Semantic Joint Source Channel Coding for Distributed Subsurface Imaging in Multi-Agent Systems

Maximilian H. V. Tillmann, *Graduate Student Member, IEEE,* Ban-Sok Shin, *Member, IEEE,* Dmitriy Shutin, *Senior Member, IEEE,* Armin Dekorsy, *Senior Member, IEEE*

***Abstract*—Multi-agent systems (MAS) are a promising solution for autonomous exploration tasks in hazardous or remote environments, such as planetary surveys. In such settings, communication among agents is essential to ensure collaborative task execution, yet conventional approaches treat exploration and communication as decoupled subsystems. This work presents a novel framework that tightly integrates semantic communication into the MAS exploration process, adapting communication strategies to the exploration methodology to improve overall task performance. Specifically, we investigate the application of semantic joint source-channel coding (JSCC) with over-the-air computation (AirComp) for distributed function computation for the application of cooperative subsurface imaging using the adapt-then-combine full waveform inversion (ATC-FWI) algorithm. Our results demonstrate that semantic JSCC significantly outperforms classical point-to-point and standard JSCC methods, especially in high-connectivity networks. Furthermore, incorporating side information at the receiving agent enhances communication efficiency and imaging accuracy, a feature previously unexplored in MAS-based exploration. We validate our approach through a use case inspired by subsurface anomaly detection, showing measurable improvements in imaging performance per agent. This work underscores the potential of semantic communication in distributed multi-agent exploration, offering a communication-aware exploration paradigm that achieves task-relevant performance gains.**

***Index Terms*—Semantic communication, multi-agent systems, autonomous exploration, subsurface imaging**

I. Introduction

A. Motivation

MULTI-AGENT systems have emerged as a crucial tool to address various tasks that cannot be handled reliably by humans such as inspection, surveying, monitoring, and exploration of hazardous or remote environments [1], [2]. For instance, multi-agent systems (MASs) have been proposed and investigated for the use in planetary exploration, gas and subsurface exploration [3]–[5]. A MAS inherently requires communication among the agents in its network to achieve the respective exploration goal such as reconstruction of a physical process. Usually, this is realized by a wireless communication system. Most works in the realm of multi-agent exploration therefore use standard communication schemes to enable reliable communication among the agents, cf. [6], [7]. Hence, exploration and communication are viewed as two separate systems that are designed separately. However, exploration and communication are often strongly entangled. Therefore, to achieve enhanced exploration performance both communication and exploration need to be considered jointly. In this work, we present an approach where the communication system is adapted to the requirements of the exploration methodology of a MAS such that the overall system's exploration performance is enhanced.

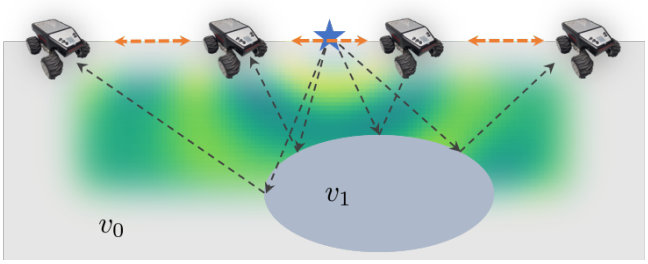

Fig. 1: Illustration of the exploration scenario: A MAS consisting of ground rovers reconstructs the velocity structure (v_0, v_1) of a subsurface using a seismic source (blue) and seismic waves reflecting from an anomaly. The rovers are connected over wireless links (orange) that enable data exchange within the MAS. The communication among rovers is realized via semantic communication to adapt the communication system to the exploration algorithm.

In this respect, semantic communication is a promising technology that enables task-specific communication beyond the Shannon limit of classical communication systems [8]–[11]. Compared to classical communication systems with the goal of accurately transmitting a message, semantic communication aims to transmit only the required "meaning" of the message for accurate task execution to save bandwidth [8]. Most research on semantic communication focuses on either task execution or data reconstruction [8]. In this work, we focus on semantic communication for task execution. Recently, modern machine learning techniques have been used to develop such semantic communication systems [8]. Such an approach becomes especially advantageous in distributed exploration scenarios, where a task needs to be solved by multiple agents. In this case, semantic communication makes

This work was supported by the German Research Foundation (DFG) under grant 500260669 (SCIL).

Maximilian H. V. Tillmann and Armin Dekorsy are with the Department of Communications Engineering, University of Bremen, 28359 Bremen, Germany (e-mail: tillmann@ant.uni-bremen.de; dekorsy@ant.uni-bremen.de).

Ban-Sok Shin and Dmitriy Shutin are with the Institute of Communications and Navigation, German Aerospace Center, 82234 Wessling, Germany (e-mail: ban-sok.shin@dlr.de; dmitriy.shutin@dlr.de).

it possible to only exchange information relevant for the exploration task between the agents instead of transmitting the complete data.

One envisioned application scenario we will address is the autonomous detection and quantification of lunar caves and lava tubes in future exploration missions using seismic techniques [5], [12], [13]. Seismic methods typically require spatially distributed, simultaneous measurements. A MAS with agents carrying seismic sensors thus perfectly exemplifies the use case of interest. Lunar caves and lava tubes are relevant for future space missions: Since they provide constant temperature and natural protection from radiation, they are promising sites to be used as habitats for space equipment or even astronauts. In this regard, the Adapt-Then-Combine Full Waveform Inversion (ATC-FWI) has been proposed that enables subsurface imaging in a distributed fashion within a MAS [5]. In essence, ATC-FWI obtains subsurface images at each agent via data exchange among neighboring agents using distributed data processing. One fundamental operation in such distributed processing is averaging of exchanged data which can be found in e.g. consensus averaging and diffusion-based schemes [14], [15]. From a semantic communication perspective the distributed averaging process of several input data at the receiver belongs to the problem of *distributed function computation* [16], [17]. It is a this point where we apply semantic communication techniques to appropriately encode the data of multiple transmitters (neighboring agents in the MAS) to perform Distributed Function Computation (DFC) for the averaging step at the decoder (receiving agent).

We show that by integrating semantic communication into the ATC-FWI, the imaging performance per agent is enhanced compared to a system that does not utilize semantic communication. In our work, we design the communication system in such a way that it enhances performance of communication-dependent processing steps in the exploration task and by that the overall exploration performance.

B. Related Works in Full Waveform Inversion

Full Waveform Inversion (FWI) is a geophysical imaging technique that enables reconstruction of subsurface structures by exploiting wave physics [18], [19]. Beside geophysics it is also applied to various other fields such as oil exploration, medical ultrasound imaging and non-destructive testing. A significant amount of research has been conducted to improve the imaging capabilities of FWI (see e.g. [20], [21] and references therein). Recently, many approaches have been proposed that use deep learning techniques to reduce computational complexity of FWI by using surrogate models for the solution of partial differential equations [22]–[25]. In its traditional form, FWI is a centralized scheme that requires all measurement data to be allocated at a central entity that performs imaging. Hence, for application in a MAS it needs to be adapted to enable imaging at each agent via exchange of data within the network. In our previous work, we proposed the ATC-FWI which enables distributed imaging within a MAS [5]. However, ATC-FWI has been considered for perfect communication among the agents, i.e., without considering any communication aspects. Other distributed or decentralized schemes for subsurface imaging have been proposed in the literature, cf. [26], [27]. However, in these works communication aspects are considered separately from the imaging problem. That is, the communication is realized by using standard wireless communication systems and an adapted design of the communication system to the subsurface imaging is not considered.

C. Related Works in Semantic Communication

For semantic communication we want to compress and channel code the data at the transmitter such that we can reconstruct the semantic information at the receiver. To this end, deep learning techniques are used to learn an end-to-end encoder decoder system for Joint Source and Channel Coding (JSCC) [8], [28]. Semantic communication systems become increasingly important in distributed settings with multiple transmitters and one or more receivers [29], quite typical for MAS applications. For example, in [30] and [31] semantic communication was investigated for the case when multiple transmitters with different observations need to cooperatively solve a classification task. In [32], the authors investigate the case where one receiver that obtains two correlated images from two transmitters has to reconstruct both images.

In the case of the ATC-FWI algorithm, high dimensional data in the subsurface image space needs to be exchanged between agents, and at each receiving agent a linear function needs to be computed from the agent's own data and the data received from the neighboring agents. These two different aspects of having side information at the decoder, i.e., the agent's locally available data, and doing DFC were investigated separately in many works. The use of side information at the receiver for communication of high-dimensional data, e.g. images, was investigated in [33] and [34]. In [33], it was shown for image transmission that the compression level for the case when side information is only available at the decoder can nearly achieve the same compression level when side information is available at both the encoder and the decoder.

DFC was studied in [35]. Its variation of DFC with Over-the-Air Computation (AirComp) was investigated in [36], [37], with deep learning based AirComp for example in [38]–[40]. Furthermore, AirComp was used in [41] for distributed convex optimization. AirComp aims at using channel resources more efficiently by letting users transmit at the same time and frequency. In this way, a function can be evaluated directly on the wireless channel as a sum of the transmitted symbols by exploiting the superposition property of electromagnetic waves [42], [43]. It was shown that despite advantages, AirComp mainly had two drawbacks: (i) the limitation to so-called nomographic and (ii) the requirement of accurate synchronization.

Related results from information theory are available for the distributed source coding problem, where channel coding is not considered [44] [45]. For the DFC problem, some results exist for the case of linear functions with two Gaussian sources [46], [47]. From the Slepian–Wolf coding theorem it is known that in the case of distributed discrete noiseless case, the

required coding sum rate to achieve an asymptotically error-free reconstruction at the decoder is the same whether the encoding is done jointly or separately [48]. Additionally, for the discrete noiseless case the required sum rate to achieve asymptotically error-free reconstruction at the decoder is the same whether side information is only available at the decoder or at both the encoder and the decoder [48].

D. Main Contributions

In this work, we propose the use of semantic communication in a distributed subsurface imaging method for a MAS to obtain reliable imaging results for noisy inter-agent communication links. To the best of our knowledge, this is the first work to employ semantic communication for distributed subsurface imaging in a MAS. In particular, we investigate the impact of bandwidth-limited, noisy communication channels between the agents on the ATC-FWI algorithm using semantic communication and compare its performance to state-of-the-art JSCC schemes.

We propose a semantic communication system for DFC. Furthermore, we consider the case when a receiving agent has correlated side information available, since the agents reconstruct the same subsurface from different positions. To take the latter into account we propose a formulation of the semantic DFC problem that includes side information. We formulate this problem as a maximization of the mutual information given the side information, meaning that the received symbols should be maximally informative about the semantic variable given the side information. However, maximizing the mutual information directly is infeasible due to the intractability of the posterior distribution of the decoder [30], [49]. In order to optimize our semantic communication system for DFC with side information, we derive a lower bound of the conditional mutual information, which can then be maximized using end-to-end training. In previous works on semantic communication, there was either no side information available, as in [30], or an evidence lower bound was used when side information was present [33].

In this work, we apply our proposed semantic communication system to an averaging function of the combine step in the ATC-FWI algorithm. Furthermore, we employ AirComp for the semantic DFC to use channel resources more efficiently. To this end, we train the encoders and the decoder for AirComp in an end-to-end manner using Deep Neural Network (DNN)s. This allows us to express the decoder as a nomographic function, which is necessary to enable addition of the encoded symbols over the communication channel.

In simulations, we show that the proposed method can improve bandwidth efficiency for noisy inter-agent connections for diffusion-based estimation algorithms in MAS systems compared to state-of-the-art JSCC methods, where the goal is to reconstruct data from each transmitting agent at the receiving agent. Our proposed approach of semantic DFC outperforms state-of-the-art JSCC methods, especially in scenarios, where many neighboring agents exchange data with each other.

II. Distributed Subsurface Imaging for Multi-Agent Systems

In the following section, we briefly describe the FWI as the classical, centralized imaging scheme from geophysics. Afterwards, we explain how FWI can be adapted to a MAS for distributed imaging, when each agent obtains a global subsurface image locally. This leads to the ATC-FWI algorithm which requires explicit data exchange among neighboring agents and thus, suitable communication.

A. Centralized Full Waveform Inversion

FWI is a geophysical imaging technique based on seismic measurements. Despite being rooted in seismic subsurface reconstruction, FWI is applied to various other modalities such as ultrasound, ground penetrating radar and electrical resistivity tomography. In its ordinary formulation, FWI aims at minimizing a least-squares cost functional of a data residual with respect to a subsurface parameter $m(\boldsymbol{x})$ that varies over space coordinate $\boldsymbol{x}$. The residual is evaluated between observed measurements $d^{\mathrm{obs}}(t)$ and predicted or synthesized measurements $d^{\mathrm{syn}}(t,m)$ over time t. The synthesized measurements $d^{\mathrm{syn}}(t,m)$ are generated by solving the acoustic wave equation and sampling the obtained wavefield $u_s(\boldsymbol{x},t)$ at the receiver positions via a sampling operator $S_{s,r}$ for a receiver r and the respective source s. The fact that the minimization is coupled to the acoustic wave equation can be formulated as a constraint in the resulting optimization problem:

$$\underset{m}{\operatorname{argmin}}\, \mathcal{L}(m) = \frac{1}{2}\sum_{s,r}\int_0^\tau [\underbrace{S_{s,r}u_s(\boldsymbol{x},t)}_{d^{\mathrm{syn}}_{s,r}(t,m)} - d^{\mathrm{obs}}_{s,r}(t)]^2\,\mathrm{d}t \tag{1a}$$

$$\text{s.t.} \quad \frac{1}{m^2(\boldsymbol{x})}\frac{\partial^2 u_s(\boldsymbol{x},t)}{\partial t^2} - \frac{\partial^2 u_s(\boldsymbol{x},t)}{\partial \boldsymbol{x}^2} = f_s(\boldsymbol{x},t). \tag{1b}$$

where $\sum_{s,r} := \sum_{s=1}^{N_{\mathrm{S}}}\sum_{r=1}^{N_{\mathrm{R}}}$ and $N_{\mathrm{S}}, N_{\mathrm{R}}$ is the number of seismic sources and receivers, respectively. In our case, the subsurface parameter $m(\boldsymbol{x})$ is a function over space and represents the spatial distribution of the P-wave velocity. As can be seen, FWI essentially aims at solving an optimization problem that is constrained by a Partial Differential Equation (PDE). In the wave equation (1b), $f_s(\boldsymbol{x},t)$ is the seismic source specific to source s. Solving (1b) for a subsurface model $m(\boldsymbol{x})$ results in the wavefield $u_s(\boldsymbol{x},t)$ which is then sampled at the receiver positions $\{\boldsymbol{x}_r\}_{r=1}^{N_{\mathrm{R}}}$ to give $d^{\mathrm{syn}}_{s,r}(t,m)$.

Problem (1) is nonlinear in $m(\boldsymbol{x})$ and therefore requires gradient-based optimization to converge to a suitable subsurface model. Commonly used methods include gradient-descent, nonlinear conjugate gradient or L-BFGS. To this end, the gradient of $\mathcal{L}(m)$ needs to be calculated. Note that $\mathcal{L}(m)$ is constrained by a PDE which complicates direct calculation of the gradient. Typically, the so-called *adjoint state method* is used to compute the gradient of cost functionals that are constrained by a PDE, cf. [50]. Applying the adjoint state

method to (1a) together with (1b) results in the following gradient computation [5]:

$$\frac{\mathrm{d}\mathcal{L}(m)}{\mathrm{d}m} = -\frac{2}{m(\boldsymbol{x})^3}\sum_{s=1}^{N_\mathrm{S}}\int_0^\tau q_s(\boldsymbol{x},\tau-t)\frac{\partial^2 u_s(\boldsymbol{x},t)}{\partial t^2}\mathrm{d}t. \quad (2)$$

Here, the function $q_s(\cdot,\cdot)$ is a so-called adjoint wavefield, which is obtained by solving the following PDE which is an adjoint wave equation:

$$\left(\frac{1}{m(\boldsymbol{x})^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \boldsymbol{x}^2}\right) q_s(\boldsymbol{x},t) = \sum_{r=1}^{N_\mathrm{R}} S_{s,r}(d_{s,r}^{\mathrm{syn}}(\tau-t) - d_{s,r}^{\mathrm{obs}}(\tau-t)). \quad (3)$$

Note that to solve (3), the data residuals $(d_{s,r}^{\mathrm{syn}}(\tau-t)-d_{s,r}^{\mathrm{obs}}(\tau-t))$ are now used as a source signal that is reversed in time and injected at the receiver positions by a sampling operator $S_{s,r}$. After obtaining the adjoint wavefield, $q_s(\boldsymbol{x},t)$ is again reversed in time and correlated with the second time derivative of the wavefield $u_s(\boldsymbol{x},t)$ following (2).

To solve PDEs in (1b) and (3) we require a numerical solver on a discrete space and time grid. In our work, we use the Devito package that provides highly efficient finite difference solvers for PDEs [51]. By discretizing the considered 2D computational domain into a $N_\mathrm{x} \times N_\mathrm{z}$ grid we obtain vector quantities for the subsurface model $m(\boldsymbol{x})$ and the gradient function $\mathrm{d}\mathcal{L}(m)/\mathrm{d}m$. Therefore, we replace these functions by their discrete vector description, i.e.,

$$m(\boldsymbol{x}) \to \boldsymbol{m}(\boldsymbol{x}) \in \mathbb{R}^{N_\mathrm{x}N_\mathrm{z}}, \quad \mathrm{d}\mathcal{L}(m)/\mathrm{d}m \to \delta\boldsymbol{m} \in \mathbb{R}^{N_\mathrm{x}N_\mathrm{z}}. \quad (4)$$

Given a gradient by (2), we can now minimize $\mathcal{L}(m)$ iteratively using gradient descent. This results in the following update procedure for the estimated subsurface model:

$$\boldsymbol{m}^{[k+1]} = \boldsymbol{m}^{[k]} - \alpha^{[k]}\delta\boldsymbol{m}^{[k]} \quad (5)$$

The parameter $\alpha^{[k]} > 0$ is a step size that can be adapted over the iterations e.g. by a simple exponential decay or a line search method.

B. Adapt-Then-Combine Full Waveform Inversion (ATC-FWI)

FWI in its classical setting is centralized algorithm, i.e., it requires all measurement data $d_{s,r}^{\mathrm{obs}}$ to be available at a central entity to perform the inversion procedure. However, in the MAS no central entity is available that collects all measurement data from the agents. Instead, we demand a computation within the MAS by exploiting agent-to-agent communication.

The objective of a distributed imaging scheme is to obtain an image locally at each agent/receiver via cooperation with neighboring agents, i.e., those that are in the communication range of the respective agent. To this end, the ATC-FWI has been proposed in [5]. In the following, we give a short summary of the algorithm.

Let us first introduce assumptions about the network topology of the MAS. We assume a network of N_R agents, where each agent is equipped with a geophone or a seismic sensor. Each agent r can communicate with neighboring agents ℓ that are contained in a neighborhood set $\mathcal{N}_r \subseteq \{1,\dots,N_\mathrm{R}\}$ and vice versa. We assume that the neighborhood set $\mathcal{N}_r$ contains the agent r itself. Furthermore, the network graph is undirected and connected, i.e., each agent can be reached by any other agent in the network over multiple hops.

The first step in deriving ATC-FWI is to represent the global cost in (1a) over all N_R agents in the network such that

$$\mathcal{L}(m) = \sum_{r=1}^{N_\mathrm{R}} \mathcal{L}_r(m), \quad (6)$$

with each agent's local cost given now as

$$\mathcal{L}_r(m) = \frac{1}{2}\sum_{s=1}^{N_\mathrm{S}}\int_0^\tau \left(d_{s,r}^{\mathrm{syn}}(t,m) - d_{s,r}^{\mathrm{obs}}(t)\right)^2 dt. \quad (7)$$

Note that each local cost $\mathcal{L}_r(m)$ depends only on locally observed measurement data $d_{s,r}^{\mathrm{obs}}(t)$. Based on its local cost each agent r is able to compute a gradient w.r.t. the subsurface model m in the same manner as in centralized FWI, however, with local measurement data $d_{s,r}^{\mathrm{obs}}(t)$. To ensure that global subsurface information is available to all agents in the network, we introduce N_R different (local) subsurface models which are then combined using diffusion strategies [14]. Specifically, we assign individual subsurface models $m_r(\boldsymbol{x})$ to each agent. Then each agent solves the wave equation (1b) and the adjoint equation (3) using its own subsurface model $m_r(\boldsymbol{x})$ to obtain the wavefield $u_{s,r}(\boldsymbol{x},t)$ and the adjoint wavefield $q_{s,r}(\boldsymbol{x},t)$ according to

$$\frac{1}{m_r(\boldsymbol{x})^2}\frac{\partial^2 u_{s,r}(\boldsymbol{x},t)}{\partial t^2} - \frac{\partial^2 u_{s,r}(\boldsymbol{x},t)}{\partial \boldsymbol{x}^2} = f_s(\boldsymbol{x},t) \quad (8)$$

$$\left(\frac{1}{m_r(\boldsymbol{x})^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \boldsymbol{x}^2}\right) q_s(\boldsymbol{x},t) = S_{s,r}^{\mathsf{T}}(d_{s,r}^{\mathrm{syn}}(\tau-t) - d_{s,r}^{\mathrm{obs}}(\tau-t)), \quad (9)$$

followed by a local gradient computation $\delta m_r(\boldsymbol{x})$ via correlation:

$$\delta m_r(\boldsymbol{x}) = -\frac{2}{m(\boldsymbol{x})^3}\sum_{s=1}^{N_\mathrm{S}}\int_0^\tau q_{s,r}(\boldsymbol{x},\tau-t)\frac{\partial^2 u_{s,r}(\boldsymbol{x},t)}{\partial t^2}\mathrm{d}t \quad (10)$$

Now, each agent has an individual subsurface model $m_r(\boldsymbol{x})$ and a local gradient $\delta m_r(\boldsymbol{x})$ available. To enable distributed imaging such that each agent's model approaches the centralized FWI result, we apply the adapt-then-combine (ATC) technique from [14]. Assuming again discretization of the spatial domain using finite differences we represent gradients and subsurface models via vectors. Applying ATC results in the following computations per agent r:

$$\widetilde{\boldsymbol{m}}_r^{[k+1]} = \boldsymbol{m}_r^{[k]} + \alpha^{[k]}\sum_{\ell\in\mathcal{N}_r} b_{\ell r}\delta\boldsymbol{m}_\ell^{[k]} \quad (11a)$$

$$\boldsymbol{m}_r^{[k+1]} = \sum_{\ell\in\mathcal{N}_r} a_{\ell r}\widetilde{\boldsymbol{m}}_\ell^{[k+1]} \quad (11b)$$

The coefficients $a_{\ell r}, b_{\ell r}$ need to be chosen appropriately such that the centralized FWI result can be approached by all agents, cf. [14]. In the simplest case, one can select these coefficients to be $1/|\mathcal{N}_r|$, where $|\cdot|$ denotes cardinality of

the neighborhood set. This results in a uniform averaging of neighboring data.

From (11) we see that in each iteration k each agent r needs to exchange its gradient $\delta\boldsymbol{m}_r$ and its intermediate subsurface model $\widetilde{\boldsymbol{m}}_r$ with its neighbors $\ell \in \mathcal{N}_r$. In particular, the fusion process of neighboring data in ATC is given by an averaging process. It is here, where we will integrate semantic communications to exploit the averaging structure in the data exchange among the agents.

III. Semantic Communication System Model and Problem Formulation

In what follows we now discuss the semantic communication approach towards computation of ATC over the network. In particular, we will outline the learning of corresponding encoder-decoder structures that are specifically designed for computing the averaging operation in (11).

Let us now consider a particular agent r that needs to compute (11). In our context this agent plays a role of the receiver, while its neighbors act as transmitters. As it can be seen in (11), for the ATC-FWI there are two types of quantities that need to be communicated to this agent: gradients $\delta\boldsymbol{m}_\ell$ and model estimates $\widetilde{\boldsymbol{m}}_\ell$, for all agents in the neighborhood $\ell \in \mathcal{N}_r$. Although gradients and model estimates are different quantities, the concept of semantic communication will be applied to each of them in a similar fashion.

Specifically, we will refer to the data available at each agent $\ell \in \mathcal{N}_r$ as the observation $\boldsymbol{s}_\ell$, $\ell \in \mathcal{N}_r$. Depending on the context, $\boldsymbol{s}_\ell$ is either the gradient $\delta\boldsymbol{m}_\ell$, or the model estimate $\widetilde{\boldsymbol{m}}_\ell$. The goal of the semantic receiver is to recover the semantic variable $\boldsymbol{z}$, which is a result of (11a) when gradients are communicated, or (11b) when model estimates are combined. Let us point out that the neighborhood $\mathcal{N}_r$ includes (by definition) the receiving agent r as well. Naturally, this "local" information should not be communicated since the agent is already aware about it. Instead, we will treat this information differently and utilize it to properly "inform" the decoder at the receiver. We will therefore refer to this data as *side information* $\boldsymbol{s}_0$ in the following, using the subscript index 0 to distinguish it from the other semantic sources. Furthermore, we define $N_r \triangleq |\mathcal{N}_r| - 1$ to count the number of neighbors of the receiving agent r, excluding the agent itself.

A. Semantic Communication System Model

Our semantic communication system model is shown in Fig. 2. We define a semantic source as the joint probability distribution $p(\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_\ell}, \boldsymbol{z})$ of the observations $\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_\ell}$ and the semantic variable $\boldsymbol{z}$. The goal at the receiver is to reconstruct the semantic variable $\boldsymbol{z}$, which together with the spatially distributed observations $\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_\ell}$ follows the joint distribution of the semantic source $p(\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_\ell}, \boldsymbol{z})$.

In our notation, the observation $\boldsymbol{s}_0$ is already available at the decoder (e.g., the local gradient or the local subsurface model) and used as side information for the decoding. The encoders $p_{\boldsymbol{\theta}_\ell}(\boldsymbol{c}_\ell|\boldsymbol{s}_\ell)$ are modeled as Neural Networks (NNs) with parameters $\boldsymbol{\theta}_\ell$ to encode the observation data $\boldsymbol{s}_\ell$ into symbols $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{N_\ell}$, which are transmitted over a wireless channel

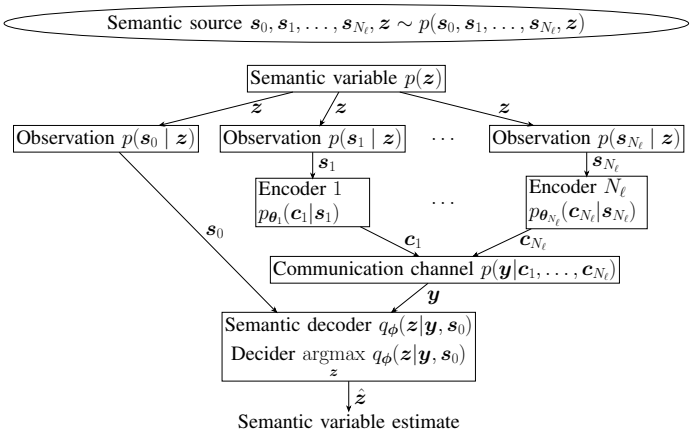

Fig. 2: Block diagram of multi-agent semantic communication system for distributed function computation with side information at the receiving agent r. The receiving agents gets the index $\ell = 0$, and the neighboring agents are indexed from 1 to N_ℓ, with $N_\ell = |\mathcal{N}_r| - 1$. Here, the observations $\boldsymbol{s}_\ell$ are for the two cases the velocity estimates $\widetilde{\boldsymbol{m}}_\ell$ or the gradients $\delta\boldsymbol{m}_\ell$, and the semantic variable $\boldsymbol{z}$ are the sums $\sum_{\ell\in\mathcal{N}_r} a_{\ell r}\widetilde{\boldsymbol{m}}_\ell$ or $\sum_{\ell\in\mathcal{N}_r} b_{\ell r}\delta\boldsymbol{m}_\ell$, respectively.

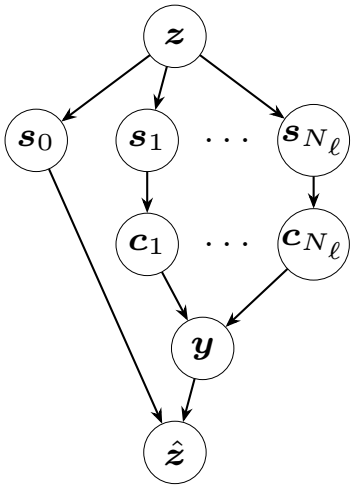

Fig. 3: The statistical dependencies of the semantic variable $\boldsymbol{z}$, the observations $\boldsymbol{s}_\ell$, the transmitted symbols $\boldsymbol{c}_\ell$, the received symbols $\boldsymbol{y}$, and the estimated sematic variable $\hat{\boldsymbol{z}}$.

$p(\boldsymbol{y}|\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{N_\ell})$, where $\boldsymbol{y}$ is a vector containing all received symbols. We use a variational approximation $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)$ with NN parameters $\boldsymbol{\phi}$ to model the semantic decoder, which outputs the distribution of the semantic variable given the received symbols $\boldsymbol{y}$ and the side information $\boldsymbol{s}_0$. Finally, to get an estimate of the semantic variable $\hat{\boldsymbol{z}}$, we select the most likely $\boldsymbol{z}$ from the posterior distribution $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)$. The statistical dependencies of all random variables is summarized in Fig. 3.

B. Semantic Communication Problem Formulation

Our objective is to reconstruct the semantic variable $\boldsymbol{z}$ at the decoder. We formulate the optimization problem of the semantic communication system such that the received symbols $\boldsymbol{y}$ are maximally informative about the semantic variable $\boldsymbol{z}$, given the side information $\boldsymbol{s}_0$. This means that we want to find the parameters $\boldsymbol{\theta}_\ell$ for the encoders $p_{\boldsymbol{\theta}_\ell}(\boldsymbol{c}_\ell|\boldsymbol{s}_\ell)$

that maximize the conditional mutual information $I(\boldsymbol{z};\boldsymbol{y}|\boldsymbol{s}_0)$. To have a realistic communication scenario, we limit the rate of our communication channel by a limited number of channel uses N_{ch} and an average power constraint per channel. This gives us the optimization problem for receiving agent r of

$$\begin{aligned}&\underset{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{N_\ell}}{\operatorname{argmax}}\, I(\boldsymbol{z};\boldsymbol{y}|\boldsymbol{s}_0) \\ &\text{s.t.}\quad \boldsymbol{y}\in\mathbb{R}^{N_\ell\cdot N_{\text{ch}}},\, P_c\leq 1,\end{aligned} \tag{12}$$

where N_ℓ is the number of transmitters, and P_c is the average signal power available at the decoder per channel use, which is defined differently depending on the particular channel model used (see Section IV-B for the definitions).

The definition of the conditional mutual information between $\boldsymbol{z}$ and $\boldsymbol{y}$ given $\boldsymbol{s}_0$ is given as [52]

$$I(\boldsymbol{z};\boldsymbol{y}|\boldsymbol{s}_0)=E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(\frac{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}{p(\boldsymbol{z}|\boldsymbol{s}_0)}\right)\right], \tag{13}$$

where $\log$ denotes the natural logarithm. When evaluating (13) for the optimization of the conditional mutual information, the optimal decoder $p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$ is intractable in practice [53]. Thus, the mutual information cannot be maximized directly, but a lower bound on (13) can be derived, which in turn can be optimized. One way of deriving such a lower bound is to use a variational approximation $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$, with parameters $\boldsymbol{\phi}$, that approximates $p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$ [30], [54], [55]. With this approach, a lower bound on the conditional mutual information can be derived (given in the Appendix), which leads to

$$\begin{aligned}&I(\boldsymbol{z};\boldsymbol{y}|\boldsymbol{s}_0)\geq \\ &-E_{p(\boldsymbol{y},\boldsymbol{s}_0)}[H(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0),q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0))]+E_{p(\boldsymbol{s}_0)}[H(p(\boldsymbol{z}|\boldsymbol{s}_0))]\,.\end{aligned} \tag{14}$$

Instead of maximizing the conditional mutual information, now the conditional mutual information lower bound (14) can be maximized. From the derivation it can be seen that the lower bound on the mutual information is tight, if the Kullback-Leibler (KL) divergence $D_{\text{KL}}\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)||q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)=0$, i.e., if the modeled probability distribution of the decoder $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$ matches the true posterior distribution $p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$ almost everywhere [52]. For an intuitive understanding of (14) we can use (34) to get

$$\begin{aligned}&-E_{p(\boldsymbol{y},\boldsymbol{s}_0)}[H(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0),q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0))]+E_{p(\boldsymbol{s}_0)}[H(p(\boldsymbol{z}|\boldsymbol{s}_0))] \\ &=\underbrace{I(\boldsymbol{z};\boldsymbol{y}|\boldsymbol{s}_0)}_{\text{encoder objective}}\underbrace{-E_{p(\boldsymbol{y},\boldsymbol{s}_0)}[D_{\text{KL}}\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)||q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)]}_{\text{decoder objective}},\end{aligned} \tag{15}$$

which shows that maximizing the conditional mutual information lower bound can be seen as optimizing an implicit encoder and decoder objective. The objective of the encoder is to maximize the mutual information of the encoded symbols after transmission over the channel $\boldsymbol{y}$ to be maximally informative about the semantic variable $\boldsymbol{z}$ given the side information $\boldsymbol{s}_0$. The objective of the decoder is to find the distribution $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$ with the minimum KL divergence to the true decoder distribution $p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)$.

For the optimization, the conditional entropy term $E_{p(\boldsymbol{s}_0)}\left[H(p(\boldsymbol{z}|\boldsymbol{s}_0))\right]$ can be omitted, as it is independent of the encoder and decoder parameters. This gives us the following minimization problem of the encoder and decoder parameters $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{N_\ell},\boldsymbol{\phi}$ of the encoders $p_{\boldsymbol{\theta}_\ell}(\boldsymbol{c}_\ell|\boldsymbol{s}_\ell)$ and the decoder $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y})$ with

$$\begin{aligned}&\underset{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{N_\ell},\boldsymbol{\phi}}{\operatorname{argmin}}\, E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[H\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0),q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \\ &\text{s.t.}\quad \boldsymbol{y}\in\mathbb{R}^{N_\ell\cdot N_{\text{ch}}},\, P_c\leq 1.\end{aligned} \tag{16}$$

In the following section, it will be discussed how the minimization of (16) can be implemented in practice using DNNs.

C. Implementation Considerations

In this section, we show how the gradients with respect to the encoder and decoder parameters can be estimated efficiently from the training data. To train the NN, training data samples of the observations $\boldsymbol{s}_0,\boldsymbol{s}_1,\ldots,\boldsymbol{s}_{N_\ell}$ are used as the inputs, and training data samples of the semantic variable $\boldsymbol{z}$ are used as the desired output. Furthermore, a suitable type of parameterized distribution for the decoder is selected such that NN training is computationally feasible.

Using Stochastic Gradient Descent (SGD) based optimization of the autoencoder, the gradient with respect to the decoder parameters $\boldsymbol{\phi}$ can be approximated for a batch of training data by

$$\frac{\partial}{\partial\boldsymbol{\phi}}E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[H\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0),q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \tag{17}$$

$$=-E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\frac{\partial\log\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)}{\partial\boldsymbol{\phi}}\right] \tag{18}$$

$$\approx-\frac{1}{L}\sum_{l=1}^{L}\frac{\partial\log\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}_l|\boldsymbol{y}_l,\boldsymbol{s}_{0_l})\right)}{\partial\boldsymbol{\phi}}, \tag{19}$$

where L is the batch size and $\boldsymbol{z}_l$, $\boldsymbol{y}_l$, and $\boldsymbol{s}_{0_l}$ are training data samples from the distribution $p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)$.

To estimate the gradients with respect to the encoders with parameters $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_{N_\ell}$, we sample from $p_{\boldsymbol{\theta}_\ell}(\boldsymbol{c}_\ell|\boldsymbol{s}_\ell)$, which depends itself on $\boldsymbol{\theta}_\ell$. This leads to high variance in the gradient estimation [53].

To get less noisy gradient estimates for the SGD optimizer, the reparametrization trick can be used by separating $\boldsymbol{y}\sim p(\boldsymbol{y}|\boldsymbol{c}_1,\ldots,\boldsymbol{c}_{N_\ell})$ into a deterministic and differentiable function $f_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{n})$ with random variable $\boldsymbol{n}$, which is independent of the $\boldsymbol{\theta}_\ell$, and $\boldsymbol{\theta}=\left[\boldsymbol{\theta}_1^\top,\ldots,\boldsymbol{\theta}_{N_\ell}^\top\right]^\top$, $\boldsymbol{s}=\left[\boldsymbol{s}_1^\top,\ldots,\boldsymbol{s}_{N_\ell}^\top\right]^\top$ [53]. In our case, we can apply the reparametrization trick by assuming deterministic encoders $p_{\boldsymbol{\theta}_\ell}(\boldsymbol{c}_\ell|\boldsymbol{s}_\ell)$ and additive noise $\boldsymbol{n}$ for the wireless channel. Then the gradients with respect to $\boldsymbol{\theta}_\ell$ can be written as

$$\frac{\partial}{\partial\boldsymbol{\theta}_\ell}E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[H\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0),q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \tag{20}$$

$$=-E_{p(\boldsymbol{z},\boldsymbol{s}_0,\boldsymbol{s}_1,\ldots,\boldsymbol{s}_{N_\ell})p(\boldsymbol{n})}\left[\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{s},\boldsymbol{n})}{\partial\boldsymbol{\theta}_\ell}\frac{\partial\log\left[q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right]}{\partial\boldsymbol{y}}\right] \tag{21}$$

$$\approx-\frac{1}{L}\sum_{l=1}^{L}\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{s}_l,\boldsymbol{n}_l)}{\partial\boldsymbol{\theta}_\ell}\left.\frac{\partial\log\left[q_{\boldsymbol{\phi}}(\boldsymbol{z}_l|\boldsymbol{y},\boldsymbol{s}_{0_l})\right]}{\partial\boldsymbol{y}}\right|_{\boldsymbol{y}=f_{\boldsymbol{\theta}}(\boldsymbol{s}_l,\boldsymbol{n}_l)}, \tag{22}$$

where $\boldsymbol{s}_l$ and $\boldsymbol{n}_l$ are training data samples of $\boldsymbol{s}$ and $\boldsymbol{n}$, respectively.

In many works of semantic communication, where the task is to solve a classification problem, e.g. as in [30], [31], [56], the posterior distribution of the decoder $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)$ is a low dimensional discrete probability distribution. However, in our case $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)$ is a continuous, or more precisely, as we operate on float32 numbers on digital computer hardware, a high dimensional discrete distribution. As an example, in our case we have a grid of 100 by 50 float32 numbers for a subsurface image. This means that the number of events is about $2^{32 \cdot 5000}$, which makes it impossible to estimate a probability for each event. Even assuming that the values at each grid point are independent, it is still infeasible as there are approximately $5000 \cdot 2^{32} \approx 2 \cdot 10^{10}$ different discrete probability events.

As a consequence, simplification is required to estimate the density $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)$. Different parametric and non-parametric methods for density estimation methods exist [57]. Here, we use a common approach of assuming independent Gaussian distributions for every i-th element z_i of $\boldsymbol{z} \in \mathbb{R}^{N_\mathrm{x} N_\mathrm{z}}$ for $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)$ with fixed variance σ^2 and means $\boldsymbol{\mu}(\boldsymbol{y}, \boldsymbol{s}_0)$. The i-th mean $\mu_i(\boldsymbol{y}, \boldsymbol{s}_0)$ is the i-th output of the decoder $\boldsymbol{\mu}(\boldsymbol{y}, \boldsymbol{s}_0)$, which depends on the decoder inputs $\boldsymbol{y}$ and $\boldsymbol{s}_0$. Our simulation results show that sufficiently good communication performance can be achieved under this assumption. We leave the investigation of different parametric distributions as further research work.

Given this assumption, our optimization objective of the cross entropy (16) becomes

$$E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[H\left(p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0), q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)\right)\right] \tag{23}$$

$$= E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_0)\right)\right] \tag{24}$$

$$= E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(\prod_{i=1}^{N_\mathrm{x} N_\mathrm{z}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(z_i - \mu_i(\boldsymbol{y}, \boldsymbol{s}_0)\right)^2}{2\sigma^2}\right)\right)\right] \tag{25}$$

$$= E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\frac{1}{2\sigma^2} \sum_{i=1}^{N_\mathrm{x} N_\mathrm{z}} \left(z_i - \mu_i(\boldsymbol{y}, \boldsymbol{s}_0)\right)^2\right] + N_\mathrm{x} N_\mathrm{z} \log\sqrt{2\pi\sigma^2} \tag{26}$$

$$= \frac{1}{2\sigma^2} E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\|\boldsymbol{z} - \boldsymbol{\mu}(\boldsymbol{y}, \boldsymbol{s}_0)\|_2^2\right] + N_\mathrm{x} N_\mathrm{z} \log\sqrt{2\pi\sigma^2}. \tag{27}$$

We see that under the given assumption our optimization objective becomes the Mean Square Error (MSE) loss $E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\|\boldsymbol{z} - \boldsymbol{\mu}(\boldsymbol{y}, \boldsymbol{s}_0)\|_2^2\right]$ [58]. Estimating the gradient for a batch of training data for (27) with respect to the encoder and decoder parameters $\boldsymbol{\phi}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{N_\ell}$ is done as shown in (19) and (22).

IV. Simulation results

In the following, we evaluate the ATC-FWI algorithm together with our proposed communication system. First, we introduce the communication systems for comparison of the proposed semantic communication system. Then our channel models, the used NN architectures, and the used training dataset are explained. In Section IV-E the communication performance, and in Section IV-F the ATC-FWI algorithm performances are evaluated for the different communication system designs.

A. Description of Compared Methods

In the following, the compared communication methods are introduced, which are shown in Figs. 4 to 8.

We note again, that in the case of transmission of the velocity model, the observations are given as $\boldsymbol{s}_\ell = \boldsymbol{m}_\ell$, and the semantic variable is given as $\boldsymbol{z} = \sum_{\ell \in \mathcal{N}_r} a_{\ell r} \widetilde{\boldsymbol{m}}_\ell$. And in the case of transmission of the gradients, the observations are given as $\boldsymbol{s} = \delta\boldsymbol{m}$, and the semantic variable is given as $\boldsymbol{z} = \sum_{\ell \in \mathcal{N}_r} b_{\ell r} \delta\boldsymbol{m}_\ell$. In our simulations, we set the coefficients $a_{\ell r}$ and $b_{\ell r}$ to be $1/|\mathcal{N}_r|$, such that the function F is simply the arithmetic mean.

First, a classic JSCC architecture is used where no side information is used at the decoder, which is shown in Fig. 4. For the case of JSCC without side information, each encoder decoder pair is optimized to individually reconstruct the observation $\hat{\boldsymbol{s}}_i$. To see how the impact the availability of side information in the encoder decoder design, we compare the case of JSCC without side information to JSCC with side information, which is shown in Fig. 5. For JSCC (with side information), the encoder pairs are optimized to individually reconstruct the observation $\hat{\boldsymbol{s}}_i$. As we consider the side information for the encoder decoder design for all further compared methods, we do not explicitly mention the consideration of side information to simplify the notation, and only mention it explicitly if no side information is used. The next approach, referred to as distributed JSCC, shown in Fig. 6, uses a joint decoder to decode all observations $\hat{\boldsymbol{s}}_1, \ldots, \hat{\boldsymbol{s}}_{N_\ell}$ to exploit the correlation between the observations 1 to N_ℓ.

Furthermore, we compare the proposed semantic JSCC and the semantic JSCC AirComp methods, which are shown in Figs. 7 and 8, respectively. The proposed methods both directly reconstruct the semantic variable $\boldsymbol{z}$ as distributed function computation, instead of reconstructing the individual observations, as only the semantic variable is required for executing the ATC-FWI algorithm. The benefit of the semantic JSCC AirComp approach compared to the semantic JSCC approach is that all agents can transmit on the same channel resources, potentially using the channel resources more efficiently, with the constraint that the transmitted data needs to be written as a nomographic function, and with the drawback that synchronization is required [42], [43].

B. Channel Models

For the evaluation, we consider N_ℓ channels from each neighboring agents to the receiving agent as independent and identically distributed (i.i.d.) Additive White Gaussian Noise (AWGN) channels. The average signal power per channel use at the decoder P_c is given as $P_c = \frac{1}{N_\ell N_\mathrm{ch}} \sum_{l=1}^{N_\ell N_\mathrm{ch}} |c_i|^2$, where c_i is the i-th entry of the vector $\boldsymbol{c}$, which is given as

$$\boldsymbol{c} = \begin{cases} \boldsymbol{c}_1 + \cdots + \boldsymbol{c}_{N_\ell} & \text{for AirComp channel,} \\ \begin{bmatrix} \boldsymbol{c}_1^\top & \ldots & \boldsymbol{c}_{N_\ell}^\top \end{bmatrix}^\top & \text{otherwise.} \end{cases} \tag{28}$$

For the AirComp channel, the power of the additive noise is σ_n^2. Otherwise, for the N_ℓ channels we assume noise powers

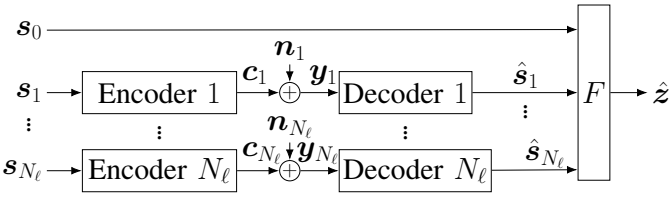

Fig. 4: JSCC without side information: Each encoder decoder pair i is optimized to reconstruct the observation $\hat{\boldsymbol{s}}_i$ without considering side information.

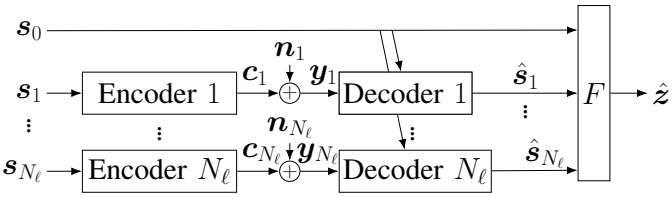

Fig. 5: JSCC: Each encoder decoder pair i is optimized to reconstruct the observation $\hat{\boldsymbol{s}}_i$ using the side information $\boldsymbol{s}_0$.

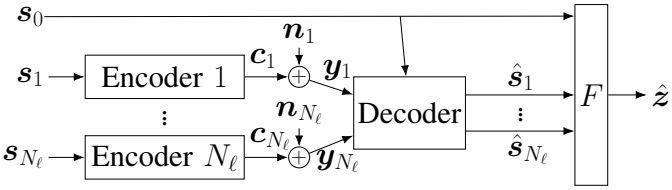

Fig. 6: Distributed JSCC: All encoders and the decoder are optimized together to reconstruct the observations $\hat{\boldsymbol{s}}_1, \ldots, \hat{\boldsymbol{s}}_{N_\ell}$ using the side information $\boldsymbol{s}_0$.

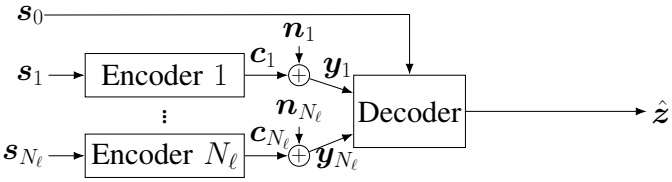

Fig. 7: Semantic JSCC: All encoders and the decoder are optimized together to reconstruct the semantic variable $\hat{\boldsymbol{z}}$ using the side information $\boldsymbol{s}_0$.

of $\sigma_{n_1}^2 = \cdots = \sigma_{n_{N_\ell}}^2 = \sigma_n^2$. The noise power σ_n^2 is scaled according to $\text{SNR} = \frac{P_c}{\sigma_n^2}$.

Finally, the number of total channel uses for neighboring agents transmitting to the receiving agent is always $N_\ell N_{\text{ch}}$, meaning that for the case AirComp channel, each user transmits has $N_\ell N_{\text{ch}}$ channel uses, and otherwise each of the N_ℓ transmitting agents has N_{ch} channel uses. For the case of AirComp, we neglect synchronization offset errors in this work, providing an upper bound on the communication performance of AirComp.

C. Neural Network Architecture

For the encoder and decoder we use a Convolutional Neural Network (CNN) architecture [59]. Each encoder consists of five 2D-convolution layers with 60 5×5 convolution filters each with batch normalization and PReLU activation function, except the last layer, where no activation function is used, and with respective strides of $\{2, 2, 2, 1, 1\}$. The same setup

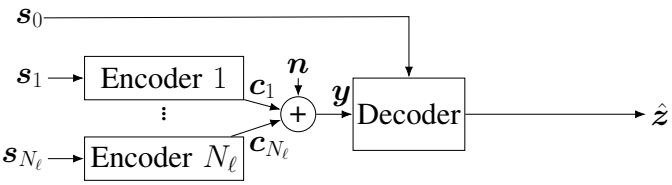

Fig. 8: Semantic JSCC AirComp: All encoders and the decoder are optimized together to reconstruct the semantic variable $\hat{\boldsymbol{z}}$ using the side information $\boldsymbol{s}_0$ for a transmission over an AirComp channel.

is used for the decoder, but with transposed convolution layers with respective strides of $\{1, 1, 2, 2, 2\}$. The number of filters of the last encoder layer is selected to match the number of channel uses N_{ch}, and the last decoder layer has just one filter to output a 100 by 50 grid of the velocity model/the gradients. Furthermore, data normalization is done before the first convolution layer and denormalization after the last decoder layer, and power normalization after the last encoder layer to match the average power constraint.

For simplification, we use parameter sharing for the encoders and decoders, such that not all combinations of sending and receiving agents need to be trained individually. For the transmissions of the velocity models and the gradients the same neural network architecture is used, but the training is done with different datasets.

D. Training Data Generation and Training

As we use NNs for the encoder and decoder and do end-to-end training, we require training data. To generate the training data, we first generate several scenarios of velocity models as ground truth P-wave velocity models. We consider a $200\,\text{m}$ wide and $100\,\text{m}$ deep two dimensional slice of the ground with associated velocity values at grid points every two meters in both directions. We consider three different scenarios of ground truth velocity models. The first two scenarios are rectangular and elliptically shaped velocity anomalies with a constant background, and the third scenario consists of three layers with varying width with different velocities.

To then generate the training data for the communication system consisting of the velocity models $\widetilde{\boldsymbol{m}}_\ell$ and the gradients $\delta\boldsymbol{m}_\ell$ for all agents ℓ, the ATC-FWI algorithm is executed with perfect communication links, and the respective velocity models and gradients are stored as training data.

$14\,000$ different velocity ground truth scenarios were generated, and the ATC-FWI algorithm was run for 20 iterations with an initial model of a Gaussian filtered ground truth for each scenario. For the training in total only $80\,000$ training data samples were used due to memory constraints. For the velocity dataset, samples were removed where all agents had nearly identical data to avoid overfitting to the side information, and for the gradient dataset $80\,000$ samples were selected randomly. All models are trained for 300 epochs with a learning rate of 10^{-4}. For the last few epochs the learning rate is lowered stepwise until 10^{-7}.

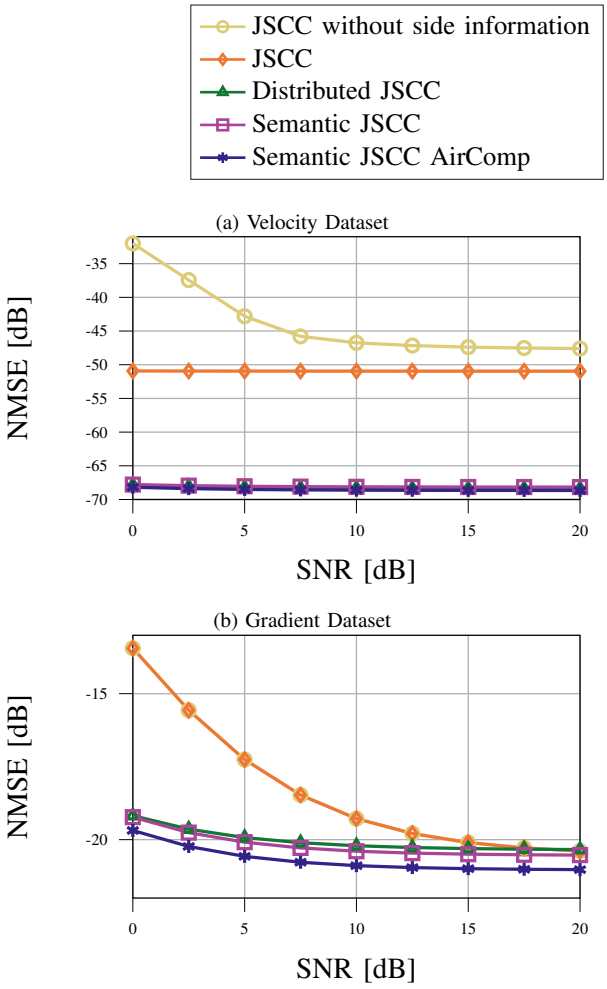

Fig. 9: Communication performance: NMSE of the semantic variable in dB for different channel SNRs is shown for the velocity and gradient evaluation datasets with 91 channel uses, training SNR 10 dB, and $N_\ell = 2$.

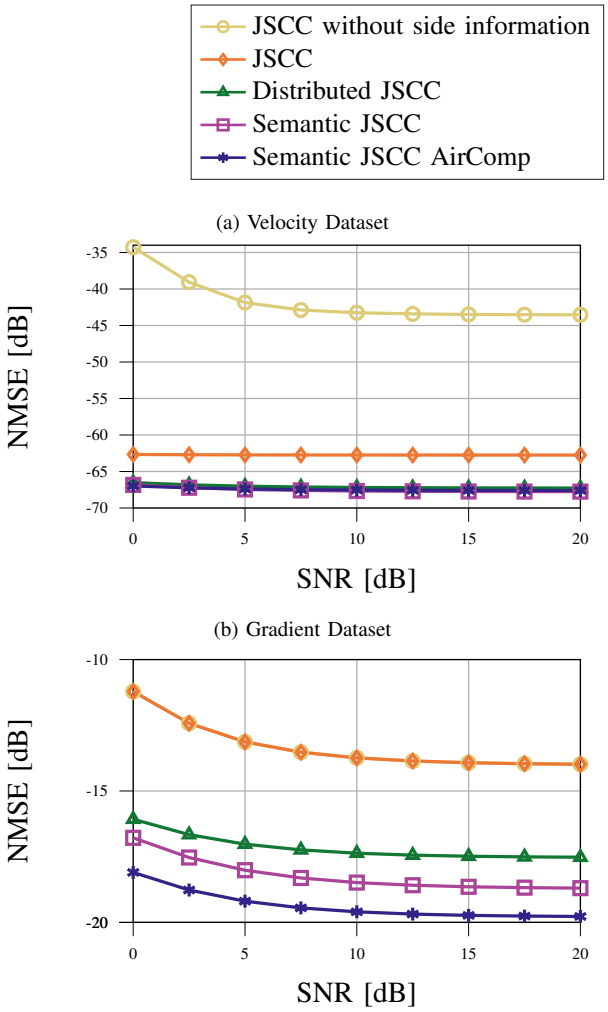

Fig. 10: Communication performance: NMSE of the semantic variable in dB for different channel SNRs is shown for the velocity and gradient evaluation datasets with 91 channel uses, training SNR 10 dB, and $N_\ell = 7$.

E. Results - Communication Performance

Our performance metric to evaluate the communication performance is the Normalized Mean Square Error (NMSE) of the semantic variable $\frac{\|\hat{\boldsymbol{z}}-\boldsymbol{z}\|_2^2}{\|\boldsymbol{z}\|_2^2}$, where $\hat{\boldsymbol{z}}$ is the reconstructed semantic variable at the receiving agent, and $\boldsymbol{z}$ is the semantic variable ground truth. Figs. 9 and 10 show the NMSE of the semantic variable over different channel SNRs for 15 000 evaluation data samples that were not used for training.

The training SNR is set to 10 dB, and the number of channel uses is $N_{\text{ch}} = 91$. Fig. 9a shows the results for the velocity dataset, and Fig. 9b shows the results for the gradient dataset for the case of two transmitting agents. Analogously, Fig. 10 shows the results for seven transmitting agents.

For the velocity datasets, the results are shown for two transmitting agents ($N_\ell = 2$) in Fig. 9a, and for seven transmitting agents ($N_\ell = 7$) in Fig. 10a. It is observed that considering the side information at the decoder significantly improves the NMSE of the semantic variable for the velocity dataset. This is expected as the correlation between the velocity models from the transmitting agents to the receiving agent is high. The distributed JSCC, the semantic JSCC, and the semantic JSCC AirComp approach achieve a lower NMSE than the JSCC approach, as they consider the correlation of the velocity models between all agents. The three models all achieve nearly the same NMSE.

The achieved NMSE for the semantic JSCC approach is expected to be the same as the distributed JSCC approach, as the pixel-wise correlation of the velocity models is always non-negative. For the special case of Gaussian sources it is known that for linear function computation of two variables, which can be written as a sum of two variables with positive weights, the required sum rate for asymptotically error free reconstruction is the same whether the linear function is reconstructed or the individual variables are reconstructed, if the correlation between the two variables is non-negative [46], [47]. Furthermore, if the correlation between the two variables is negative, the required sum rate is smaller if the decoder directly reconstructs the function [47].

For the gradient dataset, the absolute value of the correlation of the gradients to be transmitted is lower compared to the

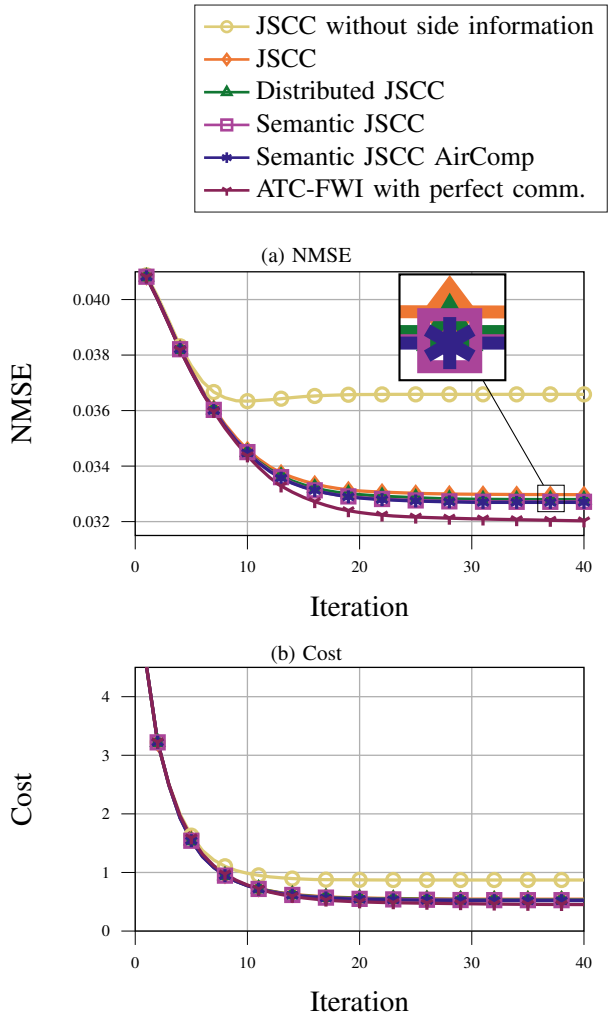

Fig. 11: ATC-FWI algorithm performance: NMSE of the estimated velocity averaged over all agents and the ATC-FWI cost function over the algorithm iterations is shown for a MAS of eight agents with a line communication topology (two communicating neighbors for each agent, except at the edges only one communicating neighbor).

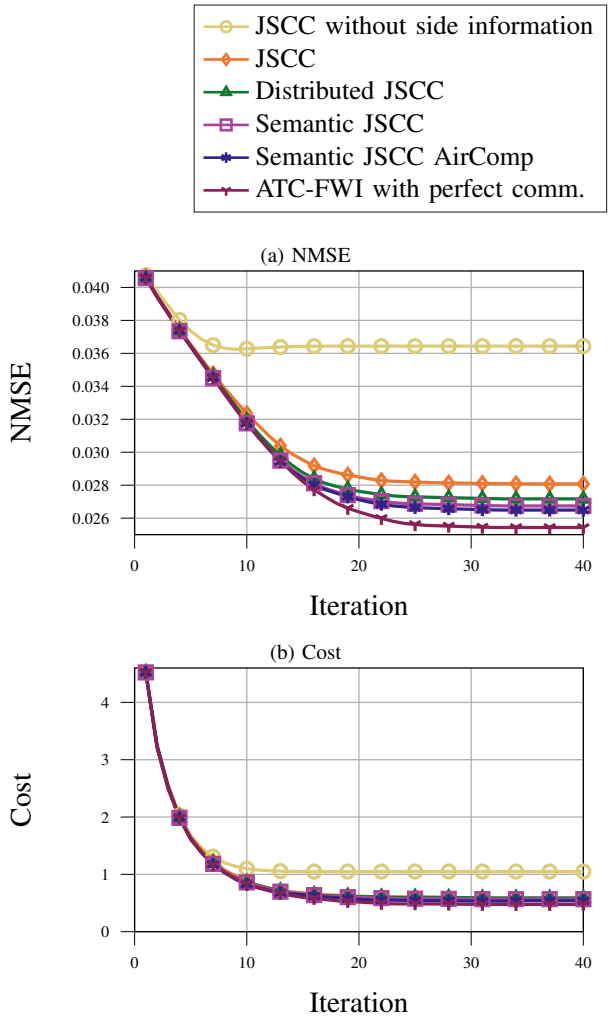

Fig. 12: ATC-FWI algorithm performance: NMSE of the estimated velocity averaged over all agents and the ATC-FWI cost function over the algorithm iterations is shown for a MAS of eight agents with a full mesh communication topology (seven communicating neighbors for each agent).

velocity dataset. However, there is some negative pixel-wise correlation for the gradient dataset. The results are shown in Fig. 9b for two transmitting agents, and in Fig. 10b for seven transmitting agents. For the gradient dataset, using the side information at the decoder resulted in overfitting of the encoders and decoder during training. As a consequence, the side information was not used for the JSCC approach. For the gradient dataset, the improvement in NMSE of the semantic JSCC approach to the distributed JSCC approach is significantly larger compared to the case of the velocity dataset. As discussed above, for the special case of Gaussian sources, the required sum rate can be smaller if the decoder directly reconstructs the function in the semantic JSCC approach compared to the distributed JSCC approach, where all individual variables are reconstructed if the correlation is negative. This effect is observed to be larger for the case of seven transmitters in Fig. 10b. Finally, the semantic JSCC AirComp approach outperforms all other methods for the gradient dataset, which is expected here, as AirComp is more advantageous for a larger number of agents, as the shared bandwidth for all agents increases [17].

F. Results - ATC-FWI Algorithm Performance With Communication

For the simulation of the ATC-FWI algorithm with the different communication systems, we use the trained end-to-end communication systems that were evaluated for their communication performance in the previous section. Furthermore, we compare the proposed communication systems with the case where no errors during communication occur, which is referred to as perfect communication in the following.

Our evaluation metric for the ATC-FWI algorithm at iteration k is the NMSE of the estimated velocity averaged over all agents, given as

$$\frac{1}{N_{\mathrm{R}}} \sum_{r=1}^{N_{\mathrm{R}}} \frac{\|\boldsymbol{m}_r^{[k]} - \boldsymbol{m}^*\|_2^2}{\|\boldsymbol{m}^*\|_2^2}, \tag{29}$$

where $\boldsymbol{m}^*$ is the velocity model ground truth. Furthermore, we show the global cost of the ATC-FWI algorithm (6) over the number of iterations k. The results are averaged over 100 independent runs with different velocity ground truths from the evaluation dataset. The seismic source locations are fixed with equidistant positions for all simulations with $N_{\mathrm{S}} = 16$. We employ $N_{\mathrm{R}} = 8$ agents for all simulations, where for each run, the agents' positions are selected randomly on a line of the surface of the two dimensional slice of the ground. We investigate a line and a full mesh communication topology. For the line topology every agent is able to communicate only with its direct neighbors, meaning that each agent can communicate with two neighbors, which corresponds to the investigation in the previous section of the case of two transmitting agents. The agents at the end of the line can only communicate with one neighbor. For the full mesh topology, every agent can communicate with every other agent, meaning that for our setting of eight agents, this corresponds to the investigation in the previous section of the case of seven transmitting neighbors.

As the ATC-FWI algorithm is sensitive to initialization, the less precise distributed traveltime tomography algorithm, could be run for a few iterations to serve as the input for the ATC-FWI algorithm, as the distributed traveltime tomography is not sensitive to initialization [60]. Here, we simply use a Gaussian smoothed version of the true velocity model for the initialization of the ATC-FWI algorithm. Furthermore, we locally stop the ATC-FWI algorithm prematurely at each agent if the local cost (7) increases to avoid divergence of the algorithm, with the averaged cost shown in Figs. 11b and 12b.

The results for the line topology, where the agents can only communicate with their direct neighbors, are shown in Fig. 11. If no side information is used at the decoder, the performance is significantly limited by the compression from $N_{\mathrm{x}}N_{\mathrm{z}} = 5000$ grid point values to $N_{\mathrm{ch}} = 91$ channel uses. The semantic JSCC and the semantic JSCC AirComp perform similarly and slightly outperform the other methods in terms of NMSE of the estimated velocity model.

For the case of the full mesh topology (Fig. 12), the overall NMSE is much lower compared to the line topology after 40 iterations, as the gradients and velocity models can be exchanged more easily between the agents. Furthermore, the advantage of the semantic JSCC and the semantic JSCC AirComp approaches is significantly larger, which can be explained by the previously observed lower NMSE for the gradient transmission in Fig. 10b.

Finally, we show the estimated velocity models for one velocity model with a squared anomaly of about $3.20\,\mathrm{km/s}$ with a background of about $1.45\,\mathrm{km/s}$. The results shown in Fig. 13 are after 40 iterations of the ATC-FWI algorithm. For all compared methods the estimated velocity models are shown in absolute terms, and for the methods without perfect communication, the difference of each case to the case of perfect communication is shown, which shows that without perfect communication, the velocity of the anomaly is underestimated on average. It can be seen that semantic JSCC AirComp outperforms the other methods.

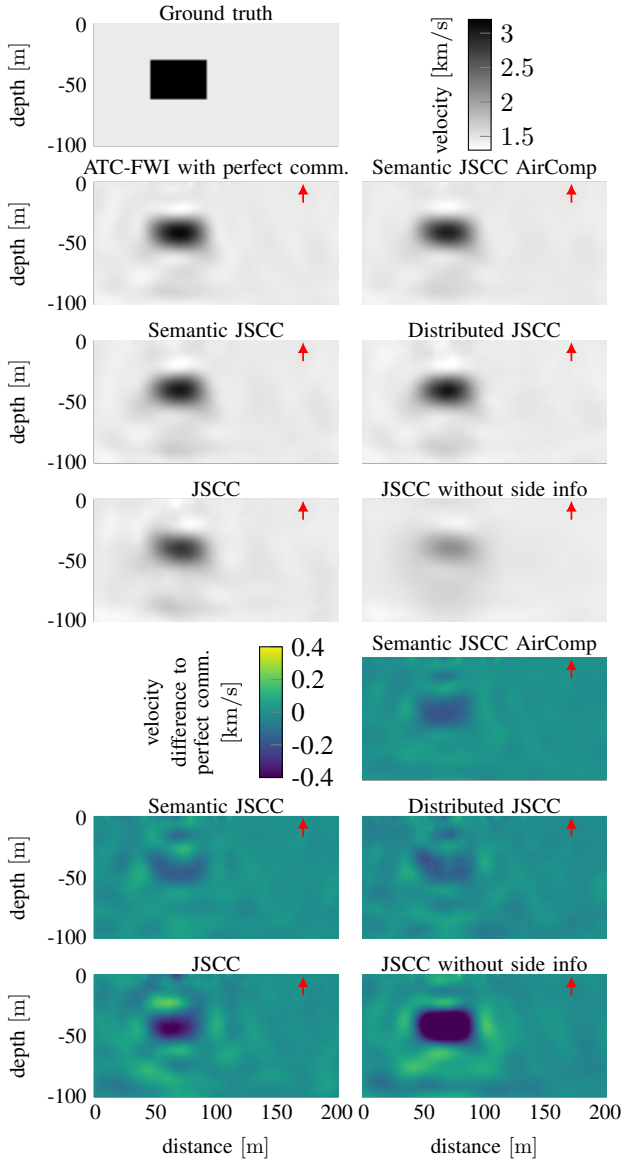

Fig. 13: The estimated velocity for the agent positioned at the red arrow after 40 iterations of the ATC-FWI algorithm is shown for a block anomaly shown in the top left (ground truth). A MAS of eight agents is used with a full mesh communication topology. For the case of the ATC-FWI with perfect communication, the estimated velocity is shown, and for the other methods, the difference to the case of ATC-FWI with perfect communication is shown.

V. Conclusion

We proposed a semantic communication system for distributed function computation for the exploration/imaging methods for MAS, investigated in this work for the ATC-FWI algorithm. It was shown that our proposed semantic JSCC AirComp can outperform state-of-the-art JSCC methods.

We showed that for a strict compression rate, where a generic JSCC approach fails, our proposed semantic methods using side information significantly outperform the generic JSCC communication system without side information in the imaging task. For the case of seven transmitting neighboring agents it was demonstrated that the distributed JSCC approach could be outperformed by the proposed semantic JSCC

approach, which again was outperformed by the proposed semantic JSCC AirComp approach. The investigation on how much a synchronization offset would limit the performance of the proposed semantic JSCC AirComp approach is left as future work.

APPENDIX
DERIVATION OF LOWER BOUND OF CONDITIONAL MUTUAL INFORMATION

$$I(\boldsymbol{z};\boldsymbol{y}|\boldsymbol{s}_0) = E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(\frac{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}{p(\boldsymbol{z}|\boldsymbol{s}_0)}\right)\right] \tag{30}$$

$$= E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(\frac{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}{p(\boldsymbol{z}|\boldsymbol{s}_0)}\frac{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}\right)\right] \tag{31}$$

$$\begin{aligned} &= E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \\ &\quad - E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(p(\boldsymbol{z}|\boldsymbol{s}_0)\right)\right] \\ &\quad + E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(\frac{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}\right)\right] \end{aligned} \tag{32}$$

$$\begin{aligned} &= E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[E_{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right]\right] \\ &\quad - E_{p(\boldsymbol{z},\boldsymbol{s}_0)}\left[\log\left(p(\boldsymbol{z}|\boldsymbol{s}_0)\right)\right] \\ &\quad + E_{p(\boldsymbol{z},\boldsymbol{y},\boldsymbol{s}_0)}\left[\log\left(\frac{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}{q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)}\right)\right] \end{aligned} \tag{33}$$

$$\begin{aligned} &= -E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[H\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0), q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \\ &\quad + E_{p(\boldsymbol{s}_0)}\left[H(p(\boldsymbol{z}|\boldsymbol{s}_0))\right] \\ &\quad + E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[D_{\mathrm{KL}}\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)||q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \end{aligned} \tag{34}$$

$$\begin{aligned} &\geq -E_{p(\boldsymbol{y},\boldsymbol{s}_0)}\left[H\left(p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0), q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{s}_0)\right)\right] \\ &\quad + E_{p(\boldsymbol{s}_0)}\left[H(p(\boldsymbol{z}|\boldsymbol{s}_0))\right], \end{aligned} \tag{35}$$

where $E_{p(x)}[\cdot]$ is the expected value with respect to $x \sim p(x)$, $H(\cdot)$ is the Shannon entropy, $H(q,p)$ is the cross entropy between p and q, and $D_{\mathrm{KL}}(p||q)$ is the KL divergence from p to q. The step for (35) uses the fact that the KL divergence is non-negative [52].

REFERENCES

[1] S. Zhang, R. Pohlmann, T. Wiedemann, A. Dammann, H. Wymeersch, and P. A. Hoeher, "Self-aware swarm navigation in autonomous exploration missions," *Proceedings of the IEEE*, vol. 108, no. 7, p. 1168–1195, Jul. 2020. [Online]. Available: http://dx.doi.org/10.1109/JPROC.2020.2985950

[2] R. J. A. A. Nathan, S. Strand, D. Mehrwald, D. Shutin, and O. Bimber, "An autonomous drone swarm for detecting and tracking anomalies among dense vegetation," 2024. [Online]. Available: https://arxiv.org/abs/2407.10754

[3] G. Rabideau, J. Russino, A. Branch, N. Dhamani *et al.*, "Planning, scheduling, and execution on the moon: the cadre technology demonstration mission," 2025. [Online]. Available: https://arxiv.org/abs/2502.14803

[4] T. Wiedemann, D. Shutin, and A. J. Lilienthal, "Model-based gas source localization strategy for a cooperative multi-robot system—a probabilistic approach and experimental validation incorporating physical knowledge and model uncertainties," *Robotics and Autonomous Systems*, vol. 118, p. 66–79, Aug. 2019. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2019.03.014

[5] B.-S. Shin and D. Shutin, "Joint distributed traveltime and full waveform tomography for enhanced subsurface imaging in seismic networks," *IEEE Transactions on Computational Imaging*, vol. 10, p. 600–612, 2024. [Online]. Available: http://dx.doi.org/10.1109/TCI.2024.3366754

[6] E. Rizk, K. Yuan, and A. H. Sayed, "Diffusion learning with partial agent participation and local updates," 2025. [Online]. Available: https://arxiv.org/abs/2505.11307

[7] J. Zhang, Z. Liu, Y. Zhu, E. Shi, B. Xu, C. Yuen, D. Niyato, M. Debbah, S. Jin, B. Ai, and S. Xuemin, "Multi-agent reinforcement learning in wireless distributed networks for 6g," 2025. [Online]. Available: https://arxiv.org/abs/2502.05812

[8] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.

[9] 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 Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213–250, 2022.

[10] D. Wheeler and B. Natarajan, "Engineering semantic communication: A survey," *IEEE Access*, vol. 11, pp. 13 965–13 995, 2023.

[11] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for b5g networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE journal of selected topics in signal processing*, vol. 17, no. 1, pp. 9–39, 2023.

[12] H. Kalita, A. Quintero, A. Wissing, B. Haugh, C. Angie, G. Nail, J. Wilson, J. Richards, J. Landin, K. Kukkala, M. Vazquez, N. Tan, Q. Lamey, R. Lu, R. Peralta, V. Vilvanathan, and J. Thangavelautham, "Evaluation of lunar pits and lava tubes for use as human habitats," in *Earth and Space 2021*. American Society of Civil Engineers, Apr. 2021, p. 944–957. [Online]. Available: http://dx.doi.org/10.1061/9780784483374.086

[13] J.-P. de la Croix, F. Rossi, R. Brockers, D. Aguilar, K. Albee, E. Boroson, A. Cauligi, J. Delaune, R. Hewitt, D. Kogan, G. Lim, B. Morrell, Y. Nakka, V. Nguyen, P. Proença, G. Rabideau, J. Russino, M. S. da Silva, G. Zohar, and S. Comandur, "Multi-agent autonomy for space exploration on the cadre lunar technology demonstration," in *2024 IEEE Aerospace Conference*. IEEE, Mar. 2024, p. 1–14. [Online]. Available: http://dx.doi.org/10.1109/AERO58975.2024.10521425

[14] A. H. Sayed *et al.*, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Processing Magazine*, pp. 155–171, 2013.

[15] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, p. 944–966, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1137/14096668X

[16] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6180–6195, 2011.

[17] O. Abari, H. Rahul, and D. Katabi, "Over-the-air function computation in sensor networks," *arXiv preprint arXiv:1612.02307*, 2016.

[18] J. Virieux and S. Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, 2009.

[19] A. Fichtner, *Full Seismic Waveform Modelling and Inversion*. Springer, Berlin, Heidelberg, 2009.

[20] H. S. Aghamiry, A. Gholami, and S. Operto, "Full waveform inversion by proximal newton method using adaptive regularization," *Geophysical Journal International*, vol. 224, pp. 169–180, 2021.

[21] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, "Total variation regularization strategies in full-waveform inversion," *SIAM Journal on Imaging Sciences*, vol. 11, pp. 376–406, 2018.

[22] Q. Zhao, D. B. Lindell, and G. Wetzstein, "Learning to solve pde-constrained inverse problems with graph networks," in *ICML*, 2022.

[23] B. Li, H. Wang, S. Feng, X. Yang, and Y. Lin, "Solving seismic wave equations on variable velocity models with fourier neural operator," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, p. 1–18, 2023. [Online]. Available: http://dx.doi.org/10.1109/TGRS.2023.3333663

[24] X. Huang and T. Alkhalifah, "Learned frequency-domain scattered wavefield solutions using neural operators," *Geophysical Journal International*, vol. 241, no. 3, p. 1467–1478, Mar. 2025. [Online]. Available: http://dx.doi.org/10.1093/gji/ggaf113

[25] C. Zou, K. Azizzadenesheli, Z. E. Ross, and R. W. Clayton, "Deep neural helmholtz operators for 3-d elastic wave propagation and inversion," *Geophysical Journal International*, vol. 239, no. 3, p. 1469–1484, Sep. 2024. [Online]. Available: http://dx.doi.org/10.1093/gji/ggae342

[26] F. Li, M. Valero, Y. Cheng, T. Bai, and W. Song, "Distributed sensor networks based shallow subsurface imaging and infrastructure monitoring," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, p. 241–250, 2020. [Online]. Available: http://dx.doi.org/10.1109/TSIPN.2020.2975349

[27] S. Wang, M. P. Panning, S. D. Vance, and W. Song, "Underground microseismic event monitoring and localization within sensor networks," *Sensors*, vol. 21, no. 8, p. 2830, Apr. 2021. [Online]. Available: http://dx.doi.org/10.3390/s21082830

[28] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gündüz, "Deep joint source-channel coding for semantic communications," *IEEE communications Magazine*, vol. 61, no. 11, pp. 42–48, 2023.

[29] D. Gündüz, M. A. Wigger, T.-Y. Tung, P. Zhang, and Y. Xiao, "Joint source–channel coding: Fundamentals and recent progress in practical designs," *Proceedings of the IEEE*, pp. 1–32, 2024.

[30] E. Beck, C. Bockelmann, and A. Dekorsy, "Semantic information recovery in wireless networks," *Sensors*, vol. 23, no. 14, p. 6347, 2023.

[31] A. H. Razlighi, M. H. Tillmann, E. Beck, C. Bockelmann, and A. Dekorsy, "Cooperative and collaborative multi-task semantic communication for distributed sources," *arXiv preprint arXiv:2411.02150*, 2024.

[32] S. Wang, K. Yang, J. Dai, and K. Niu, "Distributed image transmission using deep joint source-channel coding," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 5208–5212.

[33] J. Whang, A. Nagle, A. Acharya, H. Kim, and A. G. Dimakis, "Neural distributed source coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 5, pp. 493–508, 2024.

[34] N. Mital, E. Özyilkan, A. Garjani, and D. Gündüz, "Neural distributed image compression with cross-attention feature alignment," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2023, pp. 2498–2507.

[35] H. Yang, T. Ding, and X. Yuan, "Federated learning with lossy distributed source coding: Analysis and optimization," *IEEE Transactions on Communications*, vol. 71, no. 8, pp. 4561–4576, 2023.

[36] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2022–2035, 2020.

[37] O. Abari, H. Rahul, and D. Katabi, "Over-the-air function computation in sensor networks," *arXiv preprint arXiv:1612.02307*, 2016.

[38] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep over-the-air computation," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[39] S. Wei, C. Feng, C. Guo, B. Zhang, and J. Chen, "Deep over-the-air computation based multi-user semantic communication for smart cities," in *2024 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2024, pp. 622–626.

[40] P. Yi, Y. Cao, X. Kang, and Y.-C. Liang, "Integrated distributed semantic communication and over-the-air computation for cooperative spectrum sensing," *IEEE Transactions on Communications*, pp. 1–1, 2024.

[41] N. Agrawal, R. L. G. Cavalcante, M. Yukawa, and S. Stańczak, "Distributed convex optimization "over-the-air" in dynamic environments," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 10, pp. 610–625, 2024.

[42] Z. Wang, Y. Zhao, Y. Zhou, Y. Shi, C. Jiang, and K. B. Letaief, "Over-the-air computation for 6g: Foundations, technologies, and applications," *IEEE Internet of Things Journal*, vol. 11, no. 14, pp. 24 634–24 658, 2024.

[43] A. Şahin and R. Yang, "A survey on over-the-air computation," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 3, pp. 1877–1908, 2023.

[44] K. Eswaran and M. Gastpar, "Chapter 1 - foundations of distributed source coding," in *Distributed Source Coding*, P. L. Dragotti and M. Gastpar, Eds. Boston: Academic Press, 2009, pp. 3–31. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123744852000056

[45] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1912–1923, 1997.

[46] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1938–1961, 2008.

[47] Y. Yang and Z. Xiong, "Distributed compression of linear functions: Partial sum-rate tightness and gap to optimal sum-rate," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2835–2855, 2014.

[48] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[49] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." *Journal of machine learning research*, vol. 11, no. 12, 2010.

[50] R. E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophys. J. Int.*, 2006.

[51] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman, "Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration," *Geoscientific Model Development*, vol. 12, no. 3, p. 1165–1187, Mar. 2019. [Online]. Available: http://dx.doi.org/10.5194/gmd-12-1165-2019

[52] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[53] O. Simeone, *Machine learning for engineers*. Cambridge university press, 2022.

[54] B. Poole, S. Ozair, A. Van Den Oord, A. Alemi, and G. Tucker, "On variational bounds of mutual information," in *International conference on machine learning*. PMLR, 2019, pp. 5171–5180.

[55] D. Barber and F. Agakov, "The im algorithm: a variational approach to information maximization," *Advances in neural information processing systems*, vol. 16, no. 320, p. 201, 2004.

[56] A. H. Razlighi, C. Bockelmann, and A. Dekorsy, "Semantic communication for cooperative multi-task processing over wireless networks," *IEEE Wireless Communications Letters*, 2024.

[57] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.

[58] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[59] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999–7019, 2021.

[60] B.-S. Shin and D. Shutin, "Distributed traveltime tomography using kernel-based regression in seismic networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, p. 1–5, 2022.