

# Tapping Into Better Digital Audio

KEITH HOWARD EXAMINES THE BACKGROUND AND PHILOSOPHY BEHIND ROB WATTS' ULTRA-LONG FILTERS, FOUND IN CHORD ELECTRONICS DACS, ESPECIALLY THE M SCALER

It's an interesting piece of historical coincidence that 1948 was both the year that Columbia Records announced the LP and the year in which Claude Shannon's famous paper *A Mathematical Theory of Communication* [1] was published, describing the mathematical basis of signal sampling and thus digital audio. It is now acknowledged that Shannon wasn't actually the first to visit this territory; others, including English mathematician Edmund Whittaker, had elucidated all or part of the theory earlier. But Shannon's paper really launched the field of information theory and paved the way – awaiting the necessary technological advances – to the digitisation of all manner of continuous signals, not just audio waveforms.

Few audiophiles knew of this area of mathematics or its potential application to audio signals until, starting in 1972, the BBC began to replace audio landlines to its transmitters with 13-bit Nicam PCM digital audio links running at 32kHz sampling rate with companding (compression for transmission and subsequent expanding on receipt). The first, from Broadcasting House to the Wrotham transmitter in North Kent, began operating on 14 September that year, and the network was then progressively expanded. Later in the 1970s the first digital recorders were developed of sufficient quality (arguably...), to be used by pioneering record labels such as Denon and Decca. Then, in 1982/3, the Compact Disc arrived – the first digital audio music carrier.

Given that it's 35 years since CD went on sale in Europe, you might suppose that most audiophiles would now have a firm grasp of the basics of sampling theory, but in truth it remains widely misunderstood. And not just by audio amateurs but audio professionals too, if Rob Watts, Chord Electronics' Digital Design Consultant, is correct. For years Watts has been bucking the conventional audio industry practice of using relatively short – sometimes very short – digital filters in high oversampled DACs. As technology has allowed it, his filters have become longer and longer. And each time filter length is increased, he says, sound quality improves.

Watts' long quest to achieve a sufficiently long

filter that no further increase is of subjective benefit reached its latest apotheosis with the announcement, at the London CamJam show in July, of the Chord Electronics *M Scaler* (see Box-out), a digital in, digital out upsampler that for the first time features over one million filter taps, *ie* the linear-phase FIR (finite impulse response) interpolation filter employed has over one million coefficients: 1,015,808 to be precise. To put this number into perspective, most oversampling DACs use filters that are a few hundred taps (coefficients) in length at most. Even *Dave*, Chord's best current standalone DAC, uses 'only' 164,000 filter taps.

Understanding Watts' relentless pursuit of longer interpolation filters – each step of which has opened a still wider gap between him and accepted industry practice – requires going back to grass roots: to Shannon's sampling theory, and particularly to the 'sinc' function.

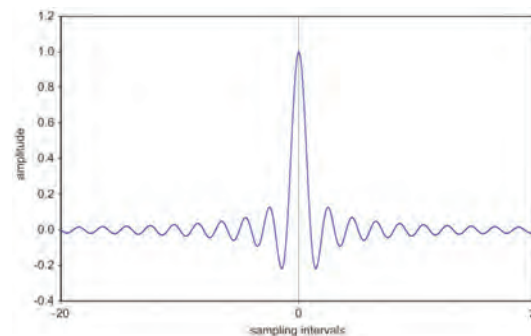


Fig 1. Central portion of the  $\text{sinc}(x)$  function

$\text{Sinc}(x)$  is mathematical shorthand for the function  $\sin(x)/x$ , which looks (over its central part) like Fig 1. The reason for the x-axis (horizontal axis) being labelled 'sampling intervals' will become apparent shortly. What Shannon showed in his famous paper was that any bandlimited continuous signal – that is, any analogue signal with a strict limit on its maximum frequency – can be exactly described as a sum of time-spaced  $\text{sinc}(x)$  waveforms.

Not only that, if the signal is sampled, *ie* if its amplitude is measured, at regular intervals, at a rate at

least double that of the highest signal frequency, each sample amplitude represents the amplitude of the associated  $\text{sinc}(x)$  waveform centred on that sampling point. At all other sampling points the value of that particular  $\text{sinc}(x)$  function is zero, just as the value of the  $\text{sinc}(x)$  functions centred on each other sampling point are zero here. So sampling the waveform as described extracts all the information necessary to reconstruct it, and to do so *with complete accuracy*.

This is the analogue-to-digital conversion process, and it's entirely practicable. All we have to add to make it realisable in practice is quantisation of the amplitude measurement, so that each sample value can be represented by a number of finite length.

By contrast, the waveform reconstruction process – digital-to-analogue conversion – is not so simple. In an ideal world it would be achieved by generating an impulse of appropriate amplitude for each sampling point and passing the train of regularly spaced impulses through an ideal low-pass filter with its passband upper edge set to half the sampling rate. The impulse response of such a filter is the sinc function, so each impulse would generate a sinc waveform of the necessary amplitude, and the train of sinc functions would sum to recreate the original waveform.

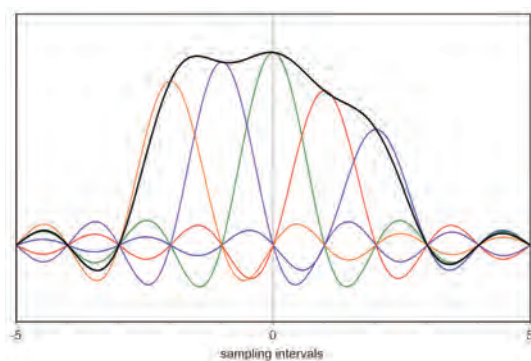


Fig 2. How the summing of  $\text{sinc}(x)$  functions – one per signal sample – builds the waveform between sampling points

This is illustrated in Fig 2, which shows six successive sinc functions of different amplitude and their sum (the black trace). Each sinc function contributes nothing to the summed signal amplitude at other sampling points, but does contribute to the waveform *between the sampling points*. It's a common misunderstanding of sampling to suppose that the waveform between sampling points is unknowable but that is not true – provided that the input signal is bandlimited and sampled at least twice as fast as its highest component frequency, as required by the Shannon sampling process. In that case the waveform between sampling points can be reconstructed unambiguously. Importantly, the

wave shape depends not just on the value of nearby samples but, ultimately, *on the pattern of samples throughout the sampled signal*.

In practice this theoretical DAC scheme is unrealisable, for two reasons. First, the ideal low-pass filter, with infinite roll-off rate at the passband edge, exists only in abstractions – in the real world, where filters always have finite rates of roll-off, it can only be approximated. Second, even if the perfect low-pass filter were not a dream, this approach would provide inadequate signal-to-noise ratio because of the tiny amount of energy contained in each impulse.

In real-world DACs two compromises have to be made. First, the amplitude of each sample is not represented as an impulse but as a step which is maintained for an entire sampling period. This 'sample and hold' process obviates the signal-to-noise issue but results in a non-flat frequency response that rolls off gently towards the Nyquist frequency (half the sample rate). The fix is trivial: the roll-off can be, and routinely is, corrected by equalisation. The second compromise I've already alluded to. Because an ideal low-pass filter is unachievable, one with a slower rate of roll-off must be substituted – and then it won't provide  $\text{sinc}(x)$  impulse response.

The way this last issue is habitually addressed is simply to ignore it. The analogue output filter (or in oversampled systems, the digital interpolation filter) is considered adequate if it achieves sufficiently good frequency domain performance, *ie* sufficiently flat passband response and adequate attenuation of the image frequencies which appear above the Nyquist frequency. Fig 3 shows the frequency response of an example interpolation filter, designed (using the well-known Parks-McClellan equiripple method) for 4× oversampling of 44.1kHz data with the following specifications:

passband upper frequency	20kHz
passband ripple	0.01dB
stopband lower frequency	24kHz
stopband attenuation	100dB

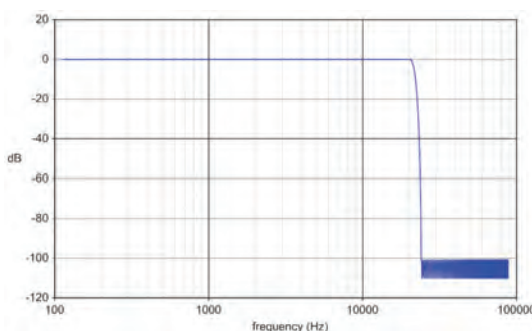


Fig 3. Frequency response of an example 4× interpolation filter

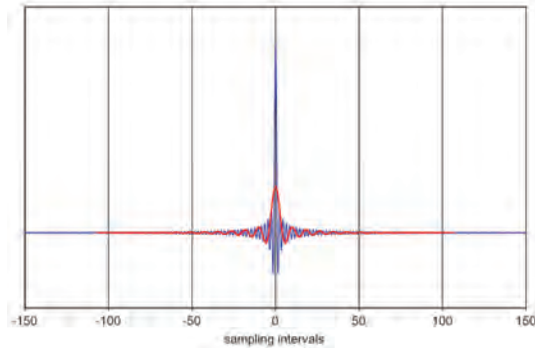


Fig 4. Impulse response of the interpolation filter of Fig 3 (red trace), overlaid on the sinc(x) function (blue trace)

The resulting FIR filter has 215 coefficients (taps) and, as it's linear-phase, a time-symmetrical impulse response. Attenuation at the Nyquist frequency (22.05kHz) is about 10dB. The filter impulse response is shown in Fig 4, overlaying the sinc function of Figure 1 but here across a wider range of sampling intervals to accommodate the number of filter coefficients. Fig 5 repeats the data of Fig 4, but represented on a decibel amplitude scale. It is clear from both graphs that the interpolation filter – though it meets representative frequency domain criteria – has an impulse response that is quite different from the sinc(x) function, and not just because it's shorter.

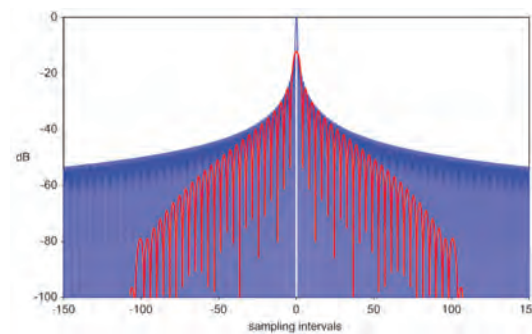


Fig 5. A repeat of Fig 4 but this time with decibel amplitude scale

Fig 5 emphasises an important point: that the envelope of the sinc(x) function – which is finite valued for values of x from minus infinity to plus infinity – decays slowly with time. At 150 sampling intervals from its central peak the envelope has only decayed by a little over 50dB. The obvious question is: by how much must it decay for its contribution to inter-sample wave shape to become insignificant? That's not a straightforward question to answer but if we say 100dB, to take the envelope amplitude below the 16-bit noise floor for a 0dBFS (full scale) sample, we can easily calculate what excerpt of the sinc function is required. The envelope of the sinc function is determined solely by the denominator

of  $\sin(x)/x$ , so it behaves as  $1/x$  where  $x = N\omega$  (angle in radians), N here being the number of sampling intervals away from the central peak. For the envelope to be 100dB down from its peak value  $1/x = 0.000001$ , which is equivalent to  $N = 31,831$ . This sample length is required either side of the central peak, so the total length of the sinc(x) excerpt is double this. In other words, for 44.1kHz sampling rate the total length of the required sinc function excerpt is 1.443 secs – pretty close to the filter length provided by Chord's *M Scaler*.

This, in a nutshell, explains Watts' pursuit of unprecedentedly long interpolation filters, which he designs using what's known as a windowed-sinc technique. As the previous paragraph suggests, this involves extracting a chunk from the centre of the sinc function, but for optimum results this process needs to be more subtle than a simple 'lift' of the sinc function values and truncation of the remainder. Better results are obtained if the excerpted sinc function is windowed, *i.e.* shaped, to avoid sudden truncation at either end. Watt's WTA (Watts Time Alignment) windowing algorithm is a closely guarded secret, and it has had to be refined as filter lengths have increased, but its name indicates Watts' principal design criterion: the maintenance of accurate transient timing.

So far as I'm aware, no other designer has followed in Watts' footsteps. And you have to suppose it would be a daunting task to do so, given the decades that Watts has been treading his lonely path. But if someone was minded to try – especially given the generally positive critical reception of Chord's digital products – how might they go about it?

The first thing to do is convince yourself that Watts' approach is right. The simple way to do that, naturally, is to listen to Chord's products, ideally a selection of them that chart the course of increased filter length. But there is another way, which is a lot cheaper than acquiring a collection of Chord hardware and much easier than taking the very major step of programming an FPGA. That's to perform sinc interpolation offline, in software.

I first wrote a software utility to do this over 10 years ago – and compared to FPGA programming it's an absolute doddle. The problem is, the program takes ages to run with anything longer than a very short audio file because it requires the calculation of  $(U-1) \times N^2 \sin(x)/x$  values, where U is the oversampling factor and N is the number of samples in the file. This is for each channel. But, like I say, it's easy, it's cheap, and it allows you to generate and listen to a file that's been oversampled using full sinc interpolation, in which respect it's even better than the *M Scaler*. Once you have the file, it can be used as a reference against which to audition others generated using

finite-length interpolation filters of different designs.

To show you an example, I ran the code using a short (0.98 second) mono, 44.1kHz/16-bit WAV file containing a single note played on a harpsichord. For 4× oversampling, the processing (which uses 64-bit floating point arithmetic and generates a 24-bit output WAV file) took 295 seconds – over 300× real time – running on a single processor core of my ageing desktop computer.

Fig 6 overlays the spectrum of the original file (red trace) and that of the oversampled file (blue trace), showing (a) that the two overlap as they should through the passband, and (b) that the sinc interpolation really does result in brick-wall low-pass filtering at 22.05kHz, the noise floor above that frequency being due to dither. (The original file was analysed using a 4096-point FFT and the interpolated file with a 16,384-point FFT, to ensure that the spectra have the same frequency resolution.)

I've suggested in the past, albeit never in print before, that someone with access to heavyweight

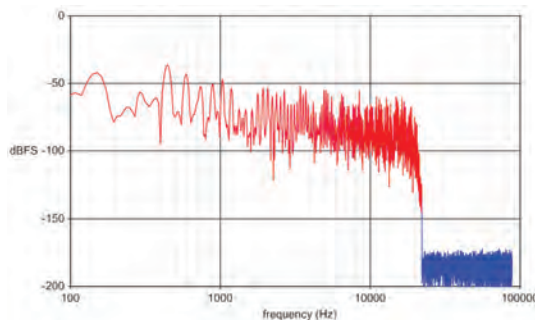


Fig 6. Spectra of a short recording of a single note played on a harpsichord. The red trace is of the 44.1kHz/16-bit original file, the blue trace of a 4× oversampled version generated using full sinc interpolation

number-crunching might use it to create a readily accessible cache of sinc-interpolated music files, precisely for the development of improved interpolation filters. This may not be an act of philanthropy likely to secure you a cover of Time magazine – but audiophiles might forever revere your name. Anyone interested?

## Reference

1) Shannon, C E. 'A Mathematical Theory of Communication', *Bell System Technical Journal* (1948). Reprinted in book form, with a layman's introduction by Warren Weaver, as 'The Mathematical Theory of Communication', University of Illinois Press.

## Chord Electronics M Scaler

I have a feeling that the *M Scaler* will prove to be one of the most significant, and probably controversial, products of 2018. It is sure to intrigue those who have already found Rob Watts' DAC designs to be a cut above, and just as likely to prompt the Monty Montgomerys of this world to dismiss it as delusional.

What the *M Scaler* does is take the oversampling technology from the Chord *Blu MkII* upscaling CD transport, and package it in a 40.5×235×236mm (hwd), 2.55kg box at less than half the price. Digital input is available *via* coaxial S/PDIF on two BNC sockets, optical S/PDIF *via* two Toslink sockets, and USB *via* a type B connector, to accommodate a wide range

of digital sources. Output is via single BNC at 352.8/384kHz, optical S/PDIF at 176.4/192kHz or, for full performance in conjunction with the *Quteq*, *Hugo TT2* or *Dave*, via dual BNC sockets at 705.6/768kHz.

The key component within the *M Scaler* is the Xilinx *XC7A200T* FPGA (field programmable gate array) chip which provides 740 DSP cores. Watts' latest filter architecture (also used in the *Hugo TT2*) employs 528 of these cores running at 4096× sampling frequency, comprises half a million lines of control code and uses 56-bit resolution in the calculations. A pass-through option is provided to allow instantaneous comparison of input and output, with gain correction to ensure that there is no disparity in levels.

[Oversampling can result in interpolated sample values that exceed 0dBFS (full scale), requiring a gain reduction to accommodate them.]

Initial testing with 512,000 taps

gave what Watts describes as a "completely unexpected" magnitude of improvement over the 164,000 taps of the *Dave*, so the target was raised to over a million taps. At the 705.6/768kHz output sampling rate (16× 44.1/48kHz) this means a latency – the delay between input and output – of around 0.6 seconds while the *M Scaler* performs its calculations on the first sample. Because this delay would cause unacceptable lip-sync issues when the audio accompanies video, the *M Scaler* offers a video input which uses an asymmetric interpolation filter instead that reduces the latency to an acceptable 0.1 secs.

What are the perceived sound quality benefits? Watts says that the improved transient accuracy of the longer filter makes instrumental timbre clearer, tightens bass and "dramatically" opens up the soundstage. "After you've listened to the *M Scaler*", he says, "it's very difficult to listen to the *Hugo TT2* or *Dave*." And has his thirst for more filter taps now been sated? No: "My gut feeling is that we need to go further."

Available this autumn, the *M Scaler* will be priced at £3495.

