TEORIE INFORMACE

I. INFORMACE

měřitelná veličina
nezávislá na fyzikálním prostředí, kterým je přenášena
"vzor" (pattern)

míra informace (nejvhodnější) je podobná míře entropie
existují důvody pro změnu znaménka – t.j. informace je opakem entropie (ve skutečnosti i v matematické formulaci)

II. PŘENOS informace kanálem, který je vystaven rušení

Největší možná rychlost přenosu informace závisí na šířce kmitočtového pásma kanálu a též na poměru signálu k šumu.
Mezi oběma veličinami – š. pásma a poměrem s/š existuje určitá možnost vzájemné převoditelnosti.

III. Lze stanovit pevná kritéria pro návrh FILTRŮ, kterými se co nejúčinněji oddělí ze signálu se šumem samotný signál.
(V soustavě sdělování na dálku, nebo v soustavě samočinného řízení)

IV. Metody teorie informace lze vedle problémů sdělovacích soustav aplikovat též na jiné problémy získávání informace ze systémů obsahujících prvky náhodnosti:
statistické plánování experimentů
třídění údajů podle požadovaného uspořádání.

Rozdíl mezi fyzikálním prostředím a informací:
před uveřejněním → C.E. Shannon: "A Mathematical Theory of Communication", Bell Syst. Techn. Journ. 27, 1948
→ C.E. Shannon: "Communication in the Presence of Noise" Proc. Inst. Rad. Eng. 37, 1949
byla pozornost věnována vlnovým průběhům el. veličin (proud, napětí)
– informační obsah zprávy nebyl zkoumán.

Úkol sdělovacího technika : rekonstruovat na přijímací straně zprávu ve formě co nejvíce shodné s formou, ve které byla vložena do ~~přijímače~~ vysílače.

Je lepší hledat, jak přenášet **INFORMACI** místo doslovné zprávy.

× Je informace úměrná počtu slov v knize?
(účtování za slovo telegramu)

"**INFORMACE**" – ve světě pojmů (nedostupná objektivnímu číselnému hodnocení)
– ve smyslu "teoretické množství informace" nebo "informačního objemu"

Tlak řady praktických problémů :

→ Rozvoj teorie informace : užití kmitočtové modulace ⟷ lepší poměr s/š

→ kódovaná impulsová modulace (PCM)
(šířková impulsní modulace, fázová imp. modulace)
je lepší než kterákoliv předešlá ⇒ Existuje nějaká hranice, za kterou již při daném výkonu vysílače nelze již zlepšovat šumový poměr použitím dokonalejších způsobů modulace, vyžadujících větší šířku pásma než prostá amplitudová modulace ?

Otázka :

Máme časovou řadu hodnot vyjadřující postupné pozorování nějakého procesu (např. poloha letadla ...) → znam procesu,
nahodilost, jistý stupeň stejnorodosti (omezené možnosti hodnot rychlosti, zrychlení, ...) zachovávání statistických vlastností
STACIONÁRNÍ ČASOVÁ ŘADA

Za předpokladu, že extrémní odchylky takové řady se vyskytují s malou pravděpodobností, očekáváme, že nejlepšího odhadu pravděpodobné příští polohy (hodnoty procesu) letadla dosáhneme tím, že výraz pro pravděpodobné odchylky podrobíme jakémusi filtračnímu (vyhlazovacímu) procesu.

Predikce (viz → N. WIENER: "Extrapolation, Interpolation and Smoothing of Stationary Time Series". John Wiley, N.Y., 1949)

→ Problém soustav samočinného řízení (servosystémů), jejichž činnost je založena na velikosti signálu odchylky, ve kterých ovšem jsou však poruchami různého druhu vyvolávány druhotné, nepravé signály, šumy. [redacted] Úkolem je nalézt nejlepší prostředek, jak oddělit signál od šumu. Problém je třeba řešit podobně, jako úkol predikce → NÁVRHEM OPTIMÁLNÍHO FILTRU.

OPTIMÁLNÍ FILTR – je takový, který dává nejlepší výsledky z hlediska uvažovaného problému (!)

→ východiska pro návrh filtru se liší
filtr pro samočinné řízení polohy (různé hodnocení výsledku)
filtr pro potlačení šumu v záznamu

Filtr může být součástí soustavy dálkového sdělování
→ problém zpracování údajů – pomocí výpočetní techniky.

1* abstraktní pojem informace, je kvantitativní, leží mezi fyzikálním prostředím a sémantickým pojetím

2* množství detailů ve fyzikální reprezentaci není synonymem pro množství informace

3* míra informace je matematicky příbuzná míře entropie (existuje pro to i fyzikální důvod: informace ↔ opak entropie)

4* Šumový poměr $\frac{P}{N}$ lze ve sdělovací soustavě zaměnit šířkou pásma W. Existuje limitní věta o rychlosti přenosu C se zanedbatelným rizikem chyby:

$$C \leq W \log\left(1 + \frac{P}{N}\right)$$

K ČEMU JE DOBRÁ TEORIE INFORMACE

a) zlepšení telekomunikačních soustav

b) studium pojmů informace, indukce, zdroj informace, zdroj

c) rozhodovací funkce při strategiích (teorie her)

d) lingvistika

e) strojový překlad cizojazyčných textů

f) automatizace administrativních prací, čtecí, poslechové stroje

g) interpretace fyziologických jevů, biologická kybernetika, mechanismy sluchu a zraku

h) konstrukce strojů s vlastnostmi živočichů, stroje schopné "učit se" ze zkušenosti

O LOGARITMICKÉ MÍŘE INFORMACE

Míra přenášené informace:

písmeno - kombinace 5 prvků; ("1", "0")

$2^5 = 32$ kombinací

(ve skutečnosti určení "začátek", "konec", číslice, písmena - pomocí 2 kombinací z 32)

sestavujeme-li slova ze 6-ti znaků $(2^5)^6 = 2^{30}$

.... ▷ atd.: ze slov zprávy

počet kombinací roste mocninně (exponenciálně s časem)

Intuitivně - množství informace roste úměrně s počtem slov (s různou konstantou úměrnosti - podle původce zprávy!)

Slova, která se vyskytují ve zprávě později, obsahují více informace (tento druh souvztažnosti mezi úseky zprávy je však malý...)

Informace - roste-li s trváním zprávy, je úměrná logaritmu počtu kombinací, které by bylo možné vytvořit v rámci daného trvání zprávy.

<u>lineární nárůst s časem:</u> inf. objem $\underline{I} = K \cdot m$

m - počet úseků

s - počet alternativ v každém úseku; s^m - počet různých kombinací za danou dobu

$K \neq f(m)$ $K = f(s)$ (může být)

Množství informace,

($\underline{I}$) závisí pouze na celkovém počtu kombinací, které lze vytvořit

Mějme 2 různé soustavy s počty alternativ S_1 a S_2

$S_1^{n_1} = S_2^{n_2}$ (přenesení stejného množství informace)

$I = K_1 n_1 = K_2 n_2$

$n_1 \log S_1 = n_2 \log S_2$

$\frac{K_1}{\log S_1} = \frac{K_2}{\log S_2} = K_0 \implies I = K_0 n_1 \log S_1 = K_0 n_2 \log S_2$

základ logaritmů ani velikost K_0 nebyly specifikovány

$K_0 = 1$, základ logaritmů: 2 $\longrightarrow$ [míra informace] = bit
"binary digit" (J.W. Tukey)

$I = n \log_2 S$ [bit]

pro $S = 2$ $I = n$ bit

Stejně, jako se přesnost číselného údaje vyjádřit na 1%, lze vyjádřit jako přesnost jako $\frac{1}{2^7}$ (lepší než 1%). Takové číslo pak obsahuje množství informace "7 bitů".
($2^{10} = 1024$)

TOTO BYLA BEZŠUMOVÁ SOUSTAVA!

Tvrzení o rozlišitelnosti je vysloveno jako definitivní.

Jinak je tomu v PRAVDĚPODOBNOSTNÍ soustavě $\longrightarrow$ kde přijatý signál sice ovlivňuje relativní pravděpodobnost možných vyslaných zpráv, ale neurčuje jedinou z nich s úplnou jistotou.

ekvivalentní $\rightarrow$ v procesu měření je určování přesnosti měření na základě "pravděpodobné chyby" (směrodatné odchylky) namísto absolutních mezí.

Pro zahrnutí těchto pravděpodobnostních situací je nutno zavést pojem ENTROPIE.

ENTROPIE A INFORMACE

(i) informační objem kanálu (bit/s)

(ii) množství informace přeneseného signálu (závisí na počtu možných signálů, z nichž byl vybrán)

(iii) míra spolehlivosti, že příjemce zprávy obdržel zprávu ve správném znění (věrohodnost "authenticity")

authenticity factor <!!> "likelihood ratio" (při testování hypotéz (!))

ŠUMOVÝ BINÁRNÍ KANÁL (50% fr. chyby)

Přijatý signál	Pravděpodobnost vyslání "0"	Pravděpodobnost vyslání "1"
0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	$\frac{1}{2}$
nic	$\frac{1}{2}$	$\frac{1}{2}$ stav věcí v případě, že přijatý signál nebyl přečten.

Lze předpokládat, že při 50% frekvenci chyb lze přenést "dostatek" informace?

→ Není rozdíl mezi odhadem vysl. signálu v případě, že přijatý signál byl čten jako "0" nebo "1", nebo – byl-li ignorován – ve skutečnosti to odpovídá situaci, jako by žádný signál nebyl vyslán.

t.j. <u>informace přenesená binárním symetrickým kanálem s 50% frekvencí chyby se rovná přesně nule.</u>

Jak to lze vyjádřit?

n jevů : $p_1, p_2, p_3 \dots p_n$ – úplná soustava $\sum p_i = 1$

$H = -\sum_{i=1}^{n} p_i \log p_i$ se nazývá **ENTROPIÍ SYSTÉMU**

H je kladné

Položme změnu informace $\Delta I = -\Delta H$

→ vztah pro entropii dobře vyjadřuje informaci přenesenou kanálem za přítomnosti chyb.

($-\Delta H = H_1 - H_2$ $H \geq 0$ při růstu entropie je změna informace záporná)

Lze matematicky dokázat (A. FEINSTEIN: Foundation of Inf. Theory, Mc Graw-Hill, 1958), že entropický charakter funkce H je jediný, který splňuje tyto 3 podmínky:

1) Pro binární systém s pravděpodobnostmi p a 1-p je funkce $H(p, 1-p)$ <u>spojitou funkcí argumentu p</u>, pokud $0 \leq p \leq 1$ (p je pravděpodobnost)

2) Pro mnohonásobný systém je $H(p_1, p_2, \dots, p_n)$ <u>symetrickou funkcí všech argumentů</u>

3) Je-li jedna z pravděpodobností, např. p_n, rozdělena na q_1 a q_2, kde $q_1 + q_2 = p_n$, lze entropii nově vzniklého systému vyjádřit jako součet entropie původního systému a příspěvku zjemnění:

$$H(p_1, p_2, \dots p_{n-1}, q_1, q_2) = H(p_1, p_2, \dots p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Pozn.: Aplikace vzorce $H = \sum_i p_i \log p_i$ pro entropii na binární kanál: (viz tabulka), 3. řádek

pravděpodobnostní systém vyjádřený 3. řádkem má entropii

$H = -2 \cdot \frac{1}{2} \log \frac{1}{2}$ 1 bit. (před vysláním signálu)

(Entropie je nulová, pokud jedno z čísel p_1 resp. p_2 je rovno 1)

Vyslání jednoho bitu <u>bezšumovým</u> kanálem anuluje entropii. Entropie na přijímacím konci se sníží o 1 bit vzhledem k přírůstku 1 bitu informace.

Kanál s šumem:

Pro binární symetrický kanál se $100q$ procentní [redacted] frekvencí chyby, kde $q = 1 - p$, je maximální množství přenesení informace rovno:

$$C = (1 + p\log_2 p + q\log_2 q) \quad \text{bit / znak} \quad \text{resp. prvek.}$$

Výraz představuje úbytek entropie, vyjadřuje také správné množství informace, které je nejvýše rovno 1 bitu na znak.

Pro 10% frekvenci chyby dostaneme $C = 0.531$ bit / znak

Pozn.: Pravděpodobnost $\Leftrightarrow$ informace

přijetí zprávy o velmi pravděpodobném jevu $\longrightarrow$ MÁLO INFORMACE

málo pravděpodobném jevu $\longrightarrow$ MNOHO INFORMACE

$\Rightarrow$ pravděpodobnost může být vhodnou mírou informace

$I(x) = \varphi\left[\frac{1}{P(x)}\right]$ množství informace je nepřímo úměrné pravděpodobnosti výskytu zprávy.

3 zprávy: A, B, C .... $P(A), P(B), P(C)$, jsou nezávislé; přijetí zpráv B a C považujeme za novou, složenou zprávu D

$I(D) = \varphi\left[\frac{1}{P(D)}\right]$, pp. že $I(D) = I(B) + I(C)$

$P(D) = P(B \wedge C) = P(B) \cdot P(C)$ (nezávislost)

$I(D) = \varphi\left[\frac{1}{P(B)\cdot P(C)}\right] =$ také $= \varphi\left[\frac{1}{P(B)}\right] + \varphi\left[\frac{1}{P(C)}\right]$

tato rovnost platí v případě, že φ je logaritmickou funkcí.

$$I(D) = \log \frac{1}{P(B)\cdot P(C)} = \log \frac{1}{P(B)} + \log \frac{1}{P(C)}.$$

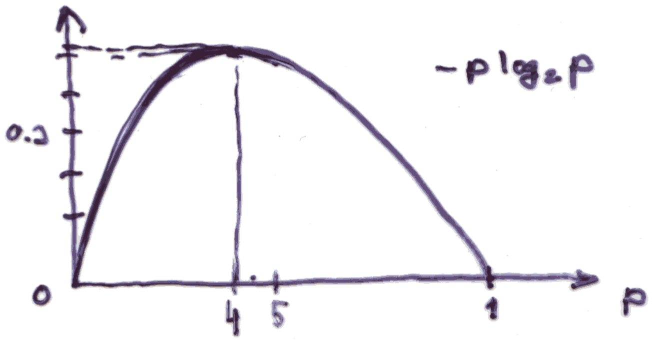
$\Rightarrow$ **Množství informace lze vyjádřit záporným logaritmem pravděpodobnosti výskytu zprávy**

$$I(x_i) = \log \frac{1}{P(x_i)} = -\log P(x_i)$$

Pojmy zdroj zpráv, nebo množství informace lze spojovat obecně s jakýmikoliv náhodnými jevy, tvořícími úplnou soustavu jevů.

Informační obsah jednoho prvku
Obecně je jeho hodnota pro každý prvek různá.
Po zavedení průměrné hodnoty $\overline{I(x_i)} = \sum_{i=1}^{m} I(x_i) \cdot P(x_i) = -\sum_{i=1}^{m} P(x_i) \log P(x_i)$

$H(x)$ – <u>entropie zdroje</u> (Shannonova kladná entropie)

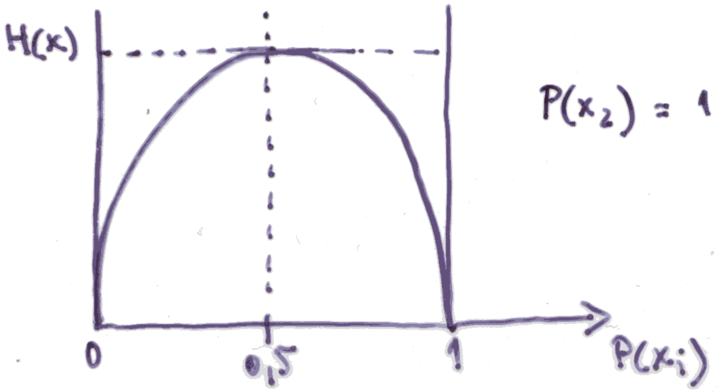


BINÁRNÍ SYSTÉM, ZDROJ $\quad P(x_1) = P(x_2) = 0.5$

$H(x) = \log 2 \qquad H(x) = 1\,\text{bit / prvek}$

interpretace: informační obsah 1 bit má zpráva, která byla vybrána ze dvou možných, stejně pravděpodobných zpráv.
($\log_2$ – bit; $\log_e$ – nat; $\log_{10}$ – Hartley)

informační obsah zprávy délky n prvků, produkované binárním zdrojem při stejně pravděpodobných prvcích (symetrický zdroj) bude dán přímo $I = n \cdot H(x) = n\,\text{bit}$



$P(x_2) = 1 - P(x_1)$

Průměrná hodnota informace, entropie, nabývá maximální hodnoty při stejně pravděpodobných prvcích

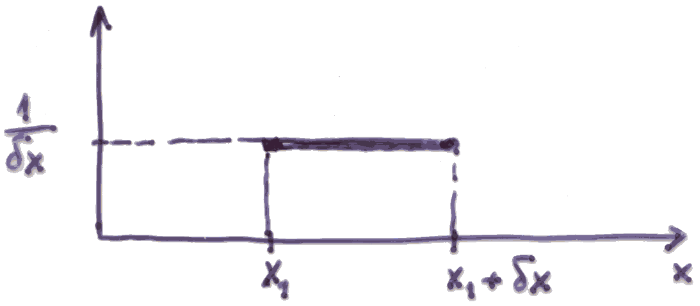
t.j. $P(x_1) = P(x_2) = \dots = P(x_s)$

$$H(X) = \log_2 S \quad \text{bit/prvek}$$

Měření pomocí tolerancí:

měření s přesností $\frac{1}{2^7}$ .... 7 bitů

Víme-li, že naměřená hodnota např. délky leží v "intervalu" δx, pak pravděpodobnost, že přesná hodnota leží mimo interval δx se rovná nule.



$p(x) = 0 \qquad x < x_1 \text{ a } x > x_1 + \delta x$

$p(x) = \frac{1}{\delta x} \qquad x_1 < x < x_1 + \delta x$

$$-H = \int_{-\infty}^{\infty} p(x) \log p(x)\, dx = \frac{1}{\delta x} \log \frac{1}{\delta x} \int_{x_1}^{x_1+\delta x} dx =$$

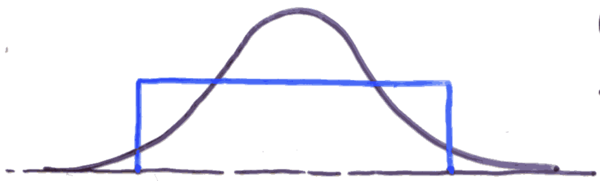
$$= \log \frac{1}{\delta x} \; ; \text{ při } \delta x = \frac{1}{2^7} \qquad H = 7$$

Gaussovo rozložení chyb:

Množství informace měření, které má směrodatnou odchylku σ

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-x_0)^2}{2\sigma^2}} \; ; \qquad I = \log_e \frac{1}{\sqrt{2\pi e}} = 1{,}443 \log_e \frac{1}{4{,}132\,\sigma} \; [\text{bit}]$$

$\pm 3\sigma$ (výskyt téměř jistý)



Gaussovo a obdélníkové rozložení chyb se stejným informačním obsahem.

Při záměně δx za σ $\Longrightarrow$ korekční činitel $\sqrt{2\pi e} \approx 4{,}132$

Entropie ve fyzice

$$\delta S = \frac{\delta Q}{T}$$

n molekul, q různých energetických stavů $\sum_{k=1}^{q} n_k = n$

Počet způsobů, jak vybrat r předmětů z celk. počtu n $C_{n,r} = \frac{n!}{(n-r)!\,r!}$

–"– , jak rozmístit r předmětů do q skupin

po $n_1, n_2 \ldots n_q$ $C_{n,q} = \frac{n!}{n_1!\,n_2!\ldots n_q!}$

Pravděpodobnost kteréhokoliv uspořádání $n_1, n_2 \ldots n_q$ je úměrná počtu způsobů, jimiž lze toto uspořádání dosáhnout výměnou jednotlivých částic mezi různými energetickými stavy.

$W(q) \sim C_{n,q}$ $-K \sum_{k=1}^{q} n_k \log_e n_k$

Statistická definice ENTROPIE $K \cdot [\log_e W(q) - n \log_e n]$

$K = k_B$ $\Rightarrow$ shoda s makroskopickou termodynamickou definicí

Pravděpodobnost, že molekula má energii E je úměrná $e^{-\frac{E}{kT}}$,

střední energie na stupeň volnosti $\frac{1}{2}kT$

Inverzní pravděpodobnost:

přístroj "neví", zda přijímá signály nebo šum

Pravděpodobnost, že se jedná o signál, sestavený s cílevědomým záměrem. VĚROHODNOST přijatého signálu.

Časový průběh signálu nechť obsahuje N jednotek energie, rozdělených do q oddělených úseků.

Náhodný šum pochází z mnoha oddělených zdrojů vz. nezávislých → rozdělení energie do časových intervalů bude náhodné.

Pravděpodobnost, že daný úsek – průběh – vznikl náhodou, je úměrná počtu vzájemně odlišných seskupení jednotlivých jednotek energie, které vedou k vytvoření tohoto průběhu:

$$p(0) = \text{const.} \cdot C_{N,q} = \text{const} \frac{N!}{\underbrace{n_1!\, n_2! \ldots n_q!}_{\text{počty jednotek energie v jednotlivých úsecích } 1, 2 \ldots q}}$$

Věrohodnost je úměrná nepravděpodobnosti náhodného původu signálu.

tj. $\sim \dfrac{1}{p(0)}$

$$\text{log. věrohodnost} \sim \sum n_k \log n_k - \underbrace{N \log N}_{\text{nemůže mít vliv na věrohodnost}}$$

$$A = \sum n_k \log n_k = -H$$

Věrohodnost je záporná entropie.

(→ Bayesův vztah)

Úvahy o rychlosti přenosu informace.

Hartley: možnost volby mezi S stavy $0 \dots E(S-1)$
(např. S úrovní "baterie")

v nejjednodušším případě, kdy odezva obvodu je po přivedení impulsu exponenciální, bude proud

$$i_t = \frac{n\cdot E}{R}\, e^{-\alpha t} \qquad \text{(tzv. zbytkový proud)}$$

Zb. proud znemožní rozlišení impulsu o menší amplitudě v nějakém pozdějším okamžiku t.

<u>Nejnepříznivější případ:</u>

impulsy s časovým odstupem τ o maximální amplitudě.

celkový rušivý proud
$$i_t = \frac{(S-1)E}{R}\sum_{q=1}^{\infty} e^{-q\alpha\tau} = \frac{(S-1)E}{R}\cdot\frac{1}{e^{\alpha\tau}-1} \quad (*)$$

$$\left(a_1 = e^{-\alpha\tau};\ p = e^{-\alpha\tau};\ \Sigma = \frac{a_1}{1-p}\right)$$

Pokud nemá být zbytkový proud ~~větší~~ než nejmenší proudový impuls, $\frac{E}{R}$

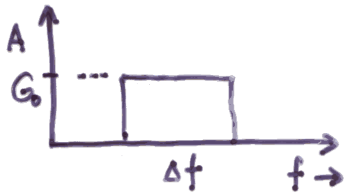
$i_t = \frac{E}{R}$ (**); z (*) a (**) $\quad S = e^{\alpha\tau}$

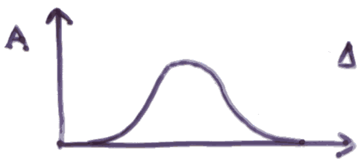
Množství informace v přir. jednotkách bude změněno o $I = n\cdot\log_e S = n\cdot\alpha\cdot\tau$

$I = \alpha\cdot T$

OBECNĚ:

<u>ef. šířka pásma</u>:





$$\Delta f = \frac{1}{G_0}\int_0^{\infty} G(f)\,df$$

zisk na ref. kmit.

zisk

pro pouhou čas. konstantu RC $\quad G(f) = \frac{G_0}{1+\omega^2R^2C^2}$

$$\Delta f = \int_0^{\infty} \frac{1}{1+\omega^2R^2C^2}\,\frac{d\omega}{2\pi} = \frac{1}{4CR} = \frac{\alpha}{4} \qquad (\text{viz} \sim e^{-\alpha t})$$

$I = \alpha \cdot T = 4T\Delta f$

impulsy .... "půlvlnky"  $|\sin x|$

počet půlvlnek – dvojnásobek nejvyšší frekvence

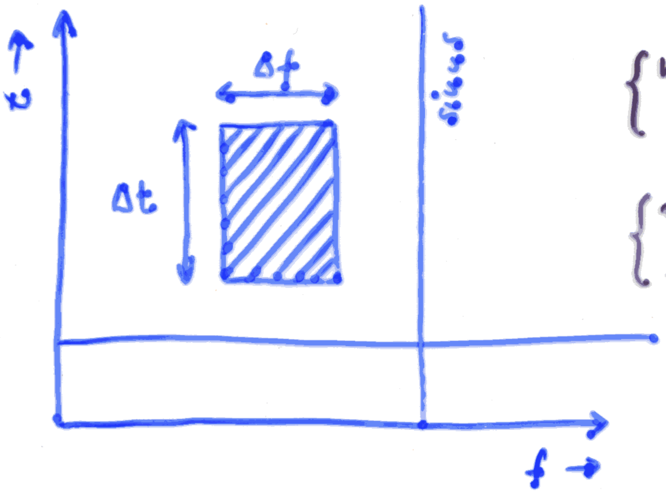
počet impulsů přenesených v pásmu W za dobu T ... 2WT

pro případ telegrafního signálu značka – mezera $I = 2WT$ bitů

GABOROVA VĚTA

vztah kmitočtu a času bez závislosti na speciálních obvodech.

(D. GABOR, Theory of Communication, Journ. Inst. Electr. Engr., 93, Pt III., 1946, p. 429)



{ monochromatická (sin) vlna – harmonické musí mít nekonečné trvání

{ ideální impuls (nulové trvání) – nekonečné frekvenční spektrum

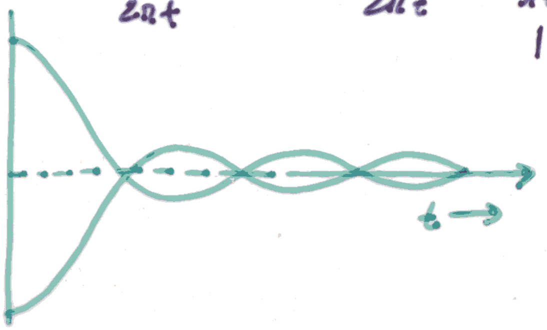
impuls

Reálný sdělovací kanál má kmitočtovou šířku pásma a je v činnosti omezenou dobu: $\Delta f \cdot \Delta t$.

Δf je obdélníková → odezva na ideální impuls:

$$\int_{f_1}^{f_2} \cos 2\pi f t \, df = \frac{\sin 2\pi f_2 t}{2\pi t} - \frac{\sin 2\pi f_1 t}{2\pi t} = \frac{1}{\pi t} \cos \frac{2\pi (f_2 + f_1) t}{2} \sin \frac{2\pi (f_2 - f_1) t}{2}$$

tlumení   nosná   modulace



$t \geq 0$

fyzikálně reálné přiblížení má velké zpoždění, odezva je posunutá a začíná později, než byl vyslán impuls.

Obdobně → k přenosu obdélníkového impulsu o trvání Δt je zapotřebí kmitočtového spektra, které je několikrát širší než pásmo $\frac{1}{\Delta t}$.

Jaký kmitočet má daný harmonický signál?

Kdy se vyskytl signál?

(gaussovský průběh F.T. ⟶ gauss. pr.) harm. signál modulovaný gauss. kř.

<u>Gabor</u>: $\Delta f = \sqrt{2\pi \overline{(f-\bar{f})^2}}$; $\Delta t = \sqrt{2\pi \overline{(t-\bar{t})^2}}$ (***)

$\left(\bar{f} = \int f\, w(f)\, df \Big/ \int w(f)\, df \ldots \right)$ $w(f)$ - výkonové spektrum.

$\Rightarrow \Delta f \cdot \Delta t \geq \frac{1}{2}$ (přesná obdoba Heisenbergovy relace neurčitosti)

$= \frac{1}{2}$ pokud časový průběh signálu má tvar gauss. kř.

Chceme-li daným sdělovacím kanálem, jehož šířka pásma je definována v (***), přenést co největší počet nezkreslených impulsů, musíme vysílat impulsy ve tvaru Gaussovy křivky. Celkové trvání zprávy lze rozdělit na jednotlivé časové úseky Δt, jejichž počet bude $N = \frac{T}{\Delta t}$, dosadíme-li za Δt ze vztahu $\Delta f \cdot \Delta t = \frac{1}{2}$.

Potom $N = 2T\Delta f = 2TW$

<u>Věta o výběru</u> (Shannon)

Požadavek dvou impulsů na každý cykl modulačního kmitočtu je nutnou a zároveň postačující podmínkou:

Jestliže $f(t)$ neobsahuje kmitočty vyšší než W Hz, pak je plně určena svými hodnotami v řadě bodů, vzájemně vzdálených $\frac{1}{2}W^{-1}$ sekund.