Delay Minimization for Hybrid Semantic-Shannon Communications

Nikos G. Evgenidis[1], Nikos A. Mitsiou[1,2], Sotiris A. Tegos[1,2], Panagiotis D. Diamantoulakis[1], Panagiotis Sarigiannidis[2], Ioannis Krikidis[3] and George K. Karagiannidis[1,4]

[1]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[2]Department of Informatics and Telecommunication Engineering, University of Western Macedonia, 50100 Kozani, Greece
[3]Department of Electrical and Computer Engineering, University of Cyprus, 1678 Nicosia, Cyprus
[4]Artificial Intelligence & Cyber Systems Research Center, Lebanese American University (LAU), Lebanon

nevgenid@ece.auth.gr, nmitsiou@auth.gr, sotiristegos@ieee.org
padiaman@auth.gr, psarigiannidis@uowm.gr, krikidis@ucy.ac.cy, geokarag@auth.gr

2024 IEEE Wireless Communications and Networking Conference (WCNC) | 979-8-3503-0358-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/WCNC57260.2024.10571303

***Abstract*—Semantic communications offer a promising approach to decrease network congestion and improve reliability, leading to more sustainable and energy-efficient wireless networks. However, the design of semantic transceivers constrain their effectiveness. This paper introduces a novel multi-carrier system that combines both semantic and Shannon communications, with a focus on text transmission. We formulate an optimization problem that jointly selects the transmission method and allocates power to reduce the transmission delay. Despite the challenges of solving this non-convex problem, we employ alternating optimization techniques to address it and the closed-form solution of the power allocation is extracted. The simulation results verify that jointly selecting semantic and Shannon communications decreases the transmission delay compared to using only one of the schemes.**

***Index Terms*—semantic communications, 6G, delay minimization, multi-carrier, resource allocation**

I. INTRODUCTION

Next-generation wireless networks will be designed to enable applications such as Industry 5.0, and the metaverse, which pose new challenges to current communication systems [1], [2]. Although capacity increase is a common direction to address these challenges, the usual way to achieve this is based on bandwidth increase, which leads to an inevitable bottleneck due to severe path loss and the inefficiency of power amplifiers, while requiring more sophisticated user-side hardware to handle the increased bandwidth [2]. Given that in many scenarios the meaning of the data transferred is more important than the data themselves, semantic-based approaches have been gaining ground. To this end, semantic communications can provide an alternative way of communicating by taking into account the difference between the meaning of the original and recovered messages [1]. Semantic communications have become a viable possibility by recent advances in deep neural networks (DNNs). Image processing and natural language processing (NLP) are some of those semantic enabling DNN techniques, which allow for the recognition of original semantic information [1]. In addition, a promising way to facilitate the robustness and reliability of semantic communication systems is the use of autoencoders as joint source-channel coding schemes [3].

Some fundamental principles aiming to facilitate semantic communications have been studied in [4], but the majority of the related works focus on their applicability. The acquisition of information from text and images are the most characteristic. In [5] perceptual understanding was investigated by optimizing the mean squared error (MSE) and the learned perceptual image patch similarities (LPIPS). In [6], a hybrid system utilizing both semantic and conventional transmission was considered, where the latter is used for partial compression of the image and the former is used for other characteristics transmission, both of which are combined to achieve better perceptual similarity. Moreover, a the problem of quantizing the implemented DNN outputs was studied in [7]. Text transmission has also attracted a lot of attention in many works especially due to their practical applications in everyday data. A DNN named DeepSC was introduced in [3] and further utilized in [8] with the aim of minimizing the data transmitted while using end-to-end training to optimize mutual information. In [9], practical approach was considered for DeepSC by applying quantization to the developed unstructured constellation.

Nevertheless, a drawback of these architectures is that DNNs are not always able to achieve the desired accuracy, thus the performance of current semantic frameworks is constrained by the design and performance of semantic DNN transceivers, like DeepSC [3]. In case where a high degree of input-output similarity is essential, relying solely on semantic communications might not be suitable, making Shannon communications indispensable. To the authors' best knowledge, no current work studies the resource allocation of the hybrid semantic-Shannon communication scheme. Driven by this gap, in this work, we focus on the simultaneous operation of Shannon and semantic communications, introducing a multi-carrier system that merges both able to choose between semantic and Shannon communication per subcarrier, aiming at the minimization of the average transmission delay of all subcarriers under

979-8-3503-0358-2/24/$31.00 ©2024 IEEE

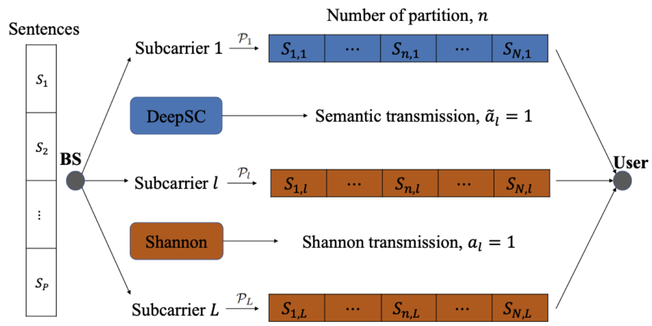

Fig. 1: System model.

similarity constraints. The optimization problem is solved via alternate optimization and useful insights are provided through simulations demonstrating the superiority of the hybrid scheme over the traditional Shannon communications, while semantic communications are not always preferred to Shannon communications.

II. System Model

We assume a point-to-point communication between a BS and a wireless connected user, as shown in Fig. 1, with the aim of completing the task S. Without sacrificing generality, the task S is text consisting of P separate sentences. Moreover, the available number of subcarriers that can be used during the text transmission is denoted as L. Therefore, the overall number of sentences matched with each subcarrier equals $N = \left\lfloor \frac{P}{L} \right\rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor operator. We also assume that P is divided exactly, thus $N = \frac{P}{L}$. The volume of data that are associated to each subcarrier are not equal by design, due to the data size of the sentences varying from sentence to sentence. Also, let S_j denote the j-th sentence that appears in the text, where $j \in \{1, \cdots, P\}$, while $S_{n,l}, n \in \{1, \cdots, N\}, l \in \{1, \cdots, L\}$ reflects the n-th sentence that will be transmitted to l-th subcarrier, as shown in Fig. 1. We note that both notations eventually indicate the same sentence, i.e., $S_j = S_{n,l}$, while the relation between the index j and the pair (n, l) is given as $j = (n-1)L + l$. The available bandwidth of all L subcarriers is given as W, while h_l denotes the channel coefficient of the l-th subcarrer, which also contains the path loss of the transmission. Furthermore, the BS is assumed to have perfect knowledge of the channel state information (CSI), while both the BS and the user can communicate by choosing to utilize either Shannon or semantic communications to the l-th subcarrier. This is shown in Fig. 1. It is important to highlight that a switch between the two available communication methods cannot be done during the coherence time of the channel, nonetheless, the transmission method can potentially be different at two different channel instances.

A. Shannon Communications

In conventional Shannon communications, the information carried by the sentences is mapped into bits, and then wirelessly transmitted. In consequence, the size of the sentence $S_{n,l}$ will equal $b_{n,l}$, where the value of $b_{n,l}$ is subject to the used character encoding standard. For practicality, the American standard code for information interchange (ASCII) will be utilized. Thus, it holds that $b_{n,l} = 8u_{n,l}$, where the number 8 occurs since 8 bits are required to represent a text character, while $u_{n,l}$ denoted the number of text characters in the $S_{n,l}$ sentence.

Shannon-Hartley's theorem dictates that the capacity, i.e., the maximum achievable transmission rate, between a point-to-point wireless communication equals

$$C_l^{\max} = W \log_2 \left(1 + \frac{\mathcal{P}_l |h_l|^2}{N_0 W} \right), \tag{1}$$

where N_0 is the power spectral density of additive white Gaussian noise (AWGN), and $\mathcal{P}_l$ denotes the transmission power at the l-th subcarrier. Despite the fact that there exist capacity-achieving codes, in practice, a gap between the capacity limit and the maximum achievable data rate exists, which is attributed to the bit error rate (BER) of data transmissions. As such, let us consider the case of an uncoded M-QAM constellation, whose BER has been shown in [10] to be upper bounded as follows

$$\text{BER} \leq \frac{1}{5} \exp \left(-\frac{1.5 \mathcal{P}_l |h_l|^2}{N_0 W} \frac{1}{M-1} \right), \tag{2}$$

where M is the modulation order. Consequently, the maximum achievable data rate of an uncoded M-QAM scheme, which guarantees a required BER threshold is given by [10], [11]

$$C_l = W \log_2 \left(1 + \frac{\mathcal{P}_l |h_l|^2}{N_0 W \Gamma} \right), \tag{3}$$

where $\Gamma = {}^{-\ln(5\text{BER})}/_{1.5}$. We note that $\Gamma \geq 1$, while for $\Gamma = 1$ (3) reduces to the Shannon capacity limit. We note, that without loss of generality, any other constellation could have been chosen. Therefore, the overall transmission time until all N sentences are transmitted from the l-th subcarrier will be given as follows

$$D_l = \frac{U_l}{C_l}, \tag{4}$$

where $U_l = 8 \sum_{n=1}^{N} u_{n,l}$. It should be highlighted that since different sentences have a varying number of characters, the total data volume to be transmitted from two different subcarriers is in general unequal, i.e., $\sum_{n=1}^{N} b_{n,l} = \sum_{n=1}^{N} b_{n,l'}$ does not have to hold.

B. Semantic Communications

In the case of semantic transmission we adopt the DeepSC semantic transceiver, which was proposed in [3]. DeepSC makes use of a semantic encoder (decoder) which maps text sentences to real-valued numbers and vice versa. Following the semantic encoder, a channel encoder is utilized to find the optimal constellation which is optimal for semantic data transmission, subject to the channel conditions. The designed constellations are made of infinite points, which implies that the data transmission of the DeepSC tranceiver relies on the discrete time analog transmission (DTAT) [12]. In consequence, the data transmission rate of the semantic transmission is given below

$$\tilde{C}_l = W. \tag{5}$$

Based on the analysis of [3], for a sentence $S_{n,l}$, we denote its semantic counterpart as $S'_{n,l}$, which is the output of the DNN encoder when its input is the sentence $S_{n,l}$. Also, to minimize the impact of both AWGN and channel fading, $S'_{n,l}$ is encoded, through the channel encoder, into the vector $\mathbf{x}_j = [x_1, \cdots, x_{kO_j}]$, where $\mathbf{x}_j$ contains all the encoded symbols that will be wirelessly sent, while O_j is the number of words that consist the j-th sentence, and k is the overall output dimensionality of the DNN semantic encoder. Also, let us denote with $s_{n,l}$ the symbols that will be transmitted after the channel encoder. Then, given the dimensionality k of the semantic encoder output, the semantic symbols to be transmitted per sentence are given as $s_{n,l} = kO_j$. Therefore, the delay of the the semantic transmission is given below

$$\tilde{D}_l = \frac{k \sum_{n=1}^{N} O_{(n-1)L+l}}{\tilde{C}_l}, \tag{6}$$

due to the association between j and (n, l).

It is highlighted that the semantic transmission is degraded by AWGN and channel fading as well. However, rather than calculating the absolute loss of information, i.e., bits in Shannon communications, semantic communications are interested in preserving the semantic similarity, i.e. the meaning resemblance, between the transmitted and the received data. For text transmission, we utilize the cosine similarity metric given by [8]

$$M_{n,l} = \frac{\mathbf{B}(S_{n,l})\mathbf{B}(S'_{n,l})^{\mathrm{T}}}{\|\mathbf{B}(S_{n,l})\|\|\mathbf{B}(S'_{n,l})^{\mathrm{T}}\|}, \tag{7}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operator, and $\mathbf{B}(\cdot)$ is the bidirectional encoder representations from transformers (BERT) for each sentence $S_{n,l}$, which is a vector representation of the original sentence, after its forward pass through the DeepSC encoder. For Shannon communications, it is assumed that all bit errors that occur during transmission can be corrected by coding implementations, such as LDPC codes, that can achieve data rates close to Shannon capacity. Therefore, the similarity between the transmitted and the received sentence equals to one, i.e., $M_{n,l} = 1$ [8]. In contrast, for semantic communications it is that $\tilde{M}_{n,l} \in (0, 1]$, where $\tilde{M}_{n,l}$ reflects the similarity between $S'_{n,l}$ and $S_{n,l}$, due to the fact that the original and the reconstructed sentences can be different in terms of semantic meaning. In fact, the DeepSC model attains a great range of similarity levels, but its similarity is bounded below one. This similarity threshold we will denote as M_{sat}. The attainable similarity levels of the DeepSC model, for different channel conditions can be obtained by offline training and testing, as proposed in [8], and are shown in Fig. 2.

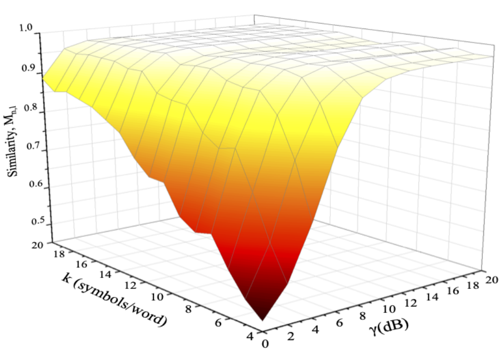

Fig. 2: The performance of the DeepSC model.

From Fig. 2, it can be seen that for a specific value of k, the function of the achievable similarity values of DeepSC is an increasing function with respect to the transmit SNR. As such, the similarity function is injective and each similarity threshold $M_{n,l}^{\mathrm{th}}$ can be uniquely associated to a minimum SNR threshold $\gamma_{n,l}^{\mathrm{th}}$. Therefore, to secure that the received SNR of all semantic symbols satisfies the minimum required threshold, all symbols which belong to one sentence have to be transmitted from one subcarrier, by utilizing the exact same power allocation policy.

III. Maximum Delay Minimization

Transmission delay metrics are very significant for the timely communication of data, and thus, it is also significant to investigate the proposed hybrid semantic-Shannon scheme with respect to the transmission delay. In our work, the maximum delay metric describes the total transmission duration of the proposed multi-carrier scheme, and it has also been widely studied in the literature [13]. Combining (4) and (6), the delay of the data transmission at the l-th subcarrier is equal to

$$\mathcal{D}_l = a_l D_l + \tilde{a}_l \tilde{D}_l, \tag{8}$$

where $a_l, \tilde{a}_l \in \{0, 1\}$ are binary variables such that $a_l + \tilde{a}_l = 1$, $\forall l \in \{1, \cdots, L\}$, which allow selection between Shannon and semantic communication with $a_l = 1$ indicating utilization of the former and $\tilde{a}_l = 1$ indicating utilization of the latter at the l-th subcarrier. Each sentence $S_{n,l}$ is assumed to have a similarity threshold of at least $M_{n,l}^{\mathrm{th}}$, which can be considered as a semantic quality of service (QoS) constraint, meaning that failure to achieve this desired threshold leads to communication outage. Thus, for the sentence $S_{n,l}$ it has to hold,

$$a_l + \tilde{a}_l \tilde{M}_{n,l} \geq M_{n,l}^{\mathrm{th}}, \forall n \in \{1, \cdots, N\}, \forall l \in \{1, \cdots, L\}. \tag{9}$$

In the l-th subcarrier, all sentences are bounded by the similarity constraints given by $M_{n,l}^{\mathrm{th}}$. Within the coherence time, all N sentences associated with a specific subcarrier are subject to equal channel fading. Thus, the power allocated at the l-th subcarrier needs to guarantee the necessary similarity QoS requirement of all N sentences, meaning that the maximum similarity these sentences must be achievable. Taking this into account, the following similarity constraint has to be satsfied

$$M_l^{\max} = \max_{1 \leq n \leq N} \left\{ M_{n,l}^{\mathrm{th}} \right\} \tag{10}$$

for the l-th subcarrier to be able to use the DeepSC. It is highlighted that due to the upper bound of the DeepSC model, regarding its similarity, whenever the desired similarity levels exceed the maximum attainable similarity $M_l^{\max} > M_{\text{sat}}$, Shannon communications is the only viable option. For brevity, we denote the set, that contains all the subcarriers that potentially can utilize semantic communications as

$$\mathcal{S} = \{l | M_{\text{sat}} \geq M_l^{\max}, \forall l\}, \tag{11}$$

while the set that contains all the subcarriers that select the semantic transmission is given as

$$\mathcal{S}' = \{l | \tilde{a}_l = 1, \forall l\}. \tag{12}$$

Since for all $M_l^{\max}$, an one-to-one mapping to a value of $\gamma_l^{\max}$ exists, it is simple to prove that by using the relations (9) and (10), the following power constraint needs to hold

$$\mathcal{P}_l \geq \gamma_l^{\max} c_l, \forall l \in \mathcal{S}', \tag{13}$$

where $c_l = \frac{N_0 W}{|h_l|^2}$.

From the problem above, it is evident that there are three sets of variables that must be optimized, namely the transmission power, denoted as $\boldsymbol{\mathcal{P}} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_L\}$ and the variables concerning the joint selection between Shannon and semantic communications, denoted as $\boldsymbol{a} = \{a_1, a_2, \cdots, a_L\}$ and $\tilde{\boldsymbol{a}} = \{\tilde{a}_1, \tilde{a}_2, \cdots, \tilde{a}_L\}$, respectively. The former variables are constrained by a maximum joint transmission power, $\mathcal{P}_{\text{tot}}$, while the the latter are binary variables. Thus, the formulated optimization problem of minimizing the maximum transmission delay time of all L subcarriers can be described as

$$\begin{aligned} \min_{\boldsymbol{\mathcal{P}}, \boldsymbol{a}, \tilde{\boldsymbol{a}}} \quad & \max_{1 \leq l \leq L} \mathcal{D}_l \\ \textbf{s.t.} \quad & \mathrm{C}_1 : a_l + \tilde{a}_l \tilde{M}_{n,l} \geq M_{n,l}^{\text{th}}, \forall (n, l) \\ & \mathrm{C}_2 : a_l + \tilde{a}_l = 1, \forall l \\ & \mathrm{C}_3 : a_l, \tilde{a}_l \in \{0, 1\}, \forall l \\ & \mathrm{C}_4 : \textstyle\sum_{l=1}^{L} \mathcal{P}_l = \mathcal{P}_{\text{tot}}. \end{aligned} \tag{\textbf{P1}}$$

The arising problem (**P1**) is not convex and, thus, alternating optimization can be used to separately optimize $\mathcal{P}_l$ and $a_l, \tilde{a}_l$. For the optimization of $\mathcal{P}_l$ we first observe that any subcarrier that utilizes semantic communications will satisfy (13) and its delay is not subject to optimization as shown by (6). Therefore, semantic communications can be utilized only by subcarriers whose delay is less than that of the subcarriers that utilize Shannon communications, since the latter must have equal delays amongst them. As such, we present a closed-form solution for the optimization of $\mathcal{P}_l$ along with a heuristic algorithm for the optimal selection between semantic and Shannon utilization. The subcarriers that utilize semantic communications have to allocate such power so that the similarity constraint is satisfied with equality, meaning that whenever holds that $\tilde{a}_l = 1$, it also has to hold that

$$\tilde{\mathcal{P}}_l^* = \gamma_l^{\max} c_l, \forall l \in \mathcal{S}'. \tag{14}$$

Then, for the subcarriers that utilize Shannon communications a closed-form solution can be obtained, because the minimization of the maximum delay problem is reduced in an equality problem between all subsequent D_l. With this in mind, the following analysis holds for any subcarriers such that $m, l \notin \mathcal{S}'$:

$$\frac{U_l}{W \log_2\left(1 + \frac{\mathcal{P}_l}{c_l \Gamma}\right)} = \frac{U_m}{W \log_2\left(1 + \frac{\mathcal{P}_m}{c_m \Gamma}\right)}, \tag{15}$$

which yields that

$$\mathcal{P}_m = c_m \Gamma \left(\left(1 + \frac{\mathcal{P}_l}{c_l \Gamma}\right)^{\left(\frac{U_m}{U_l}\right)} - 1 \right). \tag{16}$$

From the overall power constraint of the problem, the following condition must hold:

$$\mathcal{P}_{\text{tot}} - \sum_{\substack{m=1 \\ m \in \mathcal{S}'}}^{L} \gamma_m^{\max} c_m = \sum_{\substack{m=1 \\ m \notin \mathcal{S}'}}^{L} c_m \Gamma \left(\left(1 + \frac{\mathcal{P}_l}{c_l \Gamma}\right)^{\left(\frac{U_m}{U_l}\right)} - 1 \right). \tag{17}$$

The last one can be solved in terms of $\mathcal{P}_l$ and the rest of the power allocations for the other subcarriers can be found recursively by (16). Using the aforementioned analysis, we propose a heuristic algorithm to solve (**P1**) aiming to find the optimal selection between semantic and Shannon utilization and their joint power allocation problem.

Algorithm 1: Algorithm for selection between semantic and Shannon communication

---

Set k, repetitions I.
Evaluate $M_l^{\max}$ of each subcarrier l.
Map maximum similarity and SNR $M_l^{\max} \longleftrightarrow \gamma_l^{\max}$.
Set initial values in $\boldsymbol{a}, \tilde{\boldsymbol{a}}$ to $a_l = 1$ and $\tilde{a}_l = 0$, $\forall l$.
Find initial solution for (**P1**) applying (16) and (17) to get $\mathcal{P}_l$, $\forall l$.
for $i = 1 : L$ **do**
  Find the delay obtained from the Shannon utilization subset of the problem, Δ_i.
  Find semantic delays $\tilde{\Delta}_l$, $\forall l$ that achieve better delay than Δ_i.
  Concatenate the subcarriers' index that satisfy the condition above into the vector $\mathbf{v} = [m_1 \, m_2 \, \cdots \, m_{|M|}]$, where $|M|$ is the number of elements in $\mathbf{v}$ and order is taken with regard to the corresponding delay $\tilde{\Delta}_m$.
  Set $i' = i$.
  for $m = 1 : |M|$ **do**
    Let $\tilde{\mathcal{P}}_m$ be the power at the m-th subcarrier that is mandatory for utilizing semantic communication.
    if $\tilde{\Delta}_m < \Delta_i$ and $\mathcal{P}_m > \tilde{\mathcal{P}}_m$ **then**
      Set $a_m = 0$ and $\tilde{a}_m = 1$.
      break
  Solve (**P1**) with fixed $\boldsymbol{a}^{(i')}, \tilde{\boldsymbol{a}}^{(i')}$ to get the optimal $\mathcal{P}_l$, $\forall l$.
Output $\boldsymbol{a}^{(i')}, \tilde{\boldsymbol{a}}^{(i')}$ and $\mathcal{P}_l^{(i')}$.

TABLE I: Simulation Parameters

Parameter	Value (Unit)
Noise power spectral density, N_0	-174 dBm/Hz
Total bandwidth, W_{tot}	20 MHz
Frequency, f_c	2.4 GHz
Distance, R	100 m
Path loss exponent, ν	2
Semantic symbols, k	16 symbols/word
Similarity threshold, $M_{n,l}^{\text{th}}$	$[0.6, 1]$
Similarity Upper Bound, M_{sat}	0.98
Sentence Length, L_j	$4 - 32$
Number of subcarriers, L	64
Number of sentences, P	7296

Note that the condition $\tilde{\Delta}_m < \Delta_i$ and $\mathcal{P}_m > \tilde{\mathcal{P}}_m$ for each iteration ensures that if a subcarrier prefers to utilize semantic communication due to its lower delay, the remaining power for the subcarriers that utilize Shannon communication will be greater than the previous iteration. As such, the previous power allocation solution is achievable for the remaining subcarriers utilizing Shannon communications and the new power allocation will necessarily achieve better overall delay due to the larger available power. Therefore, the delay of the Shannon utilizing subcarriers will gradually reduce until no further semantic utilization is possible.

IV. Simulations and Discussion

In the following section the simulation results of the proposed hybrid multi-carrier system are presented. With regards to the communication conditions, we assume that the channel is susceptible to Rayleigh fading that follows a complex Gaussian distribution, i.e., $h \sim \mathcal{CN}(0, l_p)$, where by $l_p = \left(\frac{\lambda_c}{4\pi R}\right)^{\nu}$ we denote the path loss factor. The parameters R, ν and λ_c symbolize the transmitter-receiver distance, the exponent of the path loss and the wavelength of the central carrier frequency f_c, respectively. We also assume an equally distributed bandwidth among all L subcarriers, hence for each subcarrier it is $W = \frac{W_{\text{tot}}}{L}$. DeepSC has been pre-trained, according to [3], to derive its feasible similarity levels with respect to SNR and k.

The parameters used for the simulations are presented in Table I. Monte Carlo analysis has been performed over 500 different channel realizations each with 10 QoS realizations for better averaging. The performance of the proposed hybrid scheme is evaluated by emphasizing on two different measures. The first is the improvement of the transmission delay, occurred by using the hybrid scheme, compared to the traditional Shannon-only multi-carrier protocol and the second is the semantic utilization, i.e., the percentage of subcarriers that select to transmit data based on the semantic protocol. All results are shown with respect to the total average received SNR, which is given as $\text{SNR} = (\mathcal{P}_{\text{tot}}\mathbb{E}[|h|^2])/(N_o W)$, with $\mathbb{E}[\cdot]$ denoting expectation.

In Fig. 3, the semantic utilization is plotted. It is observable that maximum semantic utilization is achieved for medium SNR values, while reaching zero in the small values of SNR and slowly decreasing for high SNR, implying that

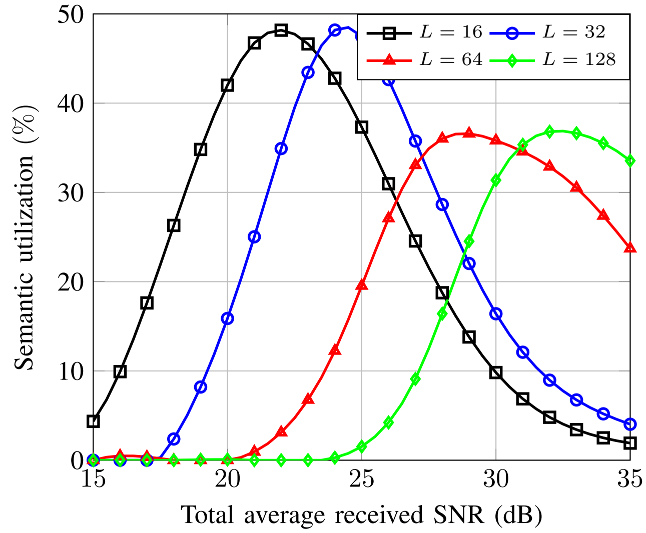

Fig. 3: Semantic utilization for $k = 16$.

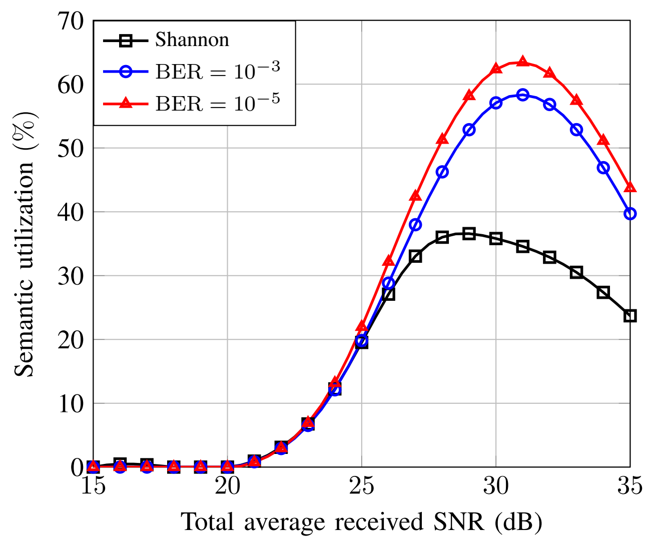

Fig. 4: Semantic utilization for $L = 64$ and $k = 16$.

only Shannon communications are used in these regimes. This behavior is expected, because in the low SNR regime, DeepSC cannot achieve the required similarity, since each subcarrier has not the necessary transmit SNR corresponding to the semantic QoS constraint. In the high SNR regime, the data rate of DeepSC is smaller than that of the Shannon communication scheme, hence the latter's delay gradually decreases and it will eventually achieve better delay than the semantic counterpart. This is a result of the increase of data rates in digital communications as the available transmit power increases, in contrast to the fixed data rate achieved by DeepSC due to its DTAT transmission. As such, semantic transmission is selected mainly in the medium SNR regime. Also, we notice that as the available number of subcarriers grows, the SNR region, where the maximum semantic utilization is achieved, deviates to the right, since increased available transmission power is needed to ensure the semantic QoS at all subcarriers. It is worth noting that maximum semantic utilization ranges in the interval $35 - 50\%$, implying that at least one third of the subcarriers choose to utilize DeepSC. It is important to point out that semantic communications can be utilized even in higher SNR values, because subcarriers that can achieve considerably small transmission delay will favor semantic utilization over Shannon until the latter is preferable for each subcarrier.

In Fig. 4, the semantic utilization is presented under different BER thresholds. It is noted that when BER is taken into account, the semantic communications utilization increases, due to the fact that BER limits the maximum achievable data rate of digital communication as shown in (3). Nevertheless, in the low and high SNR regimes, the semantic utilization again drops towards zero. Therefore, semantic communications can be used to reduce the overall transmission delay, especially in the medium SNR regime. We note that this behavior is similar to the one under the assumption of capacity achieving data rate transmission, since increasing the SNR eventually allows Shannon communications to outperform the delay achieved by semantic communications.

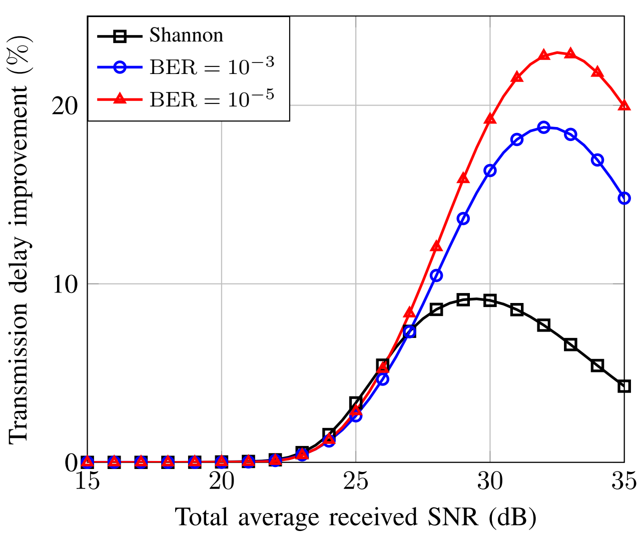

Fig. 5: Transmission delay improvement for $L = 64$ and $k = 16$.

In Fig. 5, the transmission delay improvement and different BER thresholds is presented. The improvement is associated with the semantic utilization, since the larger the latter is, the larger the improvement will be. However, it is observed that the transmission delay improvement has a similar behavior to that of Shannon's capacity case, because the same data rates will eventually be achieved by the uncoded M-QAM schemes, but for greater values of SNR. This ensures that at some point the transmission delay time of each subcarrier under Shannon communications will become smaller that the one achieved by the semantic communications and, thus, as suggested by Fig. 4, the latter will not provide any improvement.

V. Conclusions

In this work, the coexistence of conventional Shannon communications with a semantic communication system for text transmission was studied. The minimization of the maximum delay was investigated, subject to strict similarity levels between the original transmitted and the received data, and an algorithm has been proposed to find the optimal power allocation and whether to choose Shannon or semantic utilization for each subcarrier. The simulation results illustrate that semantic communications are not always the preferable way of transmission and Shannon communications still achieve better transmission delay, even for non capacity-achieving data rates, for specific SNR regimes. Our results indicate that further research is needed to facilitate the coexistence of both technologies, under different metrics, while it is necessary to study in which scenarios semantic communications may not be the preferable mean of communication.

Acknowledgment

This research has received funding from the European Union's Horizon Europe Framework Programme under grant agreement No 101096456.

References

[1] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 213–250, Firstquarter 2023.

[2] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G Wireless Communications: Vision and Potential Techniques," *IEEE Netw.*, vol. 33, no. 4, pp. 70–75, Jul./Aug. 2019.

[3] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.

[4] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," 2022. [Online]. Available: https://arxiv.org/abs/2201.01389

[5] E. Erdemir, T.-Y. Tung, P. L. Dragotti, and D. Gündüz, "Generative joint source-channel coding for semantic image transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2645–2657, Aug. 2023.

[6] X. Niu, X. Wang, D. Gündüz, B. Bai, W. Chen, and G. Zhou, "A hybrid wireless image transmission scheme with diffusion," 2023. [Online]. Available: https://arxiv.org/abs/2308.08244

[7] T.-Y. Tung, D. B. Kurka, M. Jankowski, and D. Gündüz, "DeepJSCC-Q: Constellation constrained deep joint source-channel coding," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 4, pp. 720–731, Dec. 2022.

[8] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1394–1398, Jul. 2022.

[9] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, Jan. 2021.

[10] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[11] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[12] Y. Shao and D. Gunduz, "Semantic communications with discrete-time analog transmission: A papr perspective," *IEEE Wireless Commun. Lett.*, vol. 12, no. 3, pp. 510–514, Mar. 2023.

[13] Y. Zhou, C. Pan, P. L. Yeoh, K. Wang, M. Elkashlan, B. Vucetic, and Y. Li, "Communication-and-computing latency minimization for UAV-enabled virtual reality delivery systems," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1723–1735, Mar. 2021.