IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 19, NO. 5, MAY 2020

Deep Learning-Based End-to-End Wireless Communication Systems With Conditional GANs as Unknown Channels

Hao Ye, *Student Member, IEEE*, Le Liang, *Member, IEEE*, Geoffrey Ye Li, *Fellow, IEEE*, and Biing-Hwang Juang, *Life Fellow, IEEE*

Hao Ye, Geoffrey Ye Li, and Biing-Hwang Juang are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: yehao@gatech.edu; liye@ece.gatech.edu; juang@ece.gatech.edu).

Le Liang was with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA. He is now with Intel Labs, Hillsboro, OR 97124 USA (e-mail: lliang@gatech.edu).

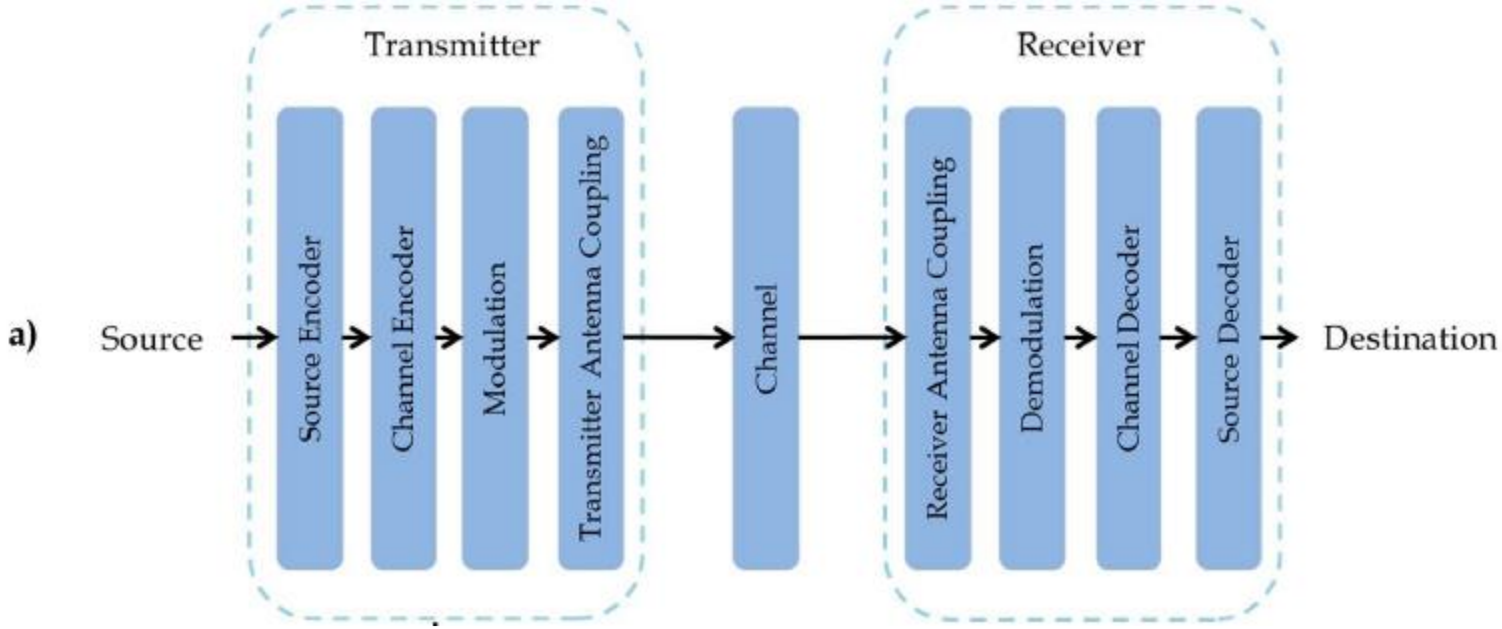
2021. 12. 11.

Hyunsoo, Yu

INDEX

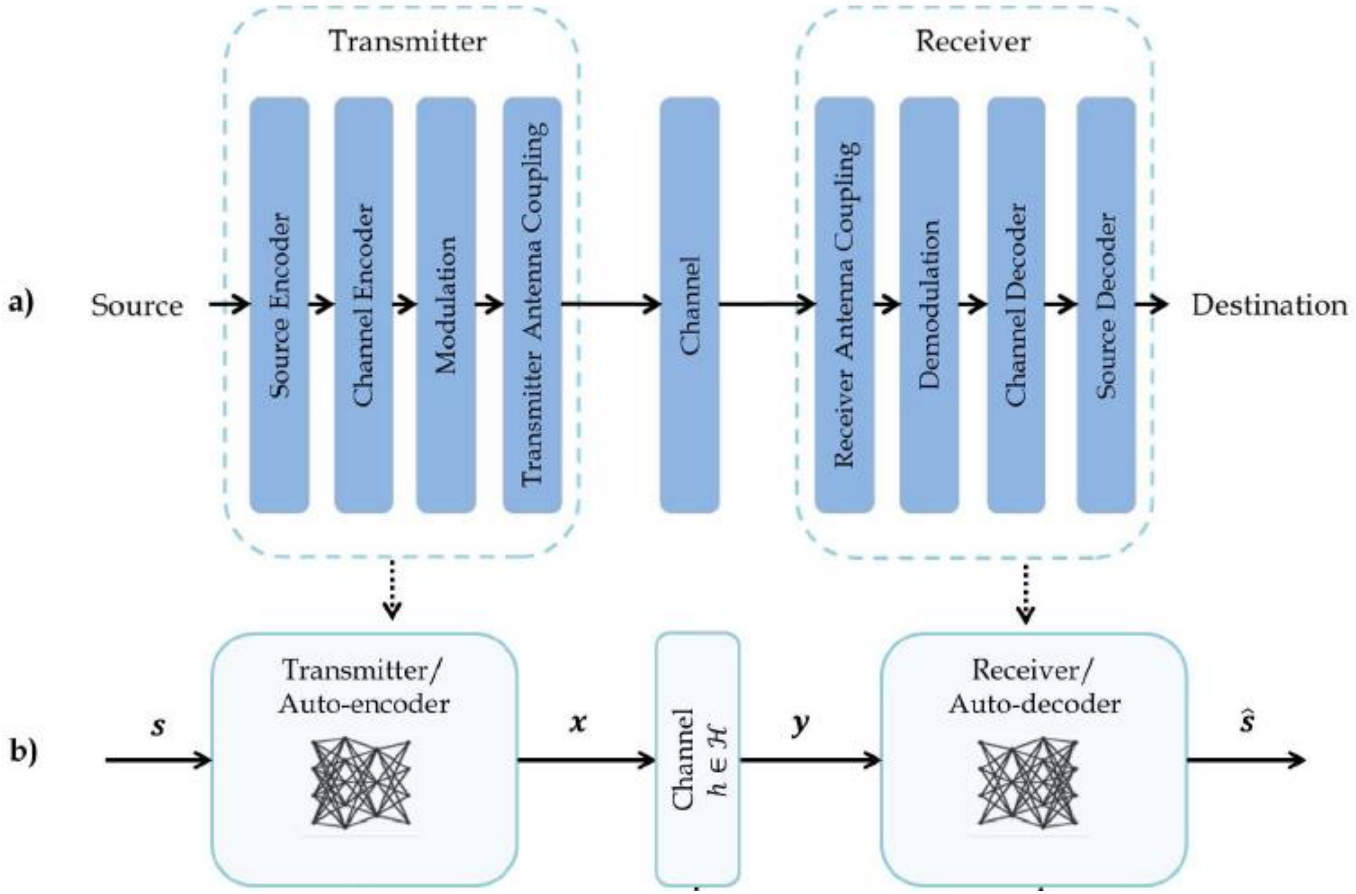
1. INTRODUCTION

- Conventional Wireless System
  - Transfer data with **multiple signal processing blocks**
  - Each block is developed and optimized **individually**
    - It is hard to measure total performance
  - Mathematical assumption on the channel propagation is not accurate => damage the system performance



the weight/parameter set of the transmit DNN $\boldsymbol{\theta_T}$
the weight/parameter set of the receiver DNN $\boldsymbol{\theta_R}$
the channel realization $\mathbf{h}$

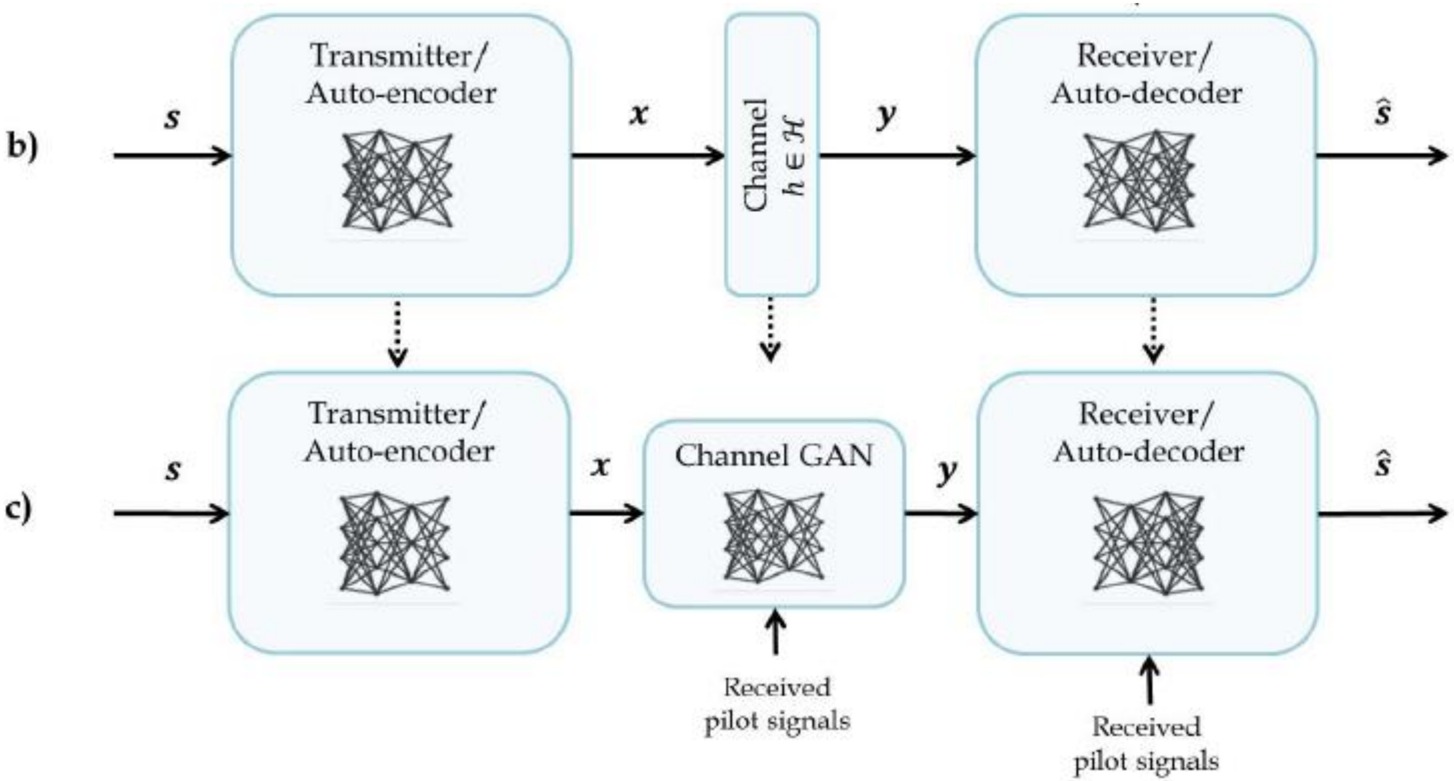
1. INTRODUCTION



- Deep learning method on communication
  - Feature & parameter can be learned from the data directly, (Data-driven methods)
  - Tx & Rx can be interpreted as an auto-encoder and auto-decoder
- Offline training
  - Loss $\mathbb{E}_{(\mathbf{h}\in\mathcal{H})}\{\mathcal{L}(\boldsymbol{\theta_T}, \boldsymbol{\theta_R}, \mathbf{h})\}$
    - h is unknown, only know it is from a sample space H
- Problem
  - We don't know $\mathbf{y} = f_{\mathbf{h}}(\mathbf{x})$
  - The curse of dimensionality; long codeword is not available

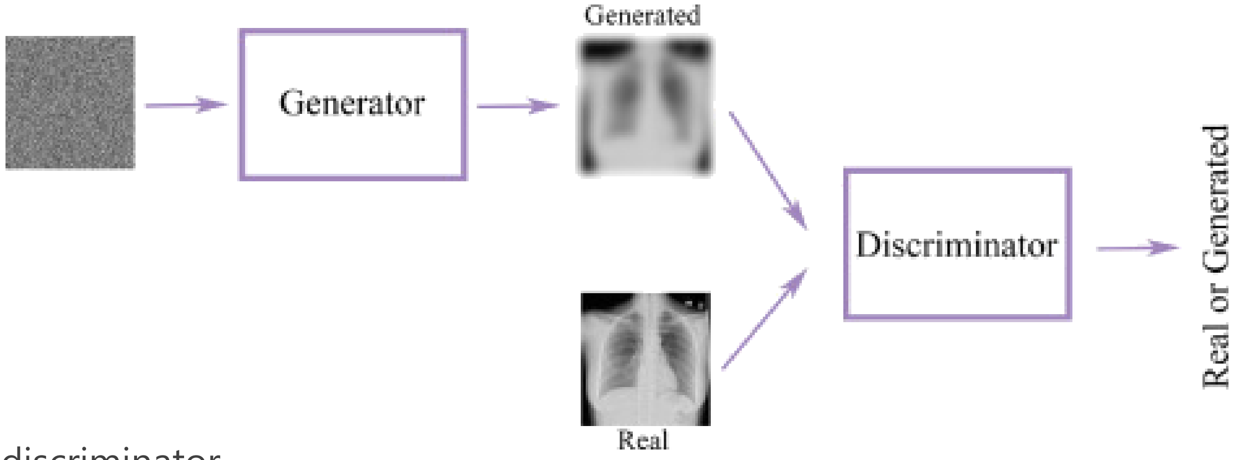
1. INTRODUCTION

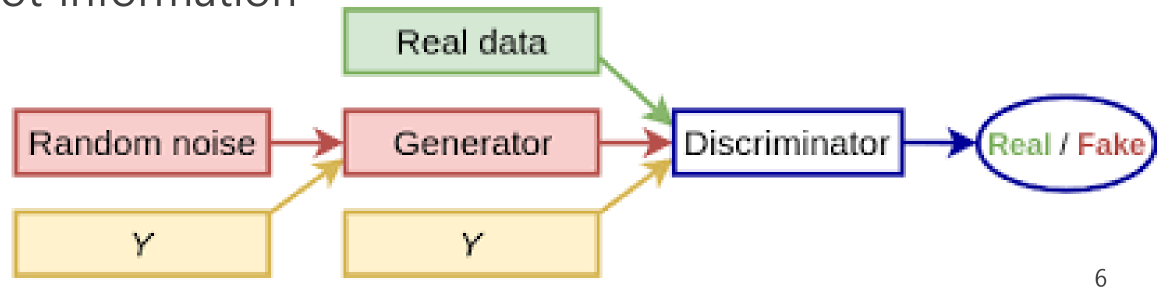
- Wireless communication system with Conditional GAN is proposed to challenge those problems.
  - Learn **the distributions of channel output**
  - Condition : encoded signals from Tx and the received pilot information used for estimating the channel
  - Optimize the Tx loss, Rx loss, and the end-to-end loss in a supervised way
  - CNN alleviates curse of dimensionality



2. RELATED WORKS
2.1 GAN AND CONDITIONAL GAN

- GAN
  - Generator
    - Generate fake sample
  - Discriminator
    - Discriminate whether fake or real
- Conditional GAN
  - Context information(Label) is added to the generator and discriminator
- When generating the channel output, the received pilot information is used as a conditional information





2. RELATED WORKS
2.2 DNN BASED END-TO-END COMMUNICATIONS

- [1] propose the idea of end-to-end learning communication system
- In [2], CNNs are employed for modulation and demodulation
- In [3], Reinforcement learning is used to optimize the end-to-end communication system

[1] T. O’Shea and J. Hoydis, "An introduction to deep learning for the physical layer", *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563-575, Dec. 2017.
[2] B. Zhu, J. Wang, L. He and J. Song, "Joint transceiver optimization for wireless communication PHY with convolutional neural network", *arXiv:1808.03242*, 2018,
[3] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model", *arXiv:1804.02276*, 2018,

2. RELATED WORKS
2.3 LEARNING BASED DECODERS

- DNN has been utilized in decoding with a wide range of application
- RNN is used in [4] for decoding the convolutional and turbo codes(error correction code).
- The traditional belief-propagation decoding algorithm is extended as deep learning layers to decode linear codes in [5]

[4] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh and P. Viswanath, "Communication algorithms via deep learning", *arXiv:1805.09317*, 2018,
[5] E. Nachmani, Y. Be'ery and D. Burshtein, "Learning to decode linear codes using deep learning", *Proc. 54th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, pp. 341-346, Sep. 2016.

3. MODELING CHANNELS WITH CONDITIONAL GAN

- To make the back propagation algorithm available in the DNN end-to-end communication system, Conditional GAN is employed.
- The output distribution of the channel can be learned in a data-driven manner.

3. MODELING CHANNELS WITH CONDITIONAL GAN
3.1 CONDITIONAL GAN

GAN

$$\mathcal{L}_G = \min_{\boldsymbol{\theta}_G} \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z})))],$$

$$\mathcal{L}_D = \max_{\boldsymbol{\theta}_D} \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z})))],$$

Conditional GAN

$$\mathcal{L}_G = \min_{\boldsymbol{\theta}_G} \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z},\mathbf{m}),\mathbf{m}))],$$

$$\mathcal{L}_D = \max_{\boldsymbol{\theta}_D} \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log(D(\mathbf{x},\mathbf{m}))] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z},\mathbf{m}),\mathbf{m}))],$$

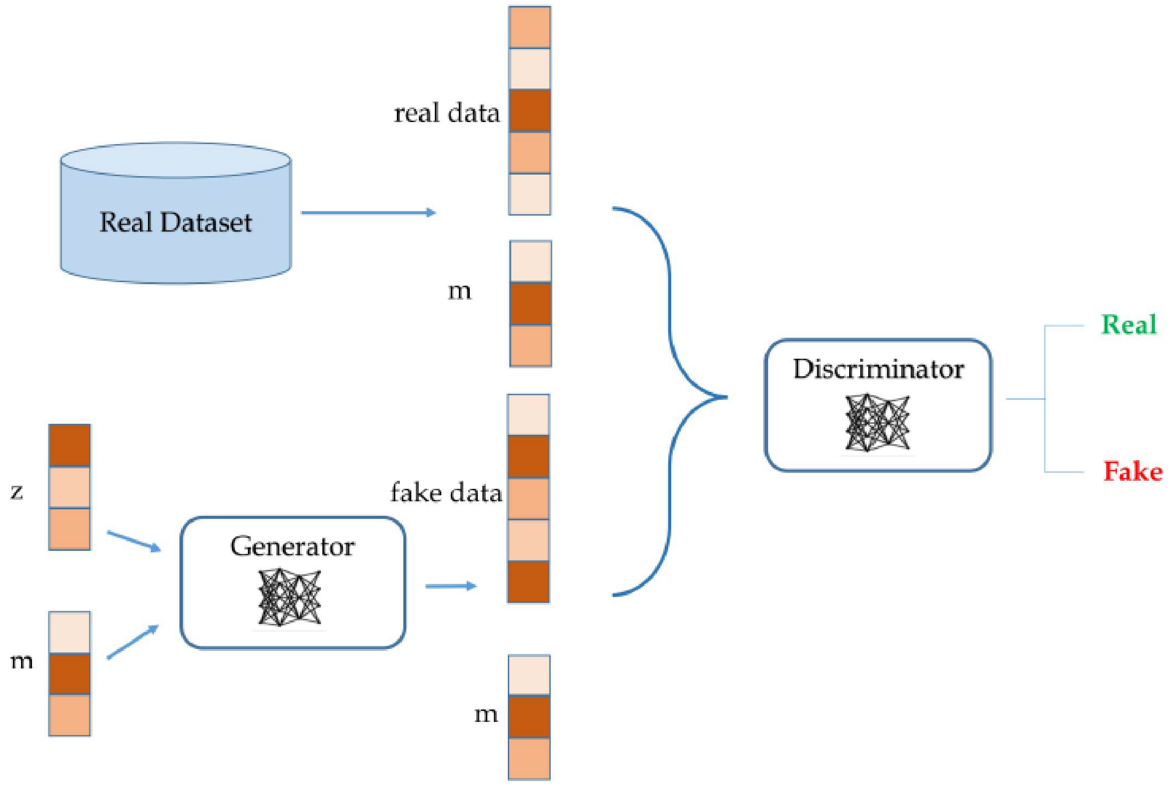
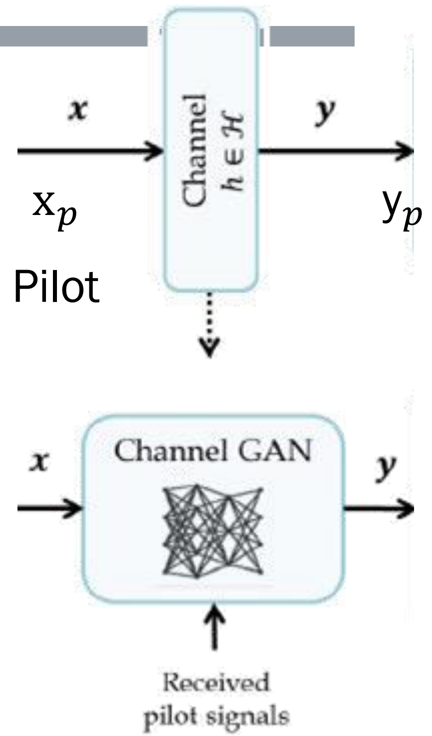


Fig. 2. Structure of conditional GAN.

m : conditioning information

3. MODELING CHANNELS WITH CONDITIONAL GAN
3.2 MODELING CHANNELS



- To obtain the Channel State Information (CSI), h,
  - some pilot information is sent to Rx
    - so that the channel information is inferred based on the received signal corresponding to the pilot symbols y_p.
- The received signal is added as a part of conditioning information
  - The output samples follow the distribution of y given the input x and the received pilot data y_p.
- With y_p as the conditional information, the output distribution of CGAN can stick to the current channel

3. MODELING CHANNELS WITH CONDITIONAL GAN
3.3 CONVOLUTIONAL LAYERS-BASED CHANNEL GAN

- Fully connected layer
  - Each neuron connect all neurons in previous layer
- Convolutional layer
  - Each neuron connect **a few nearby neurons** in previous layer
  - Same set of weights is shared by all neuron in a layer

- ※ Convolutional code
  - Encoding process is like convolutional transform

- In the channel GAN, 1D convolutional layer is used.
  - => Reduce complexity

Fully connected layer

$$u^{(i)}[n] = \sigma(\sum_k w_{nk}^{(i)} u^{i-1}[k]),$$

$u^{(i)}[n]$

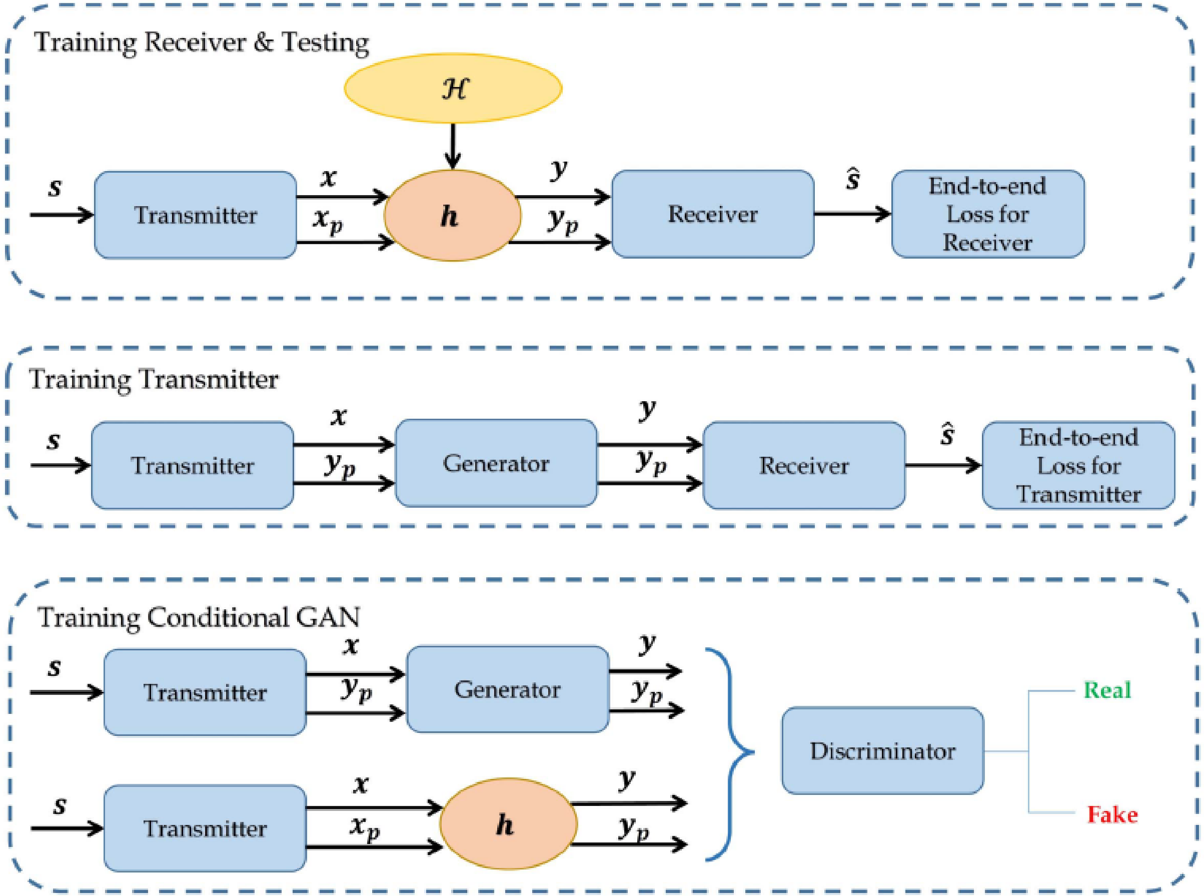
$\sigma(\cdot)$ Activation function

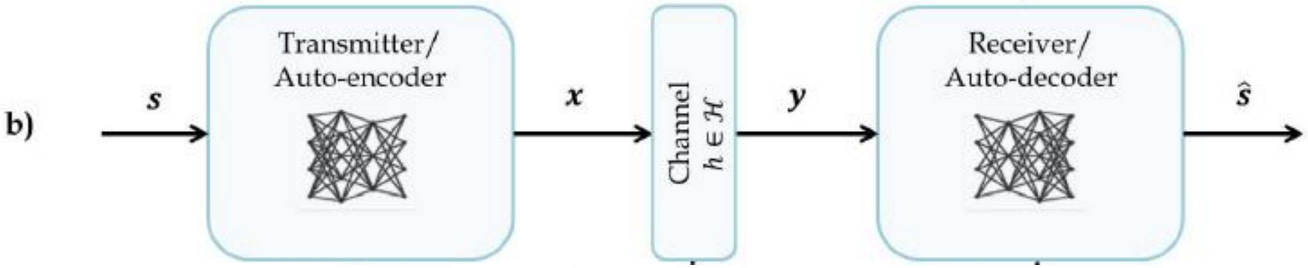
Convolutional layer

$$u^{(i)}[n] = \sigma(\sum_{k=1}^{L} w_k^{(i)} u^{(i-1)}[n-k])$$

The number of weights : L
: fewer connection than DNN

4. END-TO-END COMMUNICATION
4.1 SYSTEM OVERVIEW





$\mathbf{s} \in \{0,1\}^N$ $\quad$ $\mathbf{x} \in \mathbb{R}^K$ $\quad$ $\hat{\mathbf{s}} \in [0,1]^N$

$$L = \sum_{n=1}^{N} (s_n \log(\hat{s}_n) + (1 - s_n) \log(1 - \hat{s}_n))$$

s_n : n_th *element of* s

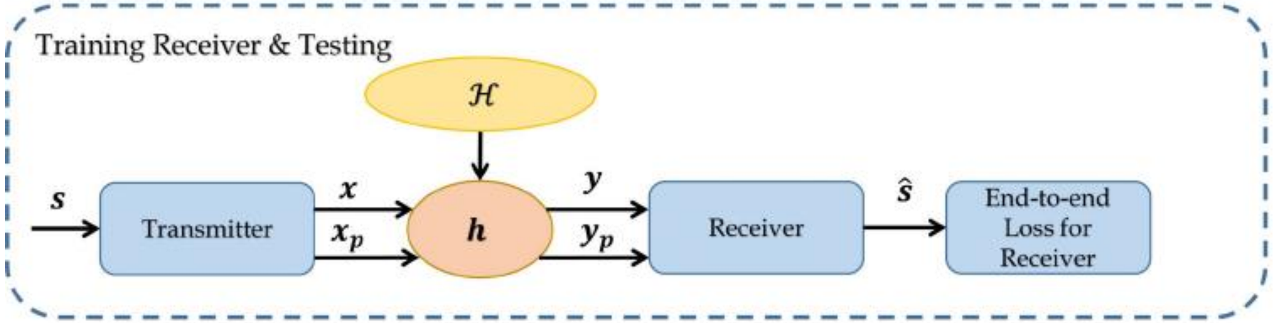
- Training
  - Tx, Rx, and C-GAN is trained with same training data
  - When train Tx & Rx, objective = minimizing 'End-to-End Loss'
  - When Train C-GAN, are used

$$\mathcal{L}_G = \min_{\boldsymbol{\theta}_G} \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z},\mathbf{m}),\mathbf{m}))],$$

$$\mathcal{L}_D = \max_{\boldsymbol{\theta}_D} \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log(D(\mathbf{x},\mathbf{m}))] + \mathbb{E}_{\mathbf{z}\sim p_z}[\log(1 - D(G(\mathbf{z},\mathbf{m}),\mathbf{m}))],$$

- Test
  - Evaluate 'end-to-end reconstruction performance with real channel

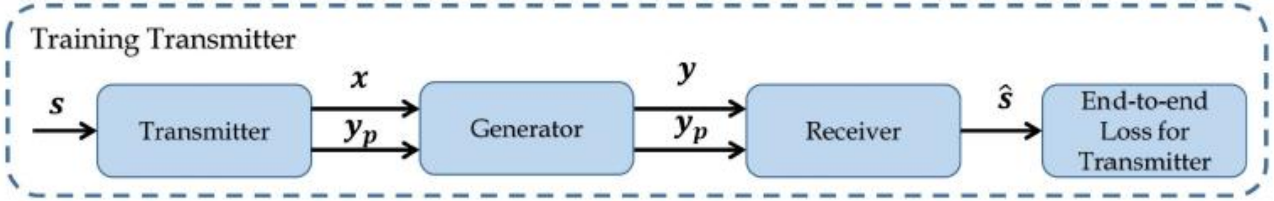
4. END-TO-END COMMUNICATION SYSTEM
4.2 TRAINING RECEIVER



$$L = \sum_{n=1}^{N} (s_n \log(\hat{s}_n) + (1 - s_n) \log(1 - \hat{s}_n))$$

- Based on loss, the receiver can be trained
- For the time-varying channels
  - Receiver can infer the channel condition
  - Perform the channel estimation and detection simultaneously
    - By putting received signal(y) & received pilot signal(y_p)

4. END-TO-END COMMUNICATION SYSTEM
4.3 TRAINING TRANSMITTER



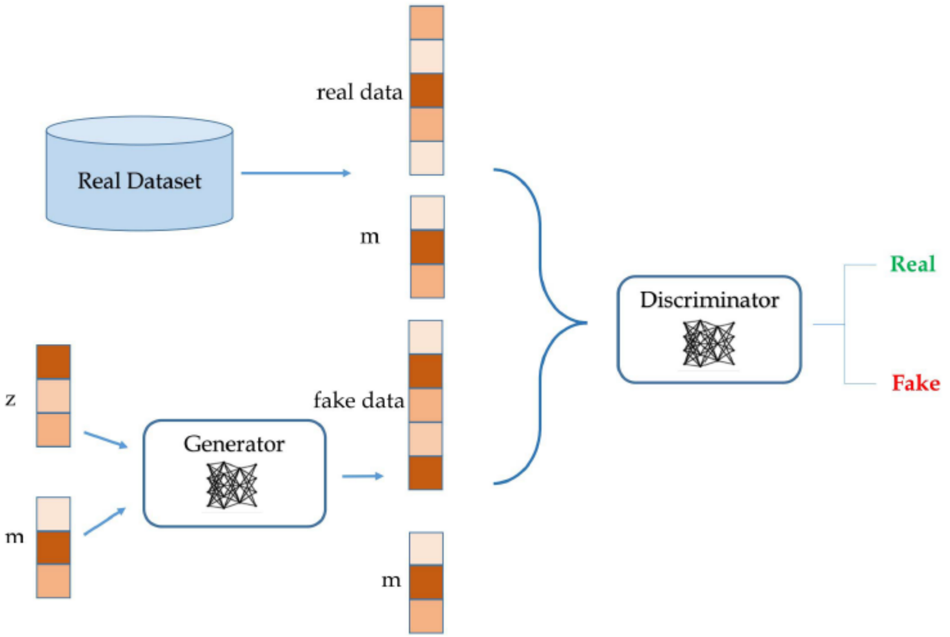
- Tx, generator and Rx can be viewed as a whole DNN
- End-to-End cross-entropy loss is computed at the Rx
- Backpropagation
  - Weights of the Tx is updated based on SGD
  - While the weighs of C-GAN and Rx remain fixed
- Tx learn the constellation of the embedding x

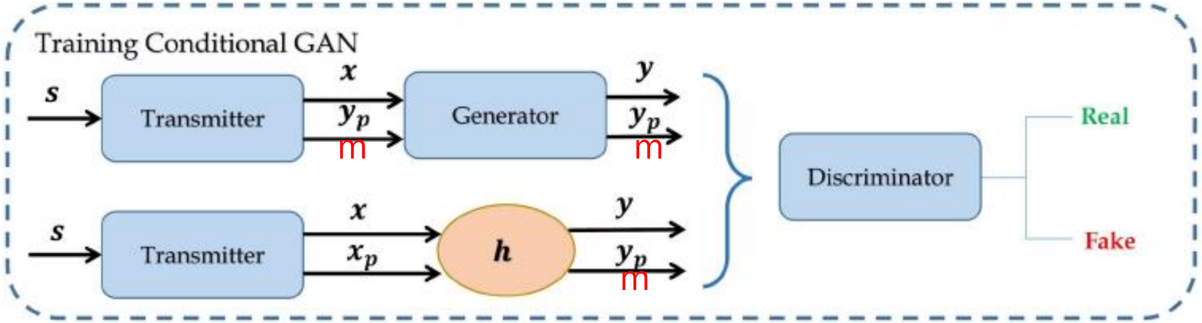
$$L = \sum_{n=1}^{N} (s_n \log(\hat{s}_n) + (1 - s_n) \log(1 - \hat{s}_n))$$

4. END-TO-END COMMUNICATION SYSTEM
4.4 TRAINING CHANNEL GAN

$$\mathcal{L}_G = \min_{\boldsymbol{\theta}_G} \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(1 - D(G(\mathbf{z},\mathbf{m}),\mathbf{m}))],$$

$$\mathcal{L}_D = \max_{\boldsymbol{\theta}_D} \mathbb{E}_{\mathbf{x}\sim p_{data}}[\log(D(\mathbf{x},\mathbf{m}))] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(1 - D(G(\mathbf{z},\mathbf{m}),\mathbf{m}))],$$





Algorithm 1 Channel GAN Training Algorithm

1: **for** number of training iterations **do**
2: *% Updating the Generator*
3: Sample minibatch of transmit data $\{\mathbf{s}\}$ and channel $\{\mathbf{h}\}$.
4: Get the minibatch of condition information $\{\mathbf{m}\}$ from the output of the transmitter DNN and the received pilot signal from the channel $\{\mathbf{h}\}$.
5: Sample minibatch of samples $\{\mathbf{z}\}$.
6: Update the generator by ascending the stochastic gradient of the loss function (3).
7: *% Updating the Discriminator*
8: Sample minibatch of transmit data $\{\mathbf{s}\}$ and channel $\{\mathbf{h}\}$.
9: Get the minibatch of condition information $\{\mathbf{m}\}$ from the output of the transmitter DNN and the received pilot signal from the channel $\{\mathbf{h}\}$.
10: Sample minibatch of examples real data by collecting the output of channel.
11: Sample minibatch of samples $\{\mathbf{z}\}$.
12: Update the discriminator by descending the stochastic gradient of the loss function (4).
13: **end for**

5. EXPERIMENTS

- Channel GAN has shown the ability to model the channel effects in a data-driven way
- End-to-End communication system achieved similar or better results even the channel information is unknown

5. EXPERIMENTS
5.1 EXPERIMENTAL SETTINGS
5.1.1 IMPLEMENTATION DETAILS

- FCN is used for a small block size
- CNN is used in the large block size to avoid the curse of dimensionality
- Weights are updated by Adam
- Batch size for training is 320

MODEL PARAMETERS OF FCN

Parameters	Values
Transmitter: Neurons in each hidden layers	32, 32
Learning rate	0.001
Receiver: Neurons in each hidden layers	32, 32
Learning rate	0.001
Generator: Neurons in each hidden layers	128, 128, 128
Discriminator: Neurons in each hidden layers	32, 32, 32
Learning rate	0.0001

MODEL PARAMETERS OF CNN

Type of layer	Kernel size/Annotation	Output size
Transmitter		
Input	Input layer	$K \times 1$
Conv+Relu	5	$K \times 256$
Conv+Relu	3	$K \times 128$
Conv+Relu	3	$K \times 64$
Conv	3	$K \times 2$
Normalization	Power normalization	$K \times 2$
Receiver		
Conv+Relu	5	$K \times 256$
Conv+Relu	5	$K \times 128$
Conv+Relu	5	$K \times 128$
Conv+Relu	5	$K \times 128$
Conv+Relu	5	$K \times 64$
Conv+Relu	5	$K \times 64$
Conv+Relu	5	$K \times 64$
Conv+Sigmoid	3	$K \times 1$
Generator		
Conv+Relu	5	$K \times 256$
Conv+Relu	3	$K \times 128$
Conv+Relu	3	$K \times 64$
Conv	3	$K \times 2$
Discriminator		
Conv+Relu	5	$K \times 256$
Conv+Relu	3	$K \times 128$
Conv+Relu	3	$K \times 64$
Conv+Relu	3	$K \times 16$
FC +Relu	100	100
FC+Sigmoid	1	1

5. EXPERIMENTS
5.1 EXPERIMENTAL SETTINGS
5.1.2 CHANNEL TYPES

- **AWGN channels**, **Rayleigh channels**, and **frequency-selective multipath channels** are considered in these experiments.
- AWGN channels
  - $y = x + w(gaussian\ noise)$
- Rayleigh channels
  - Rayleigh fading is a reasonable model for narrowband wireless channels when many objects scatter the radio signal before arriving at the receiver.
  - $y = h_n * x + w(gaussian\ noise)$
    - h_n : channel coefficient – time varying & unknown
- Frequency-selective channels
  - Radio signal propagates via multiple paths which cause undesired frequency-selective fading and time dispersion of the received signal

- Baseband complex channel impulse response of frequency-selective channels

$$h(t) = \sum_{k=0}^{K_p} b_k e^{j\theta_k} p(t - \tau_k),$$

  - K_p : # of paths
  - b_k : path gain
  - θ_k : phase shift
  - τ_k : time delay of the k-path
  - $p(t)$: shaping pulse in the communication system

5. EXPERIMENTS
5.1 EXPERIMENTAL SETTINGS
5.1.3 BASELINES

- Proposed system is compared with the conventional communication system
- BER(bit-error rate) and BLER(block-error rate) are compared under each type of channel
- Modulation
  - 4-QAM
- Channel Coding
  - Hamming code
  - Convolutional codes
    - Viterbi algorithm is used for maximum likelihood sequence decoding
    - Rate-1/2 RSC(Recursive systematic convolutional) code is used as common

5. EXPERIMENTS
5.2 MODELING THE CHANNEL EFFECTS

- To model the effects of **Rayleigh fading channels**, FCN is used.
  - Rayleigh fading channels : Time-varying
  - => As conditional information, **encoded signal** and **received pilot data** are used.
  - The effectiveness of the C-GAN in learning the distribution of the channel with standard 16-QAM as the encoded symbol can be shown

- KL divergence of generated distribution and the real distribution $KL(G(z)||p_{data})$
  - The generated distribution converges to the target distribution

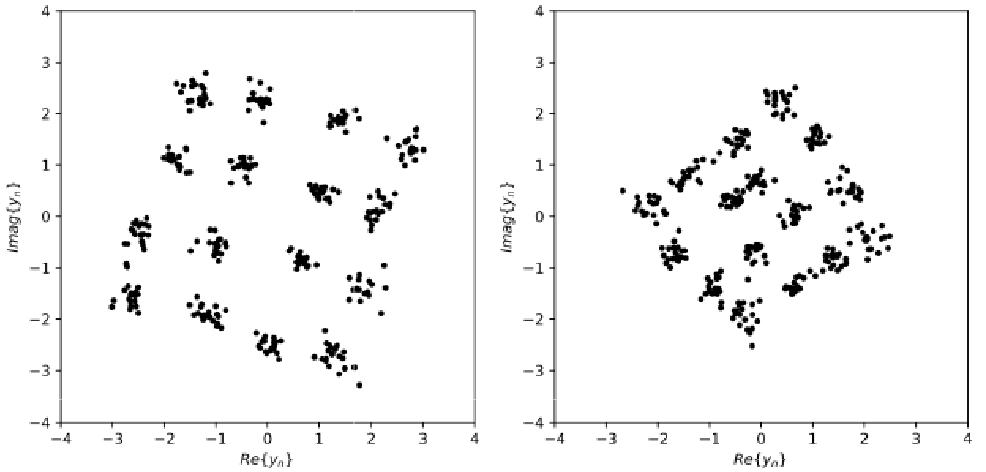


Fig. 4. Signal constellations at the output of a Rayleigh channel represented by a conditional GAN, where channel gains and phase rotations vary according to conditioning information $\mathbf{y}_P$.

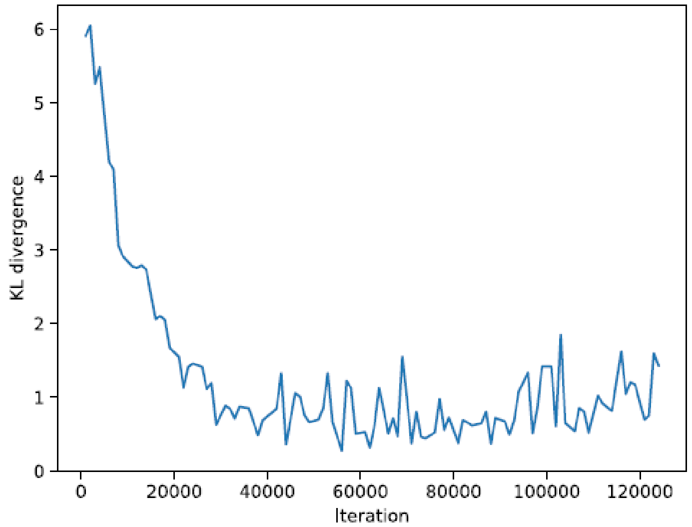


Fig. 5. KL divergence of the generated channel distribution and real channel distribution.

5. EXPERIMENTS
5.3 END-TO-END COMMUNICATION SYSTEM
5.3.1 AWGN CHANNEL

- FCN is used for small block size
  - 4 information bits are transmitted
  - Length of Transmitter output : 7
  - Performance
    - Similar to Hamming(7,4) code with maximum-likelihood decoding(MLD)
- CNN is used for large block size
  - 1. train CNN under AWGN channel (SNR : 3dB)
  - 2. test in different SNR
  - Performance
    - similar to RSC in the low SNR
    - Outperform RSC in the high SNR

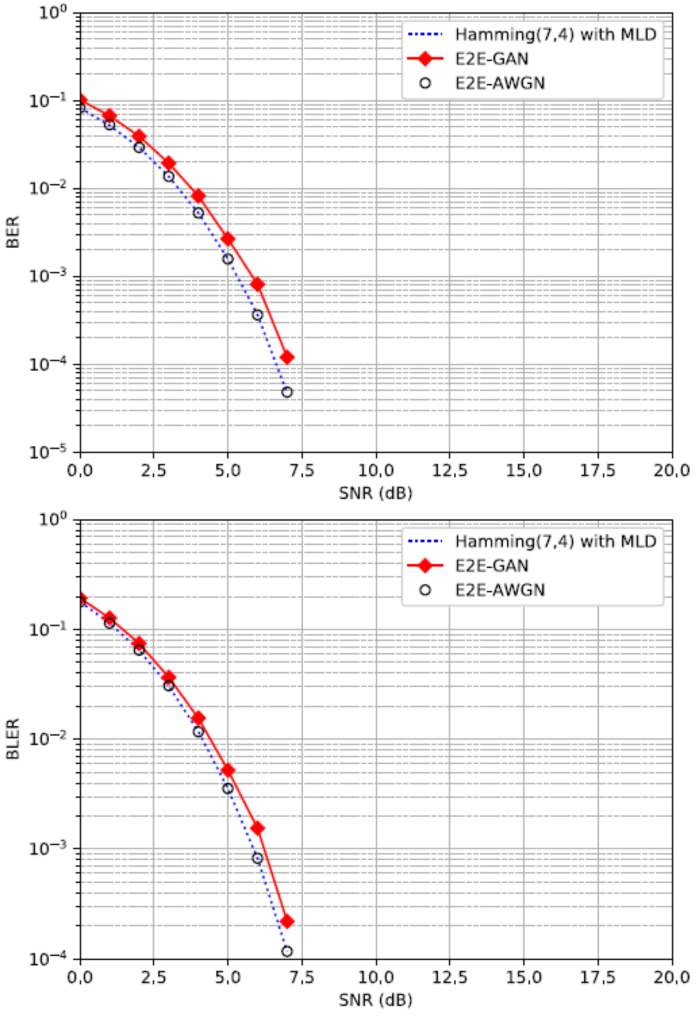


Fig. 6. BER and BLER with a block size of 4 bits under a AWGN channel.

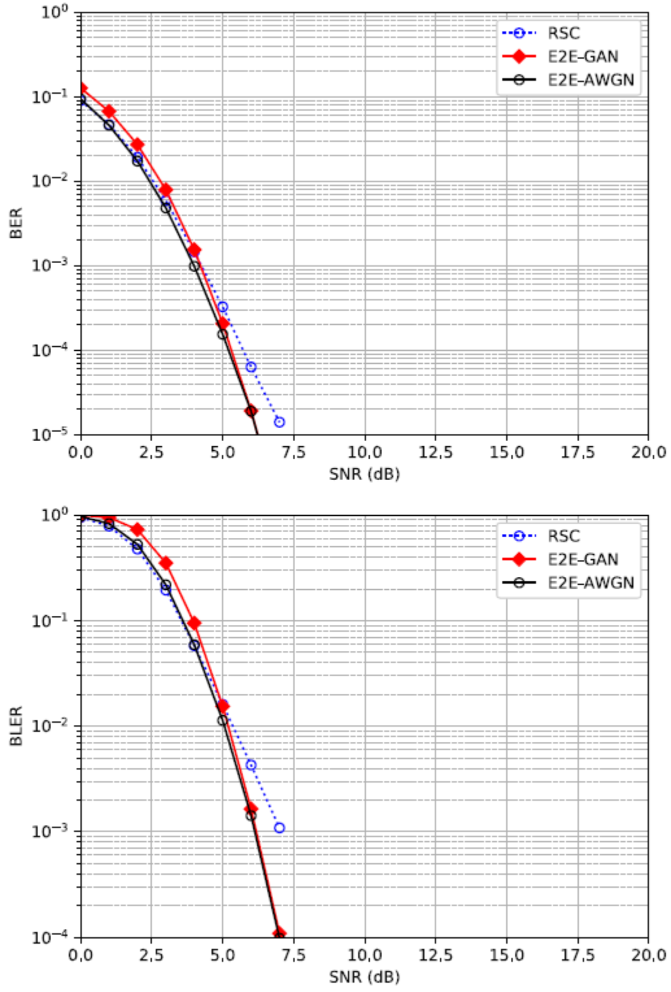


Fig. 7. BER and BLER with a block size of 128 bits under an AWGN channel.

5. EXPERIMENTS
5.3 END-TO-END COMMUNICATION SYSTEM
5.3.2 RAYLEIGH FADING CHANNEL

- CNN is used for large block
  - Channel encoding is included
  - Modulation : QAM / coding : RSC
- Performance
  - Similar to QAM+RSC & model-aware end-to-end system trained with Rayleigh channel

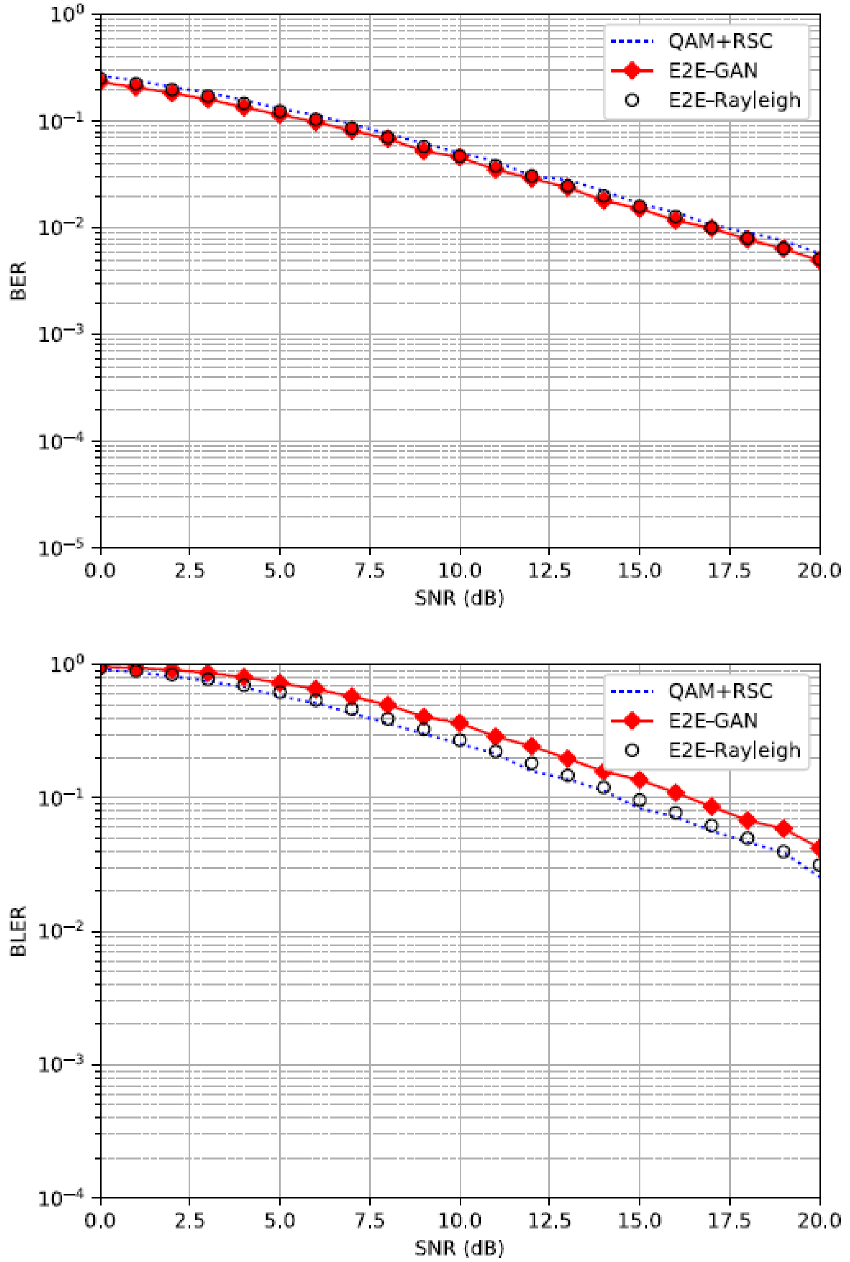


Fig. 8. BER and BLER under a Rayleigh channel.

5. EXPERIMENTS
5.3 END-TO-END COMMUNICATION SYSTEM
5.3.3 FREQUENCY-SELECTIVE FADING CHANNEL

- CNN is used
- OFDM system is used as the baseline
  - Orthogonal frequency-division multiplexing
- Modulation : 4 QAM / Channel coding : RSC
- Performance
  - Outperforms the OFDM system
  - With pilot information,
    - Rx can obtain information on a specific channel realization
  - Without pilot information
    - Channel generator learns to produce the mixture channel output distributions of all channel realizations
    - Performance degrades significantly

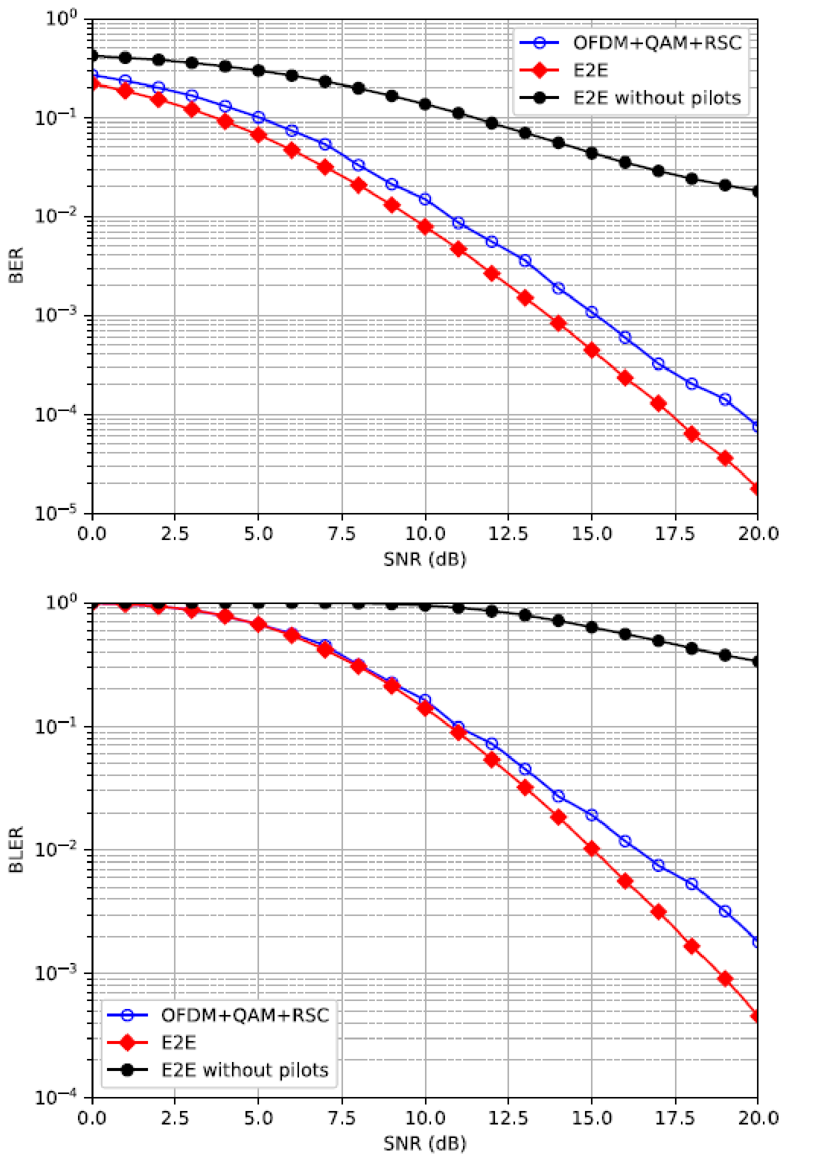


Fig. 9. BER and BLER under a frequency-selective multipath channel.

6. CONCLUSION AND DISCUSSION

- The end-to-end learning of a communication system without prior information of the channel is proposed
- By a C-GAN, the conditional distribution of the channel can be modeled
- C-GAN can generate data corresponding to the specific instaneous channel by adding pilot information
- The performance showed similar or better than traditional approaches