AURO®-CX™

Advanced Next Generation Audio Codec

Rev.: 1, Date: 2023-07-14

Contact Information

BELGIUM

NEWAURO BV

Oplombeekstraat 6

B-1755 Gooik

Tel: +32 2 793 25 50

info@auro-3d.com

www.auro-3d.com

Copyright notice

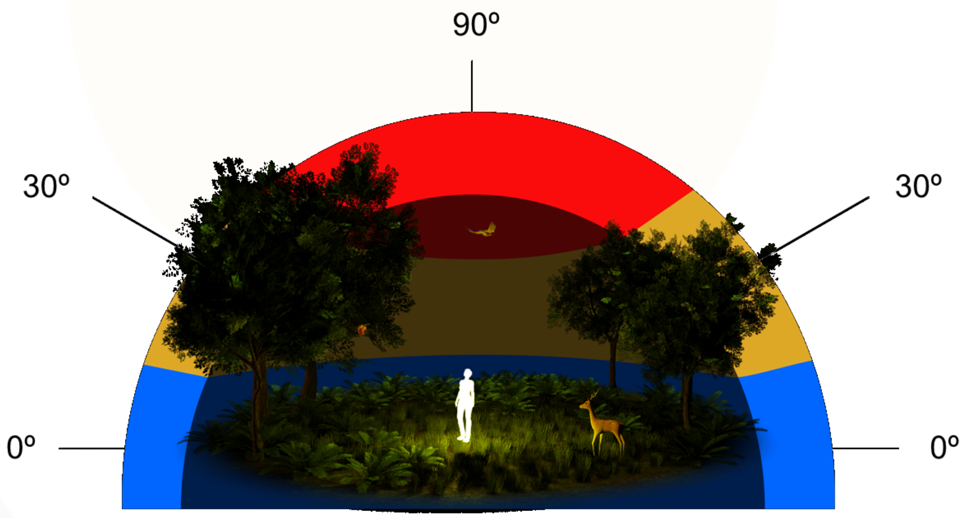
Table of Contents

1 AURO-3D® BACKGROUND

The reproduction of sound has seen an enormous evolution, even though the original intent has not changed: to deliver the *most natural, immersive listening experience* to the audience. Over the years, the technology and formats to achieve this have been evolving, starting with mono. Stereo recording was the first major revolution and has been (and still is) the main audio format for many decades. Only since the mid 1990's Surround recording became popular in consumer homes, with the 5.1 Surround format being the standard for more than 20 years. Amongst audio professionals, however, it was always felt that this established format, originally developed for cinema theatres, was still not able to achieve the old-age goal for natural sound.

In 2005 the AURO-3D listening format was introduced and finally the ultimate listening experience became available: 3-dimensional sound, sometimes referred to as multichannel sound with Height. Indeed, the introduction of the *Height* layer in the AURO-3D format is considered to be the ultimate step in providing a lifelike immersive experience to the audience, both in movie theatres and consumer homes.

Since 2011, the breathtaking AURO-3D immersive sound can be experienced at home thanks to the revolutionary Auro-Codec® decoder, integrated in over 2 million AVR devices, along with the deployment of speakers over 3 different layers – Surround, Height and Top – to make the most of the latest AURO-3D productions.



Full-fledged AURO-3D speaker setups that make optimal use of all three layers include up to 13.1 speakers to maximize the experience. More information on how to set up a home theater system can be found in the “Home Theater Setup Guidelines” on www.auro-3d.com.

2 AURO-CX – NEXT GENERATION AUDIO CODEC

With the advanced Auro-Cx NGA Codec, the high-quality AURO-3D immersive audio experience is now entering into the streaming era, with many new possibilities, previously not available on the market.

2.1 NGA – Next Generation Audio: Definition

The term 'Next Generation Audio', or in short NGA, was defined to identify the requirements for audio codecs with features going beyond traditional distribution of stereo and surround sound, to be integrated in new broadcast standards such as ATSC 3.0 and DVB-I.

Generally speaking, it was agreed that NGA codecs should provide the following typical features:

1. Support for Streaming for OTT and next-gen broadcast applications
    - OTT = Over-the-top; Broadcast = OTA = Over-the-air; cable, Satellite
    - Transmission in small, independent chunks of data
    - Using standard internet protocols (TCP/IP; HTTP)
    - Adaptive Bit Rate (ABR) support
2. Support for Immersive Audio formats, next to Stereo, Surround
3. Support for Object-based Audio (OBA)
    - Individual audio elements, paired with metadata for rendering in the playback device
4. Support for Personalized/Interactive Audio
    - Additional metadata in bitstream to enable user-controllable features
5. Support for Loudness-based Normalization and DRC
    - DRC = Dynamic Range Control
    - Alignment of perceived loudness depending on playback environment

2.2 AURO-Cx – More than just an NGA codec

During the development of Auro-Cx, all of the NGA requirements were taken into account, and more… The goal was to provide a streaming codec that can not only compete with other NGA codecs, but also brings the unique high-quality immersive experience that AURO-3D is known for. The result is a highly advanced codec bringing amazing audio quality, high flexibility and a great user experience. It supports:

- Scalable audio coding: Multiple audio waveform coding techniques, from high-resolution lossless to low-bitrate perceptual and parametric coding
- Support for Adaptive Bit Rate (ABR) with seamless switching between a wide range of bitrates
- Multiple audio formats: channel-based, object-based and scene-based audio

- Patented high-quality object rendering technology
- Personalization / Interactivity features: multiple languages, alternative audio tracks
- Scalable Channel Coding and Advanced Downmixing
- Support for Loudness and DRC metadata
- Built-in binaural rendering with the acclaimed Auro®-Headphones™ technology

2.3 Channels, objects…

As an NGA, Auro-Cx provides support for various ways to distribute Stereo, Surround and Immersive Audio:

- Channel-based: traditional distribution method, providing one signal for each speaker in the playback configuration. Supports Stereo, Surround (5.1/7.1), AURO-3D® (e.g. Auro 9.1 and Auro 13.1), up to 22.2.
- Object-based: separate elements of the content are transmitted as individual audio streams accompanied with metadata describing how to reproduce (render) the object in the playback system. Also supports new features such as personalization.
- Scene-based: support for transmission of Higher Order Ambisonics (up to 3rd order). Decoding support for FOA (First Order Ambisonics) only

Using these different paradigms, or a mixture thereof, Auro-Cx is the ideal codec for use in Home Entertainment (Music and Movies), Broadcast and VR/AR applications.

3 SUPPORTED AUDIO FORMATS

Whereas the original Auro-Codec only supported channel-based immersive audio, the Auro-Cx NGA codec also support object-based and scene-based (Ambisonics) audio formats. In this chapter, an overview is given of the different supported formats.

3.1 Channel-based audio (CBA)

In Channel-based audio (CBA) mixes are transmitted as they created and monitored in the recording studio by the engineers. Each audio stream represents a signal that is dedicated for a specific speaker. While this is the way audio has been recorded, mixed and distributed for many decades, it is still the preferred method for many audio professionals as it has the advantage of predictability and allows for specific operations such as the mastering process.

3.1.1 Channels and Configurations

Auro-Cx can be used for the transmission of channel-based audio with support for many formats up to 24 channels. The following channels are supported:

Channel Label	**Channel Name**	**ITU-R BS.2051 SP Label**	**ITU-R BS.2051 Ch Label**
L	Left	M+030	L
R	Right	M-030	R
C	Center	M+000	FC
LFE	Low Frequency Effects	LFE1	LFE1
Ls	Left Surround	M+110	Ls
Rs	Right Surround	M-110	Rs
Cs	Center Surround	M+180	BC
Lb	Left Back	M+135	BL
Rb	Right Back	M-135	BR
HL	Height Left	U+030	TpFR
HR	Height Right	U-030	TpFL
HC	Height Center	U+000	TpFC
T	Top	T+000	TpC
HLs	Height Left Surround	U+110	
HRs	Height Right Surround	U-110	
HCs	Height Center Surround	U+180	TpBC
HLb	Height Left Back	U+135	TpBL
HRb	Height Right Back	U-135	TpBR
LC	Left Center	M+022	
RC	Right Center	M-022	
LFE2	Low Frequency Effects 2	LFE2	LFE2
BtL	Bottom Left	B+045	BtFL
BtR	Bottom Right	B-045	BtFR
BtC	Bottom Center	B+000	BtFC

Channel Label	Channel Name	ITU-R BS.2051 SP Label	ITU-R BS.2051 Ch Label
BtLs	Bottom Left Surround	B+110	
BtRs	Bottom Right Surround	B-110	
Lw	Left Wide	M+045	
Rw	Right Wide	M-045	
TL	Top Left	U+090	
TR	Top Right	U-090	
M	Mono	M+000	

Any combination of up to 24 channels is supported. The following table shows an overview of a few typical formats that can be supported using these channels.

Configuration	Channels	ITU-R BS.2051
Stereo	L, R	System A
5.1 Surround	L, R, C, LFE, Ls, Rs	System B
7.1 Surround	L, R, C, LFE, Ls, Rs, Lb, Rb	System I
Auro 9.1 (5.1+4H)	L, R, C, LFE, Ls, Rs, HL, HR, HLs, HRs	System D
Auro 11.1 (5.1+5H+T)	L, R, C, LFE, Ls, Rs, HL, HR, HLs, HRs, HC, T	
Auro 11.1 (7.1+4H)	L, R, C, LFE, Ls, Rs, Lb, Rb, HL, HR, HLs, HRs	
Auro 13.1 (7.1+5H+T)	L, R, C, LFE, Ls, Rs, Lb, Rb, HL, HR, HLs, HRs, HC, T	
9.1.6	L, R, C, LFE, Lw, Rw, Ls, Rs, Lb, Rb, HL, HR, HLs, Hrs, TL, TR	
22.2	L, R, C, Ls, Rs, Lb, Rb, Lc, Rc, HL, HR, HC, HLs, HRs, HLb, HRb, T, BtL, BtR, BtC, LFE1, LFE2	System H

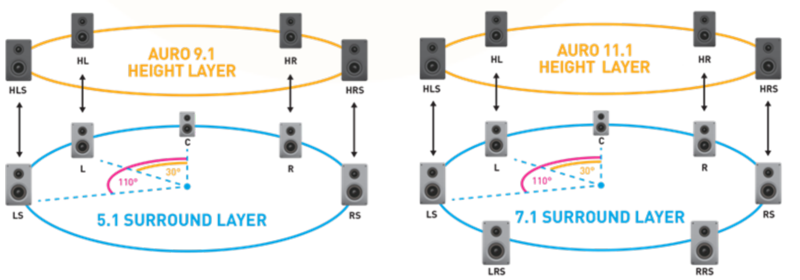


Figure 1 - Examples of 2-layer AURO-3D® Configurations: *Auro 9.1 (5.1+4H) - Auro 11.1 (7.1+4H)*

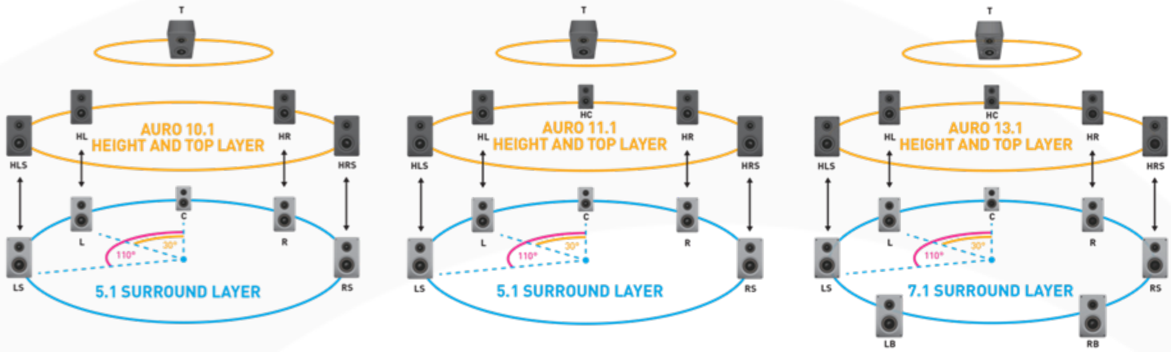


Figure 2 - Examples of 3-layer AURO-3D® Configurations:

Auro 10.1 (5.1+4H+T) - Auro 11.1 (5.1+5H+T) - Auro 13.1 (7.1+5H+T)

3.1.2 Downmix Engine

In case a transmitted channel configuration is too large for the installed playback system, e.g. when trying to playback an Auro 13.1 mix on an Auro 9.1 speaker system, Auro-Cx will provide a optimized downmix for the target playback system. The advanced Downmix Engine allows for artistically-specified downmix gains and routings, called Downmix Movements.

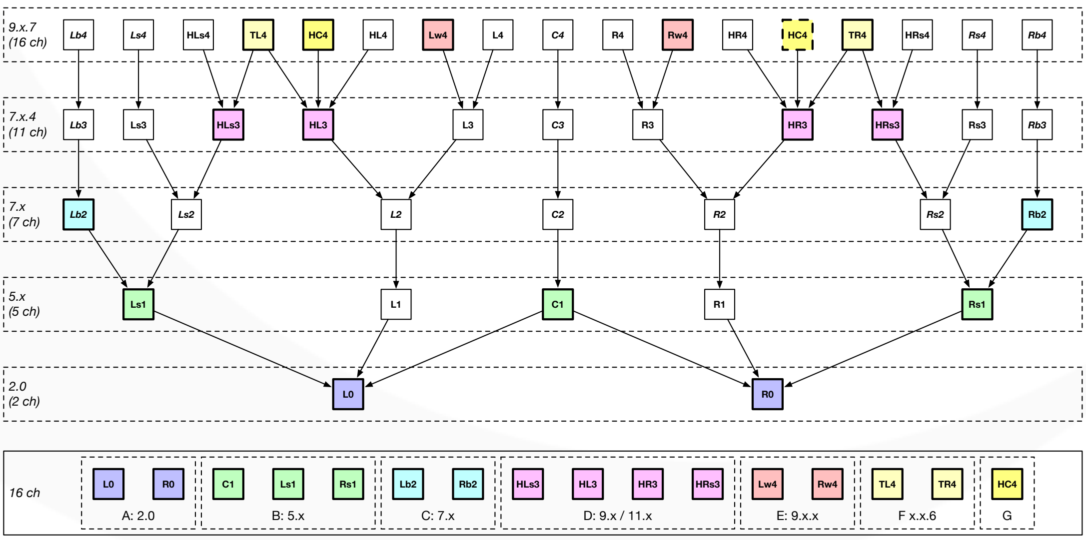
The following downmix movements are defined: Down, Front, Side, Back, Up, LFE, Sum, Split, Trio, Corners. Each source channel type can support up to three movements, depending on its original position and target layout.

3.1.3 Scalable Channel Coding

The Downmix Engine is also the foundation for the Scalable Channel Coding concept.

Instead of transmitting all individual channels of the mixes original format, Scalable Channel Coding applies a cascaded downmix scheme down to a stereo signal which can be decoded by all decoder profiles. Using additional groups of channels allows then to losslessly retrieve other formats from the stereo downmix using a pre-defined, cascaded decoding scheme.

The following diagram shows a downmix from an original 16-channel mix, down to a stereo signal.



If the decoder is requested to output a stereo signal, then it only needs to decode the two channels (L0, R0). Using the different channel groups, it is then possible to further retrieve the 'higher' format levels step-by-step, in its entirety or partially, as needed for the requested playback format, or up to the capabilities of the decoder (as defined by its profile).

3.2 Object-based Audio (OBA)

Next to the already advanced channel-based audio capabilities, Auro-Cx also provides support for object-based audio (OBA), introducing many capabilities previously not achievable with traditional codecs.

In Object-based audio, various elements of a mix are transmitted separately, typically as mono audio streams, accompanied with metadata that describes how a renderer in the playback system should reproduce the audio objects as part of the complete mix.

One potential advantage is that the mix (or parts thereof) becomes independent of the monitoring system used in the recording studio and can be optimally reproduced for the target playback system. Alternatively, audio objects provide new options for personalization and interactivity, allowing the listeners to adapt the output to their own preferences.

To achieve this, an advanced bitstream was developed for Auro-Cx providing support for:

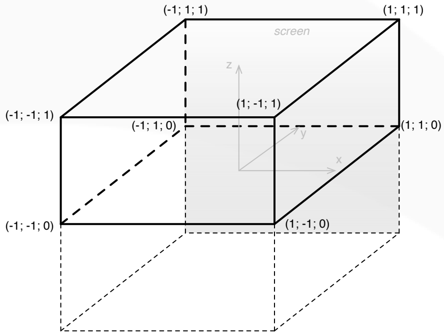
- Dynamic (moving), immersive audio objects.
- Interactive audio features, such as:
    - Multiple languages
    - Dialog Enhancement
    - Additional commentary and assistive audio tracks
    - Interactive gains and positions of objects

3.2.1 Immersive, dynamic objects

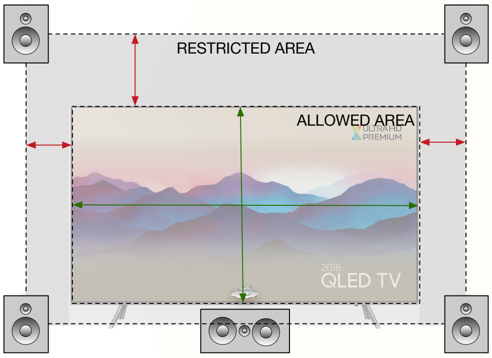
Dynamic objects are audio objects that are an inherent part of the entire mix. Thanks to the use of positional metadata, it is possible to optimally render the objects on any target playback system. This is especially useful for direct sounds that require precise positioning or movement on systems with many speakers, such as the cinematic AuroMax system with up to 32 speaker positions, or for virtualization technologies that can optimally render sounds using this information. Typical applications are sound effects, like fly-overs, laser-gun shots, etc in movies or moving synth sounds in e.g. dance music.

Note that using objects becomes less useful to 'simulate' channels. In many situations, multi-channel recordings require each microphone signal to be reproduced by a single speaker to faithfully reproduce the intended correlation between the signals. Similarly, more diffuse signals such as reverbs or ambient sounds have no particular advantage from being reproduced by an object instead of a channel. For this kind of applications, the immersive channel-based bed configurations in Auro-Cx provide a more efficient solution, both in quality as bandwidth.

For those applications where the use of objects does bring a better experience, Auro-Cx provides positional metadata using the Cartesian coordinate system (x,y,z), with a room-centric approach. This means, object positions are defined in relation to a 'nominal room' and scale with the playback system, in the same way that engineers are used to when using any traditional panning system.



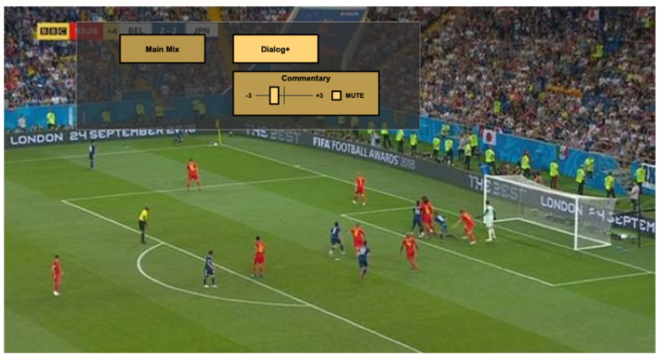
Thanks to the advanced, patented ESPCAP (Enhanced Speaker Position Compensated Amplitude Panning) object rendering algorithm, objects are rendered exactly where the mixing engineer wants them, on any compatible speaker configuration with up to 32 speakers. ESPCAP also provides additional advanced features such as Spread, Snap-to-Speaker and Zone Control that provide greater control and elevate the total experience to a new level.



Thanks to the 'on-screen' metadata flag, it is even possible to indicate whether on audio object is related to a visual event on the screen, allowing the renderer to scale the object's position and movements such that it remains locked to what happens on screen, even when the speakers are placed further away.

3.2.2 Personalized audio using static objects

With the introduction of object-based audio, new experiences can be brought to the audience in the form of Personalized or Interactive audio, allowing the listeners to adjust the audio to their own preference, within the limitations provided in the metadata.



Typical applications are:

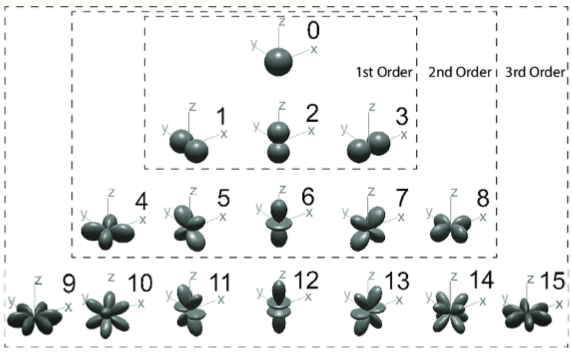
- Multiple languages within a single stream: the listener can then switch between the provided languages seamlessly.
- Dialog enhancement: the listener can adjust the level of the dialog, which is especially useful to improve intelligibility for hearing-impaired viewers.

- Additional content, such as commentary tracks or assistive audio: the listener can turn certain optional audio streams on or off, e.g., home versus visitors' team radio, directors' commentary, etc...
- Alternative mixes: the listener can choose between multiple versions of the same content within the same stream, e.g., an ambiance-only version of a sports-event or a dedicated binaural version for headphones,
- And much more…

3.3 Scene-based Audio (Ambisonics)

Next to the Channel-based and Object-based audio formats, Auro-Cx also provides limited support for Ambisonics.

In Ambisonics, a 3D spatial sound field is encoded using 'spherical harmonics', where each audio stream represents a spatial component. Combining these components in the correct way then allows to reproduce the sound field on any spherical or, within limits, so-called irregular speaker system. Thanks to the powerful mathematics behind Ambisonics, it is often the preferred choice for use in VR applications, as applying movements such as rotations can be performed as simple mathematical operations. This is especially useful for binaural reproduction with head-tracking.



In theory, Ambisonics requires an infinite amount of 'orders' to fully represent the sound field. However, good results can already be obtained with a limited number of orders. Each order requires more audio streams for transmission and decoding. The most used format is 'First Order Ambisonics' (FOA), which uses 4 harmonics: W (omni-direectional), X, Y, Z (figure-of-eight). This is because FOA is the only order that can also be recorded using standard microphones or using specific commercially available compact microphone setups. (Note: HOA Microphones also exist but are less popular.)

Higher Order Ambisonics (HOA), requires more complex representations, which can be created via software that has become more readily available in recent years. Many people agree that a very good sound can be obtained using 3rd or 5th order Ambisonics, which require 16 and 64 channels respectively (number of spherical harmonics = 2^{N+1}).

Auro-Cx has built-in support for First Order Ambisonics decoding and supports the transmission of up to 3rd order HOA. The 'HOA Transmission' mode then requires the use of an external HOA decoder. Because of this, HOA audio streams cannot be combined with other elements (CBA or OBA) within a single Auro-Cx program.

4 AURO-CX PLAYBACK

Auro-Cx brings many new possibilities for content creators and streaming services. However, a highly complex eco-system of encoders, services, players and decoders is needed to fully leverage its capabilities. The Auro-Cx decoder is therefore developed to be as flexible and efficient as possible, for integration in many playback scenarios. In this chapter, a number of these scenarios will be highlighted.

4.1 Binaural rendering on headphones

Mobile devices have become the most-used playback device for any type of content, including movies and music with immersive sound. High-quality, realistic binaural rendering on headphones is therefore an indispensable feature for any NGA system.

The built-in Auro-Headphones advanced binaural processing in the Auro-Cx decoder delivers the most natural sounding immersive experience for playback on any type of headphones. Thanks to its unique architecture, it performs real-time binaural rendering of the immersive sound on all kinds of devices, such as smartphones, tablets, PC, etc…

4.1.1 Binaural room simulation

The Auro-Headphones processing in Auro-Cx uses advanced binaural room simulation in 3D, which is critical in achieving a natural-sounding binaural experience.

Since the room simulation is performed in real time, several user presets are provided to allow the listener to select the sound that provides their best, personal experience.

4.1.2 Binaural control through metadata

The Auro-Cx bitstream also features several metadata items dedicated to optimally control the binaural rendering from an artistic point of view. In some case, creators prefer certain elements of the content, e.g. the background music, to be simply reproduced as regular stereo while other elements, such as the moving objects related to visual elements, should be represented in binaural 3D. Similarly, it might be preferable that certain binaural elements follow head-tracking info, while others do not. Auro-Cx provides all tools to define this kind of behavior.

Along the same lines, it is possible for a content creator to overrule the algorithmic room simulation of Auro-Headphones, to ensure that even the binaural 3D reflections match the intent of the content.

4.2 Auro-Cx in Home Audio

AURO-3D has always been associated with high quality immersive sound and Auro-Cx stays true to this promise.

Auro-Cx decoders will become available in all kinds of home entertainment systems such as AV-receivers, soundbars, smart speakers, TV's, etc…

4.2.1 Native Auro-Cx decoding

The Auro-Cx decoder in home audio devices provides high-quality decoding and rendering to various speaker setups, from stereo up to Auro 13.1. For high-end applications, there is even the advanced AuroMax for Home renderer, providing playback on systems with high speaker counts, up to the cinematic AuroMax 26.1 system.

In streaming applications, content encoded with Auro-Cx is played back by a media player application, often running on a settop box, smart TV or other streaming media player, which then outputs the Auro-Cx bitstream to a decoder. Typically, this is done by transmitting the bitstream via HDMI or HDMI-eARC to a decoder in the playback device. Alternatively, the media player is integrated in the decoding device itself, removing the need for the HDMI transmission.

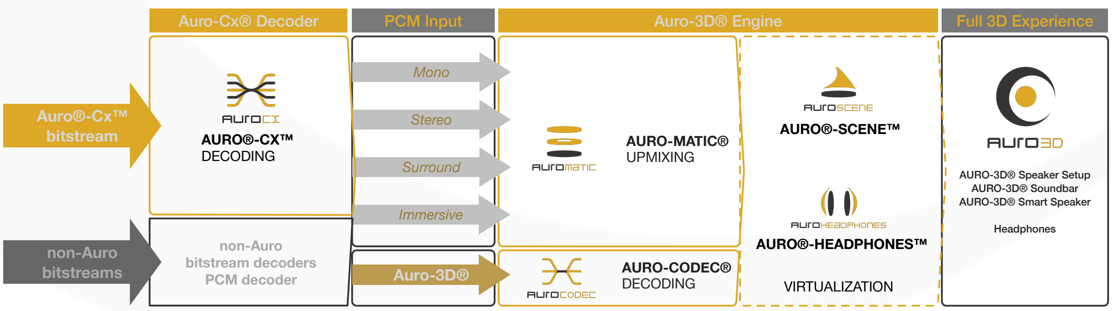
4.2.2 Auro-Codec transcoding

Integrating the Auro-Cx decoder in home audio devices enables the best possible quality and the most advanced experiences but it will require some time before these devices are readily available.

In the meantime, already more than 2 million devices support the original, premium Auro-Codec decoder. Auro-Cx decoders in media players can therefore also support real-time transcoding to an Auro-Codec encoded output, a multichannel PCM signal that can be transmitted via HDMI and decoded by any existing compatible AV-receiver, preserving the high-quality immersive sound experience provided by Auro-Cx.

4.2.3 Auro-Cx and the Auro-3D Engine

To provide the full AURO-3D experience, the Auro-Cx decoder is also tightly integrated with the Auro-3D Engine. This also includes the acclaimed Auro-Matic upmixing technology and Auro-Scene virtual speaker technologies and makes it the perfect and complete solution for advanced home audio devices such as soundbars and smart speakers.



5 AURO-CX STREAMING

Auro-Cx was developed specifically with streaming in mind, bringing the high-quality AURO-3D immersive and interactive sound experience to OTT and broadcast applications.

Thanks to the audio coding algorithms and advanced bitstream definition, Auro-Cx can achieve this at various bitrates, starting with high-resolution lossless audio at higher bitrates, down to high-quality perceptually coded audio at bit rates as low as 256 kbps for Auro 11.1 (7.1+4H).

5.1.1 File formats

For maximum compatibility, the Auro-Cx bitstream can be stored and transmitted in various file formats. The most basic format is the proriétary .a3ds file format, intended for mastering and archiving.

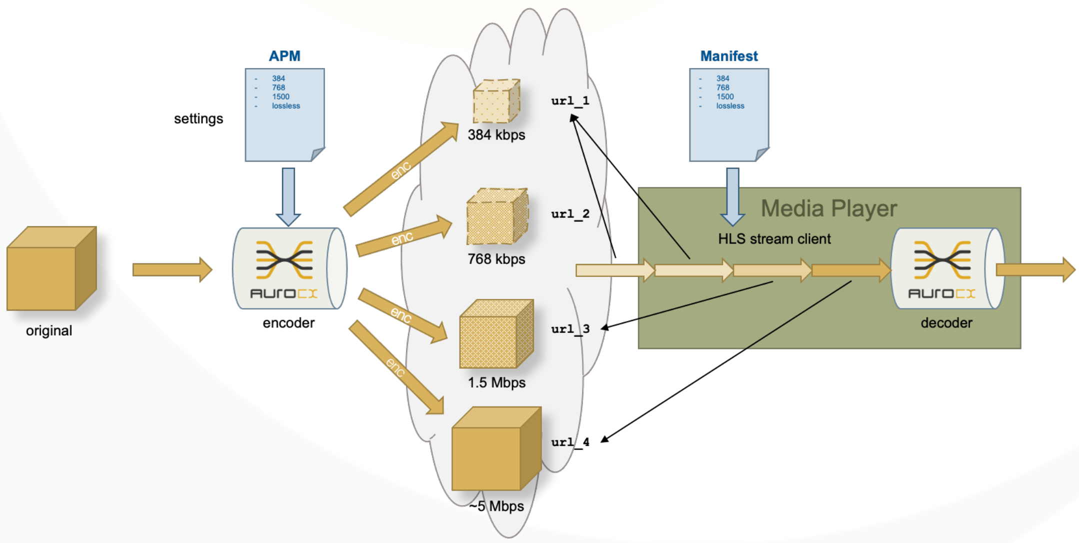
In most applications the Auro-Cx bitstream will be embedded in a (fragmented) .mp4 or .m4a file according to the ISOBMFF standard, with or without accompanying video, makgin it compatible with many standard playback systems in the market.

5.1.2 Secure and robust

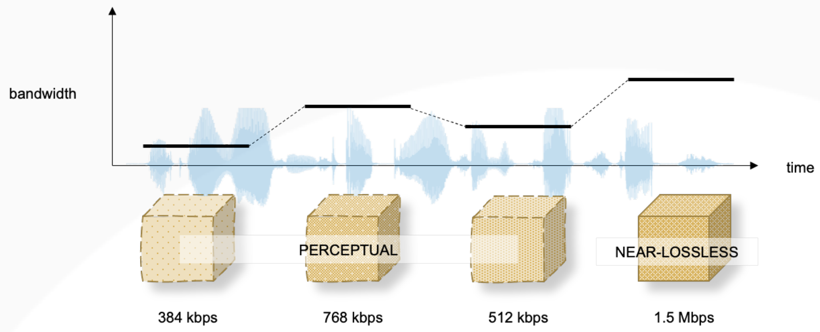
Thanks to the patented sync-algorithm, the Auro-Cx bitstream is designed to be both secure and robust. The protocol is built to reduce the risk for false-positive detection to an absolute minimum. It also contains a built-in obfuscation mechanism, making it much harder to interpret the data without deep knowledge of the protocol. This makes Auro-Cx probably the safest audio codec on the market.

5.1.3 Adaptive Bit Rate (ABR)

The Auro-Cx encoder can encode a single master at multiple target bitrates to enable full Adaptive Bit Rate support.



Using standard protocols such as MPEG-DASH or HLS, a compatible media player can seamlessly switch between the different provided bitrates, depending on the available bandwidth and/or player's capabilities, without any audible clicks, pops or other artefacts.



6 AURO-CX AUDIO CODING

The main function of any audio codec is to compress (encode/decode) one or more individual audio streams or 'waveforms', with the aim to reduce their required bandwidth for distribution or storage.

Auro-Cx supports several techniques to encode audio streams, supporting a wide range of distribution scenarios and providing solutions for even the most demanding audio professionals.

An important aspect of any audio codec is the compression factor, the ratio between the original bandwidth of the uncompressed audio signal (in Linear PCM) and the encoded signal (typically expressed as x:1). The higher the compression factor, the lower the bandwidth of the encoded signal.

Bandwidth is typically expressed in kilobits per second (kbps) or megabits per second (Mbps).

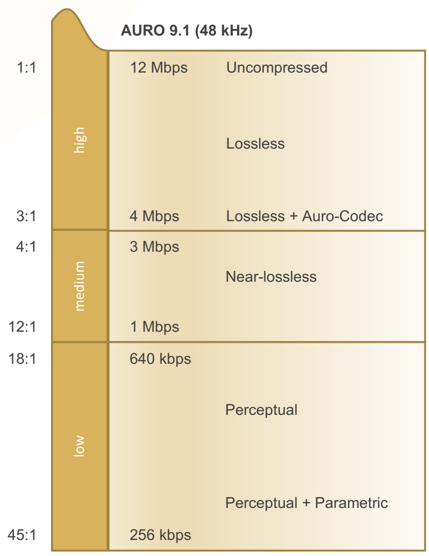
E.g., an uncompressed Auro 9.1 signal (10 channels, 48 kHz, 24 bit linear PCM), typically requires 11.520 Mbps. A highly compressed version may only require as little as 256 kbps (45 times less bandwidth!).

6.1 Three waveform coding modes

In its simplest form, one or more audio streams are encoded using techniques that allow to directly retrieve the original waveform as close as possible without the need for additional audio signals. In Auro-Cx, this kind of audio coding is referred to as waveform coding.

Auro-Cx provides a flexible, scalable way of encoding the audio at different bitrates, providing lossless, near-lossless and lossy coding modes.

- Lossless mode: The decoded audio is bit-exact to the original audio. Typical compression factors range from 1.8:1 to 2.5:1, depending on the content.
- Near-lossless or Transparent mode: This mode uses similar techniques as the Auro-Codec® to provide better bit-rate reduction, with minimal loss in audio quality (measurable, but inaudible). Typical compression factors range from 4:1 to 12:1, depending on the content.



- Lossy mode: similar to various other codecs in the market, this mode uses perceptual coding techniques to further reduce the bitrate with minimal audible loss. Compression ratios of more than 10:1 (even up to 45:1) are achieved.

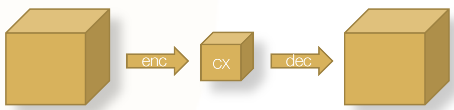
A unique feature of the Auro-Cx codec is that, depending on the encoder, different encoding modes can be used simultaneously in a single stream, e.g. to provide high-quality sound for the main program combined with lower quality audio for optional commentary tracks, etc. In the table below, an example is shown of the resulting bandwidth for an Auro 9.1 signal using the different waveform coding modes.

Waveform coding mode	Bandwidth (kbps)	Compression factor (x:1)	File size (MB per minute)
Original (L-PCM)	*11,520*	*1*	*86.4*
Lossless	6,400~4,000	1.8~3	48~30
Near-lossless (transparent)	3,000~1,000	4~12	22.5~7.5
Perceptual (lossy)	768~256	15~45	5.8~1.9

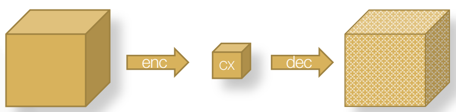
Next to the three main waveform coding modes, Auro-Cx also uses a number of parametric coding techniques, to achieve even higher compression ratios for more complex signals (see further).

6.1.1 Auto-regressive Waveform Coding (AWC)

The Lossless and Near-Lossless (Transparent) compression modes make use of the Auto-regressive Waveform Coding technique, which operates in the time-domain to provide excellent behavior with transients and avoid time-smearing and phase-distortion. AWC provides for high-resolution audio, up to 192 kHz sampling rate.



The **lossless** mode is the highest-quality coding mode for applications that require reproduction at original quality as created in the studio, without bandwidth limitations. The output of the AWC audio signal generator is then bit-for-bit identical to the original master created in the studio. The resulting average compression ratio will typically be lower than with other compression modes: between 1.6:1 and 2.5:1, depending on the complexity of the audio signal.



Auro-Cx also provides the **near-lossless** (transparent) mode, a very high-quality coding mode for applications that require reproduction at a transparent quality as close as possible to the master created in the studio. Bitrates will be lower than when using the lossless mode, but higher than with lossy, perceptual coding. The output after decoding is slightly different from the original master, but with inaudibly low differences: the introduced noise level will be lower than the masked hearing threshold.

Waveform coding mode	Max Sample rate (kHz)
Lossless	192
Near-lossless (transparent)	96

6.1.2 LFE Coding

For Low Frequency Effects (LFE) channels, Auro-Cx uses a dedicated waveform coding technique that takes advantage of the limited frequency range used by this kind of signals.

The original audio stream is subsampled to an ultra-low sample rate, typically around 300 Hz,

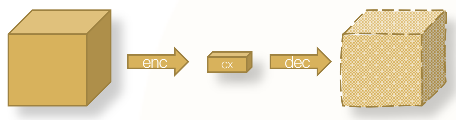
which already provides a huge reduction in bandwidth (160:1) by removing all content above 150 Hz.

This mode is not used for lossless coding, even if the original signal would not contain any signal above the cut-off frequency.

6.1.3 Perceptual Waveform Coding (PWC)

While the lossless and near-lossless modes provided by AWC already result in reduced bandwidth, the compression ratio is often rather high for streaming applications requiring much lower bitrates.

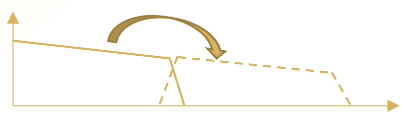
For these applications, the Perceptual (Lossy) compression mode of Auro-Cx uses the Perceptual Waveform Coding (PWC) technique. PWC operates in the frequency-domain using the MDCT



(Modified Discrete Cosine Transform) time-frequency transform, allowing easier removal of redundant information based on perceptually motivated quantization of the coefficients.

Thanks to the optimal use of psycho-acoustical principles behind the perception of spatial (= immersive, 3D) audio, and correlation between various audio streams, Auro-Cx can often achieve better audio quality than many other existing audio codecs that simply use multiple mono or stereo waveform coders simultaneously.

At the lowest bitrates, higher compression factors are achieved using the intelligent Bandwidth Extension (BWE) feature, which is capable of reconstructing the higher frequency bands from the lower audio bands, using a minimum amount of data.



6.2 Three Parametric Coding techniques

While the waveform coding modes in Auro-Cx already provide a wide range of compression factors at high-quality, additional techniques are required to further optimize the use of the bandwidth as programs become more complex, or bitrates become lower. For this reason, Auro-Cx also contains multiple Parametric Coding techniques that complement the waveform coding modes for different applications.

Parametric coding technology is used to encode/decode one or more groups of audio streams (waveforms) by combining their signals into a lower number of individual (main) waveforms with the aim to reduce their bandwidth and/or file size. A low bitrate stream of side-info (parametric coefficients) allows to extract the individual audio streams again after decoding of the 'carrier' waveforms. The carrier waveforms are encoded using one of the main waveform coding modes described above.

Auro-Cx supports the following parametric coding techniques:

- Parametric Channel Coding (PCC): allows to extract one or more channels from a downmixed 'carrier' signal.
- Parametric Object Coding (POC): allows to extract individual (dynamic) Object signals from a multi-channel Bed.
- Auro-Codec Compatible Coding (ACC): similar to PCC, but with the main difference that the 'carrier' signal is compatible with the original Auro-Codec and can therefore be sent to any existing device with an Auro-Codec decoder.

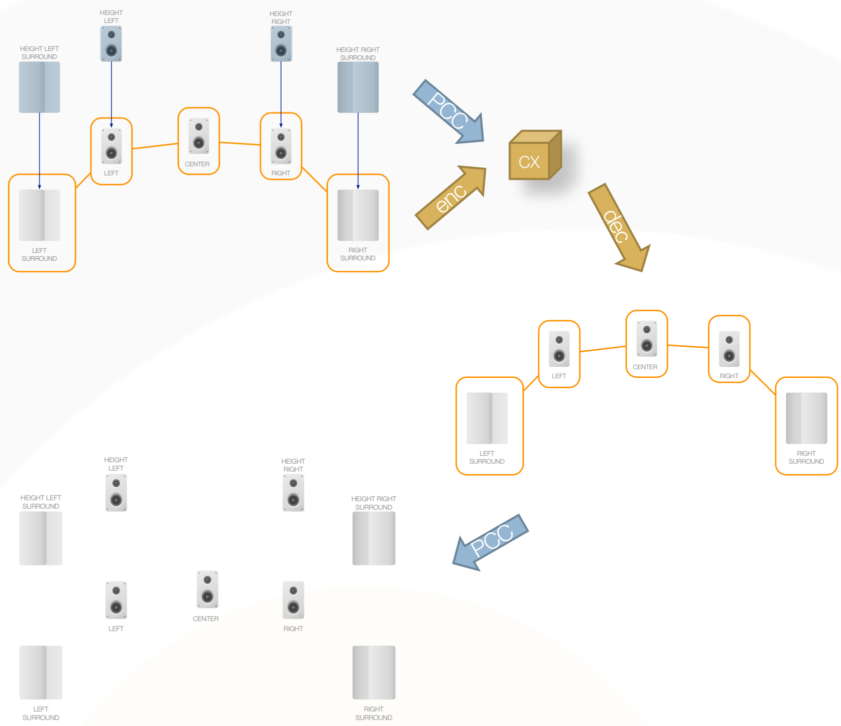
6.2.1 Parametric Channel Coding (PCC)

Similar to the Auro-Codec (see further), the Parametric Channel Coding (PCC) technology reduces the number of audio streams for encoding by downmixing the original audio and combines this with low bitrate side-info (typically around 3 kbps per channel) describing how to extract additional channels at decoding. This is especially useful to achieve higher audio quality for channel-based audio elements (Beds) at very low target bitrates.

The downmixed signal is then encoded using any of the waveform coding mode described earlier and since this now has a lower number of channels results in a lower bitrate.

A typical application is the encoding of Height and/or Surround information into a Surround or Stereo 'carrier' signal. The original format with M channels is the downmixed into a carrier with N channels, already providing a bandwidth reduction of M/N:1, independent of the waveform coding used to encode the N carrier signals.

Another advantage is that the carrier signal can directly be played if the decoded output configuration is equal to or smaller than the configuration of the carrier signal, reducing decoding complexity.



Some examples:

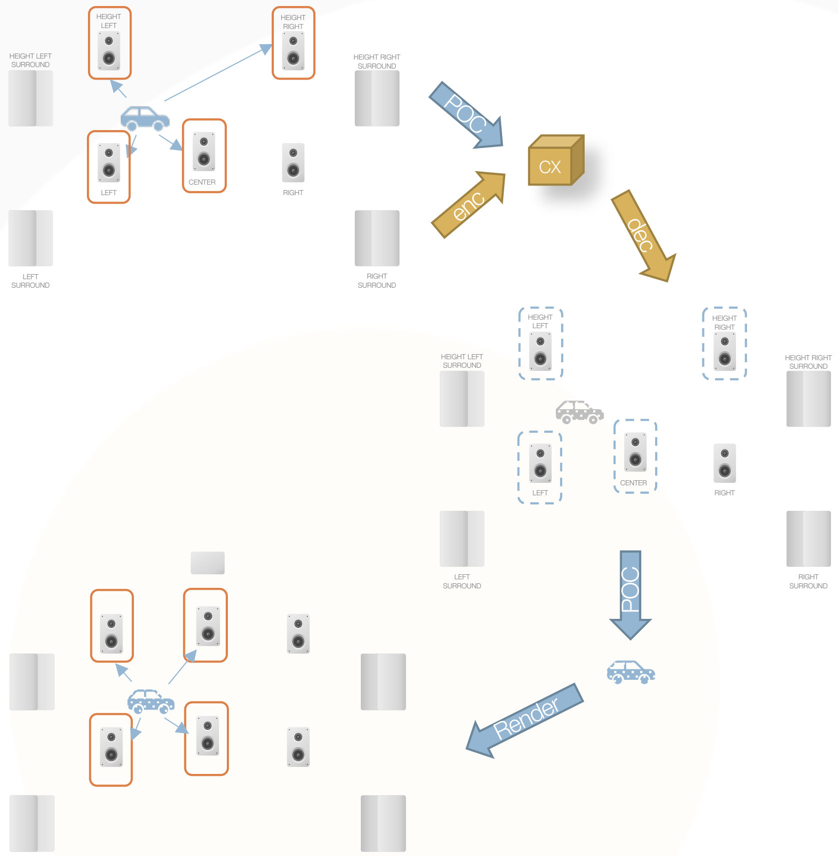
- Auro 9.0 in 5.0 carrier (M = 9; N = 5; average compression: 1.8:1)
- 5.0 original in 2.0 carrier (M = 5; N = 2; average compression: 2.5:1)
- Auro 13.0 in 7.0 Surround (M = 13; N = 7; average compression: 1.85:1)
- 22.0 in Auro 13.0 (M = 22; N = 13; average compression: 1.69:1)
- Auro 9.0 in 2.0 (M = 9; N = 2; average compression: 4.5:1)
    - *Note: theoretically possible but not used due to degraded sound quality*

6.2.2 Parametric Object Coding

When using dynamic objects (see further), the number of audio streams that needs to be individually encoded can quickly become high, requiring a very high bitrate to maintain acceptable audio quality for all streams. However, in many scenarios, it is not required to have access to the individual audio streams of each object. Indeed, it can be argued that individual rendering of each object only makes sense for playback systems with rather high speaker counts (e.g., an AuroMax 26.1 system) or for advanced virtualization systems that require access to each individual object. In most playback environments for home, mobile and car, high-quality pre-rendering to a channel-based carrier often leads to better quality results.

In Auro-Cx this is facilitated by the Parametric Object Coding (POC) technology, which reduces the number of audio streams for encoding by rendering and downmixing immersive (dynamic) audio objects into the accompanying bed, combined with low bitrate side-info (approx. 5 kbps per object) describing how to extract the objects from the bed at decoding. Again, this is especially useful to achieve higher audio quality at very low bitrates.

With the object audio now part of the bed channels' audio, the rendered and downmixed signal can be encoded using any waveform coding mode, resulting in lower bitrates as the individual audio streams of the objects do not need further encoding.

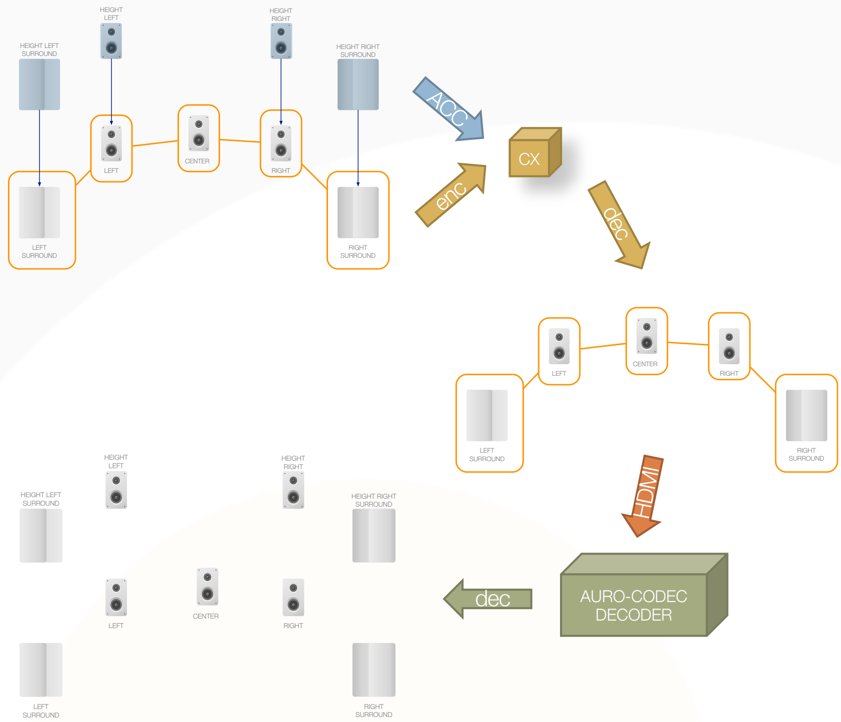


A typical application is the encoding of Immersive Objects audio information (OBA) into a related channel-based Bed (CBA). It is not applicable for interactive objects (e.g., multilanguage) due to increased crosstalk. The channel based bed can be in any format, including an AURO-3D format, which means there is no need for object-based rendering to achieve a true immersive experience. It also provides backwards-compatibility with decoders without OBA rendering.

A Surround Bed can be used to further reduce bitrate. However, height will then need to be recreated using objects at playback.

6.2.3 Auro-Codec Compatible Coding (ACC)

A third, unique parametric coding technique in Auro-Cx is the Auro-Codec Compatible Coding mode. Similar to PCC, this technology reduces the number of audio streams for waveform encoding by downmixing the original audio, combined with side-info describing how to extract additional channels at decoding. A major advantage of this technology is its backwards compatibility with existing Auro-Codec decoders, which means that the Auro-Cx decoder can create an Auro-Codec encoded surround signal that can be transmitted via HDMI to any existing Auro-Codec compatible device, providing the same experience and quality, but at lower bitrate.



A typical application is the encoding of Height and/or Surround information into an Auro-Codec encoded Surround or Stereo 'carrier' signal. The original format with M channels is the downmixed into a carrier with N channels, already providing a bandwidth reduction of M/N:1, independent of the waveform coding used to encode the N carrier signals.

A major difference with PCC is the fact that ACC also provides support for high-resolution audio up to 96 kHz/24 bit, also for immersive audio. However, where PCC can be combined with any waveform codec, including low bitrate perceptual coding, ACC can only work with lossless coding of the carrier waveforms.

Some Examples

- Auro 13.1 in 7.1 Surround (M = 14; N = 8; average compression: 1.75:1)
- Auro 9.1 in 5.1 Surround (M = 10, N= 6; average compression: 1.67:1)

7 CONCLUSION

While Auro-Cx is not the first NGA codec on the market, its unique features make it a very attractive alternative for streaming of high-quality immersive and interactive audio applications, thanks to the following features:

- Scalable audio coding, from high-resolution lossless audio to high-quality low-bit perceptual coding
- Support for flexible channel-based, object-based and scene-based audio
- Support for dynamic object-based audio with advanced rendering to many speaker setups
- Support for interactive audio applications, such as multiple languages, dialog enhancement with user-control, optional audio streams, etc…
- Support for seamless switching between different bitrates
- Built-in high-quality binaural rendering
- Full integration in the AURO-3D eco-system, including real-time transcoding to Auro-Codec encoded PCM on HDMI

Contact info@auro-3d.com for more information and request a demo now!