
Introduction

For a lot of people, dither is like transmission fluid: you've been told you need it, you've accepted that, and that's really as much as you want to know. For those that care to know more though, we've put together this document¹.

The only thing less exciting than learning about dithering is probably a documentary on the history of long division. Therefore, in what could be a futile attempt to make this more entertaining, we've titled each section using a lyric from a song. You probably never knew there were so many songs about dithering, noise shaping, and bits.

So without further ado, let's get into it.

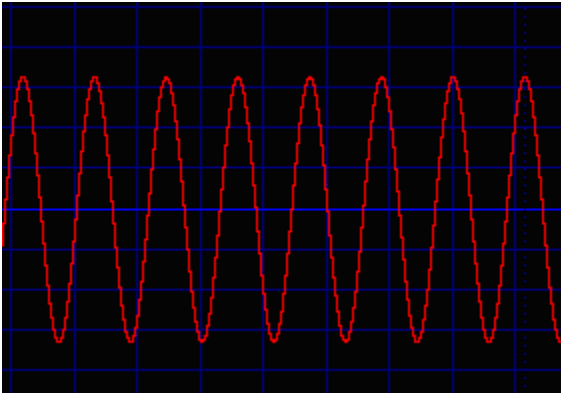
¹ To clarify, it is for people who care to know more about *dithering*. Let's just continue to accept that we need transmission fluid and leave it at that.

Section I: "The Digital Won't Let Me Go"

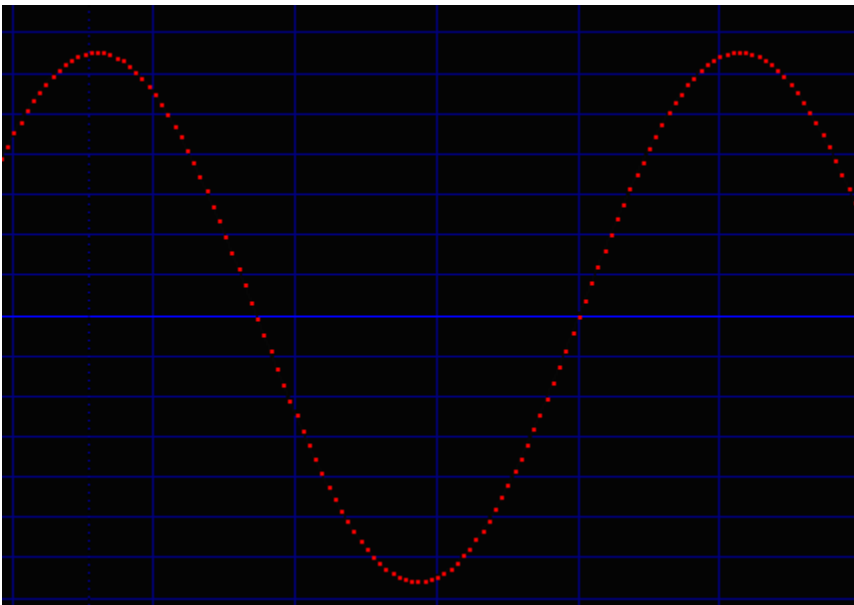
Tomorrow Comes Today by Gorillaz

The first thing to recognize is how digital audio is stored and represented. Unlike analog audio, which is "infinitely continuous", digital audio is represented by individual bits. This "quantization" means that while we might hear a continuous sound, it's really a whole bunch of discrete 1s and 0s.

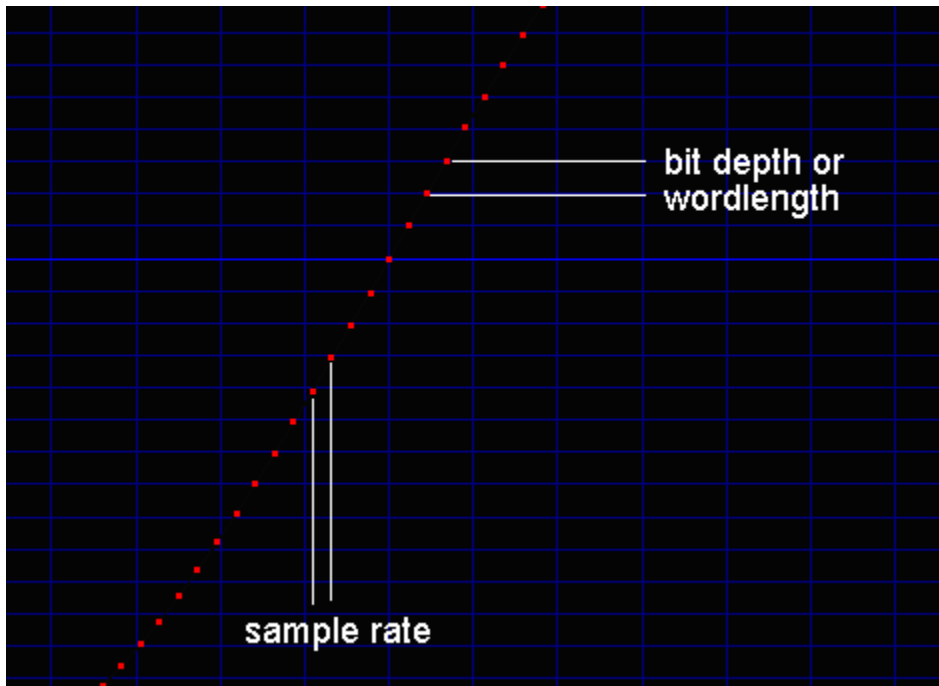
As an example, a digitally created sine wave is shown below. From a distance, it looks and sounds pretty continuous. (sound file: 24bit_sine.wav)



If we zoom in though, we can see that our continuous looking waveform is actually a bunch of individual or discrete samples



Zooming in once more...



The horizontal distance between samples represents the sample rate, that is, how often samples are taken of the audio. The vertical distance between samples is a function of the bit depth or "word length" of the audio.

As we take more samples per second (e.g. 44.1 kHz², 48 kHz, 96 kHz, etc.), we get better "horizontal" resolution, which translates to being able to represent higher frequencies. As we use more bits per sample (e.g. 16 bits, 24 bits, 32 bits, 64 bits, etc.) we get better "vertical" resolution which translates to better "dynamic resolution" (i.e. more dynamic range and/or signal to noise ratio.) In the context of dithering, it is in the "bits per sample" that we're interested.

So, better dynamic resolution is pretty easy to get on your PC; just use more bits per sample. Get a 24-bit A/D converter; mix and edit at 32 bits; process effects at 64 bits. End of story. Except... audio CDs are 16-bit. At some point, if you want to put your songs on a CD, you have to get that 32-bit word length down to 16 bits. And that's the problem.³

² Where "Hz" is literally the number of samples per second and "kHz" is the number of 1000's of samples per second.

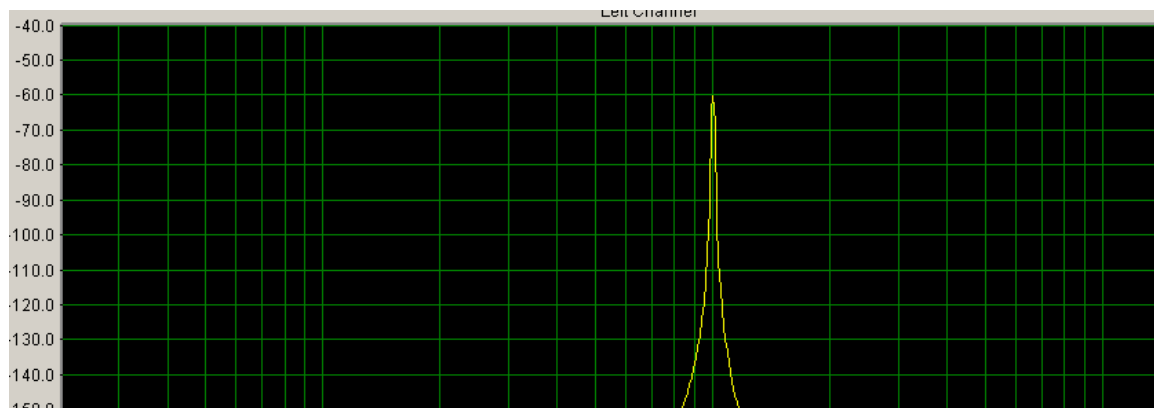
³ Granted, sometimes audio has to be delivered at 8 bits for games, answering machines, and all sorts of other things. Our focus is on the 24-bit to 16-bit issue for CD, as this conversion is what Ozone dithering is tailored for.

Section II: "We'll erase the naughty bits"

Insatiable by Prince

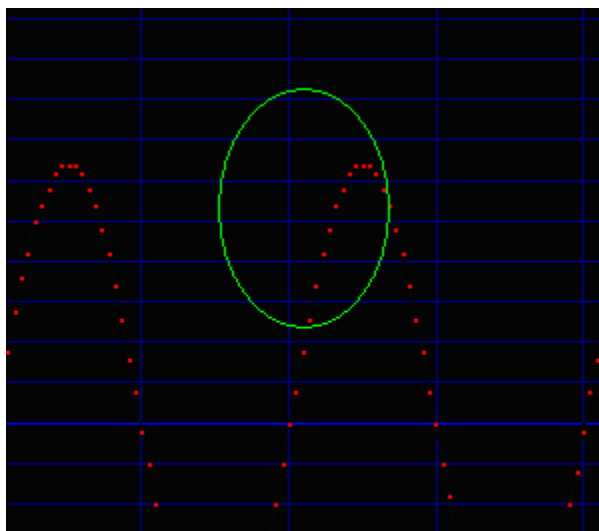
So it comes down to fitting 24 bits on a 16-bit CD. One solution, as Prince sang, "We'll just erase the naughty bits."⁴ In other words, the simplest way to convert is to simply throw away or "truncate" the lowest 8 bits: take 24 bits, throw away the lowest 8, you're left with 16, and you can make a CD.

Before you just throw the bits away though, let's check out a little comparison. First, here's a spectrum of our 1 kHz 24-bit sine wave. The spectrum of a pure tone should be single spike at the frequency of the tone, as it is below.



Now we'll convert our nice smooth 24-bit sine wave to 16 bits by simply truncating or throwing away the least significant 8 bits.

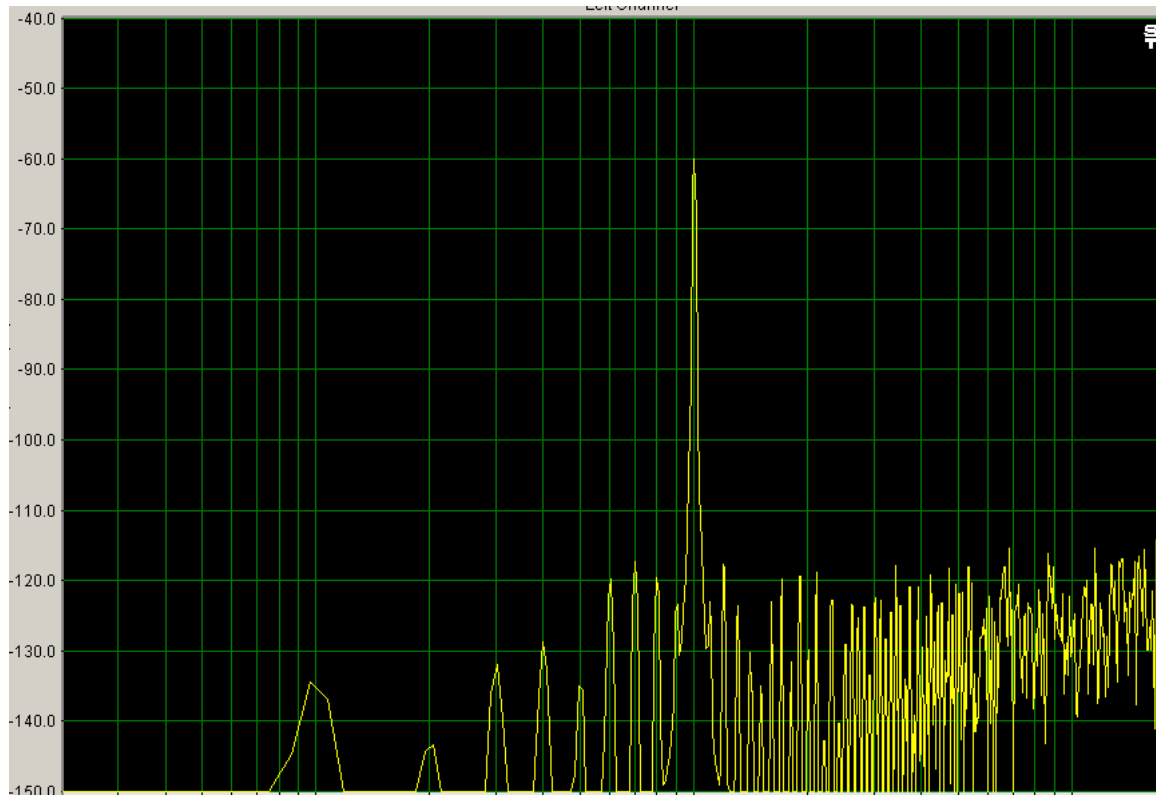
Looking first at the waveform in time, we can see that things got a little less smooth. Especially in the area with the circle, you can see that with fewer bits, we're not able to represent the original smooth curve.



⁴ Upon further listening, we realized Prince might not have been talking about word length reduction through truncation in that song.

Truncation introduces "quantization error" which is the difference between where the higher resolution sine wave had its samples and where the lower resolution sine wave has to put the samples.

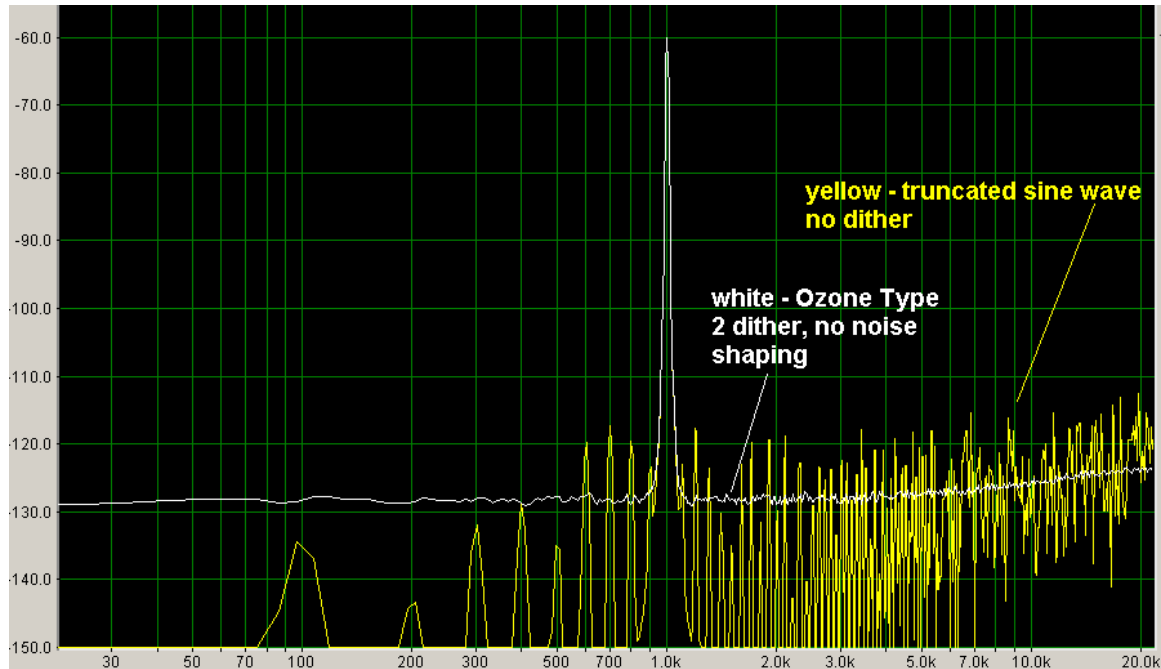
A spectrum of the truncated 16-bit sine wave is even more revealing. Again, a pure sine wave appears as a single spike. As the sine wave becomes more jagged and "squared off", the spectrum reveals artifacts related to the quantization error (e.g. noise, harmonics). Whatever you want to call it, it doesn't look or sound good. (sound file: 16bit_tone_nodither.wav)



Section III: "Over on the corner there's a happy noise"

Down on the Corner by Creedence Clearwater Revival

So what to do? We dither. We'll get into the details of dithering later, but for now consider it as adding very low-level noise to the audio before it is converted from 24 bits to 16 bits. The results of converting with and without dithering are shown below:



Our dithered tone (white spectrum) seems to look better. We've effectively done away with the jagged quantization noise. It sounds better, but we have made a tradeoff and added some low level noise. (sound file: 16bit_sine_Ozone_Type2_Clear_Dither.wav)

The terrible truth is revealed: you've spent time and money on low noise preamps, A/Ds, and everything else. But in the end, you're going to deliberately add noise to your mix when you convert it from 24 to 16-bits to put on a CD.

On the bright side, we've traded "bad noise" for "good noise". Instead of quantization error (turn up the volume and listen to 16bit_truncated_sine.wav), we've added a smoother more "continuous" noise. (turn up the volume and listen to 16bit_sine_Ozone_Type2_Clear_Dither.wav)⁵

⁵ We're simplifying not only the principles behind dithering but the benefits as well. Just understanding the tradeoff to begin with is a good place to start, though.

Section IV: "Everything Looks Worse in Black and White"

Kodachrome by Paul Simon

At the risk of confusing the issue with an analogy, here's another way to look at it.⁶ Image processing faces the same issues of bit depth, resolution and dithering. When representing a photo on a computer, each sample has a discrete number of colors or shades of gray that it can represent. The number of levels is determined by the bit depth of the file, just like the levels of samples of audio.



24-bits

A picture of an Atlantic puffin is shown above⁷. He's looking good in 24-bit gray scale, meaning that every dot that makes up the photo can be one of over 16 million shades of gray. It looks pretty "continuous". Sixteen million shades of gray is a pretty good dynamic range.

Below is what happens if we truncate the photo down to 2 bits. Now we only have 4 shades of gray to represent the photo and something is going to be lost. Same as in digital audio, each sample has to be forced to a level, and there aren't enough levels anymore to represent the data as a continuous image (or a smooth continuous signal in the case of audio).



2-bit truncated

⁶ Pun intended. We work hard for those.

⁷ No animals were harmed in the making of this guide.

As with digital audio, we can dither images. The concept is the same: we add controlled noise before we convert from 32 bits to 2 bits. The result of the conversion, using dither, is shown below.



2-bit truncated



2-bit dithered

Like 8-bit audio (or even 16-bit compared to 24-bit) our 2-bit dithered picture is not great, but it's better than when we just truncated. In fact, it shows three characteristics that are analogous to down-sampling audio with dither.

- 1) We've reduced the quantization error (the jaggedness or "blockiness") with the tradeoff of adding a different type of noise (the speckles in the dithered picture).
- 2) We've actually maintained some detail from the original higher resolution picture. In audio, this relates to the concept that reducing the number of bits per sample with dithering can give you greater perceived dynamic range than the signal to noise ratio.

OK, that doesn't make sense at first glance. Meaning, what does it mean? Well, the theoretical maximum signal to noise ratio of a 16-bit recording is about 96 dB. That means, in basic terms, that the noise floor of a 16-bit recording is around -96 dB. The noise floor should not be confused (but often is) with dynamic range, which represents the difference between the loudest and softest signal that you can hear in the recording. You can hear a dynamic range which extends past the noise floor with proper dithering. Even if a signal is "in the noise", you will still hear it with proper dithering. The breadth of the dynamic range you can hear below the noise floor can't be expressed mathematically. It's a result of the program material, and -- well -- how well you can hear, or more precisely: how well a listener can resolve signal from noise at low levels.

- 3) The picture also answers a common audio question:

"If I know I'm making a 16-bit CD in the end, should I still record, mix and master at 24-bits?"

Yes, the extra bits provide you (or more specifically the dither) with information that allows the dynamic range to be greater than the noise floor. In short, 24-bit properly dithered to 16-bit will sound better than an original 16-bit recording.⁸

⁸ Because it is a subjective judgment, we're only comfortable saying 24-bit dithered to 16-bit will have greater perceived dynamic range than an original 16-bit recording. We can't say by how much. Some other companies seem to be able to,

Section V: "Break our love to bits"

Let's Hang On by The Four Seasons

All this "perceived dynamic range" and "greater than 16 bit dynamic range in 16 bits" mumbo-jumbo is starting to sound a little fishy, and thus it calls for an audio demonstration.

Like all dither demonstrations, you need a good pair of headphones and a good D/A converter. You're listening to stuff that's 90 dB or more down in level.

1) Listen to `24_bit_Mix_Fade.wav`. It's a mix made at 24 bits, the level dropped down to begin with, and then faded out (with 24 bit calculations). This is a worst case situation, as we could have done the mixing at 32 bits and faded out using 64 bits with Ozone automation in Acid 4. There's nothing special about the mix -- it's actually just one of the Acid 4 demo songs from Sonic Foundry.

2) Now listen to `16_bit_Mix_Fade_No_Dither.wav`. This file was created by simply "throwing away 8 bits" -- we truncated the 24 bit file (no dither) to create our 16 bit version. You'll hear that around 13-14 seconds into the file the level is low enough that it just turns to quantization error and noise. You can't resolve the music anymore, and the noise itself is pretty harsh.

It's OK to turn the headphone amp up quite a bit for this demonstration. It is important in some cases to listen to dither at a "normal" level (not turned up), but we'll get into those cases later.

And please make sure that any system sounds are turned off. This lesson is quickly learned the first time you hear a system sound (dialog box message, etc.) at 110 dB through a pair of headphones. Even safer for these high volume demonstrations, burn a CD and listen at higher volumes through a CD player. Burning a CD also ensures that you're not turning up the volume with a digital volume control, but instead with an analog volume control (the CD amplifier).

3) Now listen to `16_bit_Mix_Fade_Ozone_Clear_Dither.wav`. In this example, the 24-bit master was dithered using Ozone Type 2 Dither and Clear Shaping before it was converted to 16 bits. There is still noise, but it's a different type of noise. The mix is still audible and "resolvable" even up to the very end - several seconds past when the non-dithered 16-bit mix degraded into noise.

Fun stuff. Just to mention it, this isn't some super magical feature that's specific to Ozone dithering. Any well-designed dither applied to a 24-bit mix to make a 16-bit CD master will let you hear a fade past a truncated 16-bit reference.

though - i.e. "With the Dither2000™ you can fit 22 bits of dynamic range into 16 bits!!!". Have some fun with them ask how they measured it, and then tell us, because maybe we're just the ones that don't get it.

Section VI: "How to turn down the noise in my mind"

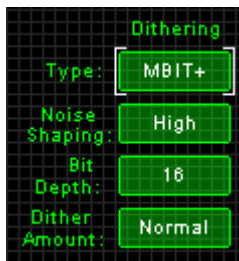
Haven't Got Time for the Pain, by Carly Simon

At this point you're probably distraught with emotional turmoil. You can hear (and see) the benefits of dither, but still -- well -- we're adding noise to your audio! We can put a man on the moon but we can't do transparent conversion from 24-bit to 16-bit?!

Fortunately, we're not done. There's a little bit of rocket science left:⁹ How you create the dither noise source and shape that noise source can give you the same benefits of dithering with a minimal cost (cost being the perceived level of the added noise).

Noise Source:

In Ozone, the noise source is selected by the Type button. Type 1 is a rectangular probability density function, Type 2 is triangular probability density function and MBIT+ is an iZotope proprietary dither technology.



Without getting into too much unnecessary detail, MBIT+ is (if you have Ozone) the preferred dither type. Beyond our own technologies, a triangular probability density function (TPDF, or Type 2 in Ozone) is a commonly used simple dither function (over other shapes such as rectangular, Gaussian, etc.). The statistical/objective level of a TPDF source is slightly higher than a rectangular function (+4.77 dB), but the perceived "annoyance" or loudness of a TPDF function is lower. We mention this fact because this is one of several cases in dealing with extremely low level noise where the mathematics go against the human perception. And since we're making music for people to listen to and not to measure, we go with the perception.

So, in general, our recommendation is to use the MBIT+ algorithm if you have Ozone, or a TPDF or triangular Type 2 if you don't have Ozone (but have that option in another dither plug-in).

There are a couple other things to mention regarding the noise source. For obvious reasons, we can't go into too much detail about how we do it in Ozone, but these constitute things to consider when evaluating dither¹⁰:

⁹ OK, it's not really rocket science, but we needed to tie together that whole "man on the moon" reference and that seemed to work.

¹⁰ Other companies also make dither and have web access.

1) The actual noise needs to be uncorrelated with the program material to be effective. How you generate the noise source and how you control the randomness is one of the tricks behind good dither.

2) There shouldn't be any periodicity or "repeating" of the noise source. Periodic clicks, hum, or predictable cycling can be detected even at low levels.

3) The noise source should be smooth; meaning we're replacing bad noise with good noise. Good noise should have some type of "pleasant" quality to it. There shouldn't be excessive spikes, breaks, or discontinuities¹¹. It should sound nice -- like those electronic boxes you can buy from The Sharper Image that play nice noise to put you to sleep.¹²

4) It should be stereo; meaning the left and right channels of noise should be uncorrelated in some way. You don't want a song to fade out and have the stereo image collapse as it falls into a mono dither noise source. Listen to the imaging as best you can as the song fades out. Does the apparent sound stage stay the same?

Noise Shape:

This is where it really gets tricky. You can actually shape the noise so that it provides effective dithering benefits and less added audible noise. Because this is a subject in itself, let's start a new section to address it.

¹¹ That's why Ozone has a "limit peaks" option in the dither module

¹² Don't quote us on that. We've never actually heard one of those boxes, but we imagine they sound nice.

Section VII: "Show me how you do that trick"

Just Like Heaven by The Cure

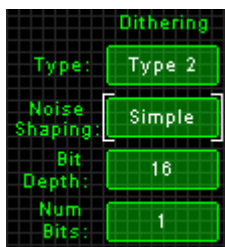
The trick here is to make the dither noise effective yet not audible. You can't just "turn it down" because then it loses its effectiveness to decorrelate the quantization error. So the noise has to exist and in some meaningful level¹³, but can you shape it so it sounds less noisy?

Yes, you can. There are some simple ways, some complex ways, and some "over-clock the Celeron cause we're preparing a CD master with Ozone dither" ways. We never were ones to be cheap on CPU cycles if you we could use them to your advantage.

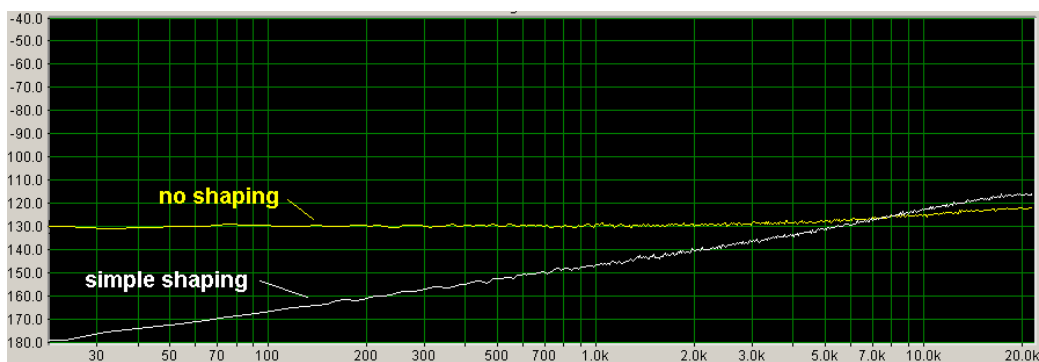
"...I am in a conventional dither..."

A Wonderful Guy by Margaret Whiting

A conventional dither shape is a simple high pass filter. This high pass filter is the shape that is provided by the Ozone Shape: Simple selection, and is also used by many sound cards, the TC Finalizer, and is at least provided as an option in most host applications. This type is also referred to as "high pass", HP, or HPF in dither modules or host applications.



Conventional doesn't necessarily mean good or bad, it just means simple. Compared to no shaping at all, simple high pass shaping can shift the dither noise into higher frequencies. This shift can make it less audible relative to no shaping at all, or at least less irritating at frequencies <5 kHz or so. This type of dither requires very little processing power (very easy to implement on inexpensive hardware) and has very little "tonal shifting"¹⁴ as the dither signal itself is filtered very broadly.



To hear for yourself:

No Shaping Dither Signal: 16bit_Dither_Noise_No_Shaping.wav

Simple (high pass) Dither Signal: 16bit_Dither_Noise_Simple_Shaping.wav

¹³ Throughout this discussion, keep in mind the noise we're talking about is noise that's more than -90 dB down even without any tricks. A meaningful level is pretty quiet to begin with.

¹⁴ Where you can actually hear the a perceived pitch in the noise

"...Tryin' ta hide up in the spectrum..."

Get Da Drop On Em by L.L. Cool J

So by combining a high pass filter with a good dither noise source, we just saw/heard that the perceived level of the noise decreased. Can we take this idea even further?

Yes, by applying more complex filters we can radically shift the majority of the dither noise up "Near Nyquist" (the Nyquist frequency)¹⁵ or around half the sample rate of our audio. In other words, a CD mix is hopefully sampled at 44.1 kHz (as that's the sample rate for CD audio). The Nyquist frequency would be 22 kHz. Since the upper limit of our hearing is in this range, an effective dithering strategy would seem to be to shift the majority of the dither noise up towards the Nyquist frequency.

And in fact, this is a very good strategy. Now, if we just said that because we thought so that would sound silly. So for comparison, below is a plot of two "high frequency" shapes. The yellow spectrum is the Waves L1 with the "moderate" shape, and the white line is Ozone with the "clear" shape.

To compare for yourself:

Ozone dither noise, clear shape:

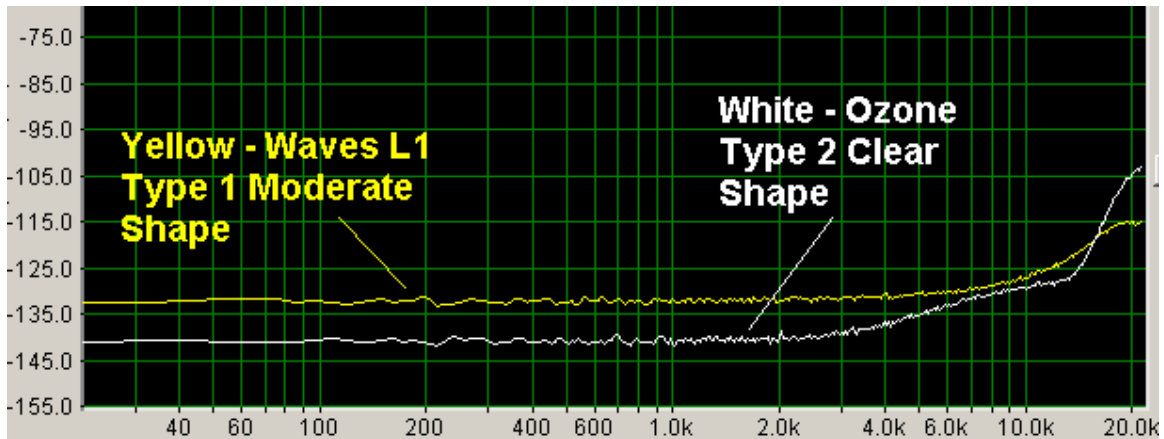
16_bit_Ozone_Clear_Dither_Noise.wav

Waves L1 dither noise, moderate shape:

16_bit_Waves_L1_Type1_Moderate_Dither_Noise.wav

And again, dither noise with no shaping:

16bit_Dither_Noise_No_Shaping.wav



As you can see, both dither shapes put the majority of the noise up in the higher than 15 kHz or so range. There are small differences beyond that - Ozone dither is higher in level above 16 kHz or so, Waves L1 dither is higher in level below 15 kHz or so. But the main point is that both offer significant improvements in the perceived loudness of the noise relative to simple shaping or no shaping at all.

¹⁵ The Nyquist frequency is defined to be half of the Nyquist sampling rate.

Just to be perfectly clear¹⁶, the purpose of presenting both an Ozone and Waves dither shape is because both (in our opinion) are excellent algorithms. They are effective at removing quantization error and still perceptually quiet in level. Any differences in the shaping is relatively minor considering we're talking about signals that are going to be heard at -95 dB or more. In short, this is meant to be a positive presentation of two similar "Near Nyquist" shapes and we hope you recognize that as well.

"...Bend me, shape me, anyway you want me..."

Paranoid by Garbage

As far as shaping dither for maximum effectiveness and minimum audibility, we've got one more trick to share with you. In concept, it's quite simple:

- 1) Dither is by definition very quiet noise - i.e. very low in level.
- 2) When listening to noise at very low levels, we are more sensitive to noise in certain frequency ranges.

So we can shape the dither so that the noise is less prominent at the frequencies that we're sensitive to at low levels. Like many simple concepts, the implementation and evaluation of this dither strategy can get very complicated very quickly. But that's the general idea behind what is called "psychoacoustic" noise shaping.

Let's explore each of the points behind the psychoacoustic shaping concept in more detail.

- 1) *Dither is by definition very quiet noise - i.e. very low in level.*

Up until now, we've been listening to dither noise at artificially high levels. For example, in hearing the difference between a mix fading out with no dither versus a mix fading out with dither, we mentioned that you should turn up the volume to a high volume to hear the last few seconds of the fade. This fade was down more than 96 dB at the end. To really hear the last few seconds, you might have turned it up 50, 60, or more dB. Nobody is going to listen to the song at that volume (if you had full scale for most of the song it would be playing at 150 dB or so).

So the point here is to recognize that the level of dither is at -90 dB levels and below. Because we know it will be heard (or even better not heard) at the threshold of our hearing, can we take advantage of that to shape it to be even less audible?

- 2) *When listening to noise at very low levels, we are more sensitive to noise in certain frequency ranges.*

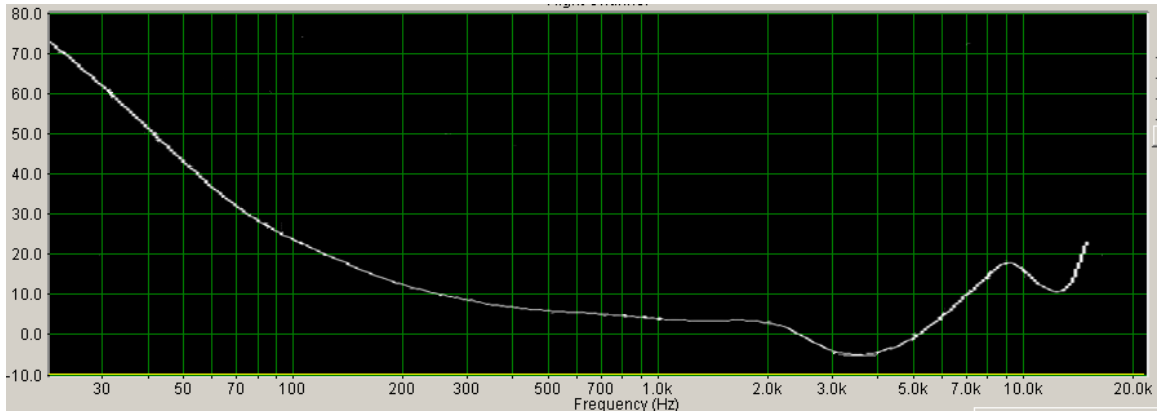
Yes, we can take advantage of the low level of dither to shape it even more based on this principle.

At high volumes, we hear frequencies more or less equally. Meaning, a 100 Hz tone played at 90 dB is perceived to be the same loudness as a 10,000 Hz tone played at 90 dB.

¹⁶ Three puns if you're keeping score...

But as sounds get quieter, our ear/brain combination starts to behave very "non-linearly". Sort of like a microphone with a bad frequency response, we perceive soft sounds at different frequencies to be very different volumes, even if from a "sound pressure level" or objective measurement standpoint they are in fact the same volume.

The plot below shows our sensitivity to different frequencies when played at a very low level (i.e. the threshold of hearing). What the plot shows is the amount that we would need to boost a frequency in level to hear it at the same perceived loudness as another frequency.



To use the "microphone with a non-flat frequency response" analogy, mentally flip the plot over and the peaks show the frequencies that we're most sensitive to, and by how much.

Zooming in on the loudness curve, we can see that we're most sensitive to sounds between 2.5-5 kHz with another dip or "sensitivity" to sounds around 12 kHz

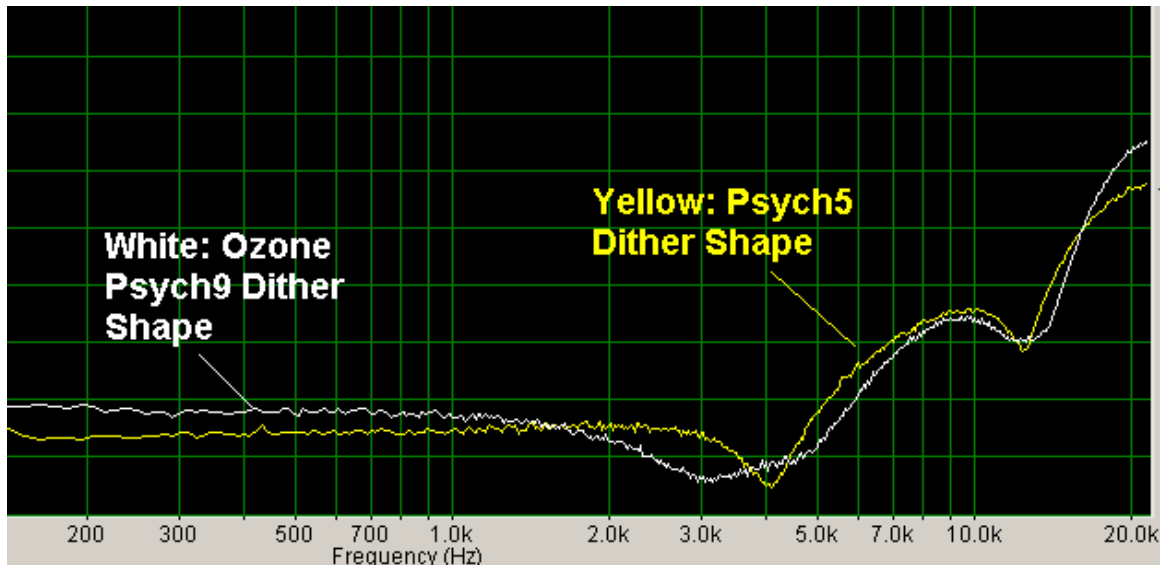


Interesting to note: 2.5-5 kHz is the predominant region for intelligibility of speech, and around 12 kHz is the frequency range that we use to localize (or determine the location of) sound. We might not have a flat frequency response at low levels, but we have a very appropriate one for life in general.

3) So shape the dither so that the noise is less prominent at the frequencies that we're sensitive to.

Makes sense, and this technique is used in the Psych5 and Psych9 shapes of Ozone, as well as being offered in dither hardware such as the Meridian 518 and Weiss POW-R.

You can see the spectrum of Ozone Psych5 and Psych9 dither shapes below



As you would expect, there are controlled dips in the noise around those frequencies that we're most sensitive to at low volumes.

So what's the upside, downside, and things to know about psycho-acoustically shaped dither?

1) The MBIT+ and Clear shapes can operate at all sample rates (and are in general the preferred modes of dither). The Psych5 and Psych9 shapes are designed specifically to work at 44.1 kHz sample rates. If you use a different sample rate with the Psych shapes, the dips and peaks will shift in proportion to the sample rate, and if you've followed this whole "we hear different frequencies differently at low volumes" concept you'll realize that putting the peaks and dips at other frequencies would be bad.

2) Psychoacoustic shaping requires significantly more calculation than simpler shaping techniques. And because of all of the calculations, rounding error becomes very important. When your desired dither signal is just 1 or 2 bits, you can't afford any rounding errors to creep into the output signal. This precision one of several reasons that Ozone takes a lot of CPU, and uses 64-bit calculations internally.

3) Last, and not least, the psychoacoustic curves are designed to be effective at the lower thresholds of hearing. If you turn up the level to an unrealistic level to "better hear" the dither, you're undoing the whole point of psychoacoustic shaping. If you convert to 12 or 8 bits, you're also turning up the level of the dither signal into a range that it wasn't designed to be effective at. What this means is that

psychoacoustic shaped dither has to be evaluated at normal listening levels -- take your mix, play it at a normal listening level so the full scale part of the mix is appropriate, and then evaluate the dither during quiet parts, fade-outs, etc. without turning up the level. We recognize that it is difficult to hear at this level -- which is the whole point.

Section VIII: Don't Believe the Hype

Don't Believe the Hype by Public Enemy

Like other dither manufacturers, we're selling you noise. Even more difficult, we're selling you noise that you can barely hear, and in the case of some shaping algorithms (psychoacoustic ones) we're telling you not to turn up the level so you can even really evaluate it. All of this would make it pretty frustrating as a user to try to get a handle on whether any of this makes a difference, or if we're all just packaging up random noise and calling it a digital audio breakthrough.

We've probably evaluated 50 combinations of third party dither implementations. We won't tell you ours is the best, because it's entirely subjective. Here's how we went about evaluating dither, though, and you have all the tools to do it yourself.

"...If my eyes don't deceive me there's something going wrong around here..."

Is She Really Going Out With Him by Joe Jackson

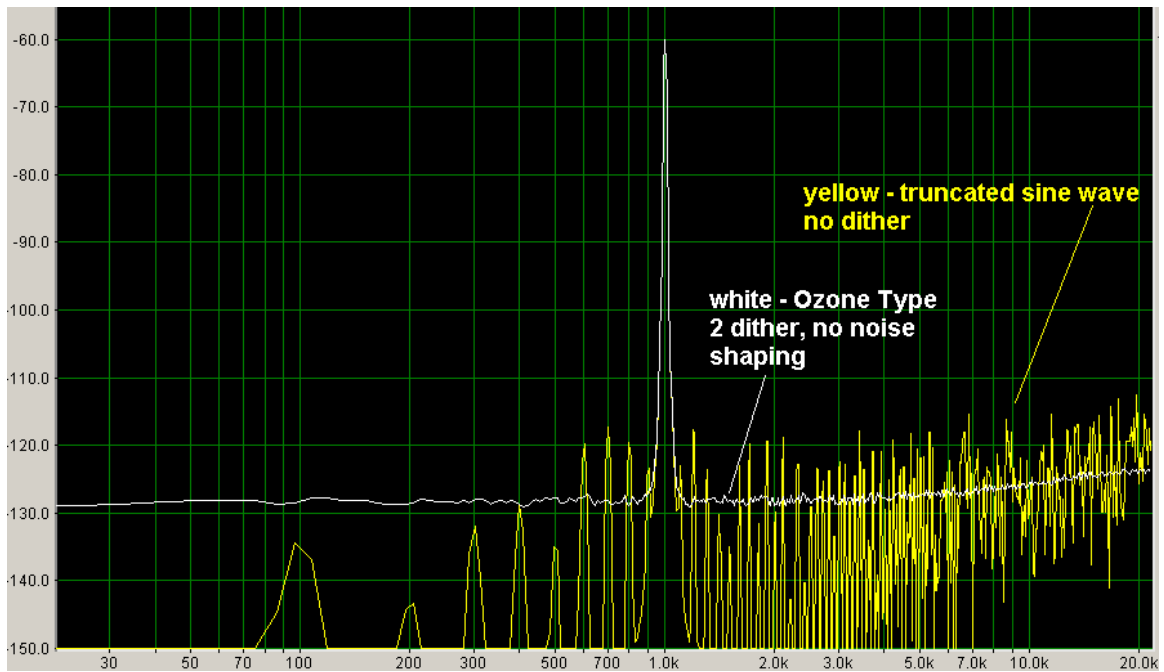
There is a bunch of different ways to evaluate "whether it works". Most of them are subjective. Here are a couple simple ways we can suggest to at least highlight an implementation that *doesn't* work.

The Sine Wave Test.

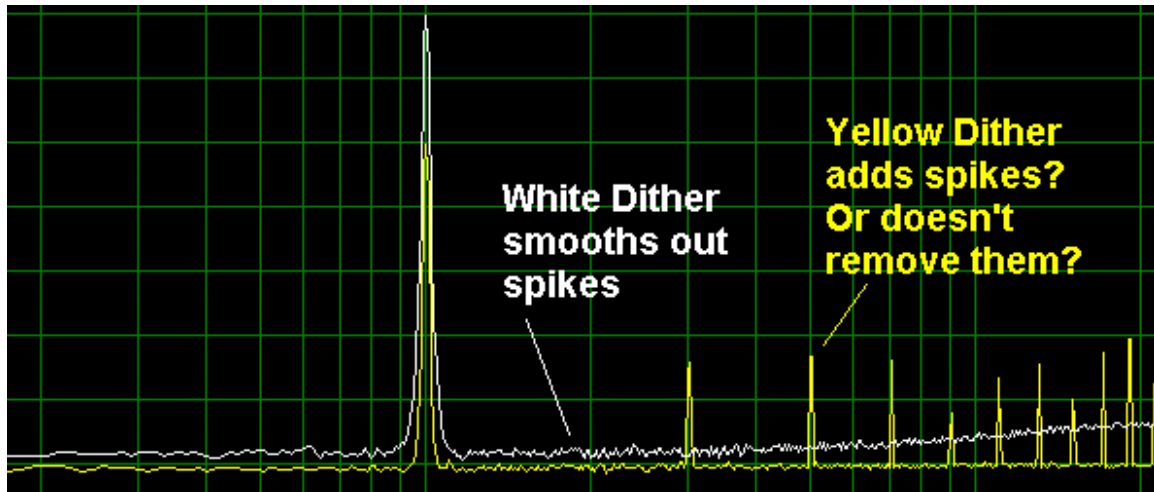
Take 24bit_sine.wav. It's a pure 1 kHz sine wave, 24 bit resolution, with a level at -60 dB relative to full scale. Convert it to 16 bits using whatever dither you want to evaluate. Then look at the dithered 16-bit file in a spectrum analyzer. We used Spectra Pro from www.SoundTechnology.com for most of these plots, but you can also use the spectrum analyzers in Sound Forge, Cool Edit, Wavelab, etc.¹⁷ Just be very careful to not have any level adjustments, effects, or other options turned on that could affect the 16-bit file you're trying to analyze.

If all is well and good, you should see the 1 kHz tone spike and a smooth line (like the white line shown below). The line might be curved (depending on the shaping you might have applied) but it should have done its job of evening out the quantization error spikes.

¹⁷ You can't use the spectrum analyzer in Ozone for this purpose, unfortunately, because the range was designed for EQ ranges (+15 to -30 dB or so) and it won't show very low (-90 dB) signals.



Inadequate dither will either not remove the spikes, or in some cases could generate its own spikes. This could be because a periodic noise source was used (that could generate its own frequency components) or it's just not effectively removing the quantization error. It's easier to explain why something works than why it doesn't work, so we'll just leave it at that.



The Mix Fade Out Test

As we showed with the mix fading out, dither can (and should) effectively even out quantization error. You can hear quantization error at the end of `16_bit_Mix_Fade_No_Dither.wav`. It's OK to turn it up. We're not evaluating the "overall level" of the dither, but just that quantization error is being replaced with dither noise.

You can hear the quantization error removed in 16bit_sine_Ozone_Type2_Clear_Dither.wav. Again, turn it up as we're just checking whether it's doing its job.

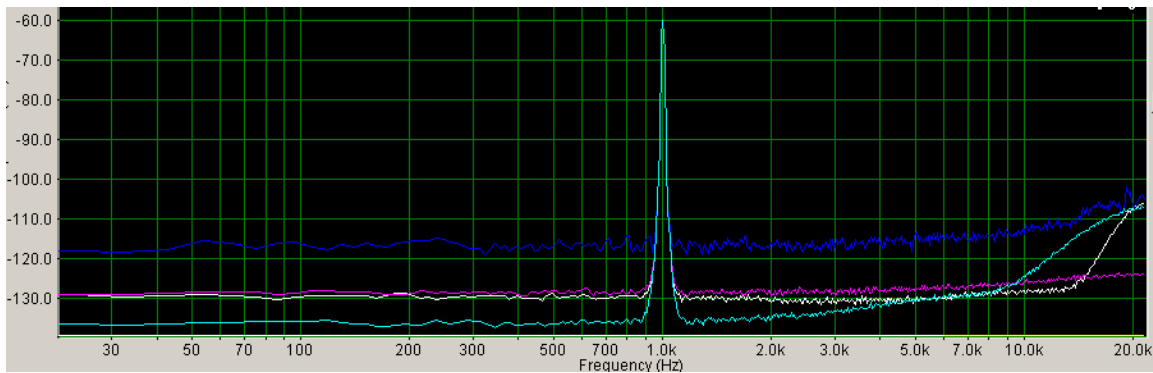
You can use 24_bit_Mix_Fade.wav for your own dither experiments. Or, use any 24 bit recording you want. Destructively adjust the level of a mix so it starts around -60 dB or so (so you can turn up your headphone amplifier and not kill your ears when it loops back to full scale) and then fade it out to nothing. We've found a linear fade or slow fade over 20 seconds or so is most effective for gradually reducing the level so you can focus in on the dither and/or quantization error.

"...Turn it up, that's enough..."

Caravan by Van Morrison

Of course, you can also have too much dither. Or, put another way, an excessive level of noise created by the dither signal.

Using a 24-bit sine tone at 1 kHz, we tried four different dither implementations and compared the results. As you can see, three out of four are in the same range of level, more or less. The blue one looks (and sounds, even at normal listening levels) to be significantly (and unnecessarily?) high. It seems to have done the job of reducing the spikes and harmonics we would see in a truncated signal, but at a level that's as much as 10 dB higher for most of the spectrum.



Of course, more important than how loud it looks on a spectrum is how loud it sounds. The purple line is higher for most of the spectrum, but not the higher frequencies. The green curve is lower for most of the spectrum, but gets very high above 10 kHz. The white line is somewhere in the middle. Which one sounds the loudest is up to you to decide; dither some tones and listen to the result.

Two things to keep in mind when listening to dither:

- 1) Some dither won't operate if it doesn't have a signal¹⁸. So you can't dither on digital silence on these boxes. It's always safest to feed it a known signal (a tone, for example) to make sure "this thing is on".
- 2) Turning up the level, as mentioned before, will give the wrong impression of any "psychoacoustic" shaped dither. If you see dips in the spectrum, especially around 3

¹⁸ Including Ozone if the "auto-blank" option is selected. This tells Ozone to only output dither noise if the input is non-zero (i.e. not complete silence)

kHz and/or 12 kHz, the dither is probably meant to be evaluated at "normal" listening levels.

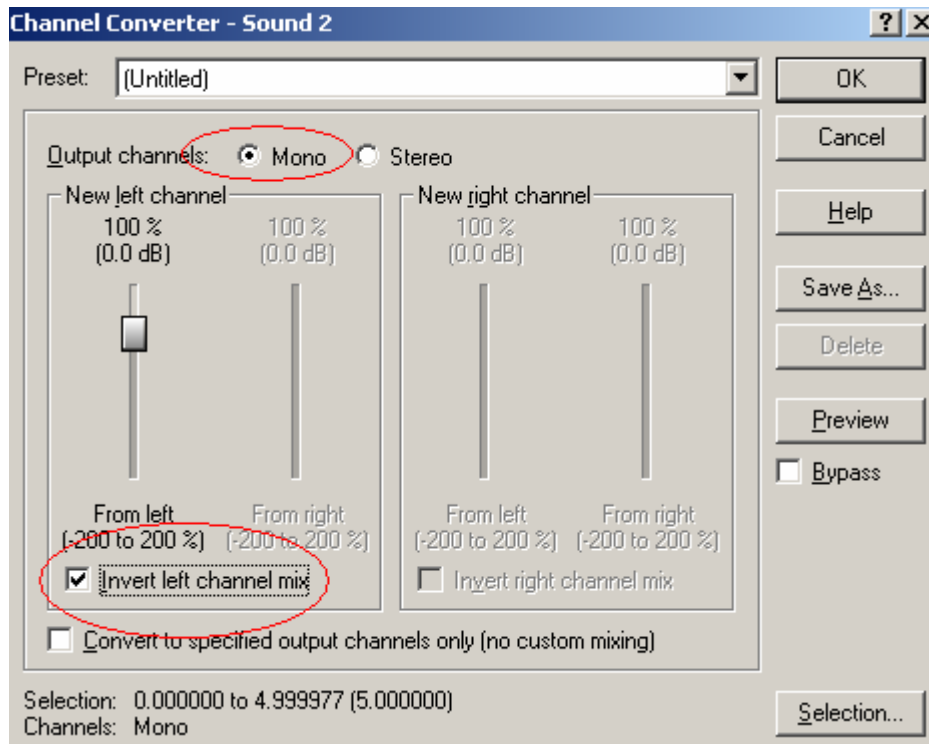
"...I'm moving in stereo..."

Moving in Stereo by The Cars

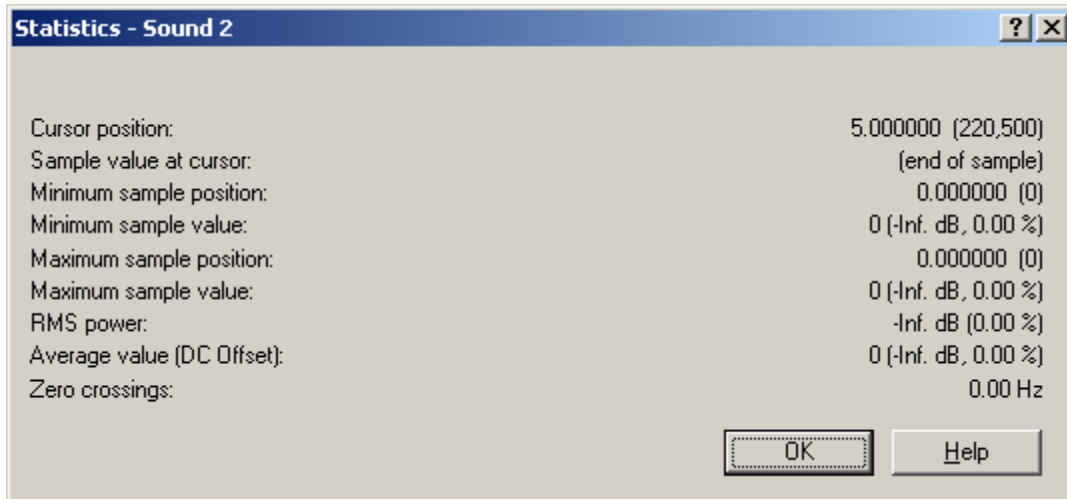
Dither noise should be uncorrelated from the program material. In addition, the left and right channels should be uncorrelated. Otherwise, when your stereo mix fades out you'll hear mono dither become the dominant sound and it will sound, well, mono.

To check for stereo or "uncorrelated left and right dither channels" just take the 24-bit sine wave, dither it, and listen to the noise. To check for stereo, it's perfectly fine to turn it up. You should hear the left and right dither "in stereo". If you don't trust your ears to know whether or not the signal is stereo or not, here's a quick way to check:

- 1) Open the dithered 16 bit file in Sound Forge (or whatever sound editor)
- 2) Invert the left channel, and do a channel conversion to sum the left and right channels into one mono channel.



- 3) Check the statistics of the file. If it was mono, the left was the same as the right, so when you flipped left and summed left and right you'll be left with zero.



You might think you could use a phase meter to check the stereo signal, but it's a little hard with most phase meters (including Ozone's) as the mono sine wave is the dominant signal, so stereo dither noise is overshadowed on a phase meter by the dominant mono sine wave.

"...Listen carefully to the sound..."

Dreams by Fleetwood Mac

The simple tests above will help you recognize that something's wrong with the dither, but to evaluate different types of good dither, there's no way to do it (that we can think of) except to listen to it on real program material at real levels. Put away the sine waves once you're sure the dither is on, you're applying it correctly, and it's working (or at least not broken). Try different program material. Listen to fades, but also listen to material with soft passages (e.g. solo piano, vocalists, etc.). Try different types of dither on different material -- we didn't implement three distinct shapes in Ozone because we were just bored. Different styles for different styles and tastes.

Section IX: “A little less talk a little more action”

A Little Less Conversation by Elvis Presley

With all the theoretical talk and spectrums, we didn't want to leave you without actually knowing how to apply dither. It's easy to make mistakes -- there was one time at iZotope when we thought we had broken our dither in an Ozone alpha version. We finally noticed that the problem was just that the level adjustment in the host app we were testing with had been turned up 1 dB (an inadvertent mouse click in the wrong spot most likely), which will completely mess up any dither. The point is, you can have done this a hundred times and still mess it up with one mouse click.

Rule 1: "...Don't forget to drop me a line..."

Mama Kin by Aerosmith

Steven Tyler was obviously referring to DC offset - having your mix offset by some small amount prior to dithering. DC offset, and even low frequencies below 10 Hz or so, are completely inaudible and only serve to take up precious bits that are needed for dither.

You should filter the DC offset line to zero before you dither. This process is more difficult than it might sound, as many effects cause DC offset and DC offset and/or sub 10Hz signal is not constant throughout a mix.

You can adjust the average DC offset to be zero in most host applications. You can alternatively activate a DC offset filter in Ozone to filter the offset in real time¹⁹. Either way, don't use bits for silence. Get rid of them.

Rule 2: You've to give a little bit

Give a Little Bit by Supertramp

Dithering is a process that changes your audio. DC offset filtering is a process that changes your audio. You've got to give a little bit before you perform these operations, meaning leave yourself a little headroom.

Don't Loudness Maximize to 0 dB and then try to add dither, or even filter DC offset. You could go over.

The order of processing in Ozone is:

- 1) All main modules - EQ, multiband modules, reverb, etc. Loudness Maximizer, if used, should be placed after these modules.
- 2) Output gain.
- 3) DC offset filter
- 4) Dither
- 5) Output meter readout.

¹⁹ The advantage of performing the DC correction with Ozone is that it's after all effects and it's real time. That is, it filters drifting DC offset as opposed to just applying an average offset correction for the mix.

This allows you to adjust the level with the Ozone output gain without ruining the dither (never apply gain change after dither). It also allows you to read out the level of the final dithered and bit reduced signal. But it also means that some processing will happen after the Loudness Maximizer. This is true in general - maximizing so you're peaking at 0 dB and then doing anything else to the signal is a bit of a risk. Leave yourself a dB of headroom, especially if you think you'll dither later. You'll still get played on the radio -- don't worry.

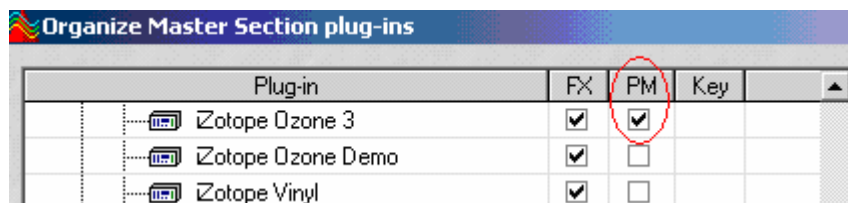
Rule 3: Now don't touch that dial...

Give it to Me by The J. Geils Band

The rule most people know about dithering is don't touch the signal at all after it's been dithered. No fades, no volume change, no normalizing, no effects, no panning. Nothing. OK, the one thing you can do is trim the ends. But no actual processing of the signal.

This is easier said than done in some cases. First of all, most host apps have their master fader after the master effects slot. So, if you adjust the volume with the output fader of the host app, that level adjustment is being applied to the dithered signal, and that's bad. Do any fades or level adjustments at 24-bit destructively before dithering.

One exception that we know of is Wavelab 4, which has a special "dither" slot which is after the output faders. Go to Options-Master Section Plug-ins and check the PM box next to Ozone and it will appear in the Dither effects section.



Another exception would be to use DirectX automation to fade Ozone's output gain (in SONAR, Sound Forge 7, Vegas, Acid 4, for example). The Ozone output gain occurs before Ozone dithering, so you can fade with Ozone and dither at the same time.

In general, though, be sure that:

- 1) Output faders in the host app are set to zero/unity gain.
- 2) Dither is turned off in the host app.
- 3) There are no effects after the dither.

Rule 4: "...One Time, One Time..."

Killing Me Softly as done by The Fugees

Only dither once. Try not to ever re-dither a dithered mix. Go back to the 24-bit mix and dither that. If you think that you might have to re-dither a mix in the future (we can't imagine why you would know that, but...) use Ozone Clear or even Simple dither. Dithering twice is bad at any rate, but applying a psychoacoustic shape dither twice is even worse. Even better, take the advice of the Fugees, "One time, one time."

The one time rule also means making sure that only one thing is dithering the signal when you mix down; meaning, if you dither with Ozone, make sure the dithering in the host app is turned off for mixdown -- which leads to a brief explanation of how dithering in a plug-in actually works.

A plug-in can not change the actual format of the audio it receives. Meaning, it can't get audio in 24 bit format and send it back to the app in true 16 bit format. What happens is that Ozone sends back the same number of bits of audio, but it zeros out all the bits below 16 (or whatever bit depth you select in Ozone). So Ozone applies dither to the 16th bit, zeros out the bits below that, and the host app doesn't care.

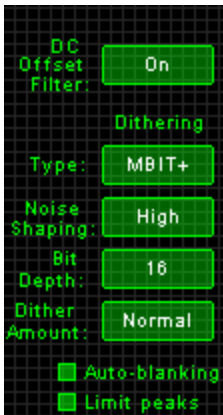
When you actually convert the file to a 16-bit wave with the host app, you can just tell the host app to truncate. Tell it to throw away the bits below the 16th one and don't try to dither the 16th bit. Ozone has already dithered the 16th bit, and the ones below that are zero so they can just be thrown away (truncated by the host app).

This rule also emphasizes why it's important to do absolutely no level changes or processing of the audio after you've applied Ozone dither. We've carefully dithered the 16th bit, and made the bits below that zero. If you do any processing, the host app will use the 16th bit, but also all the bits below it. It's all ruined.

With all those warnings and fear in place, we'll end this trip down dither lane with a step by step for a few apps. The same process should be more or less the same in other apps. The most important point is to find out how to set dithering in the app, and turn it off (assuming you're using Ozone for dithering). And don't do any level changes after dithering, including setting the master fader level to anything but 0 (unity gain).

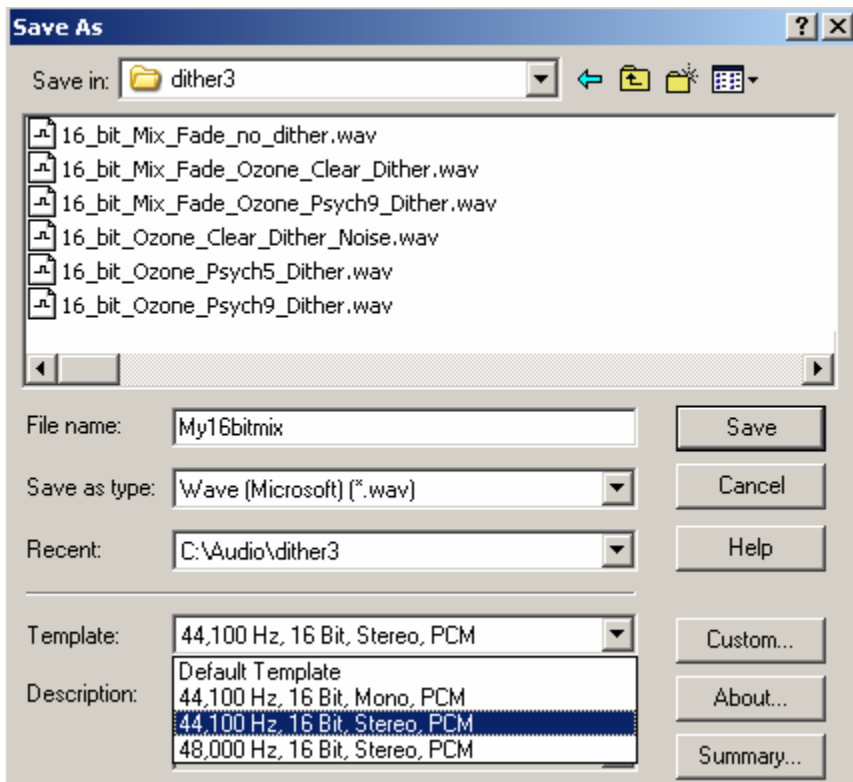
Sound Forge

- 1) Open your 24-bit file.
- 2) Insert Ozone, turn on the DC offset filter, set the dithering to MBIT+, bits to 16, and Shape to suit your taste (as shown below)



3) Click OK to apply the dither (and zero-ing of bits below 16) to your mix.

4) Select Save As, and select 16-bit stereo PCM from the Template drop down list as shown below. When you save the file it will be a 16-bit wave.

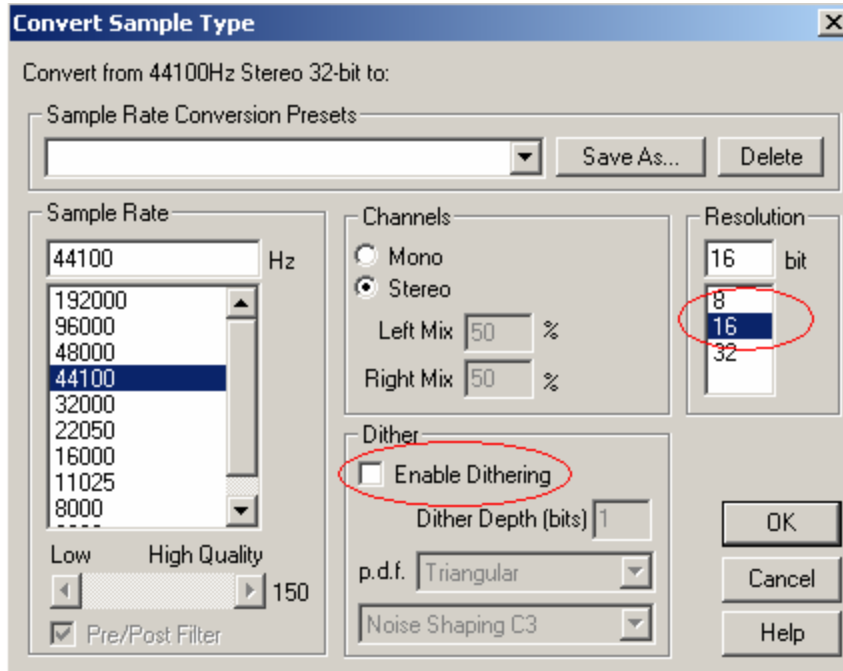


Cool Edit Pro (Adobe Audition)

Same process as Sound Forge to start - open your 24-bit wave file, apply Ozone.

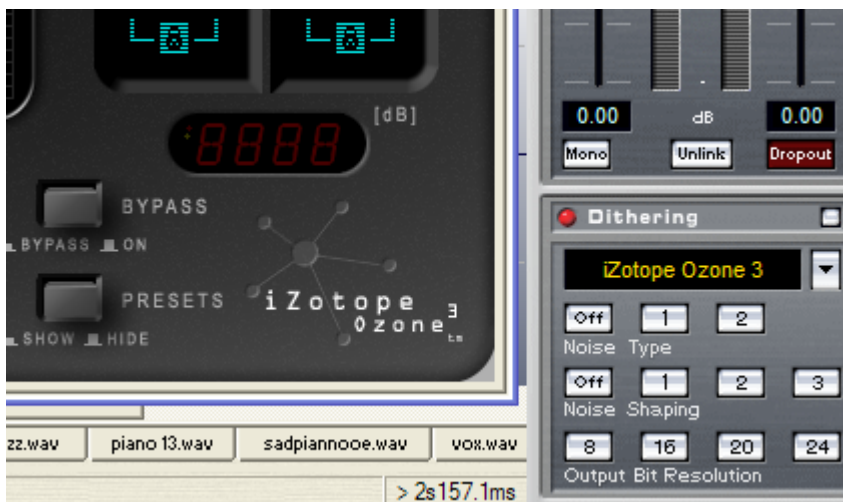
To actually perform the conversion from the 24-bit wave to the 16-bit wave, go to Edit-Convert Sample Type.

The key checkboxes here are circled. Make sure that Enable Dithering is not checked, and select 16 bits as the output format. Click OK, then save the mix.

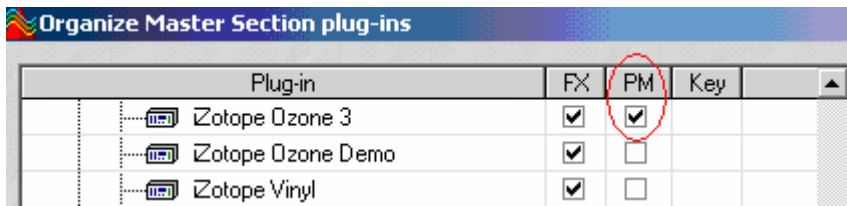


Wavelab 4

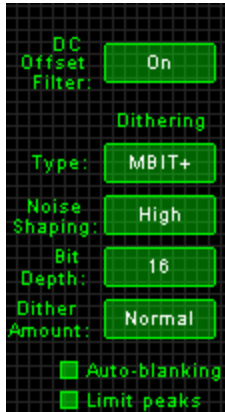
Open your 24-bit mix as a wave file. Select Ozone from the Dithering section list as shown.



If you don't see Ozone in the Dithering list, go to Options-Master Section Plug-ins and check the PM box next to Ozone. It will then appear in the Dithering list.



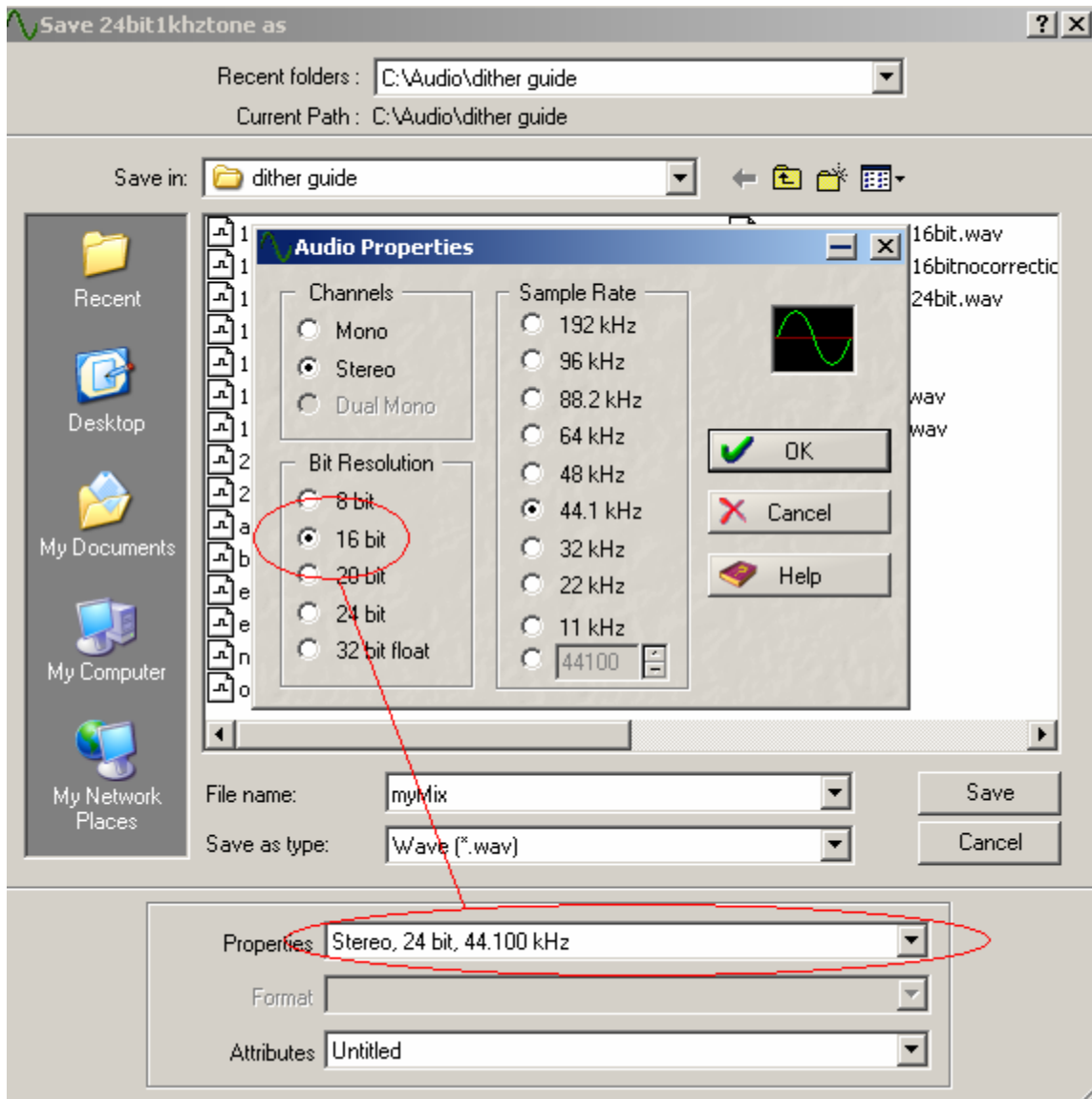
Set Ozone dithering to be as shown below (adjusting the shape to your taste if you like)



Click the render button

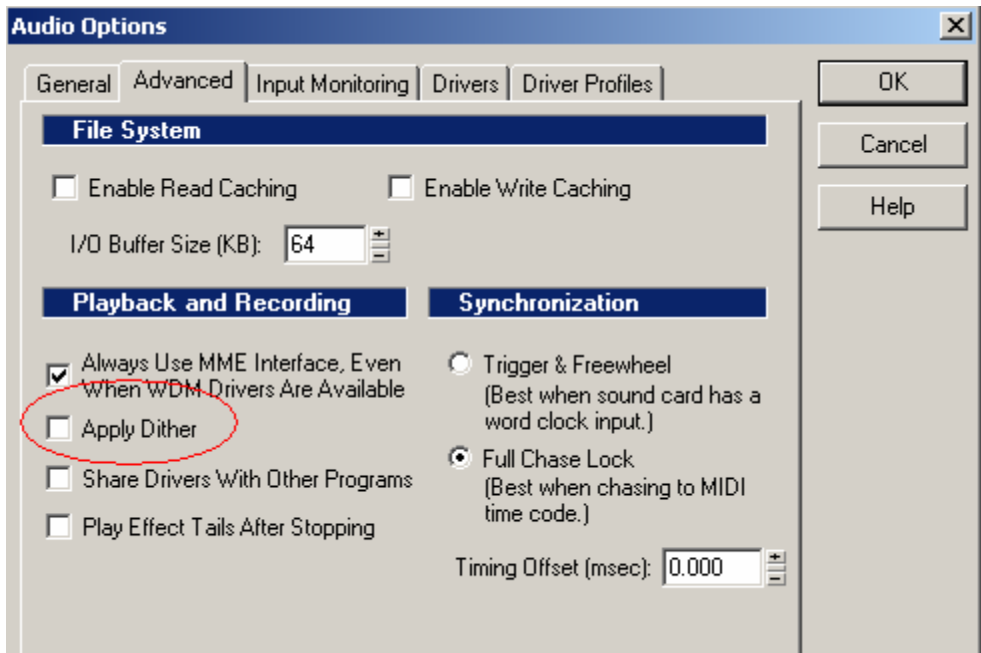


Select File-Save As, and click the "Properties" list. Change the bit depth to 16 bits as shown. Click OK, then Save

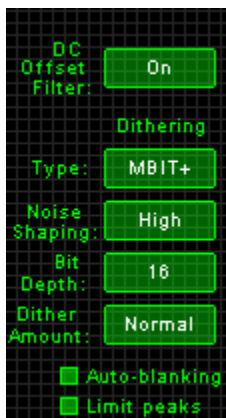


Sonar

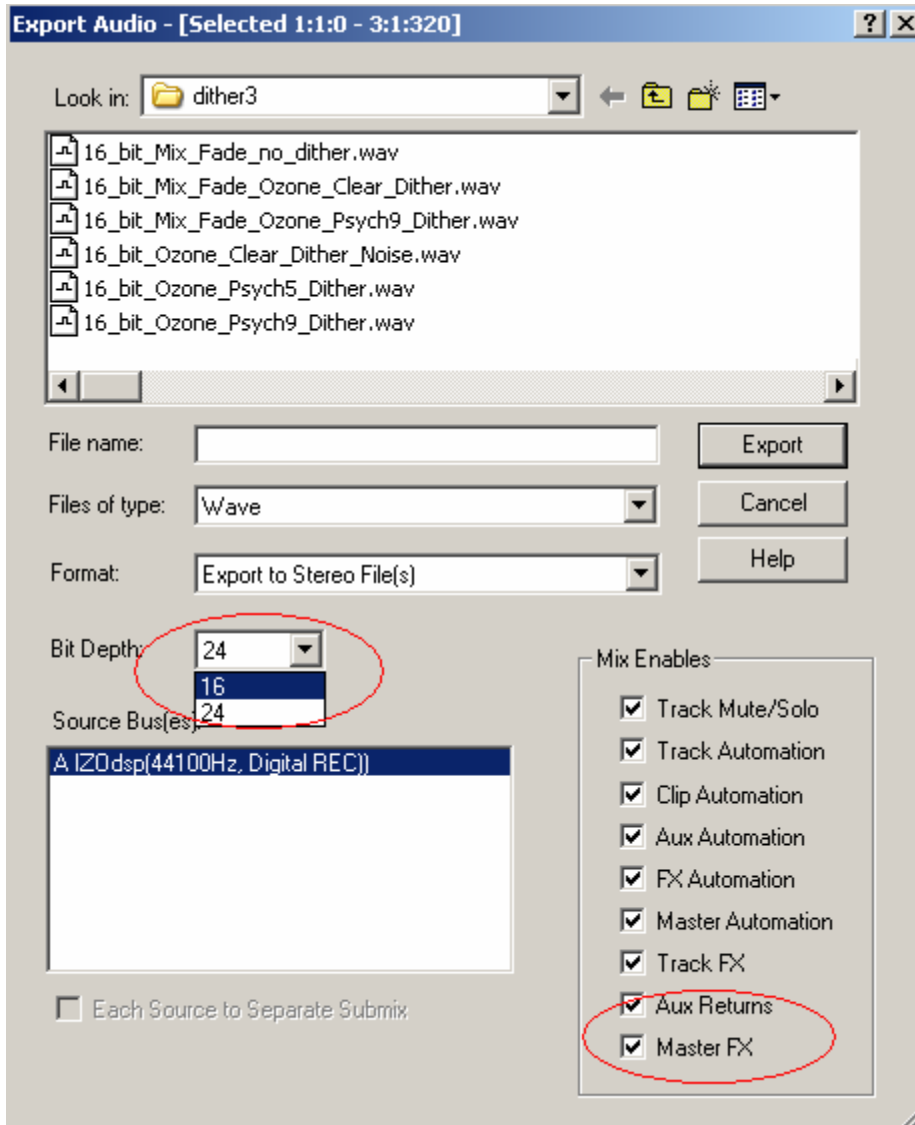
Go to Options-Audio-Advanced Tab and make sure that Apply Dither is not checked.



Insert Ozone as a Master effect, and configure it as shown below



Go to File-Export Audio and make sure that 16 bits is selected as the Bit Depth and that Master FX (Ozone) will be applied during mixdown.



In other apps the process is similar - make sure that dithering is turned off in the app, apply Ozone, and save as a 16 bit file through a Save As, Export Audio or similar Mixdown command.

Section X: Closing Time

Closing Time by Semisonic

We hope this has been of some help. We don't claim to know everything, but what we do know we try to share.

Sincerely,
The iZotope Audio Team
web: www.izotope.com
e-mail: ozone@izotope.com