BEYOND THE PALE
HIGH-RESOLUTION AUDIO

With new high-resolution standards and formats available, RICHARD ELEN revisits the hows and whys, the benefits and the cost of these technologies.

The Shannon-Nyquist theorem, put simply, indicates that the bandwidth of the information carried by a communications channel cannot exceed half the carrier frequency. In digital audio terms, this means that the highest frequency you can convert between analog and digital is half the sample rate. If you go any higher, you get information, but it is meaningless.

The call for higher sample rates goes right back to the early days of digital audio, and certainly to the beginning of Compact Disc. The rationale, as far as sample rates were concerned, was simple: we can't hear much above 18kHz, so there is no point in wasting our time trying to record anything much beyond there. A sample rate of 44.1kHz or 48kHz means a theoretical 'Nyquist limit' (the frequency above which you want to be sure not to record any signals) of 22.05kHz or 24kHz — on the face of it, satisfactorily outside the audible range. But, by the time that CD hit the streets, there were already calls for something higher.

There are several possible reasons why the call for higher sample rates began so early on. The most likely are the way in which digital conversion was performed in the early days, and the kind of anti-aliasing and anti-imaging filters that were required to avoid recording any apparent audio signals beyond the Nyquist limit, and stopping excessive, meaningless HF energy reaching the replay amps and speakers.

To get maximum audio bandwidth, you need to pass audio signals as near to the Nyquist limit as possible, but you need significant attenuation by the time you get there. So the filters have to be very steep indeed —

hence the name 'brick wall' filters. Early digital systems used analog filters for this purpose, which presented significant problems. They induce distortion as a result of ringing, and suffer major phase distortion which might result in the signal at 10kHz being hundreds of degrees out. They cause smearing of samples across time, resulting in damage to stereo imaging and other effects. It is no wonder that the analog camp felt that early digital recordings sounded dreadful, with harsh, clangy high frequencies and poor stereo imaging — they often did! Even if the filters were implemented digitally, there were still problems.

The obvious solution was to increase the sample rate, so the filters could be smoother and less like a brick wall. Unfortunately, this was not really technically feasible outside the laboratory at the time, and the 44.1kHz and 48kHz sample rates established early on remained on the books, as they do today.

However, another solution was developed over the subsequent years; this initially took the form of oversampling. In oversampling, the nominal sample rate is raised to a multiple of its actual value. To paraphrase *White's Audio Dictionary*, you can convert a 44.1kHz bitstream from a CD at, say, four times the sample rate, clocking at 176.4kHz and creating three artificial samples in between each pair of real ones. The artificial samples are at zero level and do not alter the information carried in the 'real' samples. Digital filtering is then used to interpolate the zero samples into intermediate values between the real 44.1kHz ones. But now the Nyquist limit is up at
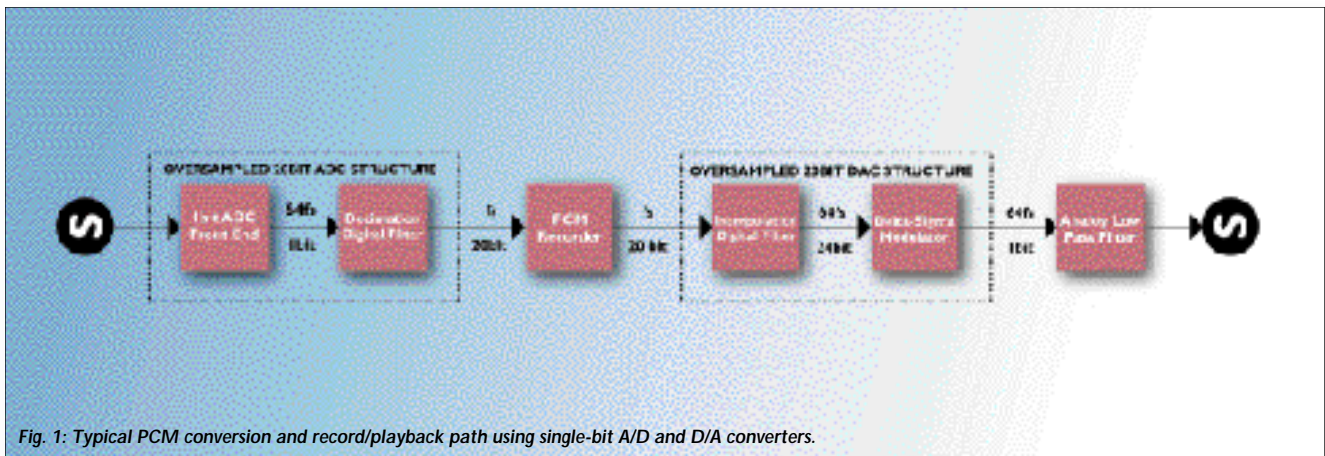
Fig. 1: Typical PCM conversion and record/playback path using single-bit A/D and D/A converters.

88.2kHz, so the filters can be much smoother and more gentle than before.

Per Ardua, Ad Astra

Even more effective than oversampling is to capture inherently more data by using a higher sample rate. Then you don't have to create artificial samples between the real ones: they're all real. The logical choice is to use twice the existing sample rates, in other words 88.2kHz and 96kHz. Logical, because having a simple integral relationship between the existing and the new rates means that sample-rate conversion is inherently easier and offers higher quality — useful if we assume that the new and older technologies will co-exist for a while, which they probably will. And, because you are really sampling at the higher rate, you have a real audio bandwidth out beyond 40kHz.

Some people, largely pushed by the audiophile fraternity one imagines, called for rates to go even higher — to 176.4kHz and 192kHz. This means a potential 'audio' bandwidth of beyond 80kHz, which, on the face of it, is an absurdity — what on Earth audio is up there?

In fact, simple non-rigorous tests seem to indicate that most people can't hear a difference in sample rate above about 64kHz or so. But 88.2/96kHz still makes sense as 'the next step' because of that integral relationship with the existing rates. And there's another benefit, too. Lossless signal 'packing' techniques, if carefully designed — like DVD-Audio's Meridian Lossless Packing (MLP) — actually operate more efficiently at high sample rates and make up for the additional data storage otherwise required. The actual data rate at 96kHz, with lossless packing, can increase by as little as 1.3 times compared with uncompressed audio sampled at 48kHz, according to Professor M. O. J. Hawksford of the University of Essex, in England.

For a long time, I was skeptical of the value of higher sample rates, because of my suspicion that there wasn't anything up there to record. What changed my mind was, basically, listening to the same converter performing at regular and high sample rates: the sound is undeniably better at the higher rate, if the converter is well-built. Listening to a single converter (or A/D-D/A converter pair, for comparison with analog) capable of multiple sample rates insures that the converter itself is not part of the problem or the solution. It is quite possible that a high-density converter will sound better than a 44.1/48 unit anyway, simply because the design required for high-density conversion is that much more exacting. But listening to a multi-rate conversion system avoids that problem — you're using the same devices all the time.

According to a paper by dCS Ltd, presented at the 20th Tonmeistertagung in 1998, the most commonly-noticed benefits of recording at 96kHz sampling over 44.1kHz include: less 'busy signal' breakup — very good quality; better separation of reverb and room acoustics from instrument output; better balanced bass; better percussion (particularly cymbals); and some stereo image formation.

Most of the raw observations presented in the paper are not very tightly defined (the exact meaning of 'some stereo image formation', for example), but the subjective improvements are clear from the above.

When we get to 192kHz, the picture gets a little cloudy, surprisingly enough. According to dCS: no 'busy signal' break up — excellent quality; very good separation of reverb/room acoustics from instrument output *but* bass can appear light and slightly out of time; *and* stereo image can be strong but widened.

In the opinion of the authors of the dCS paper, the widening of the stereo image is related to the perceived problems in the bass end, and filter impulse or transient ä
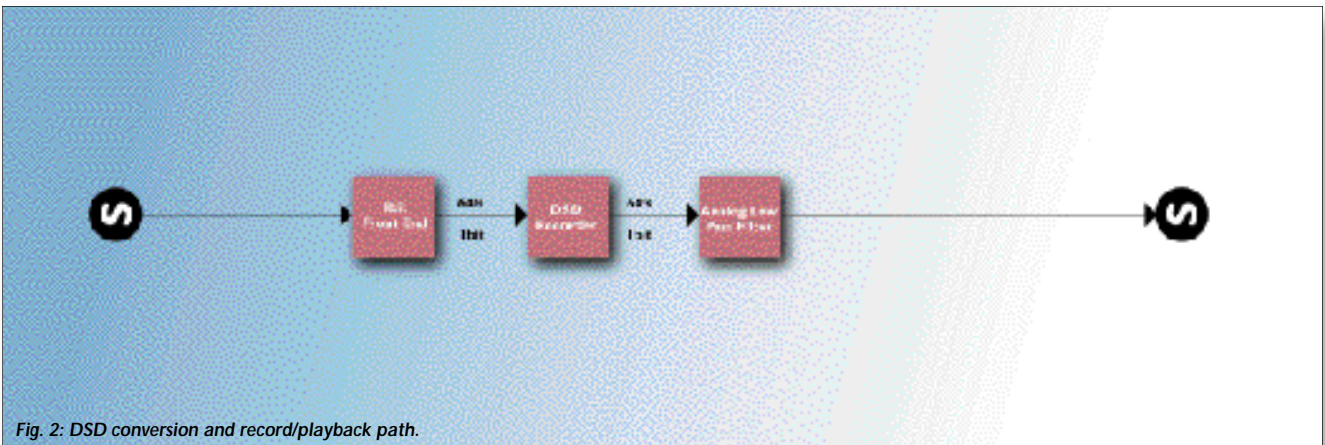

Fig. 2: DSD conversion and record/playback path.

BEYOND THE PALE

ä response may be significant in correct image formation, along with proper bass perception.

The business of alteration of apparent stereo width at 192kHz sampling raises an interesting point. On the face of it, you would right kind of filters), there are other suggestions making the rounds too.

James A. Moorer, PhD, Senior Vice President, Advanced Development and Co-Founder at Sonic Solutions, proposes in his paper people can hear a time delay of 15 microseconds or more. Under some circumstances, some people can hear time delays of 3 to 5 microseconds. Note that one sample at 48kHz is 20.833 microseconds. At
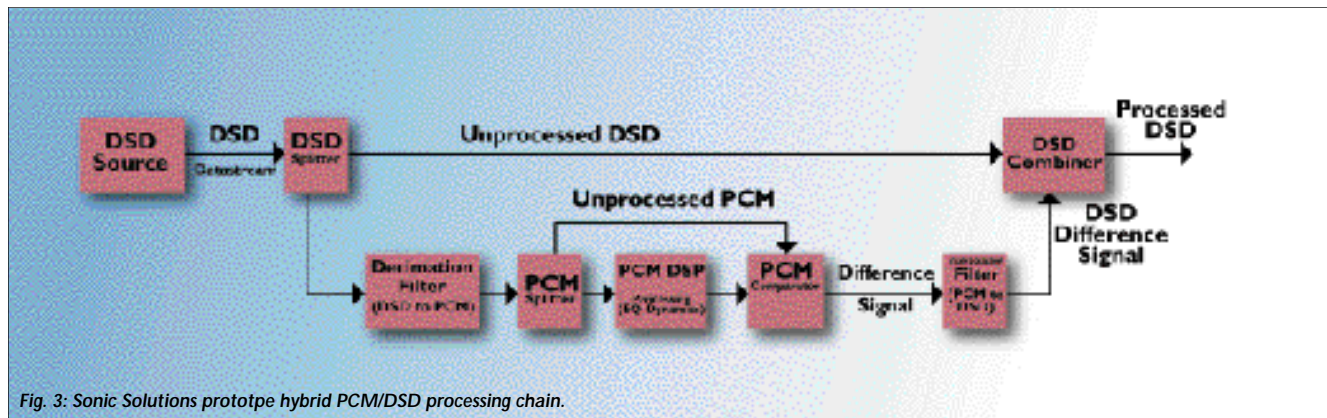

Fig. 3: Sonic Solutions prototpe hybrid PCM/DSD processing chain.

think that, if there was little going on around the upper bandwidth limit of a 96kHz conversion system (ie. at least 40Hz), you would definitely be hitting the law of diminishing returns if you had a recordable bandwidth twice that size. However, not only do the dCS subjective findings suggest that there is actually a positive difference (as well as negative differences if you don't use the

New Audio Formats — A Time of Change and a Time of Opportunity that there is one aspect of hearing where very small time intervals are readily perceived by humans, and that is binaural (two-eared) hearing, which is how we perceive the localization of sounds, such as within a stereo or surround environment. He says: "If you put a pulse into one ear, then a pulse slightly delayed into the other ear, most

96kHz, it is 10.4167 microseconds. The minimum inter-aural (across the two ears) time delay that most people can hear is *less than one sample period at 48 kHz.*"

As a result, Moorer concludes: "When listening with both ears, everyone can distinguish 96kHz recordings from 48kHz recordings, and everyone prefers the 96kHz recordings... the reason being probably because some kind of time-domain resolution between the left and right ear signals is more accurately preserved at 96kHz."

And of course, if we are venturing into surround sound, the need for more accurate localization, at least on the face of it, is even more acute (that is, if we assume that we want to do more with it than create a warm fuzzy feeling round the back or reproduce dinosaur footfalls, which I, for one, hope we do). And, as *Audio Media* readers will probably already be aware, the DVD-Audio specification allows for sample rates of up to 96kHz for six audio channels, or 192kHz for two.

If Moorer is right, then higher sample rates are better. If dCS is right, higher sample rates are better too, as long as you take care with filtering at the very highest rates. But there may be other answers, and before we consider them, we need to look at the other side of the digital conversion story, at least as far as PCM (pulse code modulation) is concerned. That other side is the word length.

How Long Is Enough?

A PCM A/D converter relies on taking measurements of the voltage of the analog waveform once each sample period (eg. 44,100 samples per second or, if you prefer, '44.1kS/s' for CD) and storing that voltage as a digital word with a specific number of binary digits (bits). The number of bits determines the dynamic range of the digital system — the distance, in dB, between the loudest and quietest sounds you can convey. A Compact Disc offers 16 bits, while modern converters

claim as high as 24-bit word lengths. (There are other methods of converting signals into digital form that do not use words at all, such as Sony and Philips' Direct Stream Digital — DSD — system, which we will consider later.)

Robert Stuart, of Meridian Audio Ltd, in his paper *Coding High Quality Digital Audio* looks at two of the most common misconceptions about PCM. The first he tackles is the misconception that PCM cannot resolve detail smaller than the LSB (least significant bit).

"What is suggested is that, because, for example, a 16-bit system defines 64kbits, that the smallest signal that can be 'seen' is 1/64k, or about -96dB. Signals dropping off because they are smaller than the smallest step, or LSB, is a process we call truncation."

However, the fact, Stuart points out, is that one of the greatest discoveries of PCM was that adding a small amount of random noise, called dither, causes the truncation effect to disappear. Even more important was finding out what was the best kind of noise to add. A great deal of the pioneering work here was performed by Peter Craven and the late Michael Gerzon.

One of the neat things about analog audio is that the noise 'floor' isn't a floor at all: it's a furry fuzz. We can hear coherent signals below the noise. With truncated digital signals this doesn't happen (the signal is cut off and distorted), and very simple forms of dither do not allow much perception of signal below the noise either. Done correctly, however, adding the right kind of noise, and at the right times (you need to dither not only whenever audio is digitized, but also whenever it is re-digitized, such as in a filter or DSP process), the resolution of the system becomes infinite, according to Stuart. He goes on: "What results from a sensible digitization or digital operation is not signal plus a highly-correlated truncation distortion, but the signal and a benign low level hiss. In practical terms, the resolution is limited by our ability to resolve sounds in noise. Just to reinforce this, we have no problem measuring [and hearing] signals of -110dB in a well-designed 16-bit channel."

What this means is that, if you dither your signal correctly and whenever you need to, the number of bits simply defines the noise floor, not how much detail you can hear — a claim made for word-length reduction systems such as Apogee Electronics' UV22 and certain other techniques.

So, where you would like your noise floor, rather than the amount of detail you want to convey, defines how many bits you need. And so does something else: the Laws of Thermodynamics.

Every electronic device produces noise. The noise generated in your converter and the systems before and after it probably defines the noise floor of a modern '24-bit' system more than the number of bits itself. We can certainly hear a noise floor at the theoretical -96dB of 16-bit, but we will never know (well, certainly not at the moment) how

the theoretical -144dB floor of a 24-bit converter sounds, because all those molecules rattling around at room temperature produce more noise than that.

As to how many bits we actually need in real-world systems to be sure that we are delivering the lowest practical noise floor: we can only think empirically right now in the absence of any firm data (that I know of anyway). A little under 24 bits sounds as if it is probably sufficient.

Stuart actually suggests that 20 bits is fine if you dither it properly. He also reasons that an audio bandwidth of 26kHz is also sufficient, again, if it's done properly, and that "further benefits would not accrue until the sound had been rendered fully 3-D."

It must be remembered that, while a minimum of 20 bits may be fine for a transmission medium such as a consumer audio disc or a recorder, if you perform any operations in the digital domain, such as EQ, compression or just changing levels, you will generate more bits — longer words — that you will need to re-dither down to the desired output resolution. This is why most DSP devices have internal busses wider than the word length either input or output. It is also not unreasonable to argue that a studio system should have higher density than a consumer distribution system, but Stuart warns that not only is well-handled, carefully delivered 20-bit data and a 24-bit processing environment good enough, but to deliver anything more is virtually to guarantee "a higher risk of inadvertent truncation in the average replay chain."

The other PCM misconception that Stuart addresses is a similar idea, but in the time domain: that PCM cannot resolve time more accurately than the sampling period. His answer is similar to the one about detail resolution: "Regarding temporal accuracy, if the signal is processed incorrectly (ie. truncated), it is true that the time resolution is limited to the sampling period divided by the number of digital levels. However, when the correct dither is used, the time resolution also *becomes effectively infinite*." (my italics)

This seems contrary to Mr. Moorer's previously stated premise, and suggests instead that stereo or surround localization depends more on correct dither than on higher sample rates. Moorer responds: "What [Stuart] says is correct — that if dither is applied properly, you can produce a waveform that you can adjust on the sub-microsecond level, and still get the waveform to change smoothly and evenly. While that works [at conventional sampling rates], it would work better at 96 or 192, or with DSD. And, that's not the same thing as reproducing sub-sample structure, which, of course, you can't do without a higher sampling rate." The dCS test results also suggest that there may be other things going on that negatively affect localization. ä

BEYOND THE PALE

For more detail on this topic, Stuart's paper is available in its entirety on the Acoustic Renaissance for Audio (ARA) website, http://www.meridian-audio.com/ara/.

The Other Way Of Doing It

There is another approach to high-quality digital audio conversion edging its way into the market, and that is Sony and Philips' Direct Stream Digital (DSD) and its related disc format, the Super-Audio CD (see *Audio Media* October 98). For this discussion, we need to look at the conversion process in rather more detail.

Modern PCM A/D conversion systems don't actually measure the instantaneous analog voltage and store it as a multi-bit word, as suggested earlier. Instead they use a one-bit 'delta-sigma' converter to produce a stream of pulses (see 'High-End Audio', p114 this issue). A negative feedback loop is used to accumulate the audio waveform. If the input accumulated over one sample period is higher than the value accumulated in the feedback loop during previous samples, the converter outputs a '1'. If it's lower, it outputs a '0'. The instantaneous amplitude of the analog waveform is represented by the density of pulses, and is sometimes called Pulse Density Modulation. In a PCM system, the PDM stream is chopped up into digital words by a decimation filter. On replay, the process is essentially undone (see Figure 1).

DSD sounds like an elegant alternative. Instead of chopping up the PDM stream, and then untangling it again later, why not simply record it (see Figure 2)? On replay, the helpful characteristic of the PDM stream — that it looks very much like the analog waveform — makes it simply require a decent analog low-pass filter to recover an analog signal. This sounds very impressive, especially when you consider that the sample rate at which the PDM stream is encoded is an enormous 64 times the usual CD sample rate: 2,822,400Hz.

Unfortunately, it isn't as simple as that. First of all, the delta-sigma conversion process is fairly noisy. Fifth-order noise-shaping filters are used to push the noise way out of the audio band, but it's still there and has to go somewhere, most likely into the replay system that may not be able to handle it. There are two ways of dealing with this: either filter it out (which might compromise the audio quality) or build replay systems that can handle it. The latter seems to be the preferred solution, and this may be the reason why some SACD replay environments include amps and speakers with 100kHz reproduction capability — not, perhaps, because you and I can hear 100kHz, but because, as they can handle the noise at those high frequencies, it doesn't get turned into distortion products. There's also the possibility, according to Hawksford, that all that high-frequency, high-level noise can risk introducing jitter, making the DSD stream more jitter-susceptible than PCM.

For those who do not record direct to two-track (or six-track for surround), there is worse to come. There is no doubt that DSD sounds great. The dCS tests described earlier are scanty on their subjective discussion of DSD but, in most senses, they tend towards thinking it sounds best of all. However, all that anyone has done with DSD so far is to record things, edit them, and play them back. If you want to perform DSP operations on a DSD signal, the only way we can do it at present is to decimate the PDM to PCM, or convert it to analog, process, then convert back. Native DSD processing will require silicon that does not yet exist. The first Sonic Solution systems to offer dynamics processing and equalization (currently in development) won't go quite as far as turning the whole signal into PCM (Figure 3). According to Moorer: "We do turn the signal into PCM, and then we run a dual chain, where we take the unmodified signal, and the DSP'd signal, and we get the difference of them and we turn that back into DSD and add that to the DSD stream." This novel approach, while ingenious, is rather complicated and is ill-suited to applications such as multi-channel recording consoles.

Hawksford also points out that, to get good results, bitstream encoders require high oversampling ratios for acceptable performance: it isn't simply impressive to have an enormous sample rate — you require one. The result is a very wide bandwidth data stream at 2.822Mbit/s. Hawksford compares this with a losslessly-compressed 96kHz PCM stream at around 1.25Mbit/s. PCM is, Hawksford alleges, simply more efficient.

Conclusions

On the PCM front, it would appear that 24/96 is going to be enough for most people, if it's done right. If it's done even more right, the audiophiles among us may prefer 192kHz sampling, although it does rather have the feeling of 'too much of a good thing' with problems in the bass end that remind one a little of 30ips analog.

The debate between PCM and DSD is not going to be resolved by technical superiority. It will hinge on marketing expertise, available titles, available machines and their features — stereo versus multichannel, stuff like that. Very likely the DSD process and its distribution medium, Super-Audio CD, will find favor among classical recordists and those employing simple digital signal paths, where it may offer levels of quality superior even to 24/192 PCM. But, for the time being, the most effective way of processing such signals may be to use high-quality analog equipment — a specific design goal behind at least one new analog console. o

Richard Elen is a frequent writer on professional audio and can be contacted at relen@igc.org. He is VP of Marketing at Apogee Electronics. The views expressed in this article are the author's own and are not necessarily those of his employers.