IEEE TRANSACTIONS ON

BROADCASTING

A PUBLICATION OF THE IEEE BROADCAST TECHNOLOGY SOCIETY

IEEE Broadcast Technology Society

bts.ieee.org

SEPTEMBER 2019 VOLUME 65 NUMBER 03 ITCTEM (ISSN 0018-9316)

REGULAR PAPERS

Physical Layer Performance Evaluation of LTE-Advanced Pro Broadcast and ATSC 3.0 Systems ..... *M. Fuentes, D. Mi, H. Chen, E. Garro, J. L. Carcel, D. Vargas, B. Mouhouche, and D. Gomez-Barquero* 477

Comparison of Low-Density Parity-Check Codes in ATSC 3.0 and 5G Standards ..... *S.-K. Ahn, K.-J. Kim, S. Myung, S.-I. Park, and K. Yang* 489

Layered Division Multiplexing for ATSC 3.0: Implementation and Memory Use Aspects ..... *J.-Y. Lee, S.-I. Park, S. Kwon, B.-M. Lim, S. Ahn, N. Hur, H. M. Kim, and J. Kim* 496

Joint Layer Prediction for Improving SHVC Compression Performance and Error Concealment ..... *X. HoangVan and B. Jeon* 504

Video Streaming Adaptation Strategy for Multiview Navigation Over DASH ..... *C. Yao, J. Xiao, Y. Zhao, and A. Ming* 521

PeQASO: Perceptual Quality Assessment of Streamed Videos Using Optical Flow Features ..... *M. A. Aabed and G. AlRegib* 534

Bitrate-Based No-Reference Video Quality Assessment Combining the Visual Perception of Video Contents .. *J. Yao and G. Liu* 546

Reduction of Padding Overhead for RLNC Media Distribution With Variable Size Packets ..... *M. Taghouti, D. E. Lucani, J. A. Cabrera, M. Reisslein, M. V. Pedersen, and F. H. P. Fitzek* 558

Cache-Based Popular Services Pushing on High-Speed Train by Using Converged Broadcasting and Cellular Networks ..... *B. Li, J. Xiong, B. Liu, L. Gui, M. Qiu, and Z. Shi* 577

A Study on Dual-Polarized MIMO-ICI Canceller With Complexity Reduction Under Mobile Reception of OFDM Signals ..... *A. Nakamura, H. Otsubo, and M. Itami* 589

Receiver Design for Alamouti Coded FBMC System in Highly Frequency Selective Channels ..... *J. Li, D. Chen, D. Qu, Y. Zhang, and T. Jiang* 601

Extremely High Frequency (EHF) Bands for Future Broadcast Satellite Services: Opportunities and Challenges ..... *C. Sacchi, T. Rossi, M. Murroni, and M. Ruggieri* 609

BRIEF PAPERS

High-Level Multiple-UAV Cinematography Tools for Covering Outdoor Events ..... *I. Mademlis, V. Mygdalis, N. Nikolaidis, M. Montagnuolo, F. Negro, A. Messina, and I. Pitas* 627

Implementation Methodologies of Deep Learning-Based Signal Detection for Conventional MIMO Transmitters ..... *M.-S. Baek, S. Kwak, J.-Y. Jung, H. M. Kim, and D.-J. Choi* 636

IEEE

Physical Layer Performance Evaluation of LTE-Advanced Pro Broadcast and ATSC 3.0 Systems

Manuel Fuentes, De Mi, Hongzhi Chen, Eduardo Garro, Jose Luis Carcel, David Vargas, Belkacem Mouhouche, and David Gomez-Barquero

***Abstract*—This paper provides a detailed performance analysis of the physical layer of two state-of-the-art point-to-multipoint (PTM) technologies: evolved Multimedia Broadcast Multicast Service (eMBMS) and Advanced Television Systems Committee - Third Generation (ATSC 3.0). The performance of these technologies is evaluated and compared using link-level simulations, considering relevant identified scenarios. A selection of Key Performance Indicators for the International Mobile Telecommunications 2020 (IMT-2020) evaluation process has been considered. Representative use cases are also aligned to the test environments as defined in the IMT-2020 evaluation guidelines. It is observed that ATSC 3.0 outperforms both eMBMS solutions, i.e., MBMS over Single Frequency Networks (MBSFN) and Single-Cell PTM (SC-PTM) in terms of spectral efficiency, peak data rate and mobility, among others. This performance evaluation serves as a benchmark for comparison with a potential 5G PTM solution.**

***Index Terms*—Benchmark, point-to-multipoint, eMBMS, MBSFN, SC-PTM, ATSC 3.0, broadcasting.**

I. Introduction

POINT-TO-MULTIPOINT (PTM) communications are the only technology enabling the delivery of the same content to a practically infinite number of users simultaneously, using just a fixed amount of resources for a given coverage area. Traditionally, PTM transmissions have been used to deliver linear content (such as TV or radio) through Digital Terrestrial Television (DTT) systems. Many first-generation DTT systems are nowadays in place over the world, such as Advanced Television Systems Committee (ATSC) [1] in North America, Integrated Services Digital Broadcasting - Terrestrial (ISDB-T) [2] in Japan and South America or Digital Terrestrial Multimedia Broadcast (DTMB) [3] in China. Among these technologies utilized in many countries, Digital Video Broadcasting - Terrestrial (DVB-T) is the most widely implemented DTT standard in the world [4]. Its evolution, DVB - Second Generation Terrestrial (DVB-T2) [5], provides a 50% increase of spectral efficiency compared to DVB-T and introduces new technologies such as the use of Low-Density Parity Check (LDPC) codes or higher orders of constellation, using 256 symbols with Quadrature Amplitude Modulation (QAM). Today, the state-of-the-art DTT standard is ATSC - Third Generation (ATSC 3.0) [6]. ATSC 3.0 provides better performance than DVB-T2 in terms of carrier-to-noise ratio (CNR) and shortens the gap to the Shannon limit, thanks to the use of more efficient constellations and very robust coding rates (CR), the aggregation of multiple radio-frequency (RF) carriers or the combined provision of fixed and mobile services through the use of non-orthogonal multiplexing techniques.

DTT systems were originally developed to support mainly fixed rooftop reception. Despite the efforts to develop mobile DTT standards such as DVB - Handheld (DVB-H) [7] or DVB - Next Generation Handheld (DVB-NGH) [8], the lack of market limited their success. In parallel, the Third Generation Partnership Project (3GPP) standardization forum developed the fourth generation (4G) standard Long Term Evolution (LTE) to provide high-speed mobile broadband for handheld services through unicast. LTE also adopted the use of evolved Multimedia Broadcast Multicast Service (eMBMS) in Release (Rel-) 9 [9] to deliver mobile video through multicast and broadcast. Today, the state-of-the-art specification for PTM is LTE-Advanced Pro Rel-14, which has included additional requirements to deliver linear services to both mobiles and fixed rooftop receivers.

Since its introduction, eMBMS has gone through a very significant set of enhancements [10]. For instance, it introduced new physical, transport and logical channels in the specification to enable MBMS over Single Frequency Networks

Manuscript received May 30, 2018; revised August 8, 2018; accepted August 13, 2018. Date of publication September 5, 2018; date of current version September 4, 2019. This work was supported in part by the European Commission through the 5GPPP project 5G-Xcast (H2020-ICT-2016-2) under Grant 761498. The views expressed in this contribution are those of the authors and do not necessarily represent the project. Parts of this paper have been partially published in [11]. *(Corresponding author: Manuel Fuentes.)*

M. Fuentes was with Samsung Electronics Research and Development U.K., Staines-upon-Thames TW18 4QE, U.K. He is now with the Institute of Telecommunications and Multimedia Applications, Universitat Politecnica de Valencia, 46022 Valencia, Spain (e-mail: mafuemue@iteam.upv.es).

D. Mi and H. Chen are with the 5G Innovation Centre, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: d.mi@surrey.ac.uk; hongzhi.chen@surrey.ac.uk).

E. Garro, J. L. Carcel, and D. Gomez-Barquero, are with the Institute of Telecommunications and Multimedia Applications, Universitat Politecnica de Valencia, 46022 Valencia, Spain (e-mail: edgarcre@iteam.upv.es; jocarcer@iteam.upv.es; dagobar@iteam.upv.es).

D. Vargas is with the BBC Research and Development, London W12 7SB, U.K. (e-mail: david.vargas@bbc.co.uk).

B. Mouhouche is with Samsung Electronics Research and Development U.K., Staines-upon-Thames TW18 4QE, U.K. (e-mail: b.mouhouche@samsung.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2866778

Comparison of Low-Density Parity-Check Codes in ATSC 3.0 and 5G Standards

Seok-Ki Ahn, Kyung-Joong Kim, Seho Myung, Sung-Ik Park, *Senior Member, IEEE*, and Kyeongcheol Yang, *Senior Member, IEEE*

***Abstract*—Recently, low-density parity-check (LDPC) codes have been adopted in Advanced Television Systems Committee 3.0 and 3rd Generation Partnership Project 5G standards. In this paper, we present their structures in detail. They are delicately designed, based on the structures of quasi-cyclic LDPC codes and multi-edge type LDPC codes. The differences in their base matrices and parity-check matrices used in both standards are highlighted from the viewpoint of the distinction between broadcasting and cellular communication systems. Numerical results show that they are very competitive in their respective areas.**

***Index Terms*—ATSC 3.0, 3GPP 5G, low-density parity-check (LDPC) codes, quasi-cyclic LDPC codes, multi-edge type LDPC codes.**

I. Introduction

In 2016, the Advanced Television Systems Committee (ATSC) approved the physical layer standard for the next generation digital terrestrial television systems [1]–[3]. During the evolution of terrestrial broadcasting standards from Digital Video Broadcasting via Satellite (DVB-S) to DVB-next generation handheld (DVB-NGH) and from ATSC 1.0 to ATSC 3.0, a lot of innovative technologies have been introduced and adopted to support efficiency and robustness. One of the innovative physical layer technologies is to employ low-density parity-check (LDPC) codes for forward error correction (FEC). Since LDPC codes were first adopted in the DVB-S2 standard, many continuous efforts have been made to improve their performance and optimize them up to the latest broadcasting standard, ATSC 3.0 [4]–[6].

Manuscript received July 27, 2018; revised September 18, 2018; accepted September 19, 2018. Date of publication October 25, 2018; date of current version September 4, 2019. This work was supported in part by the National Research Foundation of Korea under grant funded by the Ministry of Science and ICT (MSIT) of the Korea Government (2016R1A2A1A05005023), and in part by the Institute for Information and Communications Technology Promotion (IITP) under grant funded by the MSIT of the Korea Government (2018(2016-0-00123)). *(Corresponding author: Kyeongcheol Yang.)*

S.-K. Ahn, K.-J. Kim, and S. Myung are with the Network Business, Samsung Electronics Company, Ltd., Suwon 16677, South Korea (e-mail: seokki.ahn@gmail.com; kjoong.kim@gmail.com; seho78.myung@gmail.com).

S.-I. Park is with the Broadcasting Systems Research Department, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea (e-mail: psi76@etri.re.kr).

K. Yang is with the Department of Electrical Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea (e-mail: kcyang@postech.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2874541

On the other hand, turbo codes, which were first adopted as an error-correcting code in the 3rd generation partnership project (3GPP) standard, have been used for cellular communication systems until the 4G LTE standard [7]. During the recent standardization process of 5G, which develops the fundamental technologies of 3GPP Release 15, they were also considered as a candidate for channel coding, together with LDPC codes and polar codes [8]. These codes were competed with each other in the aspects of performance, complexity, and flexibility, etc. As a result, LDPC codes were eventually adopted as one of the standards for channel coding in 5G [9], thanks to their excellent advantage in supporting very-high data throughput with low complexity.

LDPC codes, which were first introduced in 1963 [10], have been extensively studied as one of modern error-correcting codes for the last two decades. They have been adopted in both ATSC 3.0 and 5G due to their capacity-approaching performance and low-complexity parallel decoding suitable for hardware implementation. Hence, they can be regarded as the most promising channel coding scheme in commercial systems, even though the requirements for broadcasting and cellular communication systems are different in general.

Broadcasting systems usually require quasi error-free performance (for example, block error rate (BLER) $\leq 10^{-6}$), and therefore, they employ LDPC codes of very long length (for example, 16,200 and 64,800 in both DVB-T2/S2 and ATSC 3.0) for both mobile and fixed services. The size of payload in these systems needs not to be dynamically changed as well as the number of the required code rates can be limited because the operating code rates are generally determined in advance according to the deployment environments. Due to these reasons, the ATSC 3.0 standardization has focused on improving the performance of LDPC codes for the fixed combinations of code lengths and rates. However, the newly designed LDPC codes for the ATSC 3.0 standard have the same basic structure as that of the LDPC codes employed in conventional broadcasting standards. A few structural constraints coming from the continuity of the standards sometimes make it difficult for newly designed LDPC codes to achieve the best performance. The LDPC codes adopted in the ATSC 3.0 standard will be named as *ATSC-LDPC codes* in this paper.

In the 5G standard, various requirements are given to support several services, such as enhanced mobile broadband (eMBB), ultra-reliable and low latency (URLLC), and massive machine type communications (mMTC). In [11],

Layered Division Multiplexing for ATSC 3.0: Implementation and Memory Use Aspects

Jae-young Lee, *Senior Member, IEEE*, Sung-Ik Park, *Senior Member, IEEE*, Sunhyoung Kwon, *Member, IEEE*, Bo-Mi Lim, *Member, IEEE*, Sungjun Ahn, *Member, IEEE*, Namho Hur, *Member, IEEE*, Heung Mook Kim, *Member, IEEE*, and Jeongchang Kim, *Senior Member, IEEE*

***Abstract*—This paper presents implementation and memory use aspects for layered division multiplexing (LDM) technology defined in the next generation terrestrial broadcast standard, called advanced television systems committee (ATSC) 3.0. As LDM becomes a new method that combines multiple broadcast contents, its practical considerations on transmitter and receiver implementations as well as memory usages are described in this paper. When multiple physical layer pipes are used, the feasibility of the implementation and memory use aspects are discussed, and the performance analysis in comparison with other multiplexing techniques that ATSC 3.0 offers is shown.**

***Index Terms*—ATSC 3.0, core layer, enhanced layer, layered division multiplexing.**

I. Introduction

THE NEXT generation terrestrial broadcasting standard, called Advanced Television Systems Committee (ATSC) 3.0 has been finalized in January 2018, by defining over 20 documents with different system protocol architecture layers. This new international wide standard has been developed taking into account broadcasters' intended services, and has distinct features over other existing standards, such as Internet protocol (IP) based transmission enabling use of broadband networks, advanced emergency alerting service, and efficient use of spectrum [1].

The physical layer standard of ATSC 3.0 has been developed to provide better flexibility, robustness, and spectrum efficiency compared to existing standards [2]. The bit-interleaved coded modulation (BICM) scheme that consists of low density parity check (LDPC) codes (2/15∼13/15 rates), bit interleavers, and non-uniform constellations (NUC, up to 4096QAM), offers not only robustness approaching to near-Shannon limits, but also flexibility enabling a wide range of operating modes for broadcasters [3]–[5]. The framing and waveform schemes based on orthogonal frequency division multiplexing (OFDM) also offer flexibility so that robust mobile and/or high capacity fixed services are feasible [6]. There are other optional features to allow further enhancement of broadcasting such as channel bonding [7], multiple-input single-output (MISO) [8], and multiple-input multiple-output (MIMO) [9]. One of novel technologies adopted in the ATSC 3.0 physical layer standard is Layered Division Multiplexing (LDM). LDM is a multiplexing technique that combines different services (e.g., mobile, indoor, and fixed services) contained in physical layer pipes (PLPs) with different power levels. Since PLPs that are layered-division multiplexed together allow more efficient use of physical layer resources than PLPs time-division multiplexed (TDM) or frequency-division multiplexed (FDM), it is well-known that LDM offers significant advantages in performance and capacity, compared to those traditional multiplexing methods [10]–[17].

As the ATSC 3.0 physical layer standard defines up to two-layer LDM consisting of Core Layer and Enhanced Layer, the standard allows to combine LDM with TDM, FDM, or both when more than two PLPs are intended by broadcasters. In the hardware implementation aspect, such combination of multiplexing methods may result in the increase of complexity and memory use in general, and therefore, a guideline is needed to implement with limited hardware resources while maintain the performance advantage of LDM. This paper presents details of hardware implementation guideline in order to achieve low complexity and low memory use when multiple PLPs are configured by ATSC 3.0 LDM. The practical concerns on ATSC 3.0 transmitter and receiver implementations as well as memory usages are described. Comparison studies in capacity are provided when different multiplexing techniques are used in the ATSC 3.0 physical layer standard.

The remaining of this paper is organized as follows: In Section II, the implementation aspects of ATSC 3.0 LDM and PLP multiplexing are described. This section also discusses practical choices of LDM injection levels depending on physical layer modes, and valid/invalid operations of LDM in practice. Section III describes detailed configurations of LDM and combination with other multiplexing technique when different time interleaver modes are selected in the physical layer. In Section IV, a performance analysis of capacity comparisons

Manuscript received September 8, 2018; revised November 27, 2018; accepted November 29, 2018. Date of publication February 25, 2019; date of current version September 4, 2019. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation Grant funded by the Korea Government (MSIT, Development of Transmission Technology for Ultra High Quality UHD) under Grant 2017-0-00081. *(Corresponding author: Jeongchang Kim.)*

J.-y. Lee, S.-I. Park, S. Kwon, B.-M. Lim, S. Ahn, N. Hur, and H. M. Kim are with the Media Transmission Research Group, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea (e-mail: jaeyl@etri.re.kr; psi76@etri.re.kr; shkwon@etri.re.kr; blim_vrossi46@etri.re.kr; sjahn@etri.re.kr; namho@etri.re.kr; hmkim@etri.re.kr).

J. Kim is with the Division of Electronics and Electrical Information Engineering, Korea Maritime and Ocean University, Busan 49112, South Korea (e-mail: jchkim@kmou.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2019.2897750

Joint Layer Prediction for Improving SHVC Compression Performance and Error Concealment

Xiem HoangVan and Byeungwoo Jeon, *Senior Member, IEEE*

***Abstract*—Scalable high efficiency video coding (SHVC) standard is expected to play a more important role in the heterogeneous landscape of broadcasting, multimedia, networks, and various services applications as it is specified as a layered coding technique in the advanced television systems committee 3.0. However, its block-based structure of temporal and spatial prediction makes it sensitive to information loss and error propagation due to transmission errors. In this context, we propose an improved SHVC with a joint layer prediction (JLP) solution which adaptively combines the decoded information from the base and the enhancement layers to create an additional reference for the SHVC enhancement encoder. To optimize the quality of the joint prediction, the minimum mean square error estimation is executed in computing a combination factor which gives weights to each contribution of the decoded information from the layers. In addition, the proposed JLP is integrated into the SHVC decoder to work as an error concealment solution to mitigate the error propagation happening inevitably in practical video transmission. Experiments have shown that the proposed SHVC framework significantly outperforms its relevant benchmarks, notably by up to 14.8% in bitrate reduction with respect to the standard SHVC codec. The proposed SHVC error concealment strategy also greatly improves the concealed picture quality as well as reducing the problem of error propagation when compared to conventional error concealment approaches.**

***Index Terms*—Compression efficiency, error concealment, joint layer prediction, MMSE, scalable video coding.**

I. Introduction

NOWADAYS, video contents can be stored or delivered in several formats in quality, frame rate, spatial resolution, chroma sub-sampling ratio, or bit-depth in order to cover a wide range of user's requirements associated most probably with available bandwidth and/or equipment resources like memory or computation as well as the quality requested by the user. Quite expectedly, encoding and delivering the video in such a way calls for scalable encoding scheme unless considerable increases of resources in both storage and bandwidth can be borne with. In addition, the increasing diversity of heterogeneous and dynamic transmission environments have expedited the development of a new scalable video coding standard to achieve a far more flexible, easy to implement solution over existing platforms but still having better compression performance than the prior standard of H.264/SVC [1]. It has led the experts group to make MPEG HEVC/ITU-T H.265 to further standardize the scalable extension of HEVC (known as SHVC) [2], [3].

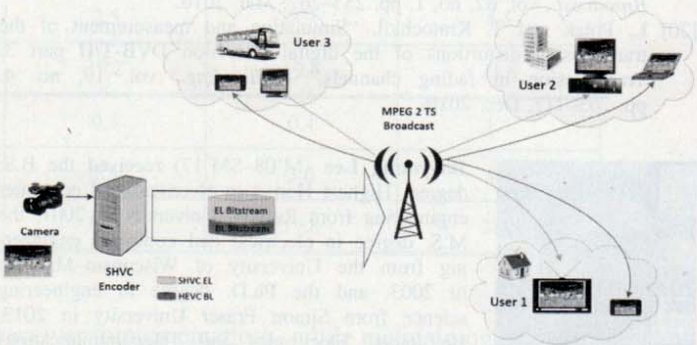

Fig. 1. An example of ATSC 3.0 broadcasting scenario.

Note that the ATSC 3.0 has considered several broadcasting scenarios, and identified the scalable video as an important functionality for combined services of HD/UHD and/or fixed/mobile services [4]–[7]. Fig. 1 shows an example of such broadcasting scenario in which a SHVC scalable compressed bitstream consisting of a base layer (BL) and an enhancement layer (EL) provides a scalable broadcasting service at the same time over different environments. The SHVC scalable bitstream can be broadcasted over the MPEG-2 TS (transport stream) [8] or through the Multi-input Multi-output (MIMO) broadcasting scheme [9]. The user can receive and decode only the base layer or both layers depending on its display, energy, and network capabilities.

SHVC employed the traditional layered coding approach again but in a slight different way. That is, different from H.264/SVC [1], the coding tools of existing HEVC (which is used as the BL) are kept intact and the scalable functionality is implemented by defining additional signaling capabilities at macroblock-level to indicate whether EL macroblock is predicted from BL or other EL layers. In SHVC, the reconstructed BL picture, taken as an inter-layer reference (ILR) picture, is put to the EL prediction buffer following the normal inter-layer processing [3], [10]. The reference index signaling mechanism in (non-scalable version of) HEVC still works perfectly to

Manuscript received March 31, 2018; revised July 10, 2018 and October 20, 2018; accepted October 28, 2018. Date of publication December 3, 2018; date of current version September 4, 2019. This work was supported by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant 102.01-2016.15. *(Corresponding author: Xiem HoangVan.)*

X. HoangVan is with the Faculty of Electronics and Telecommunications, VNU-University of Engineering and Technology, Hanoi 10000, Vietnam (e-mail: xiemhoang@vnu.edu.vn).

B. Jeon is with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, South Korea (e-mail: bjeon@skku.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2881355

Video Streaming Adaptation Strategy for Multiview Navigation Over DASH

Chao Yao, Jimin Xiao, *Member, IEEE*, Yao Zhao, *Senior Member, IEEE*, and Anlong Ming, *Member, IEEE*

***Abstract*—Video content delivery over Internet is receiving increasing attention from both industry and academia, especially for the multiview video contents, as it is the basis to support various applications, such as 3-D video, virtual reality, free view video, and so on. To cope with the dynamic nature of Internet throughput, dynamic adaptive streaming over HTTP (DASH) has been introduced to control the video streaming based on the network conditions. In this paper, we design a streaming framework to improve the user experience of the multiview video streaming over DASH, considering the user behavior of the viewpoint navigation during the streaming process. To eliminate the view switching delay, a multiple view navigation rule is introduced to pre-fetch the possible switching viewpoints. An optimal bitrate allocation scheme is proposed for the introduced rule, allowing the clients to maximize the video quality. Moreover, we found the video quality and the playback starvation probability are conflicting factors, while both are essential for the user's quality of experience (QoE). To tackle this issue, a QoE optimization solution is designed to maximize the overall performance in the proposed framework. Several experiments verify the effectiveness of the proposed framework, and the results demonstrate that the proposed framework outperforms two typical DASH methods.**

***Index Terms*—Multiview video, DASH, QoE, viewpoint navigation, rate adaption.**

I. Introduction

VIDEO streaming is gaining popularity due to the increasing Internet speed and the advancing cloud computing technologies. Users can request and play video content anywhere and at anytime as long as Internet service is available. Netflix, YouTube, Adobe OSFM and other video streaming services have been promoting adaptive streaming, which allows to deliver multiple versions of video contents (i.e., SD, HD and Ultra HD) in heterogeneous networks where network conditions are varying.

Among many adaptive video streaming techniques, the most popular one is the Dynamic Adaptive Streaming over HTTP (DASH). DASH is standardized by the ISO/IEC Moving Picture Experts Group (MPEG) [1], [2]. DASH is a client-driven streaming technology on top of TCP/HTTP without explicitly specifying the rate adaptation mechanism. In a classical DASH system, video contents are pre-encoded at different bitrates, which allows multiple stream representations stored in the server. One representation is divided into segments, each of which is corresponding to several seconds of video clip (such as 2*s*). A manifest file called Media Presentation Description (MPD) file is also provided to describe available profiles of each segment. Both video segments and MPD file are stored at the server side. After receiving the MPD file, the DASH client continuously selects suitable representations during a streaming session, depending on the buffer status, available bandwidth and other factors.

In recent years, evolved from single video streaming, multiview video streaming has emerged where the client is able to interactively switch viewpoints among the provided viewpoints from different cameras [3]–[5]. In Fig. 1, a multiview DASH system is depicted where multiple cameras are used to capture different viewpoints, and each viewpoint is encoded into multiple representations with different bitrate levels. For single view DASH system, client is able to switch different representations to meet varying networking conditions, whereas in the multiview DASH system the client is able to switch among viewpoints as well as representations. Several studies have been conducted to support the user's interactive viewpoint navigation for multiview video streaming systems [6]–[11]. However, how to provide users with seamless view switching capability as well as smooth and high playback quality is a challenging task which has not been solved yet.

In DASH systems, in order to have smooth playback, video segments are buffered before being displayed. Generally, the longer the buffer is, the less buffer starvation happens [12]. The video buffer mechanism has a major problem for multiview interactive video streaming. During the streaming process, users may freely switch to other viewpoints, and long view switching delay is generally not acceptable. However, the video segments of newly switched viewpoint may not be available among the buffered videos. To cope with this problem, one possible solution is to download and buffer video segments

Manuscript received April 16, 2018; revised July 19, 2018; accepted September 12, 2018. Date of publication October 3, 2018; date of current version September 4, 2019. This work was supported in part by the National Key Research and Development of China under Grant 2016YFB0800404, in part by the National Natural Science Foundation of China under Grant 61532005 and Grant 61501379, in part by the Jiangsu Science and Technology Programme under Grant BK20150375, and in part by the Specialized Fund for the Basic Research Operating expenses Program of Central College under Grant 2018JBZ001. *(Corresponding author: Anlong Ming.)*

C. Yao and A. Ming are with the Institute of Sensing Technology and Business, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: yaochao1986@gmail.com; mal@bupt.edu.cn).

J. Xiao is with the Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: jimin.xiao@xjtlu.edu.cn).

Y. Zhao is with the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: yzhao@bjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2871370

PeQASO: Perceptual Quality Assessment of Streamed Videos Using Optical Flow Features

Mohammed A. Aabed, *Member, IEEE*, and Ghassan AlRegib, *Senior Member, IEEE*

***Abstract*—In this paper, we introduce PeQASO, a perceptual quality assessment framework for streamed videos using optical flow features. This approach is a reduced-reference pixel-based and relies only on the deviation of the optical flow of the corrupted frames. This technique compares an optical flow descriptor from the received frame against the descriptor obtained from the anchor frame. This approach is suitable for videos with complex motion patterns. Our technique does not make any assumptions on the coding conditions, network loss patterns or error concealment techniques. In this paper, we consider both sources of artifacts and distortions in streaming, including compression artifacts. We validate our proposed metric by testing it on a variety of distorted sequences from three proposed and commonly utilized video quality assessment databases. Our results show that our metric estimates the perceptual quality at the sequence level accurately. We report the correlation coefficients with the differential mean opinion scores reported in the database. For compression artifacts, the results show Spearman's and Pearson's correlations of 0.96 and 0.94 for all the tested sequences, respectively. For channel-induced distortion, the results show Spearman's and Pearson's correlations of 0.88 and 0.89, respectively. For all other distortions, the average Spearman's and Pearson's correlations are 0.82 and 0.79, respectively.**

***Index Terms*—Video quality assessment, video quality monitoring, video coding, channel losses, optical flow, video distortion, video streaming.**

I. Introduction

OVER the past decade, video streaming services have gained massive popularity and both content and the number of users are continuously growing. The continuous growth of Internet traffic in general, and video traffic in particular, has triggered the signal processing and communication communities' concern with bandwidth and quality of experience (QoE). Global IP traffic has increased fivefold over the past five years. It is also predicted to reach 1.6 zettabytes per year by 2018 (threefold increase from 2013), out of which 79% will be video traffic (66% in 2013) [1], [2]. A million minutes of video content is estimated to cross the network every second by 2018 [2]. The Global and North America Mobile Data Traffic Forecast Update [1] released February 2014 estimates a growth of 45.7% in global mobile connections reaching 10.2 billion mobile connections. Mobile video traffic will account for over 69% of that total. Furthermore, busy-hour Internet traffic is growing more rapidly than average Internet traffic [2]. Thus, the standardization bodies are adapting to this growth by motivating technologies that increase the efficiency of bandwidth utilization, data compression and QoE.

High Efficiency Video Coding (HEVC) is the latest video coding standard approved by the JCT-VC [3]. It was also adopted by the telecommunication standardization sector of the ITU-T as its H.265 standard for video coding [4]. HEVC offers new coding features and tools to improve the compression gain of H.264/MPEG-4 AVC. In fact, it offers double the compression efficiency of AVC [3], [5], [6] for the same quality. However, the complexity of the coding operations in HEVC is much higher compared to its predecessor. The coding unit tree (CTU) structure introduced in HEVC facilitates more efficiency for coding, transform and prediction. HEVC also applies an open group of picture (GOP) format, which utilizes more inter-coded frames than AVC [7]. While this format facilitates higher compression gain, it imposes higher degree of data dependencies among frames. The GOP size in these configurations increases with frame rate. For a video encoded at 60 fps, the recommended GOP size is 64 frames. This, in turn, makes the bitstream, and consequently decoded video, more sensitive to errors.

To illustrate the sensitivity of HEVC videos to losses, we examined the quality effect of losing one packet corresponding to a single frame in a compressed bitstream. We encoded a test sequence using the HEVC test model (HM 16.2) [5] as per the recommendations in [7]. As per these configurations, every single frame is contained in one slice, which is encapsulated in a single Network Abstraction Layer (NAL) unit. For this experiment, we compressed BQMall with 60 fps and a GOP size of 64 frames. We then corrupted the first P-slice in the bitstream (frame 8), which is the second NAL unit in transmission and receiving order after the I-slice (frame 0). The decoder was configured to replace a lost frame with the temporally closest available frame in the decoding buffer. In this scenario, frame 8 (lost P-slice) will be replaced with frame 0 (I-slice). Fig. 1 shows the impact of losing this NAL unit on the objective quality of the lost frame and the resulting error propagation to the rest of the GOP.

This indicates that the effect of packet losses and errors on the perceptual quality is amplified in HEVC videos. This is

Manuscript received May 26, 2018; revised October 3, 2018; accepted October 11, 2018. Date of publication December 4, 2018; date of current version September 4, 2019. *(Corresponding author: Mohammed A. Aabed.)*

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: maabed@gatech.edu; alregib@gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2881363

Bitrate-Based No-Reference Video Quality Assessment Combining the Visual Perception of Video Contents

Juncai Yao and Guizhong Liu, *Member, IEEE*

***Abstract*—In video communication, the quality of video is mainly determined by bitrate in general. Moreover, the effect of video contents and their visual perception on video quality assessment (VQA) is often overlooked. However, in fact, for different videos, although the bitrates are the same, their VQA scores are still significantly different. Hence, it is assumed that the bitrate, video contents, and human visual characteristics mainly affect the VQA. Based on the above three aspects, in this paper, we designed a bitrate-based no-reference (NR) VQA metric combining the visual perception of video contents, namely, BRVPVC. In this metric, first an initial VQA model was proposed by only considering the bitrate alone. Then, the visual perception model for video contents was designed based on the texture complexity and local contrast of image, temporal information of video, and their visual perception features. Finally, two models were synthesized by adding certain weight coefficients into an overall VQA metric, namely, BRVPVC. Furthermore, ten reference videos and 150 distorted videos in the LIVE video database were used to test the metric. Moreover, based on the results of evaluating the videos in LIVE, VQEG, IRCCyN, EPFL-PoliMI, IVP, CSIQ, and Lisbon databases, the performance of BRVPVC is respectively compared with that of six full-reference (FR) metrics and ten NR VQA metrics. The results show that our VQA metric has a higher accuracy than six common FR VQA metrics and eight NR VQA metrics, and it is close to other two NR VQA metrics in accuracy. The corresponding values of Pearson linear correlation coefficient and Spearman rank order correlation coefficient reached 0.8547 and 0.8260, respectively. In addition, the computational complexity of proposed VQA metric is lower than video signal-to-noise ratio, video quality model, motion-based video integrity evaluation, spatiotemporal most apparent distortion, V-BLINDS, and V-CORNIA metrics. Moreover, the proposed metric has a better generalization property than these metrics.**

***Index Terms*—Video quality assessment, video contents, visual perception characteristics, temporal/spatial domain of video, correlation coefficient.**

Manuscript received May 20, 2018; revised August 1, 2018 and September 22, 2018; accepted October 10, 2018. Date of publication November 9, 2018; date of current version September 4, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61301237, and in part by the Natural Science Foundation of Shaanxi Province, China, under Grant 2015KJXX-42. *(Corresponding author: Guizhong Liu.)*

J. Yao is with the School of Electronic and Information Engineering, Xi'an Jiao Tong University, Xi'an 710049, China, and also with the School of Physics and Telecommunication Engineering, Shaanxi University Of Technology, Hanzhong 723000, China (e-mail: yaojc4782@163.com).

G. Liu is with the School of Electronic and Information Engineering, Xi'an Jiao Tong University, Xi'an 710049, China (e-mail: liugz@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2878360

I. Introduction

THE OBJECTIVE video quality assessment (VQA) plays a very important role in multimedia, network, and communication technologies. A good VQA metric not only evaluates their qualities automatically and accurately, but also monitors them in real time, guides the parameter setting, and optimizes the algorithms to better serve for video transmission. Hitherto, many VQA metrics have been proposed in the literature. However, due to the complexity of video contents, instability of network, different codecs, uncertainty in transmission conditions, and complexity of human visual system (HVS) characteristics, there is no excellent objective VQA metric at present that can accurately measure the quality of distorted video in real time. Therefore, it is essential to study a simple, effective, and practical VQA metric that satisfies the subjective perception to the extent possible [1]–[3].

Generally, VQA can be divided into three main categories: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) quality assessment [4]. So far, most of the proposed VQA models are FR and RR. Here, the typical models are peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [5], video signal-to-noise ratio (VSNR) [6], video quality model (VQM) [7], spatiotemporal most apparent distortion (ST-MAD) [8], and motion-based video integrity evaluation (MOVIE) [9]. The NR VQA does not require any source; its approach is divided further into two types: (1) NR visual type (NR-P), which works on fully decoded video frames, (2) NR coded type (NR-B), which uses the information extracted from the bit-stream [4].

A common method of constructing NR VQA model is to first extract the features of video distortion produced during the communication, then build the relationship between these features and video quality, and finally synthesize them to obtain an overall NR VQA model. Regarding precision, the best performance of VQA model is that the objective quality scores obtained using the model to evaluate videos are consistent with the subjective mean opinion scores (MOS). At this point, their correlation coefficients including the Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation coefficient (SROCC) [10] achieve a value of one. Because the source video is not needed in NR VQA, it has become a hot research topic in video communication [4], [10]. However, the precision of NR VQA model is still lower due to the following reasons:

Reduction of Padding Overhead for RLNC Media Distribution With Variable Size Packets

Maroua Taghouti, Daniel E. Lucani, *Senior Member, IEEE*, Juan A. Cabrera, Martin Reisslein, *Fellow, IEEE*, Morten Videbæk Pedersen, and Frank H. P. Fitzek

***Abstract*—Random linear network coding (RLNC) can enhance the reliability of multimedia transmissions over lossy communication channels. However, RLNC has been designed for equal size packets, while many efficient multimedia compression schemes, such as variable bitrate (VBR) video compression, produce unequal packet sizes. Padding the unequal packet sizes with zeros to the maximum packet size creates an overhead on the order of 20%–50% or more for typical VBR videos. Previous padding overhead reduction approaches have focused on packing the unequal packet sizes into fixed size packets, e.g., through packet bundling or chaining and fragmentation. We introduce an alternative padding reduction approach based on coding macro-symbols (MSs), whereby an MS is a fixed-sized part of a packet. In particular, we introduce a new class of RLNC, namely MS RLNC which conducts RLNC across columns of MSs, instead of the conventional RLNC across columns of complete packets of equal size. Judiciously arranging the source packets into columns of MSs, e.g., through shifting the source packets horizontally relative to each other, supports favorable MS RLNC coding properties. We specify the MS RLNC encoding and decoding mechanisms and analyze their complexity for a range of specific MS arrangement strategies within the class of MS RLNC. We conduct a comprehensive padding overhead evaluation encompassing both previous approaches of packing the unequal size packets into fixed size packets as well as the novel MS RLNC approaches with long VBR video frame size traces. We find that for small RLNC generation sizes that support low network transport delays, MS RLNC achieves the lowest padding overheads; while for large generation sizes, both the previous packing approaches and the novel MS RLNC approaches effectively reduce the padding overhead.**

***Index Terms*—Random linear network coding (RLNC), variable-sized packets, variable bitrate (VBR) video, video streaming, zero-padding.**

I. Introduction

A. Motivation: Padding Overhead in RLNC

MODERN media distribution scenarios involve a wide range of heterogeneous underlying communication channels, including lossy channels, e.g., wireless channels [4]–[8]. The effects of lossy channels on media distribution can be mitigated through error correction coding [9]–[14]. Recently, random linear network coding (RLNC) has emerged as a popular error correction coding scheme for a wide range of complex lossy networks. RLNC has been employed in a wide range of systems involved in media distribution, including data storage systems [15]–[20], multicast and broadcast distribution networks [21]–[25], peer-to-peer distribution networks [26]–[29], as well as general content distribution and streaming systems [30]–[34].

RLNC has been designed for equal size packets, see Section II-B. However, popular efficient multimedia compression schemes, in particular video compression schemes, are typically most efficient when operating in a variable bitrate (VBR) mode [35], [36]. Therefore, VBR video has received significant interest for media distribution [37]–[41]. Padding the unequal VBR video packets with zeros [42], [43], i.e., the so-called zero-padding to obtain equal size packets for RLNC encoding, creates a high overhead, as quantified in Section II.

B. Existing Padding Reduction Approaches

The padding overhead in RLNC for distribution of VBR encoded media has received very limited attention to date. Compta *et al.* [44] have proposed padding reduction approaches that seek to pack the unequal size packets into fixed size packets that are amenable to conventional RLNC. In particular, the simple bundling approach [44] is similar to the classical bin packing problem, i.e., strives to pack variable size objects (packets) into the least number of bins (packets) of the same capacity (fixed packet size). Bin packing is an NP-hard problem, i.e., has very high computational complexity. The chaining and fragmentation approach [44] strings the packets together back-to-back into one long string and then fragments

Manuscript received July 11, 2018; revised December 1, 2018; accepted December 22, 2018. Date of publication February 4, 2019; date of current version September 4, 2019. This work was supported in part by the German Research Foundation (DFG) through the CoSIP Project under Grant FI 1671/1-1, in part by the Collaborative Research Center 912 Highly Adaptive Energy-Efficient Computing, as well as by the Danish Council for Independent Research through the TuneSCode Project under Grant DFF - 1335-00125, and in part by the Aarhus Universitets Forskningsfond under Project AUFF-2017-FLS-7-1. Preliminary versions of Sections II and III appeared in [1]–[3]. *(Corresponding author: Martin Reisslein.)*

M. Taghouti is with the Tunisia Polytechnic School, University of Carthage, Carthage 1054, Tunisia, and also with the Deutsche Telekom Chair of Communication Networks, Technische Universität Dresden, 01062 Dresden, Germany (e-mail: maroua.taghouti@tu-dresden.de).

D. E. Lucani is with DIGIT Centre, Department of Engineering, Aarhus University, 8200 Aarhus, Denmark (e-mail: daniel.lucani@eng.au.dk).

J. A. Cabrera and F. H. P. Fitzek are with 5G Lab Germany, Deutsche Telekom Chair of Communication Networks, Technische Universität Dresden, 01062 Dresden, Germany (e-mail: juan.cabrera@tu-dresden.de; frank.fitzek@tu-dresden.de).

M. Reisslein is with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: reisslein@asu.edu).

M. V. Pedersen is with Steinwurf ApS, 9220 Aalborg, Denmark (e-mail: morten@steinwurf.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2019.2892594

Cache-Based Popular Services Pushing on High-Speed Train by Using Converged Broadcasting and Cellular Networks

Bing Li, Jian Xiong, *Member, IEEE*, Bo Liu, *Member, IEEE*, Lin Gui, *Member, IEEE*, Meikang Qiu, *Senior Member, IEEE*, and Zhiping Shi

***Abstract*—This paper presents a cache-based popular services pushing solution on high-speed train (HST) by using converged wireless broadcasting and cellular networks. Pushing and caching popular services on the HST to improve the capacity of the network is a very efficient way; and it can also bring a better user experience. The most popular services are transmitted and cached on the vehicle relay station of the train ahead the departure time in the proposed model. Then, the most popular services are broadcasted and cached on the User Equipment after all the passengers are on the train; the less popular services are delivered to the passengers by P2P mode through the relayed cellular network on the train. Specifically, we firstly use the dynamic programming algorithm to maximize the network capacity in limited pushing time, which can be converted to the 0-1 Knapsack problem. Furthermore, we propose three greedy algorithms to approximate the optimal solution on account of the high time complexity of dynamic programming when the input scale gets bigger. And simulation results show that the proposed popularity-based greedy algorithm performs well. Moreover, as the passengers may get on and off the HST when arriving at an intermediate station, a services rebroadcast algorithm is employed when more intermediate stations are considered. U-shaped distribution is adopted to indicate the number of passengers getting on and off the train. Simulations also show that the proposed rebroadcast algorithm can efficiently improve the capacity of the converged networks.**

***Index Terms*—Converged wireless broadcasting and cellular network (CWBCN), dynamic programming, high-speed train (HST), service pushing, vehicle relay station (VRS).**

Manuscript received May 7, 2018; revised June 29, 2018; accepted July 11, 2018. Date of publication August 14, 2018; date of current version September 4, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61671295, Grant 61471236, and Grant 61420106008, in part by the Shanghai Key Laboratory of Digital Media Processing, in part by the Shanghai Pujiang Program under Grant 16PJD029, in part by 111 Project under Grant B07022, and in part by the Science and Technology on Communication Networks Laboratory under Grant KX172600030. *(Corresponding author: Jian Xiong.)*

B. Li, J. Xiong, and L. Gui are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xjarrow@sjtu.edu.cn).

B. Liu is with the Department of Engineering, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: b.liu2@latrobe.edu.au).

M. Qiu is with the Department of Electrical Engineering, Columbia University, New York, NY 10027, USA (e-mail: mq2203@columbia.edu).

Z. Shi is with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: szp@uestc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2863102

I. Introduction

INCREASING popularity of smart terminals which support a wide variety of multimedia applications, e.g., smartphones, iPad, makes us come into the age of big data. Most of the data services, such as videos, live streaming, are transmitted through the cellular network. It is most likely that the pace of the existing wireless infrastructure cannot deal with the exponential growth of mobile data [1]. However, Cha *et al.* [2] point out that the multimedia contents that people are interested in submitting to long-tail distribution; that is, most people only focus on a few popular information. The long-tail rule [3] makes the hot content becoming more and more popular. Like the Matthew Effect, the cyclical effect makes good one become better. Li *et al.* [4] confirm the rule by mining data from some popular websites. Wang *et al.* [5] quote the Zipfian Law to clarify their converged network system model. Some other long tail distributions, such as Power-law and Pareto, can also describe the Rich-Get-Richer phenomenon [6]. Easley and Kleinberg [7] proves the three distributions are equivalent.

Offloading the popular services from the cellular network by heterogeneous or converged network is an efficient method, especially for video streams. Paper [8] shows the coexistence possibility of HbbTV applications and ISDB-T broadcasting. Paper [9] researches on the cooperation media transmission by both cellular network and broadcasting network. In paper [10], dynamic adaptive streaming over HTTP (DASH) format is used for both broadband and broadcast delivery; and demonstration shows the proposed ROUTE-based approach is a powerful and lean media delivery method for both streaming media and nonreal time media. Domínguez *et al.* [11] provide details on a large pilot deployment for a multi-device hybrid broadcast-Internet service for a live TV programme. A novel delay receiving approach is proposed to solve the heterogenous receiving problem with the constrain of reception bandwidth; the method needs less buffer space [12]. Moreover, more and more people are focusing their research interests on the data offloading, e.g., local cache offloading, D2D offloading, broadcasting/multicast and WiFi network offloading [13]–[17]. Furthermore, the Third Generation Partnership Project (3GPP) firstly defined the multimedia broadcast/multicast services (MBMS) in Rel-6 during 2005. The MBMS which sends the same content to different terminals at the same time makes the P2M (point-to-multipoint) transmission mechanism come true.

A Study on Dual-Polarized MIMO-ICI Canceller With Complexity Reduction Under Mobile Reception of OFDM Signals

Akira Nakamura, *Member, IEEE*, Hiroaki Otsubo, *Member, IEEE*, and Makoto Itami, *Member, IEEE*

***Abstract*—In the next Japanese digital terrestrial television broadcasting systems, the scheme that combines orthogonal frequency division multiplexing (OFDM) and dual-polarized multi-input multi-output (MIMO) is proposed. One of the major problems to be considered is inter-carrier interference (ICI) that is generated by Doppler-spread under mobile reception of MIMO-OFDM systems. In this case, reception characteristics are deteriorated significantly. The ICI canceller for MIMO-OFDM can demodulate transmitted symbols. To realize the MIMO-ICI canceller based on zero-forcing (ZF), it is necessary to reduce the complex calculations such as matrix operations. Therefore, complexity reduction of the MIMO-ICI canceller is proposed. The proposed MIMO-ICI canceller with complexity reduction based on ZF can improve the influence of ICI with low complexity. However, reception characteristics are deteriorated as compared to the MIMO-ICI canceller based on ZF. Thus, improving scheme of MIMO-ICI canceller with complexity reduction based on ZF is proposed. In the proposed scheme, the iterative detection is adopted. In addition, MIMO-ICI cancellers based on minimum mean square error (MMSE) are proposed. As the results of computer simulations, the MIMO-ICI cancellers using iterative detection based on ZF and MMSE can improve the reception characteristics with low complexity.**

***Index Terms*—OFDM, ICI, dual-polarized MIMO, complexity reduction, iterative detection.**

I. Introduction

RECENTLY, UHDTV (Ultra-High Definition Television) broadcasting systems are widely researched and developed [1]–[3]. In the USA, ATSC3.0 (Advanced Television Systems Committee standards 3.0) has been standardized to broadcast high-resolution contents [4]. Also, UHDTV terrestrial broadcasting that adopts ATSC3.0 standard has been started in South Korea, May 2017 [4]. In Europe, DVB-T2 (Digital Video Broadcasting-Terrestrial 2) that is next terrestrial broadcasting system of DVB-T has been standardized [5].

Manuscript received May 28, 2018; revised September 11, 2018; accepted November 8, 2018. Date of publication December 5, 2018; date of current version September 4, 2019. This work was supported in part by the Ministry of Internal Affairs and Communications, Japan, through the Program Research and Development for Advanced Digital Terrestrial TV Broadcasting System and in part by JSPS KAKENHI under Grant 15K18069. *(Corresponding author: Akira Nakamura.)*

The authors are with the Department of Applied Electronics, Tokyo University of Science, Tokyo 125-8585, Japan (e-mail: nakamura@te.noda.tus.ac.jp; tsubo@itmailab.te.noda.tus.ac.jp; itami@te.noda.tus.ac.jp).

Digital Object Identifier 10.1109/TBC.2018.2882784

In Japan, 8K Super Hi-Vision (SHV) terrestrial and satellite broadcasting systems are researched and developed [6]–[9]. The test 8k satellite broadcasting is already started in August 2016, Japan [10]. In the next-generation terrestrial broadcasting, the combination of dual-polarized MIMO (Multi-Input Multi-Output) and ultra-multilevel OFDM technologies to improve the transmission capacity has been proposed [6]. In addition, NU-QAM (Non-Uniform QAM) as sub-carriers modulation is adopted to improve required CNR [7].

In the mobile reception of OFDM signal, each path is affected by Doppler-shift and received under the multi-path channel. The orthogonality between sub-carriers is distorted by Doppler-shift and ICI (Inter-Carrier Interference) is generated. The reception characteristics are deteriorated by ICI. Therefore, it is necessary to compensate the influence of ICI. The ICI canceller can improve reception characteristics. However, ICI canceller has complexity because the inverse matrix calculation which dimension is the number of sub-carriers is required. Thus, the complexity reduction scheme is proposed [11]. The ICI canceller with complexity reduction can compensate the influence of ICI. However, reception characteristics of ICI canceller with complexity reduction are deteriorated as compared to ICI canceller. The improving scheme of ICI canceller with complexity reduction is proposed [12]. In the proposed scheme, the iterative detection is adopted to remove interference. The ICI canceller with iterative detection can obtain reception characteristics close to ICI canceller with low-complexity. In addition, it is necessary to consider an ICI mitigation scheme for MIMO-OFDM system. When ICI mitigation schemes for SISO/MIMO-OFDM system are operated, complexity becomes a serious problem. Thus, ICI mitigation scheme with low-complexity has been proposed [13]–[15].

In the mobile reception of dual-polarized MIMO-OFDM signal under the multi-path channel, horizontally and vertically waves of OFDM signal are affected by Doppler-shift and ICI is generated. However, the influence of ICI is not analyzed under the case that dual-polarized MIMO-OFDM is operated. Therefore, it is necessary to make a channel model of dual-polarized MIMO-OFDM and compensate the influence of ICI.

In this paper, a channel model of dual-polarized MIMO-OFDM is proposed. In addition, the compensation schemes of dual-polarized MIMO-OFDM under the mobile reception are proposed. In the proposed channel model, polarization rotations and Doppler-shift of each path are considered.

Receiver Design for Alamouti Coded FBMC System in Highly Frequency Selective Channels

Jun Li, Da Chen, Daiming Qu, Yue Zhang, and Tao Jiang

***Abstract*—Alamouti coded filter bank multicarrier systems suffer from severe performance degradation in highly frequency selective channels. To address the above problem, we propose a novel Alamouti decoding receiver based on frequency spreading/despreading, where the Alamouti decoding module is moved to the front of the frequency despreading module. Unlike the traditional subcarrier-based Alamouti decoding scheme, the Alamouti decoding in the proposed receiver could be performed on each frequency point to better combat the frequency selectivity. Moreover, the full diversity is achieved on each frequency point by introducing a phase factor before Alamouti decoding. Simulation results demonstrate the performance advantages of our proposed receiver over the traditional scheme in highly frequency selective channels.**

***Index Terms*—FBMC, Alamouti, receiver design, frequency selectivity.**

I. Introduction

BECAUSE of the robustness against multipath channels, orthogonal frequency division multiplexing (OFDM) has been adopted in many broadcasting systems [1], [2]. To obtain transmit diversity with low implementation complexity, OFDM combined with the Alamouti code [3] is a common means, such as in DVB-second generation terrestrial (DVB-T2) standard [4], [5]. However, the use of cyclic prefix (CP) in OFDM results in a spectral efficiency loss. With the emergence of new data services and explosive growth in wireless devices [6]–[10], a higher spectral efficiency modulation technique is desired for future broadcasting systems. Recently, the filter bank multicarrier (FBMC) modulation is attracting increasing attention and has been considered as an alternative to OFDM [11]–[15]. Compared with OFDM, FBMC has lower spectral sidelobes due to the well-designed prototype filter [16], [17]. Moreover, FBMC systems have higher spectral efficiency since they do not require CP. The above advantages have made FBMC a transmission technique candidate for next-generation broadcasting systems [18]. Hence, the integration of Alamouti code with FBMC is expected to boost the performance of next-generation broadcasting systems.

However, the integration of Alamouti code with FBMC is not as straightforward as that with OFDM. Unlike OFDM, the orthogonality condition of FBMC only holds in real domain, which leads to the intrinsic interference. The intrinsic interference is pure imaginary under an idea channel, which can be eliminated by taking the real part of the demodulated signal at the receiver. Nevertheless, in presence of multipath fading, the intrinsic interference turns into complex-valued form, which destroys the real orthogonality of FBMC. When the channel frequency response is flat on each subcarrier pass band, the single-tap equalizer can be used and the interference can be eliminated after taking the real part of the equalized signal. However, when the channel frequency response is not flat on each subcarrier, the inter-symbol interference and inter-carrier interference in FBMC cannot be substantially mitigated with simple signal processing techniques [19].

The presence of the intrinsic interference prevents the direct combination of FBMC with Alamouti code. Some works attempted to deal with the intrinsic interference by introducing CP. Hao *et al.* [20] proposed a pseudo-Alamouti scheme with CP to avoid the intrinsic interference. CP was also employed in the FBMC based singer-carrier frequency division multiple access transmission for the Alamouti coding scheme [21]. Besides, the Alamouti schemes were also applied to CP-FBMC system [22] and CP-FFT-FBMC system [23]. However, these schemes require quite a lot of additional resources and may destroy the superior spectral confinement of FBMC. In addition, iterative interference cancellation algorithms were investigated in [24] and [25], which require accurate estimation of interference and make the receiver complicated.

Recently, block-wise Alamouti schemes [26]–[30] were proposed for FBMC. In these schemes, the transmitted data were coded in time or frequency reversal block structure in the transmitter. The complex orthogonality could be achieved by designing the transmitted data in the block. Hence, the traditional Alamouti decoding could be employed after FBMC demodulation to decode the data in the receiver. The block-wise Alamouti schemes are easy to implement and do not destroy the structure of FBMC signal. However, severe

Manuscript received June 27, 2018; revised September 6, 2018; accepted September 18, 2018. Date of publication October 24, 2018; date of current version September 4, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61601191, Grant 61571200, and Grant 61701186, in part by the Major Program of National Natural Science Foundation of Hubei, China, under Grant 2016CFA009, in part by the Fundamental Research Funds for the Central Universities under Grant 2015ZDTD012, in part by the State Key Program of National Natural Science Foundation of China under Grant 61631015, in part by the Open Research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2014D09, and in part by EU Horizon 2020 Project loRL Funding under Grant 761992. *(Corresponding author: Da Chen.)*

J. Li, D. Chen, D. Qu, and T. Jiang are with the Wuhan National Laboratory for Optoelectronics, School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: lijun2013@hust.edu.cn; chenda@hust.edu.cn; qudaiming@hust.edu.cn; tao.jiang@ieee.org).

Y. Zhang is with the Department of Engineering, University of Leicester, Leicester LE1 7RH, U.K. (e-mail: yue.Zhang@Leicester.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2874547

Extremely High Frequency (EHF) Bands for Future Broadcast Satellite Services: Opportunities and Challenges

Claudio Sacchi, *Senior Member, IEEE*, Tommaso Rossi, Maurizio Murroni, *Senior Member, IEEE*, and Marina Ruggieri, *Fellow, IEEE*

***Abstract*—The exploitation of the bandwidth portions in the millimeter wave domain, namely extremely high frequencies (EHFs) will open new perspectives to future satellite services, in particular for high quality TV broadcasting and multimedia content delivery. The present work aims at surveying the opportunities and challenges of EHF exploitation for broadband satellite applications in the broadcast framework. Theoretical capacity evaluation confirms that EHF satellite links can offer unprecedented data-rates, clearly superior than that actually provided by Ku and Ka-bands. However, some crucial issues are still to be solved, mainly in terms of mitigation of tropospheric propagation impairments and appropriate waveform design. Link and capacity analysis will be the basis for the investigation of future application scenarios of broadcast EHF satcoms, namely high-definition TV (HDTV) and ultra-HDTV, with particular emphasis to 4K and 8K video formats.**

***Index Terms*—Next generation broadcast sat systems and standards, future technologies and services for broadcasting, extremely high frequency (EHF), UHDTV.**

I. Introduction

SATELLITE communication is crucial to support the broadcasting industry; currently, satellite broadcast digital services are available worldwide and represent an important market for the broadcasting stakeholders in near future [1]. In the last decade, the standards for satellite broadcasting have been developed beside terrestrial ones deploying the current cutting-edge technologies [2]–[4]. Furthermore, the upcoming 5G scenario is focusing on the converged satellite and terrestrial networks [5]. One of the main appeal of the satellite broadcast technology is the availability of large amount of dedicated bandwidth if compared to terrestrial services. However, the increased demand of Ultra-High Definition TV (UHDTV) is raising new challenges in term of resource allocations also in the satellite market where 4K and 8K resolutions have been already tested and are being rolled out by Japan's public broadcaster (NHK) in 2018, in preparation for the Olympics Games 2020 in Tokyo. In this case, the 8K UHDTV (4320p), called Super Hi-Vision [6], offers a frame resolution of 7680 × 4320 (33.2 megapixels), but it also offers a richer sound experience with 22.2 multi-channel sound, which has been developed by NHK Science & Technical Research Laboratories. In this context, the bandwidth spaces in the Extremely High Frequency (EHF) portion of the electromagnetic spectrum (namely: 30-300 GHz), known also as millimeter Wave (mmWave) bands, will represent, in the opinion of many researchers, "the new broadband frontier" for satellite communications [7]. Such claim is supported by the long history of satellite communications that shows an incremental shift to higher carrier frequencies in order to find additional bandwidth resources. The exploitation of Ka-band (18-30 GHz), dated since mid '90s, followed the exploitation of Ku-band (11-18 GHz), started in 1977 with the launch of Sirio satellite [8]. Nowadays, Ku-band is mainly reserved for satellite TV broadcasting, while Ka-band is used for satellite internetworking and other bandwidth-demanding services.

The further step in the exploration of novel bandwidth spaces concerns with the EHF portions beyond the Ka-band, namely: Q/V band (about 40-75 GHz) and W-band (75-110 GHz). The current allocation (ITU – Region 1) of mmWave bands for satellite fixed and broadcasting services is shown in Tab. I:

The pioneering feasibility studies about EHF for satellite communications have been carried out in early 2000s in the framework of DAVID [9] and WAVE [10] experiments. The focus of these activities – where Italian Space Agency (ASI), University of Rome Tor Vergata and Politecnico di Milano, together with industrial actors and other academic research entities, involved in synergy – was mainly addressed to the W-band utilization. In parallel, European Space Agency (ESA) and ASI supported the Alphasat experiment, under which the "Aldo Paraboni" payload was set up and launched in

Manuscript received September 4, 2018; revised December 31, 2018; accepted January 3, 2019. Date of publication February 5, 2019; date of current version September 4, 2019. A preliminary version of this paper has been published at the 12th IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB'17), Cagliari, Italy, June 7-9, 2017. *(Corresponding author: Claudio Sacchi.)*

C. Sacchi is with the Department of Information Engineering and Computer Science, University of Trento, 38150 Trento, Italy, and also with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni, 43124 Parma, Italy (e-mail: claudio.sacchi@unitn.it).

T. Rossi and M. Ruggieri are with the Department of Electronic Engineering, University of Rome "Tor Vergata," 00133 Rome, Italy, and also with the Interdepartmental Center for TeleInfrastructures, 00133 Rome, Italy (e-mail: tommaso.rossi@uniroma2.it; ruggieri@uniroma2.it).

M. Murroni is with the Department of Electrical and Electronic Engineering, University of Cagliari, 09123 Cagliari, Italy, and also with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni, 43124 Parma, Italy (e-mail: murroni@diee.unica.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2019.2892655

High-Level Multiple-UAV Cinematography Tools for Covering Outdoor Events

Ioannis Mademlis, Vasileios Mygdalis, Nikos Nikolaidis, Maurizio Montagnuolo, Fulvio Negro, Alberto Messina, and Ioannis Pitas

***Abstract*—Camera-equipped unmanned aerial vehicles (UAVs), or "drones," are a recent addition to standard audiovisual shooting technologies. As drone cinematography is expected to further revolutionize media production, this paper presents an overview of the state-of-the-art in this area, along with a brief review of current commercial UAV technologies and legal restrictions on their deployment. A novel taxonomy of UAV cinematography visual building blocks, in the context of filming outdoor events where targets (e.g., athletes) must be actively followed, is additionally proposed. Such a taxonomy is necessary for progress in intelligent/autonomous UAV shooting, which has the potential of addressing current technology challenges. Subsequently, the concepts and advantages inherent in multiple-UAV cinematography are introduced. The core of multiple-UAV cinematography consists in identifying different combinations of multiple single-UAV camera motion types, assembled in meaningful sequences. Finally, based on the defined UAV/camera motion types, tools for managing a partially autonomous, multiple-UAV fleet from the director's point of view are presented. Although the overall focus is on cinematic coverage of sports events, the majority of our contributions also apply in different scenarios, such as movies/TV production, newsgathering, or advertising.**

***Index Terms*—UAV cinematography, media production, drone swarm, shot types, intelligent shooting.**

I. Introduction

UNMANNED Aerial Vehicles (UAVs, or "drones") are a recent addition to the cinematographer's arsenal. By exploiting their ability to fly and/or hover, their small size and their agility, impressive video footage can be obtained that otherwise would have been impossible to acquire. Although UAV cinematography is expected to revolutionize A/V shooting, as Steadicam did back in the seventies [1], the topic has not yet been heavily researched and shooting is currently performed on a more or less ad-hoc basis. Employing drones in video production and broadcasting opens up numerous opportunities for new forms of content, enhanced viewer engagement and interactivity. It immensely facilitates flexibility in shot set up, while it provides the potential to adapt the shooting so as to cope with changing circumstances in wide area events. Additionally, the formation of dynamic panoramas or novel, multiview and 360-degree shots becomes easier.

However, at the current technology level, several challenges arise. Minimal battery life/flight time (typically, less than 25 minutes), limited payload and disturbing sound issues, as well as safety considerations, are all factors interacting with cinematography planning. Legal requirements in most countries, such as a demand for direct line-of-sight between the pilot and the UAV at all times, or restrictions on flight above human crowds, complicate shooting. As a result, strict manual mission planning and designation of tight flight corridors burden the director and limit the creative potential. In case multiple UAVs are employed, synchronization and collision avoidance issues, as well as a need for each UAV to avoid entering the field-of-view of the others (*FoV avoidance*), further perplex the situation. Many of the above problems can be alleviated by automating the technical/non-creative aspects of UAV filming, through intelligent/autonomous UAV shooting software. However, the related research is still in its early stages and is plagued by a lack of standardization in UAV cinematography, which severely limits the available creative possibilities.

Following early preliminary work [2]–[8], this paper is an overview of the current situation in UAV cinematography, with an emphasis on filming non-scripted, outdoor events where targets (e.g., athletes) must be actively followed. This is deemed as the most complex application, due to highly unpredictable circumstances and a typically large area to be covered. First, we present the state-of-the-art on the use of drones in media production and broadcasting, along with a brief review of relevant UAV technologies and legal restrictions on their deployment. Up to now, commercial autonomous drone shooting applications are simplistic and mostly support a single UAV. As a first step towards remedying this situation, a newly developed UAV shot type taxonomy is proposed, emphasizing the capture of moving targets. The limitations of single-UAV shooting and the advantages of employing a fleet of multiple drones are discussed from a media production perspective. Finally, based on the developed UAV shot type taxonomy, tools for managing a multiple-UAV fleet from the director's point of view are presented. Thus, we attempt to streamline the process of creating/editing/tracking UAV shooting missions, while exploiting unplanned/opportunistic shooting potentials. In the context of this paper, a UAV shooting mission consists in a meaningfully assembled sequence of the proposed multiple-UAV/camera motion types. Multiple-UAV cinematography is further detailed through example scenarios.

II. State-of-the-Art

Despite a number of legal restrictions, more and more live events in sport and entertainment are produced with the support of UAVs. Drone footage adds value to the filming of concerts and music festivals, as it gives viewers a realistic view of the size and dynamics of the event, without having to use highly expensive equipment such as helicopters or cranes.

A. Drones in Media Production

The use of UAVs in movie productions became common in the past years, as an efficient way to save time and money. Drones are

Manuscript received July 19, 2018; revised December 13, 2018; accepted January 3, 2019. Date of publication January 31, 2019; date of current version September 4, 2019. This work was supported by the European Union's Research and Innovation Programme Horizon 2020 (MULTIDRONE) under Grant 731667. *(Corresponding author: Ioannis Mademlis.)*

I. Mademlis, V. Mygdalis, N. Nikolaidis, and I. Pitas are with the Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece (e-mail: imademlis@aiia.csd.auth.gr).

M. Montagnuolo, F. Negro, and A. Messina are with the Centre for Research and Technological Innovation, Radiotelevisione Italiana, 10138 Turin, Italy.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2019.2892585

Implementation Methodologies of Deep Learning-Based Signal Detection for Conventional MIMO Transmitters

Myung-Sun Baek, *Member, IEEE*, Sangwoon Kwak, Jun-Young Jung, Heung Mook Kim, *Member, IEEE*, and Dong-Joon Choi

***Abstract*—In this paper, simple methodologies of deep learning application to conventional multiple-input multiple-output (MIMO) communication systems are presented. The deep learning technologies with deep neural network (DNN) structure, emerging technologies in various engineering areas, have been actively investigated in the field of communication engineering as well. In the physical layer of conventional communication systems, there are practical challenges of application of DNN: calculating complex number in DNN and designing proper DNN structure for a specific communication system model. This paper proposes and verifies simple solutions for the difficulty. First, we apply a basic DNN structure for signal detection of one-tap MIMO channel. Second, convolutional neural network (CNN) and recurrent neural network (RNN) structures are presented for MIMO system with multipath fading channel. Our DNN structure for one-tap MIMO channel can achieve the optimal maximum likelihood detection performance, and furthermore, our CNN and RNN structures for multipath fading channel can detect the transmitted signal properly.**

***Index Terms*—Deep learning, DNN, CNN, RNN, communication systems, MIMO, signal detection.**

I. Introduction

DEEP learning technology has been received considerable attention in wide engineering areas [1]–[3]. Especially, many deep learning researches on communication engineering area are actively investigated in recent years [4]–[6]. In the latest communication and broadcasting systems, multiple-input multiple-output (MIMO) signal processing is regarded as a promising solution to enhance the spectral efficiency [6]–[9]. Recently, MIMO transmission based on spatial multiplexing has been actively considered for various broadcasting systems [10]–[14], and some broadcasting systems include MIMO transmission techniques in their specifications [15], [16]. In MIMO system based on spatial multiplexing, interference cancellation and signal detection are the most important processes to increase spectral efficiency and ensure the system performance. Many MIMO signal detection techniques have been investigated by numerous researchers [17]–[23]. Although the maximum likelihood (ML) detection technique can give the optimal performance, the utilization of the ML technique is practically difficult because of its high computational complexity. Therefore, most researches have considered sub-optimal detection techniques which are practically implementable such as ordered successive interference cancellation (OSIC), decision feedback equalization (DFE), and so on [21]–[23]. To detect the transmitted MIMO signal, the traditional MIMO detection techniques require estimation of various communication parameters such as channel state information (CSI), signal-to-noise ratio (SNR) and so on. Performance of MIMO communication system can be degraded by not only the sub-optimal detection techniques but also imperfect communication parameter estimation. Therefore, the existing algorithm-based MIMO signal detection techniques have two weak points which are sub-optimality and imperfect estimation of communication parameters.

In single carrier communication systems, estimation and equalization processes of multi-path fading channel require complicated computation. It is hard to apply MIMO technique to the single carrier system because it needs the processes. For the reason, to apply MIMO signal processing in practical multi-path fading channel, MIMO-orthogonal frequency division multiplexing (OFDM) has been investigated because of its convenience of signal processing [14], [24], [25].

This paper proposes simple implementation methodologies of deep learning-based MIMO signal detection for conventional MIMO transmitter. The supervised classification-based deep learning technology is considered. Traditional MIMO detection techniques operate to estimate precise transmission symbols [26], so it needs complicated estimation process at each detection step. However, the proposed deep learning-based detection tries to build best detection model based on trained deep learning parameters, so it needs only simple and basic computations through the trained model. Furthermore, since the designed method requires only transmitted and received signal without any additional communication parameters for supervised learning, the performance degradation based on imperfect channel estimation can be prevented. In our investigation, two MIMO channel environments are considered.

First, dense layer-based deep neural network (DNN) is designed for single-path MIMO communication channels. The detection performance of the DNN structure is similar to that of the optimal ML detection in the single-path channel environment. Second, various deep learning structures are designed for multi-path MIMO channel environment. As mentioned, the multi-path channel is knotty problem in the single carrier MIMO system. To tackle this issue, we design convolutional neural network (CNN)-based deep learning structure [27], [28] and apply it to the multi-path MIMO channel. Furthermore, recurrent neural network (RNN) [29], [30] is also considered to enhance the performance. The proposed RNN-based detection can nearly achieve the performance of the conventional ML detection in single-path MIMO channel. Furthermore, in this paper, detailed implementation methodology is presented from input signal processing to classification. Therefore, this paper can help utilization of the deep learning techniques for communication systems.

The rest of this paper is organized as follows. Section II presents the MIMO system model with conventional algorithm-based MIMO detection. Section III describes the proposed DNN structure for single-path MIMO channel. Next, CNN and RNN-based deep learning structures are presented in Section IV. Numerical Results are

Manuscript received September 6, 2018; revised November 22, 2018; accepted December 28, 2018. Date of publication January 22, 2019; date of current version September 4, 2019. This work was supported by Electronics and Telecommunications Research Institute funded by the Korean Government (Development of Machine Learning Based Multiple Signal Communication Technology in Single Channel) under Grant 19ZR1500. *(Corresponding author: Myung-Sun Baek.)*

The authors are with the Media Transmission Research Group, Electronics and Telecommunications Research Institute, Daejeon 305-350, South Korea (e-mail: sabman@etri.re.kr; s.kwak@etri.re.kr; jungjy@etri.re.kr; hmkim@etri.re.kr; djchoi@etri.re.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2019.2891051

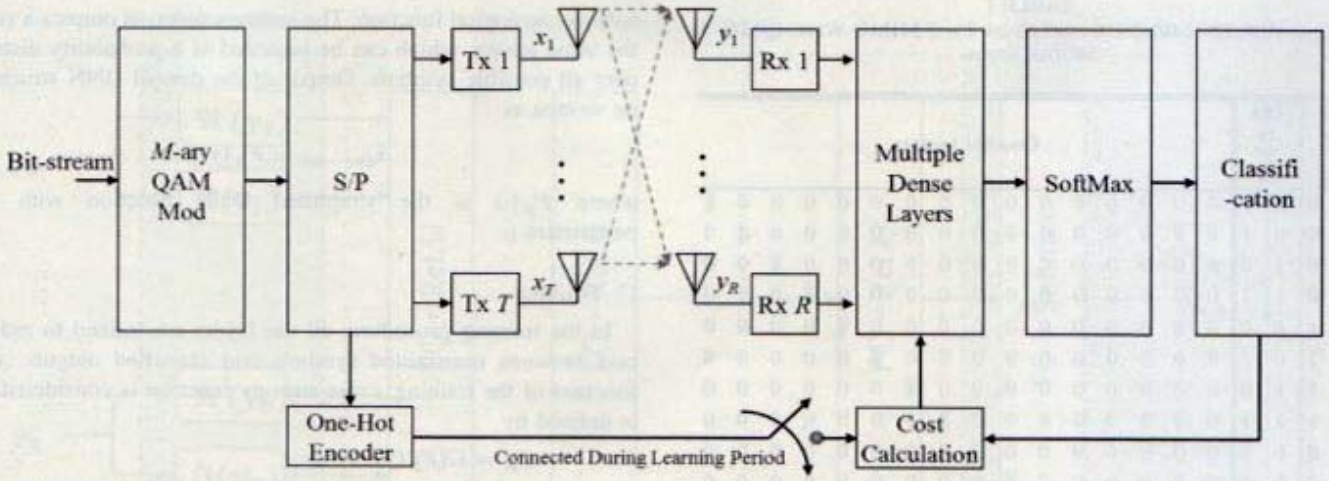

Fig. 1. MIMO communication system with deep learning-based signal detector.

reported in Section V, followed by some concluding remarks in Section VI.

II. MIMO System Model and Conventional Detection

In MIMO transmitter with spatial multiplexing, multiple symbols are transmitted simultaneously. The transmitted symbols of T transmission antennas can be written as

$$\mathbf{x} = [x_1 \; x_2 \; x_3 \; \cdots \; x_T], \tag{1}$$

where x_t is t-th transmission symbol which is M-ary modulated. The received signal in the MIMO receiver with R reception antennas is as follows:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_R \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} \cdots & h_{1,T} \\ h_{2,1} & h_{2,2} \cdots & h_{2,T} \\ \vdots & \vdots \; \ddots & \vdots \\ h_{R,1} & h_{R,2} \cdots & h_{R,T} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_R \end{bmatrix} \tag{2}$$

where $h_{r,t}$ is impulse response function of channel between t-th transmission antenna and r-th reception antenna, and w_r is additive white Gaussian noise (AWGN) in r-th reception antenna.

Under the assumption of CSI estimated by the receiver, conventional detection algorithms are executed. In the ML detection, all available combinations of transmission symbols are compared, and the likelihood test is calculated by

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\hat{\mathbf{x}}} \left\| \mathbf{y} - \tilde{\mathbf{H}} \cdot \hat{\mathbf{x}} \right\|, \tag{3}$$

where $\tilde{\mathbf{H}}$ is estimated channel by receivers. The ML detection selects the final symbol combination by checking the smallest distance metric. The ML detection is theoretically optimal way to recover the transmitted symbols, but it is not feasible for real time implementation because of its high computational complexity [17], [21], and also it requires CSI estimation process.

One of representative and commonly considered sub-optimal detection technique is OSIC. In the OSIC detector, received signal vector $\mathbf{y}$ is multiplied by filter matrix $\mathbf{G}$ which is the Moore-Penrose pseudo-inverse denoted by $(\cdot)^{\dagger}$ of channel matrix. The minimum mean square error (MMSE) Moore-Penrose pseudo-inverse matrix is defined as

$$\mathbf{G}_{\mathrm{MMSE}} = \tilde{\mathbf{H}}^{\dagger} = \left(\tilde{\mathbf{H}}^{H} \cdot \tilde{\mathbf{H}} + \sigma_n^2 \cdot \boldsymbol{I}_T \right)^{-1} \cdot \tilde{\mathbf{H}}^{H}, \tag{4}$$

where σ_n^2 is noise variance and $\boldsymbol{I}_T$ is $T \times T$ identity matrix. Let us assume that t-th symbol yields the smallest estimation error, i.e., the largest SNR after linear nulling of the interference. Then it can be concluded that the symbol is associated with the row $\mathbf{g}^{(t)}$ of $\mathbf{G}$ that has the minimum Euclidean norm since the vector brings out the smallest noise enhancement. Hence, during the first step of the algorithm, only the decision static can be written as

$$\hat{x}_t = \mathbf{g}^{(t)}\mathbf{y} = \mathbf{g}^{(t)}(\mathbf{H}\mathbf{x} + \mathbf{w}) = x_t + \eta_t, \tag{5}$$

where the effective noise $\eta_t = \mathbf{g}^{(t)} \cdot \mathbf{w}$ is used to find an estimated $\hat{x}_t$ for transmitted x_t. The interference caused by this signal is then subtracted from the received signal vector $\mathbf{y}$ and the t-th column is removed from the channel matrix, leading to a new system with only T-1 transmission antennas. This procedure consisting of nulling and cancelling is repeated for the reduced system until all signals are detected [21]. Note that this sub-optimal detection requires CSI, noise variance estimation, and SNR estimation.

III. Dense Layer-Based DNN Structure for Single-Path MIMO Channel

This section describes the implementation method of deep learning structure for received MIMO signal detection. In the proposed MIMO detection with DNN structure, the transmitted symbols are considered as one-hot encoding vector with length of M^T which is the possible symbol combinations where T transmission antennas and M-ary QAM are considered. Table I shows an example of one-hot encoding for 2 × 2 MIMO with QPSK modulation. In the traditional communication system, the one-hot encoding is not used for baseband modulation, and ot encoding is not required anymore. ared with transmitted signal with and close. rases, via gatthe traditional MIMO transmitter is considered in this research. Therefore, the one-hot encoding is used during only learning period.

Fig. 1 shows the MIMO communication system with DNN-based signal detector. The modulated signal is transmitted by the conventional MIMO transmitter, and received through MIMO channel, so the received MIMO signal is the same as (2). As shown in Fig. 1, the received signal $\mathbf{y}$ is inserted to the multiple dense layers. The dense layers are updated by learning process. The learning process should be conducted before the actual detection operation, so we assume that enough number of transmitted and received symbol sets to be learned are given in advance. For learning period, the output of Softmax function is compared with one-hot encoded transmitted signal as Fig. 1, because the output of Softmax is featured as one-hot

TABLE I
ONE-HOT ENCODING EXAMPLE FOR 2 × 2 MIMO WITH QPSK MODULATION

Bits for 1st Tx		Bits for 2nd Tx		One-hot vectors															
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

encoding. In the training procedure, hyper-parameters such as learning rate, batch size, or epoch are quit important to optimize. Although the major point of this paper is a methodology of implementation but not detail of it, some examples of parameter settings for performance evaluation will be described in Section IV.

Fig. 2 depicts the detailed DNN architecture for MIMO signal detection. It consists of a separation layer, multiple fully connected (FC) layers and a softmax classification layer. Each FC layer is followed by the rectified linear unit (ReLU) activation function [31]. The ReLU is the most general activation function which can relax the vanishing gradient problem. Following sections explain more details.

A. Data Transform

In common DNNs, it is difficult to handle complex numbers of input data because all weights and biases are real value. However, the output of MIMO channel is complex data and has both positive numbers and negative numbers. To solve the problems, we separate real and imaginary parts of complex number as shown in Fig. 2. In Fig. 2, $\Re(y_r)$ and $\Im(y_r)$ are the real and imaginary parts of the r-th received signal, respectively. The separation computation can be written as

$$\mathrm{Re}_r = \Re(y_r) \tag{6}$$

$$\mathrm{Im}_r = \Im(y_r), \tag{7}$$

where Re_r and Im_r are the r-th separated input of real and imaginary part, respectively.

B. Dense Layers and a Classification Output Layer

In the dense layers, the separated and normalized input data is processed by multiple FC layers and ReLU activation function. In the training process, parameters of theses FC layers are trained to recover the transmitted symbols. Although the layers do not actually cancel the interference as conventional receivers, it can be trained to have equivalent function with that.

For the classification, the softmax activation function is utilized, which is representative classification function for DNN. The last FC layer outputs a vector of length M^T which is the length of one-hot vector of transmitted signal [4], and the vector is inserted to the softmax activation function. The softmax function outputs a vector of the same length, which can be regarded as a probability distribution over all possible symbols. Output of the overall DNN structure can be written as

$$\hat{\mathbf{x}}_{one-hot} = F_\mu(\mathbf{y}), \tag{8}$$

where $F_\mu(\cdot)$ is the simplified DNN function with trained parameters μ.

C. Training

In the training procedure, all the layers are trained to reduce the cost between transmitted symbols and classified outputs. As cost function of the training, cross entropy function is considered, which is defined by

$$\begin{aligned} C_\mu &= C\big(F_\mu(\mathbf{y}), \mathbf{x}\big) \\ &= \mathbf{x} \cdot \log\big(F_\mu(\mathbf{y})\big) - (1-\mathbf{x}) \cdot \log\big(1 - F_\mu(\mathbf{y})\big). \end{aligned} \tag{9}$$

The cost based on cross entropy is minimized by Adam algorithm which can support large data set and high-dimensional parameter spaces as well as adaptive learning rate [32].

IV. DEEP LEARNING STRUCTURE DESIGNS FOR MULTI-PATH MIMO CHANNEL

In this section, a multi-path MIMO channel environment is considered. Since each channel has multiple paths, the simple matrix multiplication of (2) should be changed to a linear convolution. The received signal of the r-th reception antenna can be written as

$$y_r = \sum_{t=1}^{T} h_{rt} * x_t + w_r, \tag{10}$$

where $h_{rt} = [h_{rt}(0)h_{rt}(1)\cdots h_{rt}(L-1)]$ is the multi-path channel between t-th transmission antenna and r-th reception antenna. In this environment, the transmitted symbols interfere continuously with the following symbols. For the reason, interference cancellation and signal detection of considered MIMO system become more challenging. To solve the issue, we firstly consider a CNN architecture as Fig. 3. In the CNN, the same data transform as Section III-A is also considered, and we similarly assume that the neural network is trained in advance using enough number of training data. As shown in Fig. 3, the input data sets are arranged as the table with $K \times 2R$ size, and filtering computations are operated through N depths. Since the transmitted symbols are received by interference between several symbols due to the fading channel, the multiple received symbols should be considered simultaneously to detect one transmitted symbol. In the proposed CNN-based MIMO detector structure, $K \times R$ received symbols are utilized at a time to detect the transmitted symbols, and 2×2 filter with stride 1 and no padding is considered to solve the multi-path channel effect. The filter with stride 1 can process multiple symbols continuously, so the transmitted symbols can be detected with multiple symbols based on the filtering. Pooling layer is not applied because the considered matrix is not large, and all convolutional layers have the same depth with N. After the convolution operations, ReLU activation function is applied like previous DNN and the classification layer is also the same with Section III-B.

Since the transmitted symbols continuously interfere by a static multi-path MIMO channel, it can be approximately modeled as a recurrent architecture. From this point of view, we also consider an RNN architecture which can recurrently learn a relationship between times series data [29]. Since the received signal from multi-path MIMO channel can be seen as a consecutive time series data, the RNN was properly leaned in this case. Fig. 4 describes the RNN

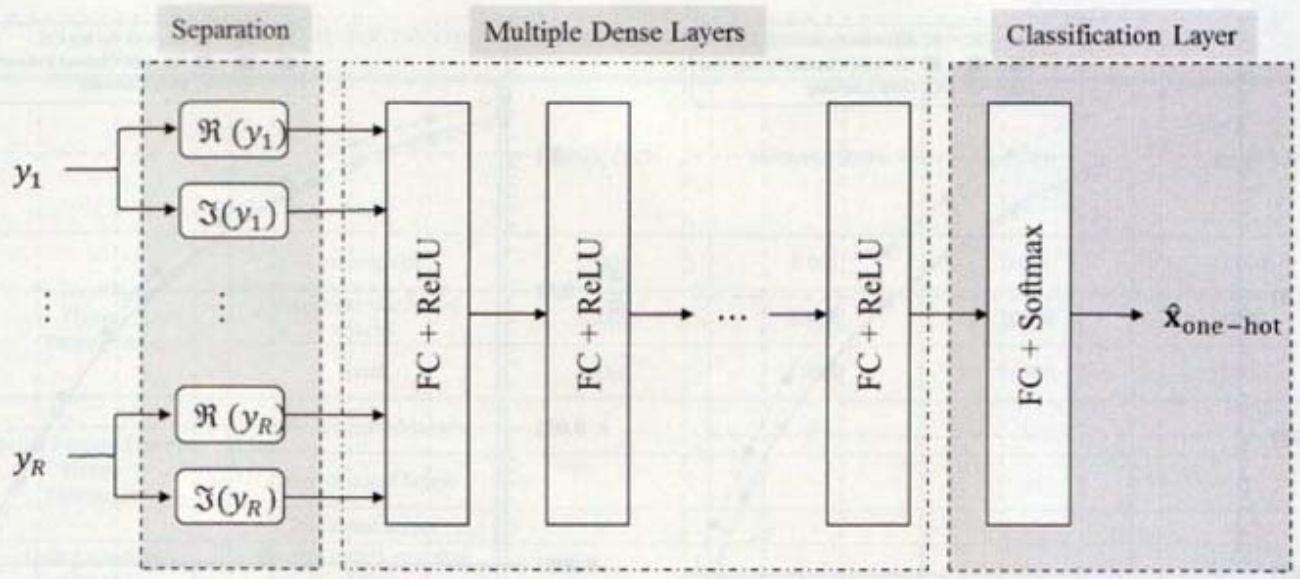

Fig. 2. DNN architecture for MIMO signal detection.

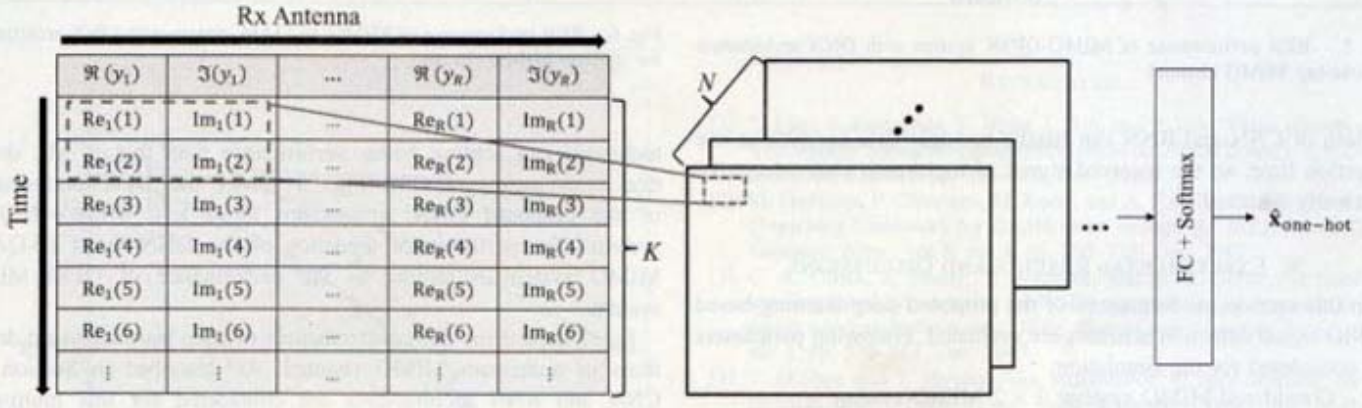

Fig. 3. CNN architecture for received MIMO signal detection.

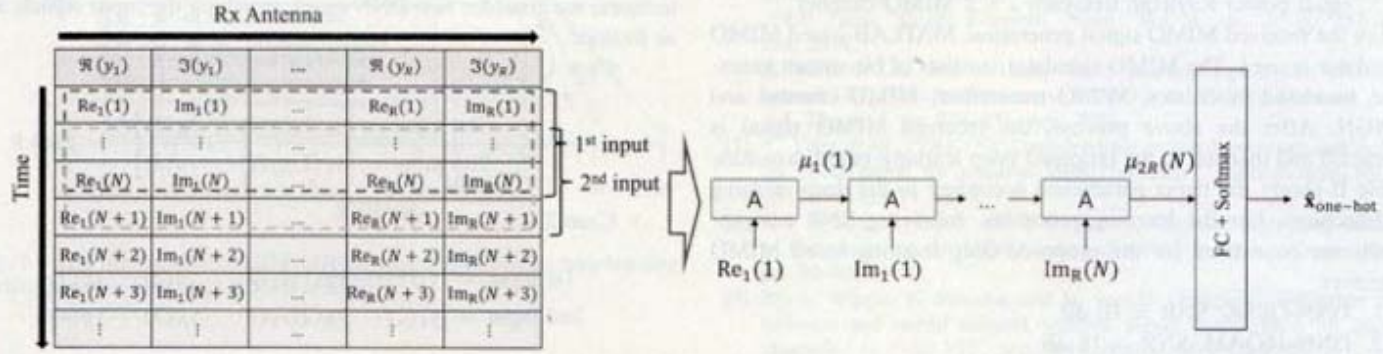

Fig. 4. RNN architecture for received MIMO signal detection.

architecture. As shown in the figure, to handle the consecutive time series data, the received symbol sequence is arranged into $K \times 2R$ matrix in chronological order. In a fading channel with L paths, one transmitted symbol is continuously received over L received symbols. Therefore, at each detection step the RNN architecture should consider at least $L \times R$ received symbols simultaneously to detect the transmitted T symbols. In the proposed detection architecture, since the first $N \times 2R$ size subset is considered at one detection step, consecutive $N \times R$ symbols are used to detect first transmitted T symbols. If $L \leq N$, the proposed architecture can detect the transmitted symbols properly because overall symbols from fading channel can be considered at a time. In the following steps, successive $N \times 2R$ size subsets are extracted and used to detect every following T symbols. Therefore, the extraction filter moves by one row at each time, and the RNN architecture outputs T estimated symbols. Therefore, the extraction filter moves by one row at each time, and the RNN architecture outputs T estimated symbols. There are variations of RNN structure such as long short-term memory (LSTM) and gated recurrent unit (GRU) [33]. In this work, LSTM is considered which can prevent the vanishing or exploding gradient problems based on a gate cell structure. Each cell makes decisions on what to store, and when to allow reads, writes and erases by opening and closing the gates. Therefore, the LSTM can give reliable MIMO signal detection performance.

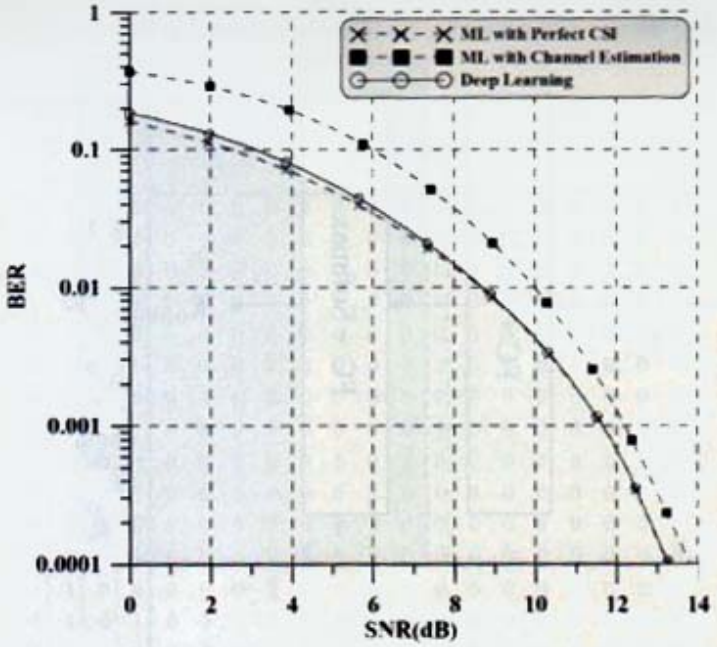

Fig. 5. BER performance of MIMO-QPSK system with DNN architecture for one-tap MIMO channel.

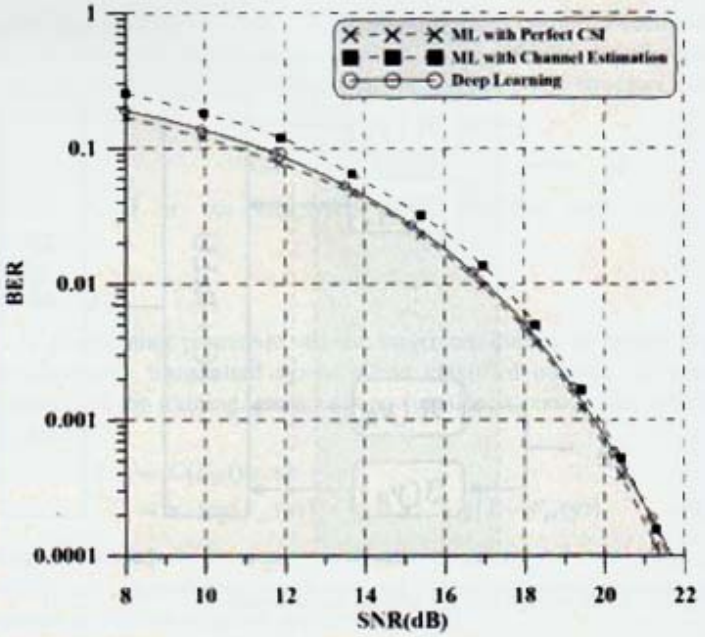

Fig. 6. BER performance of MIMO-16-QAM system with DNN architecture for one-tap MIMO channel.

Both of CNN and RNN can handle multiple data samples at one detection time, so the received signal of multi-path channel can be efficiently detected.

V. Experimental Results and Discussions

In this section, performances of the proposed deep learning-based MIMO signal detection schemes are evaluated. Following parameters are considered for the simulation:

- Considered MIMO system: 2 × 2 MIMO system
- Baseband modulation: QPSK and 16QAM
- Channel model: Rayleigh single-path 2 × 2 MIMO channel and equal power Rayleigh tree-path 2 × 2 MIMO channel

For the received MIMO signal generation, MATLAB-based MIMO simulator is used. The MIMO simulator consists of bit-stream generator, baseband modulator, MIMO transmitter, MIMO channel and AWGN. After the above process, the received MIMO signal is extracted and inserted to the proposed deep learning-based detectors. Table II shows the hyper-parameters according to the deep learning architectures. For the learning processes, following SNR environments are considered for the proposed deep learning-based MIMO detectors:

1. DNN-QPSK: SNR = 10 dB
2. DNN-16QAM: SNR = 16 dB
3. CNN-QPSK: SNR = 10 dB
4. RNN-QPSK: SNR = 10 dB.

In our test, if any SNR with BER $< 10^{-2}$ is chosen as training SNR region, the learning processes of all the considered deep learning architectures can give reasonable performance.

For comparison, the conventional ML detection performances with both perfect CSI and estimated CSI are inserted. The MMSE MIMO channel estimation technique is used for the channel estimation [35], [36].

Fig. 5 describes the 2 × 2 MIMO-QPSK system performance with proposed DNN architecture in single-path MIMO channel. For comparison, performances of ML detection schemes with both perfect CSI and estimated CSI are added. The performance of proposed DNN architecture-based deep learning is very similar to that of ML detection with perfect CSI. And the proposed technique can achieve better performance than that of ML detection with estimated CSI. Fig. 6 shows the BER performance of the proposed DNN architecture for 2 × 2 MIMO-16-QAM system. The performance tendency of the DNN-based 16-QAM-MIMO system is similar to the performance of QPSK-MIMO system.

Fig. 7 depicts the BER performances of deep learning-based detections in multi-path MIMO channel. As described in Section IV, CNN and RNN architectures are considered for this multi-path MIMO channel environment. Furthermore, for comparison, DNN architecture-based detection schemes are inserted. For the DNN architectures, we consider two DNN cases, in which the input signals are as follows:

- Case 1

$$\text{1st input} = \left[y_1(1) y_2(1) \cdots y_R(1)\right]$$
$$\text{2nd input} = \left[y_1(2) y_2(2) \cdots y_R(2)\right] \cdots$$

- Case 2

$$\text{1st input} = \left[y_1(1) \cdots y_R(1) y_1(2) \cdots y_R(2) \cdots y_R(3)\right]$$
$$\text{2nd input} = \left[y_1(2) \cdots y_R(2) y_1(3) \cdots y_R(3) \cdots y_R(4)\right] \cdots$$

where $y_r(l)$ is l-th received signal of r-th reception antenna. In case 1, the $2R$ received signals are considered for transmitted symbol detection as described in Fig. 2. Furthermore, in case 2, $3 \times 2R$ received signals are considered, in which the multi-path output symbols are considered for one detection time. As same as Fig. 5, performances of ML detection techniques in one-tap channel are added for comparisons. As expected, the BER performance of DNN architecture of case 1 is the worst among the BER performances. And the BER performance of RNN-based deep learning detection has best performance since the RNN can estimate the transmitted signal based on the relationship between the received symbols. In the low SNR, both CNN and RNN architectures can obtain similar performance of ML detection in single-path MIMO channel. Therefore, the CNN and RNN-based deep learning detection architectures can be efficiently used for the MIMO system in multi-path channel.

TABLE II
HYPER-PARAMETERS OF PROPOSED DEEP LEARNING ARCHITECTURES

		QPSK DNN	16-QAM DNN	CNN	RNN
Training Hyper-Parameters	Learning rate	0.001	0.001	0.001	0.0001
	Maximum number of epochs	20000	20000	20000	20000
	Batch	10000	10000	10000	10000
Spatial Feature Learning Hyper-Parameters	Fully connected layers	4	8	1	1
	Convolutional layers	-	-	3	-
	Recurrent layers	-	-	-	3
Data Extraction Hyper-Parameters	Overall matrix row size (K)	-	-	5	10000 (=batch size)
	Subset matrix row size (N)	-	-	-	3

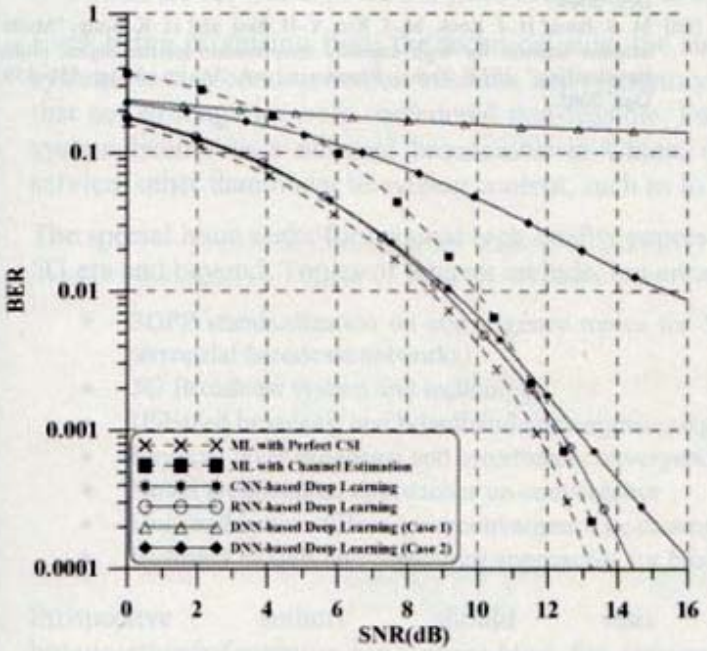

Fig. 7. BER performance of MIMO-QPSK system with various deep learning architectures for multi-path MIMO channel.

VI. CONCLUSION

In this paper, the deep learning-based MIMO signal detection techniques for conventional MIMO transmitters are proposed. The DNN based detection can achieve the performance of ML detection with perfect CSI, without estimation process of channel and SNR. Furthermore, for multi-path MIMO environment, CNN and RNN-based detection schemes are proposed. For low SNR region, the performances of both schemes are close to that of the ML detection with perfect CSI for one-tap channel.

ACKNOWLEDGMENT

The authors would like to thank Prof. Sung Kim of HKUST for his valuable online lectures.

REFERENCES

[1] Y. Liao, S. Kodagoda, Y. Wang, L. Shi, and Y. Liu, "Place classification with a graph regularized deep neural network," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 4, pp. 304–315, Dec. 2017.

[2] M. Gadaleta, F. Chiariotti, M. Rossi, and A. Zanella, "D-DASH: A deep Q-learning framework for DASH video streaming," *IEEE Trans. Cogn. Commun. Netw.*, vol 3, no. 4, pp. 703–718, Dec. 2017.

[3] C. A. Oroza, Z. Zhang, T. Watteyne, and S. D. Glaser, "A machine-learning-based connectivity model for complex terrain large-scale low-power wireless deployments," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 576–584, Dec. 2017.

[4] 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.

[5] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 257–265, Jun. 2018.

[6] K. Kim, J. Lee, and J. Choi, "Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO system," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 828–831, Apr. 2018.

[7] Y. Yang *et al.*, "DECCO: Deep-learning enabled coverage and capacity optimization for massive MIMO systems," *IEEE Access*, vol. 6, pp. 23361–23371, 2018.

[8] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Physical layer deep learning of encodings for the MIMO fading channel," in *Proc. Allerton*, 2017, pp. 76–80.

[9] M. A. Wijaya, K. Fukawa, and H. Suzuki, "Intercell-interference cancellation and neural network transmit power optimization for MIMO channels," in *Proc. VTC*, Sep. 2015, Boston, MA, USA, pp. 1–5.

[10] D. Gómez-Barquero *et al.*, "MIMO for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 298–305, Mar. 2016.

[11] T. Shitomi, E. Garro, K. Murayama, and D. Gomez-Barquero, "MIMO scattered pilot performance and optimization for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 188–200, Jun. 2018.

[12] Z. Wu, X. Gao, and C. Jiang, "Nonbinary LDPC-coded spatial multiplexing for rate-2 MIMO of DVB-NGH system," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 201–210, Jun. 2018.

[13] J.-S. Han, J.-S. Baek, and J.-S. Seo, "MIMO-OFDM transceivers with dual-polarized division multiplexing and diversity for multimedia broadcasting services," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 174–182, Mar. 2013.

[14] J. Hao, J. Wang, and C. Pan, "Low complexity ICI mitigation for MIMO-OFDM in time-varying channels," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 32–47, Sep. 2016.

[15] *Digital Video Broadcasting, Next Generation Broadcasting System to Handheld, Physical Layer Specification (DVB-NGH)*, document A160, DVB, Frankfurt, Germany, Nov. 2012.

[16] *Physical Layer Protocol*, ATSC Standard A/322, Jun. 6, 2017.
[17] M.-S. Baek, Y.-H. You, and H.-K. Song, "Combined QRD-M and DFE detection technique for simple and efficient signal detection in MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1632–1638, Apr. 2009.
[18] B. Gong, L. Gui, Q. Qin, and X. Ren, "Compressive sensing-based detector design for SM-OFDM massive MIMO high speed train systems," *IEEE Trans. Broadcast.*, vol. 63, no. 4, pp. 714–726, Dec. 2017.
[19] B. Kang, J.-H. Yoon, and J. Park, "Low-complexity massive MIMO detectors based on Richardson method," *ETRI J.*, vol. 39, no. 3, pp. 326–335, Jun. 2017.
[20] H.-S. Wang, F.-B. Ueng, Y.-K. Chang, "Novel turbo receiver for MU-MIMO SC-FMDA system," *ETRI J.*, vol. 40, no. 3, pp. 309–317, Apr. 2018.
[21] M.-S. Baek, S.-Y. Yeo, Y.-H. You, and H.-K. Song, "Low complexity ML detection technique for V-BLAST systems with DFE decoding," *IEICE Trans. Commun.*, vol. E90-B, no. 5, pp. 1261–1265, May 2007.
[22] R. P. F. Hoefel, "IEEE 802.11: On performance with MMSE and OSIC spatial division multiplexing transceivers," in *Proc. ISWCS*, Aug. 2012, pp. 376–380.
[23] B. Dhivagar, K. Kuchi, and K. Giridhar, "An iterative DFE receiver for MIMO SC-FDMA uplink," *IEEE Commun. Lett.*, vol. 18, no. 12, pp. 2141–2144, Dec. 2014.
[24] B.-S. Kim *et al.*, "16-QAM OFDM-based *K*-band LoS MIMO communication system with alignment mismatch compensation," *ETRI J.*, vol. 39, no. 4, pp. 535–545, Aug. 2017.
[25] A. Ladaycia, A. Mokraoui, K. Abed-Meraim, and A. Belouchrani, "Performance bounds analysis for semi-blind channel estimation in MIMO-OFDM communications systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5925–5938, Sep. 2017.
[26] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. SPAWC*, Jul. 2017, pp. 1–5.
[27] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8395053
[28] B. Lim, B. Yang, and H. Kim, "Real-time lightweight CNN for detecting road object of various size," in *Proc. MIPR*, Miami, FL, USA, Mar. 2018, pp. 202–203.
[29] H. Okamoto, T. Nishio, M. Morikura, and K. Yamamoto, "Recurrent neural network-based received signal strength estimation using depth images for mmWave communications," in *Proc. CCNC*, Jan. 2018, pp. 1–2.
[30] J.-S. Choi, W.-H. Lee, J.-H. Lee, J.-H. Lee, and S.-C. Kim, "Deep learning based NLOS identification with commodity WLAN devices," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3295–3303, Apr. 2018.
[31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Bolzmann machines," in *Proc. ICML*, 2010, pp. 807–814.
[32] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR2015*, May 2015, pp. 1–18.
[33] S. Kumar, L. Hussain, S. Banajee, and M. Reza, "Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster," in *Proc. EAIT*, Jan. 2018, pp. 1–4.
[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* vol. 9, no. 8, pp. 1735–1780, 1997.
[35] M.-S. Baek, M.-J. Kim, Y.-H. You, and H.-K. Song, "Semi-blind channel estimation and PAR reduction for MIMO-OFDM system with multiple antennas," *IEEE Trans. Broadcast.*, vol. 50, no. 4, pp. 414–424, Dec. 2004.
[36] M.-S. Baek, H.-J. Kook, M.-J. Kin, Y.-H. You, and H.-K. Song, "Multi-antenna scheme for high capacity transmission in the digital audio broadcasting," *IEEE Trans. Broadcast.*, vol. 51, no. 4, pp. 551–559, Dec. 2005.