Understanding PDM Digital Audio

Thomas Kite, Ph.D.
VP Engineering
Audio Precision, Inc.

Table of Contents

Introduction 3
Quick Glossary 3
PCM 3
Noise Shaping 4
Oversampling 5
PDM Microphones 6
DACs and PCM-to-PDM converters 6
PDM Modulators 7
Transmitting and Handling PDM Signals 7
Performance 8
Conclusion 9
Further Reading 9

Meta Platforms, Inc. - Ex. 1028
Page 2

Introduction

PDM stands for pulse density modulation. However, it is really better summarized as "oversampled 1-bit audio", as it is nothing more than a high sampling rate, single-bit digital system. If one increased the sample rate of audio CDs by a large factor, and reduced the wordlength from 16 bits to 1 in a reasonable way, that would serve as the basis of a PDM system.

Most current digital audio systems use multi-bit PCM (pulse code modulation) to represent the signal. PCM has the advantage of being easy to manipulate. This allows signal processing operations to be performed on the audio stream, such as mixing, filtering, and equalization.

PDM, which uses only one bit to convey audio, is simpler in concept and execution than PCM. It has become popular as a way to deliver audio from microphones to the signal processor in mobile telephones. PDM is ideally suited for this task because it brings the benefits of digital, such as low noise and freedom from interfering signals, at low cost.

This document will cover the basics of PDM: how it is generated, transmitted, and manipulated.

Quick Glossary

DAC (Digital-to-Analog Converter): a device that converts a digitally represented signal to analog.

LSB (Least Significant Bit): the smallest change that can be made in a digital word. A bit is a binary digit.

MSB (Most Significant Bit): the highest value bit in a digital word; effectively it is the sign bit in a fixed-point signed numerical representation.

PCM (Pulse Code Modulation): a system for representing a sampled signal as a series of multi-bit words. This is the technology used in audio CDs.

PDM (Pulse Density Modulation): a system for representing a sampled signal as a stream of single bits.

Sampling rate is the rate at which a signal is sampled to produce a discrete-time representation.

Wordlength is the number of bits used to represent a sample.

Quantization is a procedure for representing an arbitrary data sample using a given wordlength.

Dither is a noise-like signal added before quantization to improve performance.

Linearization is the process of mitigating the deleterious effects of data quantization, usually by adding dither.

Noise modulation is the undesirable variation of the noise floor in a system due to the signal content.

PCM

Before we tackle PDM, let's first review PCM, that is, conventional multi-bit digital audio. In PCM, the audio signal is represented as a series of samples, each a fixed number of bits long. Two factors determine the performance of the system:

- Sampling rate. This determines the bandwidth of the system.

Meta Platforms, Inc. - Ex. 1028
Page 3

- Wordlength. This determines the signal-to-noise ratio (SNR) of the system.

In particular, the bandwidth is $f_s/2$, where f_s is the sampling rate, and the SNR is given by $(6.02N + 1.76)$ dB, where N is the wordlength in bits.

A raw 16-bit system has a theoretical SNR of around 98 dB. In practice, dither is used to linearize the system and eliminate noise modulation; this reduces the SNR by about 4 dB. Using the above formula, an undithered 1-bit system has an SNR of about 8 dB, which is of course unacceptable for any real audio work. Furthermore, optimal dither needs 2 LSBs to work; since a 1-bit system only has 1 LSB total, and that's used for the audio, hence there is no room for dither.

Since the system cannot be properly dithered, a 1-bit representation would at first blush appear to be a non-starter. The solution lies in an understanding of noise shaping and oversampling.

Noise Shaping

Consider a typical PCM signal such as a 24-bit representation of a sine wave. How might this be represented in a system whose wordlength is only one bit, when such systems appear to have severe noise and distortion problems?

One might start by simply throwing away all the bits except the MSB, effectively thresholding the signal around the zero point. This will turn the sine wave into a square wave that switches at the zero crossings. This introduces a tremendous amount of distortion; over 40%, in fact. The distortion arises because the system is undithered. Quantization always introduces error, but in a dithered system, the error comes in the form of a white noise floor uncorrelated with the signal. In an undithered or under-dithered system, some of the error is in the form of distortion.

Reducing to 1-bit by retaining the MSB is therefore not the answer. However, we are all familiar with an example of wordlength reduction to one bit that works very well. It's called *halftoning*, and has been the basis for reproducing images in print media since the invention of the newspaper.

In halftoning, a continuous-tone image (such as a grayscale photograph) is converted to a series of black dots and white spaces. In other words, the wordlength is reduced to one bit, where the state of the bit corresponds to a black dot or a white space. This is done not by simple thresholding, but rather by distributing the error caused by thresholding among neighboring pixels that have yet to be thresholded. This process is known as error diffusion. (There are many other ways to create halftones, but we won't consider them here.) The effect on image quality of diffusing the error is dramatic, as shown below.

Meta Platforms, Inc. - Ex. 1028
Page 4

Original image | Thresholded image | Error diffused image

Why does diffusing the error incurred by thresholding result in a huge increase in the visual quality of the image? The answer is that error diffusion performs two functions. First, it transforms the distortion caused by simple thresholding into something more like a noise floor; and second, it shapes that noise floor so that the noise at low spatial frequencies is reduced, at the expense of noise at high frequencies. This matters in images because most of the image content is at low frequencies. Furthermore, high image frequencies are filtered by the fundamental resolution of the eye, so as long as the dots are small enough (or the image is sufficiently far away), a lot of the high frequency noise simply isn't visible.

The result is that what would have been gross distortion from thresholding becomes a fairly benign, *high-pass noise floor*. By "fairly benign", we mean that its appearance is acceptable, although it is not a true noise floor, because the system is undithered. The noise is still correlated with the signal, and exhibits tonal behavior and other artifacts. Still, the visual results are good.

Halftoning is an example of a *noise-shaping* system. The "noise" incurred by reducing the wordlength is "shaped" so that it is not flat, but high-pass. In general, noise-shaping systems can have any output wordlength, and there is no requirement that their noise transfer function be high-pass. However, the vast majority of such systems, including PDM systems, have a 1-bit output and a high-pass noise transfer function.

Oversampling

The noise incurred by reducing the wordlength is substantial. (The noise in a 1-bit system is about 90 dB higher than the noise in a 16-bit system, for example.) Noise shaping distributes that noise in a high-pass fashion, but it does not reduce the total noise level. In an imaging application, where most of the image content is at low frequency, pushing the noise to high frequencies (where it might obscure some of the signal) is not much of an issue. In audio, however, mid- and high-frequencies are very important, and very audible. It is simply not possible to achieve acceptable results if the wordlength is reduced to one bit, even with noise shaping. The resulting high-pass noise is clearly audible.

The answer is to use a higher sampling rate. This increases the bandwidth of the system, creating new spectrum above the audible range. Noise shaping can then be used to push noise into that spectrum. In

Meta Platforms, Inc. - Ex. 1028
Page 5

effect, more space has been created in which to dump noise. And since that spectrum is above the audible range, the noise cannot be heard.

A higher sampling rate can be realized in two ways:

- By using a higher sampling rate in the first place. This is the method used in PDM microphones, where the typical sampling rate is 3 MHz.
- By interpolating an existing signal that has been sampled at a low rate. This is the method used in many DACs, where a typical incoming sample rate is 48 kHz. It is also used in systems which represent audio internally as PCM, but transmit audio to external devices in PDM form.

We'll now look at both of these approaches in more detail.

PDM Microphones

A PDM microphone, also called a digital microphone, consists of the following parts:

- A microphone element. Typically this is an electret capsule.
- An analog preamplifier.
- A PDM modulator.
- Interface logic.

The analog signal from the microphone element is first amplified, and then sampled at a high rate and quantized in the PDM *modulator*. The modulator combines the operations of quantization and noise shaping; the output is a single bit at the high sampling rate. The noise shaping ensures that the noise in the audio band is relatively low, while the noise above the audio band is relatively high. The interface logic is responsible for accepting a master clock and transmitting the sampled bitstream.

The device to which the microphone connects provides the master clock to the PDM microphone. The clock rate defines the sampling rate of the system, as well as the rate at which bits are transmitted on the data line. Although there is no defined standard, typically the oversampling ratio is 64. So to achieve a bandwidth of 24 kHz (comparable to a PCM system sampled at 48 kHz), a master clock frequency of 3.072 MHz is needed.

The one-bit data is asserted on the data line on either the rising or falling edge of the master clock. Most PDM microphones support stereo operation, in which one microphone asserts the data line on the rising edge of the master clock, while a second microphone asserts on the falling edge. On the non-asserted edge, the data output has a high impedance. The data lines from the two microphones can then simply be connected together. The PDM receiver is responsible for separating the two bitstreams.

DACs and PCM-to-PDM converters

In some commercial DACs, and in systems which convert PCM to PDM, the procedure is slightly different from PDM microphones. The signal has already been sampled at a low rate, and is in PCM form. To achieve the high sampling rate needed for noise shaping to be effective, the signal must first be interpolated. Its wordlength is then reduced to one bit in a noise shaper.

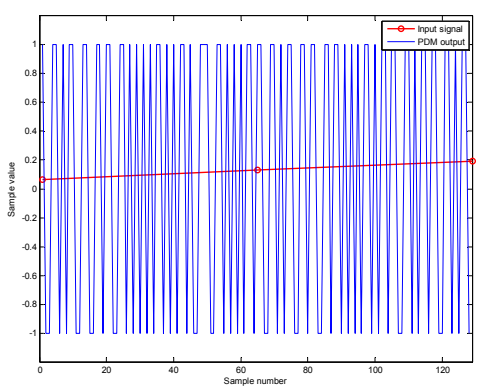
Meta Platforms, Inc. - Ex. 1028
Page 6

Interpolation is a digital filtering operation in which extra samples are generated in between the existing samples to increase the effective sampling rate. For PDM applications, the oversampling ratio is typically 64; that is, 63 new samples are generated for each input sample.

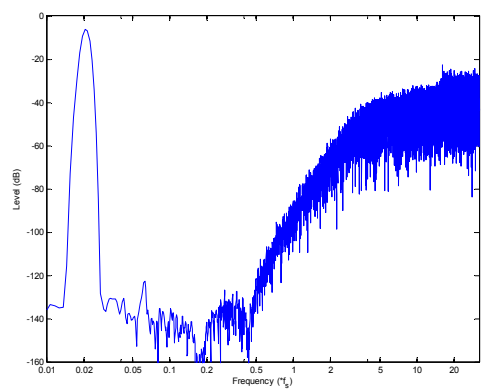
PDM Modulators

The PDM modulator (in PDM microphones) or the noise shaper (in PCM-to-PDM converters) is responsible for producing a one-bit signal which has very low noise in the passband. The complexity of the modulator is expressed by its *order*. The order of a modulator is equal to the number of integrators (accumulating nodes) it contains; in general, the higher the order, the more aggressively the noise is shaped from the passband to the stopband, and the better the noise performance. However, higher order modulators are more complex to design and manufacture; they are more likely to become unstable under certain operating conditions; and their maximum input level before overload is lower. While there is no industry standard, typical modulators in PDM microphones are fourth order. This offers a good compromise between noise performance and complexity.

Below are time domain and frequency domain views of the output of a PDM modulator when fed with a sine wave input signal. The time domain output switches at a high rate between two levels. In the frequency domain, the passband extends from 0 to 0.5 f_s on the x-axis. Above that is spectral space created by oversampling. The sharp rise of noise above the passband is clearly visible. Also visible is a small amount of third harmonic distortion (the peak at approximately 0.06 f_s).



PDM signal, time domain



PDM signal, frequency domain

Transmitting and Handling PDM Signals

A PDM bitstream is a logic-level signal typically switching at around 3 MHz, with fast edges. It therefore needs to be treated with the same care as any other fast signal (such as SPDIF, or analog video). It's important to use good quality coax cable and to terminate the signal correctly.

Ultimately a signal needs to be converted to an analog form if it is to be heard. If it is to be processed, or analyzed by test equipment, it needs to be converted to PCM. It is possible to do both of these with a PDM signal.

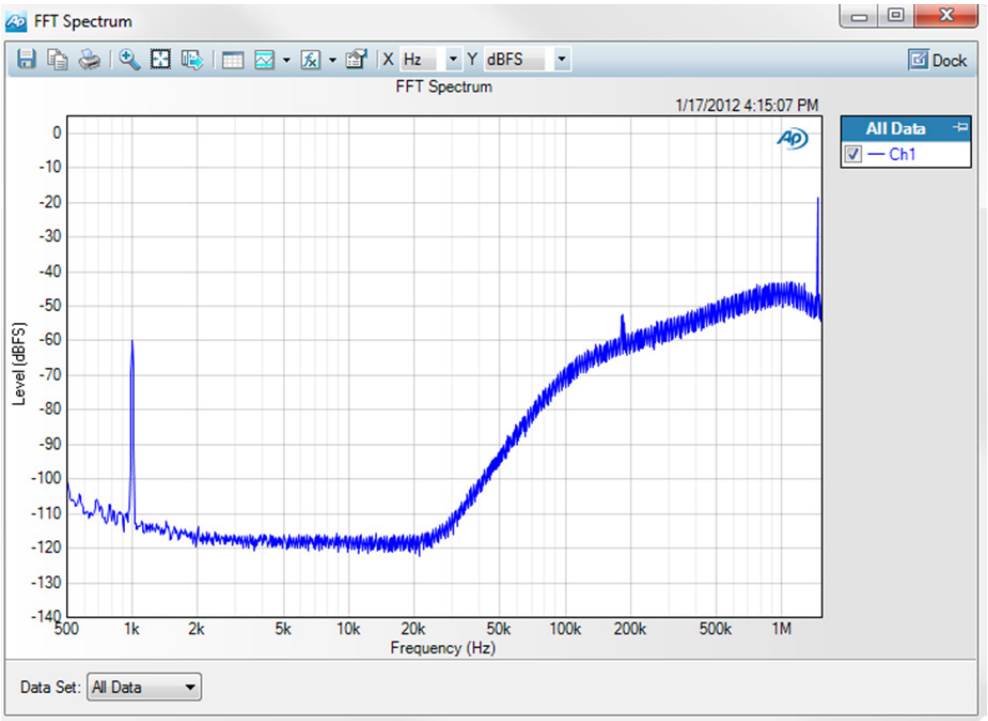
Meta Platforms, Inc. - Ex. 1028
Page 7

Converting PDM to analog is in principle very simple. The one-bit signal already contains the audio in the low part of the spectrum. All that is required to recover it is a low-pass filter. In practice, the fast switching edges in the signal require careful design of the analog filtering stages, but it is certainly possible to recover a very high quality analog signal this way.

Converting PDM to PCM is more involved. The sample rate needs to be reduced by the oversampling factor. This is accomplished in a digital filtering operation called *decimation*. Decimation is the counterpart to interpolation: samples are removed from the signal to reduce the sampling rate. It is important that the noise above the audio band in the 1-bit representation not be allowed to alias into the audio band. The decimation filters are designed to filter out this noise, leaving the baseband audio signal intact. The output of the decimator is a PCM audio stream at the baseband (non-oversampled) rate. Typically the wordlength increases from 1 bit to around 20 effective bits during the filtering.

Performance

The one-bit field is very mature. Although a 1-bit system has inherent problems, in particular the inability to add enough dither to fully linearize the system and eliminate noise modulation, it is nevertheless possible to design a system with excellent audio performance.



Output of actual MEMS (micro electromechanical system) PDM microphone captured by an AP audio analyzer, showing a 1 kHz test tone and the effects of noise shaping above the passband.

Meta Platforms, Inc. - Ex. 1028
Page 8

PDM modulators are usually proprietary; the performance therefore varies depending on the design. The modulator implemented in Audio Precision's APx PDM Interface option uses a fourth-order modulator coupled with a six-stage interpolation/decimation filter with over 120 dB of image/alias rejection. The resulting system spec is as follows:

- Maximum input level before overload: -6 dBFS
- SNR @ 1 kHz, -6 dBFS, 20 Hz – 20 kHz, unweighted: 109 dB
- THD+N @ 1 kHz, -6 dBFS, 20 Hz – 20 kHz, unweighted: -107 dB
- Third harmonic distortion @ 1 kHz, -6 dBFS: -116 dB
- Flatness, 20 Hz – 20 kHz: better than ±0.001 dB

All high-order PDM modulators have a maximum input level that is somewhat below full scale. Exceeding this level will cause modulator overload, resulting in poor noise performance. The APx user interface indicates when the modulator is in overload.

The THD+N performance of the system is dominated by the noise floor of the modulator. There is a small amount of third harmonic distortion present. This arises because the system is undithered.

Conclusion

PDM is a cost-effective way of conveying audio digitally, in mono or stereo, over a clock/data pair. Despite the inherent limitations of a one-bit representation, it is possible to achieve extremely high audio performance with careful design. The APx PDM Interface option generates and analyzes PDM signals natively, greatly simplifying the design and troubleshooting of all aspects of the PDM signal chain.

Further Reading

The following references offer more information.

A Brief Introduction to Sigma Delta Conversion, Intersil Application Note AN9504, May 1995. Retrieved from http://www.intersil.com/data/an/an9504.pdf.

Principles of Sigma-Delta Modulation for Analog-to-Digital Converters, Motorola Application Note APR8/D Rev. 1, 1990. Retrieved from http://www.numerix-dsp.com/appsnotes/APR8-sigma-delta.pdf.

Delta-Sigma Data Converters: Theory, Design, and Simulation, by Steven Norsworthy, Richard Schreier, and Gabor Temes, Wiley-IEEE Press, 1996.

Understanding Delta-Sigma Data Converters, by Richard Schreier and Gabor Temes, Wiley-IEEE Press, 2004.

5750 SW Arctic Drive
Beaverton, Oregon 97005
800-231-7350

Copyright © 2012 Audio Precision

ap.com

XII0118133800

Meta Platforms, Inc. - Ex. 1028
Page 9