example

... सूचना = $\log_2(4) = 2$ बिट्स

यदि दो स्वतन्त्र मैसेज m_1 तथा m_2 हों तो यह सिद्ध कीजिए उनकी कुल सूचना का योग उनकी अलग-अलग सूचनाओं के योग के बराबर होता है।

माना मैसेज m_1 तथा m_2 के घटित होने की प्रायिकता क्रमशः p_1 तथा p_2 है।

यदि उनकी अलग-अलग सूचना का मान लिखा जाये

$$I_1 = \log_2\left(\frac{1}{P_1}\right)$$

$$I_2 = \log_2\left(\frac{1}{P_2}\right)$$

यह दोनों मैसेज सिगनल पास्परिक रूप से स्वतन्त्र हैं। उनके एकसाथ घटित होने की प्रायिकता

$$P = P_1 \times P_2$$

पारस्परिक सूचना एक-दूसरे के प्रति

$$I_{1,2} = \log_2\left[\frac{1}{P}\right] = \log_2\left[\frac{1}{P_1 P_2}\right]$$

$$I_{1,2} = \log_2\left[\frac{1}{P_1}\right] + \log_2\left[\frac{1}{P_2}\right]$$

चूँकि

$$\log(mn) = \log m + \log n$$

$$I_{1,2} = I_1 + I_2$$

उदाहरण -माना कोई ट्रांसमीटर हमेशा केवल एक मैसेज m_k को ट्रांसमिट करता है तो सिद्ध कीजिए कि उस मैसेज से प्राप्त सूच्ना शून्य होती है।

हल—क्योंकि केवल एक मैसेज को ट्रांसमिट किया जाना है तो इसके घटित होने की प्रायिकता अधिकतम 1 हो सकती है।

$$P_k = 1$$

इस मान को सूचना में रखने पर

$$I_k = \log_2\left[\frac{1}{1}\right]$$

$$= \frac{\log_{10} 1}{\log_{10} 2}$$

$$\boxed{I_K = 0 \text{ बिट}}$$

अतः किसी सुनिश्चित घटना की सूचना 0 होती है।

औसत सूचना (इनफॉर्मेशन) या एन्ट्रोपी

एक मैसेज से प्राप्त "औसत सूचना" को एन्ट्रोपी से परिभाषित किया जाता है। इसको H से दर्शाया जाता है इसका मात्रक बिट/मैसेज होता है।

यदि आपका उद्देश्य अधिकतम सूचना को ट्रांसफर करना है तब एन्ट्रोपी सम्भवत: अधिक से अधिक होनी चाहिए। ऐन्ट्रोपी को ज्ञात करने के लिए निम्न पदों का प्रयोग किया जाता है—

पद 1—माना M विभिन्न मैसेजस् $m_1, m_2, m_3 ....$ हैं तथा उनके होने की प्रायिकता $p_1, p_2 ..... p_n$ है।

पद 2—यदि कुल मैसेज की संख्या L हो तो m_1 मैसेज की संख्या $P_1 L$ तथा m_2 मैसेज की संख्या $P_2 L$ होगी।

पद 3—m_1 मैसेज की सूचना को ज्ञात कीजिए—

$$m_1 \text{ as } I_1 = \log_2 \left[\frac{1}{P_1}\right]$$

पद 4—m_1 में होने वाली कुल सूचना को ज्ञात कीजिए—

$$I_{1(\text{total})} = p_1 L \times I_1$$

पद 5—इसी प्रकार अन्य मैसेजों की भी कुल सूचनाओं को ज्ञात कीजिए—

$$I_{2(\text{total})} \; I_{3(\text{total})} \; ..... \text{ इत्यादि}$$

पद 6—इन सभी सूचनाओं को कुल सूचना के लिए योग कीजिए।

$$I_{\text{total}} = I_{1(\text{total})} + I_{2(\text{total})} + ...$$

पद 7—ऐन्ट्रोपी का मान प्राप्त करने के लिए पद 6 से प्राप्त कुल सूचना के मान को कुल मैसेज की संख्या L से भाग दीजिए।

माना एक ट्रांसमीटर M विभिन्न मैसेज को ट्रांसमिट कर रहा है तथा स्वतन्त्र मैसेज m_1, m_2, m_3 तथा माना इनकी प्रायिकताऐं क्रमश: p_1, p_2, p_3 हैं।

माना ट्रांसमिशन के समय L मैसेज जनरेट होते हैं।

1. यदि L बहुत-बहुत बड़ी हो तो L मैसेज के लिए m_1 के $p_1 L$ मैसेज ट्रांसमिट किए जाते हैं।
   m_2 के $P_2 L$ मैसेज ट्रांसमिट किए जाते हैं।
   m_3 के $P_3 L$ मैसेज ट्रांसमिट किए जाते हैं।
   m_n के $p_n L$ मैसेज ट्रांसमिट किए जाते हैं।
2. m_1 में ज्ञात की गयी सूचना

$$I_1 = \log_2 \left(\frac{1}{P_1}\right)$$

   यदि m_1 में मैसेज की संख्या $P_1 L$ मैसेज है, तब $P_1 L$ मैसेज की सूचना $I_{1(\text{total})} = P_1 L \log_2 \left[\frac{1}{P_1}\right]$

   इसी प्रकार $P_2 L$ मैसेज में m_2 मैसेज की संख्या द्वारा रखी गयी सूचना $I_{2(\text{total})} = P_2 L \log_2 \left[\frac{1}{P_2}\right]$

   इसी प्रकार बचे हुए मैसेज के लिए भी सूचना ज्ञात की जा सकती है।

3. एक या एक से ज्यादा पारस्परिक स्वतन्त्र मैसेज की कुल सूचना उनके द्वारा ग्रहित अलग-अलग सूचनाओं के योग के बराबर होती है।

$$I_{(\text{total})} = I_{1(\text{total})} + I_{2(\text{total})} + I_{3(\text{total})} + ...$$

$$I_{(total)} = P_1 L \log_2\left[\frac{1}{P_1}\right] + P_2 L \log_2\left[\frac{1}{P_2}\right] + P_3 L \log_2\left[\frac{1}{P_3}\right]$$

$$I_{(total)} - L\left[P_1 \log_2\left(\frac{1}{P_1}\right) + P_2 \log_2\left(\frac{1}{P_2}\right) + P_3 \log_2\left(\frac{1}{P_3}\right) + \right]$$

4. ऐन्ट्रोपी की परिभाषानुसार एक मैसेज की औसत सूचना जिसको H से दर्शाया जाता है।

$$H = \frac{I_{total}}{L} = P_1 \log\left(\frac{1}{P_1}\right) + P_2 \log\left(\frac{1}{P_2}\right)$$

$$\boxed{\text{Entropy } H = \sum_{k=1}^{M} P_k \log_2\left(\frac{1}{P_k}\right)}$$

यह ऐन्ट्रोपी को दर्शाता है।

शेनन प्रमेय (Shannon's Theorem)

शेनन प्रमेय के अनुसार "डिजिटल ट्रांसमिशन में एक निश्चित bandwith (B) से एक निश्चित शोर स्तर (noise level) अर्थात् निश्चित signal to noise अनुपात $\left(\frac{S}{N}\right)$ पर बिना त्रुटि (errors) के अधिकतम डाटा बिट दर (data bits rate; C_{max}) एक सीमा से अधिक नहीं हो सकती है। इस अधिकतम डाटा बिट दर (C_{max}) को प्राय: शेनन सीमा (shannon's limit) कहते हैं तथा इसे निम्न सूत्र द्वारा प्रदर्शित करते हैं—

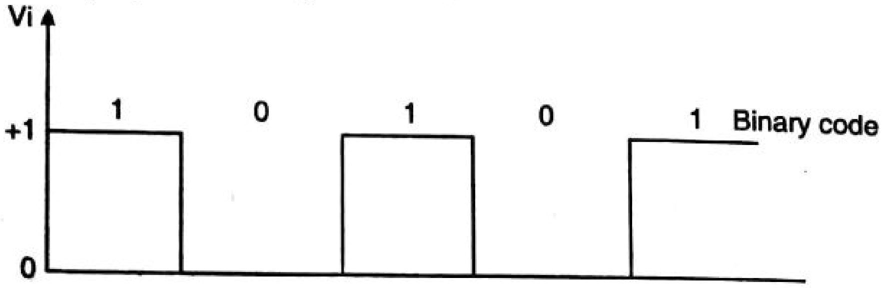
$$C_{max} = B \log_2\left(1 + \frac{S}{N}\right) \quad \ldots(1)$$

प्राय: $\frac{S}{N} >> 1$ होता है, अत: समीकरण (·) को निम्न प्रकार से भी प्रदर्शित कर सकते हैं—

$$C_{max} = B \log_2\left(\frac{S}{N}\right)$$

जहाँ C_{max} = bits per second में सूचना क्षमता (capacity) है।

B = बैण्डविड्थ (Band width) Hz में तथा



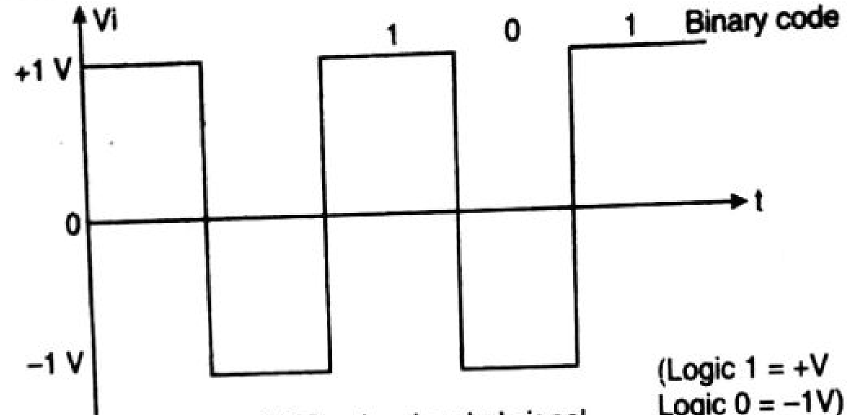
(a) Two level coded signal

(Logic 1 = + 1 V
Logic 0 = 0 V)

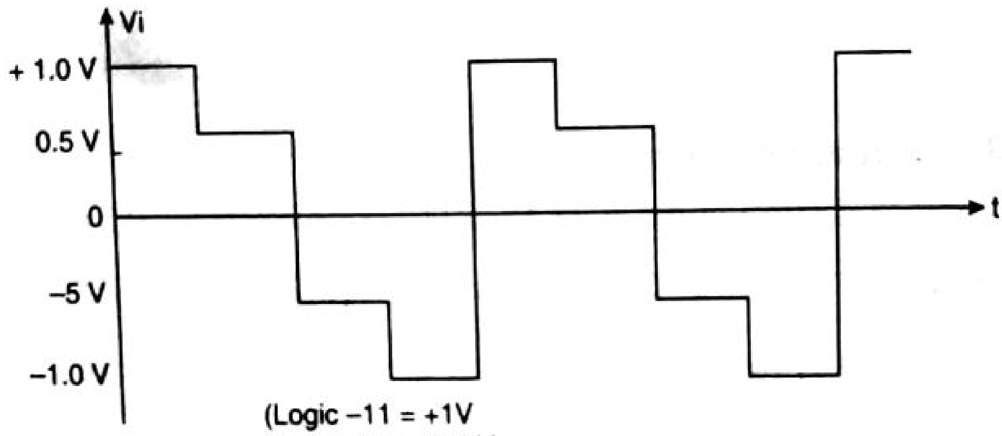
चित्र

$\frac{S}{N}$ = signal to noise ratio (यह पॉवर अनुपात है न कि dB में)

गणना की सरलता के लिए $\log_2 x$ का मान अग्र होता है—



(b) Two level coded signal



(Logic –11 = +1V
Logic –10 = +0.5 V
Logic – 01 = –0.5 V and Logic 00 = –1 V

(c) Four level coded signal

चित्र

$\log_2 x = \frac{\log_{10} x}{\log_{10} 2}$ जहाँ x एक संख्या है।

इसमें डिजिटल सिगनल किसी भी स्तर कोड (2 level code, 4 level code, 8 level code) का हो सकता है। चित्र (a) व (b) में 2 level code तथा चित्र (c) में 4 level कोड को प्रदर्शित किया गया है। यदि शोर (noise) को नगण्य मान लिया जाए तो भी डिजिटल सिगनल के स्तर के अनुसार भी एक निश्चित बैण्डविड्थ (B) में से अधिकतम डाटा बिट ट्रांसफॉर्म करने की दर (rate) एक सीमा (limit) से अधिक नहीं हो सकता है। इसे शेनन हार्टलेय (shanhon hartley) प्रमेय कहते हैं तथा इसके अनुसार,

$$C_{max} = 2B \log_2 P$$

जहाँ C_{max} = Bits per second में सूचना क्षमता (capacity) है।

B = Bandwidth Hz में तथा

P = सिगनल सूचना कोड में स्तरों (levels) की संख्या है।

अत: $P = 2$ अर्थात् 2 level digital coded सिगनल चित्र (a) व (b) होने पर

$$C_{max} 2 = 2B \log_2 2 = 2B \frac{\log_{10} 2}{\log_{10} 2} = 2B \text{ होती है।}$$

इसी प्रकार $P = 4$ अर्थात् 4 level digital coded सिगनल चित्र (c) होने पर

$$C_{max} 4 = 2B \log_2 4 = \frac{2B \log_{10} 4}{\log_{10} 2} = \frac{2B 2 \log_{10} 2}{\log_{10} 2}$$

$$= 4B$$

इससे सिद्ध होता है कि अधिक स्तरों (levels) युक्त Data signal को प्रयुक्त करने से, समान बैण्डविड्थ के चैनल से अधिक डाटा रेट द्वारा सूचना ट्रांसमिट की जा सकती है। परन्तु सिगनल कितने भी स्तरों वाला क्यों न प्रयुक्त किया जाए फिर भी डाटा बिट दर का मान समीकरण द्वारा गणना करने से अधिक नहीं हो सकता है।

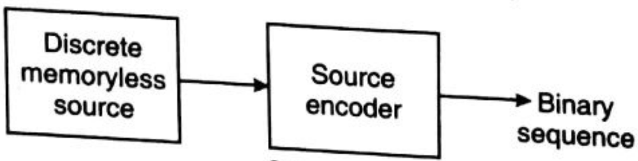
शैनन प्रथम प्रमेय (सोर्स कोडिंग प्रमेय) (Variable Length Code)

कम्यूनिकेशन में एक सबसे प्रमुख समस्या है, डिस्क्रीट (discrete) सोर्स द्वारा उत्पादित डाटा को सक्षम रूप से प्रदर्शित करना। जिस प्रक्रिया द्वारा इसे प्रदर्शित किया जाता है, उसे सोर्स कोडिंग (source coding) कहते हैं और वह युक्ति जो यह कार्य करती है, सोर्स एन्कोडर (source encoder) कहलाती है। Source encoder को बहुत ज्यादा सक्षम बनाने के लिए हमें सोर्स की सांख्यिकी को जानना आवश्यक है। इसका मतलब वह source symbol जिनकी प्रायिकता अधिक होती है, उन्हें "source" कोड वर्ड की सहायता से दर्शाया जाता है जबकि जिन source symbol की प्रायिकता कम होती है, उन्हें longer कोर्ड वर्ड से transmit किया जाता है। इसलिए source code को variable length code (परिवर्तन लम्बाई कोड) भी कहा जाता है।

सक्षम सोर्स एन्कोडर की आवश्यकता (Requirements of an efficient source encoder)

एक सक्षम सोर्स एन्कोडर को निम्न दो आवश्यकताओं को पूर्ण करना चाहिए—

1. एनकोडर द्वारा उत्पादित कोड वर्ड बाइनरी रूप में होने चाहिये।
2. सोर्स एनकोडर यूनिक रूप से डिकोडेबल होना चाहिए जिससे बाइनरी seqence की सहायता से वास्तविक सोर्स Sequence उपर्युक्त रूप से प्राप्त हो सके।



चित्र

एक ब्लॉक डायग्राम चित्र में दिखाया गया है। DMS का आऊटपुट सोर्स Encoder की सहायता से बाइनरी सिगनल में कनवर्ट हो जाता है।

माना सोर्स encoder की औसत कोडवर्ड ल० $(L) =$

$$L = \sum_{k=1}^{M} p_k \times (m_k \text{ बिट में मैसेज की ल०})$$

$$\boxed{L = \sum_{k=1}^{M} p_k \times l_k}$$

यहाँ L सोर्स symbol की औसत बिट्स संख्या को दर्शाता है जो सोर्स encoding के लिए उपयोग में लाया जाता है।

माना $L_{\min}$ L के न्यूनतम मान को दर्शाती है।

अर्थात् सोर्स कोडर की कोडिंग दक्षता (coding efficiency) को इस प्रकार परिभाषित किया जा सकता है—

$$\eta = \frac{L_{\min}}{L}$$

सोर्स encoder अगर 1 के बराबर हो तो बहुत दक्ष कहलायेगा।

$L_{\min}$ को प्राप्त करना

Source coding प्रमेय का कथन

एक discrete memoryless सोर्स जिसकी entropy H है, दिया गया है। अगर किसी सोर्स की औसत कोर्ड वर्ड ल० L हो तो एनकोडिंग के लिए

$$\boxed{L \geq H}$$

औसत कोड वर्ड ल० L हमेशा entropy H से अधिकतम या बराबर होनी चाहिए। यदि सोर्स एन्ट्रोपी H हो तो इसका अर्थ है— L_{min} का मान H के बराबर होना चाहिए।

$$L_{min} = H$$

$$\text{कोड efficiency } \boxed{= \eta = \frac{H}{L}}$$

सोर्स Encoder की दक्षता बढ़ाने के उद्देश्य से variable length coding की जाती है।

इसके लिए दो Algorithms का उपयोग किया जाता है—

1. शेनन फैनों एल्गोरिथ्म (Shanon fano algorithm)
2. हुफमैन कोडिंग (Huffmen coding)

शेनन फैनो कोड (Shanon fano codes)

माना M मैसेज की संख्या है, जो ट्रांसमिट किये जाते हैं। प्रत्येक मैसेज को N बिट वर्ड में कोडिंग किया गया है।

अर्थात् $$M = 2^N$$

एक मैसेज में औसत सूचना (अगर एकजैसे मैसेज हैं।)

$$H = \log_2 M \quad \text{...(i)}$$

$M = 2^N$ रखने पर

$$H = \log_2 2^N = N \text{ बिट मैसेज}$$

$$\frac{H}{N} = 1 \text{ bit}$$

एक मैसेज बिटों की संख्या N है। इसलिए एक बिट में ली गयी औसत सूचना $\frac{H}{N} = 1$ Bit, यह एक बिट में अधिकतम औसत सूचना है।

समस्या—

जब मैसेज M एक जैसे ना हों तब समस्या खड़ी होती है।

तब $$H < N \quad \text{or} \quad \left(\frac{H}{N}\right) < 1$$

तब हर बाइनिट (binet) में 1 बिट से कम औसत सूचना होती है तथा shanon fano algorithm का उपयोग किया जाता है।

Shanon fano algorithm (शैनन फैनो algorithm)

प्रत्येक मैसेज के लिए वर्ड प्राप्त करने के लिए Procedure निम्नलिखित है—

पद 1—सभी source symbols को उनकी घटती प्रायिकता के क्रम में व्यवस्थित कीजिये।

पद 2—Symbols के सैट को दो पदों में विखण्डित (partition) कीजिए जो एकल प्रायिकता के आसपास हो।

पद 3—प्रत्येक मैसेज के upper set को 1 assign करें तथा lower set को 0 Assign करें।

पद 4—इस प्रक्रिया को तब तक लगातार करें जब तक आगे partition सम्भव न हो।

हर बार sets को दो equal partition के सेट में विखण्डित करें।

उदाहरण. —एक Discrete memoryless सोर्स के पाँच symbols x_1, x_2, x_3, x_4 तथा x_5 हैं जिनकी प्रायिकता $P(x_1) = 0{\cdot}4, P(x_2) = 0{\cdot}19, P(x_3) = 0{\cdot}16$, $P(x_4) = 0{\cdot}15, P(x_5) = 0{\cdot}1$ है। इसके लिए shanon fano codes बनाये तथा कोड दक्षता ज्ञात कीजिए।

Shanon fano code के लिए सारणी

मैसेज	प्रायिकताएँ	I	II	III		code word	बिट्स पर कोड वर्ड
x_1	0·4	0 Partition	Stop			0	1
x_2	0·19	1	0	0 Partition	Stop	100	3
x_3	0·16	1	0 Partition	1	Stop	101	3
x_4	0·15	1	1	0 Partition	Stop	110	3
x_5	0·1	1	1	1	Stop	111	3

एक मैसेज में औसत सूचना (H)

$$H = \sum_{i=1}^{5} P_{(x_i)} \log_2\left(\frac{1}{P(x_i)}\right)$$

$$= 0{\cdot}4\log_2\left(\frac{1}{0{\cdot}4}\right) + 0{\cdot}19\log_2\left(\frac{1}{0{\cdot}19}\right) + 0{\cdot}16\log_2\left(\frac{1}{0{\cdot}16}\right)$$

$$+ 0{\cdot}15\log_2\left(\frac{1}{0{\cdot}15}\right) + 0{\cdot}1\log_2\left(\frac{1}{0{\cdot}1}\right)$$

$$\mathbf{H = 2{\cdot}15} \textbf{ बिट/मैसेज}$$

3. औसत कोड वर्ड लम्बाई (L)

$$L = \sum_{k=1}^{5} p_k (mk \text{ बिट में लम्बाई})$$

$$= (0{\cdot}4\times 1) + (0{\cdot}19\times 3) + (0{\cdot}16\times 3) + (0{\cdot}15\times 3) + (0{\cdot}1\times 3)$$

$$L = 0{\cdot}4 + 0{\cdot}57 + 0{\cdot}48 + 0{\cdot}45 + 0{\cdot}3$$

$$\mathbf{L = 2{\cdot}2} \textbf{ बिट/मैसेज}$$

3. कोड दक्षता $\eta = \frac{H}{L}\times 100 = \frac{2{\cdot}15}{2{\cdot}2}\times 100 = \mathbf{97{\cdot}22\%}$