2019 57th Annual Allerton Conference on Communication, Control, and Computing
(Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, September 24-27, 2019

Learning to Communicate with Limited Co-design

Anant Sahai Joshua Sanz Vignesh Subramanian Caryn Tran Kailas Vodrahalli

***Abstract*—In this work we examine the problem of learning to cooperate in the context of wireless communication. We consider the two agent setting where agents must learn modulation and demodulation schemes that enable them to communicate with each other in the presence of a power-constrained additive white Gaussian noise (AWGN) channel. We investigate whether learning is possible under different levels of information sharing between distributed agents that are not necessarily co-designed. We make use of the "Echo" protocol, a learning protocol where an agent hears, understands, and repeats (echoes) back the message received from another agent. Each agent uses what it sends and receives to train itself to communicate.**

To capture the idea of cooperation between agents that are "not necessarily co-designed," we use two different populations of function approximators — neural networks and polynomials. In addition to diverse learning agents, we include non-learning agents that use fixed standardized modulation protocols such as QPSK and 16QAM. This is used to verify that the Echo approach to learning to communicate works independent of the inner workings of the agents, and that learning agents can not only learn to match the communication expectations of others, but can also collaboratively invent a successful communication approach from independent random initializations. In addition to simulation-based experiments, we implement the Echo protocol in physical software-defined radio experiments to verify that it can work with real radios.

To explore the continuum between tight co-design of learning agents and independently designed agents, we study how learning is impacted by different levels of information sharing — including sharing training symbols, sharing intermediate loss information, and sharing full gradient information. The resulting learning techniques span supervised learning and reinforcement learning. We find that in general, co-design (increased information sharing) accelerates learning and that this effect becomes more pronounced as the communication task becomes harder.

I. INTRODUCTION

Machine learning is a technology and associated design paradigm that has recently seen a resurgence largely due to advances in computational capabilities. Consequently, there has been increasingly active research in the areas of supervised and reinforcement learning, both in the underlying technology as well as in the development of design paradigms appropriate to using these technologies in diverse application contexts. This paper is about seeing whether machine learning paradigms can be used to aid us with achieving interoperability in a wireless communication setting. The established paradigm for interoperation is that of standards — communication protocols are not only hand-crafted by individual humans, these hand-crafted protocols are standardized and certified by authorized committees of people. Can we use machine learning techniques to learn how to communicate with minimal assumptions on shared information, and if so how well can we learn?

Communication is a fundamentally cooperative activity between at least two agents. Consequently, communication itself can be viewed as both a special case of cooperation as well as a building block that can be leveraged to permit more effective cooperation. The fundamental limits to learning how to cooperate with a stranger have been studied in an abstract theoretical setting in [1], [2], [3], [4], [5]. By asking how two intelligent agents might understand and help each other without a common language, a basic theory of goal-oriented communication was developed in these papers. The principal claim is that for two agents to robustly succeed in the task of learning to collaborate, the goal must be explicit, verifiable, and forgiving. However the approach in these works took a fundamentally *semantic* perspective on cooperation. As Shannon pointed out in [6], the arguably simpler cooperative problem of communicating messages can be understood in a way that is divorced from the issue of what the messages mean. To see whether existing machine learning paradigms can be adapted to achieve cooperation with strangers, we consider the concrete problem where two agents learn to communicate in the presence of a noisy channel. Each agent consists of a modulator and demodulator.

This problem of learned communication has been tackled using learning techniques under different assumptions on the information that the agents are allowed to share and how tightly coordinated their interaction is. Early work in this area [7], [8], where gradients are shared among agents, demonstrated the success of training a channel auto-encoder using supervised learning when a stochastic model of the channel is known. Subsequent works relax the assumption on the known channel model by learning a stochastic channel model by using GANs as in [9], [10], [11] or by approximating gradients across the channel and using that for training. However, these approaches cannot be said to represent communication with strangers, and instead represent a way of having co-designed systems learn to communicate. If instead of sharing gradients we can only share scalar loss values then with access to a shared preamble, reinforcement learning can be used to train the system as demonstrated in [12] and [13] without having access to a stochastic channel model.

Moving closer to minimal co-design, if we further restrict ourselves to the case where the two agents only have access to a shared preamble, the "Echo" protocol, where an agent hears, understands, and repeats (echoes) back the message received from the other agent, as specified in [12] has been shown to work. By comparing the original message to the received echo, a learning agent can get feedback about how well the two agents understand each other[1] The work in [12] considered a neural network based modulator that was trained using reinforcement learning, but the demodulator was nearest neighbors based and required no training — it used small-sample-based supervised learning. Our work in the present paper builds on this and studies the case where agents do not have access to a shared preamble, and also introduces the concept of "alienness" among agents. After all, if our goal is to understand the learning of communication between strangers, we need to be able to test this. We consider modulators and demodulators represented using different types of function approximators such as neural networks and polynomials.By doing so we wish to separate the effect of the specific function approximators used from the meta protocols (specifically the Echo protocol) used to do the learning.

Our main contribution is to investigate whether the Echo protocol is universal, i.e. does it allow two agents to learn to communicate irrespective of their type, and what level of information sharing is necessary for it to work. We consider agents with different levels of "alienness" based on the hyperparameters, architectures, and techniques used in their modulators and demodulators. To explore the gradient between complete codesign and strangers, we explore different levels of information sharing, namely shared gradients, shared loss information, shared preambles (echo-shared-preamble), and finally the case where only the overall protocol is shared (echo-private-preamble). Machine learning scholarship is notorious for producing results that are not easily reproducible, and failure to identify the source of and explain the reasoning behind performance gains [14]. Keeping this in mind, in

[1] Round-trip stability is not by itself a sufficient condition to guarantee mutual comprehension. After all, one agent might just be doing raw mimicry and just repeating back the analog signal value received with no attempt to actually demodulate. However, in [12], the key insight was that intelligent agents, though strangers, are believed to be cooperative and so wish to actually understand and communicate with each other. They don't need to actually coordinate with another designer to realize that sheer mimicry would not necessarily advance their goal of cooperation. Consequently, the Echo protocol can rely on the good intention to eliminate the possibility of them just mirroring back what has been heard instead of trying to understand what was sent and repeating it back.

order to evaluate the ease, speed, and robustness of the learning task under various levels of alienness and information sharing, we conduct repeated trials for each setting using different seeds and slightly different hyperparameters sampled uniformly from a range. We report the fraction of trials that succeeded as a function of the amount of symbols exchanged as well as aggregate statistics about the bit error rate achieved at different signal to noise ratio (SNR) levels by the learned modulation/demodulation schemes under various settings. From our experiments we observe and conclude that the Echo protocol allows two agents to learn a modulation scheme even when we share the minimal amount of information, and that as we decrease the amount of shared information the learning task is harder, i.e a lower fraction of seeds succeed and the agents take longer to learn. It appears that learning to communicate with "alien" agents can be more or less difficult than learning with agents of the same type depending on the specifics of the agents. However it is significantly easier to learn to communicate if one of the agents has been pretrainedwith a good modulation and demodulation scheme (as could be achieved by self training for example). Finally, as we increase the modulation order for communication the learning task becomes harder, especially for settings with little information sharing.

Although most of the results we report in this paper were performed purely in simulations, we replicate our simulation results using USRP radios and observe similar results — two agents can learn to communicate in a decentralized way even using real hardware. More details of related work as well as everything else are in the full version of this paper.

II. Overview

A. Problem Formulation

We consider the setting where two agents communicate in the presence of an additive white Gaussian noise (AWGN) channel. Each agent consists of an encoder (modulator) and a decoder (demodulator). We treat the modulator as an abstract (black box) object that converts bit symbols to complex numbers, i.e. we treat it as a mapping $M : B \to \mathbb{C}$ where B refers to the set of bit symbols and $\mathbb{C}$ refers to the set of complex numbers. Similarly we treat the demodulator as an abstract object that converts complex numbers to bit symbols, i.e. a mapping $D : \mathbb{C} \to B$. The set of bit symbols B, is specified by the modulation order (bits per symbol). For instance, when bits per symbol is 1, $B = \{0, 1\}$ and when bits per symbol is 2, $B = \{00, 01, 10, 11\}$. For the case where bits per symbol is 1, the classic[2] BPSK (binary phase shift keying) modulation and demodulation scheme is given by:

$$M^{BPSK}(0) = 1 + 0j,$$
$$M^{BPSK}(1) = -1 + 0j.$$

These corresponding demodulator performs the demodulation as,

$$D^{BPSK}(c) = \begin{cases} 0, & Re(c) \geq 0 \\ 1, & Re(c) < 0. \end{cases}$$

In addition to agents that use fixed modulation and demodulation schemes we also consider 'learning' agents. These agents use function approximators to learn the mapping performed by modulator and demodulator, and we denote these as $M(\cdot; \theta)$ and $D(\cdot; \phi)$ where θ and ϕ denote the parameters of the underlying function approximators and are updated during training. The specifics of the learning agents and their update methods can be found in the full version of this paper.

The main focus of our work is in learning modulation and demodulation schemes, and in order to make it easier to conduct experimental simulations we make the following simplifying assumptions:

1) There are at most two agents, and they engage in perfect turn-taking
2) The two agents are separated by a unit gain AWGN channel. There is no carrier frequency offset, timing offset or phase offset.
3) Both agents encode and decode data using the same, fixed number of bits per symbol (i.e., the modulation order is preset).

[2]Here we use classic to refer to a modulation scheme that is fixed and specified identically for all communicating agents by a certain standard. One example of such a scheme is BPSK signalling as described above.

4) The environment is stationary and non-adversarial during the learning process.

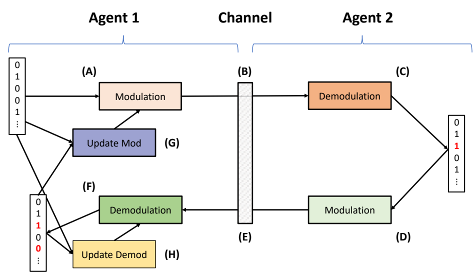

Fig. 1: Visualization of the Echo protocol. (A) Speaker Agent (A1) modulates a bit sequence and (B) sends it across a (AWGN) channel. (C) Echoer Agent (A2) receives the sequence and demodulates it; (D) A2 then modulates the recovered sequence and (E) sends it back over the channel. (F) A1 receives this echoed version of its original sequence and demodulates it. (G, H) Then A1 uses the received echo to update its modulator and demodulator. The agents switch roles and repeat until convergence.

B. Objective and Approach – Echo with Private Preamble Protocol

The main objective of our work is to specify a robust communication-learning protocol that allows two independent agents to learn a modulation scheme under minimal assumptions on information sharing beyond shared knowledge of learning protocol and ability to perform turn taking. No other information is shared a-priori or via a side channel during training. We coin the name *Echo with Private Preamble* (EPP) protocol.

The underlying premise of the Echo protocol is that an echo of the message sent by one agent repeated by another agent provides sufficient feedback for the agent to learn expressive modulation and demodulation schemes. Under the Echo protocol, one agent (the "speaker") broadcasts a message and receives back an estimate of this message(preamble), an Echo, from the other agent (the"echoer"). The passage of the original message from the speaker to the echoer and back to the speaker as an Echo is denoted as a *round-trip*. (A *half-trip* goes only from speaker to echoer.) After a *round-trip*, the speaker compares the original message to the echo and trains its modulator and demodulator to minimize the difference (usually measured in the form of bit-errors) between the two messages. The two agents then switch roles and repeat. When the difference between the original message and the echo is small, we infer that the agents can communicate with one another.

A variant of the Echo protocol, *Echo with Shared Preamble* (ESP) was introduced in [12] where both agents have access to a shared preamble (message that the "speaker" sends). Here the echo behavior is introduced to only to train the modulator, and knowledge of a shared preamble between the two agents is assumed to facilitate direct supervised training of the demodulator after a half-trip exchange. We consider the case where agents do not have access to a shared preamble and use the EPP protocol to learn a modulation scheme in this setting.

We believe that the EPP protocol minimizes the information sharing assumptions for learning modulation schemes for two reasons. First, some sort of feedback is required for learning, and the echo of the preamble provides this feedback. Second, the EPP protocol treats the environment as a regenerative channel, i.e. a channel that provides feedback without requiring assumptions about the nature of the other communicating agent. As long as the other agent cooperative in the sense of echoing back what it hears, then the environment behaves like a regenerative channel.

Next we argue that the EPP protocol is a plausible mechanism for learning modulation schemes when the channel is regenerative by considering the case of a learning agent communicating with an agent that uses fixed classic schemes. In this setting, even random exploration would eventually find a modulation scheme that successfully interfaces with the fixed agent. By using feedback to guide exploration, we expect the EPP protocol to perform much better than random guessing and quickly converge to a suitable modulation scheme. We can think of such a fixed, friendly regenerative channel as a

’game’ that the learner plays where positive reward is achieved if what the channel echoes back can be decoded as what the learner encoded and sent in. Reinforcement learning is good at optimizing behaviors for simple games like this [44]. One of our main contributions is to show that the EPP protocol works not only with fixed communication partners, but even in the case where two agents are learning simultaneously.

To verify the universality of the EPP protocol and understand its performance, we run experiments with:

1) Different learning protocols based on varying amounts of information sharing.
2) Different levels of *alienness*[3] amongst agents.
3) Different modulation order and levels of training SNR.

III. Results

Here, we include certain key figures and a key table. More details are in the full version of this paper.

	QPSK (2 BPS)	8PSK (3 BPS)	16QAM (4 BPS)
Echo, Shared Preamble	51200	76800	204800
Echo, Private Preamble	115200	665600	3379200

TABLE I: Number of symbols exchanged until $\geq 90\%$ of trials reached 3 dB off of optimal BER for the ESP and EPP protocols with Neural agents and varying bits per symbol (BPS). The agents were trained and tested at SNRs corresponding to 1% BER for the corresponding baseline modulation schemes. The results show the increased difficulty of learning at higher modulation orders. The EPP protocol is impacted much more than ESP by high modulation order.

[3]*Alienness* is a description of how different the models for the agents’ modulators and demodulators are between the two agents. Factors that determine alienness include whether the agents are fixed or learning, the class of function approximators used by the learning agents and choice of hyperparameters and initializations.

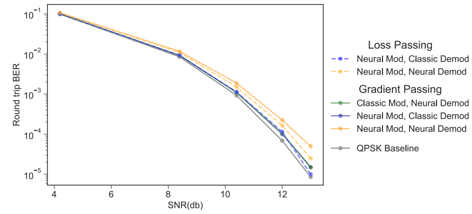

Fig. 2: Round-trip median BER curves for GP and LP protocols at 2 bits per symbol and 8.4 dB training SNR. Permutations of neural and classic modulator and demodulator models are tested in order to show individual learning components independently. (a) is the full curve at all test SNRs and (b) is zoomed in on the upper end of the testing SNRs. All of the loss and gradient passing settings show close to baseline performance.

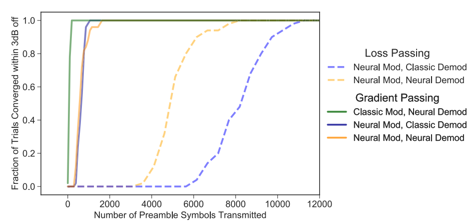

Fig. 3: Convergence of 50 GP and LP trials at 2 bits per symbol to be within 3 dB (at testing SNR 8.4 dB) at 8.4 dB training SNR. LP requires more symbols than GP to converge.

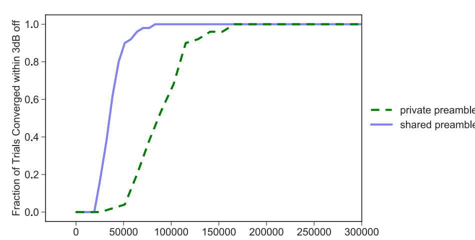

Fig. 4: Convergence of 50 Neural-vs-Clone trials to be within 3 dB (at testing SNR 8.4 dB) using the EPP and ESP protocols at 2 bits per symbol and training SNR 8.4 dB. EPP takes more symbols to converge.

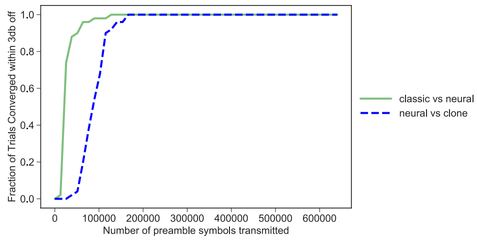

Fig. 5: Private Preamble: Neural Agent vs Clone

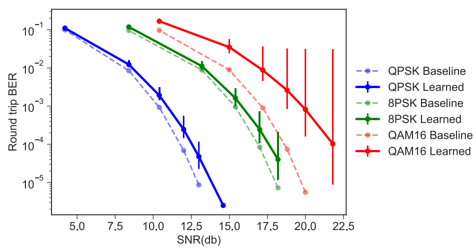

Fig. 6: Round-trip median BER curves for Neural-vs-Clone learning QPSK, 8PSK, and 16QAM under the EPP protocol at training SNRs corresponding to 1% BER. Alongside the BER curves of the learned modulation and demodulation schemes is the baseline. In all cases, modulation constellations are normalized to constrain the average signal power. The learned 16QAM agents perform worse relative to the baseline than QPSK and 8PSK agents.

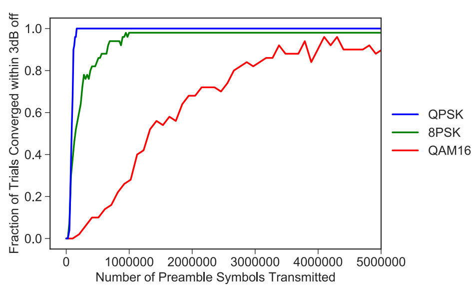

Fig. 7: Convergence of 50 trials to be within 3 dB (at testing SNR corresponding to 1% BER) of the corresponding baseline for EPP trials of Neural-vs-Clone at training SNR corresponding to 1% BER for increasing modulation order. 16QAM, with the highest modulation order 4, takes much longer to converge than QPSK (order 2) and 8PSK (order 3).

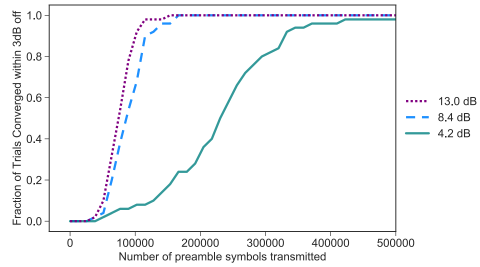

Fig. 8: Convergence of 50 Neural-vs-Clone trials to be within 3 dB (at testing SNR 8.4 dB) at 2 bits per symbol using EPP protocol at training SNRs 13.0, 8.4, and 4.2 dB. Training at higher SNR reduces the number of symbols required for most seeds to converge.

IV. Implementation in Software Defined Radios

In order to corroborate our simulation results, we implement the ESP and EPP protocols on Ettus USRP software defined radios using GNU Radio [45]. The goal of this implementation is not to provide a real-time implementation of the Echo protocol, since in general the real-time components of radio communications are implemented in ASICs, and even software components run in special real-time operating systems to achieve deterministic or bounded latencies. The focus of our work is to learn modulation schemes, so the primary goal of the GNU Radio implementation is to demonstrate that the learning protocols work not only in simulations but also when trained in real, physical systems. Other work such as [20] and [21] have also demonstrated that end-to-end learning of communication schemes is possible over the air in real radio systems. The details are in the full version of this paper.

A. Experiments

The radio experiments were conducted using two Ettus USRP X310 software defined radios (SDRs) connected to each other with SMA cables.75 dB of attenuation was added between the radios both to simulate path loss and to allow us to achieve desired SNRs with the available internal transmit and receive gains.

Figs. 9 and 10 compare the performance of the GNU Radio implementation to our simulations for EPP neural-clone training. Apart from the additional SNR required to achieve the same baseline performance, the trained neural agents show a similar spread in final BER performance across SNRs. This is another main result of our work, **EPP is successful at learning modulation schemes over the wire while using software defined radios.**

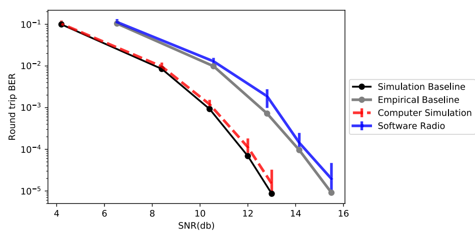

Fig. 9: Round-trip median bit error curves for Neural agent vs clone python simulation and gnuradio agent learning QPSK under the EPP protocol at training SNRs corresponding to 1% BER. Alongside the bit error curves of the learned modulation and demodulation schemes is the baseline

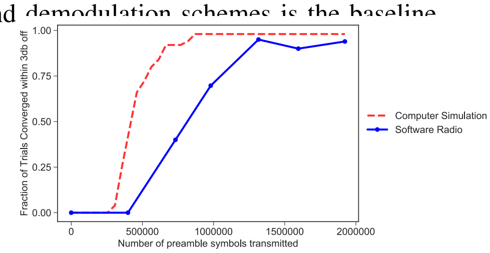

Fig. 10: Convergence of 50 simulation and 20 gnuradio trials to be within 3 dB (at testing SNR corresponding to 1% BER) of the corresponding baseline for EPP trials of Neural agent vs clone at training SNR corresponding to 1% BER for QPSK modulation. The GNU Radio agents were only trained at 1% BER SNR, equivalent to SNR_dB=8 among the simulation curves. Note that the GNU Radio agents take longer to converge to within 3 dB of optimal, but after sufficient time a similar proportion of seeds converge.

References

[1] B. Juba and M. Sudan, "Universal semantic communication i," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*. ACM, 2008, pp. 123–132.

[2] ——, "Universal semantic communication ii: A theory of goal-oriented communication," in *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 15, no. 095, 2008.

[3] I. Komargodski, P. Kothari, and M. Sudan, "Communication with contextual uncertainty," *CoRR*, vol. abs/1504.04813, 2015. [Online]. Available: http://arxiv.org/abs/1504.04813

[4] C. L. Canonne, V. Guruswami, R. Meka, and M. Sudan, "Communication with imperfectly shared randomness," *CoRR*, vol. abs/1411.3603, 2014. [Online]. Available: http://arxiv.org/abs/1411.3603

[5] O. Goldreich, B. Juba, and M. Sudan, "A theory of goal-oriented communication," *J. ACM*, vol. 59, no. 2, pp. 8:1–8:65, 2012. [Online]. Available: https://doi.org/10.1145/2160158.2160161

[6] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[7] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec 2017.

[8] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," *CoRR*, vol. abs/1608.06409, 2016. [Online]. Available: http://arxiv.org/abs/1608.06409

[9] H. Ye, L. Liang, G. Y. Li, and B. F. Juang, "Deep learning based end-to-end wireless communication systems with conditional GAN as unknown channel," *CoRR*, vol. abs/1903.02551, 2019. [Online]. Available: http://arxiv.org/abs/1903.02551

[10] H. Ye, G. Y. Li, B. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," *CoRR*, vol. abs/1807.00447, 2018. [Online]. Available: http://arxiv.org/abs/1807.00447

[11] T. J. O'Shea, T. Roy, and N. West, "Approximating the void: Learning stochastic channel models from observation with variational generative adversarial networks," *CoRR*, vol. abs/1805.06350, 2018. [Online]. Available: http://arxiv.org/abs/1805.06350

[12] C. de Vrieze, S. Barratt, D. Tsai, and A. Sahai, "Cooperative multi-agent reinforcement learning for low-level wireless communication," *arXiv preprint arXiv:1801.04541*, 2018.

[13] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," *CoRR*, vol. abs/1804.02276, 2018. [Online]. Available: http://arxiv.org/abs/1804.02276

[14] Z. C. Lipton and J. Steinhardt, "Troubling trends in machine learning scholarship," *Queue*, vol. 17, no. 1, pp. 80:45–80:77, Feb. 2019. [Online]. Available: http://doi.acm.org/10.1145/3317287.3328534

[15] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication Algorithms via Deep Learning," *arXiv e-prints*, p. arXiv:1805.09317, May 2018.

[16] T. J. O'Shea and J. Hoydis, "An introduction to machine learning communications systems," *CoRR*, vol. abs/1702.00832, 2017. [Online]. Available: http://arxiv.org/abs/1702.00832

[17] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[18] V. Raj and S. Kalyani, "Backpropagating through the air: Deep learning at physical layer without channel models," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2278–2281, Nov 2018.

[19] F. Ait Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2019.

[20] J. Schmitz, C. von Lengerke, N. Airee, A. Behboodi, and R. Mathar, "A Deep Learning Wireless Transceiver with Fully Learned Modulation and Synchronization," *arXiv e-prints*, p. arXiv:1905.10468, May 2019.

[21] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, Feb 2018.

[22] T. J. O'Shea, L. Pemula, D. Batra, and T. C. Clancy, "Radio Transformer Networks: Attention Models for Learning to Synchronize in Wireless Systems," *arXiv e-prints*, p. arXiv:1605.00716, May 2016.

[23] A. Caciularu and D. Burshtein, "Blind Channel Equalization using Variational Autoencoders," *arXiv e-prints*, p. arXiv:1803.01526, Mar 2018.

[24] J. W. Laura Brink, Anant Sahai, "Deep Networks for Equalization in Communications," Master's thesis, University of California, Berkeley, 2018. [Online]. Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-177.pdf

[25] Y. Jiang, H. Kim, H. Asnani, S. Kannan, S. Oh, and P. Viswanath, "LEARN Codes: Inventing Low-latency Codes via Recurrent Neural Networks," *arXiv e-prints*, p. arXiv:1811.12707, Nov 2018.

[26] Y. Jiang, H. Kim, H. Asnani, and S. Kannan, "MIND: Model Independent Neural Decoder," *arXiv e-prints*, p. arXiv:1903.02268, Mar 2019.

[27] V. Corlay, J. J. Boutros, P. Ciblat, and L. Brunel, "Neural lattice decoders," *CoRR*, vol. abs/1807.00592, 2018. [Online]. Available: http://arxiv.org/abs/1807.00592

[28] K. Choi, K. Tatwawadi, T. Weissman, and S. Ermon, "NECST: neural joint source-channel coding," *CoRR*, vol. abs/1811.07557, 2018. [Online]. Available: http://arxiv.org/abs/1811.07557

[29] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *CoRR*, vol. abs/1809.01733, 2018. [Online]. Available: http://arxiv.org/abs/1809.01733

[30] B. Zhu, J. Wang, L. He, and J. Song, "Joint Transceiver Optimization for Wireless Communication PHY with Convolutional Neural Network," *arXiv e-prints*, p. arXiv:1808.03242, Aug 2018.

[31] D. J. Ji, J. Park, and D. Cho, "Convae: A new channel autoencoder based on convolutional layers and residual connections," *IEEE Communications Letters*, pp. 1–1, 2019.

[32] N. Wu, X. Wang, B. Lin, and K. Zhang, "A cnn-based end-to-end learning framework toward intelligent communication systems," *IEEE Access*, vol. 7, pp. 110 197–110 204, 2019.

[33] T. Mu, X. Chen, L. Chen, H. Yin, and W. Wang, "An End-to-End Block Autoencoder For Physical Layer Based On Neural Networks," *arXiv e-prints*, p. arXiv:1906.06563, Jun 2019.

[34] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. ten Brink, "Ofdm-autoencoder for end-to-end learning of communications systems," *CoRR*, vol. abs/1803.05815, 2018. [Online]. Available: http://arxiv.org/abs/1803.05815

[35] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," *arXiv preprint arXiv:1804.07675*, 2018.

[36] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *arXiv preprint arXiv:1804.04097*, 2018.

[37] C. Lee, H. B. Yilmaz, C.-B. Chae, N. Farsad, and A. Goldsmith, "Machine learning based channel modeling for molecular mimo communications," in *Signal Processing Advances in Wireless Communications (SPAWC), 2017 IEEE 18th International Workshop on*. IEEE, 2017, pp. 1–5.

[38] S. Mohamed, J. Dong, A. R. Junejo, and D. C. Zuo, "Model-based: End-to-end molecular communication system through deep reinforcement learning auto encoder," *IEEE Access*, vol. 7, pp. 70 279–70 286, 2019.

[39] R. Raileanu, E. Denton, A. Szlam, and R. Fergus, "Modeling others using oneself in multi-agent reinforcement learning," *CoRR*, vol. abs/1802.09640, 2018. [Online]. Available: http://arxiv.org/abs/1802.09640

[40] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Perolat, D. Silver, and T. Graepel, "A unified game-theoretic approach to multiagent reinforcement learning," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4190–4203.

[41] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Basar, "Fully decentralized multi-agent reinforcement learning with networked agents," *CoRR*, vol. abs/1802.08757, 2018. [Online]. Available: http://arxiv.org/abs/1802.08757

[42] J. Foerster, I. A. Assael, N. de Freitas, and S. Whiteson, "Learning to communicate with deep multi-agent reinforcement learning," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 2137–2145. [Online].

[43] R. Siegler, J. DeLoache, and N. Eisenberg, *How Children Develop*. Worth Publishers, 2003.

[44] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," *arXiv e-prints*, p. arXiv:1312.5602, Dec 2013.

[45] GNU Radio Website, accessed July 2019. [Online]. Available: http://www.gnuradio.org