathfrak{E}} \sum_{k=1}^K T_k + h_{\mathfrak{J}\mathfrak{E}} X_i + N_{\mathfrak{E},i} \\ &\stackrel{(11)}{=} h_{\mathfrak{A}\mathfrak{E}} \sum_{k=1}^K s_k \sqrt{\mathfrak{P}_{\mathfrak{A}}/n} + h_{\mathfrak{J}\mathfrak{E}} X_i + N_{\mathfrak{E},i} \\ &\stackrel{(2)}{=} f(s_1, \dots, s_K) \cdot K \sqrt{\mathfrak{P}_{\mathfrak{A}}/n} h_{\mathfrak{A}\mathfrak{E}} + h_{\mathfrak{J}\mathfrak{E}} X_i + N_{\mathfrak{E},i}. \end{aligned}$$

From this, Lemmas 3 and 4 yield the claimed MSE-security of the scheme. \square

D. Special case $K = 1$

We conclude this section with a brief discussion of the important special case $K = 1$. While one of the main motivations of the methods developed in this paper is their scalability to large values of K , the case of low values of K can also be interesting in many practical applications and be instructive to understand the nature of our results better.

For the special case of only a single transmitter ($K = 1$), the problem reduces to a point-to-point transmission of the real number $f(s_1)$ in the presence of an eavesdropper and a friendly jammer. In our results in this paper, there is no assumption that K has to be large; in particular, they remain applicable also when $K = 1$. However, since in this case no function of *distributed* values has to be computed over the channel, it is possible to separately source and channel encode $f(s_1)$. After the source coding step has been performed, the remaining problem is very similar to jammer-aided secret communication as treated for instance in [29]–[31].

But although this approach is applicable to the same communication task, it is important to note that the way in which the friendly jammer has to be placed differs significantly. In the approach of this paper, the jamming signal has to be stronger at the legitimate receiver than it is at the eavesdropper. As long as this condition is satisfied, the legitimate receiver has the ability to almost completely cancel the jamming signal. This means that our method remains applicable even if the gap in terms of jammer signal strength between the legitimate receiver and the eavesdropper is relatively small. In [29]–[31], on the other hand, it is necessary that the jamming signal is stronger at the eavesdropper than it is at the legitimate receiver. Moreover, this gap between signal strengths has to be as large as possible since the jammer's signal strength at the legitimate receiver diminishes the capacity of the main channel. Therefore, our results in this case are more suitable for scenarios where it is possible to assure a high jamming signal strength at the legitimate receiver while results from [29]–[31] are more suitable in cases where all possible eavesdropper locations can be covered with strong jamming signals that have very low strength at the location of the legitimate receiver.

With respect to the open research questions given in Section V, we remark that methods from the literature can be used to achieve semantic security with slight adaptations; such a construction is for instance sketched in [43]. We are not aware of practically feasible schemes that achieve semantic security, but we expect that weaker guarantees such as MSE security could be derived, e.g., for the approach given in [44]. In order to accommodate a friendly jammer in the system model, all of these approaches would need to be combined with the works on friendly jamming discussed above. Therefore, it would remain necessary to also have the assumption that the jamming signal is significantly stronger at the eavesdropper than it is at the legitimate receiver. For the case in which this assumption is reversed as in the present work, to the best of our knowledge these questions remain open even for $K = 1$.

III. CODING FOR THE COMPOUND CHANNEL

In this section, we state and prove a coding result for compound channels with continuous alphabets. This result is used in the proof of Theorem 4 which in turn is a technical contribution needed to prove Theorem 1. Although similar to results already available in the literature, it is slightly more general and may therefore also be of independent interest.

A. System Model and Preliminary Definitions

We begin with some preliminary notations and definitions. Given measures μ and ν , we say that μ is absolutely continuous with respect to ν , or $\mu \ll \nu$, if all ν -null sets are also μ -null sets. If we have $\mu \ll \nu$, then the Radon-Nikodym derivative $\frac{d\mu}{d\nu}$ exists, which is an (up to a ν -null set) uniquely determined function with the property $\int_S \frac{d\mu}{d\nu} d\nu = \mu(S)$ for all measurable sets S .

For any channel W , we denote the joint input-output distribution under P and W by $Q_{P,W}$ and the marginal for \mathcal{Y} by $R_{P,W}$. Since we use Euler's number as the basis of the functions \exp and \log , all of the information quantities defined in the

following are given in nats. We define the *information density* of tuples of elements of the input and output alphabets under the channel W and an input distribution P as

$$\mathbf{i}_{P,W}(x^n; y^n) := \log \frac{dW^n(x^n, \cdot)}{dR_{P,W}^n}(y^n).$$

By convention, if $W^n(x^n, \cdot) \not\ll R_{P,W}^n$, the information density is ∞ .

Correspondingly, the *mutual information* is defined as

$$\mathbf{I}_{P,W} := \mathbb{E}_{Q_{P,W}} \mathbf{i}_{P,W}(X; Y).$$

Although the integrand is guaranteed to not be $\pm\infty$ on non-null sets, the mutual information integral can be infinite. Moreover, given two probability measures μ and ν , we define the *Rényi divergence* of order $\alpha \in (0, 1) \cup (1, \infty)$ between them as

$$\mathbf{D}_\alpha(\mu || \nu) := \frac{1}{\alpha - 1} \log \mathbb{E}_\mu \left(\left(\frac{d\mu}{d\nu} \right)^{\alpha-1} \right).$$

Again, by convention, the Rényi divergence is ∞ if $\mu \not\ll \nu$. $\mathbf{D}_1(\mu || \nu) := \lim_{\alpha \nearrow 1} \mathbf{D}_\alpha(\mu || \nu)$ is the *Kullback-Leibler divergence*.

A *compound channel* is a family $(W_s)_{s \in \mathcal{S}}$ of memoryless time-discrete point-to-point channels with common input alphabet \mathcal{X} and output alphabet \mathcal{Y} . The transmitter's channel input is passed through a fixed W_s for the entire block length, but the transmitter does not control the choice of s , nor is it governed by a probability distribution. In this work, we assume neither the transmitter nor the receiver knows s . A *compound channel code* with block length n and rate \mathcal{R} consists of an encoder $e : \{1, \dots, \exp(n\mathcal{R})\} \rightarrow \mathcal{X}^n$ and a decoder $d : \mathcal{X}^n \rightarrow \{1, \dots, \exp(n\mathcal{R})\}$. We say that it has error probability ϵ if under a uniform distribution of $\mathcal{M} \in \{1, \dots, \exp(n\mathcal{R})\}$, the following is true: Let Y^n be constructed by passing the components of $X^n := e(\mathcal{M})$ independently through W_s . Then, we have

$$\sup_{s \in \mathcal{S}} \mathbb{E}_{\mathcal{M}} \mathbb{P}_s(m \neq d(Y^n)) \leq \epsilon,$$

where ϵ is the smallest number with this property.

Our proof of Theorem 1 hinges on coding for a particular class of Gaussian compound channels. Such channels have continuous input and output alphabets, so we need an achievability result for compound channels with continuous input and output alphabets. As mentioned in Section I, it is shown in [12] that even in the case that only the output alphabet is countably infinite, the capacity expressions from the finite case [10], [11] do not carry over. It is therefore clear that an additional assumption on the compound channel is needed. In existing literature (e.g., [10], [15]), the problem is often approached by proving that the compound channel can be approximated by a finite class of channels in which case classical channel coding techniques such as joint typicality decoding can be adapted in a straightforward manner. In this work, we choose to directly pose the approximability of the compound channel by a finite class of channels as an assumption of our coding theorem.

Definition 4. Given a compound channel $(W_s)_{s \in \mathcal{S}}$ with input alphabet \mathcal{X} and output alphabet \mathcal{Y} , we say that it can be (δ, J) -approximated under a probability distribution P on \mathcal{X} if there is a sequence $(\hat{W}_{\delta,j})_{j=1}^J$ of channels from \mathcal{X} to \mathcal{Y} such that for every $s \in \mathcal{S}$, there is $j \in \{1, \dots, J\}$ such that

$$\mathbb{E}_P \mathbf{D}_1(W_s(X, \cdot) || \hat{W}_{\delta,j}(X, \cdot)) \leq \delta \quad (12)$$

$$\exists \alpha > 1 \forall x \in \mathcal{X} : \mathbf{D}_\alpha(W_s(x, \cdot) || \hat{W}_{\delta,j}(x, \cdot)) < \infty \quad (13)$$

$$\forall x \in \mathcal{X} : \hat{W}_{\delta,j}(x, \cdot) \ll W_s(x, \cdot) \quad (14)$$

$$\mathbf{I}_{P, \hat{W}_{\delta,j}} - \mathbf{I}_{P, W_s} \leq \delta, \quad (15)$$

and for every $j \in \{1, \dots, J\}$ there is $s \in \mathcal{S}$ such that

$$\mathbf{I}_{P, W_s} - \mathbf{I}_{P, \hat{W}_{\delta,j}} \leq \delta. \quad (16)$$

Conditions (12) and (15) tell us in what sense the approximating channel must be similar to the approximated channel, while the remaining conditions are of a more technical nature. (13) and (14) ensure that all moment-generating functions and Radon-Nikodym derivatives that we use to derive the exponential error bounds exist. Finally, (16) tells us that in a certain sense, the approximating sequence cannot be too rich, and it can usually be ensured that it holds by not including unnecessary channels in the sequence.

B. Feasibility of Channel Approximation

In this subsection, we provide some tools and examples to argue that many compound channels of practical interest can indeed be (δ, J) -approximated so that Theorem 3 may be applied to them. In particular, the results in this section imply that Theorem 3 can be applied to the class of Gaussian channels we need to prove Theorem 1.

We begin with an observation that shows that the approximability criterion of Definition 4 is a generalization of the assumption of finite channel alphabets that is used in [10], [11].

Remark 1. [10, Lemma 4] implies that for every compound channel $(W_s)_{s \in \mathcal{S}}$ with finite input and output alphabets and every $\delta > 0$, there is an integer $J(\delta)$ such that $(W_s)_{s \in \mathcal{S}}$ can be $(\delta, J(\delta))$ -approximated.

We repeat the construction here and discuss how this fact is proved.

Let M be an integer which satisfies

$$M \geq \max \left(\frac{4|\mathcal{Y}|^3}{\delta^2}, \frac{2|\mathcal{Y}|^2}{\delta} \right).$$

Given $s \in \mathcal{S}$, we construct a channel W'_s . To this end, given any $x \in \mathcal{X}$, we fix an enumeration $(y_k)_{k=1}^{|\mathcal{Y}|}$ such that the finite sequence $(W_s(x, \{y_n\}))_{k=1}^{|\mathcal{Y}|}$ is nondecreasing. For every $k < |\mathcal{Y}|$, we can then uniquely choose a value for $W'_s(x, \{y_n\})$ such that it is an integer multiple of $1/M$ and

$$W_s(x, \{y_n\}) \leq W'_s(x, \{y_n\}) < W_s(x, \{y_n\}) + \frac{1}{M}. \quad (17)$$

It is argued in [10] that this leaves a positive probability mass for $W'_s(x, \{y_{|\mathcal{Y}|}\})$ and therefore, this construction fully defines a channel W'_s . We define the approximation sequence $(\hat{W}_{\delta,j})_{j=1}^{J(\delta)}$ as an enumeration of the set $\{W'_s : s \in \mathcal{S}\}$. The cardinality of this set is upper bounded by $(M+1)^{|\mathcal{X}||\mathcal{Y}|}$ since all singleton probabilities are integer multiples of $1/M$.

For finite alphabets, (13) is trivially satisfied since Rényi divergence is in this case always finite [45]. Regarding the absolute continuity criterion (14), we recall that $W'_s(x, \{y_{|\mathcal{Y}|}\})$ always has a positive probability, and for $k < |\mathcal{Y}|$, the assumption $W_s(x, \{y_n\}) = 0$ immediately implies $W'_s(x, \{y_{|\mathcal{Y}|}\}) = 0$ by (17), since 0 is the only integer multiple of $1/M$ which is strictly smaller than $1/M$. The proof in [10] exploits (17) to prove that the absolute difference between the information of W_s and W'_s under any input distribution is at most $2|\mathcal{Y}|^3 M^{-1/2}$ (statement (c) of the lemma) which by our choice of M immediately implies (15) and (16). Moreover, it is shown that (17) also implies that for all $x \in \mathcal{X}, y \in \mathcal{Y}$,

$$\log \frac{W_s(x, \{y\})}{W'_s(x, \{y\})} \leq \frac{2|\mathcal{Y}|^2}{M}$$

(statement (b) of the lemma) which by our choice of M implies (12).

For many channels of interest, $(\delta, J(\delta))$ -approximability can be shown directly by going through properties (12) – (16). However, it is often easier to make an argument involving topological properties of \mathcal{S} . The following lemma provides some machinery to this end. In its statement, we use the following generalization of continuity of functions: Let \mathcal{S} be a topological space. A function $f : \mathcal{S} \rightarrow [-\infty, \infty]$ is called *upper semi-continuous* at s_0 if for every $t_0 > f(s_0)$ there exists an open set $S \subseteq \mathcal{S}$ with $s_0 \in S$ and for all $s \in S$, $t_0 > f(s)$. f is called *lower semi-continuous* at s_0 if $-f$ is upper semi-continuous at s_0 . f is called *upper semi-continuous* (lower semi-continuous) if it is upper semi-continuous (lower semi-continuous) at every point of its domain.

Lemma 5. Let $(W_s)_{s \in \mathcal{S}}$ be a compound channel with input alphabet \mathcal{X} and output alphabet \mathcal{Y} , let P be a probability distribution on \mathcal{X} and assume that there is a topology on \mathcal{S} such that \mathcal{S} is compact and

$$\forall s_0 \in \mathcal{S} : s \mapsto \mathbb{E}_P \mathbf{D}_1(W_s(X, \cdot) || W_{s_0}(X, \cdot)) \text{ is upper semi-continuous at } s_0, \quad (18)$$

$$\forall s_1, s_2 \in \mathcal{S} \exists \alpha > 1 \forall x \in \mathcal{X} : \mathbf{D}_\alpha(W_{s_1}(x, \cdot) || W_{s_2}(x, \cdot)) < \infty, \quad (19)$$

$$s \mapsto \mathbf{I}_{P, W_s} \text{ is lower semi-continuous.} \quad (20)$$

Then, for any $\delta > 0$, there is $J(\delta)$ such that $(W_s)_{s \in \mathcal{S}}$ can be $(\delta, J(\delta))$ -approximated under P .

Proof. Fix some $\delta > 0$. For a given $s \in \mathcal{S}$, consider

$$\{s' : \mathbb{E}_P \mathbf{D}_1(W_{s'}(X, \cdot) || W_s(X, \cdot)) < \delta\} \cap \{s' : \mathbf{I}_{P, W_s} - \mathbf{I}_{P, W_{s'}} < \delta\}.$$

Clearly, (18) and (20) ensure that this intersection is a neighborhood of s , so we can find an open neighborhood \mathcal{A}_s contained in it. Thus, $(\mathcal{A}_s)_{s \in \mathcal{S}}$ is an open cover of \mathcal{S} and therefore, the compactness of \mathcal{S} yields a finite subcover $\mathcal{A}_{s_1}, \dots, \mathcal{A}_{s_{J(\delta)}}$. We set $\hat{W}_{\delta,j} := W_{s_j}$ and given any $s \in \mathcal{S}$, we choose j such that $s \in \mathcal{A}_{s_j}$ and argue that $\hat{W}_{\delta,j}$ satisfies (12), (13) and (15). To this end, we note that (13) and (14) follow from (19), while (12) and (15) are ensured by the definition of \mathcal{A}_{s_j} . Finally, (16) is trivially satisfied, concluding the proof. \square

We now make use of Lemma 5 to prove that a large class of Gaussian fading multiple-input and multiple-output channels can actually be $(\delta, J(\delta))$ -approximated and thus Theorem 3 can be applied to them. The class of compound channels covered in the following theorem contains the class considered in [15, Sections 3 and 4] as a proper subset. We denote the set of symmetric, positive semidefinite $n \times n$ -matrices with Sym_+^n and the set of symmetric, positive definite $n \times n$ -matrices with Sym_{++}^n .

Theorem 2. *Let $\mathcal{X} = \mathbb{R}^j$, $\mathcal{Y} = \mathbb{R}^i$, let \mathcal{S} be a compact subset of $\mathbb{R}^{ij} \times \text{Sym}_+^{ij} \times \mathbb{R}^i \times \text{Sym}_{++}^i$ (under the topology induced by the Frobenius norm). For any $s = (\mu_H, \Sigma_H, \mu_N, \Sigma_N) \in \mathcal{S}$, let W_s be the channel given by*

$$Y = HX + N,$$

where the channel input X has range \mathbb{R}^j , the channel output Y has range \mathbb{R}^i , the entries of the $i \times j$ fading matrix H follow a multivariate normal distribution with mean μ_H and covariance matrix Σ_H and the additive noise N is independent of H and follows a multivariate normal distribution with mean μ_N and covariance matrix Σ_N . Let P be a distribution on \mathcal{X} and assume that either P is a multivariate Gaussian with positive definite covariance matrix or that the support of P is contained in some compact set. Then, given any $\delta > 0$, there is $J(\delta)$ such that $(W_s)_{s \in \mathcal{S}}$ can be $(\delta, J(\delta))$ -approximated under P .

Proof. We show that the conditions of Lemma 5 are met. [46] provides closed-form expressions for Rényi and Kullback-Leibler divergences between multivariate normal distributions. The only fact that we are going to use and which is apparent from these expressions, however, is that the Rényi and Kullback-Leibler divergences between two multivariate normal distributions are finite and continuous in the mean vectors and covariance matrices of the distributions wherever the covariance matrices are positive definite or, equivalently, both distributions are absolutely continuous with respect to the Lebesgue measure.

$\Sigma_N \in \text{Sym}_{++}^i$ and therefore, given any $x \in \mathcal{X}$, $W_s(x, \cdot)$ is absolutely continuous with respect to the Lebesgue measure and thus has a positive definite covariance matrix and a density $p_{W_s(x, \cdot)}$, which implies (19).

Next, from the well-known closed-form expression of the multivariate normal density, we know that for any x and y , $p_{W_s(x, \cdot)}(y)$ is continuous in s . The boundedness of \mathcal{S} implies a uniform upper bound on $p_{W_s(x, \cdot)}(y)$, so we can use the theorem of dominated convergence to argue that the marginal density $p_{R_P, W_s}(y) = \mathbb{E}_P p_{W_s(X, \cdot)}(y)$ depends continuously on s for any fixed y . We write

$$\mathbf{I}_{P, W_s} = \mathbb{E}_{P R_P, W_s} \left(\frac{p_{W_s(X, \cdot)}(Y)}{p_{R_P, W_s}(Y)} \log \frac{p_{W_s(X, \cdot)}(Y)}{p_{R_P, W_s}(Y)} \right).$$

Since the integrand is lower bounded by $-\exp(-1)$, (20) follows as an application of Fatou's lemma.

Finally, in order to argue (18), we distinguish between the two cases in the statement of the theorem.

First, suppose that there is a compact subset $\hat{\mathcal{X}} \subseteq \mathcal{X}$ with $P(\mathcal{X} \setminus \hat{\mathcal{X}}) = 0$. For any fixed s_0 , the map

$$(s, x) \mapsto \mathbf{D}_1(W_s(x, \cdot) || W_{s_0}(x, \cdot))$$

is continuous, therefore the image of $\mathcal{S} \times \hat{\mathcal{X}}$ is compact and hence bounded. We can therefore invoke the theorem of dominated convergence and argue that (18) is satisfied.

Now, suppose that P is multivariate Gaussian with positive definite covariance matrix. We write

$$\begin{aligned} \mathbb{E}_P \mathbf{D}_1(W_s(X, \cdot) || W_{s_0}(X, \cdot)) &= \mathbb{E}_P \mathbb{E}_{W_s(X, \cdot)} \log \frac{p_P(X) p_{W_s(X, \cdot)}(Y)}{p_P(X) p_{W_{s_0}(X, \cdot)}(Y)} \\ &= \mathbb{E}_{Q_{P, W_s}} \log \frac{p_{Q_{P, W_s}}(X, Y)}{p_{Q_{P, W_{s_0}}}(X, Y)} \\ &= \mathbf{D}_1(Q_{P, W_s} || Q_{P, W_{s_0}}). \end{aligned}$$

From our arguments above, given any s , the distribution Q_{P, W_s} is multivariate Gaussian with positive definite covariance matrix, which implies that (18) is satisfied. \square

C. Coding Result

We use the same random codebook construction that was originally employed by Shannon [47]: Given a channel input alphabet \mathcal{X} , a distribution P on \mathcal{X} , a block length n and a rate \mathcal{R} , we define the (P, n, \mathcal{R}) -ensemble of code books as a random experiment in which $\exp(n\mathcal{R})$ code words of length n are drawn randomly and independently according to P for each component of each code word.

Theorem 3. *Let $(W_s)_{s \in \mathcal{S}}$ be a compound channel with input alphabet \mathcal{X} and output alphabet \mathcal{Y} , and let P be a probability distribution on \mathcal{X} such that for every $\delta > 0$, there is a $J(\delta)$ such that $(W_s)_{s \in \mathcal{S}}$ can be $(\delta, J(\delta))$ -approximated under P . Let*

$$0 < \mathcal{R} < \inf_{s \in \mathcal{S}} \mathbf{I}_{P, W_s}, \quad (21)$$

and let \mathcal{C} be a random codebook from the (P, n, \mathcal{R}) -ensemble. Define an encoder $m \mapsto \mathcal{C}(m)$. Then there is a decoder such that the error probability ϵ of the resulting compound channel code satisfies

$$\mathbb{E}_{\mathcal{C}}(\epsilon) < \exp(-n\gamma), \quad (22)$$

for some $\gamma > 0$ and sufficiently large n .

Proof. We first pick parameters $\delta, \varepsilon, \beta_1$ and β_2 in sequence according to the following scheme, where (21) and the previous choices ensure that these intervals are all nonempty.

$$\delta \in \left(0, \frac{\inf_{s \in \mathcal{S}} \mathbf{I}_{P, W_s} - \mathcal{R}}{3}\right) \quad (23)$$

$$\varepsilon \in \left(2\delta, \inf_{s \in \mathcal{S}} \mathbf{I}_{P, W_s} - \mathcal{R} - \delta\right) \quad (24)$$

$$\beta_1 \in (\delta, \varepsilon - \delta) \quad (25)$$

$$\beta_2 \in (0, \varepsilon - \delta - \beta_1) \quad (26)$$

Fix a sequence $(\hat{W}_{\delta, j})_{j=1}^{J(\delta)}$ which $(\delta, J(\delta))$ -approximates $(W_s)_{s \in \mathcal{S}}$.

We use a joint typicality decoder, i.e., if there is a unique m such that

$$\exists j \in \{1, \dots, J(\delta)\} : \mathbf{i}_{P, \hat{W}_{\delta, j}}(\mathcal{C}(m); Y^n) \geq n(\mathbf{I}_{P, \hat{W}_{\delta, j}} - \varepsilon),$$

the decoder declares that message m has been sent; otherwise it declares an error (or that message 1 has been sent).

We denote the transmitted message with \mathcal{M} , the message declared by the decoder with $\hat{\mathcal{M}}$ and define error events

$$\mathcal{E} := \{\mathcal{M} \neq \hat{\mathcal{M}}\} \quad (27)$$

$$\mathcal{E}_1 := \left\{ \forall j \in \{1, \dots, J(\delta)\} \mathbf{i}_{P, \hat{W}_{\delta, j}}(\mathcal{C}(\mathcal{M}); Y^n) < n(\mathbf{I}_{P, \hat{W}_{\delta, j}} - \varepsilon) \right\} \quad (28)$$

$$\mathcal{E}_2 := \left\{ \exists m \neq \mathcal{M} \exists j \in \{1, \dots, J(\delta)\} \mathbf{i}_{P, \hat{W}_{\delta, j}}(\mathcal{C}(m); Y^n) \geq n(\mathbf{I}_{P, \hat{W}_{\delta, j}} - \varepsilon) \right\}. \quad (29)$$

We note that $\mathcal{E} \subseteq \mathcal{E}_1 \cup \mathcal{E}_2$ and consequently

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2). \quad (30)$$

So we can bound these two errors separately and then combine them.

We start with bounding the expectation of the first summand, using the definition (28) and \mathcal{C} , as well as an addition of zero. Pick j such that $\hat{W}_{\delta, j}$ satisfies (12) – (15) with respect to the realization W_s of the compound channel. Then we have

$$\begin{aligned} \mathbb{E}_{\mathcal{C}}(\mathbb{P}(\mathcal{E}_1)) &\leq \mathbb{E}_{\mathcal{C}} \left(\mathbb{P} \left(\mathbf{i}_{P, \hat{W}_{\delta, j}}(\mathcal{C}(\mathcal{M}); Y^n) < n(\mathbf{I}_{P, \hat{W}_{\delta, j}} - \varepsilon) \right) \right) \\ &= Q_{P, W_s}^n \left(\mathbf{i}_{P, \hat{W}_{\delta, j}}(X^n; Y^n) < n(\mathbf{I}_{P, \hat{W}_{\delta, j}} - \varepsilon) \right) \\ &= Q_{P, W_s}^n \left(\sum_{i=1}^n \log \left(\frac{d\hat{W}_{\delta, j}(X_i, \cdot)}{dR_{P, \hat{W}_{\delta, j}}}(Y_i) \right) n(\mathbf{I}_{P, \hat{W}_{\delta, j}} + \mathbf{I}_{P, W_s} - \mathbf{I}_{P, W_s} - \varepsilon) \right) \end{aligned} \quad (31)$$

The Radon-Nikodym derivative can be split as

$$\frac{d\hat{W}_{\delta, j}(X_i, \cdot)}{dR_{P, \hat{W}_{\delta, j}}} = \frac{d\hat{W}_{\delta, j}(X_i, \cdot)}{dW_s(X_i, \cdot)} \cdot \frac{dR_{P, W_s}}{dR_{P, \hat{W}_{\delta, j}}} \cdot \frac{dW_s(X_i, \cdot)}{dR_{P, W_s}}. \quad (32)$$

This is possible because $\hat{W}_{\delta, j}(x, \cdot) \ll W_s(x, \cdot)$ by (14), $R_{P, W_s} \ll R_{P, \hat{W}_{\delta, j}}$ by (12) and the joint convexity of Kullback-Leibler divergence in its arguments, and $W_s(x, \cdot) \ll R_{P, W_s}$ for P -almost all x by the properties of the marginalization.

We next bound tail probabilities corresponding to the three factors in (32) separately, starting with the first. To this end, we introduce a number $\alpha_1 > 1$ and argue, using Markov's inequality and the definition of Rényi divergence, that

$$\begin{aligned}
Q_{P,W_s}^n \left(\sum_{i=1}^n \log \frac{dW_s(X_i, \cdot)}{d\hat{W}_{\delta,j}(X_i, \cdot)}(Y_i) \geq n\beta_1 \right) &= Q_{P,W_s}^n \left(\exp \left((\alpha_1 - 1) \sum_{i=1}^n \log \frac{dW_s(X_i, \cdot)}{d\hat{W}_{\delta,j}(X_i, \cdot)}(Y_i) \right) \exp((\alpha_1 - 1)n\beta_1) \right) \\
&\leq \mathbb{E}_{Q_{P,W_s}^n} \left(\left(\prod_{i=1}^n \left(\frac{dW_s(X_i, \cdot)}{d\hat{W}_{\delta,j}(X_i, \cdot)}(Y_i) \right)^{\alpha_1 - 1} \right) \right) \cdot \exp(-(\alpha_1 - 1)n\beta_1) \\
&= \exp \left(\sum_{i=1}^n \log \left(\mathbb{E}_{Q_{P,W_s}^n} \left(\left(\frac{dW_s(X_i, \cdot)}{d\hat{W}_{\delta,j}(X_i, \cdot)} \right)^{\alpha_1 - 1} \right) \right) \right) \cdot \exp(-(\alpha_1 - 1)n\beta_1) \\
&= \exp \left(-(\alpha_1 - 1)n \cdot \left(\beta_1 - \mathbb{E}_P \mathbf{D}_{\alpha_1} \left(W_s(X, \cdot) \parallel \hat{W}_{\delta,j}(X, \cdot) \right) \right) \right). \tag{33}
\end{aligned}$$

For the second factor, we argue in an analogous way, but using $\alpha_2 > 0$.

$$\begin{aligned}
R_{P,W_s}^n \left(\sum_{i=1}^n \log \frac{dR_{P,\hat{W}_{\delta,j}}}{dR_{P,W_s}}(Y_i) \geq n\beta_2 \right) &= R_{P,W_s}^n \left(\exp \left(\alpha_2 \sum_{i=1}^n \log \frac{dR_{P,\hat{W}_{\delta,j}}}{dR_{P,W_s}}(Y_i) \right) \exp(\alpha_2 n\beta_2) \right) \\
&\leq \mathbb{E}_{R_{P,W_s}^n} \left(\prod_{i=1}^n \left(\frac{dR_{P,\hat{W}_{\delta,j}}}{dR_{P,W_s}}(Y_i) \right)^{\alpha_2} \right) \exp(-\alpha_2 n\beta_2) \\
&= \exp \left((\alpha_2 - 1)n \mathbf{D}_{\alpha_2} \left(R_{P,\hat{W}_{\delta,j}} \parallel R_{P,W_s} \right) - \alpha_2 n\beta_2 \right). \tag{34}
\end{aligned}$$

Finally, for the third factor, we use $\alpha_3 < 1$.

$$\begin{aligned}
&Q_{P,W_s}^n \left(\mathbf{i}_{P,W_s}(X^n; Y^n) < n(\mathbf{I}_{P,W_s} - \varepsilon + \beta_1 + \beta_2 + \delta) \right) \\
&= Q_{P,W_s}^n \left(\exp((\alpha_3 - 1)\mathbf{i}_{P,W_s}(X^n; Y^n)) > \exp((\alpha_3 - 1)n(\mathbf{I}_{P,W_s} - \varepsilon + \beta_1 + \beta_2 + \delta)) \right) \\
&\leq \mathbb{E}_{Q_{P,W_s}^n} \left(\prod_{i=1}^n \left(\frac{dW_s(X_i, \cdot)}{dR_{P,W_s}}(Y_i) \right)^{\alpha_3 - 1} \right) \cdot \exp(-(\alpha_3 - 1)n(\mathbf{I}_{P,W_s} - \varepsilon + \beta_1 + \beta_2 + \delta)) \\
&= \exp \left(-(1 - \alpha_3)n(\mathbf{D}_{\alpha_3}(Q_{P,W_s} \parallel PR_{P,W_s}) + \varepsilon - \mathbf{I}_{P,W_s} - \beta_1 - \beta_2 - \delta) \right), \tag{35}
\end{aligned}$$

Clearly, by (32), the union bound and (15), (31) is upper bounded by the sum of (33), (34) and (35). Next, we argue that these expressions all vanish exponentially with $n \rightarrow \infty$, using the continuity of Rényi divergence in the order which is shown in [45, Theorem 7].

From (13), the theorem of monotone convergence and (12), we can conclude that

$$\lim_{\alpha_1 \searrow 1} \mathbb{E}_P \mathbf{D}_{\alpha_1} \left(W_s(X_i, \cdot) \parallel \hat{W}_{\delta,j}(X_i, \cdot) \right) = \mathbb{E}_P \mathbf{D}_1 \left(W_s(X_i, \cdot) \parallel \hat{W}_{\delta,j}(X_i, \cdot) \right) \leq \delta,$$

so, (25) allows us to fix α_1 at a value greater than 1 such that $\beta_1 - \mathbb{E}_P \mathbf{D}_{\alpha_1} \left(W_s(X_i, \cdot) \parallel \hat{W}_{\delta,j}(X_i, \cdot) \right) > 0$ and hence, (33) vanishes exponentially.

(34) is true for all $\alpha_2 < 1$. Since the inequalities are not strict, we can take the limit $\alpha_2 \nearrow 1$ and argue that the statement is also valid for $\alpha_2 = 1$.

$\mathbf{D}_{\alpha_3}(Q_{P,W_s} \parallel PR_{P,W_s})$ converges to \mathbf{I}_{P,W_s} from below for $\alpha_3 \nearrow 1$ and so (26) allows us to fix α_3 at a value less than 1 such that $\mathbf{D}_{\alpha_3}(Q_{P,W_s} \parallel PR_{P,W_s}) + \varepsilon - \mathbf{I}_{P,W_s} - \beta_1 - \beta_2 - \delta > 0$ and therefore, (35) also vanishes exponentially.

For the second summand in (30), we use the definition (29) to argue that $\mathbb{E}_C(\mathbb{P}(\mathcal{E}_2))$ is upper bounded by

$$\exp(n\mathcal{R}) \sum_{j=1}^{J(\delta)} P^n R_{P,W_s}^n \left(\mathbf{i}_{P,\hat{W}_{\delta,j}}(X^n; Y^n) \geq n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right) \right). \tag{36}$$

We define the indicator function

$$\text{ind}(x^n, y^n) := \begin{cases} 1, & \mathbf{i}_{P,\hat{W}_{\delta,j}}(x^n; y^n) \geq n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right) \\ 0, & \text{otherwise.} \end{cases}$$

Using the definition of information density for a change of measure and multiplying one, we rewrite the probability that appears in (36) as

$$\begin{aligned}
& P^n R_{P,W_s}^n \left(\mathbf{i}_{P,\hat{W}_{\delta,j}}(X^n; Y^n) \geq n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right) \right) \\
&= \int_{\mathcal{X}^n \times \mathcal{Y}^n} \text{ind}(x^n, y^n) \cdot P^n R_{P,W_s}^n(dx^n, dy^n) \\
&= \int_{\mathcal{X}^n \times \mathcal{Y}^n} \exp(-\mathbf{i}_{P,W_s}(x^n; y^n)) \text{ind}(x^n, y^n) Q_{P,W_s}^n(dx^n, dy^n) \\
&= \int_{\mathcal{X}^n \times \mathcal{Y}^n} \exp(-\mathbf{i}_{P,W_s}(x^n; y^n) + \mathbf{i}_{P,\hat{W}_{\delta,j}}(x^n; y^n) - \mathbf{i}_{P,\hat{W}_{\delta,j}}(x^n; y^n)) \cdot \text{ind}(x^n, y^n) Q_{P,W_s}^n(dx^n, dy^n)
\end{aligned}$$

Because of the presence of the indicator, we can uniformly bound

$$\mathbf{i}_{P,\hat{W}_{\delta,j}}(x^n; y^n) \geq n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right)$$

and the indicator itself can be upper bounded by 1. This yields

$$\begin{aligned}
& P^n R_{P,W_s}^n \left(\mathbf{i}_{P,\hat{W}_{\delta,j}}(X^n; Y^n) \geq n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right) \right) \\
&\leq \exp \left(-n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right) \right) \\
&\quad \int_{\mathcal{X}^n \times \mathcal{Y}^n} \exp \left(-\mathbf{i}_{P,W_s}(x^n; y^n) + \mathbf{i}_{P,\hat{W}_{\delta,j}}(x^n; y^n) \right) \\
&\quad \cdot Q_{P,W_s}^n(dx^n, dy^n).
\end{aligned}$$

We expand the definition of information density and apply Fubini's Theorem to rewrite the integral as

$$\int_{\mathcal{Y}^n} \left(\int_{\mathcal{X}^n} \frac{d\hat{W}_{\delta,j}^n(x^n, \cdot)}{dR_{P,\hat{W}_{\delta,j}}^n}(y^n) P^n(dx^n) \right) R_{P,W_s}^n(dy^n)$$

and observe that it equals 1.

Combining with (36) and applying (16), we obtain

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}(\mathbb{P}(\mathcal{E}_2)) &\leq \exp(n\mathcal{R}) \sum_{j=1}^{J(\delta)} \exp \left(-n \left(\mathbf{I}_{P,\hat{W}_{\delta,j}} - \varepsilon \right) \right) \\
&\leq \exp(n\mathcal{R}) \sum_{j=1}^{J(\delta)} \exp \left(-n \left(\inf_{s \in \mathcal{S}} \mathbf{I}_{P,W_s} - \varepsilon - \delta \right) \right) \\
&= \exp \left(-n \left(\inf_{s \in \mathcal{S}} \mathbf{I}_{P,W_s} - \varepsilon - \mathcal{R} - \delta - \frac{\log J(\delta)}{n} \right) \right). \tag{37}
\end{aligned}$$

We observe that by (24), $\inf_{s \in \mathcal{S}} \mathbf{I}_{P,W_s} - \varepsilon - \mathcal{R} - \delta > 0$.

Finally, we pick

$$\begin{aligned}
\gamma \in & \left(0, \min \left((\alpha_1 - 1) \cdot \left(\beta_1 - \mathbb{E}_P \mathbf{D}_{\alpha_1} \left(W_s(X_i, \cdot) \parallel \hat{W}_{\delta,j}(X_i, \cdot) \right) \right), \right. \right. \\
& \beta_2, \\
& \left. \left. (1 - \alpha_3) \left(\mathbf{D}_{\alpha_3}(Q_{P,W_s} \parallel PR_{P,W_s}) + \varepsilon - \mathbf{I}_{P,W_s} - \beta_1 - \beta_2 - \delta \right), \right. \right. \\
& \left. \left. \inf_{s \in \mathcal{S}} \mathbf{I}_{P,W_s} - \varepsilon - \mathcal{R} - \delta \right) \right).
\end{aligned}$$

Since the exponent in (37) is then negative for sufficiently large n , we can combine it with (33), (34) and (35) to obtain (22). \square

D. Cost Constraint

In this section, we use standard techniques to extend Theorem 3 to the case of cost-constrained code books. We define an *additive cost constraint* (c, C) for an input alphabet \mathcal{X} consisting of a function $c : \mathcal{X} \rightarrow [0, \infty)$ and a number $C \in [0, \infty)$. Given any n , we say that $x^n \in \mathcal{X}^n$ satisfies the cost constraint if $\sum_{i=1}^n c(x_i) \leq nC$.

The specialization of this definition to a usual average power constraint would be to pick the square function as c and the maximum admissible average power as C .

As long as there is at least one $x^n \in \mathcal{X}^n$ which satisfies the cost constraint (c, C) , given any codebook \mathcal{C} of block length n , we can define an associated *cost-constrained codebook* $\mathcal{C}_{c,C}$ which is generated from \mathcal{C} by replacing all code words that do not satisfy the cost constraint with x^n . Obviously, all code words in a cost-constrained codebook satisfy the cost constraint. We say that a cost constraint (c, C) is *compatible* with an input distribution P if for a random variable X distributed according to P , $c(X)$ has a finite moment generating function in an interval containing 0 in its interior and $C > \mathbb{E}_P c(X)$.

With these preliminary definitions, we can now state the compound channel coding result under an additive cost constraint.

Corollary 1. *In the setting of Theorem 3, and given an additive cost constraint (c, C) compatible with P , there are $\gamma_1, \gamma_2 > 0$ such that for sufficiently large n ,*

$$\mathbb{P}_{\mathcal{C}_{c,C}}(\epsilon \geq \exp(-n\gamma_1)) < \exp(-n\gamma_2). \quad (38)$$

The approach used in the proof of Corollary 1 is similar to the one in [48, Section 3.3], but we include the adapted derivations in full here for the sake of self-containedness. Our approach is based on the idea that in probabilistic constructions, the union bound assures that even exponentially many constraints that are satisfied with a super-exponential error bound individually are simultaneously satisfied except for an error event of super-exponentially small probability. Such ideas have already been used in earlier works of information theory, such as [49] and, in the context of Gaussian channels, [17]. We begin with a series of preliminary lemmas and conclude the section with the proof of Corollary 1.

Lemma 6. *Let $(U_k)_{k \geq 1}$ be a sequence of independent and identically distributed random variables such that the moment generating function $\varphi(\lambda) := \mathbb{E} \exp(\lambda U_1)$ exists on an interval containing 0 in its interior. Let $C > \mathbb{E} U_1$. Then there exists $\gamma > 0$ such that*

$$\mathbb{P} \left(\sum_{k=1}^n U_k > nC \right) \leq \exp(-n\gamma).$$

Proof. We can without loss of generality assume that $C = 0$ and $\mathbb{E}(U_1) < 0$, because otherwise we could consider the random variables $(U_k - C)_{k \geq 1}$ instead.

Clearly, $\varphi(0) = 1$ and $\varphi'(0) = \mathbb{E}(U_1) < 0$, so we can find some $\lambda > 0$ sufficiently small such that $\varphi(\lambda) < 1$. With this choice of λ , we can apply Markov's inequality and get

$$\begin{aligned} \mathbb{P} \left(\sum_{k=1}^n U_k > 0 \right) &= \mathbb{P} \left(\exp \left(\lambda \sum_{k=1}^n U_k \right) > 1 \right) \\ &\leq \mathbb{E} \left(\exp \left(\lambda \sum_{k=1}^n U_k \right) \right) \\ &= \varphi(\lambda)^n \end{aligned}$$

so the lemma follows by choosing $\gamma := -\log \varphi(\lambda)$. \square

Lemma 7. *Let \mathfrak{N} be a Bernoulli random variable with $\exp(n\mathcal{R})$ trials and success probability $p \leq \exp(-n\beta_1)$ where $\beta_1 < \mathcal{R}/2$. Then there are $\gamma_1, \gamma_2 > 0$ such that for sufficiently large n ,*

$$\mathbb{P}(\mathfrak{N} > \exp(n(\mathcal{R} - \gamma_1))) \leq \exp(-\exp(n\gamma_2)). \quad (39)$$

Proof. We choose γ_1, γ_2 and β_2 such that $0 < \gamma_1 < \beta_1 < \beta_2 < \mathcal{R}/2$ and $\gamma_2 < \mathcal{R} - 2\beta_2$. Then

$$\begin{aligned} \mathbb{P}(\mathfrak{N} > \exp(n(\mathcal{R} - \gamma_1))) &= \mathbb{P}(\mathfrak{N} > p \exp(n\mathcal{R}) + (\exp(-n\gamma_1) - p) \exp(n\mathcal{R})) \\ &\leq \mathbb{P}(\mathfrak{N} > \mathbb{E}\mathfrak{N} + (\exp(-n\gamma_1) - \exp(-n\beta_1)) \exp(n\mathcal{R})) \\ &\leq \mathbb{P}(\mathfrak{N} > \mathbb{E}\mathfrak{N} + \exp(-n\beta_2) \exp(n\mathcal{R})) \end{aligned} \quad (40)$$

$$\leq \exp \left(-2 \frac{(\exp(-n\beta_2))^2 (\exp(n\mathcal{R}))^2}{\exp(n\mathcal{R})} \right) \quad (41)$$

$$= \exp(-2 \exp(n(\mathcal{R} - 2\beta_2))) \quad (42)$$

$$\leq \exp(-\exp(n\gamma_2)), \quad (43)$$

where (41) follows by the Chernoff-Hoeffding bound as stated for instance in [50, Theorem 1.1, eq. (1.6)]. \square

Lemma 8. Let P be a probability distribution on \mathcal{X} . Assume moreover that $c(X)$ has a moment generating function defined on an interval containing 0 in its interior and that $C > \mathbb{E}_P c(X)$. Denote the number of bad code words in \mathcal{C} with

$$\mathfrak{N} := \sum_{m=1}^{\exp(n\mathcal{R})} \mathbf{1}_{\sum_{i=1}^n c(\mathcal{C}(m)(i)) > nC}.$$

Then there are $\gamma_1, \gamma_2 > 0$ such that

$$\mathbb{P}_{\mathcal{C}}(\mathfrak{N} > \exp(n(\mathcal{R} - \gamma_1))) \leq \exp(-\exp(n\gamma_2)). \quad (44)$$

Proof. Since the code word components are independently and identically distributed, we can apply Lemma 6 and obtain an arbitrarily small $\beta_1 > 0$ such that for all m ,

$$p := \mathbb{P}_{\mathcal{C}}\left(\sum_{i=1}^n c(\mathcal{C}(m)(i)) > nC\right) \leq \exp(-n\beta_1).$$

So since the code words are independent, \mathfrak{N} is a Bernoulli variable with $\exp(n\mathcal{R})$ trials and success probability p , and an application of Lemma 7 proves the conclusion. \square

Proof of Corollary 1. Assume throughout the proof that n is sufficiently large. By Lemma 8, we have $\hat{\gamma}_1, \hat{\gamma}_2 \in (0, \infty)$ with

$$\mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}}) \leq \exp(-\exp(n\hat{\gamma}_2)), \quad (45)$$

where

$$\hat{\mathcal{E}} := \{\mathbb{P}_{\mathcal{M}}(\mathcal{C}(\mathcal{M}) \neq \mathcal{C}_{c,C}(\mathcal{M})) > \exp(-n\hat{\gamma}_1)\}.$$

We denote the error of \mathcal{C} with $\epsilon_{\mathcal{C}}$ and the error of $\mathcal{C}_{c,C}$ with $\epsilon_{\mathcal{C}_{c,C}}$. By Theorem 3 and Markov's inequality, we have, for some $\hat{\gamma} \in (0, \infty)$ given by the theorem and with choices $\tilde{\gamma}_1 \in (0, \min(\hat{\gamma}, \hat{\gamma}_1))$, $\tilde{\gamma}_2 \in (0, \hat{\gamma} - \tilde{\gamma}_1)$,

$$\begin{aligned} \mathbb{P}_{\mathcal{C}}(\epsilon_{\mathcal{C}} \geq \exp(-n\tilde{\gamma}_1)) &\leq \mathbb{E}_{\mathcal{C}} \epsilon_{\mathcal{C}} \exp(n\tilde{\gamma}_1) \\ &\leq \exp(-n(\hat{\gamma} - \tilde{\gamma}_1)) \\ &\leq \exp(-n\tilde{\gamma}_2). \end{aligned} \quad (46)$$

Conditioned on the complement of $\hat{\mathcal{E}}$, we have

$$\begin{aligned} \epsilon_{\mathcal{C}_{c,C}} &\stackrel{(a)}{=} \sup_{s \in \mathcal{S}} \mathbb{E}_{\mathcal{M}} \left(\mathbb{P}_s(\mathcal{M} \neq d(Y^n) | X^n = \mathcal{C}_{c,C}(\mathcal{M})) \right) \\ &= \sup_{s \in \mathcal{S}} \sum_{m=1}^{\exp(n\mathcal{R})} \exp(-n\mathcal{R}) \mathbb{P}_s(m \neq d(Y^n) | X^n = \mathcal{C}_{c,C}(m)) \\ &\stackrel{(b)}{\leq} \sup_{s \in \mathcal{S}} \sum_{\substack{m=1 \\ \mathcal{C}_{c,C}(m) = \mathcal{C}(m)}}^{\exp(n\mathcal{R})} \exp(-n\mathcal{R}) \mathbb{P}_s(m \neq d(Y^n) | X^n = \mathcal{C}_{c,C}(m)) + \sum_{\substack{m=1 \\ \mathcal{C}_{c,C}(m) \neq \mathcal{C}(m)}}^{\exp(n\mathcal{R})} \exp(-n\mathcal{R}) \\ &\stackrel{(a)}{\leq} \epsilon_{\mathcal{C}} + \exp(-n\hat{\gamma}_1), \end{aligned} \quad (47)$$

where the steps marked with (a) are by the definition of compound coding error, and (b) is by upper bounding some of the probabilities in the sum with 1. We can now choose $\gamma_1 \in (0, \tilde{\gamma}_1)$ and obtain

$$\begin{aligned} \mathbb{P}_{\mathcal{C}}(\epsilon_{\mathcal{C}_{c,C}} \geq \exp(-n\gamma_1)) &\stackrel{(a)}{\leq} \mathbb{P}_{\mathcal{C}}(\epsilon_{\mathcal{C}_{c,C}} \geq \exp(-n\gamma_1) | \neg \hat{\mathcal{E}}) + \mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}}) \\ &\stackrel{(47)}{\leq} \mathbb{P}_{\mathcal{C}}(\epsilon_{\mathcal{C}} + \exp(-n\hat{\gamma}_1) \geq \exp(-n\gamma_1) | \neg \hat{\mathcal{E}}) + \mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}}) \\ &\stackrel{(a)}{\leq} \frac{\mathbb{P}_{\mathcal{C}}(\epsilon_{\mathcal{C}} \geq \exp(-n\gamma_1) - \exp(-n\hat{\gamma}_1))}{1 - \mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}})} + \mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}}) \\ &\stackrel{(b)}{\leq} \frac{\mathbb{P}_{\mathcal{C}}(\epsilon_{\mathcal{C}} \geq \exp(-n\tilde{\gamma}_1))}{1 - \mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}})} + \mathbb{P}_{\mathcal{C}}(\hat{\mathcal{E}}) \\ &\stackrel{(45), (46)}{\leq} \frac{\exp(-n\tilde{\gamma}_2)}{1 - \exp(-\exp(n\hat{\gamma}_2))} + \exp(-\exp(n\hat{\gamma}_2)) \\ &\stackrel{(c)}{\leq} \exp(-n\gamma_2), \end{aligned}$$

where the steps marked with (a) are by the law of total probability, step (b) is by the choices of $\gamma_1, \tilde{\gamma}_1$, and step (c) is valid for any choice of $\gamma_2 \in (0, \tilde{\gamma}_2)$. \square

IV. JAMMING STRATEGIES INDUCED BY RANDOM CODE BOOKS

In this section, we leverage the results of Section III in conjunction with a known channel resolvability result to establish the main technical contribution that goes into the proof of Theorem 1. The results and arguments in this section (except for the proof of Theorem 1) are not specific to AWGN channels. In this section, we therefore use the system model described in Section II-A without the specializations from Section II-B. We fix an arbitrary admissible DFA scheme as defined in Section II-A. Such a scheme will induce effective channels for \mathfrak{B} and \mathfrak{E} as outlined in Fig. 2. We denote the legitimate user's effective channel, which is a stochastic kernel mapping from $\mathcal{S}_1 \times \dots \times \mathcal{S}_K \times \mathcal{X}$ to \mathcal{Y} , by $W_{\mathfrak{B}}$ and the eavesdropper's effective channel, which is a stochastic kernel mapping from $\mathcal{S}_1 \times \dots \times \mathcal{S}_K \times \mathcal{X}$ to \mathcal{Z} , by $W_{\mathfrak{E}}$.

In this section, we analyze jamming strategies that are induced by a codebook in the following sense: The jammer draws a code word index \mathcal{M} uniformly at random and transmits $\mathcal{C}(\mathcal{M})$, the code word in \mathcal{C} indexed by \mathcal{M} . Therefore, the number of code words in the codebook controls the amount of randomness contained in the jamming signal. We use the same random ensemble of code books that is defined at the beginning of Section III-C.

With these concepts and notations defined, we are ready to state the main result of this section, which gives sufficient conditions for the existence of a jamming scheme that can simultaneously ensure that the legitimate receiver is able to reconstruct the full jamming signal and limit the usefulness of the eavesdropper's received signal.

Theorem 4. *Let P be a jammer input distribution. Suppose that for every $\delta > 0$, there is some $J(\delta)$ such that the compound channel $(W_s)_{s \in \mathcal{S}}$ defined by $\mathcal{S} := \mathcal{S}_1 \times \dots \times \mathcal{S}_K$ and $W_{(s_1, \dots, s_K)} := W_{\mathfrak{B}}(s_1, \dots, s_K, \cdot, \cdot)$ can be $(\delta, J(\delta))$ -approximated under P . Suppose further that for all $s_1 \in \mathcal{S}_1, \dots, s_K \in \mathcal{S}_K$, the moment-generating function*

$$\mathbb{E} \exp(t \cdot \mathbf{i}_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot, \cdot)}(X; Z))$$

of the information density exists and is finite at some point $t > 0$. Let (c, C) be an additive cost constraint compatible with P , and let \mathcal{C} be a random codebook from the (P, n, \mathcal{R}) -ensemble. Let $\mathcal{R} \in (0, \infty)$ such that

$$\sup_{s_1 \in \mathcal{S}_1, \dots, s_K \in \mathcal{S}_K} \mathbf{I}_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot, \cdot)} < \mathcal{R} < \inf_{s_1 \in \mathcal{S}_1, \dots, s_K \in \mathcal{S}_K} \mathbf{I}_{P, W_{\mathfrak{B}}(s_1, \dots, s_K, \cdot, \cdot)}. \quad (48)$$

Then there are numbers $\gamma_1, \gamma_2, \gamma_3, \gamma_4 > 0$ such that for sufficiently large n ,

$$\mathbb{P}_{\mathcal{C}} \left(\left\| \hat{R}_{W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot, \cdot), \mathcal{C}_{c, C}}^n - R_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot, \cdot)}^n \right\|_{\text{TV}} \geq \exp(-n\gamma_1) \right) < \exp(-\exp(n\gamma_2)), \quad (49)$$

where $\hat{R}_{W^n, \mathcal{C}}$ denotes output of a channel W^n given that a uniformly random code word from the codebook \mathcal{C} is transmitted, and

$$\mathbb{P}_{\mathcal{C}}(\mathcal{E}) < \exp(-n\gamma_4), \quad (50)$$

where \mathcal{E} is the event that the jamming strategy induced by $\mathcal{C}_{c, C}$ does not allow reconstruction of the jamming signal with error at most $\exp(-n\gamma_3)$.

Remark 2. *The bound (49) compares two probability distributions. The first one, $\hat{R}_{W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot, \cdot), \mathcal{C}_{c, C}}^n$, is the distribution the eavesdropper observes if the jamming strategy follows the approach we propose in this paper. The second one, $R_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot, \cdot)}^n$, is the distribution the eavesdropper observes if the jammer transmits white noise. In the sense made explicit in Theorem 4, these two cases are “almost” the same. This comparison is exploited in the proof of Theorem 1 to prove the MSE-security guarantee.*

In order to prove this theorem, we decompose the system depicted in Fig. 2 into smaller (and more easily analyzed) subsystems by considering only a subset of the depicted terminals at a time.

a) *Considering the terminals $\mathfrak{A}_1, \dots, \mathfrak{A}_K, \mathfrak{B}$:* This is the DFA system model. This part of the system consists of transmitters $(\mathfrak{A}_k)_{k=1}^K$ each of which holds a value $s_k \in \mathcal{S}_k$ and a receiver \mathfrak{B} which has the objective of estimating $f(s_1, \dots, s_K)$. To this end, each transmitter \mathfrak{A}_k passes s_k through a pre-processor F_k independently n times yielding a sequence T_k^n of channel inputs. These are transmitted through n independent uses of the channel, generating a sequence Y^n of channel outputs. The receiver passes this sequence through a post-processor D^n which generates an approximation \tilde{f} of $f(s_1, \dots, s_K)$. As mentioned, the design of the pre- and post-processors depends heavily on the channel model and a particular class of functions f . The idea is that the pre-processors, the channel and the post-processor work together to mimic the function f , and any approach following this idea will be highly dependent on the particular structure of the channel and f . In Theorem 4, it is assumed that such a system is already in place and an augmentation is proposed which makes it more secure. A property of the system described in Section II-A necessary for our purposes and heavily exploited in this work is that the pre-processing is i.i.d., i.e., each pre-processor F_k is a stochastic kernel mapping from \mathcal{S}_k to \mathcal{X}_k and an n -fold product F_k^n of it is used to generate the channel input sequence.

b) *Considering the terminals $\mathfrak{A}_1, \dots, \mathfrak{A}_K, \mathfrak{J}, \mathfrak{E}$:* In this setting, we assume that the transmitters $\mathfrak{A}_1, \dots, \mathfrak{A}_K$ run a scheme of the kind described under a). Instead of the legitimate receiver, there is now an eavesdropper \mathfrak{E} . The objective is then to limit the usefulness of the eavesdropper's received signal Z^n . To this end, we add a friendly jammer \mathfrak{J} to the system which transmits, according to a certain strategy, a word X^n . In this work, any jamming strategy we consider is induced by a codebook \mathcal{C} of words of length n through the rule that the jammer chooses an element of the codebook uniformly at random and transmits it. We use existing results on *channel resolvability* to derive a bound on the usefulness of the signal Z^n received at \mathfrak{E} .

c) *Considering the terminals $\mathfrak{A}_1, \dots, \mathfrak{A}_K, \mathfrak{J}, \mathfrak{B}$:* This is the setting from a) with an additional transmitter \mathfrak{J} . Here we assume that \mathfrak{J} uses a jamming strategy induced by a codebook \mathcal{C} as described under b) and use Theorem 3 on compound channel coding to argue that for suitable choices of \mathcal{C} , \mathfrak{B} is able to fully reconstruct the jamming signal X^n . This enables \mathfrak{B} to perform a cancellation of the jamming signal before it applies the post-processor D^n it would use in setting a). How this cancellation works depends on the particularities of the channel considered, but if, e.g., the jamming signal is simply added to the channel output as in the AWGN example in Section II-B, it is possible to cancel it entirely by subtracting it from the received signal. So in this case, the post-processor would consist of a reconstruction of the jamming signal, the subtraction of this signal from the received one and a post-processing step identical to that from a).

d) *Combining settings b) and c):* The goal here is to argue the existence of a codebook \mathcal{C} which achieves both of the objectives described under b) and c). It will turn out that this can be achieved by a standard random codebook construction.

Theorem 4 formulates conditions under which there are code books in the (P, n, \mathcal{R}) -ensemble of which the (c, C) -cost constrained versions simultaneously achieve the goals set forth under b) and c).

As a technical ingredient for our proof, we recall a result on channel resolvability from [24] that will be applied in order to guarantee the virtual indistinguishability of the jamming signal from white noise for the eavesdropper.

Theorem 5. [24] *Given a channel W from \mathcal{X} to \mathcal{Y} , an input distribution P such that the moment-generating function $\mathbb{E}_{Q_{P,W}} \exp(t \cdot \mathbf{i}_{P,W}(X; Y))$ of the information density exists and is finite for some $t > 0$, and $\mathcal{R} > \mathbf{I}_{P,W}$, there exist $\gamma_1 > 0$ and $\gamma_2 > 0$ such that for large enough block lengths n , the (P, n, \mathcal{R}) -ensemble satisfies*

$$\mathbb{P}_{\mathcal{C}} \left(\|\hat{R}_{W^n, \mathcal{C}} - Q_{P,W}^n\|_{\text{TV}} > \exp(-\gamma_1 n) \right) \leq \exp(-\exp(\gamma_2 n)), \quad (51)$$

where $\hat{R}_{W^n, \mathcal{C}}$ is the output distribution of channel W given that a uniformly random code word from \mathcal{C} is transmitted.

Similarly as with the compound channel coding theorem, we can use known methods to incorporate an additive cost constraint and argue the following corollary.

Corollary 2. *Let P be an input distribution on \mathcal{X} and (c, C) an additive cost constraint compatible with P . Then the statement of Theorem 5 is valid even if the codebook \mathcal{C} is replaced with its associated cost-constrained version $\mathcal{C}_{c,C}$.*

Proof. By Lemma 8, we pick $\hat{\gamma}_1, \hat{\gamma}_2$ satisfying (44) and by Theorem 5, we pick $\tilde{\gamma}_1, \tilde{\gamma}_2$ satisfying (51).

We use the observation that $\mathfrak{N} \leq \exp((\mathcal{R} - \hat{\gamma}_1)n)$ implies

$$\|\hat{R}_{W^n, \mathcal{C}_{c,C}} - \hat{R}_{W^n, \mathcal{C}}\|_{\text{TV}} \leq \frac{\mathfrak{N}}{\exp(n\mathcal{R})} \leq \exp(-\hat{\gamma}_1 n) \quad (52)$$

and observe that, as long as $\gamma_1 < \hat{\gamma}_1, \tilde{\gamma}_1$ and $\gamma_2 < \hat{\gamma}_2, \tilde{\gamma}_2$ and n is sufficiently large,

$$\begin{aligned} & \mathbb{P}_{\mathcal{C}_{c,C}} \left(\|\hat{R}_{W^n, \mathcal{C}_{c,C}} - Q_{P,W}^n\|_{\text{TV}} > \exp(-\gamma_1 n) \right) \\ & \stackrel{(a)}{\leq} \mathbb{P}_{\mathcal{C}} \left(\|\hat{R}_{W^n, \mathcal{C}_{c,C}} - \hat{R}_{W^n, \mathcal{C}}\|_{\text{TV}} + \|\hat{R}_{W^n, \mathcal{C}} - Q_{P,W}^n\|_{\text{TV}} > \exp(-\gamma_1 n) \right) \\ & \stackrel{(b)}{\leq} \mathbb{P}_{\mathcal{C}} \left(\|\hat{R}_{W^n, \mathcal{C}_{c,C}} - \hat{R}_{W^n, \mathcal{C}}\|_{\text{TV}} > \exp(-\hat{\gamma}_1 n) \right) + \mathbb{P}_{\mathcal{C}} \left(\|\hat{R}_{W^n, \mathcal{C}} - Q_{P,W}^n\|_{\text{TV}} > \exp(-\tilde{\gamma}_1 n) \right) \\ & \stackrel{(c)}{<} \exp(-\exp(\hat{\gamma}_2 n)) + \exp(-\exp(\tilde{\gamma}_2 n)) \\ & \stackrel{(d)}{\leq} \exp(-\exp(\gamma_2 n)), \end{aligned}$$

where (a) is by the triangle inequality, (b) is by the union bound and the choice of γ_1 , (c) is due to (44), (52) and (51), and (d) is by the choice of γ_2 . \square

Given the previous observations, the proof of the main result of this section is now straightforward.

Proof of Theorem 4. An application of Corollary 1 yields (50), and (49) follows from Corollary 2. \square

We can now put everything together and prove the main theorem of this paper.

Proof of Theorem 1. For the pre-processing at the transmitters, we use the same scheme as in the proof of Lemma 2 and begin by verifying that the resulting effective channels $W_{\mathfrak{B}}$ and $W_{\mathfrak{E}}$ with the input distribution P chosen to be Gaussian with mean

0 and variance \mathfrak{P}_3 satisfy the assumptions of Theorem 4. Since the defined compound channel is a class of Gaussian channels with different means taking values in the compact set $[-1, 1]$, the approximability of the channel is an immediate consequence of Theorem 2. The finiteness of the moment-generating function of the information density can be seen by straightforward applications of the definitions of information density and Rényi divergence:

$$\begin{aligned}\mathbb{E} \exp(t \cdot \mathbf{i}_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot)}(X; Z)) &= \mathbb{E} \left(\left(\frac{dW_{\mathfrak{E}}(s_1, \dots, s_K, X, \cdot)}{dR_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot)}}(Z) \right)^t \right) \\ &= \exp \left(t \cdot \frac{1}{t} \log \mathbb{E} \left(\left(\frac{dW_{\mathfrak{E}}(s_1, \dots, s_K, X, \cdot)}{dR_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot)}}(Z) \right)^t \right) \right) \\ &= \exp \left(t \mathbf{D}_{t+1} (Q_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot)} || P R_{P, W_{\mathfrak{E}}(s_1, \dots, s_K, \cdot)}) \right)\end{aligned}$$

The Rényi divergence appearing at the end is between two multivariate Gaussian distributions and can be seen to be finite from the expressions given in [46]. In order to verify (48), we first note that the information expressions appearing are the capacities of the effective channels $W_{\mathfrak{B}}$ and $W_{\mathfrak{E}}$. Since s_1, \dots, s_K change the mean of the channel only, they do not influence the capacity. Therefore, the infimum and supremum are over singleton sets. Consequently, the condition $h_{\mathfrak{B}}/\sigma_{\mathfrak{B}} > h_{\mathfrak{E}}/\sigma_{\mathfrak{E}}$ ensures that there is some \mathcal{R} satisfying (48).

Fix γ_1', γ_3' as claimed to exist in Theorem 4, and also fix γ_1, γ_2 with $0 < \gamma_2 < \gamma_1'$ and $0 < \gamma_1 < \gamma_3'$.

Note that in the AWGN channel, s_1, \dots, s_K correspond to a shift of the output distribution of the channel, and therefore, the variational distance that appears in (49) is independent of s_1, \dots, s_K . For sufficiently large n , we can therefore fix a codebook \mathcal{C} from the (P, n, \mathcal{R}) -ensemble such that for all s_1, \dots, s_K , neither one of the error events described in (49) and (50) occurs.

Let the jamming strategy be induced by $\mathcal{C}_{C,c}$ and let $d: \mathcal{Z}^n \rightarrow [-1, 1]$ be an estimator for \mathfrak{E} . We bound the MSE of d as

$$\begin{aligned}\mathbb{E}_{\hat{R}_{W_{\mathfrak{E}}^n(s_1, \dots, s_K, \cdot)}, \mathcal{C}_{C,c}} \left((d(Z^n) - f(s_1, \dots, s_K))^2 \right) &= \int_0^\infty \hat{R}_{W_{\mathfrak{E}}^n(s_1, \dots, s_K, \cdot), \mathcal{C}_{C,c}} \left((d(Z^n) - f(s_1, \dots, s_K))^2 > t \right) dt \\ &\stackrel{(a)}{=} \int_0^4 \hat{R}_{W_{\mathfrak{E}}^n(s_1, \dots, s_K, \cdot), \mathcal{C}_{C,c}} \left((d(Z^n) - f(s_1, \dots, s_K))^2 > t \right) dt \\ &\stackrel{(49)}{\geq} \int_0^4 \left(R_{P, W_{\mathfrak{E}}^n(s_1, \dots, s_K, \cdot)} \left((d(Z^n) - f(s_1, \dots, s_K))^2 > t \right) - \exp(-n\gamma_1') \right) dt \\ &= \mathbb{E}_{R_{P, W_{\mathfrak{E}}^n(s_1, \dots, s_K, \cdot)}} \left((d(Z^n) - f(s_1, \dots, s_K))^2 \right) - 4 \exp(-n\gamma_1').\end{aligned}$$

where step (a) is due to the fact that both $d(Z^n)$ and $f(s_1, \dots, s_K)$ are restricted to the interval $[-1, 1]$. Taking the lower bound for the MSE under $R_{P, W_{\mathfrak{E}}^n(s_1, \dots, s_K, \cdot)}$ from Lemma 2 and noting $\gamma_2 < \gamma_1'$, we arrive at the expression in (8) for sufficiently large n .

For the reconstruction strategy at \mathfrak{B} , we first let \mathfrak{B} reconstruct the jamming signal as is possible by Theorem 4 and then post-process the received signal as is possible with knowledge of the jamming signal by Lemma 2. Using the error bound in Lemma 2 and observing that the maximum instantaneous square error is 4 since we are constrained to an interval of length 2 and that $\gamma_1 < \gamma_3'$, for sufficiently large n we arrive at (7). \square

V. CONCLUSION

In this work, we have introduced a framework for distributed function approximation with jamming (DFA-J). We have shown how well-known information theoretic tools can be used to improve security by means of a jammer whose signal is stronger at the legitimate receiver than it is at the eavesdropper. In the process, we have proved a compound channel coding result which is a generalization of similar results from the literature.

This work is intended as an initial step towards providing security against eavesdropping for OTA computation schemes. Our theoretical analysis derives MSE guarantees both for the eavesdropper's and the legitimate receiver's reconstruction of the objective function for the case in which an arithmetic average is computed over an AWGN channel. However, a gap between this theoretical work and its implementation for the envisioned practical applications remains. In particular, we are interested in the following questions for future research:

- Can the secrecy guarantees in this work be achieved with structured codes which allow for practically feasible encoding and decoding?
- Can the secrecy guarantees be strengthened to full semantic security?
- Can the approach be generalized to a larger class of channels?

APPENDIX

In this appendix, we prove the two lemmas used for the proof of Lemma 2.

Proof of Lemma 3. It is known [51, eq. (6.92)] that the MSE is minimized by the mean of the posterior probability distribution. We can therefore calculate the minimum MSE estimator given the observations v_1, \dots, v_n as follows, where we use p with random variables in the index to denote (conditional) densities.

$$\begin{aligned}
\hat{U} &= \int_a^b u p_{U|V_1, \dots, V_n}(u|v_1, \dots, v_n) du \\
&\stackrel{(a)}{=} \int_a^b u \frac{p_{V_1, \dots, V_n|U}(v_1, \dots, v_n|u) p_U(u)}{p_{V_1, \dots, V_n}(v_1, \dots, v_n)} du \\
&= \frac{\int_a^b u p_{V_1, \dots, V_n|U}(v_1, \dots, v_n|u) p_U(u) du}{\int_a^b p_{V_1, \dots, V_n|U}(v_1, \dots, v_n|u) p_U(u) du} \\
&\stackrel{(b)}{=} \frac{\int_a^b u \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (v_i - u)^2\right) du}{\int_a^b \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (v_i - u)^2\right) du} \\
&= \frac{\int_a^b u \exp\left(-\frac{1}{2\sigma^2/n} \left(\frac{1}{n} \sum_{i=1}^n v_i^2 - 2u\bar{v} + u^2\right)\right) du}{\int_a^b \exp\left(-\frac{1}{2\sigma^2/n} \left(\frac{1}{n} \sum_{i=1}^n v_i^2 - 2u\bar{v} + u^2\right)\right) du} \\
&\stackrel{(c)}{=} \frac{\int_a^b u \exp\left(-\frac{1}{2\sigma^2/n} (\bar{v} - u)^2\right) du}{\int_a^b \exp\left(-\frac{1}{2\sigma^2/n} (\bar{v} - u)^2\right) du}
\end{aligned}$$

For (a), we have applied Bayes' rule. (b) is by observing that $p_U(u) = 1/(b-a)$ is independent of u in $[a, b]$ and $p_{V_1, \dots, V_n|U}$ is the normal density. (c) is by multiplying

$$\exp\left(-\frac{1}{2\sigma^2/n} \left(\bar{v}^2 - \frac{1}{n} \sum_{i=1}^n v_i^2\right)\right)$$

on both sides of the fraction to complete the binomials.

The term we have calculated for \hat{U} is the mean of a normal distribution centered at \bar{v} with variance σ^2/n truncated in $[a, b]$. This is a distribution with a known mean [52, eq. 13.134], and hence we arrive at (9). \square

Proof of Lemma 4. Based on the representation (9), we calculate the MSE as follows. We use the substitution rule, substituting $v' := \frac{\bar{v}-a}{\sigma/\sqrt{n}}$ in (a) and $u' := \frac{u-a}{\sigma/\sqrt{n}}$ in (b).

$$\begin{aligned}
\mathbb{E}\left(\left(U - \hat{U}\right)^2\right) &= \int_a^b \int_{-\infty}^{\infty} \left(\bar{v} + \frac{\sigma}{\sqrt{n}} \cdot \frac{\varphi_N\left(\frac{a-\bar{v}}{\sigma/\sqrt{n}}\right) - \varphi_N\left(\frac{b-\bar{v}}{\sigma/\sqrt{n}}\right)}{\Phi_N\left(\frac{b-\bar{v}}{\sigma/\sqrt{n}}\right) - \Phi_N\left(\frac{a-\bar{v}}{\sigma/\sqrt{n}}\right)} - u\right)^2 \cdot \frac{1}{b-a} \cdot \frac{1}{\sigma/\sqrt{n}} \varphi_N\left(\frac{u-\bar{v}}{\sigma/\sqrt{n}}\right) d\bar{v} du \\
&\stackrel{(a)}{=} \int_a^b \int_{-\infty}^{\infty} \left(\frac{\sigma}{\sqrt{n}} \left(v' + \frac{\varphi_N(-v') - \varphi_N\left(\frac{b-a}{\sigma/\sqrt{n}} - v'\right)}{\Phi_N\left(\frac{b-a}{\sigma/\sqrt{n}} - v'\right) - \Phi_N(-v')}\right) + a - u\right)^2 \cdot \frac{1}{b-a} \cdot \varphi_N\left(\frac{u-a}{\sigma/\sqrt{n}} - v'\right) dv' du \\
&\stackrel{(b)}{=} \int_0^{\frac{b-a}{\sigma/\sqrt{n}}} \int_{-\infty}^{\infty} \left(v' + \frac{\varphi_N(-v') - \varphi_N\left(\frac{b-a}{\sigma/\sqrt{n}} - v'\right)}{\Phi_N\left(\frac{b-a}{\sigma/\sqrt{n}} - v'\right) - \Phi_N(-v')}\right)^2 \cdot \left(\frac{\sigma}{\sqrt{n}}\right)^3 \cdot \frac{1}{b-a} \cdot \varphi_N(u' - v') dv' du' \\
&= \frac{\sigma^2}{n} \Psi\left(\frac{b-a}{\sigma/\sqrt{n}}\right),
\end{aligned}$$

concluding the proof of the lemma. \square

REFERENCES

- [1] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2155–2169, 2020.
- [2] K. Ralinovski, M. Goldenbaum, and S. Stańczak, "Energy-efficient classification for anomaly detection: The wireless channel as a helper," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.
- [3] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Information Processing in Sensor Networks*, F. Zhao and L. Guibas, Eds. Berlin and Heidelberg, Germany: Springer, 2003, pp. 162–177.
- [4] M. Goldenbaum and S. Stańczak, "Robust analog function computation via wireless multiple-access channels," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3863–3877, 2013.
- [5] I. Bjelaković, M. Frey, and S. Stańczak, "Distributed approximation of functions over fast fading channels with applications to distributed learning and the max-consensus problem," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2019, pp. 1146–1153.

- [6] W. Liu, X. Zang, Y. Li, and B. Vucetic, "Over-the-air computation systems: Optimization, analysis and scaling laws," *IEEE Transactions on Wireless Communications*, vol. 19, pp. 5488–5502, 2020.
- [7] M. Frey, I. Bjelaković, and S. Stańczak, "Over-the-air computation in correlated channels," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5739–5755, 2021.
- [8] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, 2007.
- [9] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radio Engineering and Electronics*, vol. 4, no. 12, pp. 1–8, 1959.
- [10] D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, pp. 1229–1241, 1959.
- [11] J. Wolfowitz, "Simultaneous channels," *Archive for Rational Mechanics and Analysis*, vol. 4, pp. 371–386, 1959.
- [12] H. Kesten, "Some remarks on the capacity of compound channels in the semicontinuous case," *Information and Control*, vol. 4, no. 2-3, pp. 169–184, 1961.
- [13] K. Yoshihara, "Coding theorems for the compound semi-continuous memoryless channels," in *Kodai Mathematical Seminar Reports*, vol. 17, no. 1. Department of Mathematics, Tokyo Institute of Technology, 1965, pp. 30–43.
- [14] R. Ahlswede, "Certain results in coding theory for compound channels," in *Proceedings of the Colloquium on Information Theory Debrecen (Hungary)*, 1967, pp. 35–60.
- [15] W. Root and P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM Journal on Applied Mathematics*, vol. 16, no. 6, pp. 1350–1393, 1968.
- [16] I. Bjelaković, H. Boche, and J. Sommerfeld, "Secrecy results for compound wiretap channels," *Problems of Information Transmission*, vol. 49, no. 1, pp. 73–98, 2013.
- [17] X. He and A. Yener, "MIMO wiretap channels with unknown and varying eavesdropper channel states," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6844–6869, 2014.
- [18] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.
- [19] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.
- [20] I. Csiszár, "Almost independence and secrecy capacity," *Problems of Information Transmission*, vol. 32, no. 1, pp. 40–47, 1996.
- [21] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 44–55, 2005.
- [22] M. Hayashi and R. Matsumoto, "Secure multiplex coding with dependent and non-uniform multiple messages," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2355–2409, 2016.
- [23] P. Cuff, "Soft covering with high probability," in *2016 IEEE International Symposium on Information Theory*. IEEE, 2016, pp. 2963–2967.
- [24] M. Frey, I. Bjelakovic, and S. Stanczak, "Resolvability on continuous alphabets," in *2018 IEEE International Symposium on Information Theory*. IEEE, 2018, pp. 2037–2041.
- [25] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1562–1575, 2006.
- [26] I. Csiszar and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge: Cambridge University Press, 2011.
- [27] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Advances in Cryptology—CRYPTO 2012*. Springer, 2012, pp. 294–311.
- [28] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8077–8098, 2013.
- [29] R. Negi and S. Goel, "Secret communication using artificial noise," in *IEEE Vehicular Technology Conference*, vol. 62, no. 3. IEEE, 2005, p. 1906.
- [30] J. P. Vilela, M. Bloch, J. Barros, and S. W. McLaughlin, "Friendly jamming for wireless secrecy," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–6.
- [31] —, "Wireless secrecy regions with friendly jamming," *IEEE Transactions on Information Forensics & Security*, vol. 6, no. 2, pp. 256–266, 2011.
- [32] I. Stanojev and A. Yener, "Improving secrecy rate via spectrum leasing for friendly jamming," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 134–145, 2012.
- [33] E. Tekin and A. Yener, "The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2735–2751, 2008.
- [34] A. J. Pierrot and M. R. Bloch, "Strongly secure communications over the two-way wiretap channel," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 595–605, 2011.
- [35] X. He and A. Yener, "The role of feedback in two-way secure communications," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8115–8130, 2013.
- [36] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.
- [37] A. D. Wyner, "The wire-tap channel," *Bell Labs Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [38] U. M. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography: Two Sides of One Tapestry*, R. E. Blahut, D. J. Costello, U. Maurer, and T. Mittelholzer, Eds. Boston: Springer, 1994, pp. 271–285.
- [39] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 601–605.
- [40] R. G. D'Oliveira, S. El Rouayheb, and M. Médard, "The computational wiretap channel," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2018, pp. 1136–1140.
- [41] G. Bassi and M. Skoglund, "On the mutual information of two boolean functions, with application to privacy," in *2019 IEEE International Symposium on Information Theory*. IEEE, 2019, pp. 1197–1201.
- [42] D. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, 2nd ed. Belmont: Athena Scientific, 2008.
- [43] Z. Utkovski, P. Agostini, M. Frey, I. Bjelakovic, and S. Stanczak, "Learning radio maps for physical-layer security in the radio access," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications*. IEEE, 2019.
- [44] D. Kline, J. Ha, S. W. McLaughlin, J. Barros, and B.-J. Kwak, "LDPC codes for the gaussian wiretap channel," *IEEE Transactions on Information Forensics & Security*, vol. 6, no. 3, pp. 532–540, 2011.
- [45] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.
- [46] M. Gil, "On Rényi divergence measures for continuous alphabet sources," Master's thesis, Queen's University Kingston, Ontario, Canada, 2011.
- [47] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [48] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.
- [49] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 44, no. 2, 1978.
- [50] D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge: Cambridge University Press, 2009.
- [51] E. Jaynes, *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press, 2003.
- [52] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York, Chichester, Brisbane, Toronto, Singapore: Wiley, 1994, vol. 1.