2.Teorie informace

2.1 Vznik a vývoj teorie informace

Revoluční zvrat v potřebě zkoumat vědecky informaci a její přenos způsobilo využití elektřiny. Postupně byl objeven drátový telegraf, telefon, rádiové, družicové spojení atd. Dříve než se vývoj poznávání informace a vývoj samotných prostředků pro její přenos dostal na současnou úroveň, bylo třeba vyřešit hodně teoretických a praktických otázek. ty se zpočátku týkaly více problému přenosu informace, než problému, co to vlastně informace je. Už S.Morse přišel k závěru, že rychlost přenosu je možné zvýšit využíváním vhodné soustavy kódování (písmenům, které se v textu vyskytují nejčastěji přiřadit nejkratší kódové složky, tj. vytvoření nerovnoměrného kódu). Řešení problému kódování se již dnes neobejde bez využití matematiky.
Podstatný zvrat v chápání informace zaznamenala publikace R.V.L.Hartleyho „Přenos informací“ v roce 1928. V Hartleyho chápání odesílatel zprávy disponuje soustavou symbolů, z kterých vytváří posloupnosti. Užitečnost takového chápání vyplývá z jeho všeobecnosti: podobnou posloupnost může vytvářet každý náhodný jev.
Zevšeobecnění myšlenek Hartleyho pomocí matematické statistiky a teorie pravděpodobnosti Shannonem, Wienerem a Fisherem koncem 40.let tohoto století se pokládá za vznik kvantitativní teorie informace. Myšlenku vztahu pravděpodobnosti a informace vyslovil jako první R.Fisher, který pracoval v oblasti matematické statistiky. Zakladatelé kvantitativní teorie informace se zabývali problémem množství informace, ale ne co je to vlastně informace. Naznačili však, že je třeba hledat odpověď v obsahových a kvalitativních projevech informace. V této oblasti vedle anglického vědce W.R.Ashbyho udělali kus práce i sovětští vědci počínaje rokem 1958. O vznik a rozvoj kvantitativní teorie informace se velkou měrou přičinil zejména V.Kotelnikov, A.Kolmogorov, A.Chinčin a A.Jaglomov.

2.2 Základní pojmy teorie informace

Odmyslíme-li konkrétní obsah jevu nebo události, o němž se vydávají informace, zbude nám jen nějaká množina vzájemně odlišných stavů. Každý objekt může být v každém časovém okamžiku v jednom z mnoha možných stavů. Různé jevy se vzájemně liší počtem možných stavů a zvláštnostmi jejich výběru (pravděpodobností výskytu jednotlivých stavů).
Teorie informace tvoří část obecné kybernetiky. Slovo informace pochází z latiny, kde znamená „uvádět ve tvar“. V obecném pojetí je velikost informace dána uspořádaností (neurčitostí) soustavy prvků (systému). Informace o systému je tím větší, čím je pravděpodobnost výskytu jednotlivých jeho stavů menší. Informace je větší, obsahuje-li zpráva něco nového, co předtím nebylo známo, nebo co nelze snadno uhodnout (je málo pravděpodobné). Pojem informace je velmi široký a těžko se definuje. Přesná definice pojmu informace neexistuje, pouze lze určit, je-li informace větší či menší, tj.míru informace. Velmi často se tento pojem používá volně v nejasném intuitivně chápaném smyslu ve vztahu k pojmům, poznatek, novinka, údaj a pod. Informace má nehmotný charakter, neboť vznikla abstrakcí, ale je spojena vždy s nějakým fyzikálním pochodem, který ji nese (signálem).
Aby mohla být informace předána, musí být nějakým způsobem zakódována, tj.převedena do vhodných symbolů nebo signálů. Signál je fyzikální veličina, která nese informaci. Zpráva je způsob vyjádření informace posloupností symbolů (znaků). Zpráva může být tedy např. číslo (posloupnost číslic), text (posloupnost písmen), ale i řídící signál (posloupnost napěťových úrovní) apod.
Každá zpráva má syntaxi (skladbu), sémantiku (obsah) a důležitost. Říkáme, že zpráva má syntaktický, sémantický a pragmatický obsah.

Syntaktický obsah vyjadřuje kvantitativní míru informace v dále definovaných jednotkách zvaných bit. Informační obsah vyjadřuje míru „novosti“, tj. míru neurčitosti objektu (entropii). Počítá se podle vztahu: $I=H-\log_2 p_i = \log_2(1/p_i)$. Ve vzorci znamená H entropii systému, I je velikost informace o tom, že je systém v nějakém stavu x_i, který se vyskytuje s pravděpodobností p_i. Přitom příjemce informace musí předem znát, jaké zprávy mohou být produkovány (musí znát množinu všech možných zpráv, aby byl schopen pochopit jejich význam) a zdroji informace (vysílači) je ponechána volnost výběru z této množiny možných zpráv. U příjemce existuje tedy neurčitost o výběru a vyslání některé zprávy z množiny všech možných. Přijetím zprávy může být tato neurčitost odstraněna. Z hlediska příjemce produkce zpráv ze zdroje (vysílače) náhodným procesem. Velikost informace tedy můžeme měřit tím, kolik neurčitosti bylo dešifrováním zprávy odstraněno. Tatáž zpráva může mít různou velikost informace, neboť to záleží již na předchozích znalostech příjemce. Např. zpráva „Karel zemřel“ má zcela rozdílné velikosti informace v těchto případech:
a)Karel byl vážně nemocen a jeho smrt se dala očekávat. Zprávou byla odstraněna jen malá neurčitost.
b)Karel byl zcela zdráv. Zpráva nese informaci o stavu, který byl málo pravděpodobný; byla tedy odstraněna velká neurčitost, zpráva má velký informační obsah
c)Zprávu o smrti Karla slyšíme již podruhé od jiné osoby. Zpráva pak nenese žádnou informaci, neboť již byla odstraněna veškerá neurčitost, zprávou se nedovíme nic nového.

Sémantický obsah vyjadřuje významovou stránku zprávy a nedá se měřit (sémantika-nauka o významu slov). Zprávu se stejnou velikostí informace můžeme napsat v několika různých jazycích. Sémantická stránka zprávy říká, čeho se informace týká (např. „Karel je mrtev“ nebo „Karl ist tot“). Sémantika vyjadřuje kvalitativní stránku zprávy.

Pragmatický obsah zprávy určuje významnost (důležitost, užitečnost, cennost) sdělení a prioritu jednotlivých zpráv pro příjemce. Jinou důležitost má zpráva „Karel zemřel“ pro rodiče Karla a jinou pro osobu, která Karla neznala. Slovo pragmatický má význam „dbající všech vnějších souvislostí“.

Syntaktický obsah však může existovat i sám bez sémantického a pragmatického. Zpráva nabývá sémantický obsah ve vztahu k objektu a pragmatický obsah ve vztahu k příjemci informace. Syntaxe se týká vzájemné uspořádanosti znaků jako nositelů informace a svým charakterem v klasickém smyslu vývoje teorie informace patří pod kvantitativní stránku informace. Sémantický a pragmatický obsah tvoří dohromady kvalitativní stránku informace.

Rozvoj teorie informace a její využívání nastoluje množství otázek filozofického a metodologického charakteru. Důležitým filozofickým problémem zůstává vymezení pojmu informace a vysvětlení jejího obsahu. I v současnosti se hledají nejvšeobecnější a nejpodstatnější hlediska, která tvoří základ vymezení obsahu pojmu informace. Informace je odrazem objektu na přenosovém či záznamovém médiu (popis nějakých vlastností, stavu nebo situace objektu).

Informace v podobě odrazu vzájemného působení (interakce) objektů, jevů a procesů je především výsledkem pravděpodobnostního charakteru těchto interakcí. Pohyb světla i jeho existence je také výsledkem pravděpodobnostního charakteru těchto interakcí. Pravděpodobností charakter pohybu hmoty formuluje i neurčitý charakter odrazu jako předpoklad vzniku informace. Informační působení a tedy i samotná informace se chápe jako samostatná podstata, schopná spolu s hmotou a energií existovat (působit) při interakci hmotných objektů. Známá je tato interpretace: hmotu představuje masa, energie představuje pohyb masy a informace představuje uspořádanost (resp. neuspořádanost, neurčitost) pohybu

masy. Naznačený přístup při zkoumání informačního působení umožňuje na jedné straně tvrzení o nemateriálnosti informace, na druhé straně tvrzení o její materiálnosti, které však samo o sobě není materií, ale pouze její specifickou vlastností.

Informační působení, tj. signálové působení s informačním obsahem je takové, které v objektu odrazu (u příjemce) zesiluje, stupňuje. Zesílení charakteru signálu je možné proto, že je působením na příjemce informace, který je schopen změny (přeměny) a disponuje volnou energií. V pojmu informace (signálu s informačním obsahem) odhalila kybernetika materiální podstaty procesu řízení. Ta spočívá v nesrovnatelném účinku signálu oproti jeho vlastní energii. Na samotné řízení je třeba menší množství energie než je to, které se v systému řídí. Kdyby tomu tak nebylo, nemohlo by ani samotné řízení existovat. Z hlediska účinnosti a efektivnosti řízení jsou nejzajímavější takové signály, které vyvolávají přeměnu velkých energií. Např. slabé slabé stlačení tlačítka střelmistrem v kamenolomu, vyvolá pohyb celého skalního masívu, přestože energie potřebná pro stlačení tlačítka je nesrovnatelně menší než jakou umožní uvolnit.

Informace je abstraktní pojem. Táž informace může být různě uložena. Neobsahuje žádnou energii a je tudíž i různě přenášena. Tento fakt není na první pohled patrný, ale např. tutéž informaci lze získat ve formě dopisu, telegramu, telefonátu, poslechu televize apod., je jasné, že informace nezávisí na druhu přenosu, který je používán jejím nositelem. Energie je nutná pouze k transportu (přenosu) informace. na druhu a množství energie samozřejmě závisí kvalita přenosu informace. Důležitou skutečností je rozdíl mezi větvením toku energie (nositele informace) a toku informace.

Při kopírování informace, tj. při rozdělování (dělení) toku informace nedochází ke zmenšení velikosti informace. Dochází však ke zmenšení energie, která je jejím nositelem. Např. z hlediska velikosti informace není rozhodující, zda je televizní zpráva sdělována jednomu nebo miliónu televizních diváků. Při vyšetřování toku informace v systému obvykle sledujeme tok energie nesoucí tuto informaci. Musíme si dát pozor na to, aby nedošlo k záměně směru toku energie, popřípadě k záměně energií, neboť tok informační energie může být různý od hlavního energetického toku soustavy.

Informace je v určité zprávě obsažená jen tehdy, jestliže u přijímacího subjektu odstraňuje neurčitost. To umožňuje chápat informaci jako něco nové, co momentálně nebo v budoucnosti ovlivní v nějaké formě konání příjemce.

Pro hodnocení vlastností přenosových systémů z hlediska přenosu informace je nutné zavést vhodnou kvantitativní míru pro vyjádření množství informace. Jednotlivé přenosové systémy lze pak navzájem porovnat z hlediska množství přenášené informace. Pro vyjádření množství informace v nějaké zprávě musíme nalézt nějakou vhodnou společnou měřitelnou vlastnost, která nezávisí na smyslovém obsahu zprávy, užitečnost pro příjemce, ani na fyzikální podstatě vyjádření zprávy.

Množství informace obsažené ve zprávě X souvisí s pravděpodobností jejího výskytu tak, že množství informace I(X) je přímo úměrné pravděpodobnosti P(X) s jakou může příjemce uhodnout obsah zprávy X neboli jaká je pravděpodobnost výskytu dané zprávy u příjemce před jejím přijetím. Platí:

$$\mathrm{I(X)} = f\left[\frac{1}{\mathrm{P(X)}}\right] \tag{2.1}$$

Např. zpráva „V ruletě padlo číslo 17“ přináší větší množství informace než zprávy „V ruletě padlo liché číslo“. Pravděpodobnost padnutí lichého čísla je 0,5, číž příjemce mohl tento stav snadněji uhodnout, ne padnutí čísla 17, neboť tento stav (jev) má mnohem menší pravděpodobnost. První zpráva přináší tedy větší množství informace.

Příjemce zprávy musí předem znát, jaké zprávy mohou být produkovány (musí znát množinu všech možných zpráv) a zdroji zpráv je ponechána volnost výběru z této množiny možných zpráv. U příjemce existuje tedy neurčitost o tom, kterou zprávu obdrží. Přijetím zprávy je pak tato neurčitost odstraněna. Např. přijetím zprávy o padnutí čísla 27 (v ruletě) se odstraní velká neurčitost, kdežto zprávou o padnutí hlavy při hodu korunou jen malá neurčitost. Je tedy množství informace zprávy I(X) rovno míře neurčitosti zprávy H(X) (míra neurčitosti = entropie), kterou tato zpráva odstraní:

$$\mathrm{I(X)} = \mathrm{H(X)} \quad (\mathrm{I_X} = \mathrm{H_X}) \tag{2.2}$$

Množství informace je rovno entropii daného jevu za předpokladu, že neurčitost je po přijetí plně odstraněna.

Nechť zpráva X se skládá ze dvou nezávislých výskytů zpráv A, B (obecně z n nezávislých zpráv). Celková informace I(X), kterou získá příjemce je rovna součtu jednotlivých informací I(A) a I(B):

$$\mathrm{I(X)} = \mathrm{I(A)} + \mathrm{I(B)} \tag{2.3}$$

přičemž pravděpodobnost zprávy X dvou nezávislých jevů je rovna:

$$\mathrm{P(X)} = \mathrm{P(A)} + \mathrm{P(B)} \tag{2.4}$$

Po dosazení vztahu (2.4) do (2.1) a po rozepsání rovnice (2.3) dostaneme:

$$\mathrm{I(X)} = f\left[\frac{1}{\mathrm{P(A)}}\right] + f\left[\frac{1}{\mathrm{P(B)}}\right] + f\left[\frac{1}{\mathrm{P(A)}\cdot\mathrm{P(B)}}\right] \tag{2.5}$$

Nyní zbývá najít vhodnou funkci f, která zaručuje platnost rovnice (2.5). Zmíněným požadavkům vyhovuje logaritmická funkce. V teorii informace se nejčastěji používá dvojkový logaritmus. S uvážením rovnice (2.1) platí:

$$\boxed{\mathrm{I(X)} = \mathrm{H(X)} = \log_2 \frac{1}{\mathrm{P(X)}} = -\log_2 \mathrm{P(X)}} \tag{2.6}$$

Množství informace zprávy X je rovno entropii a je dáno logaritmem míry neurčitosti

$\frac{1}{\mathrm{P(X)}}$, tj. záporně vzatým logaritmem pravděpodobnosti výskytu této zprávy. Např. zpráva X, která nás informuje o výsledku hodu dvěmi kostkami nese informaci I(X). Zpráva o výsledku hodu první kostky nese informaci I(A) a zpráva o výsledku hodu druhou kostkou informaci I(B). Protože výsledky hodů obou kostek jsou na sobě nezávislé, je celková informace o výsledku hodu obou kostek rovna součtu informací, které nesou zprávy o výsledku jednotlivých poloh obou kostek. Platí tedy I(X) = I(A) +I(B). Nechť např. výsledkem hodu je $\boxed{2}$ na jedné kostce a $\boxed{5}$ na druhé kostce.

Pravděpodobnost toho, že současně padne 2 a 5 je:

$$P_{2,5} = P(X) = P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

a pravděpodobnost, že na první kostce padne 2 je $P_2 = P(A) = \frac{1}{6}$ a pravděpodobnost, že na druhé kostce padne 5 je $P_5 = P(B) = \frac{1}{6}$. Platí tedy:

$$I(X) = \log_2 \frac{1}{\frac{1}{6} \cdot \frac{1}{6}} = \overbrace{-\log_2 \frac{1}{6}}^{I(A)} + \overbrace{\left(-\log_2 \frac{1}{6}\right)}^{I(B)} = \log_2 6 + \log_2 6 = \log_2 36 = 2{,}585 + 2{,}585 = 5{,}17 \text{ bit.}$$

Jednotkou množství informace je bit. Je to informace, kterou nese zpráva o stavu systému, který může nabývat pouze dvou stavů, které jsou stejně pravděpodobné. Informaci 1bit nese např. zpráva o výsledku hodu korunou. Pravděpodobnost, že padne koruna je $P(X) = \frac{1}{2}$ a množství informace, kterou nese zpráva: „Padla hlava“ je podle (2.6):

$$I(X) = -\log_2 \frac{1}{2} = \log_2 2 = 1 \text{ bit}$$

Poznámka 1: Výpočet dvojkového logaritmu můžeme uskutečnit pomocí desítkového logaritmu podle vztahu :

$$\log_2 x = \frac{\log x}{\log 2} = \frac{\log x}{0{,}301} = 3{,}32 \log x$$

(je použita symbolika $\log x = \log_{10} x$),

který dostaneme logaritmováním (při základu 10) vztahu pro definici logaritmů:

$$x = 2^{\log_2 x} = 10^{\log x}$$

Poznámka 2: Poněkud jiný význam má bit (binary digit) ve výpočetní technice, kde znamená dvojkovou číslici 0 nebo 1.

Poznámka 3: Zpráva, která nese informaci o stavu systému, který je jistý (jednoznačně určený), nenese žádnou informaci. Pravděpodobnost výskytu jisté zprávy je P(X)=1 a množství informace:

$$I(X) = -\log_2 1 = 0 \text{bit}$$

Determinované systémy mají nulovou entropii (neurčitost) a zprávy o jejich stavech nesou nulovou informaci. Naopak největší neurčitost má systém, jehož stavy jsou stejně pravděpodobné tj. při rovnoměrném rozdělení pravděpodobností. Neurčitost systému (entropie) závisí tedy na počtu stavů systému i na jejich pravděpodobnosti.

Často se vyjadřuje průměrná entropie systému jako míra průměrné neurčitosti jednoho stavu systému. Může-li systém nabývat s možných stavů a pravděpodobnostmi $p_1, p_2, \ldots, p_s$ pak průměrná entropie $\overline{H(X)}$ je rovna:

$$\overline{H(X)} = H_{stř} = \frac{p_1 H_1 + p_2 H_2 + \ldots + p_s H_s}{p_1 + p_2 + \ldots + p_s} = \frac{\sum_{i=1}^{s} p_i H_i}{\sum_{i=1}^{s} p_i} \tag{2.7}$$

Z počtu pravděpodobnosti je známo, že $\sum_{i=1}^{s} p_i = 1$, neboť se jedná o stavy, které jsou nesoučasné (disjunktní). Vztah (2.7) se zjednoduší na:

$$\overline{H(X)} = \sum_{i=1}^{s} p_i H_i$$

a využitím(2.6) přejde na:

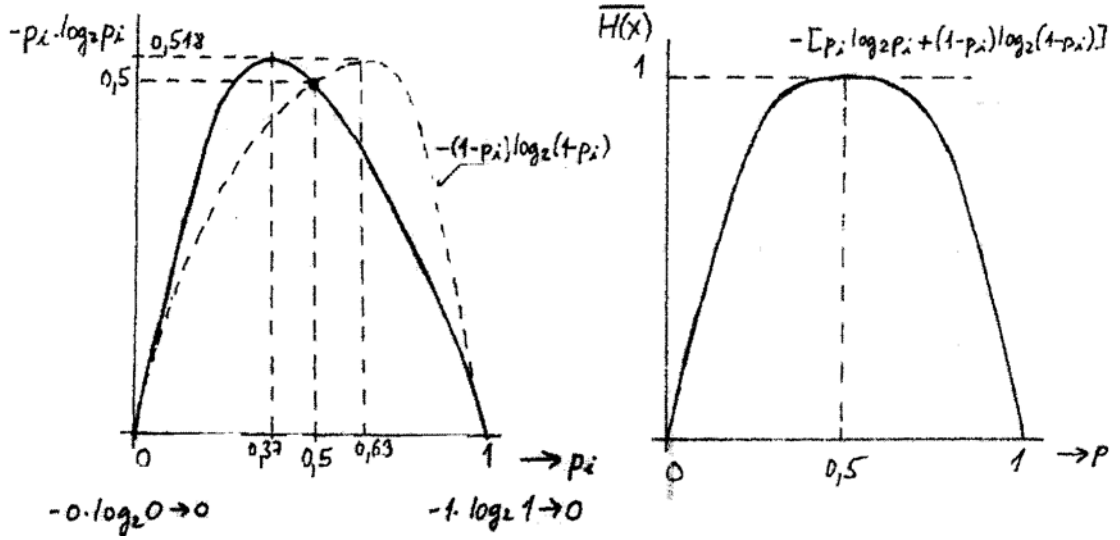
$$\overline{H(X)} = -\sum_{i=1}^{s} p_i \log p_i \ [\text{bit}] \tag{2.8}$$

Pro $p_1 = p_2 = \ldots = p_s = p_i = \frac{1}{s}$ přejde (2.8) na:

$$\overline{H(X)} = -\log_2 \frac{1}{s} \cdot \underbrace{\sum_{i=1}^{s} p_i}_{=1} = -\log \frac{1}{s} \cdot s \cdot \underbrace{\frac{1}{s}}_{=1} = -\log_2 \frac{1}{s} = \log_2 s$$

$$\boxed{\overline{H(X)} = \log_2 s} \tag{2.9}$$

To, že vztah (2.8) nabývá maximální hodnoty pro stejné pravděpodobnosti, vyplývá běhu $p_i \cdot \log_2 p_i$ na obr. 2.3, např. pro $p_1 = p_2 = \frac{1}{2}$ z



prů

Obr. 2.3 Závislost entropie $\overline{H(X)}$ na pravděpodobnostech $p_1 = p$, $p_2 = 1 - p$ (pro jednoduchost vybrána pouze dvojrozměrná funkce). Pro tři možné etapy p_1, p_2, p_3 by zobrazení $\overline{H(X)}$ bylo třírozměrné.

Máme-li zprávu o n znacích se stejným množstvím průměrné informace na znak (se stejnou pravděpodobností výskytů jednotlivých znaků s), je průměrná informace zprávy o n znacích:

$$H_{max} = \overline{I(X)}' = n \cdot \overline{H(X)}' = n \cdot \log_2 s \tag{2.10}$$

Jsou-li pravděpodobnosti výskytů jednotlivých stavů stejné, má zpráva maximální entropii.

Naopak zpráva o n znacích, z nichž každý může nabývat s stavů a různou pravděpodobností:

$$H = \overline{I(X)}' = -n \sum_{i=1}^{s} p_i \cdot \log_2 p_i \le n \cdot \log_2 s \ \left[\text{bit}\right] \tag{2.11}$$

Je-li skutečná entropie H menší než maximální entropie H_{max}, znamená to, že zdroj nevyužívá plně své abecedy. Nevyužití abecedy se kvantitativně hodnotí redundancí neboli nadbytečností zdroje:

$$R = 1 - \frac{H}{H_{max}} = 1 - \mu \qquad \mu = \frac{H}{\log_2 s} \tag{2.12}$$

kde μ je účinnost zdroje.

Redundance je z hlediska přenosu zpráv kanálem s rušením v podstatě žádoucí vlastností zdroje. Dává určitou možnost příjemci zpráv přijaté zprávy opravit (např. telegram s chybami). Na druhé straně je většinou původní nadbytečnost nevýhodná, neboť oprava chyb by vyžadovala složité zpracování. Je proto výhodnější původní nadbytečnost odstranit a místo ní zavést nadbytečnost novou, která efektivně zabezpečuje zprávy při přenosu hlukovým kanálem (využití bezpečnostních kódů).

Množství informace produkované zdrojem za jednu sekundu s průměrnou entropií na prvek $\overline{H(X)}$:

$$\overline{I(X)}'' = v_m \cdot \overline{H(X)} = \frac{1}{\tau} H(X) \ \left[\text{bit/s}\right] \tag{2.13}$$

kde v_m je počet prvků (znaků) za sekundu (s modulační rychlostí),

τ je průměrný časový interval zaujímaný jedním prvkem.

V telegrafní technice se modulační rychlost (počet modulačních stavů za sekundu) vyjadřuje v jednotkách Bd (Baud):

$$v_m = \frac{1}{\tau} \ \left[\text{Bd}\right] \tag{2.14}$$

V případě binárního zdroje se stejně pravděpodobnými prvky ($\overline{H(X)} = 1$) je: $\overline{I(X)}'' = v_m$

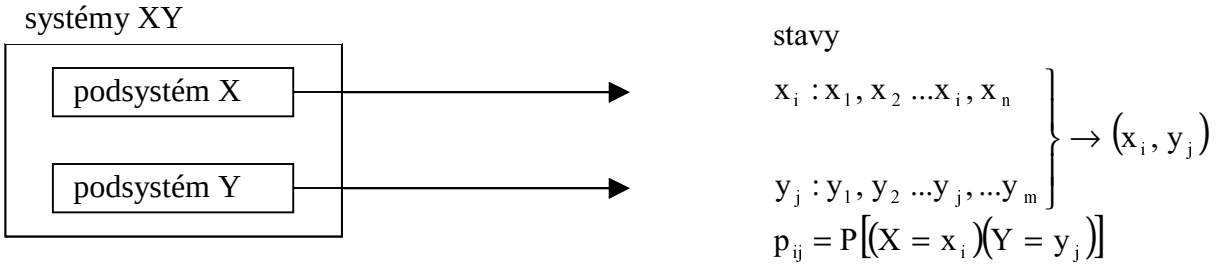
a v případě vícestavového signálu je přenosová rychlost $v_m = \overline{I(X)}''$ rovna:

$$v_p = v_m \cdot \log_2 s \tag{2.15}$$

Může-li např.signál nabývat v každém modulačním stavu některou z 8-mi fází (např. osmistavová fázová modulace) je pak přenos třikrát rychlejší než u dvojstavové modulace, tj. s = 8 a tedy:

$$v_p = v_m \cdot \log_2 s = v_m \cdot \log_2 8 = 3 \cdot v_m \tag{2.16}$$

V praxi často potřebujeme určit entropii systému, který je tvořen z několika podsystémů. Např. dva podsystémy X a Y, které mohou nabývat stavy: $x_1, \ldots, x_n$ a $y_1, \ldots, y_m$, sloučené do jednoho systémy XY, který teď může nabývat stavy určené libovolnými dvojicemi (x_i, y_j). Počet stavů tohoto nového systému je $n \times m$. Pravděpodobnost p_{ij} je pravděpodobnost, že systém XY bude ve stavu x_i, y_j.



Entropie složeného systému XY je:

$$H_{XY} = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \cdot \log_2 p_{ij} \tag{2.17}$$

a) <u>Vzájemně nezávislé podsystémy X, Y</u>

Jsou-li podsystémy X, Y vzájemně nezávislé, je z teorie pravděpodobnosti známo, že:

$$p_{ij} = P(X = x_i) \cdot P(Y = y_j) \tag{2.18}$$

kde $P(X=x_i)$ je pravděpodobnost, že podsystém X bude ve stavu x_i,

$P(Y=y_j)$ je pravděpodobnost, že podsystém Y bude ve stavu y_j.

Dosadíme-li (2.18) do (2.17) dostaneme pro entropii H_{XY} složeného systému:

$$H_{XY} = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i) \cdot P(Y=y_j) \cdot \left[\log_2 P(X=x_i) + \log_2 P(Y=y_j)\right] =$$

$$= -\sum_{i=1}^{n} \underbrace{\sum_{j=1}^{m} P(Y=y_j)}_{=1} \cdot P(X=x_i) \cdot \log_2 P(X=x_i) - \sum_{j=1}^{m} \underbrace{\sum_{i=1}^{n} P(X=x_i)}_{=1} \cdot P(Y=y_j) \cdot \log_2 P(Y=y_j) =$$

$$= -\sum_{i=1}^{n} P(X=x_i) \cdot \log_2 P(X=x_i) - \sum_{j=1}^{m} P(Y=y_j) \cdot \log_2 P(Y=y_j) = H_X + H_Y$$

$$\boxed{H_{XY} = H_X + H_Y} \tag{2.19}$$

Podobný důkaz lze provést i pro systém, který obsahuje větší množství podsystémů. Entropie systému, který je tvořen několika nezávislými podsystémy, je dána součtem entropií těchto podsystémů.

b) Vzájemně závislé podsystémy X, Y

Mějme dva závislé podsystémy X, Y, což znamená, že když jsme již informováni o stavu jednoho z nich, změní se rozdělení pravděpodobností stavu druhého. Např. víme, že podsystém X nabyl stav x_i, pak podmíněná pravděpodobnost $P(y_j/ x_i)$ vyjadřuje pravděpodobnost stavu y_j systému Y za podmínky, že systém X je ve stavu x_i:

$$P(y_j/x_i) = P(Y=y_j/X=x_i) \qquad (2.20)$$

Pro podmíněnou entropii podsystému Y za podmínky, že se podsystém X nachází ve stavu x_i platí:

$$H_{Y/x_i} = -\sum_{j=1}^{m} P(y_j/x_i)\cdot \log_2 P(y_j/x_i) \qquad (2.21)$$

Neurčitost $H_{Y/Xi}$ charakterizuje neurčitost podsystému Y, když známe, jaký je konktrétní stav podsystému X. Nyní můžeme určit průměrnou entropii podsystému Y při znalosti stavu podsystému X. Tato entropie už nezávisí na konktrétním stavu podsystému X a je určena následujícím vzorcem:

$$H_{Y/X} = \sum_{i=1}^{n} P(X=x_i)\cdot H_{Y/x_i} \qquad (2.22)$$

Lze dokázat, že entropie složeného systému, který je tvořen ze vzájemně závislých podsystémů, je určena následovně:

$$\boxed{H_{XY} = H_X + H_{Y/X} = H_Y + H_{X/Y}} \qquad (2.23)$$

Ve zvláštním případě, kdy podsystémy jsou nezávislé, platí:

$$P(y_j/x_i) = P(y_j) \Rightarrow H_{Y/x_i} = H_Y = H_{Y/X}$$

$$H_{XY} = H_X + H_Y$$

Lze dokázat, že platí:

$$H_{Y/X} \leqslant H_Y \qquad H_{X/Y} \leqslant H_X \qquad \boxed{H_{XY} \leqslant H_X + H_Y}, \qquad (2.24)$$

neboť neurčitost systému se po sdělení stavu jeho části nemůže zvětšit, ale jen zmenšit. Budou-li oba podsystémy plně závislé, tj. ze znalosti stavu podsystému X můžeme přesně určit stav podsystému Y, bude platit:

$$H_{Y/X} = \emptyset \qquad (H_{X/Y} = \emptyset) \qquad (2.25)$$

Dosazením (2.25) do (2.23) dostaneme:

$$H_{XY} = H_X = H_Y \qquad (2.26)$$

To znamená, že entropie složeného systému XY z plně závislých systémů je určena úplně entropií jednoho z podsystémů.

Entropie systému složeného ze vzájemně závislých podsystémů je menší než součet entropií jednotlivých podsystémů.

Množství informace, které získáváme zprávou budeme měřit jako úbytek entropie systému, o kterém zprávu dostáváme. Systém pak považujeme za zdroj informace. Mějme nějaký systém X. Než dostaneme zprávu o jeho stavu, je jeho entropie H_X. Po obdržení

přesné zpávy o stavu systému X je jeho entropie $H'_X = 0$. Informace I_X o systému X, kterou jsme získali, je:

$$I_X = H_X - H'_X = H_X - 0 = H_X = -\sum_{i=1}^{s} p_i \log_2 p_i \qquad (2.27)$$

Množství informace, které dostaneme ve zprávě sdělující nám přesný stav nějakého systému je rovno entropii tohoto systému.

Chápeme-li vzorec (2.27) jako střední hodnotu, pak $-\log_2 p_i$ představuje dílčí množstvé informace, které získáváme sdělením, že se systém nachází ve stavu x_i

$$I_{x_i} = -\log_2 x_i \qquad (2.28)$$

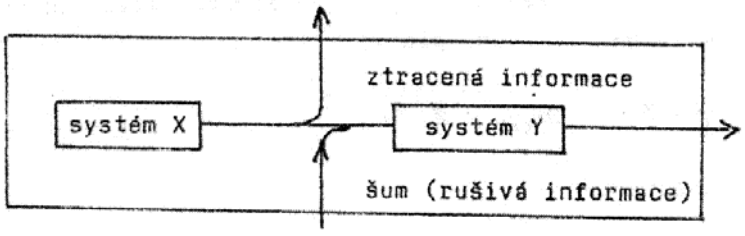
Vzorec (2.27) nám tedy udává průměrnou informaci ve sdělení o stavu systému (na symbol jeho abecedy).

2.3 Přenos informace

Až dosud jsme získávali informaci o systému jeho přímým pozorováním. V praxi získáváme informaci o stavu systému X zprostředkovaně pozorováním systému Y, který je se systémem X nějak spojen. Zpravidla je to způsobeno tím, že systém X je pro nás nedostupný, takže nám nezbývá než jeho stav zjistit na základě zprávy o jiném systému, který je nám dostupný. Např. systém X je tvořen textem telegramu, který je podán na poště v Brně a systém Y je tvořen textem přijatého telegramu na poště v Praze.

Je zřejmé, že stav systému Y nemusí být naprosto totožný se stavem systému X.
Rozdíly mezi pozorovaným systémem Y a nedostupným systémem X, který nás zajímá, mohou být dvojího druhu:

a) Některé stavy systému X se zobrazí do jednoho stavu Y, neboť systém X může nabýt více stavů než Y, tj. systém Y nedokáže rozlišit jemnosti ve stavech X, je hrubší (např. rozdíly způsobené zaokrouhlováním čísel).
b) Chybami při přenosu zprávy mezi systémem X a Y (např. zkreslení signámů způsobené šumy v kanále při přenosu zprávy).



Jestliže se od sebe systém X a Y liší, zajímá nás jak velké množství informace o systému X získáváme pozorováním systému Y.
Velikost informace, kterou získáme, je zmenšena úbytkem entropie systému X v důsledku stavu systému Y.

$$I_{X \to Y} = H_X - H_{X/Y} \qquad (2.29)$$

Do pozorování systému Y byla entropie systému X rovna H_X. Po pozorování systému Y zbyla systému X zbytková entropie $H_{X/Y}$. Rozdíl těchto entropií je informace $I_{X \to Y}$, kterou nám systém Y o systému X poskytl.

1) Pro nezávislé systémy X, Y

$$H_{X/Y} = H_X \quad \text{a tedy} \quad I_{X \to Y} = H_X - H_{X/Y} = H_X - H_X = \emptyset \tag{2.30}$$

Výsledek odpovídá našim zkušenostem, neboť nezískáme žádnou informaci o nějakém systému pozorováním úplně jiného systému, který není se sledovaným nějak spojen.

2) Systémy X a Y jsou si navzájem ekvivalentní
Stav systému Y plně určuje stav systému X a naopak, proto:

$$H_{X/Y} = \emptyset \quad \text{a tedy} \quad I_{X \to Y} = H_X \tag{2.31}$$

3) Více stavů X se zobrazuje do jednoho stavu systému Y
Stav systému X plně určuje stav systému Y, ale podle stavu Y nemůžeme jednoznačně určit stav systému X, neboť systém Y je chudší, tj. může nabývat méně stavů než X. Např. z vyslaného slova jsou vypuštěny všechny souhlásky. Pak z přijatého slova „ p r j d “ nemůžeme s určitostí říci, zda jde o slovo „přijedu“ , „přijde“ nebo „projede“. Entropie chudšího systému Y musí být menší než entropie systému X. Protože stav systému X plně určuje stav systému Y; platí:

$$H_{Y/X} = \emptyset \quad \text{a tedy} \quad I_{X \to Y} = H_Y - H_{Y/X} = H_Y \tag{2.32}$$

Na oba systémy X, Y se můžeme také dívat jako na jeden systém a pak můžeme informaci $I_{X \to Y}$ vyjádřit také pomocí entropie složeného systému XY. Ze vztahu (2.23) vyplývá, že:

$$H_{X/Y} = H_{XY} - H_Y \quad \text{a po dosazení do (2.29) dostaneme:} \tag{2.33}$$

$$I_{X \to Y} = H_X + H_Y - H_{XY} \tag{2.34}$$

Na základě vztahu (2.34) se dá dokázat, že informace $I_{X \to Y}$ může být vyjádřena pomocí pravděpodobností následovně:

$$I_{X \to Y} = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \cdot \log_2 \frac{p_{ij}}{p_i \cdot r_j} \tag{2.35}$$

kde

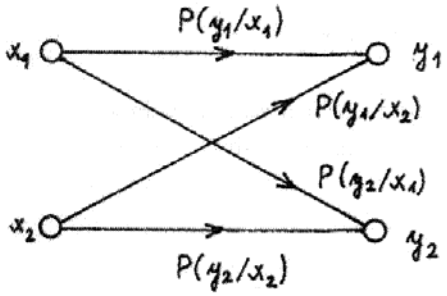
$$p_{ij} = P[(X = x_i)(Y = y_j)]$$

$$p_i = P(X = x_i) \qquad r_j = P(Y = y_j)$$

2.4 Diskrétní kanály

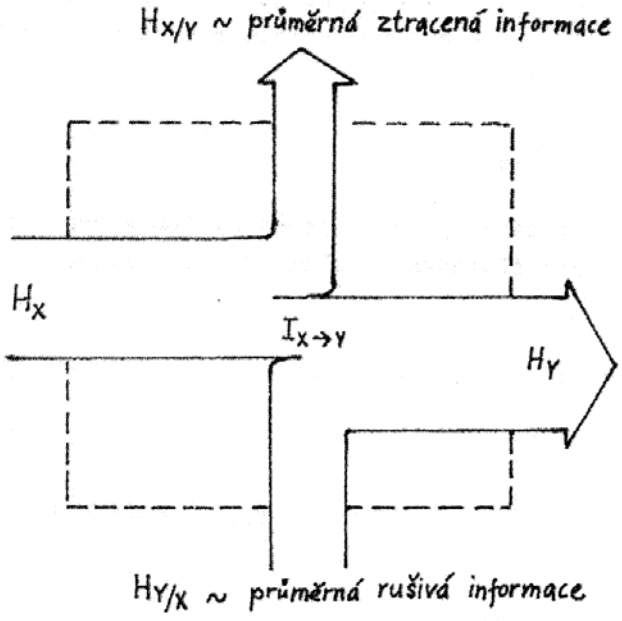
Kanálem rozumíme souhrn prostředků sloužících k přenosu signálu od zdroje k příjemci signálu. Diskrétní signály jsou určené pro přenos diskrétních zpráv.

Uvažujeme nejprve dvojkový kanál, který je určen pro přenos dvojkových signálů. Množina X, z níž jsou vybírány odesílané prvky, je tvořena prvky x_1 a x_2, množina Y přijatých prvků je tvořena prvky y_1 a y_2. V ideálním kanále by bylo přiřazení prvků z množiny Y prvkům množiny X jednoznačné. Prvku x_1 by vždy odpovídal prvek y_1, prvku x_2 prvek y_2. Kanál tohoto typu se nazývá kanál bezhlukový. Skutečný dvojkový kanál přiřazuje prvkům



Obr. 2.4 Dvojkový kanál

množiny X prvky z množiny Y více či méně náhodně. Je-li přiřazování prvků nezávislé na tom, jaké prvky byly na předchozích místech posloupnosti vysílaných prvků – jde o kanál bez paměti. Je-li přiřazování prvků popsáno soustavou podmíněných pravděpodobností, která je neměnná, tj. nezávisí na pořadí prvku v posloupnosti vysílaných prvků, jedná se o kanál stacionární. Schéma dvojkového stacionárního kanálu bez paměti je nakresleno na obr. 2.4 . Při vyslání prvku x_1 je přijat buď prvek y_1 nebo y_2. Při vyslání prvku x_2 je situace obdobná. Podmíněné pravděpodobnosti typu $P(y_j/x_i)$ udávají pravděpodobnost přijetí prvku y_j za podmínky, že byl vyslán prvek x_i. Pravděpodobnosti $P(y_1/x_1)$ s $P(y_2/x_2)$ popisují případy správného přenosu, pravděpodobnosti $P(y_2/x_1)$ s $P(y_1/x_2)$ popisují případy chybného přenosu.



Obr. 2.5 Přenos informace hlukovým kanálem

Od dvojkového kanálu zobecněním přijdeme k obecnému diskrétnímu kanálu. Na jeho vstup jsou přiváděny prvky množiny X, tj. prvky $x_1,x_2 \dots x_i, \dots, x_n$ a na výstupu kanálu se objevují prvky množiny Y, která je tvořena prvky $y_1,y_2 \dots y_j, \dots, y_m$. Vlastnosti stacionárního kanálu bez paměti jsou popsány entropiemi H_X, H_Y, $H_{X/Y}$, $H_{Y/X}$ a H_{XY} (viz vztahy 2.17 až 2.24). Entropie H_X na obr. 2.5 udává průměrné množství informace nesené jedním prvkem odesílané zprávy. Objeví-li se na výstupu kanálu určitý prvek y_j, nejsme si nikdy zcela jisti, zda byl vyslán prvek jemu odpovídající nebo nějaký jiný. Naše nejistota je vytvářena podmíněnou entropií $H_{Y/X}$, která udává, jaké množství užitečné informace se v průměru ztrácí při přenosu jednoho prvku kanálem. Vzájemná informace $I_{X \rightarrow Y}$ je určena rozdílem entropií (viz.2.29):

$$I_{X \rightarrow Y} = H_X - H_{X/Y} \qquad (2.36)$$

a udává průměrné množství informace přenesené kanálem při odeslání prvku z množiny X a přijetí prvku z množiny Y. Vzájemnou informaci můžeme vyjádřit i pomocí entropií H_Y, $H_{Y/X}$. Vztah (2.23) upravíme do tvaru:

$$H_X = H_{XY} - H_{Y/X} \qquad (2.37)$$

a dosadíme za H_X do vztahu (2.36):

$$I_{X \to Y} = H_{XY} - H_{Y/X} - H_{X/Y} \qquad (2.38)$$

Dosazením za H_{XY} z (2.23) dostaneme:

$$I_{X \to Y} = H_Y + H_{X/Y} - H_{Y/X} - H_{X/Y} = H_Y - H_{Y/X} \qquad (2.39)$$

Entropie HY udává průměrné množství informace přinesené objevením se jednoho prvku z množiny Y na výstupu kanálu, tj.:

$$H_Y = I_{X \to Y} + H_{Y/X} \qquad (2.40)$$

Část $I_{X \to Y}$ představuje užitečnou informaci, část $H_{Y/X}$ je tzv. entropie hluku. Na základě rovnic (2.36) a (2.39) lze odvodit vztah:

$$\boxed{H_X + H_{Y/X} = H_Y + H_{X/Y}} \qquad (2.41)$$

který vyjadřuje skutečnost, že součet entropií do kanálu vstupujících se rovná součtu entropií z kanálu vystupujících (tzv. zákon o zachování entropie)

Schopnost kanálu přenášet informaci se popisuje veličinou nazývanou propustnost neboli kapacita kanálu. Propustností kanálu rozumíme maximální množství informace, které je kanál schopen v průměru přenést jedním prvkem, nebo které je kanál schopen v průměru přenést za jednotku času. Na základě znalostí kapacit kanálů lze provést vzájemné porovnání kanálů. Propustnost kanálu je dána maximem vzájemné informace $I_{X \to Y}$. Maximum se hledá tak, že se mění rozdělení pravděpodobností $P(x_i)$ prvků x_i množiny X. Podmíněné pravděpodobnosti $P(y_j/x_i)$, tj.: prvky matice kanálu měnit nelze, protože vlastnosti kanálu jsou pevně dány. Propustnost kanálu C je tedy dána vztahem:

$$C = \max_{P(x_i)} \frac{1}{\tau} I_{X \to Y} = \max_{P(x_i)} \frac{1}{\tau} (H_Y - H_{Y/X}) \; [bit/s] \qquad (2.42)$$

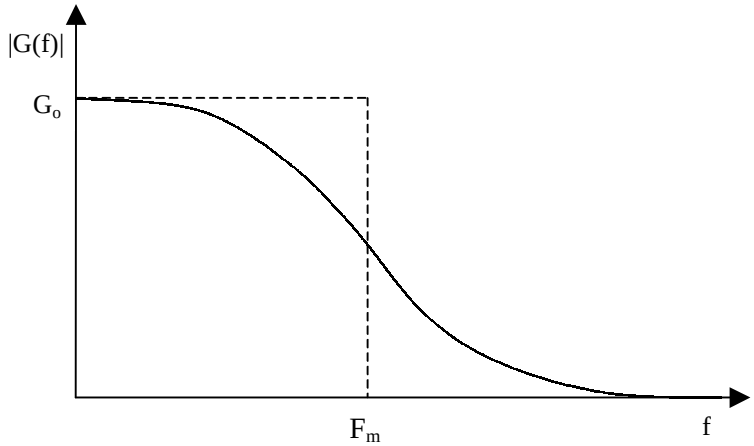
kde τ je průměrný časový interval zaujímaný jedním prvkem. Úkolem teorie informace je určení nejekonomičtějšího kódování informace tak, aby příslušná informace mohla být co nejrychleji předána přes nějaký spojovací kanál. Převážná většina primárních signálů má povahu signálů analogových. Tyto signály jsou sice převáděny do číslicové podoby, pro zpracování počítačem, ale jejich přenos se z části realizuje pomocí spojitých kanálů (např. pomocí fázové nebo frekvenční modulace).

Minimální možná délka signálových prvků τ (modulačních stavů), při které lze ještě realizovat bezchybný přenos takových prvků, souvisí pouze s fyzikálními parametry kanálu. Přibližně ji stanovíme následovně. Modul přenosové charakteristiky kanálu |G(f)| nahradíme pravoúhlou charakteristikou ideálního kanálu s ekvivalentním mezním kmitočtem F_m, jak je uveden na obr. 2.6 .

Na základě Kotelnikovova teorému dostaneme pro minimální možnou délku signálových prvků τ_0 vztah:

$$\tau_0 = \frac{1}{2 \cdot F_m} = \frac{G_0}{2 \cdot \int_0^{\infty} |G(f)| \, df} \qquad (2.43)$$

U kanálu, který má charakter souměrné, relativně úzkopásmové propustě, bude potom minimální délka signálových prvků dána dvojnásobkem hodnoty τ_0 podle (2.43).



Obr. 2.6 Ekvivalentní šířka přenosové charakteristiky kanálu

Nyní zbývá stanovit maximální hodnotu informace $I_{X\to Y}$. V případě kanálu bez rušivých procesů bude platit vztah (2.44), takže propustnost takového kaná-lu bude:

$$C = \frac{1}{\tau_0}\max I_{X\to Y} = \frac{1}{\tau_0}\max H_X \qquad (2.44)$$

Z předcházejících poznatků víme, že entropie H_X nabývá maximální hodnoty v případě stejně pravděpodobných prvků x_i, takže:

$$H_X = -\sum_{i=1}^{s} P(x_i)\cdot \log_2 P(x_i) = \log_2 s \qquad (2.45)$$

kde S je celkový počet možných prvků x_i. Propustnost kanálu bez rušivých procesů bude:

$$C = \frac{1}{\tau_0}\cdot \log_2 s = 2\cdot F_m \cdot \log_2 s = B\cdot \log_2 s \quad [bit/s] \qquad (2.46)$$

kde B je šířka úzkopásmové propustě.

Maximální množství informace, které lze takovým kanálem přenést za dobu ΔT bude:

$$I_{\max} = C\cdot \Delta = 2\cdot F_m \cdot \Delta T \cdot \log_2 s \quad [bit/s] \qquad (2.47)$$

V mnoha případech je přenos realizován binárním signálem (s = 2). Potom v předchozích vztazích bude $\log_2 s = 1$.

Stanovení propustnosti kanálu za přítomnosti rušivých procesů je v obecném případě velmi složité.

2.5 Spojité kanály

Při informačním popisu spojitých kanálů se snažíme využít postupů a veličin zavedených pro popis diskrétních signálů. Budeme postupovat tak, že nahradíme analogový signál diskrétním v čase tak, aby bylo zachováno množství informace nesené analogovým signálem (dodržení vzorkovacího Kotelnikovova teorému).

Diskretizaci signálu v časové oblasti říkáme vzorkovaní signálu a diskretizaci signálu vzhledem k amplitudě říkáme kvantování signálu.

Obdobně jako u nespojitých tak i u spojitých kanálů lze definovat propustnost kanálu. Odvození vztahu pro propustnost spojitého kanálu však vyžaduje hlubší znalosti z tohoto oboru, proto si uvedeme pouze výsledný vztah:

$$C = F_m \cdot \log_2 \frac{P+N}{N} \quad [bit/s] \qquad (2.48)$$

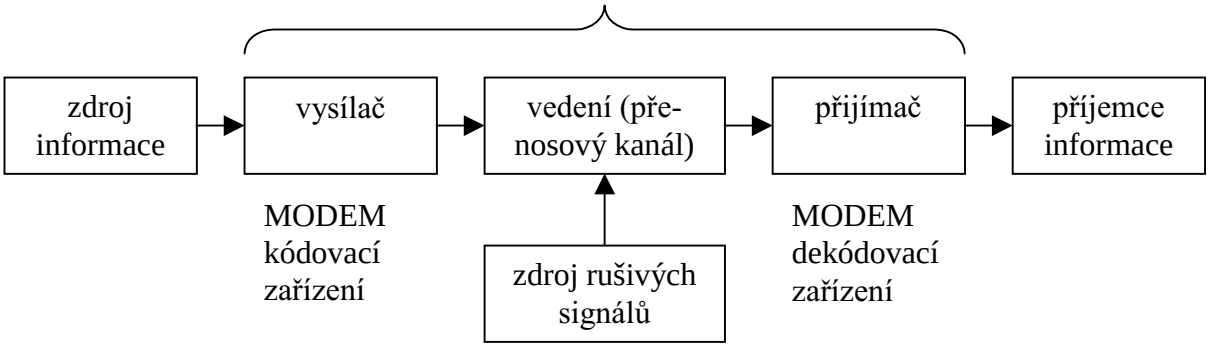
kde F_m je mezní frekvence kanálu (spektrální složky signálu nad touto frekvencí jsou zanedbatelně malé), (kanál má charakter ideální propustnosti), P je střední hodnota výkonu

užitečného signálu, N je střední výkon rušivého signálu. Číslo $\frac{P+N}{N}$ můžeme interpretovat jako počet rozlišitelných úrovní signálu u_p při působení šumu u_N. Uvedený vztah platí pouze za podmínky, že hustota rozdělení pravděpodobnosti vstupního signálu je dána normálním rozdělením a kanál je rušen normálním aditivním šumem. Maximální množství informace přenesené za dobu ΔT bude tedy:

$$I_{\Delta T} = \Delta T \cdot F_m \cdot \log_2 \frac{P+N}{N} = \Delta T \cdot \Delta F \cdot \Delta P = V \qquad (2.50)$$

kde ΔT je doba trvání signálu (doba využití kanálu),
ΔF je šířka spektra signálu (kanálu),
ΔP je odstup signál-šum, určující počet rozlišitelných úrovní signálu (dáno maximálně přípustnou úrovní signálu v kanálu, např. z důvodů přeslechu signálu a úrovní šumu),
V je objem signálu (kanálu).

Pro nezkreslený přenos signálu kanálem musí platit, že objem signálu je menší než objem kanálu. Jestliže např. ve vztahu (2.50) zmenšíme ΔP tak, že N > P, ale na druhé straně úměrně zvětšíme ΔF, můžeme realizovat tzv. systémy s rozprostřeným spektrem. Je-li N>P, pak takový užitečný signál z radiového vysílače není možné zachytit obyčejným přijímačem (slouží k utajení informace).



Obr. 2.7 Schéma přenosu zpráv

Vzájemný vztah rychlosti produkce informace zdrojem zpráv, propustnosti kanálu a pravděpodobnosti správného přenesení zprávy zformuloval Shannon. Shannonova věta o kódování v šumovém kanále říká: v šumovém kanále s danou kapacitou C můžeme pro zdroj informace s entropií H, za předpokladu H<C, nalézt takový způsob kódování dlouhých zpráv, že pravděpodobnost výskytu chyby P (e) je menší než předem dané libovolně malé číslo ε> 0, avšak nikoliv rovno nule.

Věta o kódování je větou existenční, neboť nám říká, že existuje takový způsob kódování zpráv, při kterém lze dosáhnout libovolně malé pravděpodobnosti chybného přenesení zprávy i při rychlosti produkce informace blížící se propustnosti kanálu. Z uvedené věty však nevyplývá žádný návod, jak takový kód, který je toho schopen, sestrojit. To je jedna z příčin existence velkého množství různých kódů, z nichž některé lépe, jiné hůře, slouží k zabezpečení zpráv pro přenos hlukovým (šumovým) kanálem.