

Underwater Adaptive Video Transmissions using MIMO-based Software-Defined Acoustic Modems

Mehdi Rahmati, *Senior Member, IEEE*, Zhuoran Qi, *Student Member, IEEE*,
and Dario Pompili, *Fellow, IEEE*

Abstract—Achieving reliable acoustic wireless video transmissions in the extreme and uncertain underwater environment is a challenge due to the limited bandwidth and the error-prone nature of the channel. Aiming at optimizing the received video quality and the user’s experience, an adaptive solution for underwater video transmissions is proposed that is specifically designed for Multi-Input Multi-Output (MIMO)-based Software-Defined Acoustic Modems (SDAMs). To keep the video distortion under an acceptable threshold and to keep the Physical-Layer Throughput (PLT) high, cross-layer techniques utilizing diversity-spatial multiplexing and Unequal Error Protection (UEP) are presented along with the scalable video compression at the application layer. Specifically, the scalability of the utilized SDAM with high processing capabilities is exploited in the proposed structure along with the temporal, spatial, and quality scalability of the Scalable Video Coding (SVC) H.264/MPEG-4 AVC compression standard. The transmitter broadcasts one video stream and realizes multicasting at different users. Experimental results at the Sonny Werblin Recreation Center, Rutgers University-NJ, are presented. Several scenarios for unknown channels at the transmitter are experimentally considered when the hydrophones are placed in different locations in the pool to achieve the required SVC-based video Quality of Service (QoS) and Quality of Experience (QoE) given the channel state information and the robustness of different SVC scalability. The video quality level is determined by the best communication link while the transmission scheme is decided based on the worst communication link, which guarantees that each user is able to receive the video with appropriate quality.

Index Terms—Scalable video coding; software-defined modem; underwater acoustic MIMO communications; video transmissions.

I. INTRODUCTION

Overview: Video transmissions enable a wide range of applications in the underwater environment such as coastal and tactical multimedia surveillance, marine debris detection and monitoring, undersea/offshore exploration, oil pipe/bridge inspection, monitoring of geological/biological processes from the seafloor to the air-sea interface. In order to enable these applications, which all require (near) real-time video acquisition and transmissions [2], and to pave the way towards the futuristic Internet of Underwater Things (IoUTs) paradigm [3],

M. Rahmati is with the Dept. of Electrical Engineering and Computer Science, Cleveland State University, OH, which he joined after graduating from Rutgers in July 2020, where this work was done. Z. Qi and D. Pompili are with the Dept. of Electrical and Computer Engineering (ECE), Rutgers University–New Brunswick, NJ, USA. E-mails: {mehdi.rahmati, zhuoran.qi, pompili}@rutgers.edu.

A preliminary shorter version of this work appeared in the Proc. of *IEEE/MTS OCEANS Conference*, Oct’18 [1].

This work was supported by the NSF NeTS Award No. CNS-1319945.

achieving *reliable multimedia transmissions* is a necessity, especially from places where humans cannot easily/safely go. Moreover, any communication solution aiming at enabling these applications should support different Quality of Service (QoS) requirements ranging from delay sensitive to delay tolerant and from loss sensitive to loss tolerant [4].

In practical scenarios, underwater Remotely Operated Vehicles (ROVs) are usually used, which are tethered to the supporting ship by a high-speed cable. This constrains the mission as well as the number of ROVs that can operate simultaneously in the same body of water. This is a serious limitation in the (i) development of underwater systems for future applications; (ii) maneuverability and range of the vehicles engaged in the mission; and (iii) coordination of multiple vehicles in the mission. In other cases, when not tethered, the vehicles have to rise periodically to the surface to communicate with a remote station via Radio-Frequency (RF) signals. Resurfacing periodically does not guarantee interactivity as well and leads to considerable energy/time inefficiencies.

Motivation: Having a reliable and high-speed wireless transmission underwater is a challenge in such an environment in which RF waves are absorbed for distances above a few tens of meters, optical waves suffer from scattering and ocean wave motion, and acoustic waves—while being able to propagate up to several tens of kilometers—lead to a communication channel that is dynamic, prone to fading, spectrum limited with the bandwidth of only a few tens of kHz due to high transmission loss at frequencies above 50 kHz, and affected by the ambient non-white noise [5]. While conventional underwater acoustic modems with their fixed-hardware designs [6] hardly meet the required video streaming bitrate and flexibility to support video requirements for futuristic applications, recently other solutions [7]–[9] based on open-source and reconfigurable architectures employing software-defined modems have been proposed. Using software-defined modems helps scientists and engineers explore different protocols and techniques on a single hardware, perform in-network analysis, and transmit a high volume of data to a remote node depending on environment and system specifications.

Our Approach: To adapt to the underwater channel with variable video qualities and also leverage the benefits of using a software-defined modem, Scalable Video Coding (SVC) is proposed [10], which provides scalability in the processing of video and adaptation to the video quality requirements of end-users as well as to the varying characteristics of the acoustic channel. Common types of scalability include temporal (frame rate), spatial (frame size), and quality (fidelity), which can all

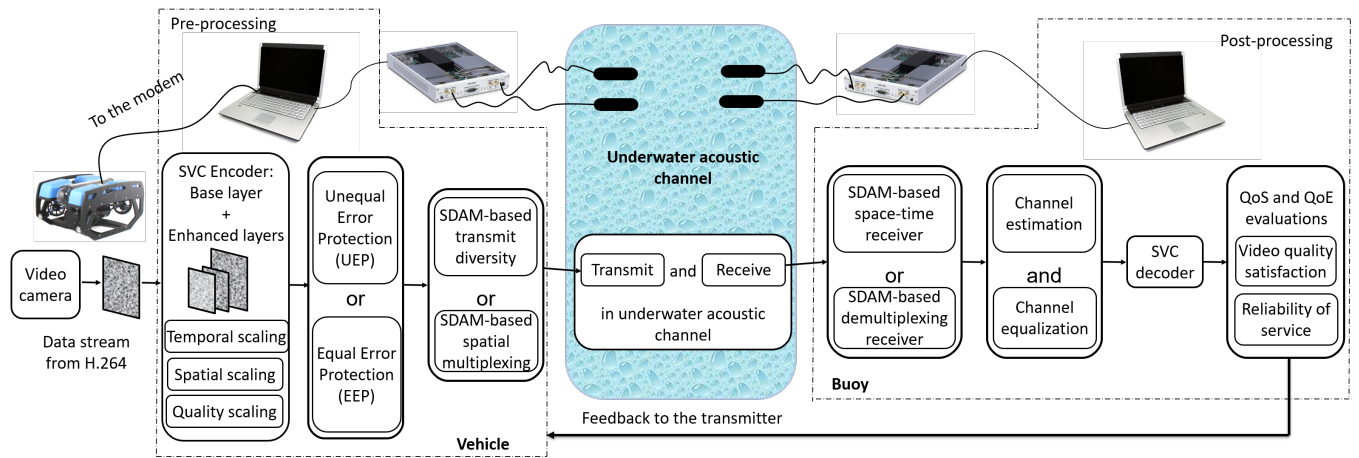


Fig. 1: Proposed system model for the MIMO-based software-defined acoustic transmissions. Transmission techniques that utilize diversity and spatial multiplexing are the modalities. The transmitter selects the optimal transmission schemes and SVC video structure based on the feedback from the receiver.

be adaptively chosen according to the channel conditions. An SVC video can be decoded with a high flexibility based on the knowledge of the receiver's channel. The SVC encodes a high-quality video stream to one or more video substreams with different video quality (one base layer and several enhancement layers). The temporal scalability enhancement layer provides the video stream sample subset by encoding the video with a different frame rate. The spatial scalability enhancement layer provides the video stream sample subset by encoding the video with a different resolution. The quality scalability enhancement layer provides the video stream sample subset by encoding the video with a different fidelity. Thanks to the SVC layering technique, an SVC video can reach high error robustness and video quality even with limited bandwidth. The SVC introduces error resilient coding and error concealment [11]. The error resilient coding injects redundancy to bit streams to detect data losses and stop error propagation, so that receivers take advantages in recovery or concealment from random error bits when discarding packets. While the error concealment provides receivers with an estimation of transmission errors based on the correctly decoded samples at the enhancement layers.

The limited capacity of the underwater acoustic channel leads to a low transmission data rate and thus a limited video streaming bitrate and a restrained utilization of SVC video standard. To make full use of this channel, our approach consists in exploiting spatial diversity and multiplexing in a Multi-Input Multi-Output (MIMO) structure in cooperation (i.e., in a cross-layer manner) with the SVC and Unequal Error Protection (UEP). Given these limitations, the video should be reconstructed without much distortion notwithstanding a limited transmission data rate. An optimization is thus required to select the optimal video transmission scheme in an adaptive manner. Traditional cross-layer techniques share information among layers in Open Systems Interconnection (OSI) model, such as the physical layer and the application layer. In this work, we exploit a new operating layer, called error-control layer, to improve the system robustness. Our cross-layer (application, physical, and error control) algorithm is

evaluated via field experiments to verify the efficiency of video transmissions, in terms of quality and bitrate, when the Signal-to-Noise Ratio (SNR) varies.

Our Contributions: We propose an adaptive cross-layer solution for underwater video transmissions using a MIMO-based reconfigurable Software-Defined Acoustic Modem (SDAM) given the latest Universal Software Radio Peripheral (USRP) family product designed by the National Instrument (NI) [12]. For the application layer, we apply videos with different types of SVC scalability, which show different error robustness with varying levels of environment SNR. For the physical layer, given the underwater channel-compatible scalable coded video with a user-defined tolerable level of distortion, we navigate the *multiplexing-diversity tradeoff* with the MIMO structure to balance transmission data rate and reliability. Experiment results show that “multiplexing” improves the transmission data rate significantly at high SNRs, while “spatial diversity” enhances the video quality at low SNRs. We also add an error-control layer, where the UEP is applied to improve the system robustness by encoding the video header packets with low channel coding rates and encoding the video body packets with high channel coding rates, since the video header packets is more important than the video body packets. Our proposed UEP seldom reduces the Physical-Layer Throughput (PLT) (which influences the achievable video streaming bitrate) and achieves higher reliability so as to avoid error propagating from important parts to less important parts. Different from terrestrial channels, the underwater channels are time-varying and the variation is hard to estimate. Therefore, we do not fix the MIMO schemes or the combination of channel coding rates but select the appropriate scheme according to the time-varying channel and users' requirements. With this cross-layer solution, the channel capacity is improved by joint work of MIMO and UEP within the limited underwater acoustic channel, the video reaches the optimal PLT within the channel capacity and video scalability, and video distortion is reduced. While optimizing the video quality, results shows that the optimal QoS cannot stand for Quality of Experience (QoE) completely, so we take

both the objective and subjective metrics into our decision process in the algorithm to make the optimization results closer to the human experience. All these are well-designed in a software-defined radio platform which can be installed on an underwater vehicle, and which is capable of switching the mode from one MIMO scheme to another adaptively and selecting the combination of channel coding rates based on the channel conditions and the target desired QoS and/or QoE. Several experiments have been conducted with USRP X300 Software-Defined Radio (SDR) at the Sonny Werblin Recreation Center at Rutgers University on a camera-equipped SDAM-based underwater vehicle, and the results are presented in this work. The adaptivity of our system is discussed based on the experimental results in different scenarios.

Article Organization: In Sect. II, we present the relevant publications. In Sect. III, we propose our solution. In Sect. IV, we present the performance results based on the conducted experiments; finally, in Sect. V, we draw the main conclusions.

II. RELATED WORK

Conventional video coding does not meet the underwater video transmission requirements for the futuristic applications. This goal is even harder to achieve in distances above hundred meters through the acoustic channel, as acoustic waves suffer from attenuation, Doppler spreading, high propagation delay, and time-varying propagation characteristics [4]. To achieve higher PLT in the bandwidth-limited underwater acoustic channel, several techniques should be combined holistically. In [13], a Hybrid Automatic Repeat Request (HARQ)-based solution is proposed that exploits the diversity gain offered by independent links in an underwater acoustic MIMO system. An Orthogonal Frequency-Division Multiplexing (OFDM) modulated dynamic coded cooperation scheme is proposed for the underwater relay network without altering the transmission procedure in [14]. In contradistinction to these works, we consider the tradeoff of spatial diversity and multiplexing in MIMO to make full use of channel information and enhance the channel capacity. Authors in [15] discuss the relationship between underwater acoustics and optics for long-range and short-range distances, respectively, to determine the correlation between the properties and the reliability of the acoustic/optical links. In [16], a signaling method for video transmissions is proposed that makes use of multiple domains to leverage the benefits of Acoustic Vector Sensors (AVSs). In this research, we apply SVC, an extension of H.264/MPEG-4 AVC, as the solution for video delivery in underwater environment, which offers higher flexibility via different modalities—temporal (frame rate), spatial (frame size), and quality (fidelity or SNR)—to compensate for the lossy video compression and erroneous transmission environments, and also supports the scalability in the complexity and in the Region Of Interest (ROI) [10]. Related works such as an adaptive mechanism based on Scalable High Efficiency Video Coding (SHVC) is proposed for surveillance video coding [17], which achieves an enhancement of bitrate compared with the traditional SHVC video coding benchmark. The effect of scalability in SVC with the goal of providing guidelines for

an adaptive strategy to select the optimal suggestion for a given bandwidth is discussed in [18]. To evaluate the video quality, an automatic tool for measuring the subjective metric—Mean Opinion Score (MOS)—of SVC video and for improving the QoE by using a random neural network [19] is considered. Authors in [20] propose an algorithm to estimate the SVC video distortion by assessing an objective metric, the Structural Similarity (SSIM). A public database for image and video quality evaluation with subjective and objective metrics is at [21]. However, these related works only focus on the performance of application layer and ignore the influence of the physical layer. Conversely, in our research, we design a cross-layer solution to analyze the interplay between physical and application layers. Authors in [22] and [23] work on Joint Source Channel Coding (JSCC) schemes and realize UEP by launching important SVC layer streams with high-SNR antennas and launching less important SVC layer streams with low-SNR antennas in the MIMO system. Authors in [24] propose a link adaptation technique to provide perceptually-optimized UEP by selecting different modulation and coding schemes for different SVC layers. In [25] and [26], a scalable decoder distortion algorithm is proposed to estimate decoder distortion and optimally select the application-layer parameter and the physical-layer parameter with orthogonal Space-Time Block Code (STBC). In [27], the bit streams are distributed to multiple spatial channels and UEP is achieved by allocating different application-layer forwarded error correction coding rates on each video layer. These works [22]–[27] are for terrestrial systems. In underwater environments, the channel state varies over time and the variation is hard to estimate, so an adaptive error-control method is in demand. In our proposed solution, UEP is realized by combining different channel coding rates, where the video header packets are encoded with a lower channel coding rate and the video body packets are encoded with a higher channel coding rate. The UEP parameters (channel coding rates) and MIMO schemes are not fixed but chosen adaptively according to the video quality demand and the channel conditions. Moreover, our proposed UEP only applies low channel coding rates to the video header packets instead of one entire SVC layer, which hardly reduces the PLT and adapts to the limited bandwidth in underwater acoustic communications.

The other software-defined platform, presented in [7], called Underwater Acoustic Networking pLATFORM (UANT), uses open-source softwares, i.e., GNU Radio and TinyOS, together with the USRP version 1 and is reconfigurable at the physical layer, the MAC layer, and the application layer. The UANT uses one transducer per node for Single-Input Single-Output (SISO) system, and must be run on one personal computer and cannot be geared toward underwater sensor networks or underwater autonomous vehicles. Authors in [8] demonstrate a high transmission data rate real-time reconfigurable modem in software-defined underwater acoustic networks based on USRP N210 and GNU Radio. Authors in [9] propose a software-defined networking platform for short-range acoustic SDAMs, called SEANet, where the physical layer is implemented on a reconfigurable Field-Programmable Gate Array (FPGA). All the above works are SISO systems

with different OSI layers working separately. Compared with these related works, our modem and testbed have a MIMO structure and can be installed on an underwater vehicle in multi-hydrophone scenarios. Our proposed platform is specially designed for underwater joint video coding and transmissions with cross-layer design. Furthermore, we exploit high performance USRP X300 SDR, whose speed of computation and processing is several times faster than USRP N210 and USRP version 1.

III. PROPOSED SOLUTION FOR VIDEO TRANSMISSIONS

In this section, we describe our system model, followed by the proposed cross-layer multimedia communication approach that leverages the MIMO structure and scalability characteristic of the compressed video to mitigate the overall distortion.

System Model: As illustrated in Fig. 1, a camera-equipped underwater robot initially records and encodes the video in the pre-processing block using an SVC encoder. Data is protected against the noisy channel with a proper channel coding technique, i.e., UEP, as well as an appropriate MIMO scheme using either spatial diversity or spatial multiplexing. At the receiver, post-processing will be performed, and the users will participate dynamically in a closed-loop manner to tune the system based on the video quality satisfaction and the reliability of service in the received video stream. The decision is optimized, and the transmitter is notified accordingly.

We consider an SVC-based video bitstream, divided into chunks/segments, consisting of a *base layer* plus *L enhancement layers* adopting different communication modalities. The chunk size is determined by the base and enhancement layer Group of Pictures (GoP) of the SVC file. The modality is being decided at the pre- and post-processing blocks, based on the Rate-Distortion (RD) requirements of the system. For a compressed video [28],

$$D_e(R_e) = \frac{\theta}{R_e - R_0} + D_0, \quad (1)$$

where D_e represents the distortion of the encoded video and R_e is the output bitrate of the encoder; the other remaining variables, θ , R_0 , and D_0 , depend on the encoded video and on the model, and are estimated empirically. To quantify and measure the video distortion over the underwater acoustic channel, the Peak Signal-to-Noise Ratio (PSNR) is used as a metric for measuring the distortion D_e based on the overall Mean Square Error (MSE). Other metrics, such as PLT, SSIM, and MOS are also used to predict the perceived quality of the video. To reduce the amount of distortion, SVC provides hierarchical prediction structures for temporal scalability, inter-layer prediction of motion for spatial and quality scalability, and key pictures definition for drift control in packet-based quality scalable coding with hierarchical prediction structures [10]. Note that the total amount of distortion is composed of the errors caused by the lossy compression D_e and the errors caused by the underwater acoustic channel, which can be alleviated by choosing an appropriate scheme.

While sound travels through the underwater medium, part of the acoustic energy is absorbed. An expression that models the medium absorption coefficient as a function of frequency f

is, $a(f) = (0.11f^2)/(1 + f^2) + (44f^2)/(4100 + f^2) + 2.75 \times 10^{-4}f^2 + 0.003$ [29]. In this empirical formula, $10 \log_{10} a(f)$ represents the channel attenuation. Propagation loss can be modeled via $P_a = \varsigma \Delta^\varpi e^{a(f)\Delta}$, in which ς , Δ , and ϖ stand for the scattering loss, distance, and spreading loss, respectively [30]. When considering multiple propagation, in which the signal at the receiver is the outcome of several delayed signals of the original signal, the Channel Transfer Function (CTF) of each path p is $H_p(f) = \Lambda_p / \sqrt{P_a}$, where Λ_p is the cumulative reflection coefficient of surface and bottom reflections for each path. The overall CTF is calculated as $H(f) = \sum_p H_p(f) e^{j\theta_p(f)}$, in which $\theta_p(f)$ is the phase response characteristic for path p . Delay characteristic can be defined as $\tau_p = -\frac{1}{2\pi} \frac{d\theta_p(f)}{df}$, and it represents the propagation delay associated with path p . This delay is highly related to the sound speed profile, which is a function that increases with the increase of water pressure (i.e., depth), salinity, and temperature [5].

Diversity and Multiplexing Modalities: For an underwater acoustic MIMO system with m transmit and n receive hydrophones, the received signal in a flat-fading channel can be represented by $y = \mathbf{H}x + \mathcal{N}$, where $\mathbf{H}$ is the $n \times m$ channel matrix, x is the transmit signal, $\mathcal{N}$ is a zero-mean Gaussian noise. We utilize Space-Time Coding (STC) and Spatial Multiplexing (SM) to achieve spatial diversity and multiplexing gains, respectively, in order to adapt to the acoustic channel's conditions. Using SM, multiple data streams are transmitted simultaneously and the transmission data rate is improved without extra bandwidth occupation [31]. However, for a MIMO system with m transmitters, each data stream interferes with the other $m - 1$ streams; hence, the receiver should be capable of eliminating this interference. Using spatial diversity, one single data stream is space-time coded over multiple hydrophones. Thus, communication channels with different fading and interference characteristics can be utilized to collect different versions of the received data so as to improve the system's reliability [32]. Given this fundamental tradeoff, the achievable diversity-multiplexing equation can be written as, $q(r) = (m - r)(n - r)$, where $q(r)$ shows the diversity gain and $r \in \mathbb{Z}$ represents the multiplexing gain, which can be defined as, $r = 0, 1, \dots, \min(m, n)$. As two special cases, we have $q_{max} = mn$ and $r_{max} = \min\{m, n\}$. The tradeoff curve confirms that while the transmission data rate increases by r bps/Hz over an increase of 3 dB in SNR, the error rate is reduced by order of $2^{-q(r)}$.

This tradeoff, however, is achieved only under *ideal conditions*, i.e., assuming that the SNR approaches infinity for i.i.d. Rayleigh-fading channels. This asymptotic definition breaks if the SNR is limited, as is the case in real scenarios [33]. The realistic diversity and multiplexing gains for a low SNR γ , array gain g , spectral efficiency R , and outage probability $P_{out}(r, \gamma)$, are calculated as follows,

$$r = \frac{R}{\log_2(1 + g\gamma)}, \quad q(r, \gamma) = -\frac{\partial \ln P_{out}(r, \gamma)}{\partial \ln \gamma}, \quad (2)$$

$$P_{out}(r, \gamma) = \Pr[\log_2 \det(\mathbf{I}_n + \frac{\gamma}{m} \mathbf{H}^* \mathbf{H}) < R], \quad (3)$$

where $\mathbf{I}_n$ represents the $n \times n$ identity matrix and superscript $*$ stands for the conjugate-transpose operation. When STC is exploited to achieve diversity, the outage probability can be approximated given the fading distribution of the channel. It is shown in [34] that for uncorrelated MIMO channels, $\mathbf{H}$ can be represented by the variances of the power gains of channel as $\text{var} \|\mathbf{H}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m \text{var} |h_{ij}|^2$. These values can be obtained by estimating the mean powers of the channel matrix in the experiment.

At the receiver, a zero-forcing equalizer is utilized, where the demultiplexed signal is expressed as,

$$\hat{x} = (\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^* y. \quad (4)$$

To estimate the channel, pilot symbols X_p are inserted after every two data symbols and Channel State Information (CSI) is calculated by analyzing the received pilot $\hat{X}_p$. Therefore, the estimated channel $\hat{\mathbf{H}}$ can be calculated as,

$$\hat{\mathbf{H}} = (X_p^* X_p)^{-1} X_p^* \hat{X}_p. \quad (5)$$

In practical scenarios, in which the underwater channel is unknown at the transmitter, we estimate a lower bound for the outage probability given only the statistics of a statistically-equivalent channel with the same distribution and with the eigenvalue set $\{\zeta_i\}_1^m$ to initiate the process as,

$$P_{out}(r, \gamma) \sim \Pr [\log_2 \prod_{i=1}^m (1 + \frac{\gamma}{m} \zeta_i) < R]. \quad (6)$$

Some underwater acoustic channels show the behavior of Rayleigh fading [35] or Rician channel, especially in short distances (saturation condition due to heavy multipath). Therefore, a lower bound on the outage probability for finite SNRs in a Rician fading channel with the equivalent channel matrix $\mathbf{H}_{eq} = (K/(K+1))^{-0.5} \mathbf{H}_{LOS} + (K+1)^{-0.5} \mathbf{H}_w$ with line of sight ($\mathbf{H}_{LOS}$) and non-line of sight ($\mathbf{H}_w$) components, and with parameter K , can be estimated as in [33],

$$P_{out}(r, \gamma) > \prod_{i=1}^m F_i(\epsilon), \quad (7)$$

where $F_i(x)$ is a Cumulative Distribution Function (CDF) with the following description while no full CSI is assumed at the transmitter,

$$F_i(x) = \begin{cases} \Phi_i(x) & i = 1, \dots, m-1 \\ e^{-Knm} \sum_{j=0}^{\infty} \frac{(Knm)^j}{j!} \Phi_{m+j/2}(x) & i = m. \end{cases} \quad (8)$$

Here, $\Phi_i(x) = \frac{\hat{\Gamma}(n-m+2i-1, (K+1)x)}{\Gamma(n-m+2i-1)}$, where $\epsilon \propto (m, \gamma, g, \vartheta)$. Furthermore, $\Gamma(\cdot)$ and $\hat{\Gamma}(\cdot)$ are gamma and incomplete gamma functions, respectively.

The diversity gain can be initially estimated as,

$$q(r, \gamma) = \sum_{i=1}^m \frac{F'_i(\epsilon)}{F_i(\epsilon)} \left[\epsilon - \frac{mg}{1+g\gamma} (\vartheta_i^* (1+g\gamma))^{\vartheta_i^*} - \vartheta_{i-1}^* (1+g\gamma)^{\vartheta_{i-1}^*} \right]. \quad (9)$$

Here $(\cdot)'$ stands for the derivative operation, ϑ_i^* is the value that maximizes the lower bound of the outage probability in (7). Note that when the SNR is high, the diversity gain

follows the asymptotic diversity in (2) for both Rayleigh fading with a full-rank transmit covariance matrix and Rician fading [33]. The low SNR analysis is essential in MIMO systems in realistic propagation conditions. With SNR and diversity gain known, the estimated bit error rate at the physical layer and the corresponding video distortion with different video layer reconstructions can be calculated.

Pre-processing and Optimization: Let $R_e(c, l)$, with SVC layers $\{l = 1, \dots, L+1\}$, denote the bitrate for layer l of video chunk c . An appropriate bitrate $R_i \geq R_e(c, l)$ for reliable communication should be assigned to layer l in order to maximize the total bitrate, i.e., transmitting as many video layers as possible without getting an outage or erroneous reception, given the bandwidth limitations and the quality of the underwater channel as well as the max distortion allowed.

The following optimization problem justifies the aforementioned discussion,

$$\max_{\alpha_i} \mathcal{F}_{\mathcal{R}} = R_e(c, 1) + \sum_{l=2}^{L+1} \alpha_l \alpha_{l-1} R_e(c, l) \quad (10a)$$

$$\text{s.t.} \quad \sum_{i=2}^{L+1} \alpha_i \alpha_{i-1} R_i \leq R_{max}, \quad (10b)$$

$$\alpha_1 = 1, \alpha_i \in \{0, 1\}, \forall i \in \{2, \dots, L+1\}. \quad (10c)$$

The first problem is a *knapsack program*, which defines the enhancement layers of bitrate $R_e(c, l)$ that could be transmitted over the underwater channel with maximum achievable bitrate R_{max} , which depends on the channel capacity and can be enhanced by MIMO and error protection. Coefficients $\{\alpha_i\}$ determine the set of enhancement layers that can be passed through the channel given the mentioned constraints. Selecting each layer depends on the presence of the preceding layer. This optimization guarantees that the base and enhancement layers can be correctly transmitted (and received) given the limited capacity of the underwater acoustic channel. As stated in (3) and (6), the MIMO transmission scheme takes full advantage of the channel, leading to an improvement of the channel capacity as well as R_{max} .

Post-processing and Video Quality Decision: To ensure the desired quality is achieved, an optimization problem finds the required parameters for the minimum possible distortion. Video header packets hold the general information of the H.264/SVC file, parameter set packets define the syntax structure of video, and slice data packets contain the detailed messages in the video. We define the distortion vector as $\mathbf{d} = [d_e \ d_c]^T$, where d_e is the encoder distortion and d_c is the channel distortion. Specifically, $d_e = D_e \left(\sum_{l=1}^{L+1} \alpha_l R_e(c, l) \right)$, where D_e is the distortion imposed by the codec at each SVC layer as presented in (1); d_c is determined via experiments as it is related to the channel effective loss rate (λ). If we assume that $\mathbf{d}_h = [d_e^h \ d_c^h]^T$ is the distortion at the stream header and that $\mathbf{d}_b = [d_e^b \ d_c^b]^T$ is the distortion at the stream body of the transmitted video, then the total distortion is modeled as,

$$D = \boldsymbol{\mu}^T [\mathbf{d}_h \ \mathbf{d}_b] \boldsymbol{\nu}, \quad (11)$$

where $\boldsymbol{\mu} = [\mu_e \ \mu_c]^T$ is a weighting vector specifying the influence of the encoder distortion and channel distortion; and $\boldsymbol{\nu} = [\nu_h \ \nu_b]^T$ is a weighting vector specifying the influence

of the distortion at the header and the body of video stream. We can cast the following optimization problem,

$$\min_{R_i} \mu^T [\mathbf{d}_h \ \mathbf{d}_b] \nu \quad (12a)$$

$$\text{s.t.} \quad \sum_{i=1}^{L+1} \alpha_i R_i \geq R_{min}, \quad (12b)$$

$$R_i \geq R_e(c, l), \quad (12c)$$

$$D \leq D_T. \quad (12d)$$

In the constraints, R_{min} stands for the minimum required bitrate to avoid P_{out} in video transmissions, and D_T represents the acceptable end-user distortion threshold. The problem can be optimized through a piecewise linear approximation method, which leads to a convex approximation function for (12a). The encoder distortion d_e is determined by the video codec and the channel distortion d_c will be alleviated by selecting an appropriate UEP scheme based on the weighting vector ν .

Unequal Error Protection (UEP): Given the structure of the video, if an error occurs in the stream header packets, the video cannot be decoded. Similarly, if the error occurs in the parameter set packets, the structure of the video will be damaged, which will lead to an extremely low-quality video. However, if the error occurs in the slice data packets, the video can be decoded successfully with relatively high quality. To achieve a high-quality video transmission, the received stream header with a negligible bit error rate is required, or the transmission of the whole video stream will fail. Since the video is much more sensitive to errors in the stream header and in the parameter set than those in the stream body, i.e., $\nu_h \gg \nu_b$, the value of ν is set as $\nu = [\nu_h \ \nu_b]^T = [5, 1]^T$ in this work. We utilize the UEP scheme to improve the received video quality by adding more redundancy in the header and parameter set. Therefore, the receiver will have the capability to recover the header and parameter set more accurately than the body [36].

Cross-layer Optimization: Algorithm 1 describes the procedure for transmitting the underwater video adaptively, where the transmitter decides on the channel coding scheme, MIMO scheme, and type of video scalability. Given the objective and subjective metrics, our SDAM adaptively self-reconfigures by solving the optimization problem so as to be able to switch between the two MIMO transmission modes, i.e., diversity-based and multiplexing-based, and decides on the number of video layers to achieve the required goals. The base-layer stream, which contains the highest priority information of the video, requires the highest reliability, while the enhancement layers require a higher bitrate, R_{max} . This, on the other hand, might result in more communication errors if the channel condition is not good. To estimate the value of R_{max} , the receiver feeds back the CSI to transmitter to estimate the channel capacity as well as R_{max} in the optimization process. In our algorithm, we consider objective and subjective metrics jointly, given the fact that the QoE is more related to the user's experience. Given similar QoS but different scalability, the QoE might be different.

Algorithm 1 Adaptive video transmissions.

```

1: Layers = scalableVideoCoder( );           % Decide video layers
2: Transmit(baseLayer); s ← 1                % s is the number of trials
3: while t < Chunk Time do
4:   Receive(feedback)
5:   if channelState.rollingAverage > threshold.distortion then
6:     transmitter.switchTo('Multiplexing')
7:   else
8:     transmitter.switchTo('Space-Time Coding')
9:   end if
10:  Estimate(diversityGain, outageProbability)
11:  if MeanOpinionScore < threshold.opinionScore then
12:    Decide(channelCoding)
13:    Reconstruct(Layers); s ← s + 1
14:  end if
15:  Transmit(Layers)
16:  channelState.update( )
17:  if s = ∑ iαi % αi stands for the layer coefficient then
18:    Goto 1 % Done Transmitting this chunk
19:  end if
20: end while

```

SVC-based Multicasting: With the SVC standardization, the low-quality video subset bitstreams can be derived and decoded from a high-quality SVC video bitstream by dropping packets. Therefore, video bitstreams with different quality levels can be received at different users by decoding the broadcast video bitstreams adaptively according to their experienced acoustic channels. When the bit error rate is high, the low-quality video stream will be decoded; whereas when the bit error rate is low, the high-quality video stream can be decoded. The video quality level will be determined by the feedback from the best communication link, while the transmission scheme will be determined by the feedback from the worst communication link, which guarantees that each user is able to receive the video stream with the appropriate quality.

Objective and Subjective Metrics: Some objective metrics are efficient to assess automatically and are of low computational cost, including PSNR, PLT, and SSIM. PLT is a physical-layer performance metric that shows the actual amount of transmitted data in one second and is calculated as,

$$PLT = \frac{MK_{FFT}R_{chc}R_{st}}{2L_T L_F T_{OFDM}} (1 - p_c), \quad (13)$$

where K_{FFT} is the Fast Fourier Transform (FFT)-size of the OFDM system and M is the order of baseband modulation, with $M = 1$ representing Binary Phase Shift Keying (BPSK). R_{chc} is the channel coding rate, R_{st} is the number of streams transmitted simultaneously. $R_{st} = 1$ for SISO and 2-by-2 STBC, $R_{st} = 2$ for 2-by-2 Vertical-Bell Laboratories Layered Space-Time (V-BLAST) [37]; p_c is the bit error rate of the received data stream, and T_{OFDM} represents the period of one OFDM symbol.

On the other hand, subjective metrics will correlate better with the human perceived video quality. The SSIM measures the fidelity of the video signals and is calculated based on the similarity of the local area luminance, local area contrast, and local patch structure. The subjective metric we apply in our study is MOS, which has a scale from 0, i.e., cannot play, to 100, i.e., fully satisfied. MOS is calculated based on the

TABLE I: Hardware Specifications.

Part	Parameter	Value
Transducer	Frequency range	1–180 kHz (Omnidirectional)
	Receiving sensitivity	−211 dB ± 3 dB re 1 V/μPa
	Transmit sensitivity	130 dB ± 3 dB re 1 V/μPa
PreAmp.	Frequency (Gain)	0.5–500 kHz (0–50 dB)
	HP/LP filters	1 Hz–250 kHz/1 kHz–1 MHz
PowerAmp.	HP filters (Gain)	1 Hz–20 kHz (0–36 dB)
Modem	Mainboard	Kintex-7 FPGA
	Frequency (Clock)	0–30 MHz (10 MHz/1 PPS)
	ADC sample rate	2 channels, 200 MS/s (14 bits)
	DAC sample rate	2 channels, 800 MS/s (16 bits)
MIMO	Uplink Structure	Up to 2x2 MIMO
	Feedback Structure	1X1 SISO (FDD Duplexing)
Camera	Standard	H.264 1080p (1X1.7 mm lens)
	Tilt range & H. FOV	±90° & 110°

Pseudo Subjective Quality Assessment (PSQA) module [19], which does subjective tests for distorted SVC videos and uses the results of this evaluation to train a Recurrent Neural Network (RNN) on the relation between the parameters that cause the distortion and the perceived quality. This module is for a normal SVC video and can be used in underwater or other environments. The MSE of this module is as low as 0.0071. The existing dataset [21] includes objective and subjective evaluations of both image and video. We use the dataset and the PSQA to get the estimated MOS value.

IV. EXPERIMENTS

To evaluate our proposal, we conduct several rounds of field experiments in a swimming pool based on the computer simulation results in [1] to validate the SVC layer structure and the SNR and to observe the video distortion. We present here both the objective and subjective assessments of the received video on the application- and physical-layer design and discuss the adaptivity of our solution to balance MIMO transmissions and channel coding as well as SVC video scalability.

Testbed Description: We modify an existing tethered ROV, called BlueRov2 [38], as shown in Fig. 1, to operate in the autonomous mode while capturing video with its 1080p camera. The video feed is passed to the acoustic modem and transducer to be sent to the buoy on the other side of the link, as shown in Fig. 1. A high-performance and scalable platform using a programmable Kintex-7 FPGA, called X-300, designed by Ettus Research Group with the NI Corporation [12], is utilized in this research. This platform contains a main-board that provides the basic functionalities of the modem and daughter-boards that take care of signal up/down conversions and other required bandpass signal processing procedures. Teledyne Marine RESON TC4013 omnidirectional transducers [39] with a frequency range from 50 to 150 kHz are used in our testbed. The specifications of the system are summarized in Table I.

Joint Scalable Video Model (JSVM) software is used as the reference package for implementing SVC. Using the FixedQPEncoder program, test videos were down-sampled and encoded into layers of different qualities. Each layer has a target bitrate, and the Quantization Parameter (QP) is varied in order to optimize the PSNR while staying under the target bitrate. OpenSVC Decoder is used for decoding due to its

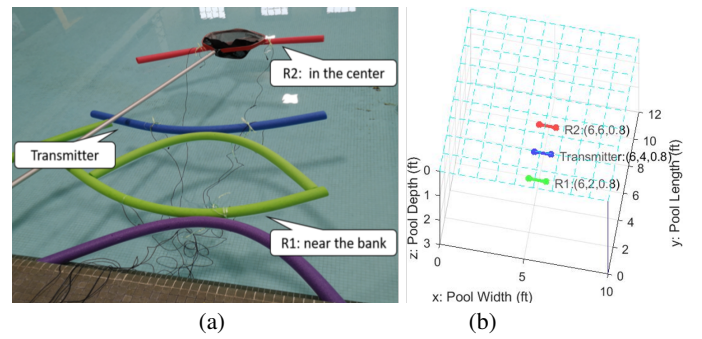


Fig. 2: (a) Testbed in the pool experiments at Sonny Werblin Recreation Center. The receivers near the bank and in the center are named as R1 and R2, respectively. (b) 3D topology in the pool. The depth of the pool is 3 ft, while the depth of the transmitter and the receivers are equal to 0.8 ft. The distance is about 2 ft between R1 and the 10-ft wall, is 4 ft between the transmitter and the 10-ft wall, and 6 ft between R2 and the 10-ft wall. The distance between R1 and the 12-ft wall is about 6 ft. So are R2 and the transmitter. The distance between the two transducers on one node (R1/R2/Transmitter) is about 1 ft.

implementation of error concealment and its integration with Mplayer for video streaming.

Pool Experiments: For our extensive experiments, hydrophones are placed in a large pool and variable distances and depths are tested. The transmission was then done with the maximum symbol rate of 100 kBaud and with H264/AVC codec JSVM signals with unknown channel. We consider placing the hydrophones near the wall and also in the center of the swimming pool with clear water (Fig. 2), which changes the results because of the multipath effect. The transmitter and receivers are not strictly fixed but floated slightly with the water wave in the pool. This situation can slightly emulate the real-world scenarios. The transmit power is adjusted mutually to get different levels of SNR as we anticipate to experience different SNRs in real-world scenarios. The stream bits are modulated with BPSK and different MIMO schemes are tested. For the frame structure, the pilot symbols are inserted after every two data symbols for channel estimation. Assume the coherent time is 3 ms with a symbol rate of 100 kBaud; therefore, the interval time between two pilot symbols is 20 μs, which is far less than the coherence time of the channel. To mitigate the multipath effect as well as to enhance the spectrum efficiency, the OFDM modulation is applied in the underwater transmissions. The OFDM FFT size is chosen to be 6144. Given a bandwidth of 100 kHz, the symbol rate is 100 kBaud and the FFT duration is $6144/100 = 61.44$ ms. We choose the cyclic prefix length to be 10.24 ms. Overall the OFDM symbol length is $61.44 + 10.24 = 71.68$ ms, and the subcarrier spacing is $1/71.68\text{ms} = 16.28$ Hz. The specifications of the SVC encoder are summarized in Table II, where layer 0 stands for the base layer and layer 1 stands for the enhancement layer. Due to the limited bandwidth in underwater acoustic networks, the video streaming bitrate is set as 30 kbps.

Figs. 3-6 show the performances of different transmission schemes with Equal Error Protection (EEP) code rate 1/3. Fig. 3 shows the PSNR of different reconstructed videos with

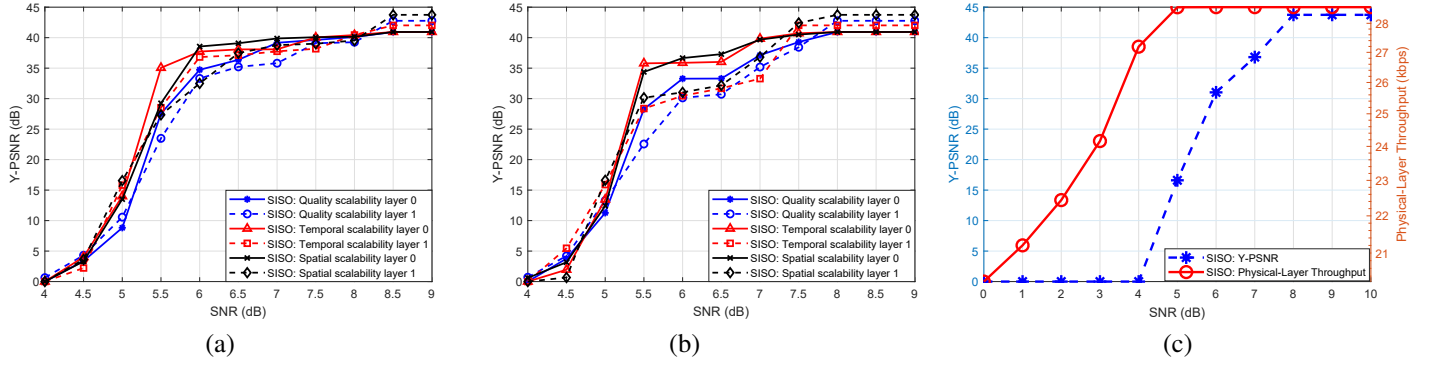


Fig. 3: Applying SISO scheme with EEP code rate 1/3, layer 0 stands for the base layer and layer 1 stands for the enhancement layer. (a) PSNR of the video received near the bank of the pool; (b) PSNR of the video received in the center of the pool; (c) PLT for the video with spatial scalability layer 1 received in the center of the pool.

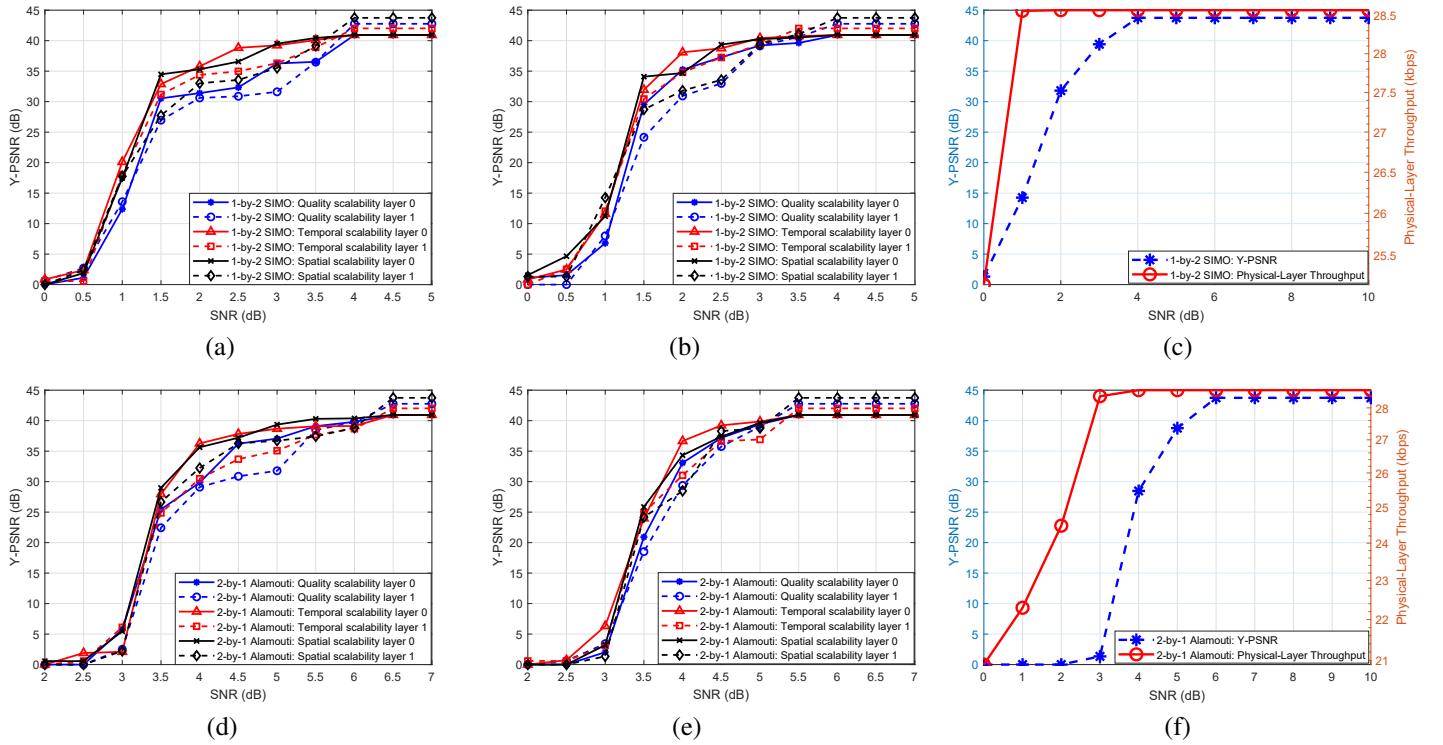


Fig. 4: With EEP code rate 1/3, (a) PSNR of the video received near the bank of the pool with 1-by-2 SIMO; (b) PSNR of the video received in the center of the pool with 1-by-2 SIMO; (c) PLT for the video with spatial scalability layer 1 received in the center of the pool with 1-by-2 SIMO. (d) PSNR of the video received near the bank of the pool with 2-by-1 Alamouti; (e) PSNR of the video received in the center of the pool with 2-by-1 Alamouti; (f) PLT for the video with spatial scalability layer 1 received in the center of the pool with 2-by-1 Alamouti.

the SISO scheme with spatial scalability, and also the PLT in the center of the pool with spatial scalability layer 1. We observe that when the SNR is low, the video with layer 0 has a higher PSNR than that with layer 1. When the SNR is high, the video with layer 1 has a higher PSNR than that with layer 0, so we apply lower-quality videos in bad channels and higher-quality videos in good channels. It should be noted that the parameter settings of the base layers (layer 0) are the same for different scalability, but the received PSNR of layer 0 is different with different enhancement layers (layer 1), since the work of error correction is different with different

enhancement layers. For example, when the SNR is 5 dB, the video stream with spatial scalability performs better than that with quality scalability. When the SNR is 5.5 dB, the video stream with temporal scalability layer 0 performs best. When the SNR ranges in 6 – 7.5 dB, the stream with spatial scalability layer 0 performs best. When the SNR is higher than 8.5 dB, the data with a PLT of 100 kbps is transmitted without errors. The stream with spatial scalability layer 1 reaches the highest PSNR.

Fig. 4 shows the PSNR and PLT of the received video with 1-by-2 Single-Input Multiple-Output (SIMO) and 2-by-

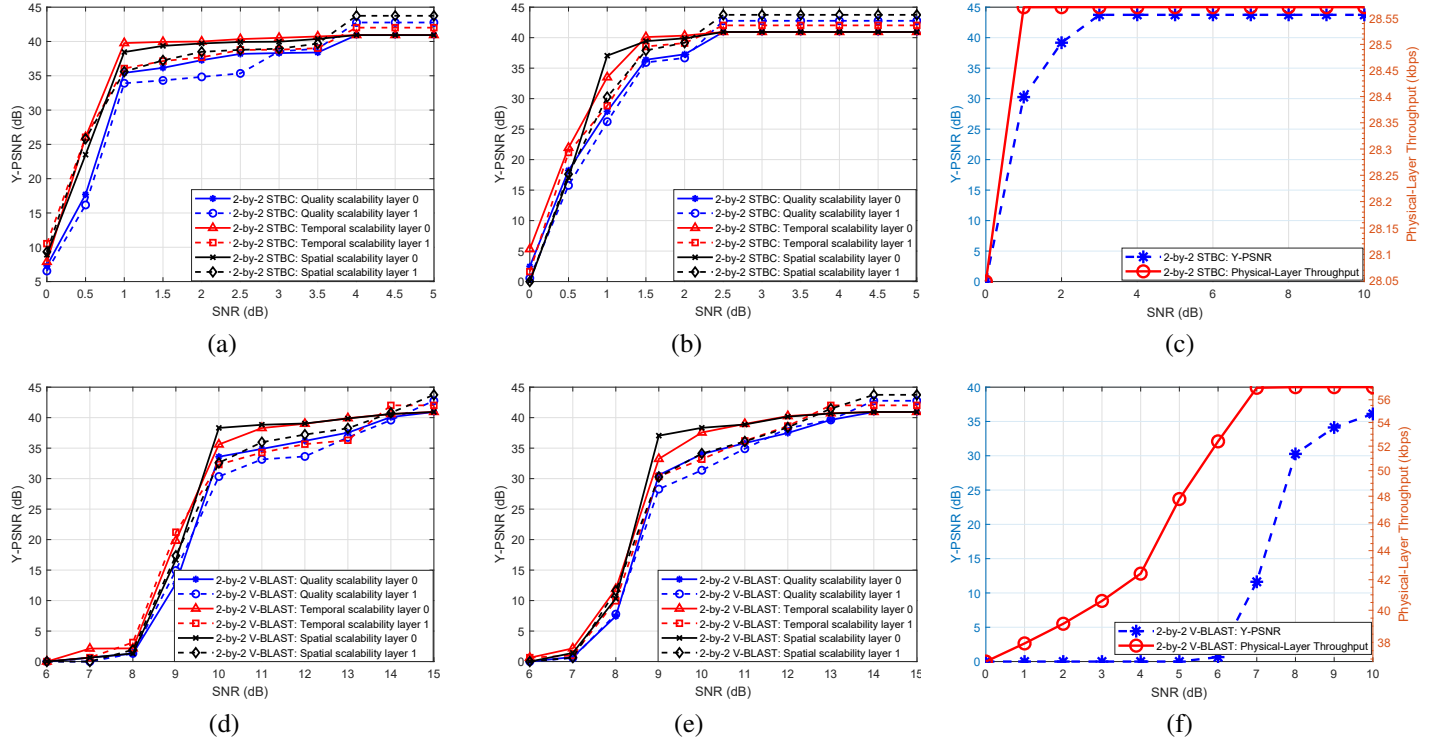


Fig. 5: With EEP code rate 1/3, (a) PSNR of the video received near the bank of the pool with 2-by-2 STBC; (b) PSNR of the video received in the center of the pool with 2-by-2 STBC; (c) PLT for the video with spatial scalability layer 1 received in the center of the pool with 2-by-2 STBC. (d) PSNR of the video received near the bank of the pool with 2-by-2 V-BLAST; (e) PSNR of the video received in the center of the pool with 2-by-2 V-BLAST; (f) PLT for the video with spatial scalability layer 1 received in the center of the pool with 2-by-2 V-BLAST.

TABLE II: SVC Encoder Specifications.

Part	Parameter	Value
Layer 0 (QP = 32)	Frame rate	15 fps
	Spatial Resolution	640×368
	Bitrate	30 kbps
Quality scalability layer 1 (QP = 30)	Frame rate	15 fps
	Spatial Resolution	640×368
	Bitrate	30 kbps
Temporal scalability layer 1 (QP = 32)	Frame rate	30 fps
	Spatial Resolution	640×368
Spatial scalability layer 1 (QP = 32)	Frame rate	15 fps
	Spatial Resolution	1280×720
	Bitrate	30 kbps

1 multi-hydrophone Alamouti schemes [40]. Compared with SISO in Fig. 3, both SIMO and Alamouti improve the robustness of the system for each receiver that gets the transmitted stream with diversity order of 2. Compared with Alamouti, SIMO has an SNR gain of 2.5 dB for R1 (the hydrophone in the center) and 1.5 dB for R2 (the hydrophone in the side). While the Alamouti scheme is able to transmit two streams simultaneously, it suffers more distortion than SIMO.

The redundancy we add into the video streams will improve the system robustness, but will reduce the PLT. Fig. 5 shows the PSNR and the PLT of the video received with 2-by-2 STBC and 2-by-2 V-BLAST [37]. Compared with SISO in Fig. 3, we can observe that the STBC improves the robustness significantly with an SNR gain of 4.5 dB for R1 and 5.5 dB

for R2. While the V-BLAST suffers more distortion than SISO, but it almost doubles the PLT. Hence, we require the multiplexing-diversity tradeoff. We note that STBC is efficient when the SNR is low, whereas V-BLAST is efficient when the SNR is high.

Fig. 6 presents the SSIM and MOS of the received video with different transmission schemes, where we can observe that the proposed cross-layer design improves the objective and subjective metrics. It shows that when the SNR is high, the SSIM with different SVC scalability is almost the same, while the MOS performances are quite different. Moreover, the spatial scalability with layer 1 has the highest MOS while the quality scalability with layer 0 has the lowest MOS, even though the PSNR (Fig. 5) and SSIM (Fig. 6) performances are close to each other. Similarly to Fig. 5, STBC improves the video quality and system robustness in the low-SNR environment, while V-BLAST can only work in the high-SNR environment but offers a higher transmission data rate.

To enhance the quality of the video, performances of UEP and EEP are compared. Fig. 7 shows the PSNR of the received video in the center of the pool, using all layers with spatial scalability. We observe that the EEP with 1/4 code rate performs the best; however, the PLT is reduced significantly. In contrast, the UEP with 1/4 code rate for the header and 1/3 code rate for the body introduces less redundancy than EEP with 1/4 code rate, which provides a lower video quality but a higher PLT. In Figs. 7(d-f), the UEP with 1/2 code rate

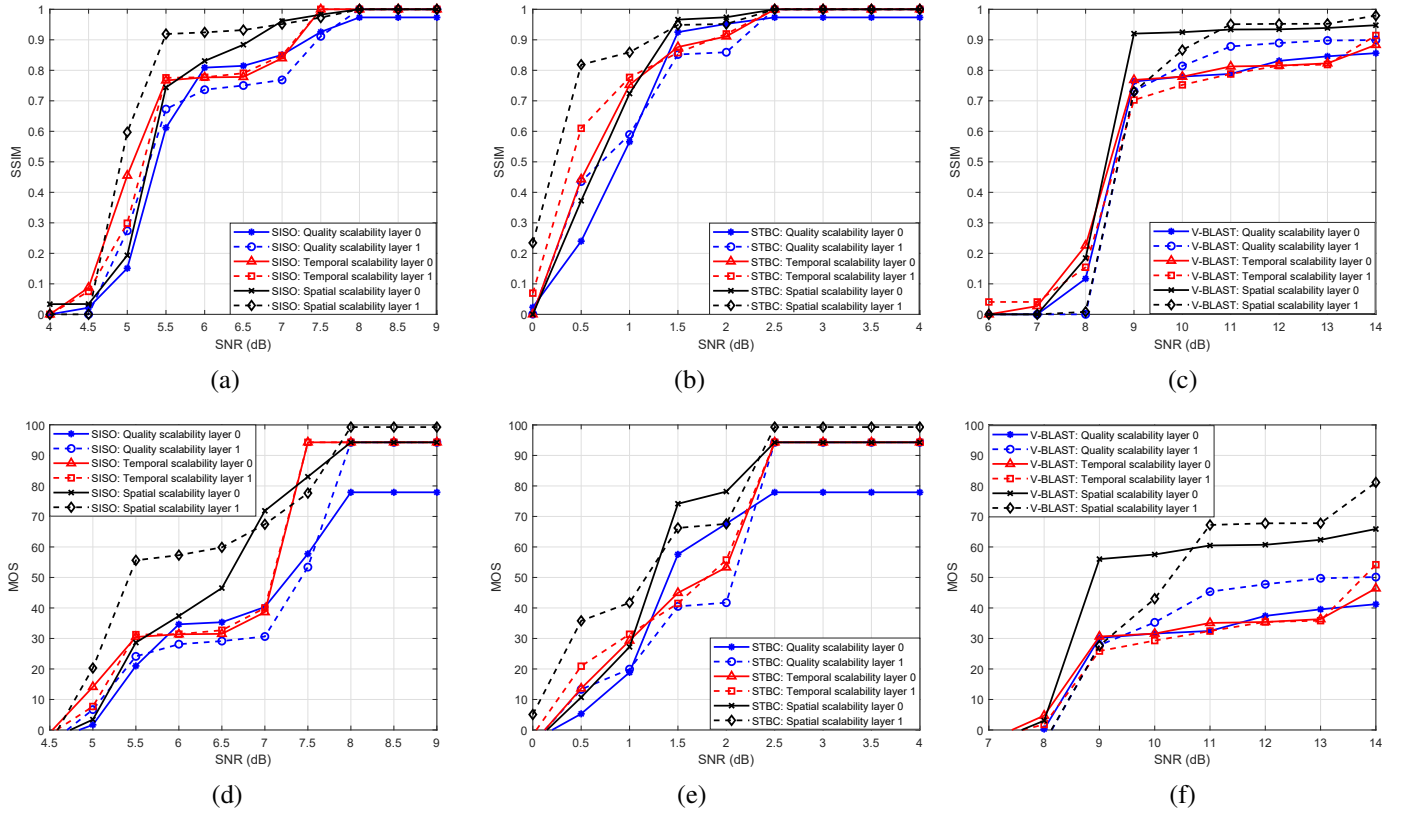


Fig. 6: SSIM of the video with different transmission schemes with EEP code rate 1/3; (a) SISO; (b) 2-by-2 STBC; (c) 2-by-2 V-BLAST. MOS of the video with different transmission schemes; (d) SISO; (e) 2-by-2 STBC; (f) 2-by-2 V-BLAST.

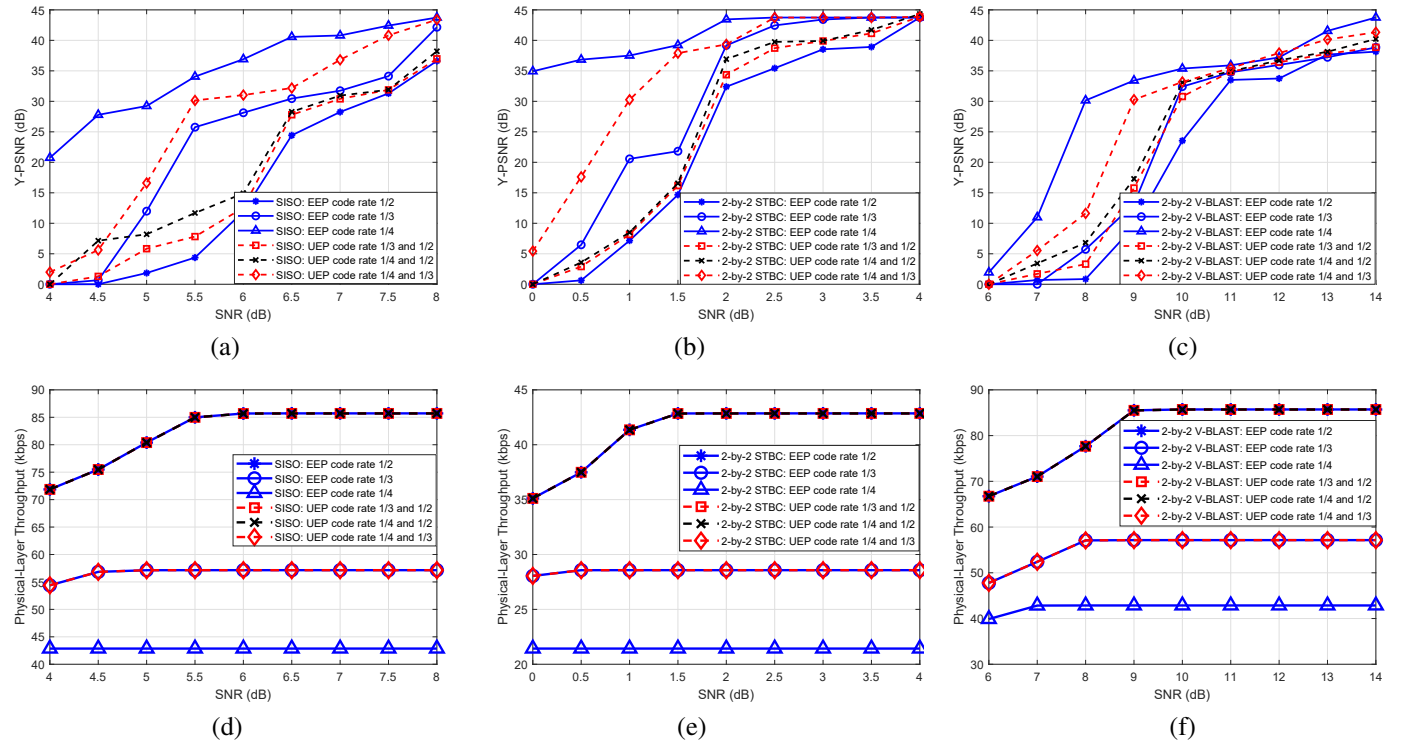


Fig. 7: PSNR of the video with different error protection schemes; (a) SISO; (b) 2-by-2 STBC; (c) 2-by-2 V-BLAST. PLT of the video with different error protection scheme; (d) SISO; (e) 2-by-2 STBC; (f) 2-by-2 V-BLAST.

only for the body are almost overlapped with the EEP with 1/2 code rate. The UEP with 1/3 code rate only for the body are almost overlapped with the EEP with 1/3 code rate. Even though we add some redundancy in the header, the length of the header is far shorter than that of the body, so the UEP reduces the PLT slightly but improves the PSNR greatly. In Fig. 7(a), we find that when the SNR is low, the UEP with 1/4 code rate for the header and 1/2 code rate for the body performs better than the EEP with 1/3 code rate, for its header is protected with higher robustness by 1/4 code rate. Fig. 8 represents the channel response with different transmission schemes experienced in this testbed, containing the phase of the channel in Figs. 8(a-b) and its power spectrum in Figs. 8(c-d). Due to the high transmission loss in high frequency band, the spectrum limited with passband bandwidths of only a few tens of kHz.

Adaptivity of Our Solution: With the cross-layer optimization algorithm described in Sect. III, we can jointly improve the PLT, the system robustness, and the video quality. Based on the optimization process in Figs. 3-6, we can select the optimal video transmission scheme. As shown in these figures, R2 performs better than R1 since it suffers from less multipath delay due to the reflected signals from the bank. Given the PSNR threshold of 30 dB and the EEP initial channel coding with Turbo coding rate of 1/3, when the SNR is 5.5 dB the optimal video transmission scheme is SISO with temporal scalability layer 0 for R1 and R2, so the transmitter only needs to transmit the video stream with temporal scalability layer 0. When the SNR is 1 dB, the optimal scheme is 2-by-2 STBC with temporal scalability layer 0 for R1 and spatial scalability layer 0 for R2. With the STBC scheme, each receiver gets up to 4 versions of received signals, which improves the reliability after gain combining. The transmit stream needs to contain both the temporal and spatial enhancement layers. When the SNR is 10 dB, the optimal video transmission scheme is 2-by-2 V-BLAST with spatial scalability layer 0 for R1 and R2, as the V-BLAST transmission scheme enables the transmitter to transmit two different streams simultaneously and achieves the highest transmission data rate. When the SNR is 9 dB, the channel coding will be switched to UEP for the V-BLAST scheme based on Fig. 7(c). The UEP puts more redundancy in the stream header, which almost doubles the PLT compared with the STBC. The detailed composition of different transmission schemes and channel coding schemes is reported in Tables III and IV.

V. CONCLUSION

We proposed a novel scheme to layerize and transmit a video stream underwater using a MIMO-based SDAM. The balance between transmission data rate and reliability, i.e., the *multiplexing-diversity tradeoff*, as well as SVC is achieved to transmit a video with a pre-defined level of distortion, which is caused by the coder and the error-prone underwater acoustic channel. The proposed optimizing algorithm provides the scalability in the video bitstream processing to adapt to the video quality requirements of end users as well as to the varying characteristics of the network. The optimal video

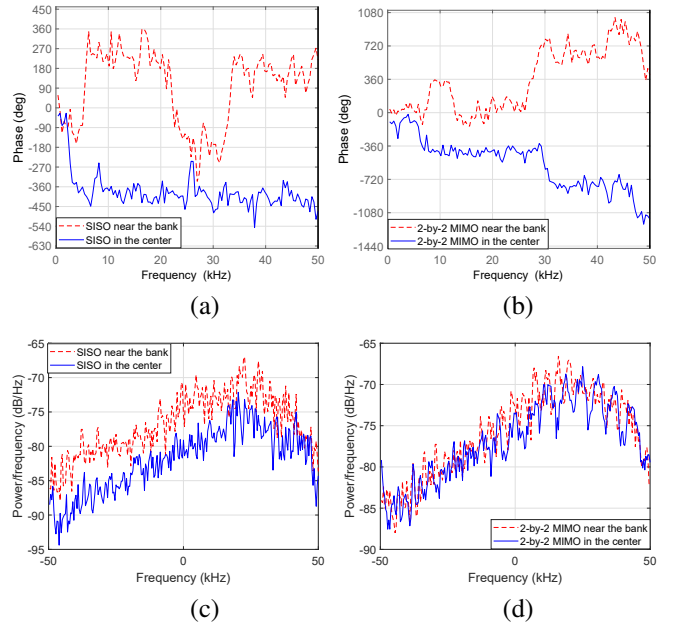


Fig. 8: Channel response in the swimming pool. Phase of (a) SISO; (b) 2-by-2 MIMO; Power spectrum of (c) SISO; (d) 2-by-2 MIMO.

TABLE III: R2 with 1/3 Turbo Coding Rate.

Scheme	SNR (dB)	PLT (kbps)	PSNR (dB)
SISO	6 – 9	28.57	30.13 – 43.74
1-by-2 SIMO	2 – 5	28.57	30.94 – 43.74
2-by-1 Alamouti	4 – 7	28.57	28.47 – 43.74
2-by-2 STBC	1 – 5	28.57	26.22 – 43.74
2-by-2 V-BLAST	9 – 15	57.14	28.28 – 43.74

TABLE IV: R2 with Spatial Scalability Layer 1.

Scheme	Channel Coding	SNR (dB)	PSNR (dB)
SISO	EEP 1/2	7.5 – 8	31.30 – 36.62
	EEP 1/3	6.5 – 8	30.45 – 42.15
	EEP 1/4	5.5 – 8	34.06 – 43.74
	UEP 1/3 – 1/2	7 – 8	30.40 – 36.97
	UEP 1/4 – 1/2	7 – 8	30.94 – 38.21
	UEP 1/4 – 1/3	5.5 – 8	30.15 – 43.74
2-by-2 STBC	EEP 1/2	2 – 4	32.39 – 43.74
	EEP 1/3	2 – 4	39.34 – 43.74
	EEP 1/4	0 – 4	34.93 – 43.74
	UEP 1/3 – 1/2	2 – 4	34.36 – 43.74
	UEP 1/4 – 1/2	2 – 4	36.93 – 43.74
	UEP 1/4 – 1/3	1 – 4	30.26 – 43.74
2-by-2 V-BLAST	EEP 1/2	11 – 14	33.51 – 38.14
	EEP 1/3	10 – 14	32.41 – 38.85
	EEP 1/4	8 – 14	30.15 – 43.74
	UEP 1/3 – 1/2	10 – 14	30.79 – 38.85
	UEP 1/4 – 1/2	10 – 14	33.18 – 40.23
	UEP 1/4 – 1/3	9 – 14	30.27 – 41.33

transmission scheme and the UEP are selected according to an optimization. The adaptivity of our system is discussed under different scenarios and both objective and subjective metrics are considered to optimize the user QoS and QoE. Experimental results at Sonny Werblin Recreation Center, Rutgers University were presented that corroborated our analysis and intuitions.

Future Work: The real-world experiments with our proposed novel scheme will be done in harsher underwater conditions, such as murky-water experiments and at-sea experiments. The influence of real-world factors (such as the temperature, the water wave speed, and different kinds of natural background noise) on the underwater acoustic communications will be explored and an adaptive solution for underwater video transmissions with the real-world factors will be designed.

Acknowledgment: The authors thank Jeffrey Zeszotarski, aquatics coordinator at Rutgers University for his help and support during the many field experiments.

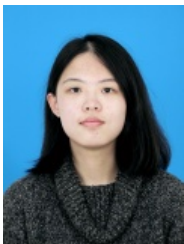
REFERENCES

- [1] M. Rahmati, A. Gurney, and D. Pompili, "Adaptive underwater video transmission via software-defined mimo acoustic modems," in *Proceedings of MTS/IEEE OCEANS*. IEEE, 2018, pp. 1–6.
- [2] M. Rahmati, R. Petrocchia, and D. Pompili, "In-network collaboration for cdma-based reliable underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 4, pp. 881–894, 2019.
- [3] M. Rahmati, V. Sadhu, and D. Pompili, "Eco-uw iot: Eco-friendly reliable and persistent data transmission in underwater internet of things," in *Proceedings of International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2019, pp. 1–9.
- [4] D. Pompili and I. F. Akyildiz, "A multimedia cross-layer protocol for underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2924–2933, 2010.
- [5] M. Rahmati and D. Pompili, "Unisec: Inspection, separation, and classification of underwater acoustic noise point sources," *IEEE Journal of Oceanic Engineering*, vol. 43, pp. 777–791, 2018.
- [6] B. Chen, P. C. Hickey, and D. Pompili, "Trajectory-aware communication solution for underwater gliders using whoi micro-modems," in *Proceedings of IEEE Conference on Sensor, Mesh and Ad-Hoc Communications and Networks (SECON)*. IEEE, 2010, pp. 1–9.
- [7] D. Torres, J. Friedman, T. Schmid, M. B. Srivastava, Y. Noh, and M. Gerla, "Software-defined underwater acoustic networking platform and its applications," *Ad Hoc Networks*, vol. 34, pp. 252–264, 2015.
- [8] E. Demirors, G. Sklivanitis, T. Melodia, S. Batalama, and D. Pados, "Software-defined underwater acoustic networks: Toward a high-rate real-time reconfigurable modem," *IEEE Communications Magazine*, pp. 64–71, 11 2015.
- [9] E. Demirors, J. Shi, R. Guida, and T. Melodia, "SEANet G2: Toward a high-data-rate software-defined underwater acoustic networking platform," in *Proceedings of the International Conference on Underwater Networks & Systems*. ACM, 2016, p. 12.
- [10] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.
- [11] Y. Guo, Y. Chen, Y.-K. Wang, H. Li, M. Hannuksela, and M. Gabbouj, "Error resilient coding and error concealment in scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 6, pp. 781–795, 2009.
- [12] USRP X Series. [Online]. Available: <https://www.ettus.com>
- [13] M. Rahmati and D. Pompili, "uwMIMO-HARQ: Hybrid ARQ for reliable underwater acoustic mimo communications," in *Proceedings of the 10th International Conference on Underwater Networks and Systems (WUWNET)*. ACM, 2015, pp. 1–8.
- [14] Y. Chen, Z. Wang, L. Wan, H. Zhou, S. Zhou, and X. Xu, "OFDM-modulated dynamic coded cooperation in underwater acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 1, pp. 159–168, 2015.
- [15] R. Diamant, F. Campagnaro, M. D. F. De Grazia, P. Casari, A. Testolin, V. S. Calzado, and M. Zorzi, "On the relationship between the underwater acoustic and optical channels," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 8037–8051, 2017.
- [16] M. Rahmati and D. Pompili, "SSFB: Signal-space-frequency beamforming for underwater acoustic video transmission," in *Proceedings of International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 2017, pp. 180–188.
- [17] T. H. Le Dao, P. V. Giap, and H. V. Xiem, "Adaptive long-term reference selection for efficient scalable surveillance video coding," in *Proceedings of International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc)*. IEEE, 2018, pp. 69–73.
- [18] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation via paired comparison: Application to scalable video coding," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 882–893, 2011.
- [19] K. D. Singh, A. Ksentini, and B. Marienal, "Quality of experience measurement tool for svc video coding," in *Proceedings of International Conference on Communications (ICC)*. IEEE, 2011, pp. 1–5.
- [20] F. Babich, M. Comisso, and R. Corrado, "Fast distortion estimation based on structural similarity for h.264/svc encoded videos," in *Proceedings of Vehicular Technology Conference (VTC Spring)*. IEEE, 2015, pp. 1–5.
- [21] "Laboratory for image and video engineering - the university of Texas at Austin." [Online]. Available: <http://live.ece.utexas.edu/research/Quality/>
- [22] D. Song and C. W. Chen, "Scalable H.264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1218–1226, 2007.
- [23] C. Zhou, C.-W. Lin, X. Zhang, and Z. Guo, "A novel JSCC scheme for UEP-based scalable video transmission over MIMO systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 6, pp. 1002–1015, 2015.
- [24] A. Abdel Khalek, C. Caramanis, and R. W. Heath, "A cross-layer design for perceptual optimization of H.264/SVC with unequal error protection," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 7, pp. 1157–1171, 2012.
- [25] M. K. Jubran, M. Bansal, R. Grover, and L. P. Kondi, "Optimal bandwidth allocation for scalable H.264 video transmission over MIMO systems," in *MILCOM 2006 - 2006 IEEE Military Communications conference*, 2006, pp. 1–7.
- [26] M. K. Jubran, M. Bansal, and L. P. Kondi, "Low-delay low-complexity bandwidth-constrained wireless video transmission using SVC over MIMO systems," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1698–1707, 2008.
- [27] X. Chen, J.-N. Hwang, J. A. Ritcey, C.-N. Lee, and F.-M. Yeh, "Quality-driven joint rate and power adaptation for scalable video transmissions over MIMO systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 2, pp. 366–379, 2017.
- [28] K. Stuhlmüller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, 2000.
- [29] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007.
- [30] S. Zhou and Z. Wang, *OFDM For Underwater Acoustic Communications*. John Wiley & Sons, 2014.
- [31] A. A. Kalachikov and N. S. Shelkunov, "Performance evaluation of the detection algorithms for MIMO spatial multiplexing based on analytical wireless MIMO channel models," in *Proceedings of International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, October 2018, pp. 180–183.
- [32] I. Ahamed and M. Vijay, "Comparison of different diversity techniques in mimo antennas," in *Proceedings of Conference on Communication and Electronics Systems (ICCES)*, October 2017, pp. 47–50.
- [33] R. Narasimhan, "Finite-SNR diversity-multiplexing tradeoff for correlated rayleigh and rician mimo channels," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3965–3979, 2006.
- [34] J. Pérez, J. Ibáñez, L. Vielva, and I. Santamaria, "Closed-form approximation for the outage capacity of orthogonal stbc," *IEEE Communications Letters*, vol. 9, no. 11, pp. 961–963, 2005.
- [35] M. Stojanovic and J. Preisig, "Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization," *IEEE Communications Mag.*, vol. 47, no. 1, pp. 84–89, 2009.
- [36] S. Shahhosseini and S. A. Hosseini, "A visual base image coding using unequal error protection in high error rate channels," in *2010 International Conference on Computer Applications and Industrial Electronics*, Dec 2010, pp. 537–542.
- [37] J. H. Chong, C. K. Ng, N. K. Noordin, and B. M. Ali, "A low computational complexity V-BLAST/STBC detection mechanism in MIMO system," *Human-centric Computing and Information Sciences*, vol. 4, no. 1, p. 2, 2014. [Online]. Available: <https://doi.org/10.1186/s13673-014-0002-1>
- [38] Blue Robotics. [Online]. Available: <http://www.bluerobotics.com>.
- [39] RESON TC4013 Hydrophone Product Information. [Online]. Available: <http://www.tedynemarine.com/reson-tc4013>
- [40] N. Amdouni, O. Ben Rhouma, and A. Bouallegue, "Alamouti scheme's generalization," in *2011 11th Mediterranean Microwave Symposium (MMS)*, 2011, pp. 169–172.



Mehdi Rahmati is an Assistant Professor at Cleveland State University, OH since Aug. 2020. He received his PhD in ECE from Rutgers University working with Prof. D. Pompili. Previously, he worked as a faculty at Bahar Institute of Higher Education and as a lecturer at Azad University of Tehran and Mashhad, Iran, after receiving his BS and MS in electrical engineering from Ferdowsi University and Iran University of Science and Technology in 2001 and 2009, respectively. Currently, he works on underwater communications and human-

robot interaction in uncertain environments. He has received the best demo award at the IEEE International Conference on Sensing, Communication and Networking (SECON'19), the best paper award at the IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS'17), and the best paper runner-up award at the ACM International Conference on Underwater Networks & Systems (WUWNet'15). In 2019, he was the recipient of the Rutgers Raritan River Consortium (R3C) award. He was selected as the winner of the FIRST PRIZE (ex aequo) of the 2019 IEEE ComSoc Student Competition. He has served as a TPC member and as an Organizing Committee member in several international conferences.



Zhuoran Qi is a PhD student in ECE at Rutgers University working with Prof. D. Pompili in the CPS Lab, which she joined in 2017. She received the BE degree from the University of Electronic Science and Technology of China in 2018 and MS degree from Rutgers in 2019. She is currently working on underwater acoustic and optical communications at the physical and network layer. She has received the best paper award at the 2019 ACM International Conference on Underwater Networks and Systems (ACM WUWNet), which was held in Atlanta, GA.



Dario Pompili is an associate professor with the Dept. of ECE at Rutgers University. Since joining Rutgers in 2007, he has been the director of the CPS Lab, which focuses on mobile edge computing, wireless communications and networking, acoustic communications, and sensor networks. He received his PhD in ECE from the Georgia Institute of Technology in 2007. He had previously received his 'Laurea' (combined BS and MS) and Doctorate degrees in Telecommunications and System Engineering from the U. of Rome "La Sapienza," Italy, in

2001 and 2004, respectively. He has received a number of awards in his career including the NSF CAREER'11, ONR Young Investigator Program'12, and DARPA Young Faculty'12 awards. In 2015, he was nominated Rutgers-New Brunswick Chancellor's Scholar. He served on many international conference committees taking on various leading roles. He published about 200 refereed scholar publications, some of which received best paper awards: with more than 12K citations, Dr. Pompili has an h-index of 44 and an i10-index of 111 (Google Scholar, October'21). He is a Fellow of the IEEE Communications Society (2020) and a Distinguished Member of the ACM (2019). He is currently serving as Area Editor for IEEE Transactions on Mobile Computing (TMC).