Chapter 3 Zero Band Dithering Method

The proposed dithering method patches spectral nullity to ease annoying fishy noise. Because of the lack of prior information, it is almost impossible to make the patched spectrum signals approximate accurately to the original ones. Furthermore, since zero bands usually occur at low frequencies, hence it is unsuitable to patch spectrum by duplicate some part of the spectrum signals. Once the unfit additives, such as unexpected tone or noise signals with exceeding energies are placed at low frequency, it is very easy to cause other perceptual artifacts. On the assumption, the method adopts random noises to dither zero bands. This is because human hearing is not very sensitive to random noise. Furthermore, the method exploits the information of the quantization and extracts the amplitude range of dithering noise.

3.1 Quantization Model in MP3 and AAC

For MP3 or AAC encoder, the non-uniform quantizer is used to handle the weights of distortion effectively. Furthermore, an overall spectrum of a time frame is separated into several quantization bands with non-uniform bandwidths. Every quantization band owns individual quantization step size Δ_q to fit different perceptually tolerable distortion allowed by psychoacoustic model. More specific, the quantization model introduced in MP3 (MPEG1—Layer3) [22] and MPEG-2/4 AAC (Advanced Audio Coding) standard [20] [23] is given as follow.

$$S[k] = \mathrm{int}\left(\frac{X[k]^{\frac{3}{4}}}{\Delta_q}\right). \qquad (56)$$

where $X[k]$ is a frequency line, $S[k]$ is the quantization value, and the operator int(.) denotes the nearest integer operation. Besides, step size Δ_q is defined as (57).

$$\Delta_q = 2^{c\cdot(g-s_q)}, \qquad (57)$$

where g is global gain used for all quantization band, s_q is scale factor for qth quantization band (also known as scalefactor band), and c is a constant. A few distinctions between MP3 and AAC are the numbers of quantization bands and the definition of constant c. The numbers of quantization bands for long window is 21 and 49, for short window is 12 and 14 respectively, and constant c is defined as 3/4 and 3/16 respectively.

3.2 Zero Bands Occurring Condition

In decoders, the encoded frequency signal $X[k]$ will be inversely quantized as $\tilde{X}[k]$ by the formula (58).

$$\tilde{X}[k] = (S[k]\cdot\Delta_q)^{\frac{4}{3}} \quad (58)$$

That is equivalent to (59) where $\tilde{\Delta}_q$ is defined as $\Delta_q^{\frac{4}{3}}$.

$$\tilde{X}[k] = S[k]^{\frac{4}{3}}\cdot\tilde{\Delta}_q \quad (59)$$

In fact, the original $X[k]$ value should be given as

$$X[k] = R[k]^{\frac{4}{3}}\cdot\tilde{\Delta}_q, \quad (60)$$

where $R[k]$ is a real number, and there exists a relation between $R[k]$ and $S[k]$ as follows

$$S[k] = \mathrm{int}(R[k]) \quad (61)$$

From the definition of zero bands, the requantized frequencies $\tilde{X}[k]$ in zero bands must be zero. From (59), it implies that the relative $S[k]$ must be also set zero. Hence, from (61), it shows that $|R[k]|$ should be less than 1/2. Substituting the result to (60) illustrates that the occurring of zero bands is due to the relation

$$|X[k]| < \left(\frac{1}{2}\right)^{\frac{4}{3}}\cdot\tilde{\Delta}_q \quad (62)$$

3.3 Dithering Model

According to (62), the original frequencies $X[k]$ in a zero band can be expressed as

$$X[k] = r\cdot\left(\frac{1}{2}\right)^{\frac{4}{3}}\cdot\tilde{\Delta}_q, \quad (63)$$

where r should be a real number between *-1* and *1*. Let $X_d[k]$ be the dithering frequencies. The formula (63) suggests a well dithering model:

$$X_d[k] = \tilde{r} \cdot \left(\frac{1}{2}\right)^{\frac{4}{3}} \cdot \tilde{\Delta}_q \quad (64)$$

By substituting a random number $\tilde{r}$ of uniform distribution from *-1* to *1* to *r*, $X[k]$ can be effectively simulated. However, the zero band phenomenons are mainly due to unsuitable bit allocation policies or excessive masking energy estimation. The abnormal factors usually result in an enormous step size $\tilde{\Delta}_p$ to make the simulated value of (64) excessive. To handle the risk, a gain g needs to be considered to constrict the magnitude range of random noise. The modified model is given as.

$$X_d[k] = \tilde{r} \cdot \left(\frac{1}{2}\right)^{\frac{4}{3}} \cdot \tilde{\Delta}_q \cdot g \quad (65)$$

For simplification, combining g and $\left(\frac{1}{2}\right)^{\frac{4}{3}}$ to a parameter g_p, say patch gain, the dithering model for zero bands is defined as (66).

$$X_d[k] = \tilde{r} \cdot \tilde{\Delta}_q \cdot g_p \quad (66)$$

3.4 Dithering Algorithm

This section presents the algorithm for zero band dithering based on (66). The algorithm consists of three components that include patch gain determining module, zero band searching module, and zero band dithering module. The block diagram of the dithering algorithm is illustrated in Figure 14. At first, according to the content of the spectrum signal, patch gain determining module will adaptively choose a suitable value for patch gain. In turn, the searching module will detect where zero bands exist on the spectrum. Ultimately, the dithering module will patch the zero bands following the dithering model (66).These following subsections will exploit the three components in detail.

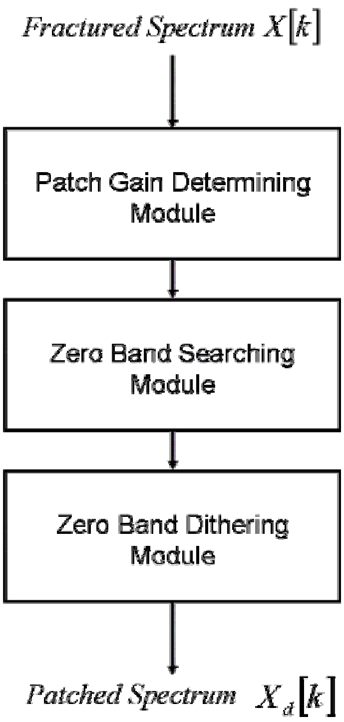

Figure 14: A block diagram of zero band dithering algorithm.

3.4.1 Patch Gain Determining Module

The spectral contents of audio signals are constantly varying whether in time or in music category. Therefore, to control a feasible distribution range of random noise by a fixed patch gain value is not effective. Especially, for a signal containing much tone component, it is very likely to harm the original quality due to an unsuitable patch gain. Figure 15~ Figure 18 illustrates the phenomenon. By comparing Figure 15 and Figure 17 ,it shows that the random noise added destroys seriously the energy ratio of tone and noise components due to an excessive patch gain. Hence, the original tone components are masked by the noise components, and the perceptual quality decays greatly. Decreasing the patch gain can improve the problem as illustrated in Figure 18. Therefore, an adaptive mechanism to change patch gain is required. Figure 19 lustrates the block diagram of the adaptive patching mechanism.

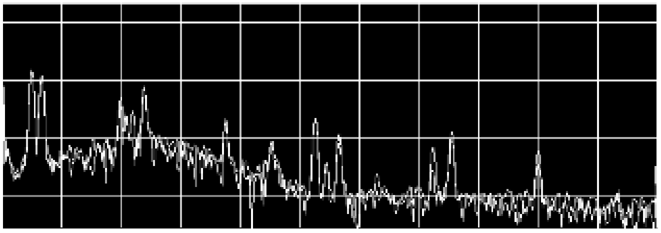

Figure 15: The spectrum of the original audio signal.

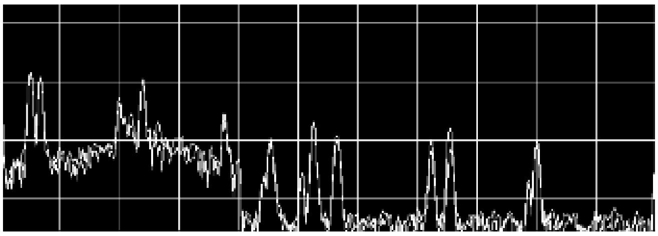

Figure 16: The spectrum of the compression audio signal.

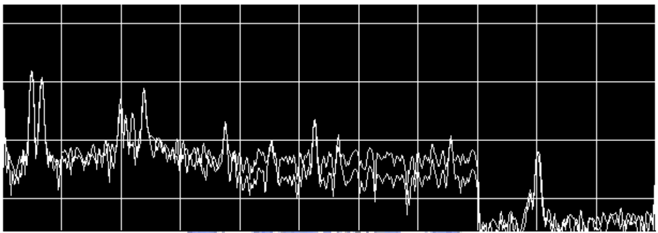

Figure 17: The spectrum of the compression audio signal with zero band dithering that sets patch gain as $\left(\frac{1}{2}\right)^{\frac{4}{3}}$.

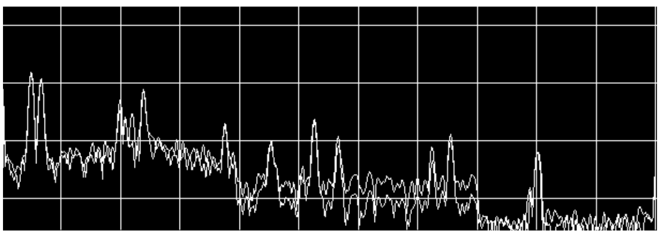

Figure 18: The spectrum of the compression audio signal with zero band dithering that sets patch gain as 1/32.

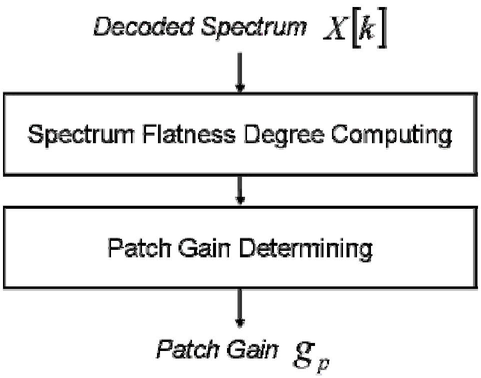

Figure 19: The block diagram of patch gain determining module.

To measure the ratio of the quantities of tone and noise components in a spectrum, flatness degree is an effective guide. This thesis calculates the flatness degree by the ratio of arithmetic average and geometric average of the frequency magnitude means of the successive spectral bands. Assume that a spectrum is separated into *M* uniform spectral bands, and each band has *m* frequency lines. The flatness degree is calculated by the formula (67).

$$\text{Flatness Degree } F = \frac{\sqrt[N]{\prod_{b=i}^{i+N-1} S_b}}{\frac{1}{N}\sum_{b=i}^{i+N-1} S_b}, \tag{67}$$

where

$$S_b = \frac{1}{m}\sum_{j=0}^{m-1}\left|X[j+m\cdot b]\right|. \tag{68}$$

Take MP3 decoder for example, this thesis sets 32 for *M*, sets 18 for *m*, and uses the ten uniform bands over about 10K~16K to compute *F*. By the flatness degree *F*, we can change the patch gain dynamically. In this thesis, the patch gain is set as

$$\text{patch gain } g_p = \begin{cases} \frac{1}{8\sqrt{2}}, & \text{for } 0.9 \le F < 1 \\ \frac{1}{16}, & \text{for } 0.0025 \le F < 0.9 \\ \frac{1}{16\sqrt{2}}, & \text{for } 0.001 \le F < 0.0025 \\ \frac{1}{32}, & \text{for } F < 0.001 \end{cases}. \tag{69}$$

The patch gain for AAC is set as follows:

$$
patch\ gain\ g_p = \begin{cases} \frac{1}{4}, & for\ \ 0.9 \le F < 1 \\ \frac{1}{4\sqrt{2}}, & for\ \ 0.0025 \le F < 0.9 \\ \frac{1}{8}, & for\ \ 0.001 \le F < 0.0025 \\ \frac{1}{8\sqrt{2}}, & for\ \ F < 0.001 \end{cases} \quad .
\tag{70}
$$

The decay degree of the patch gain adopted for AAC is lower than that for MP3 obviously. This is because the number of the spectral lines in a single quantization band for AAC is usually fewer than that for MP3.Therefore the distribution range of magnitude in a quantization band of AAC is shorter and the risk of the model (66) is lower.

3.4.2 Zero Band Searching Module

A zero band is not always located a single quantization band. On the contrary, it is usually located over two or several quantization bands. Hence, besides searching where the zero bands are located, the relative quantization bands index q also need to be found out to compute respective $\tilde{\Delta}_q$. The block diagram of zero band searching module is illustrated by Figure 20.

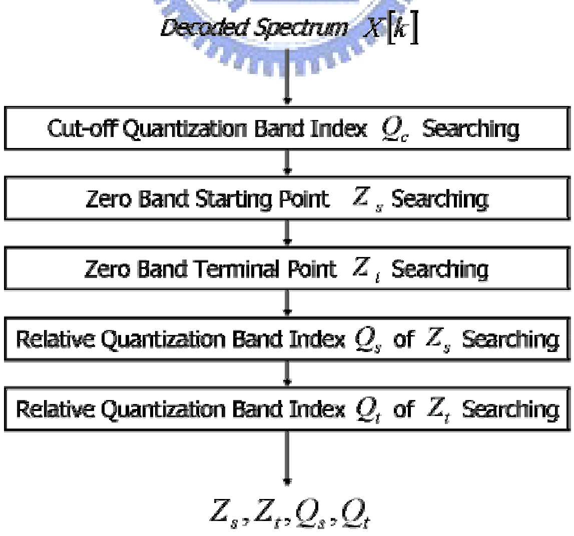

Figure 20: The block diagram of zero band searching module.

The searching module generates four indexes that are Z_s, Z_t, Q_s, and Q_t. The four indexes denote the starting point and terminal point of a zero band, and the indexes of the relative quantization bands where Z_s and Z_t locate, respectively.

Before searching the four indexes, the index Q_c of cut-off quantization band should be searched. Cut-off quantization band means the eventual quantization band containing nonzero energy. If Q_c does not exist, it implies the processed time frame is silent and the dithering processing is skipped.

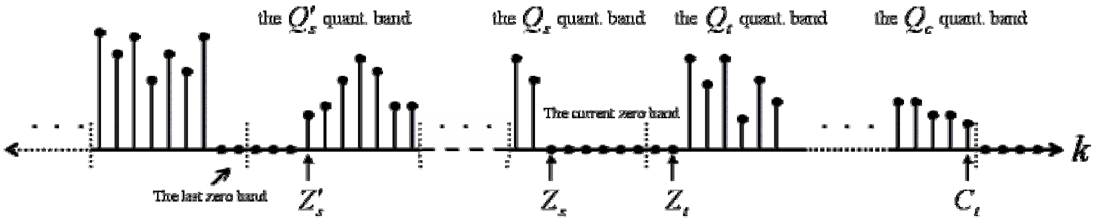

Figure 21: The relative relation between the last zero band and the current zero band.

Let Z'_s be the point that is exactly next the terminal point of the last zero band, and Q'_s be the relative index of the quantization band where the terminal point of the last zero band locates. Figure 21 illustrates the relative relation between the last zero band and the current zero band. For searching Z_s, we need to find out the first frequency k such that $X[k]$ is zero from Z'_s to C_t, where C_t denotes the terminal point of Cut-off quantization band. If such k does not exist, it shows all the range has been searched, and hence the dithering processing is completed. To continue, for searching Z_t, we need to find out the first frequency k such that $X[k+1]$ is not zero from Z_s to C_t. If such k does not exist, then set Z_t as C_t.On the other hand, similarly, we need find Q_s between Q'_s and Q_c. Finally, Q_t needs to be found between Q_s and Q_c.

3.4.3 Zero Band Dithering Module

A zero band containing a few frequency lines needs not to be dithered. It is likely a normal situation, not due to abnormal artifacts. This thesis uses the two conditions. First one is that the bandwidth BW_Z of the zero band must more than 1/4 of the bandwidth of the first quantization band which is associated with the zero band. Second is that the zero band must has at least six frequency lines. If neither of the two conditions holds, the dithering processing is skipped. On the other hand, if there is at least one non-zero frequency line $X[k']$ in a quantization band, it ensures the relative step size $\tilde{\Delta}_q$ must be lower than 3 times the magnitude of the non-zero frequency line. More accurate,

$$\tilde{\Delta}_q \le 2^{\frac{4}{3}} \cdot |X[k']| \quad (71)$$

On the contrary, when an overall quantization band is null, the risk of dithering increases highly without the guarantee of a suitable step size. Hence, a mechanism to handle the situation is required.

3.4.3.1 *Null Quantization Band Protection*

Whether in MP3 or in AAC decoders, a null quantization band occurs under two situations. The first situation is that if $Q_s \neq Q_t$, then all quantization bands of which indexes are between Q_s and Q_t are null. For the $Q_s th$ quantization band, it is a null quantization band if Z_s is equal to the starting frequency of the $Q_s th$ quantization band. For the $Q_t th$ quantization band, it is null if Z_t is equal to the ending frequency of the $Q_t th$ quantization band. The second situation is that if $Q_s = Q_t$, the $Q_s th$ quantization band is null if Z_s and Z_t are equal to the starting and ending frequency of the $Q_s th$ quantization band respectively. For a null quantization band, we let the patch gain as the half of the original one in MP3 case to decrease the risk of dithering.

3.4.3.2 *Pseudo Quantization Step Size in AAC Decoder*

In AAC encoder, a special Huffman codebook “ZERO_HCB” [20] [23] is used for a null quantization band (scalefactor band). The scalefactor is not transmitted for quantization bands which are coded with the Huffman codebook “ZERO_HCB” to save available bits. In other word, in AAC decoder, there is no information about quantization step size of a null quantization band. Under the situation, a pseudo step size for a null quantization band is required for the dithering processing. A suitable pseudo step size is given as (72) according to the step sizes of the two neighbor quantization band and referring to the distances to the null quantization band as inverse weights for linear combination.

$$\tilde{\Delta}_q = \frac{d_2}{d_1 + d_2} \cdot \tilde{\Delta}_{q1} + \frac{d_1}{d_1 + d_2} \cdot \tilde{\Delta}_{q2} , \qquad (72)$$

where *q1*, *q2* are the left and right neighbor quantization band indexes, respectively, and $d_i = |q - qi| + 1$ for i =1,2.

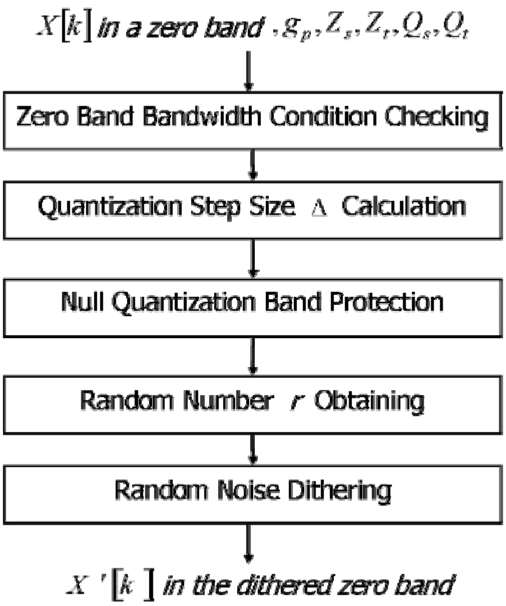

Figure 22: The block diagram of zero band dithering module.

3.5 Summary

The algorithm can be summarized as follows:

Input data: The basic sources to zero band dithering are described below.

(a) $X[k]$: the spectrum signal.

(b) N_q: the number of the quantization bands.

(c) $BW[q]$: the bandwidths of the quantization bands.

(d) $Y[q]$: the starting points of the quantization bands.

(e) *M*: the number of uniform spectral bands.

(f) *m*: the bandwidth of the uniform spectral bands.

(g) α : the index of the first uniform spectral band for flatness degree computing.

(h) *n*: the number of uniform spectral band for flatness degree computing.

There are total sixteen steps of the algorithm expressed as follow:

Step1: Calculate S_b, for b = $\alpha \sim \alpha + n - 1$

Step2: Calculate flatness degree F.

Step3: Determine patch gain g_p

Step4: Determine cut-off quant. band index Q_c

Step5: If Q_c does not exist, then the algorithm is completed.

Otherwise, go to Step 6.

Step6: Let $Z_s' = 0$, $Q_s' = 0$.

Step7: Search Z_s from Z_s' to C_t.

Step8: If Z_s does not exist or $Z_s > C_t$, then the algorithm is completed.

Otherwise, go to Step 9.

Step9: Search z_t from z_s to C_t.

Step10: Search Q_s from Q'_s to Q_c.

Step11: Search Q_t from Q_s to Q_c.

Step12: Bandwidth condition checking.

If $BW_Z > \frac{1}{4} \cdot BW[Q_s]$ or BW_Z >5, then go to Step 13.

Otherwise, go to Step 15.

Step13: Null quantization protection.

Step14': Pseudo step size computing for AAC case.

$$\tilde{\Delta}_q = \frac{d_2}{d_1 + d_2} \cdot \tilde{\Delta}_{q1} + \frac{d_1}{d_1 + d_2} \cdot \tilde{\Delta}_{q2}$$

Step14: Zero band dithering

Let $X_d[k] = \tilde{r}_k \cdot \tilde{\Delta}_q \cdot g_p$, for $k = \Psi$ to Ω, q = Q_s to Q_t,

where Ψ = the maximum of $Y[q]$ and z_s.

Ω = the minimum of $Y[q+1]-1$ and z_t.

Step15: Let z'_s = z_t+1, Q'_s = Q_t. Go to Step 7.

The associated flow chart of the algorithm is illustrated by Figure 23.

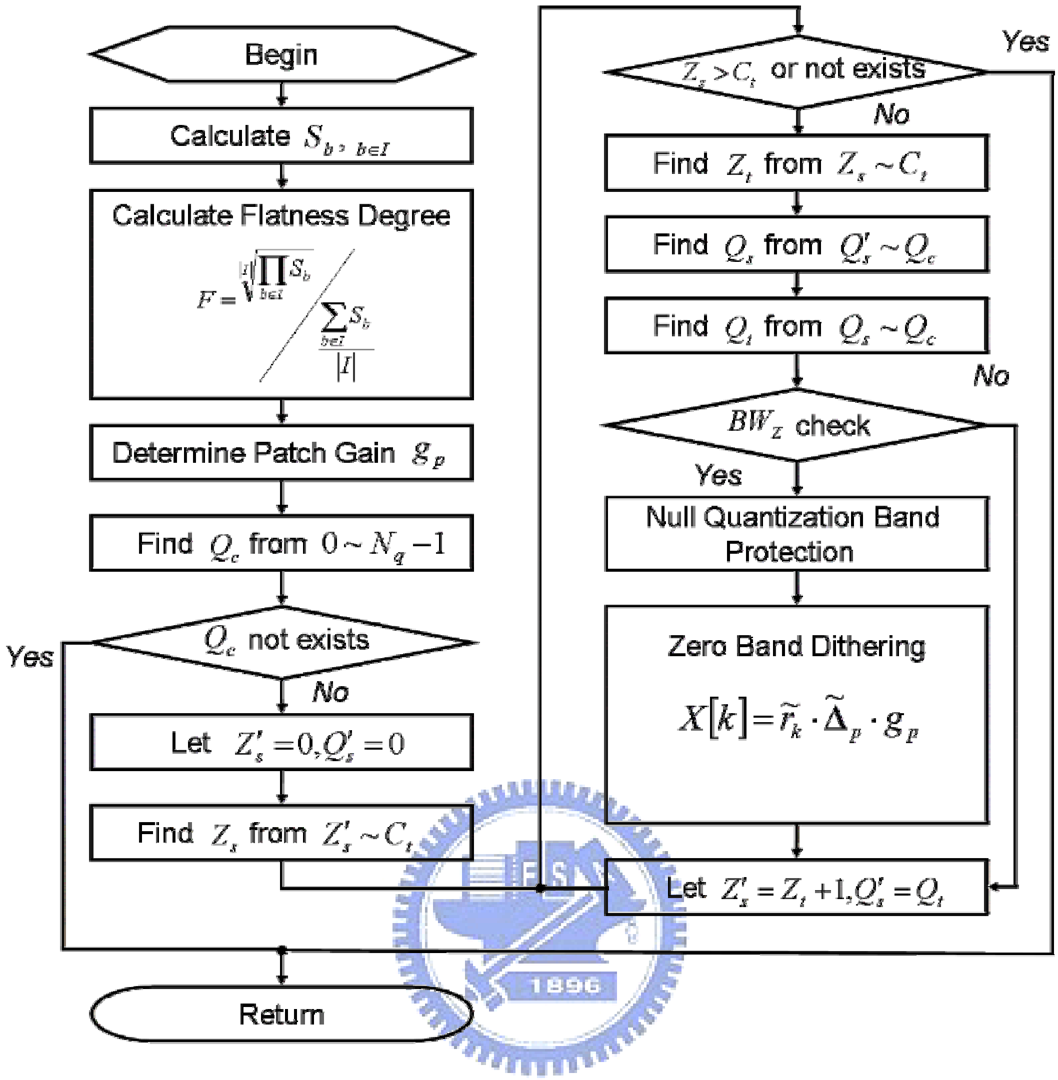

Figure 23: The flow chart of zero band dithering.