Is The AESEBU / SPDIF Digital Audio Interface Flawed ?

Chris Dunn and Malcolm Hawksford

Audio Research Group, Department of Electronic Systems Engineering
University of Essex

email: chris.dunn@scalatech.co.uk

ABSTRACT

It is a requirement of high quality digital audio systems that all digital interfaces in the signal path exhibit signal transparency. The widely adopted AESEBU/SPDIF interface has received criticism from some quarters for a lack of signal transparency; this paper addresses possible problems with such interfaces and presents methods for improving the interface standard.

0 INTRODUCTION

In a correctly functioning (uniformly quantised and sampled) digital audio system the only observable signal impairments should be attributable to band-limitation and an additive noise residue. Thus although the subjective sound quality of digital audio has received some criticism since the launch of the compact disk (CD) medium 10 years ago, there can be little doubt that the *theoretical* performance obtainable from the 16 bit linear PCM format sampled at 44.1 kHz is superb compared to any analog sources available to the consumer. When correctly dithered using triangular PDF dither, a 16 bit digital audio signal possesses a dynamic range of 93.3 dB with zero distortion and zero noise modulation. Indeed, the 16 bit format holds the possibility of even higher subjective dynamic range when minimally audible noise shaping is employed at the mastering stage of the CD manufacturing process; Lipshitz et al. [1] show that an increase in subjective dynamic range of up to 18 dB is readily achievable when making the final truncation to 16 bits.

Since any practical digital audio system will err from this `ideal' performance, much effort has been spent in the last few years in attempting to minimise measurable errors in digital components. In the case of digital to analog converters (DACs), advances in circuit architecture including oversampling, noise shaping and 1 bit conversion have resulted in much improved low level resolution over early devices; we are now at the stage where the theoretical performance obtainable from the CD medium can be realised upon replay at a relatively low cost. As the quest for resolution has progressed, many `outboard' DAC units have appeared on the consumer market, where the sensitive DA conversion process is removed from the harsh electromagnetic environment inside the typical CD transport unit. Digital data is transmitted from the transport to the DAC in a serial format known as the Sony/Philips Digital Interface Format (S/PDIF) along a coaxial or optical link (Fig. 1). The S/PDIF standard is very similar to the AES/EBU format commonly used to interconnect professional digital components, and differs only in details including transmission amplitude and subcode format. For much of this paper both interface standards will simply be referred to as the `digital audio interface'.

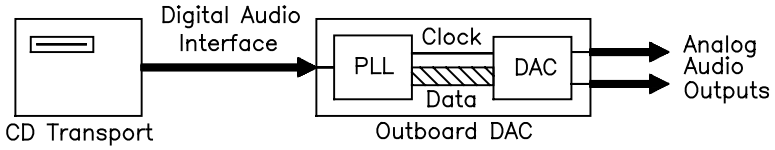


Fig. 1. 2 box CD replay system with transport and DAC linked using SPDIF digital audio interface.

As use of the digital audio interface has increased amongst consumers, some users have reported subjective differences between various implementations of the interface. Van Willenswaard [2] was among the first to note a change in outboard DAC sound quality when switching between different CD transport units, and linked this to measurable differences in interface signal rise time. Audio reviewers have made other claims concerning digital interface sound quality, including differences between optical links and wired coaxial connections, and changes in sensitivity to interface quality depending on DAC architecture.

Is the digital audio interface flawed ? Specifically, how can these claimed subjective differences occur in what is a *digital* data link ? - after all, "bits is bits". This paper examines possible impairments that can occur when audio data is transmitted over the digital audio interface. We will commence with an examination of the digital interface standard and follow this with a study of its susceptibility to both amplitude (`bit') and timing (`jitter') errors. Our analysis indicates that amplitude errors are extremely unlikely, but that jitter remains a problem, and indeed shows certain characteristics which are unique to the adopted digital audio interface standard. Jitter error models are developed for different DAC architectures and these are used to determine the audibility of interface errors. Finally techniques for reducing the audibility of interface errors are discussed.

1 THE DIGITAL AUDIO INTERFACE STANDARD

The AES/EBU and S/PDIF digital interface standards use biphase-mark encoding to transmit two channel audio data, synchronisation information and subcode data over a single serial information channel [3]; this coding scheme allows clock information to be embedded in the serial data stream. Fig. 2 shows the serial subframe structure consisting of 32 bit cells, each subframe carrying code for one audio channel. The subframe begins with a 4 bit synchronisation signal (`preamble') followed by a 4 bit auxiliary data block. Up to 20 bits of audio data can be transmitted, LSB first and the MSB occupying the last audio cell position. Finally subcode information comprises validity, user, channel status and parity bits. The biphase-mark encoding technique places cell transitions at the beginning and end of each cell if a `0' bit is transmitted, while a `1' has transitions at the beginning, end *and* midpoint of the cell; the preamble violates this coding rule so that interface receiver circuitry can detect when each subframe begins. If the audio data sampling rate f_s = 44.1 kHz then the cell (`0') width is equal to 354 ns while the half cell (`1') width is 177 ns; hence the maximum rate of transitions is equal to 1/177e-9 = 5.6 MHz, though it is true to say that harmonics of the interface signal will extend to frequencies far above this. Fig. 3 shows an example time domain simulation of a single subframe carrying a left channel audio sample of value 255, equal to 1111111100000000 in 16 bit 2's complement notation with MSB last. The mid-cell transitions can be seen at each `1' bit position, while biphase-mark violation displaces local cell transition positions in the preamble.

The biphase-mark signal can be transmitted using either a coaxial or optical connection, while the interface decoder at the receiver has to extract clock, audio data and subcode information from the serial data stream. The clock signal embedded in the serial data stream is usually used to control a phase lock loop (PLL), which in turn should provide a stable reference frequency for conversion circuitry which is interfaced to the analogue world. There are now a number of dedicated ADIC (Audio Digital Input Circuit)

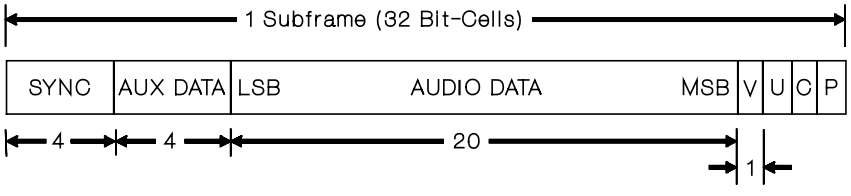


Fig. 2. Digital audio interface subframe format.

integrated circuits available which will perform these functions. An example circuit used in some of the experiments to be described below is shown in Fig. 4. This circuit uses the Philips SAA7274 ADIC; negative going edges on the SPDIF input signal are detected and compared to edges on the system clock derived from the PLL 11.2896 MHz crystal oscillator. A difference signal is fed to a varicap diode which pulls the PLL oscillator frequency to match the clock frequency embedded in the incoming interface signal. The phase lock loop has a 1st order loop filter with a break frequency of approximately 1 kHz, allowing clock recovery to reject short-term variations in the input frequency (*i.e.* high frequency jitter).

If the interface decoder supplies data to a DAC then the analogue audio output will be corrupted if either the samples are the wrong value (amplitude or `bit' errors) or they are output at the wrong times (jitter). The following sections will consider these distinct error mechanisms by simulating a hypothetical system comprising a noisy, band-limited digital audio interface decoded by an ADIC and fed to an ideal DAC.

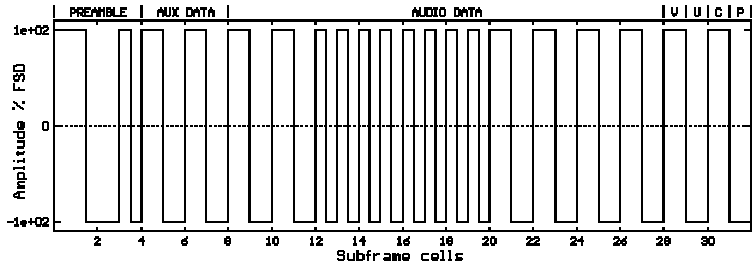


Fig. 3. Left channel subframe with audio data word equal to 255.

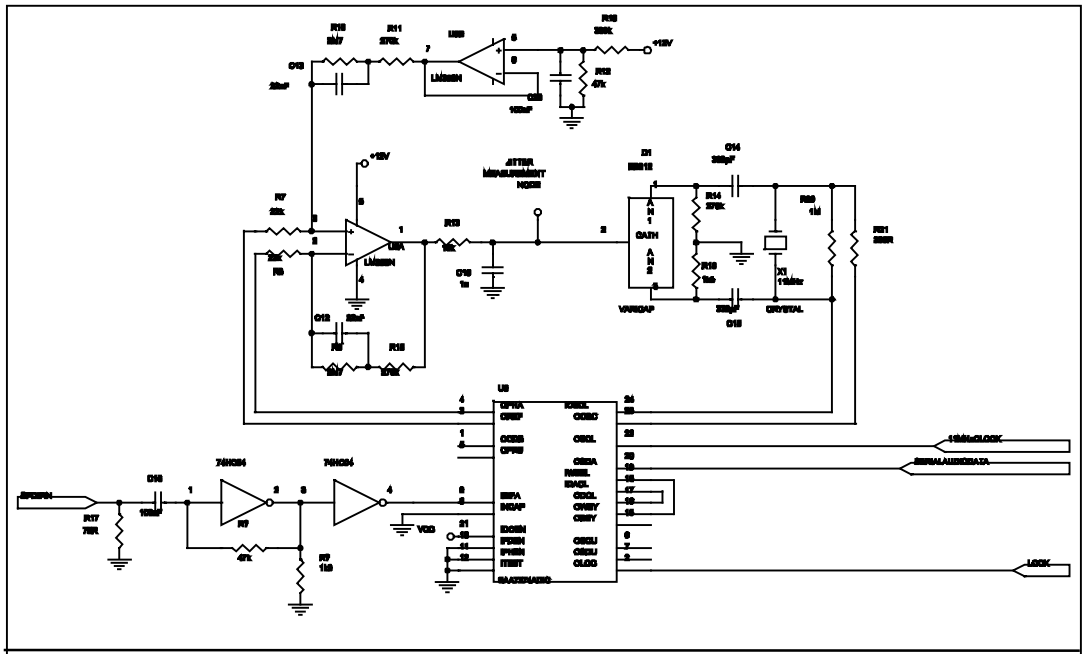


Fig. 4. Experimental interface receiver circuit using Philips SAA7274 ADIC.

2 AMPLITUDE ERRORS IN THE DIGITAL AUDIO INTERFACE

The unfiltered digital interface waveform is a binary signal where all transmitted information is determined by the *transitions* in the signal. One of the benefits of biphase-mark encoding is that the interface signal has only a small DC component, allowing interface signals to be AC coupled and edge detection to be performed using a comparator referenced to ground. If an audio data cell transition is missed at the receiver then a bit error occurs, and a DAC connected to the receiver will output an incorrect sample value.

We now proceed to model a bandwidth limited link by filtering the subframe signal with a first order (RC) lowpass filter and determine what degree of filtering will result in bit errors. Using the first order filter model is a gross simplification of the time domain behaviour of a real link where an accurate analysis requires the use of transmission line theory at the high frequencies involved, but is nonetheless a good starting point for investigation. Consider Fig. 5(a) which shows a simulation of the subframe signal carrying an audio word value 255 and filtered using a time constant of 100 ns, corresponding to a -3 dB frequency of 1.6 MHz (for all of the simulations and measurements presented in this paper, the audio sampling rate is 44.1 kHz). Although the filtered subframe now has edges with finite rise and fall times, no transitions are missed and hence no amplitude errors will occur as long as the receiver can latch the data correctly following transitions. Fig. 5(b) indicates the same subframe but filtered more severely with a 400 ns time constant, corresponding to a -3 dB frequency of 400 kHz. In this simulation the transition at the edge of cell 3 *is* missed; this will *definitely* result in a bit error in the decoded subframe. This particular example also indicates that receiver bit errors are most likely to occur during preambles where the largest variation in transition times occur (*i.e.* 1 half cell width to 3 half cell widths). This is interesting in that many practical ADIC ICs indicate full lock to the received signal when the preambles are correctly detected; hence if lock is achieved, bit errors are unlikely to occur. However, most practical interface decoders will exhibit an upper time constant lock limit considerably less than 400 ns, due to the finite `time aperture' about the average zero crossing point during which a transition is allowed without latched data errors. In Sec. 3 we derive an expression for the peak to peak variation in zero crossing times t_x in terms of the range of interface signal pulse widths; if we let the large pulse width be $3t_c/2$ and the smaller be $t_c/2$ (where t_c is the cell width in the biphase-mark coding) then [*c.f.* Eq. (11)]:

$$t_x = RC \ln \left[\frac{1 + e^{-\frac{t_c}{2RC}}}{1 + e^{-\frac{3t_c}{2RC}}} \right] . \qquad (1)$$

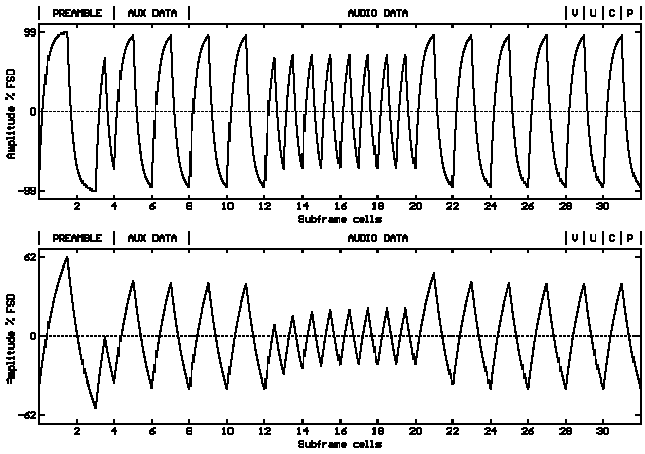


Fig. 5. (a) Subframe signal filtered with time constant of 100 ns. (b) Same subframe signal filtered with 400 ns time constant.

Using the experimental ADIC circuit shown in Fig. 4, full signal lock was achieved for interface time constants up to *RC* = 170 ns, indicating that for this particular example the maximum zero crossing time aperture t_x is approximately 45 ns. In practice time constants greater than 100 ns are excessive for digital links which should be designed with bandwidths well above 6 MHz. We have measured 2 m interface links with a characteristic impedance of 75 ohms, correctly terminated both at the transmitter and receiver, with 10 - 90 % rise and fall times of less than 10 ns; this level of performance corresponds to a bandwidth of 35 MHz.

In this section we have shown that bit errors in the received subframe occur when transitions in the interface signal are incorrectly latched. This will not occur in most interface receivers for interface time constants of less than 100 ns, and when bit errors do occur they are most likely to occur in the preamble which will also usually result in the receiver failing to lock onto the incoming signal. In essence we can state that audio bit errors due to band-limitation alone are extremely unlikely to occur. Of course this simple analysis does not consider the effects of noise upon the bandwidth-limited link. As the bandwidth of the link decreases then the eye diagram representation of the received signal suffers from a decreasing opening; this results in more time spent in the threshold region and the probability of noise induced errors increases. Nevertheless, Cabot [4] presents an interface example where bit errors are negligible for noise levels up to 20 dB below the interface signal level with *RC* filtering up to 160 ns, and claims to have achieved zero error rate transmissions over an unmatched digital audio link of 100 m length.

3 JITTER IN THE DIGITAL AUDIO INTERFACE

The second interface error mechanism to consider is that of modulation of the recovered clock frequency. If the clock fed to an ideal DAC varies in frequency then the reconstructed analogue output from the DAC will include error artifacts even if the sample *values* fed to the DAC are correct. The easiest way to analyse such an effect is by examining the *jitter* on the recovered clock (in this paper we define jitter as the instantaneous timing deviation of clock transitions from their correct positions). In this section we will look at jitter mechanisms in clock signal recovery and attempt to quantify the level of jitter in practical interface decoders.

3.1 Interface Bandwidth Limitation

Consider the subframe carrying an audio word value of 255 in Fig. 6. In the upper diagram the unfiltered subframe represents the signal transmitted while in the lower diagram the received signal at the interface decoder has been filtered with an *RC* time constant of 200 ns. If we define zero crossing time t_x as the time taken for the received interface signal to cross the 0 V detection axis after a transition has occurred at the transmitter, then t_x depends upon the voltage at the receiver at transmitter transition time and inter-symbol interference occurs, *i.e.* zero crossing time depends on the values of previous pulse widths. This phenomenon is shown more clearly with the expanded time scale of Fig. 7; the zero crossing time at the end of cell 4 is smaller than that at the end of cell 6 (where the voltage at the receiver has had time to fall to a lower value). When both transmitted and received signals are known we can compute the change in zero crossing times at each subframe cell boundary by searching for the change in polarity of the filtered signal, and a simple computer program was written to perform this task using the filtered data of Fig. 6(b). The results shown in Fig. 8 indicate that the zero crossing time variation across the filtered subframe is about 50 ns in this example; when a series of 1's are transmitted the peak voltage received at the end of the interface falls resulting in a reduction in zero crossing time. The variation in zero crossing time results in a modulation of edge timing in the clock recovered from the interface signal, and this edge modulation is clearly dependent upon the number of ones and zeros transmitted in each subframe; *i.e. instantaneous recovered clock jitter is dependent upon the audio word value transmitted over the interface*.

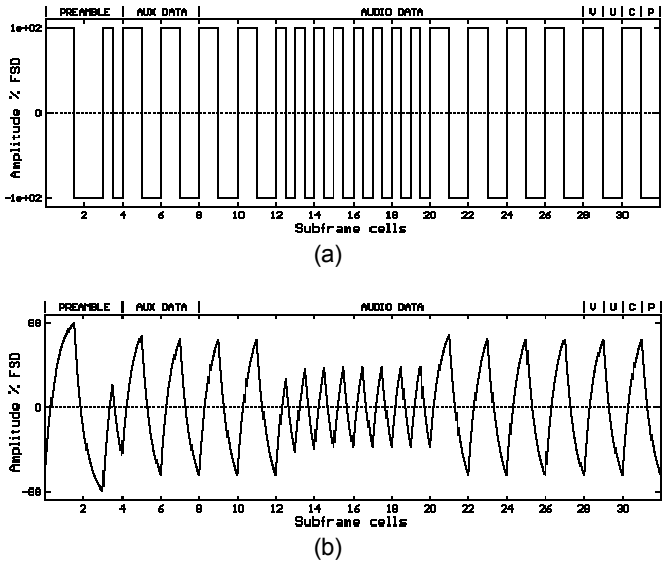


Fig. 6. (a) Unfiltered subframe signal carrying audio word value of 255. (b) As (a) but filtered with time constant of 200 ns.

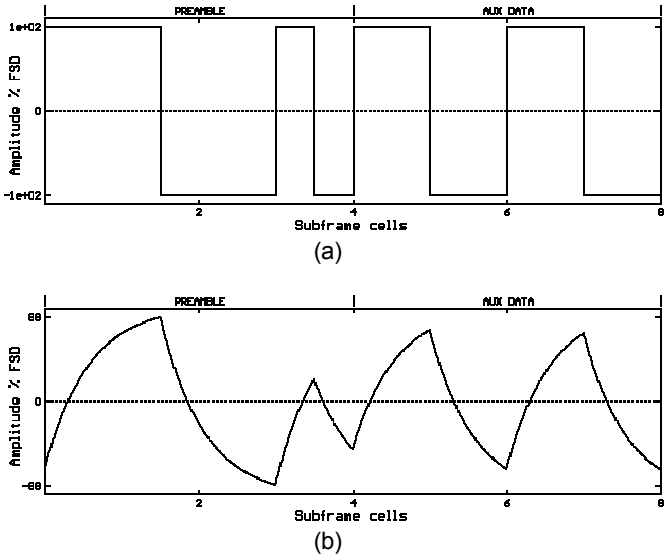


Fig. 7. Expanded view of Fig. 6 (b) showing the first 8 cells of the subframe.

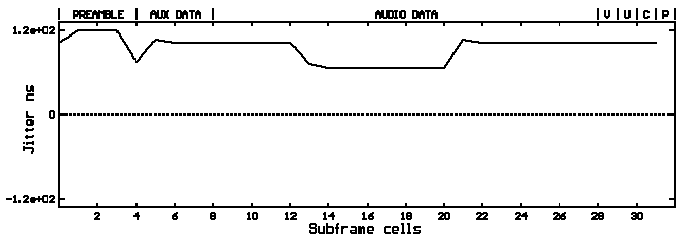


Fig. 8. Variation in cell edge zero crossing times across the subframe signal shown in Fig. 6 (b).

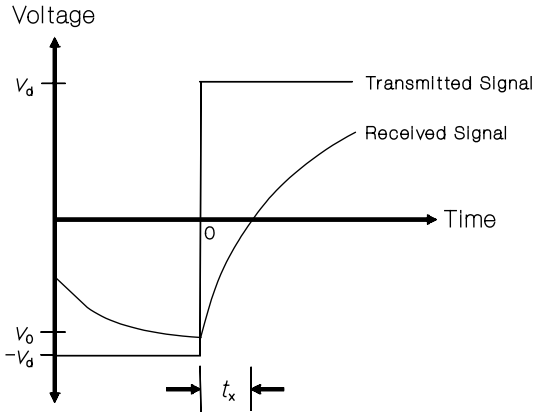


Fig. 9. Exponential rise of filtered interface signal following transmitter transition.

We will now develop an expression for the detected zero crossing time at a given transition with a known history of previous transition times. Consider Fig. 9 which shows the exponential rise of a filtered transition with time constant RC. The transmitted signal has a peak `driving' voltage V_d while the received voltage has an initial value V_0 at the time of the transition. If we denote the transition time as 0 s then the behaviour of the received signal following the transition can be described by a simple exponential time equation:

$$\begin{aligned} V &= V_d - (V_d - V_0)e^{-\frac{t}{RC}} \\ &= V_d\left(1 - e^{-\frac{t}{RC}}\right) + V_0\, e^{-\frac{t}{RC}} \quad . \end{aligned} \tag{2}$$

Setting V to zero and rearranging gives the zero crossing time:

$$t_x = RC \ln\left[1 + \left|\frac{V_0}{V_d}\right|\right] \quad . \tag{3}$$

The zero crossing time evidently has a dependency upon the initial voltage V_0 at transition time, and this in turn will depend upon previous pulse widths. If the previous transition in the interface signal occurred at $-t_1$ seconds then V_0 can be written in terms of t_1 and V_1 (the voltage at the previous transition):

$$V_0 = -V_d + (V_d + V_1)e^{-\frac{t_1}{RC}} \quad . \tag{4}$$

Substituting into Eq. (2):

$$V = V_d\left(1 - 2\,e^{-\frac{t}{RC}} + e^{-\frac{(t+t_1)}{RC}}\right) + V_1\, e^{-\frac{(t+t_1)}{RC}} \quad . \tag{5}$$

This process can be continued with the next transition time at $-(t_1+t_2)$ seconds, the next at $-(t_1+t_2+t_3)$ etc., to give:

$$V = V_d \left[1 - 2e^{-\frac{t}{RC}} \left(1 - e^{-\frac{t_1}{RC}} + e^{-\frac{(t_1+t_2)}{RC}} - e^{-\frac{(t_1+t_2+t_3)}{RC}} + \ldots \right) \right] . \quad (6)$$

Hence we can write the zero crossing time t_x for the transition at 0 s in terms of the previous pulse widths t_1, t_2, t_3, ... :

$$t_x = RC \ln \left[2 \left(1 - e^{-\frac{t_1}{RC}} + e^{-\frac{(t_1+t_2)}{RC}} - e^{-\frac{(t_1+t_2+t_3)}{RC}} + \ldots \right) \right] . \quad (7)$$

Using Eq. (7) we can now compute the zero crossing time at each transition in a filtered interface signal containing several subframes; this yields a signal with a sampling rate equal to the maximum rate of interface transitions, *i.e.* 5.6 MHz. In our simulations we want to map the zero crossing times of the filtered interface signal to jitter on the recovered clock at the output of the ADIC. However, 5.6 MHz is too high a sampling rate with which to efficiently compute the effect of such a jitter signal upon any reasonable length of audio data. In order to reduce the jitter sampling rate to a useful value we take advantage of the following argument. In a practical interface receiver circuit the PLL will usually employ a loop filter with a break frequency less than 20 kHz. If we assume that the audio sampling rate is much greater than the PLL loop frequency, then the zero crossing time to jitter mapping operation can be performed by computing a running *average* of the zero crossing times across two subframes. Thus the jitter value associated with a pair of adjacent subframes can be written:

$$t_{jRC} = \frac{RC}{M} \sum_{m=0}^{M-1} \ln \left[2 \left(1 - e^{-\frac{t_{m1}}{RC}} + e^{-\frac{(t_{m1}+t_{m2})}{RC}} - e^{-\frac{(t_{m1}+t_{m2}+t_{m3})}{RC}} + \ldots \right) \right] , \quad (8)$$

where M is the number of transitions averaged across two subframes;
t_{m1}, t_{m2}, ... are the pulse widths prior to the mth transition.

Note how the effect of previous transitions upon the zero crossing time at a given transition diminishes as we move further away from the transition. This allows a further reduction in jitter computation time since we can limit the transition history taken into account without compromising accuracy.

Before going on to present the results of simulations which calculate the jitter on a band-limited interface, it is worthwhile introducing the idea of a *jitter transfer function*. Eq. (8) indicates that the instantaneous jitter value associated with each interface signal frame depends upon the time between transitions (*i.e.* pulse widths) during that frame. The pulse widths (*i.e.* either half or full cell width) in turn depend upon the position of 1's and 0's in the transmitted interface signal. Given that each audio sample value is represented by a unique bit pattern when biphase-mark encoded by the interface transmitter, it is possible to calculate a jitter value for each audio sample value transmitted across the band-limited interface and hence represent instantaneous interface jitter as a *deterministic* function of audio sample value. Fig. 10 shows such a mapping for 16 bit audio sample values in the range -32000 to 32000 in steps of 100 for an RC time constant of 100 ns. Here we have assumed that both channels carry the same audio value and the parity subcode bit is active; the 64 zero crossing times at each cell boundary across the frame are calculated and then averaged to obtain a jitter value for the audio sample value. Fig. 11 displays the same transfer function but with an expanded horizontal scale over the audio word range -400 to 400 in steps of 1.

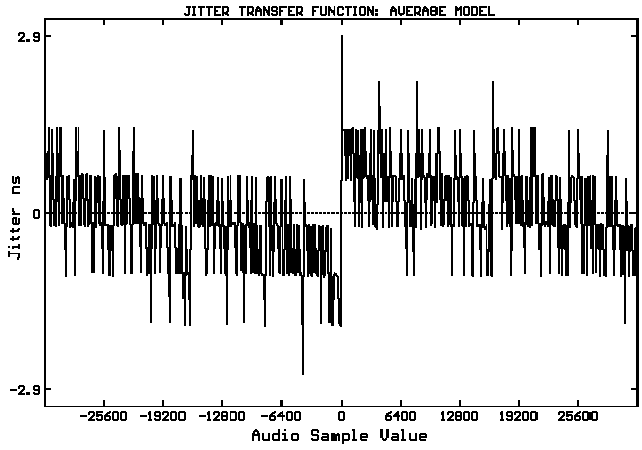


Fig. 10. Jitter transfer function over audio range -32000 to 32000 in steps of 100 for interface time constant of 100 ns.

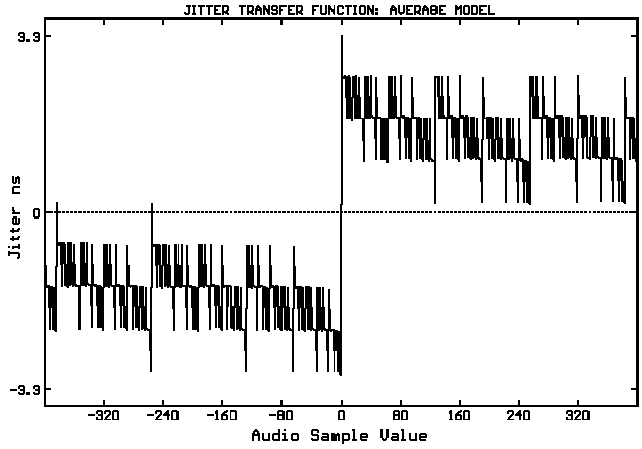


Fig. 11. Expanded view of central portion of Fig. 10.

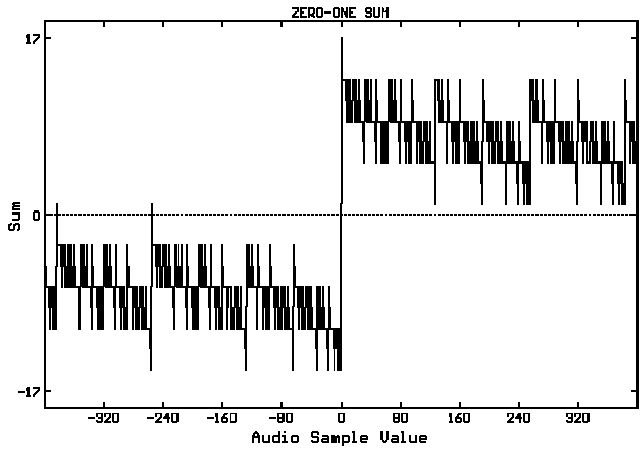


Fig. 12. Zero-one sum for two's complement coded 16 bit audio word (compare to Fig. 11).

The first point to note is that the largest change in zero crossing time occurs as the audio word value moves through zero, *i.e.* changes sign. This is due to the two's complement representation of the audio word within each subframe. The jitter exhibits a staircase-like transfer function as the audio word value moves away from zero; in fact the zero crossing time exhibits a strong dependence upon the difference between the number of 0's and 1's in the 16 bit audio word. Fig. 12 shows such a `zero-one sum' across the word range -400 to 400, with clear similarities to Fig. 11 (the sum extends to ±17 for the 16 bit audio word because the parity subcode bit is active). We can utilise this approximation to develop a simpler expression for interface jitter attributable to a given interface time constant.

Consider the case when we transmit only 0's across the interface (even in the preamble); t_1, t_2, t_3 etc. will all equal t_c and Eq. (7) reduces to:

$$t_{x0} = RC \ln\left[\frac{2}{1 + e^{-\frac{t_c}{RC}}}\right] . \qquad (9)$$

A similar expression occurs for the all 1's case where t_1, t_2, t_3 etc. = $t_c/2$:

$$t_{x1} = RC \ln\left[\frac{2}{1 + e^{-\frac{t_c}{2RC}}}\right] . \qquad (10)$$

Subtracting Eq. (9) from Eq. (10) yields the approximate peak jitter due to a band-limited interface:

$$t_{jRC} \approx \frac{RC}{2} \ln\left[\frac{1 + e^{-\frac{t_c}{2RC}}}{1 + e^{-\frac{t_c}{RC}}}\right] . \qquad (11)$$

This function is plotted for *RC* up to 400 ns in Fig. 13. We can use this approximation to scale the zero-one sum shown in Fig. 12 to obtain the approximate jitter transfer function of Fig. 14. Note how similar this plot is to Fig. 11 [calculated using the average jitter model of Eq. (8)] both in scale and shape, indicating the fundamental dependence of interface jitter upon the zero-one sum of the transmitted audio word.

Having developed expressions for jitter based upon the bit pattern across adjacent interface subframes, we are now in a position to simulate a band-limited interface transmission and examine the resultant jitter signal for sinusoidal audio data. Fig. 15(a) shows a dithered sinusoidal audio signal of 997 Hz and peak amplitude 0 dB, and Fig. 15(b) shows the corresponding jitter signal for an interface time constant of 100 ns. At first glance the jitter signal appears to be noise-like, but when it is lowpass filtered (simulating a hypothetical PLL filter) the jitter can be seen to be highly correlated with the audio (Fig. 16), despite choosing an audio frequency which maximises the number of PCM codes excited in the time domain [5]. This is confirmed by computing the Fourier transform of the windowed and filtered jitter signal (Fig. 17) revealing strong spectral lines at the fundamental and third harmonics (throughout this paper the 0 dB reference level in the jitter spectra is set to 1 ns peak jitter). A similar jitter spectrum computed for a -70 dB dithered 997 Hz audio signal is shown in Fig. 18, where even stronger 1st and 3rd harmonics are indicated. The increasing correlation between audio and jitter signals as the audio level is reduced is expected since the audio signal spends a proportionately longer time in the crossover region of the jitter

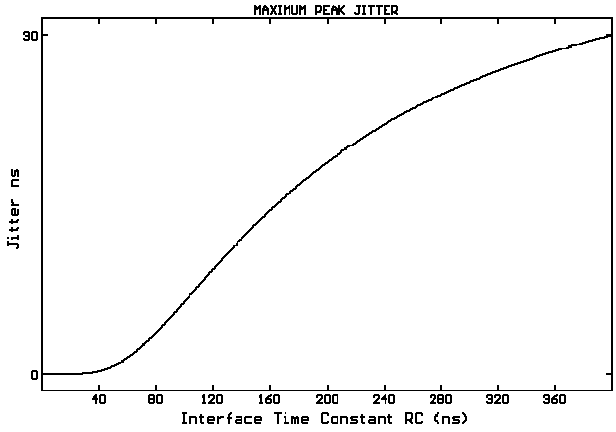


Fig. 13. Maximum peak jitter plotted against interface RC time constant.

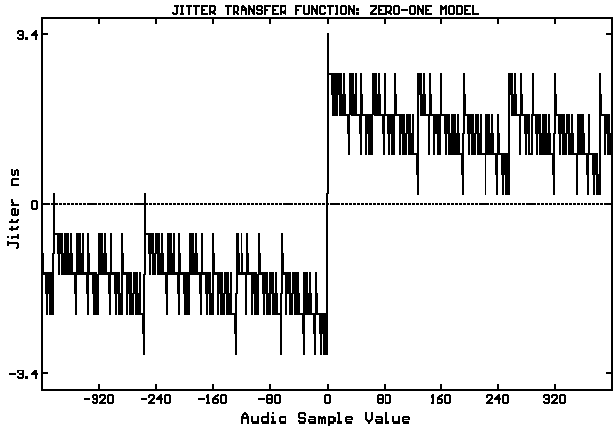


Fig. 14. Jitter transfer function calculated using zero-one model for RC = 100 ns.

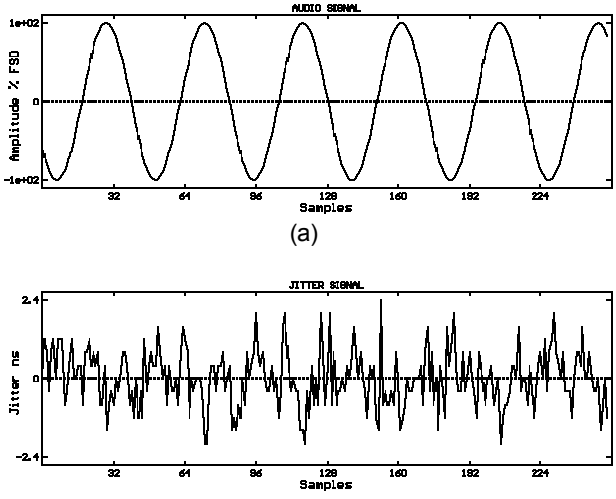


Fig. 15. 997 Hz audio signal and corresponding jitter signal for interface time constant of 100 ns and no PLL filtering.

transfer function. Meitner and Gendron [6] have also found that the jitter spectrum in a decoded interface signal has a strong dependency upon audio level but account for this behaviour in terms of power supply artifacts or `logic induced modulation'. In truth, power supply related jitter in an interface decoder will show similar characteristics to jitter due to band-limitation, though the results presented below suggest that the band-limitation model compares well to jitter measured in practical circuits.

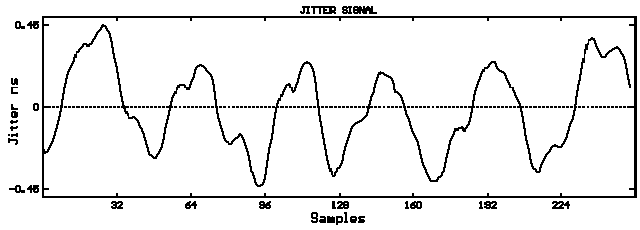


Fig. 16. Jitter signal of Fig. 15 filtered using 2nd order Butterworth lowpass filter with 1 kHz cutoff frequency.

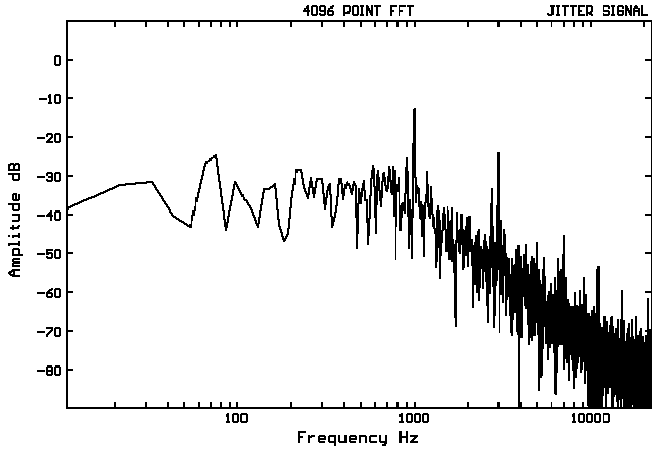


Fig. 17. Spectrum of filtered jitter signal shown in Fig. 16.

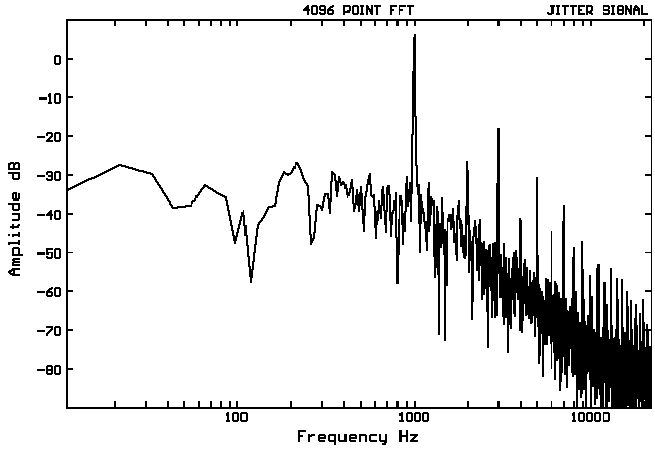


Fig. 18. Jitter spectrum corresponding to -70 dB 997 Hz sinusoidal audio signal. Note stronger jitter components at 1 and 3 kHz.

3.1.1 Comparison of Measured Results with Simulations

In this section we will compare the results of simulations using the techniques discussed above with *measured* results from the experimental interface receiver of Fig. 4. This circuit allows the *instantaneous frequency* of the recovered clock to be measured by monitoring the control voltage on the varicap diode. In order to recover the jitter signal on the clock we must convert the instantaneous frequency signal to a timing error. Consider a frequency deviation Δf in a recovered clock of nominal frequency f_o; over a time period dt the timing error t_j can be written:

$$t_j = \frac{\Delta f}{f_o} \mathrm{d}t \quad . \tag{12}$$

Now if the measured frequency deviation is sinusoidal then the corresponding jitter frequency is also sinusoidal and of the same frequency, and Δf can be written:

$$\Delta f = k V \sin\left(2\pi f_j t\right) \quad , \tag{13}$$

where k = PLL varicap control voltage sensitivity (Hz / V);
V = peak measured voltage on varicap control pin;
f_j = jitter frequency.

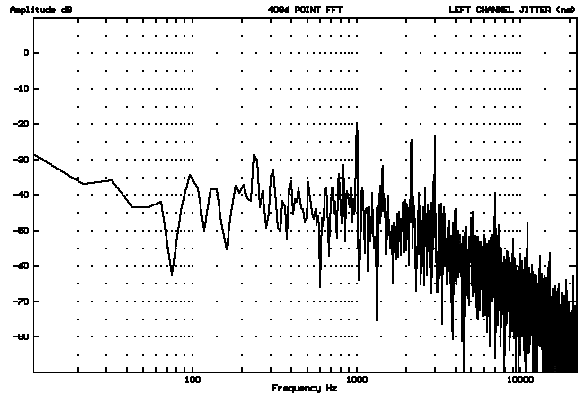
Integrating across one quarter cycle of f_j yields the peak jitter t_j at this frequency:

$$\begin{aligned} t_j &= \int_0^{\frac{1}{4f_j}} \frac{\Delta f}{f_o} \mathrm{d}t \\ &= \frac{kV}{2\pi f_o f_j} \quad . \end{aligned} \tag{14}$$

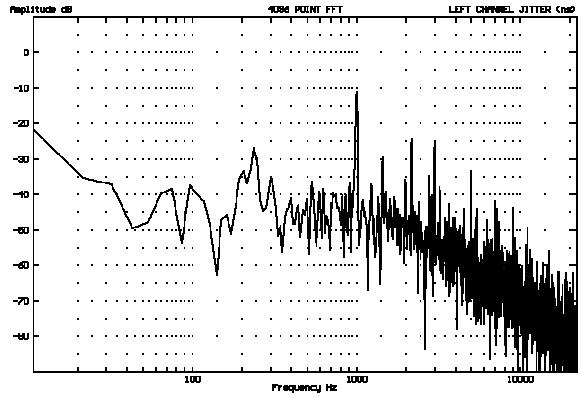
The varicap sensitivity k can be determined by decoupling the PLL feedback loop and monitoring clock frequency under direct control of a DC voltage applied to the control pin. Hence by measuring the control voltage on the varicap using an analog to digital converter (ADC) and scaling and integrating this signal to follow the law of Eq. (14) we have an indirect measure of interface jitter. (At this point it is worth noting that although the control voltage in the PLL provides a useful indicator of interface jitter, it will not reveal jitter in the PLL VCO itself above the closed loop cutoff frequency of the loop, or jitter sources between PLL and converter).

Figs. 19 show measured jitter spectra when the digital output from a CD player is connected to the experimental interface receiver. The interface signal carries the same audio data in both channels corresponding to a 1 kHz tone at (undithered) levels 0, -20, -40, -60 dB and (dithered) -80.65 dB respectively (these signals were obtained using the Hi-Fi News Test Disc 2). Each jitter spectrum consists of a large fundamental harmonic of the audio signal superimposed upon higher harmonics and noise which follow the frequency response of the PLL loop filter. Note the changes in the relative magnitudes of the harmonics as well as the absolute jitter level as the magnitude of the audio signal changes.

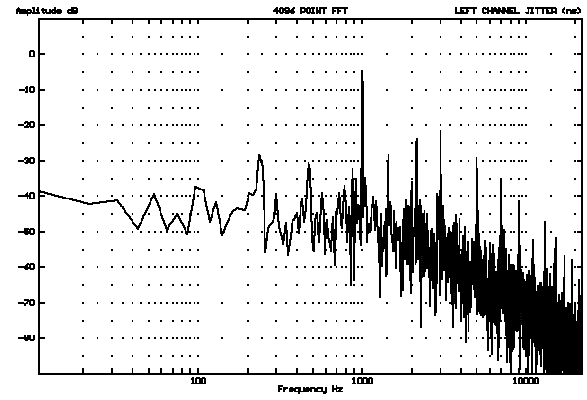
These measured results can be directly compared to *simulations* of the recovered clock jitter spectra for the same audio signals (Figs. 20). The simulations were obtained using the theory developed in Sec. 3.1 with a minor change; the average jitter model outlined above calculates the jitter based upon interface signal transitions at *every* cell edge, while the Philips SAA7274 ADIC employed in the experimental



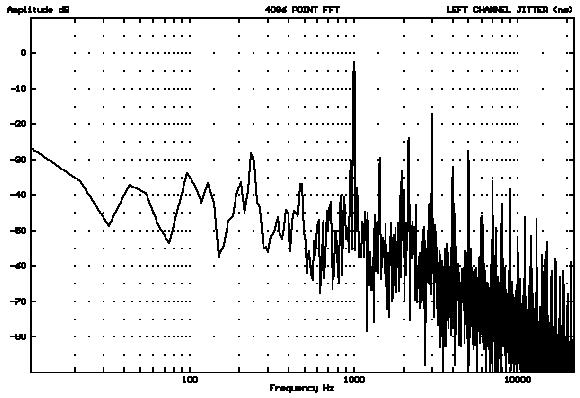
(a)



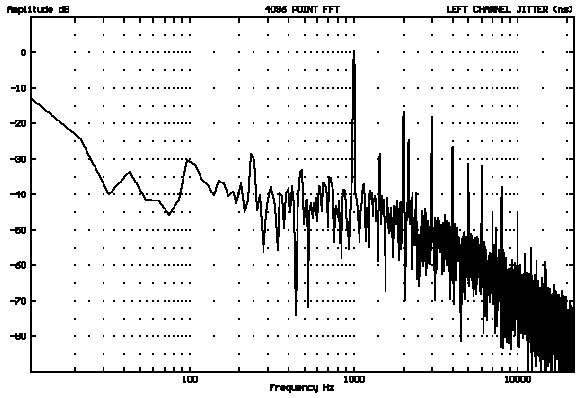
(b)



(c)

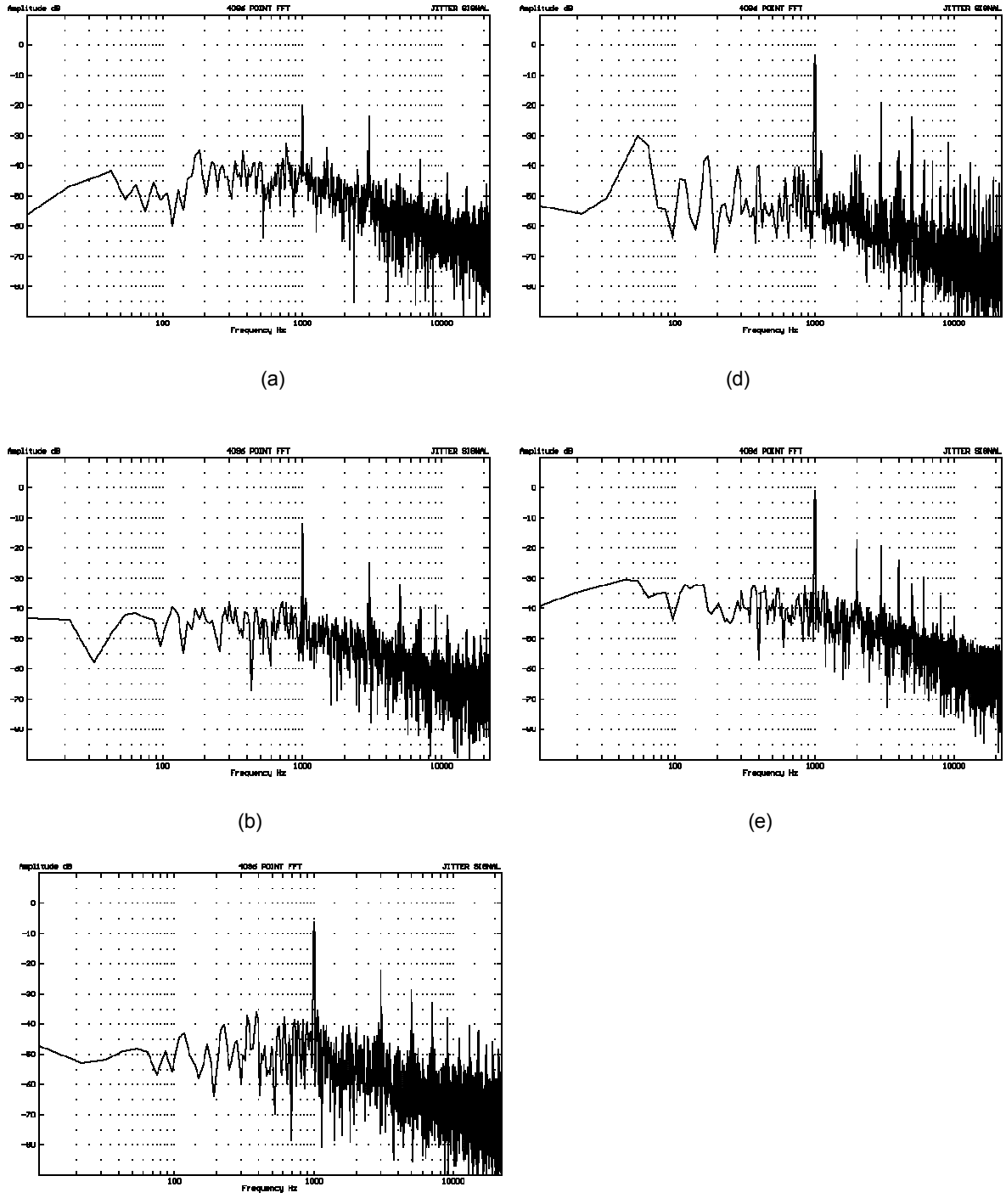


(d)



(e)

Fig. 19. Measured jitter spectra at interface receiver for 1 kHz audio signals of amplitude (a) 0 dB (b) -20 dB (c) -40 dB (d) -60 dB (e) -80.65 dB.



(a) (d)

(b) (e)

(c)

Fig. 20. Simulated interface jitter spectra for interface time constant of 65 ns and 1 kHz audio signal of amplitude (a) 0 dB (b) -20 dB (c) -40 dB (d) -60 dB (e) -80.65 dB.

receiver controls the PLL via timing errors at each *negative going* transition. The simulation software was adjusted accordingly and results obtained using an interface time constant of 65 ns and a 1st order PLL loop filter with break frequency set to 1 kHz. In general the simulations show good agreement with the measured results, the only discrepancies appearing as low level tones in the measured jitter spectra at 230 Hz, 1.6 kHz and 2.1 kHz. These tones are also present in a jitter measurement taken while no CD is playing (but the receiver is still locked to the digital interface signal) shown in Fig. 21. These frequencies are due to factors not taken into account in the software model; the first frequency can be accounted for by considering the change in the subframe preamble when a channel status block begins every 192 frames [3], while it is believed that the 1.6 and 2.1 kHz components are jitter artifacts inherent to the ADIC employed. Finally Fig. 22 shows the jitter noise floor of the measurement system, obtained with the input of the ADC connected to ground (note the expanded amplitude scale in this diagram). The low level of `jitter' in this measurement indicates that the measurement system employed has not compromised the accuracy of the results obtained above.

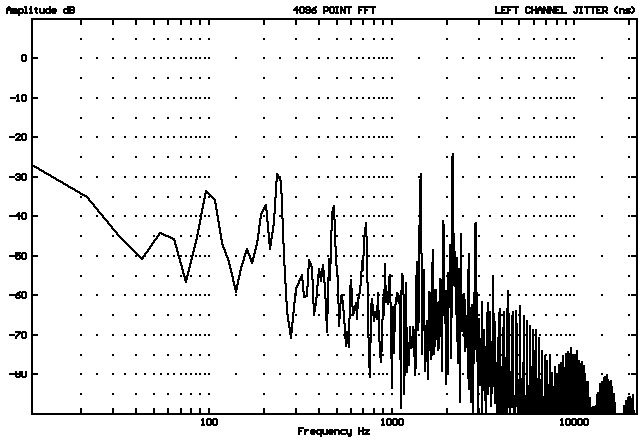


Fig. 21. Measured jitter spectrum with no CD playing.

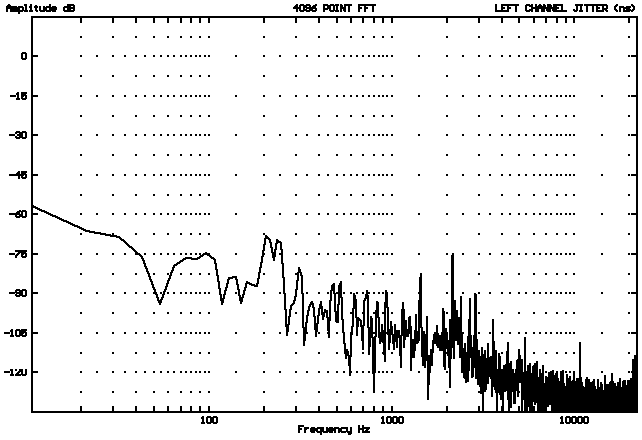


Fig. 22. Jitter measurement noise floor.

The good agreement achieved between practice and theory is important because it allows us to make predictions about the audibility of jitter errors in conversion electronics based upon the band-limited interface jitter model developed above. An important result to note from the measurements is the high degree of correlation that can occur between audio and jitter signals. Indeed it is instructive to audition the PLL control voltage in an interface receiver after suitable amplification; the audio signal transmitted over the interface can clearly be heard (albeit noisy and highly distorted) as the PLL attempts to track jitter on

the received interface signal (a phenomenon also reported by van Willenswaard [7]). The Authors' have found that Track 2 of the Hi-Fi News Test Disc 2 is particularly useful in this experiment since it includes sections where identical signals are recorded in- and out-of-phase across two stereo channels; the in-phase version is heard as a much louder signal at the PLL control voltage since the PCM 2's complement coded out-of-phase signal has a zero zero-one sum across one interface frame. A clock used in an AD or DA conversion process which suffers from jitter that is highly correlated with the audio signal can cause distortion tones which are audibly more objectionable than noise modulation which would otherwise occur for noise-like jitter (see Sec. 5).

The measured results for the experimental interface imply a time constant of about 65 ns. This is interesting in that the interface signal used in the experiments exhibited rise times in the order of 10 ns at the receiver input when observed using a high bandwidth oscilloscope. However, a band-limitation at any stage in the interface will cause jitter in the embedded clock signal, even if the interface signal transition edges are subsequently cleaned up. Every stage in the interface - transmitter, link and receiver - is a possible source of band-limitation and thus jitter. It is believed that an equivalent time constant of 65 ns within the ADIC IC was responsible for the observed jitter in the experiments.

3.2 Interface Noise

As well as increasing the (low) possibility of amplitude errors, interface noise can also be the cause of timing jitter in a band-limited interface. Consider a link with a time constant of RC where at the zero crossing points the rate of change of the received interface signal will be equal to V_d/RC (where V_d is the transmitter driving voltage). Thus peak interface noise of v_n results in a jitter noise of peak amplitude given by:

$$t_{jnoise} = RC\left|\frac{v_n}{V_d}\right| \quad . \tag{15}$$

Hence a peak interface noise 20 dB below the driving voltage and a time constant of 100 ns will result in 10 ns peak jitter due to the noise source. This is of the same order of magnitude of the jitter due to inter-symbol interference t_{jRC} given by Eq. (11) and plotted in Fig. 13. In practice jitter due to noise will be wideband and hence will be largely attenuated by the PLL filter at the receiver; this behaviour can be contrasted with jitter due to inter-symbol interference which will not be heavily attenuated for moderate PLL cutoff frequencies. This argument is given weight by examining Fig. 21 which shows the measured jitter spectrum in the experimental receiver for constant audio data (*i.e.* no inter-symbol interference). Here the jitter can be seen to be lower than for the measurements taken with CD audio tone tracks transmitted over the interface (Figs. 19) even though the same noise jitter mechanism is present in both circumstances. It should however be noted that the noise jitter mechanism can also cause problems with the recovered clock at the receiver output after the PLL, and all clock circuitry between the PLL and converter requires high-speed, low-noise characteristics.

3.3 Interface Slew Rate Imbalance

Another jitter mechanism that can cause problems with biphase-mark coded interface signals is due to asymmetrical slew rates across the link. If the receiver recovers the embedded interface clock by detecting transitions at every cell edge (as in our original detection model discussed in Sec. 3.1) then the difference between the number of positive going and negative going detection transitions across a frame will depend upon the number of 1's and 0's (and hence the audio word) transmitted. In the limit the jitter due to slew rate imbalance will equal the difference between the negative and positive going (slew limited) zero crossing times. Consider an interface signal with positive and negative going slew rates V_{SR+} and V_{SR-} respectively. The peak jitter due to slew rate imbalance will be:

$$t_{jSR} = \frac{|V_d|}{2} \left| \frac{1}{|V_{SR+}|} - \frac{1}{|V_{SR-}|} \right| \quad . \tag{16}$$

Thus for an interface signal detected using 74HC logic circuitry where V_d = 2.5 V and positive and negative going slew rates are 0.5 V/ns and 1 V/ns respectively, then Eq. (16) gives the peak jitter due to slew imbalance as 1.25 ns. Moreover, since the ratio of positive to negative going cell edges is dependent upon the transmitted audio word value then the jitter is likely to have strong components at audio frequencies.

The problem of slew rate imbalance interface jitter can be solved by only detecting interface transitions in one direction. This is the approach adopted in the SAA7274 ADIC used in the experimental receiver circuit, where monitoring the signal on pin 4 using an oscilloscope reveals only negative going interface transitions are used to recover the embedded clock.

4 JITTER ERROR MODELS

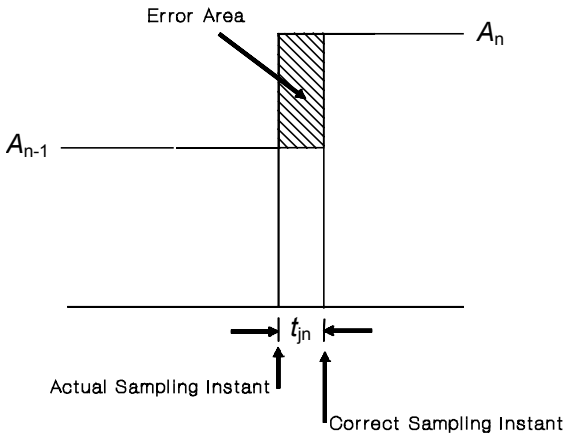
In order to assess the consequences of jitter in a DA conversion process we require a DAC jitter error model. Harris [8] has developed an analytical model for jitter in ADCs but the error mechanism in a DA process is different and depends upon the specific DAC architecture employed. In this section we present error models for two classes of DAC (essentially these models are the same as those presented in [9]). Firstly we consider a Nyquist-sampling DAC with a sample and hold unit where the individual sample values input to the DAC are held until the next sample arrives; we term this a `100% sample DAC'. Secondly an `impulsive' DA conversion process is considered where consecutive output samples do not overlap; this model approximates a pulse density modulation (PDM) conversion strategy as used by Philips in their oversampled and noise shaped `Bitstream' converters. The accuracy of the error models developed is established by comparing simulated results against actual measurements on physical DACs where jitter has been purposefully introduced into the interface. This is important since the error models are used in Sec. 5 to make predictions about the audibility of jitter artifacts.

4.1 100% Sample DAC

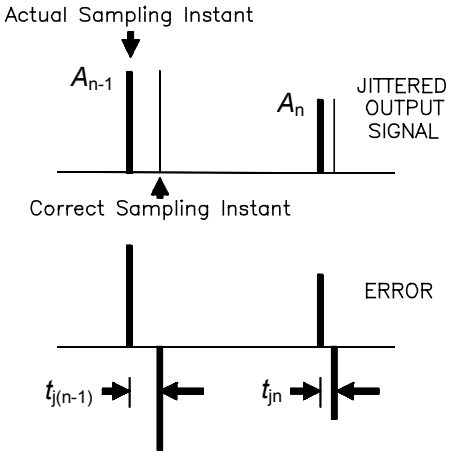
A 100% sample DAC holds the value of a given sample at its output until a new sample arrives. Thus timing error at the transition between adjacent samples results in a reconstructed analog signal with an `error area' directly proportional to the product of the sample timing jitter and the difference between the sample values [see Fig. 23(a)]. If we denote the normalised sample values as A_n and corresponding jitter values t_{jn} then, if the jitter amplitude is small compared to the sampling period t_s, we can form an error sequence e_n by scaling the error area by the sampling period:

$$e_n = [A_n - A_{n-1}] \frac{t_{jn}}{t_s} \quad . \tag{17}$$

This analysis is similar to the approach adopted by Blesser in examining slew limiting in DAC output stages [10]. The frequency domain representation $E_{100\%}(f)$ of this error sequence can be obtained by taking the discrete Fourier transform of e_n:



(a)



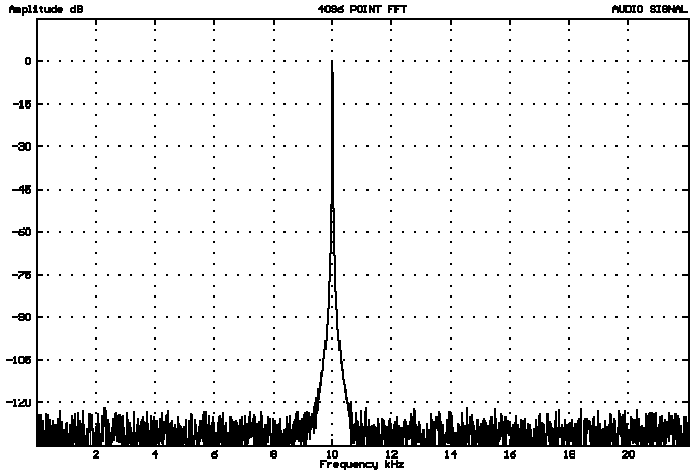
(b)

Fig. 23. DAC jitter error models. (a) 100% sample DAC. (b) Impulsive sample DAC.

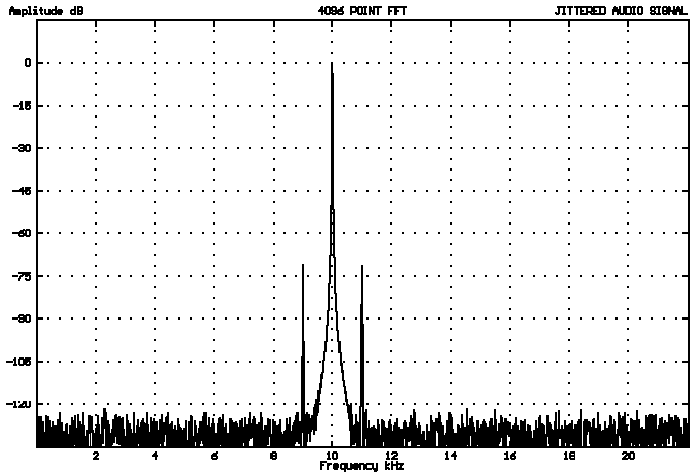
$$
\begin{aligned}
E_{100\%}(f) &= \mathrm{F}(e_n) \\
&= f_s \mathrm{F}\left([A_n - A_{n-1}]t_{jn}\right) \\
&= f_s \sum_{n=0}^{N-1} [A_n - A_{n-1}]t_{jn}\, e^{-i2\pi fnt_s}
\end{aligned}
\tag{18}
$$

where F is the discrete Fourier operator;
N is the number of samples forming the discrete error sequence;
f can take on values of kf_s/N where k is an integer;
i is the complex variable $\sqrt{(-1)}$.

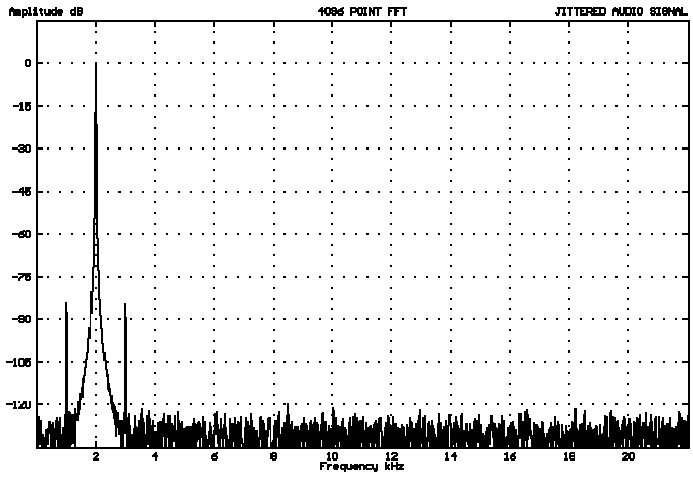
Eq. (18) indicates that the error spectrum for the 100% model should involve intermodulation components at signal/jitter sum and difference frequencies; this can be confirmed by simulating a jitter error signal using the model. Fig. 24(a) shows the *unjittered* spectrum of a 16 bit quantised 0 dB 10.001 kHz audio signal, where the choice of excitation frequency has led to a flat quantisation noise floor even though no dither has been used. Fig. 24(b) shows the same signal but now corrupted by a 1 kHz jitter signal of peak



(a)



(b)



(c)

Fig. 24. Simulated jitter error spectra for 100% DAC model. (a) 0 dB 10.001 kHz audio, no jitter. (b) 0 dB 10.001 kHz audio, 10 ns peak 1 kHz jitter. (c) 0 dB 2.001 kHz audio, 10 ns peak 1 kHz jitter.

amplitude 10 ns. As expected, the error components in the jittered spectrum occur as sidebands at 10 kHz ± 1 kHz at approximately -71 dB. The $[A_n\text{-}A_{n\text{-}1}]$ factor in Eq. (18) makes the magnitude of the error spectrum roughly proportional to the frequency as well as the amplitude of the audio signal. This is illustrated in the simulated error spectrum of Fig. 24(c) for a 0 dB 2.001 kHz sinusoid again jittered at 1 kHz; the sidebands are now approximately 14 dB lower than the 10 kHz example due to the fivefold drop in excitation frequency. Additional simulations (not shown here) have shown that the error spectrum for a 100% DAC is almost identical to that obtained from the ADC jitter model developed by Harris [8].

The accuracy of the error model can be verified by performing measurements with physical DACs where a known jitter signal is introduced into the digital interface signal; this technique has also been used by van Willenswaard [11]. Fig. 25 shows the test arrangement used. A CD transport digital output is connected to a custom digital interface unit which comprises a receiver circuit very similar to that shown in Fig. 4 and a transmitter. The receiver locks onto the incoming digital interface signal and passes the unscrambled data to the transmitter which performs a complementary function and outputs an interface signal to the DAC under examination. Now the edge timing on the output of the transmitter is controlled by the recovered clock in the receiver, hence applying signals to the control voltage on the receiver PLL allows direct injection of jitter into the digital interface signal. In the following tests we introduce sinusoidal jitter to the interface timing by connecting a signal generator to the PLL control voltage (it should be noted that the PLL control voltage governs instantaneous clock frequency and so we must follow the law of Eq. (14) in order to accurately predict injected jitter amplitude). The DAC output is then digitised using a 16 bit ADC *with an independent clock* and analysed for jitter related artifacts using a PC. Fig. 26(a) shows the output spectrum from the test DAC using a 0 dB 10 kHz CD test tone with no jitter introduced into the interface (the test unit in this example was the Musical Fidelity Digilog DAC of 16 bit 4 times oversampling design). The spectrum is quite pure with a 2nd harmonic at -96 dB relative to the fundamental (the line at 2 kHz is due to an idle tone in the ADC). Now consider the measured spectra shown in Fig. 26(b) where 1 kHz 10 ns peak jitter has been introduced into the digital data stream, causing sidebands at approximately -73 dB relative to the fundamental. A second measurement is presented in Fig. 26(c) using the same jitter signal but this time employing a CD test tone of 2 kHz. Both measurements show good agreement with the simulations in terms of both relative levels (within 3 dB) and spectral shape, and verify the accuracy of the 100% DAC jitter model. Note that this jitter injection method can also be used to assess the performance of the PLL filter employed in the interface receiver inside the DAC; if the jitter frequency is increased whilst maintaining constant amplitude then the amplitude of the error sidebands also remains constant until the break frequency of the loop is encountered. This technique was used to determine the PLL cutoff frequency of the Musical Fidelity DAC at approximately 5 kHz, implying that any jitter components below this frequency will be unattenuated by this particular unit and will contribute to jitter error at the DAC.

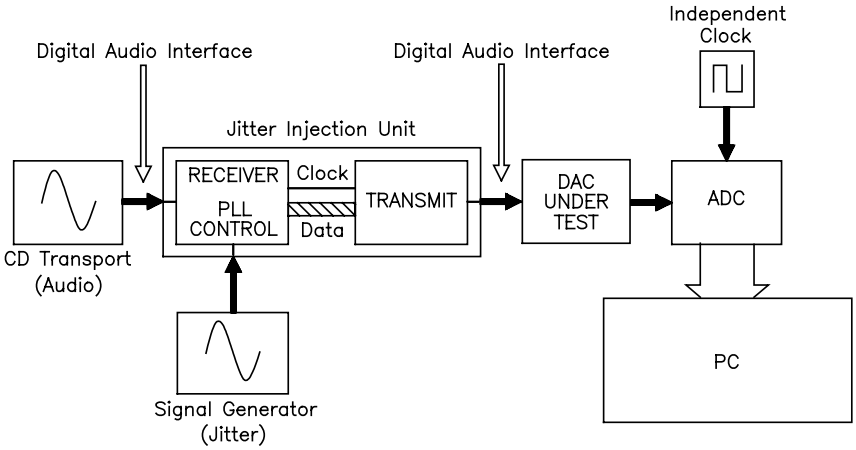
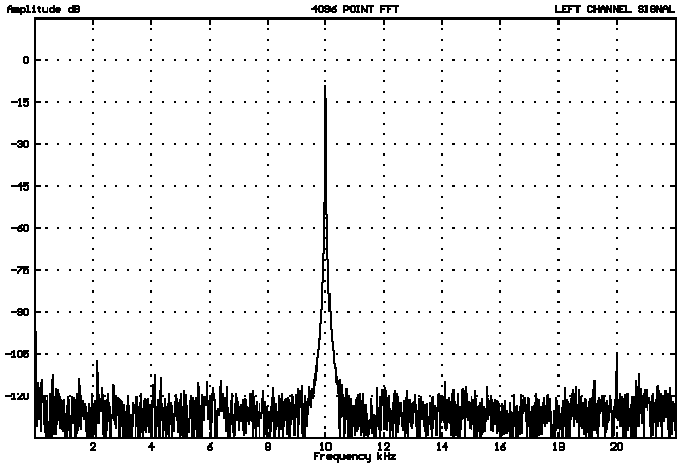
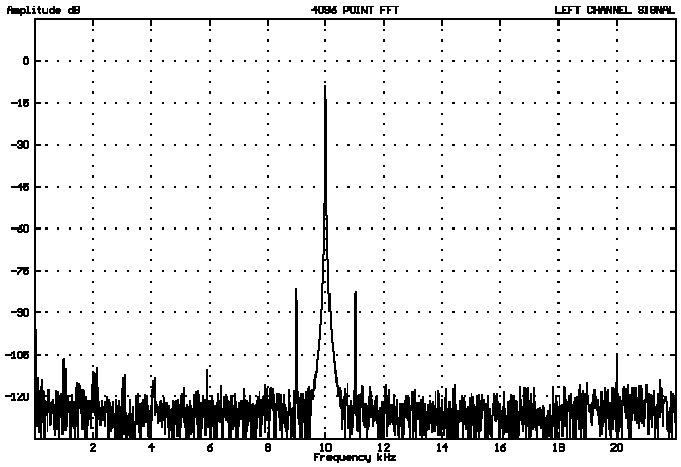


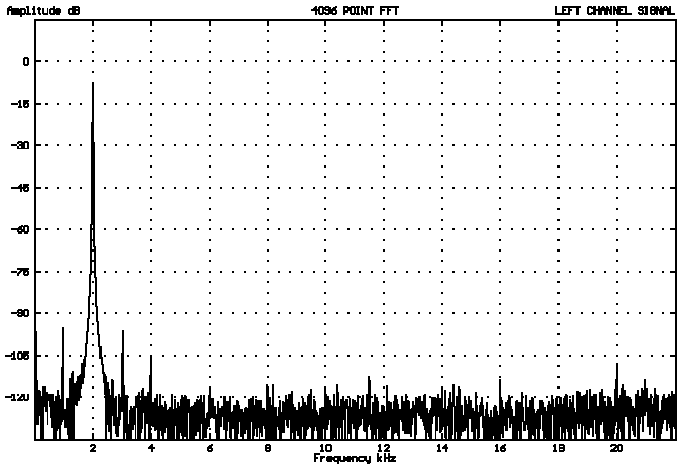
Fig. 25. Test set up for measuring DAC sensitivity to interface jitter.



(a)



(b)



(c)

Fig. 26. Measured jitter error spectra using Musical Fidelity Digilog 16 bit 4 times oversampling DAC. (a) 0 dB 10 kHz audio, no jitter. (b) 0 dB 10 kHz audio, 10 ns peak 1 kHz jitter. (c) 0 dB 2 kHz audio, 10 ns peak 1 kHz jitter.

4.2 Impulsive Sample DAC

We now progress to the impulsive DAC model where digital data samples are output as weighted impulses with no interaction between adjacent samples [Fig. 23(b)]. The output impulses are infinitesimally narrow and occur at a sampling rate f_s (which in our simulations is 44.1 kHz). Of course this is a theoretical construct and no practical DAC would be able to create such an output pulse train. Nevertheless, the impulsive model is a fair approximation to a heavily oversampled PDM DAC (e.g. Philips Bitstream) subject to a jitter process, where each 44.1 kHz audio sample is represented by many discrete output pulses each of unit amplitude and each subject to the same timing error. Referring to Fig. 23(b), the continuous error signal e_n due to *one* audio sample A_n jittered by t_{jn} can be written:

$$e_n = A_n \left[\delta(nt_s - t_{jn}) - \delta(nt_s)\right] \quad , \tag{19}$$

where $\delta(t)$ corresponds to a dirac impulse at time t.

Hence the discrete error spectra due to a sequence e_n is:

$$\begin{aligned} E_{imp}(f) &= \sum_{n=0}^{N-1} e_n \, e^{-i2\pi fnt_s} \\ &= \sum_{n=0}^{N-1} An\left[\delta(nt_s - t_{jn}) - \delta(nt_s)\right] e^{-i2\pi fnt_s} \\ &= \sum_{n=0}^{N-1} An\left[e^{i2\pi ft_{jn}} - 1\right] e^{-i2\pi fnt_s} \quad . \end{aligned} \tag{20}$$

Now if we assume that $t_{jn} << t_s$ then:

$$\begin{aligned} E_{imp}(f) &= i2\pi f \sum_{n=0}^{N-1} A_n \, t_{jn} \, e^{-\frac{i2\pi fn}{f_s}} \\ &= i2\pi f \, F\left(A_n \, t_{jn}\right) \quad . \end{aligned} \tag{21}$$

Thus the error spectra is simply the Fourier transform of the product sequence $A_n t_{jn}$ scaled to rise with frequency. Eq. (21) can be compared to the error spectra obtained for the 100% DAC model [Eq. (18)]; intermodulation between signal and jitter frequencies is still expected in the impulsive error spectra but the intermodulation components will have different weights than those obtained from the 100% model. This is best illustrated by comparing error spectra for the two models when a 1 kHz tone is jittered by wideband white noise (Fig. 27). The error spectra for the 100% model is essentially flat with frequency while the impulsive error rises with frequency at 6 dB/octave (the exact frequency at which the two curves intersect depends upon the audio signal frequency and the statistical properties of the jitter noise).

The accuracy of the impulsive DAC jitter error model can be confirmed by comparing simulated (Figs. 28) and measured (Figs. 29) results for the same signals used in Sec. 4.1. The measured results were made using a Meridian 203 Bitstream DAC and clearly show good agreement (within 2 dB) with the simulations.

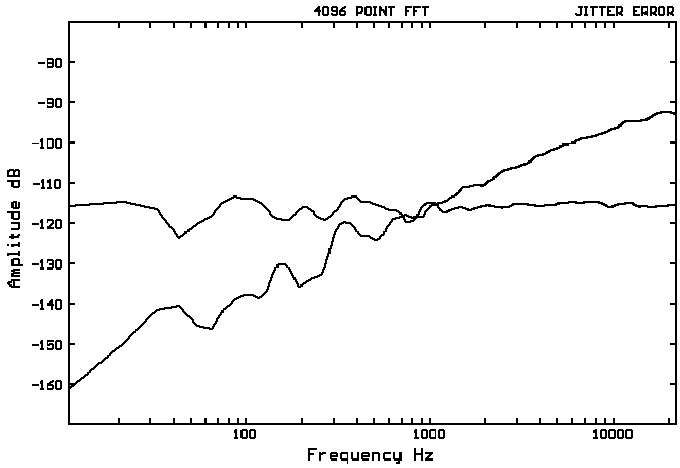


Fig. 27. Simulated third-octave smoothed error spectra for 1 kHz 0 dB tone jittered by 10 ns rms white noise; horizontal trace is 100% DAC model, upwardly sloping is impulsive model.

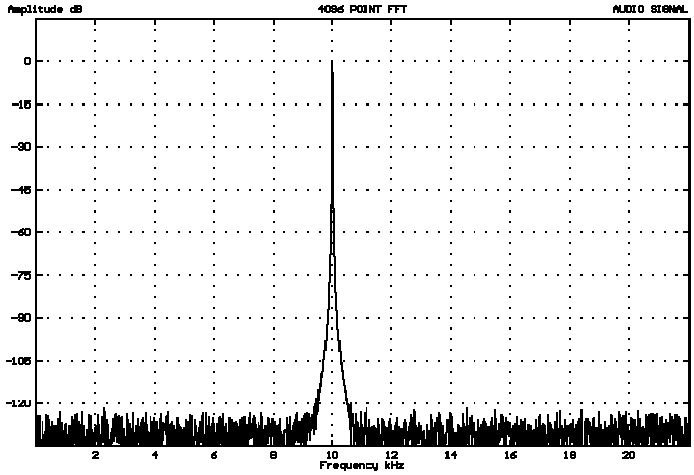
The jitter injection interface was also employed to determine the cutoff frequency of the Meridian PLL at higher than 10 kHz. The same measurement made on another Bitstream model, the Audiolab 8000 DAC, reveals no jitter error at all (Fig. 30); this is due to the superb PLL performance of this unit, offering a (claimed) closed loop cutoff frequency of 13 Hz where all audio frequency interface jitter will be attenuated to inaudible levels (note that the noise seen at 400 Hz in this diagram is due to a ground loop problem in the test apparatus).

The greatest differences between 100% and impulsive DAC jitter errors occur when the signal frequency is low. Figs. 31 and 32 show simulations and measurements respectively for a 0 dB 100 Hz audio signal jittered by 10 ns 4 kHz sinusoidal jitter. The jitter error is about 30 dB higher in the impulsive (Bitstream) DAC. Thus audible noise modulation can occur when impulsive DACs are jittered while reproducing low frequency audio signals, and this may well be the reason why some reports have suggested that Bitstream DACs lack in `dynamics' and `rhythm-and-pace' [12] since most of the rhythmic content in music occurs at low frequencies and can be of high peak level. In addition, practical low bit DACs often possess a high level of high frequency quantisation noise which has been shaped away from the audio band, and if very high frequency jitter is introduced into such a conversion process then intermodulation products can fall back down into the audio band causing a further degradation of dynamic range and noise modulation [13] (our simulations employ a sampling rate of 44.1 kHz and thus cannot model such secondary jitter artifacts).

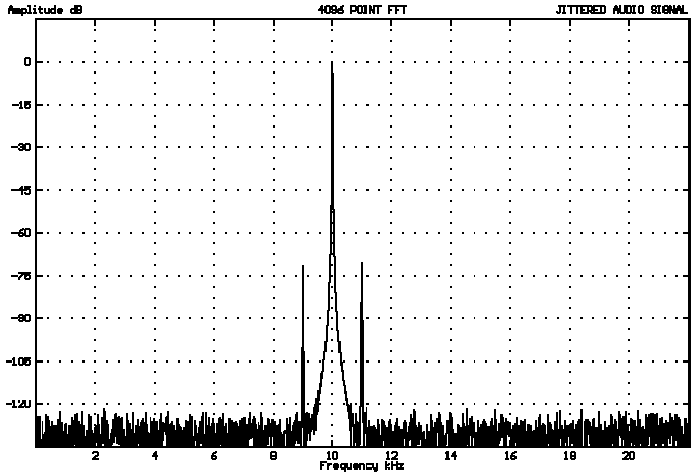
4.3 Analogy to Phase Intermodulation Distortion in Audio Amplifiers

Amplitude errors caused by timing jitter at the ADC or DAC gateway can be examined in a wider perspective by comparing the jitter error mechanism with artifacts found in analog electronics. Hawksford [9] has shown that jitter errors in DACs can be compared to slew rate limiting in transimpedance amplifiers located at DAC outputs. The jitter error mechanism can also be likened to phase intermodulation distortion or PID in analog amplifiers. Otala [14], [15] has shown that PID occurs when open loop amplitude non-linearity in a feedback amplifier is mapped to a closed loop phase non-linearity. Cordell [16] refined the PID model by writing the timing error $t(x)$ in the phase distorted output voltage x from a feedback amplifier in terms of the normalised open loop non-linearity $e(x)$ and closed loop cutoff frequency f_{co};

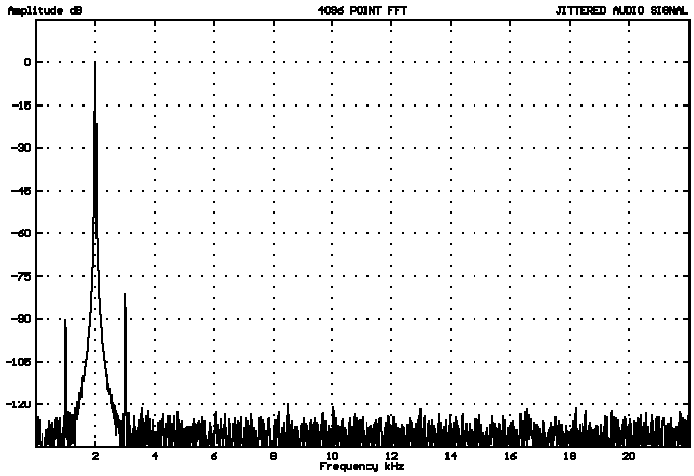
$$t(x) = \frac{e(x)}{2\pi f_{co}} \quad . \tag{22}$$



(a)

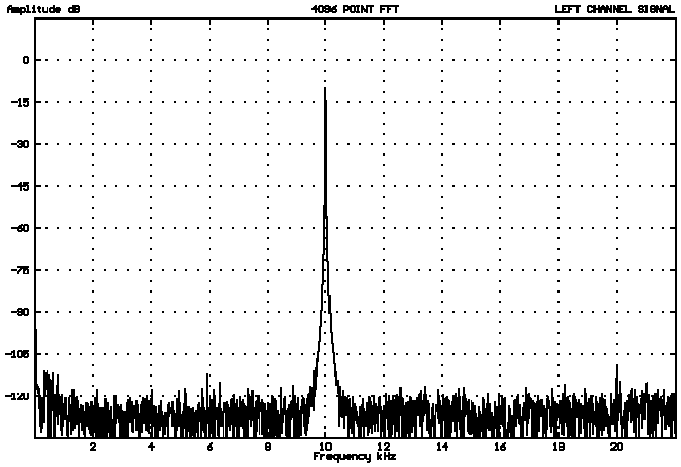


(b)

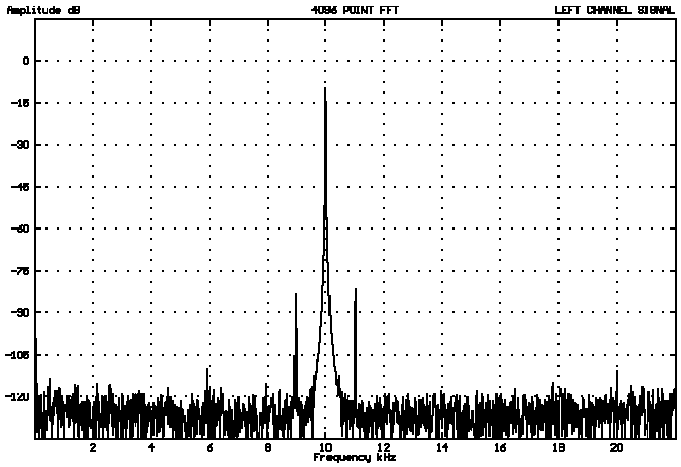


(c)

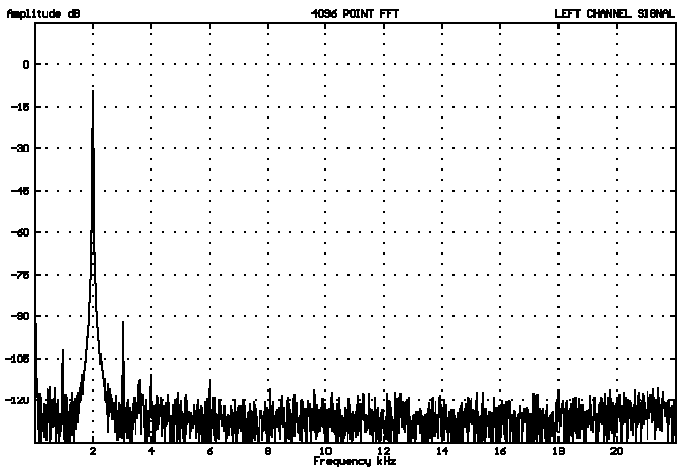
Fig. 28. Simulated jitter error spectra for impulsive DAC model. (a) 0 dB 10.001 kHz audio, no jitter. (b) 0 dB 10.001 kHz audio, 10 ns peak 1 kHz jitter. (c) 0 dB 2.001 kHz audio, 10 ns peak 1 kHz jitter.



(a)



(b)



(c)

Fig. 29. Measured jitter error spectra using Meridian 203 Bitstream DAC. (a) 0 dB 10 kHz audio, no jitter. (b) 0 dB 10 kHz audio, 10 ns peak 1 kHz jitter. (c) 0 dB 2 kHz audio, 10 ns peak 1 kHz jitter.

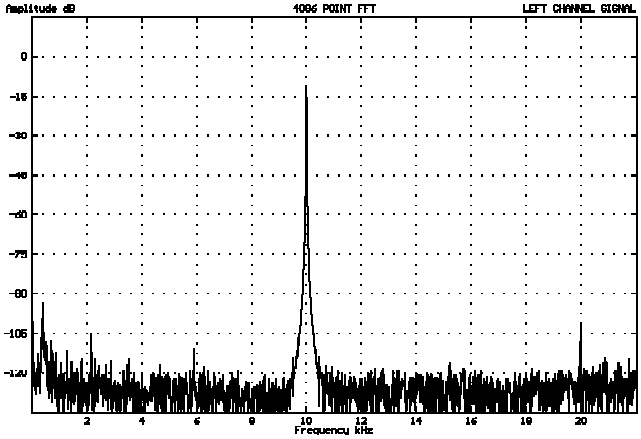


Fig. 30. Measured jitter error spectrum using Audiolab 8000 DAC (0 dB 10 kHz audio, 10 ns peak 1 kHz jitter) showing complete absence of error sidebands at 9 kHz and 11 kHz.

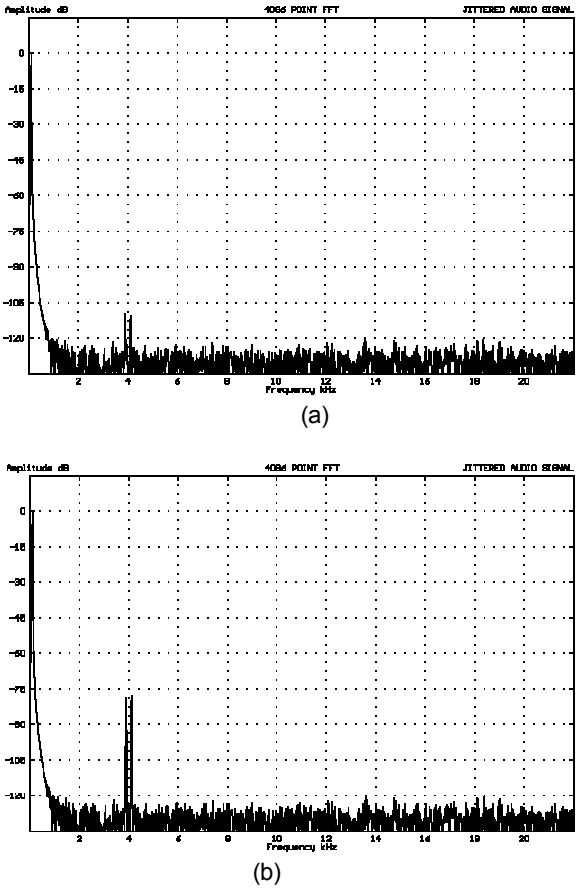


Fig. 31. Simulated jitter error spectra for 0 dB 100.01 Hz audio, 10 ns peak 4 kHz jitter. (a) 100% DAC model. (b) Impulsive model.

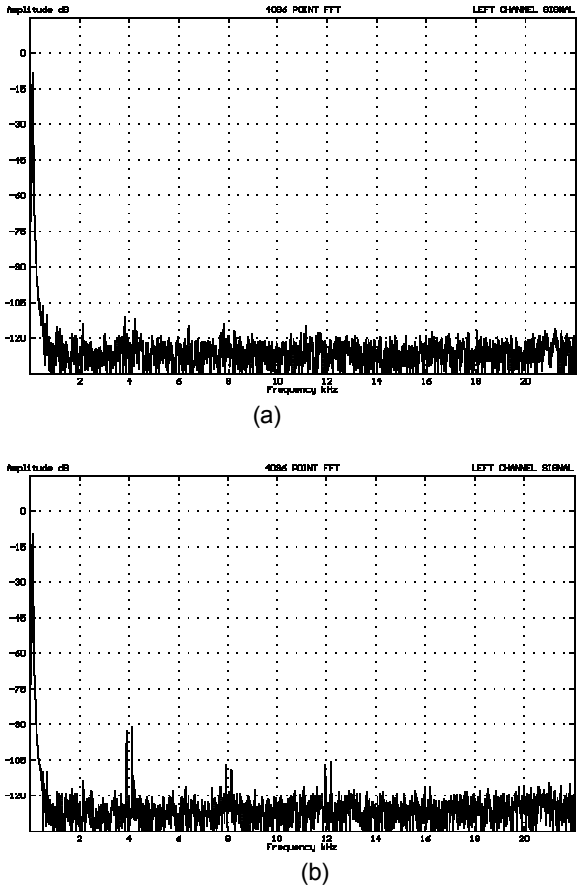


Fig. 32. Measured jitter error spectra for 0 dB 100 Hz audio, 10 ns peak 4 kHz jitter. (a) Musical Fidelity 16 bit 4 times oversampling DAC. (b) Meridian 203 Bitstream DAC.

If we substitute typical values of $e(x_{max})$ = 1% (0.01) and f_{co} = 1 MHz for an IC operational amplifier then we find that the peak timing error is equal to 1.6 ns. We have seen that this is of the same order of magnitude as the jitter found in digital audio interfaces. It is also of interest to note that a typical analog transfer function non-linearity will result in PID where the timing error is highly correlated with the audio signal. This observation lends weight to the analogy made between PID and digital audio interface jitter, where we have shown that the jitter resulting from a band-limited digital interface can also show a high degree of correlation with the transmitted audio signal. Of course PID in analogue amplifiers is intimately linked with closed loop amplitude non-linearity and is *not* in itself a cause of additional error, while jitter in digital audio interfaces *is* a source of error in AD or DA conversion. Nevertheless the analogy between PID timing error and interface jitter is useful if the overall timing error in a system sense is to be minimised; there is little point in minimising digital audio interface jitter if the analogue circuitry preceding or following conversion is of poor quality. In general it is rewarding to make comparisons between analog and digital system artifacts, an exercise which has shown interesting results before [17], [18]. In truth the boundaries between analogue and digital systems are not as clear cut as they may at first appear, and we encourage system appraisal in a global sense.

5 AUDIBILITY OF JITTER ERRORS

How much jitter is tolerable in a reconstructed stereo PCM transmission? One of the earliest studies of the audible consequences of jitter is due to Manson [19] who carried out a series of tests using a *monophonic* analogue signal input to a sampling device with a stable clock. The sampled (but unquantised) audio signal is then converted back to the analogue domain using a sample and hold unit with a controlled amount of clock jitter and auditioned. Manson suggests 35 ns rms jitter represents the threshold of subjective impairment using critical source material. However, we believe that several factors make this an unrealistically high figure of `minimum audible jitter'. Firstly, Manson's experiments were carried out using a monophonic test signal. There is much evidence to suggest that the audibility of low level distortion (as would be caused by jitter) increases when music is reproduced in stereo, as acoustic objects are now perceived in two-dimensional space and masking of errors by the primary signal is not so effective. Secondly, these tests were carried out some time ago using tape recordings, and the advances in reproduction equipment now available to the consumer should result in a lower threshold. A better estimate of the audible jitter threshold can be obtained by examining the jitter error sequence and assuming that it will be inaudible if below the level of quantisation noise present in the system for any possible excitation frequency. Lidbetter [20] thus arrives at a value of 120 ps for a 16 bit 100% sample DAC and an incredibly low 8 ps for 20 bit system. Shelton [21], Fourre [22], Harris [8], van Willenswaard [7] and the recommendations embodied in AES11-1991 [23] all quote similar values. Are these lower limits reasonable - *i.e.* how audible is jitter when its level is at the quantisation noise floor ? An attractive approach to answering this question is to model the hearing process itself, *i.e.* find whether a given jitter error is below the masked threshold due to the jittered audio signal, a method adopted by Julian Dunn [24]. Correspondingly a simple hearing model was developed in order to assess the audibility of jitter in a band-limited interface.

Our model assumes that the error due to jitter is inaudible if it is below the threshold of hearing (*i.e.* minimally audible field) at all frequencies. This approach will yield pessimistic results as far as error audibility is concerned since the additional masking effect of signal tones is not considered, although it should be remembered that masking of low frequency noise by high frequency tones is minimal and recent work by Stuart [25] suggests that the audibility of errors in isolation may well be of higher significance than has previously been thought. We define the threshold of hearing in the frequency domain by passing a cubic spline through the threshold data of ISO226 [26] and scaling by the gain of a typical audio system under critical listening conditions such that `0 dB' refers to a sound pressure level of 112 dB at 1m per speaker [27]. The error signal is then integrated at each frequency bin across a bandwidth defined by the equivalent rectangular noise bandwidth [27] at that frequency and compared to the threshold. Fig. 33 indicates the audibility of 16 bit triangular probability density (TPD) flat dither noise

assessed in this manner, clearly showing the dither to be audible in the frequency range 2 kHz to 6 kHz - this is similar to the result achieved by Stuart in Fig. 2 of [28].

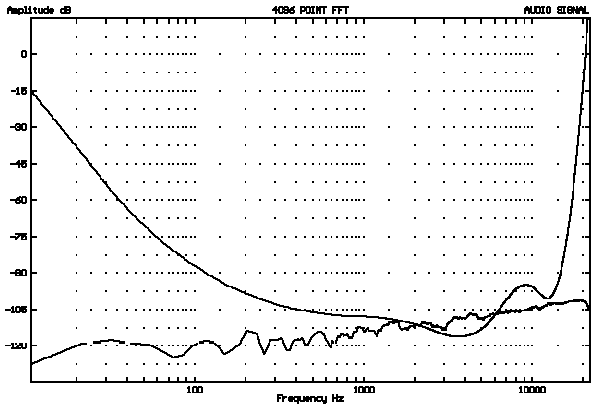
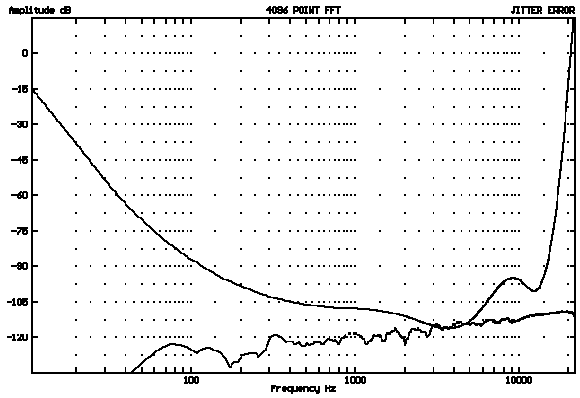


Fig. 33. Audibility of spectrally flat TPD 16 bit dither.

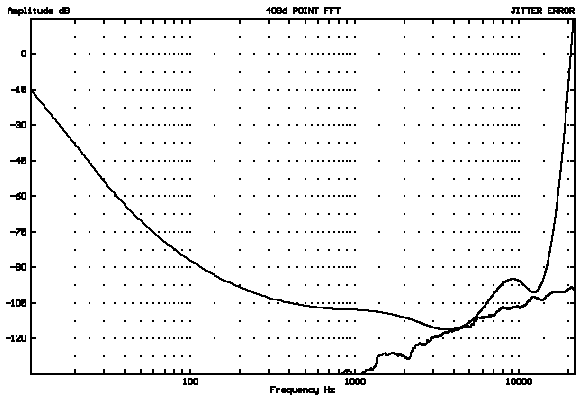
We can use the error audibility model to assess the validity of these claimed limits to jitter audibility. Consider the case where an audio tone is corrupted by spectrally white jitter. Fig. 34(a) shows that for a 100% DAC reproducing a 0 dB 20 kHz sinusoid, RPD jitter of peak amplitude 180 ps should be on the threshold of audibility, although it should be noted that the error level reduces as the audio sinewave frequency is reduced. This can be compared to Fig. 34(b) for an impulsive DAC where 550 ns peak jitter noise can be tolerated, but note that the error curve is *constant* with audio frequency (this diagram was obtained using 100 Hz). The problem with making predictions about the audibility of jitter artifacts using noise-like jitter is that the error tends to be spread across the audio band; more stringent jitter specifications are required when the jitter is sinusoidal. Fig. 35(a) shows the worst case 100% DAC jitter error resulting from a 22 kHz audio signal and 18.5 kHz jitter - only 20 ps peak jitter is required for audibility. The 75 ps limit for an impulsive DAC occurs when reproducing a low frequency tone and both jitter sidebands are coincident due to reflection about DC [Fig. 35(b)]. Are there any circumstances under which these critical combinations of audio and jitter signals could occur at the same time? In Sec. 3 we have shown that digital audio interface jitter can be highly correlated with the transmitted audio data and, when it is remembered that digital filters with delays up to a few ms often follow interface receivers before DA conversion (causing the jitter to precede the associated audio signal at the DAC), such combinations may indeed occur.

We now progress to an examination of simulated DAC errors due to band-limited interface jitter. Fig. 36(a) shows the jitter error spectrum for a 100% DAC reproducing a 0 dB 20 kHz audio signal with an interface RC time constant of 40 ns; the PLL cutoff frequency is set to infinity so that any jitter on the interface is unattenuated before reaching the DAC. The error contains discrete frequency components rising from the noise floor - this is because bandwidth limited interface jitter can contain components that are well correlated with the audio signal (see Sec. 3.2). Fig. 36(b) shows a similar simulation for an impulsive DAC fed a 0 dB 9.8 kHz audio signal; in this case an interface time constant of 50 ns results in an audible error. Of course most interface receivers will employ PLL filters such that interface jitter above the PLL cutoff frequency is attenuated. This results in the jitter error forming `skirts' around the audio signal; examples for 100% and impulsive DACs reproducing full scale 7 kHz sinewaves are shown in Figs. 37. The phase lock loop 2nd order filter cutoff has been set to 1 kHz with the result that the interface time constant can be increased to 120 ns before jitter audibility occurs. Finally Figs. 34 compare the jitter errors from 100% and impulsive DACs respectively whilst replicating a 200 Hz full-scale sinusoid from an interface with a 100 ns time constant (the PLL filter is set to 1 kHz 2nd order). The higher jitter error from the impulsive model whilst reproducing low frequency signals is clearly evident, an interesting result considering the claimed higher sensitivity of Bitstream DACs to digital interface quality compared to their multibit counterparts [29].

It is clear that the lower the bandwidth of the PLL, the more wideband jitter can be tolerated on the digital interface link. Our simple audibility model has not considered masking due to the audio signal itself, and when this is taken into account it is evident that reducing the PLL break frequency has the additional benefit of narrowing the jitter error skirt around the audio signal and hence making it less audible. There is a clear analogy here to data reduction systems where efficient coding of audio signals places the error directly under the audio signal where, because of masking, it is least audible [30]. The worst case for interface jitter audibility is when the receiver PLL has a high jitter bandwidth and it should be noted that several ADIC integrated circuits on the market have jitter bandwidths up to 5 kHz [31]. In the limit (no filter at all) white jitter noise of up to 180 ps peak amplitude and even less sinusoidal jitter can be tolerated, more or less justifying the recommendations in AES11-1991. Similarly we have seen that for no PLL filter the interface time constant must be lower than 40 ns or equivalently possess a bandwidth higher than 4 MHz, a requirement which makes the performance offered by Toslink-type optical interface links marginal [32], [33]. The revised professional interface standard AES3-1992 [34] stipulates a reasonable upper limit of 30 ns on the 10-90 % rise time for correctly terminated interface transmitters, corresponding to a time constant of 14 ns. However, we have seen that time constants need to be minimised in each part of the interface (transmitter, link and receiver), and *observing sharp rise and fall times at one point in the chain does not guarantee waveform fidelity elsewhere.* For example, our experimental receiver circuit suffers from a 65 ns time constant internal to the ADIC IC used.

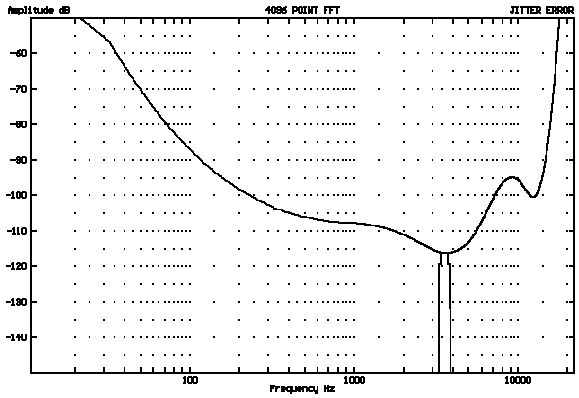


(a)

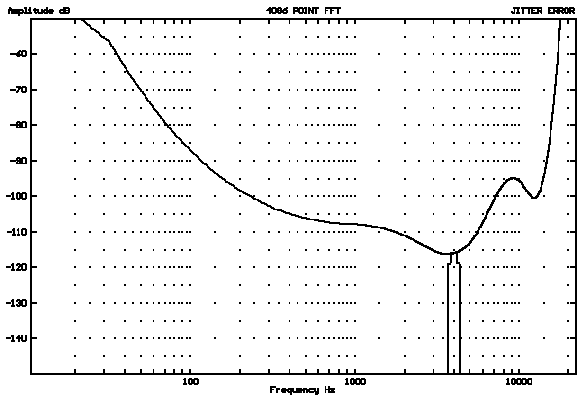


(b)

Fig. 34. Simulated jitter errors for RPD white jitter noise. (a) 0 dB 20 kHz audio, 180 ps peak jitter, 100% DAC. (b) 0 dB 100 Hz audio, 550 ps peak jitter, impulsive DAC.

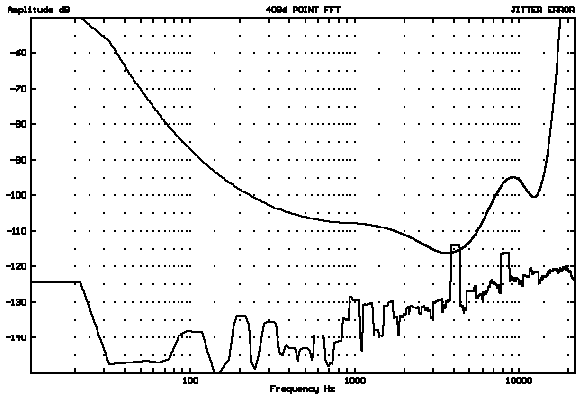


(a)

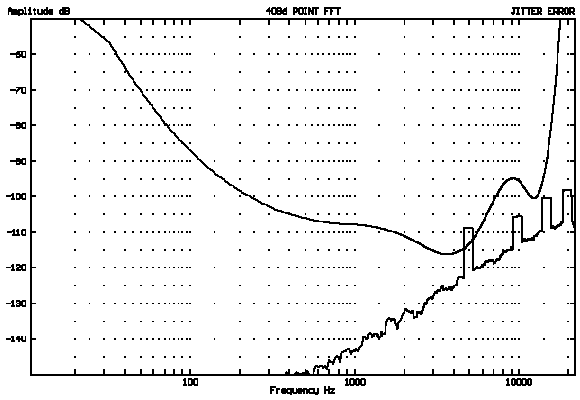


(b)

Fig. 35. Simulated worst case jitter errors for sinusoidal jitter. (a) 100% DAC, 0 dB 22 kHz audio, 20 ps pk. 18.5 kHz jitter. (b) Impulsive DAC, 0 dB 100 Hz audio, 75 ps pk. 4 kHz jitter.

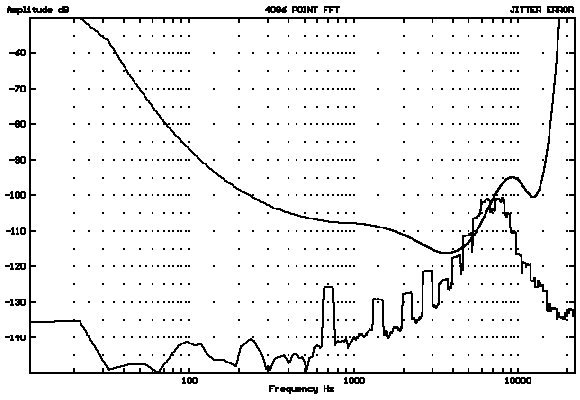


(a)

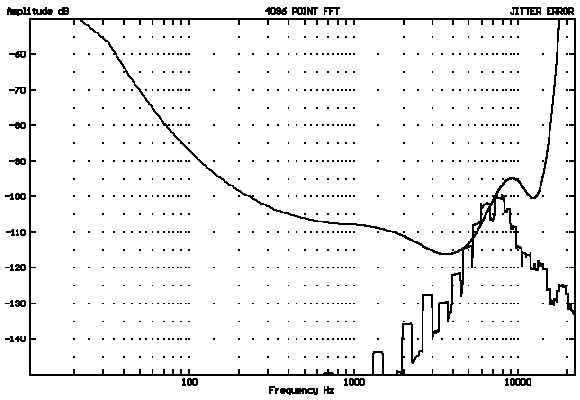


(b)

Fig. 36. Simulated band-limited interface jitter errors with no PLL filtering. (a) 0 dB 20 kHz audio, interface time constant 40 ns, 100% DAC. (b) 0 dB 9.8 kHz audio, interface time constant 50 ns, impulsive DAC.



(a)



(b)

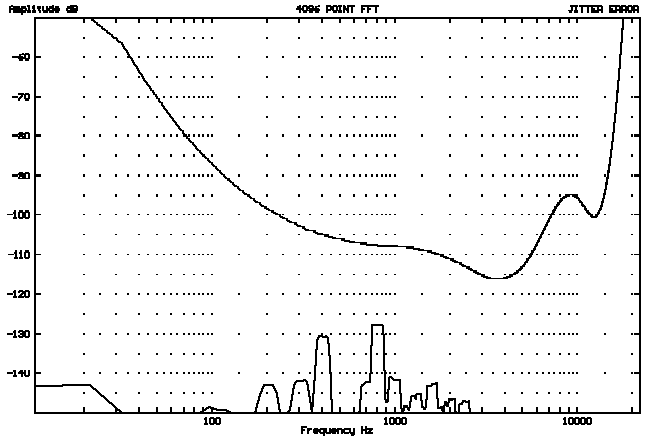
Fig. 37. Simulated band-limited interface jitter errors with 2nd order 1 kHz PLL filter. (a) 0 dB 7 kHz audio, interface time constant 120 ns, 100% DAC. (b) 0 dB 7 kHz audio, interface time constant 120 ns, impulsive DAC.

5.1 Reducing the Audibility of Interface Jitter Errors

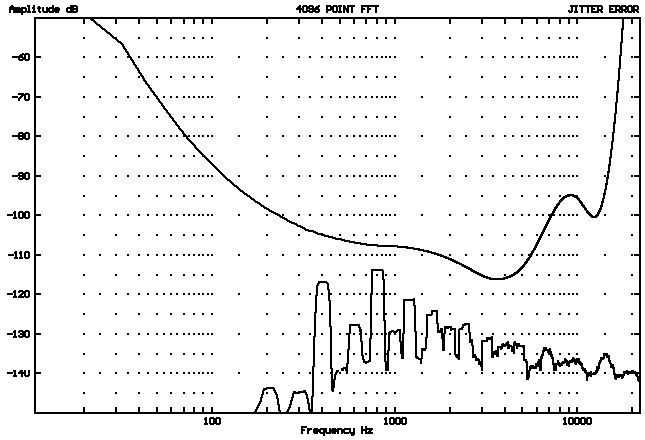
We have shown that embedded clock jitter in a band-limited digital audio interface is fundamentally a problem due to the digital audio interface standard in its current form. In particular the jitter signal is intimately linked with the zero-one bit sum of the serially transmitted audio words; a lower zero-one sum range would result in reduced jitter. One way of lowering the 'sum using the present interface standard is to pad the unused bits in the auxiliary data and low bit sections of each subframe (see Fig. 2). For example, with 16 bit audio words the sum can be reduced from ±16 to ±8 by appropriately filling these unused bits according to the value of each transmitted audio word. A second option is to transmit one of the two audio channels supported by the interface out of phase such that, for stereo programme with a strong monaural component, the zero-one sum is reduced. However, although the absolute jitter level may be reduced using these techniques, we will still be left with the undesired high correlation between jitter and audio. In order to break this correlation we must somehow randomise the sign structure of the transmitted bits, perhaps by modulating each bit sign with a known pseudo-random binary sequence which is synchronised at the start of each interface block.

However the AES/EBU-S/PDIF digital audio interface is now in widespread use and we are unlikely to see a new interface standard welcomed by the majority of users for a considerable time. Accepting that jitter will occur in a band-limited interface we must examine methods of reducing recovered clock jitter in

receivers. Evidently the first step to take in minimising interface jitter is to maximise the bandwidth of all components in the interface. However, there are limits to achieving high bandwidth at low cost, especially for long links, and conflicts with RFI may arise as rise and fall times become very fast. Alternatively the audibility of interface jitter in a conversion process can be reduced by making the receiver PLL cutoff frequency as low as possible. However it should be noted that PLLs with extremely low cutoff frequencies may be compromised in other performance aspects such as initial lock-on time and the frequency range over which lock is maintained. These problems can be overcome by employing two phase lock loops in series where the first has high bandwidth and fast start up and the second low bandwidth module switched into circuit after conditions have stabilised. Low loop bandwidths can be achieved whist maintaining wide lock range by using random access memory (RAM) to buffer the audio data before being clocked out to the converter unit.



(a)



(b)

Fig. 38. Simulated interface jitter errors with low frequency audio signal (0 dB 200 Hz), 100 ns interface time constant and 2nd order 1 kHz PLL filter. (a) 100% DAC. (b) Impulsive DAC.

A more fundamental solution to the problem of interface jitter is to locate a high precision conversion clock inside the `receiver' (*i.e.* DAC). This then creates the problem of how to match the transmitter and receiver data rates. One solution is to employ a RAM buffer for audio data where the rate at which the data is clocked out of the interface is under the control of the local clock. Enough memory must then be available to account for the long-term differences between the transmitted interface clock and the local oscillator

frequencies. However, this is hardly an acceptable solution in a studio environment where equipment must operate with identical sampling frequencies. A more general solution is to use the receiver clock to slave the transmitter, an approach adopted both by Sony and Linn in 2 box CD systems. Details concerning how to implement the additional clock interface vary, but perhaps the most sensible and universal approach is to provide a second independent digital audio interface output at the receiver from which the transmitter can derive a clock signal (Fig. 39). This scheme is also compatible with the recommendations made in AES11-1991 [23].

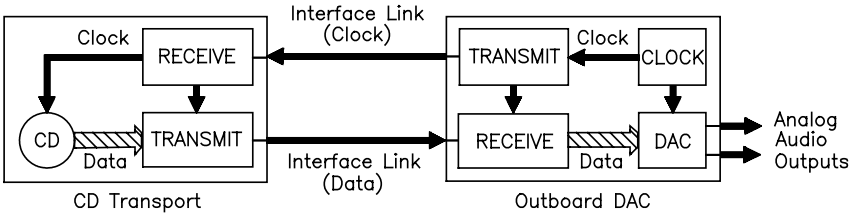


Fig. 39. Proposed system connection for master clock at receiver utilising two digital interface links.

5.2 Measuring Jitter in the Digital Audio Interface

The simulations presented above have indicated that interface jitter may be a real problem in practical audio systems. How can we measure the effects of such jitter? In general jitter errors in DACs will lie close to the noise floor of the conversion process, resulting in two basic problems for measurement strategies. Firstly there is the problem of resolution in the test instrument itself; since jitter errors are most easily revealed when exercising the test system with full scale audio signals then errors in the ADC of a digital measurement system, also excited to close to full scale, may well swamp those due to jitter in the test device. Secondly there is the problem of determining the source of measured errors; we have seen that jitter errors can contain both noise-like components *and* spectral lines, and these can be confused with distortion and noise modulation due to DAC non-linearity. Noise modulation tests where the exciting audio frequency is chosen such that all DAC non-linearity falls on a limited number of FFT bins are perhaps the best suited to revealing jitter errors. More work is required to develop a suitable test for jitter errors, although we can at this stage identify the tests currently in widespread use that will *not* reveal jitter errors. Firstly the popular `fade to noise' test which measures level-dependent logarithmic gain [35]; in this test the output of the test device is filtered with a third octave band-pass filter centred on the test frequency at 1 kHz, where jitter errors will not be detected since no jitter error occurs at the fundamental frequency in either of the DAC error models discussed above. Secondly the low-level noise modulation test where a 41 Hz sinewave is applied to the DUT at levels of -40 dB and below [36]; the problems here lie in signal levels which are too low to result in large jitter errors for either DAC model and a test frequency which will not excite large jitter errors in the 100% model.

6 CONCLUSIONS

Is the digital audio interface flawed ? We have examined the possibilities of both amplitude and timing errors corrupting audio data transmitted across an interface. The probability of received amplitude errors is not high, and indeed are most likely to occur in the preamble of each interface subframe. This means that if a receiver can lock onto an incoming interface signal, then the audio word values are safe! However jitter remains a concern; several jitter mechanisms exist for the biphase-mark encoded signal, the biggest problem being that of bandwidth limitation at *any* stage of the interface. We have shown that band-limited interface jitter has a strong relationship to the bit structure of the serial interface code, and hence can be highly correlated with the transmitted audio data. Measurements have confirmed jitter levels of higher than 1 ns in an above average interface circuit.

The effects of jitter can be predicted by forming error models for different DAC architectures. It can be shown that, compared to low-oversampling multibit designs, pulse density modulation converters are much more sensitive to jitter when producing low frequency audio signals. This phenomenon may explain certain subjective characteristics of PDM DACs which cannot otherwise be rationalised. A simple model of jitter error audibility has shown that white jitter noise of up to 180 ps can be tolerated in a DAC, but that even lower levels of sinusoidal jitter may be audible. These limits place tough constraints upon digital interface design, and it is recommended that interface receiver PLLs have closed loop cutoff frequencies as low as possible. For the ultimate immunity to the effects of jitter, a second digital audio interface employed at the receiver can be used to slave the transmitter.

7 ACKNOWLEDGEMENT

This work has been funded by the Science and Engineering Research Council, UK.

8 REFERENCES

[1] S. P. Lipshitz, J. Vanderkooy and R. A. Wannamaker, "Minimally Audible Noise Shaping," *J. Audio Eng. Soc*., Vol. 39, pp. 836 - 852 (1991 Nov.).

[2] P. van Willenswaard, *Stereophile*, November 1988, pp. 51 - 53.

[3] AES3-1985, "AES Recommended Practice for Digital Audio Engineering - Serial Transmission Format for Linearly Represented Digital Audio Data," *J. Audio Eng. Soc*., Vol. 33, pp. 979 - 984 (1985 Dec.).

[4] R. C. Cabot, "Measuring AES-EBU Digital Audio Interfaces," presented at the 87th Convention of the Audio Engineering Society, October 1989, preprint 2819.

[5] R. A. Finger, "Review of Frequencies and Levels for Digital Audio Performance Measurements," *J. Audio Eng. Soc*., Vol. 34, pp. 36 - 48 (1986 Jan/Feb.).

[6] E. Meitner and R. Gendron, "Time Distortions Within Digital Audio Equipment Due to Integrated Circuit Logic Induced Modulation Products," presented at the 91st Convention of the Audio Engineering Society, October 1991, preprint 3105.

[7] P. van Willenswaard, *Stereophile*, November 1990, pp. 78 - 83.

[8] S. Harris, "The Effects of Sampling Clock Jitter on Nyquist Sampling Analog-to-Digital Converters, and on Oversampling Delta-Sigma ADCs," *J. Audio Eng. Soc*., Vol. 38, pp. 537 - 542 (July/Aug 1990).

[9] M. O. Hawksford, "Digital-to-Analog Converter with Low Inter-Sample Transition Distortion and Low Sensitivity to Sample Jitter and Transresistance Amplifier Slew Rate," to be presented at the 93rd Convention of the Audio Engineering Society, October 1992.

[10] B. A. Blesser, "Digitisation of Audio," *J. Audio Eng. Soc*., Vol. 26 (1978 Oct.).

[11] P. van Willenswaard, *Stereophile*, October 1991, pp.63-69.

[12] M.Colloms, Hi-Fi News and Record Review, October 1989, pp. 83 - 85.

[13] B. Adams, "Comments on "Chaos, Oversampling and Noise-Shaping in Digital-to-Analog

Conversion"," *J. Audio Eng. Soc. (Letters to the Editor)*, Vol. 38, pp. 766 - 768 (1990 Oct.).

[14] M. Otala, "Feedback-Generated Phase Intermodulation in Audio Amplifiers," presented at the 65th Convention of the Audio Engineering Society, February 1980, preprint 1576.

[15] M. Otala, "Phase Modulation and Intermodulation in Feedback Audio Amplifiers", presented at the 68th Convention of the Audio Engineering Society, March 1981, preprint 1751.

[16] R. R. Cordell, "Phase Intermodulation Distortion - Instrumentation and Measurement Results," presented at the 70th Convention of the Audio Engineering Society, October 1981, preprint 1842.

[17] M. J. Hawksford, "Fuzzy Distortion in Analog Amplifiers: A Limit to Information Transmission ?," *J. Audio Eng. Soc.*, Vol. 31, pp. 745 - 754 (1983 Oct.).

[18] M. J. Hawksford, "Nth.Order Recursive Sigma-ADC Machinery at the Analogue-Digital Gateway," presented at the 78th Convention of the Audio Engineering Society, May 1985, preprint 2248.

[19] C. Manson, "Digital Sound Signals: Subjective Effect of Timing Jitter," BBC Research Department Engineering Report 1974/11, March 1974.

[20] P. S. Lidbetter, "Basic Concepts and Problems of Synchronisation of Digital Audio Systems," presented at the 84th Convention of the Audio Engineering Society, March 1988, preprint 2605.

[21] T. Shelton, "Synchronisation of Digital Audio," Proc. AES 7th Int. Conf.: Audio in Digital Times (Toronto, Ont., Canada, 1989).

[22] R. D. Fourre, "Testing 20 Bit Audio Digital-to-Analog Converters," Proc. AES 7th Int. Conf.: Audio in Digital Times (Toronto, Ont. Canada, 1989).

[23] AES11-1991, "AES Recommended Practice for Digital Audio Engineering - Synchronisation of Digital Audio Equipment in Studio Operations," *J. Audio Eng. Soc.*, Vol. 39, pp. 156 - 162 (1991 Mar.).

[24] J. Dunn, "Considerations for Interfacing Digital Audio Equipment to the Standards AES3, AES5 and AES11," Proc. AES 10th Int. Conf.: Images of Audio (London 1991).

[25] J. R. Stuart, *Hi-Fi News and Record Review* (Letter to the Editor*),* January 1991, p.15.

[26] ISO226:1987, "Acoustics - Normal Equal-Loudness Level Contours for Pure Tones under Free-Field Listening Conditions," 1987.

[27] J. R. Stuart, "Predicting the Audibility, Detectability, and Loudness of Errors in Audio Systems," presented at the 91st Convention of the Audio Engineering Society, October 1991, preprint 3209.

[28] J. R. Stuart, "A Search for Efficient Dither for DSP Applications," presented at the 92nd Convention of the Audio Engineering Society, March 1992, preprint 3334.

[29] M. Colloms, *Hi-Fi News and Record Review*, March 1991, pp. 69 - 71.

[30] M. A. Krasner, "The Critical Band Coder - Digital Encoding of Speech Signals Based on the Perceptual Requirements of the Auditory System," Proc. 1980 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 327 - 331 (1980).

[31] Crystal Semiconductor Corporation, CS8411/CS8412 Digital Audio Interface Receiver Data Sheet, April 1991.

[32] S. Sakura, S. Onobuchi, M. Ito and S. Katagiri, "Fiber Optic Link for Digital Audio Interface," *IEEE Trans. Consumer Electronics*, Vol. 34, pp. 667 - 669 (1988 Aug.).

[33] Toshiba Corporation, Toslink Fiber Optic Devices Catalogue, 1990.

[34] AES3-1992, "AES Recommended Practice for Digital Audio Engineering - Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data," *J. Audio Eng. Soc.*, Vol. 40, pp. 146 - 165 (1992 Mar.).

[35] AES17-1991, "AES standard method for digital audio engineering - Measurement of digital audio equipment," *J. Audio Eng. Soc.*, Vol. 39, pp. 962 - 975 (1991 Dec.).

[36] R. C. Cabot, "Noise Modulation in Digital Audio Equipment," presented at the 90th Convention of the Audio Engineering Society, February 1991, preprint 3021.