การสื่อสารดิจิทัล

ทฤษฎีข่าวสารและการเข้ารหัสแหล่งต้นทาง (12 - 13)

Assoc.Prof.**Piya Kovintavewat**, Ph.D.

Data Storage Technology Research Center

Nakhon Pathom Rajabhat University

http://home.npru.ac.th/piya

"Hope for the best, expect the worst"

Outline

- การวัดปริมาณข่าวสาร
- เอนโทรปี
- ช่องสัญญาณที่ไม่มีหน่วยความจำแบบไม่ต่อเนื่อง
  - ข่าวสารร่วมกัน
- ความจุช่องสัญญาณ
- ขีดจำกัดแชนนอน
- การเข้ารหัสแหล่งต้นทาง
  - ประเภทของรหัส
  - อสมการของคราฟต์
  - ความยาวเฉลี่ยน้อยสุดของคำรหัส
- รหัสฮัฟฟ์แมน
- รหัส ZIV-LEMPEL

บทนำ

- ระบบสื่อสารถูกออกแบบมาเพื่อใช้ส่งข่าวสารจากแหล่งต้นทาง (source) ผ่านช่องสัญญาณ (channel) ไปยังแหล่งปลายทาง (sink)
  - ข่าวสารที่ส่งเป็นได้ทั้งข่าวสารที่เป็นข้อความ ข่าวสารแอนะล็อก และข่าวสารดิจิทัล
- ระบบสื่อสารดิจิทัลถูกออกแบบมาเพื่อใช้ส่ง**เฉพาะ**ข่าวสารดิจิทัล
  - ข่าวสารที่เป็นข้อความ $\Rightarrow$ เข้ารหัสเพื่อแปลงให้เป็นข่าวสารดิจิทัล
  - ข่าวสารแอนะล็อก $\Rightarrow$ แปลงเป็นข่าวสารดิจิทัลโดยผ่านการกล้ำรหัสพัลส์ (PCM)
- นอกจากนี้ข่าวสารดิจิทัล $\Rightarrow$ เข้ารหัสเพิ่มเติมด้วย**วงจรเข้ารหัสแหล่งต้นทาง** (source encoder) เพื่อกำจัดความซ้ำซ้อนของข้อมูลที่จะส่ง
  - ช่วย**ลด**เวลาในการส่งข่าวสาร
  - **ลด**ปริมาณการใช้แบนด์วิดท์ของตัวกลาง

- ในทางปฏิบัติการเข้ารหัสแหล่งต้นทาง (source coding) ⇒ **ไม่ทำให้เอนโทรปีของแหล่งต้นทาง** (source entropy) หรือจำนวนเฉลี่ยของบิตข่าวสารต่อสัญลักษณ์ที่ใช้ส่งผ่านช่องสัญญาณ เปลี่ยนแปลง
- **เอนโทรปี** ⇒ เป็นพารามิเตอร์ที่สำคัญของทฤษฎีข่าวสาร (information theory)
  - สามารถนำมาใช้เป็นตัวบ่งชี้สมรรถนะของระบบสื่อสารดิจิทัลได้

การวัดปริมาณข่าวสาร

- ปริมาณข่าวสาร $\Rightarrow$ สัมพันธ์กับความสามารถในการทำนาย (predictability)
  - ปริมาณข่าวสาร**แปรผันตรง**กับ**ความไม่แน่นอน** (uncertainty) หรือ**แปรผกผัน**กับความน่าจะเป็นของการเกิดขึ้น
  - Pr[ข้อความ] = 1 ถือว่า**ไม่มี**ข่าวสาร
  - Pr[ข้อความ] = 0 ถือว่า**มี**ข่าวสารเป็นอนันต์
- ถ้าให้ I = ปริมาณข่าวสาร

$$I \triangleq \log_a (1/P)$$

เมื่อ a คือฐาน (base) ของฟังก์ชันลอการิทึม

- ถ้า $a = 2 \Rightarrow$ ปริมาณข่าวสารมีหน่วยเป็น**บิต** (bit)

$$I = \log_2 (1/P) = -\log_2 (P) \quad \text{(บิต)}$$

เอนโทรปี

- พิจารณาแหล่งต้นทางที่ไม่มีหน่วยความจำแบบไม่ต่อเนื่อง X ที่ส่งสัญลักษณ์ $\{x_1, x_2, \ldots, x_M\}$ ด้วยความน่าจะเป็น $\{p(x_1), p(x_2), \ldots, p(x_M)\}$ เมื่อ M คือจำนวนสัญลักษณ์ทั้งหมดของ X (แต่ละสัญลักษณ์เป็นอิสระต่อกัน) และ $\sum_{i=1}^{M} p(x_i) = 1$

- **ปริมาณข่าวสาร**ของสัญลักษณ์ x_i ที่มี $p(x_i)$ มีค่าเท่ากับ $I_i = \log_2\left(1/p(x_i)\right)$ บิต และ **ปริมาณข่าวสารเฉลี่ยต่อสัญลักษณ์**ของแหล่งต้นทาง X เรียกว่า**เอนโทรปี** (entropy)

$$H(X) = \sum_{i=1}^{M} p(x_i) I_i = \sum_{i=1}^{M} p(x_i) \log_2\left(\frac{1}{p(x_i)}\right) = -\sum_{i=1}^{M} p(x_i) \log_2\left(p(x_i)\right)$$

มีหน่วยเป็นบิตต่อสัญลักษณ์ (bit per symbol)

- $H(X)$ จะมีค่าสูงสุดก็ต่อเมื่อแต่ละสัญลักษณ์มีความน่าจะเป็นเท่ากัน

- โดยทั่วไป $0 \le H(X) \le \log_2(M)$
- ถ้าให้แหล่งต้นทางส่งข้อมูลด้วยอัตราสัญลักษณ์ R_s สัญลักษณ์ต่อวินาที $\Rightarrow$ อัตราข่าวสาร (information rate) หรืออัตราบิต หาได้จาก

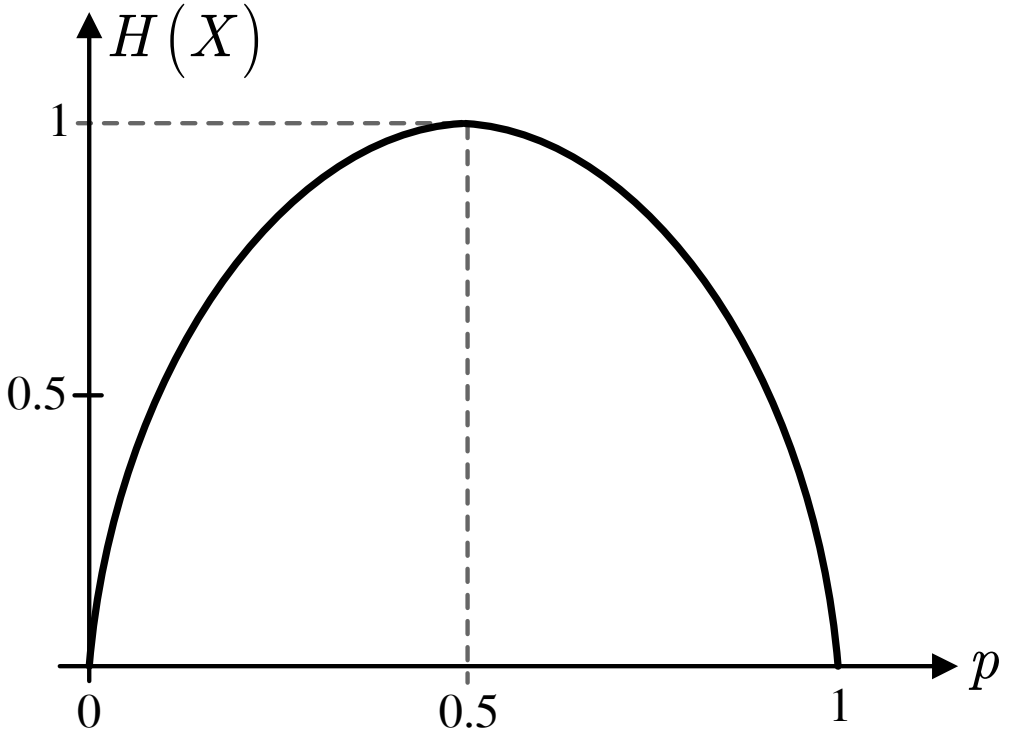
$$R_b = R_s H(X)$$ บิตต่อวินาที (bps)

- **ส่วนซ้ำซ้อน** (redundancy) คือผลต่างระหว่างเอนโทรปีสูงสุดที่เป็นไปได้ของแหล่งต้นทางและเอนโทรปีจริง ซึ่งมีค่าเท่ากับ

$$\tilde{R} = \log_2(M) - H(X)$$ บิตต่อสัญลักษณ์

Example 1

พิจารณาแหล่งต้นทางแบบไบนารี X ที่ส่งข้อมูล 0 และ 1 ด้วยความน่าจะเป็น p และ $1-p$ ตามลำดับ จงหาความน่าจะเป็น p ที่ทำให้เอนโทรปีของแหล่งต้นทางนี้มีค่าสูงสุด



วิธีทำ เอนโทรปีของแหล่งต้นทาง X มีค่าเท่ากับ

$$H(X) = p\log_2(1/p) + (1-p)\log_2(1/(1-p))$$

หาอนุพันธ์ของ $H(X)$ เทียบกับตัวแปร p โดยอาศัย $\frac{d(\log_a(u))}{dx} = \frac{\log_a(e)}{u}\frac{du}{dx}$ เมื่อ $a \neq \{0, 1\}$

$$\frac{dH(X)}{dp} = \left\{ p\frac{\log_2(e)}{1/p}\left(\frac{-1}{p^2}\right) + \log_2\left(\frac{1}{p}\right)\right\} + \left\{ (1-p)\frac{\log_2(e)}{1/(1-p)}\left(\frac{1}{(1-p)^2}\right) + (-1)\log_2\left(\frac{1}{1-p}\right)\right\}$$

$$= \left\{-\log_2(e) - \log_2(p)\right\} + \left\{\log_2(e) - \log_2\left(\frac{1}{1-p}\right)\right\}$$

$$= \log_2(1-p) - \log_2(p)$$

จากนั้นกำหนดให้ $\frac{dH(X)}{dp} = 0$ ก็จะได้ $\log_2(1-p) = \log_2(p)$ ซึ่งเป็นจริง ก็ต่อเมื่อ $p = 0.5$ ดังนั้น แหล่งต้นทางแบบไบนารี $(M = 2)$ จะมีเอนโทรปีสูงสุด เมื่อข้อมูลแต่ละตัวมีความน่าจะเป็นเท่ากัน

Example 2

พิจารณาระบบโทรเลขที่ใช้รหัสมอร์ส (Morse code) ซึ่งประกอบด้วยจุดและขีด (dot and dash) โดยที่การส่งจุดใช้เวลา 0.2 วินาที, การส่งขีดใช้เวลา 0.6 วินาที, ความน่าจะเป็นของการส่งจุดมีค่าเป็น 2 เท่าของความน่าจะเป็นของการส่งขีด, และระยะห่างระหว่างจุดและขีดใช้เวลา 0.2 วินาที จงหาอัตราข่าวสารของโทรเลขนี้

วิธีทำ เนื่องจาก $\Pr[\text{dot}] = 2\Pr[\text{dash}]$ และ $\Pr[\text{dot}] + \Pr[\text{dash}] = 1$ ดังนั้นจะได้

$$\Pr[\text{dot}] = 2/3 \quad \text{และ} \quad \Pr[\text{dash}] = 1/3$$

และเอนโทรปีของข่าวสารนี้มีค่าเท่ากับ

$$H(X) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9183 \quad \text{(บิตต่อสัญลักษณ์)}$$

เนื่องจากเวลาที่ใช้ส่งจุด ขีด และช่องว่างระหว่างจุดและขีด คือ $t_{\text{dot}} = 0.2$ วินาที, $t_{\text{dash}} = 0.6$ วินาที, และ $t_{\text{space}} = 0.2$ วินาที ตามลำดับ ดังนั้นเวลาเฉลี่ยที่ใช้ในการส่งหนึ่งสัญลักษณ์คือ

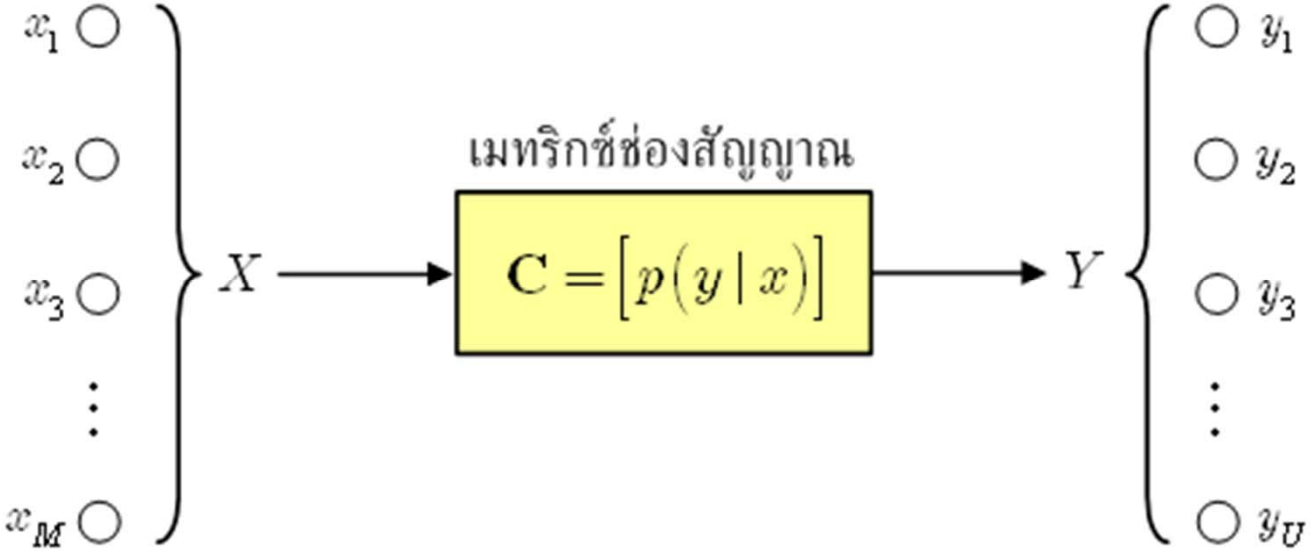
$$T_s = \Pr[\text{dot}] \times t_{\text{dot}} + \Pr[\text{dash}] \times t_{\text{dash}} + t_{\text{space}} = 0.5333 \quad \text{(วินาทีต่อสัญลักษณ์)}$$

ซึ่งจะได้อัตราสัญลักษณ์ $R_s = 1/T_s = 1.875$ สัญลักษณ์ต่อวินาที เพราะฉะนั้นอัตราข่าวสารของโทรเลขมีค่าเท่ากับ

$$R_b = R_s H(X) = 1.875 \times 0.9183 = 1.7218 \quad \text{(บิตต่อวินาที)}$$

ช่องสัญญาณที่ไม่มีหน่วยความจำแบบไม่ต่อเนื่อง

- ช่องสัญญาณ (channel) คือเส้นทางหรือตัวกลาง (medium) สำหรับส่งผ่านสัญลักษณ์จากแหล่งต้นทางไปยังแหล่งปลายทาง
- พิจารณาเฉพาะช่องสัญญาณที่ไม่มีหน่วยความจำแบบไม่ต่อเนื่อง (DMC) ซึ่งหมายถึงแบบจำลองทางสถิติของช่องสัญญาณที่มีข้อมูลอินพุต X และข้อมูลเอาต์พุต Y เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง



$\{x_1, x_2, \ldots, x_M\}$

$\{y_1, y_2, \ldots, y_U\}$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_U \end{bmatrix}}_{\text{output}} = \underbrace{\begin{bmatrix} p(y_1 \mid x_1) & p(y_1 \mid x_2) & \cdots & p(y_1 \mid x_M) \\ p(y_2 \mid x_1) & p(y_2 \mid x_2) & \cdots & p(y_2 \mid x_M) \\ \vdots & \vdots & \ddots & \vdots \\ p(y_U \mid x_1) & p(y_U \mid x_2) & \cdots & p(y_U \mid x_M) \end{bmatrix}}_{\text{channel matrix} = \mathbf{C}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix}}_{\text{input}}$$

$$\sum_{i=1}^{M} p(y_j \mid x_i) = 1 \quad \text{สำหรับทุกค่า} \quad j \in \{1, 2, \ldots, U\}$$

- โดยทั่วไปถ้าช่องสัญญาณ**ไม่มี**สัญญาณรบกวน $\Rightarrow$ สัญลักษณ์ที่ได้รับ $\{y_j\}$ สามารถใช้บ่งบอกถึงสัญลักษณ์ที่ส่ง $\{x_i\}$ ได้

- ถ้าช่องสัญญาณมีสัญญาณรบกวน $\Rightarrow$ พบความไม่แน่นอนเกิดขึ้นใน $\{y_j\}$
  - ความไม่แน่นอนของ x_i เมื่อทราบค่า y_j มีค่าเท่ากับ $\log_2\left(1/p\left(x_i \mid y_j\right)\right)$
  - เมื่อนำความไม่แน่นอนนี้มาหาค่าเฉลี่ย (เทียบกับ x_i และ y_j) ก็จะได้ผลลัพธ์เป็น **เอนโทรปีแบบมีเงื่อนไข** (conditional entropy) ซึ่งนิยามโดย

$$H\left(X \mid Y\right) = \sum_i \sum_j p\left(x_i, y_j\right) \log_2 \left(\frac{1}{p\left(x_i \mid y_j\right)} \right) \quad \text{บิตต่อสัญลักษณ์}$$

  - ถ้าช่องสัญญาณไม่มีสัญญาณรบกวน $\Rightarrow$ $H\left(X \mid Y\right) = 0$

- ในทางปฏิบัติ $H\left(X \mid Y\right)$ เรียกว่า**ความกำกวม** (equivocation) ของ X เมื่อเทียบกับ Y ซึ่งหมายถึงความไม่แน่นอนเฉลี่ยของช่องสัญญาณกับ X เมื่อทราบ Y
- การหา $H\left(X \mid Y\right)$ $\Rightarrow$ ต้องทราบค่าความน่าจะเป็นแบบมีเงื่อนไข $p\left(x_i \mid y_j\right)$ ซึ่งหาได้จาก

$$\text{Bayes's rule} \Rightarrow \quad p\left(x_i \mid y_j\right) = \frac{p\left(y_j, x_i\right)}{p\left(y_j\right)} = \frac{p\left(y_j \mid x_i\right) p\left(x_i\right)}{\sum_i p\left(x_i, y_j\right)} = \frac{p\left(y_j \mid x_i\right) p\left(x_i\right)}{\sum_i p\left(y_j \mid x_i\right) p\left(x_i\right)}$$

- $H(Y)$ คือเอนโทรปีของแหล่งปลายทางหรือความไม่แน่นอนเฉลี่ยของข้อมูลเอาต์พุต Y

$$H(Y)=\sum_i p(y_i)\log_2\left(\frac{1}{p(y_i)}\right)$$

- ความไม่แน่นอนรวม (total uncertainty) ที่เกี่ยวข้องกับคู่สัญลักษณ์ (x_i, y_j) ทั้งหมดที่เป็นไปได้ $\Rightarrow$ **เอนโทรปีประกอบ** (composite entropy)

$$H(X,Y)=\sum_i\sum_j p(x_i,y_j)\log_2\left(\frac{1}{p(x_i,y_j)}\right)$$

$$=H(Y)+H(X\mid Y)$$

$$=H(X)+H(Y\mid X)$$

เมื่อ $H(Y\mid X)=\sum_i\sum_j p(x_i,y_j)\log_2\left(\frac{1}{p(y_j\mid x_i)}\right)$ คือความกำกวมของ Y เมื่อเทียบกับ X

ข่าวสารร่วมกัน

- ถ้าช่องสัญญาณ**ไม่มี**สัญญาณรบกวน ข่าวสารเฉลี่ยที่วงจรภาครับได้รับมีค่าเท่ากับ $H(X)$
- เนื่องจากช่องสัญญาณ**มี**สัญญาณรบกวน จึงทำให้เกิดการสูญเสียข่าวสารเป็นปริมาณเฉลี่ย $H(X|Y)$ บิตต่อสัญลักษณ์ $\Rightarrow$ ปริมาณข่าวสารคงเหลือที่วงจรภาครับได้รับคือ

$$I(X;Y) = H(X) - H(X|Y)$$ บิตต่อสัญลักษณ์

หรือเรียกว่า**ข่าวสารร่วมกัน** (mutual information) ของ X และ Y ซึ่งมีค่าเท่ากับ

$$I(X;Y) = \sum_i p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right) - \sum_i \sum_j p(x_i, y_j) \log_2 \left(\frac{1}{p(x_i | y_j)} \right)$$

$$= \sum_i \sum_j p(x_i, y_j) \log_2 \left(\frac{1}{p(x_i)} \right) - \sum_i \sum_j p(x_i, y_j) \log_2 \left(\frac{1}{p(x_i | y_j)} \right)$$

$$= \sum_i \sum_j p(x_i, y_j) \log_2 \left[\frac{p(x_i \mid y_j)}{p(x_i)} \right] \quad = \sum_i \sum_j p(x_i, y_j) \log_2 \left[\frac{p(x_i, y_j)}{p(x_i) p(y_j)} \right]$$

$$= \sum_i \sum_j p(x_i, y_j) \log_2 \left[\frac{p(y_j \mid x_i)}{p(y_j)} \right] \quad = \sum_i \sum_j p(y_j \mid x_i) p(x_i) \log_2 \left[\frac{p(y_j \mid x_i)}{\sum_i p(y_j \mid x_i) p(x_i)} \right]$$

- $I(X;Y)$ มีคุณสมบัติสมมาตร $\Rightarrow$ $I(X;Y) = I(Y;X)$
- คุณสมบัติที่น่าสนใจมีดังนี้ $I(X;Y) \geq 0$

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

ความจุช่องสัญญาณ

- ถ้ากำหนดช่องสัญญาณมาให้ $\Rightarrow$ สามารถหาค่า $p(x_i)$ ที่ทำให้ $I(X;Y)$ มีค่าสูงสุดได้
- ค่าสูงสุดของ $I(X;Y)$ เรียกว่า**ความจุช่องสัญญาณ** (channel capacity) ซึ่งนิยามโดย

$$C_s = \max_{p(x_i)} \{I(X;Y)\}$$ บิตต่อสัญลักษณ์

- โดยทั่วไปความจุช่องสัญญาณ C_s บอกให้ทราบถึงปริมาณข่าวสารมากสุดที่สามารถส่งผ่านช่องสัญญาณโดยใช้หนึ่งสัญลักษณ์
- ถ้าแหล่งต้นทางส่งข้อมูลด้วยอัตรา R_s สัญลักษณ์ต่อวินาที $\Rightarrow$ ความจุช่องสัญญาณต่อวินาที C มีค่าเท่ากับ $C = R_s C_s$ บิตต่อวินาที

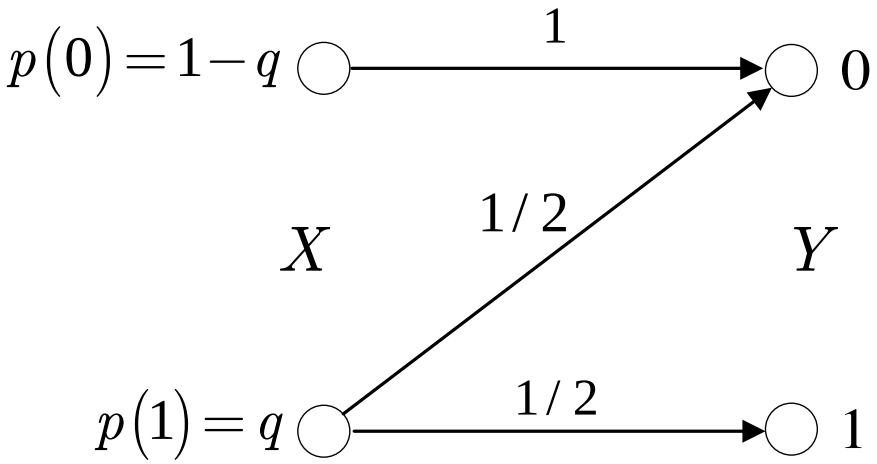
Example 3

- พิจารณาช่องสัญญาณที่มีความน่าจะเป็นการเปลี่ยนสถานะ (transition probability) ตามรูปต่อไปนี้ จงหา

  ก) เอนโทรปีของข้อมูลเอาต์พุต Y นั่นคือหาค่า $H(Y)$

  ข) ค่า q ที่ทำให้ความจุช่องสัญญาณ C_s มีค่ามากสุด

  ค) ค่า q ที่ทำให้ข้อมูลอินพุต $H(X)$ มีค่ามากกว่า $H(Y)$



วิธีทำ

ก) จากรูปพบว่าความน่าจะเป็นของ $p(y)$ สำหรับ $y \in \{0,1\}$ หาได้จาก

$$p(y=0) = p(y=0 \mid x=0)p(x=0) + p(y=0 \mid x=1)p(x=1)$$

$$= (1)(1-q) + (1/2)q = 1 - q/2$$

$$p(y=1) = p(y=1 \mid x=0)p(x=0) + p(y=1 \mid x=1)p(x=1)$$

$$= (0)(1-q) + (1/2)q = q/2$$

ดังนั้นเอนโทรปีของข้อมูลเอาต์พุต Y มีค่าเท่ากับ

$$H(Y) = -(1-q/2)\log_2(1-q/2) - (q/2)\log_2(q/2)$$

ข) ความจุช่องสัญญาณหาได้จาก $I(X;Y)=H(Y)-H(Y|X)$ และ

$$H(Y|X)=-\sum_i\sum_j p(y_j|x_i)p(x_i)\log_2\left(p(y_j|x_i)\right)$$

$$=-(1)(1-q)\log_2(1)-0-(1/2)(q)\log_2(1/2)-(1/2)(q)\log_2(1/2)=q$$

ซึ่งจะได้ $I(X;Y)=-\left(1-\frac{q}{2}\right)\log_2\left(1-\frac{q}{2}\right)-\left(\frac{q}{2}\right)\log_2\left(\frac{q}{2}\right)-q$ ดังนั้นค่า q ที่ทำให้ $I(X;Y)$ มีค่าสูงสุดหาได้โดยการหาอนุพันธ์ของ $I(X;Y)$ เทียบกับ q แล้วให้ผลสัพธ์มีค่าเท่ากับศูนย์จะได้

$$\frac{\partial I(X;Y)}{\partial q}=-\left(1-\frac{q}{2}\right)\left(\frac{1}{1-q/2}\right)\left(-\frac{1}{2}\right)\frac{1}{\ln(2)}+\left(\frac{1}{2}\right)\log_2\left(1-\frac{q}{2}\right)$$

$$-\left(\frac{q}{2}\right)\left(\frac{2}{q}\right)\left(\frac{1}{2}\right)\frac{1}{\ln(2)}-\left(\frac{1}{2}\right)\log_2\left(\frac{q}{2}\right)-1$$

$$=\frac{1}{2}\log_2\left(\frac{2-q}{2}\right)-1=0$$

ก็จะได้ $q=2/5$ ที่ทำให้ $I(X;Y)$ มีค่าสูงสุด

ค) เนื่องจาก $H(X) = -(1-q)\log_2(1-q) - (q)\log_2(q)$ ดังนั้นค่า q ที่ทำให้ $H(X) > H(Y)$ หาได้โดย

$$-(1-q)\log_2(1-q) - (q)\log_2(q) > -(1-q/2)\log_2(1-q/2) - (q/2)\log_2(q/2)$$

$$-2(1-q)\log_2(1-q) - q\log_2(q) > 2 - (2-q)\log_2(2-q)$$

$$\log_2(1-q)^{-2+2q} - \log_2(q)^q > \log_2(4) - \log_2(2-q)^{2-q}$$

$$\frac{(2-q)^{2-q} \times (1-q)^{-2+2q}}{q^q} > 4$$

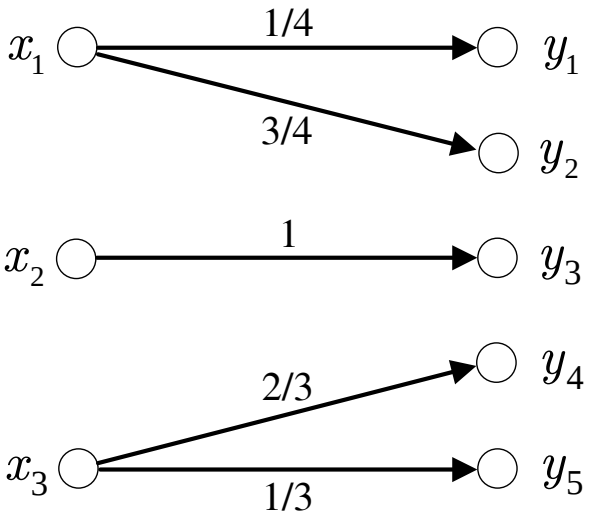
ซึ่งจะได้ $q < 2/3$ ที่ทำให้ $H(X) > H(Y)$

ช่องสัญญาณที่ไม่มีการสูญเสีย

- อธิบายได้ด้วยเมทริกซ์ช่องสัญญาณที่ในแต่ละแนวนอน (row) มีค่าไม่เป็นศูนย์เพียงค่าเดียว
- ข่าวสารร่วมกัน $\Rightarrow$ $I(X;Y) = H(X)$
- ความจุช่องสัญญาณ $\Rightarrow$ $C_s = \max_{p(x_i)} \{I(X;Y)\} = \log_2(M)$ เมื่อ M คือจำนวนสัญลักษณ์ทั้งหมดของข้อมูลอินพุต X

ช่องสัญญาณเชิงกำหนด

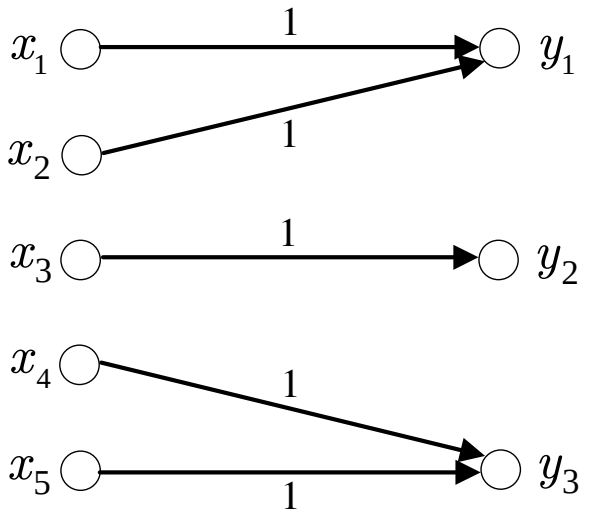
- อธิบายได้ด้วยเมทริกซ์ช่องสัญญาณที่ในแต่ละแนวตั้ง (column) มีค่าไม่เป็นศูนย์เพียงค่าเดียว
- ข่าวสารร่วมกัน $\Rightarrow$ $I(X;Y) = H(Y)$
- ความจุช่องสัญญาณ $\Rightarrow$ $C_s = \max_{p(x_i)} \{I(X;Y)\} = \log_2(U)$ เมื่อ U คือจำนวนสัญลักษณ์ทั้งหมดของข้อมูลเอาต์พุต Y



(ก)

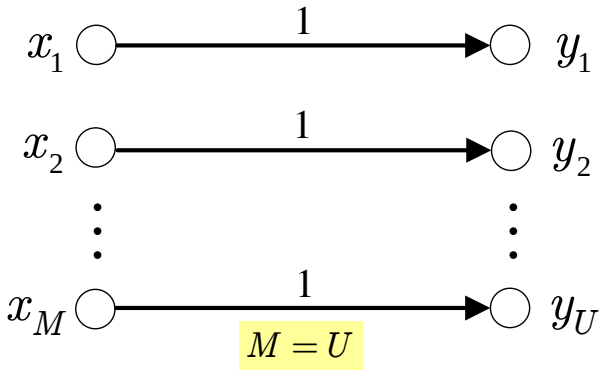
เมทริกซ์ช่องสัญญาณ

$$\begin{bmatrix} 1/4 & 0 & 0 \\ 3/4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2/3 \\ 0 & 0 & 1/3 \end{bmatrix}$$



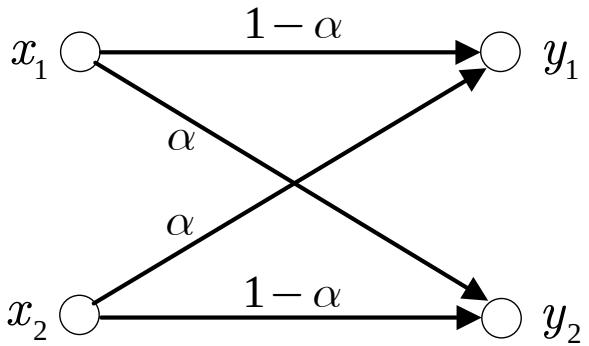
(ข)

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$



(ค)

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$



(ง)

$$\begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}$$

ช่องสัญญาณที่ไม่มีสัญญาณรบกวน

- อธิบายได้ด้วยเมทริกซ์ช่องสัญญาณที่มีค่าไม่เป็นศูนย์เฉพาะในแนวเส้นทแยงมุมหลัก
- ข่าวสารร่วมกัน $\Rightarrow$ $I(X;Y)=H(X)=H(Y)$
- ความจุช่องสัญญาณ $\Rightarrow$ $C_s=\max_{p(x_i)}\{I(X;Y)\}=\log_2(M)=\log_2(U)$

ช่องสัญญาณสมมาตรแบบไบนารี

- หมายถึงความน่าจะเป็นของการรับข้อมูลบิต 0 เมื่อส่งบิต 1 มีค่าเท่ากับความน่าจะเป็นของการรับข้อมูลบิต 1 เมื่อส่งบิต 0 ซึ่งมีค่าเท่ากับความน่าจะเป็นตัดข้าม α
- ข่าวสารร่วมกัน $\Rightarrow$ $I(X;Y)=H(Y)+\alpha\log_2(\alpha)+(1-\alpha)\log_2(1-\alpha)$
- ความจุช่องสัญญาณ $\Rightarrow$ $C_s=\max_{p(x_i)}\{I(X;Y)\}=1+\alpha\log_2(\alpha)+(1-\alpha)\log_2(1-\alpha)$

ความจุช่องสัญญาณ AWGN

- นายแชนนอน (1948) $\Rightarrow$ ความจุของช่องสัญญาณที่มีแบนด์วิดท์ W (เฮิรตซ์) และถูกรบกวนด้วยสัญญาณรบกวนเกาส์สีขาวแบบบวก (AWGN) เป็นได้ตามทฤษฎีบท Shannon-Hartley ซึ่งมีค่าเท่ากับ

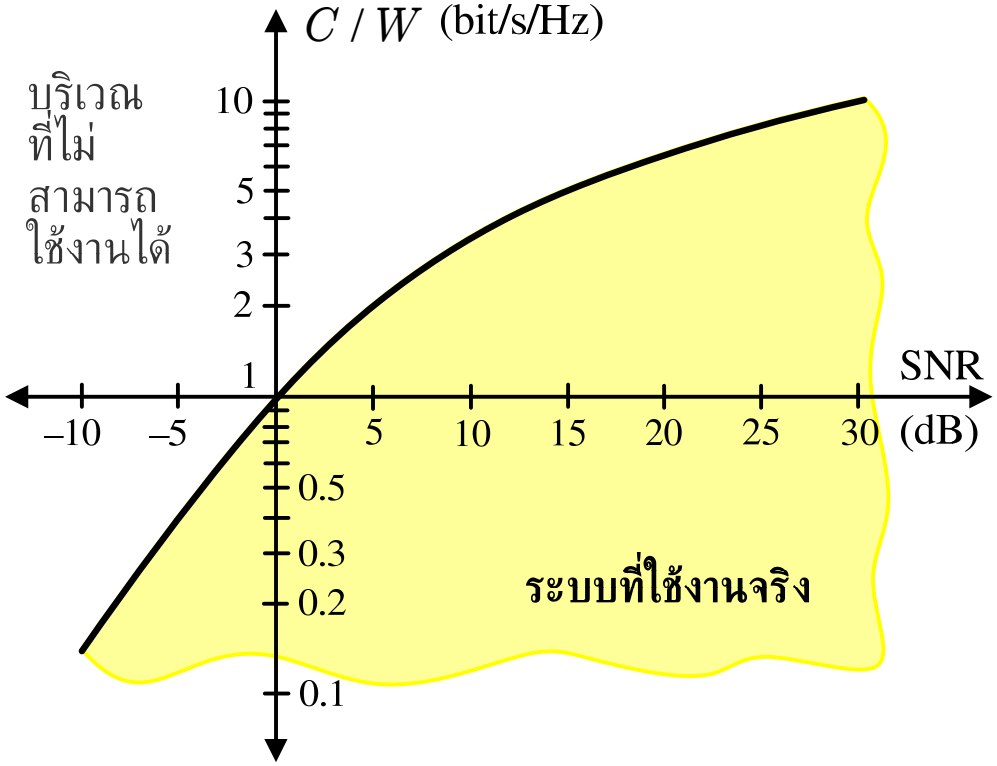
$$C = W \log_2 \left(1 + \frac{S}{N}\right) \quad \text{(bps)}$$

เมื่อ S คือกำลังเฉลี่ยของสัญญาณ
และ N คือกำลังเฉลี่ยของสัญญาณรบกวน

- ตราบใดที่ระบบใช้อัตราบิต $R_b \leq C$ ก็**มีความเป็นไปได้**ที่จะส่งข้อมูลจากต้นทางผ่านช่องสัญญาณ AWGN ไปถึงปลายทาง โดยมีข้อผิดพลาดน้อยเท่าใดก็ได้ $(P_e \to 0)$ ถ้าอาศัยความช่วยเหลือของรหัสแก้ไขข้อผิดพลาด (ECC: error correction code)
- ถ้าระบบใช้ $R_b > C$ จะ**ไม่สามารถ**ส่งข้อมูลจากต้นทางไปถึงปลายทางโดยไม่มีข้อผิดพลาด
- พารามิเตอร์ W, S และ N เป็นตัวกำหนดขีดจำกัด (limit) ของ**อัตราการส่งผ่าน** (transmission rate) **ไม่ใช่**เป็นตัวกำหนดขีดจำกัดของ P_e ของระบบ

- ขอบเขต (bound) สำหรับสมรรถนะที่สามารถทำได้ของระบบสื่อสารที่ใช้งานจริง
- ระบบที่ใช้งานจริงสามารถทำงานได้เฉพาะในบริเวณที่แรเงาเท่านั้น
- เนื่องจากกำลังเฉลี่ยของสัญญาณรบกวน N เป็นสัดส่วนกับแบนด์วิดท์ $W \Rightarrow N = N_0 W$
  เมื่อ $N_0 / 2$ คือ PSD แบบสองด้านของสัญญาณรบกวนเกาส์สีขาว ดังนั้น

$$\frac{C}{W} = \log_2\left(1 + \frac{S}{N_0 W}\right)$$ (บิตต่อวินาทีต่อเฮิรตซ์)

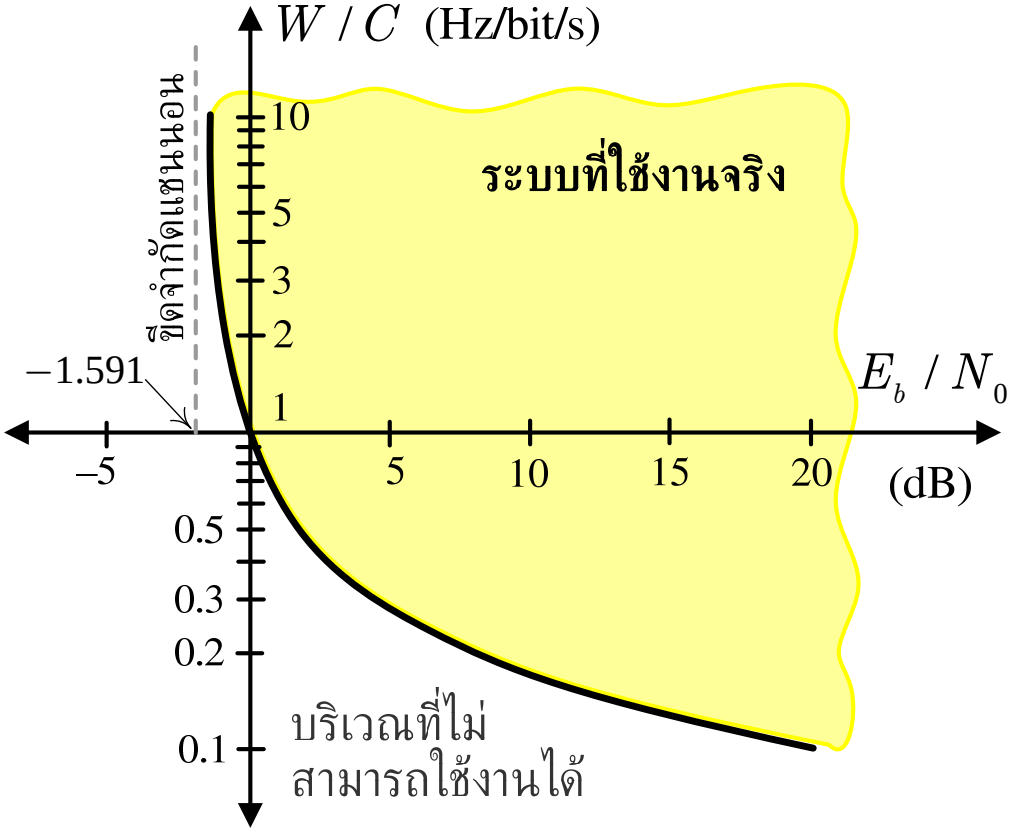


normalized channel capacity

- ค่า C/W ให้อยู่ในรูปที่เป็นฟังก์ชัน E_b/N_0 เมื่อ E_b คือพลังงานเฉลี่ยต่อบิต
- ในกรณีที่อัตราบิตมีค่าเท่ากับความจุช่องสัญญาณจะได้ว่า $E_b/N_0 = S/(N_0 C)$ ดังนั้น

$$\frac{C}{W} = \log_2\left(1 + \frac{C}{W}\frac{E_b}{N_0}\right) \quad \Rightarrow \quad \frac{E_b}{N_0} = \frac{W}{C}\left(2^{C/W} - 1\right)$$

- พฤติกรรมเชิงเส้นกำกับ (asymptotic behavior) โดยที่ $W/C \to \infty$ (หรือ $C/W \to 0$)



normalized channel bandwidth

ขีดจำกัดแชนนอน

- **ขีดจำกัดแชนนอน** (Shannon limit) $\Rightarrow$ ค่า E_b / N_0 น้อยสุดที่ยังคงทำให้สามารถส่งข้อมูลได้โดยไม่มีข้อผิดพลาด ณ อัตราส่งข้อมูลใดๆ (โดยไม่ต้องคำนึงถึงแบนด์วิดท์ที่ใช้)
- ถ้าให้ $x = (C/W)(E_b/N_0)$ จะได้

$$\frac{C}{W} = x\log_2(1+x)^{1/x} \quad \Rightarrow \quad 1 = \frac{E_b}{N_0}\log_2(1+x)^{1/x} \quad \Rightarrow$$

อาศัยเอกลักษณ์ $\lim_{x\to 0}(1+x)^{1/x} = e$ และเมื่อ $C/W \to 0$ $\Rightarrow$ $\frac{E_b}{N_0} = \frac{1}{\log_2(e)} = 0.693$

หรือเท่ากับ **–1.591** dB

- ผลงานของนายแชนนอนยังบอกให้ทราบถึง**การมีอยู่จริง**ของเทคนิคการเข้ารหัส (coding) และการกล้ำสัญญาณ (modulation) ที่จะช่วยทำให้สามารถบรรลุความจุช่องสัญญาณนี้

Example 4

พิจารณาสัญญาณแอนะล็อกที่มีแบนด์วิดท์ 4000 เฮิรตซ์ ถูกชักตัวอย่างด้วยอัตรา 1.25 เท่าของอัตราไนควิสต์ โดยแต่ละแซมเปิลที่ได้จากการชักตัวอย่างจะถูกแบ่งเป็น 256 ระดับเท่าๆ กัน สมมุติว่าแต่ละแซมเปิลเป็นอิสระต่อกัน

ก) จงหาอัตราข่าวสารของกระบวนการชักตัวอย่างนี้

ข) ถ้าส่งข้อมูลแซมเปิลเหล่านี้ผ่านช่องสัญญาณ AWGN ที่มีแบนด์วิดท์ 10000 เฮิรตซ์ และมี $S/N = 20$ dB จงหาว่าข้อมูลแซมเปิลเหล่านี้จะสามารถส่งผ่านช่องสัญญาณ AWGN โดยไม่มีข้อผิดพลาดได้หรือไม่ (ถ้าไม่จงหาค่า S/N ที่ทำให้การส่งผ่านนี้ไม่มีข้อผิดพลาดเกิดขึ้น)

ค) จงหาแบนด์วิดท์ของช่องสัญญาณที่ต้องการ ที่ทำให้การส่งผ่านข้อมูลแซมเปิลไม่มีข้อผิดพลาดเกิดขึ้น เมื่อ $S/N = 20$ dB

วิธีทำ

ก) เนื่องจากสัญญาณแอนะล็อกมีแบนด์วิดท์ $W = 4000$ เฮิรตซ์ ก็จะได้อัตราไนควิสต์มีค่าเท่ากับ $f_N = 2W = 8000$ แซมเปิลต่อวินาที ดังนั้นความถี่การชักตัวอย่างจะมีค่าเท่ากับ

$$f_s = 1.25 f_N = 10000 \quad \text{(แซมเปิลต่อวินาที)}$$

และเนื่องจากแต่ละแซมเปิลถูกแบ่งเป็น 256 ระดับ ดังนั้นเอนโทรปีของแซมเปิลมีค่าเท่ากับ

$$H(X) = \log_2(256) = 8 \quad \text{(บิตต่อแซมเปิล)}$$

เพราะฉะนั้นอัตราข่าวสารของกระบวนการชักตัวอย่างนี้มีค่าเท่ากับ

$$R_b = f_s H(X) = 10000 \times 8 = 80000 \quad \text{(บิตต่อวินาที)}$$

ข) ความจุของช่องสัญญาณ AWGN หาได้จาก

$$C = W\log_2(1+S/N) = 10^4 \times \log_2(1+100) = 66582.115 \quad \text{(บิตต่อวินาที)}$$

จากข้อ ก) จะได้ $R_b = 80000$ บิตต่อวินาที ซึ่งมีค่ามากกว่าค่า $C = 66582.115$ บิตต่อวินาที จึงทำให้**ไม่**สามารถส่งข้อมูลแซมเปิลเหล่านี้ผ่านช่องสัญญาณ AWGN โดยไม่มีข้อผิดพลาดได้

เพื่อให้สามารถส่งข้อมูลแซมเปิลผ่านช่องสัญญาณ AWGN โดยไม่มีข้อผิดพลาดเกิดขึ้น ระบบจะต้องใช้ $R_b \le C$ ดังนั้นค่า S/N ที่ทำให้ $R_b \le C$ หาได้โดย

$$C = 10^4 \log_2(1+S/N) \ge 80000$$

เมื่อแก้สมการนี้ก็จะได้ $S/N \ge 255$ หรือ 24.065 dB

ค) แบนด์วิดท์ของช่องสัญญาณที่ทำให้การส่งผ่านข้อมูลแซมเปิลไม่มีข้อผิดพลาดเกิดขึ้น $\Rightarrow$ ต้องใช้ $R_b \le C$) เมื่อ $S/N = 20$ dB หาได้โดย $C = W\log_2(1+100) \ge 80000$ ซึ่งเมื่อแก้สมการนี้ก็จะได้ $W \ge 12015.239$ เฮิรตซ์

การเข้ารหัสแหล่งต้นทาง

- Source coding เป็นการเปลี่ยนข่าวสารที่สร้างโดยแหล่งต้นทางที่ไม่มีหน่วยความจำแบบไม่ต่อเนื่องให้เป็นลำดับไบนารีหรือที่เรียกว่า**คำรหัส** (codeword)

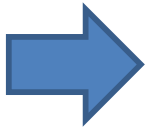
$$X = \{x_1, x_2, \ldots, x_M\}$$

  - เพื่อ**ลด**จำนวนบิต (ตัดบิตที่ซ้ำซ้อนทิ้งไป) ของข้อมูลที่จะส่งไปยังปลายทาง
  - **ลด**เวลาในการส่งข่าวสารและ**ลด**ปริมาณการใช้แบนด์วิดท์ของตัวกลาง
- เข้ารหัสแหล่งต้นทางจะ**ไม่**เปลี่ยนเอนโทรปีของแหล่งต้นทาง แต่จะช่วย**เพิ่ม**เอนโทรปีของ**สัญลักษณ์ที่ถูกเข้ารหัส** (coded symbol)

ประเภทของรหัส

- **รหัส** ⇒ เป็นการจับคู่หรือการแทน**ข้อความ**หนึ่ง (ลำดับสัญลักษณ์ที่ประกอบด้วย n ≥ 1 สัญลักษณ์) ให้อยู่ในรูปของอีกข้อความหนึ่งที่เรียกว่า**คำรหัส**
- **รหัสเอกฐาน** (singular code) ⇒ สัญลักษณ์ต้นทางมากกว่าหนึ่งสัญลักษณ์จะให้คำรหัสเหมือนกัน
  - รหัสที่ไม่มีลักษณะเป็นเอกฐาน ⇒ **รหัสไม่เอกฐาน** (non-singular code)

สัญลักษณ์ต้นทาง	รหัส I	รหัส II
A	0	0
B	1	10
C	01	11



รหัสไม่เอกฐาน

รหัส II เป็นรหัส UDC

  - ถ้านำสัญลักษณ์ A, B หรือ C มาต่อกัน เช่น ABC จะได้คำรหัส 0101 ซึ่งเหมือนกับข้อความ CC ⇒ รหัส I และรหัส II ถือว่าเป็นรหัสที่ไม่ดี

- **ภาคขยายอันดับที่ *n*** (*n*-th order extension) ของรหัส ⇒ คำรหัสที่เกิดขึ้นเมื่อนำสัญลักษณ์ต้นทางจำนวน n สัญลักษณ์มาต่อกันในทุกแบบที่เป็นไปได้

ภาคขยายอันดับที่ 2 ของรหัส I และรหัส II

คู่สัญลักษณ์	รหัส I	รหัส II	คู่สัญลักษณ์	รหัส I	รหัส II
AA	00	00	BC	101	1011
AB	01	010	CA	010	110
AC	001	011	CB	011	1110
BA	10	100	CC	0101	1111
BB	11	1010			

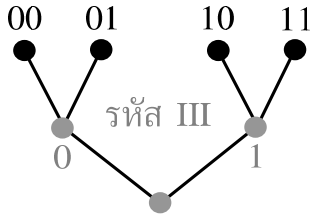
- **รหัสที่ดี** (good code) ต้องสามารถ**ถอดรหัสได้อย่างเดียว** ⇒ เป็นจริงเมื่อภาคขยายอันดับที่ *n* ของรหัสไม่มีลักษณะเป็นเอกฐานสำหรับทุกค่า n

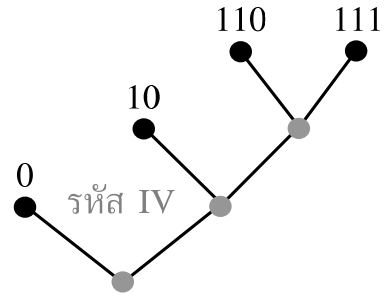
- **รหัสที่สามารถถอดรหัสได้อย่างเดียว** (UDC) ⇒ ทุกคำรหัสต้องสอดคล้องกับสัญลักษณ์ต้นทางเพียงสัญลักษณ์เดียวเท่านั้น รหัส II ในตารางหน้า 34
- วงจรถอดรหัส UDC ⇒ ค่อนข้างซับซ้อน (ต้องเก็บคำรหัสทั้งหมดที่สอดคล้องกับการต่อกันของสัญลักษณ์ต้นทางจำนวน n สัญลักษณ์ทุกรูปแบบที่เป็นไปได้ทั้งหมดไว้ในหน่วยความจำ)
- **รหัสขณะหนึ่ง** (instantaneous code) ⇒ รหัสไม่เอกฐานที่ทุกคำรหัส **ไม่เป็น**ส่วนเติมหน้า (prefix) ของคำรหัสใดๆ
  - รหัสขณะหนึ่งทุกคำรหัส ⇒ แสดงให้อยู่ในรูปของ**แผนภาพต้นไม้**ได้
  - แต่ละคำรหัส ⇒ อยู่ที่ปลายของแต่ละสาขาในแผนภาพต้นไม้

สัญลักษณ์ต้นทาง	รหัส III	รหัส IV	รหัส V
A	00	0	0
B	01	10	01
C	10	110	011
D	11	111	0111

รหัส III และรหัส IV เป็นรหัสขณะหนึ่ง

รหัส V ไม่เป็นรหัสขณะหนึ่ง





อสมการของคราฟต์

- **รหัสขณะหนึ่ง**ทุกคำรหัสต้องมีคุณสมบัติสอดคล้องกับ**อสมการของคราฟต์** (Kraft's inequality) ดังนี้

$$\sum_{i=1}^{N} D^{-l_i} \leq 1$$

เมื่อ D คือจำนวนสัญลักษณ์ที่เป็นไปได้ทั้งหมดที่ใช้ในคำรหัส, l_i คือความยาวของคำรหัสตัวที่ i, และ N คือจำนวนคำรหัสทั้งหมดที่ใช้ในระบบ

- ผู้ใช้สามารถสร้างรหัสขณะหนึ่งที่มีคำรหัสยาว l_i บิตจากเซตของตัวอักษรที่มี D สัญลักษณ์ ตราบใดที่ความยาวของคำรหัส l_i มีความยาวเพียงพอที่ทำให้สมการข้างต้นเป็นจริง

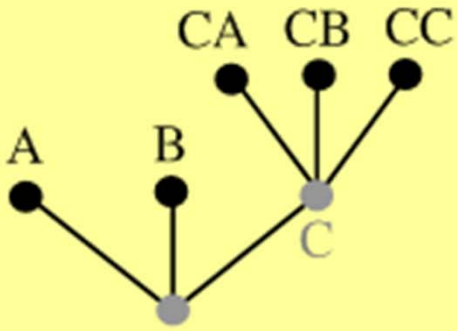
Example 5

ถ้าให้แหล่งต้นทางใช้เซตของตัวอักษร {A, B, C} นั่นคือ $D = 3$ จงพิจารณาว่าคำรหัสจำนวน 4 คำรหัสที่มีความยาวเท่ากับ 1, 2, 2 และ 2 บิต สามารถเป็นรหัสขณะหนึ่งได้หรือไม่

วิธีทำ จากสมการจะได้

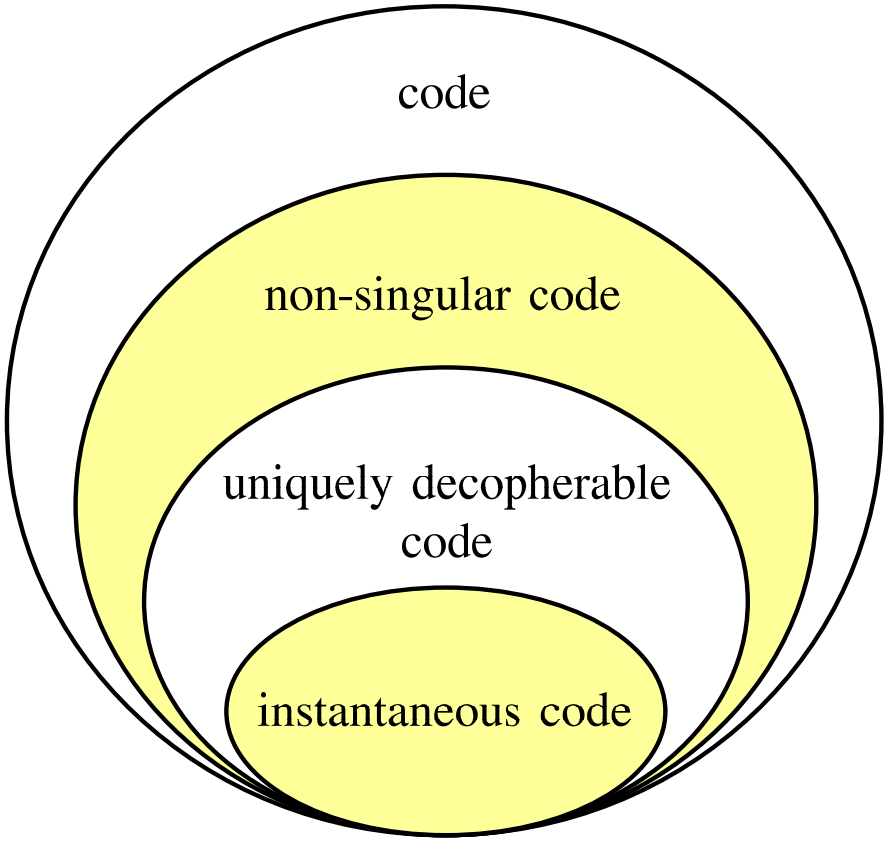
$$\sum_{i=1}^{N} D^{-l_i} = 3^{-1} + 3^{-2} + 3^{-2} + 3^{-2} = 1 \leq 1$$

ดังนั้นคำรหัสทั้งหมดเป็นรหัสขณะหนึ่ง



ตัวอย่างคำรหัสทั้ง 4 แบบที่เป็นรหัสขณะหนึ่ง

- รหัสขณะหนึ่งที่ทำให้ $\sum_{i=1}^{N} D^{-l_i} = 1$ เรียกว่า**รหัสกระชับ** (compact code)
- อสมการของคราฟต์ ⇒ ใช้อธิบายคุณสมบัติของรหัส UDC ได้ทุกคำรหัสเช่นกัน ซึ่งในกรณีนี้จะเรียกว่า**อสมการแมคมิลแลน** (McMillan's inequality)



ความยาวเฉลี่ยน้อยสุดของคำรหัส

- การเปรียบเทียบสมรรถนะของ**รหัสต้นทาง**แบบต่างๆ ⇒ พิจารณาจากความยาว (หรือจำนวนบิต) เฉลี่ยน้อยสุดของคำรหัส
  - เป็นตัวกำหนดปริมาณหน่วยความจำที่ต้องใช้ในการเก็บข่าวสารและปริมาณเวลาที่ใช้ในการส่งข่าวสารจากต้นทางไปยังปลายทาง

- ถ้าให้ $X \in \{x_1, x_2, \ldots, x_M\}$ คือแหล่งต้นทางที่ไม่มีหน่วยความจำแบบไม่ต่อเนื่องที่มีเอนโทรปี $H(X)$ และมีความน่าจะเป็น $p(x_i)$ สำหรับ $i = \{1, 2, \ldots, M\}$
- ถ้าให้ l_i คือความยาวของคำรหัสที่ได้จากการเข้ารหัสสัญลักษณ์ $\{x_i\}$ ดังนั้น**ความยาวเฉลี่ยของคำรหัส** (average length of codeword) มีค่าเท่ากับ

$$\bar{L} = \sum_{i=1}^{M} p(x_i) l_i \quad \text{(บิต)}$$

- ในทางปฏิบัติความยาวเฉลี่ยของคำรหัส UDC ต้องสอดคล้องกับความสัมพันธ์ดังนี้

$$\bar{L}_{\min} \leq \bar{L} < \bar{L}_{\min} + 1$$

เมื่อ $\bar{L}_{\min} = H(X)$ คือความยาวเฉลี่ยน้อยสุดของคำรหัส

- **ประสิทธิภาพรหัส** (code efficiency) นิยามโดย $\eta = \dfrac{H(X)}{\bar{L}}$
- **ความซ้ำซ้อนของการเข้ารหัสมีค่าเท่ากับ** $\gamma = 1 - \eta$

Example 6

เครื่องกราดภาพ (scanner) แปลงเอกสารแบบขาวดำทีละเส้นให้เป็นลำดับไบนารีก่อนจะส่งผลลัพธ์ที่ได้ไปยังปลายทาง ถ้าเครื่องกราดภาพสร้างสัญลักษณ์ 6 แบบ ซึ่งถูกแทนด้วยจุดภาพ (pixel) ที่เรียงต่อกันตั้งแต่ 1 ถึง 6 จุดภาพ โดยแต่ละแบบมีความน่าจะเป็นที่จะเกิดขึ้นดังนี้

จำนวนจุดภาพที่เรียงต่อกัน $\{l_i\}$	1	2	3	4	5	6
ความน่าจะเป็นที่จะเกิดขึ้น $\{p(x_i)\}$	0.2	0.4	0.15	0.1	0.06	0.09

จงหาความยาวเฉลี่ยของจำนวนจุดภาพที่เรียงต่อกันและหาอัตราข่าวสารของเครื่องกราดภาพที่ทำงานด้วยความเร็ว 1000 จุดภาพต่อวินาที

วิธีทำ เอนโทรปีของเครื่องกราดภาพหาได้จาก

$$H(X)=\sum_{i=1}^{6}p(x_i)\log_2\left(\frac{1}{p(x_i)}\right)=2.29 \quad \text{(บิตต่อสัญลักษณ์)}$$

ในทำนองเดียวกันความยาวเฉลี่ยของจำนวนจุดภาพที่เรียงต่อกันหาได้จาก

$$\bar{L}=(0.2)(1)+(0.4)(2)+(0.15)(3)+(0.1)(4)+(0.06)(5)+(0.09)(6)=2.69 \quad \text{(จุดภาพ)}$$

ดังนั้นประสิทธิภาพของการเข้ารหัสข้อมูลของเครื่องกราดภาพมีค่าเท่ากับ

$$\eta=\frac{H(X)}{\bar{L}}=\frac{2.29}{2.69}=0.8513$$

ณ ความเร็ว 1000 จุดภาพต่อวินาที เครื่องกราดภาพจะทำงานด้วยอัตรา 1000/2.69 = 372 สัญลักษณ์ต่อวินาที ซึ่งจะได้อัตราข่าวสารของเครื่องกราดภาพนี้มีค่าเท่ากับ 2.29×372 = 852 บิตต่อวินาที

ทฤษฎีบทการเข้ารหัสแหล่งต้นทาง

การเข้ารหัสข้อความที่ประกอบด้วย n สัญลักษณ์ ซึ่งมาจากแหล่งต้นทาง X ที่มีเอนโทรปี $H(X)$ และแต่ละสัญลักษณ์เป็นอิสระต่อกัน จำเป็นต้องใช้จำนวนบิตน้อยสุดเท่ากับ $nH(X)$ บิตต่อข้อความ หรือคิดเป็น $H(X)$ บิตต่อสัญลักษณ์ เมื่อ $n \to \infty$ เท่านั้น

- ความยาวเฉลี่ยน้อยสุดของคำรหัสที่ได้จากการเข้ารหัส 1 สัญลักษณ์มีค่าเท่ากับ $\bar{L} = H(X)$ บิตต่อสัญลักษณ์
- ความยาวเฉลี่ยน้อยสุดของคำรหัสที่ได้จากการเข้ารหัสข้อความที่ประกอบด้วย n สัญลักษณ์ (เมื่อ $n \to \infty$) จะมีค่าเท่ากับ $\bar{L}_n = nH(X)$ บิตต่อข้อความ
- โดยทั่วไปการเข้ารหัสข้อความที่มีความยาวจำกัดจะทำให้ $\bar{L}_n > nH(X)$ เสมอ $\Rightarrow$ การเลือกใช้รหัสต้นทางที่ทำให้ $\bar{L}_n \to nH(X)$ เมื่อ n มีค่าจำกัด จะช่วยเพิ่มสมรรถนะรวมของระบบได้ (เช่น รหัสฮัฟฟ์แมน และรหัส Ziv-Limpel)

รหัสฮัฟฟ์แมน

- พัฒนาขึ้นโดย David A. Huffman ในปี ค.ศ. 1952
- เป็นอัลกอริทึมการเข้ารหัสของแหล่งต้นทางที่ใช้สำหรับการบีบอัดข้อมูลที่ **ไม่มีการสูญเสีย** (lossless data compression)
- เป็น**รหัสกระชับ** เพราะทำให้อสมการของคราฟต์มีค่าเท่ากับ 1
- เป็น**รหัสเหมาะที่สุด** (optimum code) ที่ใช้เข้ารหัสลำดับสัญลักษณ์ที่มีความยาวจำกัดและให้ผลลัพธ์เป็นคำรหัสที่มีความยาวเฉลี่ยน้อยสุด นั่นคือ

$$\bar{L}_n \rightarrow nH(X)$$

เมื่อ n มีค่าจำกัด

ขั้นตอนการหารหัสฮัฟฟ์แมนแบบไบนารี (คำรหัสอยู่ในรูปของลำดับบิต)

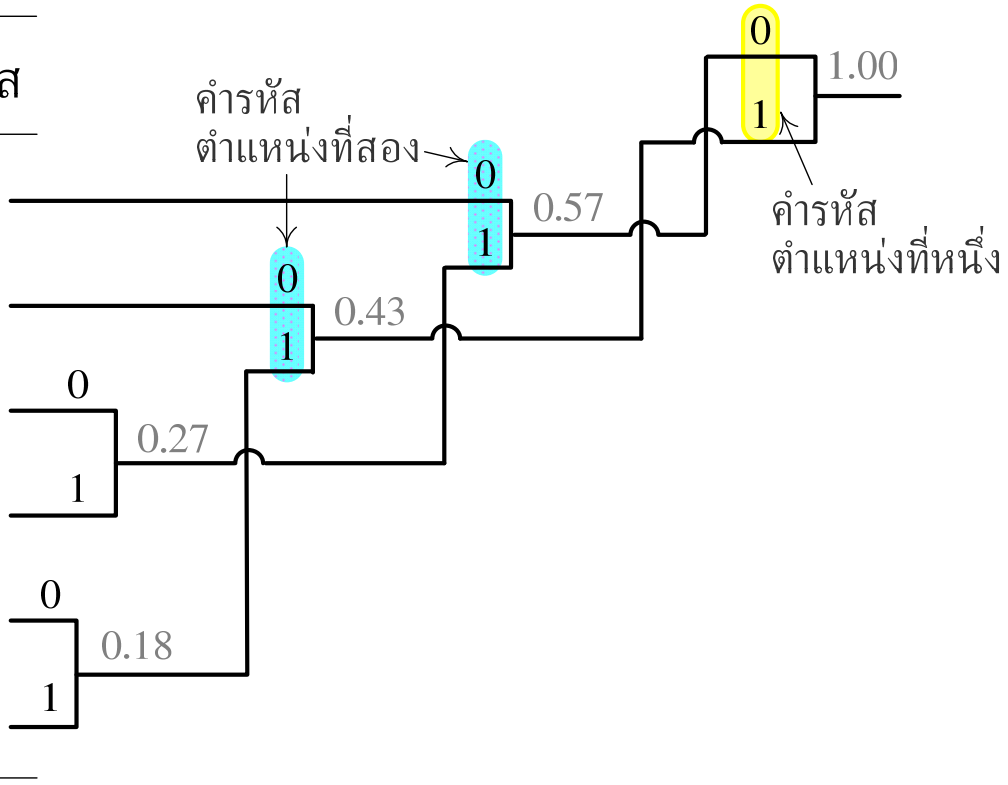
1) นำสัญลักษณ์ทั้งหมดของแหล่งต้นทางมาเรียงลำดับตามค่าความน่าจะเป็นของแต่ละสัญลักษณ์จากมากไปหาน้อย
2) รวมความน่าจะเป็นที่มีค่าน้อยสุดสองค่าเข้าด้วยกัน แล้วจัดเรียงลำดับความน่าจะเป็นใหม่ (จากมากไปหาน้อย) ให้ทำเช่นนี้ไปเรื่อยๆ จนกระทั่งเหลือค่าสุดท้ายเพียงสองค่า
3) เริ่มต้นทำการเข้ารหัสจากด้านขวามือสุด โดยให้ความน่าจะเป็นตัวแรก (ตัวบน) มีค่าเป็น 0 และความน่าจะเป็นตัวที่สอง (ตัวล่าง) มีค่าเป็น 1
4) ย้อนกลับมากำหนดให้บิตตำแหน่งที่สองมีค่าเป็น 0 หรือ 1 ที่ได้จากการรวมกันในขั้นตอนที่ 2
5) ทำซ้ำในลักษณะนี้ไปเรื่อยๆ จนถึงด้านซ้ายมือสุด ก็จะได้คำรหัสทั้งหมด

- จากทฤษฎีบทการเข้ารหัสแหล่งต้นทาง ประสิทธิภาพรหัส η จะมีค่าเข้าใกล้ 1 ก็ต่อเมื่อใช้เข้ารหัสข้อความที่มีความยาว $n \to \infty$

Example 7

พิจารณารหัสแหล่งกำเนิด X ที่ใช้สัญลักษณ์ {A, B, C, D, E, F} ที่มีความน่าจะเป็น {0.15, 0.25, 0.30, 0.12, 0.08, 0.10} จงหารหัสฮัฟฟ์แมนแบบไบนารีของสัญลักษณ์เหล่านี้ พร้อมทั้งหาประสิทธิภาพรหัส และตรวจสอบว่าเป็นรหัสกระชับหรือไม่

ข้อความ	ความน่าจะเป็น	คำรหัส
C	0.30	00
B	0.25	10
A	0.15	010
D	0.12	011
F	0.10	110
E	0.08	111



วิธีทำ

เริ่มต้นให้นำสัญลักษณ์ทั้งหมดมาเรียงลำดับตามค่าความน่าจะเป็นของแต่ละสัญลักษณ์จากมากไปหาน้อย และทำการหารหัสฮัฟฟ์แมน ก็จะได้คำรหัสตามรูปหน้า 46

เนื่องจากความยาวเฉลี่ยของคำรหัสหาได้จาก

$$\bar{L} = 0.3\times 2 + 0.25\times 2 + 0.15\times 3 + 0.12\times 3 + 0.1\times 3 + 0.08\times 3 = 2.45 \quad \text{(บิต)}$$

และเอนโทรปีของแหล่งต้นทางมีค่าเท่ากับ

$$H(X) = \sum_{i=1}^{6} p(x_i)\log_2\left(\frac{1}{p(x_i)}\right) = 2.4224 \quad \text{(บิต)}$$

ดังนั้นประสิทธิภาพรหัสมีค่าเท่ากับ $\eta = 2.418/2.45 = 0.9887$

เนื่องจากรหัสกระชับจะต้องทำให้ $\sum_{i=1}^{N} D^{-l_i} = 1$ ซึ่งรหัสฮัฟฟ์แมนแบบไบนารี $(D=2)$

ในที่นี้มีค่า $\sum_{i=1}^{N} D^{-l_i} = 2^{-2} + 2^{-2} + 2^{-3} + 2^{-3} + 2^{-3} + 2^{-3} = 1 \Rightarrow$ **รหัสกระชับ**

- จากทฤษฎีบทการเข้ารหัสแหล่งต้นทาง ประสิทธิภาพรหัส η จะมีค่าเข้าใกล้ 1 ก็ต่อเมื่อใช้เข้ารหัสข้อความที่มีความยาว $n \to \infty$

ข้อเสียของรหัสฮัฟฟ์แมน

- วงจรเข้ารหัสฮัฟฟ์แมนต้องทราบความน่าจะเป็นของแต่ละสัญลักษณ์
- คำรหัสแต่ละคำมีความยาวไม่เท่ากัน ซึ่งจะส่งผลเสียต่อการทำงานแบบเรียลไทม์ เพราะอัตราสัญลักษณ์ของข้อมูลอินพุตมีค่าแตกต่างจากอัตราสัญลักษณ์ของข้อมูลเอาต์พุต

- แก้ปัญหาได้โดย $\Rightarrow$ รหัส Ziv-Lempel

รหัส Ziv-Lempel

- รหัส Ziv-Lempel หรืออัลกอริทึม Lempel-Ziv (LZA) ได้ถูกพัฒนาโดย Jacob Ziv และ Adam Lempel ในปี ค.ศ. 1977
- คุณสมบัติที่สำคัญ

1) วงจรเข้ารหัส Ziv-Lempel ไม่จำเป็นต้องทราบความน่าจะเป็นของแต่ละสัญลักษณ์
2) ทุกคำรหัสที่ได้จากการเข้ารหัสข้อความด้วยอัลกอริทึม Lempel-Ziv จะมีความยาวเท่ากัน
3) รหัส Ziv-Lempel ถือว่าเป็นรหัสเหมาะที่สุด (optimal) เมื่อใช้เข้ารหัสข้อความที่ประกอบด้วย $n \gg 1$ สัญลักษณ์ นั่นคือความยาวเฉลี่ยของคำรหัสจะมีค่าเท่ากับ $H(X)$ บิตต่อสัญลักษณ์ ก็ต่อเมื่อ $n \to \infty$

- เนื่องจากคำรหัสตัวที่ k หรือ c_k ที่ได้จากอัลกอริทึม Lempel-Ziv ประกอบด้วยข้อมูลสองส่วน คือ ส่วนแรกคือเลขดัชนี i_k และส่วนที่สองคือสัญลักษณ์ d_k นั่นคือ $c_k = [i_k, d_k]$ โดยข้อมูลทั้งสองส่วนนี้จะมีความยาวคงที่ และเป็นที่ทราบกันระหว่างวงจรภาคส่งและวงจรภาครับก่อนเริ่มกระบวนการส่งผ่านข้อมูล
- เลขดัชนี i_k จะสัมพันธ์กับส่วนเติมหน้า $p(i_k)$ ที่เก็บอยู่ใน**หนังสือประมวลรหัสคำ** (codebook) ซึ่งจะถูกสร้างขึ้นมาทันทีที่วงจรภาคส่งและวงจรภาครับในระหว่างการส่งข้อมูล
- คำรหัส c_k จะสอดคล้องกับสายอักขระที่เป็นได้อย่างเดียวคือ $s_k = p(i_k) + d_k$ โดยสายอักขระ $p(i_k)$ ได้มาจากหนังสือประมวลรหัสคำซึ่งเก็บสายอักขระที่ถูกส่งไปก่อนหน้านี้ และ d_k คือสัญลักษณ์ตัวแรกของ**สายอักขระที่คงเหลือ**
- สายอักขระ s_k นี้ก็จะทำหน้าที่เป็นส่วนเติมหน้าตัวถัดไปที่บันทึกอยู่ในหนังสือประมวลรหัสคำของวงจรภาคส่งและวงจรภาครับ

กระบวนการเข้ารหัส

1) เริ่มต้นวงจรเข้ารหัสจะกำหนดดัชนี 0 ให้กับสายอักขระว่าง (null string) นั่นคือ " " และกำหนดดัชนี 1 ให้กับสัญลักษณ์ตัวแรกของ S

2) ณ การเข้ารหัสครั้งที่ k วงจรเข้ารหัสจะดูส่วนที่ยาวสุดหรือส่วนเติมหน้า $p(i_k)$ ที่มีอยู่ในหนังสือประมวลรหัสคำ ว่าสัมพันธ์กับส่วนแรกของสายอักขระที่คงเหลือหรือไม่

   2.1) ถ้าสัมพันธ์กัน วงจรเข้ารหัสก็จะให้ข้อมูลเอาต์พุตเป็นคำรหัส $c_k = [i_k, d_k]$ ซึ่งประกอบด้วยดัชนี i_k และสัญลักษณ์ตัวถัดไป d_k ที่อยู่ตามหลัง $p(i_k)$ ในสายอักขระที่คงเหลือ

   2.2) ถ้า**ไม่**สัมพันธ์กัน วงจรเข้ารหัสก็จะให้ข้อมูลเอาต์พุตเป็นคำรหัส $c_k = [0, d_k]$ ซึ่งกรณีนี้จะเกิดขึ้นเพียงครั้งเดียวในกระบวนการเข้ารหัส เมื่อสัญลักษณ์ตัวแรกของสายอักขระที่เหลือไม่ตรงกับสัญลักษณ์ที่กำหนดในเลขดัชนี 1

3) วงจรเข้ารหัสจะบันทึกสายอักขระ $s_k = p(i_k) + d_k$ ลงในหนังสือประมวลรหัสคำ เพื่อทำหน้าที่เป็นส่วนเติมหน้าตัวถัดไป

4) จากนั้นคำรหัสทั้งหมด (แต่ละคำรหัสจะมีจำนวนบิตเท่ากัน) ที่ได้จะถูกส่งไปยังวงจรภาครับ

กระบวนการถอดรหัส

- โดยทั่วไปก่อนเริ่มต้นใช้งานอัลกอริทึม Lempel-Ziv วงจรเข้ารหัสและวงจรถอดรหัสจะต้องตกลงกันเพื่อกำหนดจำนวนบิตให้เพียงพอสำหรับใช้แทนดัชนี $\{i_k\}$ ในคำรหัส ซึ่งจะช่วยในการถอดรหัสข้อมูลที่วงจรถอดรหัส
- พิจารณาการถอดรหัสครั้งที่ k วงจรถอดรหัสจะรับคำรหัสที่ k แล้วแบ่งออกเป็นสองส่วนคือ ส่วนแรกเป็นเลขดัชนี i_k และส่วนที่สองเป็นสัญลักษณ์ d_k
- จากนั้นจัดให้อยู่ในรูปของสายอักขระ $s_k = p(i_k) + d_k$ และบันทึกลงในหนังสือประมวลรหัสคำภายใต้เลขดัชนีใหม่ (ข้อมูล s_k ทั้งหมดก็คือข้อมูลที่ถอดรหัสได้นั่นเอง)

Example 8

การเข้ารหัสสายอักขระ S = [110000100001100011101] ทำได้ดังนี้

1) หนังสือประมวลรหัสคำอยู่ในสถานะว่าง $s_0 =$ " " (สายอักขระว่าง)

2) ส่วนแรกของ S คือ $"1" = s_0 + "1"$ ดังนั้น $s_1 = "1"$ และ $d_1 = 1$

3) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "10" ดังนั้น $s_2 = s_1 + "0"$ และ $d_2 = 0$

4) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "0" ดังนั้น $s_3 = s_0 + "0"$ และ $d_3 = 0$

5) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "00" ดังนั้น $s_4 = s_3 + "0"$ และ $d_4 = 0$

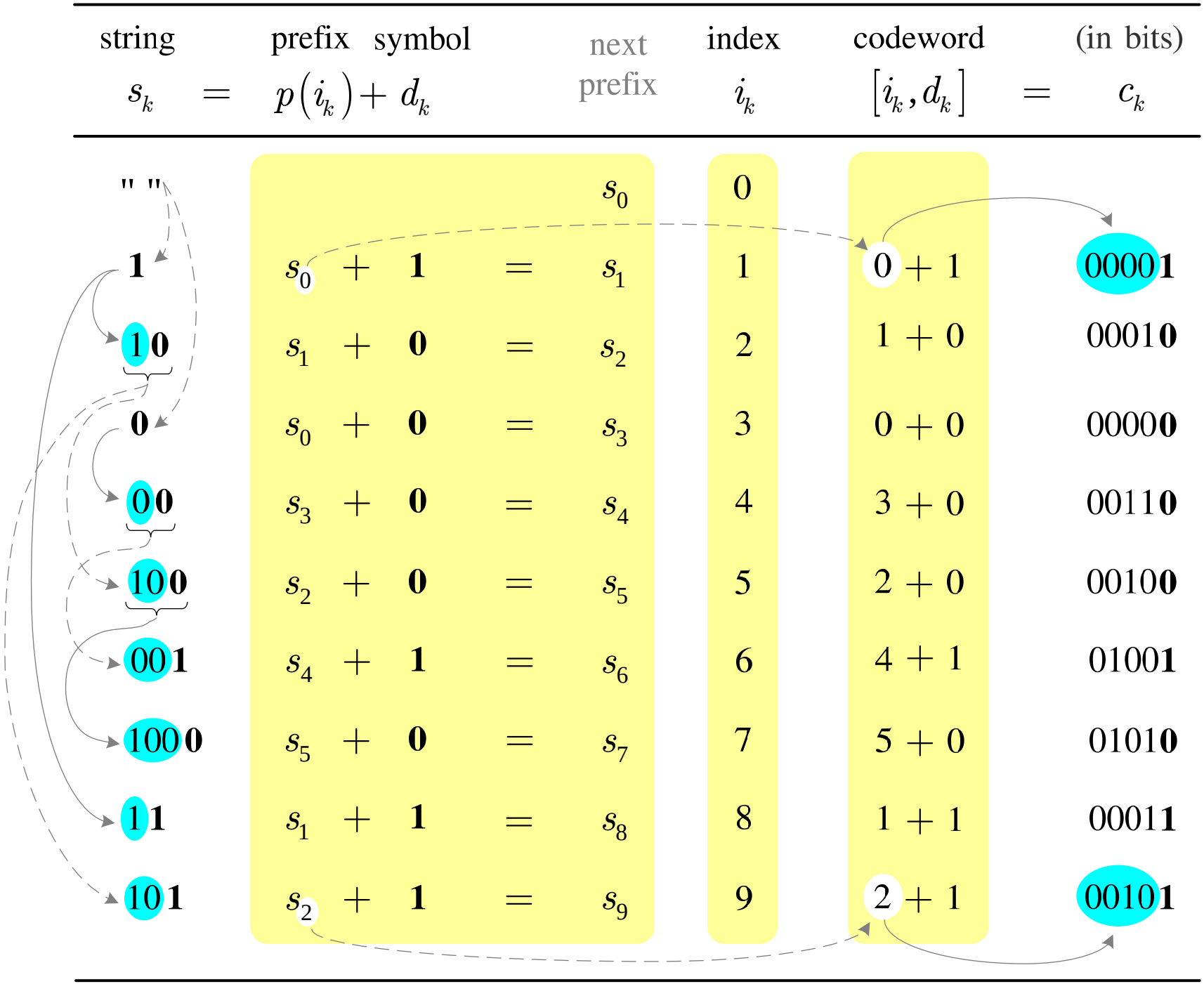
6) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "100" ดังนั้น $s_5 = s_2 + "0"$ และ $d_5 = 0$

7) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "001" ดังนั้น $s_6 = s_4 + "1"$ และ $d_6 = 1$

8) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "1000" ดังนั้น $s_7 = s_5 + "0"$ และ $d_7 = 0$

9) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "11" ดังนั้น $s_8 = s_1 + "1"$ และ $d_8 = 1$

10) ส่วนแรกของสายอักขระ S ที่เหลือ คือ "101" ดังนั้น $s_9 = s_2 + "1"$ และ $d_9 = 1$


string s_k = prefix $p(i_k)$ + symbol d_k
next prefix
index i_k
codeword $[i_k, d_k]$ = (in bits) c_k
" "
1
10
0
00
100
001
1000
11
101
s_0 0
$s_0 + 1 = s_1$ 1 $0 + 1$ 00001
$s_1 + 0 = s_2$ 2 $1 + 0$ 00010
$s_0 + 0 = s_3$ 3 $0 + 0$ 00000
$s_3 + 0 = s_4$ 4 $3 + 0$ 00110
$s_2 + 0 = s_5$ 5 $2 + 0$ 00100
$s_4 + 1 = s_6$ 6 $4 + 1$ 01001
$s_5 + 0 = s_7$ 7 $5 + 0$ 01010
$s_1 + 1 = s_8$ 8 $1 + 1$ 00011
$s_2 + 1 = s_9$ 9 $2 + 1$ 00101

- ตัวอย่างนี้ต้องใช้ข้อมูลจำนวน 4 บิตสำหรับแทนส่วนที่เป็นเลขดัชนี ($10 < 2^4$)
- ดังนั้นแต่ละคำรหัสจะประกอบด้วยข้อมูล 5 บิต (4 บิตสำหรับแทนเลขดัชนี i_k และ 1 บิตสำหรับสัญลักษณ์ d_k) $\Rightarrow$ วงจรเข้ารหัสจะให้ข้อมูลเอาต์พุตทั้งหมด 50 บิต จากการเข้ารหัสข้อมูลอินพุตจำนวน 21 บิต
- จะเห็นว่าอัลกอริทึม Lempel-Ziv ไม่เหมาะสำหรับนำมาใช้บีบอัดข้อความที่มีความยาวน้อยๆ ซึ่งสอดคล้องกับคุณสมบัติข้อที่ 3 ของรหัส Ziv-Lempel ที่กล่าวว่าอัลกอริทึม Lempel-Ziv จะมีประสิทธิภาพรหัสมากสุด ($\eta \rightarrow 1$) ก็ต่อเมื่อใช้เข้ารหัสข้อความที่มีความยาวมากๆ

*** END