Causal Semantic Communication for Digital Twins: A Generalizable Imitation Learning Approach

Christo Kurisummoottil Thomas, *Member, IEEE* and Walid Saad, *Fellow, IEEE*,
and Yong Xiao, *Senior Member, IEEE*

***Abstract*—A digital twin (DT) leverages a virtual representation of the physical world, along with communication (e.g., 6G), computing (e.g., edge computing), and artificial intelligence (AI) technologies to enable many connected intelligence services. In order to handle the large amounts of network data based on digital twins (DTs), wireless systems can exploit the paradigm of semantic communication (SC) for facilitating informed decision-making under strict communication constraints by utilizing AI techniques such as causal reasoning. In this paper, a novel framework called causal semantic communication (CSC) is proposed for DT-based wireless systems. The CSC system is posed as an imitation learning (IL) problem, where the transmitter, with access to optimal network control policies using a DT, teaches the receiver using SC over a bandwidth-limited wireless channel how to improve its knowledge to perform optimal control actions. The causal structure in the transmitter's data is extracted using novel approaches from the framework of deep end-to-end causal inference, thereby enabling the creation of a semantic representation that is causally invariant, which in turn helps generalize the learned knowledge of the system to new and unseen situations. The CSC decoder at the receiver is designed to extract and estimate semantic information while ensuring high semantic reliability. The receiver control policies, semantic decoder, and causal inference are formulated as a bi-level optimization problem within a variational inference framework. This problem is solved using a novel concept called network state models, inspired from world models in generative AI, that faithfully represents the environment dynamics leading to data generation. Furthermore, the proposed framework includes an analytical characterization of the performance gap that results from employing a suboptimal policy learned by the receiver that uses the transmitted semantic information to construct a model of the physical environment. The CSC system utilizes two concepts, namely the integrated information theory principle in the theory of consciousness and the abstract cell complex concept in topology, to precisely express the information content conveyed by the causal states and their relationships. Through this analysis, novel formulations of semantic information, semantic reliability, distortion, and similarity metrics are proposed, which extend beyond Shannon's concept of uncertainty. Simulation results demonstrate that the proposed CSC system outperforms conventional wireless and state-of-the-art SC systems by achieving better semantic reliability with reduced bits and enabling better control policies over time thanks to the generative AI architecture.**

I. INTRODUCTION

Digital twins (DTs) are replicas of the physical world [1] created through the use of simulation software, data analytics, and sensor data. These virtual replicas rely on real-time data and advanced algorithms to model and predict the behavior and performance of physical systems, enabling better decision-making and improved efficiency across a range of industries, such as manufacturing, healthcare, transportation, and aerospace. DTs can also be used to assist wireless systems [2] to enable self-configuration, proactive online learning, and support for *connected intelligence (CI)* applications such as haptics, brain-computer interaction, flying vehicles, extended reality (XR), and the metaverse. In order to create a DT within a wireless system (e.g., at a base station or mobile edge computing server), there is need for significant computing resources and transmission of large volumes of data. Additionally, the DT data collected at the base station will be employed to develop a model of the physical world, which can assist in making better control decisions or aiding CI applications. Transmitting such a staggering amount of information back from the base station to the end users can be highly inefficient from a resource utilization perspective, leading to increased latency, high power consumption, and reduced spectrum efficiency. To meet the demands of high-rate, high-reliability, and time-criticality for the aforementioned CI applications, DT-powered 6G systems could transmit only essential information relevant to the end-user. This concept forms the basis of *semantic communication (SC)* systems [3]–[5]. Clearly, integrating DTs with SC offers a more effective solution for handling large amounts of data and optimizing resource utilization. Moreover, by employing SC, DTs can leverage contextual knowledge to better interpret and process data generated from diverse sources, leading to more precise predictions for event monitoring, optimal control decisions, and improved au-

---

Christo Kurisummoottil Thomas and Walid Saad are with the Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA, USA.(emails:{christokt,walids}@vt.edu).

Yong Xiao is with the School of Electronic Information and Communications at the Huazhong University of Science and Technology, Wuhan 430074, China, also with the Peng Cheng Laboratory, Shenzhen, Guangdong 518055, China, and also with the Pazhou Laboratory (Huangpu), Guangzhou, Guangdong 510555, China (e-mail: yongxiao@hust.edu.cn).

The work of Christo Kurisummoottil Thomas and Walid Saad are supported by US National Science Foundation under Grant CNS-2210254.

tonomous agent capabilities. However, integrating DTs with SCs poses unique *challenges* particularly on the SC side. For instance, the transmitter that manages the DT must ensure that the model is trained on representative data, which can be challenging in non-stationary and dynamic wireless environments. In addition, transmitting data from a DT-enabled transmitter to facilitate real-time prediction, control, or reconstruction tasks at the end-users can be challenging due to the need for ultra-low-latency and ultra-reliable communication. To address the challenges of DT-based semantic communication, one promising approach is to extract the causal structure inherent in the network data. This enables the development of an accurate physical environment model at the transmit and receive nodes with fewer training samples. By leveraging generative artificial intelligence (AI) methods, this approach ultimately enables real-time prediction, reconstruction, and control.

A. *Prior Works*

Despite recent AI-based SC designs [6]–[11], these prior works failed to address a few critical aspects that hinder the integration of DT and SC. Firstly, most of these designs ignore the need for a rigorous formulation of semantics, instead considering the semantic concepts as being mapped to data via a probabilistic transformation [6]. For a DT, this probabilistic design could lead to inaccurate physical models at the end nodes, which in turn would lead to inaccurate decisions. Second, the solutions of [6]–[11] suffer from limited generalizability to scenarios not encountered during training. These approaches typically use variational auto-encoders (VAEs) or transformers without structure extraction, which may reduce the ability to handle unseen wireless environments. This lack of generalizability necessitates significant retraining efforts, leading to increased communication overheads and delays, which can hinder real-time prediction, reconstruction, and control under communication constraints for DT-based wireless systems. Meanwhile, the works in [12]–[17] investigated the use of DTs to enhance the performance of wireless systems. For instance, the authors in [12] proposed a distributed DT-powered federated learning framework to support edge computing in industrial IoT, while the work in [13] suggested constructing DTs at edge networks by using blockchain and federated learning. In a recent work [14], the authors investigated the joint synchronization of DTs and sub-metaverses within a distributed metaverse framework. Despite being interesting, the prior art [12]–[17] has not adequately addressed modeling the environment dynamics that capture the data observed at the DT side. The only exception is the work in [18] that proposed the use of a DT that builds a Bayesian model of the communication system environment using conventional signal processing techniques. While the Bayesian modeling approach presented is intriguing, the solution of [18] is not generalizable, i.e., it is limited to the specific data distribution used to create the DT, and thus it cannot generalize to multiple wireless environment. Moreover, an important aspect missing in the prior art [12]–[18] is the efficient transmission of network state and control information from DT-based nodes to edge users, which poses a challenge to achieving the high-reliability and low-latency goals of future wireless networks. To the authors' knowledge, this work will be the *first* to combine SC and DTs to address the aforementioned challenges.

B. *Contributions*

In contrast to the state-of-art that lacks a rigorous definition of semantics and generalizable SC system design, the main contribution of this paper is a novel framework for designing a causal SC (CSC) system consisting of two components: 1) a transmitter (called teacher) based on DTs that identifies the semantic content elements (SCEs) present in the data and performs causal discovery of the state transitions being transmitted, and 2) a receiver node (acting as apprentice) that learns the environment dynamics using the history of transmitted semantic representations and designs better control policies to maximize semantic effectiveness. While imitation learning (IL) [19] is a promising method for implementing this envisioned teacher-apprentice framework for SC [20], practical challenges arise when implementing existing IL methods [21] in DT-based SC systems. These challenges stem from the resource-intensive nature of demonstration data available at the DT side and the imperfections that may arise when transmitting data over a wireless channel, leading to inaccurate policy learning. The authors in [20] used model-free reinforcement learning (MFRL) to infer implicit semantic entities and relations from explicit semantics or observables. However, the testing performance of this approach is dependent on the specific data distribution used during training, and thus not generalizable. To overcome these issues, we propose to advance IL utilizing model-based RL (MBRL) thereby enabling the development of a model of the wireless environment dynamics using structural causal models (SCM) [22]. The proposed AI-based components in the SC chain rely on novel semantic information measures inspired by the concept of *integrated information theory (IIT)* in the literature on the *theory of consciousness* [23]. IIT measures help identify distinct *SCEs*, which hold meanings or semantics and are present within the network states observed by the DT. In summary, our key contributions include:

- We introduce *novel information measures* for the learned *SCM* at the imitator, inspired from the IIT measures [24]. We also propose a new semantic state abstraction concept that utilizes the intrinsic information concept from IIT. Semantic state abstraction plays a crucial role as it helps eliminate irrelevant information observed at the transmitter side, thus reducing transmitted information. Moreover, it enhances the ability of the agent to generalize to previously unseen areas of the state space, which could correspond to

different wireless environments. This allows us to develop generalizable native AI-based wireless SC systems.
- We use the concept of abstract simplical complexes in topology to theoretically characterize causal relations among SCEs in the data observed at the DT. Moreover, we demonstrate that the SCEs and their associations can be algebraically defined as a cell complex structure. By defining the SCEs in accordance with IIT, these topological characterizations enable us to define semantic metrics such as similarity and information that go beyond conventional information theory concepts. This capability enables the creation of DT-based SC systems that can precisely identify the causal structure and consequently develop causally invariant semantic representations.
- At the receiver, we create a *"network state model"* similar to the concept of world models in generative AI [25]. We design a semantic decoder that extracts the maximum semantic information. However, due to limited computing resources at the imitator node, the model learns a sub-optimal version of the environment dynamics. This model is then used to learn imitator policies for controlling communication tasks. To solve for the neural network (NN) parameters, we propose a bi-level optimization method within an MBRL framework. This approach enables us to learn both the state transition model and the imitator policy.
- We analytically characterize the performance shortfall (compared to the expert agent's policy) in terms of the quality-of-experience (QoE) associated with the sub-optimal network state models created at the receiver node. The QoE is defined as a function of the semantic effectiveness resulting from the transmitted semantic representation.
- Simulation results demonstrate the superiority of the proposed CSC in semantic reliability and throughput delivered by a DT-based SC system, in contrast to similar wireless systems that rely on traditional AI approaches. Furthermore, even when the channel quality is very poor, the proposed CSC achieves a semantic reliability that is 3.4 times better than classical DT-based SC systems due to its generative AI architecture at the receiver that allows it to reconstruct data. The simulations conducted in our study provide further evidence that the proposed CSC can adapt quickly to non-stationary conditions in wireless environments. The CSC requires approximately 3K fewer samples for retraining, which validates its generalizability.

The rest of this paper is organized as follows. In Section II, we present the proposed CSC system model. Section III describes the proposed IIT measures for quantifying the semantic information. Section IV introduces the MBRL based IL for learning the transmitter and receiver modules. Section IV provides simulation results. Finally, conclusions are drawn in Section VI[1].

Notations: Lower-case letter a is a scalar, boldface lower-case $\boldsymbol{a}$ and upper-case letter $\boldsymbol{A}$ represent a vector and a matrix, respectively. A set (either discrete or continuous entries) or a topological space is represented using Calligraphic font $\mathcal{X}$. $\mathcal{R}^M$ represents an $M-$dimensional vector whose entries belong to real numbers $\mathcal{R}$. $[a, b]$ represents the real number range between a and b. $\boldsymbol{A} \odot \boldsymbol{A}$ represents the element-wise multiplication of two matrices. $|a|$ represents the absolute value of a.

II. SYSTEM MODEL

Consider a point-to-point SC system that employs a DT to aid in making control decisions. This system can capture many practical applications. For example, in the context of 6G and open radio access network (ORAN) systems, a network DT located at a cloud data center would replicate various latent aspects of a network, including signals, coverage, interference, traffic behavior, and user mobility, across different frequency layers. The DT optimizes sensitive parameters, such as radiated power or multi-user scheduling, by providing a safe simulation environment without any actual risks to the real network. Within the ORAN architecture, the radio unit (RU) component is located at the base station and has limited computational power [26]. To compensate for this, the RU relies on the transmission of information from the control unit (CU) component that implements the DT. The CU has more knowledge about the network architecture and user profiles, aided by the cloud data center, and can transmit this information to the RU via a fronthaul wireless link. By implementing a network DT-based SC system over the fronthaul link, the network's efficiency (bandwidth utilization) and security can be significantly enhanced, making it an essential tool for network management. Another key example that is captured by our model is in CI applications, where DTs can improve user experience in various autonomous wireless services, such as robotic surgery, telehealth, driverless vehicles, and industrial robotic manufacturing sites. By creating a virtual representation of the physical system, a DT can simulate and predict its behavior in real-time, enabling proactive decision-making and optimized system performance to enhance the user experience. For instance, a DT can reduce latency and ensure reliable communication in telehealth services, or enhance safety and efficiency in driverless vehicles and industrial robotic manufacturing sites.

The functionalities of the DT are summarized in Fig. 1. The DT is created by an expert agent who analyzes observations from various sensing elements (defined as $(\boldsymbol{s}^t, \boldsymbol{\theta})$) in the wireless environment. These observations are used

---

[1] This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes all the appendices that derive the proofs of Lemmas and Theorems in the paper. Contact christokt@vt.edu for further questions about this work.

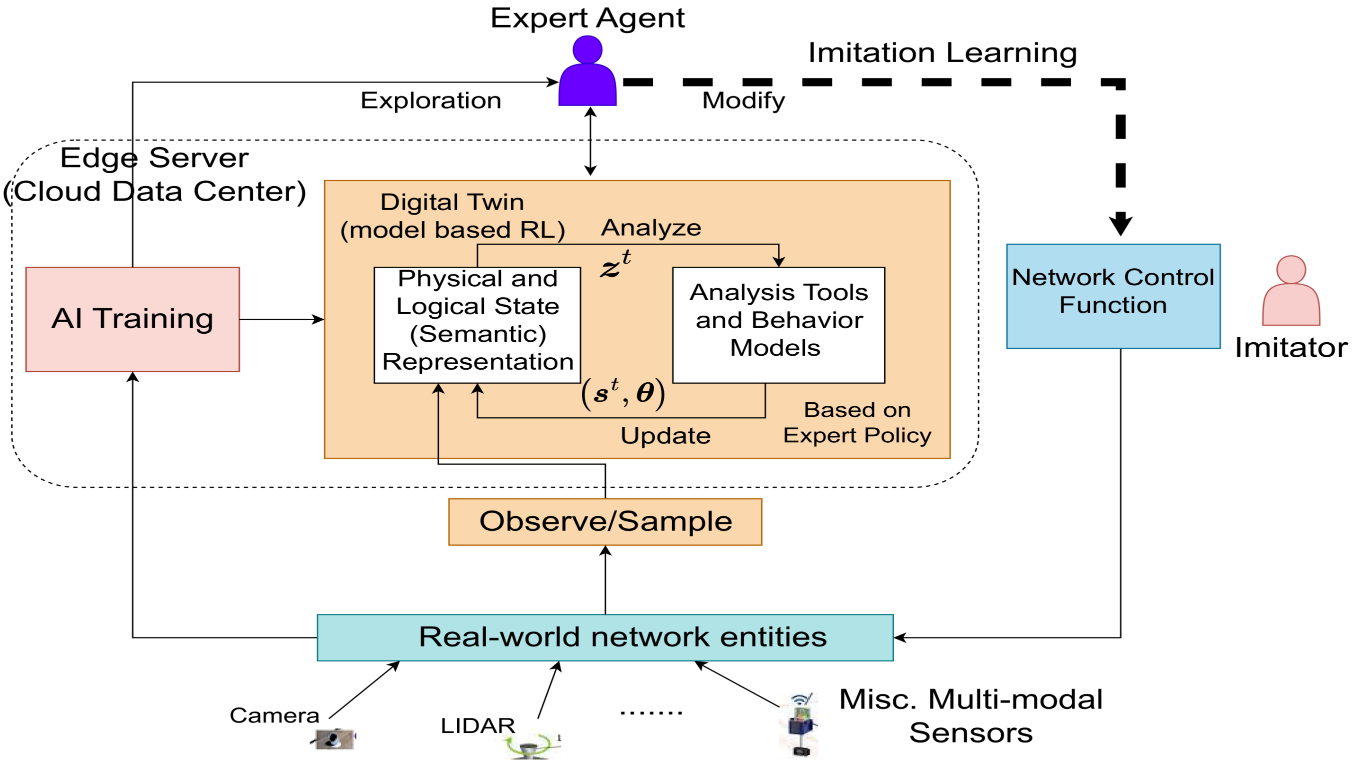

Expert Agent
Imitation Learning
Exploration
Modify
Edge Server
(Cloud Data Center)
Digital Twin
(model based RL)
Analyze
z^t
Physical and Logical State (Semantic) Representation
Analysis Tools and Behavior Models
AI Training
(s^t, θ)
Based on Expert Policy
Update
Network Control Function
Imitator
Observe/Sample
Real-world network entities
Camera
LIDAR
....
Misc. Multi-modal Sensors

Fig. 1: Illustration of our digital twin model showing expert (transmitter that has access to different sensor observations) and imitator agent interactions.

to model the physical and logical state of the network represented by the transition probability $p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{\theta}, \boldsymbol{a}^t)$, which is then used to formulate optimal control decisions $\boldsymbol{a}^t$. The expert agent communicates either a condensed version of the network state or the optimal control decisions defined as $\boldsymbol{z}^t$ to the imitator agent. Subsequently, the imitator extracts the state information and performs the necessary low-level actions. The interactions between expert and imitator agents in the SC system, as described here, can be classified as an example of IL. By utilizing a DT, accurate physical system models and real-time data can be made available to inform decision-making, facilitating effective control decisions. We next explain the details of our DT-based SC system as shown in Fig. 2. The source side of the transmission consists of a resource-constrained node aided by an edge server with significant computational and memory capabilities. The source user and the edge server together act as the expert agent (that implements the DT) to teach an apprentice or imitator agent, serving as the receiver, with no prior knowledge of the expert's data. Here, one naturally wonders whether communicating only the expert actions determined using the policy evaluated by the DT could be sufficient. While communicating only actions may be relevant when the imitator agent has limited capabilities, for cases in which the agent must possess more advanced intelligence, additional information beyond actions may be necessary for effective imitation. In the case of a SC system, building an intelligent imitator agent is crucial. For instance, during channel outages lasting over several communication time slots, the receiver must still generate policy information by leveraging the historical states received over the network and potentially using limited sensing information. Merely communicating the actions does not offer significant information about the observed states at the transmitter, which, in turn, fails to facilitate the environment modeling at the imitator. This, in turn, motivates designing systems that communicate meaningful semantic representations of the states observed at the expert agent to the imitator. Further, the imitator can utilize the potential of generative AI algorithms [27] to reconstruct the network states. Next, we outline the different AI-based components present in the proposed SC system.

A. Causal imitation learning for digital twin-based SC systems

Given the IL model, we implement the SC functionalities as follows. We divide the transmissions into training and communication phases. During training, the source user samples random initial states from a particular distribution, $p(\boldsymbol{s}^0)$. The sampler implements this functionality during training while the states are observable from the environment during the communication phase. Given the initial state, the state transition is determined by the action of the source, and it is sampled using the *expert policy*, $\boldsymbol{a}^t \sim \pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^t, \boldsymbol{\theta})$. The optimal action $\boldsymbol{a}^t$ depends on the state at the current time t and *confounding variables*

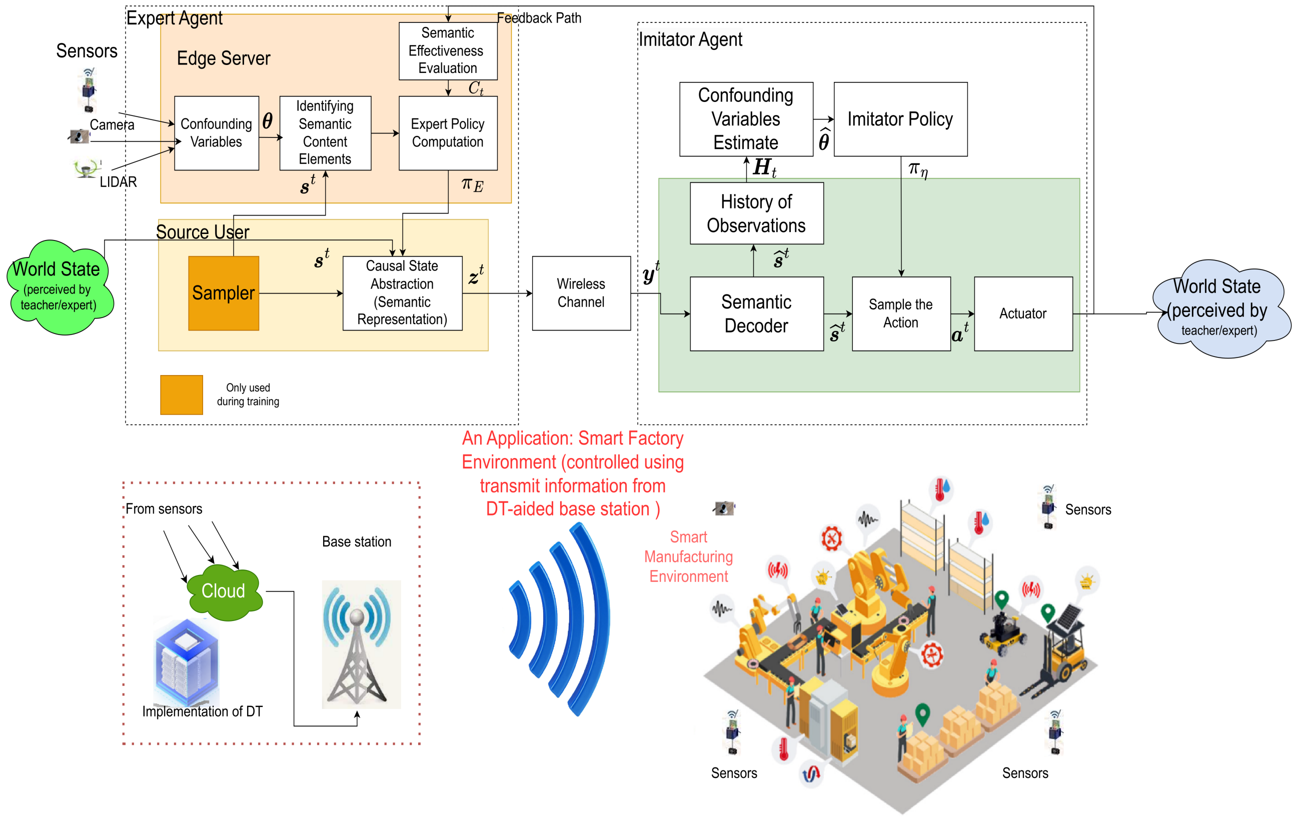

Expert Agent
Feedback Path
Sensors
Edge Server
Semantic Effectiveness Evaluation
C_t
Camera
Confounding Variables
$\boldsymbol{\theta}$
Identifying Semantic Content Elements
Expert Policy Computation
LIDAR
$\boldsymbol{s}^t$
π_E
Source User
World State (perceived by teacher/expert)
Sampler
$\boldsymbol{s}^t$
Causal State Abstraction (Semantic Representation)
$\boldsymbol{z}^t$
Wireless Channel
$\boldsymbol{y}^t$
Only used during training
Imitator Agent
Confounding Variables Estimate
$\widehat{\boldsymbol{\theta}}$
Imitator Policy
$\boldsymbol{H}_t$
π_η
History of Observations
$\widehat{\boldsymbol{s}}^t$
Semantic Decoder
$\widehat{\boldsymbol{s}}^t$
Sample the Action
$\boldsymbol{a}^t$
Actuator
World State (perceived by teacher/expert)
An Application: Smart Factory Environment (controlled using transmit information from DT-aided base station)
From sensors
Cloud
Base station
Implementation of DT
Smart Manufacturing Environment
Sensors
Sensors
Sensors

Fig. 2: Proposed SC system framework based on IL. The figure also illustrates an example application as an autonomous wireless system that includes a DT designed explicitly for a smart manufacturing environment. The application is based on sensing information gathered at the central node and is tailored to the proposed system model.

(from various sensing devices) $\boldsymbol{\theta}$ that could represent background information about the communication environment accessible only to the edge server. The expert agent computes the future state $\boldsymbol{s}^{t+1} \sim p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta})$, using the information from $\boldsymbol{\theta}$ and the current state $\boldsymbol{s}_t$. In our DT-based system, efficient transmission is of utmost importance due to the vast amounts of data the expert agent processes. To address this issue, the expert agent could exploit a *causal discovery AI* method [28] to extract data structure (inherent in the state transition dynamics) while identifying the SCEs. In this setup, a potential question is how to identify the *SCEs* (defined as $\boldsymbol{u}^t$) comprised in any subset of the state $\boldsymbol{s}_i^t$. SCEs are entities in the source data that hold *meaning (semantics)* or significance, aiming to improve the knowledge learned by the imitator. We aim to identify the SCEs in the data and characterize the semantic information in the causal structure. Causality is apt here since it helps with generalization in SC. It allows the communicating nodes to understand the underlying mechanisms and processes that drive the relationships between network variables rather than just observing their associations. This deeper understanding of the underlying causal relationships can help the nodes to make more accurate predictions or control decisions and develop more robust models that can generalize well to new, unseen data. To represent the causal structure that underlies the transition from $\boldsymbol{s}^t$ to $\boldsymbol{s}^{t+1}$, our CSC system employs an MBRL approach. The causal model underlying the state transitions based on Pearl's representation of causality [22] is defined next.

Definition 1. An *SCM* is a collection of elements represented as $< \boldsymbol{U}, \boldsymbol{V}, \mathcal{F}, P(\boldsymbol{U}) >$, where $\boldsymbol{V}$ represents endogenous variables (cause and effect variables), and $\boldsymbol{U}$ represents exogenous variables (random, unknown noise). The set of structural functions $\mathcal{F}$ maps each $f_v \in \mathcal{F}$ to $\boldsymbol{v} \in \boldsymbol{V}$ such that $f_v(Pa_v, \boldsymbol{u}_v)$ determines $\boldsymbol{v}$ where the set of parents $Pa_v \subset \boldsymbol{V}$ and $\boldsymbol{u}_v \subset \boldsymbol{U}$. The exogenous distribution $P(\boldsymbol{U})$ determines the values of $\boldsymbol{U}$, and thus the distribution of endogenous variables $\boldsymbol{V}$.

The apprentice (imitator) can observe only a subset of endogenous variables, partitioned into $\boldsymbol{O}$ and $\boldsymbol{L}$, where $\boldsymbol{O} = \boldsymbol{s}^t$ (in reality, an encoded version communicated by the teacher) and $\boldsymbol{L} = \boldsymbol{\theta}$ (learned by the apprentice). Here, $\boldsymbol{s}_i^t \in \mathcal{R}^D$. The marginal distribution $P(\boldsymbol{O})$ is called the observational distribution. MBRL allows communicating the *causal state abstraction* (i.e., the "semantic representation") $\boldsymbol{z}^t$ at any given time over a wireless channel susceptible to errors. The confounding variables are not transmitted due to the communication and memory constraints associated with the wireless channel and imitator, respectively. The design of the semantic encoder (that maps the states to actual transmit signals) must be done in a way to reduce the amount of data transmitted while maintaining *semantic effectiveness*, defined as a measure of the accuracy of the actions performed using the semantic reconstruction process at the imitator's end.

1) MBRL Model: Next, we formulate the problem of expert policy computation as sequential decision-making in a finite-horizon episodic Markov decision process (MDP), defined as $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, p(\boldsymbol{s}^0), H, \boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ is a vector of confounding variables that the imitator does not have access to and $P(\boldsymbol{\theta})$ is the probability distribution of $\boldsymbol{\theta}$. $\boldsymbol{\theta}$ is fixed within each episode but can vary across different episodes. The $M-$dimensional vector $\boldsymbol{\theta}$ represents the sensing information available from M sensors which may provide different modality information to supplement the optimal policy. The MDP framework comprises a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, a deterministic reward function $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to [0, 1]$ that provides evaluative feedback to the imitator agent, a transition function $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ that describes the distribution of future states, an initial state distribution $p(\boldsymbol{s}^0) \in \Delta(\mathcal{S})$, and a maximum episode length or horizon H. The unit interval constraint for the reward function is adopted for ease of analysis, and our framework can handle general reward functions. The imitator agent cannot access neither the transition function $\mathcal{T}$ nor the reward function $\mathcal{R}$; hence, both are treated as random variables (known as the *environment*). The N-dimensional causal state description $\boldsymbol{s} \in \mathcal{S} \subset \mathcal{R}^N$ is extracted from raw observations and represented as an SCM. Let $\mathcal{M}^*$ be the true MDP that the agent interacts with and attempts to solve over K episodes, with the environment model $\mathcal{E} = (\mathcal{R}^*, \mathcal{T}^*, \boldsymbol{\theta})$. However, the imitator agent's computing capabilities and time constraints limit the accuracy of the true environment that can be learned, resulting in a learned MDP $\widehat{\mathcal{M}}$ (the respective learned environment is $\widehat{\mathcal{E}}$). Each episode involves the agent taking exactly H steps, starting from an initial state $\boldsymbol{s}^0 \sim p(\boldsymbol{s}^0)$. For each $h \in \{1, \cdots, H\}$, the agent observes the current state $\boldsymbol{s}^h \in \mathcal{S}$, selects an action $\boldsymbol{a}^h \sim \pi^{(h)}(\cdot \mid \boldsymbol{s}^h) \in \mathcal{A}$, receives a reward $r_h = \mathcal{R}(\boldsymbol{s}^h, \boldsymbol{a}^h) \in [0, 1]$, and transitions to the next state $\boldsymbol{s}^{h+1} \sim \mathcal{T}(\cdot \mid \boldsymbol{s}^h, \boldsymbol{a}^h) \in \mathcal{S}$. We let $\tau_k = (\boldsymbol{s}^1_{(k)}, \boldsymbol{a}^1_{(k)}, r^1_{(k)}, \cdots, \boldsymbol{s}^H_{(k)}, \boldsymbol{a}^H_{(k)}, r^H_{(k)}, \boldsymbol{s}^{H+1}_{(k)})$ be the random variable that represents the trajectory experienced by the agent in the episode k. $\boldsymbol{H}_k = (\tau_1, \cdots, \tau_{k-1}) \in \mathcal{H}_k$ is the random variable representing the entire history of the agent's interaction with the environment at the start of the episode k. The regret of the RL algorithm over K episodes is: $r_g(K, \pi^{(1)}, \cdots, \pi^{(K)}, \mathcal{E}^*) = \bar{V}^*_{\mathcal{E}^*,1} - \bar{V}^{\pi^{(k)}}_{\mathcal{E}^*,1}$, with the value function $\bar{V}^{\pi^{(k)}}_{\mathcal{E},h} = \mathbb{E}_{a \sim \pi^{(k)}(\cdot|\boldsymbol{s})} \left[Q^{\pi}_{\mathcal{E},h}(s, a) \right]$. The associated action-value function is $Q^{\pi}_{\mathcal{E},h}(s, a) = \mathbb{E}\left[\sum_{h'}^{H} \mathcal{R}(\boldsymbol{s}^{h'}, \boldsymbol{a}^{h'}) \mid \boldsymbol{s}^h = \boldsymbol{s}, \boldsymbol{a}^h = \boldsymbol{a} \right]$, where the expectation integrates over randomness in the action selections and transition dynamics. The Bayesian regret is defined as the expected value of the sum of the episodic regrets: $\bar{r}_g(K, \pi^{(1)}, \cdots, \pi^{(K)}, \mathcal{E}^*) = \mathbb{E}\left[\sum_{k=1}^{K} \left[\bar{V}^*_{\mathcal{E}^*,1} - \bar{V}^{\pi^{(k)}}_{\mathcal{E}^*,1} \right] \right]$. Note that the regret is a random variable due to the uncertainty in $\mathcal{E}$. However, we emphasize again here that using a classical MFRL algorithm is not generalizable to multiple wireless environments, which necessitates us revisiting the SC design approach using causal machine learning (ML) tools as explained in later sections. To incorporate causality, the IL framework can be reformulated using an MBRL approach, as proposed in [29]. Compared to MFRL, MBRL is generally more efficient, making it particularly well-suited for SC. This is because MBRL employs a learned model of the environment $p(\widehat{\boldsymbol{s}}^{t+1} \mid \widehat{\boldsymbol{s}}^t, \boldsymbol{a}^t, \widehat{\boldsymbol{\theta}})$ (part of the imitator's job here), aided by a history of received information, to plan and update its policy. However, similar to other supervised learning methods, MBRL faces the challenge of generalization. The data used for training may not match the one encountered during testing, and even small inaccuracies in the dynamics model or changes in the control policy can lead to exploring new parts of the state space. Thus, we need new models that have strong generalization capabilities in MBRL. To achieve the SC goal of improved resource efficiency, we propose to bridge this gap in the literature for IL by developing generalizable (across several wireless environments) solutions that are causality aware.

The mismatch in knowledge between expert agent ($\mathcal{E}$) and imitator agent ($\widehat{\mathcal{E}}$) that gets learned using MBRL can lead to incorrect inferences due to auto-suggestive delusions [30], i.e., false beliefs generated within one's mind. To address this issue, we present a novel approach by applying a new IL principle, in which we treat actions as causal interventions, as described in [22]. In supervised learning, one can condition or intervene on data using factual and counterfactual error signals, respectively, which helps resolve auto-suggestive delusions. To account for imperfections in the imitator's model, we employ confounding variables in causality. We begin by using SCMs to exploit sparsity in state transitions, which helps generalize to unseen parts of the state-action space. We then extend the IL framework to include some variables $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ observed by the expert but not the imitator. This setup, introduced in [21] for RL, allows modeling of the unknown part of the environment that the imitator learns from its previous state-action sequences. The method in [21] assumes that the imitator has perfect knowledge of $\boldsymbol{s}^t$, which differs from our SC system. Also, their AI approach relies on conventional data-driven techniques, such as VAE (without any structure extraction), that may not be sufficient to achieve generalizable wireless systems. Given the MBRL model, in the next step, we discuss how the expert agent computes its expert policy using the state and confounding variables to which it has access. Additionally, we will explore how a causally invariant semantic language is defined.

2) Expert policy and causally invariant semantic language: The source user does not know the confounding variables because it is a computationally constrained node. In this case, the source user communicates the state information to the edge server. Along with confounding variables, the edge server computes an optimal policy, called the *expert policy*, π_E. π_E is computed to maximize the average reward over time, $\pi_E^* = \underset{\pi_E}{\arg\max}\, \mathbb{E}\left[\mathcal{R}(\boldsymbol{s})\right]$, where $\mathcal{R}(\boldsymbol{s}) = \mathbb{E}_{\boldsymbol{a} \sim \pi_E(\boldsymbol{a}^t | \boldsymbol{s}^t, \boldsymbol{\theta})} \mathcal{R}(\boldsymbol{s}, \boldsymbol{a})$. π_E is communicated to the source user via a highly reliable backhaul connection. Using π_E, the source user samples transitioned state information. Now, the source user must extract the causality between the state transitions which can be computed as the distribution (see [31] for more details on causal representation as posterior factorization):

$$\begin{aligned} \mathcal{P}_1 : \text{Compute } & p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}) = \prod_i \underbrace{p(\boldsymbol{s}_i^{t+1} \mid Pa(\boldsymbol{s}_i^t))}_{\text{Causal graph structure}} \\ & \text{subject to } \boldsymbol{a}^t = \pi_E^*(\boldsymbol{a}^t \mid \boldsymbol{s}^t, \boldsymbol{\theta}). \end{aligned} \tag{1}$$

The factorization of the transition probability in (1) results from causal discovery. This concept of identifying causal features is closely linked to sparsity, which is a form of *inductive bias* that shapes the behavior of the learning agent. An important aspect of this approach is the continuous improvement of the learning policy, which helps us achieve both causal invariance and sparsity [32]. With this in mind, we can ask whether we can leverage the sparsity of transition dynamics to develop a model that generalizes better to unseen parts of the state-action space. Moreover, using the associational, interventional, and counterfactual levels of reasoning, the source user can generalize to distinct data distributions. Training AI models to generalize over the wireless environment, rather than specializing in fitting specific conditions (e.g., a single cell), provides the necessary robustness to deploy a single AI model for a given task across the entire network. This improves scalability and reduces the complexity (in terms of training + distinct AI models) of the AI functionalities and operations. Additionally, we establish the semantic representation that emerges from the proposed semantic language, which is defined below.

Definition 2. A *semantic language* $\mathcal{L} = (\boldsymbol{s}^t, \boldsymbol{z}^t)$, is a mapping from the observed states $\boldsymbol{s}^t$, to their corresponding semantic representation $\boldsymbol{z}^t$, based on the identified SCEs and the causal graph. This mapping is described as the encoder probability distribution $p(\boldsymbol{z}^t \mid \boldsymbol{s}^t, \boldsymbol{s}^{t+1})$.

To achieve the goal of reducing transmission in the SC system, it is optimal to compute the encoder distribution by determining the NN parameters $\boldsymbol{\Psi}$ that result in the maximum received semantic information at the imitator node, as follows:

$$\begin{aligned} \mathcal{P}_2 : \underset{p_{\boldsymbol{\Psi}}(\boldsymbol{z}^t | (\boldsymbol{s}^t, \boldsymbol{s}^{t+1}))}{\arg\max} \quad & \mathbb{I}(\boldsymbol{s}^t; \boldsymbol{y}^t) \\ \text{subject to } & \mathbb{I}(\boldsymbol{s}^t; \boldsymbol{z}^t) \leq \mathbb{I}_b, \end{aligned} \tag{2}$$

where $\mathbb{I}_b$ represents the bandwidth limitations of the wireless channel. The information measures above correspond to the novel semantic information measures detailed in Section III (see equation (7) and (9)). The abstracted state information $\boldsymbol{z}^t$ depends on two factors: 1) the IIT, which helps identify SCEs and rigorously formulate semantic information, and 2) semantic awareness, which is determined by user-defined key performance indicators (KPIs) such as QoE. We next define the QoE metric below.

3) Semantic effectiveness as QoE metric: On the imitator's side, the semantic decoder module should be designed to maximize semantic effectiveness defined based on our work [33]. We thus introduce the metric C_t, which captures the causal impact of the expert agent's message (via the imitator's actions) as observed through a channel with a response characterized using $p(\boldsymbol{y}^t \mid \boldsymbol{z}^t)$ (this distribution could capture the fading and interference in the wireless environment). In other words, C_t measures the semantic effectiveness (inversely proportional to C_t) of the transmitted message to the end-user. We define $C_t(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t)$ in (3), where $\text{KLD}(p \;||\; q)$ represents the KL divergence between p and q. $\pi_\eta(\boldsymbol{a}^t \mid \widehat{\boldsymbol{s}}^t, \widehat{\boldsymbol{\theta}}^t)$ is the sub-optimal policy computed by the imitator using the limited information it received over the air. $p(\widehat{\boldsymbol{s}}^t \mid \boldsymbol{y}^t)$ denotes the semantic decoder distribution. The decoder component can thus be designed using the game-theoretic framework we developed in [33]. The second term in the quantity C_t is an average measure across the channel realization. Since the semantic effectiveness metric in (3) captures the channel effect on the transmitted semantics, the learned policy (for the imitator) parameters is robust to the channel errors (in terms of the best the imitator can do). The semantic effectiveness measure is the distance in the semantic space (distance between policy distributions) and not a conventional metric like Euclidean distance. However, computing the semantic effectiveness at the imitator requires it to know π_E. This necessitates communicating the imitator policy to the edge server via the backhaul. Subsequently, the edge server assesses C_t and transmits it to the imitator, which then adapts its policy based on the quality of semantic effectiveness. The communication link between the imitator and the edge server is assumed to be highly reliable compared to the wireless connection to the source user.

Having defined semantic effectiveness, we look at how to define the reward signal mentioned earlier. The agent may encounter extremely sparse or entirely absent extrinsic rewards in numerous real-world situations [34]. In these cases, the intrinsic reward signal based on semantic effectiveness can be used to facilitate collaborative exploration and communication skill learning (specifically, semantic encoder and decoder) between the expert and imitator agents. This skill becomes helpful in subsequent communication

$$C_t(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t) = \text{KLD}\left(\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^t, \boldsymbol{\theta}) \mid\mid \sum_{\boldsymbol{y}^t,\boldsymbol{z}^t} p(\boldsymbol{z}^t \mid \boldsymbol{s}^t)p(\boldsymbol{y}^t \mid \boldsymbol{z}^t)p(\widehat{\boldsymbol{s}}^t \mid \boldsymbol{y}^t)p(\widehat{\boldsymbol{\theta}}^t \mid \boldsymbol{H}_t)\pi_\eta(\boldsymbol{a}^t \mid \widehat{\boldsymbol{s}}^t, \widehat{\boldsymbol{\theta}}^t)\right). \tag{3}$$

phases. Motivated by these considerations, we can define $r_t = r_t^i + r_t^e$, where r_t^e is the extrinsic reward which is mostly (if not always) zero and the intrinsic reward, $r_t^i = \frac{1}{1+C_t(\boldsymbol{s}^t,\widehat{\boldsymbol{s}}^t)}$. By incorporating semantic effectiveness when computing π_E, the prediction of state transitions and causal structure is improved. This leads to a better design of the semantic encoder on the expert side, which adjusts its transmission strategy to maximize the semantic information extracted by the receiver. Having outlined the different IL components for CSC, we next develop a novel semantic information measure that helps us identify the SCEs and that is critical in formulating various objective functions ($\mathcal{P}_1, \mathcal{P}_2$ in (1) and (2)) elaborated in Section IV.

III. CHARACTERIZING SEMANTIC CONCEPTS VIA INTEGRATED INFORMATION THEORY

An SCM with respect to the formulations in IIT can be described as a stochastic system $\mathcal{U} = \{U_1, \cdots, U_n\}$ of n interacting units (atomic) with state space $\Omega_{\mathcal{U}} = \prod_i \Omega_{U_i}$, where $\Omega_{U_i} \subset \mathcal{R}$, and the current state is $\boldsymbol{u} \in \Omega_{\mathcal{U}}$. Each network state $\boldsymbol{s}_i^t$ in (1) that is observed at DT can be composed of multiple such atomic units, whose state space we denote as $\Omega_{s_i} \subset \Omega_{\mathcal{U}}$. Our system is updated in discrete steps such that the state space $\Omega_{\mathcal{U}}$ is finite, and the individual random variables $U_i \in \mathcal{U}$ are conditionally independent given the preceding state of $\mathcal{U}$: $p(\boldsymbol{u}^{t+1} \mid \boldsymbol{u}^t) = \prod_i p(u_i^{t+1} \mid \boldsymbol{u}^t)$. The IIT-based concept of semantic information draws inspiration from the theory of consciousness measures proposed by [35]. This concept is constructed on the fundamental principles of intrinsic information, information integration, and exclusion, which we elaborate on next.

A. Intrinsic Information for State Abstraction

In IIT, *intrinsic information* refers to the inherent cause-and-effect structure within a system that produces the particular set of observed states and transitions. Put simply, it is the semantic information built into our DT-based system that leads to the shift from one state ($\boldsymbol{s}^t$) to the next ($\boldsymbol{s}^{t+1}$). Moreover, in IIT, information present in $\boldsymbol{s}^t$ is considered to be causal only if it has selective causes and selective effects within the system. This selective nature of information distinguishes it from mere correlation or statistical association between variables (which is the extrinsic notion of information used in Shannon's theory of communication), which may not be causal in nature. We now proceed to analytically define intrinsic information. The amount of information that the current state $\boldsymbol{s}_i^t$ (which represents any subset of causal variables part of the SCM) specifies about the past, i.e., its cause information $\mathbb{I}_c$, is measured as the distance between the cause repertoire $p(\boldsymbol{s}_i^{t-1} \mid \boldsymbol{s}_i^t)$ and the unconstrained past repertoire $p(\boldsymbol{s}_i^{t-1})$, and is defined as follows:

$$\mathbb{I}_c(\boldsymbol{s}_i^{t-1} \mid \boldsymbol{s}_i^t) = \mathbb{D}\left(p\left(\boldsymbol{s}_i^{t-1} \mid \boldsymbol{s}_i^t\right) || p\left(\boldsymbol{s}_i^{t-1}\right)\right). \tag{4}$$

Just like cause information $\mathbb{I}_c$, the effect information $\mathbb{I}_e$ of $\boldsymbol{s}_i^t$ is quantified as the distance between the effect repertoire of $\boldsymbol{s}_i^t$ and the unconstrained future repertoire $p(\boldsymbol{s}_i^{t+1})$ and written as

$$\mathbb{I}_e(\boldsymbol{s}_i^{t+1} \mid \boldsymbol{s}_i^t) = \mathbb{D}\left(p\left(\boldsymbol{s}_i^{t+1} \mid \boldsymbol{s}_i^t\right) || p\left(\boldsymbol{s}_i^{t+1}\right)\right). \tag{5}$$

Next, we define an appropriate measure for $\mathbb{D}$. The Kullback-Leibler divergence (KLD) is a useful metric from classical information theory, but it is not a true metric (it is not symmetric) and it is unbounded. Additionally, KLD only measures how "sharp" a distribution is compared to another, without taking into account whether some states of the system are closer than others (in the sense of the Euclidean distance metric). Compared to the literature in IIT [23], a more appropriate measure that aligns better with the IIT notion of information as "differences that make a difference" is the *earth mover's distance (EMD) or Wasserstein distance*. This distance is formulated in the context of optimal transport problems as follows:

Definition 3. For any two probability measures μ_s, μ_t, the optimal cost of transporting from μ_s to μ_t can be formulated as the *Wasserstein distance* $W_p^p(\mu_s, \mu_t) = \min\limits_{\gamma\in\mathcal{P}} \int_{\Omega_x \times \Omega_y} \|x-y\|^p \gamma(x,y)dxdy$, where Ω_x and Ω_y are the domain of μ_s and μ_t, respectively. $\mathcal{P}$ is the set of joint probability distributions.

$W_p^p(\mu_s, \mu_t)$ is symmetric and bounded. Finally, having calculated $\mathbb{I}_c$ and $\mathbb{I}_e$, the total amount of cause-effect information $\mathbb{I}_{ce}$ specified by $\boldsymbol{s}_i^t$ over the purview defined as $\{\boldsymbol{s}_i^{t+1}, \boldsymbol{s}_i^{t-1}\}$, is:

$$\mathbb{I}_{ce}(\boldsymbol{s}_i^{t+1}, \boldsymbol{s}_i^{t-1} \mid \boldsymbol{s}_i^t) = \min(\mathbb{I}_c, \mathbb{I}_e). \tag{6}$$

From the intrinsic information perspective, each causal mechanism (includes the causes and effects of any state) in the system acts as an information bottleneck. This means that its cause information only exists for the system to the extent that it also specifies effect information, and vice versa.

Lemma 1. *We define the intrinsic information learned by the expert and imitator as $\mathbb{I}_{ce}^E$ and $\mathbb{I}_{ce}^\eta$, respectively. In the presence of confounding variables $\boldsymbol{\theta}$, the imitator learns less intrinsic information, $\mathbb{I}_{ce}^\eta < \mathbb{I}_{ce}^E$, compared to the expert. Under the assumption of perfect $\boldsymbol{s}^t$ being extracted by the imitator, the difference $|\mathbb{I}_{ce}^\eta - \mathbb{I}_{ce}^E|$ can be shown to be strictly smaller than ϵ when $D_{TV}\left(p(\boldsymbol{\theta} \mid \boldsymbol{H}_t), \delta_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right) < \epsilon$, where D_{TV} is the total variation distance (see Appendix A) and $\delta_{\boldsymbol{\theta}^0}(\boldsymbol{\theta})$ is the Dirac-delta distribution, with $\boldsymbol{\theta}^0$ the true value.*

Proof: See Appendix B. ■

Lemma 1 emphasizes the significance of precise modeling of the confounding variables (known at the DT side) at the imitator to extract the relevant semantics. Moreover, we introduce the following concept, which enables the DT-enabled transmitter to generate a semantic representation that captures only the relevant and distinctive causal states with nonzero intrinsic information.

Definition 4. We define the *causal-invariant state abstraction* as the $D-$dimensional embedding of the states, $\phi_i : \boldsymbol{s} \rightarrow \mathcal{R}^D$. ϕ_i is causal-invariant if $\forall \boldsymbol{s}_c, \boldsymbol{s}_e, \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$, $\boldsymbol{a} \in \mathcal{A}$, $\phi_i(\boldsymbol{s}_1) = \phi_i(\boldsymbol{s}_2)$ if and only if the cause and effect set are same for both $\boldsymbol{s}_1, \boldsymbol{s}_2$ while satisfying $P(\boldsymbol{s}_1 \mid \boldsymbol{s}_c, \boldsymbol{a}) = P(\boldsymbol{s}_2 \mid \boldsymbol{s}_c, \boldsymbol{a}), P(\boldsymbol{s}_e \mid \boldsymbol{s}_1, \boldsymbol{a}) = P(\boldsymbol{s}_e \mid \boldsymbol{s}_2, \boldsymbol{a})$.

From Definition 4 we can conclude the following: if $\phi_i(\boldsymbol{s}_1) = \phi_i(\boldsymbol{s}_2)$, then the intrinsic information conveyed by $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ is the same. In other words, from an SC perspective, state abstraction enables a unique representation for states that share comparable cause-and-effect repertoires. This leads to a reduction in the amount of semantic information conveyed compared to systems that do not employ state abstraction. The concept of state abstraction presented here differs significantly from the traditional ML definitions such as those in [32] and [36]. Next, using the intrinsic information concept, we analyze the information content provided by a state $\boldsymbol{s}_i^t$ composed of several SCEs $U_j \in \mathcal{U}$.

B. Information Integration (via Compositionality, for Identifying Semantic Content Elements)

At the level of an individual causal mechanism (state + its cause and effect mechanisms), the integration postulate [35] states that only mechanisms that specify integrated information can contribute to consciousness. Inspired by this, *integrated information* for an SCM is information that is generated by the whole mechanism beyond the information generated by its parts, meaning that the mechanism is irreducible with respect to information. Similar to cause-effect information, integrated information (denoted by $\mathbb{I}_\phi$) is calculated as the Wasserstein distance between two probability distributions.

To compute the quantity of information that is integrated across the m parts of a system, we can partition the system into m parts such as $\boldsymbol{M}_1^t, \boldsymbol{M}_2^t, \cdots, \boldsymbol{M}_m^t$. This partition $p_k \in \mathcal{P}_S$ (the set of all partitions of $\mathcal{S}$) is defined such that $\cup_i \boldsymbol{M}_i^t = \boldsymbol{s}_i^t$ and $\boldsymbol{M}_i^t \cap \boldsymbol{M}_j^t = \emptyset$. The measure of integrated information with respect to the cause and effect mechanisms can be expressed as (7). The integrated information for partition p_k will be:

$$\mathbb{I}_\phi^{p_k} = \min(\mathbb{I}_{\phi,c}^{p_k}, \mathbb{I}_{\phi,e}^{p_k}). \tag{8}$$

$\mathbb{I}_\phi$ satisfies $0 \leq \mathbb{I}_\phi \leq \min\left(\mathbb{I}_c\left(\boldsymbol{s}_i^{t-1}; \boldsymbol{s}_i^t\right), \mathbb{I}_e\left(\boldsymbol{s}_i^{t+1}; \boldsymbol{s}_i^t\right)\right)$. Thus, the SCM is reducible if at least one partition $p_k \in \mathcal{P}_S$ makes no difference to the cause or effect probability, i.e., $\mathbb{I}_\phi^{p_k} = 0$. If it is zero, then the partition does not contribute any shared information to the system as a whole. Next, we define the integrated information of an SCM as given by it irreducibility over its minimum partition $p_k \in \mathcal{P}_S$:

$$\begin{aligned} &\mathbb{I}_\phi = \mathbb{I}_\phi^{p_k^*}, \\ &\text{s.t. } p_k^* = \arg\min_{p_k} \frac{\mathbb{I}_\phi^{p_k}}{\max\limits_{p_i \in \mathcal{P}_S} \mathbb{I}_\phi^{p_i}}. \end{aligned} \tag{9}$$

The normalization above is over the maximum possible value that $\mathbb{I}_\phi^{p_k}$ could take for any partition.

Having defined the integrated information measure, we consider the IL setup. The imitator relies on the probability distributions obtained from the received semantic information to measure its $\mathbb{I}_\phi$. To extract the maximum information, the decoding must be performed optimally using the "true" conditional distribution, which is also a function of confounding variables.

$$p(\boldsymbol{s}_i^t, \boldsymbol{\theta} \mid \boldsymbol{s}_i^{t-1}) = p(\boldsymbol{M}_1^t, \cdots, \boldsymbol{M}_m^t, \boldsymbol{\theta} \mid \boldsymbol{M}_1^{t-1}, \cdots, \boldsymbol{M}_m^{t-1}). \tag{10}$$

To decode $\boldsymbol{s}_i^{t-1}$, the imitator uses a "false" conditional distribution, $q(\widehat{\boldsymbol{s}}_i^t, \widehat{\boldsymbol{\theta}} \mid \boldsymbol{s}_i^{t-1})$ (false, since $\boldsymbol{\theta}$ is unknown at the imitator), and hence it is an instance of "mismatched" decoding. $\widehat{\boldsymbol{s}}_i^t$ is the received semantic information. To quantify integrated information, we specifically consider the mismatched decoding that uses the "partitioned" probability distribution $q(\widehat{\boldsymbol{s}}_i^t, \widehat{\boldsymbol{\theta}} \mid \boldsymbol{s}_i^{t-1}) = p_\eta(\boldsymbol{\theta} \mid \boldsymbol{H}_t) p(\boldsymbol{s}_i^t \mid \boldsymbol{s}_i^{t-1}, \boldsymbol{\theta})$. Similarly, we can define the probability distribution $q(\widehat{\boldsymbol{s}}_i^t, \widehat{\boldsymbol{\theta}} \mid \boldsymbol{s}_i^{t+1})$. The imitator should learn (see Section IV) these cause and effect probabilities to extract the semantic information (not just look to reconstruct the current state $\boldsymbol{s}_i^t$ from $\boldsymbol{y}^t$). The corresponding learned integrated information, defined as $\mathbb{I}_\phi^\eta$, is called as the extrinsic information (measured using the imitator's observations, which are the received signals). Next, we quantitatively analyze the error in $\mathbb{I}_\phi$ between expert and imitator agents for our DT-based SC system.

Lemma 2. *The error between the true integrated information and the extrinsic information learned by the imitator can quantified as the bias of the learned estimator of cause and effect entropy at the imitator. Here,* $\mathbb{H}$ *refers to Shannon's entropy, which is a measure of uncertainty.*

$$\begin{aligned} \mathbb{I}_{\phi,c} - \mathbb{I}_{\phi,c}^e = \mathbb{E}_q\left(\mathbb{H}\left(\boldsymbol{s}_i^{t-1} \mid \boldsymbol{s}_i^t, \boldsymbol{\theta}\right) - \mathbb{H}\left(\boldsymbol{s}_i^{t-1} \mid \boldsymbol{s}_i^t, \boldsymbol{\theta}\right)\right). \\ \mathbb{I}_{\phi,e} - \mathbb{I}_{\phi,e}^e = \mathbb{E}_q\left(\mathbb{H}\left(\boldsymbol{s}_i^{t+1} \mid \boldsymbol{s}_i^t, \boldsymbol{\theta}\right) - \mathbb{H}\left(\boldsymbol{s}_i^{t+1} \mid \boldsymbol{s}_i^t, \boldsymbol{\theta}\right)\right). \end{aligned} \tag{11}$$

Proof: See Appendix C. ■

Lemma 2 means that the error in integrated information is the minimum of the estimator (of entropy, with respect to the mismatched probability distribution q) bias among cause and effect transitions. In other words, the error in semantic information at the imitator is the same as the error in the transition ($\mathcal{T}$, part of $\mathcal{E}$) modeling.

C. Semantic Concepts using Exclusion

A maximally irreducible cause-effect repertoire (MICE) is specified by a subset of elements, referred to as a

$$\mathbb{I}^{p_k}_{\phi,c} = \mathbb{I}_c(\boldsymbol{s}_i^{t-1}; \boldsymbol{s}_i^{t}) - \sum_j \mathbb{I}_c(\boldsymbol{M}_j^{t-1}; \boldsymbol{M}_j^{t}), \quad \mathbb{I}^{p_k}_{\phi,e} = \mathbb{I}_e(\boldsymbol{s}_i^{t+1}; \boldsymbol{s}_i^{t}) - \sum_j \mathbb{I}_e(\boldsymbol{M}_j^{t+1}; \boldsymbol{M}_j^{t}). \tag{7}$$

concept. To find the core cause of any state $\boldsymbol{s}_i^t$, we compute $\mathbb{I}_c(\boldsymbol{M}_i^{t-1} \mid \boldsymbol{s}_i^t)$ for all possible partitions $\boldsymbol{M}_i$, and take the maximum among them. That is, $\mathbb{I}_c^{\max}(\boldsymbol{s}_i^t) = \max_{\boldsymbol{M}_i} \mathbb{I}_c(\boldsymbol{M}_i^{t-1} \mid \boldsymbol{s}_i^t)$. Similarly, we can compute the maximum $\mathbb{I}_e^{\max}(\boldsymbol{s}_i^t)$ in the effect direction. The partitions that represent the maximum of those values are $\boldsymbol{m}_c^* = \arg\max_{\boldsymbol{M}_i} \mathbb{I}_c(\boldsymbol{M}_i^{t-1} \mid \boldsymbol{s}_i^t)$ and $\boldsymbol{m}_e^* = \arg\max_{\boldsymbol{M}_i} \mathbb{I}_e(\boldsymbol{M}_i^{t-1} \mid \boldsymbol{s}_i^t)$. The core cause and effect of mechanism $\boldsymbol{s}_i^t$ are $\boldsymbol{m}_c$ and $\boldsymbol{m}_e$, respectively. Together, they specify the "what" of the concept of $\boldsymbol{s}_i^t$. A mechanism that specifies MICE constitutes a concept, or more specifically, a core concept. A concept $\boldsymbol{s}_i^t$ is composed of several atomic units U_i, referred as *SCEs*.

Definition 5. A *semantic content element* can be formally defined as an atomic mechanism, with possible minimum integrated information among all partitions p_i.

1) Sub-Optimal Strategy for SCE Identification: To correctly compute (7), it is necessary to know the true transition probabilities that are used in the evaluation of intrinsic information. However, since the true transition dynamics are unknown, we must rely on estimating an empirical distribution from the available data. As a result, the approach outlined here represents a suboptimal scheme for identifying the SCEs. We define the input data as a sequence of entities, each of which is represented by a $d-$dimensional value, with each entity being $\boldsymbol{e}_i \in \mathcal{U}$. To extract the entities, we can utilize the β-VAE [37] which is a promising approach (the discussion of which is beyond the scope here) to extract statistically independent entities present in the data. We store the entities for further processing. Further, we define a function f_ϕ to compute the integrated information for a given subset of entities, $\boldsymbol{u} \subseteq \mathcal{U}$. The function should take the subset of entities $\boldsymbol{u}$ as input and return the integrated information, $\mathbb{I}_{\phi,\boldsymbol{u}}$ as in (8). To calculate the integrated information, the procedure utilizes the analytical approach outlined in Section III-B. First, we define a function to generate all possible partitions of a given subset of entities. The function should take the subset of entities as input and return a list of all possible partitions. Further, we iterate over all partitions and compute the integrated information for each partition as in (7) by calling the function f_ϕ. This returns the identified SCEs, which consist of subsets of entities with non-zero integrated information and their corresponding partitions. Among the partitions with nonzero integrated information $\mathbb{I}^{p_k}_{\phi,\boldsymbol{u}}$, we compute the maximally irreducible partition $\boldsymbol{s}_i^t$. The concept as captured by $\boldsymbol{s}_i^t$, and its constituent atomic units, U_i, form the SCEs. The computation of the integrated information and generation of partitions can be computationally expensive for large datasets, so it may be necessary to optimize these functions for performance.

The amount of integrated information generated by concept $\boldsymbol{s}^t$ is the minimum between past and future:

$$\mathbb{I}_\phi^{\max}(\boldsymbol{s}_i^t) = \min\left(\mathbb{I}_c^{\max}(\boldsymbol{m}_c^* \mid \boldsymbol{s}_i^t), \mathbb{I}_e^{\max}(\boldsymbol{m}_e^* \mid \boldsymbol{s}_i^t)\right). \tag{12}$$

The set of all concepts within an SCM constitutes its conceptual structure, which can be represented in concept space. Concept space is a high-dimensional space, with one axis for each possible past and future state of the system. In this space, each concept is represented as a "star," with its coordinates given by the probability of past and future states in its cause-effect repertoire. The size of the star represents the value of its $I_\phi^{\max}(\boldsymbol{s}_i^t)$. If $I_\phi^{\max}(\boldsymbol{s}_i^t)$ is zero, the concept does not exist and if it's small, the concept exist to a minimal extent. The observed state $\boldsymbol{s}^t$ at the DT-based expert agent comprises various semantic concepts, each designated as $\boldsymbol{s}_i^t$. Even though previous literature, e.g., [38], has briefly mentioned the data observed at the transmitter as composed of multiple concepts, purely as a probabilistic transformation, the proposed rigorous formulation based on theory of consciousness described in this section is the first of its kind. Up to this point, we have identified the SCEs present in the data and analyzed the information they contain. Our next step is to comprehend the associations between the different SCEs and how they collectively form a semantic concept, ultimately creating a causal graph that gives structure to the data.

D. Defining Causal Relationships via Topological Perspective

For simplicity, we remove the dependence on t in this subsection. Any concept $\boldsymbol{m} \subseteq \boldsymbol{s}$ with $\mathbb{I}_\phi(\boldsymbol{m}) > 0$ specifies a candidate distinction $d(\boldsymbol{m}) = (\boldsymbol{m}, \boldsymbol{z}^*, \mathbb{I}_\phi(\boldsymbol{m}))$ within the system $\mathcal{S}$ in state $\boldsymbol{s}$. $\boldsymbol{z}^* = \{\boldsymbol{z}_c^*, \boldsymbol{z}_e^*\}$ represents the maximal cause-effect pair for $\boldsymbol{m}$. For any state $\boldsymbol{s}$, the causal distinctions that represent its subsets with nonzero integrated information are defined as (13), where $d(\boldsymbol{m})$ is the set of nonzero integrated information for $\boldsymbol{m}$ and $\boldsymbol{s}' = \{\boldsymbol{s}_c', \boldsymbol{s}_e'\}$ is the maximal cause-effect pair for the system as a whole. *Causal relationships* refer to the way in which the causes and/or effects of a set of distinctions $\mathcal{D}(\mathcal{T}_S, \boldsymbol{s})$ in a complex system overlap. In the same way that a distinction $\boldsymbol{m}$ identifies which units/states constitute a cause and the resulting effect, a relationship identifies which units/states correspond to which units/states among the purviews of a set of distinctions. These relationships reflect how the cause-and-effect power of its concepts is interconnected within the complex system. In short, they help to attribute *meaning (semantics)* to the state $\boldsymbol{s}$. Understanding causal relationships is crucial, as these models enable DT-based wireless systems to distill knowledge and experiences,

$$\mathcal{D}(\mathcal{T}_S, \boldsymbol{s}) = \{d(\boldsymbol{m}) : \boldsymbol{m} \subseteq \boldsymbol{s}, \mathbb{I}_{\phi}(\boldsymbol{m}) > 0, \boldsymbol{z}_c(\boldsymbol{m}) \subseteq \boldsymbol{s}'_c, \boldsymbol{z}_e(\boldsymbol{m}) \subseteq \boldsymbol{s}'_e\}, \tag{13}$$

similar to the human mind, to make reliable predictions, generalizations, and analogies. These qualities are essential for enabling compositional and counterfactual reasoning, active intervention in the world to test hypotheses, and the ability to articulate one's understanding to others. We now look at how to quantify the causal relationships for our system. The degree of irreducibility resulting from this binding of cause-and-effect power is measured by the irreducibility of the relationships, denoted by $(\mathbb{I}_{\phi_r} > 0)$. In the presence of confounding variables, the learned causal distinctions by the imitator follows the relation, $\mathcal{D}(\mathcal{T}_S, \boldsymbol{s} \mid \widehat{\boldsymbol{\theta}}) \subseteq \mathcal{D}(\mathcal{T}_S, \boldsymbol{s} \mid \boldsymbol{\theta})$. As in [24], the relations between any two distinctions can be defined as follows. For any state $\boldsymbol{s}$, consider two causal distinctions $d(\boldsymbol{m}_1)$ and $d(\boldsymbol{m}_2)$, that have a possibly non-empty overlapping atomic units as their constituents. The relation between $d(\boldsymbol{m}_1)$ and $d(\boldsymbol{m}_2)$ can be defined as the maximally irreducible subset that is common to both of them. For each candidate overlap, the intrinsic difference is assessed at the maximum information partition (MIP) for each partition, and summed together, which represents the irreducibility measure. If we consider a specific set of distinctions $\boldsymbol{d} \subseteq \mathcal{D}(\mathcal{T}_S, s)$, it is possible that there are multiple sets of causes and/or effects, denoted as z, such that

$$z : z \cap \{z_c^*(d), z_e^*(d)\} \neq \emptyset \, \forall d \in \boldsymbol{d}, \cap_{z \in \boldsymbol{z}} z \neq \emptyset, |\boldsymbol{z}| > 1, \tag{14}$$

with maximal overlap (called the "faces" of the relation), $o^*(z) = \cap_{z \in \boldsymbol{z}} z \neq \emptyset$. A relation $r(\boldsymbol{d})$ thus consists of a set of distinctions $\boldsymbol{d} \subseteq \mathcal{D}(\mathcal{T}_S, \boldsymbol{s})$, with an associated set of faces $f(\boldsymbol{d}) = \{f(\boldsymbol{z})\}_{\boldsymbol{d}}$ and irreducibility measure $\phi_r > 0$ and can be written as $w(\boldsymbol{d}) = (\boldsymbol{d}, f(\boldsymbol{d}), \phi_r)$. $w(\boldsymbol{d})$ can be topologically represented as shown below.

Lemma 3. *The structure of causes and effects in an SCM can be represented using an* abstract simplical complex *[39], in which causes and effects are the vertices and the relations are simplices; relations between pairs of causes and effects are represented as 2-simplices (edges); relations between trios are represented as 3-simplices (faces) and so on.*

The Lemma 3, whose proof follows directly from its definition states that a relation $r(\boldsymbol{d})$ can be topologically characterized as a $|\boldsymbol{d}|$-simplex, where $|\boldsymbol{d}|$ is the order of a relation. Relations between two purviews are known as 2-relations, relations involving three purviews are referred to as 3-relations and so on. More details on simplex is provided in Appendix A. The degree of a relation is defined as the number of causal distinctions present in $\boldsymbol{d}$. The "faces" of the relation is same as the concept of faces in simplical homology. Further, we look at defining the irreducibility measure ϕ_r. For any relation $w \in \mathcal{W}$, the irreducibility can be defined as

$$\phi(\mathcal{R} \mid \mathcal{O}) = \sum_i d_{\mathcal{O}}(\mathcal{S}_i), \tag{15}$$

where $\mathcal{O}$ represents a candidate overlap between two causal distinctions, $\mathcal{S}_i$ represents a particular partition of $\mathcal{O}$ and $d_{\mathcal{O}}(\mathcal{S}_i) = D(p(\mathcal{O}) \mid\mid p(\mathcal{S}_i))$. Hence, $d_{\mathcal{O}}(\mathcal{S}_i)$ represents the information the partition $\mathcal{S}_i$ provides within a specific overlap $\mathcal{O}$. Finally, the maximally irreducible overlap is the $\mathcal{O}$ that maximizes ϕ:

$$\begin{aligned} \mathcal{O}_{\text{max}} &= \arg\max_{\mathcal{O}} \phi(\mathcal{S} \mid \mathcal{O}). \\ \phi_R &= \phi(\mathcal{S} \mid \mathcal{O}_{\text{max}}). \end{aligned} \tag{16}$$

We now formally define semantics using the abstract simplical complex-based definition of causal relations.

1) Topological characterization of semantics: Semantics are represented using the relations defined in the previous subsection, as follows:

Theorem 1. *An abstract cell complex $\mathcal{C} = (\mathcal{D}, \prec_w, \dim)$ is a set $\mathcal{D}$ of abstract SCEs, which are the cells here, equipped with a bounding relation $\prec_w$ and a dimension function assigning to each $\boldsymbol{d} \in \mathcal{D}$ a non-negative integer (*$\dim$*), satisfying the following properties*

- *if $\boldsymbol{d}_1 \prec_w \boldsymbol{d}_2$ and $\boldsymbol{d}_2 \prec_w \boldsymbol{d}_3$, then $\boldsymbol{d}_1 \prec_w \boldsymbol{d}_3$ (called as the* transitivity property*)*
- *if $\boldsymbol{d}_1 \prec_w \boldsymbol{d}_2$, then $\dim(\boldsymbol{d}_1) < \dim(\boldsymbol{d}_2)$ (called as the* monotonicity property*).*

Proof: The proof is given in Appendix D. ■

Further, we look at what Theorem 1 entails for the DT-based SC system. $\boldsymbol{d} \in \mathcal{D}$ represents a causal mechanism (concept) with non-zero integrated information. The dimension of $\boldsymbol{d}$, denoted $\dim(\boldsymbol{d})$, represents the number of distinct SCEs part of $\boldsymbol{d}$. The relation between two causal mechanisms $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ can be defined as the causal relationship (the overlap in the causal distinctions) between them. The usefulness of the Theorem 1 for our DT-based SC system can be two-fold: Firstly, just like in natural language, where syntax refers to the grammatical forms used to express content and semantics refers to the meaning attributed to those syntactic expressions, in communication between intelligent agents, the encoded representation refers to the syntax part [33]. The transmit encoders, which form a component of the semantic language used to convey meaning, can be connected to the specific problem formulation. The design of the AI architecture employed for this purpose may be constrained by the complexity of the transmitter, such as the maximum number of parameters or layers it can support. Regardless of this dependency on language, if the receiver can extract the semantic content defined using a topological construct such as a cell complex, it should be able to communicate seamlessly with multiple transmitters without requiring any protocol changes. In other words, regardless of the encoded representations used to convey the states observed at DT, the topological characterization of semantics acts as a "bridge" [40] that unifies different semantic contents and facilitates communication between intelligent agents. The abstract cell complex structures in Theorem 1 can be seen as "universal translators" and bridges across different knowledge representations. Additionally, the topological characterization presented in Theorem 1 allows a rigorous formulation of semantic metrics, such as similarity and reliability, as defined in Section IV-A, compared to the current state of the art. Having defined SCEs present in the observed states, we further move on to discovering the causal structure in the data using novel AI tools.

IV. Causal Discovery and Inference via Imitation Learning and model based reinforcement learning

In the context of DT-based SC, the causal dynamics model $p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta})$ cannot be accessed by the imitator due to hidden confounding variables (known only to DT). Instead, the imitator can only estimate the transition distribution based on its interactions with the expert. To enable the imitator to make better control decisions, it must have the ability to develop representations of the world based on past experiences (received semantics from the DT node), which allow for generalization to novel situations. To achieve this, the transmitter design must incorporate causal discovery, state transition probability, and semantic encoding such that the QoE is maximized (i.e., expected imagined rewards $\mathbb{E}(\sum_{\tau=t}^{H} \gamma^{\tau-t} r_\tau)$), as discussed next.

A. *Causal Imitation Learning Problem: Transmitter Design*

Herein, at the expert node, our objective is to throw away irrelevant state variables while learning the causal dynamics and hence design an abstract causal state description (represented using $p(\boldsymbol{z}^t \mid \boldsymbol{s}^t, \boldsymbol{s}^{t+1})$) which gets transmitted. Towards this objectives, we adopt the concept of deep end-to-end causal inference (DECI) from [28]. We specifically selected DECI because other contemporary works treat causal discovery and causal inference as distinct problems. In contrast, DECI is the only approach that simultaneously addresses both problems by conducting causal discovery through the estimation of a sparse directed graph and performing causal inference by estimating specific quantities based on a given set of inputs (interventions) – such as estimating $\boldsymbol{s}^{t+1}$ in this particular case. Note that the objectives for causal inference in an SC system differ from those in [28]. In contrast to [28], we suggest a constrained optimization approach to achieve a minimum QoE. DECI employs a Bayesian perspective in its causal discovery process, where the causal graph G is modeled in conjunction with the observations $\boldsymbol{s}^t, \boldsymbol{s}^{t+1}$. First, we look at the joint distribution (17), where $\boldsymbol{G}$ is defined as the adjacency matrix (corresponding to the graph $\mathcal{G}$), with entries $G_{j,i} \in \{0,1\}$ indicating the presence of an edge from $j \to i$. $\boldsymbol{\Psi}$ is the set of NN parameters. The causal relationships can be modeled using a non-linear additive noise model (ANM) as follows: $\boldsymbol{s}_i^{t+1} = \boldsymbol{f}_i(\boldsymbol{s}_i^t, Pa_i) + \boldsymbol{n}_i$, where $\boldsymbol{n}_i$ is an exogenous noise variable that is independent of all other variables. We propose a flexible NN parameterization that satisfies the graph adjacency constraints by setting $\boldsymbol{f}_i(\boldsymbol{s}_i^{t+1}) = \boldsymbol{\zeta}_i \left(\sum_j G_{j,i} \kappa_j(\boldsymbol{s}_j^t) \right)$, where the each entry of $\boldsymbol{\zeta}_i$ and κ_j are multilayer perceptrons (MLPs). Our objective is to use observational data (that involves network state $\boldsymbol{s}^t$) to fit the parameters $\boldsymbol{\Psi}$ of our non-linear ANM. After fitting the model, the posterior $p_{\boldsymbol{\Psi}}(\boldsymbol{G} \mid \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{a}^t, \boldsymbol{\theta})$ reflects our understanding of the causal structure. The graph prior $p(\boldsymbol{G})$ should characterize the graph as a *directed acyclic graph (DAG)*. We choose it as $p(\boldsymbol{G}) \propto exp(-\lambda_s \|\boldsymbol{G}\|^2 - \rho\, h(\boldsymbol{G})^2 - \alpha\, h(\boldsymbol{G}))$, where the DAG penalty $h(\boldsymbol{G}) = \mathrm{tr}(e^{\boldsymbol{G} \odot \boldsymbol{G}}) - D$, which is non-negative and zero only if $\boldsymbol{G}$ is a DAG. Variables α and ρ represent the weights for the DAG penalty [41], and they are optimized during training as in [28, Appendix B]. The prior knowledge about graph sparseness is modeled by the term $\lambda_s \|\boldsymbol{G}\|^2$, where λ_s is a fixed known quantity. The model has two difficulties: first, the actual posterior over $\boldsymbol{G}$ cannot be computed efficiently, and second, the maximum likelihood approach cannot be applied to estimate the model parameters due to the presence of the latent variable $\boldsymbol{G}$. To address both challenges simultaneously, we employ variational inference, as suggested in previous literature [42]. Specifically, we introduce a variational distribution $q_{\boldsymbol{\Psi}}(\boldsymbol{G})$ to approximate the intractable posterior $p_{\boldsymbol{\Psi}}(\boldsymbol{G} \mid \boldsymbol{z}^t, \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{a}^t, \boldsymbol{\theta})$, and utilize it to construct the evidence lower bound (ELBO) given by (18). Here $\boldsymbol{\chi}$ is the NN parameters for the $p_{\boldsymbol{\chi}}(x^{(n)} \mid G)$. (18) can be rewritten as (19). The causal discovery architecture includes a graph NN (GNN) [43] which accepts as input $\boldsymbol{s}^t$ and propagates information across a fully connected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. This graph includes vertices $v_i \in \mathcal{V}$ for each $\boldsymbol{s}_i^t$, and each pair of vertices (v_i, v_j) is connected by an edge whose embedding is represented by the MLP $\boldsymbol{e}_{i,j} = \boldsymbol{f}_{i,j}(\boldsymbol{s}_i^t, \boldsymbol{s}_j^t)$, and:

$$p_{\boldsymbol{\Psi}}(G_{j,i} \mid \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{a}^t, \boldsymbol{\theta}) = \mathrm{Softmax}(\boldsymbol{e}_{i,j} \mid T), \tag{20}$$

where T represents the softmax temperature variable [44]. DECI aims to demonstrate that the optimization of the variational objective in equation (18) can recover the ground truth data generation mechanism. To establish this statistical guarantee under the correct specification of DECI, which assumes the existence of $\boldsymbol{\chi}^0$ and $\boldsymbol{G}^0$ that correspond to the true data generating process and the absence of unobserved confounding variables, two technical assumptions must be satisfied. Firstly, function $\boldsymbol{f}_i$ must be non-invertible (since a DAG), 3rd-order differentiable, and not constant with respect to any of its inputs. Secondly, the proper noise densities must have bounded likelihood. These assumptions

$$p_{\boldsymbol{\Psi}}(\boldsymbol{z}^t, \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{a}^t, \boldsymbol{\theta}, G) = p(G)p(\boldsymbol{\theta})p(\boldsymbol{s}^t)\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^t, \boldsymbol{\theta})p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta})p(\boldsymbol{z}^t \mid \boldsymbol{s}^t, \boldsymbol{s}^{t+1}), \tag{17}$$

$$g_E(\boldsymbol{\chi}, \boldsymbol{\Psi}) = \mathrm{E}_{q_{\boldsymbol{\Psi}}(\boldsymbol{G})}\left[\log p((\boldsymbol{G}) \prod_{n:x^{(n)} \in \{\boldsymbol{z}^t, \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{a}^t, \boldsymbol{\theta}\}} p_{\boldsymbol{\chi}}(x^{(n)} \mid \boldsymbol{G})\right] + \mathbb{H}(q_{\boldsymbol{\Psi}}) \leq \log p_{\boldsymbol{\chi}}(x^{(1)}, \cdots, x^{(n)}). \tag{18}$$

$$\begin{aligned} g_E(\boldsymbol{\chi}, \boldsymbol{\Psi}) = \mathrm{E}_{q_{\boldsymbol{\Psi}}(\boldsymbol{G})}&\left[\prod_{n:x^{(n)} \in \{\boldsymbol{z}^t, \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{a}^t, \boldsymbol{\theta}\}} p_{\boldsymbol{\chi}}(x^{(n)} \mid \boldsymbol{G})\right] - \mathrm{KLD}\left(q_{\boldsymbol{\Psi}}(\boldsymbol{G} \mid \boldsymbol{x}) || p(\boldsymbol{G})\right) \\ &\leq \log p_{\boldsymbol{\chi}}(x^{(1)}, \cdots, x^{(n)}). \end{aligned} \tag{19}$$

rule out non-identifiable ANM. Further, the ELBO maximization problem $\mathcal{P}_1$ in (1) that involves expert policy computation, causal discovery and state transition dynamics can be reformulated as follows.

$$\begin{aligned} \mathcal{P}_1 : & \underset{\pi_E(\boldsymbol{a}^t|\boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta}),\, \boldsymbol{G},\, p(\boldsymbol{s}^{t+1}|\boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta})}{\arg\max} \quad g_E(\boldsymbol{\chi}, \boldsymbol{\Psi}) \\ & \text{subject to } \mathbb{E}(r_t) \leq R_b \end{aligned} \tag{21}$$

We now explain how to obtain the reward constraint bound, denoted as R_b in this context. To accomplish this, we must first define the concept of "semantic reliability". The expert agent's goal is to select an optimal semantic representation (encoder) that accurately represents the causal structure at the transmit side. The imitator, on the other hand, aims to improve semantic effectiveness on its side to achieve a desired level of semantic reliability. *Semantic reliability* is quantified by the expression, $p\left(E_t\left(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t\right) < \delta\right) \geq 1 - \epsilon$, where $E_t(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t) = \left\|\boldsymbol{s}^t - \widehat{\boldsymbol{s}}^t\right\|^2$ represents semantic distortion and ϵ is arbitrarily small. This metric reflects the imitator's ability to reliably reconstruct all the causal aspects in the decoded causal structure. Unlike classical reliability measures, in semantics, we can recover the actual meaning of transmitted messages even with a higher bit error rate (BER), as long as the semantic distortion remains within the set limit. This is illustrated by the choice of δ here, which depends on the concept of semantic space. The *semantic space* $\mathcal{K}$ is defined as an N-dimensional topological ball (since $\boldsymbol{s}^t$ belongs to an abstract cell complex as proved in Theorem 1) centered at the actual state $\boldsymbol{s}^t$, where all points inside the ball corresponding to the same semantic information. Formally, we can express this as $E_t(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t) \leq \delta$ such that $\mathbb{I}_\phi(\boldsymbol{s}^t) = \mathbb{I}_\phi(\widehat{\boldsymbol{s}}^t)$, where E_t is the topological distance between states $\boldsymbol{s}^t$ and $\widehat{\boldsymbol{s}}^t$, δ is the radius of the topological ball, and $\mathbb{I}_\phi$ is a mapping function that maps states to their corresponding semantic information. The topological distance here refers to the irreducibility measure quantifying the causal relationships between any two states $\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t$. The upper bound R_b is defined by the reward obtained at $(\widehat{\boldsymbol{s}}^t)^*$ which corresponds to $(\widehat{\boldsymbol{s}}^t)^* = \underset{\widehat{\boldsymbol{s}}^t \in \mathcal{K}}{\arg\min}\, C_t(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t)$.

The second challenge pertains to resolving the semantic encoder $(p_{\boldsymbol{\Psi}}(\boldsymbol{z}^t \mid (\boldsymbol{s}^t, \boldsymbol{s}^{t+1})))$ problem using a revised version of the information bottleneck principle [45]. In contrast to [45], we leverage semantic information measures derived from IIT principles. As a result, our metrics go beyond conventional information theory approaches. Our objective is to determine the distribution $p_{\boldsymbol{\Psi}}(\boldsymbol{z}^t \mid (\boldsymbol{s}^t, \boldsymbol{s}^{t+1}))$, which converts the input signal $\boldsymbol{s}$ into a representation $\boldsymbol{z}$, such that $\boldsymbol{z}$ discloses as little information as possible about $\boldsymbol{s}$, while extracting the maximum amount of semantic information about $\boldsymbol{s}$ in the output signal $\boldsymbol{y}$ $(\mathbb{I}_\phi\left(\boldsymbol{s}; \boldsymbol{y}\right))$. We reformulate $\mathcal{P}_2$ in (2) as:

$$\begin{aligned} \mathcal{P}_2 : & \underset{p_{\boldsymbol{\Psi}}(\boldsymbol{z}^t|(\boldsymbol{s}^t, \boldsymbol{s}^{t+1}))}{\arg\max} \quad \mathbb{I}_\phi(\boldsymbol{s}^t; \boldsymbol{y}^t) \\ & \text{s.t. } \mathbb{I}_\phi(\boldsymbol{s}^t; \boldsymbol{z}^t) \leq \mathbb{I}_b \end{aligned} \tag{22}$$

According to Definition 4, causally invariant states are represented using the same semantic representation $\boldsymbol{z}^t$, which is ensured by the constraint imposed on $\mathbb{I}_\phi(\boldsymbol{s}^t; \boldsymbol{z}^t)$ in (22). Next, we discuss how to solve the above two optimization problems $\mathcal{P}_1$ and $\mathcal{P}_2$.

1) Proposed Solution: The resulting bi-level optimization involving $\mathcal{P}_1$ and $\mathcal{P}_2$ can be alternatively solved as follows. We first write the Lagrangian corresponding to (21) and (22) as follows.

$$\begin{aligned} &\mathcal{L}_1 = \\ &\min_{\lambda_1} \max_{\pi_E(\boldsymbol{a}^t|\boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta}),\, \boldsymbol{G},\, p(\boldsymbol{s}^{t+1}|\boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta})} g_E + \lambda_1\left(R_b - \mathbb{E}(r_t)\right), \\ &\mathcal{L}_2 = \min_{\lambda_2} \max_{p_{\boldsymbol{\Psi}}(\boldsymbol{z}^t|(\boldsymbol{s}^t, \boldsymbol{s}^{t+1}))} \mathbb{I}_\phi(\boldsymbol{s}^t; \boldsymbol{y}^t) + \lambda_1\left(\mathbb{I}_b - \mathbb{I}_\phi(\boldsymbol{s}^t; \boldsymbol{z}^t)\right). \end{aligned} \tag{23}$$

To solve the resulting bi-level optimization, the technique is described in Appendix G. The resulting alternating updates for outer optimization (solving $\mathcal{L}_1$) can be obtained as (where (k) denotes iteration) (24). Our approach involves aggressively improving the model to maximize the ELBO based on the current semantic encoder, followed by implementing a more conservative policy and causal discovery. The detailed algorithmic process is presented in Algorithm 1 and the model architecture in Fig. 3, which concludes the SC system design at the expert agent side. The algorithm convergence follows similar theoretical arguments as in [28, Theorem 1]. The authors therein demonstrate that maximizing the ELBO in (18) can recover both the ground truth data generating process, $p(\boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{\theta}, \boldsymbol{z}^t, \boldsymbol{a}^t; \boldsymbol{G}^0)$, and the true causal graph, $\boldsymbol{G}^0$, in the infinite data limit. Apart from the ANM model, the convergence guarantees are under the assumption that 1) there are no latent confounders (which is true from the perspective of expert agent) and 2) the log-likelihood is regular,

$$\begin{aligned}
(\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta}))_{k+1} &= \underset{\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta})}{\arg\max}\ \mathcal{L}_1\left(\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta}), \boldsymbol{G}_{(k)}, p_{(k)}(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}), \lambda_{1,(k)})\right), \\
\boldsymbol{G}_{(k+1)} &= \arg\max_{\boldsymbol{G}} \mathcal{L}_1\left(\left(\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta})\right)_{(k+1)}, \boldsymbol{G}, p_{(k)}(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}), \lambda_{1,(k)})\right), \\
p_{(k+1)}\left(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}\right) &= \underset{\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta})}{\arg\max}\ \mathcal{L}_1\left(\left(\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta})\right)_{k+1}, \boldsymbol{G}_{(k+1)}, p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}), \lambda_{1,(k)}\right), \\
\lambda_{(k+1)} &= \arg\min_{\lambda} \mathcal{L}_1\left(\left(\pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^{t+1}, \boldsymbol{s}^t, \boldsymbol{\theta})\right)_{k+1}, \boldsymbol{G}_{(k+1)}, p_{(k+1)}(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}), \lambda_1\right).
\end{aligned} \tag{24}$$

Algorithm 1 Proposed Bi-level Optimization for Causal Discovery and Semantic Encoder

```
Given: $p(\boldsymbol{G}), p(\boldsymbol{s}^0)$
Define: $\mathcal{E} = \left(\boldsymbol{G}, p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{\theta}, \boldsymbol{a}^t), \pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^t, \boldsymbol{\theta})\right)$
Initialize: Policy $\pi_E$, causal graph $\boldsymbol{G} \sim p(\boldsymbol{G})$, data buffer $\mathcal{D} = \{\}$.
for $(t = 0, 1, 2, \dots)$ do
    Collect data $\mathcal{D}_t = (\boldsymbol{s}^t, \boldsymbol{\theta})$ by executing $\pi_E$ in the environment.
    Build local (policy-specific) dynamics model: $\left[\widehat{\mathcal{E}}_{t+1}, \widehat{\lambda}_{t+1}\right] = \arg\min_{\lambda} \max_{\mathcal{E}} g_E + \lambda(R_b - \mathbb{E}(r_t))$.
    Improve policy: $(\pi_E)_{k+1} = (\pi_E)_k + \alpha_k \nabla_{\pi_E} g_E(\pi_\eta, \widehat{\mathcal{E}}^{\pi}_{k+1})$. with a conservative algorithm like policy search with natural gradient (NPG) [46].
    Compute the encoder distribution, given the causal graph and state transitions as $p(\boldsymbol{z}^t \mid \boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{G})$ using VAE.
    Simulate the receiver part using Algorithm 2.
    Evaluate semantic effectiveness metric $C_t(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t)$.
end for
```

which means that $\mathbb{E}_p\left[\left|\log p_{\boldsymbol{\chi}}(\boldsymbol{s}^t, \boldsymbol{s}^{t+1}, \boldsymbol{z}^t, \boldsymbol{a}^t \mid \boldsymbol{G})\right|\right] < \infty$. We next look at the causal IL problem for the receiver design.

B. Causal Imitation Learning Problem: Receiver Design

The distribution of the trajectory τ of the expert's interactions with the environment can be written as:

$$P_E(\tau \mid \boldsymbol{\theta}) = p(\boldsymbol{s}^0 \mid \boldsymbol{\theta}) \prod_{t=0}^{H} p(\boldsymbol{s}^{t+1} \mid \boldsymbol{s}^t, \boldsymbol{a}^t, \boldsymbol{\theta}) \pi_E(\boldsymbol{a}^t \mid \boldsymbol{s}^t, \boldsymbol{\theta}) \tag{25}$$

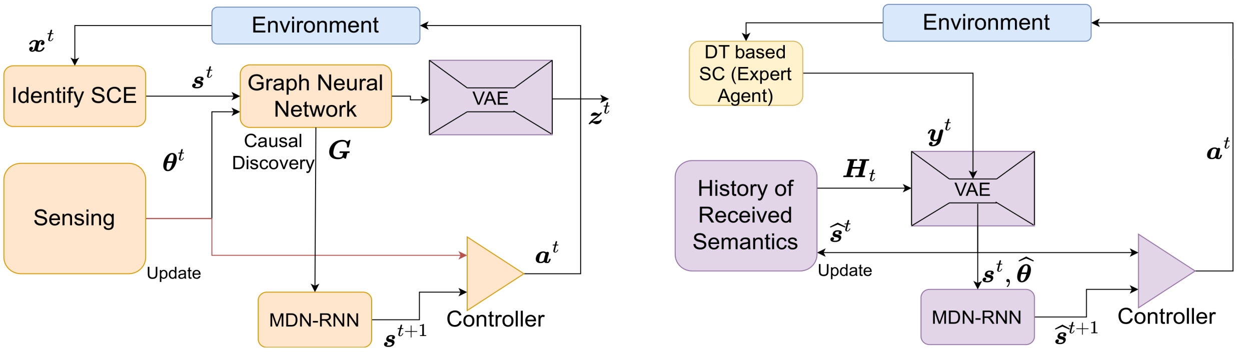
Since the imitator does not have access to the confounding variables $\boldsymbol{\theta}$, it must learn the distribution over trajectories based on the state transitions history that is available to it. This history is constructed from the imitator's interactions with the environment. The imitator can then use this learned distribution to guide its own actions and improve its performance on the task.

$$\begin{aligned}
&P_\eta(\tau \mid \boldsymbol{\theta}) = \\
&p(\widehat{\boldsymbol{s}}^0 \mid \boldsymbol{\theta}) \prod_{t=0}^{H} \underbrace{p(\widehat{\boldsymbol{s}}^{t+1} \mid \widehat{\boldsymbol{s}}^t, \boldsymbol{a}^t, \boldsymbol{\theta})}_{\text{causal dynamics model}} \underbrace{\pi_\eta(\boldsymbol{a}^t \mid \widehat{\boldsymbol{s}}_0, \boldsymbol{a}_0, \cdots, \widehat{\boldsymbol{s}}^t)}_{\text{imitator policy}}.
\end{aligned} \tag{26}$$

The goal of IL here is for the imitator to compute the policy such that the average reward (which is essentially QoE here) is maximized. The average is taken across all possible trajectories and the confounding variables, i.e., $\mathbb{E}_{\boldsymbol{\theta} \sim P(\boldsymbol{\theta} \mid \tau)} \mathbb{E}_{\tau \sim P(\tau \mid \boldsymbol{\theta})}\left[\mathcal{R}(\widehat{\boldsymbol{s}}^t, \boldsymbol{a}^t; \boldsymbol{\theta})\right]$. The resulting policy learned at the imitator is $\pi_\eta(\boldsymbol{a}^t \mid \widehat{\boldsymbol{s}}^t, \widehat{\boldsymbol{\theta}})$. However, this approach assumes that the imitator is ideal, i.e., it has infinite memory capacity and perceives the state and actions in the same way as indicated by the expert. Specifically, the optimal policy is often a deterministic function of the environment $\mathcal{E} = (\mathcal{T}, \mathcal{R}, \boldsymbol{\theta})$ such that if the imitator is able to identify the environment, it has all the necessary information to determine the optimal policy. The imitator reflects its initial uncertainty about the environment through a prior distribution $P(\mathcal{E} \in \cdot)$. As the history unfolds, the imitator's current knowledge of the environment can be represented by the posterior probabilities $P(\mathcal{E} \in \cdot \mid \boldsymbol{H}_t)$. $\boldsymbol{H}_t$ represents the history of state-action pairs available at the imitator node until time t. The total number of bits needed to identify the environment is $\mathbb{H}(\mathcal{E})$. Assuming the reward function is also unknown to the imitator (note that the expert agent evaluates the semantic effectiveness, which is the intrinsic reward), the total number bits required to learn the environment is

$$\mathbb{H}(\mathcal{E}) = \mathbb{H}(\mathcal{T}) + \mathbb{H}(\mathcal{R}). \tag{27}$$

In other words, the DT-based SC system may need to invest $\mathbb{H}(\mathcal{E})$ bits for the initial training. However, in practice, the imitator can only acquire a learning target $\widehat{\mathcal{E}}$, which is a stochastic function of the actual environment. This means that due to communication and computational limitations, the amount of information acquired by the agent about $\mathcal{E}$ is only $\mathbb{I}(\mathcal{E}; \widehat{\mathcal{E}})$, which is derived from the semantic information received (states here). This can be quantified as only a subset of history being stored at the agent due to memory limitations, i.e., $P(\mathcal{E} \in \cdot \mid \boldsymbol{H}_t^{t-K})$. Here, $\boldsymbol{H}_t^{t-K}$ represents the history of state transitions corresponding to the most recent K time instants. The states perceived by the imitator can be erroneous due to the presence of a wireless channel between the expert and the imitator. Thus, we can consider that the imitator only knows an estimate $\widehat{\boldsymbol{H}}_t$ and, hence, the environment $P(\widehat{\mathcal{E}} \in \cdot \mid \widehat{\boldsymbol{H}}_t)$. This limits the space over which the learned target environment lies. As a result, aiming for $\widehat{\mathcal{E}}$ incurs a bounded degree of performance, $\mathbb{E}[\bar{V}^* - \bar{V}^{\pi_{\widehat{\mathcal{E}}}}]$. A natural measure of distortion here is the expected squared regret between the optimal and target policies:



Transmitter (Expert Agent) Architecture

Receiver (Imitator Agent) Architecture

Fig. 3: Generative AI components of the proposed network state model.

$$d(\mathcal{E},\widehat{\mathcal{E}} \mid \boldsymbol{H}_t) = \mathbb{E}\left[\mathbb{E}_{\pi_{\widehat{\mathcal{E}}}}\left(\sum_{t'=t}^{H} \bar{V}^*(\boldsymbol{H}_t) - \bar{V}^{\pi_{\widehat{\mathcal{E}}}}(\boldsymbol{H}_t)\right)^2 \mid \mathcal{E}, \boldsymbol{H}_t\right]. \tag{28}$$

We now look at how to estimate the different components of $\widehat{\mathcal{E}}$. The receiver begins by extracting the semantics $\boldsymbol{s}^t$ from the received signal $\boldsymbol{y}^t$, which are modeled using the channel probability distribution $p(\boldsymbol{y}^t \mid \boldsymbol{z}^t)$. The resulting $\widehat{\boldsymbol{s}}^t$ is then stored in the observation history. Using this history, the receiver obtains an estimate of the confounding variables $\widehat{\boldsymbol{\theta}}^t$ that helped the expert agent to analyze the optimal policy. The imitator then uses both the confounding variable information and the extracted state to derive the policy $\pi_\eta(\boldsymbol{a}^t \mid \widehat{\boldsymbol{s}}^t, \widehat{\boldsymbol{\theta}}^t)$. We now look at the details of the proposed semantic decoder and imitator policy design at the receiver. First, we define the loss function that captures the difference in learned state transition dynamics compared to that at the expert node.

Definition 6. Given the inferred model at the imitator $\widehat{\mathcal{E}}$ and the original sampling distribution $\mu(\boldsymbol{s},\boldsymbol{a})$, the *model approximation loss* is defined as follows.

$$l(\widehat{\mathcal{E}},\mu) = \mathbb{E}_{(\boldsymbol{s},\boldsymbol{a})\sim\mu}\left[\mathrm{KLD}\left((P(\cdot \mid \boldsymbol{s},\boldsymbol{a}) \mid\mid \widehat{P}(\cdot \mid \boldsymbol{s},\boldsymbol{a})\right)\right]. \tag{29}$$

(29) is used to quantify the semantic reconstruction loss on the imitator's end. As explained next, we aim to address the issue of determining the state transition probability and imitator policy.

1) Variational Inference Framework for Receiver Design: For simplicity, we reuse the same notation $\boldsymbol{\Psi}$ from Section IV-A for the NN parameters here. Our proposed method here learns confounding variables by maximizing an information theoretic objective defined as follows

$$\begin{aligned} F(\boldsymbol{\Psi}) &= \mathbb{I}(\boldsymbol{s}^t;\boldsymbol{\theta} \mid \boldsymbol{H}_t) + \mathbb{H}\left[\boldsymbol{A} \mid \boldsymbol{s}^t;\boldsymbol{H}_t\right] - \mathbb{I}\left(\boldsymbol{A};\boldsymbol{\theta} \mid \boldsymbol{s}^t;\boldsymbol{H}_t\right) \\ &= \left(\mathbb{H}[\boldsymbol{\theta} \mid \boldsymbol{H}_t] - \mathbb{H}[\boldsymbol{\theta} \mid \boldsymbol{s}^t;\boldsymbol{H}_t]\right) + \mathbb{H}\left[\boldsymbol{A} \mid \boldsymbol{s}^t;\boldsymbol{H}_t\right] \\ &\quad - \left(\mathbb{H}\left[\boldsymbol{A} \mid \boldsymbol{s}^t;\boldsymbol{H}_t\right] - \mathbb{H}\left[\boldsymbol{A} \mid \boldsymbol{s}^t,\boldsymbol{\theta};\boldsymbol{H}_t\right]\right) \\ &= \mathbb{H}\left[\boldsymbol{\theta} \mid \boldsymbol{H}_t\right] - \mathbb{H}\left[\boldsymbol{\theta} \mid \boldsymbol{s}^t;\boldsymbol{H}_t\right] + \mathbb{H}\left[\boldsymbol{A} \mid \boldsymbol{s}^t,\boldsymbol{\theta};\boldsymbol{H}_t\right]. \end{aligned} \tag{30}$$

We particularly derived $F(\boldsymbol{\Psi})$ (where $\boldsymbol{\Psi}$ captures the NN parameters) for the following reasons. The first term encourages our prior distribution over $p(\boldsymbol{\theta} \mid \boldsymbol{H}_t)$ to have high entropy, which implies that the imitator ideally prefer to obtain maximum information about $\boldsymbol{\theta}$. The second term suggests that it should be easy to infer the confounding variable $\boldsymbol{\theta}$ from the current state and history. The third term suggests that each learned target should act as randomly as possible (random actions), which we achieve by using a maximum entropy policy to represent each confounder. As we cannot integrate over all states and confounders to compute $p(\boldsymbol{\theta} \mid \boldsymbol{s}^t,\boldsymbol{H}_t)$ exactly, we approximate this posterior with a learned discriminator $q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{s}^t,\boldsymbol{H}_t)$. By using Jensen's Inequality, replacing $p(\boldsymbol{\theta} \mid \boldsymbol{s}^t,\boldsymbol{H}_t)$ with $q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{s}^t,\boldsymbol{H}_t)$ gives us a variational lower bound $G(\boldsymbol{\phi},\boldsymbol{\psi})$ on our objective $F(\boldsymbol{\phi})$. Hence, we rewrite (30)

$$\begin{aligned} F(\Psi) \geq \mathbb{H}\left[\boldsymbol{A} \mid \boldsymbol{s}^t,\boldsymbol{\theta};\boldsymbol{H}_t\right] + \mathbb{E}_{\boldsymbol{\theta}\sim p(\boldsymbol{\theta})}[\log q_{\boldsymbol{\phi}}(\boldsymbol{\theta} \mid \boldsymbol{s}^t;\boldsymbol{H}_t) \\ - \log p(\boldsymbol{\theta} \mid \boldsymbol{H}_t)] = G(\boldsymbol{\phi},\boldsymbol{\psi}). \end{aligned} \tag{31}$$

Our goal here is to maximize (31) while simultaneously reducing the average regret (that captures the performance shortfall associated with the suboptimal learning target $\widehat{\mathcal{E}}$) in (28). To incorporate the latter aspect, we look at the principle of information-directed sampling (IDS) [47]. IDS is an abstract objective for sequential decision-making agents where, at each time period, an agent computes a policy based on the current history $\boldsymbol{H}_t$ that minimizes the following ratio of average regret to information:

$$J(\pi_\eta,\widehat{\mathcal{E}}) = \frac{\mathbb{E}_\pi\left[d(\mathcal{E},\widehat{\mathcal{E}} \mid \boldsymbol{H}_t)\right]}{\mathbb{I}(\widehat{\mathcal{E}};\boldsymbol{a}^t,\boldsymbol{s}_{t+1} \mid \boldsymbol{H}_t)}. \tag{32}$$

IDS is particularly relevant for CSC since it has the dual objective of (a) maximizing both the QoE by narrowing the gap between its performance and (b) that of an expert policy and enhancing the understanding of the environment $\mathcal{E}$ by utilizing the current extracted semantic state and its history. Unlike in [47], we propose to replace the IDS via

$$\left[\pi_{\eta}^{\widehat{\mathcal{E}}}, \widehat{\boldsymbol{\theta}}, \widehat{p}(\boldsymbol{s}^t \mid \boldsymbol{y}^t)\right] = \underset{\pi_{\eta},\, \boldsymbol{\theta},\, p(\boldsymbol{s}^t|\boldsymbol{y}^t)}{\arg\min} \frac{\mathbb{E}_{\pi}\left[d(\mathcal{E}, \widehat{\mathcal{E}} \mid \boldsymbol{H}_t)\right]}{\mathbb{E}_{\boldsymbol{y},\boldsymbol{s}} G(\boldsymbol{\phi}, \boldsymbol{\psi})} \quad \text{s.t. } p\left(E_t\left(\boldsymbol{s}^t, \widehat{\boldsymbol{s}}^t\right) < \delta\right) \geq 1 - \epsilon, \tag{33}$$

where the constraint represents the semantic reliability measure defined earlier. For the denominator term, the expectation is over $p(\boldsymbol{y}^t)p(\boldsymbol{s}^t \mid \boldsymbol{y}^t)$, where $p(\boldsymbol{s}^t \mid \boldsymbol{y}^t)$ is the semantic decoder distribution. Next, we look at how to solve (33) using MBRL.

2) Model Based RL as a Bi-level Optimization: MBRL is formulated as a bi-level optimization in order to capture the interactions between model and policy learning. Further, we look at the optimization problem for the transition model learning and the imitator policy computation.

$$\mathcal{P}_3 : \underbrace{\max_{\pi_\eta} J(\pi_\eta, \widehat{\mathcal{E}}),}_{\text{policy-learner}} \quad \underbrace{\min_{\mathcal{E}} l\left(\widehat{\mathcal{E}}, \mu_{\mathcal{E}}^{\pi_\eta}\right) \text{subject to}\, c(\widehat{\mathcal{E}}) \leq C}_{\text{model-learner}} \tag{34}$$

We use $\mu_{\mathcal{E}}^{\pi_\eta} = \frac{1}{T}\sum_{t=0}^{T} P(\boldsymbol{s}^t = s, \boldsymbol{a}^t = a)$ to denote the average state visitation distribution. $c(\widehat{\mathcal{E}}) \leq C$ captures the resource constraints at the imitator. The MBRL can be viewed as a two-level optimization process here, and the technique to solve it is described in Appendix G. At the outer level, the objective is to optimize the policy π_η to achieve the best possible performance within the learned model. At the inner level, an optimization problem is solved to minimize the prediction error for $\mathcal{E}$ under the induced state distribution of the policy. This is a bi-level optimization because each objective depends on the parameters of both the problems discussed above. The formulation above decomposes MBRL into policy learning and generative model learning components, emphasizing that they are interdependent and must be addressed together for success. The technique to solve $\mathcal{P}_3$ is discussed in Appendix G and detailed steps are described in Algorithm 2. Next, we look at the NN architecture that we adopt for various learned components in $\mathcal{P}_3$.

C. Generative AI Architecture for "Network State Model"

The imitator model considered here is based on the concept of world models proposed in generative AI [27] and is inspired by our cognitive system. Generative AI is especially advantageous for DT-based SC systems for the following reasons. First, it helps to improve the amount of semantic information acquired by the receiver through accurate modeling of the physical world with the help of DT. Second, as the information to be transmitted becomes more complex, such as 2-D images and videos, and eventually 3-D holograms or even higher-dimensional objects in future connected intelligence systems, generative AI can help the receiver generate more information with less semantics transmitted, compared to classical syntactic communication. It comprises three components that work together to process received semantic information and make decisions based on past experiences. The first component is the *visual sensory component*, which VAE represents. This component is responsible for decoding the received information received by the agent into a respective network state $\widehat{\boldsymbol{s}}^t$. This decoder distribution is represented by $p(\boldsymbol{s}^t \mid \boldsymbol{y}^t, \boldsymbol{H}_t)$. The second component is the *memory component*, represented by mixture-density network (MDN) combined with a recurrent NN (MDN-RNN). It predicts future states based on historical information, allowing the imitator to anticipate and prepare for potential future events. To anticipate the system's future states, we use a RNN as a predictive model of future $\boldsymbol{s}^{t+1}$ vectors. However, since many complex environments are stochastic, we train our RNN to output a probability density function $p(\boldsymbol{s}^{t+1})$ instead of a deterministic prediction of $\boldsymbol{s}^{t+1}$. This allows us to capture the uncertainty inherent in the environment and make more informed decisions based on possible future states. In our approach, we approximate $p(\boldsymbol{s}^{t+1})$ as a mixture of Gaussian distributions. We then train the memory component (MDN-RNN) to output the probability distribution of the next latent vector $\boldsymbol{s}^{t+1}$, based on its current and past information. To be more specific, the RNN, which has N_h hidden units, models $P(\boldsymbol{s}^{t+1} \mid \boldsymbol{a}^t, \boldsymbol{s}^t, \boldsymbol{h}^t)$, where $\boldsymbol{h}^t$ is the hidden state of the RNN at time step t. We can adjust a temperature parameter T during sampling to control model uncertainty, as in [27]. We have found that adjusting T is useful for training our controller later. The third and final component is the decision-making component, which is the *controller*. It makes decisions based solely on the representations created by the vision and memory components, enabling the agent to take appropriate actions based on past experiences and current situations. At the time t, the imitator takes action $\boldsymbol{a}^t \in \mathcal{R}^{N_a}$, where N_a is the dimension of the action space. The proposed imitator model has been named the *"network state model"*, which shares similarities with the concept of world models in AI. However, in contrast to traditional world models in AI, the various generative AI components in the network state model (that mimic the physical environment semantics) are optimized to achieve semantic effectiveness (QoE) as close to 1 as possible, which is critical for a DT-based SC system. The training procedure to optimize the generative AI parameters is detailed in Algorithm 2.

Next, we assess the performance shortfall associated with the suboptimal environment (and hence a suboptimal policy compared to expert agent) learned at the imitator.

Theorem 2. *(Global performance of equilibrium pair) Consider a pair of policy and environment model at the imitator, $(\pi_\eta, \mathcal{E})$, such that simultaneously*

$$D_{TV}(P(\cdot \mid \boldsymbol{s}, \boldsymbol{a}, \widehat{\boldsymbol{\theta}}), \widehat{P}(\cdot \mid \boldsymbol{s}, \boldsymbol{a}, \widehat{\boldsymbol{\theta}})) \leq \epsilon_{\mathcal{E}}, \forall \boldsymbol{s}, \boldsymbol{a}, \widehat{\boldsymbol{\theta}} \tag{35}$$

and for simplicity, we assume that the extrinsic reward is always zero and the semantic effectiveness (intrinsic reward) is bounded, such that $\mathcal{R}(\boldsymbol{s}) \leq R_{max}, \forall \boldsymbol{s} \in \mathcal{S}$. Also,

Algorithm 2 Proposed Bi-level Optimization for Receiver Design

```
Given: $\boldsymbol{y}$
Initialize: Sample $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$, and $\boldsymbol{s}^t \sim p(\boldsymbol{s}^t)$. Collect $10,000$ rollouts from a random policy
for $(t = 0, 1, 2, \dots)$ do
    Train VAE to decode $\boldsymbol{y}^t$ into $\widehat{\boldsymbol{s}}^t \in \mathcal{R}^N$
    Update the history of observations with $\widehat{\boldsymbol{s}}^t$.
    Train MDN-RNN to model $p(\boldsymbol{\theta} \mid \boldsymbol{H}_t, \widehat{\boldsymbol{s}}^t)$. Sample the point estimate $\widehat{\boldsymbol{\theta}} \sim \arg\max p(\boldsymbol{\theta} \mid \boldsymbol{H}_t, \widehat{\boldsymbol{s}}^t)$ as its mode.
    Train the transition model $p(\widehat{\boldsymbol{s}}^{t+1} \mid \widehat{\boldsymbol{s}}^t, \widehat{\boldsymbol{\theta}}, \boldsymbol{H}_t)$ using RNN.
    Evolve controller to maximize the expected cumulative reward.
end for
```

assume that the information learned about the environment at the imitator is incremental over time, i.e. $\mathbb{I}(\widehat{\mathcal{E}}; \boldsymbol{a}^t, \boldsymbol{s}^{t+1} \mid \boldsymbol{H}_t) = \mathbb{I}_0 t$, *where* $\mathbb{I}_0$ *is information learned during the first communication instance. Then, we show that*

$$|J(\pi_\eta, \mathcal{E}) - J(\pi_\eta, \widehat{\mathcal{E}})| \le \frac{2\gamma \epsilon_{\mathcal{E}} R_{max}}{(1-\gamma)^2 \mathbb{I}_0 t} \ \forall \pi_\eta \tag{36}$$

Proof: See Appendix E. ■

Theorem 2 indicates that, with time, the discrepancy between the IDS objective and a hypothetical scenario where the imitator has perfect knowledge of $\mathcal{E}$ decreases to zero. This means that as communication between the expert and the imitator continues, the imitator can accurately replicate the expert's policies.

Theorem 3. *Suppose we have policy-model pair* $(\pi_\eta, \mathcal{E})$ *such that the following conditions hold simultaneously:*

$$l(\widehat{\mathcal{E}}, \mu_{\mathcal{E}}^{\pi,t}) \le \epsilon_{\mathcal{E}} \ \forall t \text{ and } J(\pi, \widehat{\mathcal{E}}) \ge \sup_{\pi'} J(\pi', \widehat{\mathcal{E}}) - \epsilon_\pi. \tag{37}$$

Let π^* *be an optimal policy so that* $J(\pi^*, \mathcal{E}) \ge J(\pi', \mathcal{E}) \ \forall \pi'$. *Then, at time t, the performance of the imitator (evaluated using QoE) will be bounded, as follows:*

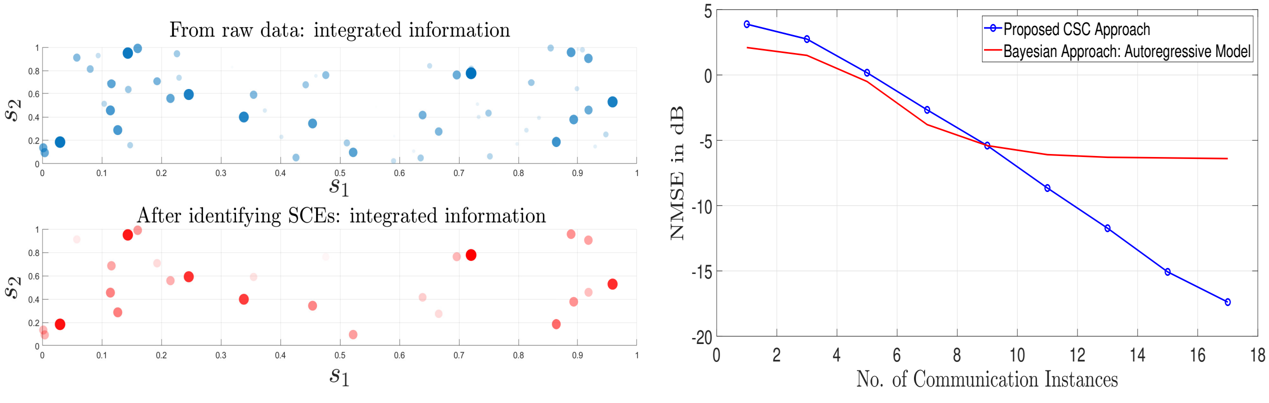
$$\begin{aligned} J(\pi^*, \mathcal{E}) - J(\pi, \mathcal{E}) \le & \frac{2R_{max}}{(1-\gamma)\mathbb{I}_0 t} D_{TV}(\mu_{\mathcal{E}}^{\pi^*}, \mu_{\widehat{\mathcal{E}}}^{\pi^*}) \\ & + \frac{2\gamma\sqrt{\epsilon_{\mathcal{E}}} R_{max}}{(1-\gamma)^2 \mathbb{I}_0 t}. \end{aligned} \tag{38}$$

Proof: See Appendix F. ■

Theorem 3 shows that, following a finite number of communication instances, if $\mathbb{I}_0 t$ is a finite value, the discrepancy in transition probability modeling (denoted by $D_{TV}(\mu_{\mathcal{E}}^{\pi^*}, \mu_{\widehat{\mathcal{E}}}^{\pi^*})$) dominates the error in the IDS objective with regards to the expert and imitator nodes. Therefore, accurately modeling the physical environment, specifically the network state transitions in our DT-based system, is crucial to ensure that the imitator's policy closely matches the optimal policy generated by the expert. As we observed in Theorem 2, using a generative AI architecture at the imitator can help to close the gap between the physical models of the expert and imitator.

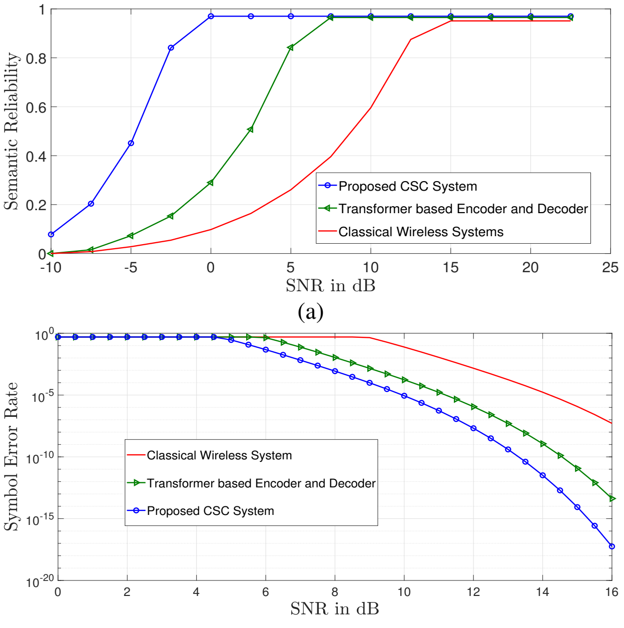
V. SIMULATION RESULTS AND ANALYSIS

To validate the effectiveness of our CSC approach, we must first confirm that our learning methods can accurately identify causal relationships in the world model. We must also ensure that our objective functions and semantic metrics facilitate this process and speed up the discovery of causal relationships. We begin with a dataset $\mathcal{X} = \{\boldsymbol{x}_s\}_{s=1}^S$ of S samples where each sample $\boldsymbol{x}_s$ consists of N stationary time-series $\boldsymbol{x}_s = \{\boldsymbol{x}_{s,1}, \boldsymbol{x}_{s,2}, \cdots, \boldsymbol{x}_{s,N}\}$ across time-steps $t = \{1, \cdots, T\}$. This dataset represent the states observed at the expert agent. We denote the t-th time-step of the i-th time-series of $\boldsymbol{x}_s$ as $\boldsymbol{x}_{s,i}^t \subset \mathcal{R}^D$. We consider an SCM captured by an associated DAG $\mathcal{G}_s^{1:T} = \{\mathcal{V}_s^{1:T}, \mathcal{E}_s^{1:T}\}$. underlying the generative process of each sample. The SCM's endogenous (observed) variables are vertices $v_{s,i}^t \in \mathcal{V}_s^{1:T}$ for each time-series i and each time-step t. Every set of incoming edges to an endogenous variable defines inputs to a deterministic function $g_{s,i}^t$ which determines that variable's value. Using DECI, our learned NN is assumed to follow the generation of $\boldsymbol{x}_s^t$ as $\boldsymbol{x}_s^{t+1} = \boldsymbol{g}_s(\boldsymbol{x}_s^{\le t}, \mathcal{G}_s, \boldsymbol{\theta}, \boldsymbol{a}^t) + \boldsymbol{v}_s^{t+1}$. $\boldsymbol{x}_s^t$ is the network state (physical environment) observed by the DT. The edges are defined by ordered pairs of vertices $\mathcal{E}_s^{1:T} = \{(v_{s,i}^t, v_{s,j}^{t'})\}$. Here, we consider $t' = t - 1$. and $\boldsymbol{g}_{s,i}^t = \boldsymbol{g}_{s,i}^{t'}$. In other words, our causal graphs and process dynamics are invariant across time for each time series sample and can vary across different samples. Our expert agent NNs must model $\boldsymbol{g}_s$ via the transition probability and the causal graph $\mathcal{G}_s$. An encoded version $\boldsymbol{z}^t$ is extracted using a VAE, which gets communicated to the imitator agent. At the receiver, the objective is to recover the causal graph accurately and regenerate the states. The actions at the receiver belong to a discrete set (of cardinality N_a), with a random chosen probability. The entire state space is divided into multiple nonoverlapping subsets, where for each subset there is a unique action probability, which is a discretized Gaussian distribution centered around the mean $\widetilde{\boldsymbol{a}}$, where the mean is distinct for different subsets. We test the CSC system on three datasets: two fully-observed physics simulations (Kuramoto and Particles) and the Netsim dataset of simulated fMRI data that originally appeared in [48] and was utilized for causal discovery in [44]. Throughout the experiments, we compare the performance of the proposed CSC framework to the two following benchmarks. The first is a classical wireless system that directly transmits the received state information (after encoding using a VAE) without any semantic extraction. The second baseline is an SC system that is causality unaware (and hence non-generalizable) [6]. Herein, the SCEs are identified using our proposed approach and further encoded/decoded using transformer modules as in [6] without any causal structure extraction. Fig. 4a presents a scatter plot of the integrated information in the data, which was calculated using (9) and taking into account two-dimensional states $\boldsymbol{s}_i^t$. The brightness of each point in the plot corresponds



(a) (b)

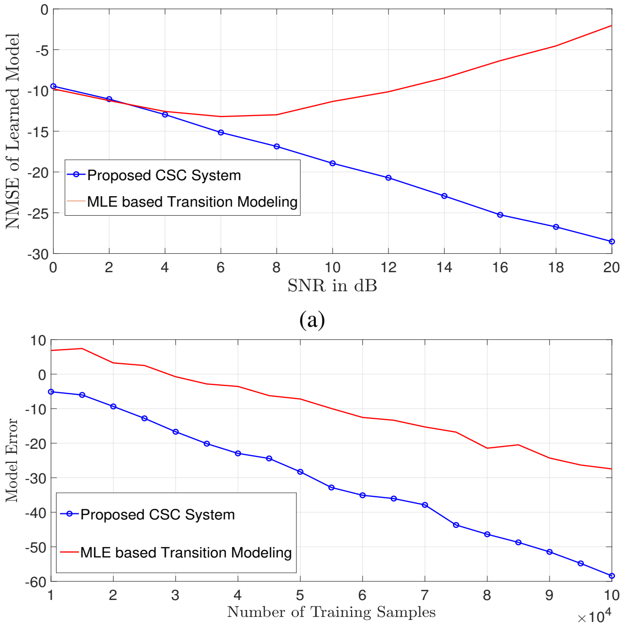
Fig. 4: (a) Integrated information of different subset of states s_i^t. The brightness of the circles represents the magnitude of the integrated information, which is shown in a two-dimensional format. (b) Error in confounding variable estimate between expert and imitator agent.



(a)

(b)

Fig. 5: (a) Semantic reliability vs SNR. (b) Symbol error rate vs SNR.



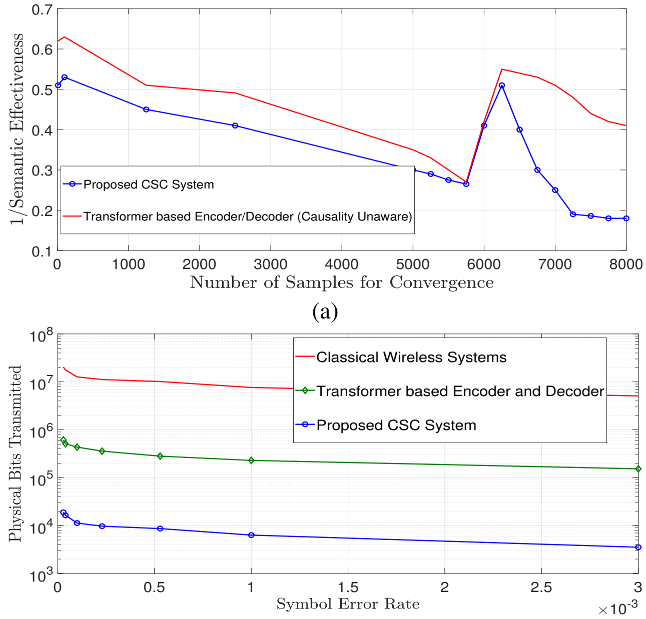
(a)

(b)

Fig. 6: (a) Environment model error between expert and imitator agents. (b) Generalization performance.

to the level of integrated information. Additionally, the second row of the same figure shows the SCEs identified through the algorithm described in Section III-C1. The number of SCEs comprises approximately two-thirds of the total number of states present in the data, representing a significant reduction in the amount of data that does convey any semantic information. Fig. 4b illustrates that as communication progresses, the error in confounding variable estimation decreases, in contrast to a model-based Bayesian approach that experiences a performance floor of approximately -5 dB due to inaccurate modeling of the physical environment.

Fig. 5a illustrates the superior performance of the proposed CSC system, which achieves a semantic reliability of 0.95 at 0 dB. In contrast, the SC system that does not consider causal structure can achieve the same level of reliability only at 7 dB. This is attributed to the fact that the proposed CSC system requires fewer samples to achieve the desired reliability on the test data set compared to the SC system, which fails to leverage causality. Fig. 5b shows the evaluation results for symbol error rate across different baselines. The results indicate that the proposed CSC system significantly outperforms the traditional SC system, achieving a 10 dB decrease in SER (in dB) at an SNR of 8 dB. Furthermore, compared to classical wireless systems, the proposed CSC system demonstrates a 40 dB decrease in SER, highlighting its superior capabilities in



(a)

(b)

Fig. 7: (a) Semantic effectiveness under non-stationary learning environments (corresponding to change in dynamics $\boldsymbol{g}_s$). (b) Number of physical bits to be transmitted to achieve a particular SER (corresponding to same semantic reliability).

achieving higher data rates and lower latency.

Fig. 6a presents the evaluation results for normalized mean squared error (NMSE) in the learned model between expert and imitator agents. The results demonstrate that the proposed CSC system significantly outperforms the maximum likelihood (MLE) baselines that utilizes a linear approximation [32] (with just a history of length 1) of the autoregressive model assumed for the dataset, achieving higher model accuracy across all SNRs. This is particularly important for complex datasets, where linear models may perform poorly due to potential nonlinearities. These findings highlight the importance of improving physical model accuracy using advanced AI algorithms, such as causal discovery. In Fig. 6b, we evaluate the generalization performance of the proposed MBRL scheme at the imitator compared to MLE based learner. Here, we computed the model error ($\boldsymbol{g}_s^E - \boldsymbol{g}_s^\eta$) between expert and imitator agents for the proposed CSC and the MLE. We observe that our MBRL performs much better than the standard model learner (MLE), especially when the number of samples available is low. This is attributed to the causal structure extraction of the environment dynamics, which leads to better generalizability.

Fig. 7a shows the performance of the proposed CSC system and the causality unaware baseline in a non-stationary learning environment. The goal here is to demonstrate the quick adaptability (with minimal training effort compared to state-of-the-art) of the CSC approach proposed to changes in the transition dynamics $\boldsymbol{g}_s$ of the time series data used in the simulations. In this case, the transition dynamics $\boldsymbol{g}_s$ change after around 5.7K samples. The results demonstrate that the proposed CSC system efficiently recovers from dynamics perturbations, requiring approximately three times fewer samples compared to the causality-unaware system, which necessitates more retraining efforts. Hence, the proposed approach is effective for generalizing to multiple environments. In Fig. 7b, we observe that the CSC system outperforms conventional communication systems and other SC baselines in terms of efficiency (measured in the number of bits required to convey the same amount of semantic information). Specifically, we analyze the number of bits transmitted to achieve a particular SER for $10,000$ communication instances. The CSC system requires significantly fewer bits (by a factor of 1000) compared to the classical system (without reasoning part), highlighting the significance of our proposed approach. Furthermore, when compared to state-of-the-art algorithms that do not consider causal reasoning, the proposed CSC approach is more robust to SER.

VI. CONCLUSION

In this paper, for the first time in the literature, we have presented a new vision of a DT-based SC system entitled CSC that relies upon the recently emerged theory of consciousness measures based on IIT and AI tools such as IL and MBRL. This approach enables causal discovery of the network state, allowing it to be generalized across multiple wireless environments. We have formulated a bi-level optimization based on variational inference and information bottleneck principle to learn causal discovery, state transitions, and semantic representation at the transmitter (expert agent). The optimized semantic representation is such that the receiver is able to extract maximum semantic information while at the same time revealing minimum information about the state. At the receiver (the imitator node in IL), using a generative AI architecture originally proposed for "world models" in RL, the node improves its knowledge about the network state transitions and causality over time. The proposed bi-level optimization is formulated using the principles of variational inference and IDS. We have shown analytically the performance shortfall associated with a suboptimal environment learning at the imitator. Simulation results demonstrate our proposed CSC's superiority in improving communication efficiency (minimal transmission) and reliability compared to classical communication and state-of-the-art SC systems.

REFERENCES

[1] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital Twin in Industry: State-of-the-Art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.

[2] L. U. Khan, Z. Han, W. Saad, E. Hossain, M. Guizani, and C. Seon Hong, "Digital Twin of Wireless Systems: Overview, Taxonomy, Challenges, and Opportunities," *IEEE Communications Surveys and Tutorials*, vol. 24, no. 4, 2022.

[3] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 96–102, Jun. 202.

[4] E. Uysal, O. Kaya, A.Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assad, G. Liva, A. Munari, B. Soret, and T. Soleymani, “Semantic Communications in Networked Systems: A Data Significance Perspective,” *IEEE Network*, vol. 36, no. 4, pp. 233–240, Oct. 2022.

[5] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, “Less Data, More Knowledge: Building Next Generation Semantic Communication Networks,” *arXiv preprint arXiv:2211.14343*, Nov. 2022.

[6] H. Xie, Z. Qin, G. Y. Li, and B-H. Juang, “Deep Learning Enabled Semantic Communication Systems,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, Apr. 2021.

[7] S. Seo, J. Park, S-W Ko, J. Choi, M. Bennis, and S-L Kim, “Towards Semantic Communication Protocols: A Probabilistic Logic Perspective,” *arXiv:2207.03920v1*, Jul. 2022.

[8] J. Liu, W. Zhang, and H. V. Poor, “A Rate-Distortion Framework for Characterizing Semantic Information,” in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Grenoble, France, Jul. 2021.

[9] M. K. Farshbafan, W. Saad, and M. Debbah, “Curriculum Learning for Goal-Oriented Semantic Communications with a Common Language,” *arXiv preprint arXiv:2111.08051*, Feb. 2022.

[10] M. K. Farshbafan, W. Saad, and M. Debbah, “Common Language for Goal-Oriented Semantic Communications: A Curriculum Learning Framework,” in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2022.

[11] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, “Is Semantic Communications Secure? A Tale of Multi-Domain Adversarial Attacks,” *arXiv preprint arXiv:2212.10438*, Dec. 2022.

[12] W. Sun, S. Lei, L. Wang, Z. Liu, and Y. Zhang, “Adaptive Federated Learning and Digital Twin for Industrial Internet of Things,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5605–5614, Aug. 2021.

[13] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, “Low-latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G Networks,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5098–5107, Jul. 2021.

[14] O. Hashash, C. Chaccour, W. Saad, K. Sakaguchi, and T. Yu, “Towards a decentralized metaverse: Synchronized orchestration of digital twins and sub-metaverses,” in *IEEE International Conference on Communications (ICC), Next-generation Networking and Internet Symposium*, Rome, Italy, May 2023.

[15] Y. Dai, K. Zhang, S. Maharjan, and Y. Zhang, “Deep Reinforcement Learning for Stochastic Computation Offloading in Digital Twin Networks,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4968–4977, Jul. 2021.

[16] Y. Lu, S. Maharjan, and Y. Zhang, “Adaptive Edge Association for Wireless Digital Twin Networks in 6G,” *IEEE Internet of Things Journal*, vol. 8, no. 22, pp. 16219–16230, Nov. 2021.

[17] D. V. Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong, “Edge Intelligence-Based Ultra-Reliable and Low-Latency Communications for Digital Twin-Enabled Metaverse,” *IEEE Wireless Communications Letters*, vol. 11, no. 8, Aug. 2022.

[18] C. Ruah, O. Simeone, and B. Al-Hashimi, “A Bayesian Framework for Digital Twin-Based Control, Monitoring, and Data Collection in Wireless Systems,” *arXiv preprint arXiv:2212.01351*, 2023.

[19] O. Takayuki, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, and J. Peters, “An Algorithmic Perspective on Imitation Learning,” *Foundations and Trends in Robotics*, vol. 7, no. 1-2, pp. 1–179, Mar. 2018.

[20] Y. Xiao, Z. Sun, G. Shi, and D. Niyato, “Imitation Learning-based Implicit Semantic-aware Communication Networks: Multi-layer Representation and Collaborative Reasoning,” *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 1, Mar. 2023.

[21] R. Vuorio, J. Brehmer, H. Ackermann, D. Dijkman, T. Cohen, and P. de Haan, “Deconfounded Imitation Learning,” *arXiv preprint arXiv:2211.02667*, 2022.

[22] J. Pearl, “Causal inference,” *International Journal on Biostat*, vol. 2, pp. 39–58, 2010.

[23] L. Albantakis, L. Barbosa, G. Findlay, M. Grasso, A. M. Haun, W. Marshall, Mayner W. G, A. Zaeemzadeh, M. Boly, and B. E. Juel amd S. Sasai, “Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms,” *arXiv preprint arXiv:2212.14787*, 2022.

[24] A. Haun and G. Tononi, “Why does space feel the way it does? Towards a principled account of spatial experience.,” *Entropy*, vol. 21, no. 12, 2019.

[25] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, M. Sugiyama, E. Uchibe, and J. Morimoto, “Deep Learning, Reinforcement Learning, and World Models,” *Neural Networks, Special issue on Artificial Intelligence and Brain Science*, vol. 152, pp. 267–275, Aug. 2022.

[26] V. D. Nguyen, T. X. Vu, N. T. Nguyen, D. C. Nguyen, M. Juntti, N. C. Luong, D. T. Hoang, D. N. Nguyen, and S. Chatzinotas, “Network-Aided Intelligent Traffic Steering in 6G ORAN: A Multi-Layer Optimization Framework,” *arXiv preprint arXiv:2302.02711*, Feb. 2023.

[27] D. Ha and J. Schmidhuber, “Recurrent World Models Facilitate Policy Evolution,” in *Advances in Neural Information Processing Systems 31 (NeurIPS)*, 2018.

[28] T. Geffner, J. Antoran, A. Foster, W. Gong, C. Mao, E. Kiciman, A. Sharma, A. Lamb, M. Kukla, M. Pawlowski, M. Allamanis, and C. Zhang, “Deep End-to-end Causal Inference,” *arXiv preprint arXiv:2202.02195*, 2022.

[29] T. M. Moerland, J. Broekens, A. Plaat, and C. M. Jonker, “Model-based reinforcement learning: A survey ,” *Foundations and Trends® in Machine Learning*, vol. 16, no. 1, pp. 1–118, 2023.

[30] Pedro A Ortega, Markus Kunesch, Grégoire Delétang, Tim Genewein, Jordi Grau-Moya, Joel Veness, Jonas Buchli, Jonas Degrave, Bilal Piot, Julien Perolat, Tom Everitt, Corentin Tallec, Emilio Parisotto, Tom Erez, Yutian Chen, Scott Reed, Marcus Hutter, Nando de Freitas, and Shane Legg, “Shaking the foundations: delusions in sequence models for interaction and control,” *arXiv preprint arXiv:2110.10819*, 2021.

[31] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, “Toward Causal Representation Learning,” *Proceedings of the IEEE*, vol. 109, no. 5, Feb. 2021.

[32] M. Tomar, A. Zhang, R. Calendra, M. E. Taylor, and J. Pineau, “Model-Invariant State Abstractions for Model-Based Reinforcement Learning ,” *arXiv preprint arXiv:2102.09850*, 2021.

[33] C. K. Thomas and W. Saad, “Neuro-Symbolic Causal Reasoning Meets Signaling Game for Emergent Semantic Communications,” *IEEE Transactions on Wireless Communications*, Oct. 2023.

[34] D. Pathak, P. Agrawal, A. Efros, and T. Darrell, “Curiosity-driven Exploration by Self-supervised Prediction,” in *In International conference on machine learning (ICML), PMLR*, 2017, pp. 2778–2787.

[35] M. Oizumi, L. Albantakis, and G. Tononi, “From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0,” *PLoS computational biology*, vol. 10, no. 5, 2014.

[36] D. Abel, D. Arumugam, L. Lehnert, and M. Littman, “State Abstractions for Lifelong Reinforcement Learning,” in *International Conference on Machine Learning, PMLR*, July 2018, pp. 10–10.

[37] I. Higgins et al., “β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework,” in *ICLR*, Apr. 2017.

[38] Y. Shao, Q. Cao, and D. Gunduz, “A Theory of Semantic Communication,” *arXiv preprint arXiv:2212.01485*, 2022.

[39] A. Hatcher, “Algebraic Topology,” in *Cambridge University Press: Cambridge, UK*, 2002.

[40] O. Caramello, “Theories, Sites, Toposes: Relating and Studying Mathematical Theories Through Topos-theoretic’bridges,” *Oxford University Press*, 2018.

[41] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing, “DAGs with no tears: Continuous optimization for structure learning,” *arXiv preprint arXiv:1803.01422*, 2018.

[42] D. M Blei, A. Kucukelbir, and J. D. McAuliffe, “Variational Inference: A Review for Statisticians,” *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[43] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The Graph Neural Network Model,” *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 561–80, Dec. 2008.

[44] S. Löwe, D. Madras, R. Zemel, and M. Welling, “Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data,” *Proceedings of Machine Learning Research*, vol. 140, pp. 509–525, Jun. 2022.

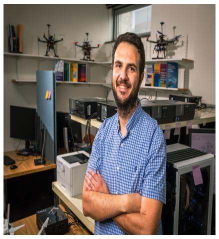
[45] E. Beck, C. Bockelmann, and A. Dekorsy, "Semantic Information Recovery in Wireless Networks," *arXiv:2204.13366v3*, Mar. 2023.

[46] A. Rajeswaran, K. Lowrey, E. Todorov, and S. Kakade, "Towards Generalization and Simplicity in Continuous Control," in *Neural Information Processing Systems*, 2017.

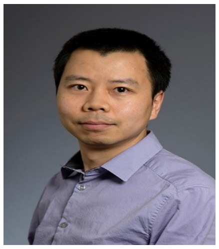
[47] D. Arumugam and B. V. Roy, "The value of information when deciding what to learn," *Advances in Neural Information Processing Systems*, vol. 34, pp. 9816–9827, May 2021.

[48] T. Kipf, E. Fetaya, K-C. Wang, M. Welling, and R. Zemel, "Neural Relational Inference for Interacting Systems," *arXiv preprint arXiv:1802.04687*, 2018.

Christo Kurisummoottil Thomas (S'17, M'20) received the B.S. degree in electronics and communication engineering from the National Institute of Technology, Calicut, India, in 2010, the M.S. degree Institute of Science, Bengaluru, India, in 2012, and the Ph.D. degree from EURECOM, France, in 2020. From 2012 to 2014, he was a Staff Design Engineer on 4G LTE at Broadcom Communications, Bengaluru. From 2014 to 2017, he was a Design Engineer at Intel Corporation, Bengaluru. From November 2020 to June 2022, he was a Staff Engineer on 5G modems at the Wireless Research and Development Division, Qualcomm Inc., Espoo, Finland. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, Virginia Tech. His research interests include semantic communications, statistical signal processing, and machine learning for wireless communications. He was a recipient of the Best Student Paper Award from the IEEE SPAWC 2018, Kalamata, Greece. He also received third prize for his team titled "Learned Chester" ML5G-PHY channel estimation challenge, as part of the ITU AI/ML in 5G challenge, conducted at NCSU, USA, in 2020.



Walid Saad (S'07, M'10, SM'15, F'19) received his Ph.D degree from the University of Oslo, Norway in 2010. He is currently a Professor at the Department of Electrical and Computer Engineering at Virginia Tech, where he leads the Network sciEnce, Wireless, and Security (NEWS) laboratory. He is also the Next-G Wireless Faculty Lead at Virginia Tech's Innovation Campus. His research interests include wireless networks (5G/6G/beyond), machine learning, game theory, security, UAVs, semantic communications, cyber-physical systems, and network science. Dr. Saad is a Fellow of the IEEE. He is also the recipient of the NSF CAREER award in 2013, the AFOSR summer faculty fellowship in 2014, and the Young Investigator Award from the Office of Naval Research (ONR) in 2015. He was the (co-)author of eleven conference best paper awards at IEEE WiOpt in 2009, ICIMP in 2010, IEEE WCNC in 2012, IEEE PIMRC in 2015, IEEE SmartGridComm in 2015, EuCNC in 2017, IEEE GLOBECOM (2018 and 2020), IFIP NTMS in 2019, IEEE ICC (2020 and 2022). He is the recipient of the 2015 and 2022 Fred W. Ellersick Prize from the IEEE Communications Society, of the IEEE Communications Society Marconi Prize Award in 2023, and of the IEEE Communications Society Award for Advances in Communication in 2023. He was also a co-author of the papers that received the IEEE Communications Society Young Author Best Paper award in 2019, 2021, and 2023. Other recognitions include the 2017 IEEE ComSoc Best Young Professional in Academia award, the 2018 IEEE ComSoc Radio Communications Committee Early Achievement Award, and the 2019 IEEE ComSoc Communication Theory Technical Committee Early Achievement Award. From 2015-2017, Dr. Saad was named the Stephen O. Lane Junior Faculty Fellow at Virginia Tech and, in 2017, he was named College of Engineering Faculty Fellow. He received the Dean's award for Research Excellence from Virginia Tech in 2019. He was also an IEEE Distinguished Lecturer in 2019-2020. He has been annually listed in the Clarivate Web of Science Highly Cited Researcher List since 2019. He currently serves as an Area Editor for the IEEE Transactions on Network Science and Engineering and the IEEE Transactions on Communications. He is the Editor-in-Chief for the IEEE Transactions on Machine Learning in Communications and Networking.



Yong Xiao (S'06, M'12) received his B.S. degree in electrical engineering from China University of Geosciences, Wuhan, China in 2002, M.Sc. degree in telecommunication from Hong Kong University of Science and Technology in 2006, and his Ph. D degree in electrical and electronic engineering from Nanyang Technological University, Singapore in 2012. He is now a professor in the School of Electronic Information and Communications at the Huazhong University of Science and Technology (HUST), Wuhan, China. He is also with Peng Cheng Laboratory, Shenzhen, China and Pazhou Laboratory (Huangpu), Guangzhou, China. He is the associate group leader of the network intelligence group of IMT-2030 (6G promoting group) and the vice director of 5G Verticals Innovation Laboratory at HUST. Before he joins HUST, he was a research assistant professor in the Department of Electrical and Computer Engineering at the University of Arizona where he was also the center manager of the Broadband Wireless Access and Applications Center (BWAC), an NSF Industry/University Cooperative Research Center (I/UCRC) led by the University of Arizona. His research interests include machine learning, game theory, distributed optimization, and their applications in semantic communications, semantic-aware networking, cloud/fog/mobile edge computing, green communication systems, and Internet-of-Things (IoT).