An Introduction to Digital Audio

by Julio Alvarez and Richard Elen

What are digital converters, anyway?

The topic of digital converters often makes people's eyes glaze over. The purpose of dedicated external converters can be even less clear. "If digital audio is 'just numbers' and digital gear already has built-in converters, why use external converters?" Indeed, an external converter unit doesn't record audio or mix or add any effects to the sound, so why do they exist? Why have you just bought one? These are the questions this section is intended to address, and while answering them, we'll try and de-mystify the whole concept of digital audio.

Why Digital?

By now, most of us know about the important advantages of digital audio over analog: storage on inexpensive media; non-destructive editing; transmission via phone lines and the Internet; multiple copying without quality loss even across different formats, etc. Unlike analog audio, which is prone to change (degradation) when carried from medium to medium, digital data, if carried without numerical error, will not change at all. Well, that's the theory, anyway.

The proliferation of low cost, professional-quality digital recorders, consoles, and digital audio workstations has brought recording to a whole new sector of the music community. Digital audio has transformed the way we record music: digital samplers, digital reverberation and effects have brought seeming impossibilities (or at least extreme improbabilities) into the realm of reality.

What do converters do?

All digital recorders, digital mixers, and digital processors must use either built-in converters or outboard "dedicated" converters somewhere in the audio chain to input audio and play it back.

To understand the advantages of high quality conversion, let's back-up and explore just what converters do in the first place. Very simply put, an audio analog to digital converter converts analog audio signals into digital data so that they can be stored, mixed with other digital signals, manipulated by digital signal processors, and transferred across interconnects as numerical information. Digital to analog converters do the same thing in reverse – they accurately reconstruct an analog signal from the digital data-stream of numbers.

High quality converters are difficult to make. As a result, you will quite often find that the converters built into a piece of equipment such as a recorder or a PC audio card are cheap and relatively nasty as far as quality is concerned. This is hardly surprising, because in a complex system the converters are just one part of the whole. Yet the converter – especially the analog to digital converter – determines the maximum level of quality that the digital system can attain. If the converter is poor, the results will be poor – whatever the quality of the rest of the chain. The solution is to use a high-quality external converter to insure that the maximum digital quality is delivered to the rest of the system. The question of what makes a high-quality converter is something we're going to address here.

Where an outboard converter can make a difference is that all the effort has gone into producing *only the converter* — it can hardly help being better. Often outboard converters are made by companies that do virtually nothing else: like Apogee, for example.

Analog versus digital – the basics

An important concept to understand is that digital is simply an alternative method of carrying, transmitting and storing audio. There are several ways to convert analog audio into digital data. The most widely used method is called **pulse code modulation (PCM)**.

In a PCM system, an analog signal is received at the input of an A/D converter where the instantaneous voltages of the signal are measured at regular intervals. For example, in a system using a sample rate (a.k.a. sampling frequency or Fs) of 48 kilohertz (kHz), 48,000 measurements are taken per second. To accurately sample an alternating current (AC) signal (i.e., a signal with a positive and negative phase – all audio signals are alternating), an A/D converter must measure it *at least twice per cycle* (we'll look briefly at why in a moment). Therefore, the highest frequency that an A/D converter can sample is equal to half the sampling rate – this is derived from the "Shannon-Nyquist Theorem" – and the maximum audio frequency a digital system can capture with a given sample rate is often called the "Nyquist Limit". So a 48 kHz sample-rate recording can accurately capture a 24 kHz signal; and a 44.1 kHz sample rate, a 22.05 kHz signal.

The reason for those slightly odd numbers, by the way, is historical. The Compact Disc uses a sample rate of 44.1 kHz, which means that it can carry audio frequencies up to about 20 kHz (it's not quite 22.05, because there needs to be a little extra room at the top to apply a filter to cut off any signal above half the sample rate). 20 kHz is generally regarded as beyond the highest frequency most people can hear so, the designers reasoned, there was no need to sample at a higher rate. And in fact 44.1 (for CD) and 48 kHz (the "professional" standard sample rate for many years) were about as high as the technology would allow at the time the standard was originally written, in the early Eighties, and have a simple relationship to frequencies used in video.

Digital Measurement Accuracy

So once every sample period, an A/D converter measures the instantaneous voltage appearing at the analog input, and it represents this voltage as a number. Binary numbers (0,1) are used instead of decimal numbers (0,1,2,3…9) because binary numbers are easy to represent in a circuit as either an "OFF" (no) voltage (0) or an "ON" voltage (1). A string of BInary digiTS, or bits, form a word and the number of bits in a word is called the word-length. The highest amplitude that can be represented in a given word is when every binary digit is a 1 (e.g. 1111111111111111 for a 16-bit word). This is called *full-scale* audio.

In decimal form there are 65,536 discrete steps in 16 bit audio (2^{16}), the standard for Compact Disc; 1,048,576 steps in 20 bit audio (2^{20}); and 16,777,216 steps in 24 bit audio (2^{24}). When an A/D converter measures an analog voltage that surpasses the digital ceiling (i.e., when the amplitude is out of the measurement range of the converter – above full scale), harsh digital distortion occurs. This is because when all the bits are set to "1", there's simply nowhere else to go. The earliest converters actually "wrapped round", so the next highest level above all ones was all zeroes – and this sounded dreadful. Nowadays, it's more common for an excessive signal level simply to "peg" at all ones, like a VU meter pegged to the end of its scale by an over-level signal.

Digital levels in excess of "all ones" are referred to as "overs". Some converters let you decide how many consecutive instances of "all ones" actually constitute an over. For example, if there's only one sample that's reached all ones, it's likely that the signal level went up to the maximum and then fell back – a quick transient perhaps. But if there's a whole string of consecutive "all ones", it probably means that the signal went way over before coming back to more acceptable levels. In most cases a couple of digital full-scale samples in succession is OK, so some converters let you decide that an "over" can mean up to four consecutive full-scale samples. But when you send a recording for digital mastering, you should be sure that it contains no more than one full-scale sample in succession.

The number of bits in the digital word defines the *dynamic range* – the difference in level between the loudest signal (full-scale) and the quietest (when all the bits are set to zero except the smallest one). A 16-bit system has a dynamic range of about 92 dB. A 24-bit system on the other hand *in theory* will give you 144 dB. Whether you can hear 144 dB, and whether a real piece of audio equipment can actually deliver it, is another question.

The word-length defines something else, too: the size of the smallest *difference* in level that you can capture, and that influences the perceived quality of the recording. Let's look at the issue of *quantization* in a little more detail. A "quantum" in atomic physics is a single, tiny, perhaps indivisible particle of energy. Quantization, therefore, refers to the tiny incremental steps in digital level between a zero and a one. Imagine an analog signal gradually and smoothly getting louder. As it does so, the digital data output by an A/D converter also increases, but instead of increasing smoothly like the analog signal, it increases in a series of steps. The size of the steps is determined by the number of bits, the smallest step being the difference between the smallest, or "least significant" bit being set to zero and set to 1. (Don't worry: you can't really hear the steps. A filter in the D/A converter smooths them all out. But you *can* hear the difference made by the size of the steps! At least, you can most of the time.)

This is a difficult idea to get across, so let's try some other examples.

On the quiet end of the dynamic range, if an analog signal is low enough that it cannot be quantized up to the first discrete value, it is quantized to 0 (silence). This is called the digital noise floor. If you were measuring the height of a 10-story building and the only measurement unit that was available to you was the meter (i.e, you were unable to use centimeters) you would have a pretty good idea of how tall the building was. You could say it was roughly 40 meters tall. If you were allowed to use centimeters, you could more accurately say that the building was 39 meters and 67 centimeters tall. However, if you were measuring the height of your cat and were unable to use centimeters in your measurement, you could only say that your cat was either 1 meter tall (a big cat) or 0 meters tall (an impossibly small cat). With centimeters available to you, you could obviously be much more accurate.

In another example, if you owed me $1,682.43 and you only had thousand-dollar bills (assuming you couldn't get change), you would either have to pay me $1,000 or $2,000. If you had $100 bills you could pay me either $1,600 or $1,700. If you had tens, you could pay me $1,680 or $1,690 and so on. If you owed me a dollar and you only had hundreds, you would probably round it off to the nearest denomination, in this case 0. The smaller the denomination you have available, the more accurate you can get in your payment or measurement. The same applies to digital audio – the more measurement values you have available (i.e., the higher the resolution, the greater the number of steps, and the greater the number of bits) the more accurately the original signal can be reproduced.

It turns out that each additional bit of resolution doubles the number of discrete measurement steps and, thus, adds 6db of dynamic range. With 1 bit resolution, you can get two measurements (2^1) – 0, 1; with 2 bits you can get 4 measurements (2^2) – 00, 01, 10, 11; with three bits you can get 8 measurements (2^3) – 000, 001, 010, 011, 100, 101, 110, 111; and so on.

Digital Audio is like film (kind of)

So the sample rate determines the highest *frequency* you can capture with an A/D converter, while the word length determines how much *detail* you can capture. Let's look at this another way for a moment, by thinking about movie film.

There are loads of different widths of movie film – from 8mm to 70mm. The wider the film, the better the picture quality. 8mm is fine for grainy little home movies but when you want to make that wide-screen epic that looks like real life, only 70mm will do – if you can afford it. If you can't, 35mm gives very nice results and is a good compromise between quality and cost. In the same way, the word length of a digital signal defines the sound quality: 8-bit is these days even too scrappy for computer games and the Internet; 16-bit has been a good average for a while – a good CD can sound very good, for example, and they are 16-bit – but today's 24-bit DVDs give superior results. Few people can contemplate the expense and difficulty of making 32-bit recordings: it's doubtful that anyone would notice the difference, and it may not even be possible to hear the subtle difference over the noise produced by components in the equipment.

You probably know that a reel of unedited movie film consists of a huge number of frames, each subtly different from the one before, and each a still photograph – a snapshot of what was happening at a given instant in front of the camera. We get the impression of a moving picture only when the film is moved past a shutter which shows us each frame in succession – and the frames must go past fast enough for us to see them as fluid, continuous movement rather than a jerky succession of still photos. The speed of motion is usually specified in frames per second, and the more frames per second, the smoother the movement appears to be.

A digital sample captured by an A/D converter is like a single frame of action captured by a movie camera. The faster the samples are taken – the higher the sample rate – the more information is acquired. In film, a higher rate makes the action more smooth and fluid – although above a certain speed we can't really notice the difference. In digital audio, the faster the sample rate, the better the frequency response – although above a certain point, we probably can't hear the difference.

Keep in mind that the sample rate defines the frequency response, and that word length (a.k.a. bit rate, bit depth, resolution) limits the detail – the more bits, the more subtle the nuances of level changes that can be captured.

On the other end of the signal chain, digital to analog converters reconstruct the analog waveform by outputting a voltage that corresponds to the values being fed to it.

Problems

In theory, this process seems perfect, but as always, there are several problems associated with it.

Jitter – Take 1

Jitter is the enemy of all digital systems. It is defined as any irregularity in the timing of samples being received at a D/A converter's input or in the sample timing of an A/D converter. Let's rewind a bit again and analyze the analog to digital conversion process.

The timing regularity of the sampling process – in other words, exactly *when* each voltage sample is digitized – is controlled by a crystal clock when the converter is in "internal" sync mode, and/or a phase-locked loop or PLL (which essentially causes the clock to lock to the timing provided by an external signal) when the converter system is being clocked from an external sync source such as Word Clock, a digital input, or video sync. In either case, the signal from the clock tells the A/D converter when to take each sample measurement.

The accuracy of the timing between each measurement is directly proportional to how accurately those measurements represent the analog signal. To see why, let's go back to the movie film analogy for a moment. A film camera might take 24 frames per second. As an analogy, imagine a film camera filming a horse running by at a constant speed. If the time between each shot is the same, the motion of the horse looks fluid and correct. Each shot would have the horse advance by the same distance, say 2 feet. If the time between each shot is not the same, the motion of the horse would look "jittery" and unnatural — the "image" would not be correct. In one shot the horse might have advanced 2 feet, in the next it could have advanced 4 feet, and in the next 1 foot. This is an exaggerated example, but you get the picture (no pun intended).

To think of it another way, imagine a drummer playing exactly in 4/4 time. Sounds good. Then imagine that the second beat comes a little late, and the fourth beat a little early. Does it sound as good? Well, it might sound interesting, but it won't sound the same. Imagine that you *recorded* the drummer playing in perfect time and it *played back* all over the place. *Definitely* not a good idea. This, on a super-magnified scale, is an idea of what jitter can do. It can mess up the timing – but on a subtle microscopic level. Unfortunately it turns out that the human ear and brain are much more sensitive to these tiny timing irregularities than was previously thought.

What does jitter sound like? Amongst other things, jitter interferes with the brain's ability to experience a stereo soundstage – to gain an impression of the relative positions of instruments and effects when a recording is played back. Jitter smears the audio soundstage; the sense of width and depth is skewed, narrowed or even lost altogether – because the arrival time of individual samples is being skewed – smeared across time.

Jitter is produced in a variety of ways. It can be the result of bad analog design, electromagnetic interference getting into digital audio interconnects, the wrong kind of digital audio cable, and other culprits. The problem is that *once a jittery signal is recorded, that's it* – it's that way forever. There are steps (such as reclocking, or even clocking the signal into RAM memory and out again) that can be taken to prevent *further* degradation of the signal, but those techniques cannot fix the quality of the initial recording. The recorded digital data represents the analog signal exactly as the A/D converter measured it – in theory, at least (you can also get jitter introduced in the recording process, for example by speed variations). And the quality of the converter is defined by how accurately that measurement was taken.

Dither and Word length

What is dither? Why do digital systems need it?

We've seen how the longer the word length, the better the quality of the digital signal – up to the point when we can no longer hear the difference (which is probably just short of 24 bits in most cases). The number of bits not only defines the smallest difference in signal that can be captured – i.e. the amount of detail, the "size" of the "steps" – but also the "height" of the "staircase", in other words the dynamic range, or the difference in level between the loudest and quietest sounds which can be converted. What happens when a bit changes between a one and a zero is essentially inaudible at high levels, because there's so much going on. But if you're dealing with low-level signals, such as reverb tails or a fade, the transition of bits from zero to one and back again becomes increasingly important.

In the analog world, as a signal dies away, it does so smoothly (assuming that nothing is wrong with the system). As the level drops, the signal gets progressively quieter. At some point it reaches the same level as the noise. But importantly, if the signal level continues to drop, you can still hear it, despite the fact that it is below the noise level. This is an important aspect of the way that analog signals behave – you can hear coherent audio information even when it is significantly lower in level than random noise.

In the raw digital environment, everything is different. As the level of a signal drops, it is represented by fewer and fewer binary digits, and the changing of these bits becomes increasingly noticeable – it's called "quantization distortion". Ultimately, you simply run out of bits, and when this happens, the signal simply stops, and in a 16-bit system, this happens at an audible level. This behavior is another of the several factors that gave early digital recordings a bad name, and led to some pundits claiming that digital audio was fundamentally inferior to analog.

A solution to the problem was to add noise to the signal. At low levels, the effective result of this procedure is to turn the last few bits on and off at random, smoothing out the sound and ensuring that everything will not simply disappear as the level falls. This noise is referred to as "dither noise" or simply "dither". The word literally means to tremble or quiver – a reference presumably to the least significant bits being turned on and off at random.

The down side of this process, however, is that you are introducing noise into the system and therefore degrading its performance by raising the effective noise floor. More than that: the noise is actually quite objectionable. Truly random (white) noise contains all frequencies and is particularly obnoxious as a dither signal – called "flat dither". More commonly, a noise spectrum where the highs and lows are rolled off ("triangular dither") is used to make the dither noise smoother and less obvious.

Dither of some sort is essentially a "necessary evil" of digital recording, and virtually every modern converter includes it, whether the documentation says so or not. Without it, the resulting quantization distortion can sound very nasty.

Several manufacturers and researchers have attempted to improve the quality of dithered signals by developing methods of hiding or "shaping" the noise created by the dithering process.

This becomes even more important when a signal is recorded, say, in 24-bit form and must be reduced to 16-bit, for example for Compact Disc. If it was possible to preserve the detail of a 24-bit recording by making the noise floor more transparent – more like analog – then the result would be an audible improvement in the quality of the final CD, and there would be audible benefits to be gained by recording beyond the 16-bit level, even if the end result was a conventional 16-bit Compact Disc.

Most “noise shaping” techniques rely on the fact that the ear is more sensitive to midrange frequencies (around 4 kHz) than it is to either low or high frequencies. In transferring a 20-bit recording to the 16-bit world of Compact Disc, for example, the last four bits of the 20-bit signal are removed and fed back into the input signal via a filter that both adds dither and changes the spectral shape.

Originally, the filter shape proposed by researchers (primarily at the Audio Research Group of Waterloo University, Ontario) was one based on psychoacoustic principles, adding more noise in the upper frequencies while lowering the noise floor at around 4 kHz, where the ear is most sensitive. (In fact, even better results would be achieved by adding the noise back in at low frequencies, where the ear is even less sensitive – remember “loudness” controls on your hi-fi? – but it’s more difficult to do). However, a number of manufacturers have since claimed that their own proprietary filter shapes are audibly superior.

Unfortunately, these noise shaping techniques can cause problems. First, although they lower the noise floor at the most audible frequency, they actually increase the overall noise. In addition, they can add audible artifacts to the sound.

A different approach is taken by Apogee with the UV22 process, incorporated originally in the $6800 UV-1000 Super CD Mastering System and subsequently made available in all Apogee’s A/D converters (and occasionally in Apogee D/A systems such as the DA-2000).

UV22 is not a different flavor of dither noise. Instead, the data from the least significant bits of a signal are essentially modulated on to the 16-bit signal according to a special algorithm, which adds an inaudible high-frequency ‘bias’ to the digital bit stream, placing a ‘clump’ of energy at around 22 kHz (the origin of the name “UV22”). This results in an essentially flat noise floor, which is at the theoretical 16-bit level – 4 to 5 dB below that of conventional ‘flat dither’. In addition, the noise floor does not have the distinctive and annoying ‘hissiness’ of conventional dither. Thus the UV22 noise floor is audibly quieter and less objectionable than other techniques. In addition, there are no audible artifacts. Yet, as with analog, you can hear coherent audio signals several dB below the noise “floor” – thus retaining much of the detail and audio quality inherent in the original signal.

Today, some imitators are trying to make their own copy of the UV22 process by using DSP technology – an approach that Apogee tried and rejected long ago as inferior. In the meantime, the Apogee process has been redesigned and refined to capture even more high-resolution detail for output at either 16 or 20 bits. The latest version, UV22 HR (“high resolution”) offers the best detail possible – and still out-performs the competition.

Word length reduction options

The point of using a long word length in an A/D converter is to capture all the subtle nuances of an audio signal. Today's Apogee converters offer 24-bit resolution, which means that the steps between bits are small enough to capture virtually all the detail you can hear. If your output medium does not offer the same or greater resolution, however – such as 16-bit CD or DAT – you have a decision to make: how to reduce the word-length to the maximum resolution the medium can handle, while retaining the highest quality.

The simplest choice is to simply throw the full 24-bit signal at your DAT machine and hope for the best. In the case of most DAT recorders, this will result in the *truncation* of the digital word length from 24 to 16 bits. The remaining eight bits are simply discarded. So there was almost no point in using a 24-bit converter in the first place (although it could be argued that a 24-bit converter has probably been designed more carefully than an old or cheap 16-bit one, so it may still sound better than if you had used the DAT machine's on-board converters).

Another option is some kind of *dither* or *noise shaping* as discussed above. This will sound better than simple truncation, because of the lack of quantization distortion, but you will have lost the fine detail.

We would suggest that UV22 is the best word-length reduction option of all. It reduces the word length and delivers the noise level you would theoretically expect for the output resolution you desire (there's no dither noise to raise the noise floor above the theoretical level). Yet it captures virtually all the detail of the original, and you can hear that detail even below the noise floor – just like analog. Digital audio that sounds like analog – what could be better?

So, now you've recorded that signal

Jitter – Take 2

An AES or S/PDIF digital signal (and some other kinds of digital signal) contains embedded timing information so that all of the samples can be received correctly at the other end. The clock inside the receiving device provides the timing information to the D/A converter so that it can accurately recreate analog audio from the data. So, the receiving clock must line-up ("resolve" or "phase lock") its clock signal to the clock information coming from the sending device. If this is done poorly, *voilà*, more jitter – more smearing of samples across time.

But it's not over yet. The signal can be further degraded when it is sent digitally across cabling to other digital gear. Induced noise or Radio Frequency Interference (RFI: you know, those TV/radio/cell phone/etc. waves that often make your guitar amp play your favorite tune – by itself!) affects jitter performance because it mingles with the digital signal and changes it.

Balanced lines help reduce interference. That's why the AES/EBU format uses 110 Ω twisted-pair, shielded cable. The cable shield effectively "shields" the conductors from most RFI and flushes any that it picks up to ground. Any RFI that gets through to the conductors gets phase-cancelled because each conductor is exactly 180 degrees out of phase with the other. Since the pairs are run together, any noise will be induced in both conductors in phase with each other and, thus, cancelled.

It's also important to insure that the cable has the bandwidth needed for digital signals. Those ones and zeroes are essentially square waves, with a very fast rise time between the two. A fast rise time translates to a very high frequency – up in the MegaHertz range. Regular mic cable simply isn't up to it, and quality digital audio cable doesn't have to cost the earth. In fact Apogee's Wyde Eye A/D cable is not much more expensive than microphone cable (because of extra shielding and more highly-specified materials, it isn't quite as flexible), yet it performs as well as some audiophile digital cables costing as much as $750 per meter! It's fine for analog interconnects, too.

High Density Conversion and DVD

Sample rates of 44.1 and 48 kHz have been the standard for over a decade, and in that time there have been persistent calls for higher sample rates. But if we can only hear up to 18 kHz or so, and 44.1 kHz sampling delivers at least 20 kHz of frequency range, why bother?

We rightfully want to make records with the highest quality gear available – preferably quality beyond that offered by consumer equipment. This is at least one reason why there is a quest for ever-higher sampling rates and longer word lengths. For most people, perceptible improvements in audio quality due to increased sampling rates for conventional PCM tail off when the sample rate exceeds about 60kHz – but 88.2 and 96 are nice simple multiples of current practice, which makes sample-rate conversion easier, so why not? 24 bits is an improvement over 20, which is a significant step forward over 16-bit. Of course, most of us will exploit all that dynamic range only during fades and in reverb tails.

48 vs 96

We can probably hear the difference between 48 and 96 kHz sampling in a quiet, modern studio, but it is difficult to say whether record buyers can.

And are 24/96 converters real anyway? Yes, but they are difficult to do well. You may be able to get 24 or so bits to wiggle 96,000 times per second, but that doesn't mean that the data itself carries any additional real information. Clock jitter is more difficult to deal with, for example, and noise levels in the analog stages – more than the digital circuitry – reduces the actual achievable dynamic range well below the theoretical 144 dB. But switch a converter between 44.1 and 88.2 kHz sampling and see what you think.

However, converters with higher sampling rates can still sound better. Let's see why.

Filters

The usual rationale given for the need for higher sample rates involves anti-imaging and anti-aliasing filters. These are required to insure that no audio information above the Nyquist limit (half the sample rate) passes through the system. You need at least two digital samples per analog cycle to accurately quantize a waveform, and if you have only one (for example because the waveform is at a higher frequency than the Nyquist limit), then you will sample a value, but it will be meaningless.

Imagine, using the film analogy again, that we make a movie of a flashing light (an independent art movie, evidently). We make the light flash on and off faster and faster as time goes by. At some point it will be flashing at half the film speed – say the film is running at 24 frames per second and the light is flashing 12 times per second. If we look at the individual film frames at this point we will see that one frame shows the light on, and the next shows the light off. The next, it's on again. This is fine. But now imagine the light has sped up so that it flashes 24 times per second. Now, each time a frame is shot, the light is on. Or maybe off. Ooops – if we look at a succession of frames we will see that the light seems to be on or off all the time, depending on what part of the cycle the light was in when the shot was taken – which is not what the light is really doing. Evidently, the film record is meaningless.

If you try sampling a waveform whose frequency is the same as the sample rate, the waveform behaves as if its frequency was 0 Hz! In more general terms, if you sample frequencies higher than the Nyquist limit, they behave like a mirror image of the frequencies below the limit. So if you're sampling at 44.1 kHz, a 30 kHz tone sounds the same as one at 14 kHz!

This is obviously undesirable, and as a result, digital systems from time immemorial (well, the last thirty years, anyway) have included anti-imaging and anti-aliasing filters to stop this very problem.

Unfortunately, these filters have traditionally sounded horrible. They need to pass all frequencies up to as close to the Nyquist limit as possible, but not a bit more. This makes their rolloffs very steep, and if implemented in the analog domain, as early ones were, they introduced enormous phase shifts into the audio – 1000 degrees out at 10 kHz was not uncommon. No wonder, then, that the top end sounded clangy and harsh and people said digital would never catch on.

The founders of Apogee Electronics were some of the first to tackle this problem. Our first product was a line of replacement filters for digital systems which used smoother slopes (less phase error) and superior components. The smoother slopes worked because, it turned out, there wasn't much audio up above the 20 kHz mark anyway, so imaging didn't matter as much as had first been thought.

A more complete solution would be to sample at a higher rate, so the Nyquist limit would be well out of the way, and thus the filters could be smoother and operate way above anything audible. Sounds good in theory, and this may be where the original push for higher sample rates came from. We would probably have used higher sample rates earlier, had it been feasible outside the laboratory.

Today, however, such filters are implemented digitally, and phase errors aren't such a problem. In addition, the type of converters used today multiply the effective sample rate internally, so that the apparent Nyquist limit is much higher than half the "real" sample rate. As a result of this "oversampling" technique, the filter rolloffs and frequencies can be kept well out of the way of the audio. So there's no need for higher sample rates. After all, there's nothing higher than around 20 kHz to record. Is there...?

Why Record Ultrasonics?

As is widely recognized, most of us can't hear much above 18 kHz, but that does not mean that there isn't anything up there that we need to record – and here's another reason for higher sampling rates. Plenty of acoustic instruments produce usable output up to around the 30 kHz mark – something that would be picked up in some form by a decent 30 in/s half-inch analog recording. A string section, for example, could well produce some significant ultrasonic energy.

Arguably, the ultrasonic content of all those instruments blends together to produce audible beat frequencies which contribute to the overall timbre of the sound. If you record your string section at a safe distance with a stereo pair, for example, all those interactions will have taken place in the air before your microphones ever capture the sound. You can record such a signal with 44.1 kHz sampling and never worry about losing anything – as long as your filters are decent and you have enough bits.

If, however, your idea of recording a string section is with a couple of 48-track digital machines, a mic on each desk feeding its own track so that you can mix it all later, you are doomed. Your close-mic technique does not pick up any interactions, so the only time they can happen is when you mix – by which time the ultrasonic stuff has all been knocked off by your 48 kHz multitrack recorders, so that will never happen. So if we want to be uncharitable, we could say that high sampling rates allow you to use bad mic technique with better results.

Pick A Number

Having established that higher sampling rates are a good idea – or at least a fact of modern life – there is a question as to what the sample rate should actually be in a studio environment. On the face of it, 96kHz takes care of capturing any audio that might ever happen, and 24 bits offer quite enough quantization steps. Is that enough?

Yes, in theory – more than enough. But there are some potential problems, real or imaginary, to having a production environment that has no better resolution than the consumer distribution format, and the emerging DVD-Audio standard offers not just 24-bit, 96kHz sampling: it even goes beyond that to support 192 kHz sampling in stereo.

[On the face of it this is quite absurd. Do we need to capture "audio" signals at up to 96 kHz? Obviously not – such signals don't exist. However recent research suggests that the human brain can discern a difference in a sound's arrival time between the two ears of better than 15 microseconds – around the time between samples at 96 kHz sampling – and some people can even discern a 5µS difference! So while super-high sample rates are probably unnecessary for frequency response, they *may* be justified for stereo and surround imaging accuracy.]

Think of higher sample rates and longer word lengths as a kind of "headroom". We need higher resolution in the studio than consumers so we can start with a higher level of quality in case some gets lost on the way – which might well happen.

And what happens when you modify a digital signal in the digital domain, say by EQing it, or fading it out? You create more bits – more data. You ought to have spare bits so you have room to work. You can always lose resolution, but you can't easily get it back again.

So, after all this, why use external converters?

In his widely-respected book, *The Art of Digital Audio* (Focal Press, 1994), John Watkinson writes, "The quality of reproduction of a well-engineered digital audio system is independent of the medium [tape, disk, etc.] and depends *only on the quality of the conversion processes*" (our italics). So, clearly, the primary reason for the existence of external converters is *better sound quality*. Whereas the internal converters of most audio equipment are just one part of the overall design picture, external converter manufacturers focus solely on the quality of the conversion process itself.

Hopefully the previous discussion has given you an insight into what converters do, and as a result, why converters are difficult to do well.

What do external converters do better?

There are several factors that determine the quality of the A/D and D/A conversion process. Among these are:

Analog design

The analog stage of a digital converter system is one of its most important design aspects. Self-noise, for example, can instantly reduce the resolution of an A/D converter. For every 3 dB of self-noise, a signal through an A/D loses 1 bit of resolution (did you ever wonder why 18- and 20- bit converters existed when the only available recording resolution was 16-bit?). The low-level resolution will be lost in noise. Be careful when you read specifications. The theoretical performance of the converter chip itself is never the real-world performance of the entire unit in use. A 24-bit converter will have smaller quantization steps than a 20-bit unit, but whether you can hear the difference in the noise floor is a different matter.

Most people believe that the important part of a converter is the converter chip itself – we are often asked "Which converter chips do you use?", the inference being that this is the primary factor in determining conversion quality. The fact is that this is not the case. Most converter chips from the small number of major manufacturers offer similar performance as far as conversion is concerned. A designer chooses one over another because of availability and specific features which suit their design. Of course, they'll measure the chip to see how it performs, and to insure it lives up to its claims, but in general the quality of a converter does not lie in the chip; it lies in *how you use it* – in other words in good design. And in that arena, analog design is the most often overlooked, yet it can make all the difference.

Additionally, poor analog design can produce noise on the clock lines which adds jitter to the signal. This brings us to the second important design consideration – the system's clock or phase-lock-loop (PLL).

Clock design

As we've seen, the accuracy and stability of the clock is vital in reducing jitter. This true whether the clock is running on its own or it is being locked to an external source. In the latter case, it's important that the synchronizing signal is *re-clocked* to remove incoming jitter on the sync signal – a particular problem when synchronizing to video.

Digital design

Because of their high frequency, digital signals behave in some ways like radio broadcasts, and the traces on the printed circuit board act as antennas. As a result, digital signals can easily leak into other circuits in a converter, creating noise and significantly degrading the performance – unless digital circuitry and boards are laid out carefully, and the right relationship is established between analog and digital areas.

Where will this all lead?

At the close of the 1990s, we already have a “high density” consumer digital distribution medium, in the form of DVD - the Digital Versatile Disc. In its video form, it can handle 5.1 surround and up to 24-bit, 96 kHz sampling – although the vast majority of discs around at the time of writing (early 1999) use 48 kHz sampling and seldom more than 20-bit word lengths.

The next incarnation of DVD, DVD-Audio, will certainly make use of high-density digital audio as a matter of course. DVD-Audio uses the same physical disc technology as DVD-Video, but focuses on high-quality multi-channel audio. From the start, it will offer over an hour of at least six channels of full-bandwidth up to 24-bit audio sampled at up to 96 kHz, or a similar length of 2-channel 192 kHz, 24-bit audio – thanks to the brilliance of Meridian Lossless Packing (MLP), the compression scheme mandated in the DVD-Audio spec. MLP also supports a number of innovative surround-sound encoding schemes including full three-dimensional replay, which will take us to the next level in our ability to create, or recreate, an acoustic environment for the listener at home.

It is for this future that your Apogee converter was designed: a future whose audio productions you will be creating.