Digital-to-analogue conversion using pulse width modulation

Pulse width modulation (PWM) has found application in various situations where efficient delivery of high power is required. In this paper its use in the conversion of signals from digital to analogue representations is addressed, i.e. as a digital-to-analogue converter (DAC), and it is shown that there are many advantages in using PWM over conventional DAC methods and that high resolution and accuracy can be achieved. One advantage is that, provided a suitable output stage can be provided, a complete, high-power, high-resolution digital-to-analogue power converter can be constructed. Because this work was motivated by the author's interest in digital audio, such a device has come to be known as a digital (power) amplifier. The article describes PWM, particularly in its digital forms, and shows how to make it practical for use as a DAC. This involves various digital signal processing stages prior to the final modulator. Results are presented from simulations and from demonstrator hardware.

by M.B. Sandler

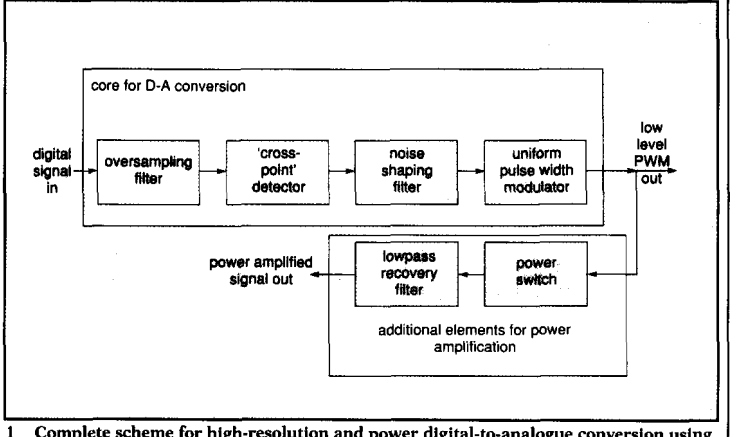
1 Introduction

Digital audio components are now commonly found in living rooms, broadcast studios and even cars. These include compact disc (CD) players, digital audio tape (DAT) and other digital tape recorders, room equalisers, delay lines and so on. In fact, virtually any processing stage which would correspond to channel encoding/decoding or the channel itself in a communications system can be, or has been, implemented with digital technology. However, even now, the source encoder/decoder stages (microphones, power amplifiers and loudspeakers) still normally use analogue technology. The motivation behind the work described here was to find a digital alternative to analogue power amplification. Along the way a new technique for digital-to-analogue conversion (DAC) was developed.

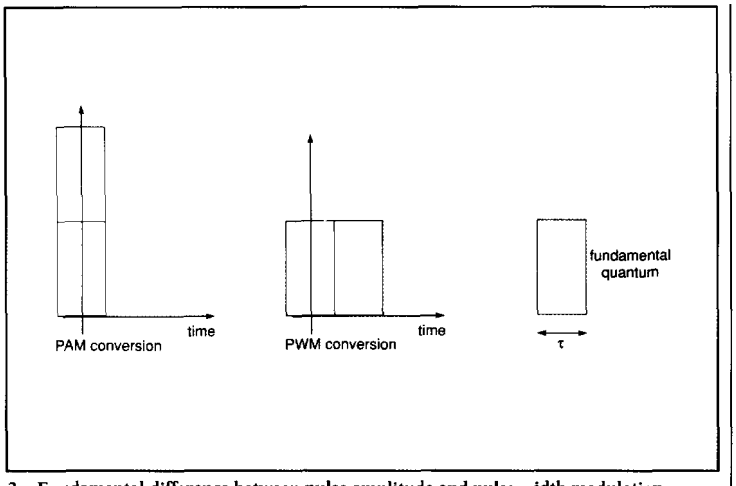
The overall scheme in which the core conversion process is pulse width modulation (PWM) is shown in Fig. 1. This Figure also shows the additional components needed to turn the PWM DAC into a power amplifier. The digital input is first processed by an 'oversampling filter' (see Section 3), which effectively increases the sampling rate (for example from 44·1 kHz to 325·8 kHz). The primary purpose of this stage is to help reduce distortion. The 'cross-point detector' is a lineariser which improves the distortion performance of the system, as will be described in Section 5. Without it performance up to about the 16 bit level is possible: with it, we expect performance to 20 bits, and quite probably more. The output from the uniform pulse width modulator is a low-level PWM signal which can be used just like the output of any other DAC (after suitable filtering). Alternatively, as shown, the low-level pulse width modulator can be used to control a class D power switching stage from which an analogue power signal is recovered by linear, passive, lowpass filtering.

Note that no assumption is made about how the digital input signal is obtained. For audio, it would probably be from a digital source such as a compact disc (CD) or digital audio tape (DAT) player; equally, it could be from an analogue source after passing through a high-quality analogue-to-digital converter (ADC), but this is not of relevance to the concepts discussed in this paper. In other applications, such as stepper motor control, the digital signal would come directly from a controlling computer.

This paper commences with a discussion of two of the three fundamental building blocks from which a PWM DAC is constructed: an analogue pulse width modulator is discussed in



1 Complete scheme for high-resolution and power digital-to-analogue conversion using PWM



2 Fundamental difference between pulse amplitude and pulse width modulation

Section 2; sample rate increase (interpolation) and oversampled noise shaping is discussed in Section 3. Section 4 then describes briefly how a real pulse width modulator for digital-to-analogue conversion works and how it differs from the conceptual model of Section 2. Section 5 describes a novelty of the approach taken at King's College London to approximate a more benign form of PWM with better distortion characteristics and thereby to push performance to that theoretically achievable using 20-bit or more samples. Finally, Sections 6, 7 and 8, respectively, deal with the application of the techniques described to power digital-to-analogue conversion (a digital power amplifier), other potential applications of the technique and a summary.

2 Pulse width modulation

This Section aims to give sufficient background in pulse width modulation for the reader to understand how the PWM DAC works. It does this by highlighting the difference between PWM and pulse amplitude modulation (PAM) and by presenting an analogue version of PWM, which although not used in a PWM DAC illustrates the principles more clearly. Then some of the variants of PWM are discussed and compared.

Originally, digital-to-analogue converters all used pulse amplitude modulation as their means of conversion: a sample represented as a binary number is converted into a narrow pulse whose height is proportional to the sample value. This is achieved either by dividing down a voltage by a resistive ladder network or by building up a current which is fed into a load resistor.

What are the problems with these approaches? For high-resolution conversion, the resistor ladder or current source networks contain many components which must be tightly matched to a high tolerance; this is costly to achieve. Often the perfection of the PAM conversion process is compromised by static nonlinearities (e.g. nonmatching of resistors or currents) or by 'missing codes' (where a binary value converts to zero or perhaps to the same voltage as the next binary count up or down). In performance terms, the data conversion is no longer ideal and suffers from distortion. From a systems points of view, we might say that PAM converters run into problems because the conversion is performed by digital control of an analogue process.

PWM DACs differ from this as there is no analogue process and the conversion is performed by a counting procedure*. There is no longer the possibility of missing codes and a complete converter may be constructed from digital logic components alone, save for a final passive lowpass filter.

Fig. 2 illustrates the essential difference between PAM and PWM converters. In each case we can imagine that a signal sample of one least significant bit (LSB) in value gives rise to a pulse of unit amplitude and width — we will call this a fundamental quantum. In PAM, for a sample of two LSBs, the second fundamental quantum is stacked on top of the first, whereas in the case of PWM it is positioned alongside the first. It can be seen that this gives different spectral properties for the two schemes, since PAM is the sum of time-coincident fundamental quanta, whereas PWM is the sum of time-shifted fundamental quanta. The difference is easily demonstrated by taking Fourier transforms.

Fig. 3 shows in principle how pulse width modulation is performed. The core of the circuit is a comparator (earlier it was said that the key process was counting — these two statements will be reconciled shortly). To one input is fed a high-frequency sawtooth or triangular waveform, $c(t)$: to the other is fed the message signal, $i(t)$ or $i_S(t)$. Note that the Figure embellishes this basic system by adding the option that the message signal can also be sampled and fed through a zero-order hold (producing $i_S(t)$) before it appears at the comparator's input. The sampling rate is the frequency of the sawtooth comparison waveform. This demonstrates the difference between the two major categories of PWM: naturally sampled PWM (NPWM) and uniformly sampled PWM (UPWM). The former is the case where no sampling occurs prior to the comparator. Other categorisations of PWM are also important, but first the difference between NPWM and UPWM will be explained. It is only fair to say that the output as shown is an artist's impression of a true output for the message signal shown.

This paper considers the conversion of digital signals, i.e. those which have undergone a uniform sampling process, and it takes only a little thought to realise that no digital version of PWM can use anything other than UPWM in the final modulator. However, there are advantages to NPWM and a major achievement of the work undertaken by the author and various co-workers is to have found a reasonably computationally efficient way to emulate the behaviour of NPWM even though the core converter performs UPWM. This is discussed in Section 5.

Although the output waveforms are probably indistinguishable on an oscilloscope, UPWM suffers from harmonic distortion whereas NPWM does not. Full details of the tone modulation spectra for both classes of modulation can be found in many texts.[1-3] It is found that the

*This is, of course, a moot point as the counting process is analogue in the time domain and suffers from time resolution problems due to finite switching speeds (analogue effects) in digital circuitry.

harmonic distortion levels for various classes of UPWM can be quite high.

Both types of PWM suffer from another distortion mechanism, namely intermodulation between the carrier (or comparison) waveform and the signal. Both distortion mechanisms are dependent on the relative frequencies of message and carrier/comparison waveforms and it has been shown[2-4] that, if the ratio $q = \frac{\text{(carrier frequency)}}{\text{(signal frequency)}}$ is made large enough, the harmonic distortion can be made small enough to be considered negligible. Likewise, the carrier intermodulation distortion can also be made negligible. To increase the ratio for a sampled, digital signal requires the use of an interpolation or oversampling filter[5] and the question of how much oversampling is required and what effect this has on implementations later will be considered later.

Other categorisations of PWM will now be considered. In Fig. 3 the pulse waveform is unipolar, that is it swings between $\pm V$ volts or between $+V$ volts and zero. Such unipolar waveform types have been named class AD.[6] There is also a three-level pulse waveform variant known as class BD, but since it is hard to build output stages for such modulators, they will not be considered further here.

Fig. 3 depicts a sawtooth comparison waveform which will modulate the trailing edge of each pulse. There is the option of modulating instead the leading edge or both edges of a pulse. Leading and trailing edge (i.e. single-sided) modulation are effectively the same thing and the choice between them boils down to convenience of implementation. All one-sided modulation schemes perform worse than two-sided schemes, all other modulation parameters being equal, i.e. one-sided NPWM is worse than two-sided NPWM and one-sided UPWM is worse than two-sided UPWM.

In UPWM, there are two types of two-sided modulation: that in which a single sample modulates a pulse, varying its width equally either side of a central regular timing marker, and that in which consecutive samples modulate each edge. In the latter case the sampling frequency is twice the carrier or pulse repetition frequency (PRF). At the same PRF, the version with two samples per pulse performs better, and only suffers from odd-order harmonic distortion — one-sample-per-pulse and one-sided modulation schemes suffer from odd and even order distortion.

Although there are certain subtleties of the modulation schemes which we will not deal with here (see References 2, 7 and 8, for example), it is possible to make some strong generalisations and draw conclusions. One is that no one-sided UPWM scheme is suitable for digital-to-analogue conversion unless q is made large, which is normally impractical, depending on the signal's bandwidth.

From many investigations over recent years, the author feels confident that a good conversion system for 16 (or fewer) bit digital signals can be constructed from either form of two-sided UPWM. There will be some residual distortion, but, depending on the signal frequency and amplitude the level of this will rarely be more than 100 dB below peak signal level. This would be at a minimum value of q (i.e. q for the highest signal frequency) of 16, which corresponds to increasing the sampling rate eight times. This has been confirmed in hardware implementations (see Fig. 8) in which although harmonic distortion is present it is at levels commensurate with that of PAM digital-to-analogue converters. If performance better than this is required, either q has to be increased further or some version of NPWM must be used.

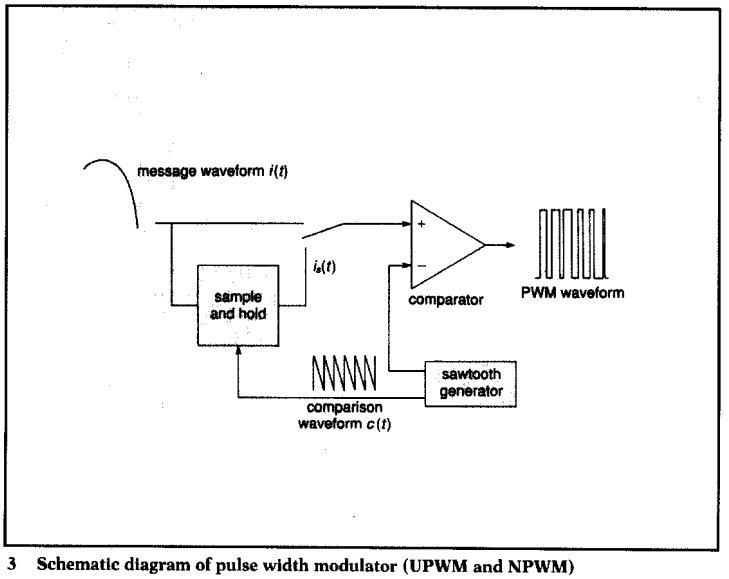
For a NPWM system, adequate baseband performance for conversion of 16 bit signals can be obtained with a one-sided modulation scheme and an oversampling ratio of between four and eight times. For a two-sided scheme the oversampling can be lower, or the signal resolution in bits can be higher. This conclusion has led the author's research team to consider ways in which a UPWM system can be converted to approximate NPWM to provide effectively distortion-free conversion. This is discussed in Section 5.

3 Interpolation and noise shaping

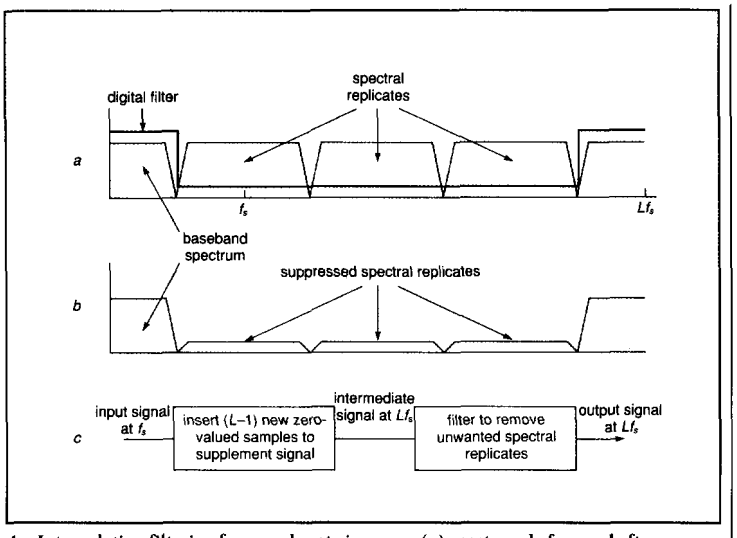
The need to increase the signal's sampling rate in order to reduce distortion has already been mentioned. This is achieved by interpolation filtering,[5] a technique that is now well understood. However, this incurs a penalty in that the modulator itself (as described in the following Section) can not readily be constructed in everyday technologies. Thus this Section also deals with oversampled noise shaping, which enables the modulator to be constructed.

The interpolation process is summarised in Fig. 4. These days, and in the hardware mentioned later, it is usual to use a commercial integrated circuit for this function and these are available in stereo versions from many major vendors.

The spectrum of the message signal after sampling at frequency f_s is shown by the coloured curve in Fig. 4*a* and comprises repeats of the baseband signal around multiples of the sampling frequency.



3 Schematic diagram of pulse width modulator (UPWM and NPWM)



4 Interpolation filtering for sample rate increase: (*a*) spectrum before and after inserting zero-valued samples but before filtering; (*b*) spectrum after filtering; (*c*) block diagram

In interpolation filtering, $(L-1)$ zero-valued, equally spaced samples are inserted between each pair of the original samples, thus increasing the sample rate to Lf_s. The sampled waveform and its spectrum are, of course, left unchanged by this process.

The aim of interpolation filtering is to make the output signal approximate as closely as possible the digital signal which would have been obtained had sampling actually been performed at Lf_s. Therefore to achieve this, all the baseband spectral replicates, as marked in Fig. 4*a*, are removed by a standard digital filtering process. Normally this is implemented by a finite impulse response (FIR) digital filter, but it is possible to use infinite impulse response (IIR) filters too. The filtered spectrum is shown in Fig. 4*b*.

The stopband suppression should be of the same order as the required signal dynamic range. For example, for a 16 bit signal, approximately 100 dB of attenuation would be required. Actually this last point is currently under debate in the audio industry, because the more extreme the filtering, the longer the impulse response of the FIR filter and the longer the 'pre-ringing' added to the signal. Thus, some people believe that the attenuation should be no more than 30–40 dB. However, this is not the place to debate the psychoacoustics of interpolation filtering; in any case, if the final DAC is nonlinear (as all are because of circuit imperfections and PWM DACs are by definition) the excess noise left outside the baseband will be mixed down into the baseband and compromise the signal-to-noise ratio.

Fig. 4*c* shows the interpolation process as a simple signal-flow block diagram.

A problem now arises. Assume (quite reasonably as it turns out) that an eight times increase in sampling frequency will give an acceptable level of performance for 16 bit signals with a two-sided UPWM modulation scheme. If the audio sampling frequency is 44 100 Hz, this is going to be raised to 8 x 44 100 = 352 800 Hz, i.e. one pulse every 2·8 μs. There are 2^{16} (=65 536) different possible pulse lengths, so the clock controlling the modulator must operate at 352 800 x 65 636 Hz ≈ 23 GHz, i.e. a time resolution of 43 ps pulse duration! This is on the very edges of achievability with modern technology and is certainly not mass-producible.

Consider Fig. 4*b*, in which the frequency occupancy of the now oversampled signal is shown. This represents a channel in which much of the frequency allocation is empty. Where there is signal (at the lowest end of the spectrum) it is represented to high accuracy, namely 16 bits; however, there is quantisation noise at the 16 bit level throughout the channel, even though the amount of noise outside the baseband is of no consequence. Effectively we have far more channel capacity than is being used. Is there a coding technique which will reduce the channel capacity so that it matches that of a realisable pulse width modulator but at the same time will not compromise the low-frequency resolution? There is, and it is called 'noise shaping' (alternatively, 'error spectral shaping' or 'sigma-delta modulation').

A noise-shaping filter structure is shown in Fig. 5a. The oversampled 16 bit data input is truncated by the quantiser Q after it has been added to the feedback error from previous truncations which has been passed through the shaping filter $H_{ns}(z)$. The operation of such a nonlinear system has been much investigated over recent years, as described in Reference 9 or more completely in Reference 10. The term 'extra noise' will be used here to describe the error added to the signal by the in-loop quantiser, Q. Without the feedback loop, this noise will be at a frequency-independent level determined by the number of remaining bits. Thus if 8 bits remain after truncation, the 'extra noise' level will be about 50 dB below peak signal level. The term 'excess noise' will also be used to refer to the 'extra noise' after it has been filtered by the action of the loop. The aim is to make the 'excess noise' lower than the 'extra noise' over the band occupied by the signal. Outside this band the 'excess noise' is greater than the 'extra noise', but it is assumed that this can be removed or that it does not affect the desired overall system response.

A simple analysis (see Reference 9, for example), facilitated by the linearising assumption that the quantisation error is noise-like and additive (see Fig. 5*b*), demonstrates that the signal and additive noise see different paths through the filter. Effectively the signal passes straight through the forward path and only the 'extra noise' passes around the loop to be filtered.

Fig. 6 depicts the typical output spectrum of a noise shaper being driven with a sinewave. It can be seen that the 'excess noise' has been shaped so that it is squeezed out of the wanted band (low frequencies) and appears (amplified) at higher frequencies not occupied by signal. In the example, the filter used was a particularly simple type for which the noise transfer function $(1-H_{ns})$ is a cascade of four differentiators. Such a scheme is known to maximise the overall signal-to-noise ratio across the whole bandwidth. It does not however maximise the signal-to-noise ratio in the baseband.

In the simplest possible form, $H_{ns} = z^{-1}$, i.e. $(1-H_{ns})$ is a first order differentiator. This has a gain at zero frequency of zero rising to a maximum value of 2 at the half sampling frequency. However, the bandwidth over which the 'excess noise' is smaller than the 'extra noise' will be small relative to the sampling frequency and this leads to a requirement for a high oversampling ratio.

To overcome this problem, higher order filters are used and the one used to produce Fig. 6 was 4th order:

$$(1-H_{ns})=(1-z^{-1})^4$$

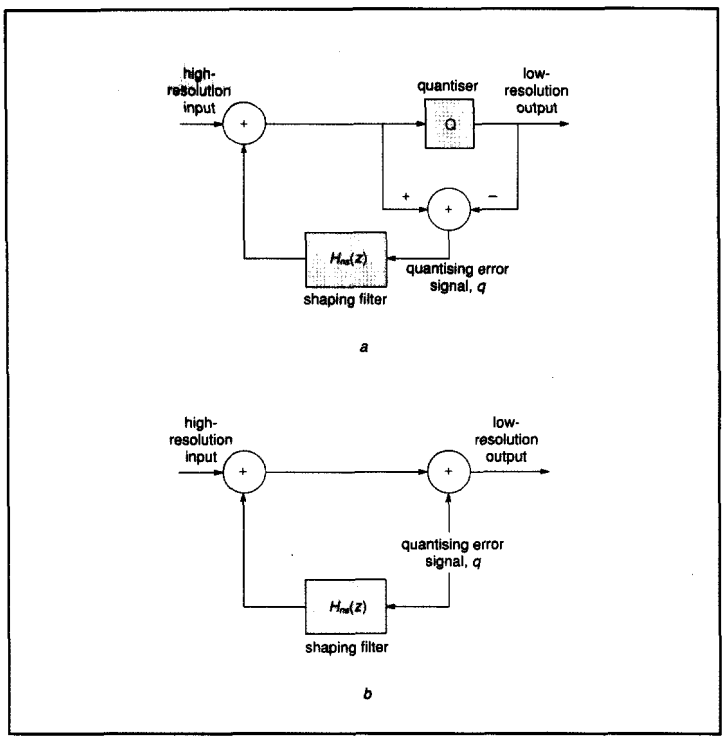
This increases the active bandwidth over which the 'excess noise' is smaller than the 'extra noise', but at the cost of pushing more noise into the higher frequency regions of the bandwidth.

The application of noise shaping has solved the realisability problem of the modulator. If noise shaping is applied to reduce the wordlength to 8 bits, the frequency of the clock needed to drive the modulator is now just $352\,800 \times 2^8$ Hz $\approx$ 90·3 MHz.

In principle it is possible to use more extreme forms of noise shaping to reduce the wordlength, and thus the clock rate, still further. However, this means that the noise shaper has to work harder to move the noise from the baseband. Not only does this incur a greater real-time processing overhead in the noise-shaping filter and a further degree of difficulty in its optimal design, but the extra noise power dumped at high frequencies can intermodulate with the carrier and 'fold back' into the baseband to diminish the signal-to-noise ratio.

This becomes a distinct problem when conversion to accuracies greater than 16 bits is required, unless the base-bandwidth is significantly less than that of audio signals, in which case more oversampling (more space to put the excess noise) can be tolerated. This problem can be solved by the use of noise shaper filters which satisfy a minimum phase requirement[11] and have been designed according to optimality criteria.

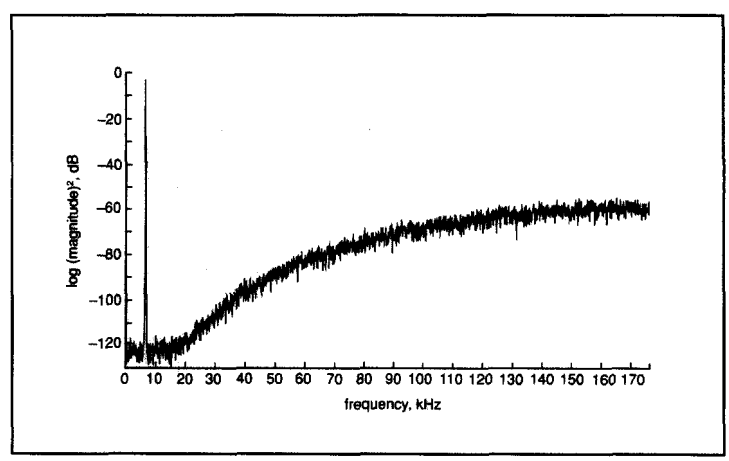
Fig. 7 is a compilation of some of the best results obtainable from a class AD UPWM functioning as a 16 bit digital-to-analogue converter. These were obtained from a simulation suite which has been developed over the course of the research. The signal is a



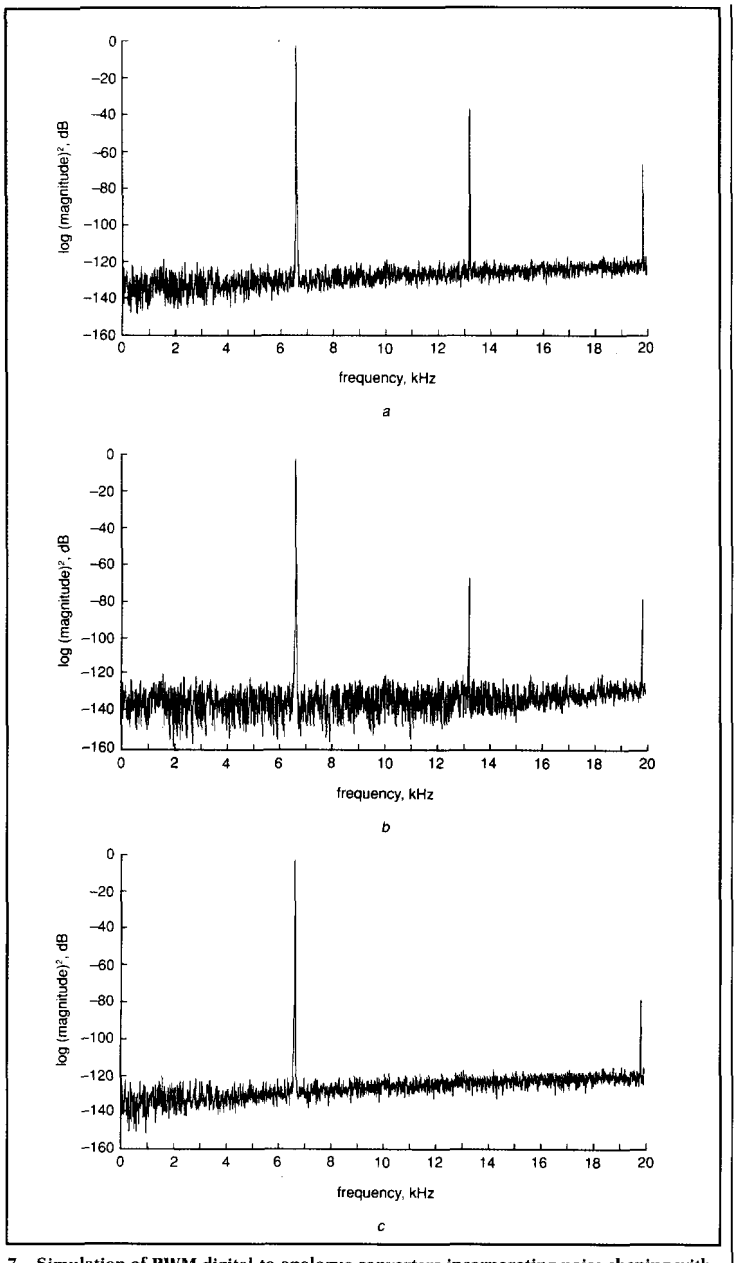
5 (*a*) Noise shaping filter schematic diagram; (*b*) linearised model of noise shaper as used in analysis and design of $H_{ns}(z)$

6·6 kHz sine wave at 3 dB below maximum level and the PRF is eight times the original sampling rate. The frequency is chosen to be the highest which generates a third harmonic inside the audio bandwidth. The analysis is performed with an 8192-point FFT and a special analysis window with low sidelobes.

The trailing edge modulation scheme (Fig. 7a) shows distortion at 13·2 and 19·8 kHz at quite high levels; this will later be compared with the performance of the enhanced scheme of Fig. 1. In Fig. 7*b* the result of using two-sided modulation is shown. Here there is still one sample per pulse but now the pulses are varied by moving each edge equally in opposite directions about a regularly spaced centre. Notice that both distortion terms have fallen dramatically — the second harmonic by about 30 dB and the third by about 10 dB. From Fig. 7*c* it can be seen that the second



6 Spectral performance of 4th order 'Tewksbury' noise transfer function



7 Simulation of PWM digital-to-analogue converters incorporating noise shaping with 6·6 kHz sinewave input at −3 dB, 8 times oversampling, and noise shaping 16 bits down to 8: (*a*) trailing edge, one-sided modulation; (*b*) double-sided, one sample per pulse modulation; (*c*) double-sided, two samples per pulse, one sample per edge, modulation

harmonic can be removed completely if the two-sided scheme is modified so that a different sample modulates each edge of a pulse (the King's College team call it 'two sample consecutive modulation').

Although the remaining distortion in Figs. 7*b* and *c* appears to be of quite a high level, it must be borne in mind that this is the worst case. As the signal level or frequency reduces, the distortion reduces. For example, if the signal frequency is halved the third harmonic of Fig. 7*c* would reduce by 12 dB. At higher signal frequencies the distortion increases, but there is an ameliorating effect in that, at least for audio signals, the signal energy reduces at higher frequencies and also the ear is less sensitive at higher frequencies, so that distortion arising from, say, 5 kHz components in a signal is unlikely to be psychoacoustically detectable.

All three plots in Fig. 7 were made by using noise shaping to reduce the resolution required in the final modulator to 8 bits. The clue that noise shaping has been used is the noise floor rising with frequency. For a more detailed discussion of the performance of these systems, see Reference 12.

4 How the modulator works

Section 3 discussed pulse width modulation in terms of an analogue equivalent. It is now necessary to highlight how a real PWM DAC is implemented as a digital pulse width modulator.

In realising the modulator of Fig. 3 using digital technology, the comparison waveform generator becomes a counter (up, down or up/down, depending on whether one or two-sided modulation is used) which cycles through its counts, and the comparator is replaced by a digital equivalent. However, it is more efficient to load a counter directly with the numerical sample value and then to count down and detect when an all-zero count is reached. In this case the comparator is replaced by a counter.

For two-sided modulation two counters are required, one counting the pulse duration and another counting a delay before the start of the pulse. This basic principle holds for all the forms of modulation: if one-sided modulation is needed the delay counter is set to zero; for two-sided, one sample per pulse modulation, the delay is calculated so as to centre the pulse in its period; and for two samples per pulse, the delay is calculated as a simple function of the two consecutive samples.

Fig. 8 presents a summary of the best currently available results obtained from a hardware demonstrator system that has been constructed. The system includes a compact disc player, an oversampling filter, a master clock, a noise shaper (application-specific IC or software on a digital signal processor) and a modulator constructed out of TTL components on a 4-layer circuit board. The master clock also drives the CD player to avoid synchronisation problems (including jitter) and has a frequency of just over 101 MHz. It is this high (rather than 90 MHz as earlier) to allow for 'guard bands',

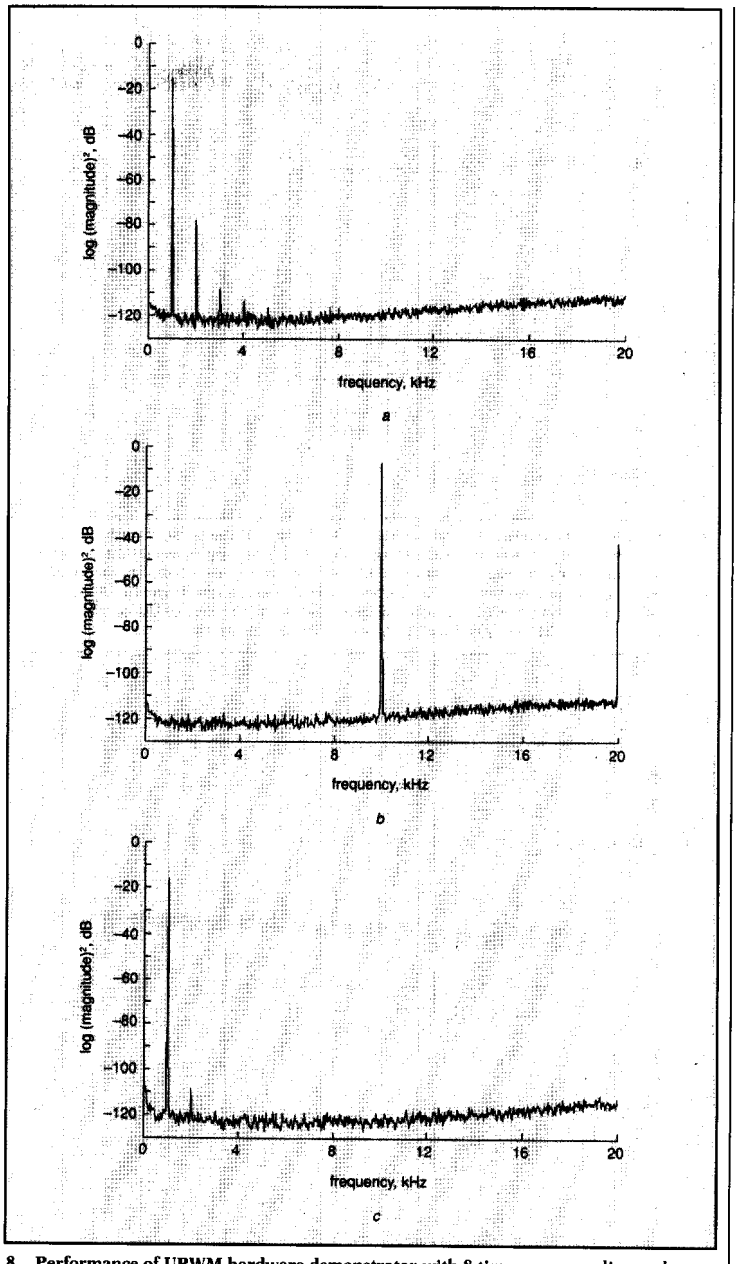
the name given to mandatory 'space' between consecutive pulses so they do not run into one another. All hardware tests are for an original sinewave sampled at 44·1 kHz and obtained from a commercial test disc. The sampled signal is interpolated to 352·8 kHz, 4th order noise-shaped in discrete hardware, modulated and finally filtered at TTL signal levels. This process is similar to that of Fig. 1 with the 'cross-point detector' block bypassed.

Fig. 8*a* shows what happens when a 1 kHz tone at 15 dB below peak is converted through a one-sided modulator. This performance is unacceptable in terms of distortion, but notice that the noise floor is close to the 16 bit level: noise shaping works! In Fig. 8*b* the signal frequency and amplitude have both been increased (to 10 kHz and −6 dB, respectively); only one distortion term is visible in the audio band and its level is increased. Fig. 8*c* shows the best the team has yet achieved in hardware. This is using the two-sided modulation scheme with one sample per pulse (as in Fig. 7*b*) and a 1 kHz tone 15 dB below peak. The second harmonic is at −108 dB, or 93 dB below the signal. A small component is apparent at 3 kHz but is at the level of the noise floor. This result has been compared with the performance of the DAC supplied in the (inexpensive) CD player: the PWM DAC is better by some 10 dB. Further details of this performance can be found in Reference 20.

5 Linearisation for distortion reduction

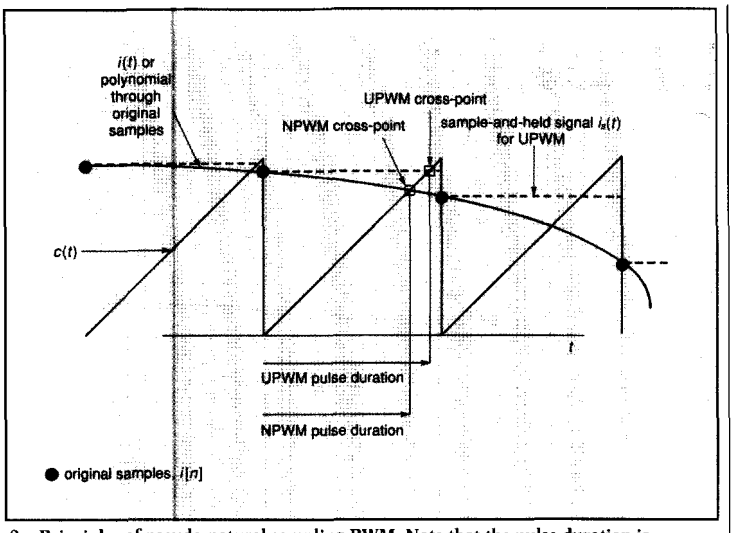
The performance presented in the previous Section is certainly good, but it is possible to do better. It was pointed out in Section 2 that both UPWM and NPWM suffer from distortion but that only UPWM suffers from harmonic distortion. Therefore, we now move on to examine a way of reducing the distortion of UPWM by additional signal processing which makes UPWM approximate the performance of NPWM. This is done by recognition of the ideal performance and the deviation from this.

In the work of the author's group and that of a team at Liverpool University[13] it is assumed that NPWM offers a performance worth striving for. In the work of Hawksford[14] and of Craven[15], PAM is seen to be the ideal and they, respectively, seek to minimise pulse-to-pulse spectral changes or pulse-to-pulse time domain changes and to come up with linearisation schemes thereby. This paper will deal only with the approach of the King's College group.

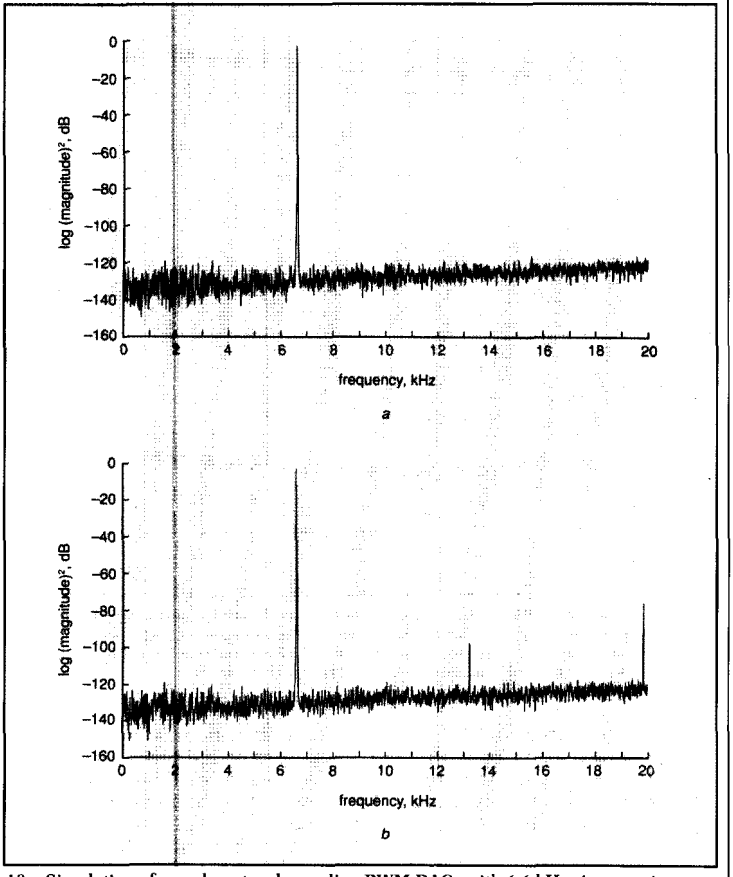


8 Performance of UPWM hardware demonstrator with 8 times oversampling and hardware noise shaping: (*a*) trailing edge, one-sided modulation of 1 kHz signal at −15 dB; (*b*) trailing edge, one-sided modulation of 10 kHz signal at −6 dB; (*c*) double-sided, one sample per pulse modulation of 1 kHz signal at −15 dB

Fig. 9 shows in close-up the difference between natural and uniform sampling pulse width modulation. The original continuous time signal is shown as a solid line and its sampled-and-held version, $i_s(t)$, as a piecewise-horizontal dashed line. Also shown are the comparison waveform, $c(t)$ (a sawtooth for clarity, to give one-sided modulation), and the samples of the original continuous time waveform, $i[n]$ (depicted by small black blobs, positioned on the verticals of the sawtooth).



9 Principles of pseudo-natural sampling PWM. Note that the pulse duration is equivalent to the signal's value at the cross-point. The 'cross-point detection' algorithm works by calculating an approximation to the NPWM cross-point from the original samples, $i[n]$, (via polynomial interpolation to give $i(t)$ and a nonlinear function solution) then sending this to the noise shaper/modulator in place of the original samples which would have produced pulses of 'UPWM' duration



10 Simulation of pseudo-natural sampling PWM DACs with 6·6 kHz sinewave input at −3 dB, 8 times oversampling, noise shaping 16 bits down to 8: (*a*) trailing edge, one-sided modulation with 3rd order cross-point detection compensation algorithm; (*b*) trailing edge, one-sided modulation with 1st order cross-point detection compensation algorithm

In the central sawtooth period, the points at which the continuous waveform, $i(t)$, and the sample-and-held waveform, $i_s(t)$, cross the rising part of the sawtooth are highlighted and called the NPWM and UPWM cross-points, respectively. At these points the PWM waveform changes state, say from high to low (it is assumed that the low-to-high transition occurs regularly at the times of the verticals of $c(t)$). It is clear, as shown, that the durations of the pulses are different.

However, the original waveform, $i(t)$, is not available; only the samples $i[n]$ are. Nevertheless the process can be approximated if an approximation to $i(t)$ can be interpolated* through the set of points $i[n]$. It will then be possible to calculate the time instant at which the continuous waveform will cross the comparison waveform and feed this to the uniform modulator's counter so that an approximation to NPWM is produced.

To do this, we appeal to numerical methods for polynomial interpolation and nonlinear equation solving. A small number, say four, of consecutive samples are fitted with a polynomial in t whose coefficients can be found by building a difference table.[16] Let this polynomial, which approximates the original analogue input signal over the short period, be called $i(t)$. Then the cross-point instant can be found from the equation $i(t) = c(t)$, where $c(t)$ represents the comparison waveform. Since $c(t)$ is a ramp over one sample epoch, it is simply equal to t, so that the problem becomes:

$$\text{Solve} \quad i(t) - t = 0 \quad \text{for } t$$

This can be done in a number of ways, the one currently being used in the work at King's College being a Newton–Raphson technique[16] which relies on computation of the derivative of the polynomial, $i'(t)$. This algorithm, when coded, corresponds to the box in Fig. 1 labelled cross-point detector. Further details can be found in References 21–23. We have called the system of Fig. 1, pseudo-natural sampling PWM (PNPWM).

Fig. 10 presents some results from simulations of a PNPWM DAC system. Only a one-sided final modulator is necessary and, as

*Here we mean interpolation in the sense of a numerical method, though this is closely related to its meaning for sample rate increase.

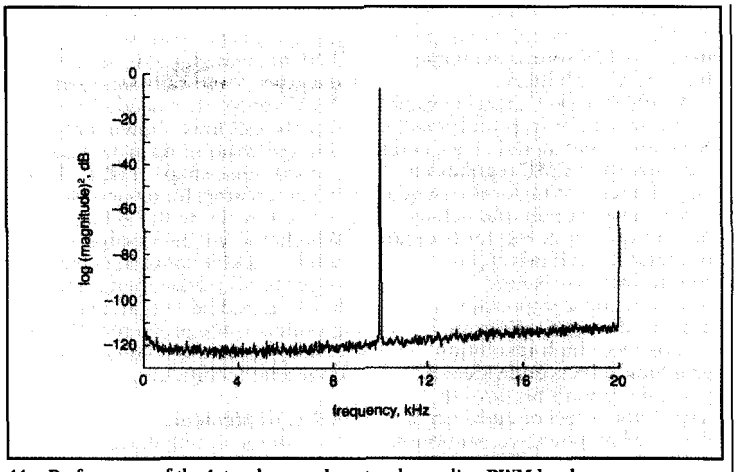
shown in Fig. 10*a*, if a 3rd order polynomial is used with 8 times oversampled data, there is no harmonic distortion; there will be some intermodulation distortion though this can be removed by using a 5th order polynomial. If the polynomial order is reduced to first order to relax computational complexity, the harmonic distortion (see Fig. 10*b*) becomes apparent. However, this should be compared to Fig. 7. The same final modulator as for Fig. 7*a* is being used, though the results compare favourably with those of Fig. 7*b*. This is achieved by a moderate amount of additional numerical computation and a reduction in modulator complexity (compared to Fig. 7*b*). The result of Fig. 7*c* is better, but is achieved with still greater modulator complexity, increased numeric computation (to determine the positions of an edge as a function of consecutive samples) and twice as much oversampling (i.e. 16 times instead of 8). More results appear in Reference 23.

As a guide to the viability of PNPWM, the hardware demonstrator has been adapted and the cross-point detection algorithm coded on a 33 MHz floating-point digital signal processor with eight times oversampling. It takes just a little more processing power than one processor provides: a second processor needed for noise shaping provides the balance.

More recently, the algorithm has been tested using a DSP description language (Silage[17]) which allows bit-true* simulations to be performed and which is the input language for a DSP silicon compiler (Cathedral 2). Synthesis down to the silicon layout stage has been performed and has shown that a 3rd order cross-point detector and noise shaper fit into 42 mm^2 (active area) of a 2·4 micron fabrication process. With routeing, metalisation and input/output pads this would become about 120 mm^2. In a 1·2 micron process these areas reduce to 12 mm^2 active area and 30–50 mm^2 with input/output pads etc. A finer geometry IC process will enable a full stereo version of the algorithm with noise shapers to be integrated into a single package.

Results from the demonstrator with a first order polynomial cross-point detector are shown in Fig. 11. The performance is not ideal but still shows a 20 dB improvement over the performance without the cross-point detector compensation (see Fig. 8*b*). The system here is a one-sided modulator operating with eight times oversampling and the signal is noise shaped from 16 to 8 bits. Before the end of 1993, the projected further improvement in performance from using higher order cross-point detection will be demonstrated.

*Fully accurate to a final hardware solution



11 Performance of the 1st order pseudo-natural sampling PWM hardware demonstrator with 8 times oversampling and hardware noise shaping: trailing edge, one-sided modulation of 10 kHz signal at –6 dB

6 Potential for power D/A conversion

If the initial aims of this work, to produce an all-digital power amplifier, are to be achieved, several points need to be addressed, all focused on the output stage. Although it has been shown that noise shaping reduces the speed of the master clock (i.e. the duration of the fundamental quantum of Fig. 2) it does not change the accuracy with which the pulse edge is to be positioned — this must be commensurate with 16 bit performance (at the near-Nyquist sampling frequency of 44·1 kHz). If this cannot be achieved there will be non-negligible jitter on pulse edges which will manifest itself as an increased noise level. Thus the output device must switch in about 10 ns and have edges that are reproducible to a few hundred picoseconds.

In addition, the complexity of the power supply must be considered. The supply must be regulated since any droop in supply rails will manifest itself as an amplitude modulation superimposed on the pulse width modulation.

Finally, there is the choice of output stage configuration and of output smoothing filter, which must recover the baseband from the switching waveform. The early stages of the *LC* filter should remain linear even though large voltage swings are occurring.

This author makes no claims to be an expert in this field, which is more like microwave engineering than data conversion and signal processing. However, from many discussions over the years, it seems that adopting a resonant output stage configuration will combat switching speed and EMC problems simultaneously and that such a stage could be integrated into the power supply design.

7 Applications

The intended application for this work when it started was all-digital amplification of digital audio signals. A current project with this aim is supported under the SERC/DTI Teaching Company Scheme in collaboration with a major UK hi-fi manufacturer. Two important aspects of this project are the output stage, as mentioned above, and the issue of local clock generation and synchronisation with other items of digital audio equipment (e.g. CD and DAT players). If the power supply problem is solved, then use of the amplifiers in all-digital active loudspeaker systems is also feasible, although the market for active loudspeakers is somewhat limited.

Because pulse width modulation is efficient, good power drive can be obtained from low-voltage power supplies. This bodes well for application to in-car stereo

systems, where the power supply problem is alleviated by simply using the 12V source available from the car's battery.

Another market, growing rapidly at the present time, is multimedia. Systems would already have had to pass stringent EMC regulations compliance-testing for the switch-mode power supply and will use high-frequency clocks for the main processor. Obviously high-resolution stereo sound reproduction is important in multimedia applications to complement high-resolution graphical effects and video. If portable systems become an important aspect of multimedia then the low-power consumption of PWM converters/amplifiers makes them doubly attractive.

The most promising application is as a high-resolution digital-to-analogue converter in place of conventional technologies. Because no significant load power is required, the power supplies can be regulated and a complete system can be integrated. The use of pseudo natural sampling to linearise the PWM process offers the promise of high resolution, perhaps close to 24 bits. The drawback will be that bandwidths much wider than those of audio signals cannot be accommodated, because the oversampling necessary will drive clock speeds unrealistically high.

If the pulse width modulator and output stage are combined with a direct digital synthesis signal generator[19] then it is possible to produce pure sinusoidal signals at high powers in an efficient way. This might be suitable for power-efficient handsets in personal and mobile communication systems.

8 Conclusions

This paper has described how digital-to-analogue converters may be constructed from pulse width modulators and further has looked at their performance both in simulation and hardware. As well as this a new technique has been described which is capable of reducing the distortion inherent in PWM to make it viable for conversion of digital signals to very high degrees of resolution. Simulation and preliminary hardware results have been shown for this technique too.

The paper has discussed the realisability of a power DAC based on pulse width modulation and the applications which PWM-based DACs might have. The author believes it to be possible to construct a 16 bit converter with no measurable distortion. To make a 20 bit converter, it is certain that the pseudo naturally sampled PWM approach is needed and simulations have shown that 16 times oversampling is necessary (for the noise shaper rather than the modulator) for a one-sided core converter with 8 bit precision. Whether it will be possible to produce 24 bit converters for signal bandwidths greater than a few hundred hertz is an open question — the problem is that of clock jitter and is common to all converter techniques.

Acknowledgments

The author would like to acknowledge financial support from SERC, BTG, DTI, B&W Loudspeakers and King's College London. He would also like to thank all the many students who have contributed to this project: Mayur Nathwani, Martin Greaves, Rod Hiorns, Jason Goldberg, Allan Paul, Rob Bowman, Peter Ziman, Michael Watson, Andy Andrzejczuk and Kevin Davis. Particular thanks are due to Rod and Allan for supplying the experimental plots.

References

1 BLACK, H.S.: 'Modulation theory' (Van Nostrand, Princeton, NJ, 1953)

2 SANDLER, M.B.: 'Investigation by simulation of digitally addressed audio power amplifier'. PhD thesis, Essex University, 1983

3 SANDLER, M.B.: 'Towards a digital power amplifier'. 76th Convention of the Audio Engineering Society, 8th–11th October 1984, New York

4 SANDLER, M.B.: 'Progress towards a digital power amplifier'. 80th Convention of the Audio Engineering Society, 4th–7th March 1986, Montreux

5 SCHAFER, R.W., and RABINER, L.R.: 'A digital signal processing approach to interpolation', *Proc. IEEE*, June 1973, **61**, (6), pp.692–702

6 MARTIN, J.D.: 'Theoretical efficiencies of class D power amplifiers', *Proc. IEE*, June 1970, **117**, (6), pp.1089–1090

7 BATURIN, N.A.: 'Full wave modulation in class D amplifiers', *Telecommun. & Radio Eng.*, Pt. 2, 1974, **29**, pp.122–123

8 BATURIN, N.A., and PLYUSNIN, V.N.: 'Combination distortion in a two-sided single ended class D power amplifier', *Telecommun. & Radio Eng.*, Pt. 2, 1973, **28**, (8), pp.123–126

9 GOLDBERG, J., and SANDLER, M.B.: 'Noise shaping for a digital power amplifier'. Presented at AES 87th Convention, 18th–21st October, 1989, New York

10 CANDY, J.C., and TEMES, G.C. (Eds.): 'Oversampling delta–sigma data converters', (IEEE Press, 1992)

11 GERZON, M., and CRAVEN, P.: 'Optimal noise shaping and dither of digital signals'. 87th AES Convention, New york, 1989, preprint 2822

12 PAUL, A.: 'Steps towards developing a high resolution DAC using PWM'. MPhil/PhD Transfer Thesis, Dept. of Electronic and Electrical Eng., King's College London, 1992

13 MELLOR, P.H., LEIGH, S.P. and CHEETHAM, B.M.G.: 'Reduction in spectral distortion in class D amplifiers by an enhanced pulse width modulation sampling process', *IEE Proc. – G*, August 1991, **138**, (4), pp.441–448

14 Hawksford, M.O.J.: 'Dynamic model-based linearization of quantized PWM for applications in digital to analogue conversion and digital power amplifier systems', *J. Audio Eng. Soc.*, April 1992, **40**, (4). pp.235–252

15 CRAVEN, P.: 'Towards the 24 bit DAC: novel noise shaping topologies incorporating correction for the non-linearity in a PWM output stage', *J. Audio Eng. Soc.*, May 1993, pp.291–313

16 KNUTH, D.E.: 'The art of computer programming: Part 2 — Seminumerical algorithms' (Addison-Wesley, Reading, Massachusetts, 1980, 2nd edn.)

17 HILFINGER, P.: 'A high level language and silicon compiler for digital signal processing'. Proc. IEEE CICC Conference, pp.213–216, May 1985

18 RABAEY, J., DE MAN, H., VANHOOF, J., GOOSENS, G., and CATHOOR, F.: 'CATHEDRAL II: a synthesis system for multiprocessor DSP systems' *in* GAJSKI, D.D. (Ed.): 'Silicon compilation' (Addison-Wesley, 1988)

19 O'LEARY, P., WOLLSCHLAGER, A., WEBER, G., PLESCHIUTSCHNIG, J., BALDUAF, H., and REININGER, F.: 'Oversampling data conversion applied to data modulation'. Proceedings of 1st IEE International Conference on Analogue to Digital and Digital to Analogue Conversion, Swansea, September 1991, pp.124–129

20 HIORNS, R.E., BOWMAN, R.G., and SANDLER, M.B.: 'A PWM DAC for digital audio conversion: from theory to performance'. IEE International Conference on Digital to Analogue and Analogue to Digital Conversion, Swansea, September 1991

21 GOLDBERG, J.M., and SANDLER, M.B.: 'Pseudo-natural PWM for high accuracy digital to analogue conversion', *Electron. Lett.*, 1st August 1991, **27**, (16), pp.1491–1492

22 UK Patent 9027503.3: 'Improvements in or relating to digital to analogue conversion', 19th December 1991. Also PCT/GB91/02279 'Digital to analogue conversion using PWM'

23 GOLDBERG, J.M.: 'Signal processing for high resolution PWM based digital to analogue conversion'. PhD Thesis, University of London, 1993

First received 1st February 1993

Dr. Mark Sandler is Reader in Digital Signal Processing at the Department of Electronic and Electrical Engineering, King's College London, Strand, London WC2R 2LS, UK. He is an IEE Member.