Hibakorlátozó kódolás

Az információtovábbításnak egy sajátosan érdekes területe a kódolás, amely egyrészt arra irányul, hogy a továbbítás során fellépő hibaforrások által okozott hibákat felismerjük, vagy esetleg kijavítsuk, de másrészt ezt úgy is megfogalmazhatjuk, hogy általa az információtovábbítás sebességét közelíthetjük az elvileg lehetséges maximumhoz, azaz a csatornakapacitáshoz.

Az előbbi megfogalmazást illetve: az nyilvánvalóan előnyös, ha meg tudjuk állapítani, hogy az információ nyelőjéhez megérkező üzenet azonos-e a forrás által kibocsátottal, vagy még ennél is kedvezőbb, ha eltérés esetén ki tudjuk javítani a hibát. Viszont mi köze van ennek a csatornakapacitáshoz? A válasz nagyon egyszerű: példaként gondoljunk egy digitális, memóriamentes csatornára (pl a megvizsgált BSC-re), ahol a kapacitás a szimbólumtévesztési valószínűség függvénye. Amennyiben csökken a tévesztés valószínűsége, nő a kapacitás.

Persze az előbbi példát tekintve a kapacitás akkor is nő, ha a digitális csatorna belsejében (ahol analóg jelek írják le a digitális szimbólumokat) növeljük a jel-zaj viszonyt. A kódolással kapcsolatos alapkérdés ezért úgy is megfogalmazható, hogy mikor, milyen feltételek esetén, milyen módszerekkel lehet növelni általa a csatornakapacitást, és ezen belül azt is meg kell válaszolni, hogy mi ennek az ára.

A válaszok megadásához induljunk ki a folytonos csatornamodellre kapott összefüggésből:

$$C_f = W \cdot \log_2\left(1 + \frac{S}{2 \cdot W \cdot 2\pi \cdot s_0}\right),$$

ami azt mondja, hogy a kapacitás növelése alapvetően két módon lehetséges: vagy növeljük a jel-zaj viszonyt, vagy növeljük a jel sávszélességét. Az előbbi hatása triviális, bár már elmondtuk, hogy alapvetően korlátos, az utóbbiról pedig kimutattuk, hogy kezdetben gyors változást eredményez, majd lelassulva egy felső korláthoz tart.

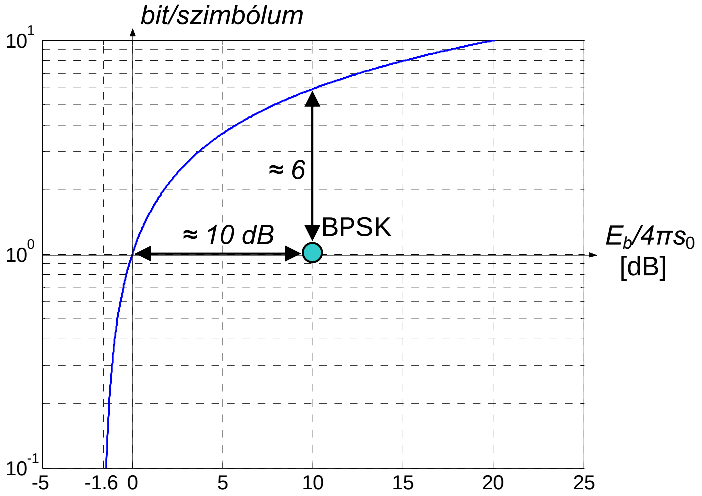
Nagyon fontos azonban arra figyelni, hogy az információt továbbító jel sávszélességének, méghozzá **szükséges** sávszélességének a növeléséről van szó. Ehhez természetesen elsősorban az kell, hogy fennálljon a nagyobb sávszélességű jel továbbításának lehetősége. Másodsorban pedig olyan sávszélesség-növelő módszer kell, ami hatékony a kitűzött cél elérése szempontjából. Ez a módszer a hibajavító kódolás. Ugyanis az információs szimbólumok kiegészítése redundáns *paritásszimbólumokkal*, nagyobb sávszélességű jelet eredményez változatlan információtovábbítási sebesség esetén.

Ez a módszer hatékonynak bizonyult azokban az esetekben, amelyeknél az alkalmazott csatorna sávszélességének csak technikai korlátai voltak, de nem állt fenn sem a „szomszédok" zavarása, sem más rendszerekkel való együttélésből származó korlát. Röviden: a nagytávolságú űrszondák kommunikációja a tipikus példa erre, ahol a teljesítmény erősen korlátozott, de a jel sávszélességének szinte nincsenek korlátai. Erre a célra szolgáló kódokat fejlesztették ki először, és itt is ezt ismerjük meg előbb.

Nagyon érdekes és izgalmas kérdés azonban az is, hogy mit lehet tenni, ha korlátoz a sávszélesség, viszont a teljesítménnyel kicsit szabadabban bánhatunk, bár korántsem korlátlanul. Van-e a kódolásnak keresnivalója ebben az esetben? Ha nem növelhető a sávszélesség, akkor csak az az út járható, hogy növeljük a továbbításra kerülő szimbólumkészlet méretét. Változatlan tévesztési arány eléréséhez ekkor nagyobb átlagteljesítményre van szükség. A *kódolás* csak úgy képzelhető el, hogy a használt szimbólumkészletet túlméretezzük, azaz a lehetséges szimbólumok közül nem mindegyiket vesszük igénybe az adás során, *(???)* amelyek a vételnél felismerhetőek. Persze felmerül a kérdés, hogy jobb eredményt kapunk-e így, szemben azzal, ha minden szimbólumot felhasználunk, de növelt teljesítménnyel (kevesebb van, így több jut egyre azonos átlag mellett), és így kisebb a tévesztési valószínűség. Ezt a kérdést fogjuk a második lépésben megválaszolni.

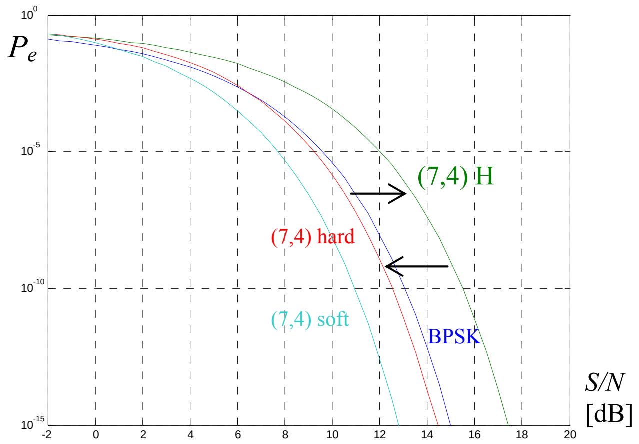
Kódolás korlátozott átlag-teljesítménynél

A kódolás céljának egyik megfogalmazásaként azt adtuk meg, hogy általa közelíthetjük az információtovábbítási sebességet a maximálisan lehetséges csatornakapacitáshoz. Az alábbi ábra segítségével röviden rámutathatunk arra, hogy mennyiségileg mekkora lehetőség rejlik a kódolásban a teljesítménykorlátos esetben, amikor elsősorban bináris szimbólumokat használunk:



Az ábra szerint az elérhető *kódolási nyereség* közel 10 dB, illetve mintegy hatszoros, attól függően, hogy a jel-zaj viszony lehetséges csökkentésének irányában, vagy az átviteli sebesség növelésének irányában akarunk eljutni a határhoz. Az előbbi azt jelenti, hogy elvileg 10 dB-el kisebb jel-zaj viszony mellett is elérhető a 10^{-6} bithibaarány, ami az ábrában lévő BPSK jelű pontra áll fenn, az utóbbi pedig azt mondja, hogy a 10 dB-es jel-zaj viszony esetén elvileg akár 6 bit információ is továbbítható 10^{-6} bithibaarány mellett.

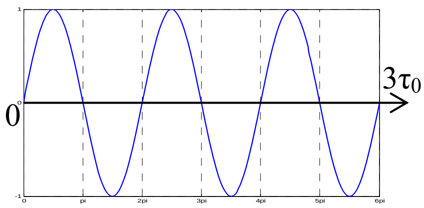
A kódolással elérhető kódolási nyereség alatt általában a megegyező bithibaarány mellett lehetséges jel-zaj viszony csökkentést értik. Tanulságos annak végiggondolása is, hogy miként éri el ezt a kódolás. Először vegyük észre, hogy a kódolással előbb rontunk a helyzeten, és csak a dekódolás után érhetünk el javulást. A mellékelt ábrán azt szemléltetjük, hogy miként változik a BPSK hibaarány görbéje, amennyiben az üzenet K bitjéhez N-K bitet a kódolás révén hozzáadunk, azaz egy $R=K/N$ sebességű, vagyis (N,K) paraméterű bináris kódot használunk. A nyíl azt mutatja, hogy pl. R=4/7 esetén a görbe ≈ 2,5 dB-el elmozdul jobbra, amire azt szokták mondani, hogy ez a „büntetés” (penalty) a kódolásért. Ahhoz, hogy nyereség legyen, először le kell győzni ezt a büntetést, illetve még azon is túl kell jutni. Ez történik meg a dekódolás során, amikor a megfelelően tervezett, hatékony kód képességei révén a hibákat sikerül kijavítani, és a hibavalószínűség kisebb lesz,

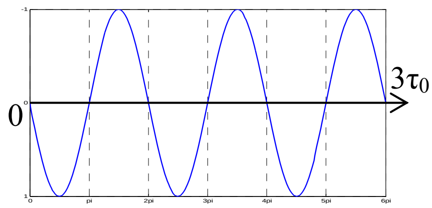


mint a kódolatlan esetben volt. Az alsó, balra mutató nyíl mutatja a lehetséges javulást, amelyet a példában alkalmazott (7,4)-es Hamming kódra elérhetünk. (Ez a kód, mint ismert blokkonként 1 hibát tud javítani, és könnyű kiszámítani a dekódolás után hibátlan blokkok valószínűségét. Ennek 4/7 része lesz a *maradék* bithiba-valószínűség.) Érdekes megfigyelni, hogy nem minden jel-zaj viszony mellett érünk el javulást, hanem csak kb 7 dB-nél nagyobb értékekre, és nem is sokat (≈ 0,4 dB lesz 10^{-6} hibavalószínűségnél).
Végül ismét emlékeztetünk, hogy a kódolási nyereség a kódolatlan és a javított görbék különbsége, ami általában (miként a példában is) nőhet a hibavalószínűség csökkenésével.
Ha kötött a bitenkénti energia (adott információtovábbítási sebesség és adott átlagteljesítmény esetén), akkor megegyező forrás- és kód-abc melletti kódolásnál kevesebb energia jut egy-egy csatornába küldött bitre, mint a kódolás nélküli esetben, és ezáltal megnő a csatornában továbbított bitek tévesztési valószínűsége. Az arány a kódsebesség lesz $R=K/N$. A dekódoló viszont kijavít a keletkező hibákból annyit, amennyit a kód képességei lehetővé tesznek, és így a maradék hiba kisebb **lehet**, mint a kódolatlan átvitelé, azaz létrejöhet a *kódolási nyereség*.

Kódszavak optimális detektálása

Az optimális detektálásról eddig megismertek továbbfejlesztésre szorulnak kódolás esetén. (Ezt említhettük volna már a részleges válaszfüggvényű átvitelnél is.) A kódolás révén keletkező szimbólumsorozatokban megszűnik az egyes szimbólumok függetlensége. Azt is mondhatjuk, hogy belső kapcsolatok jönnek létre a szimbólumok között. Ennek következtében nem az eredményezi a legjobb megoldást, ha az egyes **szimbólumokat** egymástól függetlenül a legvalószínűbben detektáljuk, hanem az egyes **kódszavak** legvalószínűbb detektálása adja a legjobb eredményt. Itt egy nagyon egyszerű példa keretében mutatjuk meg a viszonyokat, és az elérhető eredményeket.
Vizsgáljuk meg azt az esetet, amikor (3,1)-es Hamming kódot használunk egy BPSK csatornán, maximált (konstans) átlagteljesítmény mellett! Ekkor mindegyik adatbitre 3 kódbit jut, és ez változatlan jelteljesítménynél 1/3-ra csökkenti a csatornán a bitenergiát. Az adott kódra két lehetséges kódszó a (0 0 0) és az (1 1 1), amelyeket a BPSK moduláció esetén az alábbi két jel reprezentál az analóg csatornán:

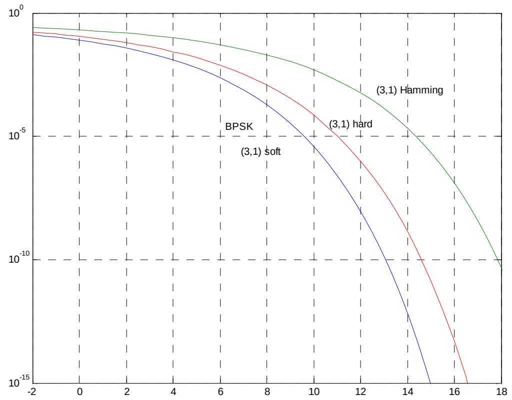




A zajos, kódolt csatorna kimenetén a legjobb döntést úgy kapjuk, ha a fenti jelekkel szorzunk, és a szozatok $3\tau_0$-ra vett integráljai közül a nagyobbat fogadjuk el, szemben azzal, hogy τ_0 időrésenként döntünk, majd a kapott háromelemű bináris sorozatokból következtetünk a legvalószínűbb adatbitre. Ez utóbbi neve *kemény döntés*, az előbb javasolt pedig *lágy döntés*, mivel az egyes időrések végén, a szorzás-integrálás révén kapott értéket vittük tovább a következő időrésre, nem pedig egy annak megfeleltethető bináris szimbólumot. Így bizonytalan (a döntési küszöb közeli) esetekben sokkal jobban járhatunk. Könnyű kiszámítani ebben az elemi példában a javasolt lágy döntéssel, vagy másképpen szemlélve, a teljes kódszóra együttesen végzett döntéssel elérhető hibavalószínűséget:

$$P_e = 0{,}5 \cdot \mathrm{erfc}\left(\sqrt{\frac{3E_b/3}{4\pi \cdot s_0}}\right),$$

ami megegyezik a kódolatlan BPSK hibavalószínűségével. A gyök alatti első 3-as a $3\tau_0$ alatt felhalmozódó energiát képviseli, míg az 1/3-ad a kódolás miatti energiacsökkenést adja meg. Ha erre az elemi példára megismételjük a korábbi, (7,4)-es kódra készített ábrát, akkor egyrészt önmagában érdekes eredményt kapunk, másrészt pedig megtesszük az első lépést a korábbi ábra eddig nem említett görbéjének a magyarázatára. A mellékelt ábra szerint a (3,1)-es Hamming kód konstans átlagteljesítmény esetén célszerűtlen, mert kemény döntésre ront a hibavalószínűségen, míg lágy döntés esetén éppen csak a kódolási veszteséget tudja behozni. Ezzel szemben már egy (7,4)-es Hamming kód is képes nyereséget elérni akár kemény döntésre is, de lágy döntéssel már 1 dB körüli jel-zaj viszonytól nyereség lép fel, amint azt a korábbi ábra balszélső görbéje mutatja, és ez a nyereség ≈ 2,4 dB lesz 10^{-6} hibavalószínűségnél. Most már csak az szorul magyarázatra, hogy miként lehet meghatározni a hibavalószínűséget lágy döntés esetén. Ehhez a jeltér nyújt segítséget.



Szimbólumsorozatok a jeltérben

Amennyiben a sorozatban használt egyes szimbólumoknak a jeltérbeli reprezentációját megoldottuk (amint ezt megtettük a legfontosabb modulációs módszerekre), könnyen tudunk olyan ortonormált jelteret konstruálni, amelyben a sorozatok megjeleníthetők.

Korlátlan sávszélesség esetén

Az egyes szimbólumokat képviselő elemi jelek egy-egy időrésre korlátozottak, így az időben egymást követő szimbólumokat képviselő, időben időrésnyivel eltólt elemi jelek ortogonálisak, a jeltér pedig annyi dimenziós lesz, mint a szimbólumsorozat hossza. Így például a (3,1)-es Hamming kódra tudunk szerkeszteni egy háromdimenziós jelteret, amelyet még ábrázolni is lehet. Ha BPSK-t tételezünk fel a csatornán, akkor a jeltér bázisfüggvényei a következők lesznek:

$$\varphi_1(t) = \sqrt{\frac{2}{\tau_0}} \sin\left(\frac{2\pi}{\tau_0} \cdot t\right), \quad 0 \le t < \tau_0$$

$$\varphi_2(t) = \sqrt{\frac{2}{\tau_0}} \sin\left(\frac{2\pi}{\tau_0} \cdot t\right), \quad \tau_0 \le t < 2 \cdot \tau_0$$

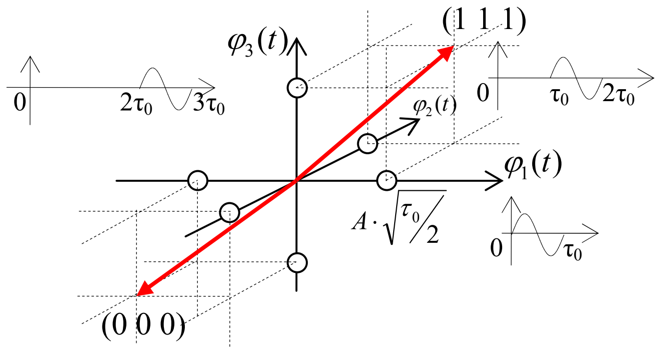
illetve

$$\varphi_2(t) = \varphi_1(t - \tau_0)$$

és

$$\varphi_3(t) = \varphi_1(t - 2 \cdot \tau_0)\,.$$

Az egyik szimbólumot $y_1 = A \cdot \sqrt{\tau_0 / 2}$ vektor írja le, míg a másikat $y_0 = -y_1$, amivel az (1 1 1) kódszó a jeltrében (y_1 y_1 y_1) lesz, a másik pedig ($-y_1$ $-y_1$ $-y_1$). A két vektort a jeltérben az alábbi ábra mutatja:



A két sorozatot (kódszót) képviselő vektorok hossza $\sqrt{3}$-szorosa az $y_1 = A\cdot\sqrt{\tau_0/2}$ „egységnek", így két kódszóvektor végpontjainak euklédeszi távolsága a jeltérben ennek kétszerese. A hibavalószínűséget pedig (bináris esetben pontosan) a jeltérbeli szomszéd távolságának a fele határozza meg, ami a korábban (más megfontolással) kapott érték lesz:

$$P_e^{(3,1)s} = 0{,}5\cdot \operatorname{erfc}\left(\frac{\sqrt{3}\cdot\sqrt{E_b/3}}{\sqrt{4\pi\cdot s_0}}\right).$$

Milyen következtetést vonhatunk le ebből a (7,4)-es Hamming kódra vonatkozóan? A Híradástechnikából tudjuk, hogy a Hamming kódok olyan *perfekt* kódok , amelyek egy hibát képesek javítani, és így a kódtávolságuk (a kódszavak közötti legkisebb Hamming távolság) három, pontosabban valamennyi szomszédos kódszónak három a távolsága. Ha erre a kódra megpróbáljuk elképzelni az előbb bemutatott jelteret, akkor az 7 dimenziós tér lesz, az eltólt elemi jelekkel, viszont a kódszavak képviselő jelvektorok nem lehetnek közelebb egymáshoz, mint az előbbi egyszerű példában. Ha feltételezzük, hogy a jeltérben akár az összes irányban lehet *effektív szomszéd*, akkor a hibavalószínűség felső korlátjaként az alábbit kapjuk:

$$P_e^{(7,4)s} = 6\cdot 0{,}5\cdot \operatorname{erfc}\left(\frac{\sqrt{3}\cdot\sqrt{4E_b/7}}{\sqrt{4\pi\cdot s_0}}\right),$$

ahol 6 az effektív szomszédszám, $\sqrt{3}$ az euklédeszi távolság fele, és 4/7 a bitenergia csökkenése a kódolás miatt. A kapott görbét tehát a (7,4)-es példa görbéi közül a balszélen találjuk.

Megjegyzés: Felmerülhet az a gondolat is, hogy a teljesítménykorlát mellett az átviteli sebesség csökkentése által is növelhetjük a bitenergiát, és ezzel a jel-zaj viszonyt. Azonban fontos észrevenni, hogy ettől nem nő az egységnyi sávszélességben továbbított információ mennyisége.

Kódolás korlátozott sávszélességnél

Ha növelhetjük a teljesítményt is

Mit lehetne elérni (3,1)-es Hammingre 3-szoros teljesítménnyel, vagy kódolás nélkül 3-szoros teljesítménnyel?

Meg kell különböztetni:

- *Kódolás: művelet, amely kódolt szimbólumsorozatot állít elő az adatsorozatból. A leírás matematikai modell alapján. Pl mátrix-művelet, polinom-művelet.*
- *Kódoló: az előbbi művelet megvalósítási módja. Hardware, vagy software leírása. Pl tömbvázlat vagy algoritmus.*

- *Kód:a lehetséges (kódolási szabály által meghatározott) kódolt sorozatok összessége.*