HIERARCHICAL AND LOSSLESS CODING OF AUDIO OBJECTS IN DOLBY TRUEHD

Vinay Melkote†, Malcolm Law‡, and Rhonda Wilson†

†Dolby Laboratories, Inc., San Francisco, CA 94103
‡Algol Applications Ltd, UK
{vmelk, rwils}@dolby.com, mjlaw@algol.co.uk

ABSTRACT

Dolby TrueHD is a lossless and hierarchical audio coding format that not only enables compact bit-exact representation of the source multichannel audio signal, but also facilitates low complexity reconstruction of downmixes thereof. The dual objective is achieved by linear transformation of input channels into internal channels coded in the bitstream, via primitive matrices that are exactly invertible with finite precision, such that a subset of the internal channels spans the subpsace of the required downmix. A decoder need only extract and linearly combine this subset of internal channels to reconstruct the downmix presentation. Until recently, TrueHD was primarily employed for lossless carriage of speaker feeds (typically, 7.1ch), in which case downmixes of interest (5.1ch or stereo) could be obtained via time-invariant linear transformations. Concurrent with Dolby Laboratories' efforts towards a next generation surround sound experience, however, the format has been extended to support lossless transmission of audio objects such that renderings (downmixes) of these moving objects to standard speaker layouts are still accessible to legacy decoders. The mandated representation of a continuously varying downmix matrix trajectory is facilitated by the novel paradigm of interpolated primitive matrices and the associated matrix decomposition strategy presented here.

***Index Terms*—** lossless audio coding, downmixing, primitive matrices, lifting matrices, audio objects

1. INTRODUCTION

The Dolby TrueHD format has been widely employed for lossless carriage of multichannel audio tracks, typically 7.1ch signals, on Blu-ray discs, and is derived from Meridian Lossless Packing [1] that was originally standardized for DVD-audio. Compared to other lossless audio coding formats such as FLAC [2], Shorten [3], MPEG's Scalable-to-Lossless AAC [4, 5, 6] and Audio Lossless Coding [6], DTS-HD Master Audio [7], etc., TrueHD is unique in that, in addition to being lossless it has a hierarchical, i.e., scalable, bitstream structure that facilitates low-complexity decoding of downmix presentations, such as 5.1ch or stereo, of the source multichannel signal.

A simplified view of the TrueHD system is provided in Fig. 1. In addition to N input audio channels, whose samples at a given time instant are denoted by the $N \times 1$ vector $\mathbf{x}$, the encoder receives an $M \times N$ downmix matrix $\mathbf{A}$ ($M < N$) that defines an M channel downmix, $\mathbf{Ax}$, of the source, and decomposes this matrix into a product of encoder-end (or input) and decoder-end (or output) primitive matrices. In the figure, $P_0, \cdots, P_{K-1}$ represent the K, $N \times N$ input primitive matrices, that are sequentially applied to transform the input channels into N 'internal' channels, each of which is then individually encoded into the bitstream via linear prediction and Huffman coding (henceforth, we will employ the symbol $\{a_n\}_{n=0}^{N}$ to denote the sequence $a_0, \cdots, a_N$). The first M internal channels are packed into a first 'substream' or layer (indexed 0) and the remaining $N - M$ channels are encoded into substream 1. While primitive matrices are discussed in more detail in Sec. 2.1, suffice to say that $\{P_k\}_{k=0}^{K-1}$ can be exactly inverted with finite precision arithmetic, and the lossless original can be recovered by decoding both substreams and applying the appropriate matrix inverses to the internal channels. When the downmix alone is desired, substream 0 has sufficient information: the M internal channels it contains are transformed via a second set of L, $M \times M$ output primitive matrices $\{Q_l\}_{l=0}^{L-1}$ to construct the downmix. The decomposition of $\mathbf{A}$ into input and output primitive matrices is such that:

$$\mathbf{A} = Q_{L-1} \cdots Q_0 \mathbf{I}' P_{K-1} \cdots P_0 \quad (1)$$

where $\mathbf{I}'$ is the $M \times N$ 'row selector' matrix composed of the first M rows of an $N \times N$ identity matrix. The condition (1) ensures that the first M rows of the product $P_{K-1} \cdots P_0$, of input primitive matrices, span the subspace of rows of $\mathbf{A}$, and the required downmix can be obtained by linear combination of the first M internal channels alone. As an aside, note that matrixing also provides the ability to exploit inter-channel correlations for improved compression efficiency, but this is not the focus here.

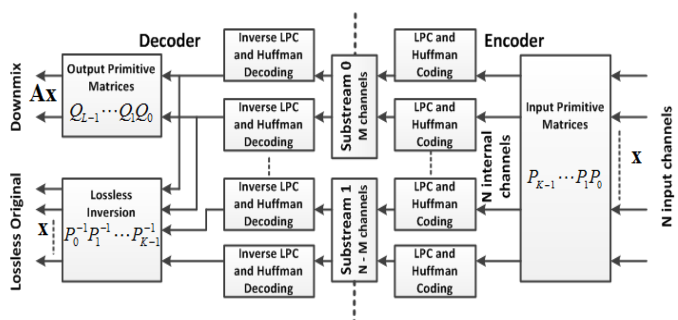

Fig. 1. Overview of the TrueHD system

In legacy TrueHD, primitive matrices are held constant unless an update is received in the bitstream. This was acceptable since traditionally the downmix was from one fixed speaker layout to another, and the matrix $\mathbf{A}$ was in fact time-invariant. An one-time decomposition of $\mathbf{A}$ into primitive matrices sufficed, and they were were only repeated as often as required for random access of the audio track (every 5120 samples at 48kHz sampling rate). The recently introduced Dolby Atmos format [8, 9], however, enables an immersive audio experience via *audio objects*: audio signals associated with time-varying spatial metadata representing sound sources with arbitrary location/motion in the room. Playback over a fixed speaker layout requires a *dynamic* linear transformation to *render* (map) the moving objects to speaker feeds. Translation of this enhanced surround sound experience, originally introduced in the cinema, to the home required lossless carriage of object audio via TrueHD, but with the caveat that existing playback devices which could not interpret these objects still be able to receive a rendering, i.e., downmix, to a standard speaker layout such as 7.1. Plausible solutions in the legacy TrueHD framework included simulcasting the audio objects and their rendering to 7.1 as independent TrueHD bitstreams, or piecewise constant approximation of the rendering matrix trajectory with rapid matrix updates, both of which methods entail significant bitrate expense.

This motivates the augmentation of the TrueHD syntax with the ability to transmit seed (intercept) primitive matrices and delta (slope) matrices, so that primitive matrices can be interpolated over time, and a continuous (per-sample) evolution of the input-to-downmix linear transformation affected without recourse to frequent matrix updates. The schematic of the new TrueHD system is effectively the same as in Fig. 1, with the understanding that the input channels are really audio objects, and the downmix matrix $\mathbf{A}$ and the primitive matrices evolve over time. Maintaining backwards compatibility, however, constrains usage of this novel paradigm to only those substreams not accessed by legacy decoders. Esssentially the output matrices $\{Q_l\}_{l=0}^{L-1}$ which are required for decoding the downmix presentation, in particular in a legacy decoder, will need to be compliant with legacy TrueHD syntax and cannot be interpolated. Nevertheless, the input primitive matrices $\{P_k\}_{k=0}^{K-1}$ are interpolatable, since their inverses carried in substream 1 will only be required to retrieve the audio objects losslessly in a new TrueHD decoder and can adhere to the modified syntax. A novel matrix decomposition strategy is therefore necessitated to represent the specified time-varying rendering matrix as a combination of interpolated input and piecewise-contant output primitive matrices. The rest of this paper discusses the challenges involved, intricacies of the solution, and provides a system evaluation.

2. BACKGROUND

2.1. Primitive Matrices

A primitive matrix is a square matrix that is identical to an identity matrix except in one row, i.e., a primitive matrix P of size $N \times N$ has the structure:

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ \alpha_{p,0} & \alpha_{p,1} & \cdots & \alpha_{p,p} & \cdots & \alpha_{p,N-1} \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (2)$$

where row p is the non-trivial row. Given a column vector $\mathbf{x}$ of N audio samples, each from a separate channel, $P\mathbf{x}$ differs from $\mathbf{x}$ only in the p^{th} element. The primitive matrix P can therefore be associated with a unique channel, indexed p, which it operates on and modifies, while leaving the remaining channels unaltered. Naturally, a sequence of primitive matrices $\{P_k\}_{k=0}^{K-1}$, can be associated with a corresponding 'row index sequence' (RIS) $\{p_k\}_{k=0}^{K-1}$.

When the coefficient on the diagonal $\alpha_{p,p} = 1$, P is referred to as a unit primitive matrix, and has the effect of adding a weighted combination of the remaining channels into channel p. Obviously the process can be reversed by simply substracting the same weighted combination from the modified channel p. In other words the inverse of P, when $\alpha_{p,p} = 1$, is simply

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ -\alpha_{p,0} & -\alpha_{p,1} & \cdots & 1 & \cdots & -\alpha_{p,N-1} \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (3)$$

Thus, P and P^{-1} can be represented with the same precision of coefficients, and constraining $\{P_k\}_{k=0}^{K-1}$ in Fig. 1 to be unit primitive matrices facilitates lossless retrieval of the input signal even with finite precision architecture. A similar argument can be made when $\alpha_{p,p} = -1$, however this case is not considered here. The output matrices $\{Q_l\}_{l=0}^{L-1}$, required to reconstruct the downmix, need not be invertible and may not be unit primitive matrices.

Each primitive matrix is represented in the TrueHD bitstream by enumeration of the index and coefficients of its non-trivial row. The syntax imposes limits on coefficient values - in legacy TrueHD they are bound to the semi-open interval $[-2, 2)$. A decomposition of the form (1) is not necessarily unique, and the challenge is to derive primitive matrices with coefficients that are representable in the TrueHD syntax. Primitive matrices have been otherwise referred to in audio coding literature as 'lifting' matrices [10, 11].

2.2. Object Audio

In the context of Dolby Atmos, an audio object could be defined as a mono audio waveform together with a time varying specification of the location of the sound source it represents. Typically, the object's x-y-z coordinates in the room are specified intermittently. An object audio rendering algorithm converts the position metadata into 'panning gains' so that the object signal may be distributed into speaker signals which, on playback, ideally create the illusion that the sound source is located/moving as intended. Given N audio objects to be played back over a layout of M speakers, the panning gains define an $M \times N$ downmix matrix at each time instant the metadata is specified. The rendering algorithm suitably interpolates between these 'sample' downmix matrices to construct a matrix trajectory that ensures continuous object motion.

3. ENHANCEMENTS TO THE TRUEHD SYSTEM TO REPRESENT OBJECT MOTION

Let the N input mono signals in Fig. 1 be audio objects. It is required that the M-channel output from substream 0 at times t_1 and t_2, respectively, is the downmix of the input objects to M speakers via given $M \times N$ matrices $\mathbf{A}(t_1)$, and $\mathbf{A}(t_2)$. These matrices could be samples of the downmix matrix trajectory of a reference renderer. At intermediate time instants t ($t_1 < t < t_2$) it is desired that the downmix correspond to a smoothly varying trajectory between $\mathbf{A}(t_1)$ and $\mathbf{A}(t_2)$. Additionally, to ensure backwards compatibility (see Sec. 1), any information required to decode the downmix should be representable in existing (i.e., legacy) TrueHD syntax.

3.1. Interpolated Primitive Matrices

Let $P(t_1)$ and $P(t_2)$ be a pair of primitive matrices, defined at times t_1 and t_2, respectively, that *operate on the same channel* p. Define a matrix $P(t)$ at an intermediate time instant by linear interpolation:

$$P(t) = P(t_1) + (t - t_1)\Delta(t_1), \text{ where } \Delta(t_1) = \frac{P(t_2) - P(t_1)}{t_2 - t_1} \quad (4)$$

Clearly, $P(t)$ is also a primitive matrix. It is a unit primitive matrix if $P(t_1)$ and $P(t_2)$ themselves are of the same type, and based on (3), computation of $P(t)^{-1}$ would require $-\Delta(t_1)$ which can be represented at the same precision as $\Delta(t_1)$, and hence $P(t)$ is exactly invertible with finite precision. Thus, a TrueHD system that employs interpolated input primitive matrices at the encoder and corresponding interpolated inverses at the decoder will continue to ensure lossless reconstruction of the source signal.

A continuous transition from downmix matrix $\mathbf{A}(t_1)$ to $\mathbf{A}(t_2)$ can be achieved via interpolation between primitive matrices in their decomposition. This is facilitated by a modification of the TrueHD syntax wherein, at an update point in the bitstream a seed primitive matrix ($P(t_1)$ in (4)) and a corresponding delta or slope matrix ($\Delta(t_1)$ in (4)), that describes its evolution until the next update, may be transmitted. Independent decompositions of $\mathbf{A}(t_1)$ to $\mathbf{A}(t_2)$ could be derived by employing the matrix decomposition routine of a legacy TrueHD encoder (for details refer to [12]), designed to deal with one matrix at a time. However, the resultant primitive matrix sequences at t_1 and t_2 are not guaranteed to share the same RIS, which is an obvious requirement for pairing of primitive matrices for interpolation. Thus a novel matrix decomposition strategy is necessitated that *suitably constrains the decomposition of one downmix matrix based on the decomposition of its neighbor*.

3.2. Matrix Decomposition to Enable Interpolation

In order to keep the description succinct, we forgo an exact enumeration of the algorithm, and only highlight the basic principles involved. We will assume that $\mathbf{A}(t_1)$ and $\mathbf{A}(t_2)$ are full rank, i.e., have rank M (handling rank deficiency is not discussed here). Let the singular value decomposition of $\mathbf{A}(t_1)$ be denoted by:

$$\mathbf{A}(t_1) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{I}'\mathbf{V} \tag{5}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices of dimensions, $M \times M$ and $N \times N$ respectively, and $\boldsymbol{\Sigma}$ is the $M \times M$ diagonal matrix containing the singular values, which are all non-zero. Setting $\mathbf{Q} = (\mathbf{U}\boldsymbol{\Sigma})^{-1}$, we will attempt to decompose the 'rotated' downmix matrices $\mathbf{B}(t_i) = \mathbf{Q}\mathbf{A}(t_i)$, $i \in \{1, 2\}$ into a product of unit (input) primitive matrices alone:

$$\mathbf{B}(t_i) = \mathbf{I}'P_{K-1}(t_i)\cdots P_0(t_i), i \in \{1, 2\} \tag{6}$$

such that $\{P_k(t_2)\}_{k=0}^{K-1}$ has the same RIS $\{p_k\}_{k=0}^{K-1}$ as $\{P_k(t_1)\}_{k=0}^{K-1}$. Clearly, then, delta matrices $\{\Delta_k(t_1)\}_{k=0}^{K-1}$ can be defined to interpolate from elements of $\{P_k(t_1)\}_{k=0}^{K-1}$ to corresponding elements in $\{P_k(t_2)\}_{k=0}^{K-1}$. A decomposition of $\mathbf{Q}^{-1}$ itself into a product of output primitive matrices:

$$\mathbf{Q}^{-1} = Q_{L-1}\cdots Q_0 \tag{7}$$

can be obtained by approaches such as in the legacy encoder (see [12]), and is fairly straightforward given $\{Q_l\}_{l=0}^{L-1}$ are not constrained to be unit primitive matrices. Note that (6) and (7) together describe a primitive matrix decomposition of the form (1) for $\mathbf{A}(t_1)$ and $\mathbf{A}(t_2)$, and the output primitive matrices are essentially designed to compensate for the rotation $\mathbf{Q}$. Holding the output matrices constant while linearly interpolating only the input matrices results in a smoothly varying downmix matrix trajectory:

$$\begin{aligned} \mathbf{A}(t) &= Q_{L-1}\cdots Q_0\mathbf{I}'P_{K-1}(t)\cdots P_0(t), \text{where} \\ P_k(t) &= P_k(t_1) + (t - t_1)\Delta_k(t_1),\ 0 \le k < K - 1 \end{aligned} \tag{8}$$

that, as required, matches the specified downmix matrices at the ends of the segment $[t_1, t_2]$, and ensures continuous object motion in between. Lossless reconstruction of the objects is guaranteed (due to exact invertibility of $\{P_k(t)\}_{k=0}^{K-1}$), but more significantly conformity of the output matrices to legacy TrueHD syntax is facilitated by dint of the invariance of $\{Q_l\}_{l=0}^{L-1}$ over the segment.

As an aside, note that $\mathbf{B}(t_1) = \mathbf{I}'\mathbf{V}$, and the rows of $\mathbf{B}(t_1)$ are orthogonal to each other and have unit norm. Since $\mathbf{B}(t_1)$ is the transfer function between input channels and the first M internal channels at time t_1, the above choice of $\mathbf{Q}$ ensures that these internal channels are considerably uncorrelated with each other ensuring improved compression efficiency. Further, these channels are bound in power, and their exact representation on system datapaths of fixed bit depth, necessary to enable lossless reconstruction, is facilitated.

3.2.1. Choosing the Input Primitive Matrices

Deriving the decomposition in (6) for a given choice of $\{p_k\}_{k=0}^{K-1}$ is now described in the context of a simple example. We will only consider $\mathbf{B}(t_1)$; handling $\mathbf{B}(t_2)$ is similar. Let $N = 3, M = 2$. Given

$$\mathbf{B}(t_1) = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \tag{9}$$

and RIS $\{2, 0, 1\}$ (hence $K = 3$), a matrix decomposition of the following structure:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ \gamma_0 & 1 & \gamma_2 \\ 0 & 0 & 1 \end{bmatrix}}_{P_2(t_2)} \underbrace{\begin{bmatrix} 1 & \beta 1 & \beta 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P_1(t_2)} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_0 & \alpha_1 & 1 \end{bmatrix}}_{P_0(t_2)} \tag{10}$$

such that the first two rows of the product are equal to $\mathbf{B}(t_1)$, is desired. Naturally, there are 6 equations in 6 unknowns, solving for which in the following sequence: β_2, α_0, γ_0, γ_2, α_1, β_1, ensures that each step only involves linear equations, and the solution (if it exists) is unique. The cumbersome but straight forward math is not detailed, but the final expression for two coefficients is highlighted below:

$$\alpha_0 = \frac{a_{00} - 1}{a_{02}},\ \alpha_1 = \frac{\alpha_0 \begin{vmatrix} a_{01} & a_{02} \\ a_{11} & a_{12} \end{vmatrix} + \begin{vmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{vmatrix} - 1}{\begin{vmatrix} a_{00} & a_{02} \\ a_{10} & a_{12} \end{vmatrix}} \tag{11}$$

where $|\cdot|$ denotes the determinant. The above approach of solving for the coefficients in a specific sequence so that only linear equations are involved is then generalized to handle an arbitrary $\mathbf{B}(t_1)$.

Observe that the denominators in (11) contain determinants of sub-matrices of (9). Clearly, it is necessary to choose an RIS that maximizes the absolute value of determinants of these 'critical' sub-matrices, so that the coefficients in the primitive matrices are well bound. Each $N \times N$ unit primitive matrix affords $N - 1$ degrees of freedom (unknown coefficients), and a decomposition of a given $M \times N$ matrix typically requires $M + 1$ such primitive matrices, implying a choice of RIS $\{p_k\}_{k=0}^{M}$ of length $M + 1$. It can be shown that the decomposition is characterized by M critical sub-matrices, $\{\mathbf{C}_k\}_{k=0}^{M-1}$, of the given $M \times N$ matrix, where the sub-matrix $\mathbf{C}_k$ is of size $(k + 1) \times (k + 1)$, and is uniquely determined by the sub-sequence $\{p_l\}_{l=0}^{k}$. This result naturally leads to a greedy method for determining a suitable RIS where, given the the sub-sequence $\{p_l\}_{l=0}^{k-1}$ a new element p_k is chosen to maximize the absolute value of $|\mathbf{C}_k|$. The last element of the RIS, p_M, can be arbitrarily chosen in the end so that $p_M \notin \{p_k\}_{k=0}^{M-1}$. For a full-rank $M \times N$ matrix there is at least one choice of the RIS that guarantees that none of the determinants of critical sub-matrices are zero.

Since, the matrices $\mathbf{B}(t_1)$ and $\mathbf{B}(t_2)$ have to be decomposed in tandem, the critical sub-matrices in both cases are simultaneously considered. The singular values of $\mathbf{B}(t_1)$ are all necessarily unity (due to the choice of $\mathbf{Q}$). Nevertheless, this decomposition strategy can lead to large coefficients, in particular when the singular values of the matrix $\mathbf{B}(t_2)$ are small. Note, however, that we are dealing here with input primitive matrices, whose coefficients are only required in new decoders that attempt to reconstruct the original objects, and hence need not conform to constraints of the legacy syntax. The updated TrueHD syntax allows an expanded range of coefficients - $[-128, 128)$ - to handle the issue.

3.3. Additional Parameters in the Decomposition

The syntax (both legacy and updated) provides additional degrees of freedom to ensure reasonable coefficient values. Specifically, a ‘channel assignment’ parameter can be sent in each of the substreams that shuffles the channels before presentation at the decoder. This operation is equivalent to pre- and post- multiplying the given downmix matrices with permutation matrices corresponding to the channel assignments, prior to their decomposition. The critical submatrices discussed previously thus become a function of the RIS as well as the channel assignments, and the greedy search strategy simultaneously finds both parameters to minimize coefficient values. Further, a per-output-channel shift (attenuation or gain in powers of 2) can be specified, which is exploited to modify the output matrix coefficients to a range representable in the syntax.

3.4. System Design

A reference renderer provides a sequence of downmix matrices by sampling its rendering matrix trajectory at a high-rate (as often as 0.83ms). A pair of consecutive matrices in this sequence are sufficiently similar that with very high probability they can be decomposed such that the system can interpolate between the resultant primitive matrices. The encoder in fact identifies segments of consecutive matrices in this sequence that can share the same decomposition parameters. The matrix at the end of a segment is the beginning of the next segment, hence needs to be decomposed into primitive matrices in two different ways. For instance, considering the pairs $(\mathbf{A}(t_1), \mathbf{A}(t_2))$, and $(\mathbf{A}(t_2), \mathbf{A}(t_3))$ separately, can provide different decompositions of $\mathbf{A}(t_2)$. Seed primitive matrices are only transmitted at segment boundaries, and within a segment only delta matrices need to be updated. Significant reduction in the number of matrix updates is realized by approximating many consecutive delta matrices within a segment by a single delta matrix, i.e., by making a straight line approximation of a piecewise linear trajectory, if the resultant deviation from the spatial intent is small.

The handling of rank deficient matrices has not been discussed here. However, modifications to the approach in Sec. 3.2 result in decompositions that are amenable for interpolation even when one or both of the two downmix matrices considered are not full-rank. Nevertheless, on rare occasion the greedy search may not be able to find any decomposition of two consecutive matrices such that they can be interpolated between. In such a situation, the system locally reverts to piecewise constant updates of the input primitive matrices, and the high sampling rate of the downmix matrix trajectory ensures that continuous motion of objects is still preserved.

4. RESULTS

The experiments described here employ as input 16 object, 48kHz, 24-bit pcm versions of the full-length movies listed in Table. 1, with position metadata specified every 32ms for each object. A comparison between two TrueHD-based systems that achieve the same objective - lossless transmission of the objects while simultaneously ensuring access in legacy decoders to a rendering of these objects to a 7.1ch speaker layout - is provided.

System A (**Simulcast**): The objects are rendered to 7.1 via the reference (in-house) Atmos renderer that linearly interpolates between downmix matrices at points where the metadata is specified. The resulting 8 channels, and the 16 input objects are independently encoded into two TrueHD streams.

System B (**Proposed Method**): The proposed augmentations to TrueHD are employed to encode the objects into a single hierarchical bitstream. The same renderer provides a sequence of sample downmix matrices which are interpolated between via primitive matrices.

The average bit-rate (Avg. Br.), peak bit-rate (Peak Br.), and compression ratio (CR) for the two systems are indicated in Table. 1. The CR for System A is based on 16+8 input channels and for System B, on 16 input channels. Clearly, significant savings in bit-rate are obtained by employing the proposed hierarchical scheme. In fact, for the item “Chicago” the peak bit-rate for the simulcast solution exceeds the mandated 18.64Mbps limit for TrueHD streams on a Bluray disc, making the solution infeasible in that case. The decrease in CRs for the proposed hierarchical system is expected - it is the price to pay for scalability. The 7.1 downmix from either approach is of very high quality, and artifacts as in a perceptual audio coder are not expected. The difference in interpolation strategies, however, results in slight differences in the downmix matrix trajectories. Neverthless, subjective comparison of the 7.1ch downmix from either method did not indicate any noticeable difference in the spatial scene.

Item $\downarrow$	**Avg. Br. (Mbps)**		**Peak Br.(Mbps)**		**CR**	
System $\rightarrow$	**A**	**B**	**A**	**B**	**A**	**B**
Chicago	12.18	8.74	19.26	13.85	2.26	2.1
Chasing Mavericks	9.86	7.01	17.72	12.55	2.80	2.62
Where the Trail Ends	9.36	6.82	15.34	11.48	2.95	2.69
Scooby-Doo	10.01	7.30	18.47	12.97	2.73	2.52

Table 1. Comparison of the simulcast solution (**A**) and the proposed system (**B**) in terms of bitrates and compression performance

5. CONCLUSIONS

The paper presents recent modifications to the Dolby TrueHD audio coder that enable lossless carriage of audio objects such that legacy playback devices incapable of rendering these objects can still access a downmix to a standard speaker layout directly from the bitstream. The dual objective is facilitated by the novel framework of interpolated primitive matrices, and an associated matrix decomposition strategy that effectively captures the dynamics of the downmix matrix trajectory in the interpolation of input primitive matrices alone and holds output matrices (needed to decode the downmix) constant, ensuring their conformance to legacy syntax. Experiments demonstrate that the proposed modifications result in significant bit-rate savings compared to an alternate solution that simulcasts the objects and their downmix.

6. REFERENCES

[1] M. A. Gerzon et al., "The MLP Lossless Compression System," in *17th International Conference: High-Quality Audio Coding*, September 1999.

[2] "Free Lossless Audio Codec," `http://xiph.org/flac/documentation.html`.

[3] A. Robinson, "Shorten: Simple lossless and near-lossless waveform compression," *Tech Rep CUED/F-INFENG/TR.156, Cambridge University*, December 1994.

[4] R. Yu et al., "A Fine Granular Scalable to Lossless Audio Coder," *IEEE Tran Audio Speech and Lang Proc.*, July 2006.

[5] R. Geiger et al., "ISO/IEC MPEG-4 High-Definition Scalable Adavanced Audio Coding," *Journal of the Audio Engineering Society*, January 2007.

[6] ISO/IEC 14496-3:2009, *"Information Technology - Coding of Audio - Visual Objects - Part 3: Audio"*, 2009.

[7] "DTS-HD Master Audio," `http://www.dts.com/professionals/sound-technologies/codecs/dts-hd-master-audio.aspx`.

[8] "Dolby Atmos Next-Generation Audio for Cinema," `http://www.dolby.com/us/en/technologies/dolby-atmos.html`.

[9] C. Q. Robinson, S. Mehta, and N. Tsingos, "Scalable Format and Tools to Extend the Possibilities of Cinema Audio," *SMPTE Motion Imaging Journal*, November 2012.

[10] R. Geiger et al., "IntMDCT - A link between perceptual and lossless audio coding," in *Proc. IEEE ICASSP*, May 2002.

[11] J. Li, "Reversible FFT and MDCT via Matrix Lifting," in *Proc. IEEE ICASSP*, May 2004.

[12] P. G. Craven, M. J. Law, and J. R. Stuart, "Matrix improvements to lossless encoding and decoding," US Patent 6611212 B1.