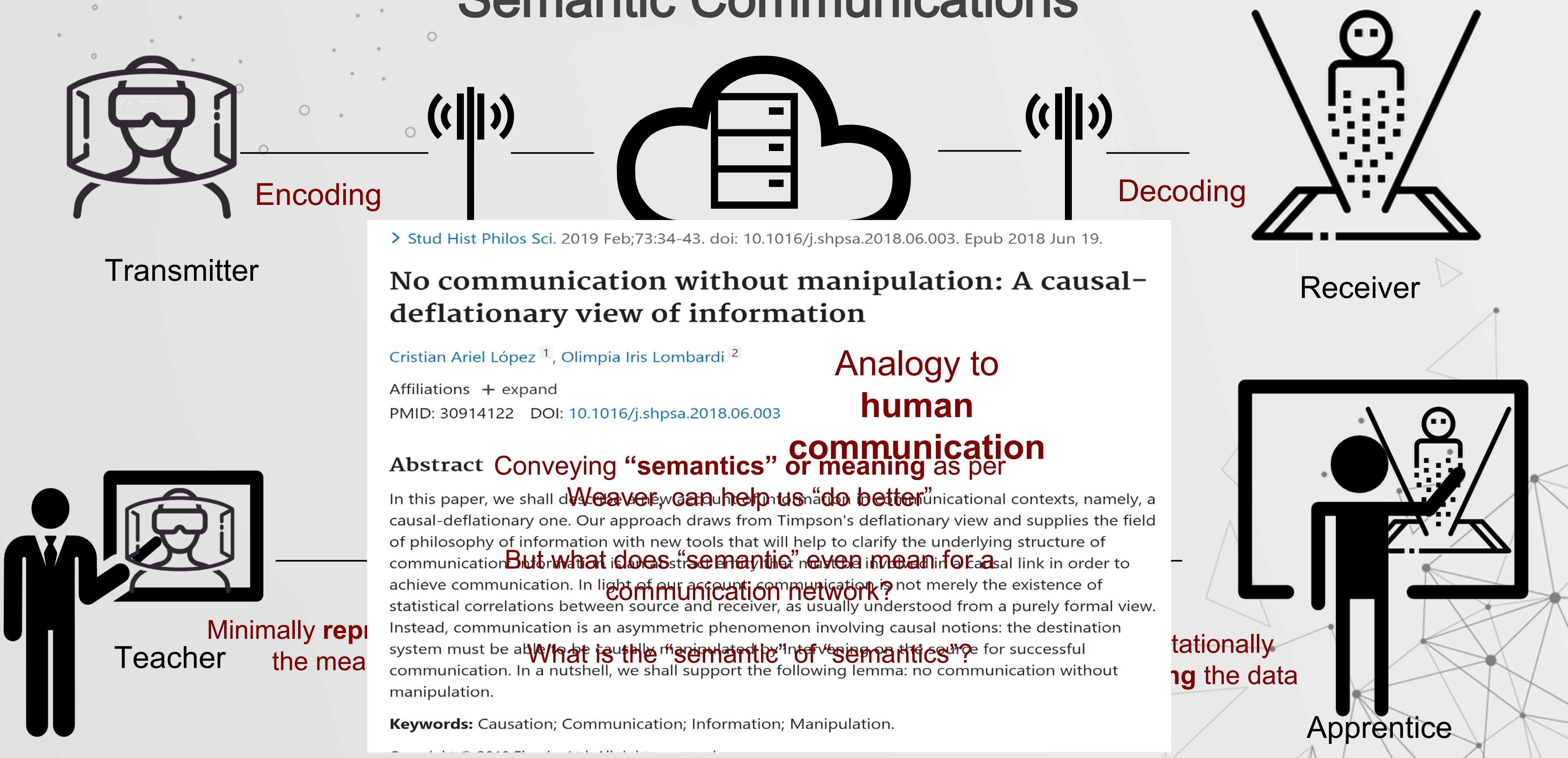
Understand Me If You Can: Reasoning Foundations of Semantic Communication Networks

Walid Saad
Electrical and Computer Engineering Department
Virginia Tech

Key paper
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, 2022.

Semantic Communications



Stud Hist Philos Sci. 2019 Feb;73:34-43. doi: 10.1016/j.shpsa.2018.06.003. Epub 2018 Jun 19.

No communication without manipulation: A causal-deflationary view of information

Cristian Ariel López [1], Olimpia Iris Lombardi [2]

Affiliations + expand

PMID: 30914122 DOI: 10.1016/j.shpsa.2018.06.003

Abstract

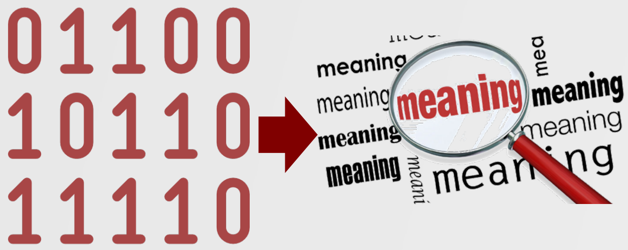
In this paper, we shall describe a new account of information in communicational contexts, namely, a causal-deflationary one. Our approach draws from Timpson's deflationary view and supplies the field of philosophy of information with new tools that will help to clarify the underlying structure of communication. Information is an abstract entity that must be involved in a causal link in order to achieve communication. In light of our account, communication is not merely the existence of statistical correlations between source and receiver, as usually understood from a purely formal view. Instead, communication is an asymmetric phenomenon involving causal notions: the destination system must be able to be causally "manipulated" by "intervening on the source" for successful communication. In a nutshell, we shall support the following lemma: no communication without manipulation.

Keywords: Causation; Communication; Information; Manipulation.

Semantic Communication System

Teacher

Apprentice



- **Transmitter ➔ Knowledge-driven Teacher**

  What is a Teacher capable of?

  ➢ **Disentangling** multiple semantic content elements within to source data.

  ➢ For every semantic content element identified, craft a semantic representation with ***desirable properties.***

- **Receiver ➔ Knowledge-driven Generative Apprentice**

  What is an apprentice capable of?

  ➢ **Understanding** the minimal semantic representation used by the teacher ➔ mapping it to a semantic content element

  ➢ **Generating**, via their computing resources, the semantic content element, with the highest fidelity possible.

- **Bit-pipeline ➔ Semantic Language**

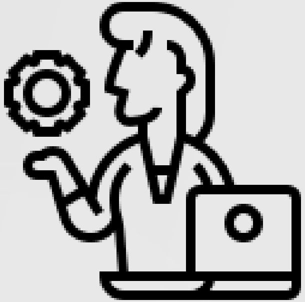
  What is a semantic language?

  ➢ A **series of representations** constitutes a semantic language.

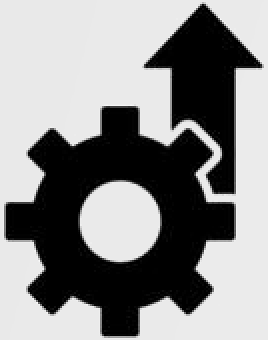
  ➢ Semantic languages will mimic natural languages but they should be **less focused on syntax and pragmatics.**

Key Characteristics of Semantic Representations and a Semantic Language

- **Minimalism**
  - The capability of characterizing the structure found in the information with the least number of language elements possible
  - Reduction of the number of exchanged messages in the long run as well.



- **Generalizability:**
  - Representing a particular underlying structure (or understanding one at the receiving end) while being invariant to changes in: **a) distribution, b) domain, and c) context.**
  - This mimics the behavior of a natural language to universally use words to describe events.



- **Efficiency:**
  - The ability of the apprentice to re-generate the information with **high fidelity**, in the **least time** possible.

➔ The resolution of the data generated at the apprentice must be equal (or better) to that which could be recovered by a classical receiver.

Semantic Communication Systems: The Bigger Picture

Reasoning via Causality

- Learn a language by identifying the root-causes of information.
- Generate semantic content and make logical decisions.

Semantic language

A map between a stream of raw data bits and their respective semantic representation.

Semantic-based KPIs

Novel metrics that can evaluate the performance of semantic networks by characterizing the intrinsic reasoning and level of symmetry of the SC system

Minimally sufficient representation

Minimal representations, yet expressive and can capture the semantic structure of the data.

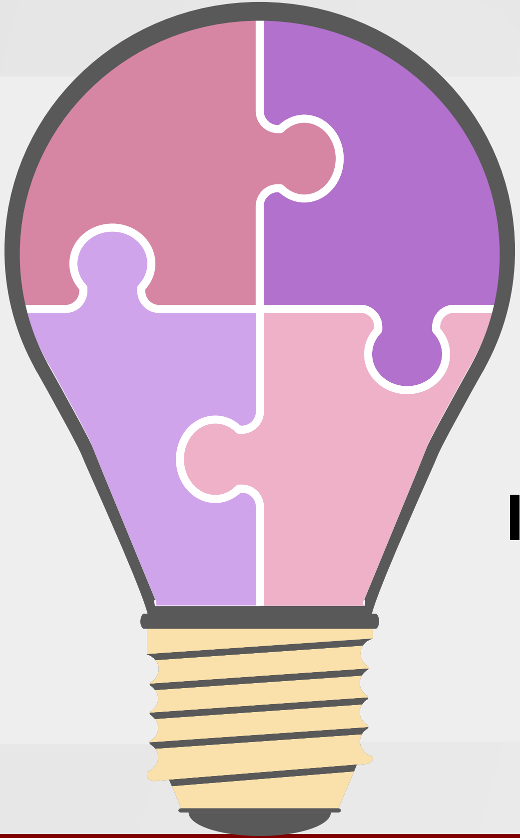
Computing resources

The efficiency of SC grows substantially with the ease, dynamicity, and availability of computing resources

Semantic Communication Backbone

(Some) Benefits of Semantic Communications

AI-Nativeness and Interoperability

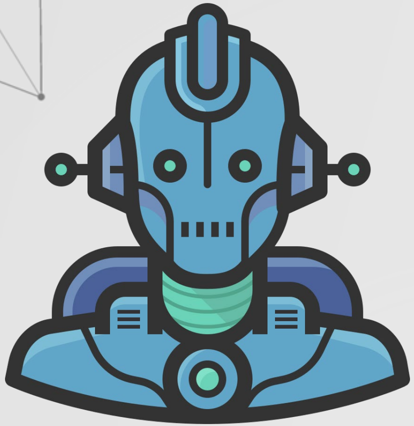


Less Data, More Knowledge

Robust/Resilient Channel Control

Intrinsic Contextual Awareness

Applications?



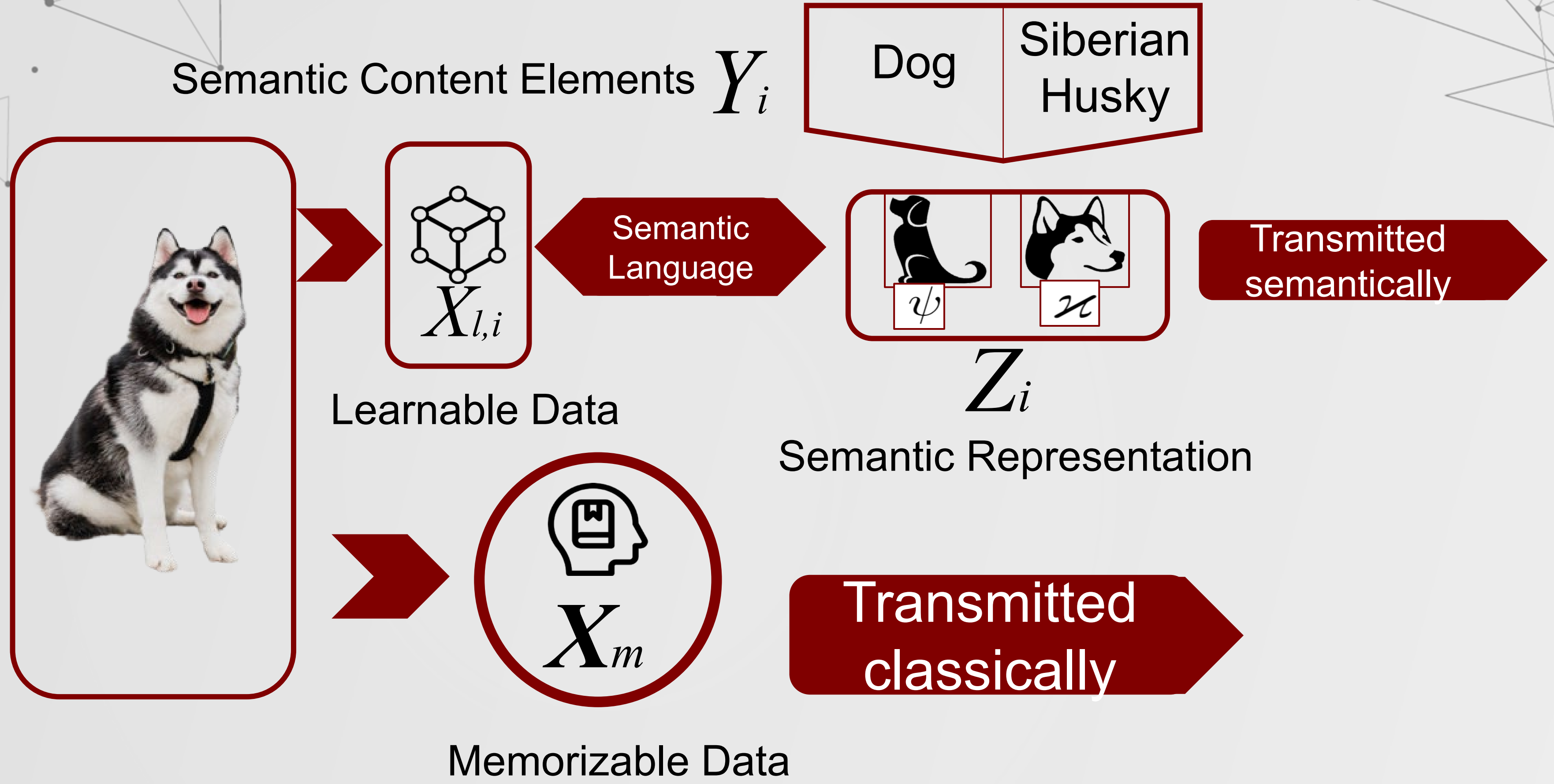
Digital twin control through semantic language

Bridging the **digital divide**: satellite **semantic showers**

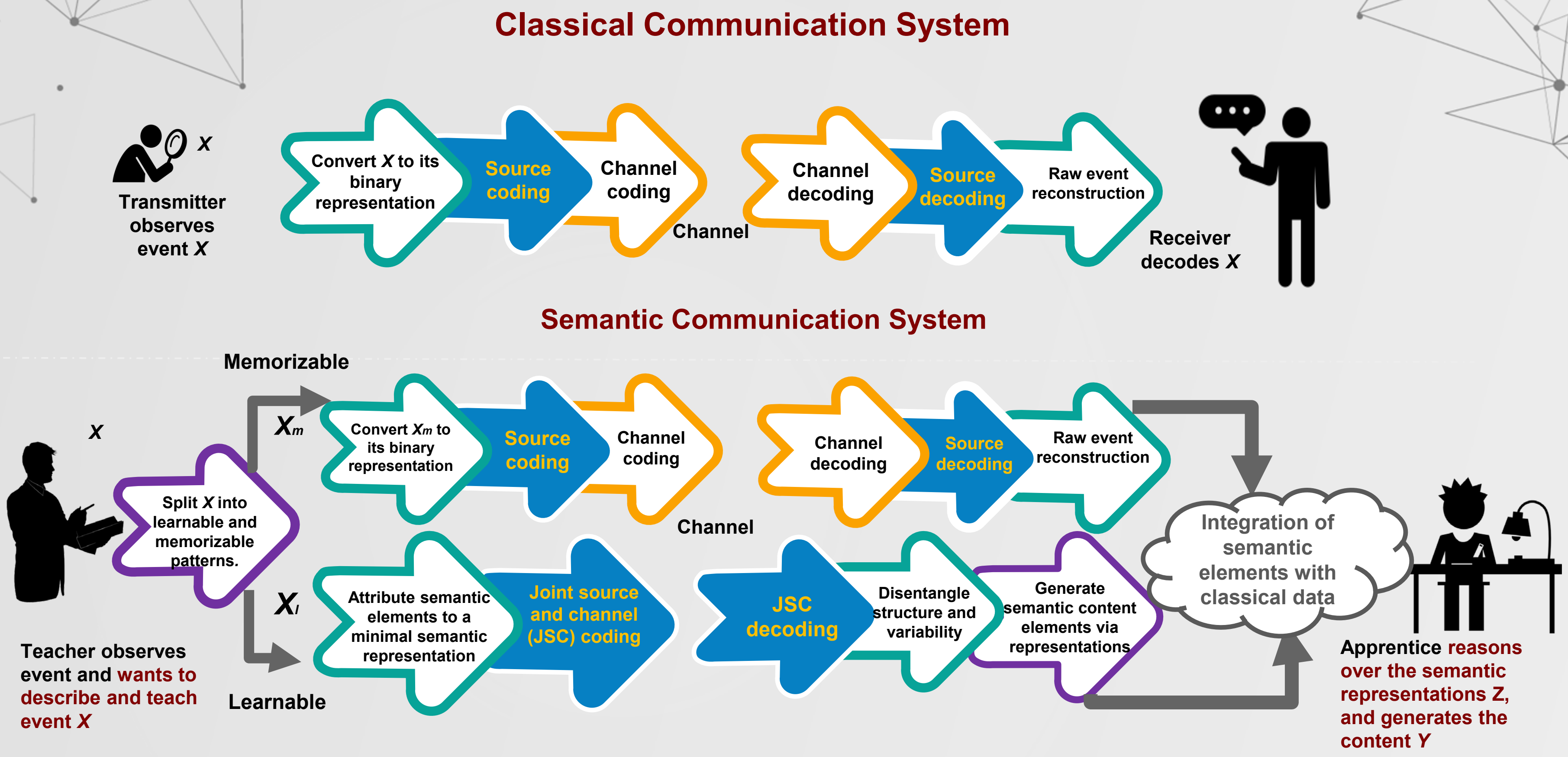
Contextual networking for **Metaverse/digital experiences**

Goal oriented communications

Disentangling Meaning/Semantics



The Bigger Picture



Semantic Language vs Natural Language

Natural Language

- **Syntax-dependent:** Set of deterministic rules to build sentences.
- **Semantics-oriented:** Can only be mapped when defining words.
- **Pragmatics:** Context-dependent features of a language

“Are you wearing a seatbelt”➔ Danger

Constrained to wording, the same way the bit-pipeline is constrained to bits.

Semantic Language

- **Syntax-independent:** Must be characterized with logic not deterministic rules.
- **Semantics-focused:** Acts more like a coded language that commands tasks.
- **Pragmatics:** Improve dynamic reasoning if present, not necessary

Reasoning and Knowledge Driven

Semantic Language: From Entropy to Language Complexity

- *A semantic language $\mathcal{L} = (X_{l,i}, Z_i)$, is a dictionary (from a data structure perspective) that maps the learnable data points $X_{l,i}$ to their corresponding semantic representation Z_i, based on the identified semantic content elements Y_i.*

Proposition 1 *The complexity of a specific language $\mathcal{L}$ adopted among a teacher and apprentice pair is given by:*

$$\Gamma(\mathcal{L}) = \min_{p(Z|\mathbf{X}_l)} \underbrace{L_{\mathcal{L}}(p)}_{\text{Cross Entropy loss}} + \underbrace{K(p)}_{\text{Kolmogorov complexity}}.$$

- Capture the **fitness of the representation** in expressing the content elements and the **Kolmogorov complexity** of the model built.
- Complexity too high➔ Service content is of complex structure **OR** X_l and X_m separation performed poorly.
- Kolmogorov complexity enables characterizing the individuality of the semantic content elements.
- The structure function achievable by a model *p* for a language $\mathcal{L}$ is given by:

$$\Psi_{\mathcal{L}}(t) = \min_{K(p) \leq t} L_{\mathcal{L}}(p).$$

- The structure function tends to zero for sufficiently high complexity ➔ data tends to purely random information that lacks structure➔ very difficult learning task ➔ easy memorization task.
- Achieving structure – complexity tradeoff via optimization

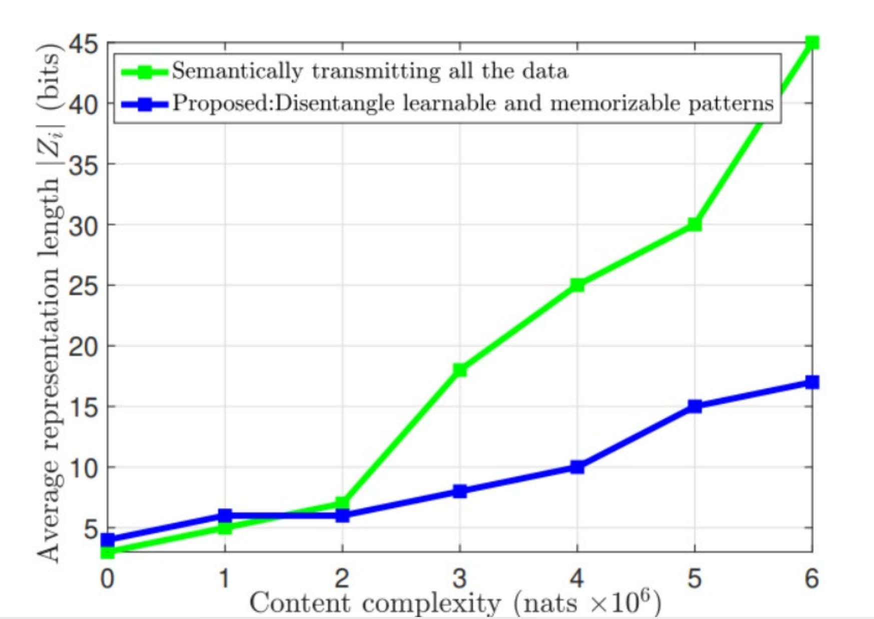
Key Result: Disentangling Learnable and Memorizable Data via Contrastive Learning for Semantic Communications

Christina Chaccour and Walid Saad

Published in the Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA

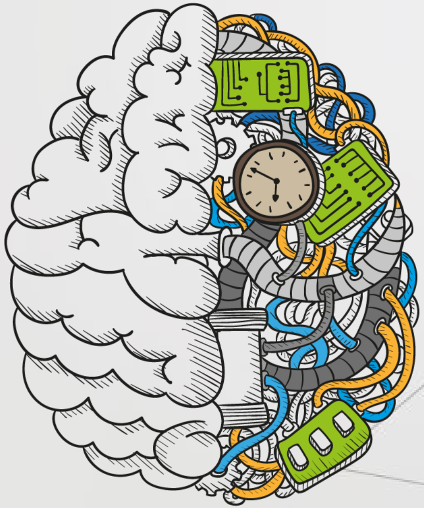
Simulation Results

- This work uses contrastive learning to perform the pre-processing/disentanglement process
- **We can see that the average representation length increases with the content complexity:**
  - For a low content complexity, semantically transmitting all the data might result in a smaller representation length. This is because the amount of random information ➔ $\boldsymbol{X_m}$ is considerably small.
  - As we ↑ the content complexity ➔ Semantically transmitting all the data is not a feasible approach ➔ Representation length steeply increases as we increase the content complexity.
  - Our **representation is minimized by 57.22%** compared to the vanilla semantic approach.
- Now that we know how to disentangle information, let's go deeper into **reasoning and causality**

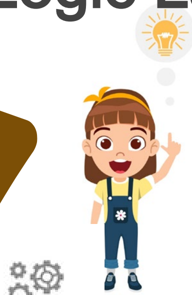


Why do we need Causality?

- Reasoning and “real” learning can only be performed by asking questions➔ Queries (counterfactuals and interventions), the emerging framework of causality enable this.
- Reasoning mainly relies on *characterizing causal and associational logic in the data.*
- We cannot rely on state-of-the-art ML frameworks that make assumptions such as:
  - ❖ i.i.d. datasets
  - ❖ Stationarity scenarios
  - ❖ Data has no root-cause



Causal Logic Ladder



Reasoning

Congregating associative, interventional, and counterfactual logic to understand the representations conveyed and generate representations with their proper semantic connotation.

Counterfactual Logic

Learning with retrospection and imagination. The apprentice is attempting to ask the "Why?" questions when it comes to the current representations used by the teacher and their respective semantics. "What is the root cause of a particular representation?"

Interventional Logic

Learning while invoking questions with the do operator. That is, the apprentice is attempting to learn what would happen in case the causes were different. In other words, the apprentice is asking "What if?". "What would the representation be if the semantics were different?"

Associative Logic

Learning information based on purely statistical relationships without invoking any causality or semantics within the data. This is a purely observational task on the datastream.

Fundamentals of Causal Reasoning

How is causality important for semantic communications?

Constructing a semantic language with causal reasoning capabilities requires mapping the language to a structural causal model (SCM) $\mathcal{L} := (\psi_L, p(\epsilon))$ *where* $\psi_L = \{s_i\}_{i=1}^{N}$. *The learnable data can now be written:*

$$X_{l,i} := s_i(\epsilon_i, \rho_i).$$

Set of direct causes leading to $X_{l,i}$

Exogeneous variable
➔ related to variability

Insights:

- Defining a language that can map to an SCM is a key step
  ➔ Such a language can implement counterfactuals and interventions.
- That is the apprentice can ask questions via *do*-operators

Interventions➔ What if we change the cause…?
Counterfactuals ➔ Why is the current causal link leading to…?

Fundamentals of Causal Reasoning

Causality enables disentangling semantic content elements!
➔Allowing the teacher and apprentice to reason every each meaning and its cause separately

Building a semantic language $\mathcal{L}$ that can be mapped to an SCM model enables disentangling each data stream and its respective representation from other established representations. In other words, the model describing the language can be written as:

$$P(\boldsymbol{X}_l) = P(X_{l,1}, \ldots, X_{l,N}) = \prod_{i=1}^{M} P(X_{l,i}|\rho_i),$$

where $M \leq N$.

Insights:

- **Performing an intervention or a counterfactual on one mechanism, does not change any of the others.**
- **Acquiring information about a specific mechanism $P(X_{l,i}|\rho_i)$ does not give us any information about the other $P(X_{l,i}|\rho_j)$.**

Fundamentals of Causal Reasoning

How do we define a generalizable reasoning system?

A semantic representation is dubbed, **generalizable**, if it fulfills the general causal invariant prediction criterion. That is, despite different "what if"s posed on the causal model, the same representation results in describing its respective content elements in data:

$$p^{do(\kappa_i)}(\boldsymbol{Y}|\boldsymbol{Z}) = p^{do(\kappa_j)}(\boldsymbol{Y}|\boldsymbol{Z}) \forall \kappa_i, \kappa_j \in \mathcal{K},$$

Insights:

- **If one executes different queries ("What ifs" with the different subject), and the exact same learned causal model remains unchanged ➔ The representation and subsequent semantic language is mature.**
- ***This mimics the behavior of words to represent universal events in our daily lives.***

KPI 1: Communication Symmetry Index

The communication symmetry index between a teacher *b* and apprentice *d*, for a transmission session τ is given by:

$$\eta_{b,d,\tau} = \frac{\zeta_{d,\tau}}{\nu_{b,\tau}} \times \iota_{\tau,Y_i},$$

Number of query packets (e.g., interventions, counterfactuals, etc.) needed to reason

Number of raw data packets accompanying semantics

The semantic impact generated by a semantic representation Z_i during a time τ is the number of packets that would have been needed to be transmitted to regenerate the semantic content element Y_i

Insights:

- Based on the values of $\eta_{b,d,\tau}$ and, ι_{τ,Y_i} , one can determine the level of symmetry between the teacher and the apprentice.
- E.g.: A high $\eta_{b,d,\tau}$ ➔ high level of symmetry between teacher and apprentice ➔ the apprentice has generative capabilities. (A high $\eta_{b,d,\tau}$, with low ι_{τ,Y_i}) ➔ reverse mentorship, teacher's capabilities are also weak.

KPI 2: Reasoning Capacity

The reasoning capacity between a teacher b and an apprentice d is given by:

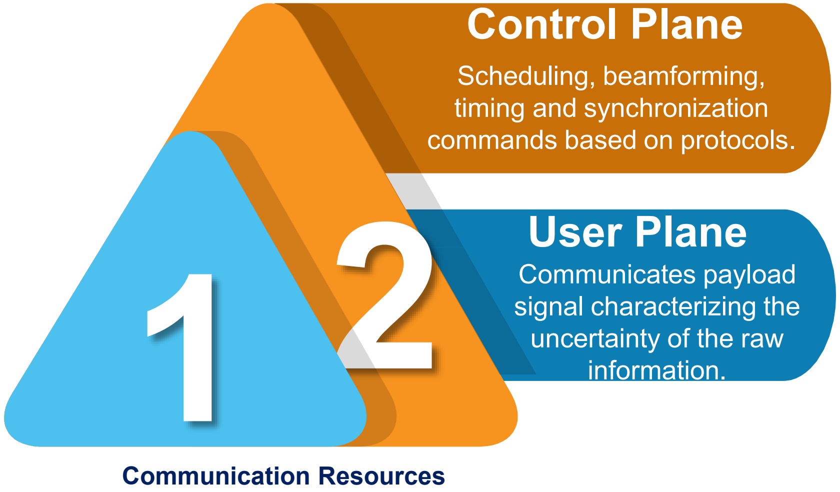
$$C_R = \Omega \log_2(1 + \eta_{b,d}),$$

Maximum Computing Resources

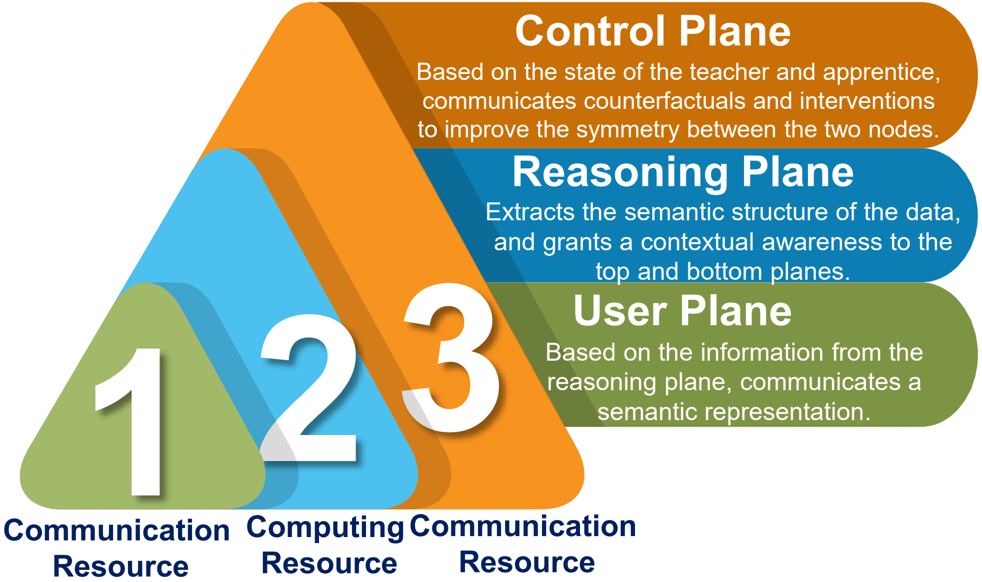
➔ $$C_T = C_C + C_R = W \log_2(1+\gamma) + \Omega \log_2(1 + \eta_{b,d}),$$

The total capacity is no longer limited by Shannon's bound only as a result of the convergence of computing and communications!

How does semantic communication affect the network planes?



5G and Beyond Communication Planes



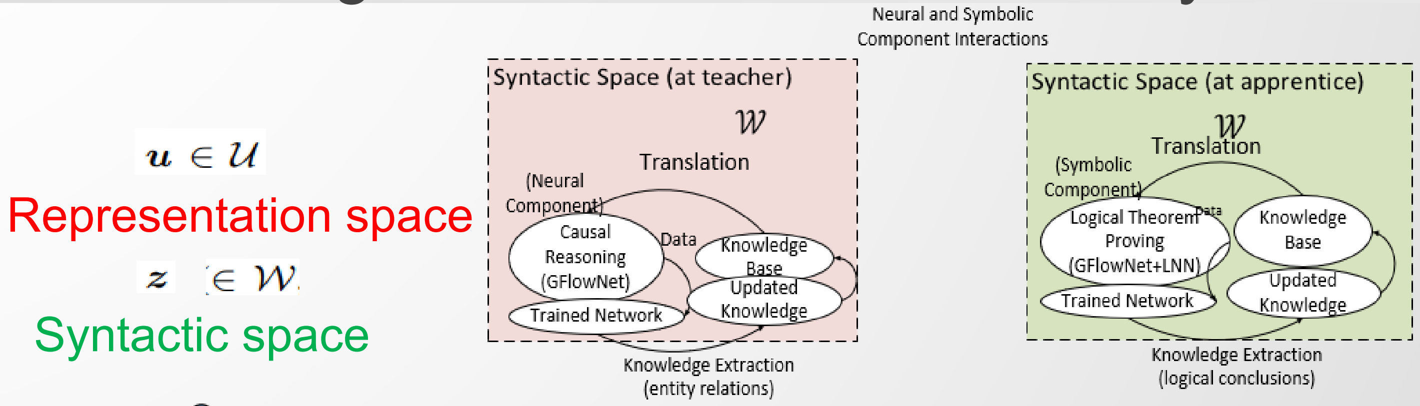
6G and Beyond AI-Native Planes

Key Result: Neuro-Symbolic Causal Reasoning Meets Signaling Game for Emergent Semantic Communications

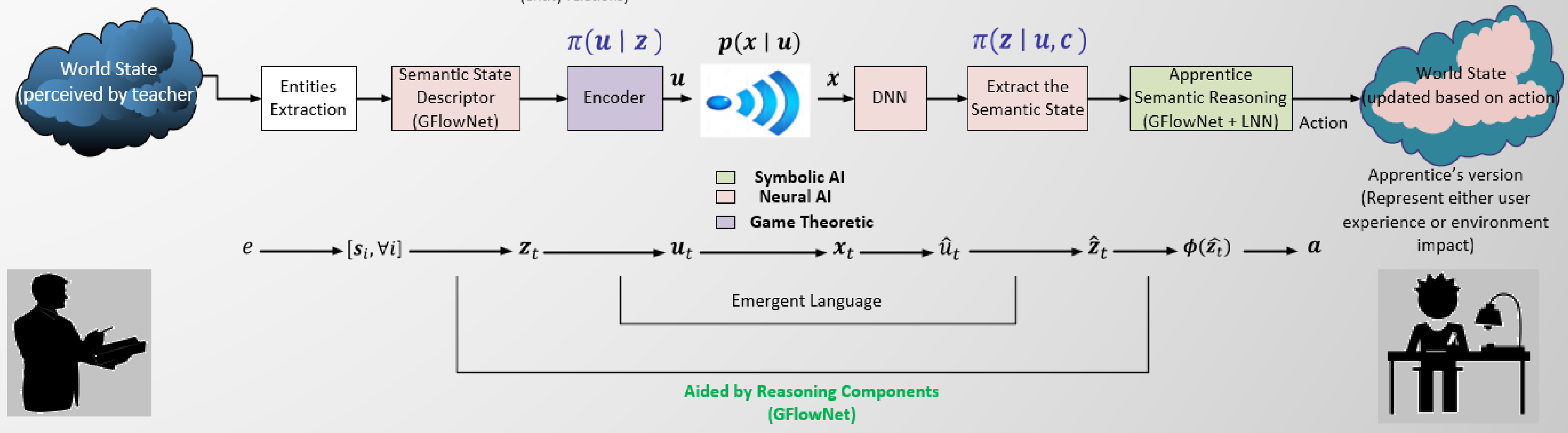
Christo Thomas and Walid Saad

Under review: : https://arxiv.org/abs/2210.12040

Reasoning Semantic Communication System: Overview

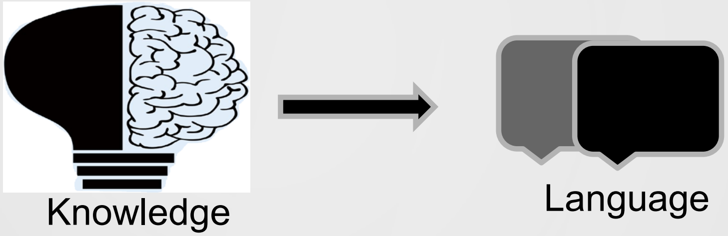


- How to **build the language** => game theory
- **KPIs** => category theory
- How to **reason** over the data (teacher) and **generalize** (apprentice) => neuro-symbolic AI + generative flow networks



Proposed ESC System: Addressing the Challenges

Language Problem



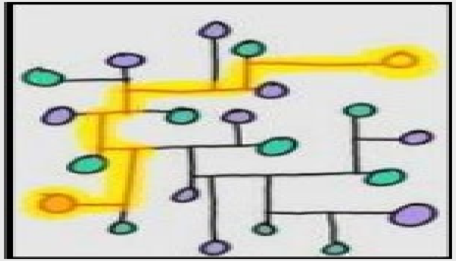
Emergent Language (communicating language emerges)

- Compute teacher transmit strategy (encoder) - $\pi(u \mid z)$ and apprentice inference strategy (decoder) – $\pi(z \mid u, c)$

Encode based on semantics, benefits

- transmitting semantically similar messages as same signal thus saving bits/BW
- removing redundant semantics

Causal Reasoning Problem



Semantic state descriptor:

- Infer the hidden relations among the entities (the causal sequence that best explains the event observed)

Parent nodes in the graph

$$p(s_0, \cdots, s_N \mid e) = \prod_i p(s_i \mid \mathrm{pa}(s_i)),$$

Apprentice semantic reasoning:

Set of logical formulas

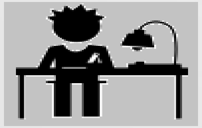
$$p(\phi_i = \phi \mid \hat{z}), \ \forall \phi_i \in \Phi,$$

- evaluate the logical formulas

Two-Player Signaling Game

Semantic notion of information

Teacher objective



Apprentice wants to

$$
\begin{aligned}
\pi^*_{s,t} &\in \arg\max_{\pi_{s,t}} -\mathbb{E}_{\boldsymbol{u}_t}\left[S_s(\boldsymbol{z}_t;\boldsymbol{u}_t \mid \pi_{s,t}, [\boldsymbol{z}_{t-1}], \boldsymbol{c})\right], \\
\text{s.t.} &\qquad \mathbb{E}[V(\pi_{s,t}, \pi_{l,t})] \leq D \\
\pi^*_{l,t} &\in \arg\max_{\pi_{l,t}} \mathbb{E}_{\boldsymbol{u}_t}\left[S_l(\widehat{\boldsymbol{z}}_t;\boldsymbol{u}_t \mid \pi_{l,t}, \pi^*_{s,t}, [\widehat{\boldsymbol{z}}_{t-1}])\right],
\end{aligned}
$$

Transmit less

Extract more

$$V(\pi_{s,t}, \pi_{l,t}) = c(\boldsymbol{u}_t) - \log \pi_{l,t}$$

Computing or Communication costs

Apprentice surprise (to minimize)

- "Category theory to define semantic information: A more general approach compared to set-theoretic methods [1], and it can represent deductive and logical theorem proving properties
  - Syntax category ($\mathcal{L}$) – category of state descriptions (entity or entity-relations).
  - Semantic category (category of copresheaves of all state descriptions part of ($\mathcal{L}$) – represents plausible logical conclusions that entail from any state description. Represented as the functor $\mathcal{F}: \mathcal{L} \to \hat{\mathcal{L}}$.

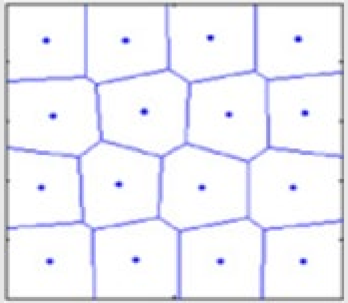
[1] R. Carnap and Y. Bar-Hillel, "An Outline of a Theory of Semantic Information," Technical Report No. 247, Oct. 1952.

Nash Equilibrium Analysis

- Theoretical results on the derivation of the Nash equilibrium
  - Pooling equilibrium $\boxed{(|\mathcal{U}| = 1)}$ and separating equilibrium ($\boxed{|\mathcal{U}| = |\mathcal{W}|}$), not of interest for ESC.
  - $|\mathcal{U}| < |\mathcal{W}|$
  - Partial pooling is realistic ➔ listener extracts max. semantic information when the speaker partition its semantic category space into a Voronoi tessellation, (each => distinct partition). Optimal strategies below:

Speaker – Transmit signal partition $\boldsymbol{u}_k$

Voronoi tessellation of syntactic space $\mathcal{W}$ s.t. avg. semantic info. extracted at listener is maximum among all possible partitions.



Listener

Decoding Strategy: Bayesian estimator

$$\boxed{\arg\min_{z \in \widehat{\mathcal{W}}} \int_{\widehat{\mathbf{W}}} \|z - \hat{z}\|_S \, \pi(d\hat{z} \mid \boldsymbol{u}),}$$

Not in the Euclidean space but in the semantic space!

Potential Gains of ESC vs Classical Wireless : Reduced Bits

- **Theorem 1:** For a particular syntactic space, $\boxed{\mathcal{W}}$ and context distribution $\boxed{p(\boldsymbol{c})}$ over $\boxed{\mathcal{C}}$, the average amount of bits to represent the state description in an ESC system can be bounded as follows.

$$\sum_{\boldsymbol{c}_i} \pi(\boldsymbol{c}_i) H(\boldsymbol{z}_i \mid \boldsymbol{c}) \leq \sum_{\boldsymbol{u}_i \in \mathcal{U}} \pi(\boldsymbol{u}_i) l_i \leq -\sum_{\boldsymbol{c}_i} \pi(\boldsymbol{c}_i) \sum_{\boldsymbol{z}_i} \pi(\boldsymbol{z}_i \mid \boldsymbol{c}_i) \lceil \log \pi(\boldsymbol{z}_i \mid \boldsymbol{c}) \rceil,$$

Codeword length

And for a classical communication system (which directly encodes the entities)

Shannon entropy

$$H(\boldsymbol{s}_i) \leq \sum_{\boldsymbol{u}_i \in \mathcal{U}} \pi(\boldsymbol{u}_i) l_i \leq \max_{\boldsymbol{c}_i} \sum_{\boldsymbol{z}_i} \pi(\boldsymbol{z}_i \mid \boldsymbol{c}_i) [\sum_{\boldsymbol{s}_i \in \boldsymbol{z}_i} \lceil \log \pi(\boldsymbol{s}_i \mid \boldsymbol{c}) \rceil].$$

- **Key point:** for an ESC system, the lower and upper bounds for a physical representation of the semantics are smaller compared to a classical system justifying the transmission efficiency of an ESC system

Potential Gains of ESC vs Classical Wireless : Improved Reliability

Theorem 2: For a given representation space $\boxed{\mathcal{U}}$, the lower bound on the semantic error probability ($\boxed{S_e}$ −representing reliability) is always less than or equal to the lower bound on the probability of bit error ($\boxed{P_e}$) measure achieved using classical communication system.

State descriptions (or just entities) in classical sense

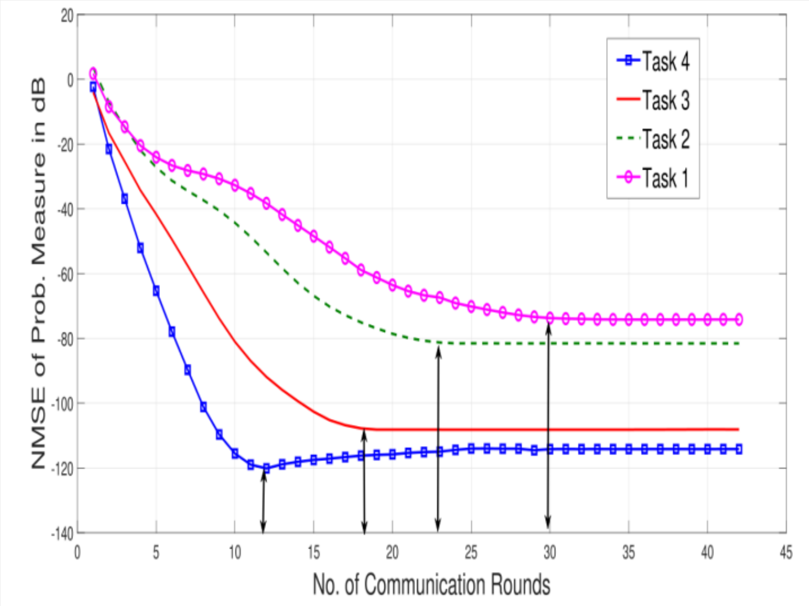
$$P_e \geq \frac{H(\hat{z}_c|z_c)-1}{\log|\mathcal{W}|}, \quad S_e \geq \frac{H(z|\hat{z})-H(e|\hat{z})}{\log|\mathcal{W}|}$$

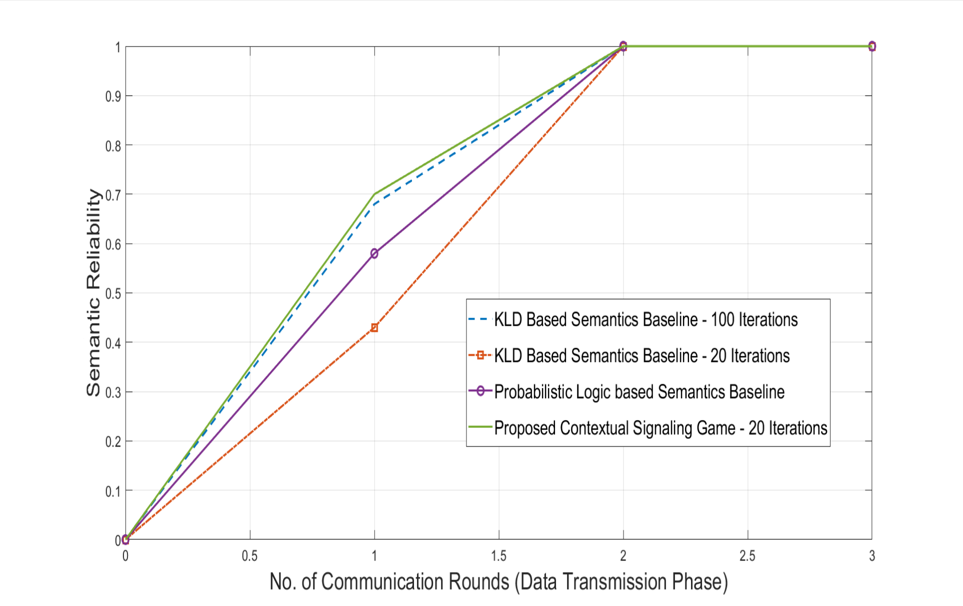
$$where, \frac{H(\hat{z}_c|z_c)-1}{\log|\mathcal{W}|} \geq \frac{H(z|\hat{z})-H(e|\hat{z})}{\log|\mathcal{W}|}.$$

Error in reconstructed semantics

Key point: Inducing reasoning + emergent language at teacher and apprentice can improve semantic reliability compared to a classical system that uses the same number of bits to communicate.

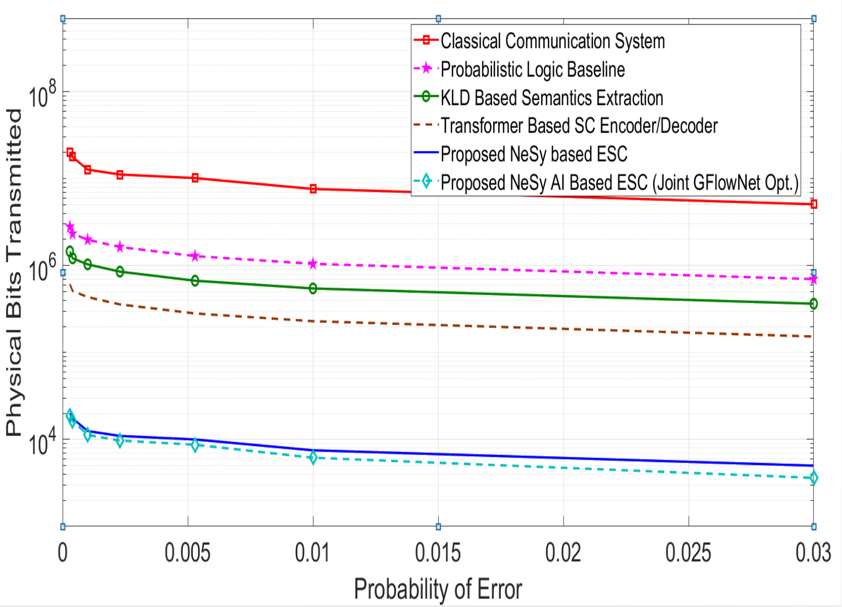
Simulation Results

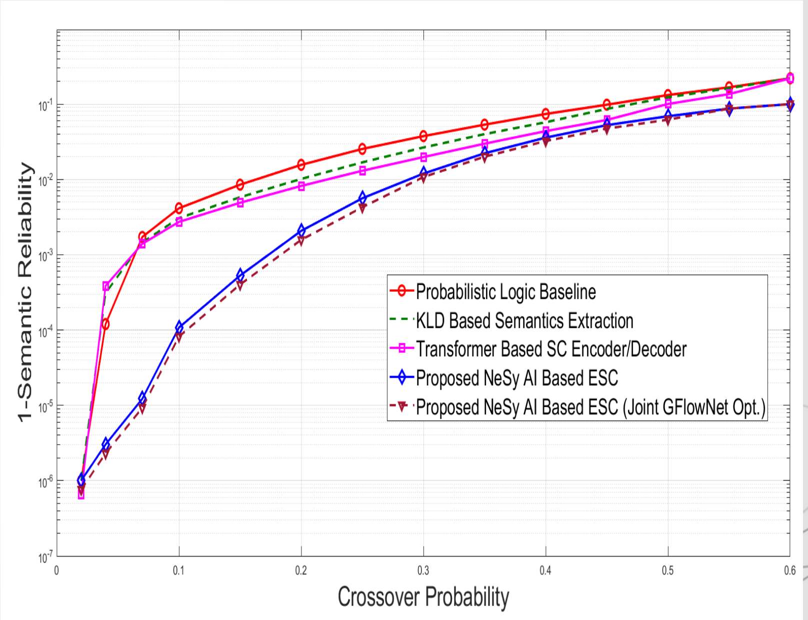




- Number of communication rounds, decreases over time which demonstrates how the generalizable aspects of proposed approach help over time.
- Emergent language gives a much better reliability.

Simulation Results





- Using ESC, the system transmits less compared to state of the art and achieves better semantic reliability
- Semantic error probability (1-reliability) is quite low for ESC compared to SotA until crossover probability 0.3, after which performance becomes
  worse for all since the channel inverts almost half of the bits

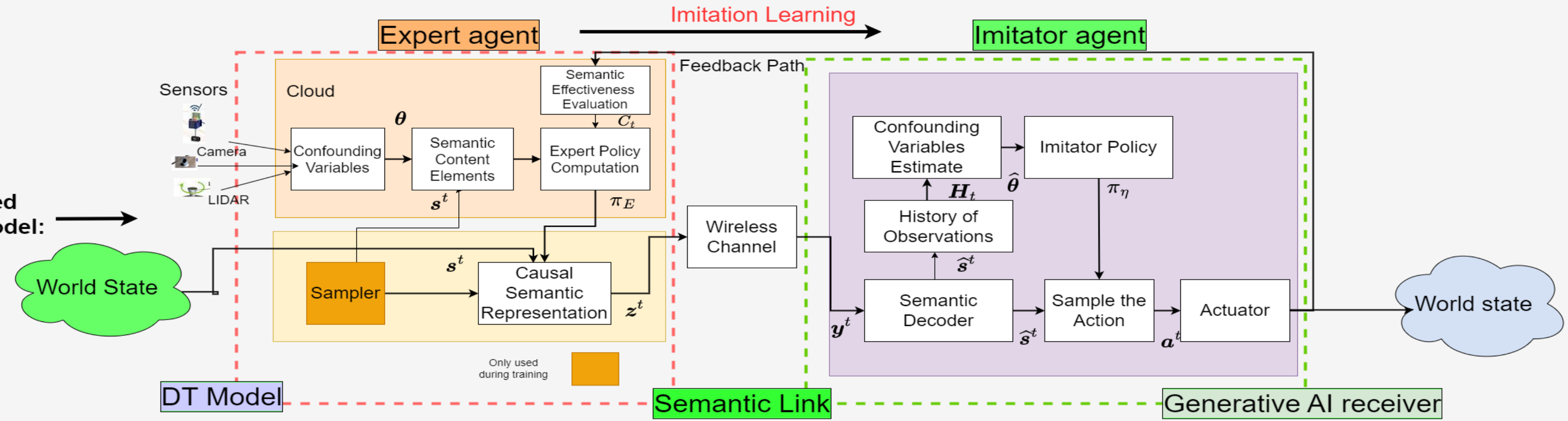
Key Result: Causal Semantic Communication for Digital Twins: A Generalizable Imitation Learning Approach

Christo Thomas, Walid Saad, and Yong Xiao

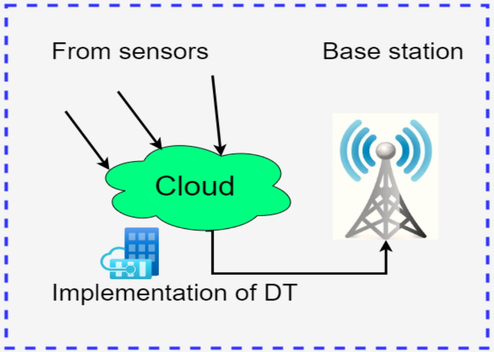
Under review: https://arxiv.org/abs/2304.12502

Causal Semantic Communication for Digital Twins (DTs)

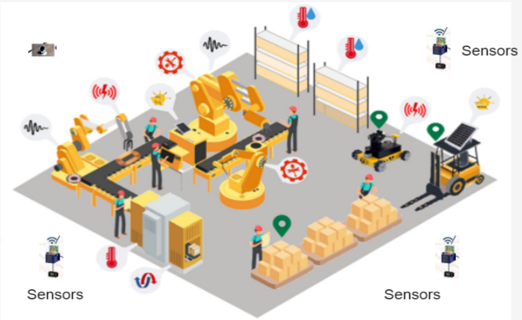
Proposed system model:



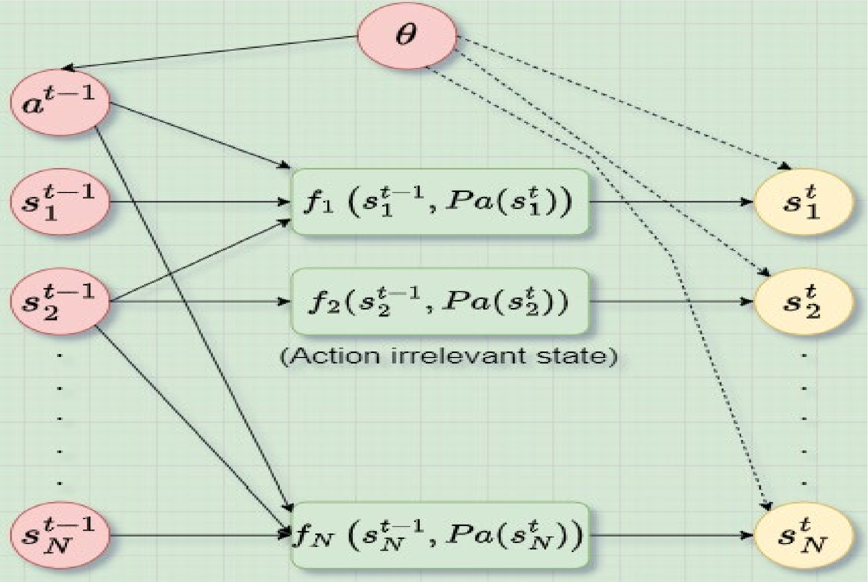
An application:

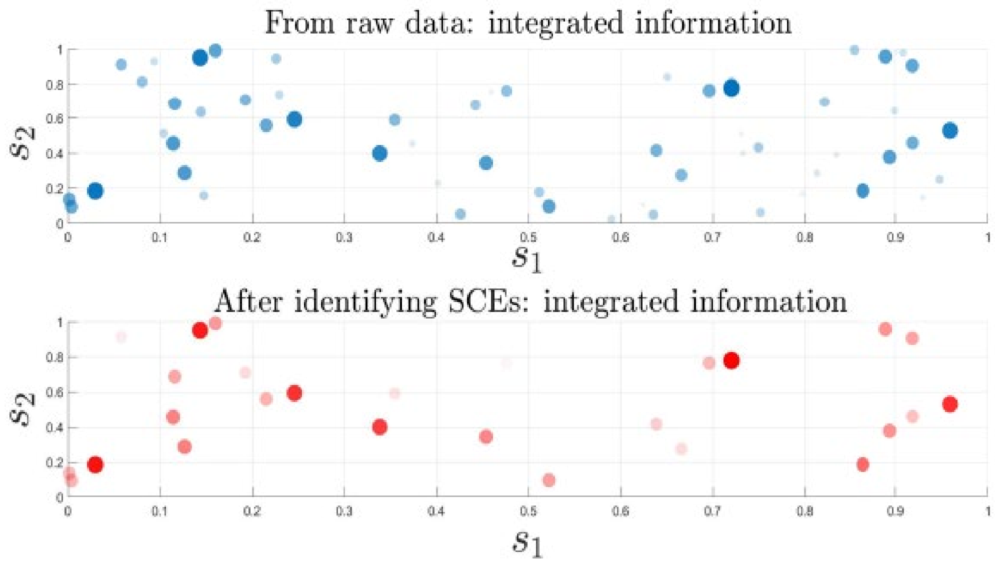


Smart factory environment (controlled using transmit information from DT-aided base station)



Causal Transition Models: Structural Causal Model





- **Semantic Information: uses Integrated information theory (IIT)** from theory of consciousness in neuroscience
- Intrinsic Information for State Abstraction
  - Cause and effect information conveyed by any s_i^t (Impact under confounding variables as theoretical result)
- Information Integration (via Compositionality, for Identifying Semantic Content Elements - SCEs)
  - Information conveyed by a subset of SCEs, as a whole and beyond sum of information of its parts
- Semantic concepts, causal relations among concepts, topological characterization (abstract cell complex) as a theoretical result

Key Analytical Results

Confounding variable estimate

The **error in semantic information** learned between expert and imitator nodes can be made arbitrarily small if the learned posterior about the confounding variables has its peak around the true value of confounders.

Causal relations as simplical complex

The structure of causes and effects in an SCM can be represented using an **abstract simplical complex**, in which causes and effects are nodes and relations are simplices.

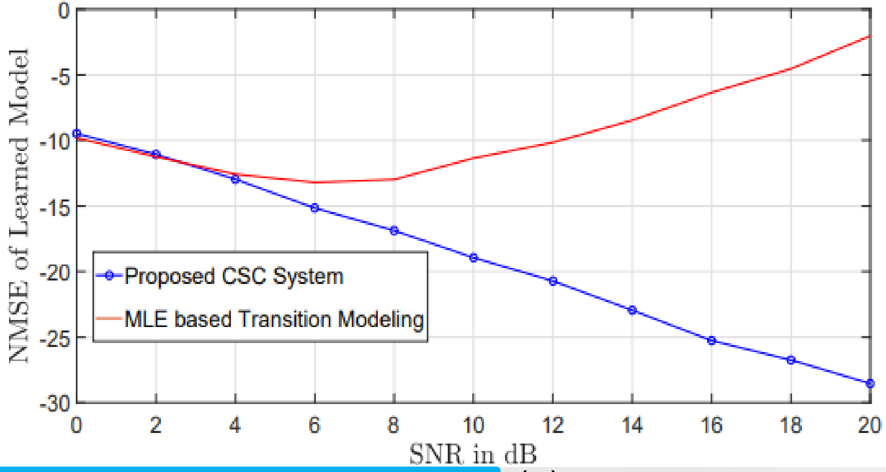
Semantics and Topology

- Topological characterization of semantics as **abstract cell complex**→ represents meaning
- Seamless communication across multiple nodes, without dependency on transmit encoders
- Allows a rigorous formulation of **semantic metrics** → reliability, similarity.

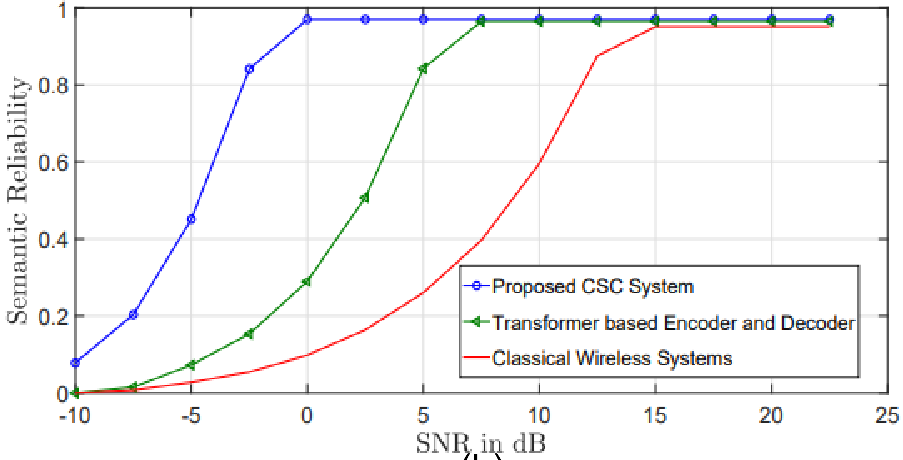
Performance characterization

- Discrepancy in transition probability modeling dominates the error in performance (average regret) → accurate physical environment modeling is crucial.
- **Generative AI** can help to close this gap at the receiver as communication progresses.

Simulation Results



(a)



(b)

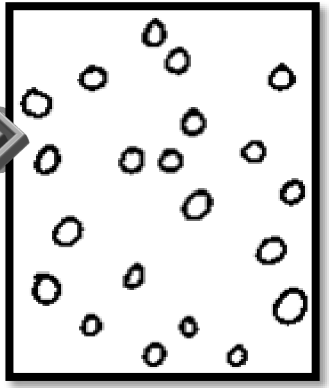
Key Points

- (a) Proposed CSC system significantly outperforms the maximum likelihood (MLE) baselines that uses linear autoregressive models → improved physical model accuracy using advanced AI algorithms, such as causal discovery
- (b) 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

From Data-driven to Reason-driven Wireless Networks

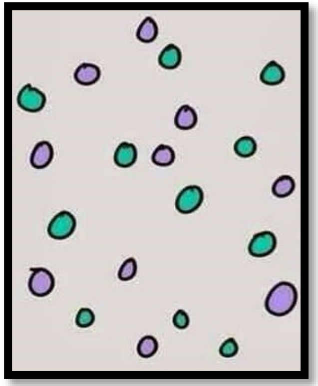
Data

Data-driven



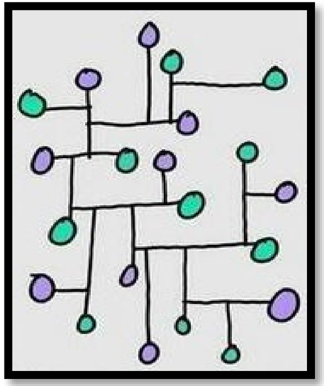
Discrete elements
(scalars, vectors, matrices, or databases that enable *visualization*):
E.g. QoS values, antenna alignment angles, fading coefficients, etc.
Wireless networks are "tied" to data.

Information-driven



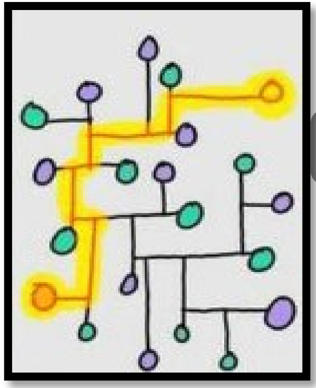
Linked elements
(concepts, metrics that facilitate the *design and optimization*):
E.g. Information-centric networking, Age of Information, AI-augmented wireless.
Wireless networks are application-aware and quasi-autonomous.

Knowledge-driven



Organized information
(AI frameworks that exhibit generalizability and invariance):
E.g. Meta-learning, transfer learning, and continual learning for wireless.
Wireless networks are AI-aware with a generalizable intelligence (first step to semantic communications).

Reasoning-driven



Applied knowledge
(higher reasoning capabilities that can perform logic, associations, and linkage):
E.g. Semantic language, neuro-symbolic or causal structural models of reasoning.
Wireless networks are semantic-based AI-native and can perform versatile logic.

Human-like reasoning

Conclusion and Future Recommendations

1. Semantic communications may significantly enhance network performance

It is not merely a form of minimalism as existing works allude; it can enhance **resilience, reliability**, and overall **capacity** of a network

2. Advances in AI and computing are necessary

More efforts needed on **generalizable, reasoning** and knowledge driven AI as well as judicious computing resouces

3. Semantic communications is not here to replace classical communications

Nor to solve all of its problems.
➔ Memorizable datastreams are more efficiently sent via classical channels.

4. Less spectrum reliance via the convergence of computing and communications

This could help alleviate technical and regulatory burdens associated with the need to open new spectrum bands for every wireless cellular generation.

Thank you
Q&A