ON ERROR RESILIENT DESIGN OF PREDICTIVE SCALABLE CODING SYSTEMS

Ahmed ElShafiy, Tejaswi Nanjundaswamy, Sina Zamani, Kenneth Rose

Department of Electrical and Computer Engineering, University of California Santa Barbara, CA 93106
{a_elshafiy, tejaswi, sinazmn, rose}@ece.ucsb.edu

ABSTRACT

Scalable coding is potentially useful in content distribution over unreliable channels, as it enables meaningful reconstruction when the hierarchical bitstream is only partially received. However, its deployment in conjunction with predictive coding may result in considerable performance degradation when errors due to packet loss propagate through the prediction loop. Despite this, most existing predictive scalable coding techniques employ components whose design completely ignores the effects of unreliable channel conditions. This paper proposes an efficient design technique for predictive scalable coding systems, which effectively accounts for: *i)* all available information at a given layer by optimizing its prediction parameters within an estimation-theoretic framework; *ii)* the uncertainty due to packet loss via estimation and minimization of end-to-end distortion. It further leverages an asymptotic closed loop design technique for the predictor and quantizer modules, which provides the stability benefits of open-loop design, while ultimately optimizing the system for closed-loop operation. Experimental results provide compelling evidence for the effectiveness of the approach, with considerable performance gains over existing design techniques, in terms of end-to-end signal-to-noise ratio.

Index Terms— Scalable Coding, End-to-End Distortion, Asymptotic Closed Loop design, Linear Prediction

1. INTRODUCTION

In many applications, it is beneficial to compress the source information into a scalable bitstream. This paper focuses on scalability in receiver signal-to-noise ratio (SNR), which allows decoding at different SNR quality levels. The scalable coding framework generates multiple streams at differing target rates, such that a lower information bitstream is embedded into a higher information bitstream in a way that minimizes redundancy. The lowest information layer is called the base layer (B-layer), while higher information layers are called enhancement layers (E-layers). Unlike the wasteful alternative of encoding each layer independently at the desired target rate [1], scalable coding exploits the redundancies across layers. In [2, 3, 4], each E-layer simply compresses the reconstruction error of the preceding (lower) layer. However, the information provided by previous reconstructed samples at the current E-layer is ignored. In [5], both the B-layer and E-layer predictions are based on the previous E-layer reconstructed samples, but since the B-layer decoder does not have access to E-layer reconstructed samples, this results in a drift between encoder and decoder at the B-layer. The estimation-theoretic (ET) approach was developed in our lab [6] to effectively combine all available information, wherein E-layer prediction is based on previous E-layer reconstructions, while also accounting for the quantization interval specified by the B-layer. Note that none of the above design techniques account for the effects of packet loss in the network. When predictive compression systems are deployed over lossy networks, a packet loss results in significant error propagation through the prediction loop and may seriously degrade the quality of the reconstructed signal. Another early contribution of our lab [7], in the area of video networking, is an optimal technique to iteratively estimate, at pixel level precision, the end-to-end distortion (EED) experienced at the decoder. The EED estimate is then used to optimize encoder decisions including switching between inter/intra-modes. For a single-layer (non-scalable) setting, we recently proposed in [8, 9] techniques to design a first order predictor and quantizer to minimize EED. This paper proposes a framework to accurately estimate and minimize EED for the scalable coder setting. Note that the scalable coder setting poses major additional EED challenges given the complex inter-layer dependencies, and the non-linear operations required for optimal prediction at enhancement layers, which must all be accounted for to exploit all available information.

Predictor and quantizer design for predictive coding, is an old problem for which many techniques have been developed [10, 11, 12]. The open-loop (OL) approach computes the prediction error training sequence, used to design the quantizer, based on the original source sequence, while the closed-loop (CL) approach uses decoder reconstructed samples as reference for prediction in an iterative manner. OL suffers from mismatch between the statistics used for training and statistics seen during operation. CL suffers from instability in the predictor and quantizer design procedure, often causing considerable performance degradation, especially at low bit rates. The asymptotic closed-loop (ACL) design [12], which we leverage in this work, employs a subterfuge to leverage the advantages of both OL and CL, where the prediction sequence is based on the reconstructed samples generated at a previous iteration. Hence, in each iteration the ACL quantizer is effectively designed in a stable OL fashion, while at convergence the reconstructed samples are unchanged from iteration to iteration, i.e., ACL is asymptotically equivalent to CL. Therefore, ACL design enjoys the stability of OL approach as well as ultimately optimizing the system for CL operation.

In summary, this paper proposes a novel technique for designing scalable coder predictors and quantizers at all layers, which accounts for potential packet loss by estimating and minimizing EED. The E-layer predictor effectively combines all available information by employing the ET paradigm, and ACL quantizer design is employed to circumvent CL instabilities. Experimental results in Section 5 show considerable performance gains over existing design methods, which grow with increase in packet loss rate. The remaining of this paper is organized as follows; the scalable coding framework and problem statement are discussed in Section 2. Relevant background is briefly reviewed in Section 3. Section 4 provides a detailed description of the proposed scalable coding framework. Experimental results and conclusions are presented in Sections 5 and 6, respectively.

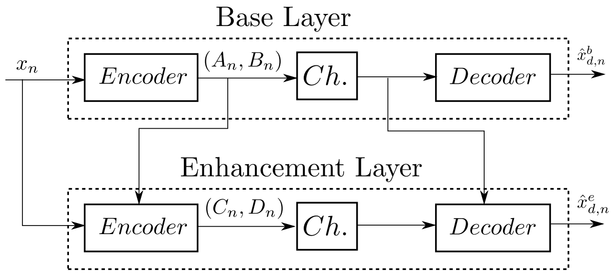

Fig. 1. High level Architecture of Scalable Coder.

2. PROBLEM STATEMENT

Without loss of generality, a two-layer predictive scalable coding framework is considered (as shown in Fig. 1). The source sequence is denoted by x_n, $0 \leq n < N$. The B-layer structure follows [8, 9] to account for potential packet loss. Let (A_n, B_n) be the quantizer interval corresponding to $\hat{e}^b_{e,n}$, the B-layer quantized prediction errors. Throughout this paper, superscripts specify the relevant layer: B-layer (b) or E-layer (e), and subscripts are used to specify location: encoder (e) or decoder (d). B-layer packets are assumed to be lost independently with probability p_b. If the packet is lost, an error concealment method is assumed where $\hat{e}^b_{e,n}$ is set to zero. The available information at the E-layer includes the B-layer current reconstructed sample $\hat{x}^b_{d,n}$, as well as the previous E-layer reconstructed samples $\left(\hat{x}^e_{d,n-1}, \hat{x}^e_{d,n-2}, \dots\right)$. The E-layer encoder combines the available information to generate the predicted sequence $\tilde{x}^e_{e,n}$. The prediction errors are quantized into $\hat{e}^e_{e,n}$ and the indices of E-layer quantizer intervals (C_n, D_n) are transmitted over the channel. The E-layer packets are dropped independently with probability p_e. The E-layer decoder combines the available information from the received packets to generate the reconstructed sequence $\hat{x}^e_{d,n}$. Considering the uncertainty due to packet loss, similar to [7], the EED at the E-layer can be computed according to

$$\begin{aligned}\mathbb{E}\{D\} &= \sum_{n=0}^{N-1} \mathbb{E}\left\{\left(x_n - \hat{x}^e_{d,n}\right)^2\right\} \\ &= \sum_{n=0}^{N-1} x_n^2 - 2x_n \mathbb{E}\left\{\hat{x}^e_{d,n}\right\} + \mathbb{E}\left\{\left(\hat{x}^e_{d,n}\right)^2\right\} \cdot\end{aligned} \tag{1}$$

Note that due to the lossy nature of the channel, $\hat{x}^e_{d,n}$ is viewed as a random variable by the encoder. The problem at hand is to design the scalable coder components (predictors and quantizers) to minimize EED experienced at each layer, while effectively utilizing all the information available at that layer.

3. RELEVANT BACKGROUND

3.1. Scalable Coding

In scalable coding, the B-layer structure is similar to a standard predictive compression system. The more challenging problem is that of how to combine the available information at the E-layer to minimize the reconstruction distortion. An optimal ET approach was proposed in [6] for the case of an ideal channel, where encoder and decoder are fully synchronized. The ET approach is optimal in the sense that it minimizes the mean square prediction and reconstruction errors. Given the B-layer current quantizer interval (A_n, B_n) and the previous E-layer reconstructed samples $\{\hat{x}^e_{n-1}, \hat{x}^e_{n-2}, \dots\}$, the predicted sequence at the E-layer is calculated according to

$$\tilde{x}^e_n = \mathbb{E}\left\{x_n | x_n \in \left(\tilde{x}^b_n + A_n, \tilde{x}^b_n + B_n\right), \hat{x}^e_{n-1}, \hat{x}^e_{n-2}, \dots\right\}, \tag{2}$$

where $\tilde{x}^b_n$ is the current sample prediction at B-layer. The E-layer prediction error is quantized, and let (C_n, D_n) be the E-layer quantizer interval. Thus, all the available information can be compactly represented by,

$$\begin{aligned}E_n &= \max\left[\tilde{x}^b_n + A_n, \tilde{x}^e_n + C_n\right], \\ F_n &= \min\left[\tilde{x}^b_n + B_n, \tilde{x}^e_n + D_n\right],\end{aligned} \tag{3}$$

$$x_n \in (E_n, F_n). \tag{4}$$

Therefore, the optimal E-layer reconstruction is obtained as,

$$\hat{x}^e_n = \mathbb{E}\left\{x_n | x_n \in (E_n, F_n), \hat{x}^e_{n-1}, \hat{x}^e_{n-2}, \dots\right\} \cdot \tag{5}$$

3.2. End-to-end distortion estimation and prediction

In [7], an iterative approach was proposed to accurately estimate the decoder distortion at pixel level precision in the presence of packet loss in video coding applications. EED (1) is computed for pixel j in frame n, based on the first and second moments of the decoder reconstruction, i.e., $\mathbb{E}\{x^j_{d,n}\}$, $\mathbb{E}\{(x^j_{d,n})^2\}$. In [8, 9] we showed how EED can be estimated and minimized for a first order linear predictive system when operating over unreliable channels. Specifically, the prediction at the encoder side is based on the first moment of the decoder reconstructed samples, i.e.,

$$\tilde{x}^b_{e,n} = \alpha_b \mathbb{E}\left\{\hat{x}^b_{d,n-1}\right\}. \tag{6}$$

The first and second moments were iteratively calculated accounting for possible packet loss as follows

$$\begin{aligned}\mathbb{E}\left\{\hat{x}^b_{d,n}\right\} &= (1-p_b)\hat{e}^b_{e,n} + \alpha_b \mathbb{E}\left\{\hat{x}^b_{d,n-1}\right\}, \\ \mathbb{E}\left\{\left(\hat{x}^b_{d,n}\right)^2\right\} &= (1-p_b)\left(\hat{e}^b_{e,n} + 2\alpha_b \mathbb{E}\left\{\hat{x}^b_{d,n-1}\right\}\right) + \\ &\quad \alpha_b^2 \mathbb{E}\left\{\left(\hat{x}^b_{d,n-1}\right)^2\right\},\end{aligned} \tag{7}$$

where α_b is the prediction coefficient. It was shown that the optimal prediction coefficient, minimizing EED, is

$$\alpha^*_b = \frac{\sum\limits_{n=0}^{N-1} \mathbb{E}\left\{\hat{x}^b_{d,n-1}\right\}\left(x_n - (1-p_b)\hat{e}^b_{e,n}\right)}{\sum\limits_{n=0}^{N-1} \mathbb{E}\left\{\left(\hat{x}^b_{d,n-1}\right)^2\right\}} \cdot \tag{8}$$

4. PROPOSED SCALABLE CODING FRAMEWORK

4.1. Base layer operation

To account for potential packet loss, the B-layer employed in our design is similar to the error resilient predictive compression system of [8]. The prediction coefficient used at the B-layer is calculated according to (8). The architecture of the B-layer is shown in Fig. 2. The ACL design technique is used at both layers to estimate the predictors and quantizers parameters.

4.2. Enhancement layer encoder operation

At the E-layer the encoder predictor that combines all the information provided by the current sample B-layer quantizer interval as well as the first moments of the previous E-layer reconstructed samples, is given by

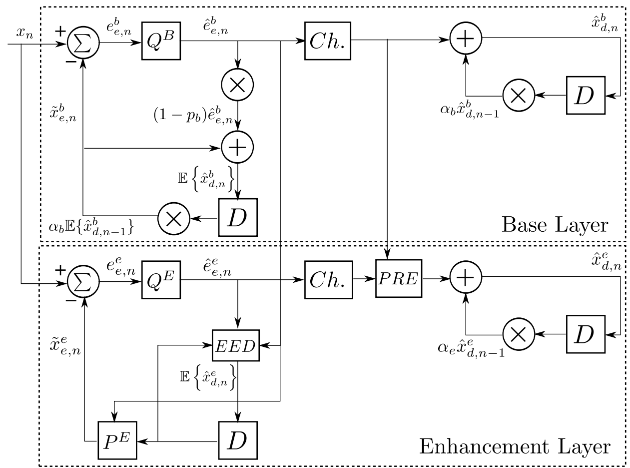

Fig. 2. Architecture of the Proposed Scalable Coder.

$$\tilde{x}^e_{e,n} = \mathbb{E}\Big\{ x_n | x_n \in \Big(\tilde{x}^b_{e,n}+A_n, \tilde{x}^b_{e,n}+B_n\Big), \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\}, \mathbb{E}\left\{\hat{x}^e_{d,n-2}\right\}, \dots \Big\}. \tag{9}$$

The expression in (9) can be rewritten as first order linear prediction based only on previous E-layer reconstructed sample, plus an innovation term as follows

$$\begin{aligned} \tilde{x}^e_{e,n} =& \alpha_e \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\} + \\ & \mathbb{E}\left\{k_{e,n} | k_{e,n} \in (L_1, R_1), \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\}, \mathbb{E}\left\{\hat{x}^e_{d,n-2}\right\}, \dots\right\}, \\ L_1 =& \tilde{x}^b_{e,n} + A_n - \alpha_e \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\}, \\ R_1 =& \tilde{x}^b_{e,n} + B_n - \alpha_e \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\}, \end{aligned} \tag{10}$$

where α_e is the E-layer linear prediction coefficient. The sequence $k_{e,n}$ is the encoder residual errors after linear prediction operation, i.e.,

$$k_{e,n} = x_n - \alpha_e \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\}. \tag{11}$$

The prediction errors $e^e_{e,n} = x_n - \tilde{x}^e_{e,n}$ are quantized, and the E-layer quantizer intervals (C_n, D_n) are obtained. The decoder receives the base and enhancement layers' packets with probabilities $(1-p_b)$, and $(1-p_e)$ respectively. Let

$$\begin{aligned} E_{e,n} &= \max\left[\tilde{x}^b_{e,n} + A_n, \tilde{x}^e_{e,n} + C_n\right], \\ F_{e,n} &= \min\left[\tilde{x}^b_{e,n} + B_n, \tilde{x}^e_{e,n} + D_n\right]. \end{aligned} \tag{12}$$

Thus, all the available information at the decoder can be summarized as

$$x_n \in (E_{e,n}, F_{e,n}) \cdot \tag{13}$$

Note that even if the current sample packets were received, the decoder is not guaranteed to compute correctly the interval $(E_{e,n}, F_{e,n})$ due to the possible prediction mismatch between encoder and decoder. The decoder is not expected to be perfectly synchronized with the encoder in unreliable networks.

4.3. Enhancement Layer decoder operation

The proposed E-layer reconstruction can be obtained according to

$$\begin{aligned} \hat{x}^e_{d,n} =& \mathbb{E}\left\{x_n | x_n \in (E_{d,n}, F_{d,n}), \hat{x}^e_{d,n-1}, \hat{x}^e_{d,n-1}, \dots\right\}, \\ E_{d,n} =& \max\left[\tilde{x}^b_{d,n} + A_n, \tilde{x}^e_{d,n} + C_n\right], \\ F_{d,n} =& \min\left[\tilde{x}^b_{d,n} + B_n, \tilde{x}^e_{d,n} + D_n\right], \end{aligned} \tag{14}$$

with

$$\begin{aligned} \tilde{x}^b_{d,n} =& \alpha_b \hat{x}^b_{d,n-1}, \\ \tilde{x}^e_{d,n} =& \alpha_e \hat{x}^e_{d,n-1} + \\ & \mathbb{E}\left\{k_{e,n} | k_{e,n} \in (L_2, R_2), \hat{x}^e_{d,n-1}, \hat{x}^e_{d,n-2}, \dots\right\}, \\ L_2 =& \tilde{x}^b_{d,n} + A_n - \alpha_e \hat{x}^e_{d,n-1}, \\ R_2 =& \tilde{x}^b_{d,n} + B_n - \alpha_e \hat{x}^e_{d,n-1}. \end{aligned} \tag{15}$$

Rewriting the reconstruction expression as a linear prediction term plus an innovation term we obtain

$$\begin{aligned} \hat{x}^e_{d,n} =& \alpha_e \hat{x}^e_{d,n-1} + \\ & \mathbb{E}\left\{k_{d,n} | k_{d,n} \in (L_3, R_3), \hat{x}^e_{d,n-1}, \hat{x}^e_{d,n-2}, \dots\right\}, \\ L_3 =& E_{d,n} - \alpha_e \hat{x}^e_{d,n-1}, \\ R_3 =& F_{d,n} - \alpha_e \hat{x}^e_{d,n-1}, \end{aligned} \tag{16}$$

where $k_{d,n}$ is the decoder residual errors after linear prediction operation, i.e.,

$$k_{d,n} = x_n - \alpha_e \hat{x}^e_{d,n-1}. \tag{17}$$

For the ease of notation denote the centroid of the conditional PDF of $k_{d,n}$ in the interval $(L - \alpha_e \hat{x}^e_{d,n-1}, R - \alpha_e \hat{x}_{d,n-1})$ as

$$\begin{aligned} \bar{k}^{(L,R)}_{d,n} =& \mathbb{E}\Big\{k_{d,n} | k_{d,n} \in \left(L - \alpha_e \hat{x}^e_{d,n-1}, R - \alpha_e \hat{x}^e_{d,n-1}\right), \\ & \hat{x}^e_{d,n-1}, \hat{x}^e_{d,n-2}, \dots\Big\}, \end{aligned} \tag{18}$$

where $L, R \in \mathbb{R}$. Therefore, if both layers' packets are received, combining (16) and (17) leads to

$$\hat{x}^e_{d,n} = \alpha_e \hat{x}^e_{d,n-1} + \bar{k}^{(E_{d,n}, F_{d,n})}_{d,n}. \tag{19}$$

Next we consider the operation of the decoder when packets are dropped. If the B-layer packet is lost, the interval $(E_{d,n}, F_{d,n})$ cannot be computed by the decoder, and a simple error concealment is assumed where $(E_{d,n}, F_{d,n})$ is set to $(-\infty, \infty)$, i.e., no information about x_n is available. If only the E-layer packet is lost, then $(E_{d,n}, F_{d,n})$ is set to $(\tilde{x}^b_{d,n} + A_n, \tilde{x}^b_{d,n} + B_n)$, which captures the information provided by the B-layer. Define

$$(A_{d,n}, B_{d,n}) = (\tilde{x}^b_{d,n} + A_n, \tilde{x}^b_{d,n} + B_n). \tag{20}$$

Hence, denoting "with probability" by $w.p.$, we write:

$$\hat{x}^e_{d,n} = \begin{cases} \alpha_e \hat{x}^e_{d,n-1} + \bar{k}^{(E_{d,n}, F_{d,n})}_{d,n} & w.p. \quad (1-p_b)(1-p_e) \\ \alpha_e \hat{x}^e_{d,n-1} + \bar{k}^{(A_{d,n}, B_{d,n})}_{d,n} & w.p. \quad (1-p_b)p_e \\ \alpha_e \hat{x}^e_{d,n-1} + \bar{k}^{(-\infty,\infty)}_{d,n} & w.p. \quad p_b \end{cases} \tag{21}$$

4.4. End-to-end distortion and prediction coefficient estimation

It is worth noting that the encoder does not have access to $\hat{x}^e_{d,n}$ in unreliable networks, and must treat it as a random variable. The first and second moments of $\hat{x}^e_{d,n}$ are investigated in the following analysis. Considering (21), the first moment of the E-layer reconstructed samples can be obtained as

$$\begin{aligned} \mathbb{E}\left\{\hat{x}^e_{d,n}\right\} =& \alpha_e \mathbb{E}\left\{\hat{x}^e_{d,n-1}\right\} + \mathbb{E}\left\{p_b \bar{k}^{(-\infty,\infty)}_{d,n}\right\} + \\ & \mathbb{E}\left\{p_e(1-p_b)\bar{k}^{(A_{d,n}, B_{d,n})}_{d,n}\right\} + \\ & \mathbb{E}\left\{(1-p_b)(1-p_e)\bar{k}^{(E_{d,n}, F_{d,n})}_{d,n}\right\}. \end{aligned} \tag{22}$$

The expectation of $\bar{k}_{d,n}^{(L,R)}$ over the statistics of the reconstructed E-layer samples can be approximated as

$$\mathbb{E}\left\{\bar{k}_{d,n}^{(L,R)}\right\} \approx \mathbb{E}\Big\{k_{d,n}|k_{d,n}\in\left(L-\alpha_e\mathbb{E}\left\{\hat{x}_{d,n-1}^e\right\},R-\alpha_e\mathbb{E}\left\{\hat{x}_{d,n-1}^e\right\}\right), \mathbb{E}\left\{\hat{x}_{d,n-1}^e\right\},\mathbb{E}\left\{\hat{x}_{d,n-2}^e\right\},\dots\Big\}, \quad (23)$$

which is a reasonable approximation if $\bar{k}_{d,n}^{(L,R)}$ is assumed to have locally linear dependence on previous reconstructed samples. For ease of notation, define

$$\hat{k}_{d,n}=p_b\mathbb{E}\left\{\bar{k}_{d,n}^{(-\infty,\infty)}\right\}+p_e(1-p_b)\mathbb{E}\left\{\bar{k}_{d,n}^{(A_{d,n},B_{d,n})}\right\} + (1-p_e)(1-p_b)\mathbb{E}\left\{\bar{k}_{d,n}^{(E_{d,n},F_{d,n})}\right\}. \quad (24)$$

Furthermore, the centroids $\bar{k}_{d,n}^{(L,R)}$ are assumed to be statistically uncorrelated with the previous E-layer reconstructed sample $\hat{x}_{d,n-1}^e$. Hence, the second moment of the E-layer reconstructed samples can be written as

$$\begin{aligned}\mathbb{E}\left\{\left(\hat{x}_{d,n}^e\right)^2\right\}=&\alpha_e^2\mathbb{E}\left\{\left(\hat{x}_{d,n-1}^e\right)^2\right\}+\mathbb{E}\left\{p_b\left(\bar{k}_{d,n}^{(-\infty,\infty)}\right)^2\right\}\\ &+\mathbb{E}\left\{p_e(1-p_b)\left(\bar{k}_{d,n}^{(A_{d,n},B_{d,n})}\right)^2\right\}\\ &+\mathbb{E}\left\{(1-p_b)(1-p_e)\left(\bar{k}_{d,n}^{(E_{d,n},F_{d,n})}\right)^2\right\}\\ &+2\alpha_e\mathbb{E}\left\{\hat{x}_{d,n-1}^e\right\}\hat{k}_{d,n}.\end{aligned} \quad (25)$$

The EED at the E-layer is given by

$$\mathbb{E}\left\{D\right\}=\sum_{n=0}^{N-1}x_n^2-2x_n\mathbb{E}\left\{\hat{x}_{d,n}^e\right\}+\mathbb{E}\left\{\left(\hat{x}_{d,n}^e\right)^2\right\}. \quad (26)$$

It follows that the prediction coefficient that minimizes the EED is

$$\alpha_e^*=\frac{\sum_{n=0}^{N-1}\mathbb{E}\left\{\hat{x}_{d,n-1}^e\right\}\left(x_n-\hat{k}_{d,n}\right)}{\sum_{n=0}^{N-1}\mathbb{E}\left\{\left(\hat{x}_{d,n-1}^e\right)^2\right\}}. \quad (27)$$

To obtain these initial results, dependence between $\bar{k}_{d,n}^{(L,R)}$ and α_e was neglected when deriving the expression in (27). While rigorous analysis of such possible dependence is currently underway, the validity of the approximations in this section is strongly supported by the experimental results of Section 5, where the proposed approach provides considerable performance gains. The architecture of the scalable coder is depicted in Fig. 2. The EED block computes $\mathbb{E}\left\{\hat{x}_{d,n}^e\right\}$ as shown in (22). The E-layer predictor is denoted as P^E and it operates according to (10). Additionally, the function of the pre-processing block at the E-layer decoder (PRE) is to compute $\bar{k}_{d,n}^{(L,R)}$, where (L,R) depends on the current channel event as shown in (21).

5. EXPERIMENTAL RESULTS

The experimental dataset consisted of the 6 speech files available in the EBU SQAM database [13]. About 75% of the speech files was used in the training phase and the remaining 25% was used as test data. It should be noted that although experimental results are presented for speech files, the approach is general and applicable to any source with memory. Lloyd's Algorithm was used to design the Entropy Constrained Scalar Quantizer (ECSQ). In our simulations, we compared the following three different scalable coders:

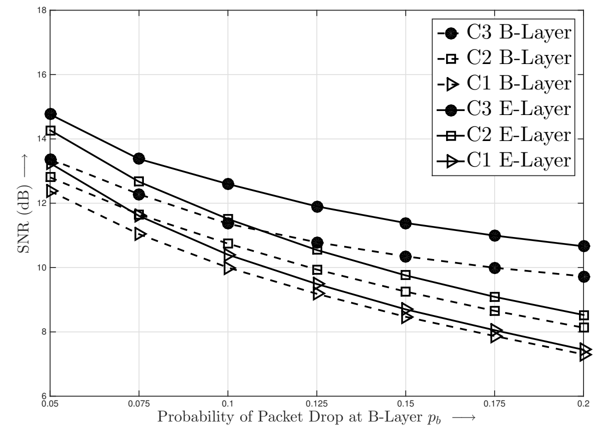

Fig. 3. SNR performance of the proposed scalable coder design (C3) compared to existing design techniques (C1 and C2). Dashed lines represent B-layer performance, while solid lines correspond to E-layer. The average rates at B-layer and E-layer for all coders are 0.8 bits/sample and 1.6 bits/sample, respectively.

- C1: The B-layer coder's predictor design completely ignores packet losses, similar to [11]. The E-layer coder employs direct quantization on the B-layer reconstruction errors, i.e, employ residual coding (e.g. [2, 3]). We consider CL approach to obtain the training sequence for designing ECSQs at both layers. This coder is the typical implementation of the scalable predictive coder.
- C2: The B-layer here is similar to C1. However, the E-layer operation follows the ET approach in [6]. ACL design approach is used at both layers. Hence, this coder adds ET and ACL components to C1.
- C3: The proposed approach discussed in Section 4. This coder adds EED estimation and minimization framework to C2.

The performance of the scalable coders is evaluated in terms of SNR observed at the decoder, thus accounting for quantization errors, packet loss and error propagation. The SNR is averaged over 20 different packet loss patterns. The loss patterns are generated independently for the two layers, while the probability of packet loss at E-layer layer is assumed to be twice the probability of packet loss at B-layer, i.e., $p_e = 2p_b$. The B-layer and E-layer quantizers in all scalable coders were designed to maintain 0.8 average bits per sample. The number of quantizer levels was allowed to vary, in order to achieve the target rate. Samples of the the sequences $k_{e,n}$ and $k_{d,n}$ were obtained during the training process; these samples were used to compute the conditional centroids in (10) and (18). Fig. 3 depicts the SNRs achieved by the competing methods versus packet loss rate. Clearly, the proposed approach consistently outperforms its competitors, offering up to 2.2 dB and 3.3 dB gains in SNR over C2 and C1, respectively. The SNR gains grow with increase in packet loss rate (p_b), as the ability to fully account for loss in the network becomes critical.

6. CONCLUSION

In this paper, a new design technique for predictive scalable coders is proposed, which effectively utilizes all available information at enhancement layers, circumvents instability challenges in the design of predictor and quantizer parameters, and incorporates optimal end-to-end distortion estimation to fully account for potential loss in the network. Experimental results show consistent and substantial gains over existing techniques, providing compelling evidence for the utility of the approach.

7. REFERENCES

[1] B. Girod, U. Horn, and B. Belzer, *Scalable Video Coding With Multiscale Motion Compensation And Unequal Error Protection*, Multimedia Communications and Video Coding, 1996.

[2] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Englewoods Cliffs, NJ: Prentice-Hall, 1984.

[3] R. Aravind, M.R. Civanlar., and A.R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Transactions on Circuits for Video Technology*, vol. 6, pp. 426–435, 1996.

[4] D. Wilson and M. Ghanbari, "Transmission of SNR scalable two layer MPEG-2 coded video through ATM networks," in *Proc. 7th Int. Workshop Packet Video*, 1997, pp. 426–435.

[5] M. Ghanbari and V. Seferidis, "Efficient H.261-based two-layer video codecs for atm networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 171–175, 1995.

[6] K. Rose and S.L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Transactions on Image Processing*, vol. 10, pp. 965–976, 2001.

[7] R. Zhang, S.L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 966–976, 2000.

[8] S. Zamani, T. Nanjundaswamy, and K. Rose, "Asymptotic closed-loop design of error resilient predictive compression systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.

[9] B. Vishwanath, T. Nanjundaswamy, S. Zamani, and K. Rose, "Deterministic annealing based design of error resilient predictive compression systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[10] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 kbits/s," *IEEE Transactions on Communications*, vol. 33, pp. 685–696, 1985.

[11] P.-C. Chang and R. Gray, "Gradient algorithms for designing predictive vector quantizers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 679–690, 1986.

[12] H. Khalil, K. Rose, and S.L. Regunathan, "The asymptotic closed-loop approach to predictive vector quantizer design with application in video coding," *IEEE Transactions on Image Processing*, vol. 10, pp. 15–23, 2001.

[13] G Waters, "Sound quality assessment material recordings for subjective tests," Tech. Rep., European Broadcasting Union (EBU), 1988.