Distortion Estimates for Adaptive Lifting Transforms with Noise

Fabio Verdicchio and Yiannis Andreopoulos*

Abstract

Multimedia signal processing algorithms often resort to adaptive transforms that exploit local characteristics of the input source. Following the signal decomposition stage, the produced transform coefficients and the adaptive transform parameters can be subject to quantization and/or data corruption (e.g. in a coding and transmission framework). As a result, mismatches between the analysis- and synthesis-side transform coefficients and adaptive parameters may occur, severely impacting the reconstructed signal. A thorough understanding of the quality degradation ensuing from such mismatches is essential for multimedia applications that rely on adaptive signal decompositions. This paper focuses on lifting-based adaptive transforms that represent a broad class of adaptive decompositions. By viewing the mismatches in the transform coefficients and the adaptive parameters as perturbations in the synthesis system, we derive analytic expressions for the expected reconstruction distortion. Our theoretical results are experimentally assessed using 1D adaptive decompositions and motion-adaptive temporal decompositions of video signals.

I. Introduction

The lifting scheme was initially introduced by Sweldens as a generalized construction of discrete wavelet transforms based on the factorization of the analysis (decomposition) or synthesis (reconstruction) polyphase matrix [1]. Recently, the lifting scheme became the vehicle for introducing signal-adaptive decompositions in a variety of coding frameworks [2]–[7]. Adaptive lifting decompositions are also used to capture edges and other directional features in image analysis [8], image enhancement [9] and in object detection [10]. Extensions to video signals apply adaptive temporal decompositions of the input sequences based on motion-adaptive prediction and update filters [6], [11]–[15].

The essential building block of lifting analysis (decomposition) is the cascade application of a predict step (using matrix $\mathbf{P}$) and an update step (using matrix $\mathbf{U}$) to the input signal. Adaptive lifting schemes in the literature [4][6][11]–[15] employ $\mathbf{P}$ and $\mathbf{U}$ matrices that perform signal-adaptive decomposition: the coefficients of each matrix are adaptively selected using signal-dependent criteria. The adaptive selection is signaled by a set of adaptive parameters [11]–[16]. When lifting synthesis is performed using lossless versions of both the decomposed

*Corresponding author. The authors are with the University College London, Dept. of Electronic & Electrical Engineering, Torrington Place, London WC1E 7JE, UK; Tel: +442076797303; Fax: +442073889325; e-mail: fverdicc@ee.ucl.ac.uk (F. Verdicchio), iandreop@ee.ucl.ac.uk (Y. Andreopoulos).

signal *and* the adaptive parameters produced by the analysis process, the input signal is perfectly reconstructed. However, in most practical application the quantization of the transform output (required to meet bandwidth or storage constraints) and the corruption of data (resulting from transmission errors and hardware faults) may impact both the decomposition coefficients *and* the adaptive lifting parameters that are available at synthesis side. As a consequence, lifting synthesis is performed using coefficients and adaptive parameters that differ from the analysis ones, thus deriving a distorted signal. As an application, consider a video stream produced by a scalable codec based on adaptive temporal decompositions [14][15]. In this case the input signal consists of a group of pictures (GOP). An adaptive lifting decomposition is performed in the temporal direction and derives the (estimated) motion trajectory of each pixel within each frame of the GOP. Therefore the lifting parameters also include the motion vectors indicating these trajectories [14][15]. Due to loss of motion-vectors and transform-coefficient data [17][18] during transmission, or due to quantization being applied on both [19][24], the decoder synthesizes the video sequence using erroneous or incomplete information.

I.A. Novel Contributions and Paper Organization

This work pursues the analytic characterization of the reconstruction error resulting from the synthesis of adaptive lifting transforms with noisy data, i.e. when the decomposition coefficients and the transform parameters are subject to both quantization and transmission errors. We begin by considering the adaptive lifting transform of 1D signals that constitutes the building block for a variety of applications to images [5][8][9] and video sequences [11]–[19]. We then extend our framework to motion-adaptive temporal transforms of video signals. Such transforms include the ones studied by [22]–[24]. Previous works [20]–[27] modelling the reconstruction errors of video coding schemes can be divided into two main groups. The first group [20]–[25] focuses on the rate-distortion aspects of scalable video coding schemes using spatiotemporal transforms. The second group [26][27] addresses system-specific features, such as the selection of the coding modes that minimize the decoding distortion in case of packet losses. Within the first group, there is research that extensively address the spatial transform [20][21], as well as thorough studies of the temporal transform [22][23]. In this paper consider noise-induced mismatches in any synthesis-side lifting parameter. This includes aspects that are neglected by [20]–[27] such as *(i)* the erroneous selection of the reference frame and *(ii)* the effect of arbitrary mismatches in the spatial displacements. Below we summarize our main contributions:

- Starting with the 1D case, we estimate the distortion in the synthesized signal considering additive noise sources (representing channel impairments and/or quantization) that affect *both* the transform coefficients and the adaptive parameters. Our results can be applied to noisy synthesis of any adaptive lifting scheme. Experimental validation is carried out using dyadic three-level lifting decompositions.

- We extend our approach to motion-adaptive temporal lifting synthesis of video. In this framework, the proposed distortion estimates retain the ability to account for the presence of noise in the transform coefficients and in the adaptive parameters. Specifically, we consider lifting parameter mismatches that impair *both* the selection of the reference frames *and* the relative spatial displacements (motion vectors) employed during synthesis. This is an aspect that, to our knowledge, has never been analytically studied before.

The paper is organized as follows. Section II introduces the notation and the mathematical formulation of the lifting synthesis with noise. Considering 1D signals, Section III derives the proposed synthesis distortion estimates. The extension of our approach to video systems is detailed in Section IV. The theoretical findings are then validated in Section V using both 1D signals and video sequences. Conclusions are drawn in Section VI.

II. Adaptive Lifting Scheme and Synthesis Mismatches

II.A. Notation and Definitions

All signals and filters are considered in the time or spatial domain. Boldface lowercase and uppercase letters indicate vectors and matrices respectively. For all signals and filters, superscripts indicate properties of the related quantities identifiable by the context (except for the superscript "T" that denotes transposition). Subscripts "even" and "odd" indicate the respective polyphase components. Notation $\hat{x}$ indicates the noisy version of the scalar x. It is applied similarly to vectors and matrices. Notations $\mathbf{Y}\{\mathrm{X}\}$, $\mathbf{y}\{\mathrm{X}\}$ and $y\{\mathrm{X}\}$ respectively indicate a matrix, a vector and a scalar that depend on $\mathbf{X}$ (the boldface notation is only applied, as appropriate, to y). The following definition is used extensively.

Definition 1: For a given $T \times T$ matrix $\mathbf{X}$, the $T \times T$ matrix $\mathbf{W}\{\mathrm{X}\}$ is defined as:

$$\mathbf{W}\{\mathrm{X}\} = \sum_{i=1}^{\mathrm{rank}\{\mathrm{X}\}} \left[\left(\varsigma_i \{\mathrm{X}\} \right)^2 \left(\mathbf{q}_i \{\mathrm{X}\} \mathbf{q}_i^{\mathrm{T}} \{\mathrm{X}\} \right) \right] \tag{1}$$

where the scalars $\varsigma_i \{\mathrm{X}\}$ and the $T \times 1$ vectors $\mathbf{q}_i \{\mathrm{X}\}$, with i=1,2,…,$\mathrm{rank}\{\mathrm{X}\}$, are respectively the *singular values* and the *right singular vectors* yielded by the singular value decomposition (SVD) of $\mathbf{X}$ [28]. The element at position (j,k) within $\mathbf{W}\{\mathrm{X}\}$ is denoted as $W^{(\mathrm{X})}[j,k]$. □

II.B. Lifting Synthesis of 1D Signals with Noise

Consider the adaptive decomposition of the $T \times 1$ input signal $\mathbf{x} = \left[x[0] \ \cdots \ x[T-1] \right]^{\mathrm{T}}$. The decomposed signal is the $T \times 1$ vector $\mathbf{x}^{\mathrm{u}}$ (comprising the low-frequency coefficients $\mathbf{x}^{\mathrm{u}}_{\mathrm{even}}$ and the high-frequency coefficients $\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}$). Most adaptive lifting schemes in the literature [4][6][11]–[15] use predict and update filters that are selected from a pre-determined set of N filters on the basis of signal-dependent criteria [15][16]. We indicate this selection by vector $\mathbf{a} = \left[\ a[0] \ \cdots \ a\left[T/2 - 1 \right] \ \right]^{\mathrm{T}}$ that identifies the filter-pair $a[t] \in \{0,\dots,N-1\}$ associated to each polyphase sample $x^{\mathrm{u}}_{\mathrm{even}}[t]$ and $x^{\mathrm{u}}_{\mathrm{odd}}[t]$, $t \in \{0,1,\ldots,T/2-1\}$. The predict and update filters have the “à-trous” structure with a unity tap placed at the position of the “current” sample [1], i.e.:

$$\mathbf{p}_{a[t]} = \left[p_{a[t]}[0] \;\; 0 \;\; \cdots \;\; p_{a[t]}[L^{\mathrm{p}}\text{-}3] \;\; 0 \;\; p_{a[t]}[L^{\mathrm{p}}\text{-}1] \;\; 1 \;\; p_{a[t]}[L^{\mathrm{p}}{+}1] \;\; 0 \;\; p_{a[t]}[L^{\mathrm{p}}{+}3] \;\; 0 \;\; \cdots \;\; p_{a[t]}[2L^{\mathrm{p}}] \right]^{\mathrm{T}} \tag{2}$$

$$\mathbf{u}_{a[t]} = \left[u_{a[t]}[0] \;\; 0 \;\; \cdots \;\; u_{a[t]}[L^{\mathrm{u}}\text{-}3] \;\; 0 \;\; u_{a[t]}[L^{\mathrm{u}}\text{-}1] \;\; 1 \;\; u_{a[t]}[L^{\mathrm{u}}{+}1] \;\; 0 \;\; u_{a[t]}[L^{\mathrm{u}}{+}3] \;\; 0 \;\; \cdots \;\; u_{a[t]}[2L^{\mathrm{u}}] \right]^{\mathrm{T}} \tag{3}$$

where L^{p} (respectively L^{u}) denote the maximum temporal span of the predict (respectively update) filter (see Table 1 for practical examples). The predict and update lifting operators $\mathbf{P}$ and $\mathbf{U}$ are given by the $T \times T$ matrices whose rows alternate between: *(i)* the unity sample on the main diagonal and *(ii)* the filters of (2)-(3) such that the unity filter tap is on the main diagonal. The adaptive lifting analysis of signal $\mathbf{x}$ is expressed as:

$$\mathbf{x}^{\mathrm{u}} = \mathbf{UPx}\ . \tag{4}$$

At synthesis side, the input signal $\mathbf{x}$ can be perfectly reconstructed from the transform coefficient vector $\mathbf{x}^{\mathrm{u}}$ and lifting matrices $\mathbf{P}$ and $\mathbf{U}$ as:

$$\mathbf{x} = \mathbf{P}^{-1}\,\mathbf{U}^{-1}\,\mathbf{x}^{\mathrm{u}}\,. \tag{5}$$

Based on the reversibility property of the lifting scheme [1], the synthesis matrices are:

$$\mathbf{P}^{-1} = 2\mathbf{I} - \mathbf{P} \quad , \quad \mathbf{U}^{-1} = 2\mathbf{I} - \mathbf{U} \tag{6}$$

with $\mathbf{I}$ the $T \times T$ identity matrix. However, when the lifting parameters are received erroneously (e.g. due to transmission noise), we have $\hat{a}[t] \neq a[t]$, for some t. Therefore the incorrect lifting synthesis matrices

$$\widehat{\mathbf{P}}^{-1} = 2\mathbf{I} - \widehat{\mathbf{P}} \quad , \quad \widehat{\mathbf{U}}^{-1} = 2\mathbf{I} - \widehat{\mathbf{U}} \tag{7}$$

are derived assuming (at synthesis side) the erroneous analysis matrices:

$$\widehat{\mathbf{P}} = \begin{bmatrix} \ddots & & & & \ddots & & & \ddots & & & \ddots & & & & \ddots \\ \cdots & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & p_{\hat{a}[t]}[0] & 0 & \cdots & p_{\hat{a}[t]}[L^{\mathrm{p}}\text{-}1] & 1 & p_{\hat{a}[t]}[L^{\mathrm{p}}{+}1] & \cdots & 0 & p_{\hat{a}[t]}[2L^{\mathrm{p}}] & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & p_{\hat{a}[t+1]}[0] & 0 & \cdots & p_{\hat{a}[t+1]}[L^{\mathrm{p}}\text{-}1] & 1 & p_{\hat{a}[t+1]}[L^{\mathrm{p}}{+}1] & \cdots & 0 & p_{\hat{a}[t+1]}[2L^{\mathrm{p}}] & 0 & \cdots \\ \ddots & & & & \ddots & & & \ddots & & & \ddots & & & & \ddots \end{bmatrix} \tag{8}$$

$$\widehat{\mathbf{U}} = \begin{bmatrix} \ddots & & & & \ddots & & & \ddots & & & \ddots & & & & \ddots \\ \cdots & 0 & u_{\hat{a}[t]}[0] & 0 & \cdots & u_{\hat{a}[t]}[L^{\mathrm{u}}\text{-}1] & 1 & u_{\hat{a}[t]}[L^{\mathrm{u}}{+}1] & \cdots & 0 & u_{\hat{a}[t]}[2L^{\mathrm{u}}] & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & u_{\hat{a}[t+1]}[0] & 0 & \cdots & u_{\hat{a}[t+1]}[L^{\mathrm{u}}\text{-}1] & 1 & u_{\hat{a}[t+1]}[L^{\mathrm{u}}{+}1] & \cdots & 0 & u_{\hat{a}[t+1]}[2L^{\mathrm{u}}] & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots \\ \ddots & & & & \ddots & & & \ddots & & & \ddots & & & & \ddots \end{bmatrix} \tag{9}$$

with $\hat{a}[t] \in \{0,\ldots,N-1\}$, $t = 0,\ldots,T/2-1$, and $\hat{a}[t] \neq a[t]$ for some $t \in \left\{0,\ldots,T/2-1\right\}$. In other words, during synthesis, *a different filter* than the one used during the analysis is selected for some time instances. In addition, the synthesis-side coefficient vector differs from its analysis-side counterpart due to quantization or transmission errors. Hence, the noisy vector $\hat{\mathbf{x}}^{\mathrm{u}} \neq \mathbf{x}^{\mathrm{u}}$ is available at synthesis-side. As a result, the incorrect signal $\hat{\mathbf{x}} \neq \mathbf{x}$ is synthesized as:

$$\hat{\mathbf{x}} = \left(2\mathbf{I} - \widehat{\mathbf{P}}\right)\left(2\mathbf{I} - \widehat{\mathbf{U}}\right)\hat{\mathbf{x}}^{\mathrm{u}} \ . \tag{10}$$

In the following section we characterize the synthesis error $\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x}$ in terms of the noise sources that affect the lifting parameters vector $\hat{\mathbf{a}}$ and the coefficient vector $\hat{\mathbf{x}}^{\mathrm{u}}$.

III. Distortion Estimate for 1D Lifting Synthesis With Noise

The predict step analysis, given by $\mathbf{x}^{\mathrm{p}} = \mathbf{P}\mathbf{x}$, and the subsequent update step analysis, given by $\mathbf{x}^{\mathrm{u}} = \mathbf{U}\mathbf{x}^{\mathrm{p}}$, are each equivalent to the linear system defined by the $T \times T$ lifting analysis matrix $\mathbf{M} \in \{\mathbf{P}, \mathbf{U}\}$ and by the $T \times 1$ vectors $\mathbf{x}$ and $\mathbf{v}$, respectively holding the input signal and the resulting transform coefficients, as follows:

$$\mathbf{v} = \mathbf{M}\,\mathbf{x} \ . \tag{11}$$

When one synthesis step is performed using erroneous adaptive parameters and a corrupted coefficient vector, the erroneous signal $\hat{\mathbf{x}}$ is reconstructed as:

$$\hat{\mathbf{x}} = \left(2\mathbf{I} - \widehat{\mathbf{M}}\right)\hat{\mathbf{v}} \tag{12}$$

where $\widehat{\mathbf{M}} = \mathbf{M} + \Delta\mathbf{M}$ is the erroneous lifting matrix, with $\Delta\mathbf{M} \in \{\Delta\mathbf{P}, \Delta\mathbf{U}\}$ representing the net effect induced by the erroneous synthesis-side parameter vector $\hat{\mathbf{a}}$, which ultimately causes the matrices of (7) to differ from the ones of (6). Similarly, the coefficient vector $\hat{\mathbf{v}} = \mathbf{v} + \Delta\mathbf{v}$ is affected by quantization or transmission errors[1].

Via simple algebraic manipulation of (12) the reconstruction error $\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x}$ is derived as:

$$\Delta\mathbf{x} = (2\mathbf{I} - \mathbf{M})\,\Delta\mathbf{v} - \Delta\mathbf{M}\,\mathbf{v} - \Delta\mathbf{M}\,\Delta\mathbf{v} \ . \tag{13}$$

The last equation shows the functional dependency of the synthesis error $\Delta\mathbf{x}$ with:

- the coefficients vector $\mathbf{v}$ and the adaptive matrix $\mathbf{M}$, which are determined by the adaptive lifting decomposition of the input signal.
- the noise sources $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$, which originate from quantization and transmission noise.

Observation 1: The term $\Delta\mathbf{M}\,\Delta\mathbf{v}$ in (13) accounts for the combined effect of the noise corrupting the transform coefficients ($\Delta\mathbf{v}$) and the noise affecting the analysis matrix ($\Delta\mathbf{M}$). Neglecting the term $\Delta\mathbf{M}\,\Delta\mathbf{v}$ in (13) yields the following approximation:

$$\Delta\mathbf{x} \cong (2\mathbf{I} - \mathbf{M})\,\Delta\mathbf{v} - \Delta\mathbf{M}\,\mathbf{v} \ . \tag{14}$$

The use of (14) significantly simplifies the analytic derivation of the distortion estimate pursued in this paper. We investigated the loss of accuracy incurred by the approximation of (14). Extensive experimental results, reported in Appendix A, show that $\left\|\Delta\mathbf{x}\right\|$ can be estimated using (14) with less than 10% average error for a variety of practical instantiations of $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$. We therefore employ the approximation (14) in the ensuing analysis. □

[1] In the case of predict step synthesis, the noise in the coefficient vector results from the previous synthesis of the update step.

Observation 2: We choose the mean squared error (MSE) as our distortion metric and derive the expected synthesis distortion $E\left\{\|\Delta\mathbf{x}\|^2\right\}\big/T$ from (14). Under the assumption that the noise sources $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$ are statistically independent stochastic processes and that $\Delta\mathbf{v}$ has zero mean, the following expression ensues:

$$\frac{E\left\{\|\Delta\mathbf{x}\|^2\right\}}{T} = \frac{E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^2\right\}}{T} + \frac{E\left\{\|\Delta\mathbf{M}\mathbf{v}\|^2\right\}}{T} \ . \tag{15}$$

In order to asses the applicability of the expression (15) in practical applications, when independence of $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$ is not guaranteed, we considered several instantiations of $\Delta\mathbf{v}$ and $\Delta\mathbf{M}$ that originate from quantization schemes and parameter erasures that are encountered in practice. Our results, reported in Appendix A, show that (15) approximates the observed data with less than 10% discrepancy on average. Therefore, we use (15) in this work as it provides a good tradeoff between model complexity vs. model accuracy. □

In Sections III.A and III.B we derive analytic expressions of the terms in (15) on the basis of the singular value decomposition (SVD) [28] of the lifting $(2\mathbf{I}-\mathbf{M})$ and perturbation $(\Delta\mathbf{M})$ matrices. These results are then used, in Section III.C, to express the distortion estimate for the lifting synthesis with noise.

III.A. Effect of Noise corrupting the Transform Coefficients: $E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^2\right\}\big/T$

Proposition 1 (SVD-based expression of $E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^2\right\}\big/T$*)*: The contribution of the noise process $\Delta\mathbf{v}$ to the expected lifting synthesis distortion of (15) is expressed as:

$$\frac{1}{T}E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^2\right\} = \frac{1}{T}\operatorname{tr}\left\{\mathbf{W}\{2\mathrm{I}\text{-}\mathrm{M}\}\,\mathbf{R}_{\Delta\mathrm{v}}\right\} \tag{16}$$

where $\mathbf{W}\{2\mathrm{I}\text{-}\mathrm{M}\}$ is given by (1) and $\mathbf{R}_{\Delta\mathrm{v}} = E\left\{\Delta\mathbf{v}\;\Delta\mathbf{v}^{\mathrm{T}}\right\}$ is the $T \times T$ autocorrelation matrix of $\Delta\mathbf{v}$.

Proof: See Appendix B. ■

Corollary 1: Assuming that $\Delta\mathbf{v}_{\mathrm{even}}$ and $\Delta\mathbf{v}_{\mathrm{odd}}$ are two mutually independent white wide-sense-stationary (WSS) processes, the synthesis distortion of (16) is:

$$\frac{1}{T}E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^2\right\} = \gamma_{\mathrm{e}}\{\mathrm{M}\}\frac{E\left\{\|\Delta\mathbf{v}_{\mathrm{even}}\|^2\right\}}{T/2} + \gamma_{\mathrm{o}}\{\mathrm{M}\}\frac{E\left\{\|\Delta\mathbf{v}_{\mathrm{odd}}\|^2\right\}}{T/2} \tag{17}$$

with $\gamma_{\mathrm{e}}\{\mathrm{M}\}$ and $\gamma_{\mathrm{o}}\{\mathrm{M}\}$ given by:

$$\gamma_{\mathrm{e}}\{\mathrm{M}\} = \frac{1}{T}\sum_{k=0}^{T/2-1} W^{(2\mathrm{I}\text{-}\mathrm{M})}[2k,2k] \tag{18}$$

$$\gamma_{\mathrm{o}}\{\mathrm{M}\} = \frac{1}{T}\sum_{k=0}^{T/2-1} W^{(2\mathrm{I}\text{-}\mathrm{M})}[2k+1,2k+1] \ . \tag{19}$$

Proof: See Appendix B. ■

The results of Proposition 1 and Corollary 1 yield estimates of the synthesis distortion introduced by noise in the transform coefficients. We remark that (16) and (17) require only:

- $\mathbf{W}\{2\mathrm{I}-\mathrm{M}\}$ or the ensuing scalars $\gamma_{\mathrm{e}}\{\mathrm{M}\}$ and $\gamma_{\mathrm{o}}\{\mathrm{M}\}$. These terms are completely known at analysis side as they depend solely on the analysis lifting matrix $\mathbf{M}$.
- the statistics of the noise process $\Delta\mathbf{v}$. When $\mathbf{R}_{\Delta \mathrm{v}}$ is available at analysis side, e.g. via statistical characterization of the quantization scheme and channel impairments, (16) is employed. When the noise sources $\Delta\mathbf{v}_{\mathrm{even}}$ and $\Delta\mathbf{v}_{\mathrm{odd}}$ can be considered mutually independent white WSS processes, then (17) applies. This requires only the knowledge of the noise power.

III.B. Effect of Noise corrupting the Synthesis Lifting Parameters: $E\left\{\|\Delta\mathbf{M}\mathbf{v}\|^2\right\}/T$

Proposition 2 (SVD-based expression of $E\left\{\|\Delta\mathbf{M}\mathbf{v}\|^2\right\}/T$*)*: The contribution of the stochastic process $\Delta\mathbf{M}$ to the expected synthesis distortion of (15) is:

$$\frac{1}{T}E\left\{\|\Delta\mathbf{M}\mathbf{v}\|^2\right\}=\frac{1}{T}\sum_{\eta=1}^{T/2}\left[\ \Pr(\eta)\ \mathrm{tr}\left\{\left(\mathbf{v}\,\mathbf{v}^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathcal{M}_\eta\right\}\right\}\ \right] \tag{20}$$

where $\Pr(\eta)$ is the probability that η out of $T/2$ synthesis lifting parameters are erroneous (i.e. $\hat{a}[t]\neq a[t]$ at η time instants) and, for any $\eta\in\left\{1,2,\ldots,T/2\right\}$:

$$\mathbf{W}\left\{\Delta\mathcal{M}_\eta\right\}=\sum_{\Delta\mathbf{M}\in\Delta\mathcal{M}_\eta}\left\{\Pr(\Delta\mathrm{M}\mid\eta)\,\mathbf{W}\{\Delta\mathrm{M}\}\right\} \tag{21}$$

where:

- The set $\Delta\mathcal{M}_\eta$ comprises all noise matrices $\Delta\mathbf{M}$ resulting from η mismatches in the lifting parameters. Notice that $\mathrm{rank}\{\Delta\mathbf{M}\}=\eta$ for any $\Delta\mathbf{M}\in\Delta\mathcal{M}_\eta$.
- For a given $\Delta\mathbf{M}\in\Delta\mathcal{M}_\eta$, $\mathbf{W}\{\Delta\mathrm{M}\}$ is given by (1) and $\Pr(\Delta\mathrm{M}\mid\eta)$ is the probability that η mismatches in the lifting parameters result in the given error matrix $\Delta\mathbf{M}$.

Proof: See Appendix B. ■

Proposition 2 derives the synthesis distortion induced by lifting parameters mismatches by linking:

- the output of the lifting analysis, i.e. the transform coefficients $\mathbf{v}$;
- the probability that η (out of $T/2$) synthesis lifting parameters differ from their analysis counterparts, which can be derived based on the channel impairments estimates.
- $\Pr(\Delta\mathrm{M}\mid\eta)$, which reflects particular mismatch patterns (e.g. as a result of grouping lifting parameters together in a certain packetization scheme) or accounts for the net effect of channel codes and unequal error protection strategies. For the simple case where any of the $N-1$ possible mismatches is equally likely to occur to each of the η positions affected lifting parameters, $\Pr(\Delta\mathrm{M}\mid\eta)=\left\{\binom{T/2}{\eta}(N(N-1))^\eta\right\}^{-1}$ for any $\Delta\mathbf{M}\in\Delta\mathcal{M}_\eta$.
- the average response of the system to η mismatches in the synthesis lifting parameters, which is represented by $\mathbf{W}\left\{\Delta\mathcal{M}_\eta\right\}$ given by (21). This matrix is the statistical average of $\mathbf{W}\{\Delta\mathrm{M}\}$ given by (1) for each $\Delta\mathbf{M}\in\Delta\mathcal{M}_\eta$, i.e. the average over all $\Delta\mathbf{M}$ resulting from η errors in the lifting parameters. The set $\Delta\mathcal{M}_\eta$

can be constructed off-line by considering all possible choices of the analysis and synthesis filters that lead to η mismatches. In practice, one can consider only a subset of $\Delta\mathcal{M}_\eta$ to derive an approximation of $\mathbf{W}\left\{\Delta\mathcal{M}_\eta\right\}$ and resort to bootstrapping techniques [29] to avoid bias. In our experiments, the derivation of $\mathbf{W}\left\{\Delta\mathcal{M}_\eta\right\}$ is performed off-line as it requires neither the knowledge of the input-dependent coefficient vector $\mathbf{v}$ nor the actual analysis matrix $\mathbf{M}$. The knowledge of the N supported lifting filters of (2) and (3) suffices.

III.C. Distortion Estimate for Lifting Synthesis and Extension to General Lifting Schemes

The following proposition considers the general case of noise corrupting the synthesis-side transform coefficients ($\hat{\mathbf{x}}^{\mathrm{u}} \neq \mathbf{x}^{\mathrm{u}}$) and the synthesis-side parameter vector ($\hat{\mathbf{a}}$), which affects both predict and update step.

Proposition 3 (Distortion Estimate for Lifting Synthesis with Noise): Assuming that $\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{even}}$ and $\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}$, i.e. the noise sources corrupting the even and odd polyphase components of the lifting analysis output vector, are independent white WSS processes and assuming that synthesis-side mismatches in the lifting parameters are independent and identically distributed with probability $\rho = \Pr(\hat{a}[t] \neq a[t])$, then the expected synthesis distortion is:

$$\frac{1}{T}E\left\{\|\Delta\mathbf{x}\|^2\right\} = \varphi_{\mathrm{e}}\{\mathrm{P},\mathrm{U}\}\frac{E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{even}}\|^2\right\}}{T/2} + \varphi_{\mathrm{o}}\{\mathrm{P},\mathrm{U}\}\frac{E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}}{T/2} + \frac{\psi\{\rho,\mathrm{P},\mathrm{x}^{\mathrm{p}},\mathrm{x}^{\mathrm{u}}\}}{T} \tag{22}$$

where $\varphi_{\mathrm{e}}\{\mathrm{P},\mathrm{U}\}$ and $\varphi_{\mathrm{o}}\{\mathrm{P},\mathrm{U}\}$ are given by:

$$\varphi_{\mathrm{e}}\{\mathrm{P},\mathrm{U}\} = 2\gamma_{\mathrm{e}}\{\mathrm{P}\}\gamma_{\mathrm{e}}\{\mathrm{U}\} \tag{23}$$

$$\varphi_{\mathrm{o}}\{\mathrm{P},\mathrm{U}\} = (2\gamma_{\mathrm{o}}\{\mathrm{U}\}-1)\gamma_{\mathrm{e}}\{\mathrm{P}\} + \gamma_{\mathrm{o}}\{\mathrm{P}\} + \xi\{\mathrm{P},\mathrm{U}\} \tag{24}$$

with $\gamma_{\mathrm{e}}\{\mathrm{P}\}$, $\gamma_{\mathrm{e}}\{\mathrm{U}\}$, $\gamma_{\mathrm{o}}\{\mathrm{P}\}$, $\gamma_{\mathrm{o}}\{\mathrm{U}\}$ as in (18)-(19) and $\xi\{\mathrm{P},\mathrm{U}\}$ and $\psi\{\rho,\mathrm{P},\mathrm{x}^{\mathrm{p}},\mathrm{x}^{\mathrm{u}}\}$ given by (B11) and (B15) in Appendix B.

Proof: See Appendix B. ■

Proposition 3 estimates the distortion of lifting synthesis with noise on the basis of:

- the power of the noise sources $\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{even}}$ and $\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}$ that affect the analysis output, e.g. due to quantization. Although quantization noise is not strictly white WSS, the estimate of (22) closely matches the measurements obtained using practical quantization schemes, as shown in Section V.A.
- $\varphi_{\mathrm{e}}\{\mathrm{P},\mathrm{U}\}$ and $\varphi_{\mathrm{o}}\{\mathrm{P},\mathrm{U}\}$, which act as gain factors in the response of the lifting system to noise in the synthesis-side transform coefficients. These terms depend solely on the analysis matrices $\mathbf{P},\mathbf{U}$ and account for the interaction of the (synthesis) predict and update steps.
- $\psi\{\rho,\mathrm{P},\mathrm{x}^{\mathrm{p}},\mathrm{x}^{\mathrm{u}}\}$, which represents the response of the lifting system to synthesis-side parameters mismatches. This term depends on:
    - o $\rho = \Pr(\hat{a}[t] \neq a[t])$, i.e. the probability that errors occur in the synthesis-side lifting parameters; when this probability is zero, $\psi\{0,\mathrm{P},\mathrm{x}^{\mathrm{p}},\mathrm{x}^{\mathrm{u}}\} = 0$;

- the noiseless transform coefficients $\mathbf{x}^{\mathrm{p}}$ and $\mathbf{x}^{\mathrm{u}}$ and the analysis matrix $\mathbf{P}$. They are all available at the analysis side;
- $\mathbf{W}\{\Delta\mathcal{P}_{\eta}\}$ and $\mathbf{W}\{\Delta\mathcal{U}_{\eta}\}$ given by (21), which are derived off-line from the filters of (2)-(3) as explained in Section III.C.

We remark that Proposition 3 can be generalized to any lifting decomposition that comprises more lifting stages [1]. Such decomposition is obtained cascading multiple stages, each comprising a pair of predict and update steps. Since each stage is defined by its own predict and update lifting matrices, the distortion induced by noise in the transform coefficients or due to parameter mismatches is given by Proposition 1 or Proposition 2 respectively. The estimated synthesis distortion can then be derived by extending Proposition 3, which accounts for the interaction between one pair of synthesis steps (e.g. via the term $\xi\{\mathrm{P},\mathrm{U}\}$), to multiple pairs. Since all adaptive lifting transforms from the literature use a single pair of lifting steps [2]-[16], we shall not pursue the extension of Proposition 3 to multiple lifting steps in this work. On the other hand, the case of multi-level decompositions is often encountered in practical applications and is discussed in the following.

In this paper we consider the common case of dyadic multi-level decompositions [1], where the lifting analysis is recursively applied on the low-frequency coefficients $\mathbf{x}_{\mathrm{even}}^{\mathrm{u}}$, each time generating a new decomposition level (comprising low- and high-frequency coefficients), until the desired number of decomposition levels is reached. Proposition 3 can be applied at the top (coarsest) level to derive the estimated synthesis distortion of the next level. Combining this estimated distortion (which characterizes the reconstructed low-frequency coefficients) with the estimated distortion affecting the high-frequency coefficients allows for extending Proposition 3 to all finer levels and eventually to the reconstructed signal. For each finer decomposition level, correlation may emerge in the noise that affects the low-frequency transform coefficients, as a result of the recursive synthesis process. This may reduce the accuracy of the estimate of Proposition 3. Nevertheless, for typical numbers of decomposition levels (e.g. up to four), the recursive application of Proposition 3 yields sufficiently-accurate estimates for the multi-level lifting synthesis distortion, as verified in Section 0.

IV. DISTORTION ESTIMATE FOR MOTION-ADAPTIVE TEMPORAL LIFTING SYNTHESIS

We extend our notation to describe a spatially-varying adaptive temporal decomposition of video. Input frames are indicated by $X[s,t]$ where $s=(r,c)$ represents the spatial location within the frame[2] and t is the time instant. The lifting decomposition produces frames $X^{\mathrm{p}}[s,t]$ after the prediction step and $X^{\mathrm{u}}[s,t]$ after the update step. An instantiation of predict-step analysis is depicted in Figure 1 (top left). In the example depicted in Figure 1, the frame $X[s,2t+1]$ is predicted from frames $X[s,2t]$ and $X[s,2t+2]$, where $t\in\{0,\dots,T/2-1\}$. The choice of

[2] Specifically, for a frame comprising $R\times C$ pixels, $s=(r,c)$, with $r\in\{0,1,\dots,R-1\}$ and $c\in\{0,1,\dots,C-1\}$.

the particular prediction filter is made from a predetermined filter set, such as the one given in (2), and the adaptation tracks the motion of pixel s between the three successive frames. Hence, apart from parameter[3] $a^{\mathrm{p}}[s,t] \in \{0,\ldots,N-1\}$ indicating the temporal filter choice for pixel s, we also need to indicate, for each tap of the prediction filter, the spatial displacement within the corresponding reference frame. Considering the example in the figure, the displacement within the previous frame is denoted as $d^{\mathrm{p}}_{-1}[s,t]$ and the displacement within the following frame is denoted as $d^{\mathrm{p}}_{1}[s,t]$. In general we denote as $d^{\mathrm{p}}_{j}[s,t]$ the spatial displacement within frame $X[s,2t{+}1-j]$, with $j \in \mathcal{J}$ and $\mathcal{J} = \{0,\pm1,\pm3,\ldots,\pm L^{\mathrm{p}}\}$. Upon completion of the predict step, the update step inverts the prediction-residual back to the reference position according to the update weights [13]. With respect to the above example, the corresponding update is shown in Figure 1 (bottom left). Similarly to the predict-step, the displacement parameter[4] $d^{\mathrm{u}}_{i}[s,t]$, $i \in \mathcal{I} = \{0,\pm1,\pm3,\ldots,\pm L^{\mathrm{u}}\}$, is used to identify the sample [associated with a tap of the update filter of (3)] within the frame $X^{\mathrm{P}}[s,2t+i]$.

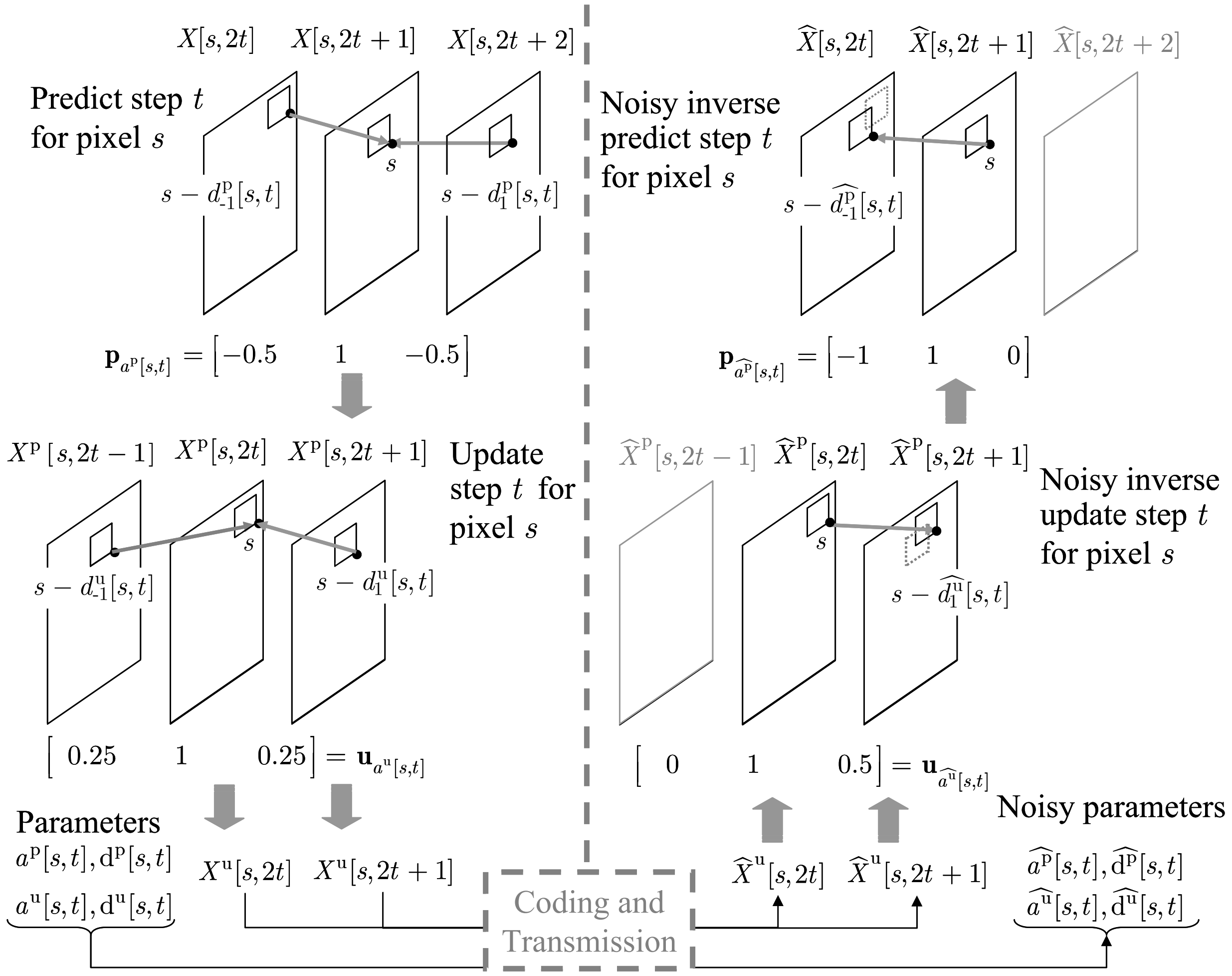

$X[s,2t]$ $X[s,2t+1]$ $X[s,2t+2]$
Predict step t for pixel s
s
$s - d^{\mathrm{p}}_{-1}[s,t]$
$s - d^{\mathrm{p}}_{1}[s,t]$
$\mathbf{p}_{a^{\mathrm{p}}[s,t]} = \begin{bmatrix} -0.5 & 1 & -0.5 \end{bmatrix}$
$\widehat{X}[s,2t]$ $\widehat{X}[s,2t+1]$ $\widehat{X}[s,2t+2]$
Noisy inverse predict step t for pixel s
s
$s - \widehat{d^{\mathrm{p}}_{-1}}[s,t]$
$\mathbf{p}_{\widehat{a^{\mathrm{p}}[s,t]}} = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix}$
$X^{\mathrm{p}}[s,2t-1]$ $X^{\mathrm{p}}[s,2t]$ $X^{\mathrm{p}}[s,2t+1]$
Update step t for pixel s
s
$s - d^{\mathrm{u}}_{-1}[s,t]$
$s - d^{\mathrm{u}}_{1}[s,t]$
$\begin{bmatrix} 0.25 & 1 & 0.25 \end{bmatrix} = \mathbf{u}_{a^{\mathrm{u}}[s,t]}$
$\widehat{X}^{\mathrm{p}}[s,2t-1]$ $\widehat{X}^{\mathrm{p}}[s,2t]$ $\widehat{X}^{\mathrm{p}}[s,2t+1]$
Noisy inverse update step t for pixel s
s
$s - \widehat{d^{\mathrm{u}}_{1}}[s,t]$
$\begin{bmatrix} 0 & 1 & 0.5 \end{bmatrix} = \mathbf{u}_{\widehat{a^{\mathrm{u}}[s,t]}}$
Parameters
$a^{\mathrm{p}}[s,t], \mathrm{d}^{\mathrm{p}}[s,t]$
$a^{\mathrm{u}}[s,t], \mathrm{d}^{\mathrm{u}}[s,t]$
$X^{\mathrm{u}}[s,2t]$ $X^{\mathrm{u}}[s,2t+1]$
Coding and Transmission
$\widehat{X}^{\mathrm{u}}[s,2t]$ $\widehat{X}^{\mathrm{u}}[s,2t+1]$
Noisy parameters
$\widehat{a^{\mathrm{p}}}[s,t], \widehat{\mathrm{d}^{\mathrm{p}}}[s,t]$
$\widehat{a^{\mathrm{u}}}[s,t], \widehat{\mathrm{d}^{\mathrm{u}}}[s,t]$

Figure 1. Spatially-varying adaptive temporal analysis (left) and noisy synthesis (right) using the lifting scheme. For the noisy inverse update and predict step, filter $\mathbf{u}_{\widehat{a^{\mathrm{u}}[s,t]}}$ and $\mathbf{p}_{\widehat{a^{\mathrm{p}}[s,t]}}$ are used. Differences in the displacement parameters $\widehat{d^{\mathrm{u}}_{1}}[s,t] \neq d^{\mathrm{u}}_{1}[s,t]$, $\widehat{d^{\mathrm{p}}_{-1}}[s,t] \neq d^{\mathrm{p}}_{-1}[s,t]$ and the modification of the temporal filter are due to noise in the transmission of the adaptive lifting parameters.

[3] In this section, all prediction and update lifting parameters are explicitly indicated by the superscripts p and u respectively.
[4] We let $d^{\mathrm{p}}_{0}[s,t] = 0$ and $d^{\mathrm{u}}_{0}[s,t] = 0$ for any s, t. In other words, the *current* sample requires no displacement.

As shown in Figure 1, the predict and update filters and their associated displacement parameters can be selected for different blocks [13]. The analysis of pixels $(s,2t)$ and $(s,2t+1)$, with $t \in \{0,1,\ldots,T/2-1\}$, is expressed as:

$$X^{\mathrm{P}}[s,2t] = X\,[s,2t] \quad ; \quad X^{\mathrm{P}}[s,2t+1] = \sum_{j\in\mathcal{J}} p_{a^{\mathrm{P}}[s,t]}[L^{\mathrm{P}}+j]\,X\left[s-d_j^{\mathrm{P}}[s,t]\,,2t+1+j\right] \tag{25}$$

$$X^{\mathrm{u}}[s,2t] = \sum_{i\in\mathcal{I}} u_{a^{\mathrm{u}}[s,t]}[L^{\mathrm{u}}+i]\,X^{\mathrm{p}}\left[s-d_i^{\mathrm{u}}[s,t]\,,2t+i\right] \quad ; \quad X^{\mathrm{u}}[s,2t+1] = X^{\mathrm{p}}[s,2t+1]\ . \tag{26}$$

The formulation of (25)-(26) is easily extended to include fractional displacements [13]. For notational simplicity we omit this case here. Fractional displacements are accounted for in the experimental validation of Section V.B.

The proposed methodology for lifting synthesis with noise is applied to video signals in Section IV.A. Then Section IV.B derives the distortion estimates for the synthesis of motion-adaptive temporal filtering with noise.

IV.A. Motion-Adaptive Temporal Lifting Synthesis with Noise

For any pixel s, the expressions of (25)-(26) can be given in matrix form analogous to the 1D case of (4). To this end, we denote the vector comprising the spatial location s within a group of T input frames as $\mathbf{x}[s] = \left[X[s,0]\ \ X[s,1]\ \cdots\ X[s,T\text{-}1]\right]^{\mathrm{T}}$. Similarly, we denote the frames produced by the predict step as $\mathbf{x}^{\mathrm{p}}[s] = \left[X^{\mathrm{p}}[s,0]\ \ X^{\mathrm{p}}[s,1]\ \cdots\ X^{\mathrm{p}}[s,T\text{-}1]\right]^{\mathrm{T}}$ and we let $\mathbf{x}^{\mathrm{u}}[s] = \left[X^{\mathrm{u}}[s,0]\ \ X^{\mathrm{u}}[s,1]\ \cdots\ X^{\mathrm{u}}[s,T\text{-}1]\right]^{\mathrm{T}}$ denote the output of update step. In order to utilize $\mathbf{P}$ and $\mathbf{U}$ to express motion-compensated lifting analysis, we need to identify the samples used to predict or update a given sample s. As a practical example, assume that the input signal comprises $T=4$ frames and consider the following instantiation of motion-compensated predict-step (25):

$$\begin{aligned} X^{\mathrm{P}}[s,0] &= X[s,0] \\ X^{\mathrm{P}}[s,1] &= -\frac{1}{2}X\left[s-d_{-1}^{\mathrm{P}}[s,0]\,,0\right] + X[s,1] - \frac{1}{2}X\left[s-d_{1}^{\mathrm{P}}[s,0]\,,2\right] \\ X^{\mathrm{P}}[s,2] &= X[s,2] \\ X^{\mathrm{P}}[s,3] &= -X\left[s-d_{-1}^{\mathrm{P}}[s,1]\,,2\right] + X[s,3]\,. \end{aligned} \tag{27}$$

If we neglect the displacement information in (27), $\mathbf{x}^{\mathrm{p}}[s] = \mathbf{P}\,\mathbf{x}[s]$ of (27) is:

$$\begin{bmatrix} X^{\mathrm{p}}[s,0] \\ X^{\mathrm{p}}[s,1] \\ X^{\mathrm{p}}[s,2] \\ X^{\mathrm{p}}[s,3] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \text{-}1/2 & 1 & \text{-}1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \text{-}1 & 1 \end{bmatrix} \begin{bmatrix} X[s,0] \\ X[s,1] \\ X[s,2] \\ X[s,3] \end{bmatrix}. \tag{28}$$

In order to incorporate the displacement information of (27) in the matrix formulation of (28), the input vector $\mathbf{x}[s]$ needs to be modified so that the appropriate displacements are considered when each predict operation occurs. In the example of (27), multiple samples belonging to *one frame* (i.e. *one element* of the vector $\mathbf{x}[s]$, e.g. $X[s,2]$), but placed at different spatial locations (i.e. $s\text{-}d_1^{\mathrm{P}}[s,0]$, s and $s\text{-}d_{-1}^{\mathrm{P}}[s,1]$ for the case of $X[s,2]$), are each involved in different predictions (the ones resulting in $X^{\mathrm{p}}[s,1]$, $X^{\mathrm{p}}[s,2]$ and $X^{\mathrm{p}}[s,3]$ respectively). Therefore each element of the input vector should contain the contributions of all the spatial locations, within the corresponding frame, that are involved in the predict step. To this end, a weighted superposition of the samples at different displaced locations is used. The weights are given by the discrete-time delta function $\delta[t_{\mathrm{cur}} - t]$, in order to select the appropriately displaced sample during the derivation of each predicted sample $X^{\mathrm{p}}[s, t_{\mathrm{cur}}]$, with $t_{\mathrm{cur}} \in \{0, \ldots, T\text{-}1\}$. The resulting input vector is denoted using the shorthand notation of $\mathbf{x}[s - \mathrm{d}^{\mathrm{p}}]$. Specifically, with respect to the example of (27), the third element ($t = 2$) of the input vector $\mathbf{x}[s - \mathrm{d}^{\mathrm{p}}]$ is given by: $X\left[s\text{-}d_1^{\mathrm{p}}[s,0], 2\right] \cdot \delta[t_{\mathrm{cur}}\text{-}1]$ $+ X[s,2] \cdot \delta[t_{\mathrm{cur}}\text{-}2]$ $+ X\left[s\text{-}d_{\text{-}1}^{\mathrm{p}}[s,1], 2\right] \cdot \delta[t_{\mathrm{cur}}\text{-}3]$. Using the same shorthand notation to the update step[5], the adaptive lifting analysis of pixel s, $s \in \{(0,0), \ldots, (R\text{-}1, C\text{-}1)\}$, is compactly expressed as:

$$\begin{aligned} \mathbf{x}^{\mathrm{P}}[s] &= \mathbf{P}\mathbf{x}[s - \mathrm{d}^{\mathrm{p}}] \\ \mathbf{x}^{\mathrm{u}}[s] &= \mathbf{U}\mathbf{x}^{\mathrm{p}}[s - \mathrm{d}^{\mathrm{u}}] \end{aligned} \tag{29}$$

where $\mathbf{P}$ and $\mathbf{U}$ are as in the 1D case with the replacement of $a[t]$ by $a^{\mathrm{p}}[s,t]$ and $a^{\mathrm{u}}[s,t]$, since (29) are applied per pixel (or per block).

When errors affect the received lifting parameters, $\widehat{\mathbf{a}}^{\mathrm{p}}[s]$ and $\widehat{\mathbf{a}}^{\mathrm{u}}[s]$, and spatial displacements, $\widehat{\mathrm{d}^{\mathrm{p}}}$ and $\widehat{\mathrm{d}^{\mathrm{u}}}$, as depicted on the right side of Figure 1, the lifting synthesis produces errors in the reconstructed video frames. For every pixel s, the synthesis process reconstructs the input sequence by cascading the following steps:

$$\begin{aligned} \hat{\mathbf{x}}^{\mathrm{p}}[s] &= \widehat{\mathbf{U}}^{-1} \hat{\mathbf{x}}^{\mathrm{u}}[s + \widehat{\mathrm{d}^{\mathrm{u}}}] \\ \hat{\mathbf{x}}[s] &= \widehat{\mathbf{P}}^{-1} \hat{\mathbf{x}}^{\mathrm{p}}[s + \widehat{\mathrm{d}^{\mathrm{p}}}] \end{aligned} \tag{30}$$

where $\widehat{\mathbf{P}}^{-1}$ and $\widehat{\mathbf{U}}^{-1}$, as given by (7), are the noisy lifting matrices and the vectors $\hat{\mathbf{x}}^{\mathrm{p}}[s + \widehat{\mathrm{d}^{\mathrm{p}}}]$ and $\hat{\mathbf{x}}^{\mathrm{u}}[s + \widehat{\mathrm{d}^{\mathrm{u}}}]$ are the noisy signals used to perform motion-adaptive synthesis. Therefore perfect reconstruction is hampered by the noise ensuing from three possible causes:

- The noisy transform coefficients $\hat{\mathbf{x}}^{\mathrm{u}}[s]$ (i.e. output frames) corrupted by quantization or transmission errors.
- The incorrect matrices $\widehat{\mathbf{P}}^{-1} = 2\mathbf{I} - \widehat{\mathbf{P}}$ and $\widehat{\mathbf{U}}^{-1} = 2\mathbf{I} - \widehat{\mathbf{U}}$ resulting from incorrect parameters $\widehat{\mathbf{a}}^{\mathrm{p}}[s], \widehat{\mathbf{a}}^{\mathrm{u}}[s]$.
- The incorrect spatial displacements $\widehat{\mathrm{d}^{\mathrm{p}}}, \widehat{\mathrm{d}^{\mathrm{u}}}$.

IV.B. Distortion Estimates and Displacement Mismatches for Motion-Adaptive Lifting Synthesis

Starting with (29)-(30) and following the line of reasoning of Section III, we can derive the expected distortion incurred by motion-compensated lifting synthesis across T frames. In the following we describe the key aspects that are specific of the video case. We consider the extension of Corollary 1 to motion-adaptive predict-step lifting synthesis. The equivalent analysis applies for the update lifting synthesis and is omitted for brevity of description.

Consider a spatial location $s \in \{(0,0), \ldots, (R\text{-}1, C\text{-}1)\}$. The expected predict-step synthesis distortion is:

$$\frac{1}{T} E\left\{\|\Delta\mathbf{x}[s]\|^2\right\} = \gamma_{\mathrm{e}}\{\mathrm{P}\} \frac{E\left\{\left\|\Delta\mathbf{x}_{\mathrm{even}}^{\mathrm{p}}\left[s + \widehat{\mathrm{d}^{\mathrm{p}}}\right]\right\|^2\right\}}{T/2} + \gamma_{\mathrm{o}}\{\mathrm{P}\} \frac{E\left\{\left\|\Delta\mathbf{x}_{\mathrm{odd}}^{\mathrm{p}}\left[s + \widehat{\mathrm{d}^{\mathrm{p}}}\right]\right\|^2\right\}}{T/2} \tag{31}$$

[5] Replacing $\mathbf{x}^{\mathrm{p}}[s]$ with $\mathbf{x}[s]$, swapping the role of $2t$ and $2t{+}1$, and replacing d^{p} with d^{u}.

where $\gamma_{\mathrm{e}}\{\mathrm{P}\}$ and $\gamma_{\mathrm{o}}\{\mathrm{P}\}$ are given by (18) and (19). Errors corrupting the synthesis-side coefficients, i.e. the frames $\hat{\mathbf{x}}^{\mathrm{p}} \neq \mathbf{x}^{\mathrm{p}}$ in the video case, contribute to the noise vector $\Delta\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]$, in analogy to the 1D case. In addition, the displacement mismatches $\widehat{\mathrm{d}^{\mathrm{p}}} \neq \mathrm{d}^{\mathrm{p}}$ introduce a contribution to the noise $\Delta\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]$ that is specific of the video case. We highlight these two contributions by rewriting the noise vector as:

$$\Delta\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}] = \hat{\mathbf{x}}^{\mathrm{p}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right] - \mathbf{x}^{\mathrm{p}}\left[s+\mathrm{d}^{\mathrm{p}}\right] = \Delta^{\mathrm{coef}}\mathbf{x}^{\mathrm{p}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right] + \Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right] \tag{32}$$

where the two noise terms in (32) are:

- The *(transform) coefficients noise* $\Delta^{\mathrm{coef}}\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}] = \hat{\mathbf{x}}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}] - \mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]$, which purely originates from the noise that affects the coefficients (i.e. frames) $\hat{\mathbf{x}}^{\mathrm{p}} \neq \mathbf{x}^{\mathrm{p}}$ at spatial location $s+\widehat{\mathrm{d}^{\mathrm{p}}}$.
- The *displacement noise* $\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}] = \mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}] - \mathbf{x}^{\mathrm{p}}[s+\mathrm{d}^{\mathrm{p}}]$, which is solely due to mismatches in the (synthesis-side) displacements $\widehat{\mathrm{d}^{\mathrm{p}}} \neq \mathrm{d}^{\mathrm{p}}$ within the noiseless frames $\mathbf{x}^{\mathrm{p}}$.

Observation 3: The power of the noise source $\Delta\mathbf{x}^{\mathrm{p}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right]$ of (32) is derived by simply adding the power of the coefficients noise and the displacement noise. We observed experimentally that (*i*) vectors $\Delta^{\mathrm{coef}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right]$ and $\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right]$ are nearly orthogonal and (*ii*) the power of $\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right]$ is almost independent of the displacement $\widehat{\mathrm{d}^{\mathrm{p}}}$. Moreover, the displacement $\widehat{\mathrm{d}^{\mathrm{p}}}$ is only applied to the even samples of the vector $\mathbf{x}^{\mathrm{p}}$, hence $\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}] = \mathbf{0}$. These remarks lead to the following expressions:

$$\frac{E\left\{\left\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2\right\}}{T/2} \cong \frac{E\left\{\left\|\Delta^{\mathrm{coef}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s]\right\|^2\right\}}{T/2} + \frac{E\left\{\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2\right\}}{T/2} \tag{33}$$

$$\frac{E\left\{\left\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2\right\}}{T/2} = \frac{E\left\{\left\|\Delta^{\mathrm{coef}}\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}}[s]\right\|^2\right\}}{T/2}. \tag{34}$$

The approximation of (33) incurs less than 10% error on average, as assessed in Appendix A experimentally. Hence, it reduces the complexity of our analysis without significant sacrifice in modelling accuracy. We incorporate (33) and (34) in our framework and experimentally verify, in Section V.B, that our estimate predicts the average synthesis distortion accurately, for various video sequences and noise conditions. □

The terms $E\left\{\left\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right]\right\|^2\right\}\Big/\left(T/2\right)$ and $E\left\{\left\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}}\left[s+\widehat{\mathrm{d}^{\mathrm{p}}}\right]\right\|^2\right\}\Big/\left(T/2\right)$ of (31) can be evaluated using (33) and (34) since these expressions separate the distortion contribution introduced by noise in the transform coefficients from the distortion induced by noisy displacement data. The quantization noise power in the transform coefficients can be computed using existing techniques, e.g. the distortion estimate proposed in [21] when quantization is applied in the spatial wavelet domain. Concerning the noise stemming from displacement mismatches, we make the following observation.

Observation 4: The power of the noise stemming from predict-step displacement mismatches can be expressed as follows:

$$\frac{E\left\{\left\|\Delta^{\text{disp}}\mathbf{x}^{\text{p}}_{\text{even}}[s+\widehat{\text{d}^{\text{p}}}]\right\|^2\right\}}{T/2} \cong \frac{2}{T}\sum_{t=0}^{T/2-1}\left[\mathcal{S}^{\text{p}}_{\mathcal{B}}\left(\text{d}^{\text{p}}[s,t]\right)\cdot \Pr\left(\widehat{\text{d}^{\text{p}}}[s,t] \neq \text{d}^{\text{p}}[s,t]\right)\right] \tag{35}$$

where $\Pr\left(\widehat{\text{d}^{\text{p}}}[s,t] \neq \text{d}^{\text{p}}[s,t]\right)$ is the probability that a displacement mismatch occurs (as a result of erroneously received motion data) during the predict-step synthesis of $X[s,2t{+}1]$. For this synthesis operation the term $\mathcal{S}^{\text{p}}_{\mathcal{B}}\left(\text{d}^{\text{p}}[s,t]\right)$, which represents the *block-based sensitivity* to incorrect displacements, is given by:

$$\mathcal{S}^{\text{p}}_{\mathcal{B}}\left(\text{d}^{\text{p}}[s,t]\right) = \sum_{\widehat{\text{d}^{\text{p}}}\neq\text{d}^{\text{p}}}\left[\frac{\mathcal{D}_B\left(\widehat{\text{d}^{\text{p}}}[s,t]\right)}{\mathcal{N}\left(\widehat{\text{d}^{\text{p}}}[s,t] \neq \text{d}^{\text{p}}[s,t]\right)}\right] - \mathcal{D}_B\left(\text{d}^{\text{p}}[s,t]\right) \tag{36}$$

where:

- $\mathcal{D}_B\left(\text{d}^{\text{p}}[s,t]\right)$ is the block-wise residual distortion (stemming from motion-compensated prediction of B samples of the frame $X[s,2t+1]$) that is associated to the displacement d^{p} used to perform analysis. Similarly $\mathcal{D}_B\left(\widehat{\text{d}^{\text{p}}}[s,t]\right)$ is the block-wise prediction-distortion associated to a displacement $\widehat{\text{d}^{\text{p}}} \neq \text{d}^{\text{p}}$. During the motion estimation phase and prior to lifting analysis, the distortion values associated with several displacements $\widehat{\text{d}^{\text{p}}}$ are computed whilst searching for the best-matching displacement d^{p}.
- $\mathcal{N}\left(\widehat{\text{d}^{\text{p}}}[s,t] \neq \text{d}^{\text{p}}[s,t]\right)$ is the number of candidate displacements (other than d^{p}) which are tested during motion estimation.

The derivation of (35)-(36) is given in Appendix A, the remainder of this section provides the necessary insight. □

The intuition behind (35) is that the energy of the error induced by displacement mismatches will depend on local signal characteristics [via the sensitivity term $\mathcal{S}^{\text{p}}_{\mathcal{B}}\left(\text{d}^{\text{p}}[s,t]\right)$] as well as on the mismatch probability over time, which reflects the channel impairments. The sensitivity term of (36) is derived by comparing, for the given block, the average distortion induced by all the displacements within the legitimate search range (all potentially used at synthesis-side in case of transmission errors) with the distortion induced by the displacement used to perform analysis. We remark that the block-based sensitivity of (36) is obtained as a by-product of block-based motion estimation that is typically used in practical systems [14][15]. As an example, the sensitivity of the blocks of one frame in the *Coastguard* sequence (using variable block-size motion estimation [12]) is given on the left side of Figure 2: dark shades of gray represent low sensitivity values, whereas light shades represent high sensitivity values. The blocks enclosed by dashed lines are highly sensitive and incur high distortion in case of synthesis with incorrect displacements. Conversely, the blocks in the upper and lower part of the picture exhibit low sensitivity. Hence, displacement mismatches during synthesis will result in low distortion for these areas. The comparison with the corresponding video frame (depicted on the right side of Figure 2) reveals that the blocks in the "high sensitivity" group correspond to dissimilar frame areas containing distinct features, whereas the blocks in "low sensitivity" group correspond to smooth areas in the frame with a lot of similar features (e.g. blocks in the water area of the video frame at the bottom).

The proposed estimate of (35) derives the distortion stemming from displacement mismatches in motion-compensated lifting synthesis. In the context of scalable motion vector coding [6][19][30] this could be used to model the impact to the reconstruction quality when transmitting a quantized version of the motion field. In this case, the sensitivity (36) could be computed by selecting the subset of incorrect displacements $\widehat{\mathrm{d}^{\mathrm{p}}} \neq \mathrm{d}^{\mathrm{p}}$ that correspond to a certain quantization interval for the motion parameters. Following the indications provided by our model, regions with low values of the sensitivity (36) can be identified and the respective motion vector field could be quantized more coarsely than the motion data relative to high sensitivity areas.

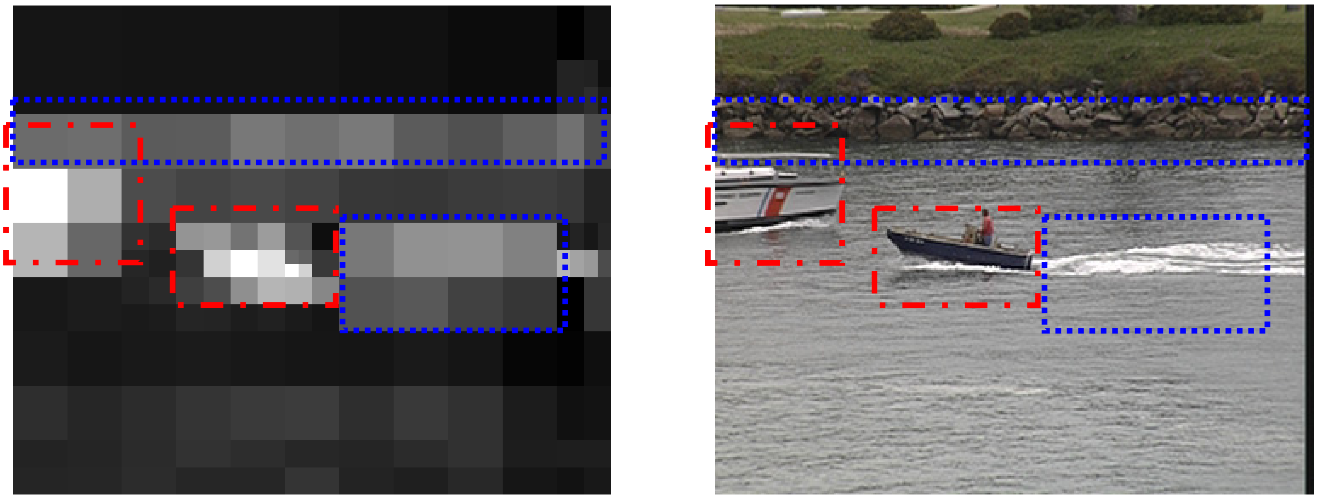


Figure 2. Block-based sensitivity (left) as given by (36) and corresponding video frame (right). Bright shades of gray correspond to high values, i.e. blocks which are highly sensitive to displacement mismatches during predict-step synthesis (areas enclosed by dashes). Dark shades correspond to low values, i.e. less sensitive blocks (top/bottom areas).

It is interesting to compare our block-based sensitivity of (36) with the sensitivity criterion introduced in [24] for a scalable video coding system featuring mesh-based motion prediction. The system of [24] incorporates a motion sensitivity factor that is derived, for each frame, from the power spectral density (PSD) of the entire frame. In turn, the PSD is estimated from the spatial discrete wavelet transform (DWT) employed by the coding system of [24]. The common treat between our block-based sensitivity of (36) and the spectral-based sensitivity of [24] is their ability to represent the characteristics of the video source. In either case, the sensitivity criterion matches the specific temporal lifting approach closely, i.e. block-based motion prediction for (36) vs. mesh-based prediction followed by spatial DWT for [24]. The two sensitivity metrics show complementary features such as spatial localization offered by (36) vs. spectral localization given by the one of [24].

The analysis of this section incorporates the effect of displacement mismatches, which are specific to the video case, in the distortion estimation framework introduced in Section III. Therefore distortion estimates, akin to those introduced for 1D signals, can be analytically derived for motion-compensated lifting synthesis with noise.

V. Experimental Results

In Section V.A, we assess the theoretical distortion estimates derived in Section III for 1D signals. In Section V.B, we compare our analytic estimates with distortion measurements relative to motion-compensated lifting synthesis of video. We then present an application to video streaming with unequal error protection.

V.A. 1D Signals

Throughout this section we assume the following setup:

- Several test input signals $\mathbf{x}$, each comprising $T = 256$ samples, are considered. They are taken from the horizontal and vertical lines of greyscale test images from the USC SIPI database.
- For each input signal, lifting analysis is performed and the lifting matrices $\mathbf{M} \in \{\mathbf{P}, \mathbf{U}\}$ and transform coefficients vector $\mathbf{x}^{\mathrm{u}}$ are derived (along with the intermediate predict-step output $\mathbf{x}^{\mathrm{p}}$). The analysis matrices $\mathbf{M} \in \{\mathbf{P}, \mathbf{U}\}$ are obtained by selecting one filter-pair out of the $N = 4$ pairs given in Table 1: for each pair of polyphase samples, the filter-pair minimizing the residual prediction energy is selected. The resulting list of $T/2$ filter indices forms the adaptive parameters vector $\mathbf{a}$. The filter set of Table 1 comprises filters that predict (or update) the current sample on the basis of *(i)* the value of either the previous or the following sample ($n = 0$ and $n = 1$) and *(ii)* via either linear or bilinear interpolation of both previous and following samples ($n = 2$ and $n = 3$). They are filters commonly used in adaptive lifting schemes [2][3][6][11]-[13].

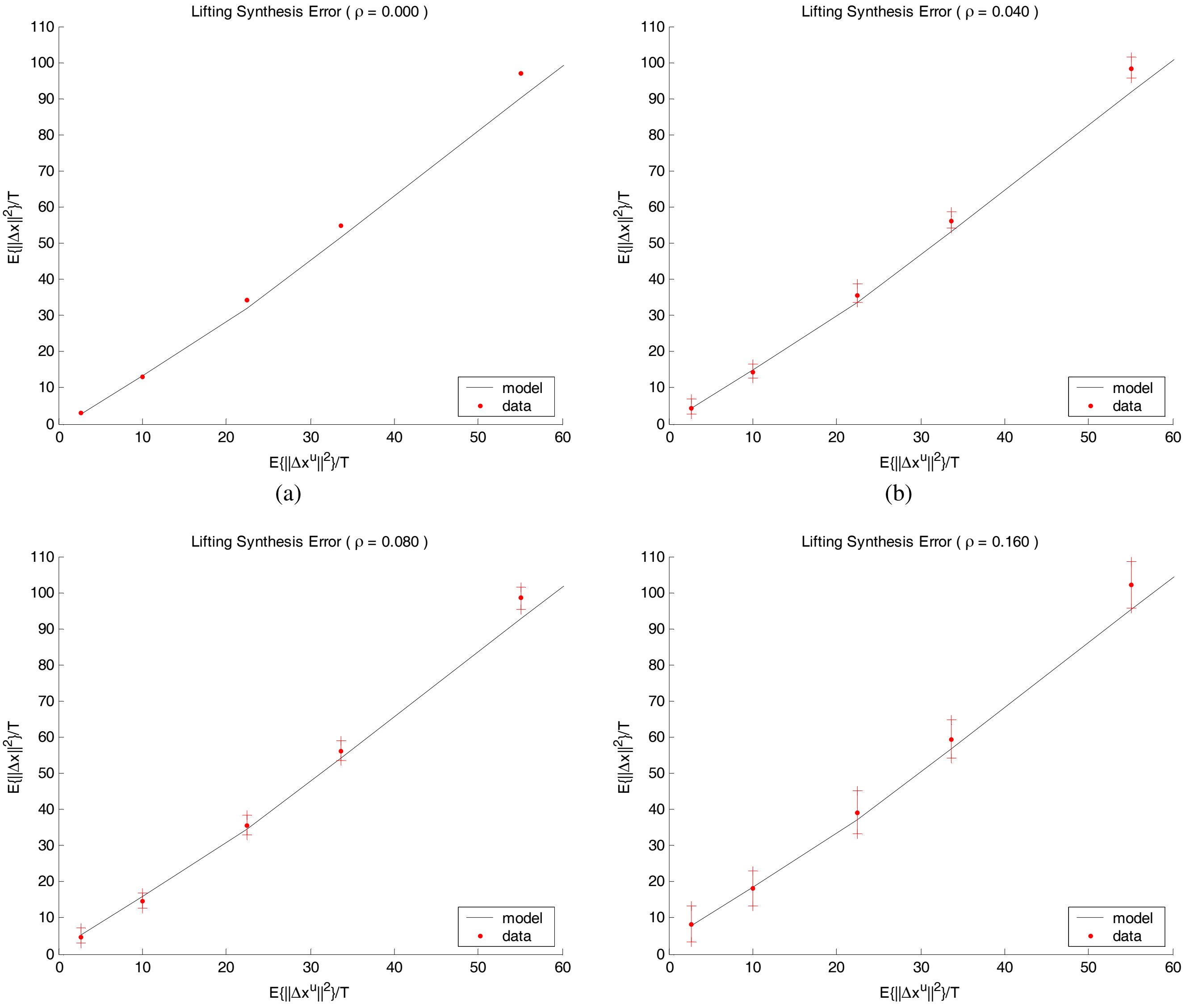
Table 1. Set of $N = 4$ pairs of predict and update lifting filters (2)-(3) with $L^{\mathrm{p}} = L^{\mathrm{u}} = 3$.

$a[t]$	$\mathbf{p}_{a[t]}$	$\mathbf{u}_{a[t]}$
0	$\begin{bmatrix} 0 & 0 & -1 & 1 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$	$\begin{bmatrix} 0 & 0 & 0 & 1 & 1/2 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$
1	$\begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$	$\begin{bmatrix} 0 & 0 & 1/2 & 1 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$
2	$\begin{bmatrix} 0 & 0 & -1/2 & 1 & -1/2 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$	$\begin{bmatrix} 0 & 0 & 1/4 & 1 & 1/4 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$
3	$\begin{bmatrix} 1/16 & 0 & -9/16 & 1 & -9/16 & 0 & 1/16 \end{bmatrix}^{\mathrm{T}}$	$\begin{bmatrix} -1/32 & 0 & 9/32 & 1 & 9/32 & 0 & -1/32 \end{bmatrix}^{\mathrm{T}}$

- The errors ($\Delta\mathbf{M}$) affecting the lifting matrices used during synthesis originate from perturbations applied to the parameters vector $\mathbf{a}$. For test purposes, we consider a uniform distribution of admissible perturbations where any of the $T/2$ adaptive parameters is equally likely to be affected with a given *mismatch probability* $\rho = \Pr\left(\hat{a}[t] \neq a[t]\right)$, for any $t \in \{0, \ldots, T/2 - 1\}$. When a parameter is affected, any of the $N - 1$ mismatches is equally likely to occur. For each value of ρ considered in the experiments, several perturbation patterns are drawn from this uniform distribution.
- Uniform scalar quantization (both with and without a double deadzone) is applied to the transform coefficients vector. Different quantization accuracies are obtained by scaling the width of the quantization bins dyadically.

V.A.(i) Distortion Estimate for Single-Level Lifting Synthesis

We focus on the adaptive lifting scheme synthesis comprising one pair of predict and update steps, which is given in (22). By selecting four representative values for the mismatch probability ρ, we perform lifting synthesis multiple times for each input signal $\mathbf{x}$, each time using increasingly-coarser quantized versions of the transform coefficients $\mathbf{x}^{\mathrm{u}}$. Figure 3 shows the synthesis distortion measured against the quantization-noise power (using dots) for each quantization accuracy of an indicative signal $\mathbf{x}$. When mismatches occur with probability $\rho > 0$, the average synthesis error power (taken over a set of 500 admissible perturbation patterns to $\mathbf{P}$ and $\mathbf{U}$) is indicated using dots and the standard deviation is shown using bars. Figure 3 shows the expected reconstruction distortion, as given by (22), using solid lines.



(a)

(b)

(c)

(d)

Figure 3. One example of average synthesis distortion vs. quantization noise power when (a) no mismatches in the adaptive parameters occur and when mismatches occur with: (b) 4%, (c) 8%, and (d) 16% probability. Dots denote the experimentally measured synthesis distortion (in the cases (b) to (d) the dot denotes the average distortion value taken over several mismatch patterns whereas bars indicate the standard deviation). The average distortion predicted by the proposed estimate is shown using solid lines. The range of synthesis distortion values shown in the figure varies between $\mathrm{SNR} \approx 50$ dB and $\mathrm{SNR} \approx 15$ dB (respectively bottom and top values on the vertical axis shown in the figure).

The plots of Figure 3 show that the theoretical estimate captures the trend of the experimental measurements successfully. As ρ increases, i.e. Figure 3(b)-(d), the distortion range associated to different perturbation patterns increases (vertical bars in the figure). However, the average values retain the quasi-linear behavior of Figure 3(a). The estimate of (22) determines the slope of this quasi-linear trend by the gain factors of (23) and (24). Hence, this slope is constant with ρ and can be determined solely on the basis of the analysis matrices $\mathbf{P}$ and $\mathbf{U}$, both available at encoding side. The results in Figure 3 confirm that the slope of the linear trend is independent of ρ. Moreover, the estimate of (22) successfully predicts the vertical offset $\psi\{\rho, \mathrm{P}, \mathrm{x}^{\mathrm{p}}, \mathrm{x}^{\mathrm{u}}\}$ for each value of ρ. We remark that the derivation of this offset requires information that is available at encoding-side along with a simple statistical characterization of the admissible perturbations to the adaptive parameters.

The experimental data reported in Figure 3 are in good agreement with the proposed estimate. In order to examine the accuracy of the distortion estimate of (22) over a large data set, we repeated the above experiments for a pool of 1000 signals and directly measured the behaviour of the synthesis distortion as a function of quantization power (five quantization accuracies were selected to span the range of distortion values shown in Figure 3). For each experimental instantiation, we compared the behaviour observed from the experimental data with the behaviour predicted by the estimate of (22). We then computed the correlation coefficient (R^2) and the average relative error between the experimentally observed and the model-predicted behaviour of the synthesis distortion. The correlation coefficient captures the similarities between the slope of the predicted curves and the trend of the observed data (as in Figure 3), but is insensitive to constant discrepancies such as large differences of the vertical offset. On the other hand, the average relative error does not capture local discrepancies in the slope, but detects large offset variations. The values of the correlation coefficient (R^2) and the average relative error resulting when averaging over the entire pool of experiments are given in Table 2.

Table 2. One-level lifting synthesis: Matching of the model-predicted vs. experimentally-measured distortion (1000 signals used)

ρ	R^2	Average Relative Error
0	0.9997	5.8 %
0.04	0.9995	6.0 %
0.08	0.9995	6.7 %
0.16	0.9993	8.4 %

The fact that $R^2 \approx 1$ for all values of the probability of mismatch indicates that the trend predicted by (22) always matches the experimentally observed behaviour closely. Although specific instantiations of the mismatch patterns may be overestimated or underestimated by the model (see Figure 3), the outcome of the extensive experiments given in Table 2 shows that the average discrepancy is below 10%. This suggests that (22), derived in Proposition 3 assuming white (quantization) noise, provides a good estimate of the synthesis distortion that remains accurate even when practical quantization schemes are involved.

V.A.(ii) Distortion Estimate for Dyadic Three-level Lifting Synthesis

For each decomposition level, Figure 4 reports the synthesis distortion vs. the power of the noise that affects the transform coefficient (due to quantization and error propagation through the coarser levels). Considering the case when no synthesis mismatches occur in the lifting parameters, the left plot of Figure 4 shows (using dots) the synthesis distortion measured at each decomposition level for one experimental instantiation. The estimated distortion, derived by applying (22) recursively, is also shown for each level (with a solid line). Similarly, the right plot of Figure 4 refers to the case when no mismatches occur at the top decomposition level and mismatches arise at the first and second decomposition levels with probability $\rho = 0.16$ (dots indicate the average value taken over several admissible mismatches and bars indicate the standard deviation where applicable). This example is in line with the idea that top-level lifting parameters, which are the least numerous and the most important, can be protected against errors more effectively than those of lower levels.

The graphs in Figure 4 show that, although increasingly correlated noise is fed from one level to the other as the recursive decomposition is synthesized (i.e. proceeding top to bottom in the figure), the prediction that is derived by recursively applying (22) captures the trend of the experimental data successfully. We seek further validation of these results by repeating the above experiments for several signals, as discussed previously for the case of a single decomposition level. The values of the correlation coefficient (R^2) and the average relative error for the three-level synthesis distortion of 1000 signals are reported in Table 3 and Table 4 for the case of mismatch-free and mismatched lifting parameters respectively.

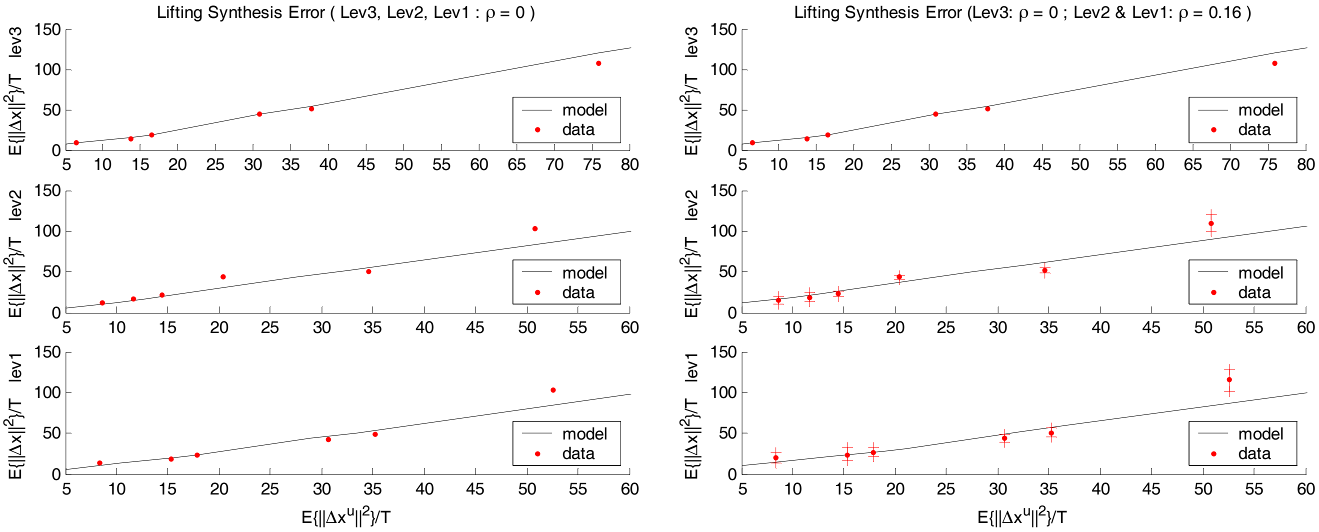


Figure 4. Example of average synthesis distortion vs. quantization noise power for dyadic three-level lifting decomposition when no mismatches occur (left) and when mismatches occur (right). Dots indicate the average error whereas bars indicate standard deviation. The average distortion predicted using the proposed estimate is shown using solid lines. The range of synthesis distortion values shown in the figure varies between $\mathrm{SNR}{>}50\ \mathrm{dB}$ and $\mathrm{SNR} \approx 15\ \mathrm{dB}$ (respectively bottom and top values on the vertical axis shown in the figure).

Table 3. Three-level lifting synthesis (with no mismatches): Matching of the model-predicted vs. experimentally-measured distortion (1000 signals used).

Decomposition Level	ρ	R^2	Average Relative Error
3	0	0.9978	7.3 %
2	0	0.9987	7.5 %
1	0	0.9996	8.1 %

Table 4. Three-level lifting synthesis (with mismatches): Matching of the model-predicted vs. experimentally-measured distortion (1000 signals used).

Decomposition Level	ρ	R^2	Average Relative Error
3	0	0.9978	7.3 %
2	0.04	0.9983	12.1 %
1	0.08	0.9992	9.1 %

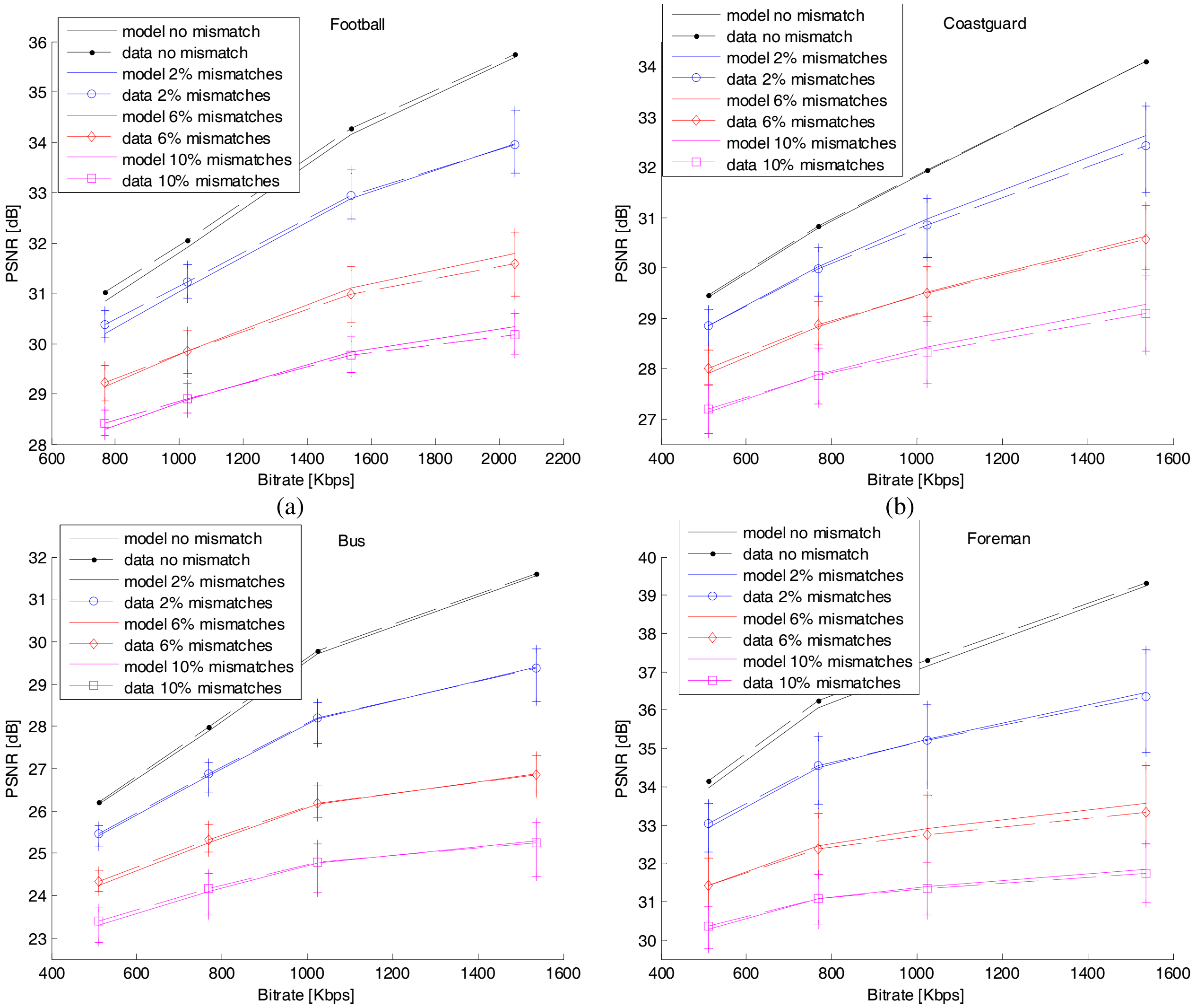
V.B. Video Signals

We employ the spatial-domain version of the scalable video codec of [12] using the following configuration:

- We perform multi-level temporal lifting decomposition of the video sequence featuring block-based motion estimation with two reference frames (corresponding to the predict filters of Table 1 with $n = 0,1$) and embedded quantization of the transformed video frames yielding seamless bitrate adaptation. Motion displacement is tracked up to quarter-pixel accuracy considering variable block-sizes, adaptively selected in the range of 2×2 to 64×64; further details on the motion estimation/compensation scheme, the entropy coding engine, and the rate allocation procedure can be found in [12][19].
- The only encoding modifications required are *(i)* the calculation of the sensitivity measurement of (36) as a by-product of motion estimation and *(ii)* the application of dequantization (inverse QTL [12]) for each extracted bitrate. The latter provides the quantization noise power $E\left\{\left\|\Delta^{\text{coef}}\mathbf{x}_{\text{even}}^{\text{p}}[s]\right\|^2\right\}$ and $E\left\{\left\|\Delta^{\text{coef}}\mathbf{x}_{\text{odd}}^{\text{p}}[s]\right\|^2\right\}$ of the video frames (of each temporal level) that is used to derive the estimate (31). Even though this power can be estimated per bitrate based on modelling [21], we opt to measure it experimentally since this requires only inverse-quantization that is a very low-complex process (no motion compensation or temporal synthesis performed). In this way we also avoid any bias that could be introduced by a rate-distortion model.
- The experiments reported below are performed using several common interchange format (CIF) video sequences recorded at 30 frames per second. We consider segments comprising $T = 24$ frames, corresponding to 0.8 seconds of video. This segmentation limits the propagation of decoding errors within the reconstructed sequence with minimal effect on the coding efficiency in the error-free scenario. It also provides estimates within frequent intervals of time, useful for a practical video processing and streaming server.

V.B.(i) Distortion Estimate for the Dyadic Three-level Temporal Synthesis

We select four bitrates and assume either no mismatch in the synthesis lifting parameters, or random mismatches occurring with 2%, 6% or 10% probability. Figure 5 reports the representative results obtained for one segment of four CIF sequences. For each bitrate, the peak-signal-to-noise ratio (PSNR) measurements of 100 decoding processes are considered, representing a variety of mismatches affecting different frames and spatial locations. The experimental averages per bitrate are indicated by markers and dashed lines, enclosed by vertical bars showing the observed range. The estimated distortion is shown using solid lines. Figure 5 shows the proposed estimates match the experimentally-measured average distortion closely. With a similar procedure to the one discussed Section V.A, we repeat the above experiments for 50 segments of $T = 24$ frames taken from four CIF sequences and report, in Table 5, the *correlation coefficient* (R^2) and the *average relative error* between the experimentally observed synthesis distortion and the behaviour predicted by our analytic estimate. The results demonstrate that the proposed distortion estimate is in very good agreement with the experimental observations.



(a) (b)

(c) (d)

Figure 5. Y-channel PSNR measurements (data) and theoretical estimate (model) for one segment of the sequences (a) *Football,* (b) *Coastguard,* (c) *Bus* and (d) *Foreman* each decoded at various bitrates and featuring synthesis mismatches occurring with various probabilities: no mismatch, 2%, 6% or 10% mismatches. When mismatches occur, the average PSNR value (marker) and the observed range (bars) are shown. The expected distortion (model) is shown with solid lines.

Table 5. Lifting synthesis of video: Matching of the model-predicted vs. experimentally-measured distortion (50 video segments of $T=24$ frames used).

	Football		**Coastguard**		**Bus**		**Foreman**	
$\Pr(\hat{\mathrm{d}} \neq \mathrm{d})$	R^2	Avg. Relative Error	R^2	Avg. Relative Error	R^2	Avg. Relative Error	R^2	Avg. Relative Error
0	1.0000	1.1 %	1.0000	0.1 %	1.0000	1.1%	0.9999	0.3 %
0.02	1.0000	1.4 %	1.0000	0.6 %	0.9999	1.6%	0.9995	2.1 %
0.06	0.9999	0.8 %	0.9999	1.0 %	0.9999	1.3%	0.9989	1.3 %
0.1	0.9998	0.9 %	0.9999	0.7 %	0.9997	0.9%	0.9972	0.7 %

V.B.(ii) Video Streaming Application

In the following experiments, our framework is used to estimate the decoding quality of video streams subject to time-varying packet-losses under different protection strategies. *The aim is to demonstrate how the proposed distortion model can predict the effect of different strategies in sender-driven error-resilient video streaming*.

Experimental setup: Using the bitstream extractor engine of the system [12], we form bitstreams providing progressive quality refinement, i.e. the video quality increases by progressively receiving and decoding more layers. The streams contain the same source data, but are assembled and channel coded following two different strategies, labelled as *Strategy1* and *Strategy2*. Each protection strategy comprises three and four unequally-protected layers respectively, as shown in Table 6 and Table 7 for the *Football,* and in Table 8 and Table 9 for the *Coastguard* sequences respectively. As shown in the tables, the layers contain both the lifting parameters (including the motion displacement[6]) and the transform coefficients (i.e. frames). Decoding an extra layer increases the knowledge of the lifting parameters (providing information relative to additional blocks) and allows refining the coefficients' quantization accuracy. Each layer is protected against packet losses using Reed-Solomon (RS) codes following an unequal error protection strategy [31]: lower layers are protected by stronger codes as they are mandatory to decode the information contained in higher layers. The protected stream is divided into temporal intervals (corresponding to 0.8 seconds of video) and is subject to time-varying packet losses. For each interval the reconstruction quality is then measured by the average PSNR. Both the experimental data and the theoretical estimate are reported in the upper part of Figure 6 and Figure 7 for the *Football* and *Coastguard* sequence respectively. The lower part of each figure shows the packet loss rate relative to each interval. During the decoding process, a layer is discarded whenever the packet loss rate has exceeded the error correcting capability of the code used to protect that layer. As a result, the quantization noise incurred by the transform coefficients varies depending on the number of available layers, which can be calculated based on the RS rate and the packet loss rate. Similarly, mismatches in the lifting operators occur when a layer containing the adaptive parameters is lost. Different reconstructions (100 decodings) are then obtained for the same loss-rate depending on the blocks affected by a loss. Figure 6 and Figure 7 show that our estimate matches the average experimental value for both strategies closely.

[6] The lifting parameters and motion data of the coarsest (highest) temporal decomposition levels are always within Layer 1.

The theoretical prediction tracks all variations of the source mismatch sensitivity and is robust to a broad range of PSNR values and loss rates. This is illustrated by comparing the results relative to intervals 1 and 5 or intervals 3 and 4 in Figure 6, which feature the same loss rate but show very different average PSNR. We conclude that, when the packet loss rate is known for a given interval, the proposed estimate can identify the strategy yielding the lowest expected distortion at the decoding side, thus guiding the sender on the layering strategy to use for each interval.

Table 6. Layers of “Strategy1” for *Football*.

Layer index	Layer bitrate (Kbps)	Layer breakdown (%)		Layer RS code rate
		Transform coefficients	Lifting parameters	
1	1015	71.5	28.5	0.75
2	290	98.3	1.7	0.9
3	540	99.6	0.4	0.95
4	517	100	0	0.99

Table 7. Layers of “Strategy2” for *Football*.

Layer index	Layer bitrate (Kbps)	Layer breakdown (%)		Layer RS code rate
		Transform coefficients	Lifting parameters	
1	1400	80	20	0.75
2	720	98.3	1.7	0.9
3	517	99.6	0.4	0.99

Table 8. Layers of “Strategy1” for *Coastguard*.

Layer index	Layer bitrate (Kbps)	Layer breakdown (%)		Layer RS code rate
		Transform coefficients	Lifting parameters	
1	680	86	14	0.75
2	280	99.8	0.2	0.9
3	270	99.9	0.1	0.95
4	516	100	0	0.99

Table 9. Layers of “Strategy2” for *Coastguard*.

Layer index	Layer bitrate (Kbps)	Layer breakdown (%)		Layer RS code rate
		Transform coefficients	Lifting parameters	
1	680	87.5	12.5	0.75
2	560	99.6	0.4	0.9
3	516	100	0	0.99

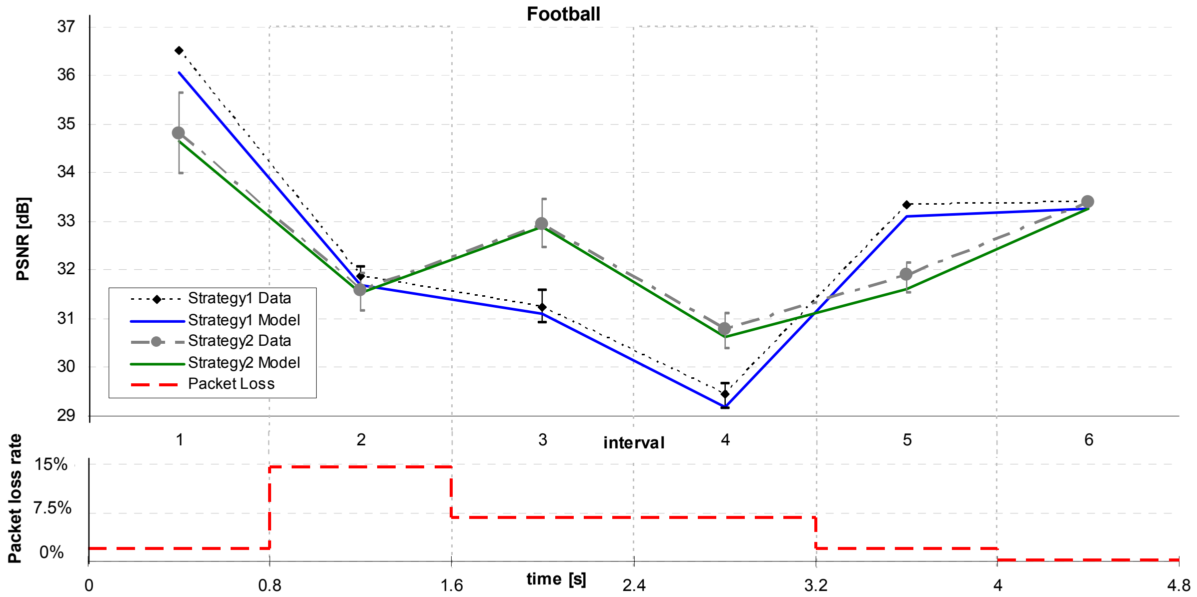


Figure 6. Simulation using the *Football* sequence. For each interval, the Y-channel PSNR experimental data and theoretical estimate are shown (top) along with the packet loss rate (bottom). Losses in layers comprising the lifting parameters result in different PSNR values: the average value (marker) and the observed range (bars) are reported.

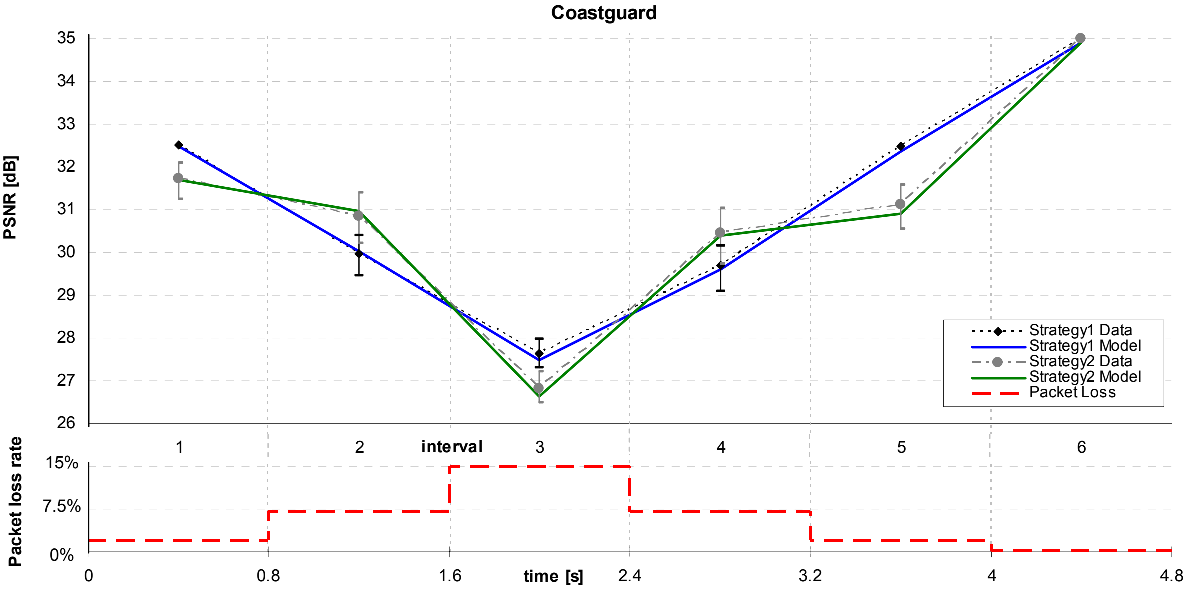


Figure 7. Simulation using the *Coastguard* sequence. For each interval the Y-channel PSNR experimental data and theoretical estimate are shown (top) along with the packet loss rate (bottom). Losses in layers comprising the lifting parameters result in different PSNR values: the average value (marker) and the observed range (bars) are reported.

Summary of application findings: The experimental results of this section demonstrate that the proposed theoretical framework is directly applicable to signal and video communications over error-prone networks. For example, video streaming servers can use the proposed framework to derive expectations of the receiver video quality for a given interval of a video stream based on the expected channel condition (packet loss rate). This can be very useful for Quality-of-Service environments where one needs to ensure appropriately high quality for a given set of clients (receivers) [17]. An interesting extension of the proposed framework would be in the design of optimal adaptive lifting decompositions under knowledge of noise conditions. Adaptive lifting is superior to non-adaptive lifting in a rate-distortion sense; however, under the consideration of transmission noise, non-adaptive lifting can be preferable. The proposed distortion estimates can be incorporated in future designs of adaptive lifting schemes as the evaluation mechanism to derive the appropriate level of signal-dependent adaptivity parametrical to the noise conditions.

VI. Conclusion

This paper presents a novel theoretical framework that characterizes the reconstruction error stemming when adaptive lifting-based transforms are synthesized using erroneous data. We considered the general case in which the adaptive parameters and the transform coefficients used during synthesis are affected by quantization noise and transmission errors. We approached the problem of noise in the synthesis of the adaptive transform from the standpoint of 1D signals and derived analytic estimates for the reconstruction error. This framework, suitable to describe a generic class of adaptive decompositions, was extended to motion-adaptive temporal lifting decompositions of video sequences. Our estimates were experimentally validated considering adaptive decompositions of both 1D and video signals under a variety of noise conditions. The method was also applied to layered video streams corrupted by time-varying packet losses. The results suggest that the proposed framework provides a useful mechanism to derive operational estimates for the average reconstruction error. Apart from the practical usefulness of the proposed approach in real-world signal and video transmission systems with unequal error protection, this work provides the means for a theoretical understanding of the trade-off between adaptivity in the lifting decomposition of a signal versus the robustness of the derived adaptive transform to noise.

References

[1] W. Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal*, vol. 3, nr. 2, pp. 186-200, 1996.

[2] Y. Wenxian, L. Yan, W. Feng, J. Cai, N. King and L. Shipeng, "4-D wavelet-based multiview video coding," *IEEE Trans Circ. Sys. Video Tech.*, vol. 16, no. 11, pp. 1385-1396, 2006.

[3] M. Flierl and B. Girod, "Multiview video compression," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 66-76, November 2007.

[4] G. Piella and H. J. A. M. Heijmans, "Adaptive lifting schemes with perfect reconstruction," *IEEE Trans. on Signal Process.*, vol. 50, no. 7, pp. 1620-1630, 2002.

[5] R. Claypoole, G. Davis, W. Sweldens and R. Baraniuk "Nonlinear wavelet transforms for image coding via lifting," *IEEE Trans Image Process.*, Vol. 12, No. 12, pp 1149-1459, 2003.

[6] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Trans. Image Process.*, vol 12, no 12, pp 1530-1542, 2003.

[7] A. Gouze, M. Antonini, M. Barlaud, B. Macq, "Design of signal-adapted multidimensional lifting scheme for lossy coding," *IEEE Trans. Image Process.*, vol. 13, pp. 1589–1603, 2004.

[8] G. Piella, B. Pesquet-Popescu, H. Heijmans and G. Pau, "Combining seminorms in adaptive lifting schemes and applications to image analysis and compression," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 2, pp. 203-226, 2006.

[9] H. Tasmaz and E. Ercelebi, "Image enhancement via space-adaptive lifting scheme exploiting subband dependency," *Digital Signal Processing*, 2010 (in press) doi:10.1016/j.dsp.2010.03.006.

[10] M. Amiri and H.Rabiee, "A novel rotation/scale invariant template matching algorithm using weighted adaptive lifting scheme transform," *Pattern Recognition*, vol 43, no. 7, pp. 2485–2496, 2010.

[11] V. Bottreau, B. Pesquet-Popescu, M. Bénetière, and B. Felts, "A fully scalable 3D subband video codec," *Proc. IEEE Int. Conf. Image Process.*, Thessaloniki, Greece, 2001.

[12] Y. Andreopoulos, *et al*, "In-band motion compensated temporal filtering," *Signal Process.: Image Comm.*, vol. 19, no. 7, pp. 653-673, 2004.

[13] J.-R. Ohm, "Advances in scalable video coding," *Proc. of the IEEE*, vol. 93, pp. 42-56, 2005.

[14] N. Adami, A. Signoroni, and R. Leonardi, "State-of-the-art and trends in scalable video compression with wavelet-based approaches," *IEEE Trans Circ. Sys. Video Tech.,* vol. 17, no. 9, pp. 1238-1255, 2007.

[15] R. Xiong; J. Xu; F. Wu, and S. Li, "Barbell-Lifting Based 3-D Wavelet Coding Scheme," *IEEE Trans Circ. Sys. Video Tech.*, vol. 17, no. 9, pp. 1256-1269, 2007.

[16] B. Girod and S. Han, "Optimum update for motion-compensated lifting," *IEEE Signal Process. Letters*, vol 12, no 2, pp150–153, 2005.

[17] Y. Andreopoulos, N. Mastronarde and M. van der Schaar, "Cross-layer optimized video streaming over wireless multi-hop mesh networks," *IEEE J. Select. Areas Comm.*, vol. 24, no. 11, pp. 2104-1215, 2006.

[18] F. Verdicchio, A. Munteanu, A. Gavrilescu, J. Cornelis, and P. Schelkens, "Embedded multiple description coding of video," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3114-3130, 2006.

[19] J. Barbarien, A. Munteanu, F. Verdicchio, Y. Andreopoulos, J. Cornelis and P. Schelkens, "Motion and texture rate-allocation for prediction-based scalable motion-vector coding," *Signal Process.: Image Comm.*, vol. 20, pp. 315-342, April 2005.

[20] M. Wang and M. van der Schaar, "Operational rate-distortion modeling for wavelet video coders," *IEEE Trans Signal Process.*, Vol. 54, No. 9, pp 3505-3157, 2006.

[21] B. Foo, Y. Andreopoulos and M. van der Schaar, “Analytical rate-distortion-complexity modeling of wavelet-based video coders,” *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 797-815, 2008.

[22] C.-L. Chang, A. Mavlankar, and B. Girod, “Analysis on quantization error propagation for motion-compensated lifted wavelet coding,” *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Shanghai, China, November 2005.

[23] T. Rusert, K. Hanke, and C. Mayer, “Enhanced interframe wavelet video coding considering the interrelation of spatio-temporal transform and motion compensation,” *Signal Process.: Image Comm.*, Vol 19, no 7, pp. 617-635, 2004.

[24] A. Secker and D. Taubman, “Highly scalable video compression with scalable motion coding,” *IEEE Trans. Image Process.*, vol 13, no 8, pp 1029-1041, 2004.

[25] R. Leung and D. Taubman, “Perceptual optimization for scalable video compression based on visual masking principles,” *IEEE Trans. Circuits Syst. Video Tech.,* vol 9, no. 3, pp. 309-322, 2009.

[26] S. Wan and E. Izquierdo, “Rate-distortion optimized motion-compensated prediction for packet loss resilient video coding,” *IEEE Trans. on Image Process.*, vol. 16, no. 5, pp. 1327-1338, May 2007.

[27] R. Zhang, S. Regunathan, K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” *IEEE J. Select. Areas Comm.*, vol. 18, no 6, pp. 966–976, 2000.

[28] V. Klema, A. Laub, “The singular value decomposition: Its computation and some applications,” *IEEE Trans Aut. Control*, vol. 25, no. 2, pp. 164-176, 1980.

[29] A. Zoubir, B. Boashash, “The bootstrap and its application in signal processing,” *IEEE Signal Process. Mag.*, vol. 15, no. 1, pp. 56-76, 1998.

[30] Y. Wu and J. W. Woods, “Scalable motion vector coding based on CABAC for MC-EZBC,” *IEEE Trans. Circuits Syst. Video Tech.*, vol. 17, no. 6, pp. 790-795, 2007.

[31] A. Mohr, E. Riskin, and R. Ladner, “Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction,” *IEEE J. Select. Areas Comm.*, vol. 18, pp. 819–828, 2000.

Distortion Estimates for Adaptive Lifting Transforms with Noise – *Support Document*

Fabio Verdicchio and Yiannis Andreopoulos[*]

APPENDIX A

A.1. Validation of Observation 1

Assuming the experimental settings described in Section V.A, we consider the approximation of (14), i.e. the relative impact of the term $\Delta\mathbf{M}\;\Delta\mathbf{v}$ in the synthesis error $\Delta\mathbf{x}$. The approximation of (14) involves only one lifting step, hence the experimental assessment is separately carried out for the predict and update step as follows. When $\mathbf{M}=\mathbf{P}$, the vector $\mathbf{x}$ comprises the samples of the input signal and the coefficient vector $\mathbf{v}=\mathbf{x}^{\mathrm{p}}$ holds the predict-step output. Conversely, when $\mathbf{M}=\mathbf{U}$, the input vector is $\mathbf{x}=\mathbf{x}^{\mathrm{p}}$ and the coefficient vector is given by $\mathbf{v}=\mathbf{x}^{\mathrm{u}}$. The objective is to assess the relative impact of neglecting the term $\Delta\mathbf{M}\;\Delta\mathbf{v}$ in the expression of $\Delta\mathbf{x}$ given by (13), hence we compute the ratio $\left\|\Delta\mathbf{M}\;\Delta\mathbf{v}\right\| / \left\|\Delta\mathbf{x}\right\|$ for several signals $\mathbf{x}$. Prior to synthesis, several perturbation patterns $\Delta\mathbf{M}$ are generated (each with a given mismatch probability ρ) and quantization is applied to the coefficient vector $\mathbf{v}$ (thereby inducing noise $\Delta\mathbf{v}$). This leads to a population of synthesis errors $\Delta\mathbf{x}$. Sample results are given in Figure A1 for several probabilities of mismatch $\rho\in\left[0.02\,,\,0.14\right]$. The graphs in the figure report both the average value of the relative approximation error $\left\|\Delta\mathbf{M}\;\Delta\mathbf{v}\right\| / \left\|\Delta\mathbf{x}\right\|$, using dots, and the standard deviation, using bars. As shown in the figure, the approximation of (14) incurs less than a 10% error on average.

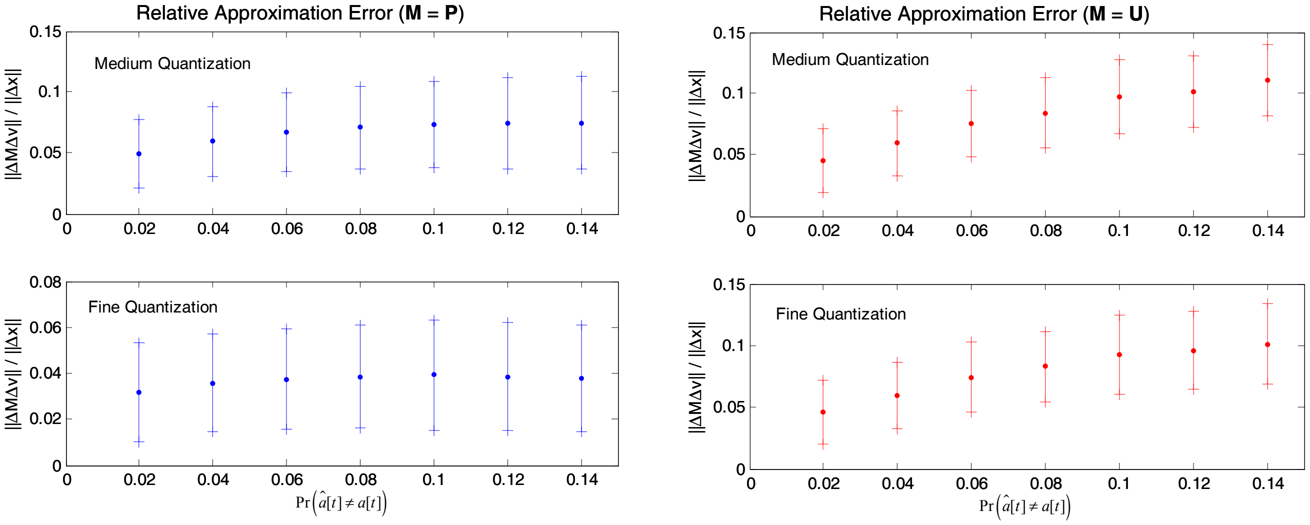


Figure A1. Relative error incurred by the approximation of (14) with $\mathbf{M}=\mathbf{P}$ (left) and $\mathbf{M}=\mathbf{U}$ (right) for the cases of medium and fine quantization of the transform coefficients. Several noise matrices $\Delta\mathbf{M}\in\{\Delta\mathbf{P},\Delta\mathbf{U}\}$ representative of different mismatch probabilities are considered: dots denote the average error whereas bars indicate standard deviation.

[*] Corresponding author. The authors are with the University College London, Dept. of Electronic & Electrical Engineering, Torrington Place, London WC1E 7JE, UK; Tel: +442076797303; Fax: +442073889325; e-mail: fverdicc@ee.ucl.ac.uk (F. Verdicchio), iandreop@ee.ucl.ac.uk (Y. Andreopoulos).

A.2. Validation of Observation 2

Assuming the experimental settings described in Section V.A, we carried out two complementary set of experiments, measured the average synthesis distortion $E\left\{\|\Delta\mathbf{x}\|^2\right\}/T$ and computed the *relative approximation error* incurred by the expression of (15), both when $\mathbf{M} = \mathbf{P}$ and $\mathbf{M} = \mathbf{U}$.

- In the first set of experiments, whose results are reported in Table A1, we consider noise signals $\Delta\mathbf{v}$ resulting from increasingly coarse quantization of the transform coefficients (this is indicated in Table A1 by the increasing values of Q, which represents the width of the scalar quantizer). Concerning the mismatches in the lifting parameters, which result in the noise matrix $\Delta\mathbf{M}$, we consider random mismatches occurring independently for any $t \in \{0,\ldots,T/2-1\}$ with probability $\rho = \Pr\left(\hat{a}[t] \neq a[t]\right)$. In essence, this scenario fits the case of *independent sources* $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$. Therefore the expression of (15) should represent the observed experimental data with good accuracy. This is confirmed by the results of Table A1, which demonstrate that the relative error incurred by the expression of (15) is 7% on average.
- The second set of experiments, whose results are reported in Table A2 below, investigates the *effect of correlation* among the mismatches in the lifting parameter and the quantization noise in the transform coefficients due to packet losses. During lifting analysis, we impose that the selection of the lifting filters is kept constant during four consecutive predict-and-update operations. Hence, one adaptive parameter tracks the adaptive decomposition of eight consecutive samples of the input signal. We form individual packets containing both the adaptive parameter and the quantized transform coefficients relative to each segment. During synthesis, the unavailability of one such packet implies that:
    - Four consecutive (and identical) mismatches occur in the adaptive parameter vector and in the resulting noise matrix $\Delta\mathbf{M}$.
    - The corresponding transform coefficients are approximated by the coarsest available representation (i.e. the DC component of the corresponding polyphase component).

The results of Table A2 show that relative error incurred by the expression of (15) is 9.5 % on average. This suggest that, although each lost packet induces noise samples that are placed at highly correlated *locations* within $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$, the *values* taken by the samples of the two noise sources are independent of each other. As a result, the weak statistical correlation between $\Delta\mathbf{M}$ and $\Delta\mathbf{v}$ does not induce a major deviation of the values predicted by (15) with respect to the observed data.

Table A1. Relative approximation error for *independent* mismatches and quantization noise

ρ	Q=20	Q=24	Q=28	Q=32	Q=36
0.02	4.8%	3.9%	3.3%	2.6%	2.3%
0.05	7.8%	6.6%	5.4%	4.5%	3.8%
0.1	13.3%	11.6%	9.8%	8.1%	7.0%
0.15	17.4%	15.3%	13.0%	10.9%	9.5%

Table A2. Relative approximation error for *correlated* mismatches and quantization noise

ρ	Q=20	Q=24	Q=28	Q=32	Q=36
0.02	9.2%	7.1%	4.9%	3.5%	2.7%
0.05	12.2%	9.9%	7.0%	5.1%	3.9%
0.1	22.4%	17.6%	12.8%	9.5%	7.4%
0.15	26.2%	22.4%	16.4%	12.3%	9.7%

A.3. Validation of Observation 3

Assuming the experimental settings described in Section V.B, we generate random mismatched spatial displacements $\widehat{\mathrm{d}^{\mathrm{p}}} \neq \mathrm{d}^{\mathrm{p}}$, which occur within each frame with a certain probability. We then measure the relative error incurred by the approximation (33) for a wide range of quantization accuracies. Sample results obtained using the *Football* sequence and 5% and 10% mismatch probability are shown in Figure A2. Each curve represents one instantiation of this experiment. As shown in the figure, the approximation (33) incurs errors in the order of 10% in the worst case. We notice that, independently of the mismatched displacement $\widehat{\mathrm{d}^{\mathrm{p}}}$, the approximation of (33) is more accurate when fine-scale quantization is applied. On the other hand, as the overall noise increases due to increasingly coarse quantization, the approximation of (33) progressively overestimates the overall noise power. This behavior agrees with the intuition that the distortion induced by displacement mismatches is “masked” by high quantization noise. In other words, the effect of pointing to the incorrect spatial location within a certain area of the frame is less evident when that area is coarsely quantized.

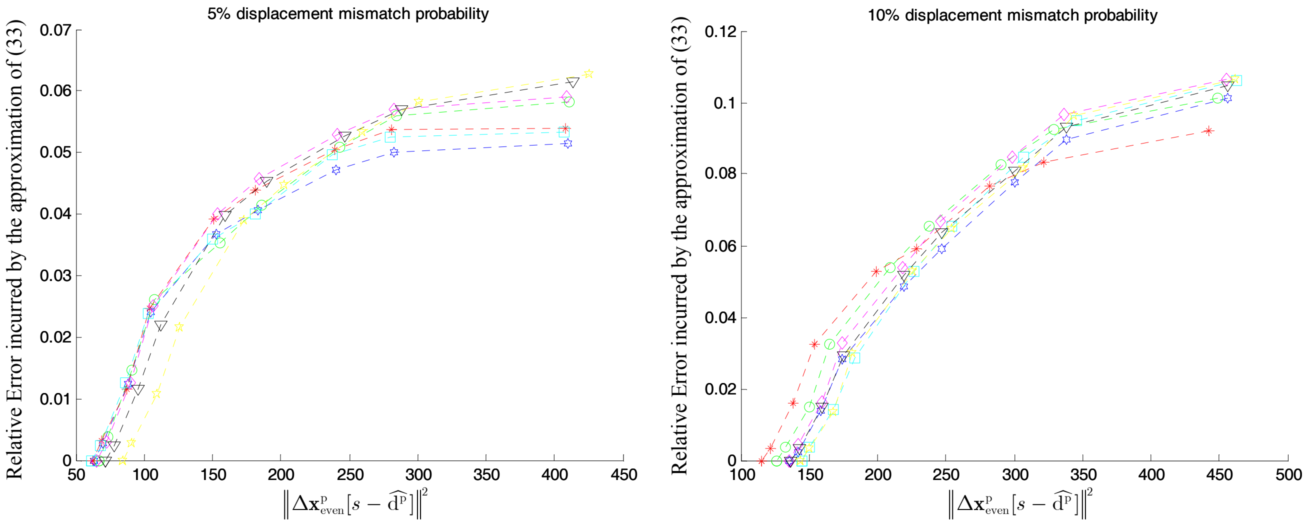


Figure A2. Relative error incurred by the approximation of (33) for a range of quantization accuracies. Displacement mismatches occur with 5% (left) or 10% probability (right). Different markers denote experiments with different mismatches $\widehat{\mathbf{d}^{\mathrm{p}}} \neq \mathbf{d}^{\mathrm{p}}$.

A.4. Derivation of Observation 4

The following formal definitions are used in subsequent derivations of the expressions (35) and (36).

First, we denote as $\mathcal{D}_B\left(\mathrm{d}^{\mathrm{p}}[s,t]\right)$ the *block-based multi-reference prediction distortion*, i.e. the mean squared error ensuing when predicting a block of B samples, comprising $X[s,2t+1]$, using the prediction filter $p_{a^{\mathrm{p}}[s,t]}$ and the displacement vector $\mathrm{d}^{\mathrm{p}}[s,t]=\left[\cdots \;\; d^{\mathrm{p}}_{-3}[s,t] \;\; d^{\mathrm{p}}_{-1}[s,t] \;\; d^{\mathrm{p}}_{1}[s,t] \;\; d^{\mathrm{p}}_{3}[s,t] \;\; \cdots\right]$:

$$\mathcal{D}_B\left(\mathrm{d}^{\mathrm{p}}[s,t]\right)=\frac{1}{B}\sum_{s'\in\mathcal{B}(s)}\left(X\left[s',2t+1\right]+\sum_{j\in\mathcal{J}'}p_{a^{\mathrm{p}}[s,t]}\left[L^{\mathrm{p}}+j\right]\cdot X\left[s'-d^{\mathrm{p}}_j[s,t],2t+1+j\right]\right)^2 \tag{A1}$$

where $\mathcal{B}(s)$ denotes the block, comprising the sample s, which is treated as a whole during motion-adaptive prediction, and $\mathcal{J}'=\mathcal{J}-\{0\}=\{\pm1,\pm3,\ldots,\pm L^{\mathrm{p}}\}$.

Similarly to the above, we denote as $\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j[s,t]\right)$ the *sample-wise single-reference prediction error*, i.e. the error resulting when the sample $X[s,2t+1]$ is predicted from the sample $X\left[s-d^{\mathrm{p}}_j[s,t],2t+1+j\right]$:

$$\mathcal{E}_{2t+1}\left(d_j[s,t]\right)=X[s,2t+1]-X\left[s-d_j[s,t],2t+1+j\right]. \tag{A2}$$

For each element of the noise vector $\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]$ in (35), i.e. each time instant $2t$, we consider the worst-case scenario of $L^{\mathrm{p}}+1$ mismatched displacements and average all the ensuing errors, thus obtaining:

$$\Delta^{\mathrm{disp}}x^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}},2t]=\frac{1}{L^{\mathrm{p}}+1}\sum_{j\in\mathcal{J}'}\left\{X^{\mathrm{p}}\left[s-\widehat{d^{\mathrm{p}}_j}[s,t_j],2t\right]-X^{\mathrm{p}}\left[s-d^{\mathrm{p}}_j[s,t_j],2t\right]\right\} \tag{A3}$$

where $t_j=(2t-j-1)/2$, with $j\in\mathcal{J}'$ as for (A1), and where the sign of the decoding-side displacements is reversed to fit the encoding-side coordinate system. We pursue an approximation of the expected predict-step *synthesis* distortion (induced by displacement mismatches) which exploits the data already gathered during predict-step *analysis*. We proceed as follows:

a) First we express the term $\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2$ as a function of the differences in the source samples between motion-compensated *neighbouring* input frames, e.g. $X\left[s,2t+1\right]$ and $X\left[s-d^{\mathrm{p}}_j,2t\right]$. The resulting expression aims to involve the prediction residuals (corresponding to odd time instants) that are typical of motion-adaptive temporal prediction.

b) Then we approximate the distortion contribution of individual samples with the average contribution of a block of samples. This links with practical motion-adaptive prediction algorithms that consider blocks rather individual pixels.

c) Finally we derive an approximation of the term $E\left\{\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2\right\}$ that incorporates a *block-based sensitivity* term (which reflects the local source characteristics) and the mismatch probability term (which reflects the transmission settings).

(a) <u>*Sample-wise distortion induced by displacement mismatches in predict-step synthesis*</u>: Under the assumption that prediction errors relative to different instants (i.e. resulting when samples within different input frames are predicted) are temporally orthogonal, the term $\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2$ can be expressed as:

$$\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2=\sum_{t=0}^{T/2-1}\sum_{j\in\mathcal{J}'}\left[\left(\frac{\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j[s,t]\right)}{L^{\mathrm{p}}+1}\right)^2+\left(\frac{\mathcal{E}_{2t+1}\left(\widehat{d^{\mathrm{p}}_j}[s,t]\right)}{L^{\mathrm{p}}+1}\right)^2-\left(\frac{2\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j[s,t]\right)\mathcal{E}_{2t+1}\left(\widehat{d^{\mathrm{p}}_j}[s,t]\right)}{\left(L^{\mathrm{p}}+1\right)^2}\right)\right]. \tag{A4}$$

The expression (A4) is obtained by first adding and subtracting the term $X\left[s,2t_j{+}1\right]$ inside the summation of (A3) yielding:

$$\Delta^{\mathrm{disp}} x^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathbf{d}}^{\mathrm{p}},2t] = \frac{1}{L^{\mathrm{p}}+1}\sum_{j\in\mathcal{J}'}\left[\mathcal{E}_{2t_j+1}\left(d^{\mathrm{p}}_j\left[s,t_j\right]\right)-\mathcal{E}_{2t_j+1}\left(\widehat{d^{\mathrm{p}}_j}\left[s,t_j\right]\right)\right] \tag{A5}$$

where the error $\mathcal{E}_{2t_j+1}\left(d^{\mathrm{p}}_j\left[s,t_j\right]\right)$ ensues as the sample $X\left[s,2t_j+1\right]$ is predicted from the sample $X\left[s\text{-}d^{\mathrm{p}}_j\left[s,t_j\right],2t_j{+}j{+}1\right]$. Similarly for $\mathcal{E}_{2t_j+1}\left(\widehat{d^{\mathrm{p}}_j}\left[s,t_j\right]\right)$. The hypothesis that prediction errors relative to different instants are temporally orthogonal implies that $\sum_{t=0}^{T/2-1}\left[\mathcal{E}_{2t_l+1}\left(d^{\mathrm{p}}_l\left[s,t_l\right]\right)\cdot\mathcal{E}_{2t_i+1}\left(d^{\mathrm{p}}_i\left[s,t_i\right]\right)\right]=0$ when $i\neq l$, irrespectively of both $d^{\mathrm{p}}_l\left[s,t_l\right]$ and $d^{\mathrm{p}}_i\left[s,t_i\right]$. Evaluating $\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2$ using (A5) then leads to (A4).

(b) <u>*Approximation to block-wise distortion*</u>: The sample-wise error terms in (A4) are approximated with the block-based equivalent formulation as follows. First we approximate the distortion relative to single pixels with the distortion contribution of small areas of the frames. In other words:

$$\sum_{t=0}^{T/2-1}\sum_{j\in\mathcal{J}'}\left[\frac{1}{L^{\mathrm{p}}+1}\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j\left[s,t\right]\right)\right]^2 \simeq \sum_{t=0}^{T/2-1}\sum_{j\in\mathcal{J}'}\left\{\frac{1}{B}\sum_{s'\in\mathcal{B}(s)}\left[p_{a^{\mathrm{p}}[s,t]}\left[L^{\mathrm{p}}{+}j\right]\cdot\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j\left[s',t\right]\right)\right]^2\right\}. \tag{A6}$$

Assuming that prediction errors from different reference frames (e.g. $\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j\left[s,t\right]\right)$ and $\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_l\left[s,t\right]\right)$, $j\neq l$) are spatially orthogonal and recalling that $\sum_{j\in\mathcal{J}'}p_n\left[L^{\mathrm{p}}+j\right]=-1$, the approximation (A6) becomes:

$$\sum_{t=0}^{T/2-1}\sum_{j\in\mathcal{J}'}\left[\frac{1}{L^{\mathrm{p}}+1}\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j\left[s,t\right]\right)\right]^2 \simeq \sum_{t=0}^{T/2-1}\mathcal{D}_B\left(\mathrm{d}^{\mathrm{p}}\left[s,t\right]\right). \tag{A7}$$

Using (A7), we approximate equation (A4) as:

$$\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2 \simeq \sum_{t=0}^{T/2-1}\left[\mathcal{D}_B\left(\widehat{\mathrm{d}^{\mathrm{p}}}\left[s,t\right]\right)-\mathcal{D}_B\left(\mathrm{d}^{\mathrm{p}}\left[s,t\right]\right)\right]+\sum_{t=0}^{T/2-1}\mathcal{X}^B_{2t+1}\left(\mathrm{d}^{\mathrm{p}}\left[s,t\right],\widehat{\mathrm{d}^{\mathrm{p}}}\left[s,t\right]\right) \tag{A8}$$

with:

$$\mathcal{X}^B_{2t+1}\left(\mathrm{d}^{\mathrm{p}}\left[s,t\right],\widehat{\mathrm{d}^{\mathrm{p}}}\left[s,t\right]\right)=\frac{2}{B}\sum_{s'\in\mathcal{B}(s)}\sum_{j\in\mathcal{J}'}\left[\left(\frac{\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j\left[s',t\right]\right)}{L^{\mathrm{p}}+1}\right)^2\cdot\left(1-\frac{\mathcal{E}_{2t+1}\left(\widehat{d^{\mathrm{p}}_j}\left[s',t\right]\right)}{\mathcal{E}_{2t+1}\left(d^{\mathrm{p}}_j\left[s',t\right]\right)}\right)\right]. \tag{A9}$$

Neglecting the contribution of the terms $\mathcal{X}^B_{2t+1}$ in the above incurs less than 15% error in practice and allows simplifying the approximation (A8) as:

$$\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2 \simeq \sum_{t=0}^{T/2-1}\left[\mathcal{D}_B\left(\widehat{\mathrm{d}^{\mathrm{p}}}\left[s,t\right]\right)-\mathcal{D}_B\left(\mathrm{d}^{\mathrm{p}}\left[s,t\right]\right)\right]. \tag{A10}$$

During the motion estimation phase, several suitable candidate displacements $\widehat{\mathrm{d}^{\mathrm{p}}}$ are tested and the corresponding block-based distortions $\mathcal{D}^B_{2t+1}\left(\widehat{\mathrm{d}^{\mathrm{p}}}\left[s,t\right]\right)$ are measured. The displacement yielding the minimum distortion (the "correct" value d^{p}) is then selected to perform analysis. Such values $\mathcal{D}^B_{2t+1}\left(\mathrm{d}^{\mathrm{p}}\left[s,t\right]\right)$ and $\mathcal{D}^B_{2t+1}\left(\widehat{\mathrm{d}^{\mathrm{p}}}\left[s,t\right]\right)$ are used in (A10).

(c) <u>*Expected distortion induced by displacement mismatches in predict-step synthesis*</u>: The expression (35), which approximates the distortion induced by displacement mismatches $E\left\{\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^2\right\}$, is derived by taking statistical expectation of (A10) over the probability that $\widehat{\mathrm{d}^{\mathrm{p}}}$ is used, thus obtaining:

$$E\left\{\left\|\Delta^{\mathrm{disp}}\mathbf{x}^{\mathrm{p}}[s+\widehat{\mathrm{d}^{\mathrm{p}}}]\right\|^{2}\right\} \simeq \sum_{t=0}^{T/2-1}\left\{\sum_{\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]}\left[\mathcal{D}_{B}\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\right)\cdot \Pr\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t] \mid \widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)\right.\right.$$
$$\left.\left.\cdot\Pr\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)\right] - \mathcal{D}_{B}\left(\mathrm{d}^{\mathrm{p}}[s,t]\right)\cdot \Pr\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)\right\} \tag{A11}$$

where $\Pr\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)$ denotes the probability that a displacement mismatch occurs and $\Pr\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t] \mid \widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)$ denotes the probability that the displacement $\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]$ is used in case of mismatch. Assume that, in case of a mismatch, any candidate displacement (as tested during motion estimation) can be used to perform synthesis. Let $\mathcal{N}\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)$ denote the number of such displacements. Therefore $\Pr\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t] \mid \widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right) = 1/\mathcal{N}\left(\widehat{\mathrm{d}^{\mathrm{p}}}[s,t]\neq \mathrm{d}^{\mathrm{p}}[s,t]\right)$ and (A11) becomes (35).

APPENDIX B

B.1. Proofs of Proposition 1 and Corollary 1

Proof of Proposition 1: Using the SVD [28] of the matrix $(2\mathbf{I}\text{-}\mathbf{M})$ to express $[(2\mathbf{I}\text{-}\mathbf{M})\Delta\mathbf{v}]$ yields:

$$\frac{1}{T}E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^{2}\right\} = \frac{1}{T}\sum_{i=1}^{T}\left[\varsigma_{i}\left\{2\mathrm{I}\text{-}\mathrm{M}\right\}\right]^{2}\mathrm{E}\left\{\left[\Delta\mathbf{v}^{\mathrm{T}}\,\mathbf{q}_{i}\left\{2\mathrm{I}\text{-}\mathrm{M}\right\}\right]^{2}\right\} \tag{B1}$$

with $\varsigma_{i}\{2\mathrm{I}\text{-}\mathrm{M}\}$ and $\mathbf{q}_{i}\{2\mathrm{I}\text{-}\mathrm{M}\}$ as given by Definition 1. We then recall that:

$$\left(\mathbf{b}^{\mathrm{T}}\mathbf{c}\right)^{2} = \mathrm{tr}\left\{\left(\mathbf{c}\mathbf{c}^{\mathrm{T}}\right)\left(\mathbf{b}\mathbf{b}^{\mathrm{T}}\right)\right\} = \mathrm{tr}\left\{\left(\mathbf{b}\mathbf{b}^{\mathrm{T}}\right)\left(\mathbf{c}\mathbf{c}^{\mathrm{T}}\right)\right\} \tag{B2}$$

where $\mathbf{b}$ and $\mathbf{c}$ are $T\times 1$ vectors. Using (B2) in (B1), interchanging the trace and expectation operators, and combining the linearity of the trace operator with (1) leads to (16). ■

Proof of Corollary 1: By expanding (16) we have:

$$\frac{1}{T}E\left\{\|(2\mathbf{I}-\mathbf{M})\Delta\mathbf{v}\|^{2}\right\} = \frac{1}{T}\sum_{k=0}^{T-1}W^{(2\mathrm{I}\text{-}\mathrm{M})}[k,k]R_{\Delta\mathbf{v}}[k,k] + \frac{2}{T}\sum_{j=1}^{T-1}\sum_{k=j}^{T-1}W^{(2\mathrm{I}\text{-}\mathrm{M})}[k,k-j]R_{\Delta\mathbf{v}}[k,k-j]\ . \tag{B3}$$

The hypothesis made for $\Delta\mathbf{v}_{\mathrm{even}}$ and $\Delta\mathbf{v}_{\mathrm{odd}}$ implies that the second term on the right hand side of (B3) is zero. Furthermore, we have $R_{\Delta\mathbf{v}}[2k,2k] = E\left\{\|\Delta\mathbf{v}_{\mathrm{even}}\|^{2}\right\}/\left(T/2\right)$ and $R_{\Delta\mathbf{v}}[2k{+}1,2k{+}1] = E\left\{\|\Delta\mathbf{v}_{\mathrm{odd}}\|^{2}\right\}/\left(T/2\right)$ for $k=0,1,\ldots,T/2\text{-}1$. Therefore separating the even and odd values of k in (B3) yields (17). ■

B.2. Proof of Proposition 2

Lemma 1: Let $\eta\in\left\{1,2,\ldots,T/2\right\}$ denote the number of synthesis lifting parameters that do not match their analysis counterpart, i.e. $\hat{a}[t]\neq a[t]$ at η distinct time instants. The induced synthesis distortion is:

$$\frac{1}{T}E\left\{\|\Delta\mathbf{M}\mathbf{v}\|^{2}\mid\eta\right\} = \frac{1}{T}\mathrm{tr}\left\{\left(\mathbf{v}\,\mathbf{v}^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathcal{M}_{\eta}\right\}\right\} \tag{B4}$$

where the $T\times T$ matrix $\mathbf{W}\left\{\Delta\mathcal{M}_{\eta}\right\}$ is defined in (21).

Proof: The distortion induced by a given matrix $\Delta\mathbf{M}\in\Delta\mathcal{M}_{\eta}$ and coefficient vector $\mathbf{v}$ is given by (16) and (B2) as $\|\Delta\mathbf{M}\mathbf{v}\|^{2}/T = \left(1/T\right)\mathrm{tr}\left\{\left(\mathbf{v}\,\mathbf{v}^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathrm{M}\right\}\right\}$. Performing statistical average over each $\Delta\mathbf{M}\in\Delta\mathcal{M}_{\eta}$, recalling the definition (21) and exploiting linearity, yields the expression (B4) for $E\left\{\|\Delta\mathbf{M}\mathbf{v}\|^{2}\mid\eta\right\}/T$. ■

Proof of Proposition 2: $E\left\{\|\Delta\mathbf{Mv}\|^2\right\}/T = \left(1/T\right)\sum_{\eta=0}^{T/2}\left[\Pr(\eta)\,E\left\{\|\Delta\mathbf{Mv}\|^2 \mid \eta\right\}\right]$. Since $\Delta\mathcal{M}_0 = \{\mathbf{0}\}$ then $E\{\|\Delta\mathbf{Mv}\|^2 \mid \eta = 0\} = 0$. Employing (B4) for $\eta > 0$ yields (20). ■

B.3. Proof of Proposition 3:

Lemma 2: Assuming that $\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{even}}$, $\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}$ and the coefficients on the even rows of $\Delta\mathbf{U}$ constitute three mutually independent white WSS noise processes, we have:

$$E\{\Delta x^{\mathrm{p}}[2i{+}1]\,\Delta x^{\mathrm{p}}[2j]\} = \begin{cases} -\left[E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}/\left(T/2\right)\right] u_{a[j]}[L^{\mathrm{u}} + (2i{+}1{-}2j)] & , \ |2i{-}2j{+}1| \le L^{\mathrm{u}} \\ 0 & , \ |2i{-}2j{+}1| > L^{\mathrm{u}} \end{cases} \tag{B5}$$

$$E\{\Delta x^{\mathrm{p}}[2i]\,\Delta x^{\mathrm{p}}[2j{+}1]\} = \begin{cases} -\left[E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}/\left(T/2\right)\right] u_{a[i]}[L^{\mathrm{u}} + (2j{+}1{-}2i)] & , \ |2j{-}2i{+}1| \le L^{\mathrm{u}} \\ 0 & , \ |2j{-}2i{+}1| > L^{\mathrm{u}} \end{cases} \tag{B6}$$

where $u_n[k]$ is as in (3). Furthermore, for $i \ne j$, we have:

$$E\{\Delta x^{\mathrm{p}}[2i]\,\Delta x^{\mathrm{p}}[2j]\} = \begin{cases} \left[E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}/\left(T/2\right)\right]\left(\sum\limits_{h\in\mathcal{I}'} u_{a[i]}[L^{\mathrm{u}}+h]\,u_{a[j]}[L^{\mathrm{u}}+(2i{-}2j)+h]\right) & , \ |i{-}j| \le L^{\mathrm{u}} \\ 0 & , \ |i{-}j| > L^{\mathrm{u}} \end{cases} \tag{B7}$$

$$E\{\Delta x^{\mathrm{p}}[2i{+}1]\,\Delta x^{\mathrm{p}}[2j{+}1]\} = 0 \tag{B8}$$

where $\mathcal{I}' = \{\pm1, \pm3, \ldots, \pm L^{\mathrm{u}}\}$.

Proof: It follows via simple algebraic derivation recalling that $\Delta\mathbf{x}^{\mathrm{p}} = (2\mathbf{I}\text{-}\mathbf{U})\,\Delta\mathbf{x}^{\mathrm{u}} \text{-} \Delta\mathbf{U}\,\mathbf{x}^{\mathrm{u}} \text{-} \Delta\mathbf{U}\Delta\mathbf{x}^{\mathrm{u}}$. ■

Proof of Proposition 3: From (15) we have $E\left\{\|\Delta\mathbf{x}\|^2\right\}/T = E\left\{\|(2\mathbf{I}\text{-}\mathbf{P})\,\Delta\mathbf{x}^{\mathrm{p}}\|^2\right\}/T + E\left\{\|\Delta\mathbf{P}\mathbf{x}^{\mathrm{p}}\|^2\right\}/T$. Expressing the first term using (B3) (since $\Delta\mathbf{x}^{\mathrm{p}}$ is not white) and the second term using (20) yields:

$$\begin{aligned} \frac{1}{T}E\left\{\|\Delta\mathbf{x}\|^2\right\} = {} & \frac{1}{T}\sum_{\eta=1}^{T/2}\left[\Pr(\eta)\ \mathrm{tr}\left\{\left(\mathbf{x}^{\mathrm{p}}(\mathbf{x}^{\mathrm{p}})^{\mathrm{T}}\right)\mathbf{W}\{\Delta\mathcal{P}_\eta\}\right\}\right] + \gamma_{\mathrm{e}}\{\mathrm{P}\}\frac{E\left\{\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}\|^2\right\}}{T/2} \\ & + \gamma_{\mathrm{o}}\{\mathrm{P}\}\frac{E\left\{\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}}\|^2\right\}}{T/2} + \frac{2}{T}\sum_{j=1}^{T-1}\sum_{k=j}^{T-1} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k, k-j]\,E\{\Delta x^{\mathrm{p}}[k]\,\Delta x^{\mathrm{p}}[k-j]\}. \end{aligned} \tag{B9}$$

Since[7] $E\left\{\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}}\|^2\right\}/\left(T/2\right) = E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}/\left(T/2\right)$ we derive $E\left\{\|\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{even}}\|^2\right\}/\left(T/2\right)$ from $E\left\{\|\Delta\mathbf{x}^{\mathrm{p}}\|^2\right\}/T$, which is in turn obtained applying (15) (17) and (20) to the update step. The last term in (B9) equals

$$\begin{aligned} & \frac{2}{T}\sum_{j=1,3,\ldots}^{T-1}\left[\sum_{k=j,j+2,\ldots}^{T-1} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\,E\{\Delta x^{\mathrm{p}}[k]\,\Delta x^{\mathrm{p}}[k\text{-}j]\} + \sum_{k=j+1,j+3\ldots}^{T-2} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\,E\{\Delta x^{\mathrm{p}}[k]\,\Delta x^{\mathrm{p}}[k\text{-}j]\}\right] + \\ & \frac{2}{T}\sum_{j=2,4,\ldots}^{T-1}\left[\sum_{k=j,j+2,\ldots}^{T-1} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\,E\{\Delta x^{\mathrm{p}}[k]\,\Delta x^{\mathrm{p}}[k\text{-}j]\} + \sum_{k=j+1,j+3\ldots}^{T-2} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\,E\{\Delta x^{\mathrm{p}}[k]\,\Delta x^{\mathrm{p}}[k\text{-}j]\}\right]. \end{aligned} \tag{B10}$$

Using (B5)-(B8) the expression (B10) becomes $\left[\xi\{\mathrm{P},\mathrm{U}\}\,E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}/\left(T/2\right)\right]$, with[8]:

$$\xi\{\mathrm{P},\mathrm{U}\} = \alpha\{\mathrm{P},\mathrm{U}\} - \beta\{\mathrm{P},\mathrm{U}\} \tag{B11}$$

[7] The odd samples of $(2\mathbf{I}\text{-}\mathbf{U})\,\Delta\mathbf{x}^{\mathrm{u}}$ and $\Delta\mathbf{x}^{\mathrm{u}}$ coincide whereas the odd samples of $\Delta\mathbf{U}\,\mathbf{x}^{\mathrm{u}}$ are zero, hence $\Delta\mathbf{x}^{\mathrm{p}}_{\mathrm{odd}} = \Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}$.

[8] The approximated expressions of $\alpha\{\mathrm{P},\mathrm{U}\}$ and $\beta\{\mathrm{P},\mathrm{U}\}$, isolating the contributions of $\mathbf{P}$ and $\mathbf{U}$, are used in practice.

$$\begin{aligned}\alpha\{\mathrm{P},\mathrm{U}\} &= \frac{2}{T}\sum_{j=2,4,\ldots}^{2L^{\mathrm{u}}}\left[\sum_{k=j,j+2,\ldots}^{T-2} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\sum_{h\in\mathcal{I}'}\left(u_{a\left[k/2\right]}[L^{\mathrm{u}}+h]\,u_{a\left[(k\text{-}j)/2\right]}[L^{\mathrm{u}}+j+h]\right)\right] \\ &\cong \frac{2}{T}\left[\sum_{j=2,4,\ldots}^{2L^{\mathrm{u}}}\sum_{k=j,j+2,\ldots}^{T-2} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\right]\left[\frac{2}{TL^{\mathrm{u}}}\sum_{t=0}^{T/2-1}\sum_{j=2,4,..}^{2L^{\mathrm{u}}}\sum_{h\in\mathcal{I}'}\left(u_{a[t]}[L^{\mathrm{u}}+h]\,u_{a\left[t-j/2\right]}[L^{\mathrm{u}}+j+h]\right)\right]\end{aligned} \tag{B12}$$

$$\begin{aligned}\beta\{\mathrm{P},\mathrm{U}\} &= \frac{2}{T}\sum_{j=1,3,\ldots}^{L^{\mathrm{u}}}\left[\sum_{k=j,j+2,\ldots}^{T-1} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\,u_{a\left[(k\text{-}j)/2\right]}[L^{\mathrm{u}}+j] + \sum_{k=j+1,j+3,\ldots}^{T-2} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\,u_{a\left[k/2\right]}[L^{\mathrm{u}}\text{-}j]\right] \\ &\cong \frac{2}{T}\left[\sum_{j=1,3,\ldots}^{L^{\mathrm{u}}}\sum_{k=j}^{T-1} W^{(2\mathrm{I}\text{-}\mathrm{P})}[k,k\text{-}j]\right]\left[\frac{2}{TL^{\mathrm{u}}}\sum_{t=0}^{T/2-1}\sum_{h\in\mathcal{I}'} u_{a[t]}[L^{\mathrm{u}}+h]\right] \quad .\end{aligned} \tag{B13}$$

Using the short-hand of (23)-(24) the above leads to:

$$\begin{aligned}\frac{1}{T}E\left\{\|\Delta\mathbf{x}\|^2\right\} &= \varphi_{\mathrm{e}}\{\mathrm{P},\mathrm{U}\}\frac{E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{even}}\|^2\right\}}{T/2} + \varphi_{\mathrm{o}}\{\mathrm{P},\mathrm{U}\}\frac{E\left\{\|\Delta\mathbf{x}^{\mathrm{u}}_{\mathrm{odd}}\|^2\right\}}{T/2} \\ &+ \frac{1}{T}\sum_{\eta=1}^{T/2}\left\{\Pr(\eta)\left[\mathrm{tr}\left\{\left(\mathbf{x}^{\mathrm{p}}(\mathbf{x}^{\mathrm{p}})^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathcal{P}_\eta\right\}\right\}+2\gamma_{\mathrm{e}}\{\mathrm{P}\}\,\mathrm{tr}\left\{\left(\mathbf{x}^{\mathrm{u}}(\mathbf{x}^{\mathrm{u}})^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathcal{U}_\eta\right\}\right\}\right]\right\} \ .\end{aligned} \tag{B14}$$

Since errors in the lifting parameters occur independently, $\Pr(\eta)=\rho^\eta(1-\rho)^{T/2-\eta}\begin{pmatrix}T/2\\ \eta\end{pmatrix}$. Therefore, replacing

$$\psi\{\rho,\mathrm{P},\mathrm{x}^{\mathrm{p}},\mathrm{x}^{\mathrm{u}}\}=\sum_{\eta=1}^{T/2}\left\{\left[\rho^\eta(1-\rho)^{T/2\text{-}\eta}\begin{pmatrix}T/2\\ \eta\end{pmatrix}\right]\left[\mathrm{tr}\left\{\left(\mathbf{x}^{\mathrm{p}}(\mathbf{x}^{\mathrm{p}})^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathcal{P}_\eta\right\}\right\}+2\gamma_{\mathrm{e}}\{\mathrm{P}\}\,\mathrm{tr}\left\{\left(\mathbf{x}^{\mathrm{u}}(\mathbf{x}^{\mathrm{u}})^{\mathrm{T}}\right)\mathbf{W}\left\{\Delta\mathcal{U}_\eta\right\}\right\}\right]\right\} \tag{B15}$$

[where $\mathbf{W}\left\{\Delta\mathcal{P}_\eta\right\}$ and $\mathbf{W}\left\{\Delta\mathcal{U}_\eta\right\}$ are given by (21)] in (B14) leads to (22). ■