TOWARDS COMMON SPECIFICATIONS FOR DIGITAL AUDIO INTERFACE JITTER

Julian Dunn, Cambridge, England
Barry A. McKibben, Tektronix, Beaverton, Oregon, USA
Roger Taylor, Crystal Semiconductor Corporation, Austin, Texas, USA
Chris Travis, British Broadcasting Corporation, London, England

ABSTRACT

An ad-hoc Task Group within the SC-2-2 (Digital Input/Output Interfacing) Working Group is developing a more detailed specification for jitter in the embedded clock of the digital audio interface signal. The discussion and the intermediate conclusions of this work are presented in this paper as an invitation for informal comment and debate about the issues raised.

1. INTRODUCTION

When digital audio equipment was relatively rare the AES3 Digital Audio Interface [] was mostly used for short-distance links between the few pieces of digitally interfaced equipment in the same studio. It is now increasingly being used over complete studio and broadcast centres. The link carries a digital signal which should be robust - but in some cases there can be problems. Most aspects of link integrity can be monitored by error detection circuits at the receiver but if necessary the digital signal quality can be assessed by examination of the data stream using an oscilloscope or a dedicated analyzer, and comparing the signal with the interface specification.

Some interface problems are related to deviation in the timing of the interface signal. As the effect of these errors is different they can be divided into three categories:

i) Short-term timing variations that can cause equipment to fail to decode the signal correctly. This is interface jitter.

ii) Static or very slowly changing timing deviations of several micro-seconds that cause problems with sample synchronisation between equipments, but should not produce data errors. These characteristics are called offset and wander.

iii) Frequency errors beyond the frequency lock-in range of the receiving equipment. This will often result in total loss of signal, or at least make the signal unusable.

Specifications for these parameters are in AES3, AES11 [], and AES5 [] respectively.

As larger systems and more equipment have become available the specification for data jitter (in AES3) has been found to be ambiguous and inadequate to guarantee interface operation. When interface problems related to jitter have arisen it has been found not specific enough to determine which equipment is at fault. In this paper the argument for revision will be presented and some new specifications will be proposed.

Please note that these specifications are primarily related to the integrity of the data communications channel formed by the interface. This is distinct from the effect on the quality of an audio signal if it is sampled or re-sampled with timing errors, as may happen with analogue to digital converters (ADCs), digital to analogue convertors (DACs) and sample rate convertors (SRCs). Such timing errors are called sampling jitter. The requirements for sampling jitter (which are related to audibility criteria) are much tighter than those possible for the interface and are not being directly considered in this document. However the clock generating circuit used in a sampling device will often derive a timing reference from the interface signal itself. For good performance these devices require a phase locked loop (PLL) that will filter out the interface jitter to make a sample clock that is pure enough to avoid potentially audible jitter modulation products. The performance required from these jitter filters depends on the worst case data jitter expected or permitted on the interface - and that parameter is considered in this paper.

2. THE DIGITAL AUDIO INTERFACE

In order to interconnect pieces of digital audio equipment without using analogue interfaces the two devices need a data communications channel and a common clock synchronisation. In the simplest case one connection can be used for both functions. In more complex cases the devices are all synchronised to a common timing reference or master clock, and the interconnecting signal is used purely to pass data. The digital audio interface format discussed here is used, in different situations, for all three functions: Passing audio programme data and synchronisation, carrying a master clock synchronisation signal, and carrying audio data which is not being used to synchronise equipment.

2.1. The interface format

The AES3 digital audio interface for professional applications - defined in [], (The EBU [] and the appropriate part of IEC-958 [] specifications are almost identical) and the similar consumer interface format, defined in IEC-958 and EIAJ CP-340 [], define a bi-phase mark coded signal carrying two channels of audio data with an embedded

clock. This allows both the data and the sample-synchronous clock to be carried on one circuit.

There are two sub-frames of data per sample period - one for each channel. Each sub-frame has 32 equal time slots. For slots 4 to 31 signal transitions always occur at time slot boundaries. Transitions will also occur at the centre of the time slot if the data bit carried in that time slot is a logic one. There are synchronising preambles, in time slots 0 to 3, that are recognised by a decoder because they omit some of the time slot boundary transitions. This is in order to identify the start of each data word.

From this it can be seen that, in addition to audio data, the interface carries an embedded `bit-clock' at 64 cycles per frame, and synchronising patterns for deriving a sample rate clock. For a sample rate of 48kHz the embedded clock is at 3.072MHz.

2.2. The digital audio reference signal, DARS

The synchronisation standard, AES11 [], recommends the use of a separate signal for maintaining common synchronisation: This is called a digital audio reference signal or DARS. Devices that can use a DARS will have a dedicated input that can be selected as the synchronisation reference. When this is used the synchronisation of the device output is derived from the DARS, which can be sourced at one point for the whole studio. This helps to control the build-up of synchronisation timing errors that can occur when signals are cascaded through a large number of units. The other significant benefit from the use of a DARS is that, as the synchronisation is independent of the programme signal routing, the signal source can be changed without disrupting the timing of the equipment.

3. DEFINITION OF TERMS

3.1. Interface jitter

In a practical system interface signals will suffer from timing inaccuracies. Transitions will occur at times slightly in advance of or later than the expected times. Such fluctuations are called jitter.

In order to differentiate jitter on the interface signal from sampling jitter (which occurs within convertors) the former is called interface or data jitter.

3.2. Sampling jitter

As mentioned in the introduction, jitter in sampling clocks affects the quality of the final reproduced signal by adding potentially audible phase modulation products to the original signal. The audibility of these products is related to the nature of the jitter and of the audio signal. One of the authors has calculated the jitter amplitudes required for audibility of these products in the presence of worst case signals []. The graph reproduced in figure is taken from that paper. This shows that even sub-nanosecond amplitudes of sampling jitter can produce an audible effect (with some audio signals) when the jitter frequency is above 600Hz. Below that frequency the modulation required for audibility rises steeply as a result of masking by the tone being modulated.

In many convertors clock recovery circuits are used to derive sampling clocks from the interface signal. If these do not properly attenuate the jitter coming from the interface - which can often be orders of magnitude above the levels required for a sample clock - then sampling jitter may be at an excessive level.

With appropriate clock recovery circuits reasonable amounts of interface jitter need not affect the quality of the finally reproduced signal - and measurements of interface jitter might not give any clue to the levels of sampling jitter in the associated equipment.

3.3. Jitter frequency

There is sometimes confusion over just what is meant by "jitter frequency". It can best be understood if the timing jitter deviation itself is considered a sampled signal. The timing error at each clock edge can be plotted against time. This error may be sinusoidal, in which case there is a discrete component frequency in the jitter. More generally the spectrum will be more complex - as are signals represented by voltages against time.

3.4. Unit interval

The digital audio interface signal consists of pulses in either direction of nominal widths of 1/128th, 2/128th and 3/128th of a sample frame period. The smallest of these nominal widths is referred to as the unit interval (UI). At a sample rate of 48kHz this corresponds to 163ns.

The unit interval is often used for specifying jitter amplitudes because, as it relates directly to the duration of the data pulses, it gives figures that scale appropriately with the nominal data pulse width. For example the CCITT telecommunications standards G.823 and G.824 [,] specify jitter amplitudes in UI. This allows the same jitter limits (specified in UI) to apply over a wide range of bit-rates. It is suggested here that for similar reasons, and for compatibility both with these standards and with the test equipment guidelines that have been written for them [,], that the unit interval should be used in any new digital audio interface jitter specifications.

3.5. Intrinsic jitter

If a unit is either free-running or synchronised with a jitter-free signal then any output jitter measured at the transmitter will be generated from within. This is referred to as the intrinsic jitter of the device. The level of intrinsic jitter is usually determined by the type of clock recovery circuit. For example, a `high Q' circuit, such as a phase locked loop (PLL) with a voltage controlled crystal oscillator (VCXO), should produce lower intrinsic jitter than a `low Q' circuit using resistors and capacitors as the timing elements.

Intrinsic jitter in a PLL is attenuated by the action of the feedback around the loop. Therefore a PLL with loop gain to high frequencies will not have much intrinsic jitter at low frequencies. In addition the phase noise with typical oscillators rises at lower frequencies. These two factors mean that loops designed to attenuate jitter by having lower loop gain (giving lower corner frequencies) tend to have more intrinsic jitter.

The intrinsic jitter of a device has taken on particular significance because the present AES3 jitter specification is often interpreted to refer only to this type of jitter.

3.6. **Jitter transfer**

When a device is synchronised with an external source, such as an audio or DARS input or a video timing reference, any jitter on this source can be passed through the clock recovery circuit(s) within the device to the digital output. The jitter at the output is, as a result, a combination of intrinsic and transferred jitter.

The transfer characteristic is often non-linear, particularly at lower jitter levels. However, at higher levels it is generally modelled as linear with the gain being a function of jitter frequency: Devices are regarded as jitter transceivers having transfer functions that are determined by the type of clock recovery circuit that they use.

The magnitude response of a typical jitter transfer function is shown in figure . At frequencies within the bandwidth of the transceiver PLL the output tracks the input and the jitter or wander is passed through with no attenuation. At higher frequencies - beyond the PLL bandwidth - there is increasing attenuation of input jitter as jitter frequencies increase until the output jitter is dominated by the intrinsic jitter of the device.

In the figure the gain initially rises to peak at a value that is actually greater than unity. This is called jitter gain peaking. It can aggravate the problem of jitter accumulation, particularly when several similar devices are connected in cascade. (More on jitter accumulation later).

Many digital audio devices suffer from jitter gain peaking of several dB. This is in contrast to the telecommunications industry where there are tight specifications to control jitter gain. (Typically these require peaking to be kept below 0.5dB.) It is true that the single-chip digital audio interface receivers are generally well-behaved in this respect. This is partly because, for data decoding, low peaking goes hand-in-hand with good jitter tolerance. However (for various reasons) many pieces of equipment also include a second stage of clock recovery, typically a PLL. This second stage is independent of the data decoding process. The applications notes provided with most general-purpose PLL chips don't focus on jitter transfer. Calculations suggest that they lead to circuits having peaking of between 3 and 6dB.

3.7. **Jitter tolerance**

The amount of interface jitter required before an AES3 receiver fails to decode the signal correctly is called the jitter tolerance. This tolerance is a function of the jitter frequency as follows:

Gradual timing variation (wander and low-frequency jitter) is usually tracked by the PLL of an AES3 receiver and causes no data errors. In fact, slow variations may have a peak amplitude (in time) of many unit intervals before the slewing capability of the pll is exceeded. If, however, timing variations are so abrupt that the receiver's PLL fails to track them (high-frequency jitter), then an amplitude as small as half a unit interval may cause a displaced edge to cross a timing threshold and cause a bit error. This shows why jitter tolerance increases as jitter frequency goes down, and why the peak amplitude of the jitter is usually more important than either the RMS or the average amplitude.

The jitter tolerance of a device is partly related to the frequency agility of the data recovery system. For example, receiver circuits with limited data-recovery clock deviation (such as those using a VCXO to recover the data) will have a limited jitter tolerance. For instance a clock jitter of 100ns peak to peak (pk-pk) at 320 Hz represents a frequency deviation of 100ppm pk-pk. This is too large for many VCXOs to track, so errors would be produced.

Typical jitter tolerances of receivers capable of small and large frequency deviation are compared in figure . Below the corner frequency the receiver will tolerate increasing amounts of jitter as the jitter frequency drops. The separation of the lines below 25kHz shows that the device capable of higher clock deviation has a higher tolerance limit in this region. This is because the clock can track the jitter. Above the corner frequency the receiver data recovery clock irons out the timing variations and does not track. Existing devices typically have a tolerance in this region of 0.25 UI pk-pk.

If, as is sometimes the case, only one PLL is used both for data recovery and output clock re-generation then the JTF and the jitter tolerance of the device are directly related. The frequency at which the jitter transfer function starts to attenuate will be

close to the frequency at which the jitter tolerance reaches the minimum value. This is unfortunate because - even with a ideal PLL damping - jitter from earlier stages will not have been attenuated and could have built up to a level that exceeds the jitter tolerance of the next stage. This close proximity of the knees on these two curves (as shown in figure) is a fundamental disadvantage with the single PLL architecture.

3.8. **Pattern-dependent jitter**

If a jitter-free interface signal is transmitted down a cable then the data pattern in the signal will modulate the signal zero-crossing timings. This modulation is such that patterns of zeroes produce more delay to the transition timings than patterns of ones, so as the data varies the timings modulate in sympathy. The amount of modulation depends on the amount of smoothing the waveform has undergone as a result of high frequency loss in the cable. Figure shows accumulated storage oscilloscope traces of this effect using 100 metres of Belden 9180 cable. This shows zero-crossings to be shifted by about 5ns dependent on the immediately preceding data pattern. Figure shows the effect is much greater on an equivalent length of Belden 8451 cable. (This lower bandwidth cable is typical of the type installed for analogue audio operation).

The effect has been analyzed by Chris Dunn and Malcolm Hawksford [], and by Julian Dunn [], both using a first-order low-pass filter model for the cable. This model shows how cable-induced jitter strongly dependent on the bandwidth of the link. For example it is less than 1ns if the cable has a 3dB bandwidth of more than 4MHz but if the bandwidth of the link is halved there will be 9ns pk-pk of modulation on these timings. If these edges are used to lock a PLL being used to generate an interface output then this jitter will transfer, subject to the PLL response, to the output and pass on to the next interface.

This need not be the case. If the data pattern immediately preceding the transition is constant the mechanism for inducing the jitter will be removed. Figures and , illustrate this. The static pattern, in this case, is one of the synchronising preambles and the traces show significantly reduced modulation. (Figure shows the Belden 9180 cable with jitter of less than 1ns compared with 4.5ns in the modulated part of the signal). It is possible to use these edges to lock a PLL to generate a clock relatively immune to data pattern dependent jitter. Note that this PLL would have to be in addition to the data recovery PLL in the receiver circuit. That first stage is required in order to isolate the preambles. This is a second reason for using a separate PLL stage for the interface output clock.

There are other mechanisms for the generation of pattern-dependent jitter. These are generally mechanisms that distort the symmetry of the waveform, such as un-equal rise and fall times of the data pulses, DC offsets in the differential line receivers, logic level miss-matches between devices and transition asymmetries within devices. Though these mechanisms are slightly different, the technique described above can also avoid this form of jitter being transferred to the output clock. The even parity of the interface format ensures that, in any continuous signal, the edges in the preambles will keep the same transition polarity, and therefore differences in transition times between positive-going and negative-going edges will not produce jitter on those edges.

3.9. **Jitter accumulation**

If we consider a short chain of digital audio devices, where each device is locked to the previous one, we have several contributions to the jitter at the end of the chain. Each device will add its own intrinsic jitter, and each interconnecting cable will make some contribution with cable-induced pattern-dependent jitter. There will also be some jitter gain or loss at each stage. The net effect of this varies with the individual device jitter characteristics and the data patterns at each stage but, with a relatively simple calculation it is possible to get some idea of the potential problem.

For the purposes of this calculation we are looking at jitter at frequencies below the roll-off frequency of the JTF, so jitter attenuation does not occur. If we assume that all the devices contribute the same amount of jitter, J, at each stage (lumping cable-induced and intrinsic jitter together), and that each device also amplifies the jitter from the previous stage by the same gain and the total number of stages is N, then the total output jitter produced at the end of the chain, as a multiple of J, can be tabulated for different gains as follows:

Gain per device	Total jitter after 3 stages	Total jitter after 4 stages	Total jitter after 5 s
0dB (ideal)	3J	4J	5J
1dB	3.8J	5.4J	7.1J
3dB	6.2J	10.2J	15.8J
6dB	13.9J	29.8J	61.4J

This table shows that with a gain of 0dB at each stage the output jitter is simply a sum of the jitter produced at each stage. (These jitter levels are peak values so they will add). Remember that this happens at frequencies below the JTF roll-over frequency. At higher frequencies the input jitter will be attenuated and so the final output jitter will grow more slowly.

The gains of greater than 0dB show the worst-case effect of jitter transfer function peaking. If peaking is present it is near to the JTF roll-over frequency. Where the jitter is wide-band only a small proportion of it will be amplified and the peaking will have little effect. But there are two mechanisms that can concentrate the jitter in the region of the peak.

Firstly, the pattern-dependent jitter can have narrow spectral components. For example with low level audio signals the jitter will become coherent with the polarity of the signal. This occurs because for small numbers the more significant bits within the data word change together as an extension of the sign bit. If the interface audio signal is a low level tone at one frequency then the cable induced jitter will concentrate at that frequency. Such a spectral peak would sometimes coincide with the peak in the JTF.

Secondly, the spectrum of intrinsic jitter from a PLL is weighted towards it's JTF roll-over frequency. So if the previous stage has a similar PLL (with the same roll-over frequency) then the intrinsic jitter from that stage will be amplified by the peak.

In a chain of devices with similar characteristics this signal will have the same effect at each stage. The figure of 6dB in the table reflects levels of peaking found in equipment that has not been designed to avoid this problem. As the table shows this can lead to a very large amount of jitter accumulation after only a few similar stages. The normal symptom of this effect is for equipment occasionally to lose data or lock, particularly with low-level audio signals.

A signal that has accumulated jitter can be 'cleaned-up' by re-timing it with a low-jitter synchronous reference such as a DARS. This can be done at every stage, as recommended in AES11 for master-clock synchronisation. If it is done less often then the level of jitter accumulation will depend on the characteristics of the devices in the un-re-timed chains that are formed. Ideally these characteristics need to be specified to keep the jitter within the jitter tolerance of the next devices and the re-clocking stage at the end of the chain.

Even if all equipment in a chain is using master clock synchronisation the problem of jitter accumulation still exists - but it is now solely in the DARS distribution chain.

With single stage clock recovery systems the clock used to recover the data is the same as the clock used to generate the output timing. This means that there has to be a compromise between the requirements for jitter tolerance, which benefit from having a high PLL roll-over frequency, and those for jitter attenuation, which benefit from having a low PLL roll-over frequency. The effect of this compromise is compounded in any system combining similar circuits. Jitter accumulation will occur at frequencies up to the roll-over frequency. At that frequency the jitter tolerance is almost at its minimum because the PLL is no longer tracking the jitter. An overlay of the jitter transfer function and the jitter tolerance of the same PLL is shown in figure to illustrate the situation.

A two-PLL system allows the two curves to be independent. At the cost of a separate PLL for device output clock it is possible to shape the JTF to ensure that jitter accumulation can only occur at frequencies where the jitter tolerance is at a higher level. This is shown as a dashed line on the same figure.

4. JITTER MEASUREMENT TECHNIQUES

The following techniques illustrate fairly rudimentary methods for measuring some of the parameters being discussed. They are not being proposed as recommended techniques but more as illustrations to show how tests could be devised.

Interface jitter can be measured by comparing the interface transition times with those of a related low jitter signal or clock. This low-jitter reference can be derived from the signal to be measured using a low-noise PLL as a low-pass filter (shown in figure , taken from []), or it can be taken from an interface signal earlier in the synchronisation chain, such as a master clock generator, if that is known to have low jitter (figure). The former approach requires dedicated hardware and can be applied universally irrespective of synchronisation source.

The latter technique is much simpler for where it can be adopted. Use a relatively jitter-free signal (such as would be provided by a source clocked directly by a free-running crystal oscillator) as the synchronisation reference. An indication of output jitter can be assessed (without regard for the low frequency cut-off) by direct comparison of the input and output signals on a two-channel oscilloscope. Within the limitation of estimating the width of a diffuse oscilloscope trace this will show the intrinsic jitter of the device, and because there is no jitter at the input it is not affected by the device jitter transfer function. (An accumulating storage oscilloscope can provide clear results - as illustrated in some of the earlier figures).

Intrinsic jitter can be measured directly using the techniques described above. Transfer function and tolerance measurements require a known jitter source, such as a sinusoidal jitter generator.

To measure the jitter transfer function, drive the synchronisation input of a device under test (DUT) with sinusoidal jitter of known amplitude while the output jitter level is measured. This should be repeated over several frequencies. A test arrangement for this is shown in figure .

Jitter tolerance is measured by driving the data input of the DUT with the jitter generator. The data stream should carry a pre-defined modulation, such as a pseudo-random sequence. For a given applied sinusoidal jitter frequency the tolerance is the maximum jitter level that can be applied before the error rate becomes unacceptable. This should be repeated with sinusoidal jitter at each frequency to be tested. As the interface has no error correction capability the error rate required needs to be zero (or

at least very low). Bit error rates (BERs) acceptable to the telecommunications industry, such as 1 in 10^6, will produce several uncorrectable errors per second on the AES3 interface.

It is possible to assess pattern-dependent jitter susceptibility by measuring the output jitter of the DUT with an unmodulated jitter-free signal at the synchronisation input. Then, to simulate cable losses, put a simple first order low pass network with a time constant of about 100ns at the DUT input, modulate the audio data with a low level, low frequency tone and make a second measurement. If the output jitter measurement worsens then the device is susceptible to pattern dependent jitter. If it stays roughly the same then it is not. (A low-level tone is used because that has the effect of modulating the twos-complement data pattern with almost all ones or almost all zeroes, dependent on the sign of the audio signal, and this will produce jitter at the same frequency as the audio. A low audio frequency is used because it is likely to be in the flat, pass-band, region of the DUT jitter transfer function)

Note that the maximum possible jitter frequency carried on the interface is determined by the rate of transitions in the signal. This depends on the data pattern. For a signal of all ones there can be up to 124 transitions per frame: 6 million per second. Through Nyquist's theorem this gives a jitter bandwidth of about 3MHz (both figures are for a sample rate of 48kHz). For a signal with all zeroes the rate of transitions is halved, so the jitter bandwidth is 1.5MHz. In normal systems jitter at these frequencies does not generally accumulate because it is strongly attenuated by almost all clock recovery systems.

It is possible for clock recovery circuit behaviour to be very sensitive to exact signal frequency. It is therefore important to assess equipment over the range of frequencies that the equipment is specified for.

5. THE CURRENT INTERFACE TIMING SPECIFICATIONS

5.1. Jitter

The AES3 standard provides a specification for jitter at the output port of an interface transmitter. This is defined in [] as follows:

> **6.2.5 Data Jitter**
>
> Data transitions shall occur within ±20 ns of an ideal jitter-free clock measured at the half-voltage points.

The ideal jitter-free clock that is used as the measurement reference is not defined but could logically be assumed to have a frequency and phase corresponding to the average frequency and phase of the clock carried in the data-stream being examined. The timescale over which these averages are taken will effectively define a lower frequency cut-off for the measurement. The phase noise in many oscillators tends to reduce with increasing jitter frequency with the effect that the value used for this cut-off frequency can have a significant effect on the results that are measured. This will impact on the clock generation circuits required to meet the specification. (In order to eliminate this inconsistency one of the authors, Dunn, suggested a brief amendment to this specification [] - and it was this suggestion that prompted the discussion which has lead to this paper).

When the device is not the master clock synchronisation source the interface transmitter will be itself synchronised from a signal which may have some jitter. Some of this jitter will be transferred to the device output. The amount of transfer will depend on its spectrum and on the jitter transfer function of the device. For any output jitter measurement it is necessary to define the jitter in the synchronisation source.

It may seem logical for the output jitter specification to be a measurement in worst-case conditions. However at present there is no specification for the worst-case jitter at the input port of the digital audio interface. One could assume the maximum allowable output jitter specified above is present on the synchronisation input but there are two problems with this approach:

Firstly, the frequency-dependence of the jitter transfer function of the device under test means that two jitters of the same amplitude, but different frequency spectra, may produce very different output jitter levels, so the shape of the jitter spectrum would also have to be defined.

Secondly, jitter at an interface receiver comes from a combination of the previous stage output jitter, and jitter induced by losses in the cable - so the specification for maximum output jitter level does not represent the worst-case input jitter amplitude and will not predict behaviour in those conditions.

5.1.1 AES11 and jitter

The current revision of the synchronisation standard, AES11, makes a reference to the interface jitter specification. This is as follows (taken from ref [] para 5.4.2):

> It should be noted that the interface specification permits a jitter level, sample to sample, of ±20 ns, irrespective of its character.

In the opinion of the authors it is incorrect to put in the clause `sample to sample'. This makes the specification appear to be one for the sample period. This is misleading and it is not correct. (At the time AES11 was produced the AES3 specification (AES3-

1985) had a less precise specification for signal timing which had several possible interpretations (ref. [] page 117)).

5.2. Frequency accuracy

The digital audio interface has been defined for operation over a range of sampling frequencies but devices are normally intended to operate within a limited range of a standard rate. This range is defined in AES5 [] and AES11 [] so that the tolerance of generators is ±10ppm and receivers are required to be able to lock to signals within ±50ppm. (Digital audio reference generators specified as "grade one" have tolerances and lock ranges of ±1ppm and ±2ppm respectively).

This tight accuracy requirement allows the use of narrow-band voltage controlled oscillators - such as those based on quartz crystals - which often have a capture range of less than ±100ppm.

5.3. Synchronisation

The synchronisation standard, AES11 [], defines synchronisation as being achieved when the timing difference between the output of a piece of equipment and its timing reference is within 5% of a sample period, and when all the input signals are timed within 25% of a sample period from the timing of the reference [Ref. , section 5.3.1]. This requires equipment to be designed so that digital audio output to sync timing variations are within the ±5% range - which is ±1042ns at a 48kHz sample rate.

Following this specification ensures that sample to sample timing relationships are maintained throughout a studio centre. While it recommends that devices have dedicated synchronisation inputs that allow all units to be timed to one common signal (the Digital Audio Reference Signal or DARS) there is a ratio of five between the output tolerance of 5% and the input tolerance of 25%. This suggests that a system with up to five successive devices synchronised in cascade (with timing errors all in the same direction) should still provide a output within the input offset tolerance of a later device that is synchronised to the same initial source. (In practice any differential delays through cables will also reduce this timing offset budget.)

6. INTERFACE CLOCK RECOVERY SCHEMES

A large number of different clock recovery techniques are in use. These range from the simple resonant `tank' circuit (commonly used in the telecoms industry) to all-digital phase- and length-locked-loops. Despite this variety we can identify two distinct classes of equipment, distinguished by the number of clock recovery stages that they incorporate.

To illustrate some of the compromises we will concentrate analogue PLL solutions. However many of the features are independent of the technology used.

The PLL characteristic most commonly discussed is the choice between a narrow and wide bandwidth PLL, and the effect of this on the output jitter of the device. The other important design considerations are jitter attenuation, jitter tolerance and susceptibility to pattern-dependent jitter. These are all related to the choice of bandwidth.

6.1. Single-stage clock recovery

The scheme used for this is shown in figure . The PLL reference is derived from data transitions in the interface signal, and the PLL output is used to clock the data into the bi-phase decoder or, if the signal does not need to be decoded, to re-clock the bitstream directly to the output. The oscillator used in the PLL may have a wide or a narrow frequency range.

A wide range oscillator, such as one based on a tuned circuit or a resistor-capacitor multi-vibrator, will typically be operated with a turn-over frequency of greater than 10kHz. This will have two benefits. The feedback around the PLL will attenuate the oscillator phase noise below that frequency and therefore reduce the intrinsic jitter. It will also allow the oscillator to track incoming interface jitter below that frequency and provide an increasing level of jitter tolerance for lower frequencies. At frequencies above the PLL roll-over frequency the jitter tolerance of the circuit will typically be ± 0.25 UI. The disadvantage of this approach is that if the oscillator is tracking the incoming jitter there is no jitter attenuation taking place. This means that jitter will accumulate with the number of devices in cascade.

A crystal oscillator with a narrow frequency range will be in a PLL with much less feedback and a roll-over frequency of less than a few kHz. This will reduce the jitter tolerance of the system to less than ±0.5UI at all but the lowest frequencies. The phase noise of a crystal oscillator is low and so the reduced feedback will not produce any significant intrinsic jitter at the output. Because of the lower corner frequency the jitter transfer function will now attenuate jitter to lower frequencies. This is an advantage for the recovery of sampling clocks in convertors - but unfortunately not for interface reliability.

With a single-stage clock recovery system reducing the turn-over frequency will not provide protection from jitter accumulation. The attenuation scales with the jitter tolerance so there will be no increased protection, other than (in the case of the VCXO) from the reduced intrinsic jitter. The intrinsic jitter level of either PLL type can be

swamped by the jitter introduced by the cable so this is not a significant advantage either. The key disadvantage of this approach is the narrow lock range. If the unit needs to operate at several sampling frequencies a separate VCXO will be required for each one.

The cost of wide-band or narrow-band receiver solutions is similarly low these two methods of implementation is similarly low. If an integrated receiver chip is used in the wide bandwidth solution, this may also provide features for buffering the audio and decoding the channel status. If a low jitter sampling clock has to be generated, for example in a digital to analogue converter, then the crystal-based PLL might provide a significant cost saving because the required jitter attenuation is provided in one stage.

6.2. Two-stage clock recovery

The two-stage architecture is shown in figure . The first stage is used to recover a clock that is used to decode the bi-phase mark data pattern. (There will be one of these for each receiver in the unit but for simplicity only one is shown). The second stage is used to generate the timing for the output(s). The first stage receiver PLL is selected to track jitter well without any requirement for jitter attenuation, while the second stage is not required to track incoming jitter and can be designed for jitter attenuation. The output PLL can be locked to either the recovered clock generated by the receiver PLL, or to some of the input signal transitions gated by the receiver logic.

One of the benefits of the two-PLL scheme was illustrated earlier in figure . This shows how the turn-over frequency of the jitter transfer function of the second PLL can be separated from the corner frequency of the jitter tolerance curve defined by the first stage PLL response. The increased vertical separation of the two lines can be increased so that for the flat portion of the JTF (where there is no jitter attenuation) the jitter tolerance can be several times better than the high frequency jitter tolerance. This will provide reliable operation in a system even when jitter accumulation has resulted in several UI of low-frequency jitter.

There are other benefits for this method: As mentioned in section a gated preamble edge from the input signal can be used to lock the output clock. This reduces the susceptibility to pattern-dependent jitter being passed on to the output. It is also possible to read the incoming data even if the sampling frequency is beyond the range of the output clock. The ancillary data decoded in this situation can be useful for diagnostic purposes.

This two stage clock recovery circuit is likely to be more expensive for the individual piece of equipment than the simpler one stage approach. This difference is small in equipment with several receivers and one output (or common output clock timing) because the second stage only occurs once in the system. However in cases where there are a large number of interface receivers each feeding a transmitter clocked from one receiver, such as distribution amplifiers and routing matrices the extra cost is significant - but the improved reliability with long cascades of equipment will often allow savings to be made in other areas.

7. THE TARGETS FOR A NEW JITTER SPECIFICATION

The following goals were considered while developing proposals for new jitter specifications.

7.1. The specified measurements should be unambiguous.

As discussed in section there are three key jitter parameters that affect performance. Intrinsic jitter, jitter transfer function and jitter tolerance. These all need to be specified to avoid ambiguity. (Pattern-dependent jitter susceptibility is relevant to jitter performance but not fundamental)

7.2. Consistent and reliable operation in the intended environment.

This includes operation over at least 100m of inexpensive twisted-pair cable. It was also considered that, like AES11, the specification should be designed to allow interface cascades (where the signal does not have to be re-synchronised to a DARS) for up to five devices with a jitter tolerance specification such that the final device in the chain will always be able to decode the signal. However the more stringent specification required for this may not be acceptable for all equipment.

7.3. It should be possible to derive good quality sampling clocks for ADCs and DACs from the interface signal.

As a result of masking effects the sampling jitter audibility threshold (figure) rises sharply for jitter frequencies below 500Hz and above that frequency it falls inversely with increasing frequency. This suggests that if total interface jitter could be controlled above 500Hz it would simplify the task of deriving adequate sampling clocks from the interface signal.

7.4. Total implementation costs should be low.

Any techniques required to conform to the specification should not result in significant cost increases compared with normal industry practice.

7.5. Well designed equipment currently in use should not be excluded.

There are interface designs in use with radically different jitter characteristics. These should not be excluded by the new specification.

7.6. Where possible measurements should use simple and/or currently available test equipment.

The choice of defined frequencies and levels should be based, where possible, on standard values on available jitter measuring equipment.

8. PROPOSED SPECIFICATIONS

8.1. Intrinsic jitter

The intrinsic jitter specification will define a limit for the peak to peak jitter amplitude present at the output of a piece of equipment that is free-running or locked to a jitter-free reference.

8.1.1 There is a case for this limit to have a weighting or template that allows the intrinsic jitter amplitude to increase below a specified frequency.

We have three reasons for this:

As mentioned before the jitter tolerance of practical receivers is such that more jitter can be tolerated at low jitter frequencies (below the receiver PLL cut-off frequency), when it can be tracked by the receiver PLL, than at high frequencies where it cannot. Therefore low frequency jitter can be allowed to increase to a higher level than that at higher frequencies.

The nature of the intrinsic jitter in continuous-time PLL designs is for the amplitude to reduce with increasing frequency (for frequencies above the PLL tracking cut-off frequency). Therefore tighter controls for higher frequency jitter are practical.

Any measurement of jitter needs to make timing comparisons with a relatively ideal, jitter-free, reference. This reference needs to have the same average frequency and phase as the signal under test. The simplest averaging circuit required to derive this already gives the measurement a similar weighting.

8.1.2 The intrinsic jitter turn-over frequency (below which jitter can increase linearly with decreasing frequency) could be 200Hz.

There are three factors contributing to the selection of this frequency:

The jitter of an inexpensive wide-range voltage controller oscillator (such as one that can lock to signals of 32, 44.1 and 48kHz sample rate) rises inversely with frequency. Selection of a lower turn-over frequency would make the specification more difficult to meet and measure using such oscillators.

The frequency is selected to be lower than the jitter tolerance turn-over frequency (specified later). This increases the safety margin between intrinsic jitter and jitter tolerance at lower frequencies where, in the absence of jitter attenuation, accumulation can be a significant problem. (see the discussion on jitter tolerance).

This frequency is one of the standard frequencies used in CCITT tests (recommendation G.823 table 1 []). Jitter measuring equipment designed for the telecommunications industry already has this frequency as an option.

(Please note that it is not enough just to specify the amplitude response of the filter. As the specification needs a peak to peak measurement the phase response of the weighting filter is also important.)

8.1.3 In the flat region (above the turnover frequency) the limit could be set to 0.04 UI pk-pk.

At 48kHz sampling frequency 0.04 UI is 6.5ns, which is significantly less than the overall jitter limit of 40ns peak to peak specified in AES3-1991. The reasons for this are as follows:

The AES3-1991 specification is for jitter at the interface output. Intrinsic jitter is only one contributor to output jitter.

To allow for jitter accumulation effects the specification for individual devices should be much less than the jitter tolerance limit for the receiver. This limit is one sixth of the proposed jitter tolerance at high frequencies. (That proposed specification is described later).

It has been shown that this limit can be easily achieved with very basic circuits and that typical equipment has intrinsic jitter well below this level [].

For most present devices, which are susceptible to cable induced jitter, there is little benefit in setting a very much lower figure for intrinsic jitter than the level of line induced jitter resulting from the use of the nominal 100m run of typical cable. Cable measurements shown in section have jitter of 0.02 to 0.06 UI (3 to 10ns).

8.1.4 With these considerations in mind we tentatively propose the following specification for consideration by the industry:

1. Intrinsic Jitter

The peak to peak intrinsic output jitter, measured at the half voltage points with the jitter weighting filter, shall be less than 0.04 UI.

Note: This applies both when the equipment is locked to an effectively jitter-free timing reference (which may be a modulated digital audio signal) and when the equipment is free-running.

The jitter weighting filter is shown in figure . It is a minimum- phase high pass filter with a 3dB frequency of 200Hz, a first order roll-off to 20Hz and with a pass-band gain of unity up to the maximum possible jitter frequency of 64 times the sampling frequency.

Measurements confirm that the vast majority of existing devices easily meet this specification.

8.2. Jitter peaking

This specification will define a limit for the amount by which jitter transferred from the timing reference input(s) of a device is amplified at the digital audio output(s). Possible reference sources might not need to be specified but include `word-clock' square wave and video signals.

The conventional approach is to avoid phase and non-linearity problems by concentrating on sinusoidal jitter [,]. Less conventional is the idea of specifying a flat limit independent of frequency. We have taken this path in the belief that we need, to start with, a simple specification that everyone can agree on. (An ancillary specification is presented in the next section).

The question of what value this limit should take has been a matter of debate amongst the authors. One argument is that the limit should not be so restrictive that it precludes the application of different technologies in the future. The majority view is that there is no good reason for equipment to be designed with jitter peaking of any more than 1dB - with the benefit of well controlled jitter accumulation. This is despite the fact that some existing equipment has peaking of more than 3dB.

With these points in mind we favour the following specification:

2. Jitter Gain

The sinusoidal jitter gain from any timing reference input(s) to the signal output(s) shall be less than 1dB at all frequencies.

8.3. Jitter transfer

This specification will indicate the minimum amount by which higher-frequency jitter on the timing reference input(s) of certain devices should be attenuated before appearing on the AES3 outputs. This attenuation is important in controlling jitter accumulation in systems.

Designers have already been building jitter attenuation (or `clock smoothing') into certain types of digital audio equipment. However the frequency at which the attenuation starts varies widely between different devices. If this frequency is too high then jitter accumulation is not adequately controlled. On the other hand, lower corner frequencies can mean slower lock-up and either higher intrinsic jitter, a restricted frequency range or higher cost.

What devices should be required to incorporate jitter attenuation ? At one extreme the attenuation function could be provided only where system jitter accumulation demanded it - with attenuator boxes being added to the installation as required. The opposite is to require it in all equipment. A compromise position is to allow for equipment with or without attenuation of a defined minimum form, and to require that the manufacturers specification state whether it is or is not provided.

The conventional way of framing this kind of specification is to require that the jitter transfer function falls below some defined mask [,]. This is the approach that we have used in the following tentative proposal:

3. Jitter Transfer Function

It is recommended that where jitter attenuation is provided it shall be such that the sinusoidal jitter gain, from any timing reference input(s) to the signal output(s), falls below the jitter transfer function mask. It is desirable that the equipment specification states whether the equipment does or does not have jitter attenuation to this specification.

(The mask is shown in figure . It imposes no additional limit on low frequency jitter gain. The limit starts at the input jitter frequency of 500Hz where it is 0dB, and falls to -6dB at and above 1kHz)

Having the attenuation start at 500Hz restricts jitter accumulation to below this frequency. Receivers can have good jitter tolerance at such low frequencies so this need not be a problem. At the same time this frequency is high enough to ensure that even the simplest types of wide-band VCOs can be used for the attenuation stage without intrinsic jitter problems.

The upper corner frequency, where the mask flattens out, is not critical. The response of practical jitter attenuators will continue to fall beyond this frequency but as well as being difficult to measure this is not a requirement for correct operation.

As in telecoms, the assumption is that the transfer function is measured with the input jitter at a reasonably high level.

We note two reservations about the above specification. One is that it does not define the input jitter stimulus levels, and the second is that by definition the jitter transfer function assumes a linear model. It may be more appropriate to measure the total output jitter that results from defined input jitter stimuli. This will include the combination of intrinsic jitter and both linear and non-linear jitter transfer from the input.

8.4. **Jitter tolerance**

This specification defines levels of jitter that an interface receiver must accept without producing errors. Quoting from the CCITT [,]:

> "the required tolerance is defined in terms of the amplitude and frequency of sinusoidal jitter which, when modulating a test pattern, should not cause any significant degradation in the operation of the equipment. It is important to recognise that the test signal is not, in itself, intended to be representative of the type of jitter to be found in practice in a network. However, the test does ensure that the Q factor associated with the timing signal recovery of the equipment's input circuitry is not excessive and, where necessary, that an adequate amount of buffer storage has been provided."

The jitter transfer specification of section would constrain the jitter accumulation problem to jitter frequencies below 500Hz. Receivers that are not sufficiently tolerant to jitter in this frequency region cannot be expected to provide robust reception in larger systems or where there is a lot of cable-induced pattern-dependent jitter.

Fortunately all the recent digital audio interface receiver chips can have good low-frequency jitter tolerance. (Receiver circuits that use VCXOs for the data recovery clock, or that use loops that have a low corner frequency for other reasons, do not have such good jitter tolerance at these frequencies.)

This specification also covers tolerance at higher frequencies where jitter is not tracked by the receiver. Typical decoding schemes have `theoretical' high-frequency tolerances of either 0.5 or 1 UI pk-pk (depending on the particular scheme). Real circuits only approach these levels of performance.

We tentatively submit the following proposal for discussion:

> *4. Jitter Tolerance*
>
> *An interface data receiver should correctly decode an incoming data stream with any sinusoidal jitter defined by the jitter tolerance template.*
>
> (The tolerance template is shown in figure . It requires a jitter tolerance of 0.25 UI pk-pk at high frequencies, increasing with the inverse of frequency below 8kHz to level off at 10 UI pk-pk below 200 Hz).

The low frequency limit is chosen for measurement convenience: CCITT Rec O.171 gives this level as a maximum level that jitter generators will be required to produce. The AES11 synchronisation standard requires that receivers have a phase lock range of at least $\pm 25\%$ of a sample period. This specification is not intended to overlap with AES11, but it should be noted that $\pm 25\%$ of a sample period is 64 UI pk-pk.

The intercept at 8kHz is selected to provide a large low-frequency tolerance without disqualifying current and future medium-Q receivers. Note that high-Q receivers, such as those that use VCXOs to generate the data recovery clock, are excluded by this choice of intercept frequency.

9. **DISCUSSION**

The draft specifications just presented should lead to consistent and reliable operation. It is reasonable to ask what the costs of implementation are, and whether it is possible to achieve the same objective with lower circuit complexity. In particular we need to ask whether jitter accumulation could reliably be constrained such that there is never any need for the additional jitter attenuation that is provided by the two stage clock recovery

architecture. If this were the case then the jitter attenuation specification could be removed and the jitter tolerance specification could be relaxed to allow single-stage VCXO based clock recovery circuits.

This hinges on the level of jitter accumulation that has to be catered for. The specifications provided here limit the jitter build up through intrinsic jitter and jitter gain but, as we have shown, the pattern-dependent jitter introduced by cable losses can swamp these levels. (It is possible to reduce susceptibility using the technique described in section but this also requires a separate clock recovery circuit so this will not produce a saving).

The choice is between an environment in which cable-induced jitter has to be carefully controlled, and one in which jitter accumulation is accommodated by more stringent specifications. The former requires a more jitter-sensitive operational practice - such as shorter or higher bandwidth cable runs - and the latter requires jitter-tolerant receivers.

Another consideration is the question of compatibility with existing equipment. Whilst this is a matter of philosophy to some it is of practical importance to the purchaser, user, and manufacturer.

The ad-hoc Task Group will welcome contributions from all parts of the audio community to this debate. The group coordinator, Julian Dunn, can be contacted at:

209 Mill Road,
Cambridge,
United Kingdom
CB1 3AN

Or by electronic mail to: 100014.1730@compuserve.com

Formal contributions should be made to the chairman of the AES Working Group on Digital Input/Output Interfacing (SC 2-2-2) at the AES international headquarters address.

10. ACKNOWLEDGMENTS

Thanks are due to Charles Meyer of Nvision and other members of the ad-hoc jitter task group who have contributed to the development of these draft proposals.

11. REFERENCES

[1] AES3-1992 - ` Recommended Practice for Digital Audio Engineering - Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data'
J. Audio Eng. Soc., vol 40 No. 3, pp 147-165 (June 1991)

[2] AES11-1991 - ` Recommended Practice for Digital Audio Engineering - Synchronisation of Digital Audio Equipment in Studio Operations' *J. Audio Eng. Soc. (Standards and Information Documents)*, vol 39, pp 155-162 (1991 March)

[3] AES5-1984 - `AES Recommended Practice for Professional Digital Audio Applications Employing Pulse Code Modulation - Preferred Sampling Frequencies' *J. Audio Eng. Soc. (Standards and Recommended Practices)*, vol 32, pp 781-785 (1984 Oct.)

[4] EBU Tech. 3250-E - `Specification of the Digital Audio Interface' EBU Technical Centre, Brussels, Belgium (1985 Nov.)

[5] IEC-958 - `Digital Audio Interface' First Edition, International Electrotechnical Commission, Geneva, Switzerland (1989 March)

[6] EIAJ CP-340 - `Digital Audio Interface', Standards of Electronic Industries Association of Japan, 1987

[7] Dunn, Julian. - `Jitter: Specification and Assessment in Digital Audio Equipment'. Preprint 3361, presented at the 93rd Convention, San Francisco, October 1992.

[8] CCITT Recommendation G.823 - `The Control of Jitter and Wander within Digital Networks which are Based on the 2048 kbit/s Hierarchy', Fascicle III.5, 1988

[9] CCITT Recommendation G.824 - `The Control of Jitter and Wander within Digital Networks which are Based on the 1544 kbit/s Hierarchy', Fascicle III.5, 1988

[10] CCITT Recommendation O.171 - `Timing Jitter Measuring Equipment for Digital Systems', Fascicle IV.4, 1988

[11] CCITT Supplement No. 3.8 - `Guidelines concerning the Measurement of Jitter', Fascicle IV.4, 1988

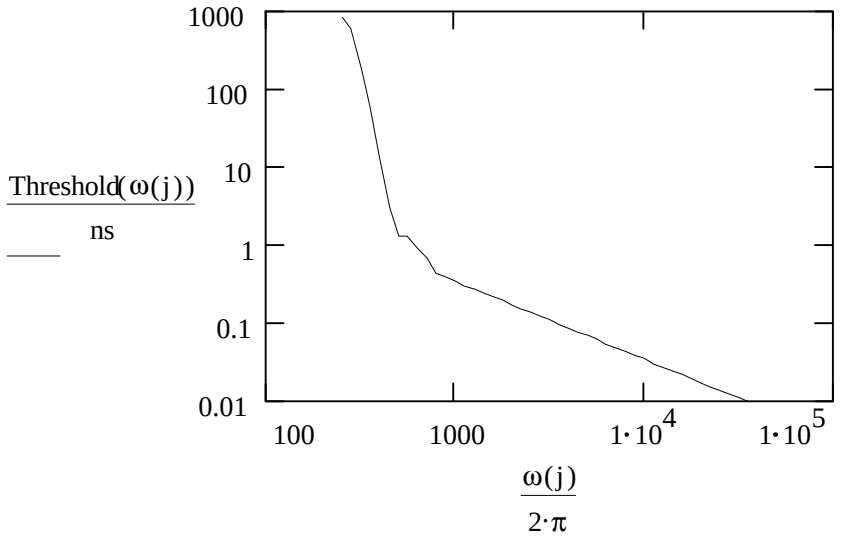
[12] Dunn, Chris and Hawksford, Malcolm Omar. - `Is the EBU/SPDIF Digital Audio Interface Flawed ?'. Preprint 3360, presented at the 93rd Convention, San Francisco, October 1992.

[13] Dunn, Julian. - `AES3 amendments to the jitter specification', letter to Chairman SC 2-2, March 23rd 1992.

[14] Dunn, Julian. - `Considerations for Interfacing Digital Audio Equipment to the Standards AES3, AES5, AES11'. Published in `Images of Audio', the *Proceedings of the 10th International Conference*, London, September 1991. pp 115-126.

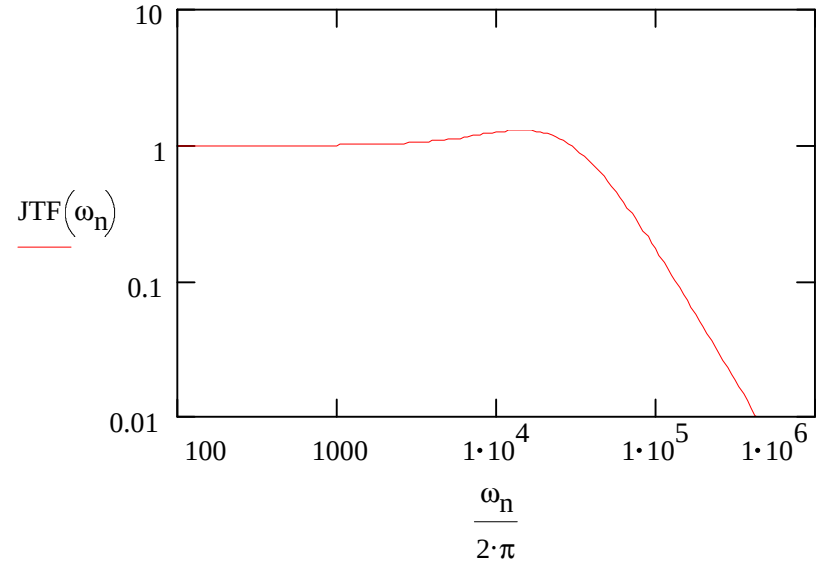
[15] Travis, Chris - `The AES/EBU Interface: Jitter Measurement Results'. Contribution to the ad-hoc jitter task group, June 1993

FIGURES



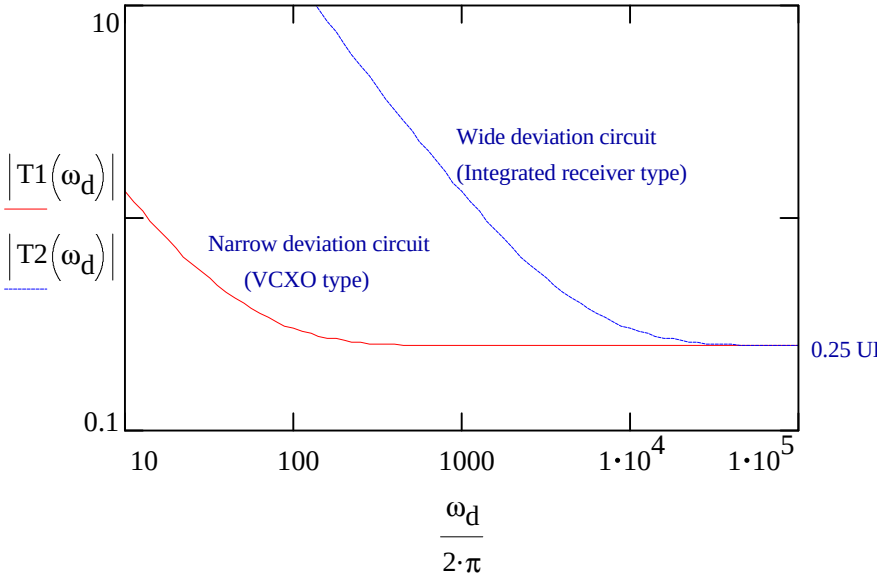
Maximum sinusoidal jitter amplitudes that will not produce audible modulation products

Figure 1



A typical jitter transfer function for a digital audio receiver-transmitter

Figure 2



Typical jitter tolerances (in UI) of receiver circuits capable of small and large frequency deviations

Figure 3

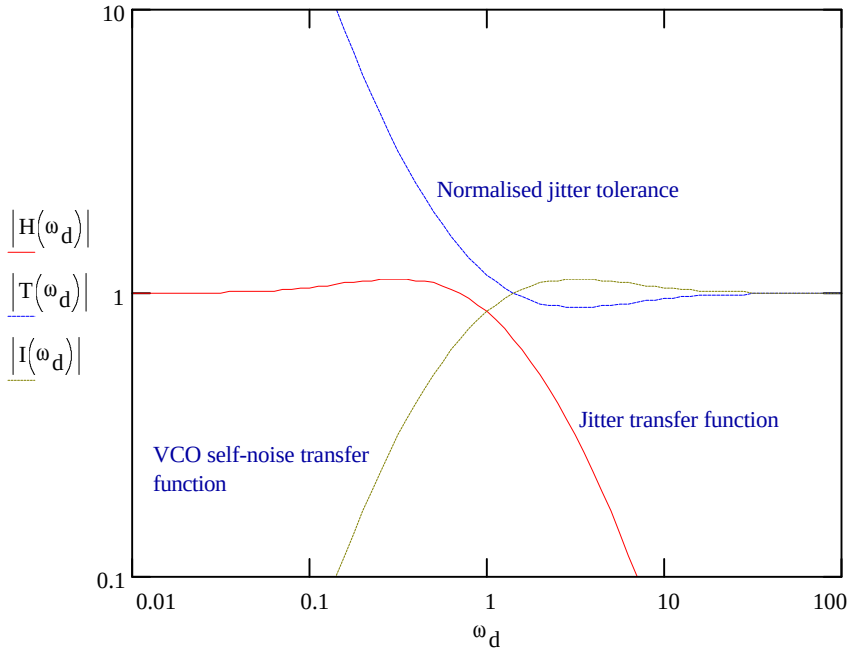
$$G(\omega) := \frac{1}{\omega^2} \cdot \left[\frac{1 + (j \cdot \omega \cdot 6)}{1 + \left(j \cdot \frac{\omega}{6} \right)} \right]$$ Normalised PLL open loop response

$$K := -\frac{1}{|G(1)|}$$ Normalisation factor

$$H(\omega) := \frac{K \cdot G(\omega)}{1 + K \cdot G(\omega)}$$ Jitter transfer function

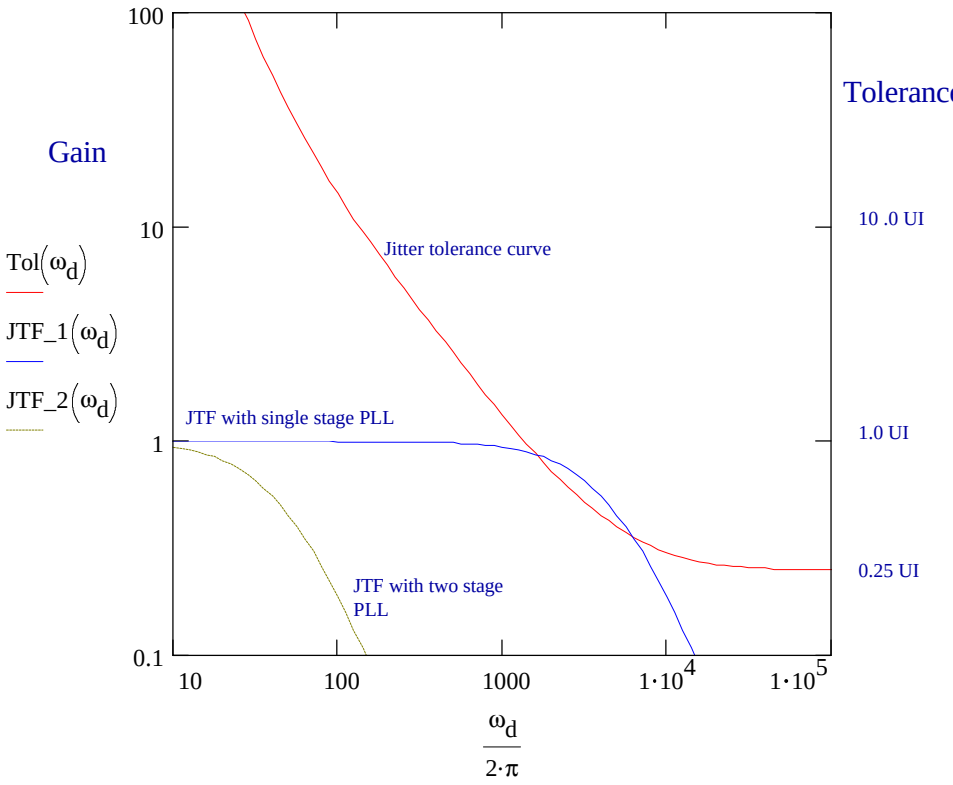
$$T(\omega) := 1 + K \cdot G(\omega)$$ Normalised jitter tolerance

$$I(\omega) := \frac{1}{1 + K \cdot G(\omega)}$$ VCO self noise transfer function



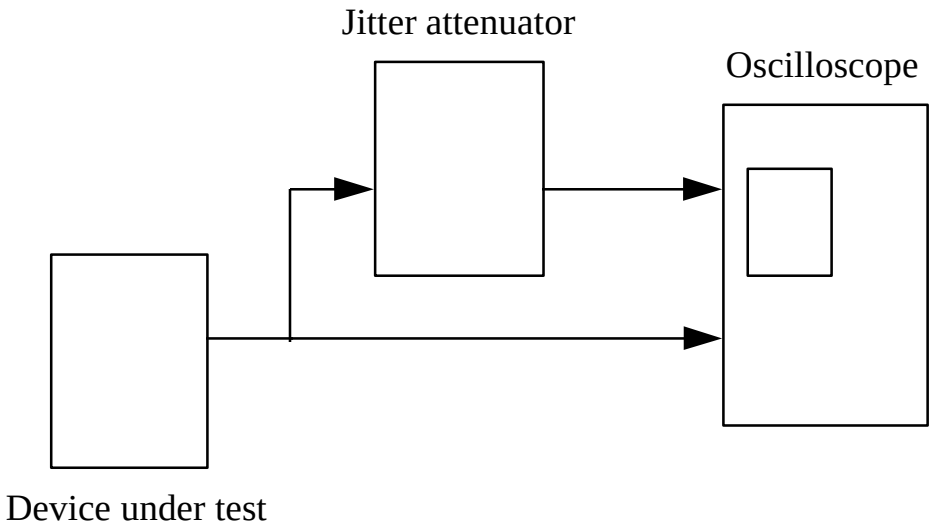
Jitter transfer function, normalised jitter tolerance and VCO self noise transfer function of a single 3rd order Phase Locked Loop recovery circuit (log scales)

Figure 4



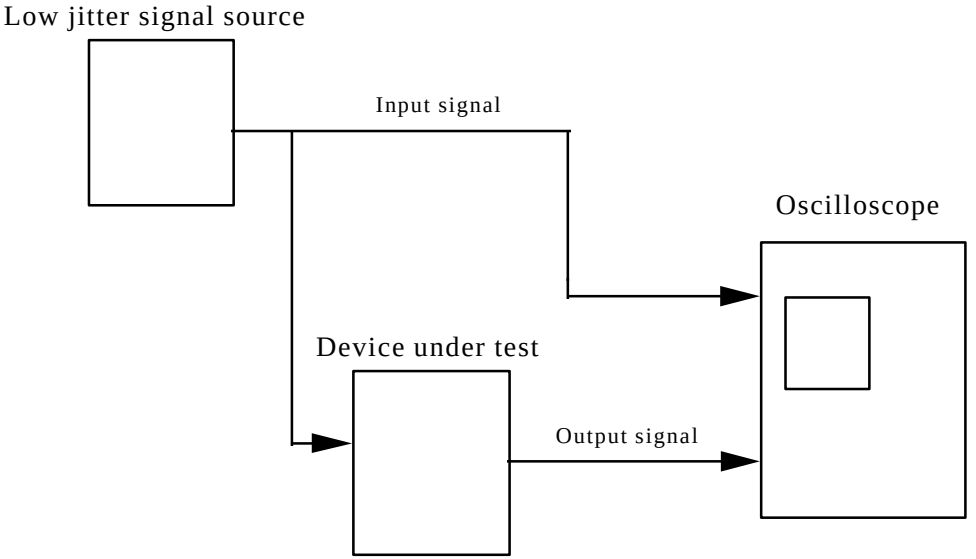
The jitter tolerance and transfer function of single and dual PLL architectures

Figure 9



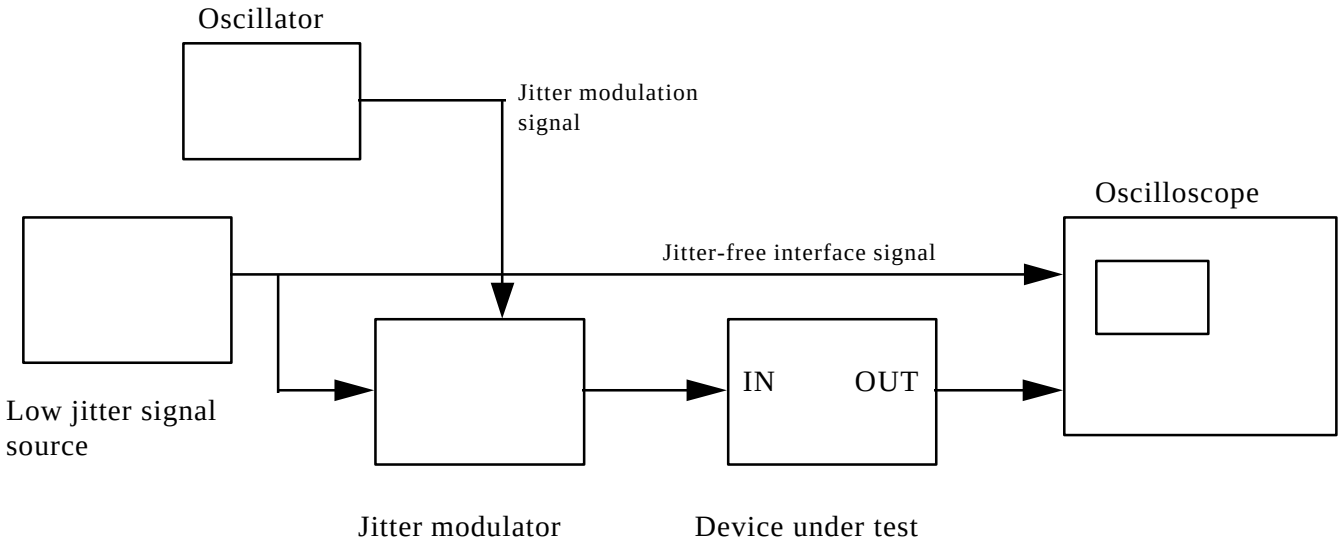
Intrinsic jitter measurement without an external reference

Figure 10



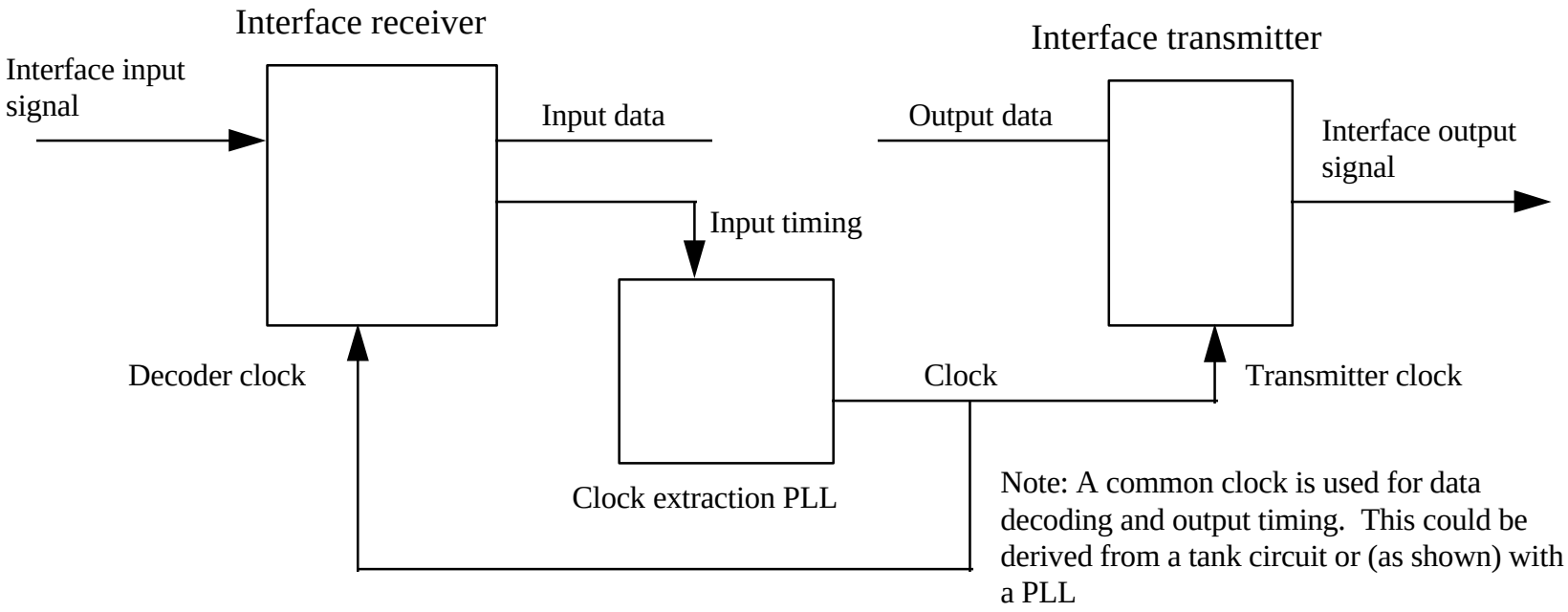
Intrinsic jitter measurement using a low jitter external reference

Figure 11

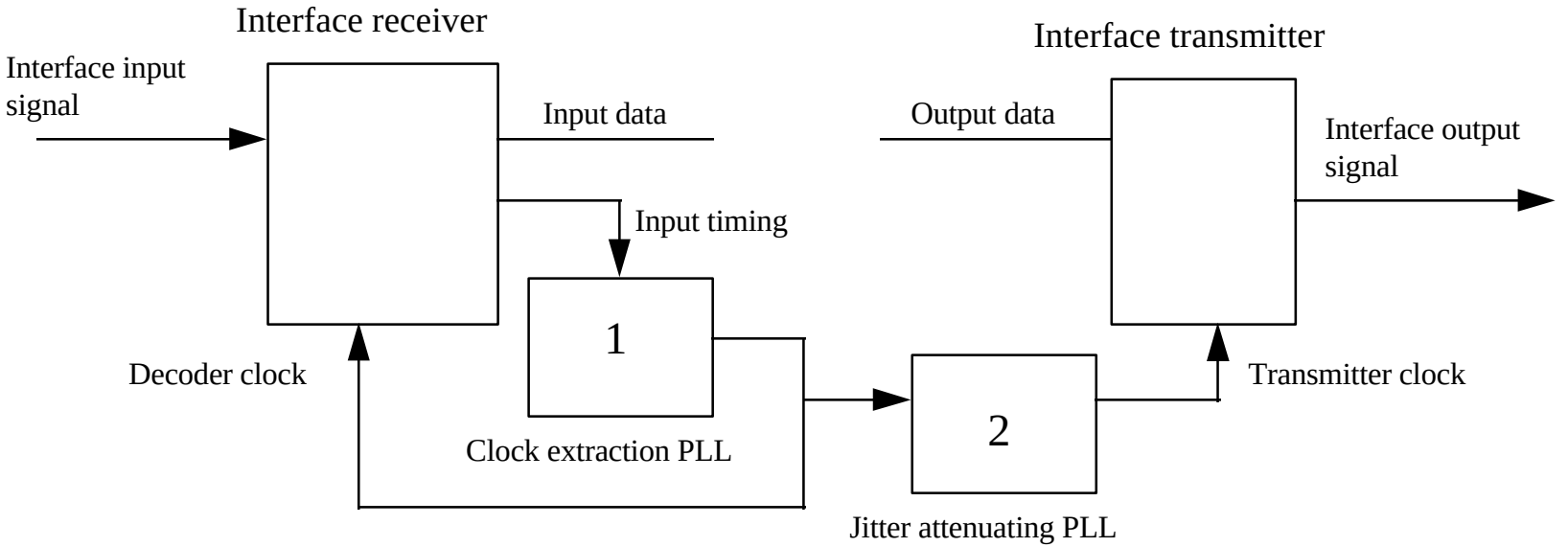


Jitter tolerance and transfer function measurement

Figure 12



Single clock AES3 transceiver architecture
Figure 13

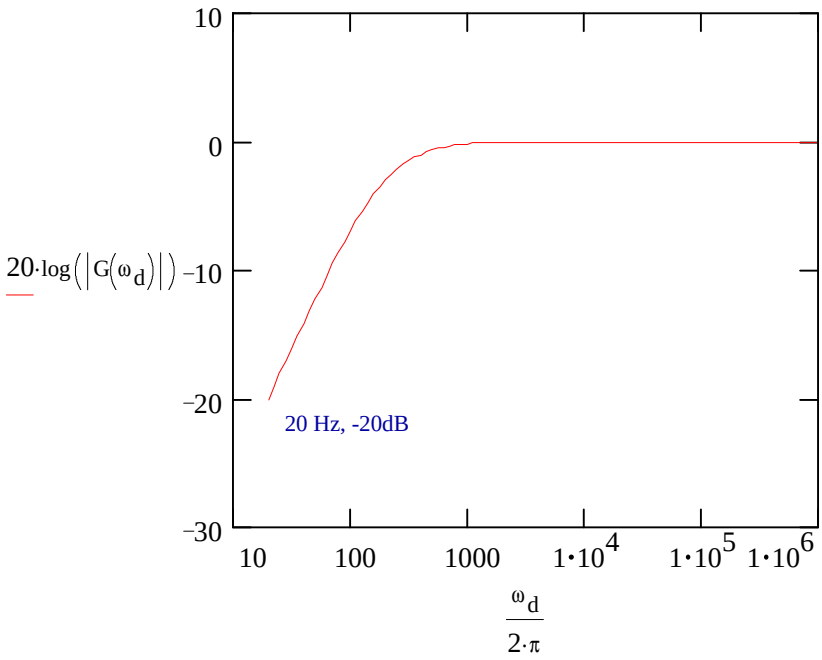


Dual clock AES3 transceiver architecture
Figure 14

$\omega 0 := 2 \cdot \pi \cdot 200$ Turnover frequency of 200Hz

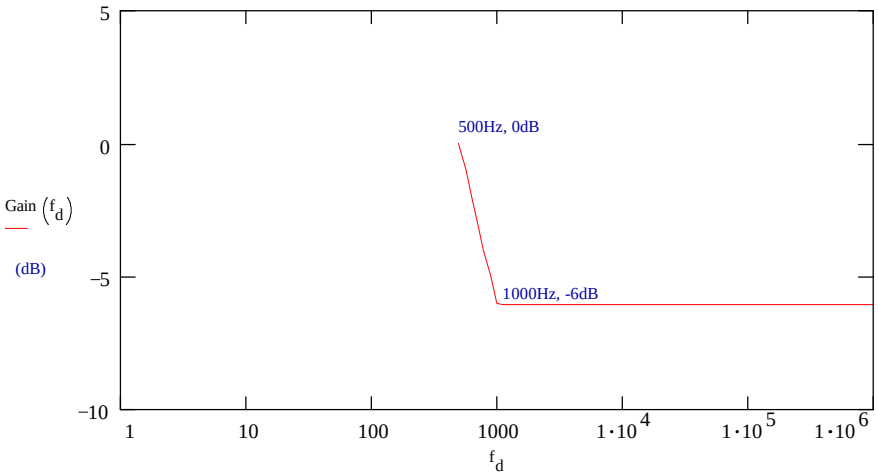
$$G(\omega) := \frac{j \cdot \frac{\omega}{\omega 0}}{1 + j \cdot \frac{\omega}{\omega 0}}$$ Filter definition

(The filter is only defined between 20Hz and 64 times the sample rate)



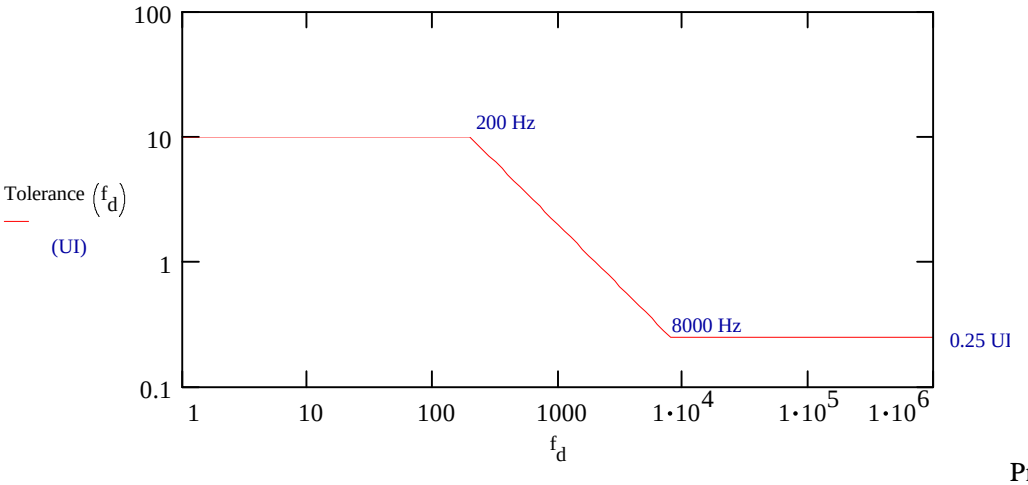
Proposed jitter measurement weighting filter

Figure 15



Proposed jitter transfer function mask

Figure 16



Proposed jitter tolerance template

Figure 17