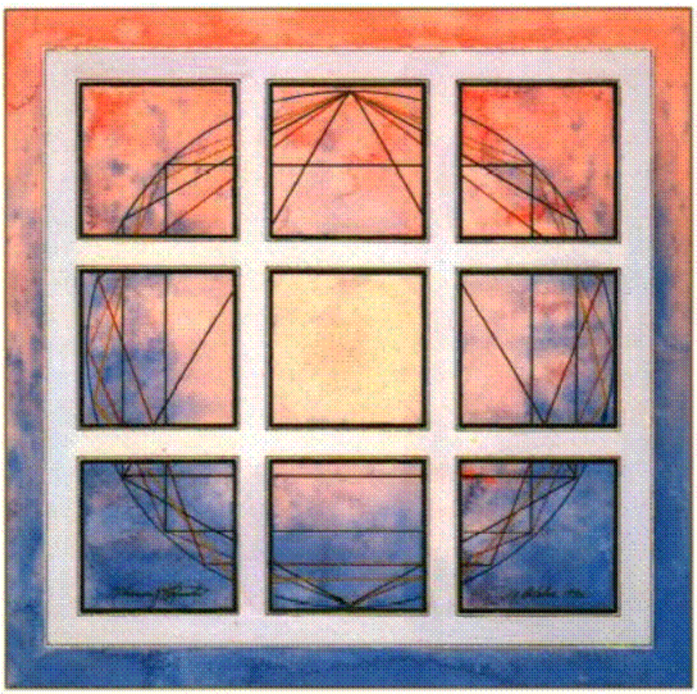
SBM
Super Bit Mapping

Technical Overview



SONY

What is Super Bit Mapping?

Super Bit Mapping (SBM) is a form of digital processing which converts high-resolution signals recorded by a 20-bit master recorder into 16-bit signals which can be recorded on a CD. Although 20-bit signals have 16 times more resolution than 16-bit signals, the SBM process can effectively convert them to 16-bit signals by taking advantage of psychoacoustic principles. As a result, SBM makes it possible to use high quality 20-bit masters in the production of 16-bit CDs which sacrifice little of the original sound quality when played back on home CD players.

In order to understand the remarkable benefits of SBM, let's look at the differences between 16-bit and 20-bit recording.

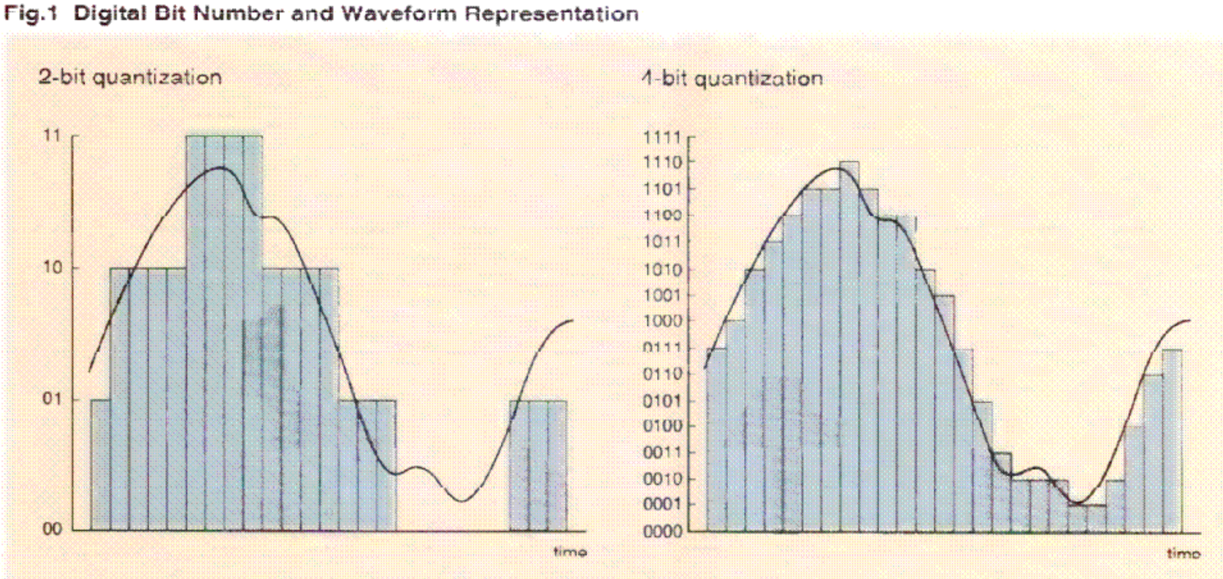
The Connection Between Sound Quality and the Number of Bits

There is a strong relationship between the number of bits (the wordlength of a PCM data word) and sound quality. In the digital realm, sound is expressed in "1" and "0." These, the smallest building blocks, are called "bits." A single bit represents only two possibilities, 1 or 0. With two bits, however, there are four combination possibilities: "11," "10," "01," and "00." Accordingly, with three bits there are eight (111, 110, 101, 100, 011, 010, 001, and 000) and with four bits the combinations double again to sixteen (1111, 1110, 1101, 1100, 1011, 1010, 1001, 1000, 0111, 0110, 0101, 0100, 0011, 0010, 0001, and 0000). As you can see, each time a bit place is added, the number of combinations doubles.

With this in mind, we can see that a word length of 16 bits provides two to the sixteenth power, or 65,536 different combinations. Similarly, adding only four bits to create 20 bits results in a startling number of combinations: 2 to the twentieth power, or 1,048,576 combinations. This is sixteen times the 65,536 combinations possible with 16-bit recording. Essentially, the greater the number, the higher the sound resolution as each combination represents one sound level. In simple terms, 20-bit recording can record a much greater amount of data, making possible a more accurate representation of the music.

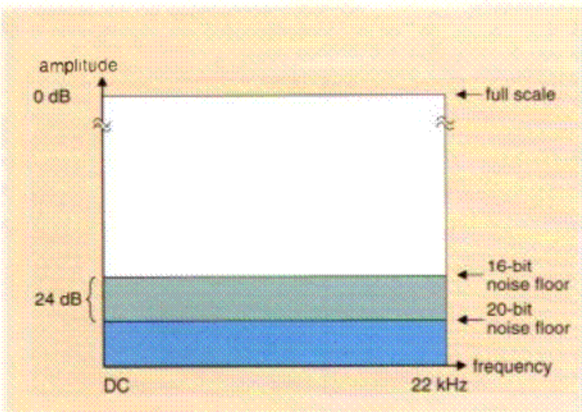
As an example, what is the difference between digitizing a sound wave by 2 bits or 4 bits? Fig.1 provides a good illustration. With 2 bits, there are only four levels with which the sound wave may be represented. But with 4 bits, there are sixteen different levels, so the waveform may be more accurately represented by the different combinations. Thus, the greater number of bits, the closer the resulting sound wave is to the original, with a smoother waveform and less conversion error, commonly referred to as quantization noise.

Fig.1 Digital Bit Number and Waveform Representation



The frequency characteristics of noise resulting from this conversion error are shown in Fig.2, which shows the noise floor levels (levels of quantization noise) for both 16 bits and 20 bits. Since 20-bit recording has much less conversion error, its noise floor is some 24 dB lower.

Fig.2 Digital Bit Number and Quantization Noise Floor



Decibels (dB) is a logarithmic unit of measure for both dynamic range and S/N ratio, with a 6 dB increase representing a doubling of the sound volume over a given level. Accordingly, a drop of 6 dB represents a halving of sound level.

Since the conversion error of 16 bits is 16 times greater than that of 20 bits, the noise is 24 dB higher. That is, the noise level is "doubled" four times.

Referring to the comparison of sound waves in Fig. 1 and the comparison of noise floor levels in Fig. 2, sound reproduction possible with 20 bits is smoother, with higher fidelity and less noise, than with 16 bits. As the fidelity to the original sound source is improved, we can also expect improvements in low-level signals which can be considered those which represent "depth" and "soundstage."

Processing when the number of bits differ

The advent of the 20-bit digital recorder represents a marked improvement in studio sound, with the output being almost identical to that of the mixing console. However, there remains the question of how best to convert 20 bits into the 16 bits required for CD. While the simplest way was to truncate the last four bits to make 16 bits, it is clear that this is not the best approach.

The problem with truncation is easily understood by looking back at the relationship between the number of bits and resulting sound quality. For example, while there are eight combinations in 3 bits, if the last digit is left off, the number of combinations is reduced to four, the same as with two bits. Thus, if the last four bits of 20 bits are left off to achieve 16 bits, then the number of combinations, and the resulting sound quality, will be the same as with 16 bits. Given this, there is no point in using a 20-bit recorder instead of a 16-bit unit.

Indeed, this truncation of 20 bits into 16 would be worse than an original 16 bit recording because the truncation recording would not have a dither signal in its 16th bit.

Psychoacoustics: human auditory perception

As audiophiles with test discs may have experienced, when listening to sine waves of different frequencies with the same amplitude, the perceived volume of sound differs. Though we may not be aware of this phenomena in everyday life, it is one distinctive aspect of human auditory perception.

These characteristics have been scientifically measured, and have been compiled onto what is known as the "Loudness Curve," which is shown in Fig.3. These are the amplitudes of different frequencies which produce the same perceived volume. In simple terms, the human ear is more sensitive at certain frequencies than others. Though there are slight variations at different sound pressure levels, the general tendency is for sensitivity to be highest between 500 and 5,000 Hz. Not shown is the data for very high frequencies where the sensitivity of the human ear is poor and essentially unmeasurable.

Interestingly, this range between 500 and 5,000 Hz centers on the range of the human voice and many musical instruments. Thus, if the resolution of this "highly audible at low volume" zone is improved, the perceived fidelity will greatly increase.

As Fig.2 shows, 20-bit recording allows the reproduction of sounds 24 dB quieter than with 16-bit recording throughout the entire band, including the "highly audible at low volume" range from 500 - 5,000 Hz. Making sound seem improved to human ears is to make it so that "even the smallest sounds are more audible" than those reproduced by 16 bits, particularly in the "highly audible at low volume" range.

Fig.3 Loudness Curve

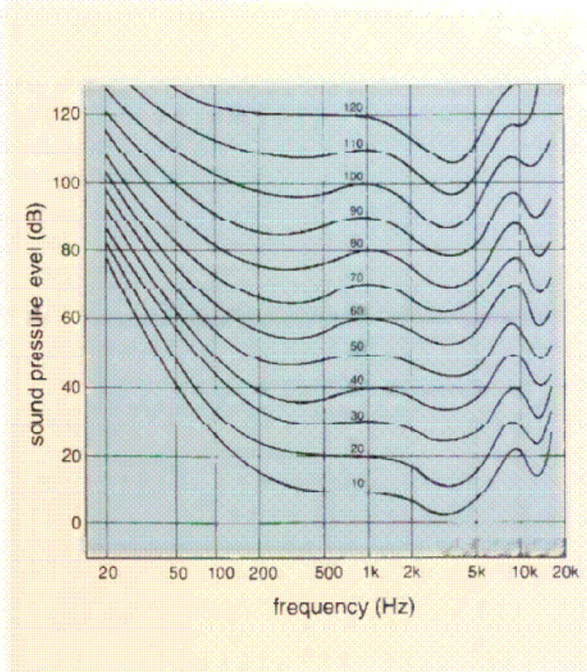
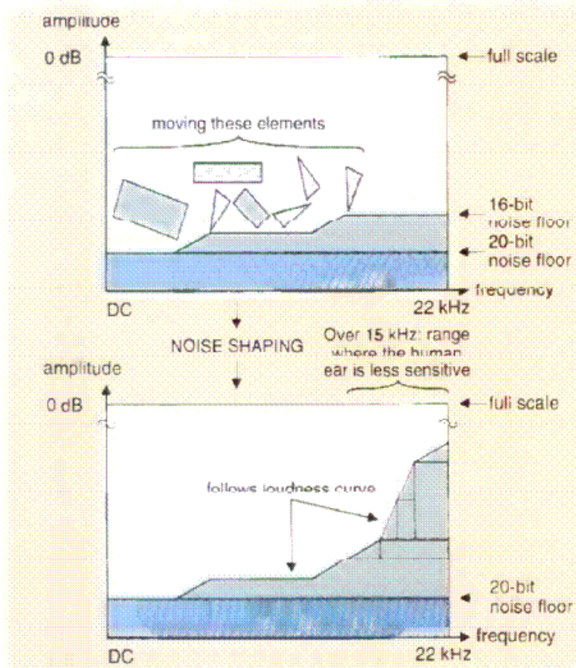


Fig.4 Principle of Super Bit Mapping



The Principle of SBM

Now we will explain how SBM converts a 20-bit signal into a 16-bit signal without losing sound quality. Fig.2 shows quantization noise floor levels for both 16-bit and 20-bit signals. The total energy of the quantization noise is calculated by adding up the quantization noise values across the frequency range. This is determined by the number of quantized bits. In the figure, the areas of the shaded rectangles for both 16 bits and 20 bits do not change size as long as the number of bits does not change. However, even though the area size (total noise energy) cannot be changed, it is possible to "reshape" the pattern of noise using Sony's original Noise Shaping technology, which has been incorporated in SBM.

Taking into account the previously explained psychoacoustic principles, it is possible to shift the quantization noise from the frequency range where the human ear is most sensitive to a range where the ear is less sensitive. This makes it possible for 16-bit reproduction to sound like 20-bit reproduction.

Specifically, SBM reorients quantization noise to the frequencies above 15,000 Hz where the human ear is far less sensitive. As a result, deep bass and mid-range sound are far more dynamic. Fig.4 illustrates how the quantization noise is re-distributed into the higher frequencies, even though its total volume remains unchanged.

Noise Shaping and SBM

Those familiar with audio technology realize, however, that one change often causes other changes. With this Noise Shaping technique, many may wonder how the musical waveforms will be affected when the frequency characteristics of the noise floor are changed.

How 20-bit sound information is obtained from 16 bits can be explained using a simplified wave. Fig.5 shows a very gently sloped wave quantized at 20 bits and at 16 bits. One question which soon arises is why there are so many small pulses in SBM. Actually, these many small pulses are part of the secret of SBM operation. For easier understanding, a part of the SBM processed wave in Fig.5 is shown in Fig.6. The solid line represents the SBM wave, while the dotted line shows the 20-bit wave.

The following explanation uses the expression LSB, meaning Least Significant Bit. Essentially, the LSB is 1 divided by $2^N - 1$, with N being the number of bits. For example, the LSB of 16 bits is: $1/2^{16} - 1$, or 1/65,535.

Fig.5 Digitizing a Smoothly Sloped Waveform

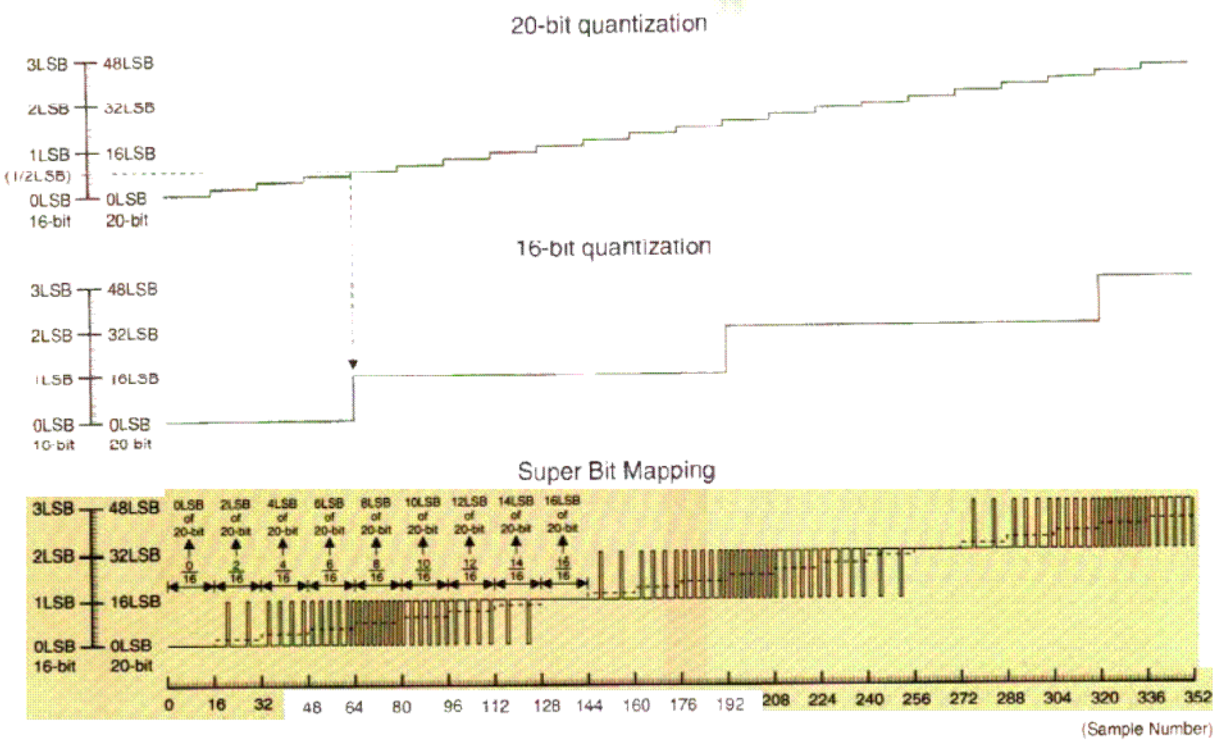
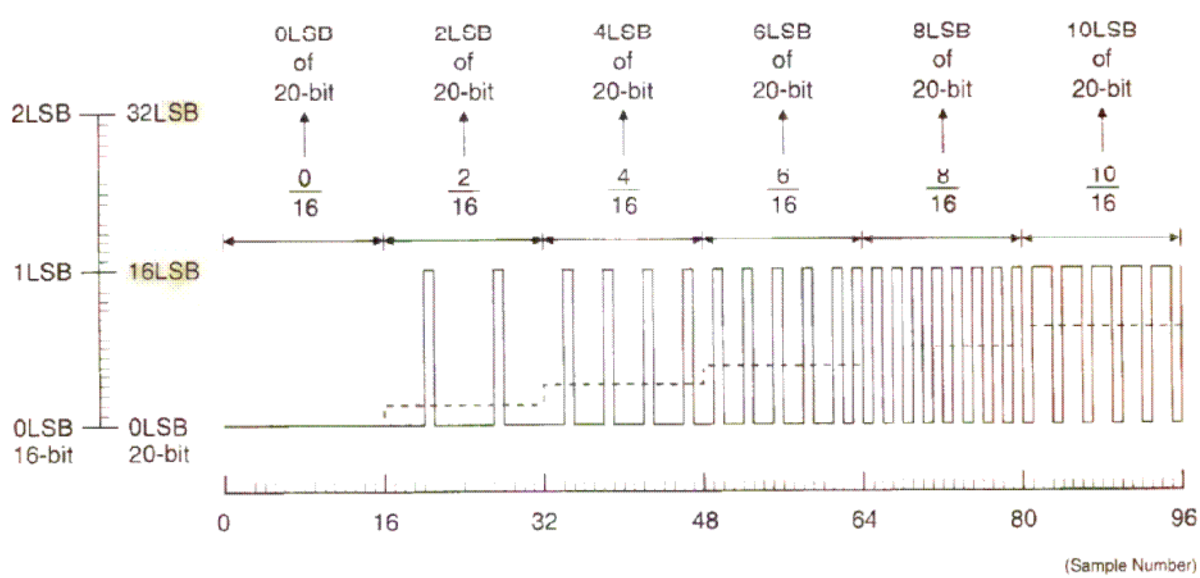


Fig.6 Super Bit Mapping Waveforms



Looking at the 16 samples, from the 16th to the 31st sample, in Fig.6, although there are two 20-bit LSBs in 20 bits (equal to 2/16 LSB of 16 bits), SBM has only two pulses with the height of the 16-bit 1 LSB sample while all the others are 0 LSBs. If the average of the SBM wave, from the 16th to the 31st sample, is calculated it will be 2 LSB - 16 = 2/16 LSB, expressing the same amount of information as with 20-bit operation. Similarly, if the average of the SBM wave from the 32nd to the 63rd sample is calculated, it will be 4 LSB - 16 = 4/16 because there are 4 pulses with the height of the 16-bit 1 LSB sample, expressing the same information as with 20-bit operation.

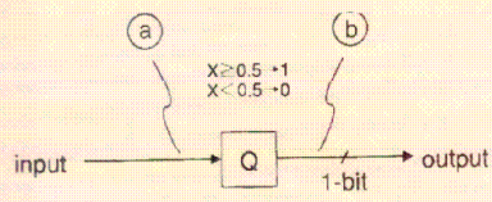
In this way, SBM expresses 20-bit information through averaging 16-bit 1 LSBs which appear in a given range. This phenomenon is similar to the reproduction of photos in newspapers wherein light and shade density are expressed by the size and density of very small dots.

Along with this high density, however, there are pulses with the height of a 16-bit 1 LSB sample. These are concentrated mainly in the high frequency range, centering on 22,050 Hz (half of the CD sampling rate of 44,100 Hz). These groups of pulses generated to express 20-bit information were neither in 16 bits nor in the original waveform, so they can be regarded as one form of noise. While noise in the high frequency range is increased by these groups of pulses, the human ear does not easily recognize them as noise since they are gathered above 15,000 Hz where human hearing is not so sensitive.

This "noise gathering" process is called Noise Shaping, wherein errors generated in re-quantization are fed back to the input in order to reduce operation error. Primary noise shaping, the simplest form, will be used to explain the principle of signal processing. Signals generated by feedback are explained to clarify the principle of its operation.

Fig.7 shows the operation of a very simple 1-bit quantization of input signals, while Fig.8 shows the same operation when primary noise shaping is employed. If a constant "0.6" is input into the simple 1-bit quantizing process, output is always "1" (as is shown in Fig.7) because the input is greater than "0.5." With this, a difference of "0.4" always arises. This is for the same reason that (as shown in Fig.5) the fine steps below 1 LSB of 16-bit operation were lost when the 20-bit signals are rounded off to 16 bits.

Fig.7 Simplified 1-bit Quantization



sample	1	2	3	4	5	6	7	8	9	10
input	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
point ⓐ	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
point ⓑ	1	1	1	1	1	1	1	1	1	1
output	1	1	1	1	1	1	1	1	1	1

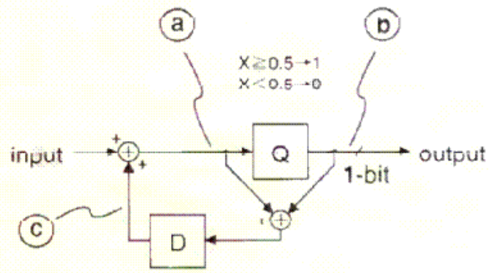
$$\text{average value } \frac{1+1+1+1+1+1+1+1+1+1}{10} = \frac{10}{10} = 1.0$$

When a constant "0.6" is input to a 1-bit quantizing operation in which primary noise shaping is performed, output will vary as shown in Fig.8, and following the changes in the signal as indicated in the table will explain the operation.

If at first the value of D (the delay element) is 0, and a value of 0.6 is then input, the signal at point a will be 0.6 and the value at point b will be 1 if this value is quantized by 1 bit. If the input, however, continues to be 0.6 and the sum of the values of d and the input will be 0.2, then the value at point b will be 0 if this value is quantized by 1 bit. At this point, the value at point c will be the difference between the values at points a and b, namely 0.2, with this signal then becoming delayed by one sample to formulate the new delay element value D.

The output thus varies by the signals fed back to reduce operation error, with the ratio (average value of 10 samples) of "1" the same as the input value of 0.6 if the average of the output signals in a given time frame (10 samples in this case) is calculated. In simple terms, SBM is a form of signal processing on the lower 4 bits of 20 bits when quantizing 20 bits as 16 bits.

Fig.8 1-bit Quantization with 1st-order Noise Shaping



sample	1	2	3	4	5	6	7	8	9	10
input	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
point ⓐ	0.6	0.2	0.8	0.4	1.0	0.6	0.2	0.8	0.4	1.0
point ⓑ	1	0	1	0	1	1	0	1	1	1
point ⓒ	-0.4	0.2	-0.2	0.4	0	-0.4	0.2	-0.2	0.4	0
output	1	0	1	0	1	1	0	1	0	1

$$\text{average value } \frac{1+0+1+0+1+1+0+1+0+1}{10} = \frac{6}{10} = 0.6$$

Actual Noise Shaping Operation in SBM

If errors are fed back, even though the fidelity of a certain frequency range is improved, the noise level is increased in other frequency regions, as shown in Fig.4. Fundamentally, the shape of the noise floor is "re-shaped" by error feedback, which is why this approach is called "noise shaping."

In the simple feedback circuit shown in Fig.8 the frequency characteristics of quantizing noise take a simple form, with a gradual increase towards the high frequencies, as shown in Fig.9, with insignificant improvements at middle and low frequencies.

Noise shaping which achieves frequency characteristics that take advantage of the principles of psychoacoustics cannot be performed by such designs, so more complicated feedback circuits are employed in SBM. Fig.10 shows such a noise shaping circuit used in SBM. In order to control the frequency characteristics of quantization noise, errors are fed back through a 12-tap digital filter. The coefficients of these 12 filters determine the frequency characteristics of the quantization noise.

In SBM, though, how are the frequency characteristics of quantization noise actually determined? This is explained in Fig.11, which shows the frequency characteristics of SBM quantization noise. When compared to the loudness curve of Fig.3 you can see that the frequency characteristics of quantization noise in the middle and high frequencies are quite similar, although the signal around 10,000 Hz is suppressed, and there is no increase in low-frequency noise.

Why does SBM use a different curve than the loudness curve? When SBM was first being developed, the standard loudness curve was adopted for noise shaping. Comparative listening tests, involving a broad spectrum of studio engineers, producers, and music industry figures revealed a number of peculiarities in the middle and high frequencies, although sound reproduction was definitely improved over ordinary 16-bit operation.

Subsequent listening tests, using a 20-bit master tape as a control, allowed us to eventually arrive at noise shaping frequency characteristics that would achieve sound reproduction closest to that of the 20-bit master tape in all parameters of music information. The final result is shown in the SBM curve in Fig.11

While there are a number of reasons why this curve differs from the standard loudness curve, one explanation is that the standard loudness curve measures human sound perception using a sine-wave sweep, while music sources have a wide spectrum with interaction causing shifts in frequency characteristics.

Fig.10 Super Bit Mapping

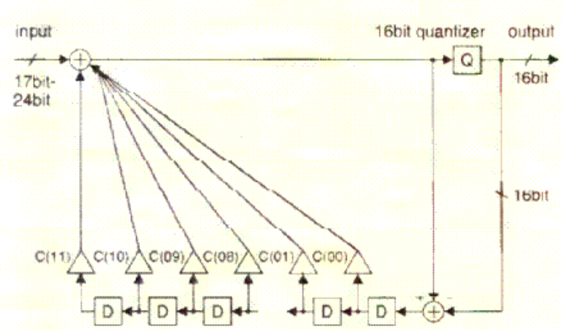


Fig.9 Frequency Characteristics of 1st-order Noise Shaping

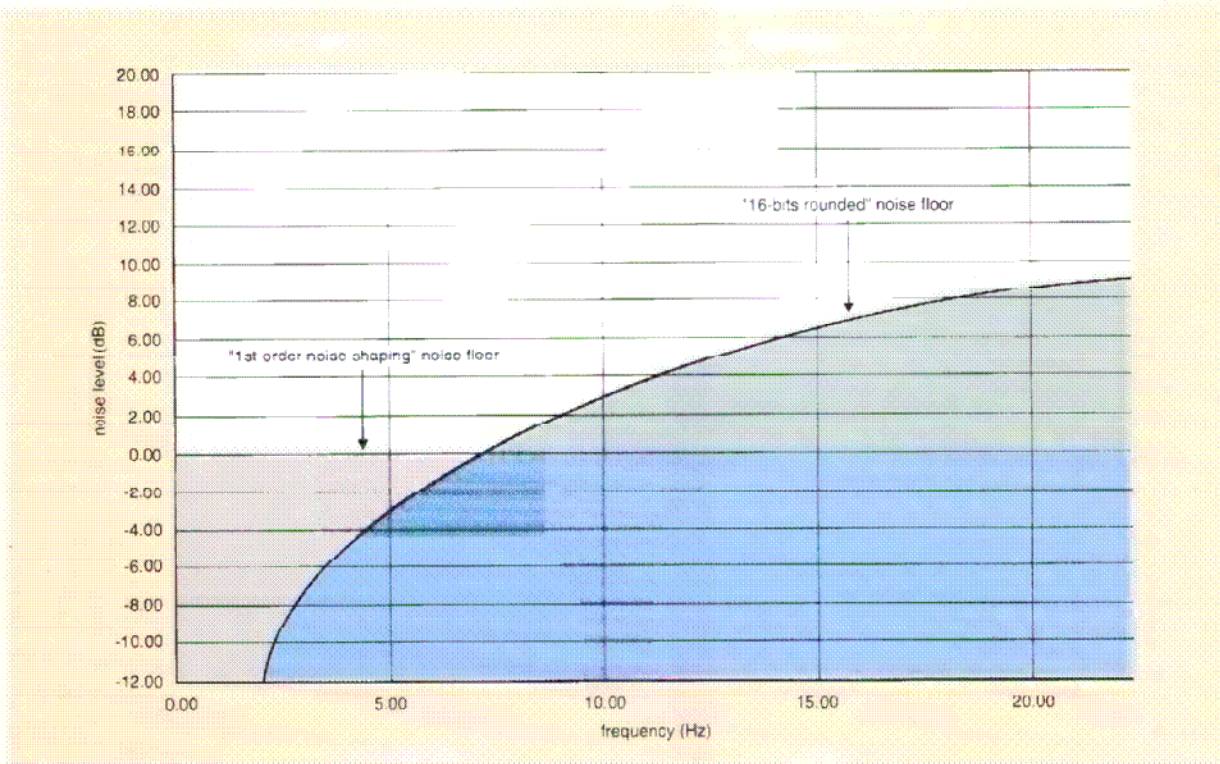
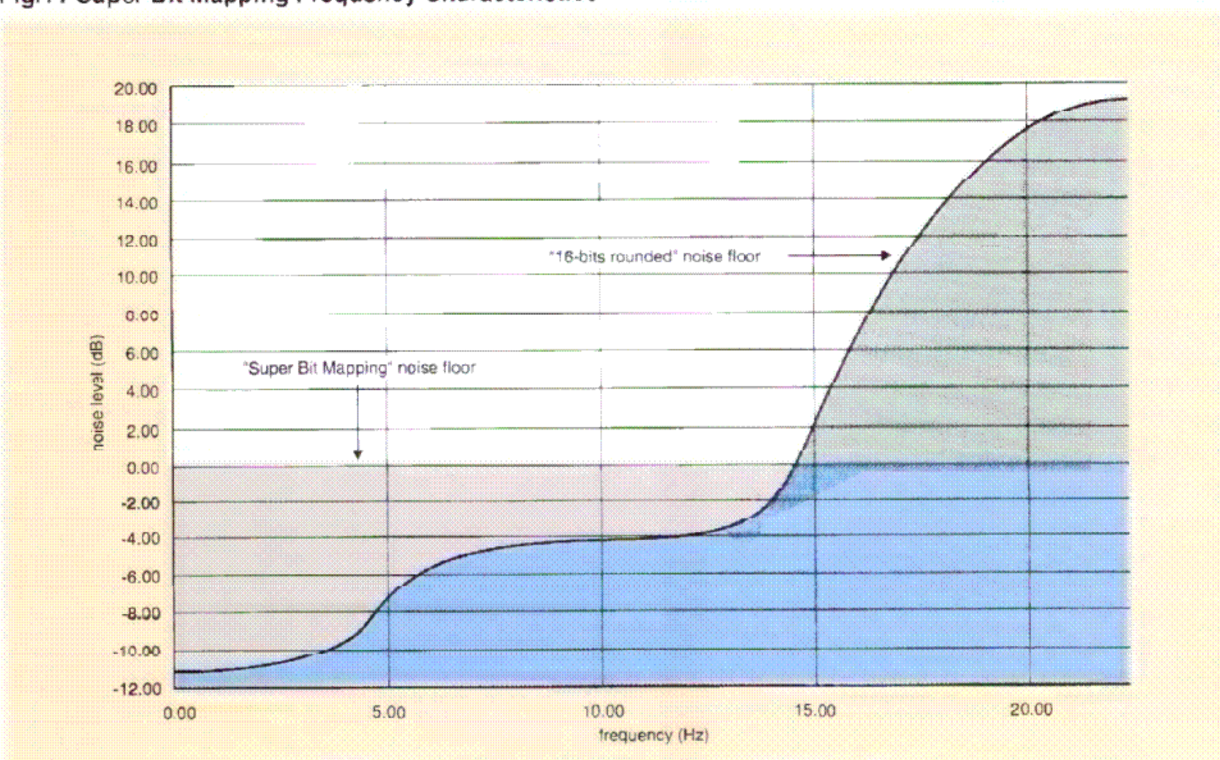


Fig.11 Super Bit Mapping Frequency Characteristics



Actual SBM Waveforms

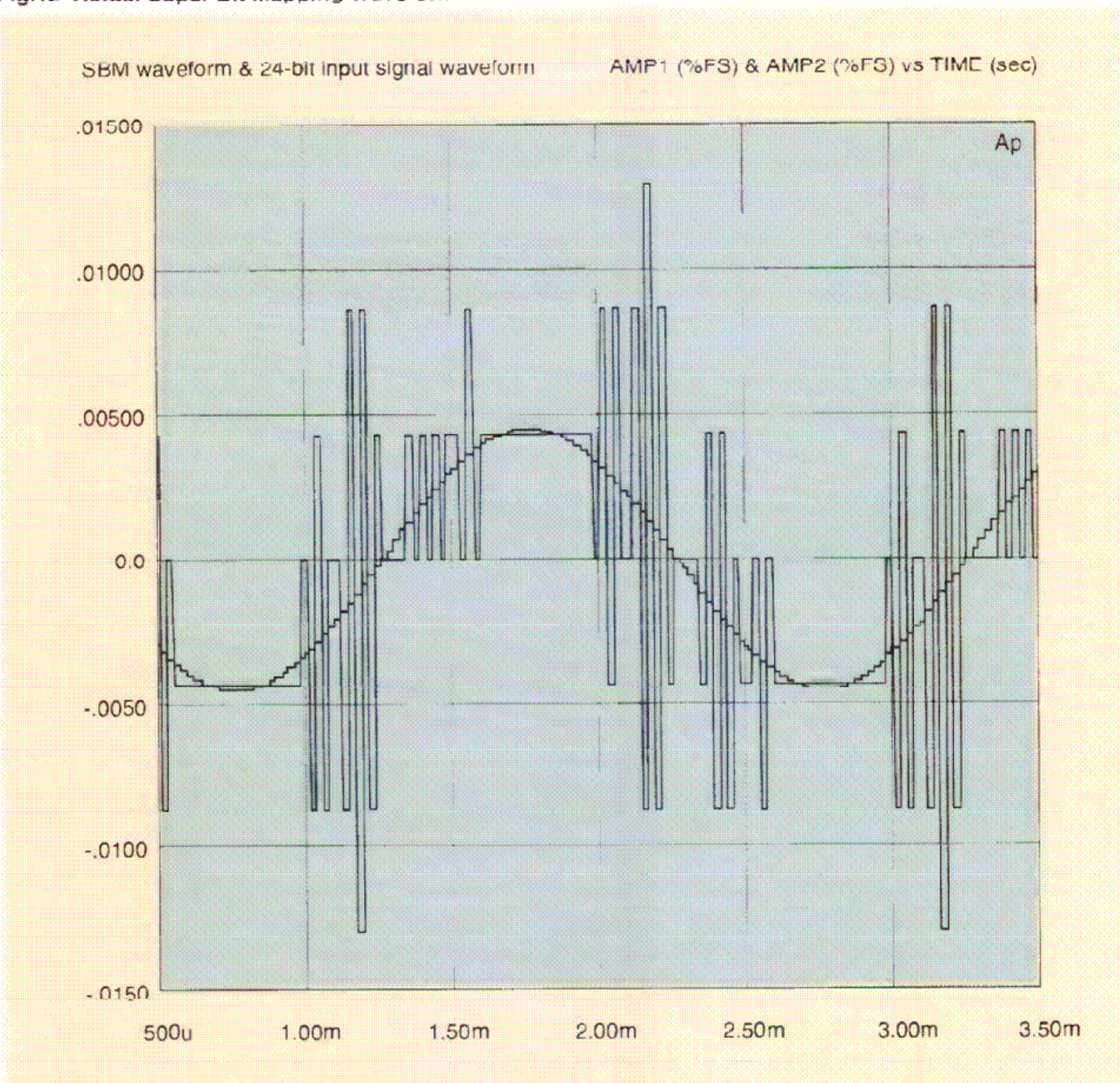
Feedback circuits become rather complicated, as shown in Fig.10, and it may be easy to conclude that the actual SBM wave signal be more complicated than what is shown in Fig.5.

In actual SBM operation, 1 LSB pulses are only generated in primary noise shaping, and pulses of different sizes are generated if the number of filter steps are increased to allow improved resolution at middle and low frequencies. In SBM noise shaping, pulses which are 1 to 5 steps in size (referred to 16 bit LSBs) are generated. As the number of steps is increased, noise in the high frequencies increases accordingly. In contrast, however, resolution at middle and low frequencies is improved by a number that can be expressed by a simple equation:

pulse group density X number of steps

Fig.12 shows a 500 Hz sine wave at -90 dB generated by 24 bits, with the audio signal converted by SBM into 16 bits. Pulses 1 to 5 steps in size are generated, which follow the 24-bit sine wave.

Fig.12 Actual Super Bit Mapping Waveform



About Characteristics

Let us now add an explanation of "16 bits rounded with dither" to our explanation of waveforms and other characteristics. Essentially, "16 bits rounded with dither" means that the original 20-bit signal is rounded to 16 bits only after random noise is added to the last 4 bits of 20 bits, while "16 bits rounded" means the signal is only rounded from 20 bits to 16 bits. Since random noise is added in "16 bits rounded with dither," the S/N ratio deteriorates slightly. One merit, however, is that signals below 16 bits are output as modulated signals of random noise.

Fig.14 shows reproduced waves of 20 bits, 16 bits rounded, SBM, and 16 bits rounded with dither. These are measured by a very low-level sine wave of -90 dB and filtered by low-pass filters with a cutoff frequency of 16,000 Hz. With 16 bits rounded, triangular dither of 16 bits + 1 LSB is used. (Measurement was performed according to the simulation shown in Fig.13)

Looking at the overall picture of the waveform, though much of the original sine wave information is retained, the dither noise is present in roughly half the amplitude. Although the result matches well with human auditory perception, it seems as though the sound is half-buried in noise.

Fig.15 shows spectrums of outputs from a 20-bit converter in 20 bits, 16 bits rounded, SBM, and 16 bits rounded with dither when playing a 1,000 Hz sine wave at the very low level of -90 dB. Note that the noise floor of the 20-bit spectrum is low and flat, without any higher harmonics.

In 16 bits rounded, the noise floor is high, not surprisingly, and many higher harmonics are generated. In 16 bits rounded with dither, the noise floor is still higher, although no higher harmonics are generated.

In SBM, no higher harmonics are generated, and the noise floor in the range below 5,000 Hz is approximately the same as with 20 bits, however noise increases above 15,000 Hz.

In Fig.16, a 1,000 Hz sine wave is input to a 20-bit converter and its level is decreased from full scale down to -120 dB. It is shown with outputs by 20 bits, 16 bits rounded, SBM, and 16 bits rounded with dither, to measure signal linearity.

While 20 bits shows nearly perfect linearity, 16 bits rounded shows that signal linearity deteriorates in low-level signals below -90 dB. Similarly, in 16 bits rounded with dither the linearity is good, but noise stands out below -90 dB.

In SBM, the input/output linearity of low level signals below -90 dB is good, and only slightly inferior to that of 20 bits.

Fig.13 Simplified Listening Model Simulation

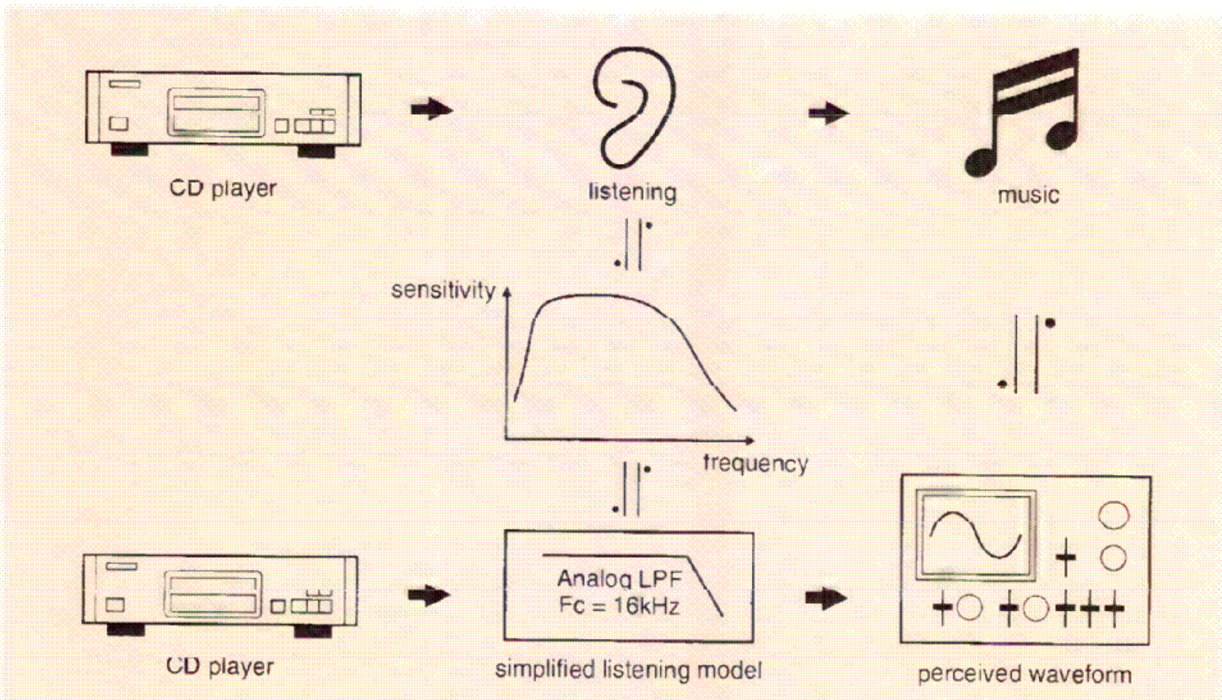
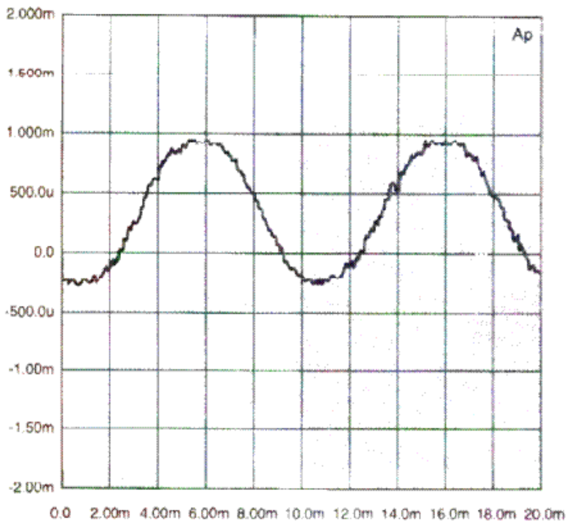


Fig.14 Waveform Comparison by Simplified Listening Model

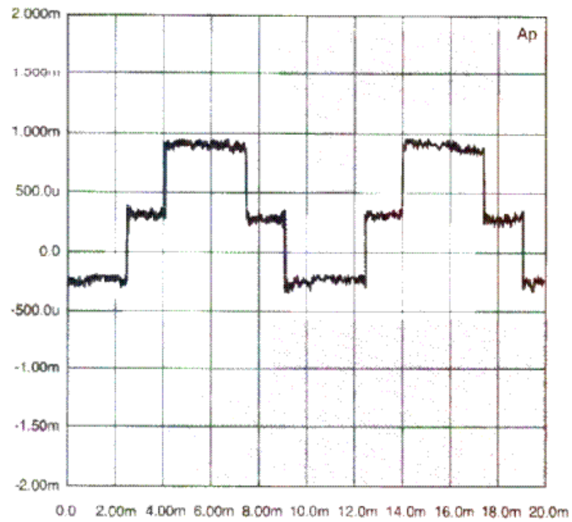
20-bit

20-bit, 100 Hz, -90 dB, DC-16 kHz bandwidth AMP1(V) vs TIME(sec)



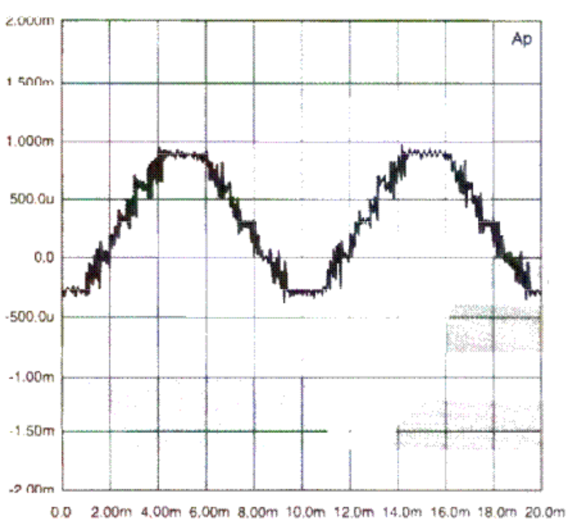
16 bits rounded

16-bit, 100 Hz, -90 dB, DC-16 kHz bandwidth AMP1(V) vs TIME(sec)



Super Bit Mapping

Super Bit Mapping, 100 Hz, -90 dB, DC-16 kHz bandwidth AMP1(V) vs TIME(sec)



16 bits rounded with dither

16-bit with dither, 100 Hz, -90 dB, DC-16 kHz bandwidth AMP1(V) vs TIME(sec)

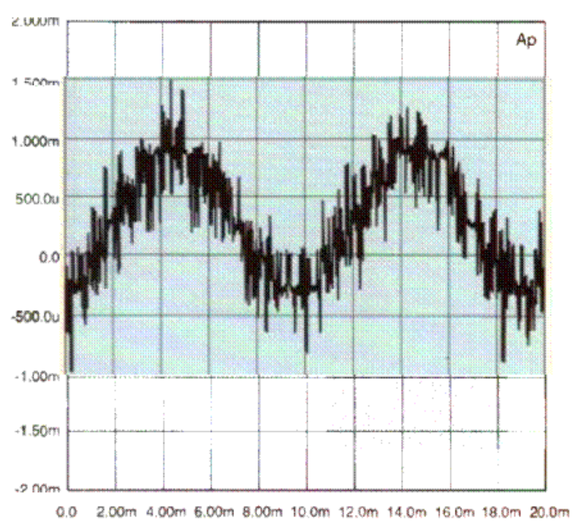
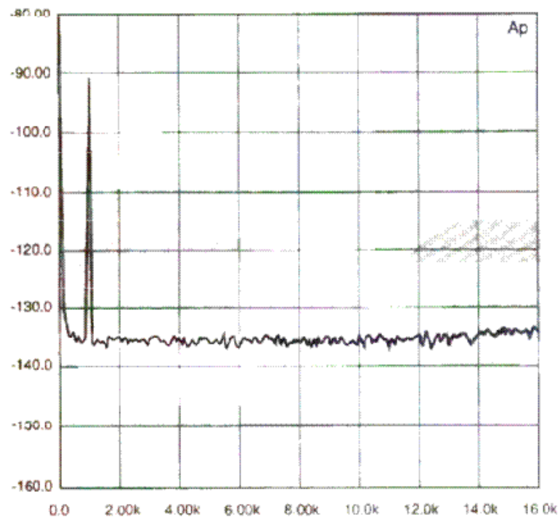


Fig.15 1 kHz / -90 dB Sine Wave Spectrum Comparison

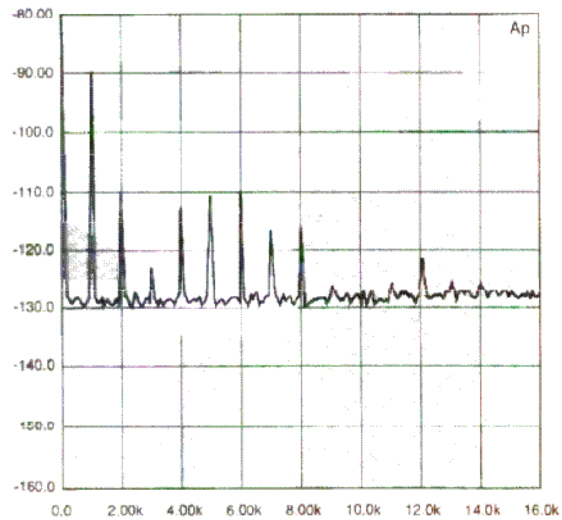
20-bit

20-bit AD (1kHz, -90 dBFS) -> 16-bit Thru. DC -16 KHz amp1 (dBFS) vs frequency (Hz)



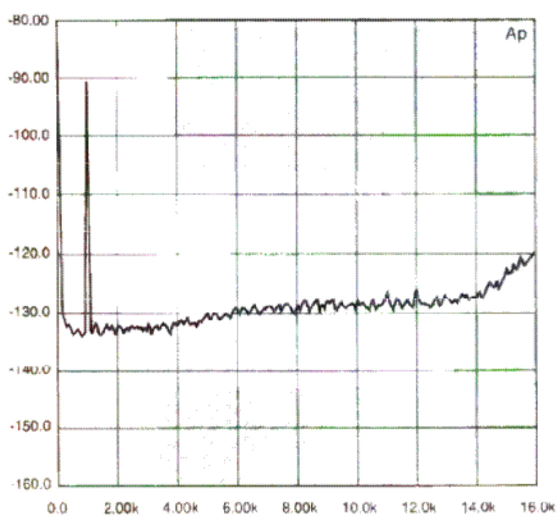
16 bits rounded

20-bit AD (1 kHz, -90 dBFS) -> 16- bit rounding, DC-16 KHz amp1 (dBFS) vs frequency (Hz)



Super Bit Mapping

20-bit AD (1 kHz, -90 dBFS) -> Super Bit Mapping, DC-16 KHz amp1 (dBFS) vs frequency (Hz)



16 bits rounded with dither

20-bit AD (1 kHz, -90 dBFS) ->16-bit rouding with dither, DC-16 KHz amp1 (dBFS) vs frequency (Hz)

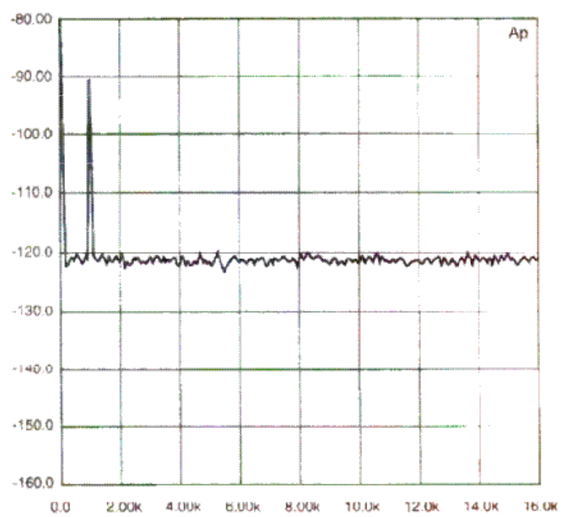
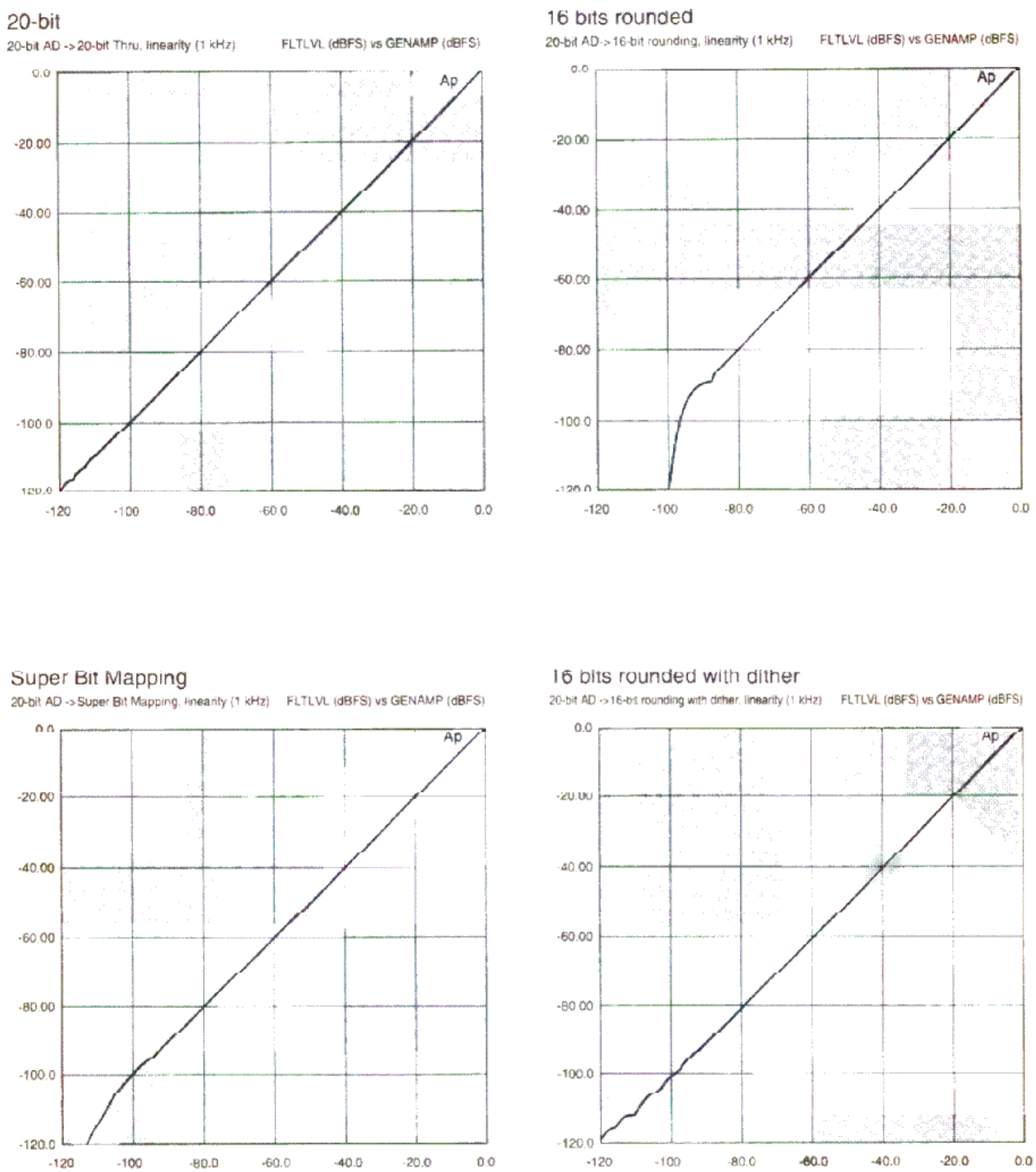


Fig.16 1 kHz Sine Wave Signal Linearity Comparison



Achieving Satisfactory SBM Reproduction

To maximize the benefits of SBM, there are two major points which require attention in relation to playback hardware.

As Fig.11. indicates, SBM improves the S/N ratio by some 12 dB in the middle and low frequency ranges over commonly used 16 bits with dither. This, however, is a characteristic in the digital state, and some amount of added noise can be expected after D/A conversion. If the noise floor after conversion is not lower than that of the SBM noise floor, the amount of S/N improvement will naturally be lower. Fig.17 explains this relationship between the two noise floors.

If the S/N ratio after conversion is 100 dB (compared to the original digital signal and S/N ratio is within the range limited by 22,000 Hz and without level weighting), the S/N ratio in the lower and middle ranges is only improved by 2 dB because the theoretical S/N threshold of 16 bits is 98 dB. However, if the S/N ratio is 110 dB after conversion, the S/N ratio of the middle and lower ranges is improved by 12 dB, as can be seen in the middle graph of Fig.17. If the S/N ratio is 120 dB after conversion, the sound source can be reproduced taking full advantage of SBM.

While it may be easy to draw the conclusion that if hardware having an S/N ratio less than 110 dB is used for playback, SBM has little to offer in terms of actual sound improvement over original 16-bit operation, note that the total S/N ratio of conventional 16-bit recording in CD production is about 90 dB, and the total S/N ratio when a CD is played back does not go over 90 dB, even if the S/N ratio after conversion is over 90 dB.

However, the S/N ratio of a CD produced using SBM processing is 100 dB, so some S/N improvement over CDs recorded with 16-bit operation is audible if played back on hardware with an S/N ratio better than 90 dB. This is shown in Fig.18.

The actual value of an S/N ratio varies slightly according to the conditions of measurement, so the above figures cannot be considered exact. In general terms, however, if an S/N ratio above 94 dB (A-weighting) is given for a CD player in its specifications, some actual S/N improvement can be expected. If a CD player with an S/N ratio over 114 dB is used, the full benefits of SBM can be realized.

The precision of the D/A converter is also important. SBM expresses the low level signals of 20-bit operation in 16 bits by using many fine pulses, as shown in Fig.12, with these SBM-processed signals converted to analog by the D/A converter. If the steps of the D/A converter are not precise enough, however, the sizes of the pulses will differ from the original values and the low-level signal information will become distorted.

Basically, if the D/A converter is as precise as 16 bits, some benefit of SBM can be expected. But it's not simply a matter of the number of bits, though, because there are 1-bit D/A converters which deliver 20-bit precision through pulse density modulation. While D/A converter precision may be hard to determine, in most cases the "dynamic range" value will be indicated in the specifications. If it is over 98 dB, then the precision of the D/A converter is above 16 bits and the player will be able to deliver the benefits of SBM. Just as with the S/N ratio, the dynamic range of conventional 16-bit sound sources is about 90 dB, so any player with a dynamic range above 90 dB will deliver sound reproduction with some improvement when playing back SBM-processed software.

In Conclusion : the sound of SBM

If performed skillfully, 20-bit information can be expressed in 16 bits. But can a sound wave with so many pulses really reproduce 20-bit sound quality? To answer this, let's simulate the human sense of hearing with a simple model, and compare the sound waves visually. Human ears are like microphones having very poor frequency response at low and high frequencies. In particular, response is extremely poor above 15,000 Hz.

Taking this into account, Fig.13 shows a CD player in which an analog low pass filter is used to remove frequencies above 16,000 Hz to produce a response that roughly matches that of the human ear. Fig.14 shows 100 Hz sound waveforms produced by 20-bit, 16-bit and SBM operation, all at the extremely low level of -90 dB. While the 20-bit wave is rather smooth, the 16-bit wave is abruptly stepped due to limitations in resolution. Using SBM, however, the wave closely resembles that of the 20-bit wave since the pulse groups are "averaged" by selective removal with a low-pass filter. The resulting SBM wave sounds quite satisfactory to the human ear, closely resembling the original waveform.

CDs recorded using the SBM process can be played back on any CD player as they are, without any special modification, and the human ear will perceive sound as though it contained the same amount of information as with 20-bit operation.

Fig.17 S/N including D/A Conversion and Effects of SBM

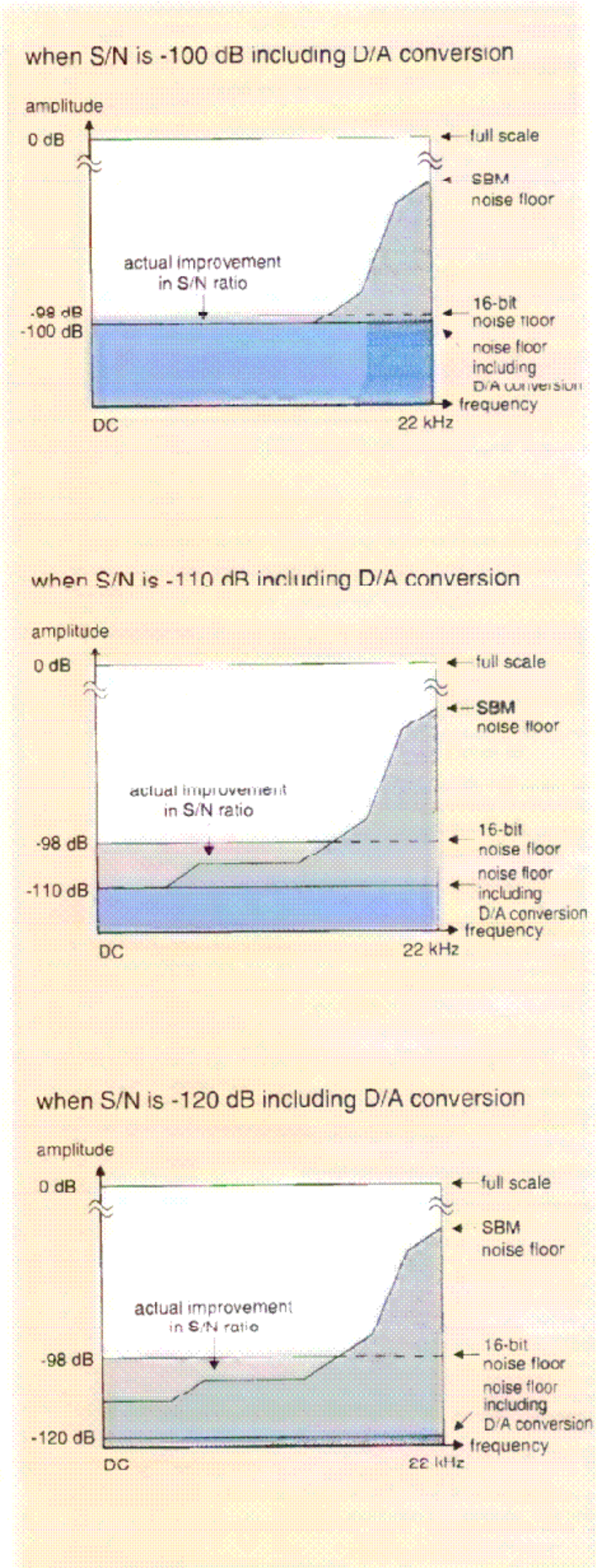
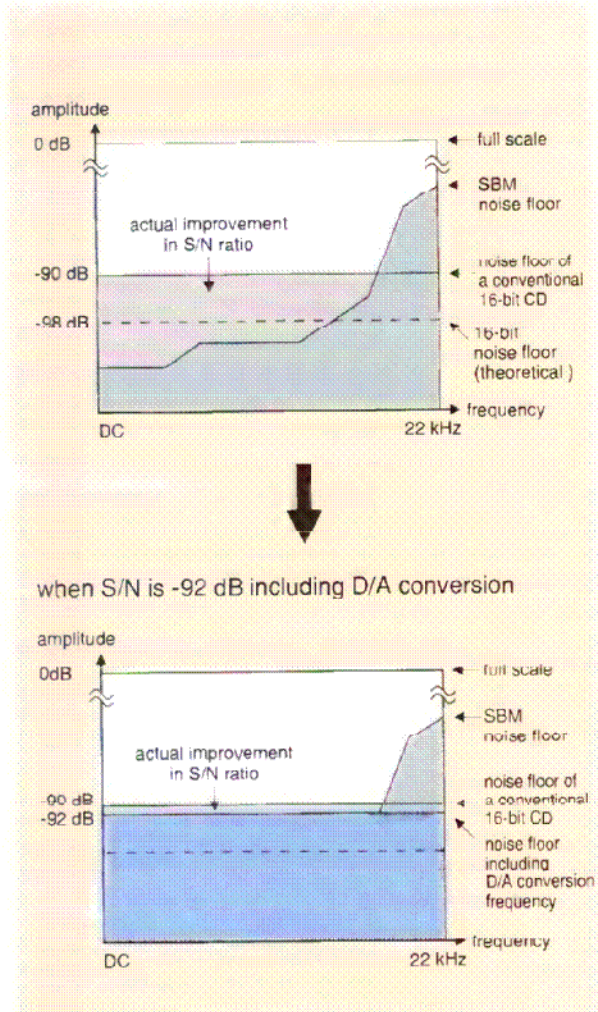


Fig.18 S/N including D/A Conversion and Effects of SBM (compared with conventional 16-bit CD)



Published by Sony Corporation

Inquiries should be directed to:

Sony Corporate Technology Group
R&D Entertainment Engineering Coordination

Fax: +81-3-5448-2207