Engineer Thomas Sandmann explains the high-resolution audio formats

What do 24 bit and 96 kHz achieve?

Those working with high-resolution audio in Europe will be unable to avoid the name Thomas Sandmann. The sound engineer whose productions cover the whole spectrum on a stylistic level, from the classical reference CD "Ave Maria" to the party hit "König von Mallorca" by Jürgen Drews, and whose studio is referred to in technical publications as one of the best state-of-the-art studios in Europe, is known as one of the first pioneers of digital audio technology. He is regularly invited as a speaker in the framework of international audio seminars and is currently mentioned in the same breath in the US as mastering icons such as Bob Katz. Thomas Sandmann is summarizing sense and nonsense of high-resolution formats for TerraTec.

The high-resolution audio format with a 24-bit word width and a sampling rate of 96 kHz has been a hot topic for manufacturers of audio technology. Everyone taking pride in oneself equips his hardware with digital interfaces in this format, uses internal 24-bit converters for analogue inputs or extends the software with the ability to process 24-bit files with high sampling rates. While a few professional studios are already preparing for forthcoming DVD that supports the 24-bit format, the majority of today's commercial productions continues to arrive on the good old CD. But, it stores only 16 bits and offers the well known sampling rate of only 44.1 kHz. In turn, the

question arises whether the purchase of equipment with high-resolution digital formats makes sense after all. To answer quickly: It does. And the following shows why this is so.

Quantization

In digitization, an analogue signal is sampled at certain intervals. At each of these intervals, the signal level is measured and represented as a numeric value. With the CD format, 16 bits are available for this numeric value, so that 2 to the power of 16 (i.e. 65,536) different, discrete numeric values can be represented. Since only integers are possible, an error occurs in those ranges in which the original analogue signal falls between the numeric values that can be represented. This error is different for every sampling value. The difference compared to the original value appears to be superimposed on the signal and can be recognized as quantization noise or as rough sound. With a 16-bit resolution, however, the quantization noise is theoretically 96 dB below the maximum control setting, that is, it is smaller than most of the other interference signals of the recording to be digitized. The necessity for increasing resolution and, therefore, the associated lowering of the quantization noise, appears to make no sense given these circumstances.

A/D conversion

At the first signal processing stage of a digital audio system, i.e. the A/D conversion, not even the slightest saturation may occur as compared to analogue recording. The produceable numeric range ends with 0dB, and every higher value is also represented as 0dB. Acoustically, this manifests itself as an extremely uncomfortable, hard clipping.

To ensure that no overmodulation occurs during a recording, A/D converters should always be modulated very carefully and reserves—the so-called headroom—be maintained. But this does not fully utilize the produceable numeric range—of the 16 bits, only 14 may be used. Consequently, a 24-bit converter presents itself for

digitization. While it also requires a headroom, its 24 bits will nevertheless leave more than 20 bits. The signal can subsequently be converted downward to 16 bits using different algorithms such as dithering and noise shaping, although they are then fully utilized up to the last bit. As we will see later on, this should be done as close to the end of the signal chain as possible—preferrably just before the master medium. This gives the recordings maximum dynamics and resolution.

Since 24 bits allow reproducing 2 to the power of 24 (i.e. 16.7 million) steps, the quantization noise of a 24-bit converter is theoretically at -144 dB. That is significantly less than the thermal noise of a single resistor connected to the input. It becomes obvious that not only the quality of the converter chip is important, but especially the quality of the analogue section of the A/D converter—and this is precisely where the wheat is separated from the chaff among the units offered on the market.

Processing

During the processing of a digital signal, the numeric values that make up the signal are used for calculation. Each calculation step creates decimal positions, which cannot be represented using the 16-bit format. Hence, the values must be changed to integers, which can be done through simple cutting, rounding or redithering. Similar to the sampling process, this creates an error which manifests itself through quantization noise or rough sound. Each calculation process creates a new error, and all created errors are added together. With intensive processing in mastering (equalizer, compressor, limiter, etc.) or with a digital audiomixer, not much of the original good signal quality would remain.

In the digital domain, the created decimal positions translate to an increase of the word width. If it is not immediately reduced back to 16 bits but instead coupled, no error is created. However, even a greater word width will be maximized at some time. But with a 24-bit word width, the error is so small that it is no longer audible, because the least significant bit of a 24-bit signal is located at minus 144 dB. For this

reason, not only the recording should be made using 24 bits as much as possible, but all intermediate steps of digital processing should be stored in this format. This also entails that the transfer to the computer for editing purposes must be carried out using 24 bits. Only after the complete processing session has concluded should the downward calculation to 16 bits required for the CD be carried out. Only in this way will the CD format actually be utilized, which would not have been possible with 16-bit technology at the beginning of the signal chain.

Compression

A special case in processing is the dynamic compression which requires that additional issues be observed. Typical for a modern production is processing with a low to medium compression ratio whose threshold is set relatively low. Lowering the louder signal portions enables a subsequent level increase to reach the maximum recording level again and, at the same time, lift the softer sections which are present and unchanged—the result is an increase in loudness, i.e. the perceived volume with the same level. Furthermore, dynamic peaks and short transients are intercepted with a limiter, and in doing so, the additional headroom can also be used to increase the level.

By using this process, data from the lower 8 bits of the 24-bit signal enter the usable dynamic range of the 16 bits of a CD. If, however, only a 16-bit recording were available as output material, the quantization noise would, instead, more distinctly move to the foreground. Since today's productions can no longer be imagined without compression, this in and by itself results in the necessity for a recording using 24 bits, if the CD format should really be utilized to its maximum capability. But the utilization of even the softest signal portions achieved in this manner also causes the cumulative character of the quantization error to take a stronger effect again, so that an additional increase of the word width within the dynamic processor becomes meaningful. For this reason, some units operate internally with a precision of, for example, 48 bits.

24-bit floating point

The 24-bit floating point format represents a specialty. Similar to a pocket calculator with floating point function, decimal points can be displayed here if fewer digits are found before the decimal point. This strongly complies with our sense of hearing, because quantization noise is covered by the signal at high levels, while it is much stronger during soft passages. Yet, soft passages mean small numbers and,

therefore, few digits in front of the decimal point, so that this allows for carrying many decimal positions. The result is an almost infinite dynamic during processing. Afterwards, the procedures described above are applied to convert to 16 bits.

At the end back to 16

For every process presented here, the final conversion to 16 bits plays an important role. As mentioned earlier, several methods are available for this purpose. The simplest option of transferring a 24-bit signal into a 16-bit signal consists of simply cutting off the surplus bits. This cut action generates an error in every sample word, which corresponds to difference between the actual numeric value represented in the 24-bit signal and the 16-bit number resulting from the truncation referred to as cutting. Naturally, the error shows the magnitude of the least significant bit (LSB) and is independent of the signal's amplitude. Hence, the relative error in reference to the signal amplitude increases with decreasing level. A constantly falling signal level, such as those occurring in a fade out or reverberant image, will then lead to an acoustic pattern that becomes rougher towards the end, and is simply "switched off" when falling below the level threshold of an LSB, i.e. breaks off suddenly.

Dithering

Our analogue hearing experience, however, knows the exact opposite. Distortions and relative harmonic content increase with increasing amplitude, not with falling amplitude. On the other hand, very soft signals come closer and closer to the background noise, and if the level continues to fall, they disappear in it. To achieve this behavior in digital systems, an artificial noise—the so-called dither noise—is added to the signal during the reduction of the word width. In this case, the background noise compulsorily increases, but the recording sounds much more natural. Dithering also causes the correlation between interference and effective signal to decrease or even eliminate completely, which is perceived by the human ear as significantly more pleasant.

Noise shaping

In order to avoid the increased noise now, clever engineers have come up with an interesting option. The frequency spectrum of the dithering noise is influenced by noise shaping by shifting the largest possible share of the noise energy into high frequency ranges in which the human ear is less sensitive. Thus, the complete noise energy remains the same, but the human ear subjectively perceives the noise to be softer. Noise shaping algorithms offer various settings from moderate shaping of the noise to the attempt of placing almost all of the noise energy before half of the sampling frequency by radically increasing the noise level.

Of course, such a drastic process offers not only advantages. Raising the noise level in the high frequency ranges can be the contributing factor in the creation of audible side effects during subsequent post-processing, for example, by using equalizers. It is, therefore, important that noise shaping algorithms are, indeed, used only as the last step in the production phase. Once again it becomes obvious that sophisticated technical processes only lead to the desired results if they are applied deliberately and with the necessary expertise.

Nyquist frequency and aliasing

We do know that an analogue signal is sampled at certain intervals during digitization. The higher the frequency of this sampling, the faster it is possible to register changes of the amplitude which translates to simply being able to represent higher frequencies. The highest representable audi frequency according to Shannon's sampling theorem is half of the sampling frequency. With 44.1 kHz, the audio signal could theoretically be as high as 22.05 kHz. Provided that higher frequencies as half of the sampling frequency (also referred to as Nyquist frequency) occur in the audio signal, unpleasant non-linear distortions occur which are referred to as aliasing (from Latin "alias," the other). Hence, it must be ensured that such frequencies are, indeed, no longer present in the spectrum prior to the A/D conversion. For this reason, every

A/D converter features an anti-aliasing filter which must, however, start the de-emphasis before the value of 22.05 kHz due to its finite edge steepness. If we assume that the audible audio range goes up to 20 kHz and should remain completely free of influences, the requirement for the ideal filter is such that the edge starts above 20 kHz and drops off sufficiently by 22.05 kHz that levels in the frequency range above are negligibly small. However, since such a filter can hardly be constructed in a satisfactory manner with traditional technology, a compromise is required: Either the filter must be designed with very steep edges so that the de-emphasis will only start above 20 kHz, or a more level edge is selected and the de-emphasis is started well below the 20 kHz. In the first case, the filter will generate a falsification due to high ripple in the audible range, in the second case the falsification is due to the filter edge. Only the newer developments of digital filters show results where demands for wide frequency response and low residual ripple are solved equally well.

96 kHz vs. Nyquist

Increasing the sampling frequency to 96 kHz increases the Nyquist frequency to 48 kHz. This allows, for example, to set the cut-off frequency of the filter to 24 kHz, which not only expands the audio frequency range, but also makes a full octave available for the filter edge between these 24 kHz and the Nyquist frequency of 48 kHz. Hence, the problem of the somewhat "narrow" design of current sampling frequencies is significantly defused. On the other hand, one has to admit that today's digital filters sound so outstanding that the advantage filter edges to be designed in a more level fashion is not the most important reason for changing the format.

And what about the expansion of the audio range itself? In the example of the level filter edge, frequencies up to 24 kHz already remained unaltered, and by designing the filter to be steeper, the linear reproduction of frequencies up to 40 kHz, for example, is no longer a problem. The audible range of the human ear only goes up to 20 kHz, and even this value is rather euphoric for most people who have outgrown

the infant stage—instead of being selected too low. There are, however, voices out there claiming that spectral shares beyond the limit of audibility contribute to the perception after all. One frequent argument is the so-called residual listening, i.e. the effect which, for example, renders a deep double bass sound audible based on its harmonics contained in the recording, even if the fundamental sound is missing. Yet, since the harmonics are located in the audible range in this case, both effects cannot be truly compared, so that there are just as many opinions out there proclaiming that an expansion of the audio range beyond 20 kHz would not be beneficial.

For these reasons, many listening tests have been conducted, which frequently led to a very interesting conclusion: Whether the test listeners were actually capable of differentiating recording with 44.1 kHz and 96 kHz, could not be established at the end since the tonal differences of the various converters were significantly larger, independent of the sampling rate used. These results show that even an expansion of the audio range cannot be the decisive reason for switching from the existing 44.1 kHz to 96 kHz.

96 kHz and Equalizer

Digital equalizers operate with algorithms that involve the adjacent sample values for calculating a single sample value. But in a 96 kHz signal, many more adjacent values are available in a time window of the same which makes the algorithm much more precise. In addition, the higher sampling rate makes it easier to also achieve the goal of analogue sound on the digital level. Again and again one can hear that analogue equalizers sound warm and musical, but digital ones cold and hard. And those who do not bother looking more closely at a few connections will forever believe in the fairy tale of evil digital and good analogue concepts. While it is indeed difficult to build a digital equalizer that sounds good, it is nevertheless possible as evidenced by some of the solutions currently offered on the market.

While an analogue equalizer goes far beyond the limit of audibility with the filter curves of its bands, a digital concept can only reproduce frequencies up to half of the sampling frequency due to the Nyquist sampling theorem. This is nothing serious because the human ear—as established earlier—does not hear anything in the higher frequency range. For this reason, the ideal digital equalizer features filter curves that are identical to the analogue model, but which end abruptly at half the sampling frequency.

In reality, digital equalizers are designed as IIR filters (Infinite Impulse Response). In this case, it is not the analogue spectrum between zero and half the sampling frequency that is projected onto this range of the digital level, but the complete spectrum up to infinitely high frequencies which is projected onto the finite range on the digital level. As a result, a compression of the bandwidths and shifting of the center frequencies results with the digital simulation. It is not surprising that such an equalizer sounds "hard," because if the Q factor at an analogue counterpart is increased, it would result in the same sound characteristic.

An algorithm that corrects the shifting of the center frequencies and the compression of the filter curves presents a remedy. While this works well in the bass and center range, compromises must be made in the high frequency range—not only to approximate the filter curve of the analogue model as closely as possible, but also to implement the principle-based drop of the curve to the value zero. These equalizer concepts typically exhibit a very analogue sound, but one which still distinguishes itself from the original for the bands with high center frequency specifications.

Only with twice the sampling frequency and, therefore, twice the audio bandwidth does the condition occur in which the analogue model can also be simulated beyond the audible range and, therefore, create an exact image within this range. Equalizer built this way sound like their analogue predecessors and even surpass them with

respect to signal quality since cascaded analogue stages always have to deal with noise. The double sampling frequency in so-called double sampling equalizers can also be generated internally, so that the input and output signals continue to be clocked with 44.1 or 48 kHz. To this extent, such equalizers may enter the digital signal chain even if it does not continuously operate with the high sampling rate.

96 kHz and word width reduction

As demonstrated above, signal processing with a word width of 24 bit and subsequent conversion to CD format with 16 bit presents an important foundation for today's production technologies. The word width reduction is often accompanied by a minor noise addition, the so-called dither noise. Here, the 96 kHz signals are one step ahead, too, since the noise output distributes itself to a band of twice the width of which only one half is audible. This, in turn, reduces the perceived noise by 3 dB.

Significantly more far-reaching are the advantages for noise shaping. This entails arranging as much of the noise output as possible in the band between limit of audibility and Nyquist frequency. With a sampling frequency of 44.1 kHz, this band is very narrow, and a high noise output can only be implemented using large level increases, quickly leading to the limits possible. But things look quite different in a 96 kHz environment because the wide range between 20 and 48 kHz is now available, which allows for displacing the majority of the noise to the non-audible range by using sophisticated filtering.

96 kHz and the Haas effect

In the directional perception of a stereo recording, we differentiate between intensity stereophony and phase-delay stereophony. The former is based on different levels of a signal in both channels and is created in the studio with the panorama controls of the audiomixer. With natural stereo recordings, particularly when using the classic dual-microphone technique, the directional perception results from the phase-delay differences. According to the law of the first wave front, also referred to

as Haas effect, we locate a signal by the direction from which the sound first reaches our ear—even if the sound level is identical at both ears.

Signals, which are only slightly shifted from the stereo center, create phase-delay differences of only a few microseconds. The distance of two samples of a 44.1-kHz signal, however, measures one 44,100th of a second, which is approx. 23 microseconds. One frequent opinion states that the higher sampling rate with its shorter time distance between two samples is better suited for replicating such phase-delay differences. However, this theory is without any foundation because it is indeed possible to present shorter time distances in a digital signal than the distance of two samples. The phase position of a digital audio signal is quite continual with respect to its value since the quantization and the resultant numeric values always apply only to the current amplitude in a discrete time pattern. The reconstruction in the D/A conversion process results in the original waveform and also in the original phase position of the signal. Here, simply raising the sampling frequency does not result in an advantage.

Conclusion

The increase of the word width of a digital signal to 24 bits offers tremendous advantages, even if the music signal is subsequently stored on a CD using 16 bits. With several digital processing steps and the use of dynamic compressors it is even essential for highest demands at the quality of a CD to work with the higher resolution, since the 16 bits of the CD format can only be utilized in this manner.

But one should also know where the limits are. If one exclusively records heavy metal with a total dynamic of 5 decibels over the complete duration of the production, the level will never reach the range of the LSBs. Likewise, it is nonsense to add dither noise with the magnitude of an LSB during the conversion of an audiomixer output featuring a noise level of -60 dB. Here, one should rather be concerned about the mute automation and a clean gating.

As far as aliasing problems are concerned, we are now well served with today's customary sampling frequencies. At most, the increase to 96 kHz brings about minor improvements and profits more from other advantages, such as improved equalizers and expanded possibilities for post-processing. Hence, those who attach great importance to highest audio quality for their productions, will appreciate the advantages of the 96-kHz format.

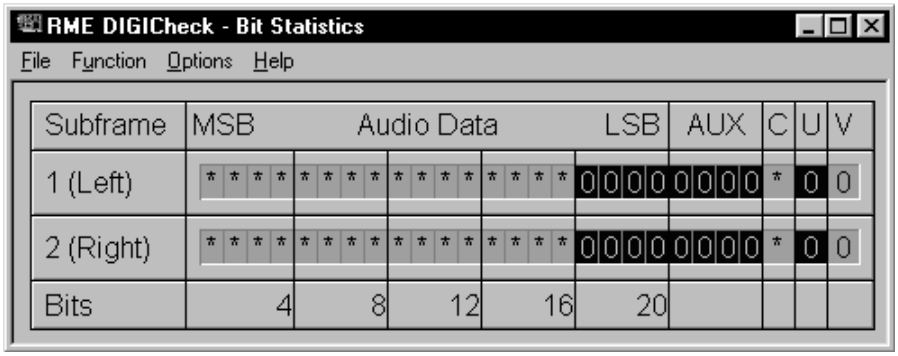
If the post-processing is done with 96 kHz and all steps, which profit from the high sampling rate, have been carried out, hardly anything speaks against the subsequent conversion to 44.1 kHz. Thus, it remains to be seen whether 96 kHz is meaningful as sound medium on the consumer market. In any event, increasing the resolution to 24 bits is much important than the higher sampling rate.

Dipl.-Ing. Thomas Sandmann

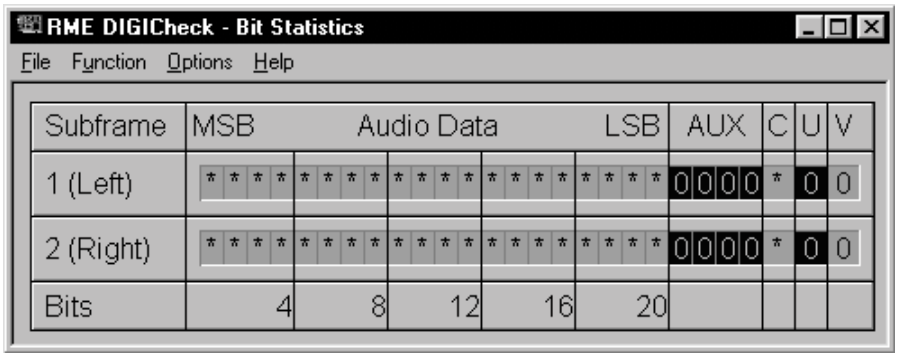
www.master-orange.de

Appendix

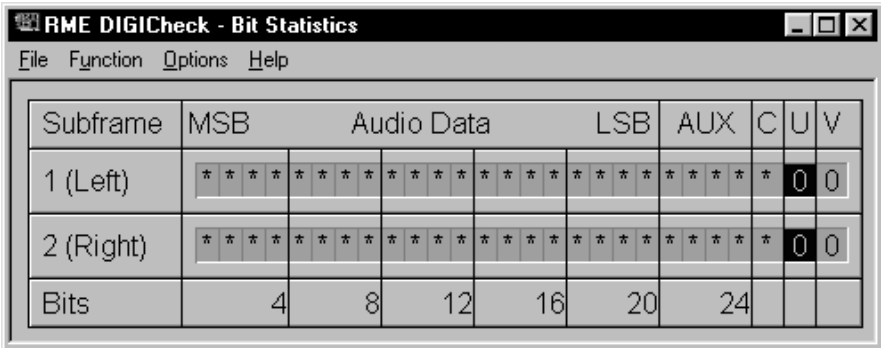
Graphics and diagrams



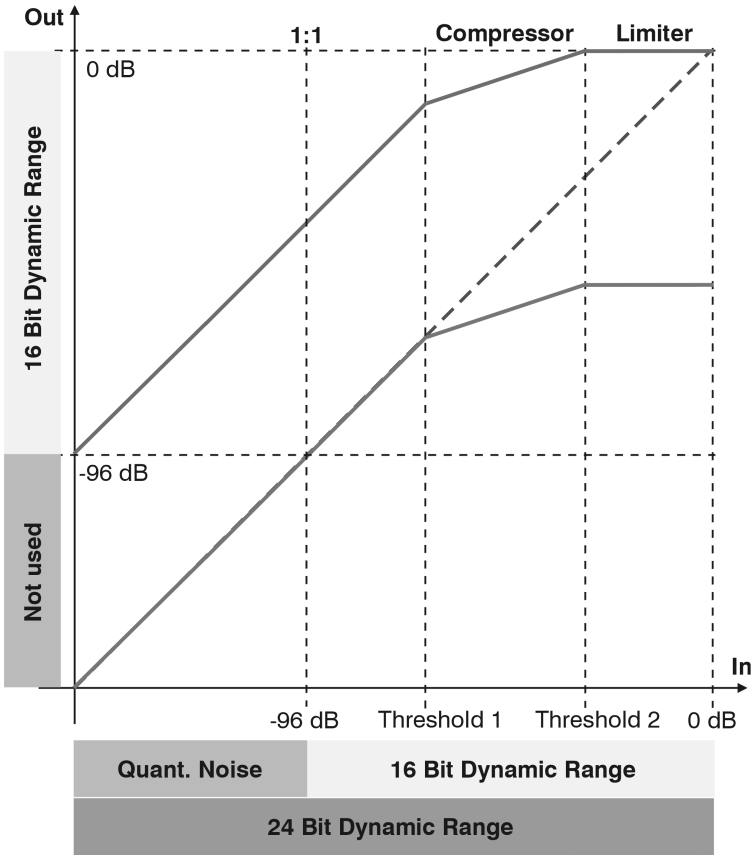
The data words of a 16-bit signal in AES/EBU format also contain four additional bits per stereo channel for the audio data as well as four bits for additional information, which are not utilized here.



With a 20-bit signal, the additional bits are used, which increases the distance between maximum recording level and quantization noise.

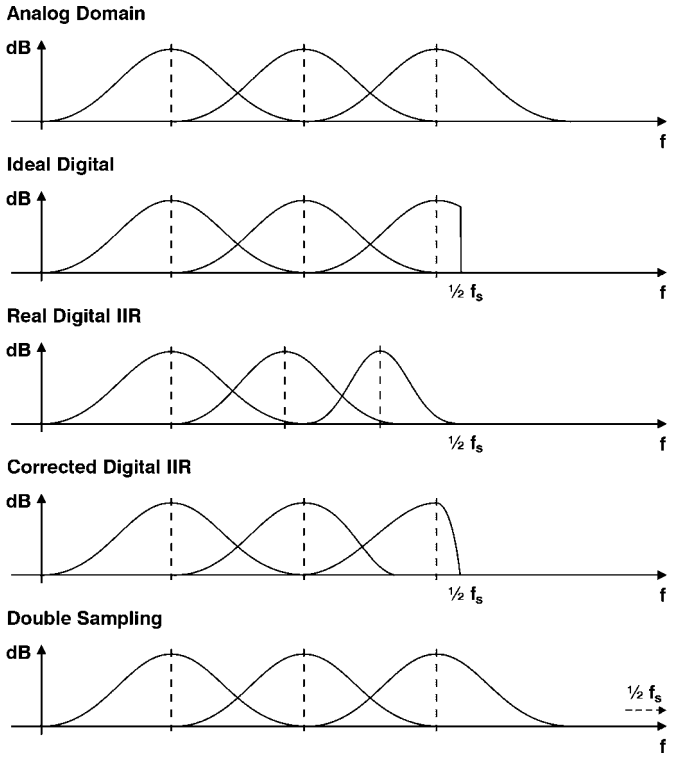


The 24-bit format uses the aux bits, which were originally intended for additional information, also for audio data so that this results in the largest possible dynamic range using the AES/EBU format. This range measures 144 dB.

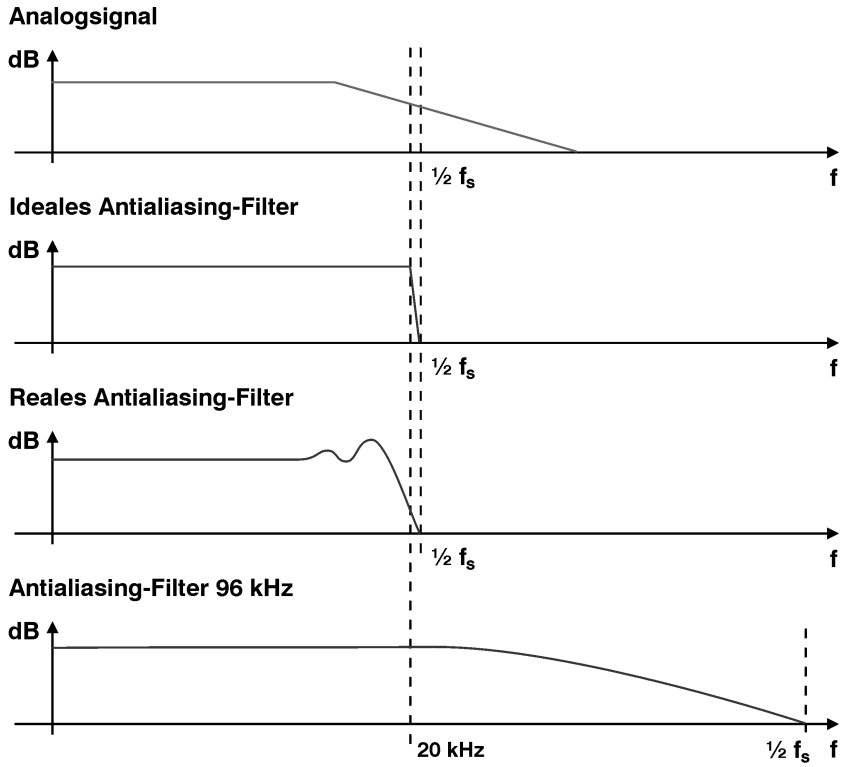


A look at the level relations during compression quickly reveals the necessity for 24-bit recordings. If a signal remains unprocessed (straight broken line), the 16 bits of a signal are projected in CD format (x-axis) or the upper 16 bits of a 24-bit signal are projected onto the 16 bits of the CD format (y-axis). The lower bits of the 24-bit signal do not offer any advantage, since they are projected onto a range below -96 dB, which is not utilized by the CD.

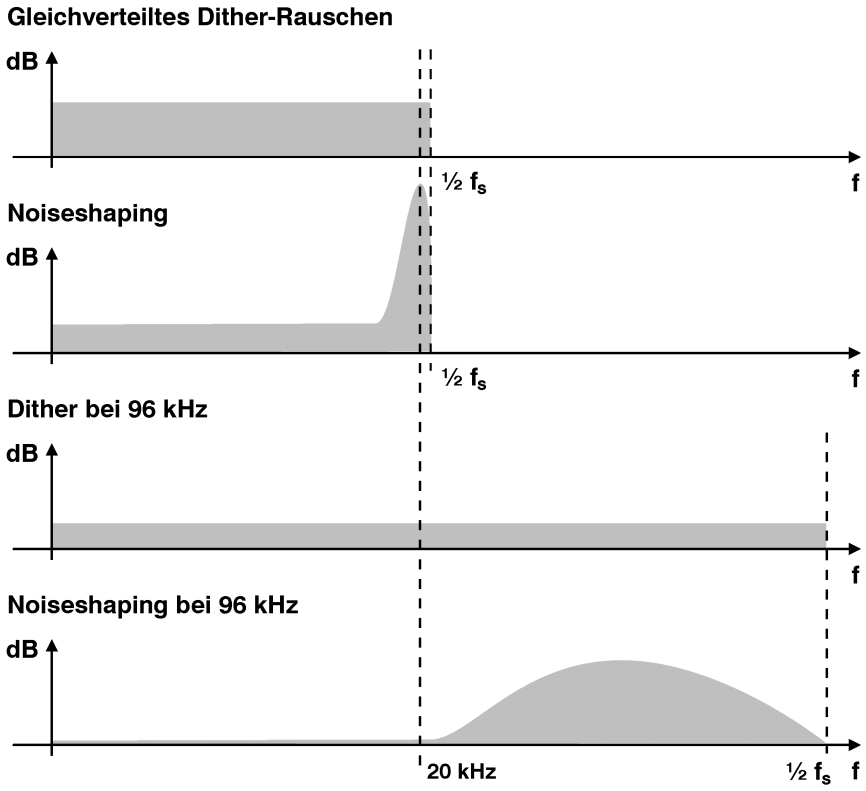
However, if compression and limiting are applied (bottom curve), the additional headroom can be used to raise the level to maximum recording level (upper curve). Now, not only the upper 16 bits, but also ranges of the lower 8 bits of the 24-bit signal reach the usable range of the CD.



The first figure shows the frequency responses of three peak bands of an analogue parametric equalizer. If this equalizer were digitally simulated, it would result in a frequency response which is shown as an ideal one in figure 2. In reality, however, an IIR filter projects the infinite spectrum in the range up to half the sampling frequency, which results in the behavior shown in figure 3. Correcting the center frequencies and band widths results in figure 4, where only the distortion of the bell curves in the upper frequency range interferes. This can be avoided by the double sampling rate since the limit of the half sampling frequency is shifted upward from the critical range.



While the frequency spectrum of a natural audio signal falls off towards the high frequencies, it generally contains components that lie above the half sampling sampling rate. To remove them, a steep edge filter is required, which does not affect the audio signal. Real filters, however, show a less steep filter edge as well as ripple in the transition region. With doubled sampling frequency, the problem is disappears since the edges are very flat and the filter can be constructed much simpler.



A characteristic factor of the noise added to the dithering is its power, can be regarded as the area below the frequency curve. Instead of a uniform distribution over the frequency range, the goal in noise shaping consists of placing the largest possible part of the area and therefore the power to the high frequency range, which causes the portion in the audible range to be reduced. In a 96-kHz environment, such a relation already occurs with evenly distributed noise, since more than half of the noise output is already located in the inaudible range between 20 and 48 kHz. If noise shaping processes are additionally applied here, it is relatively easy to displace the far largest area of the noise output into inaudible areas.