ASVTuw: Adaptive Scalable Video Transmission in Underwater Acoustic Multicast Networks

Zhuoran Qi
Rutgers University
New Brunswick, NJ, USA
zhuoran.qi@rutgers.edu

Roberto Petroccia
NATO STO CMRE
La Spezia, Italy
roberto.petroccia@cmre.nato.int

Dario Pompili
Rutgers University
New Brunswick, NJ, USA
pompili@rutgers.edu

ABSTRACT

Scalable Video Coding (SVC) has been widely used in video transmissions. However, inappropriate SVC structures may lead to received video quality lower than user's requirement or resource waste, especially in underwater time-varying channels. In this work, an adaptive cross-layering solution is proposed and validated for video transmissions in underwater acoustic multicast networks, namely Adaptive Scalable Video Transmission (ASVTuw). In ASVTuw, the transmitter collects over time the information about the channel states and the users' video quality requirements to adaptively select the SVC video structures and transmission schemes, using Machine Learning (ML). At-sea experiments were conducted to collect the required acoustic data. The collected data were then used in MATLAB simulations to validate the ASVTuw. The results show that the usage of ASVTuw avoids resource wasting from transmitting redundant SVC substreams and satisfies the multicast users' video quality requirements effectively with higher flexibility compared with the existing noncross-layering designs.

KEYWORDS

Underwater acoustic communication, video transmission, at-sea experiment, scalable video coding, cross-layer protocol

ACM Reference Format:
Zhuoran Qi, Roberto Petroccia, and Dario Pompili. 2022. ASVTuw: Adaptive Scalable Video Transmission in Underwater Acoustic Multicast Networks. In *The 16th International Conference on Underwater Networks & Systems (WUWNet'22), November 14–16, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3567600.3568137

WUWNet'22, November 14–16, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9952-4/22/11…$15.00
https://doi.org/10.1145/3567600.3568137

1 INTRODUCTION

Overview: The application of underwater video multicast has been widely used in military, commercial, and scientific fields for efficiency in the highly bandwidth-limited environment (so as to avoid duplicated transmissions) [2]. When unmanned platforms are collecting/transmitting images during mine countermeasure activities or in any distributed surveillance task, scenarios need to be envisioned with multiple vehicles and different qualities to coordinate via efficient communication. Traditional wireless radio frequency communication attenuates dramatically in seawater, and wired communication limits future development of underwater systems, such as the range of vehicles and the coordination of multiple vehicles. In comparison, the Underwater Acoustic Communication (UAC) suffers less absorption and can reach a distance of up to kilometers [13], which is the best choice to use the scarce resources available. However, the sound wave speed in the water is as slow as 1500 m/s, leading to time-varying multipath delay. Other challenges introduced by the UAC include the time-varying channels caused by the dynamic water wave and limited bandwidth [10]. To realize effective and efficient underwater acoustic video multicast, an adaptive video transmission solution is in demand.

Existing Works: Authors in [6] introduce a method of Modulation and Coding Scheme (MCS) selection in the Long Term Evolution (LTE) system and a selection table can be generated by mapping the channel quality to the MCS. However, different from the LTE system where wireless communications are in free space, UAC will not work effectively with a stationary selection table since the underwater channel varies over time. To select the MCS adaptively in the time-varying fading channel, authors in [8] develop a decision tree capable of selecting the fastest data rate from the modulation selection table depending on Channel State Information (CSI) as inputs, including detailed values of Signal-to-Noise Ratio (SNR), multipath delay lifetime and Doppler frequency shift. In [12], we propose in-network coordination utilizing the Scalable Video Coding (SVC) in underwater acoustic networks and prove SVC's efficiency and flexibility with Autonomous Underwater Vehicles (AUVs) in a multicast manner. The traditional SVC [9] is receiver-driven and works by transmitting fixed SVC video substreams and reconstructing

the substreams adaptively at the receiver. However, the underwater environment is time-varying and when the channel becomes harsh, high enhancement SVC layer substreams in a fixed SVC video are likely to be discarded.

Motivation: SVC is a viable solution to achieve video transmissions in multicast networks in the presence of limited communication bandwidth. SVC slices the video into several layers while providing scalability in time, space and quality. The transmitted video can be reconstructed at the receiver by composing the base layer with the optional enhancement layers. However, random bit errors in SVC video introduce distortion and reduce the video quality. There are two challenges to design a reliable underwater SVC video transmission system: (i) A fixed physical-layer transmission scheme cannot balance the system robustness and throughput in a time-varying channel. (ii) High enhancement layers in a fixed SVC structure are inclined to be discarded due to users' low video quality requirements, which is a waste of resources. To address these issues, we aim at (i) selecting physical-layer transmission schemes adaptively according to the CSI, and (ii) avoiding transmitting redundant enhancement layers while satisfying the video quality requirements.

Contributions: In this work, we propose and validate an adaptive video transmission strategy for underwater acoustic multicast networks, called Adaptive Scalable Video Transmission (ASVTuw), to adapt to the underwater acoustic channel dynamics and obtain the required video quality. The contributions are summarized as follows.

- According to our knowledge, our proposed strategy is the first to concern resource waste due to discarded SVC video substreams in underwater time-varying channels.
- We establish a relationship between the Bit Error Rate (BER) and SVC video quality metrics.
- The proposed strategy considers the Equal Error Protection (EEP) and Unequal Error Protection (UEP) for SVC video transmissions to allocate resources in an adaptive way to achieve an efficient data rate.
- The ASVTuw works not only for unicast scenarios but for multicast as well.
- At-sea experiments for validating our proposed ASVTuw in underwater acoustic multicast networks are conducted with software-defined acoustic networking using the NATO Science and Technology Organization (STO) Centre for Maritime Research and Experimentation (CMRE) Littoral Ocean Observatory Network (LOON) testbed [3] from August 2020 to June 2021, which is located in the Gulf of La Spezia, Italy.
- The collected experimental data are processed by MATLAB and then employed to evaluate the performance of the proposed ASVTuw over time. The results show that the proposed strategy effectively reduces resource waste and meets the quality requirements of multiple users with high flexibility.

Paper Organization: Section 2 describes the proposed ASVTuw. The results of experiments and simulations are reported in Section 3. Finally, Section 4 draws the conclusions and discusses future works.

2 PROPOSED SOLUTION

In this section, we describe the proposed ASVTuw, whose model is shown in Figure 1. The ASVTuw selects the MCS and the SVC video encoding method at the transmitter based on the receiver feedback, including the estimated CSI and the video quality requirements. As shown in Figure 2, the cross-layering design is composed of adaptive SVC selection and adaptive MCS selection. At the transmitter, the CSI and the video quality requirement are provided by the receiver to drive the selection of the MCS and the SVC encoding method. At the receiver, the CSI is computed and the bitstreams are demodulated. The SVC decoding method is selected according to the required video quality. With the scalability of the SVC system, the ASVTuw can be used to accommodate the request of a single receiver (unicast) or multiple receivers with different video quality requirements (multicast).

Adaptive SVC Selection: To meet the video quality requirement while reducing the waste of resources resulting from the transmissions of unneeded redundant bits or the transmissions of enhancement layers that will be discarded at the receiver, the SVC encoding method is selected adaptively at the transmitter. To ensure the quality of the received video, a BER threshold is set by the cross-layering strategy to limit the distortion caused by random error bits. The reason why the video structure is determined by the transmitter instead of the receiver is that when moving from the unicast to the multicast scenario with one transmitter and multiple receivers, different receivers may have different video quality requirements. Therefore, the transmitter needs to know the requirements of all the receivers and customize the video structure accordingly.

Adaptive MCS Selection: In ASVTuw, MCS selection is performed by Machine Learning (ML). The training set is composed of N_s training samples $\{F^n, C^n\}_{n=1}^{N_s}$, where F^n is the feature set and C^n is the class label of n-th training sample. The features of the training sample contain the CSI and the BER thresholds. The class label is the MCS that achieves the highest effective data rate while controlling the BER under the threshold. In this work, the ML model is trained with a simulated CSI dataset and applied to the real world. Since different scenarios lead to different CSI, we simulate a rich set of possible scenarios exploring different values of SNR, multipath delay lifetime, and Doppler frequency shift,

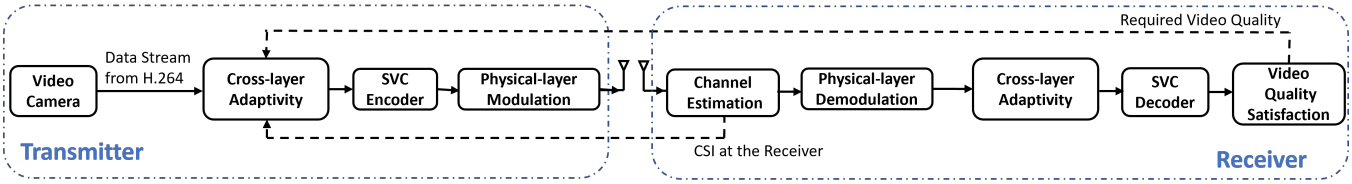
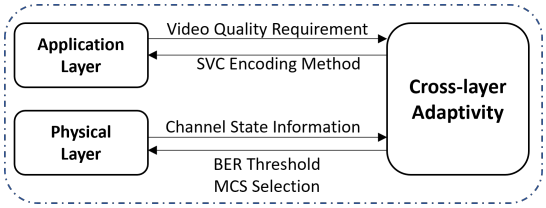
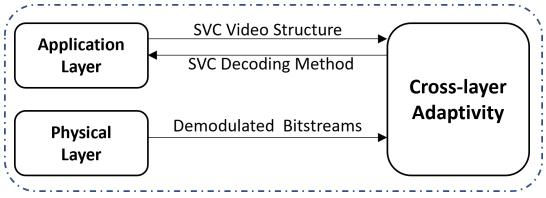


Figure 1: Proposed system model for the ASVTuw. The video camera records the video in H.264 format. At the transmitter, the cross-layer adaptivity selects the SVC encoding method and the MCS according to the feedback of required video quality and estimated CSI. At the receiver, the cross-layer adaptivity selects the SVC decoding method according to the required video quality.



(a) Transmitter



(b) Receiver

Figure 2: Cross-layer interactions at (a) Transmitter and (b) Receiver. The cross-layer adaptivity includes adaptive MCS selection and adaptive SVC selection.

based on the Rician fading channel model, which has been proved to be a good match for the short-range shallow-water channel, i.e., with a depth less than 100 m [5, 11]. After the simulated CSI dataset is trained, the ML inputs the receiver-feedback CSI from the at-sea experiments together with the BER thresholds set by the adaptive SVC selection and outputs the selected MCS.

Unequal Error Protection: The SVC generates layered bitstreams that can be modulated separately. Therefore, the BER threshold for each layer can be different according to the cross-layering interactions. When constructing the relationship between the BER and the video quality, we find that when applying low BER (e.g., 10^{-5}) at the base layer and high BER at the enhancement layers (e.g., 10^{-4}), the received video quality is still high. Therefore, the ASVTuw can select the MCS for each SVC layer according to the BER threshold per layer: The MCS with high robustness but a low data rate is selected for the base layer (low BER threshold); The MCS with low robustness but a high data rate is selected for the enhancement layers (high BER threshold). As a result, the effective data rate is improved compared with the EEP.

ASVTuw procedure: Algorithms 1 and 2 show the procedure of the ASVTuw at the transmitter (unicast case) and the receiver, respectively. The whole video transmission is

Algorithm 1 Transmitter-Side Processing for Unicast.

```
1:  TrainML();
2:  Receive(required.videoQuality);    t ← 0;
3:  BER.thresholds ← SVCselection(required.videoQuality);
4:  while t < chunkTime do
5:      videoStreams ← ScalableVideoCoding();
6:      Transmit(pilotSequence);
7:      Receive(CSI);
8:      if CSI.SNR < SNR.threshold then
9:          Increase(transmitPower);
10:         if transmitPower>maxPower then
11:             Goto 1
12:         end if
13:         Goto 6
14:     end if
15:     MCSs ← MLclassify(CSI, BER.thresholds);
16:     Transmit(MCSs, pilotSequence, videoStreams);
17:     if t ≥ chunkTime then
18:         Goto 2    % Done transmitting this chunk
19:     end if
20: end while
```

Algorithm 2 Receiver-Side Processing.

```
1:  Transmit(required.videoQuality);    t ← 0;
2:  rx.pilotSequence ← Receive();
3:  while t < chunkTime do
4:      CSI ← EstimateCSI(rx.pilotSequence);
5:      Transmit(CSI);
6:      (MCSs, rx.pilotSequence, rx.videoStreams) ← Receive();
7:      SVClayers ← Decode(rx.videoStreamsHeader);
8:      SVCdecoding(SVClayers, required.videoQuality);
9:      if t ≥ chunkTime then
10:         Goto 1    % Done receiving this chunk
11:     end if
12: end while
```

divided into several chunks. The chunk size is decided by the transmission time and decoding time, so that transmitting new chunks and decoding old chunks can be processed in parallel. After the transmitter gets the required video quality (with a packet size of about 10 bits), the BER threshold for each SVC layer is decided for EEP or UEP. The pilot sequence is transmitted to estimate the CSI at the receiver, which is composed of 64 symbols with high auto-correlation. The packet containing the CSI has a size of about 640 bits. With the most robust transmission scheme, i.e., Code Division Multiple Access (CDMA) # 1 in Table 2, the delay caused by feeding back CSI is about 1.25 s. With the feedback CSI as well as the BER threshold, the proper MCS for each SVC layer is predicted based on the trained ML model. If the SNR reported in the CSI is too low (e.g., lower than a threshold obtained in preparing the training set), the transmitter will increase the transmit power and send the pilot sequence once more, then the receiver will feed back the updated CSI to the transmitter. If the in-demand power exceeds the maximum allowed power level, the MCS with higher robustness will be introduced, e.g., CDMA with a longer spread code length. Then the training model should be updated. With the selected SVC encoding method and MCSs, the transmitter transmits the MCS information, the pilot sequence, and the video streams to the receiver. With CDMA # 1, the transmission time of each MCS packet with a size of 12 bits is 0.024 s. The receiver decodes the video streams' headers to learn the SVC structure, and reconstruct the SVC video according to the required video quality. Since the estimated CSI keeps being updated at the transmitter at the start of each transmission loop, the MCSs are always selected according to the most recent information. The video quality requirements are updated only when a chunk finishes being transmitted and received.

SVC-based Multicast: Thanks to the flexible layer structure of SVC video streams, only one video stream needs to be transmitted to meet the quality requirements of multiple users. Therefore, each user/receiver can select the SVC layers to decode according to its own need. To extend the ASVTuw to multicast, each user transmits the video quality requirement to the transmitter, and feeds back CSI in each loop of video transmission. If one of the receivers feeds back a low SNR, the transmitter will increase the transmit power. Different from the unicast scenario, this time the transmitter needs to meet the requirements and BER thresholds of multiple users. The transmitter predicts the suitable MCSs for all the receivers and generates lists of the suitable MCSs, each list corresponding to one user. The intersections of the lists are selected as the MCSs that meet all the BER thresholds. At each receiver side, the processing is the same as the unicast.

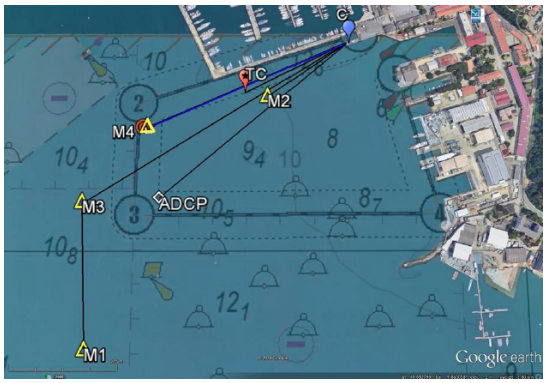


Figure 3: LOON location and deployed assets. M1, M2 and M3 are receivers, M4 is the transmitter. TC stands for the thermistor chain, which measures water temperature and sound speed. ADCP stands for acoustic Doppler current profiler. C stands for container lab. The background reports bathymetric information as well as mooring locations and working areas.

3 PERFORMANCE EVALUATION

To validate our proposed solution, we conducted a total of 12 full days of experimentation over a time window spanning from August 2020 to June 2021, using the CMRE LOON testbed, thus exploring different seasons and at-sea conditions. The results on June 10, 2021 (Figure 7) is depicted as an example. In the experiments, a multicast scenario was considered with one transmitter and three receivers at different locations, as shown in Figure 3. The collected acoustic data of received video streams and channel states were then used in simulations (MATLAB) to analyze the quality and data rate in detail and to compare with the existing noncross-layering designs. (We did not have the proposed solution running in real time on the LOON.) Different video quality metrics are investigated using Luminance Peak Signal-to-Noise Ratio (Y-PSNR), Structural Similarity (SSIM) and Mean Opinion Score (MOS). The Y-PSNR metric measures the luminance-associated distortion based on the overall Mean Square Error (MSE) of video streams. The SSIM metric instead measures the similarity of the original and received video streams. To correlate better with the human perceived video quality, MOS is applied as the subjective metric. The MOS has a scale from 0 to 100 and is calculated based on the existing dataset [**?**]. Since the different metrics focus on different video characters, the variation tendency of different metrics won't be totally same, which is the reason why we consider the three metrics when building relationships between the physical and application layers.

Three modulation schemes are considered: CDMA, Orthogonal Frequency Division Multiplexing (OFDM), and Orthogonal Signal-Division Multiplexing (OSDM). K is the number of symbols in a symbol vector. In one frame, there is

Table 1: Parameters Setting for Experiments.

Parameter	Value
Symbol rate f_s	6 kBd
DAC sampling rate	48 kHz
Frequency band of LOON testbed	8 − 14 kHz
ADC sampling rate	128 kHz
Source modulation	QPSK
Modulation order M	2
Number of symbols per symbol vector K	64
Number of symbol vectors per frame N	3 − 6
Number of Pilot vectors per frame	1
Length of GI or CP L	60
Carrier frequency f_c	11 kHz
Sound power level re 1 pW	175 − 180 dB
Turbo coding rate R_{ch}	1/3
CDMA spread code length	4

one pilot vector and $N-1$ data vectors. L is the length of zero Guard Intervals (GI) or Cyclic Prefix (CP). The frame length $L_f = KN + L$, and the effective data rate $\eta_f = \frac{MK(N-1)R_{ch}}{(KN+L)T_s}$, where M is the modulation order. $T_s = 1/f_s$ is the symbol period of the system. Table 1 states the setting of parameters for experiments, where DAC is the Digital-to-Analog Converter and ADC is the Analog-to-Digital Converter. Since the frequency band is limited in 8 kHz − 14 kHz, the effective data rate is also limited. There is one speaker at the transmitter and one hydrophone at each receiver. Table 2 describes the parameters of different MCSs. Note that the CDMA transmits signals with a spread code length of 4, so the effective data rate of CDMA is 1/4 times of other schemes' effective data rates. At the application layer, the SVC is encoded by Joint Scalable Video Model (JSVM) software, and decoded by OpenSVC Decoder [7]. In what follows, we first present the CMRE LOON testbed, used for at-sea experiments and data collection. We then illustrate the procedure of the ASVTuw strategy, including the adaptive SVC selection and the Deep Convolutional Neural Networks (DCNN)-based adaptive MCS selection. The results show that the ASVTuw strategy can select the video transmission scheme while meeting users' quality requirements effectively in a time-varying fading underwater acoustic channel.

CMRE LOON Testbed: The CMRE LOON [3] is a permanent testbed deployed in the Gulf of La Spezia, in Italy. It consists of 4 bottom-mounted tripods (M1, M2, M3 and M4 in Figure 3) with acoustic communication equipment plus oceanographic and meteorological sensors all cabled to shore and remotely accessible. All four tripods are able to transmit arbitrary waveforms while only M1, M2 and M3 are able to record data using the icListen smart hydrophone, so we consider a multicast transmission where M4 is used as the transmitter and M1, M2, M3 are the receivers[1]. The underwater depth in the area is around 10 m with transducers deployed about 1 m above the seafloor. The communication link is half-duplex. Transmissions with different signal configurations are performed in a round-robin way on the CMRE LOON to collect the required data.

Table 2: Cases of Physical-Layer Modulations.

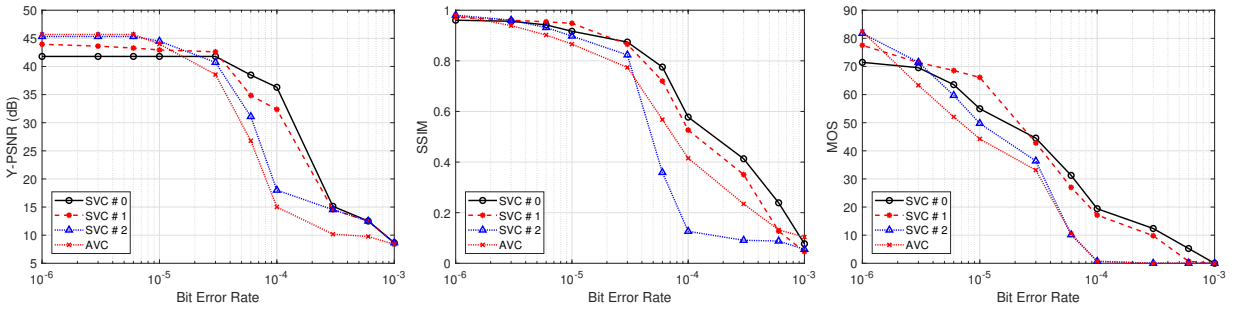
Case	Type	Parameter	L_f	η_f[kbps]
# 1	OFDM/OSDM	$N = 3$	252	2.03
	CDMA			0.51
# 2	OFDM/OSDM	$N = 4$	316	2.43
	CDMA			0.61
# 3	OFDM/OSDM	$N = 5$	380	2.69
	CDMA			0.67
# 4	OFDM/OSDM	$N = 6$	444	2.88
	CDMA			0.72

Table 3: Encoder Specifications.

Part	Parameter	Value
SVC Base Layer	Bitrate	5.6 kbps
(QP = 30)	SR	320 × 184
SVC Quality Enhancement Layer	Bitrate	16.4 kbps
(QP = 26)	SR	320 × 184
SVC Spatial Enhancement Layer	Bitrate	29.2 kbps
(QP = 26)	SR	640 × 368
AVC	Bitrate	29.2 kbps
(QP = 26)	SR	640 × 368

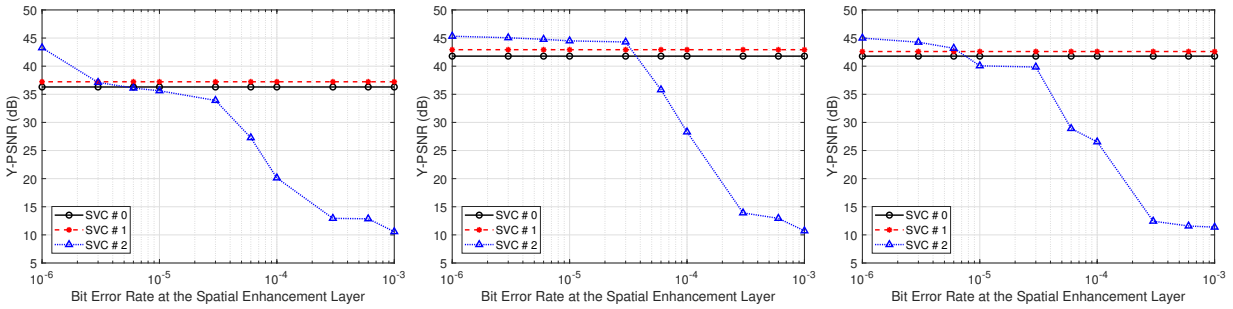
SVC Video Structure Customization: Figure 4 displays the received video quality versus BER considering both SVC and Advanced Video Coding (AVC) with EEP, which depicts the relationship between the physical-layer BER and the application-layer video quality. The SVC video stream is generated with one base layer, one quality enhancement layer and one spatial enhancement layer, while the AVC video only has one layer without any enhancement layers [9]. All the layers are set with a slow frame rate of 1.875 fps. Parameters for the SVC and AVC design are stated in Table 3. The Quantization Parameter (QP) regulates how much spatial detail is saved. The Spatial Resolution (SR) refers to the number of pixels in an image. From Figure 4, we can observe that when the BER is lower than 10^{-5}, the SVC video with more layers improves the video quality. However, when the BER is high, the SVC video with more layers has a lower quality than

[1]The distance is 310 m between M4 and M1, 280 m between M4 and M2, and 200 m between M4 and M3.



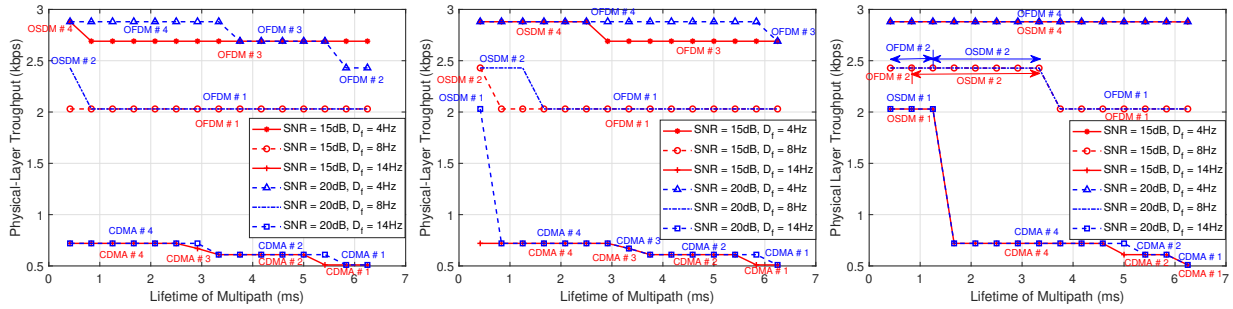
(a) (b) (c)

Figure 4: Video quality with EEP and different video structure versus BER with SVC/AVC: (a)Y-PSNR; (b) SSIM; (c) MOS. The SVC # 0 stands for SVC only with the base layer. The SVC # 1 stands for SVC with the base layer and one quality enhancement layer. The SVC # 2 stands for SVC with the base layer, one quality enhancement layer and one spatial enhancement layer.



(a) $\text{BER}_B = 10^{-4}$, $\text{BER}_Q = 10^{-5}$ (b) $\text{BER}_B = 10^{-5}$, $\text{BER}_Q = 10^{-5}$ (c) $\text{BER}_B = 10^{-5}$, $\text{BER}_Q = 10^{-4}$

Figure 5: Video quality with UEP and different video structures versus varying BER at the spatial enhancement layer. BER_B is the BER at the base layer. BER_Q is BER at the quality enhancement layer.



(a) BER threshold = 10^{-5} (b) BER threshold = 10^{-4} (c) BER threshold = 10^{-3}

Figure 6: The average physical-layer throughput of the proposed ASVTuw with different channels and BER thresholds. The parameters for different MCSs are shown in Table 2. D_f is the Doppler frequency shift. To allocate the resource adaptively and efficiently, the ASVTuw uses multiple MCSs at the physical layer and selects the MCSs according to the CSI.

that with fewer layers. The spatial enhancement layer has a larger size and introduces more errors when the BER is high, which is the reason why the curves of SVC # 2 (one base layer, one quality enhancement layer and one spatial enhancement layer) drop rapidly with increasing BER. The Y-PSNR of AVC drops even more rapidly than those of SVC # 2 when the BER is lower than 3×10^{-5}, since the AVC lacks error resilient coding and error concealment compared with SVC.

The Y-PSNR performance with UEP is shown in Figure 5. In Figure 5(c), where the BER is 10^{-5} at the base layer and is 10^{-4} at the quality enhancement layer, the performance of SVC # 1 (one base layer and one quality enhancement layer) and SVC # 2 is worse than for (b) but better than for (a), and it is able to tolerate more error bits than (b) at the quality enhancement layer.

Table 4: Required Y-PSNR (dB) at different users.

Time	M1	M2	M3
00 : 00 − 08 : 00	> 40	> 45	> 45
08 : 00 − 09 : 30	> 35	> 35	> 35
09 : 30 − 12 : 30	> 40	> 42	> 42
12 : 30 − 15 : 00	> 35	> 35	> 35

Physical-Layer MCS Customization: In this work, the MCS is processed by training a DCNN composed of convolutional layers, average pooling layers, and fully connected layers. Compared with other basic ML models, such as decision tree and adaptive boosting ensemble, the DCNN can input the BER threshold and the entire CSI sequence and learn the characteristics of the channel state without losing information [4]. The input dataset includes the CSI dataset and the corresponding proper MCSs. The CSI dataset is composed of channel impulse responses with the SNR of 15, 16, ..., 25 dB, the multipath delay lifetime of 0.42, 0.84, ..., 6.30 ms, and Doppler frequency shift of 4, 6, ..., 14 Hz, which are prepared by simulations. We find that when the SNR is lower than 15 dB, an MCS with a lower effective data rate (e.g., CDMA # 1) is always preferred. Hence, the SNR threshold is set as 15 dB. The channel impulse response is a vector of 64 complex numbers. There are 200 CSI samples for each combination of SNR, multipath delay lifetime and Doppler shift. The BER thresholds include three values: 10^{-3}, 10^{-4}, and 10^{-5}. In the at-sea experiments, the feedback estimated CSI from the real world is treated as the input to predict the optimal MCSs. The transmitter determines what MCSs to select based on DCNN classification, as shown in Figure 6. Once the prediction of MCSs is inaccurate, the BER might be increased due to a less robust MCS, or the physical-layer throughput might be decreased due to a low transmission data rate.

Adaptivity of Our Solution: Figure 6 shows the average physical-layer throughput of the proposed ASVTuw. The physical-layer throughput is equal to $\eta_f \times (1-\text{BER})$, which determines the maximum achievable video transmission bitrate. When the channel is good, the ASVTuw selects physical-layer MCS with a higher data rate (e.g., OFDM # 4). When the channel is bad, the ASVTuw selects MCS with more robustness but also with a lower data rate (e.g., CDMA # 1). Therefore, the resource is allocated adaptively to improve the system robustness and achieve an efficient data rate. Figure 7 depicts the comparison between the received video quality performance of adaptive SVC selection in the ASVTuw and that of the noncross-layering design (the SVC encoding method is fixed). The video quality requirements of three users are listed in Table 4. Figure 7(a) shows the varying BER when using EEP-based ASVTuw from 00 : 00 to 15 : 00 on June 10, 2021. The transmission schemes are selected by adaptive MCS selection based on DCNN. Figure 7(b) shows the

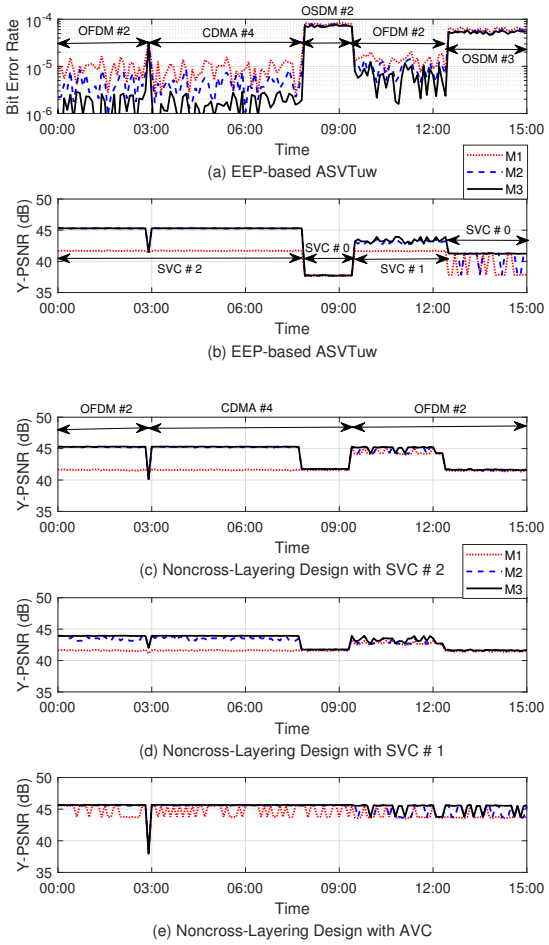


Figure 7: On June 10, 2021, with EEP, (a) The BER with adaptive MCS selection in ASVTuw. (b) The received video quality with adaptive SVC selection in ASVTuw. The performance of noncross-layering design with fixed video encoding methods: (c) SVC # 2; (d) SVC # 1; (e) AVC. The MCSs for (c)(d)(e) are the same.

received video quality at the receivers. In each loop of video transmissions, the CSI is updated and feedback to the transmitter to decide if the MCS needs to be changed. At 02 : 50, the channel state varies dramatically and the CDMA #4 is selected. There is a BER peak at 02 : 50, since the MCS is not changed accordingly in time. After 03 : 00, the BER performance goes back to the previous status, because a more robust MCS (CDMA # 4) is applied. At 08 : 00, the users change the video quality requirement requesting for a Y-PSNR higher than 35 dB. To meet this request, it is sufficient for the transmitter to encode the video using only the SVC # 0 and OSDM #2 with a BER threshold of 10^{-4}. With a noncross-layering design where the SVC/AVC structure is fixed, the physical layer cannot determine the BER threshold according to the required video quality, so a secure BER threshold

would always be selected, i.e., 10^{-5}, and the UEP would not be applied. For the noncross-layering design with SVC # 2 shown in Figure 7(c), the SVC decoding method is according to the required video quality. From 08 : 00 to 09 : 30, the video quality requirements of the three users are all Y-PSNR above 35 dB, so only the SVC base layer is decoded, while the enhancement layers are discarded, which is a waste of resources. For the noncross-layering design with SVC # 1 shown in Figure 7(d), the received video quality of M2 cannot meet the required video quality with a PSNR of 45 dB from 00 : 00 to 08 : 00, because the achievable video quality of SVC # 1 is limited. For the noncross-layering design with AVC shown in Figure 7(e), the receivers cannot select the decoding method according to their requirements, which is less flexible than SVC.

UEP-based ASVTuw: When using UEP-based ASVTuw, better results can be achieved with respect to using EEP. When using EEP, as in Figure 7(b), the SVC # 1 with OFDM # 2 is selected at 09 : 30, so the effective data rate $\eta_f = 2.43$ kbps according to Table 2. However, with UEP, the BER threshold is 10^{-5} for base layer packets, and is 10^{-4} for quality enhancement layer packets, as shown in Figure 5(c). Hence, we transmit base layer packets with OFDM # 2 and quality enhancement layer packets with OSDM # 3, which also meets the users' requirements. The size of the base layer packets is 9693 bytes in total and the size of the quality enhancement layer packets is 17947 bytes in total, so the effective data rate is $\frac{2.43\times9693+2.69\times17947}{9693+17947} = 2.60$ kbps. Therefore, the effective data rate is increased with UEP compared with EEP.

4 CONCLUSION

We proposed a novel adaptive cross-layer video transmission solution for underwater acoustic networks, namely ASVTuw, to avoid resource waste as well as to meet the requirements of video quality adaptively and effectively. The advantages of the ASVTuw include selecting MCSs adaptively with EEP/UEP by referring to the CSI based on ML, i.e., DCNN, decoding the SVC video adaptively according to users' video quality requirements, and saving resources by avoiding transmitting redundant SVC enhancement layers. The proposed ASVTuw was validated in a half-duplex acoustic multicast network with at-sea experiments using the CMRE LOON testbed. The results showed that the proposed ASVTuw had the capacity to reduce resource waste and meet users' requirements with higher flexibility compared to existing noncross-layering designs.

Future Work: A full-duplex multicast with the adaptive cross-layering strategy for video transmissions will be designed and experimented. An acoustic data collection and dataset updating will be performed spanning across different daytime/seasons and varying weather/tidal conditions, thus incurring different ambient noise levels and communication scenarios.

ACKNOWLEDGMENTS

The authors would like to thank the NATO STO CMRE team for the excellent support during the experimental activity. The research is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 731103 and NSF NeTS Award No. CNS-1763964.

REFERENCES

[1]]database [n. d.]. Laboratory for Image and Video Engineering - The University of Texas at Austin. http://live.ece.utexas.edu/research/Quality/

[2] Mohammad Ali, Dushantha Nalin Jayakody, Tharindu Perera, Abhishek Sharma, Kathiravan Srinivasan, and Ioannis Krikidis. 2019. Underwater Communications: Recent Advances. In *Conference: International Conference on Emerging Technologies of Information and Communications (ETIC)*. 1–10.

[3] Joao Alves, John Potter, Piero Guerrini, Giovanni Zappa, and Kevin LePage. 2014. The LOON in 2014: Test bed Description. In *Underwater Communications and Networking (UComms)*. 1–4.

[4] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

[5] Hovannes Kulhandjian and Tommaso Melodia. 2014. Modeling Underwater Acoustic Channels in Short-Range Shallow Water Environments. In *Proceedings of the International Conference on Underwater Networks & Systems* (Rome, Italy). 1–5.

[6] Mingfu Li and Cheng-Han Lee. 2020. Design and Analysis of CQI Feedback Reduction Mechanism for Adaptive Multicast IPTV in Wireless Cellular Networks. In *IEEE Transactions on Vehicular Technology*, Vol. 69. 2008–2020.

[7] Mickael Raulet Mederic Blestel. 2010. Open SVC Decoder: A Flexible SVC Library. In *Proceedings of the inter-national conference on Multimedia*. 1463–1466.

[8] Konstantinos Pelekanakis, Luca Cazzanti, Giovanni Zappa, and Joao Alves. 2016. Decision Tree-Based Adaptive Modulation for Underwater Acoustic Communications. In *IEEE 3rd Underwater Communications and Networking Conference (UComms)*. 1–5.

[9] Ilias Politis, Lampros Dounis, and Tasos Dagiuklas. 2012. H.264/SVC vs. H.264/AVC Video Quality Comparison under QoE-Driven Seamless Handoff. In *Signal Processing: Image Communication*, Vol. 27. 814–826.

[10] Dario Pompili, Tommaso Melodia, and Ian F. Akyildiz. 2009. Three-Dimensional and Two-Dimensional Deployment Analysis for Underwater Acoustic Sensor Networks. In *Ad Hoc Networks*, Vol. 7. 778 – 790.

[11] Andreja Radosevic, John G. Proakis, and Milica Stojanovic. 2009. Statistical characterization and capacity of shallow water acoustic channels. In *OCEANS 2009-EUROPE*. 1–8.

[12] Mehdi Rahmati, Zhuoran Qi, and Dario Pompili. 2021. Underwater Adaptive Video Transmissions using MIMO-Based Software-Defined Acoustic Modems. In *IEEE Transactions on Multimedia*. 1–13.

[13] Aleksandr Y. Rodionov, Sergey Y. Kulik, Fedor S. Dubrovin, and Peter P. Unru. 2020. Experimental Estimation of the Ranging Accuracy Using Underwater Acoustic Modems in the Frequency Band of 12 kHz. In *The 27th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*. 1–3.