

If the volume of the playback system is adjusted so that peak playback levels exceed 96 dB SPL, these -96 dB errors will exceed the threshold of hearing. These errors are -96 dB relative to the peak level that can be represented by the 16-bit digital system. They are not necessarily -96 dB relative to the level of the music! Peak levels are often 18 dB higher than average levels. In a 16-bit system, these quantization errors may be only 78 dB (96-18=78) below the average level of a loud passage of music. What is worse is that these quantization errors may exceed the level of the music being played during low-level passages. Audio often fades at the end of an audio track. Near the end of these fades, quantization errors can easily exceed the level of the music! Quantization errors can be a serious problem in a 16-bit audio system.

Staircase, Hill, and Laser Pointer – A walk inside of an A/D converter

To understand quantization errors, let's imagine that I have a hill facing the staircase. For simplicity, let's just say that the steps are all 1 foot high. I paint a number on each step riser to identify each step. The hill facing the staircase is analog – it has no steps. I can walk up and down the hill and stop anywhere I like. If I carry a laser pointer, hold it level, and point it at the staircase, my movements up and down the hill will be quantized by the staircase riser number illuminated by my laser. I will have created a giant A/D converter. I decide to test my new digital system:

If I start at the bottom of the hill and climb 10 feet, my laser pointer says I have reached step riser 10. If I climb another 5 feet, it says I have reached step riser 15 – all is good. Now, let's suppose I climb another ½ foot. My laser is still pointing to step riser 15. According to the numbering on the risers, I have not moved. I am still on step riser 15! My A/D converter ignored my ½ foot movement. I can move back down ½ foot and back up ½ foot and my A/D will completely ignore my movements. Movements as large as almost 1 foot up and then back down are completely ignored. Obviously my A/D converter has a problem – it can ignore small movements. Now let's assume I start 15 feet up the analog hill and then move down ½ foot. My laser is now pointing to step riser 14. My A/D converter now says that I have moved 1 foot (from 15 to 14). In reality, I only moved ½ foot. I then discover that very small movements around the 15 foot elevation produce a change of 1 on my staircase. Again my A/D converter has a problem – it can amplify small movements. The first ½ step up from 15 was ignored (or muted) while the ½ step down from 15 was amplified.

Quantization errors can mute audio details, or amplify audio details. For this reason, quantization errors can add very high distortion to low-level signals. Reverberation tails, and low-level passages of music are most vulnerable to quantization distortion.

Frustrated with the poor performance of my hill-side A/D converter, I take a long coffee break.

Too Many Cups of Coffee – A solution to the quantization distortion problem

After far too many cups of coffee, I venture back to my hill-side A/D. I repeat my movements from the first test, but now I am getting different results! My initial climb of 10 feet should have landed my laser on step riser 10. Instead it is randomly hitting riser 9 and 10. Half of the time it hits 9 and half of the time it hits 10. Clearly the coffee has impaired my ability to hold the laser steady. I move $\frac{1}{2}$ foot up the hill and my laser now points to 10 much more often than 9. I experiment a little and discover that very small movements on the analog hill produce changes in the distribution of numbers quantized by the steps. If I average the results, I find that my digital system knows exactly where I am on the analog hillside. The random movements of my hand “dither” the position of the laser pointer. Dither is a random noise that is added to digital systems in order to eliminate quantization distortion. When dither is applied, the quantization noise still remains, but the distortion is gone.

Killing Quantization Distortion with Dither – Musical Details are Saved

When properly dithered, digital systems behave exactly like analog systems: The low-level resolution of a properly-dithered digital system is only limited by noise. No quantization distortion is present when a digital system is properly dithered. It is a common misconception that a 16-bit system is deaf to signals that are more than 96 dB below full scale. A 16-bit system that is properly dithered with white-noise dither can just sound like an analog system having a 93 dB signal to noise ratio. The white-noise dither adds some noise reducing the signal-to-noise ratio (SNR) to 93 dB, but this dither entirely eliminates the quantization distortion. It can be shown mathematically that a properly-dithered digital system has the same resolution as an analog system having the same signal to noise ratio. Our ears have an amazing ability to hear sounds that are as much as 30 dB lower in amplitude than the noise around us. If we are listening to a properly-dithered 16-bit system, it is possible to hear musical tones that are 30 dB lower than the noise (or $30+93=123$ dB below full scale). Low-level tones are digitized and reproduced without quantization distortion. Once a digital system is properly dithered, we can focus our efforts on improving signal-to-noise ratios. Dither does not remove quantization noise, but it can remove all of the quantization distortion. Dither does not mask (or cover up) the quantization distortion, it actually eliminates the distortion by converting it to random noise.

Killing Quantization Noise with More Bits

Clearly, if we throw enough bits at the quantization “evil” we will have victory. 16, 24, 32, do I hear 64? Where do we stop? Analog seems to have an infinite number of bits. The truth is that all analog electronic systems are quantized by electrons, but this is a topic for another time and place.

A 16-bit system has 2^{16} or 65,536 levels available for quantizing a signal. A 24-bit system has 2^{24} or 16,777,216 levels available. It is easy to see why a 24-bit signal should have advantages over a 16-bit system. The 24-bit “staircase” has 16 steps for every step on our 16-bit “staircase”. Obviously accuracy should improve. If we do the

math, we see that the error in a 24-bit system is $1/16,777,216 = 0.000006\%$. Expressed in dB the error signal is $20 \cdot \log(1/16,777,216) = -144$ dB (relative to the maximum output level). Errors that are 144 dB below peak level are well below the threshold of hearing, even at very loud listening levels. Adding bits can reduce quantization errors to insignificant levels. Every additional bit reduces the error level (distortion or noise) by 6 dB.

Imagine how bad things would be if we only had a 1-bit digital system! But wait ...

Killing Quantization Noise with High Sample Rates

Sony's DSD (SACD) system is a 1-bit system and only has 2 quantization levels available. It has a quantization noise level of $20 \cdot \log(1/2) = -6$ dB. A DSD system only has a 6 dB dynamic range when measured over its entire bandwidth! How does this even work? Or, why does it work so well?

Sony chose to throw high sample rates at the problem, and they showed that a high quality digital system could be built around a 1-bit format. Actually this was nothing new; 1-bit A/D and D/A converters were readily available when DSD was proposed. Sony simply suggested that we connect our 1-bit A/D converters directly to our 1-bit D/A converters to reduce the system processing.

A 1-bit Digital Experiment – try this at home

If you are fortunate enough to have at least one old-fashioned tungsten light bulb in your house, try this: Walk over to the light switch and try dimming light by rapidly turning the switch on and off. If you are really clever (and fast) you can actually adjust the brightness of the light by varying the time spent in the on position. If the switching is fast enough, the light will cease to flicker. The now-illegal mercury-wetted "silent switches" are ideal for these experiments. Any brightness between full off and full on can be achieved with only 1-bit (a single switch). I suspect many of us experimented with 1-bit digital systems as children but were scolded for playing with the lights. I even recall watching my father replace a light switch that failed after some extended 1-bit experiments.

Most light dimmers actually control brightness by turning the lights on and off 120 times per second (100 times per second if you have 50 Hz power). We can't see the flicker, but sometimes it is possible to hear the filament of the bulb vibrate when dimmed to low brightness. The 1-bit quantization noise of the light dimmer is at too high a frequency to see (120 Hz), but it can be heard. The vibrating filament confirms that the quantization noise is present, even though our eyes cannot see any flickering. The light dimmer hides the 1-bit quantization noise by operating at a 120 Hz switching frequency. Some of the flickering is removed by the thermal inertia of the filament, and some is removed by the slow temporal response of our eyes.

Sony's DSD Systems Hides Quantization Noise at Ultrasonic Frequencies

DSD audio systems hide 1-bit quantization noise by toggling at 2.8224 MHz. Most of the quantization noise in a 1-bit DSD system is above 20 kHz so we are not able to hear it. Some of the DSD quantization noise is removed by analog lowpass filters, and some is removed by the limited frequency response of our ears. DSD operates at a sample rate of 2.8224 MHz (64 x 44.1 kHz). This means that the quantization noise in DSD can be spread across a bandwidth that is 64 times as wide as a conventional 44.1 kHz PCM system. The quantization noise is “hidden” at frequencies that we cannot hear, and at frequencies that cannot be reproduced by our playback system.

High Sample Rates Provide More Space to Hide Quantization Noise

If quantization noise is evenly distributed across the bandwidth of a system, every doubling of bandwidth reduces the noise in the audio band by 6dB. DSD doubles the bandwidth of a CD 6 times to achieve a bandwidth that is 64 times as wide. At 6dB per doubling, DSD achieves a 36 dB reduction of in-band quantization noise. By itself, this added improvement would still only give DSD a SNR of 42 dB (6 dB + 36 dB). DSD systems must make heavy use of a technique known as “noise shaping” to move much more of the quantization noise to ultrasonic frequencies where it cannot be heard.

Noise Shaping – If you can’t get rid of the noise, just hide it!

Noise shaping is sort of like cleaning house. If you aren’t ready to throw the junk out, just move it out of the way - put it in the attic where it cannot be seen. DSD has a huge “attic” where quantization noise can be hidden. DSD’s “attic” begins at 20 kHz and extends up to 1411.2 kHz. To achieve acceptable signal to noise ratios in a 1-bit DSD system, aggressive noise shaping is used to move the quantization noise out of the audio band and into ultrasonic frequencies. The result is that a 1-bit DSD system can have a 120 dB or better SNR when measured over a 20 kHz bandwidth. On playback, the ultrasonic noise can be removed with an analog low-pass filter (at the output of the D/A converter). After filtering, the resulting signal will be free from any apparent quantization errors. DSD proves that a 1-bit system with aggressive noise shaping and a very high sample rate, can rival a 96 kHz 20-bit PCM system.

Noise Shaping Has Improved the Quality of 16-bit CD’s

Noise shaping is now almost always used to master 16-bit CD’s. This noise shaping is not as effective as the noise shaping in a DSD system. CD’s are restricted to a 44.1 kHz sample rate and therefore they have a very small “attic” in which to hide junk. Noise must be moved into the rather narrow region between 18 kHz and 22 kHz. CD’s have a 4 kHz band in which to hide noise. In contrast, DSD has over 1300 kHz available to hide noise, but DSD has a lot more quantization noise to hide. The bottom line is that a noise-shaped 16-bit CD system can rival the performance of a 44.1 kHz 20-bit system that lacks noise-shaping. Properly dithered and noise-shaped CD recordings have the ability to audibly reproduce tones that are in excess of 140 dB below full scale. Because of the noise shaping, these 16-bit recordings can sound like they have a 120 dB SNR. In most cases, the noise from microphones, analog electronics, and studios, greatly exceed the

perceived noise of a noise-shaped 16-bit system. Based on signal to noise consideration, a 16-bit 44.1 kHz system should be capable of delivering extremely high-quality audio. If you have any doubts about the effectiveness of dither and noise shaping, remember that DSD only has one bit.

A Better Solution – Increase Both the Sample Rate and the Bit Depth

Combining 96 kHz with 24 bits yields an in-band SNR of over 150 dB! The quantization noise in a 96 kHz 24-bit system is well below audible levels at even the loudest playback levels. 96/24 systems do not need noise shaping, nor do they need analog lowpass filters to remove ultrasonic noise. The 24-bit word length makes digital processing simple and transparent. It is therefore an ideal format for recording, editing, and mixing. 16-bit systems degrade quickly when processed. 1-bit DSD systems are extremely difficult to process. The quality of a DSD recording can degrade very quickly when mixing and editing. Benchmark does not recommend CD or DSD systems for professional recording, editing, and mixing applications. In our opinion, these formats are only suitable for distribution of the final product.

1-bit Works, 16-bits Work, Why 24?

OK, if dither and noise shaping work so well, why do we have 24-bit audio systems? Are the extra bits just marketing hype? Are we buying something that we do not need? The answer is a definite no!

Until a few years ago, most digital audio was recorded, edited, mixed, and mastered in 16-bits. Unfortunately, digital audio degrades very quickly when subjected to 16-bit mathematical operations. The problem with this is that every mathematical process applied to the 16-bit audio creates a result that has more than 16-bits. To understand how this works, let's consider an example: Suppose I put a dollar in a savings account. If the dollar earns 1% interest I now have \$1.01. If I then earn 1% on my \$1.01, I now have \$1.0201 until my bank rounds this down to \$1.02. Money is lost in my bank account due to rounding, and in the same way, audio detail can be lost in an audio DSP system. DSP operations extend word lengths. If these extended word lengths are truncated or rounded back to their original length, we introduced another quantization process. Like the quantization process in an A/D converter, this quantization process can be dithered, and noise shaped to reduce its audibility. Nevertheless, noise will rise with every re-quantization. If dither is omitted, distortion will rise quickly with every operation. Many early professional DSP systems operated at 16-bits and lacked dither. These systems produced some of the worst sounding digital artifacts and are largely responsible for digital audio's bad reputation. 24-bit systems were rare until a few years ago, and consequently, a very large percentage of 16-bit CDs were made with 16-bit digital mixers and 16-bit digital effects that lacked dither. These recordings clearly demonstrate the need for higher resolutions in the studio. 24-bit audio is very robust when passed through digital processors. In contrast, 16-bit audio is very fragile.

Every added bit reduces the damage done per DSP operation by 6 dB. It takes 256 DSP operations at 24-bits to equal the damage done by one 16-bit DSP operation. Many high-quality professional mixing, editing, and effects systems use 32-bit DSP processing to insure that any errors are well below audibility.

Evil #3 - Aliasing

Aliasing is an effect that frequency-shifts signals so that they are incorrectly represented. In digital audio systems, ultrasonic tones may be aliased such that they are reproduced at audible frequencies. This frequency shifting may also reverse the relative locations of tones within a multi-tone signal such that a higher tone is reproduced below a lower tone. Alias tones produced from inaudible ultrasonic tones can clutter the audible band with tones that have no relationship to the music. The destructive effects of aliasing raise havoc on a musical signal. What causes aliasing? Is there a cure?

Wagon Wheels and Old Movies - A Foreshadowing of Future Problems

The wagon wheels in old western movies often seemed to turn backward. As a wagon began to move, the wheels would appear to begin turning in the forward direction, but then they would appear to slow, then stop, and then reverse as the wagon gained speed. In some cases it was possible to see the wheels reverse several times as the wagon accelerated. Was something wrong with the old wagons, or was something wrong with our movie equipment? The answer of course is that wagons are wagons, but movies are flicks. Movie cameras capture 24 still images (or frames) per second, and projectors flash 24 still images before our eyes each second. This flickering of still images gives movies their nickname (flicks).

When the images flash fast enough, we correctly perceive the motion in the images. If the spokes of the wagon wheel move less than $\frac{1}{2}$ spoke position from one image to the next, we can accurately interpret the speed and direction of the wheel. If the wheel turns exactly one $\frac{1}{2}$ spoke position the direction of motion of the wheel cannot be determined. It could be rotating forward or backward by $\frac{1}{2}$ spoke position per frame. If the wheel moves more than $\frac{1}{2}$ spoke position but less than 1 spoke position between successive images, the wheel would appear to rotate backward. As the rotational speed of the wheel increases, the perceived direction of rotation will keep changing until each image becomes sufficiently blurred such that the spokes are no longer visible.

Like Movies, Digital Audio Systems can have Aliasing Problems

Movies are sampled systems. Images are samples 24 times per second in an effort to capture motion. Similarly, digital CD systems sample music at a much faster 44,100 times per second in an effort to capture the waveform of the music. If a wagon wheel moves less than one-half spoke between frames, its motion is preserved. If an audio

signal changes less than one-half cycle between samples, its “motion” is preserved. For this reason, a digital system that samples at 44.1 kHz can only accurately reproduce audio signals having a frequency less than 22.05 kHz (one half of the sampling rate – also known as the Nyquist frequency). Tones between 22.05 kHz and 44.1 kHz will “alias” down to lower frequencies (much like the wagon wheel appeared to turn backward). Tones at exactly 44.1 kHz will disappear (much like the wagon wheel appeared to stop at a rotation rate of one spoke per frame). A 44,000 Hz input tone will alias to 100 Hz ($44,100 - 44,000$). A 43,900 Hz tone will alias to 200 Hz ($44,100 - 43,900$). A 22,150 Hz tone will alias to 19,950 ($44,100 - 22,150$).

Attempting to Kill Aliasing with High Sample Rates

If we were to speed up the frame rate of movies, they could capture a wider range of rotational speeds without aliasing. Similarly, 96 kHz, 192 kHz or even 2.8224 MHz audio systems can accurately represent a wider range of frequencies than the 44.1 kHz CD system. Nevertheless, aliasing will still occur whenever that audio input frequency exceeds one half of the sampling frequency. A high sample rate cannot guarantee that aliasing will not occur.

Killing Aliasing with Low-Pass Filters

Aliasing will not occur if we remove high-frequency signals before they are sampled. In the example of the movie, aliasing stops when the motion is so fast that the spokes are blurred to the extent that they are no longer visible. The open shutter of a camera creates a low-pass filter that blurs motion. If the blurring is sufficient, aliasing is prevented.

In a CD system, we must filter out all signals above 22.050 kHz while attempting to leave audible frequencies untouched. We would like to pass 20 kHz without loss while removing 22 kHz. This is a very difficult task, and it requires a very abrupt low-pass filter. In the early days of digital audio, these “brick-wall” filters were analog filters. Unfortunately, it is very difficult to build analog filters that have the necessary performance. Consequently, many early recording and playback systems suffered from some audible aliasing problems, and/or some loss of frequency response.

It is much easier to construct a brick-wall digital anti-alias filter. Converters can be configured to “oversample” at some multiple of the desired sample rate. A brick-wall digital filter can be applied when the audio is down-sampled to the desired output sample rate. A very simple analog low-pass filter must still be used to remove very high-frequency signals from the input of the oversampled converter. Virtually all digital audio systems now use over sampled A/D and D/A converter. Aliasing is rarely a problem in these newer “oversampled” audio converters.

Low-pass filters are applied in the A/D to prevent aliasing. Low-pass filters are applied in D/A converters to remove sampling artifacts and produce a continuous waveform

that has no evidence of passing through a sampled system. End-to-end, a digital audio system can be indistinguishable from a band-limited analog system.

In Summary

Quantization distortion can be completely eliminated with dither. The number of bits, the sample rate, and the type of dither employed will collectively determine the SNR of the digital transmission system.

Jitter-induced distortion can be reduced to levels that are well below audibility. Many newer recordings are essentially jitter-free. If reproduced through a low-jitter D/A converter, jitter-induced distortion will not even approach audibility.

Aliasing has been virtually eliminated through the use of oversampled converters. No audible aliasing artifacts should exist in a well-designed 44.1 kHz system.

The audio industry has produced thousands of CD titles that include all 3 of the “digital evils” at audible levels. These older recordings do not accurately represent the capabilities of the CD system. Clearly the CD format has not delivered “perfect audio forever”, but it can now deliver “nearly perfect audio”. We haven’t followed the wrong path for 18 years; it has just taken us that long to perfect the system. Few commercial recordings reach the performance that is achievable with the CD format. For this reason, 44.1 kHz 16-bit systems still remain a viable distribution format for high-quality recordings. While the CD format is well-suited for distribution of the finished product, Benchmark does not recommend this format for recording or production use as the quality degrades rapidly when processing is applied.

In theory, the DSD system can offer a slightly better signal to noise ratio than a noise-shaped CD system. It can also provide a slightly wider usable bandwidth. Unfortunately, neither of these capabilities may be realized in a recording that was completely recorded and produced in DSD. DSD relies on very aggressive use of noise shaping, and consequently, it is even less robust than the CD format in a production environment. DSD 1-bit signals are not easy to process and quality can degrade quickly in mixing and mastering operations. For these reasons, Benchmark does not recommend DSD for production use. We believe that DSD recordings usually fail to deliver the benefits claimed for the format.

High-resolution 96 kHz 24-bit systems are essentially artifact-free when properly designed. All distortion and noise artifacts can be held well below audibility in these systems. These high-resolution digital systems can capture, store, process, and reproduce analog signals without a hint of quantization, jitter, or aliasing. These systems have sufficient resolution to tolerate many stages of digital processing and are ideally suited for recording, production, and distribution.