WIRELESS DEEP SPEECH SEMANTIC TRANSMISSION

Zixuan Xiao *, Shengshi Yao*, Jincheng Dai*, Sixian Wang*, Kai Niu†, Ping Zhang*

* Beijing University of Posts and Telecommunications, Beijing, 100876, China
† Peng Cheng Laboratory, Shenzhen, 518067, China
Email: daijincheng@bupt.edu.cn

ABSTRACT

In this paper, we propose a new class of high-efficiency semantic coded transmission methods to realize end-to-end speech transmission over wireless channels. We name the whole system as Deep Speech Semantic Transmission (DSST). Specifically, we introduce a nonlinear transform to map the speech source to semantic latent space and feed semantic features into source-channel encoder to generate the channel-input sequence. Guided by the variational modeling idea, we set an entropy model on the latent space to estimate the importance diversity among semantic feature embeddings. Accordingly, these semantic features of different importance can be reasonably allocated with different coding rates, which maximizes the system coding gain. Furthermore, we introduce a channel signal-to-noise ratio (SNR) adaptation mechanism such that a single model can be applied over various channel states. The end-to-end optimization of our model leads to a flexible rate-distortion (RD) trade-off, supporting an adaptive rate wireless speech semantic transmission. Experimental results verify that our DSST system clearly outperforms current engineered speech transmission systems on both objective and subjective metrics. Compared with existing neural speech semantic transmission methods, our model saves up to 75% of channel bandwidth costs when achieving the same quality. Audio samples are available at https://ximoo123.github.io/DSST.

Index Terms— Semantic communications, speech transmission, nonlinear transform source-channel coding, RD trade-off.

1. INTRODUCTION

State-of-the-art (SOTA) engineered speech transmission methods over the wireless channel can usually be separated into two steps: source coding and channel coding [1]. Source coding performs a linear transform on the waveform, which removes redundancy and decorrelate, and channel coding provides error-correction against imperfect wireless channels. However, the optimality of separation-based schemes holds only in the asymptotic limit of infinite source and channel blocklengths, which is impossible in actual communication scenarios. Moreover, the cliff effect exists in separation-based schemes where the performance breaks down when the channel capacity goes below the communication rate.

Hence, it is very time to bridge source and channel coding to boost end-to-end communications performance. Conventional joint source-channel coding (JSCC) methods are based on statistical probabilities while ignoring source semantic aspects. With the development of deep learning, a growing number of researchers are devoted to using neural network to extract semantic features and realize JSCC for efficient semantic transmission [2–4]. In this paper, we focus on speech transmission. Recent study DeepSC-S [5] employs the same method to accurately transmit speech waveform, showing the potential of joint coding schemes. However, they encode pulse code modulation (PCM) signals directly without removing the decorrelate between samples as in SOTA schemes, which leads to a colossal waste of bandwidth.

To address these shortcomings, we aim to integrate the advantages of classical transfer schemes and deep learning to realize a high-efficiency end-to-end deep speech semantic transmission (DSST). Specifically, we first introduce a nonlinear transform to extract waveform semantic features [6–8]. Then, we develop the Transformer architecture [9] to encode semantic features to the channel input symbols directly. Particularly, to get a higher coding gain, we estimate the importance of semantic features by a learnable entropy model and feed the value into the encoder to guide the assignment of coding rate more reasonably accordingly. The whole system is formulated as an optimization problem aiming to minimize the end-to-end rate-distortion (RD) performance. It achieves the highest possible performance at any given transmission rate.

Finally, we verify the performance of DSST across various channel conditions. Compared with the widely used classical separate schemes adaptive multi-rate wideband (AMR-WB) + 5G LDPC, Opus + 5G LDPC [10–12], our model gets a noticeable improvement in both objective scores and subjective scores across almost bandwidth consumptions and signal-to-noise ratios (SNRs) under additive white gaussian noise (AWGN) channel and COST2100 fading channel [13]. In addition, our model outperforms the existing neural method DeepSC-S in these two evaluation metrics. With a similar objective score, our model saves up 75% of the bandwidth cost.

2. OUR PROPOSED DSST SYSTEM

2.1. Architecture

Our network architecture, shown in Fig. 1, comprises a semantic transform network, a prior codec, and a Deep JSCC codec. Given a speech sample \( x \) in time domain, it is first transformed into semantic latent representation \( y \) in the semantic feature domain by a DNN-based nonlinear analysis transform \( g_a \). Then, the latent feature \( y \) is fed into both prior encoder \( h_a \) and JSCC encoder \( f_c \). On the
one hand, \( h_\alpha \) captures the temporal dependencies of \( y \) and utilizes a learnable entropy model to characterize the semantic importance of rate allocation. On the other hand, \( f_c \) encodes \( y \) as channel-input sequence \( s \) for transmission over the wireless channel and the received sequence is \( \hat{s} \). In this paper, we consider AWGN channel and COST2100 wireless fading channel, such that \( s \) can be written as \( s = W(s|h) = h \odot s + n \), where \( h \) is the CSI vector, \( \odot \) is the element-wise product and \( n \) denotes the noise vector, it is independently sampled from Gaussian distribution, i.e. \( n \sim CN(0, \sigma^2 I) \). The structure of the receiver is with a mirrored architecture. As illustrated in Fig. 1, \( f_d \) is the Deep JSCC decoder which aims to recover latent representation \( \hat{y} \), \( \hat{y} \) is semantic synthesis transform for reconstructing source signals \( \hat{x} \) and \( h_\alpha \) is prior decoder. Hence, the total link of the DSST system is formulated as follows:

\[
x \xrightarrow{s_e(x)} y \xrightarrow{f_c(x)} s \xrightarrow{W(s|h)} \hat{s} \xrightarrow{f_d(h)} \hat{y} \xrightarrow{h_\alpha(y)} \hat{x}
\]

with the latent prior \( y \xrightarrow{h_\alpha(y)} z \xrightarrow{h_\beta(z)} (\mu, \sigma) \).

Moreover, we design an optional transmission according to the actual channel situation. When the channel capacity is large enough, we transmit the side information \( z \) to the receiver by reliable digital coding schemes as a hyperprior to acquire higher decoding gain. When there is a shortage of channel capacity, \( z \) will not be transmitted, and the system performance will be slightly degraded as a result.

### 2.2. Variational Modeling and Source-Channel Coding

In our intuition, the speech signal at the silent time almost carries few information. Therefore, signals with zero magnitudes should be allocated less channel bandwidth and vice versa. If we assign the same coding rate for semantic features of all frames, the coding efficiency cannot be maximized. To tackle this problem, we design an entropy model to accurately estimate the distribution of \( y \) such that it evaluates the importance of different semantic features. Moreover, we try to make the Deep JSCC encoder able to aware of the semantic characteristics of the source and reasonably allocate coding rate according to the semantic importance.

Our entropy model is illustrated in the middle part of Fig. 1. Following the work of [8], the semantic feature \( y \) is variational modeled as a multivariate Gaussian with the mean \( \mu \) and the standard deviation \( \sigma \). Thus the true posterior distribution of \( y \) can be modeled by a fully factorized density model as:

\[
p_{y|z}(y|z) = \prod_i \left( N(y_i|\mu_i, \sigma_i) \ast U\left( -\frac{1}{2}, \frac{1}{2} \right) \right) (y_i)
\]

with \((\mu, \sigma) = h_\beta(z)\).

where \(*\) is convolutional operation, \( U \) denotes a uniform distribution, it utilized to match the prior to the marginal such that the estimated entropy \(-\log p_{y|z}(y|z)\) is non-negative. During optimizing, the actual distribution \( d_{y|z} \) created by \( h_\beta \) will gradually approximate the true distribution \( p_{y|z} \), hence the entropy model estimates distribution of \( y \) accurately. The entropy value of each frame \(-\log p_{y|z}(y_i|z)\) will be fed into JSCC encoder and guides the allocation of coding rate accordingly. If the entropy model indicates \( y_i \) of high entropy, the encoder will allocate more resources to transmit it. The total channel bandwidth cost \( K \) for transmitting \( y \) can be written as:

\[
K = \sum_{i=1}^{B} k_{y_i} = \sum_{i=1}^{B} Q(k_{y_i}) = \sum_{i=1}^{B} Q(-\eta_y \log p_{y|z}(y_i|z)),
\]

where \( B \) is the number of frames of the speech signal, \( \eta_y \) is a hyperparameter to balance channel bandwidth cost from the estimated entropy, \( k_{y_i} \) is the bandwidth consumption of frame \( i \), \( Q \) denotes a scalar quantization whose range includes \( n \) integers.

### 2.3. Optimization Goal

The optimization goal of the DSST system is to attain an RD tradeoff between channel bandwidth cost and speech reconstruction quality. It is proven by Ballé that it is more reasonable to model the whole model as a variational autoencoder [8], such that minimizing the KL divergence is equivalent to optimizing the model for RD performance. Hence, the goal of inference model is creating a variational density \( q_{\theta,z|x} \) to approximate the true posterior \( p_{z,x|a} \). The RD function can be approximated as:

\[
L_{RD} = \mathbb{E}_{x \sim p_X} D_{K L} [q_{\theta,z|x} \| p_{z|x}] \leftrightarrow \mathbb{E}_{x \sim p_X} (\mathbb{E}_{\tilde{z} \sim q_{\theta,z|x}} [−\log p_{z}(\tilde{z}) − \log p_{z|x}(\tilde{z}|x)]) + \mathbb{E}_{y \sim p_{y|x,z}} \left[ \log p_{y}(y|x) \right].
\]

The first term is seen to be identical to the cross entropy between the marginal \( q_{\theta}(\tilde{z}) = \mathbb{E}_{x \sim p_X} \mathbb{E}_{\tilde{z} \sim q_{\theta,z|x}} ([−\log p_{z}(\tilde{z}) − \log p_{z|x}(\tilde{z}|x)]) \) and the prior \( p_{z}(\tilde{z}) \), it represents the cost of encoding the side information assuming \( p_{z} \) as the entropy model. Note \( p_{z}(\tilde{z}) \) is modeled as non-parametric fully factorized density [8] as

\[
p_{z}(\tilde{z}) = \prod_i \left( p_{z_i|\psi(z)}(\tilde{z}) \ast U\left( -\frac{1}{2}, \frac{1}{2} \right) \right),
\]

where \( \psi(z) \) encapsulates all the parameters of \( p_{z_i|\psi(z)} \). In addition, to allow optimization via gradient descent during model training phase [7], \( \tilde{z} \) is added an uniform offset \( o \) instead of scalar quantization \( \tilde{z} = \lfloor z \rceil \), i.e. \( \tilde{z} = z + o \) with \( o_i \sim U\left( -\frac{1}{2}, \frac{1}{2} \right) \).

The second term is similar to the first term, representing the cost of encoding \( \tilde{s} \) and denoting the transmission rate of the speech signal. In practice, we utilize the intermediate proxy variable \( y \) to derive \( p_{z|x} \). During the transmission process, \( y \) is directly fed into the JSCC encoder and channel without quantization. The process can be described as:

\[
p_{z|x} = W(p_{z|x|h}) = W(f_c(p_{y|x})|h).
\]
Thus, the density $p_{z|y}$ can be approximated to $p_{z|x}$ and transformed to $p_{y|z}$. Similarly, $z$ is added an uniform offset instead of scalar quantization, the transmission rate is constrained proportionally to $-\log P_{y|x}(y|z)$.

The third term represents the log-likelihood of recovering $x$. Based on the above analysis, the RD function can be simplified to

$$L_{RD} = E_{x \sim p_x}[-\eta_y \log p_{y|x}(y|x) - \eta_z \log p_{z}(z) + \lambda d(x, \hat{x})],$$

(7)

where the Lagrange multiplier $\lambda$ determines the tradeoff between the wireless transmission total bandwidth cost $R$ and the end-to-end distortion $D$. $\eta_y$ and $\eta_z$ are two hyperparameters to balance channel bandwidth consumption of $y$ and $z$. $d(\cdot, \cdot)$ is the distortion between the original signals and reconstructed signals.

In our paper, we employ the distortion function $d$ both in time domains and frequency domains. Firstly, we utilize mean square error (MSE) between the original signals $x$ and reconstructed signals $\hat{x}$ to evaluate the reconstructed error in time domain. Then we compute MFCCs [14] for the source signal and reconstructed signal and employ $l_2$ distance between MFCC vectors in frequency domains to pursue better human perceptual quality [15]. Considering the spectral tilt phenomenon [16], if we use MSE as a loss function in frequency domains, the relative error of high-frequencies will be more significant, leading to the loss of high frequency information. Hence, we finally apply normalized mean squared error (NMSE) as the loss function, which helps the model be optimized toward reconstructing high frequencies and gives speech higher human perceptual quality.

### 2.4. Modular Implementation Details

In this part, we will share some details of DSST. We take speech signals as an example. After preprocessing operations, the input frame $x$ is modeled as a vector $x \in \mathbb{R}^{B \times C \times L}$, where $B$ is the number of frames, $C$ is the number of sound channels, and $L$ is the frame length.

1. **Semantic Analysis and Synthesis Transform $g_a, g_s$:** The semantic analysis and synthesis transform are used a similar structure as the Kankanahalli network [17]. It contains multilayer 1D convolutional blocks and residual networks to extract the semantic features of the waveform efficiently. The latent feature can be written as $y \in \mathbb{R}^{B \times M \times \frac{C}{2}}$ after twice downsampling, $M$ is channel dimension.

2. **Prior Codec $h_s, h_a$:** The structure of $h_s$ is consisted of $3$ convolutional blocks with $N$ channels. The first block contains a convolutional layers with kernel size $3$ and a ReLU activation functions. Other blocks are composed of a down-sampling layer consisting in a strided convolution, with kernel size $5$ and stride $2$, followed by a ReLU activation functions. $h_a$ mirrors $h_s$. The hyperprior model summarizes the distribution of means and standard deviation in $z \in \mathbb{R}^{B \times N \times \frac{C}{4}}$. It effectively captures the semantic temporal connections of the speech waveform.

3. **Adaptive rate Deep JSCC Codec $f_e, f_d$:** The architecture of Deep JSCC codec is illustrated in Fig. 2. The encoder consists of $N_c$ Transformer blocks and FC layers to implement rate allocation. Specifically, $y$ is firstly separated into patch embedding sequence $\{y_1, y_2, \ldots, y_B\}$. To adaptively map $y_i$ to a $k_y$-dimensional channel-input vector $s_i$, we develop a rate token vector set $R = \{r_{y_1}, r_{y_2}, \ldots, r_{y_B}\}$ to represent rate information and quantization value set $V = \{k_{y_1}, k_{y_2}, \ldots, k_{y_B}\}$ to denote output dimension. By this approach, each $y_i$ is merged with its corresponding rate token $r_{y_i}$ according to the entropy $-\log P_{y_i|z}(y_i|z)$ and is fed into the Transformer block and FC layer to transform into a $k_y$-dimensional vector. In particular, we employ a group of FC layers with different output dimensions, it guided by the rate token $r_{y_i}$ accordingly.

Moreover, we perform SNR adaptation in this process. This paper assumes that the transmitter and receiver can receive the SNR feedback information for better performance. As illustrated in Fig. 2, each patch $y_i$ is concatenated with SNR token $C \in \mathbb{R}^B$, such that the Transformer can learn the SNR information. Hence, the single model can perform at least as well as the models trained for each SNR individually finally when trained under random SNRs.

4. **Optional Transmission Link:** If we transmit $z$ to obtain the decoding gain, we will quantize $z$ first and perform entropy coding and channel coding on it, which ensure an reliable transmission. At the receiver, the prior decoder reconstructs side information $\hat{\mu}, \hat{\sigma}$, and feeds it into the Deep JSCC decoder.

### 3. EXPERIMENTS

In this section, we provide illustrative numerical results to evaluate the quality of speech waveform transmission. Objective quality metrics and a result of a subjective listening test are presented to validate our designed approach.

#### 3.1. Experimental Setups and Evaluation Metrics

The waveform is sampled at 16kHz from TIMIT dataset [18]. During total training process, we use Adam optimizer [19] with $B = 100$ frames in a single batch and $L = 512$ samples in a frame.

To evaluate our model, we use an objective evaluation metric and conduct a subjective experiment as well. Concerning objective quality evaluation, we report perceptual evaluation of speech quality (PESQ) [20] scores to reflect the reconstructed speech quality. Furthermore, we implement MUSHRA subjective test [21] to evaluate audio quality by human raters. We randomly select ten reconstructed waveform signals from the test dataset. To demonstrate the superiority of our model, the comparison schemes include the traditional transmission methods: AMR-WB + 5G LDPC and Opus + 5G LDPC, a standard neural JSCC model DeepSC-S. For traditional separation methods, we choose different coding and modulation schemes according to the adaptive modulation coding (AMC) principle [22].

#### 3.2. PESQ Performance

Fig. 3(a) shows the rate-quality results among various transmission methods over the AWGN channel with channel SNR at 6dB. The proposed DSST method outperforms traditional AMR-WB + 5G LDPC and Opus + 5G LDPC under all bandwidth consumption situations. It is mainly because our scheme estimates the distribution of speech source accurately, which mitigates the mismatch between
source coding and channel coding and gets the compression gain. Moreover, the DSST scheme demonstrates its flexibility. The model can implement an efficient speech transmission under any transmission rate. Compared with traditional schemes, neither AMR-WB nor Opus schemes can effectively transmit the speech at low bandwidth.

Fig. 3(b) and Fig. 3(c) discuss the performance of the model under different SNRs. Similar to the above findings, the proposed DSST scheme outperforms separation-based traditional schemes by a large margin for all SNRs. Furthermore, illustrated in Fig. 3(a) and Fig. 3(c), when PESQ is 4.3 and the SNR is 6dB, DeepSC-S uses 32kHz bandwidth, while our model only uses 8kHz bandwidth. It saves about 75% of channel bandwidth consumption. This gain comes from both the semantic transform module and the Deep JSCC module. The former drops most of the redundancy in speech signals. The latter allocates different coding rates more reasonably, which gets the most out of coding resources.

To further verify the effectiveness of our model, we carry out experiments on the practical COST2100 fading channel. CSI samples are collected in an indoor scenario at 5.3GHz bands, and all schemes use one-shot transmission. Illustrated in Fig. 3(d), DSST shows strong robustness to channel variations. Especially under the condition of low SNRs, our model incomparably outperforms other schemes. Compared with existing neural methods DeepSC-S, our model achieves more than a 30% increase in terms of PESQ performance when they use similar bandwidth resources.

3.3. User Study
To better align with human perception, we have conducted the MUSHRA subjective test user study. The first three figures show the results under the AWGN channel. Fig. 5(a) shows the performance under the low bandwidth cost condition. The DSST significantly surpasses two separate coding schemes. Fig. 5(b) reveals the performance under a low SNR situation. We observe similar results, and our model achieves a three times higher MUSHRA score than DeepSC-S with similar bandwidth. Fig. 5(c) compares various transmission schemes with the DSST under different bandwidth costs. To match the speech quality of DSST, Opus + 5G LDPC needs to use 7kHz bandwidth, while AMR-WB + 5G LDPC needs at least 7.5kHz bandwidth. It saves more than 45% of bandwidth resources. Moreover, DSST reconstructed high-quality speech using as little as 4kHz bandwidth, with quality significantly better than DeepSC-S at 12kHz. At last, Fig. 5(d) illustrates the performance over the COST2100 fading channel. Our model still outperforms other schemes in a real fading channel.

We provide examples of reconstructed speech for an auditory comparison at https://ximoo123.github.io/DSST.

4. CONCLUSION
We present the DSST, a new class of high-efficiency semantic speech transmission model over the wireless channel. The model first extracts the semantic features of the speech waveform and then transmits it. Particularly, our model takes advantage of the nonlinear transform method, learnable entropy model, and adaptive rate Deep JSCC model to get higher coding gain. Results indicate that our model achieves better performance at various transmission rates and outperforms the separation-based traditional coding schemes and the existing neural scheme DeepSC-S under both the AWGN channel and COST2100 wireless channel. It is exciting that the DSST reconstructs high-quality speeches under harsh channel conditions. It will be of great help to future speech semantic communication.
5. REFERENCES

[1] Claude Elwood Shannon, “A mathematical theory of communication,” The Bell system technical journal, vol. 27, no. 3, pp. 379–423, 1948.

[2] David Burth Kurka and Deniz Gündüz, “Bandwidth-agile image transmission with deep joint source-channel coding.” IEEE Transactions on Wireless Communications, vol. 20, no. 12, pp. 8081–8091, 2021.

[3] Nariman Farsad, Milind Rao, and Andrea Goldsmith, “Deep learning for joint source-channel coding of text.” in 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2018, pp. 2326–2330.

[4] Jincheng Dai, Sixian Wang, Kailin Tan, Zhongwei Si, Xiaoqi Qin, Kai Niu, and Ping Zhang, “Nonlinear transform source-channel coding for semantic communications.” IEEE Journal on Selected Areas in Communications, vol. 40, no. 8, pp. 2300–2316, 2022.

[5] Zhenzi Weng and Zhijin Qin, “Semantic communication systems for speech transmission.” IEEE Journal on Selected Areas in Communications, vol. 39, no. 8, pp. 2434–2444, 2021.

[6] Johannes Ballé, Philip A Chou, David Minnen, Saurabh Singh, Nick Johnston, Eirikur Agustsson, Sung Jin Hwang, and George Toderici, “Nonlinear transform coding.” IEEE Journal Selected Topics Signal Process., vol. 15, no. 2, pp. 339–353, 2020.

[7] Johannes Ballé, Valero Laparra, and Eero P Simoncelli, “End-to-end optimized image compression.” arXiv preprint arXiv:1611.01704, 2016.

[8] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, “Variational image compression with a scale hyperprior.” in Proceedings International Conference Learning Representations, 2018.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” Advances in neural information processing systems, vol. 30, 2017.

[10] Bruno Bessette, Redwan Salami, Roch Lefebvre, Milan Jelinek, Jani Rotola-Pukkila, Janne Väinio, Hannu Mikkola, and Kari Jarvinen, “The adaptive multirate wideband speech codec (amr-wb).” IEEE transactions on speech and audio processing, vol. 10, no. 8, pp. 620–636, 2002.

[11] Tom Richardson and Shrinivas Kudekar, “Design of low-density parity check codes for 5g new radio.” IEEE Communications Magazine, vol. 56, no. 3, pp. 28–34, 2018.

[12] Jean-Marc Valin, Koen Vos, and Timothy Terriberry, “Definition of the opus audio codec,” Tech. Rep., 2012.

[13] Lingfeng Liu, Claude Oestges, Juho Poutanen, Katsuyuki Haneda, Pertti Vainikainen, François Quittin, Fredrik Tufvesson, and Philippe De Doncker, “The cost 2100 mimo channel model.” IEEE Wireless Communications, vol. 19, no. 6, pp. 92–99, 2012.

[14] Lindasalwa Muda, Muntaj Begam, and Irraivan Elamvazuthi, “Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques.” arXiv preprint arXiv:1003.4083, 2010.

[15] Yoichiro Blau and Tomer Michaeli, “The perception-distortion tradeoff.” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6228–6237.

[16] Jun-Hwey Chen and Allen Gersho, “Adaptive postfiltering for quality enhancement of coded speech.” IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 59–71, 1995.

[17] Srihari Kankanahalli, “End-to-end optimized speech coding with deep neural networks.” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 2521–2525.

[18] John S Garofolo, “Timit acoustic phonetic continuous speech corpus,” Linguistic Data Consortium, 1993, 1993.

[19] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization.” in Proceedings of the International Conference on Learning Representations (ICLR), 2015.

[20] Antony W Rix, John G Beerends, Michael P Hollier, and Andries P Hekstra, “Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs.” in 2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221). IEEE, 2001, vol. 2, pp. 749–752.

[21] B Series, “Method for the subjective assessment of intermediate quality level of audio systems,” International Telecommunication Union Radiocommunication Assembly, 2014.

[22] Fei Peng, Jinyun Zhang, and William E Ryan, “Adaptive modulation and coding for ieee 802.11 n.” in 2007 IEEE Wireless Communications and Networking Conference. IEEE, 2007, pp. 656–661.