Semantic Communications for SWIPT Systems

Nizar Khalfet, *Member, IEEE*, Constantinos Psomas, *Senior Member, IEEE*, and Ioannis Krikidis, *Fellow, IEEE*

Abstract

In this paper, we study the fundamental limits of simultaneous semantic information and power transfer in wireless networks, where we consider both the point-to-point case as well as the Gaussian multiple access channel (MAC). Specifically, for the point-to-point case, we consider a three-party communication system, where a transmitter aims to simultaneously convey semantic information to an information receiver and deliver energy to an energy harvesting receiver (ER). An achievable and a converse region in terms of information and energy rates are presented for both the discrete memoryless (DM) and Gaussian channel. For the DM channel, the achievable region is obtained by using the asymptotic equipartition property and a converse region is obtained by using outer bounds on the semantic information rates. For the Gaussian channel, we characterize an achievable region by using a power splitting technique between the information and the semantic context parts. A converse region is obtained that provides an estimate on the information-energy capacity while taking into account semantics. On the other hand, for the Gaussian MAC case, we consider a setup where a semantic transmitter and a conventional transmitter are employed subject to an energy harvesting constraint at the ER. Specifically, we characterize the semantic-bit information energy region, by providing an achievable and a converse region. Numerical results show that in both cases a higher performance can be achieved in terms of information and energy rates by considering a low semantic ambiguity code in comparison to the classical coding scheme (without semantic). Moreover, in the context of Gaussian MAC, it is shown that is better to use semantic communications in scenarios with low signal-to-noise ratio (SNR), while conventional communications is preferred at high SNRs.

Index Terms

SWIPT, semantic-energy capacity region, discrete memoryless channel, multiple access channel.

N. Khalfet, C. Psomas, and I. Krikidis are with Department of Electrical and Computer Engineering, University of Cyprus, Cyprus (e-mail: {khalfet.nizar, psomas, krikidis}@ucy.ac.cy).

I. INTRODUCTION

When Shannon established the theoretical foundation of communications engineering in 1948, he intentionally excluded semantic aspects from the system design [1]. However, Weaver [2] expanded upon Shannon's work and proposed a classification of communication into three levels. The first level involves the transmission of symbols, as defined in Shannon's classical Information Theory. At this level, the primary concern is the accuracy with which the communication symbols are transmitted. The second level pertains to the precision with which the transmitted symbols convey the desired meaning. This level recognizes the importance of semantic information in communication, encompassing the interpretation of meaning by the information receiver (IR), referred to as semantic communications. The third level focuses on the effects of semantic information exchange on communication and coding schemes. This level explores how the received meaning influences communication processes and the design of coding systems. By considering these three levels, we gain a more comprehensive understanding of communication, encompassing not only the accurate transmission of symbols but also conveying the meaning and the impact of semantic information on the overall communication process.

In the last few decades, wireless communication systems have experienced significant development from the first generation (1G) to the fifth generation (5G) with the system capacity gradually approaching the Shannon limit [3]. Nevertheless, the meaning behind the transmitted data in classical Shannon's framework is expected to play an important role in 6G communications. Therefore, semantic communication techniques can be used to enable wireless devices to extract and transmit the meaning of original data to reduce the heavy congestion of current wireless networks, thus improving the network's efficiency [4]. However, using semantic communication techniques for data transmission faces several challenges such as semantic information modeling and extraction, original data recovery, and the definition of appropriate semantic metrics that can capture the effects of wireless factors (e.g., transmit power, packet errors) on semantic communications [5]. Existing research suggests that semantic communications shows promise in scenarios where the signal-to-noise ratio (SNR) is low or the available wireless resources are limited. In other words, semantic communications typically requires less power or bandwidth resources compared to conventional communications while achieving similar performance [6]. The authors in [7] use a novel approach to model the semantics through a Bayesian game, where the semantic similarity is used as a semantic error metric. The work in [8] proposes a signal shaping method by minimizing the semantic loss, which is measured by the pretrained bidirectional encoder representation from transformers (BERT). In [9], the authors propose a deep learning based multi-user semantic communication system that can extract the semantic information of image and text from different users. The authors in [10] investigate a novel framework that enables users to communicate with a base station using a semantic communication and energy harvesting (EH) technique. Semantic data can be compressed to a proper size for transmission by using a lossless method [11], which utilizes the semantic relationship between different messages.

Given these significant advantages of semantic communications, it is reasonable to explore their application in other promising communication technologies to further enhance the performance. Among them, simultaneous wireless information and power transfer (SWIPT) is a technology that leverages the dual nature of radio frequency (RF) signals, enabling them to carry both information and energy [12]. The concept of wireless power transfer was initially proposed by Tesla in the 20th century [13], and it now presents a promising solution for future communication systems such as low-power short-range communication systems, sensor networks, machine-type networks, and body-area networks [14]. In the context of point-to-point SWIPT scenarios, the work in [15] first formalized the notion of information-energy capacity region. This work was further extended in [16] to include parallel links in point-to-point channels. More recent research has focused on integrating SWIPT into more complex network topologies, such as multiple access channels (MAC) [17], interference channels [18], multiple-input multiple-output systems [19], multiple-antenna cellular networks [20], and others. A comprehensive overview of existing results in SWIPT for various fundamental multi-user channels is provided in [21]. Ongoing research continues to explore and advance the understanding of SWIPT in various network scenarios, aiming to optimize its performance and unlock its full potential benefits in modern communication applications. Specifically, in a SWIPT communication network with semantics, users can boost their EH capabilities by compressing data using semantic encoding. This combination of semantic information and EH has applications in wireless sensor networks, internet of things, and smart grids [22]. Semantic communication helps these systems to make better decisions based on data context, leading to more efficient energy management, reduced waste, and improved EH. Using semantic communication protocols in MAC scenarios [23] allows multiple users to efficiently share communication resources, enhancing capacity and reducing interference. Integrating MAC and SWIPT enables energy-efficient communication, improved cooperation, sustainability, and extended network coverage. These advancements lead to more advanced wireless communication systems that address energy constraints and enhance EH performance [17].

With the exception of a few studies (e.g., [6]), existing works do not consider the effects of multi-user setting on semantic communication performance. In addition, most of the aforementioned studies on SWIPT systems focus on the conventional communication and therefore the impact of semantics has not been investigated. To the best of the authors' knowledge, this is the first work that takes into account the effect of semantics on SWIPT systems over multi-user setting from an information theory standpoint. The main contribution of this work is a novel framework that enables the study of the fundamental limits of SWIPT with semantic communications in wireless networks. In particular, we first consider a basic point-to-point semantic SWIPT communication system, where a transmitter simultaneously sends data to an IR and power to an EH receiver (ER). We propose an information-theoretic framework to characterize an achievable information-energy region as well as its converse for the discreet memoryless (DM) channel and the Gaussian channel by taking into account the semantic context into the communication. For both cases, an achievable region is obtained by using the asymptotic equipartition property (AEP) and a converse region is obtained to provide an estimate of the information-energy capacity while taking into account semantics. For the Gaussian multiple access channel (MAC), we consider a system where a semantic transmitter and a conventional transmitter are employed subject to an EH constraint at the ER. Specifically, we characterize the semantic-bit information energy region by providing an achievable and a converse region for the information-energy region. For the proof of achievability, the key idea is the power splitting between the two signal components: the information-carrying component and the semantic context component. The construction of the former is based on random coding arguments, whereas the latter consists of a deterministic sequence known by the first transmitter and the IR. The proof of the converse is obtained using Fano's inequality for semantic communications and appropriate concentration inequalities. By studying the DM channel, the Gaussian channel, and the Gaussian MAC, we gain a deeper understanding of the impact of noise, interference, and other factors on the performance limits of SWIPT systems with semantics. This knowledge can be leveraged to design efficient SWIPT systems that are more resilient and reliable, with improved information and energy capabilities. Numerical results show that by considering a low semantic ambiguity code, a higher performance is observed in comparison to conventional

Notation	**Description**	**Notations**	**Description**
n	Number of channel uses	X	Random variable of the input signal
M	Cardinality of the states	$\boldsymbol{T}$	Random variable of the extended alphabet
t	Time index	$\mathcal{C}(b)$	Information energy capacity region
P_{req}	Average power constraint for transmitter	X^i	Sequence of the the input distribution with length i
E_{req}	Energy required at the energy receiver	$\Pr(\cdot)$	Probability operator
h_1	Channel fading for the information link	$I(X;Y)$	Mutual information between X and Y
h_2	Channel fading for the energy link	$H(X)$	Entropy of random variable X
$F^{(n)}$	Sequence of a distribution F	$I(X;Y)$	Mutual information between X and Y
Q	Random variable of semantic context	$\mathcal{CN}(0,\sigma^2)$	Circular complex Gaussian random variable with zero mean and variance σ^2

TABLE I: Summary of notation.

communication approaches (i.e., without semantics). Moreover, in the Gaussian MAC case, it is typically better to use semantic communications when the SNR is low, whereas conventional communications is preferred at high SNRs.

Notation: The realization and the set of the events from which the random variable X takes values are denoted by x and $\mathcal{X}$, respectively. The argument $\mathbb{E}[X]$ denotes the expectation with respect to the distribution of a random variable X. $\mathcal{C}(b)$ represents the information-energy capacity region given an EH constraint b. $\tilde{Q}(\cdot)$ denotes the tail distribution function of the standard normal distribution. Table I summarizes the key notation of the paper.

The remainder of this paper is structured as follows. The fundamental limits of the point-to-point case are presented in Section II. In Section III, we characterize the information-energy capacity region for the case of Gaussian MAC with a single semantic transmitter. Finally, numerical results are presented in Section IV and V concludes the paper.

II. Point-to-point SWIPT with semantics

Consider a three-party semantic communication setup, in which a semantic transmitter aims to simultaneously convey information to an IR and power to an ER. In particular, we consider two cases: a DM channel (see Fig. 1) and a Gaussian channel (see Fig. 2). The study of the DM and the Gaussian channels is crucial as it provides the foundation for characterizing the fundamental limits of SWIPT systems with semantic communications. Specifically, gaining insights into these channels can enhance our understanding on how SWIPT systems operate and how they can

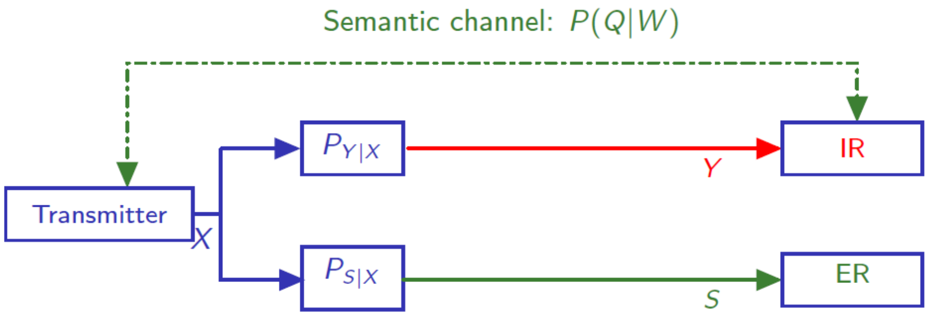

Fig. 1: A SWIPT system over a DM channel for semantic communication.

be optimized for maximum efficiency. First, we provide the following notations that will be used to characterize the semantic information energy region for the point-to-point channel with semantics:

- $I(X;Y) = H(X) - H(X|Y)$ is the mutual information between X and Y, where $H(X)$ and $H(X|Y)$ are the corresponding entropy of X and the conditional entropy of X given Y.
- $H(X|W)$ is the equivocation of the semantic encoder. Specifically, a higher $H(X|W)$ means higher semantic redundancy in the semantic coding.
- $H(Q)$ measures the semantic context source, or the local information available at the transmitter and the IR. Specifically, a higher $H(Q)$ means strong ability of the IR to interpret received messages.

A. DM channel with SWIPT and semantics

The transmitter sends a message w from a set $\mathcal{W}$ with a probability distribution $P(W = w)$ to an encoder. The message w is encoded into a channel input $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ by using an encoding function $\phi : \mathcal{W} \to \mathcal{X}^n$, where n denotes the number of channel uses. The output at the decoder is given by $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ from the set $\mathcal{Y}^n$ and is observed at the IR with probability

$$\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} \Pr(y_i|x_i), \tag{1}$$

where $\Pr(y_i|x_i)$ is the transition probability distribution. The decoder uses a decoding function $\psi : \mathcal{Y}^n \times \mathcal{Q} \to \mathcal{W}$. The output at the ER is given by $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$ from the set $\mathcal{S}^n$ and is observed at the IR with probability

$$\Pr(\mathbf{S}=\mathbf{s}\mid\mathbf{X}=\mathbf{x})=\prod_{i=1}^{n}\Pr(s_i|x_i), \tag{2}$$

where $\Pr(s_i|x_i)$ is the transition probability distribution.

In contrast to conventional communication systems, we assume the existence of side information provided via a genie-aided channel to the IR [24]. This side information is likely to be useful in helping the IR to better understand and interpret the semantic information being transmitted. Specifically, the communication between the transmitter and the IR is taking place within a specific context, which means that the transmitted information is related to a particular topic or area of interest [7]. Therefore, semantic context can influence the IR on how it decodes the received signals, depending on the tasks and/or actions to be executed. The semantic context is characterized by a random variable Q with respect to a probability distribution $P(Q|W)$ [7], which satisfies

$$\sum_{q\in\mathcal{Q}}\Pr\left(Q=q|W=w\right)=1, \tag{3}$$

where Q denotes the context random variable. Furthermore, we define the semantic distance between the words w and $\hat{w}$ as,

$$d(w,\hat{w})=1-\mathrm{sim}(w,\hat{w}), \tag{4}$$

where $0\leq\mathrm{sim}(w,\hat{w})\leq 1$ denotes the semantic similarity between w and $\hat{w}$. By following similar steps as in [7], the average semantic error denoted by P_{SE} is given by

$$\mathrm{P}_{\text{SE}}=\sum_{w\in\mathcal{W},q\in\mathcal{Q},\boldsymbol{y}\in\mathcal{Y}^n,\boldsymbol{x}=\phi(w)}\Pr(Y=y|X=x)\Pr(Q=q|W=w)\Pr(W=w)d(w,\psi(\boldsymbol{y},q)). \tag{5}$$

Let $\mathcal{E}:\ \mathcal{S}\rightarrow\mathbb{R}_+$ be the function that determines the average energy harvested, given by

$$\mathcal{E}(\boldsymbol{s})=\frac{1}{n}\sum_{i=1}^{n}g(s_i), \tag{6}$$

where $g:\mathcal{S}\rightarrow\mathbb{R}_+$ is a positive real-valued function that determines the energy harvested from the output symbols. Then, the probability of energy-shortage when transmitting the message w can be written as

$$\mathrm{P}_{\text{ES}}(b)=\Pr\left(\mathcal{E}(\boldsymbol{s})\leq b\right), \tag{7}$$

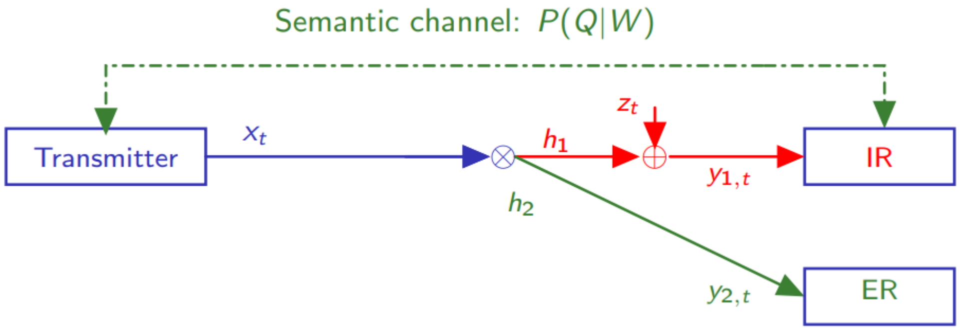

Fig. 2: A SWIPT system over Gaussian channel for semantic communication.

where b denotes the targeted energy rate at the ER. The system is said to be operating at the semantic information-energy rate $R(b) \in \mathbb{R}$ when both the transmitter and the IR use a transmit-receive configuration such that: (i) a reliable semantic communication at rate R is ensured; and (ii) a reliable energy transmission at energy rate b is ensured. A formal definition is given below.

Definition 1. *From a semantic communication standpoint, an information-energy rate $R(b)$ is achievable, if the probability of miss-interpretation of the message w, given the context Q, satisfies the limit $\mathrm{P}_{SE} \rightarrow 0$, and the energy shortage probability, $\mathrm{P}_{ES}(b)$, satisfies $\mathrm{P}_{ES}(b) \rightarrow 0$, for $n \rightarrow \infty$.*

By using Definition 1, the fundamental limits of semantic information and energy transfer over a DM channel can be described by the semantic-energy capacity region, defined as follows.

Definition 2. *The information-energy capacity region $C(b)$ is defined as the maximum rate over all the achievable rates, i.e.,*

$$C(b) \triangleq sup\{R(b) : R(b) \;\; is \; achievable\}. \tag{8}$$

B. Gaussian channel with SWIPT and semantics

In this case, the message w is mapped into a vector $\boldsymbol{x} \in \mathbb{R}^n$, where n is the number of channel uses. The channel gains for the IR and ER are considered constant and denoted by $h_1 \in \mathbb{R}$ and $h_2 \in \mathbb{R}$, respectively. The received signal at the IR during channel use t is given by

$$y_{1,t} = h_1 x_t + z_t, \tag{9}$$

and the ER observes

$$y_{2,t} = h_2 x_t, \tag{10}$$

where $Z \sim \mathcal{N}(0,1)$ is the Gaussian noise with unit variance and is assumed to be independent of the signal X. The semantic encoder is subject to an average power constraint of the form

$$\frac{1}{n}\sum_{t=1}^{n} x_t^2 \leq P_{\text{req}}, \tag{11}$$

where P_{req} is the transmit power constraint. By following similar steps as in [8], an upper bound of the semantic loss for the Guassian case when a maximum likelihood (ML) detector is used, is given by

$$P_{\text{SE}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} d(x_i, x_j)\tilde{Q}\left(P_{\text{req}}\sqrt{\frac{||x_i - x_j||^2}{2}}\right), \tag{12}$$

where $d(w, \hat{w}) = 1 - \text{sim}(w, \hat{w})$, with

$$\text{sim}(x_i, x_j) = \frac{B_\Phi(x_i)B_\Phi(x_j)^T}{||B_\Phi(x_i)||\,||B_\Phi(x_j)||}, \tag{13}$$

where $B_\Phi(\cdot)$ is the pretrained BERT model, which provides an efficient way to quantify the semantic similarity between two different messages [25] and $||\boldsymbol{x}||$ denotes the norm of vector $\boldsymbol{x}$.

For EH, let $\mathcal{E} : \mathbb{R} \rightarrow \mathbb{R}_+$, where $\mathcal{E}(y)$ determines the average harvested energy function, given by[1]

$$\mathcal{E}(y_2) = \frac{1}{n}\sum_{t=1}^{n} y_{2,t}^2. \tag{14}$$

Therefore, given an energy constraint b, the energy shortage probability denoted by P_{ES}, is written as

$$\text{P}_{\text{ES}} = \Pr\left(\mathcal{E}(y_2) < b\right). \tag{15}$$

C. SWIPT with semantics

In the following, we provide a description of the semantic information-energy region, presented in the form of an approximation in the sense of an achievable and a converse region.

[1]For the purpose of this paper, we adopt a linear approximation of the non-linear EH characteristic. The motivation of using the linear EH model is two-fold: first, it is analytically tractable, and second, it approximates the performance of practical EH circuits (e.g., linear operation regime) [26].

1) Achievable and converse regions for the DM channel: Let $b_0 = \max_{s\in\mathcal{S}} g(s)$ and $b_1 = \min_{s\in\mathcal{S}} g(s)$. By using these definitions, the following theorem introduces an achievable information-energy region for semantic communication over a DM channel.

Theorem 1. *The information-energy capacity region for semantic communication is lower bounded by the function* $\mathcal{C} : [b_1, b_0] \to \mathbb{R}_+$ *with*

$$\mathcal{C}(b) \geq \max_{\Pr(X|W)|\mathcal{E}(\boldsymbol{s})\geq b} I(X;Y) - H(X|W) + H(Q).$$

Proof: The proof is presented in Appendix A. ■

Remark 1. *From the above expression, the achievable semantic energy region could be lower or higher than the conventional Shannon capacity region* $\sup\{I(X;Y)\}$ *depending on the term* $-H(X|W) + H(Q)$. *Specifically, we may achieve a higher semantic information-energy region by considering a semantic encoder with a codebook of lower ambiguity with respect to the semantic context* Q.

The following theorem introduces an upper bound for the SWIPT semantic information-energy region.

Theorem 2. *The information-energy capacity region for semantic communications is upper bounded by the function* $\mathcal{C} : [b_1, b_0] \to \mathbb{R}_+$, *i.e.,*

$$\mathcal{C}(b) \leq \max_{\Pr(X|W)|\mathcal{E}(\boldsymbol{s})\geq b} H(X) - H(X|W) - H(X|W,Y) + H(Q).$$

Proof: The proof is presented in Appendix B. ■

Thus, Theorem 2 provides a fundamental limit on the semantic rate at which information can be transmitted over a DM channel, while ensuring a certain level of energy reliability. It establishes a lower bound on the error probability of any reliable communication scheme and provides a guideline for designing efficient coding schemes that approach the semantic capacity.

2) Achievable and converse regions for the Gaussian channel: In the case of the memoryless Gaussian channel, the alphabets are continuous. Nonetheless, information and energy transfer can be described similarly to the DM channel, where the finite input and output alphabets are replaced by $\mathbb{R}$. The following theorem introduces a lower bound on the information-energy capacity region for the Gaussian channel.

Theorem 3. *The information-energy capacity is lower bounded by the function $\mathcal{C} : \mathbb{R}_+ \to \mathbb{R}_+$, i.e.,*

$$\mathcal{C}(b) \geq \max_{\Pr(X|W)|\mathcal{E}(y_2)\geq b} \frac{1}{2}\log(1+\lambda_1 P_{\text{req}}) - h(X|W) + \frac{1}{2}\log(1+\lambda_2 P_{\text{req}}),$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ denote the fraction of the power dedicated to the input X and the semantic context Q, respectively, and $h(X|W)$ denotes the conditional differential entropy of the input X given the message W.

Proof: The proof is presented in Appendix C. ■

The following theorem introduces an upper bound for the semantic information-energy region.

Theorem 4. *The information-energy capacity region for the SWIPT semantic communication is upper bounded by the function $\mathcal{C} : \mathbb{R}_+ \to \mathbb{R}_+$, i.e.,*

$$\mathcal{C}(b) \leq \max_{\mathcal{E}(\boldsymbol{y}_2)\geq b} \frac{1}{2}\log(1+\mu_1 P_{\text{req}}) + \frac{1}{2}\log(1+\mu_2 P_{\text{req}}),$$

where $\mu_1 \geq 0$ and $\mu_2 \geq 0$ denote the fraction of the power dedicated to the input X and the semantic context Q, respectively.

Proof: The proof is presented in Appendix D. ■

The converse of the information-energy region of a Gaussian channel with semantics is a fundamental concept in information theory. It sets an upper limit on the rates of communication possible over this channel and describes the trade-off between the semantic rate and the achievable EH rate.

III. Multi-user SWIPT with Semantics

For the two-user memoryless Gaussian MAC scenario, we have a two hybrid transmitters, one semantic transmitter and one conventional (bit) transmitter and a non-collocated ER. At each channel use, $t \in \{1, 2, \ldots,\}$, we denote by $x_{1,t}$ and $x_{2,t}$ the real symbols sent by the semantic transmitters and the conventional transmitter, respectively. The IR observes the real channel output

$$y_{1,t} = h_{11}x_{1,t} + h_{12}x_{2,t} + z_t, \tag{16}$$

and the ER observes

$$y_{2,t} = h_{21}x_{1,t} + h_{22}x_{2,t}, \tag{17}$$

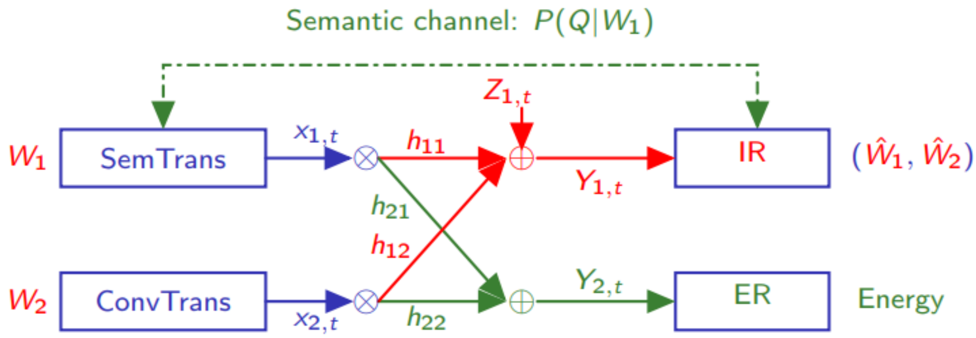

Fig. 3: A SWIPT system over the Gaussian MAC for semantic communication.

where h_{1i} and h_{2i} denotes the corresponding real channel coefficients from the semantic and conventional transmitter to the IR and the EH, respectively. The noise z_t is assumed to be a realization of identically distributed zero-mean unit variance real Gaussian random variables. The semantic transmitter sends a message W_1 from a set $\mathcal{W}_1 = [1, \ldots, 2^{nR_1}]$ to the IR. Similarly, the conventional transmitter sends a message W_2 from a set $\mathcal{W}_2 = [1, \ldots, 2^{nR_2}]$ to the IR, where R_1 and R_2 denote the semantic and information rate, respectively. The message w_i, $i \in \{1, 2\}$ is encoded into a channel input $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n}) \in \mathcal{X}_i^n$ by using an encoding function $\phi_i : \mathcal{W}_i \to \mathcal{X}_i^n$.

Following a similar approach as in the point-to-point case, we assume that the communication between the bit transmitter and the IR is taking place within a specific semantic context. The semantic context for the semantic transmitter is characterized by a random variable Q with respect to a probability distribution $\Pr(Q|W_1)$. We assume that the channel inputs for both transmitters are subject to an average power constraint, i.e.,

$$\frac{1}{n}\sum_{t=1}^{n} x_{i,t}^2 \leq P_{\text{req}}, \tag{18}$$

where P_{req} denotes the average transmit power of both transmitters. Therefore, the average harvested energy and the energy shortage probability are given by (14) and (15), respectively.

The system is said to be operating at the semantic/conventional information-energy rate $(R_1(b), R_2(b)) \in \mathbb{R}_+^2$ when both transmitters and the IR use a transmit-receive configuration such that: (i) a reliable semantic communication at rate $R_1(b)$ is ensured; (ii) a reliable conventional communication at rate $R_2(b)$ is ensured; and (iii) a reliable energy transfer at energy rate b is ensured. A formal definition is given below.

Definition 3. *From a semantic communication standpoint, an information-energy rate double* $(R_1(b), R_2(b))$ *is achievable if: 1) the probability of miss-interpretation of the message* W_1*, given the context* Q*, satisfies the limit* $\mathrm{P}_{SE} \to 0$ *2) the probability of error* P_{SE} *of the message* W_2 *satisfies the limit* $\mathrm{P}_{SE} \to 0$*, for* $n \to \infty$*. 3) the energy shortage probability,* $\mathrm{P}_{ES}(b)$*, satisfies* $\mathrm{P}_{ES}(b) \to 0$*, for* $n \to \infty$*.*

By using Definition 3, the fundamental limits of semantic information and energy transfer over Gaussian MAC can be described by the semantic-energy capacity region, defined as follows.

Definition 4. *The semantic-bit information-energy capacity region* $C_1(b)$ *is defined as the maximum rate over all the achievable rates, i.e.,*

$$C_1(b) \stackrel{\triangle}{=} sup\{(R_1(b), R_2(b)) : (R_1(b), R_2(b)) \text{ is achievable}\}. \tag{19}$$

A. Achievable semantic-bit information-energy region

The following Theorem introduces a lower bound for the semantic-bit information-energy capacity region.

Theorem 5. $\mathcal{C}(b)$ *contains all* $(R_1(b), R_2(b))$ *that satisfy*

$$\begin{aligned} R_1 &\leqslant \frac{1}{\alpha}\Big(\frac{1}{2}\log(1+\lambda_1 P_1) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1)\Big), \\ R_2 &\leqslant \frac{1}{2}\log(1+P_2), \\ R_1 + R_2 &\leqslant \frac{1}{2}\log\big(1+\lambda_1 P_1 + P_2\big) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1), \\ b &\leqslant 1 + P_1 + P_2 + \sqrt{\lambda_3 P_1 P_2}, \end{aligned}$$

with $(\lambda_1, \lambda_2, \lambda_3) \in [0,1]^3$*.*

Proof: The proof is presented in Appendix E. ■

The achievable region for a Gaussian MAC with a semantic transmitter involves considering transmission rates and meaningful information conveyed. It represents rate combinations that reliably communicate, while preserving semantics and ensuring minimum energy rate b at the ER. The achievable information energy region with semantics considers accuracy and fidelity of semantic exchange, balancing rates and semantic quality. Points within the defined region adhere to both semantic preservation and transmission requirements. These points characterize scenarios where the intended semantics are effectively transmitted and comprehended. The boundary of this region outlines the upper limits for the achievable transmission rates while upholding semantic integrity and ensuring a minimum EH rate at the ER.

Now, for the case where both transmitters are assumed to be semantic, an achievable region for two user semantic Gaussian MAC is characterized by the following theorem.

Theorem 6. *$\mathcal{E}_b$ contains all (R_1, R_2, b) that satisfy*

$$\begin{aligned} R_1 &\leqslant \frac{1}{\alpha}\Big(\frac{1}{2}\log(1+\lambda_1 P_1) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1)\Big), \\ R_2 &\leqslant \frac{1}{\alpha}\Big(\frac{1}{2}\log(1+\lambda_2 P_2) - h(X_2|W_2) + \frac{1}{2}\log(1+\lambda_1 P_2)\Big), \\ R_1 + R_2 &\leqslant \frac{1}{2}\log\big(1+\lambda_1 P_1 + P_2\big) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1), \\ b &\leqslant 1 + P_1 + P_2 + \sqrt{\lambda_3 P_1 P_2}, \end{aligned}$$

with $(\lambda_1, \lambda_2, \alpha) \in [0,1]^3$.

Proof: The proof is presented in Appendix F. ■

B. Converse semantic-bit information-energy region

In order to prove the converse of the semantic-bit information-energy region, we will use the semantic Fano's inequality introduced in [27]. The following theorem describes a converse for the semantic-bit information-energy region.

Theorem 7. *$\mathcal{E}(b)$ is contained into the set of all (R_1, R_2, b) that satisfy*

$$\begin{aligned} R_1 &\leqslant \frac{1}{2}\frac{1}{\alpha}\left(\log(1+\mu_1 P_1) + h(X_1|W_1)\right), \\ R_2 &\leqslant \frac{1}{2}\log(1+P_2), \\ R_1 + R_2 &\leqslant \frac{1}{2}\log\big(1+\mu_1 P_1 + P_2\big) + h(X_1|W_1) + \frac{1}{2}\log(1+\mu_2 P_1), \\ b &\leqslant 1 + P_1 + P_2 + \sqrt{\mu_3 P_1 P_2}, \end{aligned}$$

with $(\mu_1, \mu_2, \alpha) \in [0,1]^3$.

Proof: The proof is presented in Appendix G. ■

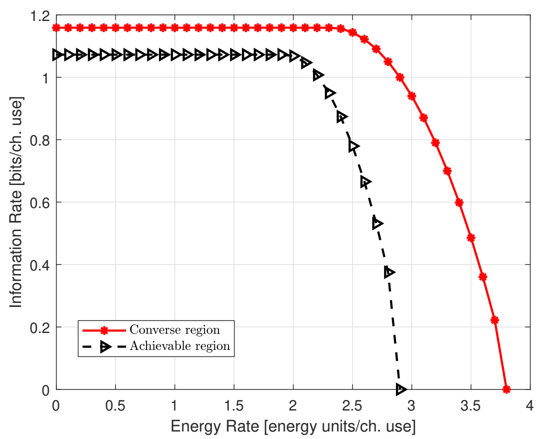

Fig. 4: Achievable semantic information-energy region versus converse region over the DM channel.

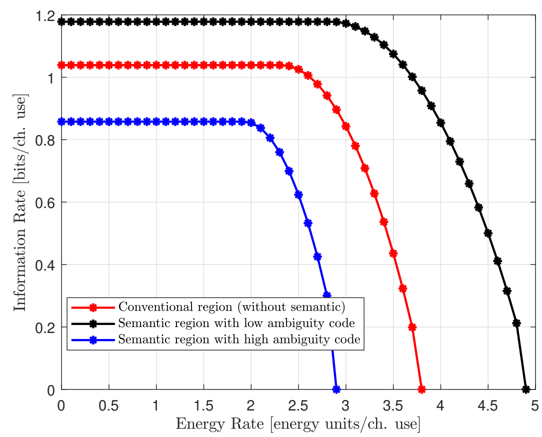

Fig. 5: Achievable semantic information-energy region versus conventional capacity over the DM channel.

IV. Numerical Results

For the DM channel, without loss of generality, we consider a binary symmetric channel (BSC) with crossover probability ρ, i.e.,

$$\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \rho^{l(\mathbf{y},\mathbf{x})}(1-\rho)^{n-l(\mathbf{y},\mathbf{x})}, \tag{20}$$

where $l(\mathbf{y}, \mathbf{x})$ is the Hamming distance between $\mathbf{y}$ and $\mathbf{x}$. For the sake of illustrating our results, we consider a binary set of context $Q = \{q_1, q_2\}$, which satisfies the following distribution [7]

$$\Pr(Q = q_1 | W = w) = \begin{cases} 1, & \text{if } w = \text{car, automobile}, \\ 0, & \text{if } w = \text{bird}, \\ 0.5, & \text{if } w = \text{crane}. \end{cases} \tag{21}$$

The average energy harvested at the ER for the BSC case is given by [28]

$$E(\rho) = \frac{1}{n}\sum_{n=1}^{n}[(1-\rho)\Pr(x_n = 1) + \rho\Pr(x_n = 0)]b_1 + [\rho\Pr(x_n = 1) + (1-\rho)\Pr(x_n = 0)]b_0,$$

with $b_0 = g(0)$ and $b_1 = g(1)$. A benchmark word set from semantic similarity literature is used with $\Pr(W = w) = \frac{1}{|\mathcal{W}|}$. Fig. 4 plots the achievable semantic information-energy capacity and its corresponding converse region. A trade-off between the information rate and energy rate is observed and becomes evident as b increases. As shown in Fig. 5, by considering a low semantic ambiguity code, an upper bound for the conventional information-energy capacity region is obtained by the semantic code.

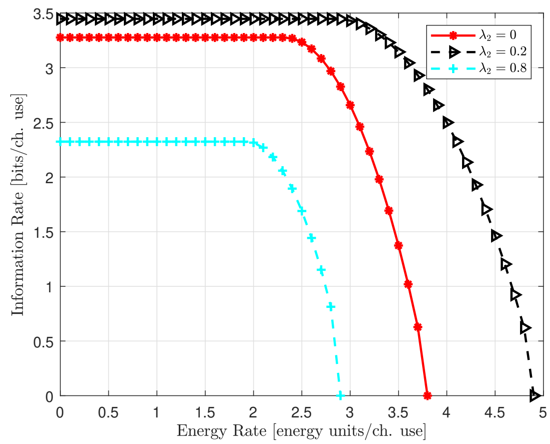

Fig. 6: Impact of semantic context on the information-energy capacity region over the Gaussian channel.

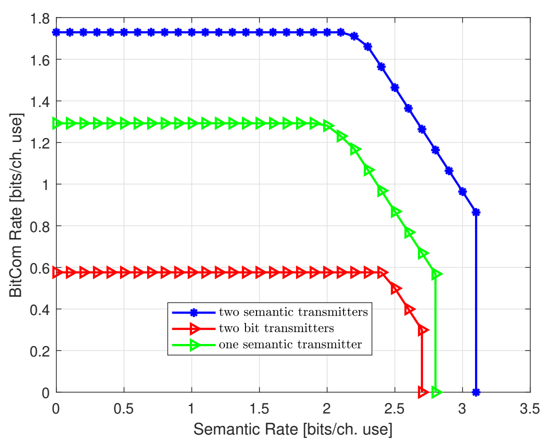

Fig. 7: Impact of the nature of the transmitters on the information-energy capacity region for the Gaussian MAC, $P_1 = P_2 = 0$ dBW (low SNR).

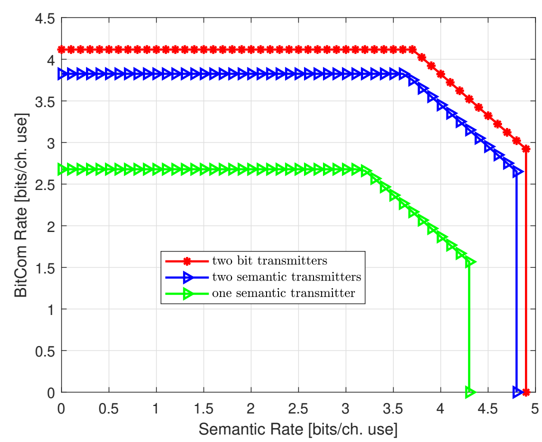

Fig. 8: Impact of the nature of the transmitters on the information-energy capacity region for the Gaussian MAC, $P_1 = P_2 = 40$ dBW (high SNR).

Fig. 6 shows the impact of the semantic context on the information-energy capacity region over the Gaussian channel. By setting $\lambda_2 = 0$ (no semantics), we obtain the same region as [29]. By slightly increasing λ_2, i.e., ($\lambda_2 = 0.2$) we observe an enlargement of the semantic information-energy region, due to the fact that semantic boost the performance of the information transfer task as well as EH task. However, by setting $\lambda_2 = 0.8$ corresponding to a higher ambiguity code, we observe a lower performance in comparison with the conventional Shannon region. Fig. 7 and Fig. 8 gives the achievable information-energy region for the one semantic transmitter case, two semantic transmitters, and two conventional transmitters for the low SNR and the high SNR regime, respectively. Moreover, as seen from Fig. 7 and Fig. 8, the two semantic transmitter MAC outperforms the other channels for the case of low SNR ($P_1 = P_2 = 0$ dB), but performs worse than the conventional scheme for ($P_1 = P_2 = 20$ dB). This is inline with the results presented in [6], where semantic communication and conventional communication are generally preferred to be employed at the high SNR and low SNRs, respectively.

V. Conclusion

In this paper, we have proposed a novel framework for exploring the fundamental limits of semantic information and energy transfer in wireless networks. Our focus has been on point-to-point and multiple access scenarios, where we incorporated the semantic context into the communication process. For the point-to-point case, we have developed an information-theoretic framework to characterize achievable information-energy regions and their converses, considering both the DM and the Gaussian channels. These regions provide insights into the information-energy capacity while taking into account semantics. For the Gaussian MAC case, we introduced a system with a semantic transmitter and a conventional transmitter, operating under EH constraints at the receiver. By analyzing achievable and converse regions for the information-energy trade-off, we showed that power splitting between information and semantic context components plays a pivotal role in achieving these limits. Furthermore, our numerical results indicate that employing low semantic ambiguity codes can lead to improved performance compared to conventional communication approaches devoid of semantics. In Gaussian MAC scenarios, our findings suggest that semantic communication is advantageous at low SNRs, while conventional communication may be preferable at high SNRs.

Appendix A
Proof of Theorem 1

The achievability scheme used to obtain the lower bound on the semantic-energy capacity region relies on the AEP [30] and random coding arguments. By assuming that the semantic context Q is independent from W, the joint probability distribution is simplified as follows

$$\mathrm{Pr}_{W,X,Y,Q}(w,x,y,Q) = \mathrm{Pr}(w)\mathrm{Pr}(q)\mathrm{Pr}(x|w)\mathrm{Pr}(y|x). \tag{22}$$

We generate 2^{nR} n-length i.i.d codewords, according to the distribution

$$\mathrm{Pr}(\boldsymbol{x}) = \prod_{i=1}^{n} \mathrm{Pr}(x_i). \tag{23}$$

According to the AEP, the set of all possible sequences is divided into typical sets, where the sample entropy is close to the entropy of individual variables with high probability, i.e., $\Pr\left(|-\frac{1}{n}\log p(x^n) - H(X)| < \epsilon\right) > 1-\eta$, with $\eta > 0$ and other non-typical sets with low probability. In the following, we discuss the typical sets, and their properties hold with high probability for all sequences. A semantic error appears, if a received message is not decoded by the IR using the context Q. Now, let n be a sufficiently large number. Assume that $Q_1, Q_2, \ldots, Q_n$ is the sequence of the observed context, $X_1, X_2, \ldots, X_n$ is the sequence of the transmitted signals, and $Y_1, X_2, \ldots, Y_n$ is the sequence of the received signals. According to the AEP, there are $2^{nH(Q)}$ typical sequences of context. By using the channel coding theorem, there are $2^{(I(X;Y)-R)n}$ typical input sequences. For a typical sequence X, there are $2^{-nH(X|W)}$ typical sequences of X, given the context. Hence, there are $2^{(I(X;Y)-H(X|W)+H(Q))n}$ typical sequences of input, given the context. Specifically, if

$$R < I(X;Y) - H(X|W) + H(Q), \tag{24}$$

the probability of semantic error $P_{\text{SE}} \to 0$, when $n \to \infty$.

Appendix B
Proof of Theorem 2

Let X, Y, and W be random variables such that $W \to X \to Y$ forms a Markov chain. Then, W and Y are independent given X, i.e,

$$I(Y;W|X) = 0. \tag{25}$$

By using the chain rule for the mutual information, the following holds [30]

$$I(Y;X,W) = I(Y;X) + I(Y;W|X) \tag{26}$$

$$= I(Y;W) + I(Y;X|W). \tag{27}$$

Then, by using (25), we have that

$$I(Y;X) = I(Y;W) + I(Y;X|W). \tag{28}$$

Now, define $f \triangleq I(X;Y) - H(X|W) + H(Q)$. Hence, f could be written as follows

$$f = I(Y;W) + I(Y;X|W) - H(X|W) + H(Q) = I(Y;W) - H(X|W,Y) + H(Q)$$
$$= \sum_{w\in\mathcal{W}}\sum_{y\in\mathcal{Y}} \Pr(w,y)\log\frac{\Pr(w,y)}{\Pr(y)\Pr_W(w)} - H(X|W,Y) + \sum_{W} p_w(W)H(Q|W)$$

$$= \sum_{w \in \mathcal{W}} \sum_{y \in \mathcal{Y}} \Pr(y|w)\Pr(w) \log \frac{\Pr(w,y)}{\Pr(y)\Pr(w)} - H(X|W,Y) + \sum_{W} p(W)H(Q|W), \tag{29}$$

which follows from the definition of $I(X;Y)$ in [30]. By using the data procession inequality [30], i.e, $I(X;W) \geq I(Y;W)$ and (29), the following holds

$$\begin{aligned} f &\leq I(X;W) - H(X|W,Y) + \sum_{W} \Pr(W)H(Q|W) \\ &= H(X) - H(X|W) - H(X|W,Y) + \sum_{W} p(W)H(Q|W) \end{aligned}$$

with equality when $I(X;W|Y) = 0$. Hence, by using Fano's inequality [30], we have

$$R \leq I(Y;W) + I(Y;X|W) - H(X|W) + H(Q), \tag{30}$$

which completes the proof.

APPENDIX C
PROOF OF THEOREM 3

We use the same approach as in the proof of the achievable information-energy region of the DM channels, i.e., random coding and joint typicality decoding. However, in this case, we must take into account the power constraint, the non-linear EH constraint as well as the fact that the variable are continuous and not discrete.

Codebook Generation: We wish to generate a codebook in which all the codewords satisfy the power constraint as well as the EH constraint. To ensure this, the elements of the codeword are chosen to be independent and identically distributed, with variance $\lambda_1 (P_{\text{req}} - \epsilon)$, where λ_1 denotes the fraction of power dedicated to the non-semantic component and satisfies $\mathbb{E}[X_i^2] = b + \epsilon = \lambda_1 P_{\text{req}}$, and $\epsilon > 0$.

Encoding: After the generation of the codebook, the context random variable Q is revealed to both transmitter and the IR, i.e., we assume that the context Q is sent through a feedback link by using a fraction λ_2 of the power P_{req}, where λ_2 denotes the fraction of the power dedicated to the semantic component.

Decoding: The IR observes the codeword list and the sequence of the context $Q_1, Q_2, \ldots, Q_n$ generated via a genie-aided link by the transmitter, and decides on w if

- $\{X^n(w), Y\}$ are jointly typical, $\{Q^n(w)\}$ is a typical sequence, and
- $\{X^n(w)|W\}$ is a typical sequence given the context.

By using the AEP, the semantic error probability is upper bounded as follows

$$P_{\text{SE}} \leq 2^{n(-I(X:Y)+H(X|Q)-H(Q)-3\epsilon)}, \tag{31}$$

which completes the proof.

APPENDIX D
PROOF OF THEOREM 4

Assuming that the message indices $W \in \{1, 2, \ldots, 2^{nR}\}$ follow a uniform distribution, the following holds

$$\begin{aligned} nR = H(W) &= I(W; \hat{W}) + H(W|\hat{W}) & (32)\\ &\leq I(W; \hat{W}|Q) + n\epsilon & (33)\\ &\leq I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{Q}) + h(\boldsymbol{Q}) + n\epsilon & (34)\\ &= h(\boldsymbol{Y}|\boldsymbol{Q}) - h(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Q}) + n\epsilon & (35)\\ &= h(\boldsymbol{Y}|\boldsymbol{Q}) - h(\boldsymbol{Z}) + n\epsilon & (36)\\ &\leq \sum_{i=1}^{n} h(Y_i|Q_i) - h(\boldsymbol{Z}) + n\epsilon, & (37) \end{aligned}$$

where (33) follows from Fano's inequality, i.e.,

$$H(W|\hat{W}) \leq 1 + nRP_{\text{SE}} = n\epsilon, \tag{38}$$

where $\epsilon \rightarrow 0$ as $P_{SE} \rightarrow 0$; (35) follows from the side information Q provided via a genie-aided channel to the IR that leads to the enhancement of the capacity. Then, we use a power splitting technique between the information component and the semantic context Q, i.e.,

$$Y = \mu_1 X + \mu_2 Q + Z, \tag{39}$$

where $\mu_1 + \mu_2 = 1$. Since $f : x \rightarrow \frac{1}{2}\log(1 + x)$ is a concave function, by applying Jensen's inequality, we obtain

$$R \leq \frac{1}{2}\log(1 + \mu_1 P) + \frac{1}{2}\log(1 + \mu_2 P), \tag{40}$$

which completes the proof.

APPENDIX E
PROOF OF THEOREM 5

The proof of achievability uses a very simple power-splitting techniques between information component and the semantic context Q for the first transmitter.

Semantic source: Given a source message $W_1 \in \left[1 : 2^{nR_1}\right]$. Let S_1 represent the inherent semantic information of W_1. Let $f_s : \left[1 : 2^{nR_1}\right] \to \left[1 : 2^{\alpha nR_1}\right]$ represent the semantic mapping from W_1 to S_1 according to the realization of the random variable Q, with α denotes the compression factor induced by the semantic encoder.

Random codebook generation: By applying random coding, we randomly and independently generate 2^{nR_s} sequences $x_1^n(w_1)$, each according to $\Pr(x_1^n) = \prod_{i=1}^n \Pr(x_{1i})$. The generated sequences constitute the codebook $\mathcal{C}_1 = \left[1 : 2^{nR_s}\right]$ as follows

$$\Pr(\mathcal{C}_1) = \prod_{m=1}^{2^{nR_s}} \prod_{i=1}^{n} \Pr\left(x_{1i}(m)\right).$$

The codebook $\mathcal{C}$ is known by both the semantic encoder and the decoder. In order to represent $2^{\alpha nR}$ semantic information losslessly, R_s must satisfy

$$R_s \geq \alpha R_1.$$

Similarly, for the conventional encoder, by applying random coding, we randomly and independently generate 2^{nR_2} sequences $x_2^n(w_2)$, each according to $\Pr(x_2^n) = \prod_{i=1}^n \Pr(x_{2i})$. The generated sequences constitute the codebook $\mathcal{C}_2 = \left[1 : 2^{nR_2}\right]$ as follows

$$\Pr(\mathcal{C}_2) = \prod_{m=1}^{2^{nR_2}} \prod_{i=1}^{n} \Pr\left(x_{2i}(m)\right).$$

Semantic encoding: Given a source message W_1, the encoder finds the corresponding semantic information set index m_1, i.e., to send semantic index $m_1 \in \left[1 : 2^{nR_s}\right]$, and transmits $x_1^n(m_1, Q)$ given the context Q.

Conventional encoding: Given a source message W_2, the encoder finds the corresponding semantic information set index m_2, i.e., to send semantic index $m_2 \in \left[1 : 2^{nR_2}\right]$, and transmits $x_2^n(m_2)$.

Decoding: Let $\mathcal{A}_\epsilon^{(n)}$ denote the set of typical $(x_1^n, x_2^n, \boldsymbol{y})$ sequences. The IR declares that the pair (m_1, m_2) is sent if

$$(x_1^n(m_1), x_2^n(m_2), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}. \tag{41}$$

Otherwise, if there is none or more than one such message, it declares an error E.

Analysis for the probability of error: By the symmetry of the random code construction, the conditional probability of error does not depend on which pair of indices is sent. Thus, the conditional probability of error is the same as the unconditional probability of error. So, without loss of generality, we assume that $(m_1, m_2) = (1, 1)$ was sent. Let E_q be the event where there is no typical sequences of the observed context and E_1 be the event where there is no typical sequences of X_1 given the observed context. Moreover, we denote by E_{ij} the event where x_1^n and x_2^n are not typical given the output $\boldsymbol{y}$, i.e.,

$$E_{ij} = \{(x_1^n(i), x_2^n(j), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}. \tag{42}$$

Thus, by the union of events bound, the probability of error denoted by $\Pr(E)$ is given by

$$\Pr(E) = \Pr\left(E_q \cup E_1 \cup E_{11}^c \cup \{E_{ij}\}_{(i,j)\neq(1,1)}\right) \tag{43}$$

$$\leq \Pr(E_q) + \Pr(E_1) + \sum_{i\neq 1, j=1} \Pr(E_{i1}) + \sum_{i=1, j\neq 1} \Pr(E_{1j}) + \sum_{i\neq 1, j\neq 1} \Pr(E_{ij}), \tag{44}$$

where $\Pr(E)$ is the conditional probability of error given that $(1, 1)$ was sent. From the AEP, we have $\Pr(E_{11}^c) \to 0$,

$$\Pr(E_q) \leq 2^{nR_s} 2^{-n(H(Q)-\epsilon)}, \tag{45}$$

and

$$\Pr(E_1) \leq 2^{nR_s} 2^{-n(H(X_1|W_1)-\epsilon)}, \tag{46}$$

and for $i \neq 1$, we have

$$\Pr(E_{i1}) = \Pr\left((x_1^n(i), x_2^n(1), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}\right) \tag{47}$$

$$= \sum_{(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}} p(\boldsymbol{x}_1) p(\boldsymbol{x}_2, \boldsymbol{y}) \tag{48}$$

$$\leq |\mathcal{A}_\epsilon^{(n)}| 2^{-n(H(X_1)-\epsilon)} 2^{-n(H(X_2,Y)-\epsilon)} \tag{49}$$

$$\leq 2^{-n(H(X_1)+H(X_2,Y)-H(X_1,X_2,Y)-3\epsilon)} \tag{50}$$

$$= 2^{-n(I(X_1;X_2,Y)-3\epsilon)} \tag{51}$$

$$= 2^{-n(I(X_1;Y|X_2)-3\epsilon)}, \tag{52}$$

where (52) follows from

$$I(X_1; X_2, Y) = I(X_1; X_2) + I(X_1; Y|X_2) = I(X_1; Y|X_2). \tag{53}$$

Similarly, for $j \neq 1$,

$$\Pr(E_{1j}) = \Pr\left((x_1^n(1), x_2^n(j), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)} \} \right) \tag{54}$$

$$\leq 2^{-n(I(X_2;Y|X_1)-3\epsilon)}, \tag{55}$$

and for $i \neq 1$, $j \neq 1$,

$$\Pr(E_{ij}) \leq 2^{-n(I(X_1,X_2;Y)-4\epsilon)}. \tag{56}$$

It follows that

$$\Pr(E) \leq \Pr(E_{11}^c) + 2^{nR_s} 2^{-n(I(X_1;Y|X_2)+H(Q)-H(X_1|W)-3\epsilon)} + 2^{nR_2} 2^{-n(I(X_2;Y|X_1)-3\epsilon)}$$
$$+ 2^{n(R_1+R_2)} 2^{-n(I(X_1,X_2;Y)-4\epsilon)} \tag{57}$$

Since $\epsilon > 0$ is arbitrary, the condition of the achievability implies that each term tends to 0 as $n \to \infty$. Hence, there exists at least one code with arbitrarily small probability of error. Furthermore, by combining $R_s \geq \alpha R$, the following holds

$$R_1 \leqslant \frac{1}{\alpha} I(X_1; Y|X_2, Q), \tag{58}$$

$$R_2 \leqslant I(X_2; Y|X_1), \tag{59}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y, Q), \tag{60}$$

$$b \leqslant \mathcal{E}(Y_2). \tag{61}$$

We consider that the joint distribution follows a Gaussian input distribution

$$Q \sim \mathcal{N}(0, \lambda_1), \;\; X_{11} \sim \mathcal{N}(0, \lambda_2), \;\text{ and }\; X_2 \sim \mathcal{N}(0, \sqrt{P}). \tag{62}$$

The input symbol is generated given the mutual independent random variables Q, X_{11}, X_2, where the input symbol at the semantic transmitter satisfies

$$X_1 = \sqrt{P_1} Q + \sqrt{P_1} X_{11}. \tag{63}$$

The choice of the Gaussian input distribution (62) yields

$$I(X_1; Y|X_2) = \log(1 + \lambda_1 P_1), \tag{64}$$

$$I(X_2; Y|X_1) = \frac{1}{2} \log(1 + P_2), \tag{65}$$

$$I(X_1, X_2; Y) = \frac{1}{2} \log\big(1 + \lambda_1 P_1 + P_2\big). \tag{66}$$

Finally, using (64), (65), and (66) into (61), yields the following result

$$\begin{aligned} R_1 &\leqslant \frac{1}{\alpha}\Big(\frac{1}{2}\log(1+\lambda_1 P_1) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1)\Big), \\ R_2 &\leqslant \frac{1}{2}\log(1+P_2), \\ R_1 + R_2 &\leqslant \frac{1}{2}\log\big(1+\lambda_1 P_1 + P_2\big) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1), \\ b &\leqslant \mathcal{E}(Y_2), \end{aligned}$$

with $(\lambda_1, \lambda_2) \in [0,1]^2$, which completes the proof.

Appendix F
Proof of Theorem 6

First, we fix a semantic information-energy achievable pair rates $(R_1(b), R_2(b))$ with a given coding scheme (see Definition 3). Denote by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ the channel inputs resulting from transmitting the independent messages W_1 and W_2 using such coding scheme, we can now bound the rates R_1, R_2 and the sum rate $R_1 + R_2$ as follows

$$\begin{aligned} nR_1 &= H(W_1) && (67)\\ &= I(\boldsymbol{X}_1; \boldsymbol{Y}) + H(W_1|\boldsymbol{Y}) && (68)\\ &\leq I(\boldsymbol{X}_1; \boldsymbol{Y}) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (69)\\ &= H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (70)\\ &\leq H(\boldsymbol{X}_1|\boldsymbol{X}_2) - H(\boldsymbol{X}_1|\boldsymbol{Y}, \boldsymbol{X}_2) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (71)\\ &= I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{X}_2) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (72)\\ &= H(\boldsymbol{Y}|\boldsymbol{X}_2) - H(\boldsymbol{Y}|\boldsymbol{X}_1, \boldsymbol{X}_2) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (73)\\ &= H(\boldsymbol{Y}|\boldsymbol{X}_2) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, \boldsymbol{X}_1, \boldsymbol{X}_2) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (74)\\ &= H(\boldsymbol{Y}|\boldsymbol{X}_2) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (75)\\ &= \sum_{i=1}^{n} H(Y_i|\boldsymbol{X}_2) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (76)\\ &\leq \sum_{i=1}^{n} H(Y_i|\boldsymbol{X}_2) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + + \Big(1-\alpha+(\gamma+\alpha-1)P_{\text{e,s}}^{(n)}\Big)\, nR_1 && (77) \end{aligned}$$

$$= \sum_{i=1} I(X_{1i}; Y_i|X_{2i}) + \left(1 - \alpha + (\gamma + \alpha - 1)P_{\text{e,s}}^{(n)}\right) nR_1, \tag{78}$$

where (69) follows from Fano's inequality [27]; (71) follows from the data-processing inequality; (75) follows from the chain rule; (76) follows from the fact that Y_i depends only on X_{1i} and X_{2i} by the memoryless property of the channel; (77) follows from the chain rule and removing conditioning. Hence, we have

$$R_1 \leq \frac{1}{\left(\alpha - (\gamma + \alpha - 1)P_{\text{e,s}}^{(n)}\right) n} + \frac{\frac{1}{n}\sum_{i=1} I(X_{1i}; Y_i|X_{2i})}{\alpha - (\gamma + \alpha - 1)P_{\text{e,s}}^{(n)}}. \tag{79}$$

Following similar steps, we obtain

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}) + \epsilon. \tag{80}$$

Finally, we bound the sum rates as follows:

$$\begin{aligned} n(R_1 + R_2) &= H(W_1; W_2) && (81)\\ &= I(W_1, W_2; \boldsymbol{Y}) + H(W_1, W_2|\boldsymbol{Y}) && (82)\\ &\leq I(W_1, W_2; \boldsymbol{Y}) + n\epsilon && (83)\\ &\leq I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}) + n\epsilon && (84)\\ &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}_1, \boldsymbol{X}_2) + n\epsilon && (85)\\ &= H(\boldsymbol{Y}) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, \boldsymbol{X}_1, \boldsymbol{X}_2) + n\epsilon && (86)\\ &= H(\boldsymbol{Y}) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon && (87)\\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_{2i}.X_{2i}) + n\epsilon && (88)\\ &= \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) + n\epsilon, && (89) \end{aligned}$$

where (83) follows from Fano's inequality; (84) follows from data-processing inequality; (85) follows from the chain rule; (86) follows from the chain rule; (87) follows from the fact that Y_i depends only on X_{1i} and X_{2i}; (88) follows from the chain rule. Hence, we have

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) + \epsilon, \tag{90}$$

which completes the proof.

APPENDIX G
PROOF OF THEOREM 7

The proof of achievability uses a very simple power-splitting techniques between information component and the semantic context Q for the first transmitter.

Semantic source: For the first transmitter, given a source message $W_1 \in \left[1 : 2^{nR_1}\right]$. Let S_1 represent the inherent semantic information of W_1. Let $f_{s1} : \left[1 : 2^{nR_1}\right] \to \left[1 : 2^{\alpha n R_1}\right]$ represent the semantic mapping from W_1 to S_1 according to the realization of the random variable Q. For the second transmitter, given a source message $W_2 \in \left[1 : 2^{nR_2}\right]$. Let S_2 represent the inherent semantic information of W_2. Let $f_{s2} : \left[1 : 2^{nR_2}\right] \to \left[1 : 2^{\alpha_1 n R_2}\right]$ represent the semantic mapping from W_2 to S_2 according to the realization of the random variable Q.

Random codebook generation: By applying random coding, we randomly and independently generate 2^{nR_s} sequences $x_1^n(w_1)$, each according to $\Pr(x_2^n) = \prod_{i=1}^n \Pr(x_{1i}|q_i)$. The generated sequences constitute the codebook $\mathcal{C}_1 = \left[1 : 2^{nR_{s1}}\right]$ as follows

$$\Pr(\mathcal{C}_1) = \prod_{m=1}^{2^{nR_{s1}}} \prod_{i=1}^{n} \Pr\left(x_{1i}(m)|q_i\right).$$

The codebook $\mathcal{C}$ is known by both the semantic encoder and the decoder. In order to represent $2^{\alpha n R}$ semantic information losslessly, R_{s1} must satisfy

$$R_{s1} \geq \alpha_1 R_1.$$

Similarly, for the conventional encoder, by applying random coding, we randomly and independently generate 2^{nR_2} sequences $x_2^n(w_2, Q)$, each according to $\Pr\left(x_2^n|Q\right) = \prod_{i=1}^n \Pr\left(x_{2i}|q_i\right)$. The generated sequences constitute the codebook $\mathcal{C}_2 = \left[1 : 2^{nR_{s2}}\right]$ as follows

$$\Pr(\mathcal{C}_2) = \prod_{m=1}^{2^{nR_{s2}}} \prod_{i=1}^{n} \Pr\left(x_{2i}|q_i(m)\right).$$

Semantic encoding: Given a source message $W_i,\ i \in \{1, 2\}$, transmitter i finds the corresponding semantic information set index m_i, i.e., to send semantic index $m_i \in \left[1 : 2^{nR_{si}}\right]$, and transmits $x_i^n(m_i, Q)$ given the context Q.

Decoding: Let $\mathcal{A}_\epsilon^{(n)}$ denote the set of typical $(Q, x_1^n, x_2^n, \boldsymbol{y})$ sequences. The IR declares that the pair (m_1, m_2) is sent if

$$(Q, x_1^n(m_1), x_2^n(m_2), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}. \tag{91}$$

Otherwise, if there is none or more than one such message, it declares an error $\mathcal{E}$.

Analysis of the probability of error: By the symmetry of the random code construction, the conditional probability of error does not depend on which pair of indices is sent. Thus the conditional probability of error is the same as the unconditional probability of error. So, without loss of generality, we assume that $(m_1, m_2) = (1, 1)$ was sent. Let E_q be the event that there is no typical sequences of the observed context, E_1 be the event that there is no typical sequences of X_1 given the observed context and E_2 be the event that there is no typical sequences of X_2 given the observed context Q_2.

$$E_{ij} = \{(\boldsymbol{q}, x_1^n(i), x_2^n(j), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}. \tag{92}$$

Thus, by the union of events bound

$$\Pr(E) = \Pr\left(E_q \cup E_1 \cup E_{11}^c \cup \{E_{ij}\}_{(i,j)\neq(1,1)}\right) \tag{93}$$

$$\leq \Pr(E_q) + \Pr(E_1) + \sum_{i\neq 1, j=1} \Pr(E_{i1}) + \sum_{i=1, j\neq 1} \Pr(E_{1j}) + \sum_{i\neq 1, j\neq 1} \Pr(E_{ij}), \tag{94}$$

where $\Pr(\cdot)$ is the conditional probability given that $(1, 1)$ was sent. From the AEP, we have: $\Pr(E_{11}^c) \to 0$;

$$\Pr(E_q) \leq 2^{nR_s} 2^{-n(H(Q)-\epsilon)}, \tag{95}$$

and

$$\Pr(E_1) \leq 2^{nR_s 1} 2^{-n(H(X_1|W_1)-\epsilon)}, \tag{96}$$

and

$$\Pr(E_2) \leq 2^{nR_s 2} 2^{-n(H(X_2|W_2)-\epsilon)}, \tag{97}$$

and for $i \neq 1$, we have:

$$\Pr(E_{i1}) = \Pr\left((x_1^n(i), x_2^n(1), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}\right) \tag{98}$$

$$\leq 2^{-n(I(X_1;Y|X_2)-3\epsilon)}, \tag{99}$$

Similarly, for $j \neq 1$,

$$\Pr(E_{1j}) = \Pr\left((x_1^n(1), x_2^n(j), \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}\right) \tag{100}$$

$$= \sum_{(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in \mathcal{A}_\epsilon^{(n)}\}} p(\boldsymbol{x}_1, \boldsymbol{y}) p(\boldsymbol{x}_2) \tag{101}$$

$$\leq |\mathcal{A}_\epsilon^{(n)}| 2^{-n(H(X_1,Y)-\epsilon)} 2^{-n(H(X_2)-\epsilon)} \tag{102}$$

$$\leq 2^{-n(H(X_2)+H(X_1,Y)-H(X_1,X_2,Y)-3\epsilon)} \tag{103}$$

$$= 2^{-n(I(X_2;X_1,Y)-3\epsilon)} \tag{104}$$

$$= 2^{-n(I(X_2;Y|X_1)-3\epsilon)}, \tag{105}$$

and for $i \neq 1$, $j \neq 1$,

$$\Pr(E_{ij}) \leq 2^{-n(I(X_1,X_2;Y)-4\epsilon)}. \tag{106}$$

It follows that

$$\Pr(E) \leq \Pr(E_{11}^c) + 2^{nR_s}2^{-n(I(X_1;Y|X_2)+H(Q)-H(X_1|W)-3\epsilon)} + 2^{nR_2}2^{-n(I(X_2;Y|X_1)-3\epsilon)} \tag{107}$$

$$+2^{n(R_1+R_2)}2^{-n(I(X_1,X_2;Y)-4\epsilon)}. \tag{108}$$

Since $\epsilon > 0$ is arbitrary, the condition of the achievability imply that each term tends to 0 as $n \to \infty$. Hence, there exist at least one code with arbitrarily small probability of error. Furthermore, by combining $R_s \geq \alpha R$, the following holds

$$R_1 \leqslant \frac{1}{\alpha}\left(I(X_1;Y|X_2,Q) + H(Q) - H(X_1|W_2)\right), \tag{109}$$

$$R_2 \leqslant \frac{1}{\alpha}\left(I(X_2;Y|X_1,Q) + H(Q) - H(X_2|W_2)\right), \tag{110}$$

$$R_1 + R_2 \leq \frac{1}{\alpha} I(X_1,X_2;Y,Q), \tag{111}$$

$$b \leqslant \mathcal{E}(Y_2). \tag{112}$$

The proof of Theorem 7 continues as follows. We consider that the joint distribution follows a Gaussian input distribution

$$Q \sim \mathcal{N}(0,1), \;\; X_{11} \sim \mathcal{N}(0,\lambda_1), \;\; \text{and} \;\; X_{22} \sim \mathcal{N}(0,\lambda_2). \tag{113}$$

The input symbol is generated given the mutual independent random variables Q, X_{11}, X_{22}, where the input symbol at both transmitters satisfies

$$X_1 = \sqrt{P_1(1-\lambda_1}Q + \sqrt{P_1}X_{11}, \tag{114}$$

$$X_2 = \sqrt{P_2(1-\lambda_2}Q + \sqrt{P_2}X_{22}. \tag{115}$$

The choice of the Gaussian input distribution, yields

$$I(X_1;Y|X_2) = \log(1+\lambda_1 P_1). \tag{116}$$

$$I(X_2;Y|X_1) = \frac{1}{2}\log(1+\lambda_2 P_2). \tag{117}$$

$$I(Q,X_1,X_2;Y) = \frac{1}{2}\log\left(1+\lambda_1 P_1 + \lambda_2 P_2\right). \tag{118}$$

Finally, using (116), (117), and (118) into (112), yields the following semantic information-energy region

$$\begin{aligned} R_1 &\leqslant \frac{1}{\alpha}\Big(\frac{1}{2}\log(1+\lambda_1 P_1) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1)\Big), \\ R_2 &\leqslant \frac{1}{\alpha}\Big(\frac{1}{2}\log(1+\lambda_2 P_2) - h(X_2|W_2) + \frac{1}{2}\log(1+\lambda_1 P_2)\Big), \\ R_1 + R_2 &\leqslant \frac{1}{2}\log\big(1+\lambda_1 P_1 + P_2\big) - h(X_1|W_1) + \frac{1}{2}\log(1+\lambda_2 P_1), \\ b &\leqslant \mathcal{E}(Y_2), \end{aligned}$$

with $(\lambda_1, \lambda_2, \alpha) \in [0,1]^3$, which completes the proof.

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[2] W. Weaver, "Recent contributions to the mathematical theory of communication," *The Mathematical Theory of Communication*, vol. 10, 2019. [Online]. Available: https://www.jstor.org/table/42581364

[3] W. Tong and G. Y. Li, "Nine challenges in artificial intelligence and wireless communications for 6G," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 140–145, Aug. 2022.

[4] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Selec. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.

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

[6] X. Mu and Y. Liu, "Exploiting semantic communication for non-orthogonal multiple access," 2022.

[7] B. Guler, A. Yener, and A. Swami, "The semantic communication game," *IEEE Trans. Cognitive Commun. Netw.*, vol. 4, no. 4, pp. 787–802, Dec. 2018.

[8] S. Guo, Y. Wang, and P. Zhang, "Signal shaping for semantic communication systems with a few message candidates," Aug. 2022. [Online]. Available: https://arxiv.org/abs/2202.02072

[9] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Selec. Areas Commun.*, vol. 40, no. 9, pp. 2584–2597, Jul. 2022.

[10] M. Chen, Y. Wang, and H. V. Poor, "Performance optimization for wireless semantic communications over energy harvesting networks," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Proc. (ICASSP)*, May. 2022, pp. 8647–8651.

[11] M. D. P. Basu, J. Bao and J. Hendler, "Preserving quality of information by using semantic relationships," *Pervasive Mobile Comput.*, vol. 11, pp. 188–202, Mar. 2014.

[12] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: opportunities and challenges," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 117–125, Apr. 2015.

[13] N. Tesla, "Apparatus for transmitting electrical energy," New York, NY, Dec. 1914.

[14] I. Krikidis, S. Timotheou, S. Nikolaou, G. Zheng, D. W. K. Ng, and R. Schober, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 104–110, Nov. 2014.

[15] L. R. Varshney, "Transporting information and energy simultaneously," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, Canada, 2008, pp. 1612–1616.

[16] P. Grover and A. Sahai, "Shannon meets tesla: Wireless information and power transfer," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, Texas, June. 2010, pp. 2363–2367.

[17] S. Belhadj Amor, S. M. Perlaza, I. Krikidis, and H. V. Poor, "Feedback enhances simultaneous wireless information and energy transmission in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5244–5265, Aug. 2017.

[18] N. Khalfet and S. M. Perlaza, "Simultaneous Information and Energy Transmission in the Two-User Gaussian Interference Channel," *IEEE J. Selec. Areas Commun.*, vol. 37, no. 1, pp. 156–170, Jan. 2019.

[19] J. Park and B. Clerckx, "Joint wireless information and energy transfer in a two-user MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4210–4221, Aug. 2013.

[20] T. Tu Lam, M. Di Renzo, and J. P. Coon, "System-level analysis of SWIPT mimo cellular networks," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2011–2014, 2016.

[21] S. M. Belhadj Amor and S. M. Perlaza, "Fundamental limits of simultaneous energy and information transmission," in *Proc.* 23*rd Int. Symp. Telecom.*, Thessaloniki, Greece, May 2016.

[22] S. N. G. Z. D. W. K. N. I. Krikidis, S. Timotheou and R. Schober, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 104–110, Nov. 2014.

[23] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 841–850, 1982.

[24] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[25] M. E. Peters, "Deep contextualized word representations," in *Proc. North Amer. Chapter. Assoc. Comput. Linguistics: Hum. Lang. Tech.*, New Orleans, Louisiana, Jun. 2018, pp. 2227–2237.

[26] S. Pejoski, Z. Hadzi-Velkov, and R. Schober, "Optimal power and time allocation for WPCNs with piece-wise linear EH model," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 364–367, 2018.

[27] S. Ma, Y. Wu, H. Qi, H. Li, G. Shi, Y. Liang, and N. Al-Dhahir, "A theory for semantic communications," 2023.

[28] N. Khalfet, S. M. Perlaza, A. Tajer, and H. V. Poor, "On ultra-reliable and low latency simultaneous information and energy transmission systems," in *Proc. IEEE 20th Int. Work. Signal Proc. Adv. Wireless Commun. (SPAWC)*, Cannes, France, Jul. 2019, pp. 1–5.

[29] R. Morsi, V. Jamali, A. Hagelauer, D. W. K. Ng, and R. Schober, "Conditional capacity and transmit signal design for swipt systems with multiple nonlinear energy harvesting receivers," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 582–601, 2020.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley-Interscience Publication, 2005.