

TECHNICAL PAPER

THE AAC-ELD FAMILY FOR HIGH QUALITY COMMUNICATION SERVICES

Since 1994, when the development of AAC in MPEG-2 was initiated, five generations of AAC codecs have evolved. The audio codecs are designed for meeting every possible need in the fields of communications, broadcast and streaming.

The AAC-ELD family consists of AAC-LD, AAC-ELD and AAC-ELD v2. The state-of-the-art MPEG-4 audio codecs are designed for maximum speech and audio quality at very low coding delay and are therefore all excellent solutions for professional and consumer communication applications.

This paper introduces the three members of the AAC-ELD family and gives a closer look at how to tackle coding delay.

Fraunhofer Institute for
Integrated Circuits IIS

Management of the institute
Prof. Dr.-Ing. Albert Heuberger
(executive)

Dr.-Ing. Bernhard Grill
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmf.fraunhofer.org

Contact China
Toni Fiedler
china@iis.fraunhofer.de

Contact Japan
Fahim Nawabi
Phone: +81 90-4077-7609
fahim.nawabi@iis.fraunhofer.de

Contact Korea
Youngju Ju
Phone: +82 2 948 1291
youngju.ju@iis-extern.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

THE MEMBERS OF THE AAC-ELD FAMILY

AAC-LD (Low Delay AAC), AAC-ELD (Enhanced Low Delay AAC) and AAC-ELDv 2 are optimized for a low algorithmic delay, which is essential for natural real-time communication. In contrast to common speech codecs, they extend the application area from clean voice to a broad variety of source material, including voice and singing, music and ambient sounds. Due to their technical superiority, the three members of the AAC-ELD family are represented all across the field of telecommunication, including Over-the-Top (OTT) services, video telephony, video conferencing and telepresence, as well as broadcast contribution services. The highly successful Apple FaceTime is just one example of a video telephony application that relies on the quality of AAC-ELD. The codec is also natively included in the operating systems iOS, Android and Mac OS X.

The AAC-ELD family delivers a new level of audio quality which is called Full-HD Voice. Unlike Plain Old Telephone Services (POTS), ISDN and mobile phone calls, Full-HD Voice offers an unsurpassed level of quality, resulting in calls that sound as clear as talking to someone in the same room, or listening to high-quality digital audio. This is possible as the codecs support the full audio bandwidth of 20 kHz.

In addition to the millions of calls already being made today by using AAC-ELD, this technology is set to enable many new Full-HD Voice applications, including telepresence at home and mobile rich media telephony.

The three family members can be regarded as a superset of each other, as they share the same coding core and each adds new coding tools [Figure 1]. Software of the AAC-ELD codec family can be expected to be fully backward compatible. The codecs can handle mono, as well as stereo and multi-channel signals - all with latencies as low as 15 ms and at a wide range of bit rates (down to 24 kbit/s) and sampling rates.

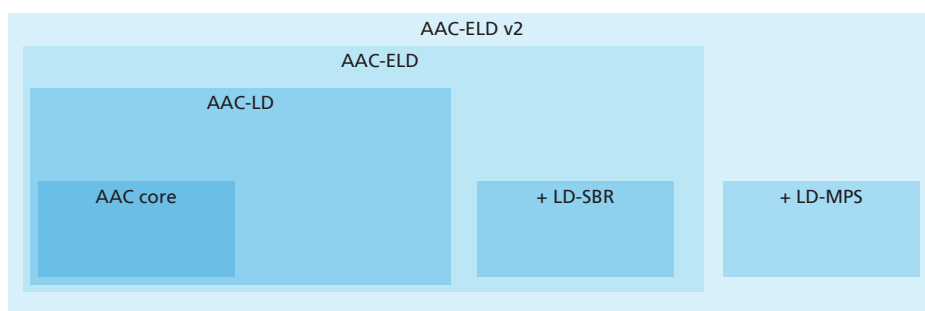


Figure 1: AAC-ELD family.

The audio quality and operating point of the AAC-ELD family members is described in Figure 2 for stereo audio. While AAC-LD is a very good choice for bit rates above 96 kbit/s, AAC-ELD improves the audio quality down to 48 kbit/s. Below this bit rate, AAC-ELDv2 is the best choice to keep the audio quality high. For mono applications, a similar relationship between AAC-ELD and AAC-LD at half bit rate can be expected, whereas AAC-ELD v2 delivers identical audio quality to AAC-ELD.

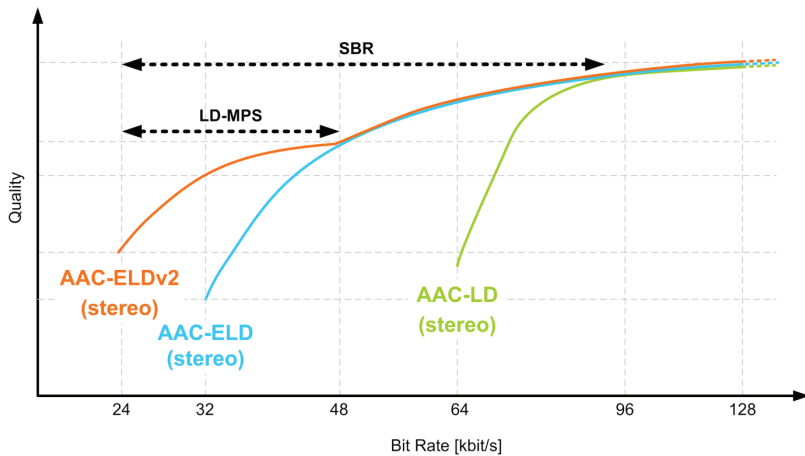


Figure 2: AAC-ELD stereo operating points.

A CLOSER LOOK AT AAC-LD

The core structure of AAC-LD is directly derived from AAC. The time domain input samples are transformed into a frequency domain representation by an MDCT (or Low Delay MDCT in case of AAC-ELD) filter bank. The 960 (or 1024) sample size of the MDCT analysis window utilizes a frequency resolution of 50 Hz and a time resolution of 10 ms. These are chosen to efficiently exploit psychoacoustic effects of frequency and time domain masking.

As natural audio signals show diverse signal characteristics, specialized tools take care of them:

- Temporal Noise Shaping allows the AAC-LD coder to exercise control over the temporal fine structure of the audio signal and improve the time resolution.
- Intensity Coupling and Mid/Side Stereo increase the coding gain for a stereo channel pair compared to encoding two mono channels separately.
- Perceptual Noise Substitution (PNS) uses a parametric representation of noise-like frequency bands for an efficient transmission.

The codec can operate in a fixed frame length mode where every packet is equal in size, or in a fixed bit rate mode where the average bit rate within a limited time frame is constant.

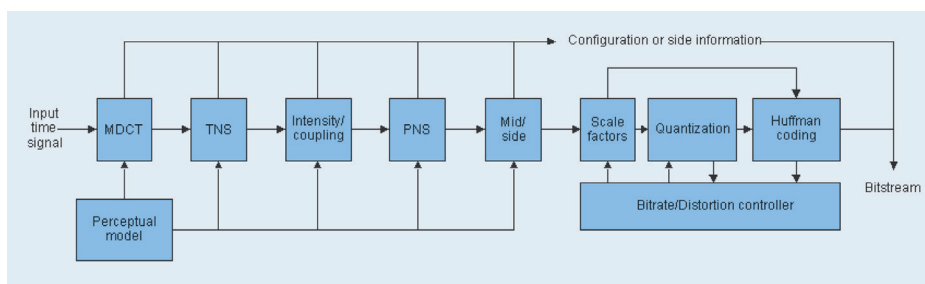


Figure 3: AAC core encoder.

A CLOSER LOOK AT AAC-ELD

AAC-ELD is the most flexible codec to suit the different needs of all possible Full-HD Voice applications. To achieve this level of flexibility, AAC-ELD can be used at three different operation modes – all of them completely compatible with standard compliant decoders:

1. AAC-ELD core: This mode can be used in all applications where high bit rates are available, for example 96 kbit/s and more for a stereo signal. A Low Delay MDCT filter bank replaces the MDCT filter bank used in AAC-LD. With this delay-optimized filter bank, AAC-ELD operates with a lower delay compared to AAC-LD.

2. AAC-ELD with SBR: This mode is the most flexible mode of AAC-ELD as it covers a very wide range of bit rates (approximately 32 to 64 kbit/s per channel) and sampling rates, and is therefore the preferred mode for video telephony applications such as Apple FaceTime. The delay stays constant over a wide range of bit rates enabling dynamically switching of bit rates without causing delay variances. In MPEG documents, this mode is typically called “down sampled mode”. It incorporates a delay-optimized version of Spectral Bandwidth Replication (LD-SBR) technology to the AAC-ELD core. LD-SBR allows the reduction of overall bit rate while maintaining excellent audio quality. The lower part of the audio spectrum is coded with AAC-ELD core, while the LD-SBR tool encodes the upper part of the spectrum. LD-SBR is a parametric approach that exploits the harmonic structure of natural audio signals. It uses the relationship of the lower and upper part of the spectrum for a guided recreation of the whole audio spectrum of the signal.

3. AAC-ELD with Dual Rate SBR: For applications demanding for even lower data rates, for example in live broadcast contribution, the “Dual Rate SBR” mode can be used. It is the most bit rate efficient mode enabling bit rates down to 24 kbit/s per channel at an increased delay compared to the other two modes. In this mode, again the LD-SBR tool is added to AAC-ELD: But the AAC-ELD core is coded with half the sampling frequency of the overall signal, instead of coding at the full sampling rate. This results in the best possible audio quality at very low bit rates. The structure of an AAC-ELD codec with Dual Rate SBR is shown in Figure 4.

Every AAC-ELD standard-compliant decoder can operate in any of the three modes, which allows the designer of the encoder side to freely choose the mode that best fits the application scenario.

The audio quality of AAC-ELD has been confirmed in several independent listening tests. In 2010, Deutsche Telekom Laboratories investigated the bit rate demands of state-of-the-art, super-wideband communication codecs [AES 129, Ulf Wüstenhagen et. al, “Evaluation of Super-Wideband Speech and Audio Codecs”.] The study showed AAC-ELD as the only codec that delivers an excellent overall quality at 32 kbit/s. Other codecs employed in the test procedure, e.g. G.722.1-C and CELT, required a minimum of 48 kbit/s to reach the same quality level, while speech codecs such as Speex and Skype’s SILK failed to deliver excellent quality at any of the selected bit rates.

AAC-ELD can be used in three operation modes:

AAC-ELD core codec for applications with 96 kbit/s and more for a stereo signal.

AAC-ELD with SBR as perfect choice for applications with 32 - 64 kbit/s per channel.

AAC-ELD with Dual Rate SBR enables bit rates down to 24 kbit/s per channel.

Independent listening tests confirm the superior audio quality of AAC-ELD.

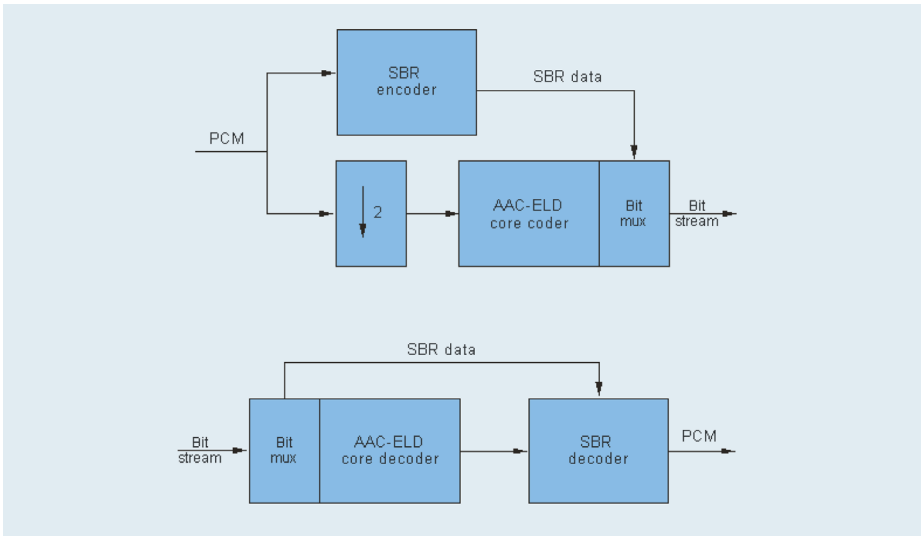


Figure 4: AAC-ELD codec with Dual Rate SBR.

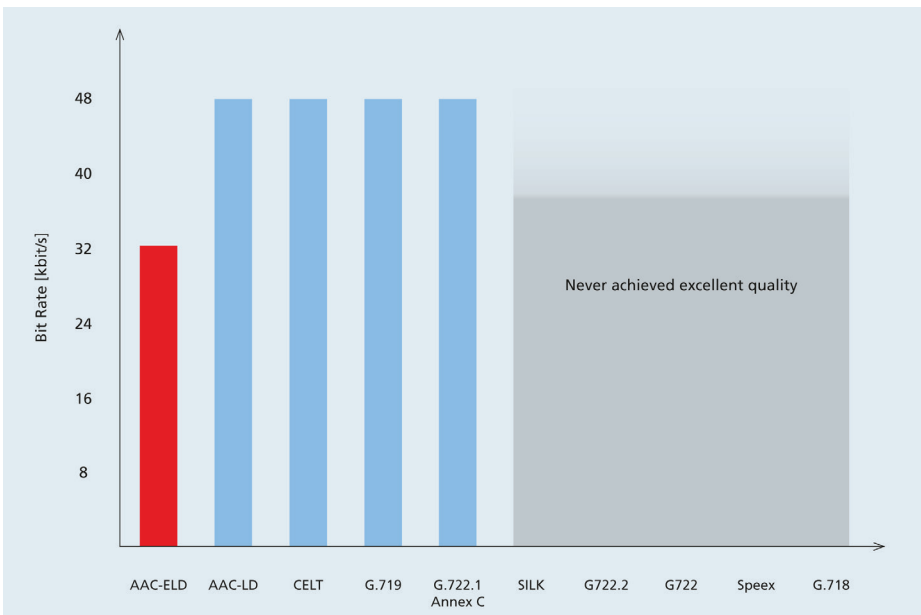


Figure 5: Minimum bit rate for excellent audio quality.

A CLOSER LOOK AT AAC-ELD v2

To achieve stereo performance at bit rates close to monophonic operation, a parametric stereo extension has been integrated into AAC-ELD v2. This parametric extension is based on a 2-channel version of Low Delay MPEG Surround (LD-MPS) that further reduces the bit rate. Instead of transmitting two channels, the LD-MPS encoder extracts spatial parameters to enable reconstruction of the stereo signal at the decoder side; the remaining mono down mix is AAC-ELD encoded. The LD-MPS data is transmitted together with the SBR data in the AAC-ELD bit stream. The AAC-ELD decoder reconstructs the mono signal and the LD-MPS decoder recreates the stereo image.

Typically, the bit rate overhead for the stereo parameters is around 3 kbit/s at 48 kHz. This allows AAC-ELD v2 to code stereo signals at bit rates significantly lower than with discrete stereo coding.

[AAC-ELD v2 for low bit rate stereo signals.](#)

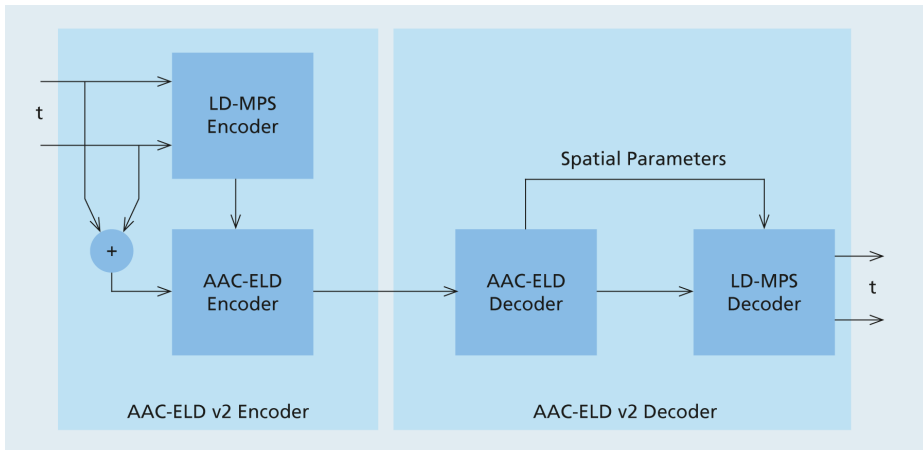


Figure 6: AAC-ELD v2.

TACKLING THE DELAY ISSUE

In a face-to-face conversation, delays in response can be interpreted in a variety of ways including hesitation, requiring time to think, or not wanting to give an answer. However, if the other party actually responds immediately, but a delay is introduced by technical shortcomings, misinterpretations can happen very quickly and the conversation can become awkward and frustrating. Therefore, it is very important to keep these delays, also called latencies, to a minimum of 150 to 200 ms end to end delay.

The end-to-end delay of a VoIP call is aggregated by several processing steps and components, such as echo cancellation, noise suppression, automatic gain control, routers, jitter buffer and speech/audio coding. As it is very important to maintain a low total latency, it becomes crucial that every component uses this resource responsibly. AAC-ELD is ideally suited in this regard as it contributes only 15 to 32 ms, depending on what bit and sampling rate are used.

Delay of AAC-LD

The only sources of AAC-LD algorithmic delay for an IP based transmission are the overlap-add of the MDCT filter bank, which generates a delay of 480 samples and the framing (audio input buffering), which adds another 480 samples. This corresponds to a minimum algorithmic delay of 20 ms at a sampling rate of 48 kHz.

Delay of AAC-ELD

In AAC-ELD core mode, the overlap-add delay of the filter bank is cut in half to 240 samples resulting in a very low delay of 15 ms. In the AAC-ELD with SBR mode, the SBR tool adds only a small delay of 64 (or 32) samples which leads to a very low delay of 15.7 ms. Finally, the dual rate mode achieves the best coding efficiency and ends up with a delay of still only 31.3 ms.

Delay-reduced mode for AAC-ELD

Typically, the algorithmic delay of AAC-ELD ranges between 15 to 32 ms at a sampling rate of 48 kHz and a block length of 480 samples. Applications that require a significantly

Aside of high quality audio, communication codecs need to feature a low coding delay to ensure a natural flow of communication.

The minimum coding delay of AAC-LD is 20 ms.

Depending on the operation mode, the coding delay of AAC-ELD ranges between 15 and 30 ms.

lower delay at broadcast quality, can increase the audio sampling rate to further reduce the algorithmic delay. However, this leads to higher workload and memory consumption.

The new delay-reduced mode of AAC-ELD which uses the downscaled filterbank mode allows to reduce the delay without changing the external sampling rate. Operated in the new delay-reduced mode, AAC-ELD has an algorithmic delay of only 7.5 ms at 48 kHz and an audio block length of 240 samples. The memory consumption and workload is reduced compared to AAC-ELD running at 96 kHz.

This mode avoids additional resampling. Instead the encoder is only virtually operating at a higher sample rate whereas the external sample rate stays at 48 kHz and the block length is reduced to 240 samples.

Bit streams created with the new mode of AAC-ELD can be decoded by all legacy AAC-ELD decoders supporting decoding of 96 kHz and 480 sample block length. Thus, the new mode can easily be integrated into existing systems.

The mode is fully conforming to ISO/MPEG standard 14996-3. It is MPEG standardized and described in detail in chapter 4.6.20.5 and DCOR 7 of ISO/IEC 14496-3.

Delay of AAC-ELD v2

With AAC-ELD v2, the Low Delay MPEG Surround tool is incorporated in a way that it only causes a small filter bank delay on the decoder end. If the core coder operates in the AAC-ELD core mode, the additional delay is 5.3 ms (sampling rate 48 kHz). In case the core codec operates in a mode with LD-SBR, the additional delay can be reduced to 4 ms. This results in a typical algorithmic delay of 35 ms.

The coding delay of AAC-ELD v2 is 35 ms.

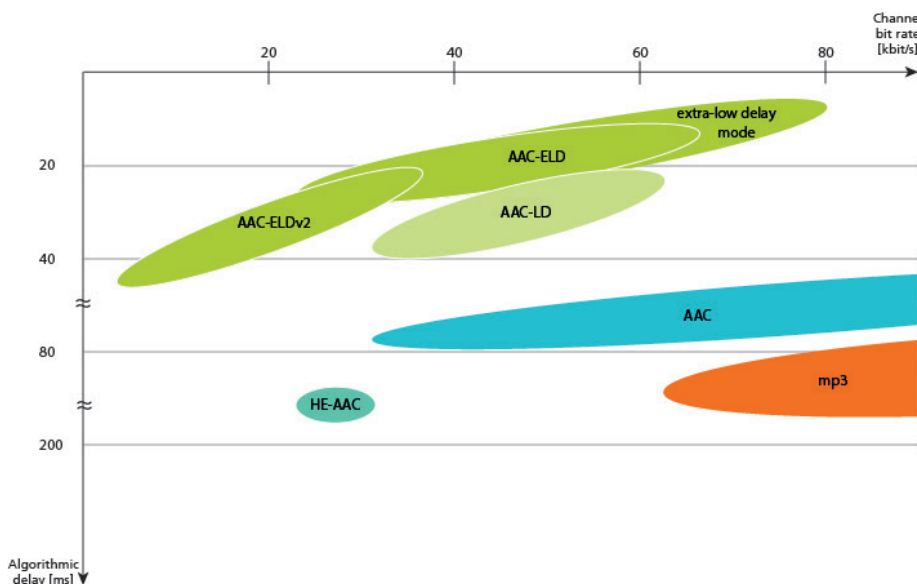


Figure 7: Delay of MPEG audio codecs.

IMPLEMENTATIONS

Software implementations of the AAC-ELD codec family are available from Fraunhofer IIS. For more information, please browse www.iis.fraunhofer.de/amm or www.full-hd-voice.com. AAC-ELD is also natively included in the popular Android and iOS operating systems. Application bulletins available on www.iis.fraunhofer.de/ammwhitepapers explain how to use the built-in AAC-ELD software in iOS and Android.

MORE INFORMATION

Read more about the AAC-ELD codec family and Full-HD Voice in the EDN Network article "Full-HD Voice: Understanding the AAC codecs behind a new era in communication" here: <http://www.edn.com/design/consumer/4405424/Full-HD-Voice--Understanding-the-AAC-codecs-behind-a-new-era-in-communication>.

ABOUT FRAUNHOFER IIS

The Audio and Media Technologies division of Fraunhofer IIS has been an authority in its field for more than 25 years, starting with the creation of mp3 and co-development of AAC formats. Today, there are more than 10 billion licensed products worldwide with Fraunhofer's media technologies, and over one billion new products added every year. Besides the global successes mp3 and AAC, the Fraunhofer technologies that improve consumers' audio experiences include Cingo® (spatial VR audio), Symphoria® (automotive 3D audio), xHE-AAC (adaptive streaming and digital radio), the 3GPP EVS VoLTE codec (crystal clear telephone calls), and the interactive and immersive MPEG-H TV Audio System.

With the test plan for the Digital Cinema Initiative and the recognized software suite easyDCP, Fraunhofer IIS significantly pushed the digitization of cinema. The most recent technological achievement for moving pictures is Realception®, a tool for light-field data processing.

Fraunhofer IIS, based in Erlangen, Germany, is one of 69 divisions of Fraunhofer-Gesellschaft, Europe's largest application-oriented research organization.

For more information, contact amm-info@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/amm.

INFORMATION IN THIS DOCUMENT IS PROVIDED 'AS IS' AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

INFORMATION IN THIS DOCUMENT IS OWNED AND COPYRIGHTED BY THE FRAUNHOFER-GESELLSCHAFT AND MAY BE CHANGED AND/OR UPDATED AT ANY TIME WITHOUT FURTHER NOTICE. PERMISSION IS HEREBY NOT GRANTED FOR RESALE OR COMMERCIAL USE OF THIS SERVICE, IN WHOLE OR IN PART, NOR BY ITSELF OR INCORPORATED IN ANOTHER PRODUCT.

Copyright © March 2017 Fraunhofer-Gesellschaft