IEG4160 Image and Video Processing

Quantization

Chang-Su Kim

Quantization

- Digitization =
  - sampling (coordinate) + quantization (value)

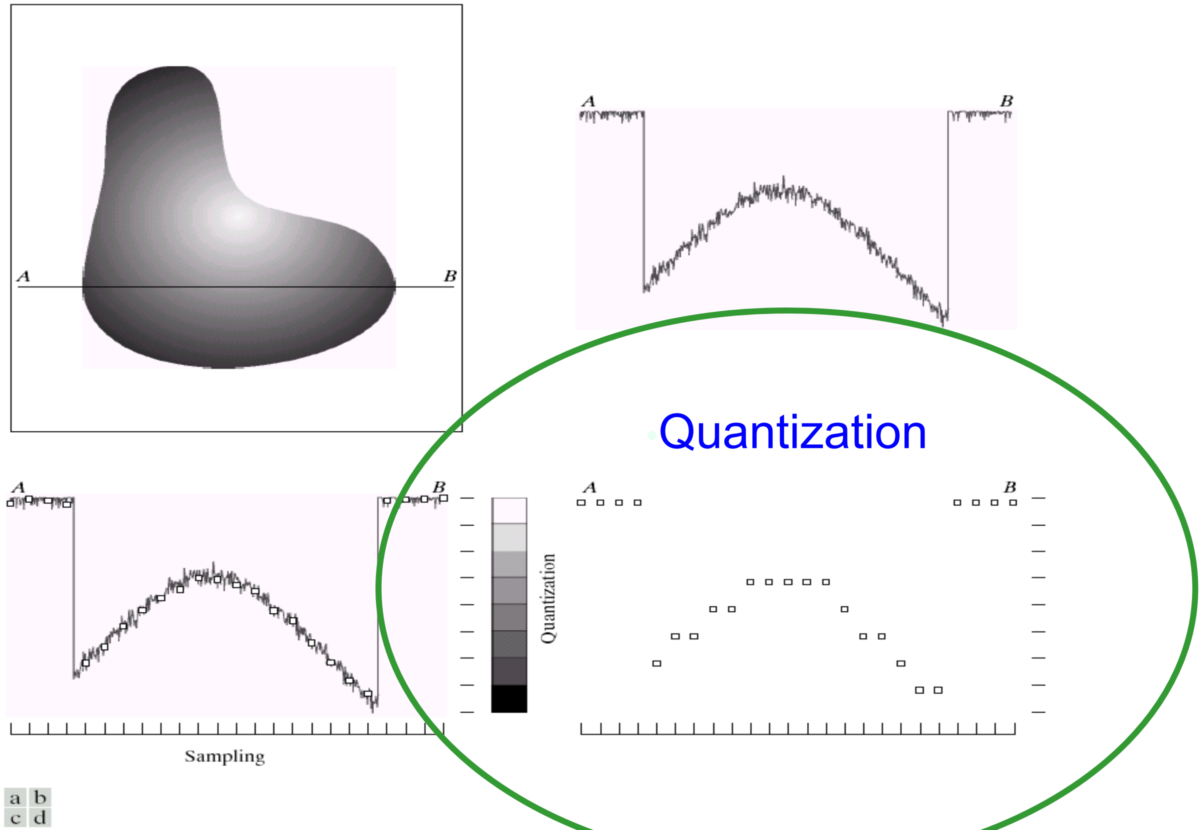
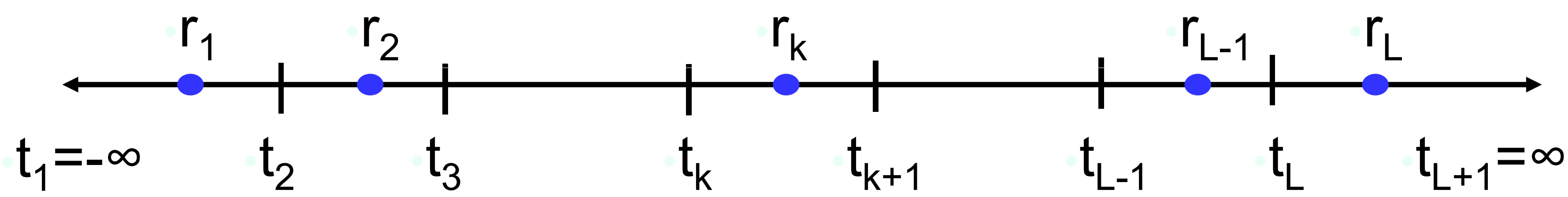


FIGURE 2.16 Generating a digital image. (a) Continuous image. (b) A scan line from *A* to *B* in the continuous image, used to illustrate the concepts of sampling and quantization. (c) Sampling and quantization. (d) Digital scan line.

Quantizer

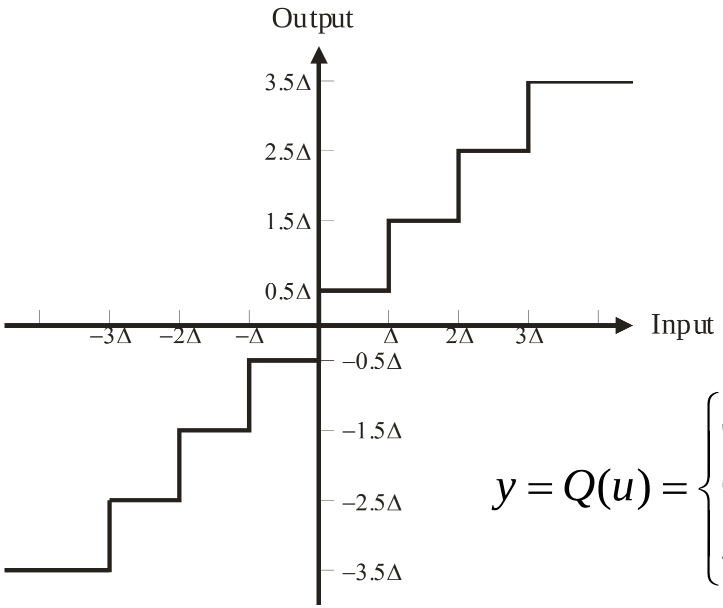
- A quantizer Q maps a continuous variable u into a discrete variable Q(u) in $\{r_1, r_2, r_3, \ldots, r_L\}$



- Partition the real line into L cells and map input values within a cell into a constant r_k
  - $Q(u) = r_k \quad \text{if} \;\; t_k \leq u < t_{k+1}$
  - r_k : reconstruction level
  - t_k : transition or decision level
  - $\Delta_k = t_{k+1} - t_k$: step size

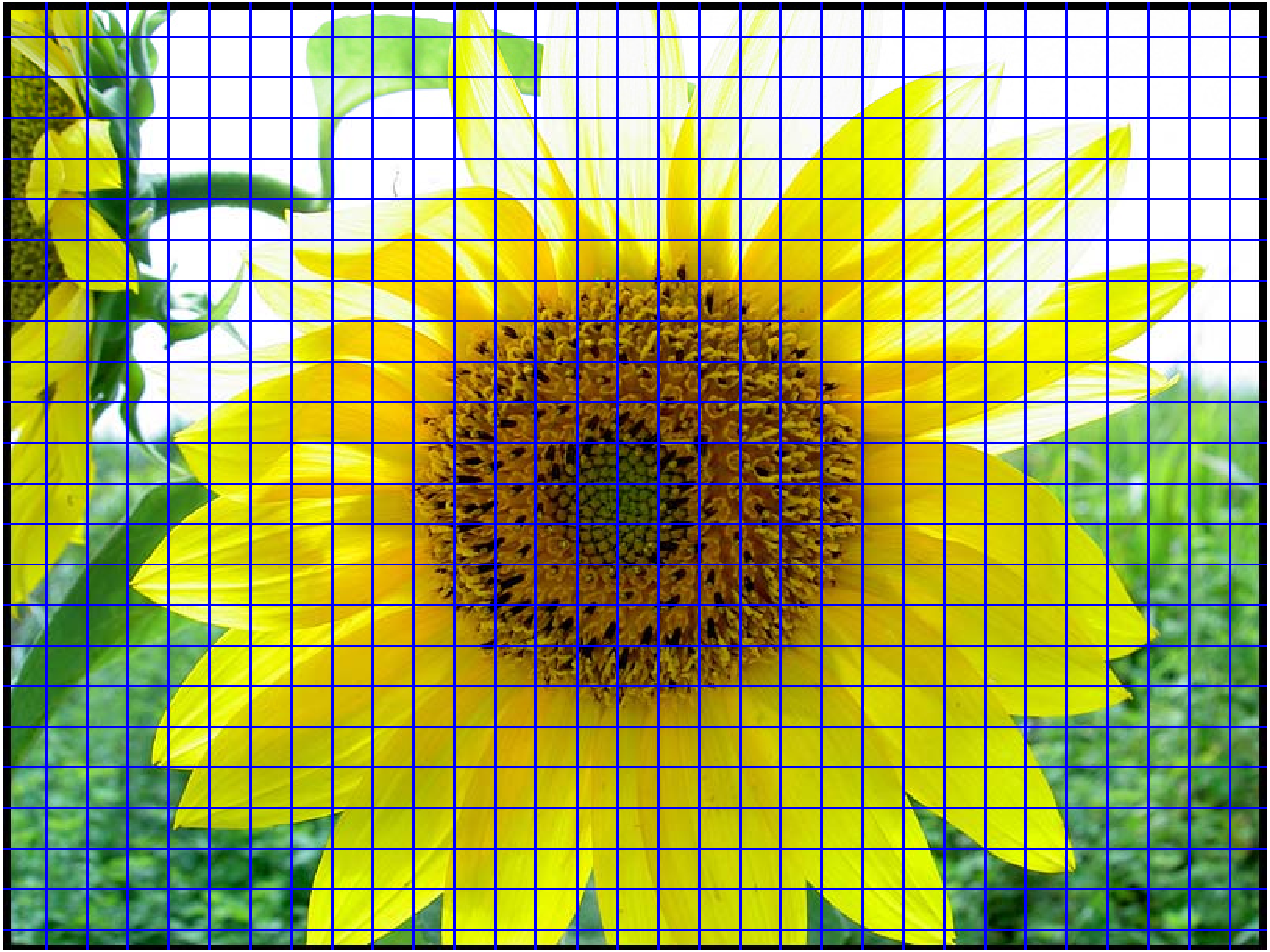
Quantizer Example

- Input-output graph of an 8-level quantizer



$$y = Q(u) = \begin{cases} 3.5\Delta & \text{if } u > 3\Delta, \\ 0.5(2n-1)\Delta & \text{if } (n\text{-}1)\Delta < u \le n\Delta \ (n = -2, -1, \ldots, 3), \\ -3.5\Delta & \text{if } u \le -3\Delta. \end{cases}$$

- Uniform quantizer
  - Except the outer two cells
    - $t_{k+1} - t_k = \Delta$ and $r_k = (t_k + t_{k+1})/2$



Lloyd-Max Quantizer

- Quantization error: $u - Q(u)$
- Probability distribution of input: $p(u)$
- Mean square error (MSE)

$$\mathcal{E} = E[(u - Q(u))^2] = \int_{t_1}^{t_{L+1}} (u - Q(u))^2 p(u) du$$

- Lloyd-Max quantizer minimizes $\mathcal{E}$,
  i.e. it is the minimum mean square error (MMSE) quantizer

Lloyd-Max Quantizer – Centroid Condition

- MSE

$$\mathcal{E} = \sum_{i=1}^{L} \int_{t_i}^{t_{i+1}} (u - Q(u))^2 p(u)du = \sum_{i=1}^{L} \int_{t_i}^{t_{i+1}} (u - r_i)^2 p(u)du$$

- For fixed transition levels t_k's, find the optimum reconstruction levels r_k's
- Minimize each $\int_{t_k}^{t_{k+1}} (u - r_k)^2 p(u)du$

$$\begin{aligned}
\mathcal{E}_k &= \int_{t_k}^{t_{k+1}} (u - r_k)^2 p(u)du \\
&= \int_{t_k}^{t_{k+1}} u^2 p(u)du - 2r_k \int_{t_k}^{t_{k+1}} up(u)du + r_k^2 \int_{t_k}^{t_{k+1}} p(u)du \\
\therefore \quad \frac{\partial \mathcal{E}_k}{\partial r_k} &= -2 \int_{t_k}^{t_{k+1}} up(u)du + 2r_k \int_{t_k}^{t_{k+1}} p(u)du = 0 \\
\therefore \quad r_k &= \frac{\int_{t_k}^{t_{k+1}} up(u)du}{\int_{t_k}^{t_{k+1}} p(u)du} = E[u | u \in [t_k, t_{k+1})]
\end{aligned}$$

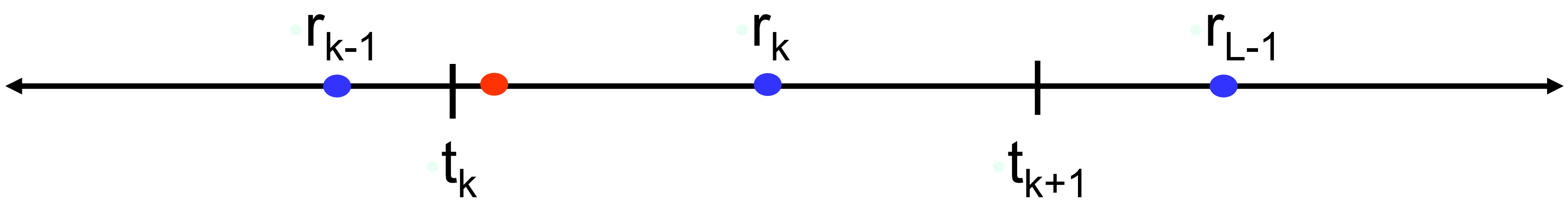
- This is called the centroid (center of mass) condition

Lloyd-Max Quanitzer – NN condition

- For fixed r_k's, find the optimum t_k's

$$\begin{aligned}
\frac{\partial}{\partial t_k}\mathcal{E} &= \frac{\partial}{\partial t_k}\sum_{i=1}^{L}\int_{t_i}^{t_{i+1}}(u-r_i)^2 p(u)du \\
&= \frac{\partial}{\partial t_k}\left(\ldots + \int_{t_{k-1}}^{t_k}(u-r_{k-1})^2 p(u)du + \int_{t_k}^{t_{k+1}}(u-r_k)^2 p(u)du + \ldots\right) \\
&= (t_k - r_{k-1})^2 p(t_k) - (t_k - r_k)^2 p(t_k) = 0 \\
& \quad \left(\because \frac{\partial}{\partial\alpha}\int_{\beta}^{\alpha} f(x)dx = f(\alpha), \frac{\partial}{\partial\alpha}\int_{\alpha}^{\beta} f(x)dx = -f(\alpha)\right) \\
\therefore & \quad (t_k - r_{k-1})^2 = (t_k - r_k)^2 \\
\therefore & \quad t_k = \frac{r_{k-1} + r_k}{2}
\end{aligned}$$

- This is called the nearest neighbor condition



Design of Lloyd-Max Quantizer

- Centroid condition

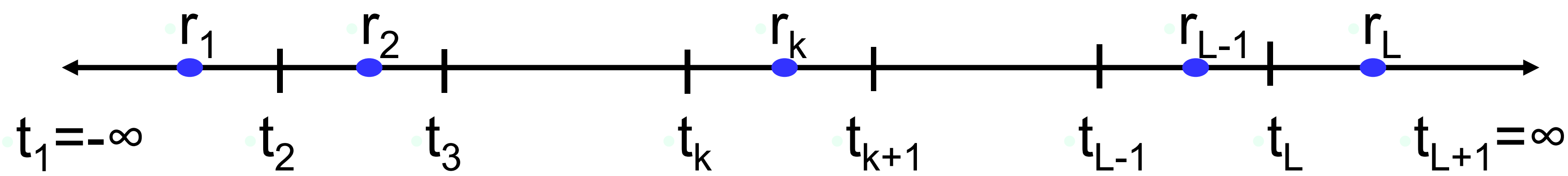
$$r_k = \frac{\int_{t_k}^{t_{k+1}} up(u)du}{\int_{t_k}^{t_{k+1}} p(u)du}$$

- Nearest neighbor condition

$$t_k = \frac{r_{k-1} + r_k}{2}$$

- These two conditions are iteratively applied to obtain the optimal quantizer

Design of Lloyd-Max Quantizer (Sequential Version)



1. Pick initial sets $\{t_2, t_3, \ldots, t_L\}$ and $\{r_1, r_2, \ldots, r_L\}$. Set k=1.
2. Update r_k such that it is the centroid of the interval (t_k, t_{k+1}).
3. Update t_{k+1} such that it is the midpoint of r_k and r_{k+1}.
4. If k=L-1 goto step 5, otherwise set k=k+1 and goto step 2.
5. Compute c, the centroid of the interval (t_L, t_{L+1}). If $|r_L - c| < \varepsilon$, stop. Otherwise, goto step 6.
6. Set $r_L = r_L - \alpha(r_L - C)$ and set k-1. Goto step 2.

Properties of Lloyd-Max Quantizer

- The quantizer output is an unbiased estimate of the input, i.e.

$$E[Q(u)] = E[u]$$

Proof)

$$\begin{aligned}
E[u - Q(u)] &= \sum_{i=1}^{L} \int_{t_i}^{t_{i+1}} (u - Q(u))p(u)du \\
&= \sum_{i=1}^{L} \int_{t_i}^{t_{i+1}} (u - r_i)p(u)du \\
&= \sum_{i=1}^{L} \left[\int_{t_i}^{t_{i+1}} up(u)du - r_i \int_{t_i}^{t_{i+1}} p(u)du \right] \\
&= 0 \qquad (\because \text{the centroid condition})
\end{aligned}$$

Properties of Lloyd-Max Quantizer

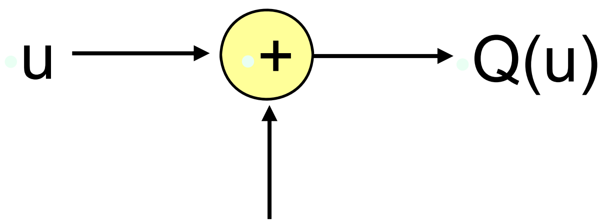
- The quantizer error is uncorrelated with the quantizer output, i.e.

$$E[(u - Q(u))Q(u)] = 0$$

Proof)

$$\begin{aligned} E[(u - Q(u))Q(u)] &= \sum_{i=1}^{L} \int_{t_i}^{t_{i+1}} (u - r_i) r_i p(u) du \\ &= \sum_{i=1}^{L} r_i \int_{t_i}^{t_{i+1}} (u - r_i) p(u) du = 0 \end{aligned}$$

- Equivalent additive noise model



η = u - Q(u)

$$\begin{aligned} \sigma_\eta^2 &= E[(u - Q(u))^2] \\ &= E[u^2] - 2E[uQ(u)] + E[Q^2(u)] \\ &= E[u^2] - E[Q^2(u)] \qquad (\because E[uQ(u)] = E[Q^2(u)]) \\ &= \sigma_u^2 - \sigma_{Q(u)}^2 \qquad (\because E[u] = E[Q(u)]) \end{aligned}$$

Lloyd-Max Quantizer for Uniform Distribution

$$p(u) = \begin{cases} \frac{1}{t_{L+1}-t_1}, & t_1 < u < t_{L+1} \\ 0, & \text{otherwise} \end{cases}$$

- The input has variance $\sigma_u^2 = A^2/12$, where $A = t_{L+1} - t_1$.
- From the centroid condition

$$r_k = \frac{\int_{t_k}^{t_{k+1}} up(u)du}{\int_{t_k}^{t_{k+1}} p(u)du} = \frac{t_{k+1}^2 - t_k^2}{2(t_{k+1} - t_k)} = \frac{t_{k+1} + t_k}{2} \quad (1)$$

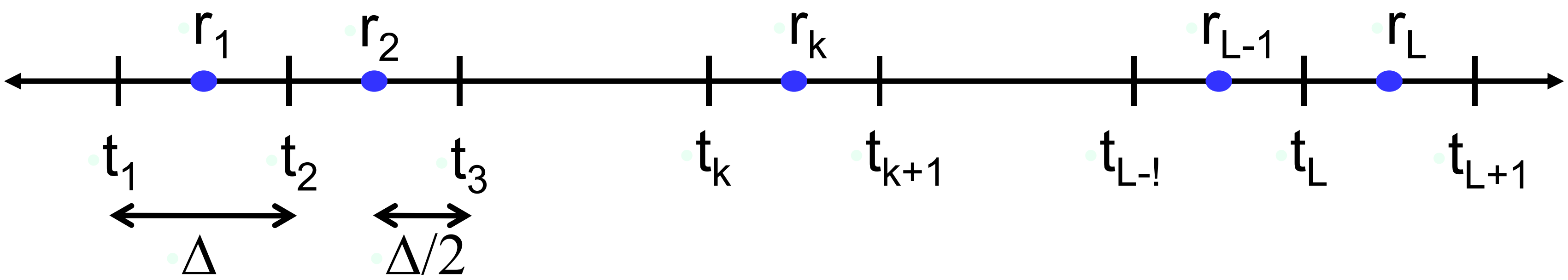
- Also, the nearest neighbor condition is

$$t_k = \frac{r_{k-1} + r_k}{2} \quad (2)$$

- By inserting (1) into (2), we have

$$\begin{aligned} t_k &= \frac{t_k + t_{k-1} + t_{k+1} + t_k}{4} \\ \Rightarrow & \quad t_k - t_{k-1} = t_{k+1} - t_k = \text{constant} \doteq \Delta \end{aligned}$$

Lloyd-Max Quantizer for Uniform Distribution



- The quantization error $\eta = u - Q(u)$ is uniformly distributed over $[-\Delta/2, \Delta/2)$.

$$\mathcal{E} = E[\eta^2] = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} x^2 dx = \frac{\Delta^2}{12}$$

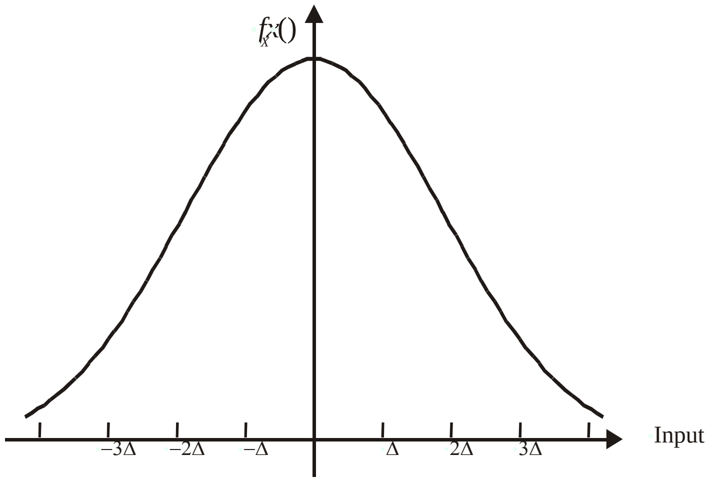
- If the quantization resolution is B bits,

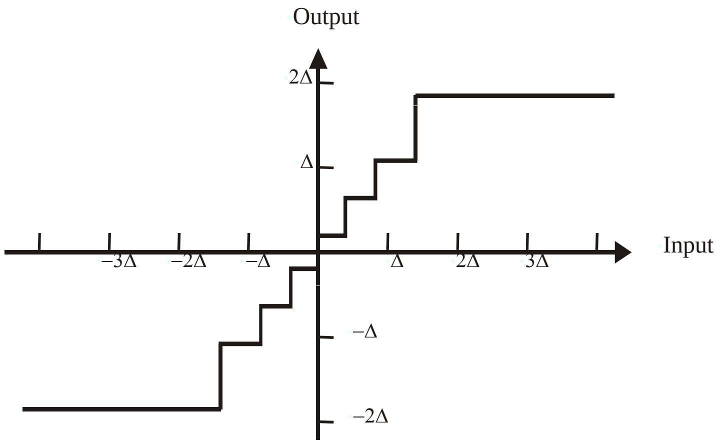
$$\Delta = \frac{A}{2^B}$$

- Thus, SNR is given by

$$\begin{aligned} 10\log_{10}\frac{\sigma_u^2}{\mathcal{E}} &= 10\log_{10}\frac{A^2/12}{\Delta^2/12} \\ &= 10\log_{10} 2^{2B} = 20B\log_{10} 2 \simeq 6B \quad \text{(dB)} \end{aligned}$$

Lloyd-Max Quantizer for Other Distributions



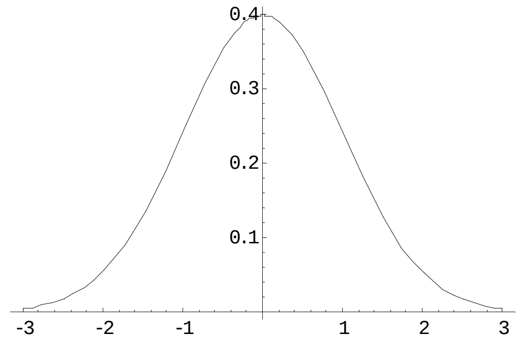


- Notice that the Lloyd-Max quantizer reduces the average distortion by approximating the input more precisely in regions of higher probability.

Lloyd-Max Quantizer for Other Distributions

- Gaussian: for pixel distribution

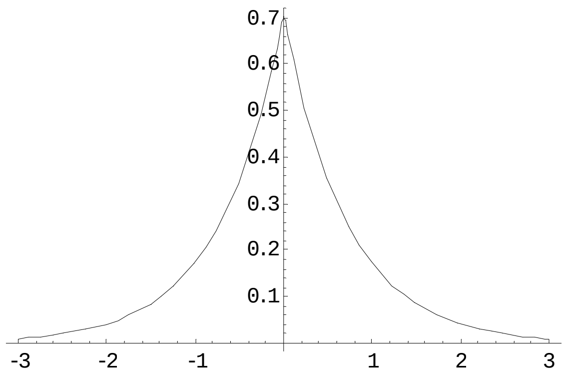
$$p(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-(u-\mu)^2}{2\sigma^2})$$



- Laplacian: for the distribution of differences between adjacent pixels

$$p(u) = \frac{\lambda}{2} \exp(-\lambda|u-\mu|)$$

where $\sigma^2 = \frac{2}{\lambda^2}$



- Look-up table of Lloyd-Max Q is available for these distributions

Mathematical Formula for Quantizer MSE

- Exact formula

$$\mathcal{E} = \sum_{i=1}^{L} \int_{t_i}^{t_{i+1}} (u - Q(u))^2 p(u) du$$

  - Not convenient for large L
  - Does not offer insight
- High resolution assumption
  - L is large
  - Maximum step size is small
  - $p(u)$ is reasonably smooth
- Approximate formula

$$\mathcal{E} = \frac{1}{12L^2} \int_{t_1}^{t_{L+1}} p(u) \lambda(u)^{-2} du$$

where $\lambda(u)$ is the density function for reconstruction levels

Derivation of Approximate Formula

- $p(u) \simeq p(r_i)$, if $u \in [t_i, t_{i+1})$
  (i.e. uniform density over a cell)
- $P_i \triangleq Pr(u \in [t_i, t_{i+1})) = \int_{t_i}^{t_{i+1}} p(u)du \simeq (t_{i+1} - t_i)p(r_i)$
  $\Rightarrow p(r_i) = \frac{P_i}{\Delta_i}$, where $\Delta_i \triangleq t_{i+1} - t_i$
- Therefore,

$$\begin{aligned} \mathcal{E} &\simeq \sum_{i=1}^{L} \frac{P_i}{\Delta_i} \int_{t_i}^{t_{i+1}} (u - r_i)^2 du \\ &\simeq \sum_{i=1}^{L} \frac{P_i}{\Delta_i} \int_{t_i}^{t_{i+1}} \left(u - \frac{t_i + t_{i+1}}{2}\right)^2 du \\ &= \sum_{i=1}^{L} \frac{P_i}{\Delta_i} \frac{\Delta_i^3}{12} = \frac{1}{12} \sum_{i=1}^{L} P_i \Delta_i^2 \end{aligned}$$

Centroid condition & uniform density over a cell

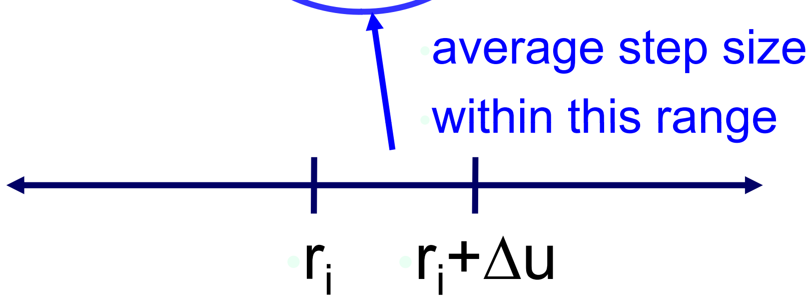
Derivation of Approximate Formula

- Consider a family of quantizers
  - with the same relative concentration of reconstruction levels
  - but with a increasing number of total levels L
- $L(u)\Delta u$: the number of levels between u and $u + \Delta u$
- Density function for levels

$$\lambda(u) = \lim_{L \to \infty} \frac{L(u)}{L}$$

  - $\int_{t_1}^{t_{L+1}} \lambda(u) du = 1$ like probability density function
  - $L\lambda(u)\Delta u$ levels within $[u, u + \Delta u)$
  - Therefore

$$\Delta_i \simeq \frac{\Delta u}{L\lambda(r_i)\Delta u} = \frac{1}{L\lambda(r_i)}$$



Derivation of Approximate Formula

- Finally,

$$
\begin{aligned}
\mathcal{E} &= \frac{1}{12}\sum_{i=1}^{L} P_i \Delta_i^2 \\
&= \frac{1}{12}\sum_{i=1}^{L} p(r_i)\Delta_i \frac{1}{(L\lambda(r_i))^2} \\
&= \frac{1}{12L^2}\sum_{i=1}^{L} p(r_i)\lambda(r_i)^{-2}\Delta_i \\
&= \frac{1}{12L^2}\int_{t_1}^{t_{L+1}} p(u)\lambda(u)^{-2}du
\end{aligned}
$$

Approximate Formula for Optimal MSE

- Objective: find the best level density function $\lambda(u)$ and the corresponding MSE $\mathcal{E}$
- Recall that $\mathcal{E} = \frac{1}{12}\sum_{i=1}^{L} P_i \Delta_i^2 = \frac{1}{12}\sum_{i=1}^{L} p(r_i)\Delta_i^3$
- Let $\alpha_i \triangleq p(r_i)^{1/3}\Delta_i$, then

$$\mathcal{E} = \frac{1}{12}\sum_{i=1}^{L} \alpha_i^3$$

- There is a constraint on α_i's, since

$$\sum_{i=1}^{L} \alpha_i = \sum_{i=1}^{L} p(r_i)^{1/3}\Delta_i = \int_{t_1}^{t_{L+1}} p(u)^{1/3} du = c$$

- Lagrangian cost function

$$\begin{aligned} \mathcal{C} &= \frac{1}{12}\sum_{i=1}^{L} \alpha_i^3 + \mu \sum_{i=1}^{L} \alpha_i \\ \Rightarrow \frac{\partial \mathcal{C}}{\partial a_i} &= \frac{1}{4}\alpha_i^2 + \mu = 0 \end{aligned}$$

Approximate Formula for Optimal MSE

- α_i^2 (and hence α_i) should be constant for all i
  - $\alpha_i = p(r_i)^{1/3}\Delta_i = c$
  - $\Delta_i \propto p(r_i)^{-1/3}$
  - Step size should be small in high input density area
- Recall that $\Delta_i \propto \frac{1}{\lambda(r_i)}$
- Therefore, $\lambda(r_i) \propto p(r_i)^{1/3}$ and

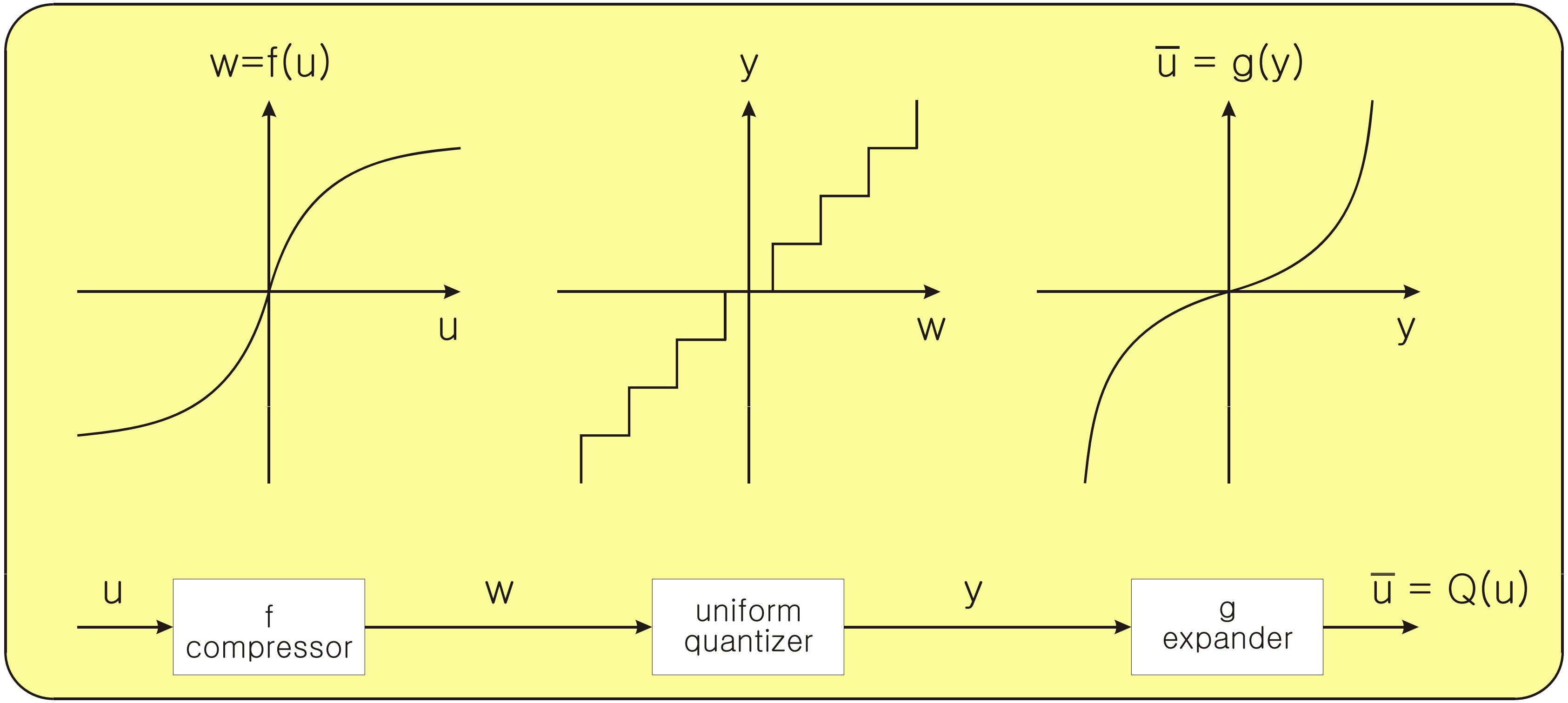
$$\lambda(u) = \frac{p(u)^{1/3}}{\int_{t_1}^{t_{L+1}} p(v)^{1/3} dv} \qquad (\because \int_{t_1}^{t_{L+1}} \lambda(u) du = 1)$$

- The optimal MSE is hence given by

$$\begin{aligned} \mathcal{E} &= \frac{1}{12L^2} \int_{t_1}^{t_{L+1}} p(u)\lambda(u)^{-2} du \\ &= \frac{1}{12L^2} \frac{\int_{t_1}^{t_{L+1}} p(u)^{1/3} du}{(\int_{t_1}^{t_{L+1}} p(v)^{1/3} dv)^{-2}} \\ &= \frac{1}{12L^2} \left(\int_{t_1}^{t_{L+1}} p(u)^{1/3} du \right)^3 \end{aligned}$$

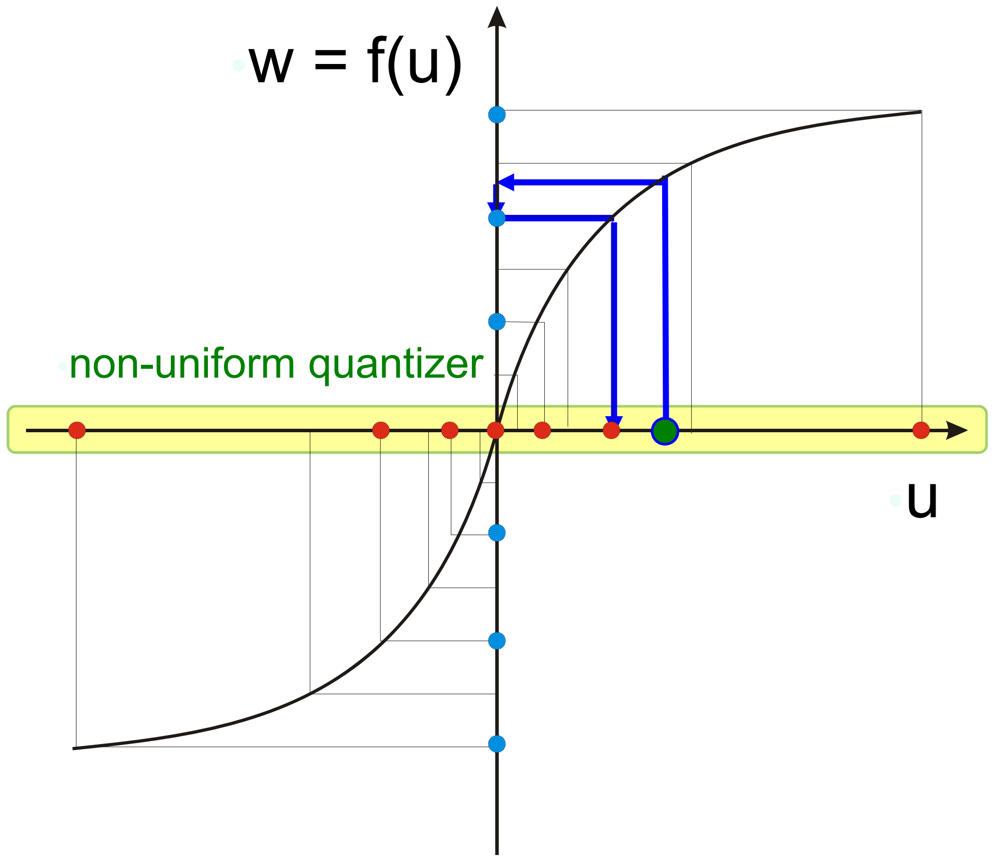
Compandor

- A way to use uniform quantizer efficiently for non-uniform input density



Compandor

- Equivalent to non-uniform quantizer



Compandor

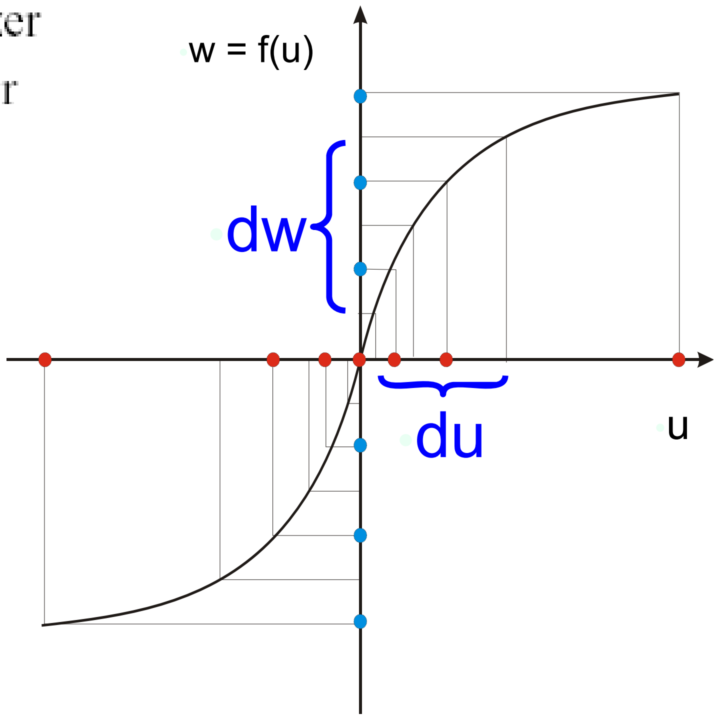
- Given an input density $p(u)$ and a uniform quantizer within range $[-a, a]$, how to design the compandor $f(\cdot)$ to minimize the MSE $\mathcal{E}$?
- To be optimum,

$$\lambda(u) = \frac{p(u)^{1/3}}{\int_{t_1}^{t_{L+1}} p(v)^{1/3} dv}$$

- $\lambda(w) = \frac{1}{2a}, \qquad w \in [-a, a]$
- $\lambda(u)du = \lambda(w)dw$

$$\Rightarrow f'(u) = \frac{dw}{du} = \frac{\lambda(u)}{\lambda(w)} = 2a\lambda(u)$$

$$\Rightarrow f(u) = 2a \frac{\int_{t_1}^{u} p(v)^{1/3} dv}{\int_{t_1}^{t_{L+1}} p(v)^{1/3} dv} - a$$



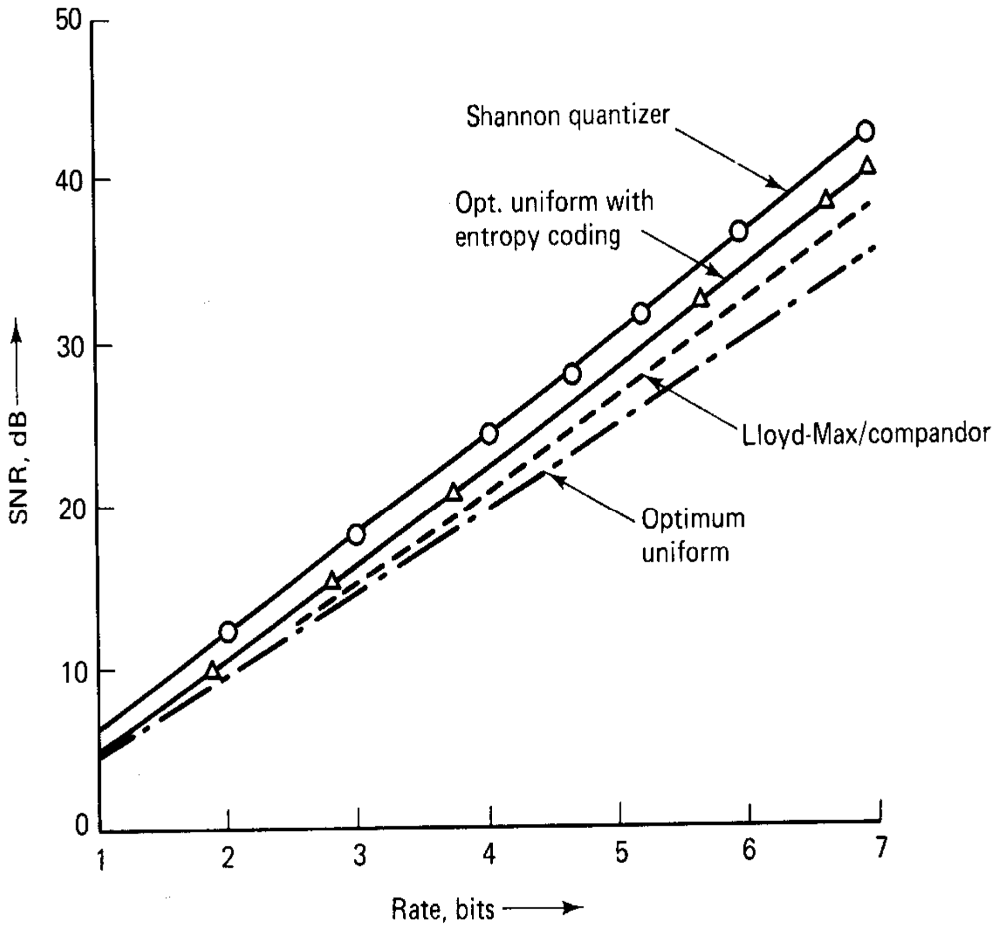
Lλ(u) du = Lλ(w) dw
= # of levels
= 2

Optimum Mean Square Uniform Quantizer for Nonuniform Densities

- Given data
  - p(u) : input density
  - L : the number of levels
- Goal
  - find the range $[t_1, t_{L+1}]$ that minimizes the MSE
- If we assume that p(u) is an even function centered around 0
  - the range should be [-a, a]
  - $2a = L\Delta$
  - Thus, the MSE can be represented as a function of a single variable Δ
- The output levels are not equi-probable, hence can be more efficiently represented using entropy coding techniques

Comparison

- For Gaussian Source



- Lloyd-Max Q provides better SNR than optimum uniform quantizer (2dB at B=6)
- Lloyd-Max Q and compandor are practically indistinguishable
- Optimum uniform + entropy coding provides better performance than Lloyd-Max Q
- Shannon Q is the theoretical limit
  - No quantizer can do better than Shannon Q.

Contouring Artifacts

- Regions of constant gray levels (visible: less than 6 bits/pixel)

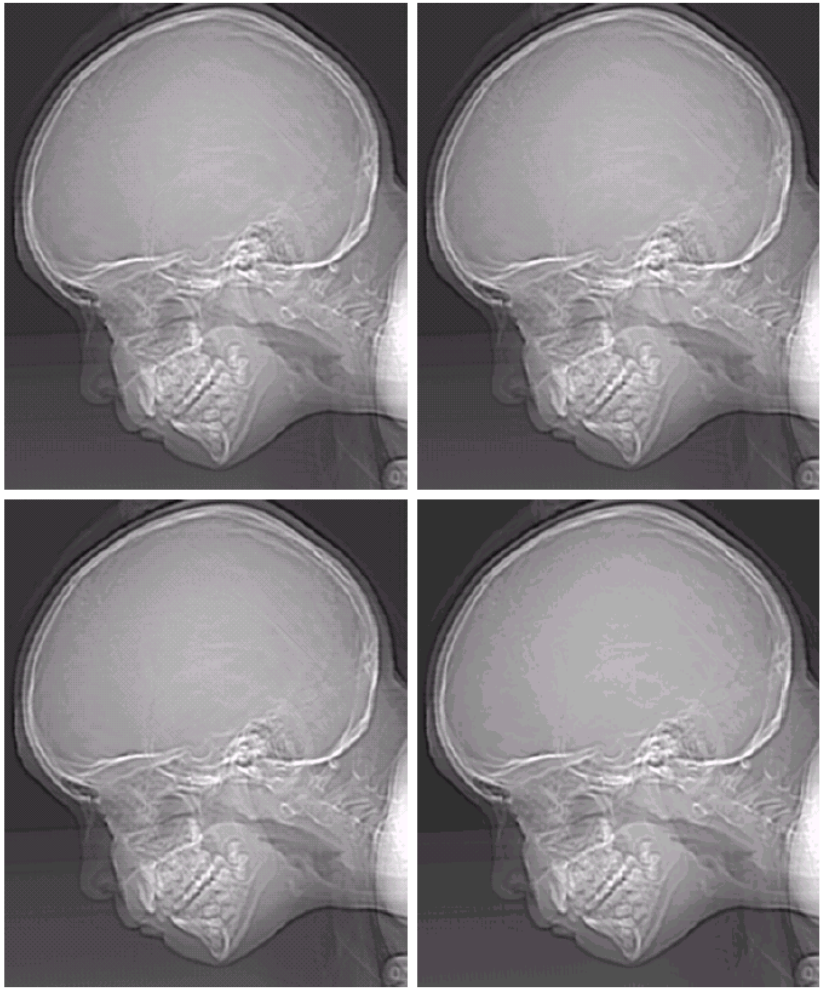
Original (8bits/pixel)

6bits/pixel

4bits/pixel

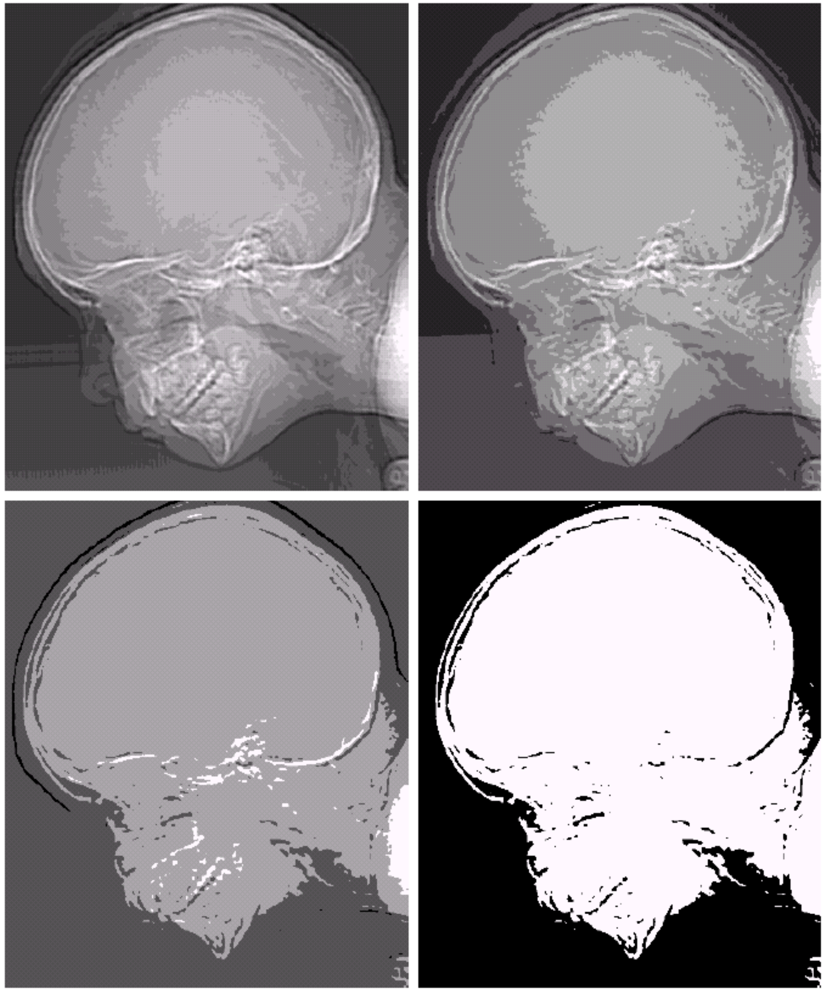
2bits/pixel

Contouring Artifacts



a b
c d

FIGURE 2.21
(a) 452 × 374, 256-level image. (b)–(d) Image displayed in 128, 64, and 32 gray levels, while keeping the spatial resolution constant.



Visual Quantization

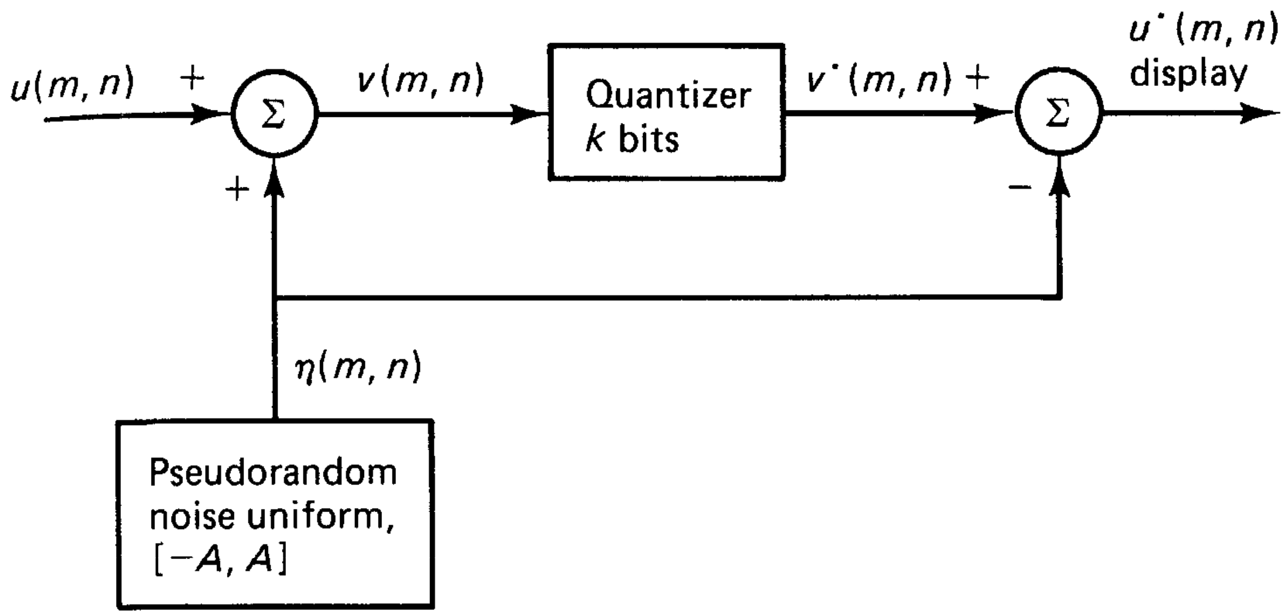
- Contouring artifacts are not well represented by MSE
  - MSE is not directly proportional to subjective quality

- There are many methods to alleviate these artifacts, including
  - Pseudo-random noise quantization

Pseudo-Random Noise Quantization

1. Add a small amount of random noise (dither) before quantization to break contours
2. Subtract the same noise after quantization

- Reasonable image quality at 3-bit quantization



Pseudo-Random Noise Quantization

a) Ordinary quantization yields contour artifacts

b) Random noise: its average should be 0 so that the overall image luminance does not change

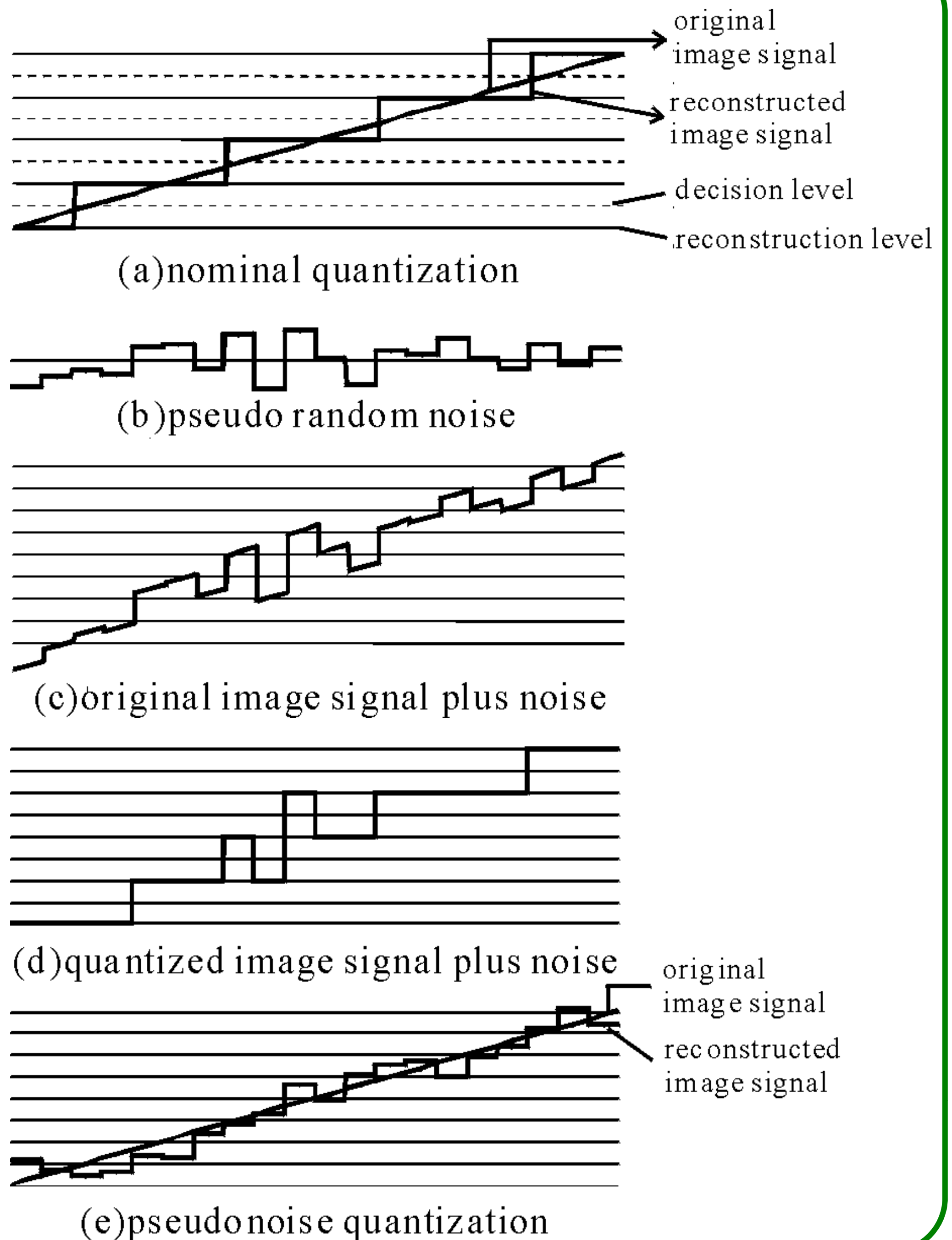
c) Signal+Noise

d) Quantization of "Signal+Noise"

- At a few points, contours are broken due to the noise

e) Subtract the same noise from quantization output

- Shaky image without contour
- Shaky effects (high frequency components) are less visible than contour artifacts



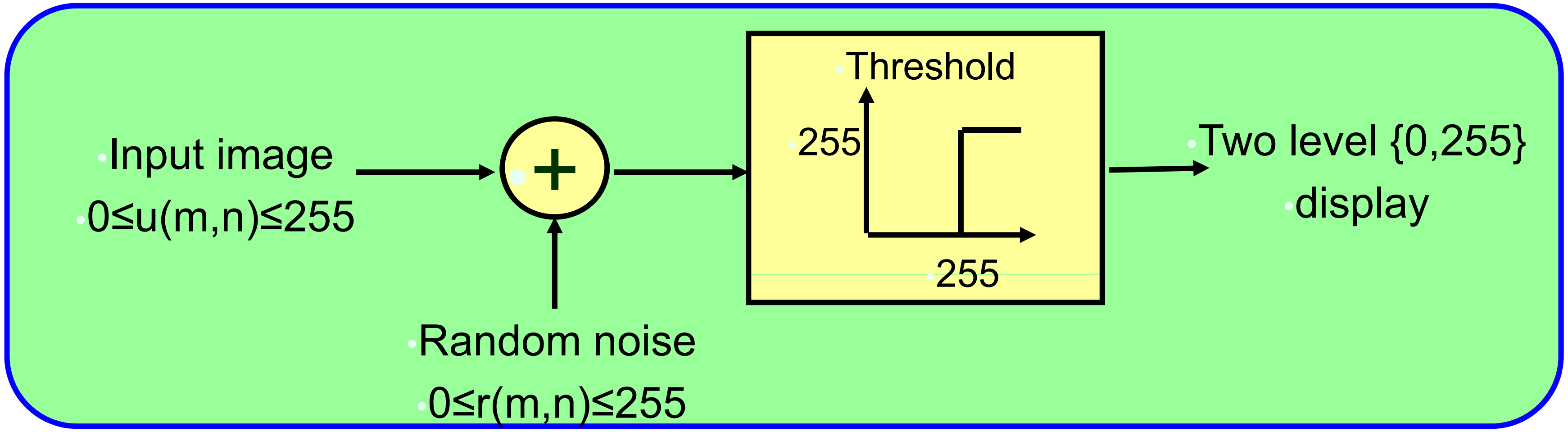
Pseudo-Random Noise Quantization

(a) (b)
(c) (d)

(a) **4-bit quantized image. Contours are visible**
(b) **Image + random noise**
(c) **4-bit quantized image of (b)**
(d) **image after subtracting the random noise**

Halftone Image Generation

- Halftone Images
  - Binary images that give a gray scale rendition



- Suppose that u(m,n) = g for every coordinate (m,n)
- Then, u(m,n)+r(m,n) will have the following values with the same probability
  - g, g+1, …, 255, 256, …, 255+g (before thresholding)
  - 0, 0, …, 0, 255, …, 255 (after thresholding)
- Thus, the average gray level will be

$$\frac{256-g}{256}\times 0+\frac{g}{256}\times 255\cong g$$

Halftone Image Generation

- Procedure
  - Optional oversampling (provides better rendition)
    - e.g.) 256x256 →1024x1024 with repetition
  - Add random number
  - Two-level quantization

- Halftone matrix (random number matrix)
  - can be repeated periodically

$$H_1 = \begin{bmatrix} 40 & 60 & 150 & 90 & 10 \\ 80 & 170 & 240 & 200 & 110 \\ 140 & 210 & 250 & 220 & 130 \\ 120 & 190 & 230 & 180 & 70 \\ 20 & 100 & 160 & 50 & 30 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 52 & 44 & 36 & 124 & 132 & 140 & 148 & 156 \\ 60 & 4 & 28 & 116 & 200 & 228 & 236 & 164 \\ 68 & 12 & 20 & 108 & 212 & 252 & 244 & 172 \\ 76 & 84 & 92 & 100 & 204 & 196 & 188 & 180 \\ 132 & 140 & 148 & 156 & 52 & 44 & 36 & 124 \\ 200 & 228 & 236 & 164 & 60 & 4 & 28 & 116 \\ 212 & 252 & 244 & 172 & 68 & 12 & 20 & 108 \\ 204 & 196 & 188 & 180 & 76 & 84 & 92 & 100 \end{bmatrix}$$

Halftone Image Generation

Halftone Image Generation
Without Upsampling

(a) (b)
(c) (d)

(a) Original 8-bit image
(b) Most significant 1-bit image
(c) Halftone screen H_2
(d) Halftone image

Halftone Image Generation

Halftone Image Generation
With Upsampling

(a) (b)
(c) (d)

(a) Halftone screen H_2 (512x512)
(b) Halftone image (512x512)
(c) Halftone screen H_2 (1024x1024)
(d) Halftone image (1024x1024)