

Ideiglenes verzió

INFOKOMMUNIKÁCIÓ

FRANKÓ ATTILA ERNŐ

Előszó

Mikor ennek a könyvnek az írása felvetődött bennem, az a cél lebegett a szemem előtt, hogy egy színvonalas, modern, színes, áttekinthető, de részletes jegyzet szülessen a Budapesti Műszaki Egyetem Villamosmérnöki és Informatikai Karán oktatott Infokommunikáció tárgyhoz. Mivel maga a képzés és így a tantárgy tematikája is időről időre átalakul, a korábbi szakmai anyagok és tankönyvek mind tartalmilag, mind struktúráisan egyre kevésbé illeszkednek az új oktatási sémába. Éppen ezért, ez a könyv kizárólag digitális formában jelenik meg, így szükség esetén átdolgozható, de mindenekelőtt bővíthető, hogy minél hosszabb ideig releváns maradjon, alapot szolgáltatva a tananyag elsajátításához.

Írás közben döbbsentem rá arra, hogy mennyi mindent nem tudunk elmondani a tárgy keretein belül. Ezen el nem mondott ismeretmorzsák egy részét a képzés (vagy a mesterképzés) többi alapozó-, illetve szakirányos tárgyaiban elszórva oktatják, más részét egyáltalán nem oktatják. Természetes, hogy az idő előrehaladtával, ahogy a szakmán belül új szakterületek jönnek létre, úgy a hallgatóknak is egyre több szakterület alapjait kell megismerni. Ez azt jelenti, hogy az egyes szakterületek megismerésére kevesebb idő jut, így az átadható tudás mértéke is csökken. Ez ellen harcolni – véleményem szerint – nem érdemes és nem is szabad, azonban meg kell teremteni annak lehetőségét, hogy a tématerület iránt érdeklődők, illetve más szakterület művelői is részletesebben megismerhessék egy adott tématerület alapjait. Éppen ezért, a könyv későbbi kiadásiban (a jelenlegiben csak minimálisan) nagy szerephez jutnak a kiegészítő anyagrészek, amelyek a tárgy keretein belül nem oktatott, de az infokommunikáció alapjait részletesebben feltáró anyagrészek. Így, ezt a könyvet nem csak a tárgy hallgatóinak ajánlom, hanem bármely érdeklődőnek, aki szeretne megismerkedni az infokommunikáció alapjaival.



Figyelem! A könyv jelenlegi verziója nem fedi le a teljes anyagot! A könyv elolvasása az előadásokon való részvételt nem helyettesíti! A számonkérésekben az előadáson elhangzott ismereteket kérjük számon!



Kérlek légy tekintettel a környezetre, a könyv kinyomtatása előtt!



A könyvvel kapcsolatos, bármilyen jellegű észrevételt a *franko@tmit.bme.hu* e-mail címre lehet küldeni!



A pictogramok a Freepik, a Vectors Market, illetve mavadee munkái. A borító eredetije vector_corp munkája.



Ezúton köszönöm a szakmai lektorok munkáját, akik átnézték és véleményezték munkámat. Kívánom, hogy tegyen ez a könyv jó szolgálatot valamennyiüknek, oktatási tevékenységük során.



Továbbá, köszönöm Buda Szabinának, hogy végig támogatott a könyv írása során, átnézte azt, kijavított több, mint kétszáz elgépelést és helyettem is odafigyelt a folyadékháztartásomra.

Szakmai Lektorok:

- Dr. Orosz Péter
- Dr. Osváth László
- Dr. Varga Pál

Tartalomjegyzék

1	Bevezetés az infokommunikációba	1
1.1.	A távközlés összetevői	1
1.1.1.	Az átviteli lánc elemei	2
1.1.2.	A híradástechnika és az információelmélet története	4
1.2.	ICT – Az infokommunikáció jelene	5
1.2.1.	Az ICT konvergencia fázisai	6
1.2.2.	ICT konvergencia jövője	7
2	Jelelméleti alapok	9
2.1.	Determinisztikus jelek	9
2.1.1.	Effektív érték	10
2.1.2.	Csúcsstényező	11
2.1.3.	Periodikus jelek Fourier-sora	12
2.1.4.	Periodikus jelek spektruma	14
2.1.5.	A logaritmikus skála használata	15
2.2.	Nem periodikus jelek	18
2.2.1.	Tranziens jelek Fourier-transzformáltja	18
2.2.2.	A Dirac-impulzus	19
2.3.	Sztochasztikus jelek	20
3	Mintavételezés és kvantálás	21
3.1.	Mintavételezés és jelrekonstrukció	21
3.1.1.	Elméleti mintavételezés	21
3.1.1.1.	Dirac-fésűs mintavételezés	22
3.1.1.2.	A mintavételi tétel	24
3.1.1.3.	Jelrekonstrukció	26
3.1.1.4.	Alul és túlmintavételezés	27
3.1.1.5.	Aliasing és szivárgás	29
4	Kódolástechnika	32
4.1.	Az információelmélet alapjai	32
4.1.1.	Diszkrét források leírása	33
4.1.2.	Az entrópia	34

4.2.	Forráskódolás	35
4.2.1.	Kódolási alapfogalmak	36
4.2.2.	Prefix-kódok	37
4.2.3.	Optimális kódok	39
4.2.4.	Bináris Huffman-kód	40
4.3.	Hibajavító kódolás	41
4.3.1.	Kódok hibajavító képességei	41
4.3.2.	Bináris hibajavító kódok	43
4.3.2.1.	Bináris, lineáris kódok előállítás	44
4.3.2.2.	Bináris kódok hibajavítási mechanizmusa	45
5	Analóg modulációs eljárások	48
5.1.	Amplitúdómoduláció	48
5.1.1.	Kétoldalsávós amplitúdómoduláció	49
5.1.1.1.	AM-DSB/SC moduláció	49
5.1.1.2.	AM-DSB moduláció	52
5.2.	Szögmodulációk	54
5.2.1.	A pillanatnyi frekvencia	55
5.2.2.	Fázis-, és frekvenciamodulált jelek konstruálása	56
5.3.	Rádiós műsorszóró rendszerek	58
5.3.1.	Szuperheterodin vevők	58
5.3.2.	AM rádióadás	60
	Irodalomjegyzék	62

1. fejezet

Bevezetés az infokommunikációba

Ha valamivel kapcsolatban olyan közhelyeket lehet puffogatni, mint: „*napjaink egyik legdinamikusabban fejlődő ágazata*”, vagy a klasszikus „*meghatározza a mindennapjainkat*”, akkor felmerül a gyanú, hogy az érintett területtel érdemes foglalkozni, annak ellenére, hogy egyesek szerint a villamosmérnök képzés egy vidámabb hely lenne e tárgy oktatása nélkül. Eltekintve a közhelyektől, mérnöki szempontból is helytálló, hogy egyre több és több műszaki megoldás és mérnöki alkotás tartalmaz valamilyen kommunikációs technológiát, miközben a témakör egyébként is széles határai egyre sebesebben tágulnak. Ez a hatalmas növekedés persze annak is köszönhető, hogy inkább egy folyamatról beszélünk, mint sem hagyományos értelemben vett szakterületről, ugyanis az **infokommunikáció**, az informatika és a távközlés konvergenciájának tekinthető. Annak idején, a betárcsázós internet (dial-up) korában, ha böngészni támadt kedvem, meg kellett várnom míg édesanyám abbahagyja a beszélgetést a vezetékes telefonon. Azóta eltelt több mint 20 év: nekem vezetékes telefonom sincs, anyuka meg telefonálás helyett inkább az Ali-expressen vásárol, de ha mobilon felhív, akkor az egy VoLTE¹ alapú hívás lesz – hacsak nem Skype-on hív. Ez a 20 év alatt bekövetkezett változás kiválóan szemlélteti, hogy miként fonódott össze az informatika és a távközlés, illetve, hogy a folyamat jelentősége mennyire alapvető.

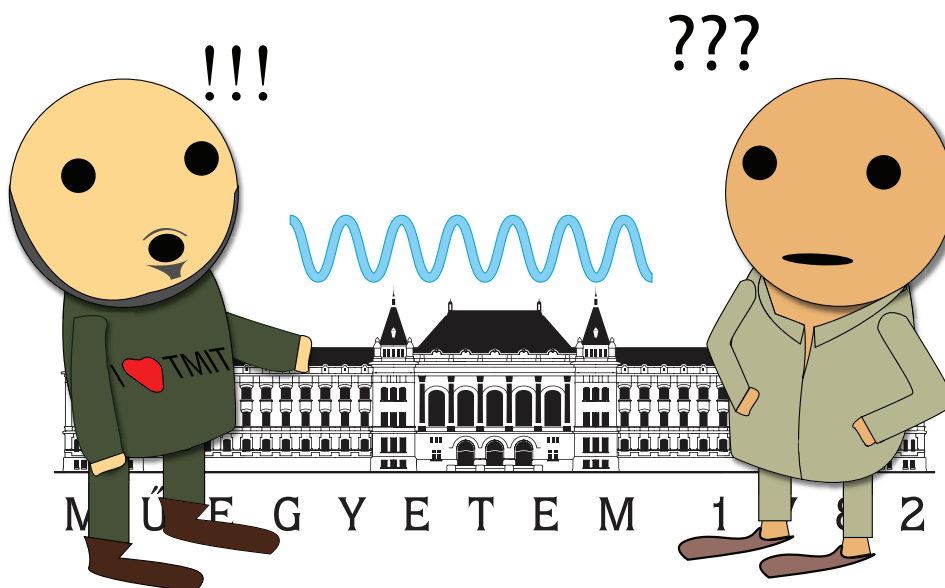
1.1. A távközlés összetevői

A **távközlés** vagy **telekommunikáció** célja röviden: információátvitel különböző – jellemzően távoli – pontok között a térben. Ehhez egészen hasonló az információátvitel problémája, ahol az információ helye változatlan, hisz az információt az ”időben” szeretnénk átvinni. Az átvitt információ jellege sokféle lehet: hagyományosan ide tartozik a beszédátvitel (telefon), a hang- és képátvitel (rádió, televízió), illetve a primitívebb szöveg,- és adatátvitel (távíró), azonban a távközlés digitalizálódása és így a fizikai tartalom uniformitása miatt ma már megszámolni is nehéz, hány különböző típusú információ vihető át a közegen. Példának okáért: egy továbbított csomag lehet egy beszédszegmens, képkocka egy videóból, de akár egy egyszerű szöveg (string) vagy bináris adat is, amelyet fizikailag ugyanaz a bytesorozat reprezentál.

Az átvitel egyszerűsített modellje megegyezik a kommunikáció általános modelljével, mint amilyen például az élő szóban folytatott beszélgetés: van egy adó (aki beszél), egy vevő (aki hallgatja), köztük a közeg, amelyen átvisszük az információt (például levegő vagy egy kábelben

¹Voice over LTE

a vezető anyag), illetve maga a hordozó amely szállítja az információt (akusztikai hullámok, elektromágneses hullámok). Ez a modell látható az 1.1 ábrán is.



1.1. ábra. Meglepődött villamosmérnök hallgató az Infokommunikáció vizsga nehézségeit eseteli kollégájának, aki nem érti miről beszél.

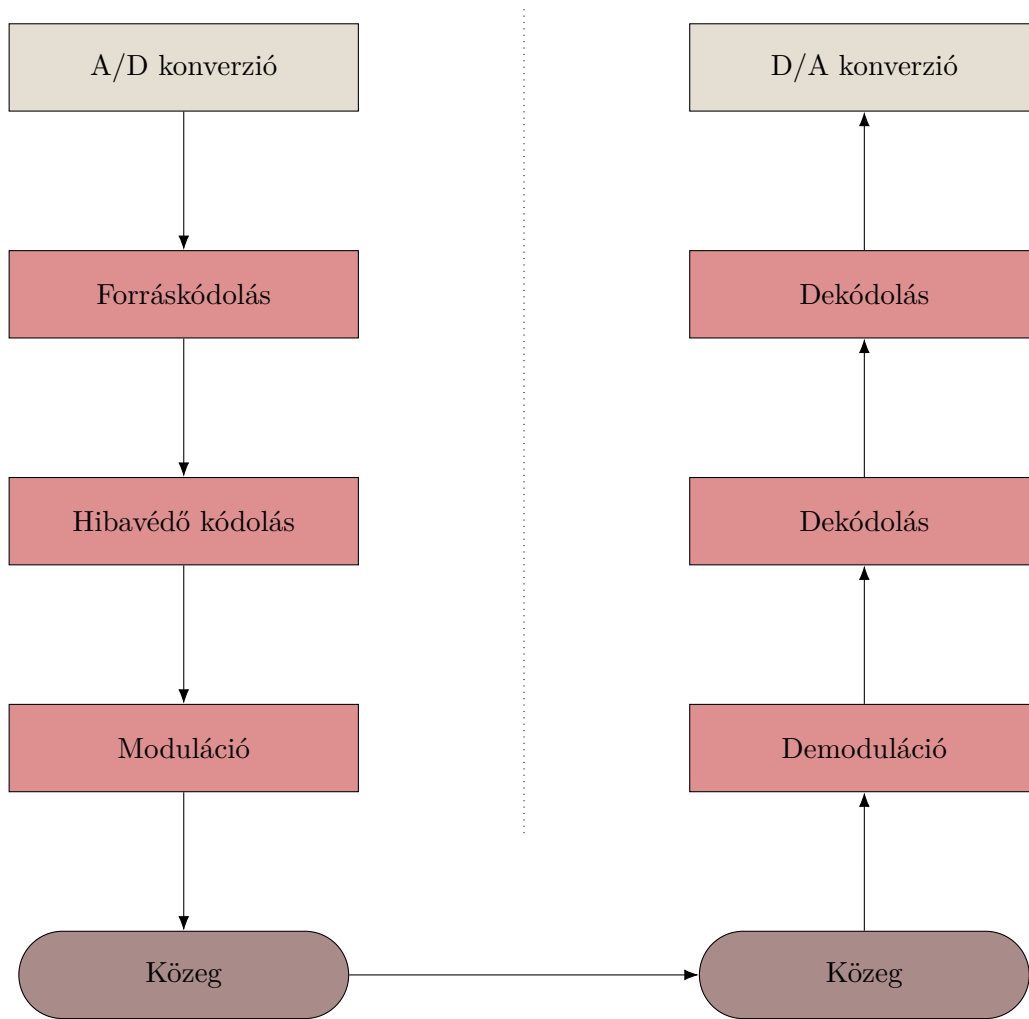
A teljes távközlési átviteli láncban számos elem található, amelyek igény szerint továbbiakra bonthatók fel. A tárgyban, az 1.2 ábrán látható – általánosan használt – távközlési modellt alkalmazzuk. A modell az információt fizikailag hordozó jel útját követi végig a forrástól kezdve a közegig, amelyen átvisszük azt. Az átviteli lánc – a fenti absztrakt blokkokat használva – szimmetrikusnak tekinthető, tehát a vételi oldalon az átviteli lánc elemei fordított sorrendben található meg. Az esetek jelentős részében mindkét fél egyaránt ad és vesz, tehát a szerepek folyamatosan cserélődnek, ez esetben a végpontokat adóvevőnek nevezzük.

1.1.1. Az átviteli lánc elemei

Az első elem az átviteli láncban az analóg-digitális átalakító, valójában opcionális, pontosabban alkalmazásfüggő. Amennyiben analóg jellel reprezentált információt szeretnénk átvinni, mint amilyen a hang is, azt először célszerűen átalakítjuk egy digitális jelfolyammá. Ennek számos oka van, ezek közül a fontosabbak:

- A digitális jel zajtűrése messzemenően jobb, mint az analóg jelé.
- A digitális jel könnyen és megbízható módon tömöríthető, módosítható illetve tárolható.
- A különböző titkosítási és autentikációs eljárások csak digitális jeleken értelmezhetők és alkalmazhatók.

Amennyiben az információ közvetlenül digitális forrásból származik, úgy nincs szükség erre az átalakításra. Itt érdemes megjegyezni, hogy a fent vázolt modell természetesen csak a mai



1.2. ábra. Tipikus távközlési átviteli modell

értelemben vett távközlésre, vagyis az infokommunikációs megoldásokra érvényes, a korábbi (analóg korszakbeli) telekommunikációs módszerekre nem.

A következő elem a forráskódolás, melynek célja a redundancia csökkentése, tehát a tömörítés, annak érdekében, hogy minél rövidebben tudjuk leírni az átvinni kívánt információt. Így csökkenthető az átvitel időtartama (amit úgy érzékelhetünk, hogy nő az átviteli sebesség, azonban ez műszaki értelemben nem igaz), illetve ezáltal a hálózat terheltsége. Amennyiben titkosítani szeretnénk az átvitt információt, azt jellemzően a forráskódolás után szoktuk megtenni, így a modellben a következő elem lehetne akár a titkosító is, azonban a tárgy keretein belül erről nem lesz szó.

Miután a forráskódolás során csökkentettük vagy akár meg is szüntettük a jelfolyamban lévő redundanciát, meglepőnek tűnhet a következő lépés. A csatornakódolás vagy hibavédő kódolás célja ugyanis, hogy tervezett redundanciát adjon a jelfolyamhoz. Tegyük fel, hogy a vevő a "vad és veszélyes pŁnă" üzenetet vette, mivel az átvitel során a jelfolyam torzult, így a "pu-ma" szó megfelelő karakterei helyett valami egészen más jelenik meg. Azonban ha hibavédő kódolást alkalmazunk, akkor a vevő a beérkezett jelfolyam alapján meg tudja állapítani, hogy

történt-e a hiba az átvitel során (ekkor újra lehet küldeni), illetve bizonyos esetekben még azt is, hogy hol. Ilyen hibadetekciós célokat szolgál a Digitális Technika 2 tárgyból megismert soros kommunikáció (UART) paritás bitje is.

Az átviteli lánc utolsó eleme a moduláció, amely tulajdonképpen egyfajta interfészként is felfogható az átviteli közeghez. Célja, hogy az információt hordozó jelfolyam fizikai reprezentációját megváltoztassa úgy, hogy az alkalmassá váljon az adott közegen való továbbításra. Maga az elv egyszerűen megérthető az analóg modulációk példáján: Jellemzően alacsony frekvenciás – úgynevezett alapsávi – jelünk nem alkalmas arra, hogy rádióhullámként kisugározzuk egy antennán a szabadtérbe, erre a célra nagyfrekvenciás jelet – más néven vivőt – kell használni. A moduláció valójában a vivő jel valamely paraméterét (pl.: amplitudó, frekvencia) változtatja meg az alapsávi jel függvényében, így a közegen könnyen továbbítható nagyfrekvenciás jel lesz a továbbiakban az információ fizikai reprezentációja (technikailag az alapsávi jelé). Fontos megjegyezni, hogy a modulációt nem csak akkor használunk, ha a jel és a fizikai közeg jellemzői összeegyeztethetők, hanem többszörös közeghozzáférés esetén is (pl.: CDMA). A modulációs technikákról egy későbbi fejezetben írunk részletesen.

A fenti modellben található egységek műszaki tervezésének és működésének kérdését, illetve az eljárások matematikai hátterét alapvetően két, egymással szorosan összefonódó szakterület a *híradástechnika* illetve a *információelmélet* tárgyalja.

1.1.2. A híradástechnika és az információelmélet története

Hagyományosan, a távközléssel, híradástechnikával foglalkozó könyvek röviden áttekintik annak útját, fejlődését, így ezt most mi is megteesszük. A távközlés története valójában az emberiség történetével egyidős, hisz az embereken mindig is megvolt az igény az információ továbbításra. A korai kommunikációs formák közé tartozik a tűz, a fényjelzés, füstjelek, illetve a dobok és kürtök ilyen célú felhasználása. Ezek a módszerek természetesen korlátosak voltak, hisz csak meghatározott üzeneteket tudtak átvinni. A később megjelenő levélküldés, illetve optikai távírás (szintén fényjelzéssel) a meglévő kommunikációs lehetőségeket jelentősen kitágították.

A híradástechnika történetét tulajdonképpen az elektromosság távközlési célú felhasználásától számítjuk: 1838-ban Morse bemutatja a telegráfot (távíró), amellyel elérhetővé vált az adattovábbítás nagy távolságokra, méghozzá a korábbi mechanikai, optikai megoldásoknál jóval olcsóbban. Később a szikratávíró, a telefon, illetve a televízió megjelenése hozott forradalmi változást a távközlésben, míg a legutolsó igazán jelentős változást a modern számítógépek, illetve számítógép-hálózatok megjelenése jelentette. Tulajdonképpen a fenti technológiákat kisebb vagy nagyobb részben kiváltotta az internet feletti kommunikáció valamely formája, így manapság a híradástechnika is a digitális átvitel irányába fordult. Fontos látni azonban, hogy végsősoron a fizikai rétegben minden jel analóg, így a fenti technológiák többsége bár elavult, de a modern technológiák részben ezekre is építenek.

A fenti évszámok közül külön kiemelendő az 1948-as év, amikor is Shannon *"A Mathematical Theory of Communication"* munkájával megalapozta az információelméletet. A mű eredeti célja annak bizonyítása volt, hogy az átviteli hiba valószínűsége közel zérusra csökkenthető, bármely átviteli sebesség esetén, ha ez utóbbi kisebb mint az átviteli csatorna kapacitása. Ez a kijelentés

A Híradástechnika Történetének Fontosabb Állomásai

- 1838 • Morse bemutatja a távírót
- 1858 • Az első transz-atlanti távíróvezeték üzembehelyezése
- 1876 • Alexander Graham Bell feltalálja a telefont
- 1877 • Puskás Tivadar megalkotja az első telefonközpont terveit
- 1878 • A Connecticut-i New Haven-ben megnyílik az első telefonközpont
- 1901 • Guglielmo Marconi bemutatja a rádióhullámok vezeték nélküli sugárzását (szikratávíró)^a
- 1926–1928 • Tihanyi Kálmán, Takayanagi Kenjiro és Philo Farnsworth egymástól függetlenül, különböző megoldásokkal megalkotják az első televíziókat
- 1948 • Claude Shannon megalapozza az információelméletet "A *Mathematical Theory of Communication*" munkájával
- 1956 • Az első transz-atlanti telefonkábel lefektetése
- 1959 • Mohamed Atalla és Dawon Kahng feltalálják a MOSFET-et
- 1964 • Az első próbálkozások az optikai elvű telekommunikációra
- 1968 • Norman Abramson vezetésével megalkotják az ALOHAnet-et
- 1969 • Négy számítógép összeköttetésével megindul az ARPANET
- 1979 • Beindul az első cellás mobiltelefonhálózat (1G) Tokyo-ban és környékén
- 1981 • Az SMTP-t publikálják az RFC 788-ben
- 1982 • Az Internet kifejezés megjelenése
- 1997 • Megjelenik az IEEE 802.11 (WiFi) szabvány
- 1998 • Több mint 300 millió (web)dokumentum található meg az interneten
- 2009 • Stockholmban elindul a világ első lakossági 4G hálózata
- 2019 • A világ több pontján (részleges lefedettséggel) beindul a lakossági 5G hálózat

^aAlekszander Popov szűk közönség előtt már 1985-ben bemutatta, míg Marconiról később kiderül, hogy Nikola Tesla munkáját használta fel

önmagában is hatalmas jelentőséggel bír – akkoriban pedig egyenesen sokkolta a szakmát – azonban Shannon emellett bevezetett számos alapvető foglalat, illetve felvonultatta a szükséges matematikai apparátust, ezzel megalapozva az információelméletet mint tudományterületet.

1.2. ICT – Az infokommunikáció jelene

A távközlés távlati történelme mellett fontos részletesebben foglalkoznunk az elmúlt pár évtized szűkebb történetével is, hisz ezek a trendek és irányvonalak határozzák meg az infokommunikáció

mai arculatát, illetve ezen folyamatok egy része ma is tart.

1.2.1. Az ICT konvergencia fázisai

Információ- és kommunikációtechnológia – angolul Information and Communications Technology – azaz *ICT* alatt tulajdonképpen az infokommunikációt magát, szűkebben véve pedig a távközlést, az informatikát és az elektronikus médiát értjük. Maga az ICT konvergenciája – néhol infokommunikációs konvergencia – tehát valamelyest redundáns fogalom, hisz az ICT kifejezést pont a fenti három terület konvergenciájának leírásakor kezdték szélesebb körben használni.

Ahhoz, hogy megérthessük a konvergencia lényegét, vissza kell tekintenünk² azokra az időkre, mikor az ICT alatt értett területek még markánsan elkülönültek: az analóg korszakba. Ekkor a telefon (sokáig csak vezetékes, majd – Magyarországon a 90-es évektől – mobil), a televízió, illetve az informatika (számítástechnika) tartalmanként elkülönült kommunikációs szektorok voltak, mely mindegyike rendelkezett saját infrastruktúrával: vezetékes és cellás telefonhálózat, kábeles és műholdas műsorszóró rendszerek illetve az internet kialakuló gerinchálózata. Az előfeltétele és indikátora ezek konvergenciájának a digitális technika robbanásszerű térnyerése (erre sokszor csak Moore-törvényeként hivatkozunk) volt.

Visszatekintve logikusnak tűnik, hogy a hálózatok konvergenciája során az eredendően digitális – és így a későbbiekben, közös technológiai alapot is nyújtó – internet (ekkor már nem csak mint számítógép-hálózat) vált az egységes infrastruktúra alapjává, azonban a konvergencia kezdeti fázisában ez pont fordítva volt. Az internet egyre szélesebb rétegeket ért el, felhasználva a többi kommunikációs szektor infrastruktúráját, úgy mint a dial-up (betárcsázós internet a vezetékes telefonvonalon) később az ADSL³, DOCSIS⁴ (internet a kábel TV-n) vagy a CSD⁵ később pedig a GPRS.⁶ Ahogy aztán ezek a szektorok is a digitalizálódás (pl.: vonalkapcsolás helyett csomagkapcsolás) útjára léptek, így vált egyre dominánsabbá köztük az internet.

A konvergencia formáit tekintve beszélhetünk *horizontális* illetve *vertikális* konvergenciáról. Vertikális konvergencia alatt értjük – műszaki megközelítésben – az egyes infokommunikációs ágazatokon belül a szolgáltatások (tartalomelőállítás), hálózatok (tartalomtovábbítás) és készülékek (tartalomfelhasználás) konvergenciáját: például, ha képessé tesszük a számítógép-hálózatokat audiovizuális tartalmak továbbítására. Szempontunkból fontosabb a horizontális konvergencia, mely a különböző ágazatok azonos fázisai vagy síkjai közti közeledés, kapcsolódás, kiegészítés vagy helyettesítés formájában jelentkezik: ennek gyakorlati vetületei az előző bekezdésben tárgyaltak, mikor is az internetbe mint közvetítő közegbe konvergálnak (illetve integrálódnak) az egyes ágazatok hálózati síkjai, illetve, hogy a mobiltelefonon lehetőség van böngészni az interneten.

²Érdekes, hogy ez egyeseknek valóban visszatekintés, míg másoknak már történelem, hiszen pont ezen jegyzet olvasóközönsége van abban a korban, hogy talán még aktív szemlélői voltak ennek az időszaknak. Természetesen az idő múlásával, ez a kis megjegyzés egyre inkább okafogyottá válik.

³Asymmetric Digital Subscriber Line

⁴Data Over Cable Service Interface Specification

⁵Circuit Switched Data

⁶General Packet Radio Service

A konvergencia első fázisában többnyire vertikális konvergenciáról, illetve a kései szakaszban integrációról beszélhetünk, mígnem a második fázisban megtörtént a kommunikáció rendszerek konvergenciája, ezzel kialakult az e-hírközlés. Ezen fázis tetőzése (volt) a *Triple Play* megjelenése, mikor a szolgáltatók az egységes digitális (IPv4 alapú) kommunikáció következtében tudtak integrált telefon, internet és televízió szolgáltatásokat kínálni az ügyfeleknek. A konvergencia harmadik (még meg nem haladott) lépésben alakult ki a digitális ICT szektor, amely a különböző médiumok integrációját jelenti: ez tulajdonképpen amit ma infokommunikációnak nevezünk.

1.2.2. ICT konvergencia jövője

A ICT konvergencia következő fázisai – például a digitális ökoszisztéma – sem tűnnek már annyira távoli jövőnek, azonban ezek áttekintésével itt nem foglalkozunk. Ehelyett érdekesebb röviden áttekinteni napjaink népszerű trendjeit, technológiai újításait és ezek felhasználási területeit, melyek a változások mozgatórugói.

Napjaink egyik legfontosabb trendje vagy paradigmája, a köztudatban is egyre inkább jelenlévő *Internet of Things* avagy a Dolgok Internete. Maga a fogalom egy meglehetősen tág területet fed le, azonban az alapkoncepció egyértelműen "intelligens" eszközök hálózatba kapcsolása minél nagyobb számban. Az intelligens alatt itt többnyire beágyazott vagy ambiens rendszereket értünk, de tulajdonképpen ide tartozik minden, aminek van processzora. A paradigma egyértelműen a digitális vagy okos ökoszisztémának nyújt alapot, hisz ennek alappillére a nagy mennyiségű adat gyűjtése és hasznosítás az élet minden területén; legyen ez akár egy gyár automatizálása, a közlekedés gördülékennyé tétele egy *okos városban* vagy akár egy okos otthon hatékony működtetése.

Természetesen ez a paradigma számos leküzdendő problémát felvetett, amely problémák köré új szakterületek szerveződtek. Ilyen például a nagy mennyiségű adat feldolgozása és elemzése, melyet összefoglaló néven *Big Data*-nak nevezünk, annak ellenére, hogy a adatok feldolgozására több eltérő megközelítés is létezik, például a stream (online vagy on-the-fly) és batch (offline) processzálas.

Jelentős szemléletváltást igényelt maga az a tény is, hogy az internet egyes (illetve egyre több) végpontjai önműködő rendszerek, nem pedig egy ember aki a számítógép mögött ül. Így például a *gép-gép kommunikáció* (M2M)⁷ is egy merőben új terület, amelynek égisze alatt számos új protokoll család jött létre: architektúra szempontjából például a szolgáltatás-orientált architektúrájú⁸ (SOA) a REST⁹, illetve az üzenet orientált protokollok, míg séma szempontjából a megszokott kérdés-válasz alapú kommunikáció mellett megjelent a publikáció-feliratkozó mintájú üzenetváltás.

Alapvető változást jelent már önmagában az IoT eszközök (várhatóan) hatalmas száma is, amelynek következtében az IPv4 szerinti címezésben nem lehet elég címzetni minden eszköznek az interneten. Ezen probléma leküzdésére (is) vállalkozik az *IPv6*, mely nem csak a nagyszámú eszközök megcímezését teszi lehetővé, hanem a mobilitás és titkosítás kérdését is ke-

⁷Machine-to-machine

⁸Service-oriented architecture

⁹Representational State Transfer

zelní kívánja. Az IP alapú címzéssel szorosan összefügg a skálázhatóság kérdése is, azaz, hogy miként lehet sok eszközt egyszerre megszólítani, a hálózat nagymértékű leterhelése nélkül. Ennek egyik lehetséges megoldása a *multicast*, melynek használatával csoportokat is lehet címezni, így elkerülve, hogy egyenként kelljen elküldeni ugyanazt az üzenetet több (esetlegesen több millió) címzettnek anélkül, hogy kigyulladna a gerinchálózat.

Szintén számos kihívást jelentenek az IoT kapcsán megjelenő új vagy meglévő de egyre hangsúlyosabb követelmények¹⁰, például az alacsony késleltetés és a nagy sávszélesség, különösen vezeték nélküli technológiák tekintetében. Napjainkban az *5G* technológiát tekintik ezen problémakör svájci bicskájának, amely ezekre a követelményekre hatékonyan tud reagálni és képes ezeket kiszolgálni. Azonban, elterjedésére és kiteljesedésére még várunk kell, így az erre épülő, egymással kommunikáló, gyors reakcióidejű önvezető autók hálózatára, és az elképesztően nagymennyiségű adatok vezeték nélküli továbbítására is.

Fontos megemlíteni, hogy a nagy mennyiségű adat gyűjtésének, társadalmi és etikai vonzatai is vannak, hisz nem csak a beágyazó (fizikai) környezetet figyelik meg ezek az okos eszközök, hanem az abban élő embert is: ez műszakilag az adatbiztonság kérdésre fordítható le. A legfontosabb kérdés, hogy az IoT széleskörű elterjedésével mi fog történni az emberek privát életterével, magánszférájával: illetve mennyit kívánunk feláldozni ezekből a technológia oltárán.¹¹ Sajnos, napjaink mérnökképzésében nem jutott hely a *mérnöketikának*, azonban nem felejthetjük el a műegyetemi mérnökeskü szövegének utolsó sorát: „*Szokmány erkölcsi szabályait megtartom, megszerzett tudásomat embertársaim, és hazám javára használom fel.*”. Aki pedig mérnökhallgatóként úgy érzi, hogy neki nem tiszte ”humán” kérdésekkel foglalkozni, emlékezzen karunk legendás oktatójára, Simonyi Károlyra, aki szerint nincs humán illetve természettudományos kultúra vagy tudomány, mivel az egy és egységes.

¹⁰Fontos követelmény még az applikációk nagy részében az itt nem tárgyalt energiahatékonyság.

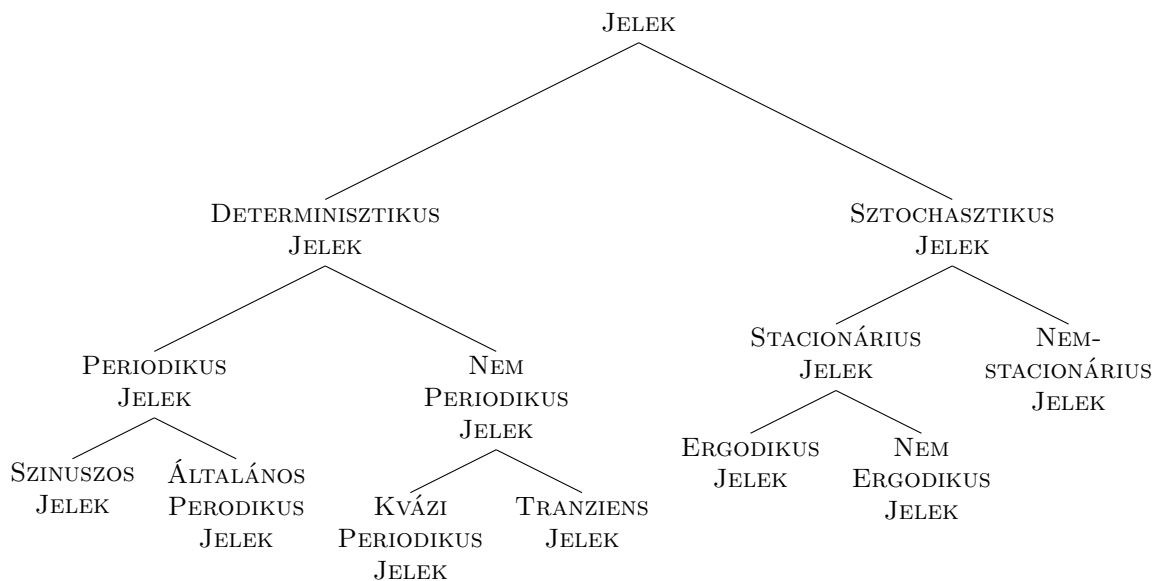
¹¹Itt fontos megérteni, hogy megvédhetjük ugyan az adatot az illetéktelen hozzáférésektől, az még létezik, tehát a privátszféra akkor is sérül, ha az adathoz csak ”gépek” férhetnek hozzá.

2. fejezet

Jelelméleti alapok

Az alábbi fejezetben azokat a jel-, és hálózatelméleti, illetve elektronikai ismereteket foglaljuk össze, melyekre a későbbiek során hivatkozni fogunk. Mivel ezeket az előzménytárgyak során oktatták, így a fejezetben található formulák mögötti mélyebb összefüggéseket nem ismertetjük.

A továbbiakban – kiváltképp a jegyzet első felében – megannyi különböző, többnyire folytonos idejű jellel fogunk foglalkozni, hiszen ezek szolgáltatják a számunkra fontos információkat az adott jelenségekről. Ezek csoportosítása nem mindig egyértelmű, hisz absztrakciós szinttől is függ, hogy hány különböző jeltípust határozunk meg. Most egy viszonylag egyszerű, ésszerű – ”hallgatóbarát” – felbontást adunk meg a 2.1 ábrán.



2.1. ábra. Különböző jeltípusok egy lehetséges csoportosítása

2.1. Determinisztikus jelek

Mivel az előkövetelmény tárgyak szinte kizárólagosan a determinisztikus jelek csoportját tárgyalják, így ezek mindenki számára ismerősek, még ha a fogalom maga nem is tudatosult mindenki-ben. **Determinisztikus jelnek** nevezzük azokat jeleket, amelyeknek értéke minden időpontban ismert vagy meghatározható, tehát a jel egy megismételhető folyamatot ír le. Érdemes megjegyezni, hogy a gyakorlatban ez így, ebben a formában nem alkalmazható, hisz az egyes jelenségek,

folyamatok és mérések minden esetben zajjal terhelték, így érdemes kikötni, hogy milyen pontossággal kell a jelnek leírnia a folyamatot, hogy determinisztikusnak tekintsük. Az esetek nagy többségében a pontosság meghatározása az alábbi heurisztikus módszerrel zajlik: „*A másik szkopon is kábé így néz ki, a tegnapi mérés eredménye is kábé így nézett ki, szóval szerintem ez így oké.*”

A periodicitás kérdése már jóval szemléletesebben kezelhető: **Periodikusnak** nevezünk egy jelet, ha azt T periódusidővel eltolva önmagát kapjuk. Ezek jó tulajdonsága, hogy Fourier-sorba fejthetők¹, amely megkönnyíti a feldolgozást illetve az analízist.

A periodikus jelek közül számunkra a legfontosabbak a **szinuszos** jelek. A szinuszos jelnek számos kedvező tulajdonsága van, azonban ezek mindegyike abból következik, hogy a lineáris rendszerek sajátfüggvényei a szinuszos jelek², így bármely periodikus jel megadható szinuszos függvénynek összegeként. Egy szinuszos jel általánosan, az alábbi formában írható fel:

$$x(t) = A \cos(2\pi ft + \phi) \quad (2.1)$$

A 2.1 formula alapján a szinuszos jelet az A amplitudó, f frekvencia és ϕ fázis egyértelműen meghatározza.

2.1.1. Effektív érték

A továbbiakban a szinuszos jel egyes – számunkra fontos – jellemzőit fogjuk bemutatni, melyek közül az egyik legfontosabb az **effektív érték** vagy négyzetes középérték, mely periodikus jelekre az alábbi módon számítható ki:

$$x_{eff}(t) = \sqrt{\frac{1}{T} \int_0^T [x(t)]^2 dt} \quad (2.2)$$

A fentiek alapján a szinuszos jel effektív értéke könnyedén kiszámítható, azonban mivel zárthelyiben megkérdezhetjük, úgy igazságos, hogy a megoldást meg is mutatjuk. A számítások során az alábbi – más témakörben is előszeretettel használt – trigonometrikus azonosságot használjuk fel; a teljesség kedvéért ennek párját is megadjuk:

$$\cos^2(x) = \frac{1 + \cos(2x)}{2} \quad (2.3)$$

$$\sin^2(x) = \frac{1 - \cos(2x)}{2} \quad (2.4)$$

A szinuszos jel effektív értékének kiszámítása, felhasználva a 2.2 és 2.3 formulákat:

¹Elméletileg lehetnek olyan periodikus jelek, melyek nem fejthetők Fourier-sorba, de a korlátosság illetve a szakaszokénti folytonosság kikötésével ezt a problémát elegánsan áthidaljuk.

²Valójában a komplex exponenciálisok, melyeket az Euler-formula segítségével hozhatunk trigonometrikus alakra. Ezért is használjuk a szinuszos jelek reprezentálására a koszinusz függvényt, hisz az adja meg a komplex exponenciális jel valós részét.

$$\begin{aligned}
x_{eff} &= \sqrt{\frac{1}{T} \int_0^T [A \cos(2\pi ft + \phi)]^2 dt} \\
&= \sqrt{\frac{A^2}{T} \int_0^T \left[\frac{1 + \cos(2[2\pi ft + \phi])}{2} \right] dt} \\
&= \sqrt{\frac{A^2}{2T} \int_0^T [1 + \cos(4\pi ft + 2\phi)] dt} \\
&= \sqrt{\frac{A^2}{2T} \left[t + \frac{1}{2} \sin(4\pi ft + 2\phi) \right]_0^{T=\frac{1}{f}}} \\
&= \sqrt{\frac{A^2}{2T} \left[T - 0 + \frac{1}{2} \sin(2\phi) - \frac{1}{2} \sin(2\phi) \right]} \\
&= \sqrt{\frac{A^2}{2}} = \frac{A}{\sqrt{2}} \approx 0,7071 \cdot A
\end{aligned}$$

Természetesen ezzel megegyező eredményt kaptunk volna, ha koszinusz helyett szinusz függvényt választunk (még a levezetés is csak minimálisan tér el). Az effektív érték (villamos jelekre) lényegében azt adja meg hogy mekkora az ekvivalens egyenszint, amely a váltakozó jellel (átlagosan) megegyező mennyiségű hőt termel egy ellenálláson. Tipikusan azért használjuk, mert így a teljesítményre vonatkozó formulák egyszerűsödnek és jobban hasonlítanak az időben állandó esetre. Az effektív érték kiszámítására egy hasznos formula a 2.5 képlet, a Parseval-tétel, amennyiben egy N felharmonikusból álló periodikus jelet vizsgálunk és az egyes felharmonikusok effektív értékei ismertek:

$$x_{eff}(t) = \sqrt{\sum_{k=0}^N (X_{eff}^{(k)})^2} \quad (2.5)$$

2.1.2. Csúcstényező

Egy másik fontos jellemezője egy jelnek (nem csak periodikus esetben) az úgynevezett **csúcstényező**, mely a jel csúcsertékeinek (korlátos jelet feltételezve) illetve effektív értékének a hányadosa:

$$C_x = \frac{A_x}{x_{eff}} \quad (2.6)$$

Ennek értéke szinusz jel esetében:

$$C_{\sin} = \frac{A_{\sin}}{\frac{A_{\sin}}{\sqrt{2}}} = \frac{A_{\sin}}{A_{\sin}} \sqrt{2} = \sqrt{2} \quad (2.7)$$

A csúcstényező egy kifejezetten fontos gyakorlati jellemző, tulajdonképpen egy konstans szorzóként fogható fel, bármely áramkörü komponens ára kapcsán. Minél nagyobb a csúcstényező értéke, annál nagyobbra kell méretezni (tűrés szempontjából) az eszközeinket, miközben az átvitt teljesítmény értéke állandó.

A 2.1 táblázatban megadjuk a számunkra legfontosabb jelek – szinusz, háromszögjel, négyszögjel – effektív értékét illetve csúcstényezőjét.

Jeltípus	Csúcsérték	Csúcstényező	Effektív érték
Szinusz	A	$\sqrt{2}$	$\frac{A}{\sqrt{2}}$
Háromszög	A	$\sqrt{3}$	$\frac{A}{\sqrt{3}}$
Négyszög	A	1	A

2.1. táblázat. Nevezetes jelek legfontosabb paraméterei

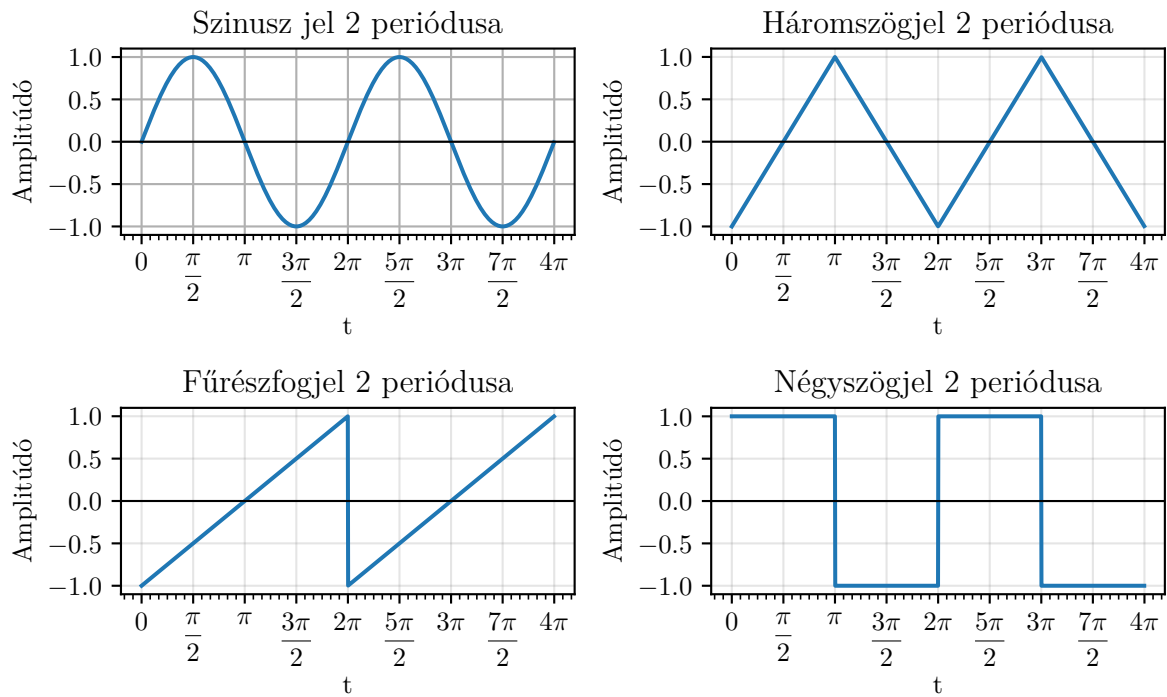
2.1.3. Periodikus jelek Fourier-sora

A jelek időtartománybeli vizsgálata mellett, foglalkoznunk kell a frekvenciatartománybeli jellemzőivel is, ahol főként a fent említett három jeltípusra szorítkozunk. Ahhoz, hogy a jelet frekvencia tartományban vizsgáljuk, elő kell állítanunk a jel spektrumát, amelyhez a Fourier-transzformáción keresztül jutunk. A Fourier-transzformációról még a tranziens jelek kapcsán fogunk értekezni a 2.2.1 fejezetben, azonban azt érdemes előrebocsájtani, hogy a periodikus jelek a szó szoros értelmében nem Fourier-transzformálhatók.³ Szerencsére ahhoz hogy néhány alapvető kérdést tisztázzunk a fenti jelek spektrumára vonatkozóan, nincs szükségünk Fourier-transzformációra. Az egyetlen ismeret amelyet felhasználunk, hogy a szinuszos függvények – ahogy azokat a 2.1 kifejezésben magadtuk és a 2.2 ábrán látható – csak a $\pm f$ frekvenciákon⁴ tartalmaznak spektrális komponenseket, ahogy a 2.3 ábra mutatja.

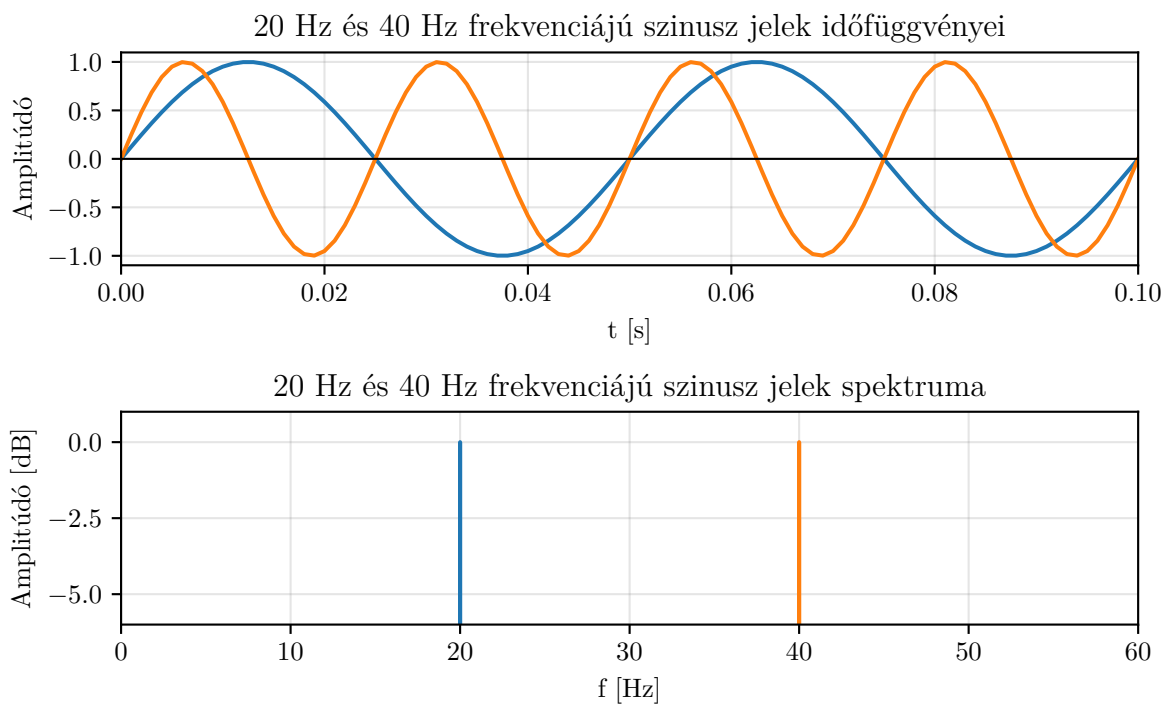
Ahhoz, hogy a további nevezetes jeltípusok spektrumáról nyilatkozzunk, elég csupán megadni az egyes jelek Fourier-sorát. Rövid emlékeztetőül álljon előttünk, hogy az egyes jelek Fourier-sorba fejtése (valós alakban) a Fourier-együtthatókkal súlyozott adott frekvenciájú szinusz illetve koszinusz jelek összegéből áll, a 2.1.3 formula értelmében.

³Ennek feloldását nem mutatjuk be, de ahogy a fejezetben csak röviden említett sok más témakörhöz, ez is található szakirodalom bőséggel, kiváltképp az előzmény tárgyak kötelező irodalma.

⁴A spektrum negatív részét ez esetben figyelmen kívül hagyhatjuk, de a további fejezetekben szinte kizárólag kétoldalas spektrumokkal foglalkozunk.



2.2. ábra. Alapvető jelek időtartományi ábrázolása



2.3. ábra. Különböző frekvenciájú szinusz jelek időfüggvényei és spektrumai

$$x(t) = X_0 + \sum_{n=1}^{\infty} [X_n^A \cos(2\pi nft) + X_n^B \sin(2\pi nft)]$$

ahol, X_n a n -edik Fourier-együtthatót jelöli

$$\begin{aligned} X_0 &= \frac{1}{T} \int_0^T x(t) dt \\ X_n^A &= \frac{2}{T} \int_0^T x(t) \cos(2\pi fnt) dt \\ X_n^B &= \frac{2}{T} \int_0^T x(t) \sin(2\pi fnt) dt \end{aligned}$$

A teljesség kedvéért bemutatjuk a Fourier-sorfejtés komplex alakját is:

$$x(t) = X_0 + \sum_{n=1}^{\infty} [X_n^C e^{jn2\pi ft}] \quad (2.8)$$

ahol, X_n a n -edik Fourier-együtthatót jelöli

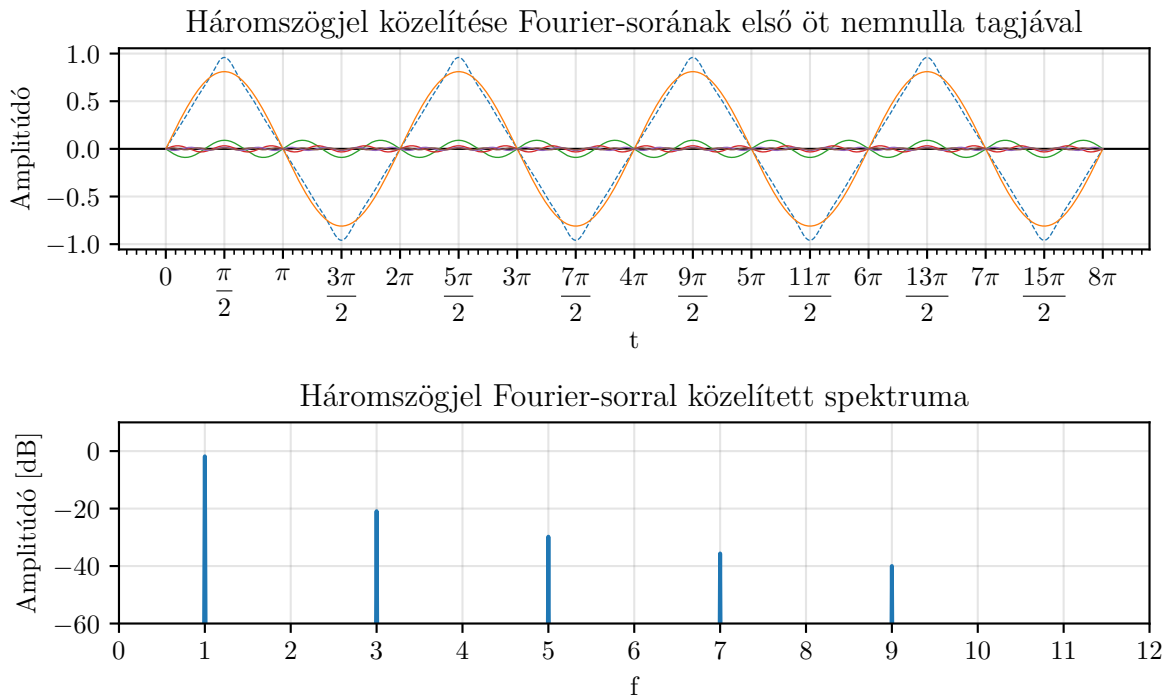
$$X_n^C = \frac{1}{T} \int_0^T x(t) e^{-jn2\pi ft} dt \quad (2.9)$$

Az egyes jeltípusok – háromszög-, fűrészfog- és négyszögjel, melyek időtartománybeli ábrázolása a 2.2 ábrán látható – spektruma fentiek alapján előáll az összetevő szinuszok spektrális komponenseinek összegeként: az ilyen jellegű spektrumokat **vonalasnak** nevezzük. Természetesen a Fourier-sor és így annak spektruma is csak abban az esetben egyezik meg az eredeti jellel, illetve annak spektrumával, ha végtelen sok tagot összegzünk. Emiatt a gyakorlatban csak N darab tagot használunk fel, ezért a Fourier-sor helyett az első N tagokból alkotott polinommal közelítjük az eredeti jelet.

2.1.4. Periodikus jelek spektruma

A 2.4, 2.5, 2.6 ábrák az egyes jelek (rendre, háromszög-, fűrészfog- és négyszögjel) Fourier-sorral való közelítését mutatják be, ahol $N = 10$ harmonikus jel összegével közelítjük az eredeti jeleket. Az ábrákról leolvasható az is, hogy a háromszög-, illetve négyszögjel esetében a páros felharmonikusok hiányoznak, azok Fourier-együtthatói zérust adnak. Az ábrák tanulsága szerint háromszögjel esetében az öt darab nem nulla harmonikussal való közelítés is az eredeti jelhez nagyon hasonló eredményt ad, míg a másik két esetben szemmel láthatóan jóval nagyobb az eltérés. Ennek egyik oka az úgynevezett **Gibbs-jelenség**: Mivel szakszosan folytonos függvényeket közelítünk folytonos függvények összegeként, ezért ahol az eredeti függvénynek szakadási helye van, ott a Fourier-sorában túllövés lép fel. Nagyobb N értékekre a túllövés mértéke csökken, így

a közelítés pontosabbá válik.

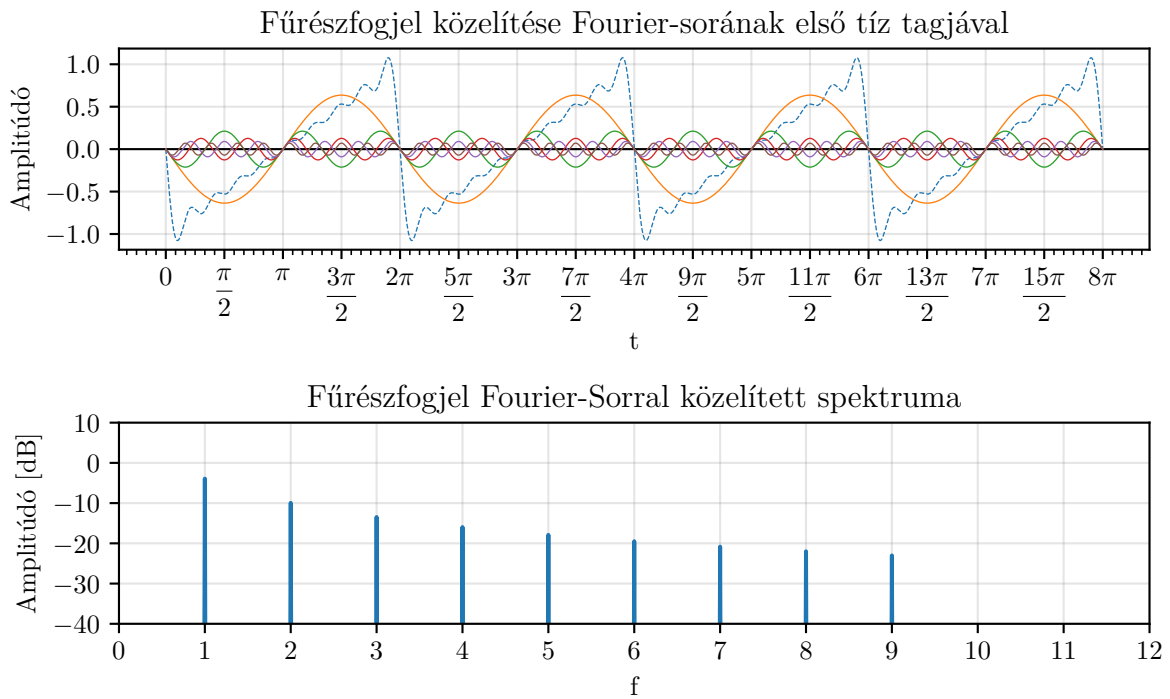


2.4. ábra. Fourier-sorának első öt nemnulla tagjával közelített háromszögjel időfüggvénye és spektruma

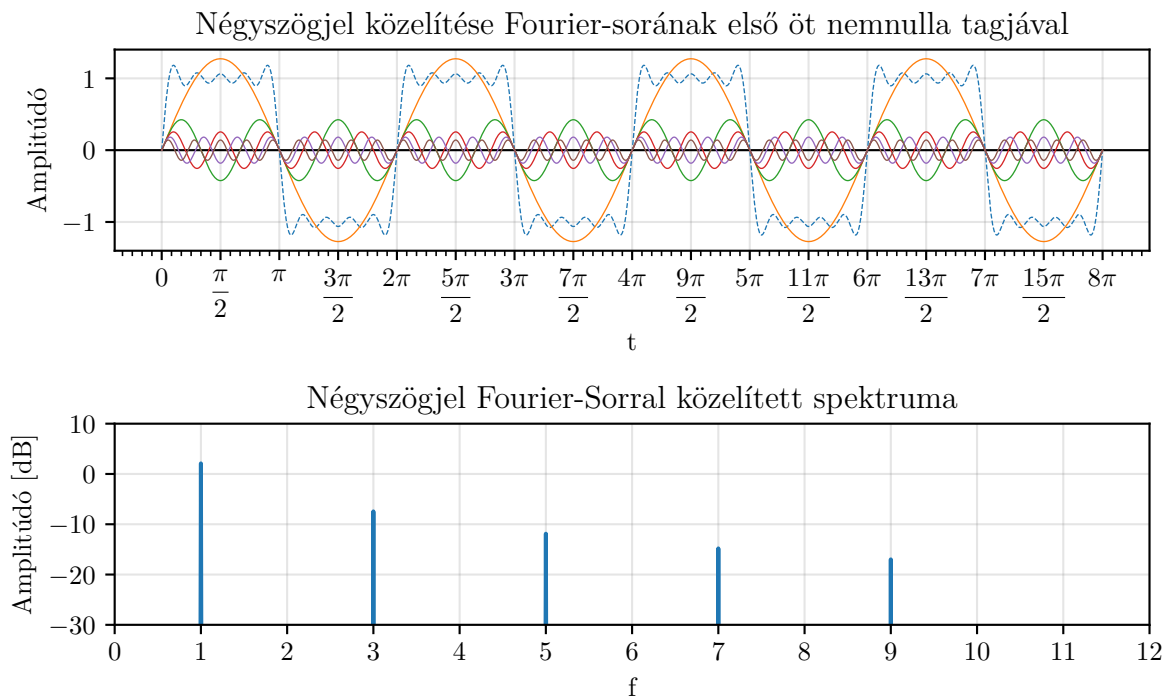
A spektrumokat vizsgálva már nem ér minket meglepetés: az egyes szinuszos összetevőkhöz tartozó spektrális komponensek összegeként áll az egyes jelek spektruma. Ennek az eredménynek a bevése kiemelten fontos, mivel sokan az alapharmonikus frekvenciáját (f) és a sáv szélességet (B) ekvivalens fogalmakként használják, holott ez – az ábrák tanulsága szerint is – hamis következtetés. Az egyes jeleink sáv szélessége (ahogy a Fourier-sorral való közelítésnél láthattuk) végtelen, hiszen végtelen sok felharmonikus állítja elő az eredeti jelet. Azonban az is látható, hogy az egyes felharmonikusok amplitúdója az alapharmonikustól távolodva szignifikánsan csökken, tehát gyakorlati szempontból megállapíthatunk egy küszöbértéket, amelynél kisebb amplitúdóval rendelkező felharmonikusok már elhanyagolhatók.

2.1.5. A logaritmikus skála használata

Az egyes ábrákon a különböző frekvenciájú komponensek amplitúdóját **decibelben** (dB) adtuk meg, azaz (10-es alapú) logaritmikus skálát használtunk az ábrázolásra. Ennek nagy előnye, hogy sokkal nagyobb tartományt tudunk ábrázolni vagy átfogni, így a több nagyságrenddel eltérő értékek összehasonlítása is meglehetősen leegyszerűsödik, hisz az ábrán a 0.1-szeres csillapítás pont olyan távolságban van a nullától, mint a 0.01-szeres csillapítás a 0.1-szerestől. A másik nagy előnye a logaritmikus skála használatának, hogy az arányokkal végzendő műveleteket összeadásra és kivonásra egyszerűsíti le. Egy mennyiség dB-ben való megadáshoz szükségünk van egy referencia szinthez, amelyhez képest viszonyítjuk: a spektrumok ábrázolásánál a referencia



2.5. ábra. Fourier-sorának első tíz tagjával közelített fűrészfogjel időfüggvénye és spektruma



2.6. ábra. Fourier-sorának első öt nemnulla tagjával közelített négyszögjel időfüggvénye és spektruma

szint az eredeti jel amplitúdója. Amennyiben teljesítményeket hasonlítunk össze ott a referencia szint megegyezés szerint 1 Watt, illetve ha kifejezetten kis mennyiségekről beszélünk és ezeket milliwattban adjuk meg akkor 1 mW, ekkor viszont a **dBm** jelölést használjuk. Teljesítményeket a 2.10 képletben megadott módon válthatunk át dB-be.

$$x^{[dB]} = 10 \lg \left(\frac{x^{[W]}}{1} \right) = 10 \lg(x^{[W]}) \quad (2.10)$$

Az arányokon végzett műveletek megértéséhez fel kell elevenítenünk a 2.11, 2.12 és 2.13 logaritmus azonosságokat. Ezek alapján látható, hogy két mennyiség szorzata, hányadosa és hatványa a dB-ben megadott mennyiségek összegeként, különbségeként és szorzataként megadható.

$$\log_a xy = \log_a x + \log_a y \quad (2.11)$$

$$\log_a \frac{x}{y} = \log_a x - \log_a y \quad (2.12)$$

$$\log_a x^k = k \cdot \log_a x \quad (2.13)$$

A decibelt eredetileg teljesítményarányok kifejezéséhez használták, azonban a mérnöki gyakorlatban más, különféle mennyiségeket is szokás decibelben megadni, például feszültséget (és így jelek amplitudóját). Ez sokakban zavart kelt, mivel a különböző dimenziójú mennyiségek decibelbe váltásakor eltérő szorzótényezőt használunk. A teljesítményére vonatkozó formula tehát a 2.10 egyenlet, míg a feszültségre vonatkozó formula (és annak levezetése a 2.14 egyenlet. Bár a dB-be való átváltás első ránézésére számológépet igényel, a fenti formulák, prímtényezőkre bontás és a 2.2 táblázatban foglaltak ismerete alapján könnyen elvégezhető akár fejben is – természetesen ez a módszer csak egy igen pontos közelítést ad.

$$\begin{aligned} x^{[dB]} &= 10 \lg \left(\frac{x^{[W]}}{1} \right) \\ &= 10 \lg \left(\frac{x^2 [V]}{\frac{R}{1^2 [V]}} \right) \\ &= 10 \lg \left(\frac{x^2 [V]}{1^2 [V]} \right) 10 \lg \left(\frac{x^2 [V]}{1^2 [V]} \right) \\ &= 10 \log \left(\frac{x^{[V]}}{1^{[V]}} \right)^2 \\ &= 2 \cdot 10 \lg \left(\frac{x^{[V]}}{1^{[V]}} \right) \\ &= 20 \lg \left(x^{[V]} \right) \end{aligned} \quad (2.14)$$

X	$\lg(X)$	dB ^[W]	dB ^[V]
1	0	0 dB	0 dB
2	0,3	3 dB	6 dB
3	0,48	4,8 dB	9,6 dB
π	0,5	5 dB	10 dB
5	0,7	7 dB	14 dB
7	0,85	8,5 dB	17 dB

2.2. táblázat. (Nem csak) prímszámok 10-es alapú logaritmus és átváltása decibelbe

2.2. Nem periodikus jelek

Miután összefoglaltuk a periodikus jelek legfontosabb jellemzőit, érdemes pár szót ejteni a nem periodikus jelekről. Elsőként tekintsük át a számunkra is igen fontos *majdnem periodikus jeleket*: ezek olyan jelek melyek periodikus komponensekből állnak, azonban azok összege már nem periodikus. A gyakorlatban tipikusan akkor áll elő ilyen ha két szinkronizálatlan jelet összeadunk vagy összeszorozunk. Hasonlóan kezelhető az az eset amikor szinkronizált jeleket adunk össze de az adódó periódushossz az egyes komponensek periódushosszainál nagyságrendekkel nagyobb; értsd: a vizsgált tartományon nem tekinthető periodikusnak. Ezekben az esetekben tipikusan nehezen tudunk nyilatkozni a korábban vizsgált jellemzőkről: például nagyon hosszú periódus idő esetén nehéz megmondani, hogy az összegjelnek mennyi is a csúcserő, ellenben egész jól becsülhetők. Például szinkronizált szinusz jelek összegeként előálló nem periodikusnak tekinthető jel (de, ez igaz bármilyen szögmodulált jelre) megadható egy "fázisban lévő" és egy "kvadratúra" komponens összegeként.

2.2.1. Tranziens jelek Fourier-transzformáltja

Nem periodikus jelek egy másik fontos csoportját alkotják a *tranziens* jelek, melyek nem tartalmaznak periodikus komponenseket (ez még nem feltétlenül jelenti, hogy a jel tranziens) és csak egy adott intervallumon különbözik az értékük zérustól (tehát előbb-utóbb eltűnnek). Matematika értelemben ezt nevezzük abszolút integrálhatóságnak, vagyis a jel abszolút értékének integrálja véges a teljes értelmezési tartományon, ahogy azt a 2.15 formula is mutatja.

$$\int_A |f(x)| dx < \infty \quad \text{ahol, } A \text{ az } f \text{ függvény értelmezési tartománya} \quad (2.15)$$

Ez egyben azt is jelenti, hogy a tranziens jeleknek létezik *Fourier-transzformáltja*, mely a 2.16 formula szerint áll elő. Természetesen nem abszolút vagy négyzetesen integrálható jelek spektruma is létezhet, de az alábbi formulával nem számolható.

$$X(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j2\pi ft} dt \quad (2.16)$$

Hasonlóképp állítható elő az eredeti jel a Fourier-transzformáltból az inverz Fourier-

transzformáció segítségével a 2.17 szerinti alakban áll elő.

$$x(t) = \mathcal{F}^{-1}\{X(f)\} = \int_{-\infty}^{+\infty} X(f) \cdot e^{j2\pi ft} df \quad (2.17)$$

Fontos megjegyezni a fenti transzformációk kapcsán, hogy mi az f frekvenciát tüntettük fel változóként a megfelelő argumentumokban, míg az irodalomban gyakorta az ω körfrekvenciát használják erre a célra. Az utóbbi esetben az inverz Fourier-transzformáció képletében megjelenik egy $\frac{1}{2\pi}$ szorzó, hisz X mértékegysége t mértékegységének inverze kell hogy legyen. Néhol, annak érdekében, hogy a formulák szimmetrikusak legyenek, mindkét transzformáció során $\frac{1}{\sqrt{2\pi}}$ szorzót alkalmaznak.

A Fourier-transzformációt ugyan direkt módon ritkán fogjuk használni a továbbiakban, azonban érdemes felidézni, az úgynevezett **modulációs tételt** melyre még szükségünk lesz, illetve szempontunkból elvi jelentősége is nagy.

$$\mathcal{F}\{x(t)e^{j2\pi f_0 t}\} = X(j2\pi(f - f_0)) \quad (2.18)$$

A fenti kifejezés tanulsága alapján a kisméretű $x(t)$ jel átvihető egy másik nagyfrekvenciás vivőjel használatával.

2.2.2. A Dirac-impulzus

A **Dirac-impulzus** gyakran használt, hasznos eszköz ha egy nagyon rövid idejű, egységnyi intenzitású impulzust szeretnénk leírni. A Dirac-delta formális kezelése azonban nem triviális, hisz matematikai értelemben nem függvényről van szó, hanem egy úgynevezett disztribúcióról avagy általánosított függvényről. Ennek egy realizációját a következő módon – az $\varepsilon(t)$ egységugrás függvényre támaszkodva – lehet bevezetni:

$$\delta(t) \approx \delta(t, \tau) = \frac{\varepsilon(t) - \varepsilon(t - \tau)}{\tau}, \quad \tau \ll \tau_0 \quad (2.19)$$

Itt τ_0 a vizsgált jelenségre jellemző legkisebb idő (pl.: időállandó, periódusidő), tehát a $\delta(t)$ minden t időre 0 értékű, kivéve a $t = 0$ helyen, ahol meghatározott módon végtelenné válik. A rendszerelmélet, gyakran használja az egységnyi Dirac-impulzust amely, szemléletesen a $\delta(t, \tau)$ impulzus határértéke $\tau \rightarrow 0$ esetén, azonban ez a megfogalmazás nem precíz, mivel a határérték nem függvény. Ezzel vissza is kanyarodtunk első állításunkhoz, miszerint a Dirac-impulzus disztribúcióval írható le, így a $\delta(t)$ egységnyi Dirac-delta egy egyezményes jelölés, amely így értendő:

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1 \quad (2.20)$$

A fentiek értelmében következik az alábbi két fontos összefüggés a Dirac-impulzussal kapcsos-

latban, amely bármely folytonos $f(x)$ függvényre igaz:

$$f(t)\delta(t - T) = f(T)\delta(t - T) \quad (2.21)$$

$$\int_{-\infty}^{+\infty} f(t)\delta(t - T) dt = f(T) \quad (2.22)$$

A 2.22 egyenletből és a Dirac-delta ama tulajdonságából, hogy páratlan disztribúció, következik az alábbi fontos tulajdonság:

$$(f(t) * \delta(t - T)) = f(t - T) \quad (2.23)$$

Ezek alapján már könnyen magadható a Dirac-impulzus Fourier-transzformáltja az alábbi módon:

$$\mathcal{F}\{\delta(t)\} = \int_{-\infty}^{+\infty} \underbrace{\delta(t)}_{\delta(t-0)} e^{-j2\pi ft} dt = e^{-j2\pi f0} = 1 \quad (2.24)$$

2.3. Sztochasztikus jelek

Sztochasztikus jelek alatt szemléletesen "véletlennek tűnő" jeleket értünk. Ez legegyszerűbben úgy fogható fel, ha ugyanazt a mérést elvégezzük többször és rendre különböző jeleket mérünk. Ez egyrészt betudható a determinisztikus ismeretek hiányából származó bizonytalanságnak, másrészt "véletlen" ingadozásoknak: a lényeg, hogy a mérési körülmények nem definiálják egyértelműen a mérések eredményét.

A sztochasztikus jelek matematikai modellje a sztochasztikus folyamatok – melyek a valószínűségi változó általánosításának tekinthetők –, amelyekkel különféle véletlen rendszerek időbeli viselkedését írjuk le. Ebben a megközelítésben az elemi események a mért függvények, amelyeket mintafüggvényeknek nevezünk. Leegyszerűsítve tehát, nem egy mért függvény ír le egy folyamatot, hanem mintafüggvények sorozata, így egy paraméterfüggő eloszlás-seregben gondolkodunk, ahol az egyes – valamilyen módon definiált – momentumokat vizsgáljuk. Maga a sztochasztikus folyamat nehezen kezelhető fogalom, ezért kiemelt szerep jut az egyes speciális folyamatoknak.

Erősen stacionáriusnak nevezünk egy folyamatot, ha tulajdonságai az idő előre haladtával nem változnak meg, tehát a felvett változók együttes eloszlása időinvariáns. Amennyiben az első és másodrendű momentumok időinvariánsak, úgy gyengén stacionárius folyamatról beszélünk. A legegyszerűbb példa gyenge stacionaritásra a fehér zaj, hisz egy adott időpontban értéke véletlen, azonban várható értéke és varianciája időben állandó.

3. fejezet

Mintavételezés és kvantálás

Emberi léptékben gondolkodva – melynek határait a XX. századi fizika¹ nagyságrendekkel átlépte – a környező világ folyamatait időben folytonosnak tekintjük. Azonban manapság jóval nagyobb szerephez jut a digitális jel, hisz a műszaki megoldások egyre nagyobb része digitális: elég csak – az egyébként eredeti értelmezésében érvényét veszítő² – Moore törvényére gondolnunk. Az *all digital* trend térhódítása miatt fontos szerephez jut a folytonos jelek diszkrét jelfolyammá alakításának illetve annak visszaalakításának kérdése. Az alábbi fejezetben az analóg-digitális átalakítás elméleti és gyakorlati alapjait fogjuk tárgyalni, nagy hangsúlyt fektetve arra, hogy az átalakítások során miként torzul az információt szállító jel.

3.1. Mintavételezés és jelrekonstrukció

Az analóg-digitális átalakítás során a jel valamilyen reprezentációjából veszünk mintákat, többnyire egyenletes lépésközönként. Tisztán elméleti szempontból – tehát feltételezve, hogy tetszőleges pontossággal tudjuk a mintáinkat tárolni, illetve egyéb kevésbé realizálható körülmények teljesülését, mint az akauzalitás, – a vett mintákból az eredeti jel maradéktalanul visszaállítható. Az alábbiak során az úgynevezett matematikai (elméleti) mintavételezésről lesz szó³, majd bemutatjuk, hogy ez milyen megszorításokkal alkalmazható a gyakorlatban.

3.1.1. Elméleti mintavételezés

A legkönnyebben értelmezhető, és az alkalmazások tekintetében is leggyakrabban használt mintavételezési eljárás az időtartományban történő mintavételezés. Ennek kapcsán, bármilyen előzetes ismeret nélkül is két benyomásunk támadhat. Először is, hogy minél több mintával rendelkezünk, annál pontosabban (vagy akár hibátlanul) helyreállítható az eredeti jel. Másodsor pedig, hogy a gyorsan változó jelek visszaállításához nyilvánvalóan több mintára van szükségünk, mint egy "lassú" jel esetében. A jel változási sebessége annak spektrumával van összefüggésben,

¹Ez csak részben igaz, hisz számos XX. századi fizikai elmélet, modell és törvény megalapozása a késő XIX. században történt.

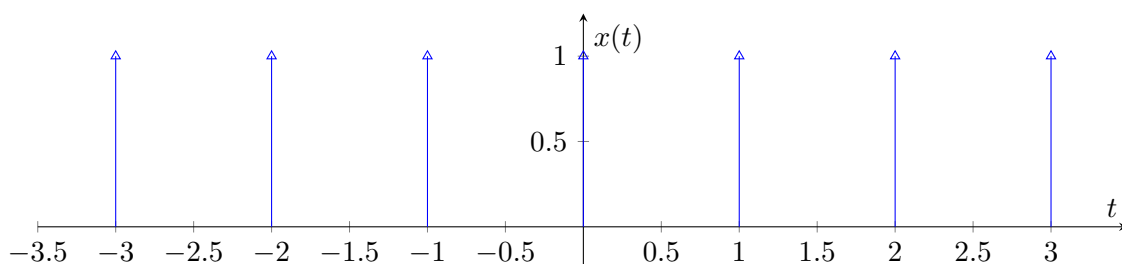
²Manapság kezdjük azt a korlátot elérni, hogy nem lehet már több tranzisztort zsúfolni egy szilícium lapra, mert ekkora csíkszélesség esetén már kvantummechanikai hatásokkal is számolni kell, így a processzorok fejlődése inkább a többmagú architektúrák irányába terelődött el.

³A mintavételezés kérdésével az előzménytárgyak (például Jelek és Rendszerek) is részletesen foglalkoztak, így itt csak a legszükségesebb részeit tekintjük át.

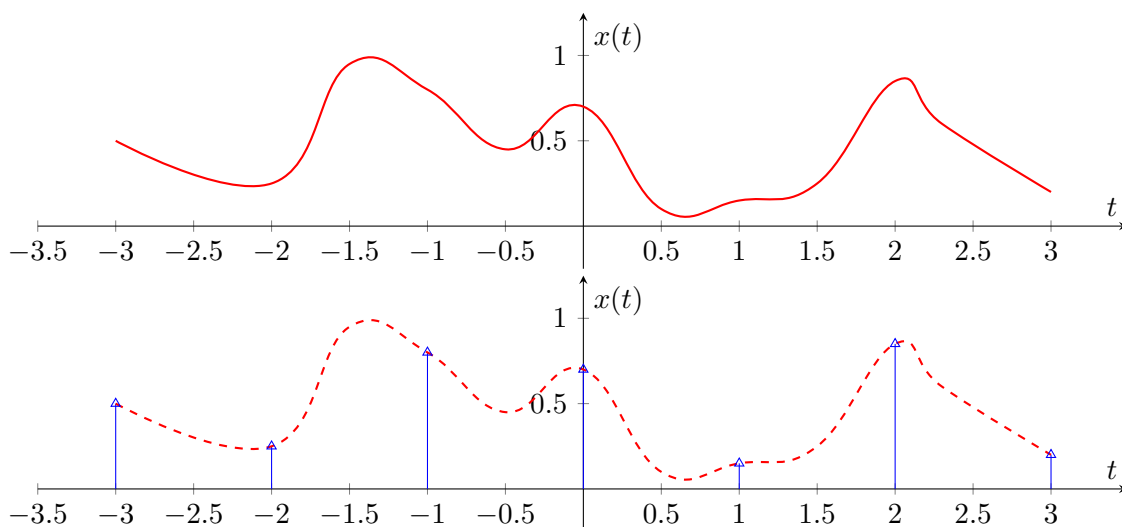
ezért a továbbiakban megpróbáljuk megadni az egyes mintavett jelek spektrumát és az alapján nyilatkozni a mintavételi eljárásunk tulajdonságairól.

3.1.1.1. Dirac-fésűs mintavételezés

Folytonos jel mintavételezésére az egyik lehetőség, hogy egy Dirac-delta impulzus sorozattal, úgynevezett Dirac-fésűvel – melyet a 3.1 egyenletben definiálunk – megszorozzuk az eredeti jelet. Gyakorlatilag ez a lépés tekinthető a mintavételnek matematika szempontból, de erre szükség is van, hiszen ha az eredeti jel egyes diszkrét pontjaiból vennénk mintát, akkor annak Fourier-transzformáltja zérus lenne, ugyanis a mintavett jel csak diszkrét pontokban különbözik nullától.⁴ Ezzel szemben – ahogy azt a 2.2.2 alfejezetben bemutattuk – a Dirac-impulzus integrálja létezik és értéke 1.



3.1. ábra. A Dirac-fésű időtartományi ábrázolása



3.2. ábra. Általános jel mintavételezése Dirac-fésűvel

$$\text{III}_T(t) \triangleq \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \quad \text{ahol, } T_s \text{ az impulzusok közti időköz} \quad (3.1)$$

Ahhoz, hogy a 3.2 szerinti mintavett jel spektrumához eljussunk a Dirac-fésűvel megszorozott eredeti jelünket kell Fourier-transzformálni. Ez kétféle módon is felírható, folytonosan és min-

⁴Szemléletesen: egy darab diszkrét pontnak nincs területe, így az integrál értéke nulla.

tánként (ez utóbbi a rekonstrukció során lesz hasznos), hisz szempontunkból csak az $x(t)$ jel azon értékei érdekesek, melyeket az nT_s időpillanatokban vesz fel.

$$x_s(t) = x(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) \quad (3.2)$$

Ahhoz, hogy a Dirac-fésűt Fourier-transzformálni tudjuk, érdemes azt először Fourier-sorba fejteni, amelyhez a sorfejtés komplex alakját használjuk fel, ahogy a lenti formában látható:

$$\text{III}_T(t) = \sum_{n=-\infty}^{+\infty} X_n^c \cdot e^{j2\pi n \frac{t}{T_s}}$$

A sorfejtéshez ki kell számolnunk az egyes komplex Fourier-együtthatókat, azonban most általánosan az n -edik együtthatót számoljuk ki.

$$X_n^c = \frac{1}{T_s} \int_0^{T_s} \text{III}_T(t) e^{j2\pi n \frac{t}{T_s}} dt$$

Mivel a jel periodikus, tetszőleges intervallumon integrálhatunk, így célszerű választás a $[-\frac{T_s}{2}; \frac{T_s}{2}]$ intervallum, hisz a Dirac-fésű itt csak egy darab Dirac-deltát tartalmaz, így elég azzal számolni.

$$X_n^c = \frac{1}{T_s} \int_{-\frac{T_s}{2}}^{\frac{T_s}{2}} \delta(t) e^{j2\pi n \frac{t}{T_s}} dt = \frac{1}{T_s} \int_{-\frac{T_s}{2}}^{\frac{T_s}{2}} \delta(t - 0) e^{j2\pi n \frac{t}{T_s}} dt$$

Jó hír, hogy az integrált sem kell kiszámolni, hisz felhasználhatjuk a Dirac-delta – a 2.2.2 fejezetben bemutatott – eltolási tulajdonságát. A könnyebb láthatóság miatt explicit is megadtuk a zérussal való eltolást.

$$X_n^c = \frac{1}{T_s} \underbrace{e^{j2\pi n \frac{0}{T_s}}}_1 = \frac{1}{T_s}$$

A fenti levezetés értelmében az összes Fourier-együttható értéke megegyezik, így a Dirac-fésű sorfejtése előáll a 3.3 szerinti alakban.

$$\text{III}_T(t) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} e^{j2\pi n \frac{t}{T_s}} \quad (3.3)$$

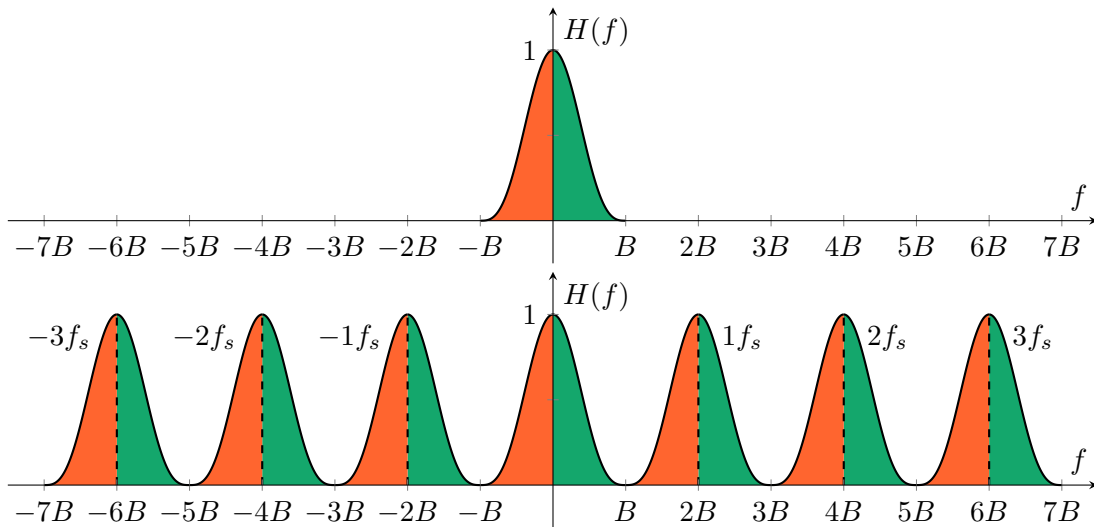
Az impulzussorozat Fourier-transzformáltját a sorfejtés eredményének segítségével tudjuk megadni. Ehhez felhasználjuk a 2.2.1 fejezetben ismertetett modulációs tételt, illetve, hogy a dualitás miatt, az 1 konstans transzformáltja Dirac-delta. A transzformáció eredményeképp eljutottunk a Dirac-fésű spektrumának 3.4 szerinti felírásához. Ezen eredmények felhasználásával előállítható a mintavételezett jel spektruma az elemi jel Fourier-transzformáltja és a Dirac-fésű Fourier-transzformáltjának konvolúciójaként, a 3.5 szerinti formában.

$$\begin{aligned}\mathbb{I}\mathbb{I}_T(f) &= \int_{-\infty}^{\infty} \underbrace{\frac{1}{T_s} \cdot e^{j2\pi n \frac{t}{T_s}} \cdot e^{-j2\pi ft}}_{\mathbb{I}\mathbb{I}_T(t)} dt \\ &= \frac{1}{T_s} \int_{-\infty}^{\infty} e^{-j2\pi ft} \cdot e^{j2\pi t \frac{n}{T_s}} dt \\ &= \frac{1}{T_s} \mathcal{F}\{1(t) \cdot e^{j2\pi t \frac{n}{T_s}}\}\end{aligned}$$

$$\mathbb{I}\mathbb{I}_T(f) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} \delta\left(f - \frac{n}{T_s}\right) \quad (3.4)$$

$$X_\delta(f) = \mathcal{F}\{x_\delta(t)\} = X(f) * \mathbb{I}\mathbb{I}_T(f) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} X\left(f - \frac{n}{T_s}\right) \quad (3.5)$$

A spektrális felírás és a 3.4 ábra alapján jól látható, hogy a kapott spektrum periodikus és ezt a tulajdonságát használjuk ki a mintavételi tétel megfogalmazása során. Néhányaknak esetleg szúrhatja a szemét az $\frac{1}{T_s}$ mértékű – a Dirac-fésűvel való szorzás jóvoltából megjelenő – skálázás, amely számos reprezentációból hiányzik. Ennek oka igen prózai: ez a mintavételezés szempontjából nem fontos. Egyrészt a hangsúly kizárólag a spektrum periodicitására esik, másrészt mivel konstans szorzó, így kiküszöbölhető, ahogy azt a 3.1.1.3 alfejezetben látni fogjuk.



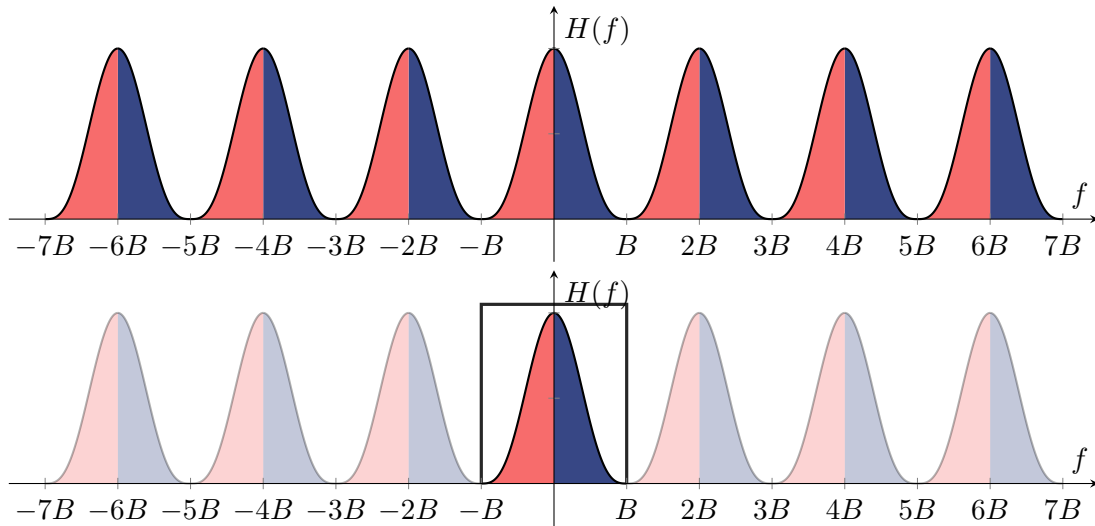
3.3. ábra. Mintavételezett jel spektruma

3.1.1.2. A mintavételi tétel

A Dirac-fésűs mintavételezés tanulságai szerint, tehát a mintavételezett jel spektruma az eredeti jel spektrumának $\frac{1}{T_s}$ távolságban vett ismétléséből áll elő. Ebből tehát az következik, hogy az

eredeti jelünk maradéktalanul helyreállítható, csupán annyit kell tennünk, hogy az ismétlődő spektrumszeleteket el kell távolítanunk egy aluláteresztő szűrővel, hogy az eredeti jel spektrumához jussunk. Nyilvánvaló, hogy ennek a műveletnek feltétele az, hogy az egyes ismétlődések során a spektrumszeletek ne lapolódjanak át. Ha, abból indulunk ki, hogy a kétoldalas spektrum $\frac{1}{T_s}$ frekvenciaközönként ismétlődik, akkor az azt jelenti, hogy $\frac{1}{2T_s}$ frekvencián már az eredeti jel spektruma zérus kell hogy legyen, hisz ott kezdődik az első ismétlődés. A fenti megfontolások alapján már felírható az úgynevezett – Shannon és Nyquist nevéhez köthető – **mintavételi tétel**:

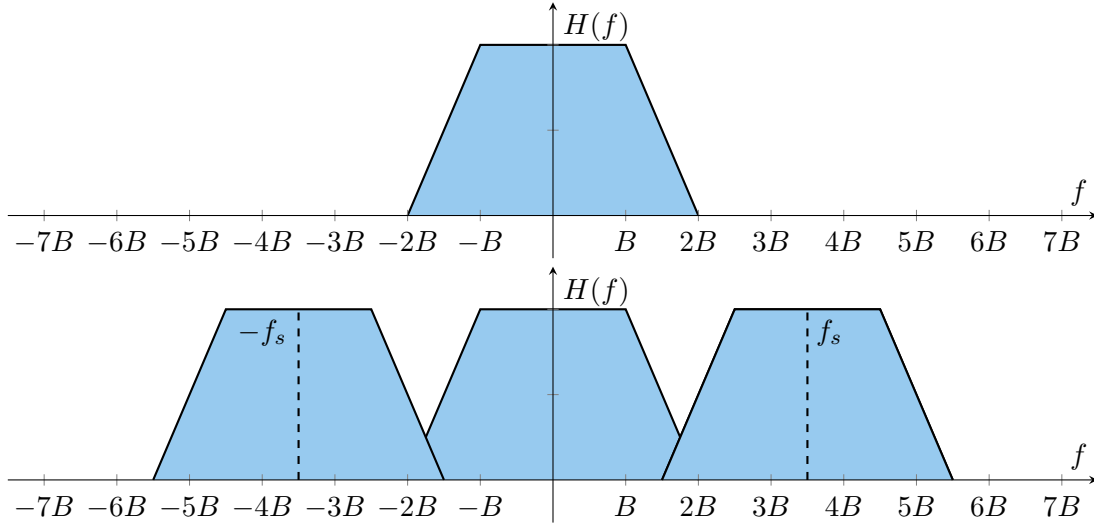
3.1 Tétel: *Ha egy jel Fourier-transzformáltja sávkorlátozott, azaz $X(f) = 0$, ha $|f| \geq B$, akkor $f_s = \frac{1}{T_s} \geq 2B$, azaz $T_s \leq \frac{1}{2B}$ esetén a folytonos időfüggvény hibátlanul visszaállítható (azaz a mintavételezéssel nem veszítünk információt).*



3.4. ábra. Jelrekonstrukció aluláteresztő szűrővel

Fontos megjegyezni, hogy a spektrum már $f = B$ értéknél sem lehet zérustól különböző értékű. Ennek jelentősége különösen lényeges a mintavételi tétel bizonyos alkalmazásainál és félrevezető lehet például szinusz jel mintavételezésénél. Ugyan egy véges függvényérték nem befolyásolná az inverz Fourier-transzformációt, azonban egy Dirac-impulzus (például $f = B$ frekvenciájú szinusz jel) estében súlyos hiba lenne, ha $f_s = 2B$ frekvenciával mintavételeznénk. Tehát szinusz jel esetében a mintavételi tétel az alábbiak szerint alkalmazható: $f_s > 2B$.

A mintavételi tétel alkalmazása során az egyik legjelentősebb akadály, hogy az összes gyakorlati jel időkorlátos, azaz nincsenek jelen az idő kezdete óta annak végéig. Ez azt is jelenti, hogy ezen jelek egyike sem sávkorlátolt, így a mintavételi tételt nem lehet betartani. Ez a későbbiekben a legkülönbözőbb gyakorlati problémákhoz vezethet, melyre a 3.1.1.5 fejezetben ki is térünk. Egy másik jelentős probléma, hogy a visszaállító, aluláteresztő szűrő impulzusválasza nem belépő jel és mint ilyen akauzális. Ez azt jelenti, hogy az ideális szűrő nem realizálható, vagyis az alkalmazott szűrők zárótartományi csillapítása véges, ami szintén számos probléma forrása.



3.5. ábra. Alulmintavételezett jel spektruma

3.1.1.3. Jelrekonstrukció

Az előzőekben egy folytonos, sávkorlátos jel időtartományi mintavételezése során előállt a mintasorozat spektruma. A célunk most az, hogy ezen spektrum alapján visszaállítsuk az időtartománybeli jelünket, méghozzá maradéktalanul.

Az elvet az előzőekben már meghatároztuk: mivel a rendelkezésre álló spektrum az eredeti jel spektrumának ismétlődéseiből áll – ahogy az a 3.5 egyenletből kiderült –, így az ismétlődő spektrumszeleteket egy aluláteresztő szűrővel levágva az eredeti jel spektrumához jutunk el, melyet inverz Fourier-transzformálva visszakapjuk az eredeti időtartományi jelünket.

Természetesen, nem hagyunk senkit az út szélén, így a mintavételezés során megjelenő $\frac{1}{T_s}$ szorzót is elimináljuk, oly módon, hogy az aluláteresztő szűrő átvitele nem egységnyi, hanem T_s -nyi lesz. Az ideális, B vágási frekvenciájú, T_d késleltetésű aluláteresztő szűrő spektruma egy téglalap⁵, egészen konkrétan:

$$H_{LPF}(f) = \begin{cases} 1 \cdot e^{j2\pi f T_d}, & \text{ha } |f| < B \\ 0, & \text{egyébként} \end{cases} \quad (3.6)$$

Az eredő spektrum a két részspektrum szorzataként áll elő, az alábbi formában (érthető módon csak a $-B$ és B közötti intervallumra felírva):

$$X(f) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} X\left(f - \frac{n}{T_s}\right) \cdot T_s \cdot e^{j2\pi f T_d}$$

A kívánt eredményhez több úton is eljuthatunk, mi a tömörség útján járva, a dualitás elvét használjuk ki, tehát a frekvencia tartományban elvégzett szorzás, időtartományi konvolúcióként való leírását. Ekkor, a konvolúció első operandusa a mintavétel során előálló spektrum inverz Fourier-transzformáltja, amely az eredeti mintavett jelünk időfüggvénye. A másik operandus

⁵Ezt a függvényt az angolszász irodalomban *rect* néven említik.

pedig az ideális aluláteresztő szűrő impulzusválasza:

$$x(t) = \underbrace{\left(\sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) \right)}_{x_s(t)} * \mathcal{F}^{-1}\{H_{LPF}(f)\}$$

A konvolúció elvégzéséhez tehát ismernünk kell az ideális aluláteresztő szűrő impulzusválaszt, amely B vágási frekvencia és T_d késleltetés esetén az alábbi módon írható fel:

$$h_{LPF}(t) = 2B \frac{\sin(2\pi B(t - T_d))}{2\pi B(t - T_d)} = 2B \operatorname{sinc}(2B(t - T_d)) \quad (3.7)$$

Jelen esetben a sinc az úgynevezett normalizált sinc függvény.⁶ A megadott impulzusválaszt $B = \frac{1}{2T_s}$ és $T_d = 0$ helyettesítéssel felhasználva, illetve alkalmazva a 2.23 formulával leírt eltolási tulajdonságát a Dirac-deltának, előállítható a rekonstruált jel időfüggvénye:

$$x(t) = \underbrace{\left(\sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) \right)}_{x_s(t)} * \underbrace{2B \operatorname{sinc}(2Bt)}_{h_{LPF}(t)}$$

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT_s) \operatorname{sinc}\left(\frac{t - nT_s}{T_s}\right) \quad (3.8)$$

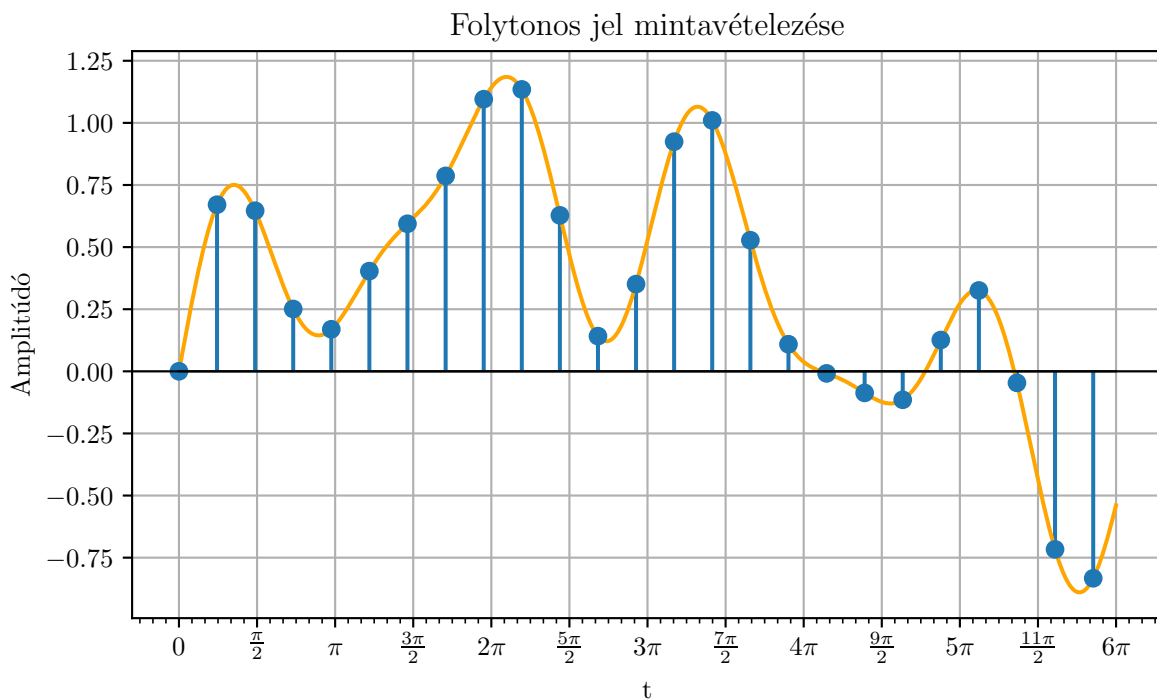
Az így kapott kifejezés az úgynevezett **Whittaker-Shannon interpolációs formula** vagy gyakran csak interpolációs formula, ugyanis a jel mintái alapján interpolálja a minták közti jelértékeket. A 3.6 ábrán látható, ahogy jelből mintákat veszünk, a 3.7 ábrán pedig, hogy a vett minták alapján rekonstruáljuk a jelet. Ahhoz, hogy a rekonstruált jel teljesen megegyezzen az eredeti jellel, az összes mintát fel kell használnunk, azonban a gyakorlatban elég jó közelítést tudunk elérni, ha minél távolabbi $x(nT_s)$ mintákat fel tudunk használni, mivel a sinc függvény a nullától távolodva egyre csökken.

A végső konklúzió tehát az, hogy elméleti úton beláttuk, egy folytonos jelből meghatározott időközönként mintákat véve, az eredeti jel maradéktalanul visszaállítható, a vett mintákból. Persze, *az élet nem habostorta*, így a gyakorlatban csak közelítőleg tudjuk visszaállítani a jelünket annak mintáiból, azonban ez így is egy jelentős eredmény.

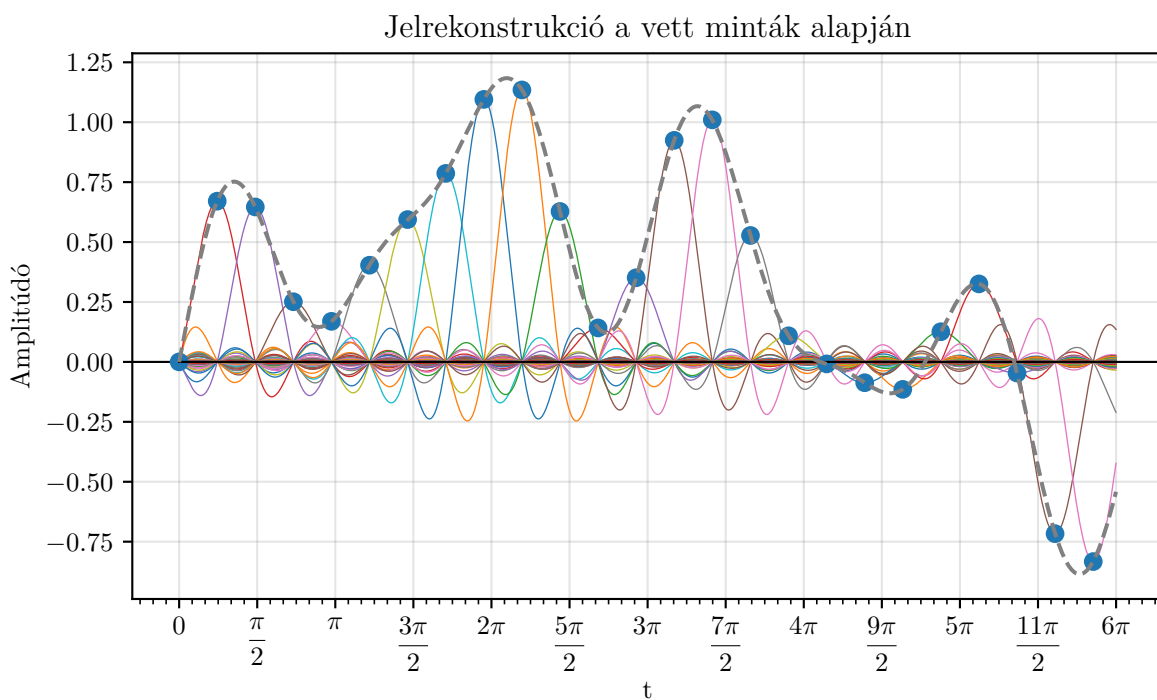
3.1.1.4. Alul és túlmintavételezés

Ugyan a mintavételi tétel precízen megfogalmazza, hogy milyen feltételei vannak a tökéletes rekonstrukciónak, előfordulhatnak estek amely során a mintavételi tétel be nem tartása esetén is tökéletes rekonstrukció végezhető. Ehhez elég egy olyan jelet elképzelnünk melynek spektruma "lyukas", azaz csak bizonyos, összefüggő, zárt tartományok léteznek zérustól különböző spektrális komponensei. Ez esetben találhatunk olyan mintavételi frekvenciát, amely nem tartja be a mintavételi tételt, azonban a mintavételezett jel spektruma mégsem lapolódik át. Ez utóbbi lát-

⁶A normálizált sinc definíciója: $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$

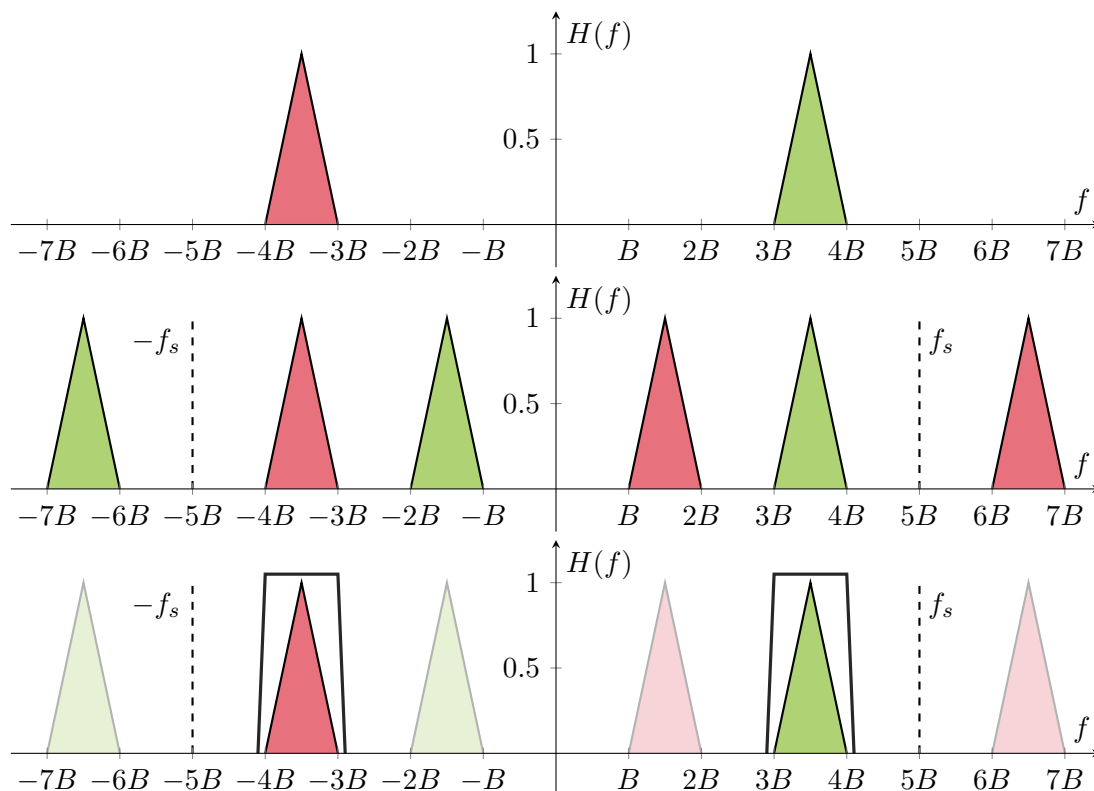


3.6. ábra. Folytonos jel mintavételezése



3.7. ábra. Jelrekonstrukció a vett minták alapján

ható a 3.8 ábrán. Ekkor a jelrekonstrukció során – egy vagy több – sávszűrőt kell alkalmaznunk, ezért is nevezik az alulmintavételezést olykor sávszűrős mintavételezésnek.



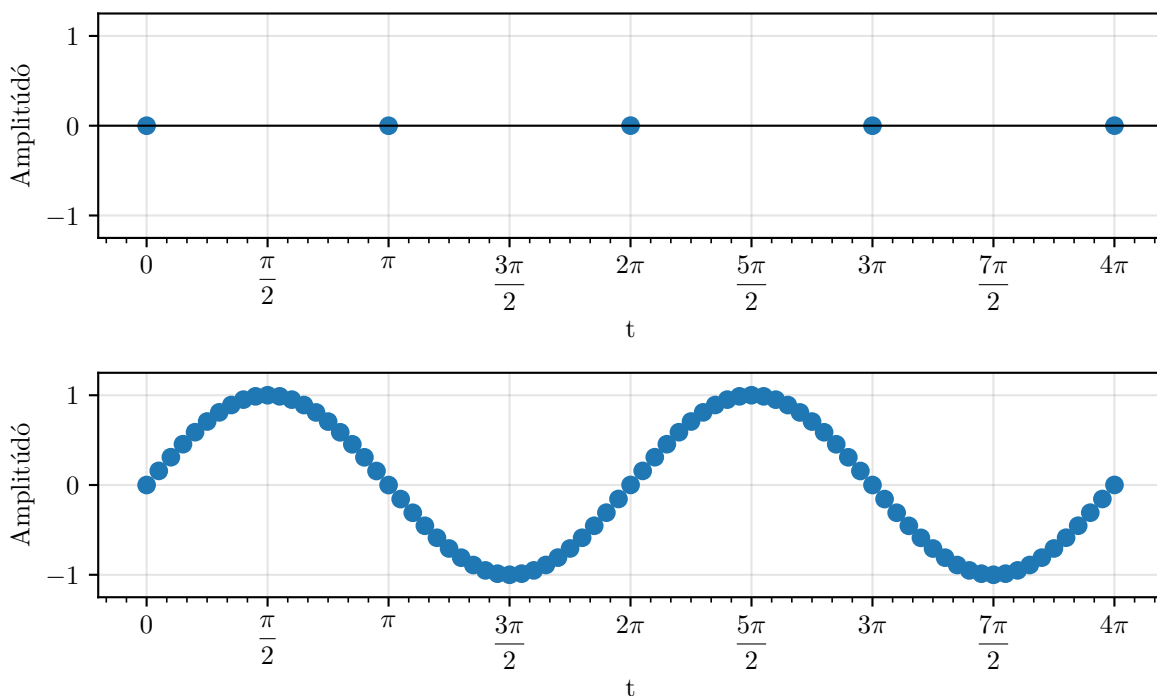
3.8. ábra. Sávszűrős mintavételezés alkalmazása

Természetesen a mintavételi tételtől máshogy is el lehet térni, mégpedig a minimális mintavételi frekvencia értékén felüli mintavételezéssel, ezt nevezzük túlmintavételezésnek. Ez elméleti szempontból nem tűnik logikus választásnak, hisz a célunk minden esetben az, hogy az erőforrás igényét a lehető legalacsonyabban tartsuk egy-egy megoldásnak (ilyenkor lehet hasznos az alulmintavételezés, amennyiben megtehető). A túlmintavételezés szerepe azonban a gyakorlatban jelentős: egyrészt a mintavételezett jel vizuális megjelenítésénél szükséges, hogy túlmintavételezzük a jelet, ahogy a 3.9 ábrán⁷ is látható. Másrészt a túlmintavételezéssel spektrumszeleteket hagyunk szabadon az ismétlődések között, amellyel növelhető a jel-zaj viszony.

3.1.1.5. Aliasing és szivárgás

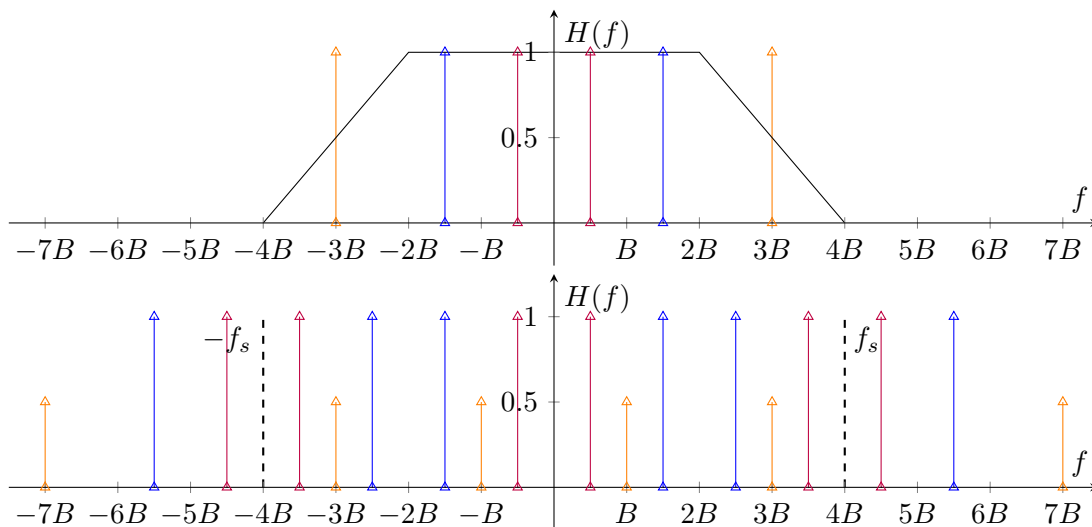
Annak érdekében, hogy a jelünk minél jobban közelítse a megadott sávkorlátot (vagy azt a spektrális képet, amit kapni szeretnénk), gyakran alkalmazunk úgynevezett bemeneti szűrőket. Mivel, ezek a szűrők nem ideálisak így a csillapításuk sem végtelen, tehát bizonyos mértékben átengedi a vágási tartományába eső frekvencia komponenseket – ugyanez igaz a kimeneti szűrőkre is melyeket a jelrekonstrukció során használunk. Ez az ideálistól eltérő viselkedés azt okozza, hogy a képzeletbeli sávkorlátnál – amely szerint meghatározzuk a mintavételi frekvenciát – nagyobb spektrális komponensek is megjelennek a jelben, tehát nem szándékos alulmintavételezéssel állunk szemben.

⁷Az ábra felső szegmenségben a mintavételi frekvencia $f_s = 2,01B$.



3.9. ábra. Folytonos jel elégséges-, és túlmintavételezése

Az alulminta vételzés egyik súlyos tünete az **aliasing**: ekkor, az alulminta vételzés miatt a spektrumszeletek átlapolódnak és oda nem illő úgynevezett áljelek vagy aliasjelek jelennek meg az egyes spektrumszeletekben. Az aliasing a jel-zaj viszony jelentős romlását is okozhatja, attól függően, hogy a bemenő szűrő milyen zárótartományi paraméterekkel (például csillapítással) rendelkezik. Ebben az esetben túlminta vételzés alkalmazása előnyös lehet a jel-zaj viszony kompenzálása érdekében.

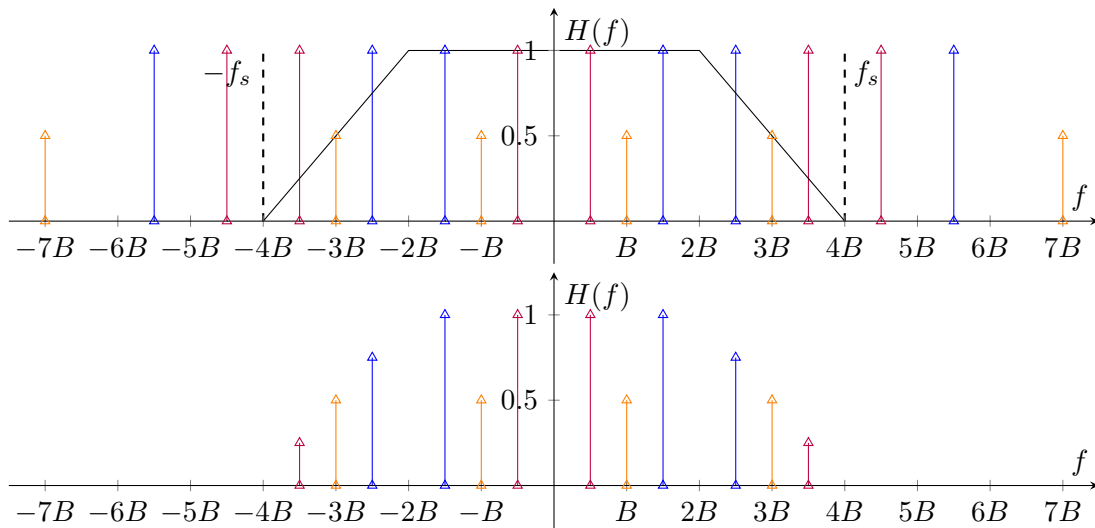


3.10. ábra. Hibás mintavételezés következtében fellépő áljelek

Szintén a nem ideális szűrők problémájához köthető egy másik jelenség is: ez a **spektrum**

szivárgás jelensége. Szivárgás alatt azt értjük, mikor a kimenő vagy helyreállító szűrő nem ideális volta miatt nem biztosítható az alakhú átvitel, mivel az ismétlődő spektrumszeleteket csak bizonyos mértékig csillapítja a szűrő. Ezek a "szivárgó" frekvencia komponensek a jelrekonstrukció során megjelennek a visszaállított időfüggvényben. Az alakhú visszaállítást torzításként vagy zajként tudjuk leírni, tehát a jel-zaj viszony ebben az esetben is csökken, a csillapítás mértéktől függően.

Természetükből fakadóan, az aliasing és a szivárgás kéz a kézben járnak, így jelentősen ronthatnak a jel-zaj viszonyon. A 3.10 és 3.11 ábrákon látható szűrő viszonylag nagy átmeneti tartománnyal rendelkezik illetve, zárótartományban teljesen elnyom, azonban még így is jelentősen romlik a jel-zaj viszony a megjelenő áljelek és szivárgó komponensek következtében (az ábrákon csak az f_s mintavételből adódó komponenseket tüntettük fel, f_s többszöröseit nem). Túlmintavételezéssel, illetve jó minőségű bemenő (antialiasing) és visszaállító szűrők használatával azonban ez mérsékelhető.



3.11. ábra. Hibás rekonstrukció után fellépő szivárgó komponensek

4. fejezet

Kódolástechnika

Ahogy az 1.1.1 fejezetben már tárgyaltuk, a digitális jel egyik nagy előnye, hogy megbízható módon tömöríthető, módosítható, illetve titkosítható. Ezen tulajdonságok kihasználásával az információtovábbítás és -tárolás folyamata sokkal hatékonyabbá tehető: egyrészt az információ adott esetben tömörebben leírható, továbbá a jelfolyamot módosítva (kiterjesztve) növelhető annak hibatűrése. Egyszerű, de szemléletes példa erre a késleltetés kérdése. Ha a kommunikáció során sérül az átvitt üzenet, akkor azt újra kell küldeni, ami további időt vesz igénybe. Azonban, ha a fogadó fél a hibásan vett üzenetből elő tudja állítani az eredeti üzenetet, akkor az újraküldés és az ezzel járó késleltetés elkerülhető.

Ebben a fejezetben kizárólag diszkrét jelforrásokkal fogunk foglalkozni. Először is bevezetjük az információelmélet alapfogalmait, melyekre a későbbiekben támaszkodni fogunk, majd ezekből kiindulva tárgyaljuk a veszteségmentes forráskódolás különböző algoritmusait, illetve a hibajavító kódolás elméletét és gyakorlati aspektusait.

4.1. Az információelmélet alapjai

Shannon 1948-as *A Mathematical Theory of Communication* munkája óta az információ, korábban meglehetősen zavaros természete megváltozott: jól definiált, mérhető mennyiséggé alakult át. Az információ matematikai eszközökkel való leírása lehetővé tette számos új, az információátvitelhez köthető elméleti eredmény elérését. Ezen eredményeket felhasználva megannyi eljárás és technika (például a kódolás) vált egzakttá, melyek alapjaiban változtatták meg a hírközlési eljárásokat és módszereket.¹

Az információ értelmezéséhez a továbbiakban valószínűségi megközelítést alkalmazunk. Ez alatt azt értjük, hogy az információ "értéke" attól függ, hogy mennyire váratlan az információ maga. Például, a „ZH-ban új típusú feladatok várhatók” üzenet információtartalma extrém magas, ahogy az általa kiváltott érzelmi hatások is, melyeket a továbbiakban elhanyagolunk. Ahhoz, hogy értelmezni tudjuk az információt magát, először meg kell vizsgálnunk, hogy az miből is származik.

¹Ehhez, természetesen szükség volt a digitális technika rohamos fejlődésére is.

4.1.1. Diszkrét források leírása

Shannon szavaival élve: „Az információ a bizonytalanság mértéke”, az információ tehát nem önmagában létezik, hanem *valami* hordozza azt. Általánosan szólva, a *valamiket* nevezzük **eseményeknek**, azaz minél vártalanabb egy esemény annál magasabb az információ tartalma. Eseményeket az úgynevezett **források** bocsátanak ki magukból, mely források lehetnek **diszkrét** vagy **folytonosak**. A továbbiakban kizárólag diszkrét forrásokkal fogunk foglalkozni.

A diszkrét források által kibocsátott elemi eseményeket a továbbiakban **szimbólumnak** nevezzük. A generált szimbólumok az $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ véges halmazból kerülnek ki, amit **forrásábécének** nevezünk, elemei pedig a szimbólumok vagy **betűk**. A forrás jellemzéséhez tartozik az úgynevezett **forráseloszlás**, mely azt adja meg, hogy az egyes szimbólumokat milyen valószínűséggel generálja a forrás. A forrás jellegéből adódóan ez egy diszkrét valószínűség-eloszlás, melyet az alábbi formában adunk meg:

$$P(\mathcal{X}) = \{p_1, p_2, \dots, p_n\}, \quad \text{ahol } \sum_{i=1}^n p_i = 1$$

A fenti jelölésekkel tehát: $P(\mathcal{X})$ az \mathcal{X} forrásábécéjű forrás eloszlása és p_n az x_n szimbólumhoz tartozó valószínűség. Annak érdekében, hogy egyszerűbbé tegyük a formuláinkat bevezetjük az X diszkrét valószínűségi változót, mely az \mathcal{X} halmazból veszi elemeit. Továbbá vezessük be a $p: \mathcal{X} \rightarrow [0; 1]$ függvényt, az alábbi módon:

$$p(x) = P\{X = x\}, \quad x \in \mathcal{X}$$

Ennek értelmében a $p(x)$ függvény az x szimbólumhoz (forrásbetűhöz) rendeli annak valószínűségét.

A jegyzetben tárgyalt források fontos tulajdonsága, hogy kivétel nélkül **stacionárius, memóriamentes** források, tehát az egyes p_i valószínűségek nem függenek az időtől, illetve az egyes elemi események függetlenek egymástól.

Az eddigiek ismeretében bevezethetjük az **információ** vagy **információtartalom** fogalmát, mint egy esemény váratlanságának mértékét a 4.1 szerinti módon²:

$$I(x_i) = -\log p_i = \log \frac{1}{p_i} \quad (4.1)$$

Miután alaposan szemügyre vettük a fenti a formulát, az a – mellesleg teljesen jogos – érzésünk támadhat, hogy nem tudtunk meg sokat az információ mibenlétéről. Mielőtt azonban végérvényesen elmenne a kedvünk a további olvasástól, érdemes kicsit alaposabban is körüljárni a benne foglaltakat. A legalapvetőbb megállapítást már úgymond „elspoilereztek”: az információtartalom a szimbólum valószínűségével van összefüggésben; minél kisebb a valószínűsége, hogy a szimbólum megjelenik, annál jobban meglepődünk, ha ez bekövetkezik. Ezért az infor-

²Annak ellenére, hogy a $-\log x$ és a $\log \frac{1}{x}$ kifejezések ekvivalensek, a legtöbb információelméleti könyv az előbbihez ragaszkodik. Mi a jegyzetben végig az utóbbit fogjuk használni.

mációtartalmat, a meglepettség mértéknek is nevezhetjük.

A logaritmus szerepét Shannon több különböző módon is indokolta, melyek közül a mi szempontunkból az additivitás kérdése a legfontosabb. Két független esemény egyidejű bekövetkezésének valószínűsége, a két elemi valószínűség szorzata, a logaritmus használatával pedig két független esemény információ tartalma összeadódik, ami az információ természetéből fakadóan is logikusnak tűnik. Az, hogy milyen logaritmust használunk lényegtelen, hisz a különböző alapú logaritmusok között könnyedén átválthatunk, azonban konvenció szerint kettes alapú logaritmust fogunk használni a jegyzetben végig, tehát végig a $\log = \log_2$ jelölést alkalmazzuk.³ Az információ mértékegysége a használt logaritmus alapjától függ, ez kettes alapú logaritmusnál a **bit**, míg tízes alapúnál a **hartley**. Fontos kiemelni, hogy a *bit*, mint információtartalom nem feleltethető meg egy az egyben a számítástechnikában használatos bitnek (*binary digit*), de a későbbiekben láthatjuk, hogy a kettő között szoros kapcsolat áll fenn. Ennek oka, hogy egy biten (*binary digit*) maximum 1 bit információ tárolható.

Fontos még egyszer kihangsúlyozni a fenti állításokat, vagyis hogy az információtartalom pusztán az egyes események valószínűségétől és nem azok értékétől függ. Tegyük fel, hogy a forrásunk nem mást reprezentál, mint egy érme feldobását, aminek kimenete lehet fej vagy írás, azonban az érme "cinkelt", ezért $p_0 = 90\%$ valószínűséggel lesz fej és $p_1 = 10\%$ valószínűséggel írás. Kiszámolva az egyes információ tartalmakat rendre $I(x_0) = 0,15$ bit és $I(x_1) = 3,32$ bit az eredmény. Ismét lényeges, hogy az információ nem váltható át egy az egyben digiteké, hisz nyilvánvalóan nincs szükségünk összesen öt digitre, a két esemény által hordozott leírásához – a számított értékek, csak azt mutatják, hogy mennyi a két esemény "hírértéke". Természetesen, ezt a hírértéket előbb-utóbb valóban szeretnénk lekódolni fizikailag, de ezt nem az egyes események egyenkénti kódolásával fogjuk elérni. Ezt az elsőre furcsának tűnő jelenséget a következő fejezetek során részletesen tárgyalni fogjuk.

4.1.2. Az entrópia

Az információelmélet egyik alapvető fogalma, melyre mi is sokszor fogunk hivatkozni a továbbiakban, az **entrópia**.

4.1 Definíció: Az X valószínűségi változó $H(X)$ entrópiáját az alábbi összeggel definiáljuk:

$$H(X) = \sum_{x \in \mathcal{X}} p(x) I(x) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}$$

Az entrópia kapcsán egy további konvenciót is fel kell vennünk a képzeletbeli listánkra, miszerint:

$$0 \log 0 = 0, \quad \text{mivel } x \log x \rightarrow 0, \quad \text{ha } x \rightarrow 0$$

Ez alapján egy zérus valószínűségű esemény nem módosítja az entrópiát, ami logikus, hisz azokból végtelen sokat fel tudnánk venni.

³Német nyelvterületen, bizonyos angol nyelvű írásokban és a tárgy korábbi segédanyagai között a kettes alapú logaritmust "ld"-vel jelölik. Mi a fenti konvenciót használjuk a jegyzet során végig, de valójában ez ízlés kérdése.

Az entrópia szemléletesen a valószínűségi változó értékének, vagyis az általa reprezentált kísérletek kimenetelének bizonytalanságát jelöli. Természetesen kapcsolatba hozható a termodinamikai entrópia fogalommal is, mint a rendezetlenség mértéke, de erre részleteiben nem térünk ki. Az entrópia fogalma axiomatikusan is bevezethető, egy diszkrét valószínűségi változó entrópiájára vonatkozó tulajdonságok megkövetelésével, azonban a mi szempontunkból az értelmezésnél fontosabb, hogy a későbbiekben számos kódolási, tömörítési kérdés megválaszolása az entrópiához köthető.

Az entrópia definíciójából közvetlenül is belátható a következő, az entrópia nem negatív voltára vonatkozó lemma:

4.2 Lemma: $H(X) \geq 0$.

BIZONYÍTÁS: Mivel, $0 \leq p(x) \leq 1$, ezért $p(x) \log \frac{1}{p(x)} \geq 0$. □

Ahhoz, hogy jobban rálássunk az entrópia mibenlétére és jelentőségére, érdemes ismét megvizsgálni a 4.1.1 alfejezet végén felvetett kérdést. A feladat ugyanaz, a cinkelt érménk $p_0 = 90\%$ valószínűséggel lesz fej és $p_1 = 10\%$ írás. Most számoljuk ki ennek a forrásnak az entrópiáját:

$$H(X) = p_0 \log \frac{1}{p_0} + p_1 \log \frac{1}{p_1} = 0.1 \log 10 + 0.9 \log \frac{10}{9} = 0.4690 \text{ bit}$$

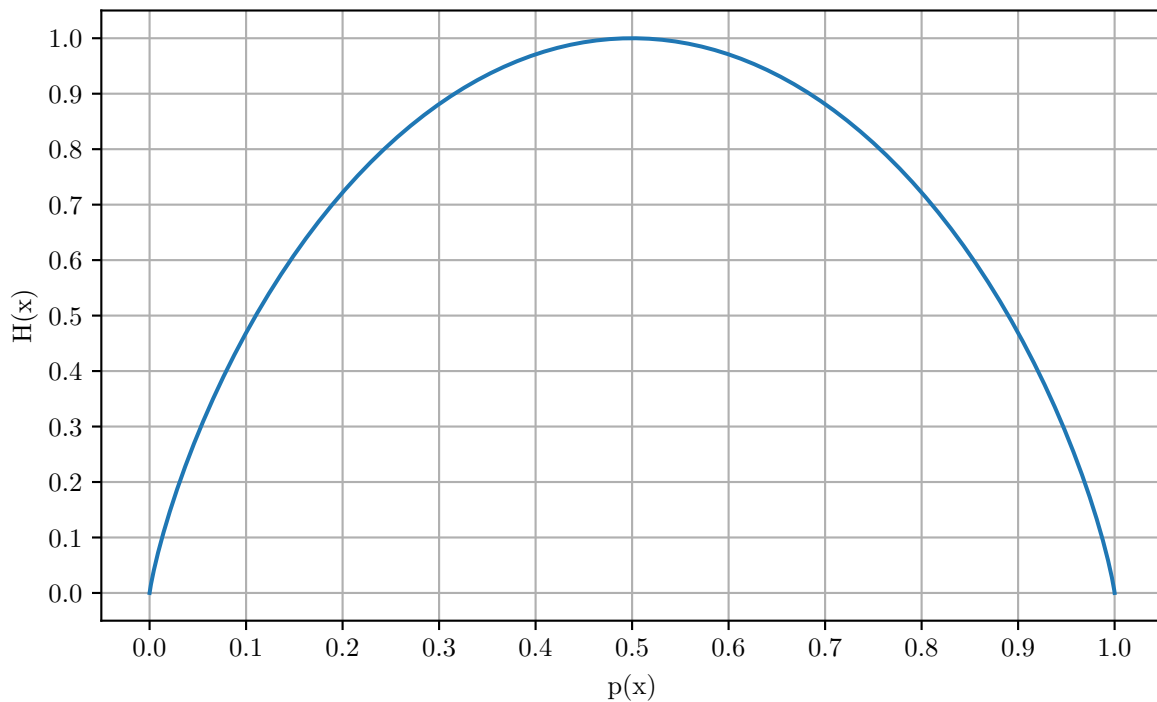
Ahogy arra számítani lehetett, a várható átlagos információtartalma a forrásunknak a két esemény saját információtartalma közé esik, nyilván a nagyobb valószínűségű eseményéhez közelebb. A korábbi okfejtésünk végén azzal szembesültünk, hogy az események egyenkénti leírása az információtartalom alapján szimplán értelmetlennek tűnt. Az egyik ok emögött, hogy az események közt a forráseloszlás teremt kapcsolatot, tehát önállóan az eseményekkel – szempontunkból – nem érdemes foglalkozni.⁴ Az entrópia ezt a furcsaságot oldja fel azzal, hogy megadja a forrás várható, átlagos információtartalmát, így rendelkezünk egy olyan mérőszámmal, amely az események összességéről állít valamit. Az entrópia ezen tulajdonsága az, ami miatt számos kódolás tételben fontos szerepet játszik.

Az entrópia definíciója alapján felrajzolható a 4.1 ábra, amely az entrópia változását ábrázolja a forráseloszlás függvényében, egy két eseményt kibocsátó forrás esetére. Eszerint az entrópia akkor maximális, ha az egyes események valószínűsége megegyezik, ami logikus, hisz az átlagos információtartalom akkor a legmagasabb.

4.2. Forráskódolás

Manapság az információtárolás, illetve -átvitel szinte elképzelhetetlen forráskódolási eljárások nélkül. Kódolásnak nevezzük, mikor egy forrás által generált szimbólumot kódszavakkal írunk le – például egy bitsorozattal. Forráskódolás alatt lényegében tömörítési eljárásokat értünk, melyekből létezik veszteségmentes és veszteséges is: veszteségmentes esetben az eredeti információ

⁴Például az a hír, hogy „ma nem fognak betonszamarak aláhullani az égből” csak akkor értékes számunkra, ha tudjuk, hogy mekkora a valószínűsége annak, hogy „ma betonszamarak fognak aláhullani az égből”.



4.1. ábra. Az entrópia alakulása két eseményű források esetén

visszállítható, míg veszteséges esetben ez nem tehető meg, ekkor információvesztés lép fel. Veszteségmentes eljárás például a *zip*, míg veszteséges az *mp3* és a *jpeg*. A továbbiakban főként szimbólumok (veszteségmentes) forráskódolásával foglalkozunk, melynek alapötlete az esetleges redundanciák eliminálása, így adott információ tartam tömörebben vihető át a csatornán.

4.2.1. Kódolási alapfogalmak

Első lépésként ismerkedjünk meg magával a **kódolás** fogalmával. Tekintsük az X valószínűségi változóval leírt forrást, illetve legyen \mathcal{Y} egy s elemű $\mathcal{Y} = \{y_1, y_2, \dots, y_s\}$ halmaz, melyet **kód-ábécének** nevezünk. Jelölje \mathcal{Y}^* az \mathcal{Y} elemeiből álló véges sorozatok halmazát, melynek elemeit **kódszavaknak** hívjuk. Egy $f : \mathcal{X} \rightarrow \mathcal{Y}^*$ leképezést (függvényt) **kódnak** nevezünk (a továbbiakban ezt C -vel jelöljük), az \mathcal{X} elemeiből alkotott véges sorozatokat pedig **üzeneteknek**. Továbbá, legyen $C(x)$ az x szimbólumhoz tartozó kódszó, $l(x)$ pedig a $C(x)$ kódszó hossza.

A forráskódolás célja, hogy a diszkrét forrásuk szimbólumait hatékonyan kódolva átvigyük egy csatornán, majd a vevő oldalon a kódot dekódolva megkapjuk a kódszavak által hordozott információt. Elsőre az a benyomásunk támadhat, hogy a kódolás folyamata jóval egyszerűbb, mint a dekódolás, hisz amíg kódolásnál egyértelmű, hogy melyik szimbólumot melyik kódszóba képezzük le, addig fordítva ez nem biztos, hogy igaz. Ezen a gondolaton továbbhaladva, vizsgáljuk meg az egyes kódok dekódolhatóságának kérdését.

4.3 Definíció: Egy C kódot **nem szingulárisnak** nevezünk, ha X minden értékét különböző \mathcal{Y}^* -beli betűsorozatokra képezzük le, azaz

$$x \neq x' \implies C(x) \neq C(x').$$

Ezzel tulajdonképpen azt írtuk elő, hogy a kódoló leképezés kölcsönösen egyértelmű legyen (ami többet jelent, mint a függvény fogalom), így a kódoló függvény invertálásával a kódunk dekódolhatóvá válik. Ezzel akár le is zárhatnánk az ügyet, de ez esetben a kódszavainkat küldés során elválasztva kell küldeni. Az elválasztás lehet egy konkrét elválasztó karakter vagy akár csak várakozás a kódszavak elküldése közt. Könnyen belátható, hogy ez egy szuboptimális megoldás, hisz a célunk az, hogy megszakítás nélkül egybefüggő karakterfolyamként továbbíthassuk a kódszavakat.

4.4 Definíció: A C kód C^* kiterjesztése egy leképezés, mely \mathcal{X} elemeiből alkotott véges sorozatokhoz rendel hozzá \mathcal{Y}^* -beli véges sorozatokat, az alábbiak szerint:

$$C(x_1x_2 \dots x_n) = C(x_1)C(x_2) \dots C(x_n)$$

ahol, $C(x_1)C(x_2) \dots C(x_n)$ az egyes kódszavak konkatenációját (egybefűzését) jelenti.

A kódszavak adatfolyamként való továbbítása során tehát egy adott kód kiterjesztését szükséges dekódolni a vevő oldalon. Ez azt is jelenti, hogy egy adott kód dekódolhatósága az attól függ, hogy annak kiterjesztése egyértelműen képezi le az üzeneteket kódszófolyamokká.

4.5 Definíció: Egy C kódot egyértelműen dekódolhatónak nevezünk, ha annak kiterjesztése nem szinguláris.

A fenti definíciókat összegezve jól látható, hogy egy kód nem attól dekódolható egyértelműen, hogy az egyes kódszavakat önmagukban képesek vagyunk visszafejteni. Ennek jelentősége éppen az – ahogy korábban már említettük –, hogyha egy csatornán egymás után küldjük a kódszavakat, akkor nem kell azokat elválasztani (pl.: várakozással, vagy elválasztó karakterrel), mert úgy is képesek vagyunk dekódolni az üzeneteket, hogy a kódszavaink "egybefolyanak". Ekkor viszont az a nehézség léphet fel, hogy nem mindig tudjuk eldönteni, hogy egy kódszónak mikor is van vége. Ez azt jelenti, hogy "előre kell néznünk" (*lookahead*) és ha felismerünk egy kódszót, akkor meg tudjuk határozni az előző kódszó végét is. Ez a dekódolási mechanizmus nyilvánvalóan nem a leghatékonyabb, ezért a következő fejezetben annak járunk utána, hogy lehetséges-e a kódszófolyam dekódolása lookahead nélkül.

4.2.2. Prefix-kódok

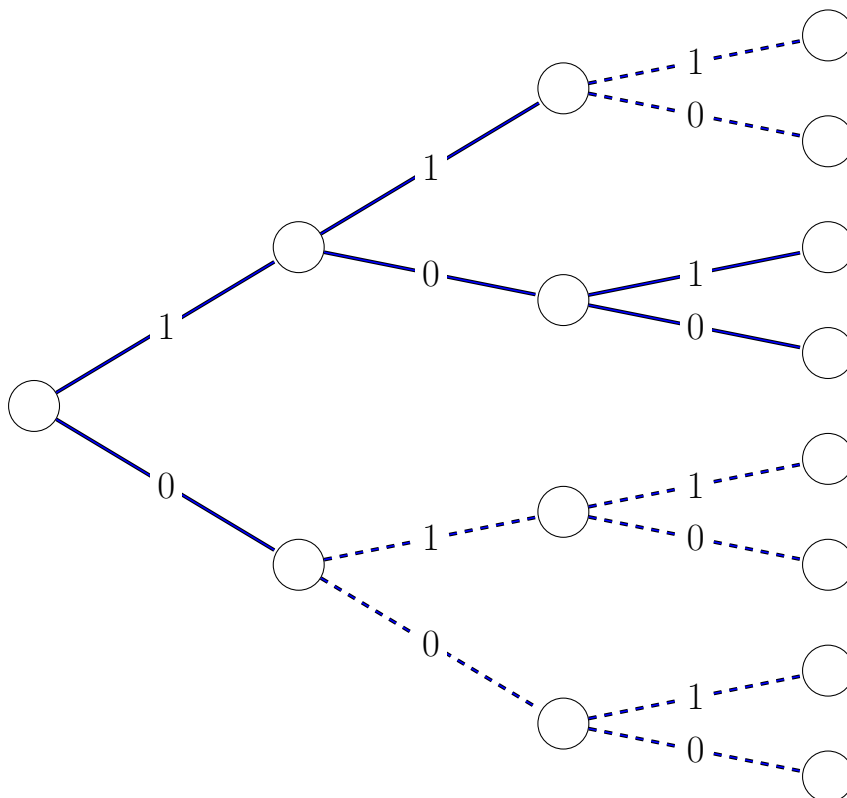
4.6 Definíció: Egy kódot **prefixmentes** kódnak vagy röviden **prefix** kódnak nevezünk, ha a lehetséges kódszavak közül egyik kódszó sem folytatása egy másik kódszónak. Minden prefix kód egyértelműen dekódolható.

A prefixmentesség egy szimpatikus tulajdonság, hisz nincs szükség referenciára, hogy egy vett karaktersorozatról eldöntsük, hogy kódszó-e vagy sem: ez pedig jelentősen gyorsíthatja a feldolgozás menetét, hisz amint beérkezett az üzenet utolsó betűje már dekódolhatjuk is. A kérdés az, hogy mi a hátránya annak, ha a kódtól az egyértelmű dekódolhatóságon kívül még a prefixmentességet is elvárjuk.

4.7 Lemma: (Kraft egyenlőtlenség) *Bármely, egy s elemű ábécé felett értelmezett prefix kód, l_1, l_2, \dots, l_n kódszóhosszokkal kielégíti az alábbi egyenlőtlenséget:*

$$\sum_i s^{-l_i} \leq 1$$

BIZONYÍTÁS: *Konstruáljunk egy s -ed fokú fát (ezt bármely prefix kódra megtehetjük), ez legyen A , így minden csomópontnak s gyereke van. Továbbá tegyük fel, hogy $l_1 \leq l_2 \leq \dots \leq l_n$. Legyen a fa minden ága (tehát a gyökér és az egyes csúcsok közti utak), egy nem feltétlenül érvényes kódszóhoz tartozó betűsorozat. Ezáltal a fa minden levele egy érvényes kódszót reprezentál, ugyanis prefix feltétel miatt, egy érvényes kódszó sem lehet szülője másik érvényes kódszónak a fában, így ha egy csúcs kódszó, akkor leszármazottja már nem lehet érvényes kódszó. (Bináris prefix kódból konstruált fára mutat példát a 4.2 ábra.)*



4.2. ábra. Prefix kódot reprezentáló bináris fa

Legyen l_{\max} a leghosszabb érvényes kódszó hossza a fában. Egy l_i mélységű/hosszú kódszónak $s^{l_{\max}-l_i}$ utódja lehetne l_{\max} mélységben, amelyek érvénytelenek. Ezt átfogalmazva: egy érvényes

kódszó l_i mélységben $s^{l_{\max}-l_i}$ csomópontot zár ki. Az egyes kódszavak által kizárt csomópontok halmaza (amely egy részfa) diszjunkt. Triviális, hogy legfeljebb $s^{l_{\max}}$ csomópont zárható ki. Az előző két megállapítás értelmében, azt mondhatjuk, hogy:

$$\sum_i s^{l_{\max}-l_i} \leq s^{l_{\max}}$$

Ezt a kifejezést átalakítva, már könnyen eljutunk a Kraft-egyenlőtlenségig:

$$\sum_i s^{l_{\max}} s^{-l_i} \leq s^{l_{\max}}$$

$$\sum_i s^{-l_i} \leq 1$$

□

4.8 Lemma: (McMillan) Minden egyértelműen dekódolható $f : \mathcal{X} \rightarrow \mathcal{Y}^*$ kódra

$$\sum_i s^{-|f(x_i)|} \leq 1$$

ahol, s a kódábécé elemszáma, $|f(x_i)|$ pedig az $f(x_i)$ kódszó hossza.

A fenti eredményeket összefoglalva azt mondhatjuk, hogy minden egyértelműen dekódolható kódhoz konstruálható vele ekvivalens prefix kód is. Ezeket a lemmákat összevonva **Kraft-McMillian tételként** szoktuk hivatkozni. Ez azt is jelenti, hogy nem veszítünk semmit, ha a kódunktól a speciálisabb prefix tulajdonságot követeljük meg, amely értelemszerűen egyszerűbben kezelhető.

4.2.3. Optimális kódok

Az előző fejezet tanulságai alapján már tudunk könnyen kezelhető, hatékonyan dekódolható kódokat készíteni. Azonban, azt nem állíthatjuk, hogy tudunk jó kódokat konstruálni: egyrészt, sose lennénk ilyen nagyképűek, másrészt a fenti tulajdonságok csak egy részét alkotják a kód minőségi fokmérőinek. A kódunk egyik legfontosabb paramétere annak tömörsége, melyet az **átlagos kódszóhosszal** tudunk leírni.

4.9 Definíció: A C kód átlagos kódszóhossza az

$$L(C) = \mathbb{E}C(x) = \sum_{x \in \mathcal{X}} p(x)l(x)$$

várható értéket értjük.

Általában elmondható, hogy *nem a méret a lényeg*, azonban ez most nem igaz. Az átlagos kódszóhossz a kód hatékonyságának egyik legfontosabb mérőszáma és célunk az, hogy ezt minél

alacsonyabban tartjuk, ezért a továbbiakban arra teszünk kísérletet, hogy minél rövidebb kódokat konstruáljunk. Ehhez az első lépés, hogy megvizsgáljuk, mekkora lehet az elméleti alsó határa a kód tömörségének.

4.10 Tétel: *Bármely, egyértelműen dekódolható C kódra*

$$\mathbb{E}|f(x)| = L(C) \geq H(X).$$

A 4.10 tétel alapján az entrópia szab alsó határt a kód tömöríthetőségének. Amire igazán kíváncsiak vagyunk, hogy az eddig meglehetősen hatékonynak bizonyuló prefix kódjainkkal milyen közel tudunk kerülni ehhez az elméleti határhoz.

4.11 Tétel: *Bármely, C prefix kódra*

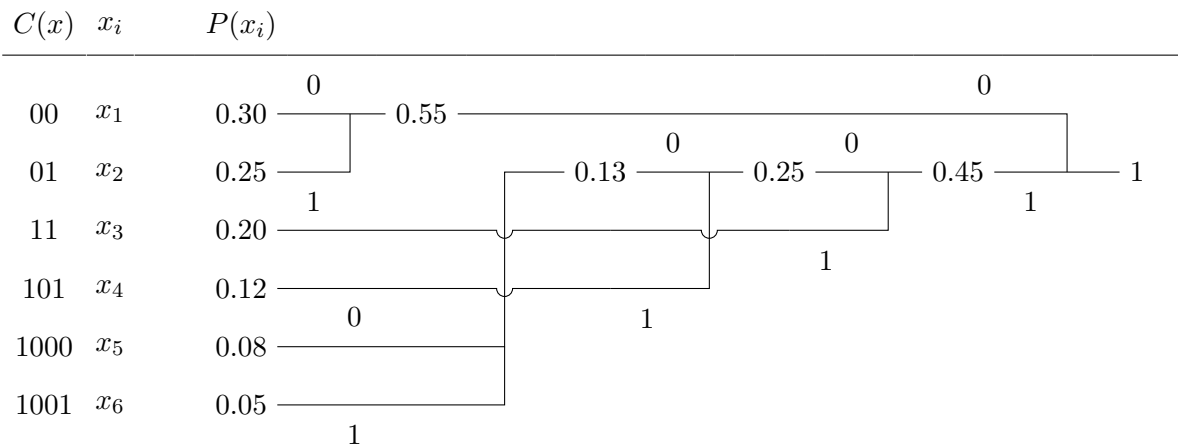
$$\mathbb{E}|f(x)| = L(C) < H(X) + 1.$$

A végső tanulság tehát az, hogy a prefix kódokkal igen jól megközelíthető az elérhető legtömörebb kódolás egy adott forrásra.

4.2.4. Bináris Huffman-kód

Egyszerű, de hatékony algoritmus optimális, prefix kódok konstruálására, ismert eloszlású forrásokhoz a **Huffman-kódolás**. A kód az egyes szimbólumokhoz (adott esetben fix hosszú blokkokhoz) rendel hozzá változó hosszúságú kódszavakat, hogy minél jobban közelítse a tömöríthetőség entrópia által szabott elméleti határát. Az algoritmus működése a következő: Először is redukáljuk a problémát a két legkisebb valószínűségű szimbólum összevonásával. Ezt egészen addig végezzük, míg végül csak két valószínűségünk marad. Ezután az összevonások megfordításával kiegészítjük a megfelelő kódszavunkat 0-val vagy 1-gyel, amíg van fel nem oldott összevonás.

A Huffman-kód konstrukcióját bináris fa építésével a legszemléletesebb demonstrálni. Ekkor, célszerű növekvő sorrendbe rendezni a szimbólumokat valószínűség szerint. Ezután egy közös csomópontba összekötjük egy-egy éllel a két legkisebb valószínűségű szimbólumot, illetve csomópontot, ameddig el nem fogynak. Célszerű az egyes csomópontokra felírni az összesített valószínűséget, így ha a végén 1 kerül az utolsó csomópontra, biztosan nem hibáztunk. Végül a fa gyökerétől lefelé haladva minden ágra felírunk egy 0 vagy 1 karaktert meghatározott sorrendben (pl.: mindig a magasabb valószínűségű élre írjuk az 1-et). Amennyiben minden élt megjelöltünk, a gyökértől az adott szimbólumig összeolvasva a karaktereket megkapjuk az egyes kódszavakat, amely alapján előállítható a kódtábla. Az algoritmus menete alapján látható, hogy a kódolás nem egyértelmű hisz, a 0 és 1 karakterek felcserélésével az előzővel ekvivalens kódot kapunk.



4.3. ábra. Huffman-kódolás bináris fával reprezentálva

4.3. Hibajavító kódolás

Amíg forráskódolás során az volt a cél, hogy egy adott forráshoz a lehető legtömörebb kódot rendeljük, azaz elimináljuk az előforduló redundanciákat, addig a **Hibajavító kódolás** során redundanciát viszünk a rendszerbe. Ennek célja, hogy az információt minél megbízhatóbb módon tudjuk továbbítani a zajos csatornákon. A redundancia hozzáadásával a vételi oldalon megállapítható, hogy történt-e hiba az átvitel során, illetve bizonyos esetekben a hiba akár javítható is.

4.3.1. Kódok hibajavító képességei

A célunk az, hogy az u **üzenetet**⁵ kódoljuk, az előálló c **kódszót** átvigyük a csatornán, a fogadó oldal pedig a v vett kódszó alapján előállítja a dekódolt u' üzenetet, amely reményeink szerint megegyezik az eredeti üzenettel. Matematikai értelemben, az üzenetek k , míg a kódszavaink v hosszú vektorok, melyek rendre az F halmazból, azaz **forrásábécéből**, illetve a Q halmazból, a **kódábécéből** veszik értékeiket. Itt érdemes megállni egy percre, hogy a félreértéseket elkerüljük: ebből a szempontból az üzenetek, a forráskódolás során előállított kódszavaink, tehát a forráskódolás és a hibajavító kódolás során előállított kódszavak nem azonosak; ezeket érdemes nem összekeverni.

A továbbiakban főként **blokk-kódokkal** fogunk foglalkozni, mely a korábban ismertetett betűnkénti kódolás természetes általánosítása. Ebben az esetben a forrásábécé betűiből alkotott rendezett n -eseket tekintjük forrásszimbólumoknak és ezekhez rendeljük hozzá a kódszavakat.

A hibajavító kódolás egyik alapvető ötlete az, hogy a kódszavaknak kellően „távolinak” kell lenniük, hogy megállapítható legyen, ha érvénytelen kódszót vettünk. A távolságok mérése során az egyik alapvető metrikánk az úgynevezett **Hamming-távolság** lesz.

⁵Idegennyelvű, főként angol szakirodalomban m betűvel jelölik az üzenetet.

4.12 Definíció: Két adott $c = (c_1, c_2, \dots, c_n)$ és $v = (v_1, v_2, \dots, v_n)$ kódszó **Hamming-távolságán** azt a számot értjük, ahány i pozícióban $c_i \neq v_i$. A \mathbf{c} és \mathbf{v} kódszavak Hamming-távolságát az alábbi módon jelöljük: $d(c, v)$.

4.13 Tétel: Bármely ábécére és n -re, a Hamming-távolság az összes n hosszú kódszó halmazán értelmezett távolság-függvény (metrika), mivel:

1. $d(a, a) = 0$ és $d(a, b) > 0$, ha $a \neq b$
2. $d(a, b) = d(b, a)$
3. $d(a, b) + d(b, c) \geq d(a, c)$ (háromszög-egyenlőtlenség)

tetszőleges \mathbf{a} , \mathbf{b} , \mathbf{c} kódszavakra és n hosszra.

BIZONYÍTÁS: Az (1) és (2) pont a Hamming-távolság definíciójából következik. A (3) pont bizonyításához legyen $d(a, b) = v$ és $d(b, c) = w$, így az i_1, \dots, i_v indexek az \mathbf{a} és \mathbf{b} vektorok közötti eltéréseket jelölik, míg a j_1, \dots, j_w indexek a \mathbf{b} és \mathbf{c} közöttieket. Továbbá, $a_i = b_i$, ha $i \neq i_1, \dots, i_v$, illetve $b_i = c_i$, ha $i \neq j_1, \dots, j_w$. Ebből következik, hogy $a_i = c_i$, ha $i \neq i_1, \dots, i_v, j_1, \dots, j_w$, tehát \mathbf{a} és \mathbf{c} $v + w$ számú indexekben különböznek. Ez alapján $d(a, b) + d(b, c) \geq d(a, c)$. \square

A hibajavítás során a célunk az, hogy a vett, érvénytelen kódszó alapján kitaláljuk, hogy mi volt az eredeti kódszó. Bizonyos feltételek teljesülése esetén feltehetjük, hogy a vett v kódszótól legkisebb Hamming-távolságra lévő érvényes c kódszó az, amit keresünk. A fenti feltétel a C kód egyik legfontosabb paraméteréhez, az úgynevezett **kódtávolsághoz** kapcsolódik.

4.14 Definíció: Egy C blokk-kód $d(C)$ (vagy d_{\min}) **kódtávolságán**, bármely különböző kódszavai közötti legkisebb Hamming-távolságot értjük:

$$d_{\min}(C) = \min_{\substack{a, b \in C \\ a \neq b}} d(a, b).$$

A továbbiakban látni fogjuk, hogy a kódtávolság valóban az egyik, ha nem a legfontosabb paramétere a kódnak, hisz annak számtalan tulajdonságával van összefüggésben, többek közt a hibajelzési képességgel. A **hibajelzés** feladata a vételi oldalon a hiba tényének detektálása, azonban annak javítása nem, tehát ez egy jelzés arról, hogy hiba történt, de ennél konkrétabb dolgot nem mond.

4.15 Tétel: Egy d_{\min} kódtávolságú kód minden, legfeljebb $d_{\min} - 1$ hibát képes jelezni.

BIZONYÍTÁS: Legyen a jelezhető hibák száma t , ekkor $d(C) > t$. Legyen c egy érvényes kódszó, míg v egy kódszó, amely kielégíti az $1 \leq d(c, v) \leq t$ egyenlőtlenséget. Ebben az esetben a v kódszó, biztosan érvénytelen. \square

Az igazán izgalmas része a hibavédő kódolásnak a kód tényleges **hibajavító** képessége. Ebben az esetben, nem csak a hiba tényét állapítjuk meg, hanem az eredeti kódszót kívánjuk visszaállítani.

4.16 Tétel: Egy d_{min} kódtávolságú kód minden, legfeljebb $\lfloor \frac{d_{min}-1}{2} \rfloor$ hibát képes javítani.

BIZONYÍTÁS: Ahhoz, hogy a vett v kódszóból az eredeti c kódszó egyértelműen visszaállítható legyen, a

$$d(v, c') > d(v, c)$$

egyenlőtlenségnek teljesülnie kell. A háromszög-egyenlőtlenséget felírva

$$d(v, c') \geq d(c, c') - d(v, c),$$

látható, hogy az eredeti egyenlőtlenség csak úgy teljesülhet, ha

$$d(c, c') - d(v, c) > d(v, c).$$

Ez azt jeletni, hogy minden $c \neq c'$ -re,

$$d(c, c') > 2 \cdot d(v, c)$$

illetve,

$$\frac{d_{min}}{2} > d(v, c). \quad \square$$

Az eddig elmondottak alapján, tehát a kód hibajelzési és hibajavítási képességei a kódtávolságból adódnak, amely igazolja azt a korábbi állításunkat, miszerint a hatékony hibajavításhoz minél távolábbi kódszavakra van szükség. A következő fejezetben nem csak azt tekintjük át, hogy miként lehet ilyen kódokat konstruálni, de meg is vizsgáljuk, hogy milyen lehetőségünk van a hibák hatékony javítására anélkül, hogy a hibás vett szó összes érvényes kódszóval vett távolságát meg kellene vizsgálnunk.

4.3.2. Bináris hibajavító kódok

A továbbiakban bináris kódokkal foglalkozunk, ahol a kódszavakat alkotó szimbólumok 0 vagy 1 értékűek lehetnek. Ugyan, a bináris kódok elmélete valójában megegyezik bármely véges tér felett értelmezett kód elméletével, de mind közül ez az eset a legkevésbé absztrakt, így megértése némileg könnyebb, mint a nem bináris kódoké.

A bináris kódok aritmetikája miatt, az összeadás illetve szorzás alatt mindig modulo 2 összeadást, illetve modulo 2 szorzást értünk. A számolásokat némileg egyszerűsíti, hogy a modulo 2 összeadás nem más, mint a logikai **XOR** – amit \oplus szimbólummal fogunk jelölni –, a modulo 2 szorzás pedig az **AND** művelet.

4.3.2.1. Bináris, lineáris kódok előállítása

Mielőtt rátérnénk a hatékony hibajavításra, meg kell ismerkednünk a hibajavító kódok előállításának módszerével. Annak érdekében, hogy a kód konstruálása és a hibajavítás a legegyszerűbb legyen, bevezetjük a **lineáris kódok** fogalmát.

4.17 Definíció: Egy C kódot **lineárisnak** nevezünk, ha C halmaza lineáris tér, azaz, ha minden $c, c' \in C$ -re, $c + c' \in C$.

A definíció értelmében bármely lineáris kódnak érvényes kódszava a csupa 0 kódszó. Továbbá, minden valós vektortérben megszokott fogalom azonos, úgy mint **bázis** és **lineáris függetlenség** alkalmazható, így kódszavak is valójában vektorok, a két fogalom tetszőlegesen felcserélhető.

4.18 Definíció: Legyen K lineáris kód, melynek bázisai $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_k \in C$. Ekkor, minden $\mathbf{c} \in C$ kódszó előállítható, az alábbi módon:

$$\sum_{i=1}^k u_i g_i$$

Ezt megadhatjuk mátrixos felírásban is:

$$c = uG$$

Ebben a felírásban G , a C kód $k \times n$ méretű **generátormátrixa**, mely az alábbi alakban áll elő:

$$G = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{pmatrix}$$

Eszerint a generátormátrix előállítja az u üzenetekhez tartozó c kódszavakat, vagyis minden u -hoz kölcsönösen egyértelműen hozzárendel egy c kódszót. Bináris esetben a k -dimenziós u vektorokkal 2^k üzenetet tudunk kifejezni, miközben a kódszavaink n -dimenziósak, ahol $n > k$. Ebben az esetben, nem tudjuk az összes lehetséges kódszavunkat üzenethez rendelni és pontosan ezt használjuk ki a hibajavítás során.

Természetesen a generátormátrix fenti megadása csak azt mutatja, hogy annak sorait a kódtér bázisai adják, azonban a bázisok sorrendje tetszőleges. Ez azt is jelenti, hogy a G mátrix többféleképp előállhat vagyis több különböző G mátrix is ugyanazokat a kódszavakat generálja. Ezt nem lenne feltétlen szükséges kiküszöbölni, de ezen kívül más előnyökkel is jár, ha bevezetjük a **szisztematikus kódok** fogalmát.

4.19 Definíció: Egy lineáris kódot **szisztematikusnak** nevezünk, ha a generátormátrixa előáll az alábbi alakban:

$$G = [I|B]$$

ahol, I a $k \times k$ méretű egységmátrix, B pedig egy $k \times n - k$ méretű mátrix.

A fenti definíció értelmében, bármely szisztematikus kódra igaz, hogy a kódszavak első k betűje az eredeti üzenetet tartalmazza, ezt üzenetszegmensnek nevezzük, míg a további $n - k$ betűt paritászegmensnek, vagy bináris esetben paritásbiteknek. Mindjárt látni fogjuk, hogy a szisztematikuságból adódó B mátrix milyen szerepet is játszik a hibajavításban. Előtte viszont érdemes megjegyezni, hogy minden lineáris kódhoz létezik egy ekvivalens szisztematikus kód is, így célszerű ezt a tulajdonságot megkövetelni a kódtól, hisz mind a dekódolás, mind a hibajavítás lényegesen egyszerűbb, ahogy mindjárt látni is fogjuk.

4.3.2.2. Bináris kódok hibajavítási mechanizmusa

4.20 Definíció: Legyen C egy lineáris kód és legyen H egy $n - k \times n$ méretű mátrix. Ha,

$$Hc^T = 0$$

akkor és csak akkor, ha $c \in C$, akkor H a C kód **paritás ellenőrző mátrixa**, röviden **paritásmátrixa**.

A paritásmátrix segítségével tehát, bármely vett kódszóról meg tudjuk mondani, hogy érvényes kódszó-e vagy sem. A kérdés az, hogy a paritásmátrixot miként tudjuk egyszerűen előállítani.

4.21 Tétel: Ha H és G ugyanazon lineáris kód paritás és generátor mátrixra, akkor:

$$HG^T = 0.$$

BIZONYÍTÁS: Legyen Q^k a k hosszú szimbólumsorozatok halmaza. Ekkor minden $u \in Q^k$ -hoz létezik $c \in C$, amire $c = uG$. Ugyanakkor $c \in C$ miatt $Hc^T = 0$, tehát

$$Hc^T = H(uG)^T = HG^T u^T = 0.$$

Az utolsó egyenlőség pedig csak úgy teljesülhet, ha minden $u \in Q^k$ -ra, ha $HG^T = 0$. \square

Ezen tétel következményeképp egyszerűen előállítható bármely szisztematikus kód generátor mátrixból, az alábbi módon:

$$H^T = \begin{bmatrix} B \\ I \end{bmatrix}$$

A tényleges hibajavítás előtt még érdemes bevezetni a vektor **súlyának** fogalmát, amiről a későbbiekben látni fogjuk, hogy jelentősen leegyszerűsíti egy adott kód kódtávolságának meghatározását.

4.22 Definíció: Egy c kódszó **Hamming-súlyán** (vagy egyszerűen csak **súlyán**) a kódszóban található nem 0 pozíciók számát értjük. Jelölése: $w(c)$.

4.23 Definíció: Egy C kód **minimális súlyán**, a legkisebb súlyú kódszó Hamming-súlyát értjük:

$$w_{\min} = \min_{\substack{c \in C \\ c \neq 0}} w(c).$$

4.24 Tétel: Bármely lineáris C kódra $d_{\min} = w_{\min}$.

BIZONYÍTÁS: Legyen $d(C)$ a C kód kódtávolsága és $w(C)$ a minimális súly. Legyen a egy C -beli kódszó, melynek Hamming-súlya $w(C)$, így a csupa 0 kódszó és az a kód Hamming-távolsága megegyezik a Hamming-súlyával, vagyis:

$$w(C) = d(a, 0) \geq d(C).$$

Most válasszunk egy szintén C -beli a és b (ahol, $a \neq b$) kódszavakat $d(C)$ Hamming-távolsággal. Ekkor $a - b$ szintén kódszó (hiszen, C lineáris kód), amely pontosan annyi 0 pozíciót tartalmaz, ahány pozícióban a és b megegyezik. Ebből következik, hogy $d(a, b) = w(a - b)$, tehát:

$$d(C) = d(a, 0) \geq w(C). \quad \square$$

A két egyenlőtlenség csak akkor teljesülhet egyszerre, ha: $d(c) = w(c)$.

A tétel tanulsága szerint a kódtávolság meghatározásához szükséges lépések számát jelentősen leredukáltuk, hiszen definíció szerint $\frac{|C|(|C|-1)}{2}$ műveletre lett volna szükség, azonban a súly fogalmának felhasználásával, mindössze $|C| - 1$ lépésre van szükség ($|C|$ jelöli a C halmaz elemszámát).

Most vizsgáljuk meg a vett szó és az eredeti kódszó közötti eltérések, legyen ez a **hibavektor** $= e = v - c$. Feltéve, hogy olyan kódot használunk, ami illeszkedik a csatornai átviteli paramétereikhez (vagyis lehetőleg annyit hibázik csak, amennyi hibát a kódunk javítani tud), a vett kódszóhoz legközelebbi érvényes kódszó az eredetileg küldött kódszó lesz. Azonban a célunk az, hogy hatékonyan tudjunk hibát javítani anélkül, hogy a vett kódszó összes érvényes kódszóval vett távolságát ki kellene számítani. Ehhez érdemes belátni, hogy:

$$\mathbf{H}\mathbf{v}^\top = H(c + e)^\top = \underbrace{Hc^\top}_0 + He^\top = He^\top$$

Eszerint a Hv^\top értéke csak a hibavektortól függ, az adott kódszótól nem. A tényleges dekódoláshoz még egy fogalmat kell bevezetnünk a **szindrómát**, ami az előbbiekből alapján a hibavektor

egy lineáris leképzése.

4.25 Definíció: Az $s = eH^T$ mennyiséget **szindrómának** nevezzük.

A dekódolás menete a következő: a vett kódszó alapján kiszámítjuk a szindrómát és a szindróma alapján becsüljük meg a hibavektort, amelyet kivonva a vett kódszóból, megkapjuk az eredeti kódszavunkat. A hibavektor megbecsülésének elméleti módjától (egyelőre) eltekintünk, a ZH megoldásához elég a H^T mátrix ismerete. Az e vektor azon pozíciójában lesz 1-es (bináris esetben), tehát ott keletkezett hiba, ahányadik sorával azonos alakú a kiszámolt s szindróma a H^T mátrixnak. A kíváncsiskodók kedvéért annyit megemlítünk, hogy az úgynevezett **standard elrendezési táblázat** konstruálása – amely alapján a hibavektorokat becsüljük – pont úgy van kitalálva, hogy a táblázat azon része, amely biztosan javítható hibák dekódolást biztosítja, egybeesik a H^T mátrixszal. Ez egyébként azt is jelenti, hogy egy adott kód a $\lfloor \frac{d_{min}-1}{2} \rfloor$ szerinti korlátnál is több hibát képes jelezni, azonban azok közül nem mindegyiket.

5. fejezet

Analóg modulációs eljárások

A kommunikációs rendszerek szerepét és jelentőségét már eleget taglaltuk a bevezetőben: az 1. fejezet újraolvasásával ismét feleleveníthetjük a modern világra gyakorolt elképesztő hatását. Ezen rendszerek minden egyes eleme kulcsfontosságú az információ közlése szempontjából, eddig mégis keveset szoltunk azon részeiről, melyek ténylegesen továbbítják az információt. Ebben a fejezetben a vizsgálatunk tárgya a kommunikációs rendszerek azon részeire irányul, amelyek az információt hordozó jelet az adó oldalon oly módon alakítják át, hogy azt a megfelelő közegen továbbítva, a vevő oldalon visszalakítható legyen eredeti formájába: ezek a **modulátorok**, illetve **demodulátorok**.

A **moduláció** célja szemléletesen, hogy az információt hordozó jelet az átviteli közeg tulajdonságaihoz igazítsa. Ennek során az eredeti, úgynevezett **alapsávi jelet**, melynek spektruma tipikusan a kHz-es tartomány aljára esik (például az audio jeleké 20 Hz és 20 kHz közé) egy nagyfrekvenciás **vivőjel** segítségével továbbítjuk. Ennek módja, hogy a vivőjel valamely paraméterét az alapsávi jel függvényében módosítjuk, így a vételi oldalon az eredeti információt hordozó jel visszaállítható a vett jel adott paraméterének változása nyomán: ezt a folyamatot nevezzük **demodulációnak**.

Az elmondottak alapján belátható, hogy az előálló jel, melyet **modulált jelnek** nevezünk, az eredeti alapsávi jel, másképp **moduláló jel** információtartalmával rendelkezik, azonban spektrális tulajdonságait – helyesebben átviteli tulajdonságait – a **vivő jeltől** örökli meg. Jellemző példa a modulációra az audio jel átvitele sávszűrő karakterisztikájú csatornán. A sávszűrő karakterisztika miatt az alacsonyfrekvenciás komponensek (esetünkben az összes) elvesznének, így az eredeti jelünket egy nagyfrekvenciás jelre modulálva, már olyan jelet kapunk amely képes, minimális torzítás mellett áthaladni a csatornán. Jelen fejezetben időtartományban folytonos (analóg) jelek modulációját fogjuk tárgyalni.

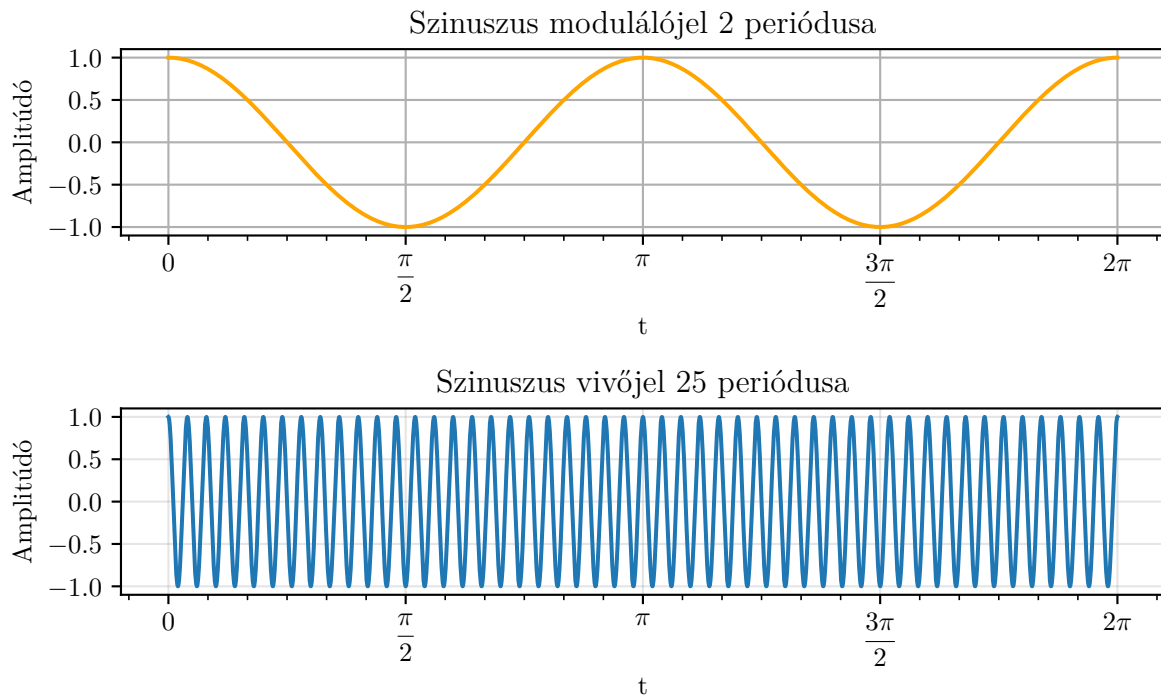
5.1. Amplitúdómoduláció

Az **amplitúdómoduláció** során a modulált jel amplitúdója hordozza az információt, tehát a vivő jel amplitúdóját modulálja az alapsávi jel. A 2.2.1 fejezetben bemutatott Fourier-transzformációs tétel alapján a vivőjel felírható komplex exponenciális jelként, azonban az Euler-tétel segítségével ez felírható egy valós koszinuszos és egy képzetes szinuszos jel összegeként is. A továbbiakban még egyszer visszatérünk erre az alakra, azonban a képzetes rész elhagyásával nem veszünk semmit (a hatékonyság szempontjából), ezért a továbbiakban a vivőjel, $(c(t))$ egy

valós koszinuszos jel lesz, ahol A_c a vivő amplitúdója, f_c pedig a vivőfrekvencia:

$$c(t) = A_c \cos(2\pi f_c t) \quad (5.1)$$

A továbbiakban az általánosság elvesztése nélkül feltehetjük, hogy a vivő fázisa $\varphi_c = 0$, mivel ennek információ tartalma nincs, így ezt a továbbiakban rendszerint nem tüntetjük fel.



5.1. ábra. Színusz moduláló és vivőjel időfüggvénye

5.1.1. Kétoldalsávós amplitúdómoduláció

Az amplitúdómoduláció általános formája a kétoldalsávós moduláció, avagy **AM-DSB** (Double Sideband). Azt, hogy ez konkrétan mit is jelent, majd a későbbiekben látni fogjuk. A modulációról szóló ismeretinket azonban egy egyszerűbb, speciálisabb eset tanulmányozásával kezdjük: ez az **AM-DSB/SC** (Supressed Carrier) vagyis az elnyomott vivőjű kétoldalsávós amplitúdó moduláció.

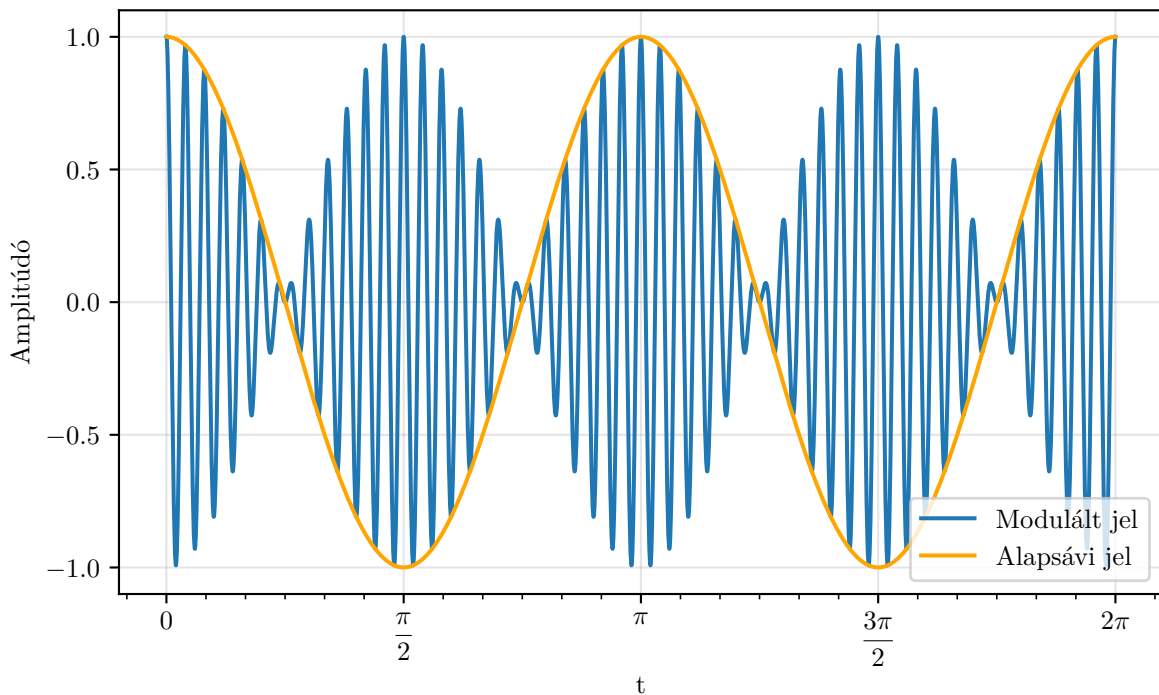
5.1.1.1. AM-DSB/SC moduláció

Először is állítsunk elő egy egyszerű amplitúdómodulált jelet. Legyen a vivő az $m(t)$ általános jel, az előálló modulált jel pedig az $S(t)$ függvény az alábbiak szerint:

$$S(t) = m(t)c(t) = A_c m(t) \cos(2\pi f_c t) \quad (5.2)$$

A fenti formula alapján (is) látható, hogy a modulált jel amplitudója $A_c m(t)$ szerint változik, míg frekvenciája látszólag megegyezik a vivő frekvenciájával. A további vizsgálatiánkhoz, tekintsünk egy speciális esetet is, mikor a moduláló jel is szinuszos:

$$m(t) = A_m \cos(2\pi f_m t) \quad (5.3)$$



5.2. ábra. AM-DSB/SC modulált jel időfüggvénye

Ezt a speciális esetet angolul **tone modulation**-nek nevezzük (magyarul ritkán: egytónusú moduláció), mivel a moduláló tisztán szinuszos. Így az előálló modulált szinuszos jel amplitudója is szinuszosan változik:

$$S(t) = m(t)c(t) = A_c A_m \cos(2\pi f_m t) \cos(2\pi f_c t) \quad (5.4)$$

A következő átalakításhoz az alábbi trigonometrikus azonosságot fogjuk felhasználni:

$$\cos(x) \cos(y) = \frac{\cos(x - y) + \cos(x + y)}{2} \quad (5.5)$$

Így az időfüggvény előáll az alábbi formában:

$$S(t) = m(t)c(t) = \underbrace{A_m \frac{A_c}{2} \cos((f_c - f_m) 2\pi t)}_{\text{LSB}} + \underbrace{A_m \frac{A_c}{2} \cos((f_c + f_m) 2\pi t)}_{\text{USB}} \quad (5.6)$$

Ahhoz, hogy nyilatkozzunk a spektrumról még Fourier-transzformálni sem szükséges a jelet,

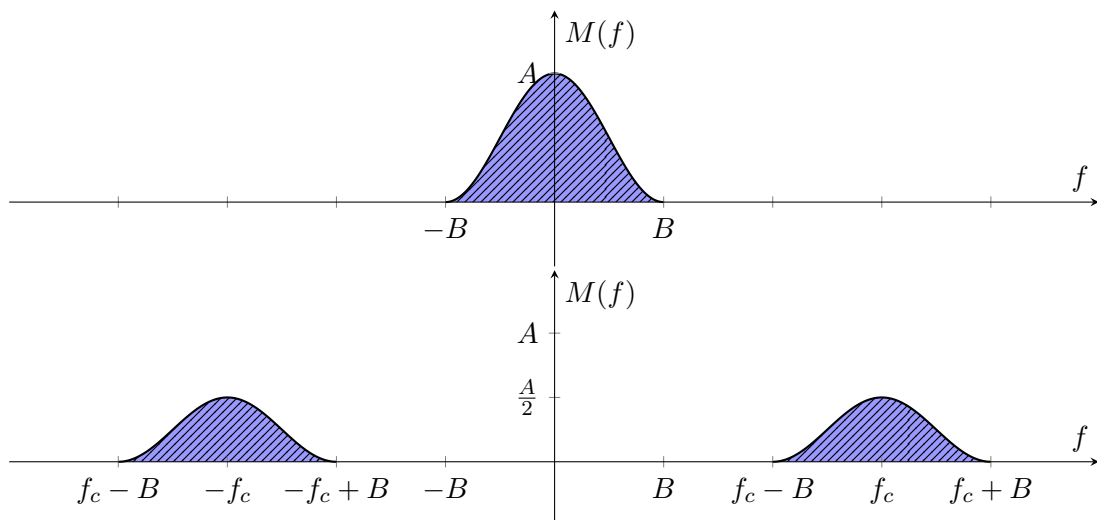
hisz kizárólag szinuszos komponensekből áll. A két komponensnek neve is van: **alsó oldalsáv** (Lower Sideband), illetve **felső oldalsáv** (Upper Sideband), ezért is hívják ezt a típusú amplitúdómodulációt kétoldalsávossáknak. Jól látható, hogy mindkét oldalsáv a moduláló jel frekvenciájával megegyező távolságban található a vivőfrekvenciától, tehát az $f_c \pm f_m$ frekvenciákon. Az egyes oldalsávok amplitúdója, az alapsávi jel amplitúdójának felével egyezik meg. Szembetűnő, hogy az f_c vivőfrekvencián nem található komponens, innen ered a másik név, azaz **elnyomott vivős** moduláció.

A teljesség kedvéért érdemes megvizsgálni egy olyan esetet is, ahol a moduláló jel spektruma általánosabb. Legyen a moduláló jel és spektruma:

$$m(t) \iff M(f)$$

Eszerint a modulált jel és annak spektruma az alábbi alakban áll elő:

$$m(t) \cos(2\pi f_c t) \iff \frac{1}{2} [M(f_c + f) + M(f_c - f)]$$



5.3. ábra. AM-DSB/SC modulált jel spektruma

A felvázolt spektrum alapján a szinuszos speciális eset tulajdonságai ebben az esetben is fennállnak. A spektrummal kapcsolatban két további kijelentést is tehetünk: egyrészt, ha a sáv-szélességhez képest aránytalanul kicsire választjuk a vivőfrekvenciát, akkor a negatív és pozitív spektrum szeletek átlapolódnak, aminek következtében nem állítható vissza az alapsávi jel a demoduláció során. Tehát a maradéktalan visszaállíthatóság feltétele: $f_c \geq B$. Másrészt, az elnyomott vivős jellege a modulációnak nem jelenti szükségszerűen, hogy az f_c frekvencián a spektrum értéke zérus, csupán azt, hogy a vivőtől származó Dirac-impulzus nincs jelen (zérus középpértékű vivőt feltételezve). Természetesen, a gyakorlatban más megkötések is vannak a vivőfrekvencia értékére. Ilyen megkötés például, hogy teljesüljön az $\frac{f_c}{B} \gg 1$ feltétel, mivel az egyes antennák tipikusan csak keskenysávban képesek torzításmentesen adni. Ennek jelentőségét az AM műsorszórásnál még látni fogjuk.

Eddig azzal foglalkoztunk, hogy miként állíthatjuk elő a modulált jelünket az adó oldalon, azonban legalább ugyanilyen fontos, hogy miként állítható vissza a vevő oldalon az alapsávi jel. Ezt a folyamatot **demodulációnak** nevezzük, és ahogy látni fogjuk nem különbözik lényegesen a modulációs eljárásunktól. Ebben az esetben is azt a jelenséget használjuk ki, hogy a moduláció eltolja a jel spektrumát, csak most az eredeti tartományba szeretnénk visszatolni a spektrumot. Viszonylag könnyen belátható, hogy ha a vett modulált jelet az eredeti vivőjellel megszorozva, a modulált jel spektruma a zérus frekvencia körül is megjelenik. Ez formálisan az alábbi módon írható le (ismét az 5.5 formulát felhasználva).

$$e(t) = m(t) \cos^2(2\pi f_c t) = \frac{1}{2} [m(t) + m(t) \cos(4\pi f_c t)]$$

Ez a megoldás, a "bal kézzel vakarom a jobb fülem" tipikus esete, ugyanis a modulált jel kétoldalas spektrumában a negatív tartomány pozitív irányba tolódik, míg a pozitív tartomány negatív irányba, így a zérus frekvencia körül ezek átlapolódnak és visszaadják az eredeti spektrumot (pontosabban annak felét). Azonban, emiatt két további spektrális komponens is megjelenik a vivőfrekvencia kétszerese körül.

$$E(f) = \frac{1}{2} M(f) + \frac{1}{4} [M(f - 2f_c) + M(f + 2f_c)]$$

Szerencsére ezek a komponensek viszonylag nagy távolságban találhatóak az alapsávi spektrumtól, így egy megfelelő aluláteresztő szűrő használatával könnyen eltávolíthatjuk őket. Tisztán matematika szempontból nézve ugyan nem kaptuk vissza az eredeti jelünket, de ezt is könnyen megoldhatjuk, ha nem az eredeti vivőjellel, hanem annak kétszeresével szorozzuk a vett jelet.

Örömünk azonban nem lehet felhőtlen, hisz ez az eljárás nem olyan egyszerűen kivitelezhető, mint ahogy hangzik. Az alapsávi jel maradéktalan visszaállításához ugyanis szükségünk van az eredeti vivőjelre, aminek frekvenciáját bár (jó esetben) tudjuk, de fázisát biztosan nem. Ezt a demodulációs eljárást egyébként az előbb leírtak miatt **koherens demodulációnak** nevezik.

5.1.1.2. AM-DSB moduláció

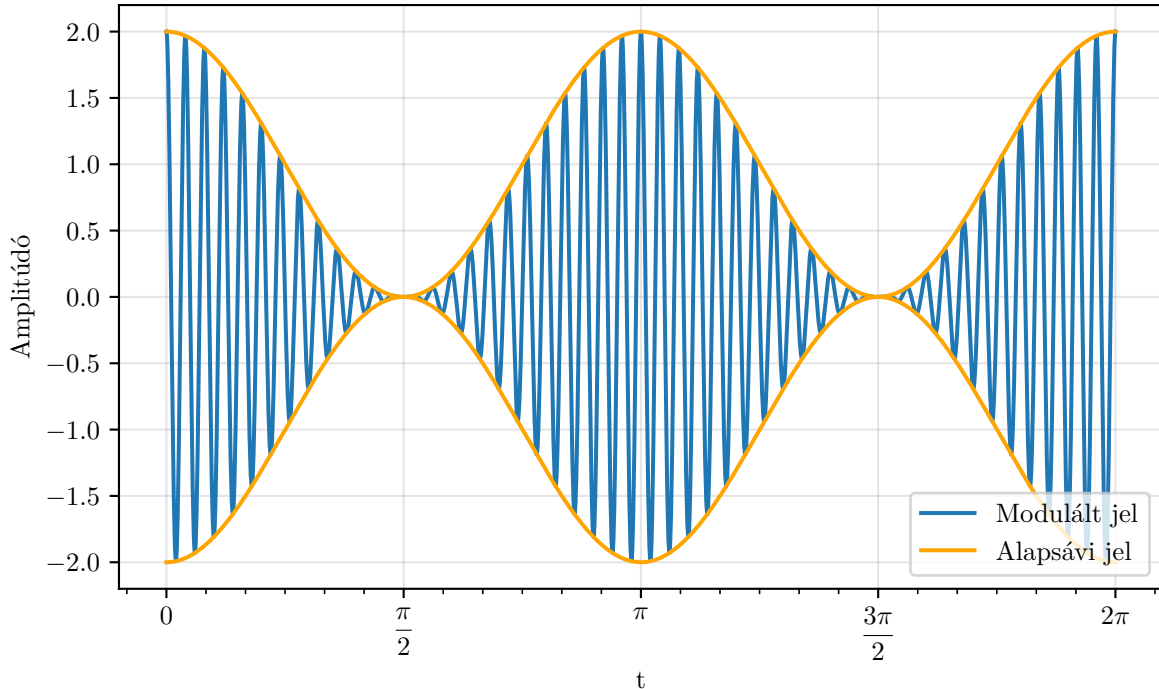
Annak érdekében, hogy az AM-DSB/SC jel demodulálásával járó nehézségeket elkerülhessük, az AM moduláció egy másik típusával kell megismerkednünk, az **AM-DSB** modulációval. A konstrukció az elnyomott vivős moduláció fényében már könnyen értelmezhető, az alábbi alakban áll elő:

$$S_{AM} = A \cos(2\pi f_c t) + m(t) \cos(2\pi f_c t) = ([A + m(t)] \cos(2\pi f_c t))$$

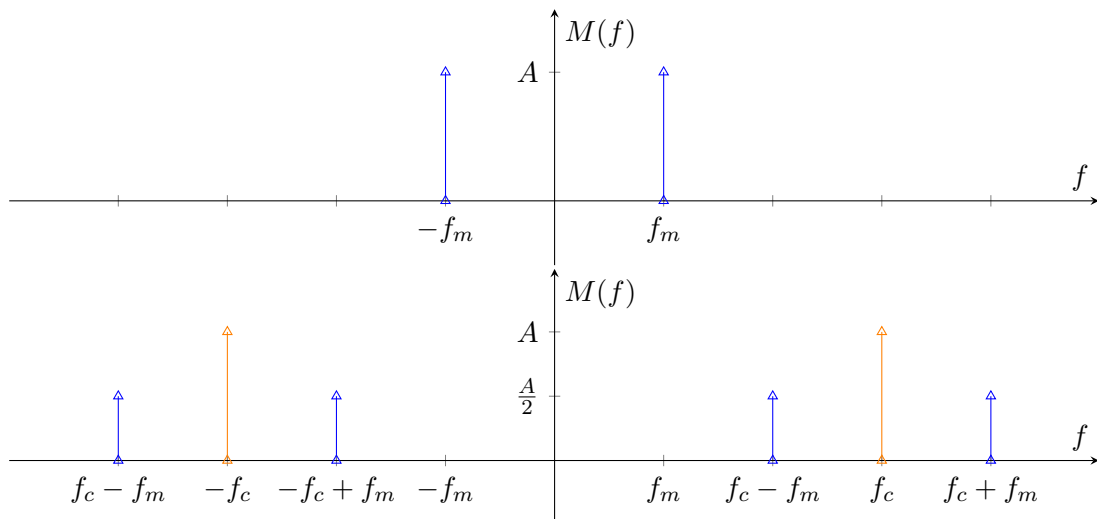
Az időfüggvény alapján is nyilatkozhatunk a spektrumról, azonban a rend kedvéért megadjuk a spektrális felírást is:

$$S_{AM} \iff \frac{1}{2} [M(f + f_c) + M(f - f_c)] + \frac{A}{2} [\delta(f + f_c) + \delta(f - f_c)]$$

Eszerint az AM-DSB moduláció majdnem teljesen megegyezik az elnyomott vivős variációval, azzal a különbséggel, hogy itt nincs elnyomva a vivő, tehát a spektrumban az f_c vivőfrekvencián megjelenik egy Dirac-delta.



5.4. ábra. AM-DSB modulált jel időfüggvénye



5.5. ábra. Egytónusú AM-DSB modulált jel spektruma

A dolog érdekessége, hogy ha ábrázoljuk a modulált jel **burkoló görbáját**, a $|A + m(t)|$ és $-|A + m(t)|$ (a burkoló tükörképe) jeleket, akkor láthatjuk, hogy a modulált jel amplitúdója a burkoló és annak tükörképe által közrefogott tartomány belül ingadozik. Az ingadozás jellegét tekintve két esetet tudunk megkülönböztetni: az első esetben $A + m(t) \geq 0$, míg a második esetben

$A + m(t) < 0$. A kettő közti alapvető különbség, hogy míg az első esetben a burkoló megegyezik az $m(t)$ moduláló jellel, addig a második esetben ez nem teljesül, ekkor az $A + m(t)$ kifejezés negatív értékeire a modulált jel fázisa megfordul. Az ilyen jeleket túlmodulálnak nevezzük és azért nem kívánatosak – szemben az első esettel –, mert ezeket nem lehet burkolódetektorral demodulálni. A burkolódetektor nem más, mint egy nagyon egyszerű és olcsó eljárás AM-DSB jelek demodulálására. Ezen felül nagy előnye, hogy nem szükséges hozzá a vevőt helyben előállítani, ugyanis a modulált jel burkológörbéje alapján állítja elő az alapsávi jelet.

A moduláció jellemzésére érdemes bevezetni egy új mérőszámot, a **modulációs mélységet**. Ez az alábbi módon adható meg:

$$\mu = \frac{\max |m(t)|}{A}$$

Így azt mondhatjuk, hogy az AM-DSB jel demodulálható, ha $0 \leq \mu \leq 1$. Amennyiben $\mu > 1$, a jel túlmodulált.

Bár, a burkolódetektoros demoduláció egyszerű és olcsó, de nagy hátránya, hogy ki kell sugározni az információt nem hordozó vivőjelet is. A kérdés, hogy a vivő továbbítása mekkora felesleges energia többletet jelent a moduláció szempontjából. A vivő jel teljesítménye $P_c = \frac{A^2}{2}$, míg az oldalsáv teljesítménye általánosan $P_s = \frac{1}{2}P_m$, arra a speciális esetre, mikor a moduláló jel szinuszos, pedig: $P_s = \frac{1}{2} \frac{(\mu A)^2}{2}$. Így a moduláció hatásfokát az alábbi módon adhatjuk meg (általánosan és a szinuszos esetre):

$$\eta = \frac{\text{hasznos teljesítmény}}{\text{összes teljesítmény}} = \frac{P_s}{P_s + P_c} = \frac{\mu^2}{2 + \mu^2}$$

$$\eta_{\text{sin}} = \frac{\mu^2}{2 + \mu^2}$$

5.2. Szögmodulációk

Eddig kizárólag olyan modulációs eljárásokkal foglalkoztunk, ahol a modulált jel amplitudója hordozta az információt. A továbbiakban azt az ötletet fogjuk boncolgatni, hogy az információt a modulált jel frekvenciája vagy fázisa hordozza, ezeket összefoglaló néven **szögmodulácónak** nevezzük.

A **frekvenciamoduláció (FM)** illetve **fázismoduláció (PM)**, mint gondolat először az AM moduláció nem túl hatékony spektrum kihasználása kapcsán jött létre. A cél az lett volna, hogy a vivőfrekvencia tetszőlegesen kicsi változtatásával át tudjuk vinni az információt, a modulált jel spektruma azonban csak a vivőfrekvencia körüli tetszőleges kicsi tartományt foglalja el. A későbbiekben látni fogjuk sajnos, hogy ez az ötlet egyáltalán nem helytálló, ennek ellenére az FM moduláció számos gyakorlati alkalmazásban a leghatékonyabban alkalmazható modulációs eljárás.

5.2.1. A pillanatnyi frekvencia

A tapasztalat az, hogy az FM moduláció megértése egy többrétegű folyamat, mivel a legtöbbször egyszerűen az AM moduláció elméletét ráhúzzák az FM modulációra és kész is. Sajnos ez a megközelítés nem a legjobb, ugyanis elsiklunk egy nagyon fontos tényező felett, így végül az FM moduláció spektrumkihasználásának értelmezése távoli ábránd marad.

A probléma viszonylag egyszerű: az információt a modulált jel frekvenciája hordozza, vagyis a vivőjel f_c frekvenciája minden időpillanatban változik. Ha kicsit jobban belegondolunk, rájövünk, hogy ennek semmi értelme. Ahhoz hogy a frekvenciát, mint jellemzőt megadhassuk, szükséges egy periodikus jel legalább egy periódusa (bizonyos esetekben elég félperiódus). Ez a jelenség nagyon hasonló a sebesség megváltozásának kérdéséhez: ha egy test gyorsul, akkor sebessége minden időpillanatban változik, azonban a sebességet a megtett út és az idő hányadosként adjuk meg, tehát egy adott pillantra vonatkoztatva nem értelmezhető. A probléma feloldásához a pillanatnyi sebesség fogalmát kell bevezetni és mi is így fogunk tenni, az úgynevezett **pillanatnyi frekvencia** fogalmának bevezetésével.

Tekintsünk egy általános szinuszos $\Phi(t)$ jelet, amely az alábbi alakban áll elő:

$$\Phi(t) = A \cos \theta(t)$$

A kifejezésben látható $\theta(t)$ az úgynevezett általánosított szög, mely az idő függvénye. Ez egy megszokott, egyszerű, konstans frekvenciájú és fázisú szinuszos jelre, t lineáris függvénye: $\theta(t) = \omega t + \varphi$. A továbbiakban körfrekvencia helyett mindenhol frekvenciát használunk, azonban ez csak egy 2π konstans szorzóban való eltérést jelent, elvi különbség nincs.

A továbbiakban legyen $\theta(t)$ egy általános függvény, jelen esetben szigorúan monoton növekvő, hogy a vizsgálatunk tárgyából kizárjuk a negatív, illetve zérus frekvenciákat – azonban ez a feltevés nem szükségszerű. Továbbá legyen az $ft + \varphi$ egyenes $\theta(t)$ érintője valamely t időpillanatban. Belátható, hogy t környezetében egy $\Delta t \rightarrow 0$ intervallumon $A \cos \theta(t) = A \cos(ft + \varphi)$. Tehát:

$$\Phi(t) = A \cos(ft + \varphi) \quad t - \frac{\Delta t}{2} < t < t + \frac{\Delta t}{2}$$

Ezen a Δt nagyságú intervallumon tehát $\Phi(t)$ frekvenciája f . Mivel $ft + \varphi$ érintője a $\theta(t)$ függvénynek, ezért $\Phi(t)$ frekvenciája a $\theta(t)$ függvény meredekségét adja meg az adott Δt intervallumon. Ha általánosítjuk ezt az állítást minden t időpillanatra, akkor azt mondhatjuk, hogy a f_i pillanatnyi frekvencia a $\theta(t)$ meredekségét adja meg minden adott t időpillanatban, azaz f_i a $\theta(t)$ függvény deriváltja.

$$f_i(t) = \frac{d\theta}{dt}$$

$$\theta(t) = \int_{-\infty}^t f_i(\alpha) d\alpha$$

5.2.2. Fázis-, és frekvenciamodulált jelek konstruálása

Fázismoduláció során az $m(t)$ modulálójel értelemszerűen a vivőjel fázisát módosítja, amely a pillanatnyi frekvenciáról szerzett ismereteink alapján az alábbi formában írható fel (immáron feltüntetve a 2π szorzót):

$$\theta(t) = 2\pi f_c t + \varphi + k_p m(t)$$

A fenti kifejezésben az f_c a vivőfrekvencia, míg a k_p egy konstans szorzó. Tegyük fel az általánosság elvesztése nélkül, hogy $\varphi = 0$, így az adódó PM modulált jel az alábbi alakban áll elő:

$$S_{PM}(t) = A \cos [2\pi f_c t + k_p m(t)]$$

Innen könnyedén belátható, hogy PM esetében a pillanatnyi frekvencia a modulálójel deriváltjának lineáris függvénye:

$$f_i(t) = \frac{d\theta}{dt} = 2\pi f_c + k_p \dot{m}(t)$$

A frekvenciamoduláció a fázismoduláció alapján könnyen leírható, hisz a kettő közel azonos alakot ölt annyi különbséggel, hogy FM esetén a pillanatnyi frekvencia a moduláló jel – és nem annak deriváltjának – lineáris függvénye:

$$f_i(t) = \frac{d\theta}{dt} = 2\pi f_c + k_f m(t)$$

A fentiek fényében a $\theta(t)$ az alábbi módon adható meg:

$$\theta(t) = \int_{-\infty}^t [2\pi f_c + k_f m(\alpha)] d\alpha = 2\pi f_c t + k_f \int_{-\infty}^t m(\alpha) d\alpha$$

Az FM modulált kimenő jelünk az alábbi alakban írható fel:

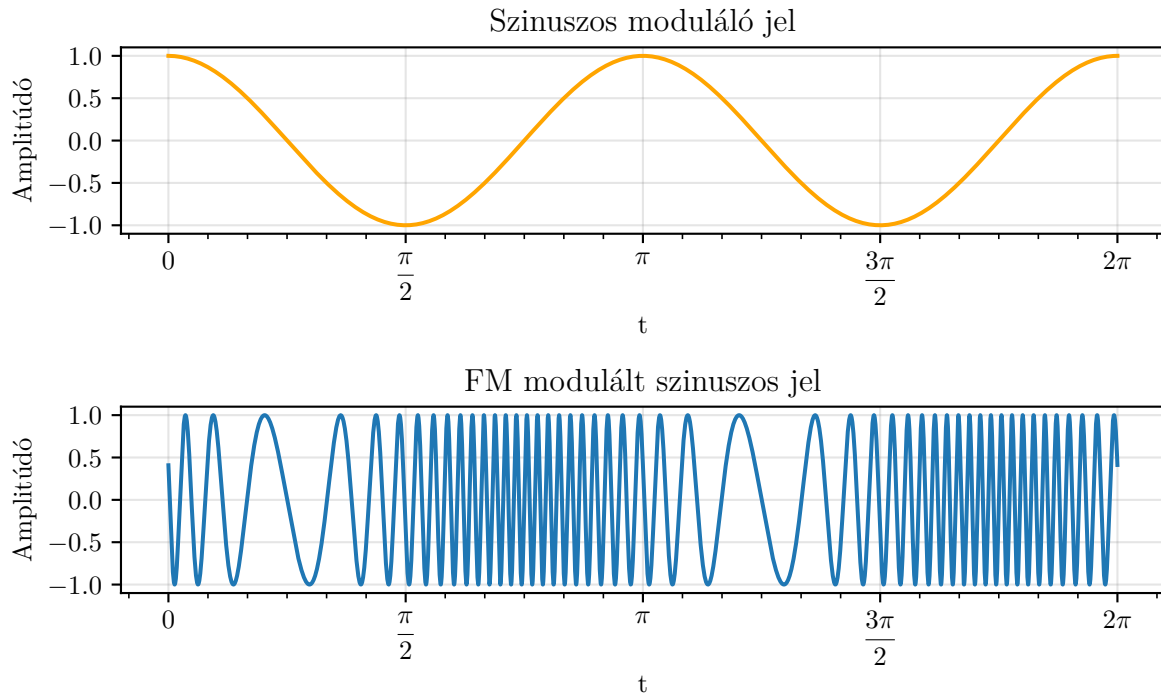
$$S_{FM}(t) = A \cos \left[2\pi f_c t + k_f \int_{-\infty}^t m(\alpha) d\alpha \right]$$

Miután felírtuk az FM modulált jel általános formuláját elmondhatjuk, hogy az FM és PM moduláció nem csak közel azonos alakú, hanem a maga módján egyforma, azaz egy szögmodulált jelről nem tudjuk megmondani, hogy FM vagy PM modulált-e. Az ok nyilvánvaló, egy $m(t)$ jellel FM modulált jel pontosan megegyezik egy $\int m(\alpha) d\alpha$ jellel PM modulált jellel. Megfordítva: egy $m(t)$ jellel PM modulált jel megegyezik egy $\dot{m}(t)$ jellel FM modulált jellel.

A rend kedvéért, felírjuk a legegyszerűbb esetét a PM, illetve FM modulációnak, mikor is a moduláló jel szintén szinuszos: $m(t) = A_m \cos(2\pi f_m t)$.

$$\text{PM: } S_{PM}(t) = A \cos [2\pi f_c t + k_p A_m \cos(2\pi f_m t)]$$

$$\text{FM: } S_{FM}(t) = A \cos \left[2\pi f_c t + k_f \frac{A_m}{2\pi f_m} \sin(2\pi f_m t) \right]$$



5.6. ábra. Szinuszos moduláló jel és azzal FM modulált jel időfüggvénye

A jegyzet jelen formájában sajnos nem jut idő részletesen elemezni a szögmodulációkkal kapcsolatos legfontosabb kérdést, mégpedig az eredő sávzélességet, ezért most csak röviden érintkezünk róla. Ellentétben az AM modulációval, a szögmoduláció nemlineáris, így a Fourier-transzformációt nem tudjuk alkalmazni ebben az esetben, a spektrum vizsgálatának céljából. Levezetés nélkül fogadjuk el, hogy a szögmodulált jelek sávzélessége végtelen. Ellenben, remek gyakorlati közelítést tudunk adni egy látszólagos sávzélességre, ugyanis a modulált jel energiája 98%-át egy adott sávban sugározza ki. Ennek kiszámításához FM esetben a moduláló jel sávzélességére, illetve az úgynevezett **frekvencialöket** (f_D) ismeretére van szükség. A frekvencialöket annak az intervallumnak a nagysága, amelyen a modulált jel frekvenciája változik, azaz:

$$f_D = \frac{k_f A_m}{2\pi}$$

A modulált jel sávzélesség a **Carson szabály** szerint becsülhető, az alábbi módon:

$$B = 2(f_D + B_m)$$

A PM jel sávzeléssége természetesen megbecsülhető a fenti módon, de érdekesebb a számí-

táshoz az úgynevezett **fázislököt** mennyiségét felhasználni:

$$\Phi = k_p A m$$

A fázislököt a modulációs mélységhez hasonló szerepet tölt be, azaz megadja, hogy a moduláló jel milyen mértékben változtatja a moduláló jel a vivő jel adott jellemzőjét.

A fázislököt felhasználásával a sáv szélesség az alábbi módon becsülhető:

$$B = 2B_m(1 + \Phi)$$

5.3. Rádiós műsorszóró rendszerek

A korábbi fejezetekben tárgyalt modulációs technikák egyik legkézenfekvőbb gyakorlati alkalmazása a rádiós műsorszórás. A legrégebbi ilyen rendszerek valójában rádiós távirók voltak, hangot nem továbbítottak. A szűkebb értelemben vett rádiós sugárzás (beszéd, zene) csak később jelent meg, azonban a mai napig töretlenül üzemel. A továbbiakban az AM és FM rádió legfontosabb aspektusait vesszük górcső alá.

5.3.1. Szuperheterodin vevők

A rádió egyik nagy előnye, hogy mindig is viszonylagosan olcsó és egyszerűen kezelhető volt a felhasználók, hallgatók számára. Elég csak arra gondolni, hogy a kezdeti időkben kizárólag (egyébként manapság is kapható) passzív vevőket használtak.¹ A rádiózás egyik legjelentősebb problémája, amelynek leküzdése jelentősen elősegítette elterjedését, a rádióadók közti váltás nehézsége. A rádiózás hajnalán egy darab rádiófrekvenciás szűrő segítségével nyomták el a szomszédos adók jelét, azonban ezek a szűrők szelektivitása meglehetősen gyenge volt, így több kaszkádosított szűrőt kellett alkalmazni. Nyilvánvalóan ezek (pontosabban a rezgőkörök) összehangolása meglehetősen időigényes feladat volt. Az idő múlásával természetesen megjelentek a jó minőségű hangolható szűrők, azonban ezek jelentősen drágábbak mint egy fix szűrő.

A fenti problémák áthidalására fejlesztették ki az úgynevezett **szuperheterodin vevőt**, melynek blokkváza az 5.7 ábrán látható. A vevő egy rádiófrekvenciás (RF) erősítőből és szűrőből, egy keverőből, egy középfrekvenciás erősítőből és szűrőből, egy burkoló detektorból, illetve egy audiófrekvenciás erősítőből áll. A továbbiakban főként a szűrők és a keverő szerepével foglalkozunk, az erősítők szerepe triviális.

A rádiófrekvenciás rész feladata a kívánt rádióadó vétele, éppen ezért az RF szűrő hangolható, azonban a szelektivitás tekintetében nem támasztunk erős követelményeket, még hozzá okkal. A szűrő kimenőjele egy keverőbe csatlakozik, amely a jelet egy úgynevezett **helyi oszcillátor** kimenő jelével szorozza össze. A helyi oszcillátor szintén hangolható, kimenő jelének frekvenciája az alábbi módon áll elő:

¹A passzív vevőknek nincs szüksége tápellátásra, az energiát a vett jel szolgáltatja. A kezdeti időkben persze nem feltétlenül az olcsóság dominált, hanem inkább a lakossági áramellátás és a fejlettebb technológia hiánya.

$$f_{LO} = f_c + f_{KF}$$

Az f_c az RF jel vivőfrekvenciája, ahova a szűrőt is hangoljuk, az f_{KF} pedig egy konstans érték, az úgynevezett **középfrekvencia**. Ez alapján felírhatjuk a keverő kimenő jelét, ahol a helyi oszcillátor jele általános szinuszos alakot ölt, míg az RF jel AM-DSB vagy FM modulált jel. AM-DSB moduláció esetén a kimenő jel:

$$A \cos(2\pi f_{LO}t) \cdot [A + m(t)] \cos(2\pi f_c t) = [A + m(t)] \cos(2\pi f_c t) \cos(2\pi f_{LO}t)$$

$$\frac{A + m(t)}{2} [\cos(2\pi(f_c - f_{LO})t) + \cos(2\pi(f_c + f_{LO})t)]$$

$$\frac{A + m(t)}{2} [\cos(2\pi(-f_{KF})t) + \cos(2\pi(2f_c + f_{KF})t)]$$

Miután behelyettesítettük f_{LO} értékét, érdekes megfigyelésben lehet részünk. A keverő tulajdonképpen AM-DSB/SC modulálta az RF jelet, melynek eredményeképp két komponens jött létre, egy középfrekvenciás komponens, illetve egy a vivőfrekvencia kétszeresétől középfrekvenciányira lévő komponens. Az eredmény legizgalmasabb megállapítása az, hogy a középfrekvenciás komponens teljesen független az eredeti vivő frekvenciájától, így a keverő minden adót a középfrekvenciára fog letranszponálni, mindössze annyi a dolgunk, hogy a helyi oszcillátort és az RF szűrőnkent ugyanarra a frekvenciára hangoljuk. A keverő után elhelyezkedő KF szűrőt így nem is kell hangolnunk, épp ezért ide választható egy olcsóbb, de megfelelő szelektivitású szűrő, amely így csak a venni kívánt jelet engedi át. A szuperheterodin vevő struktúra legnagyobb eredménye tehát, hogy képesek vagyunk indirekt módon hangolni a vevőnket, így a jó minőségű szűrőnk nem kell, hogy hangolható legyen. Miután a vett jelet letranszponáltuk középfrekvenciára már csak demodulálnunk kell, például burkoló detektorral. Ez az eredmény FM jelre következőképp néz ki:

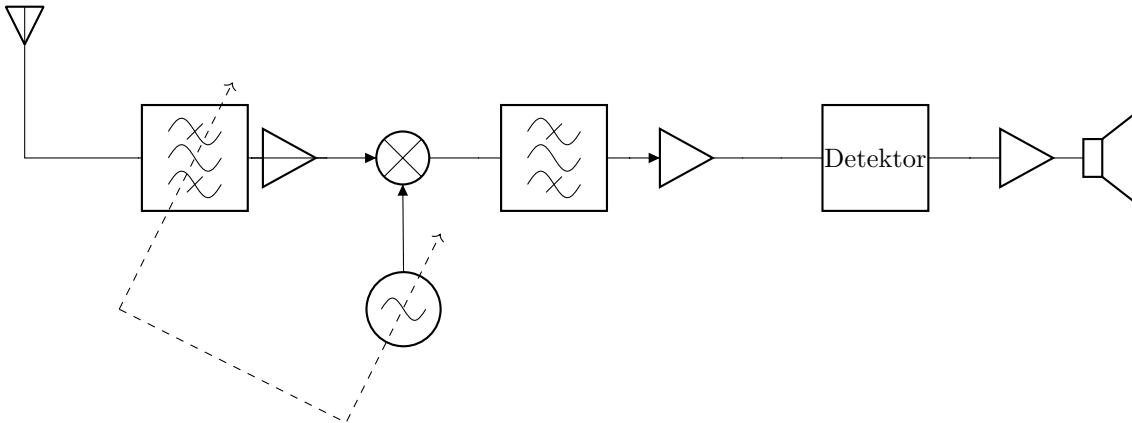
Természetesen egy ilyen frekvencia transzponálás megoldásnak is megvannak a maga hátrányai, amit látni is fogunk, ha alaposabban megvizsgáljuk a helyi oszcillátor frekvenciáját. Az $f_{LO} = f_c + f_{KF}$ frekvencia helyett használhatnánk az $f_{LO} = f_c - f_{KF}$ frekvenciát is, így is a kívánt középfrekvenciás sávra transzponálnánk az eredeti jelünket. Az első megoldást **felső keverésnek**, a második megoldást **alsó keverésnek** nevezzük, mivel a helyi oszcillátor frekvenciája magassabb, illetve alacsonyabb a vételi frekvenciánál. A gyakorlatban szinte kizárólag felső keverést alkalmazunk, mivel az oszcillátor relatív frekvenciaátfogása (tehát a legkisebb és legnagyobb frekvencia aránya) kisebb, mint alsó keveréses esetben, így az oszcillátor jelének amplitudója kevésbé változik a frekvencia függvényében, miközben a frekvenciastabilitása nem javul számottevően.

Elméletben, mind alsó és felső keverésnél jelentkezik egy igen bosszantó probléma, aminek

vizsgálatához induljunk a fenti felső keverséses struktúrából. Legyen az f_{KF} középfrekvenciás vevőnk, f_c frekvenciára hangolva, ahol venni szeretnénk az adást. A siker teljesnek tűnik, hisz transzponálás után a KF szűrő remekül végzi a dolgát, és többek között a keverőben létre jött másik oldalsávunkat, nevezetesen a $\cos(2\pi(2f_c + f_{KF})t)$ komponenst maradéktalanul kiszűri. A probléma az, hogy mivel az RF szűrő szelektivitása gyenge (gyakorlatilag csak a KF szűrő nyomja el a nem kívánt komponenseket), ezért a keverő bármilyen vett jelet letranszponál. Ez azt is jelenti, hogy az $f_c + 2f_{KF}$ frekvenciás adó jelét lekeverve az alábbi komponensek jelennek meg a KF szűrő bemenetén (amplitudó nélkül):

$$\begin{aligned} &\cos[2\pi(f_c + 2f_{KF} - (f_c + f_{KF}))t] + \cos(2\pi[f_c + 2f_{KF} + (f_c + f_{KF})]t) \\ &\cos(2\pi f_{KF}t) + \cos[2\pi(2f_c + 3f_{KF})t] \end{aligned}$$

Az f_{KF} frekvenciás komponens jelenléte azt jelenti, hogy van egy másik adó, ami letranszponálva, pont a KF szűrő áteresztőtartományában jelenik meg, az egyébként venni kívánt jelünk mellett. Azokat a frekvenciákat, melyek a venni kívánt jelünkre transzponálódnak **tükörfrekvenciáknak** vagy **tüköradóknak** nevezzük. A tükörfrekvenciák felsőkeverésnél az $f_c + 2f_{KF}$, míg alsó keverésnél az $f_c - 2f_{KF}$ frekvenciák. Szavakkal kifejezve: felső keverés esetén tükörfrekvencia az, amelyet a helyi oszcillátorunk alsó keveréssel transzponálna a középfrekvenciára és vice versa. A tükörfrekvenciák kiszűrése az RF szűrő feladata, ezért azok szelektivitását mindig a középfrekvencia függvényében határozzuk meg.



5.7. ábra. Szuperheterodin vevő

5.3.2. AM rádióadás

Rádiózás tekintetében a frekvenciatartomány különböző intervallumait, az adódó hullámhosszak szerint szoktuk felosztani², így létezik:

- HH - Hosszúhullám – Ez a 30 - 300 kHz tartományt jelenti, kb. 1 - 10km hullámhosszal

²Ez tulajdonképpen az elektromágneses hullámok besorolásának, rádiózásra használt részét fedti le.

- KH - Középhullám – Ez a 300 - 3000 kHz tartományt jelenti, kb. 0,1 - 1km hullámhosszal
- RH - Rövidhullám – Ez a 3 - 30 MHz tartományt jelenti, kb. 10m - 100m hullámhosszal
- URH - Ultrarövidhullám – Ez a 30 - 300 MHz tartományt jelenti, kb. 1m - 10m hullámhosszal

Mivel a frekvencia természeti erőforrásnak számít, így a frekvenciagazdálkodás kormányzati hatáskörbe tartozik. Ez azt jelenti, hogy az egyes frekvenciákon csak az arra engedéllyel rendelkező állomások adhatnak. Ez alól kivételt képeznek az ISM³ sávok, melyek közül a legismertebb (és kihasználtabb) a 2,45 GHz és 5,8 GHz körüli 100, illetve 150 MHz szélességű sávok, melyekben a WiFi és a Bluetooth is üzemel. Az egyes frekvenciákat vagy frekvenciatartományokat különböző célra használják (ebben vannak nemzetközileg és nemzetileg kijelölt sávok), például a 450 MHz-es sáv 11-es csatornáján (457.450 MHz) Rajka település vasútállomásával vehetjük fel a kapcsolatot (de, inkább ne tegyük). Középhullámú AM rádiózásra Európában a 531 kHz és 1611 kHz közötti sávot használják (például a Kossuth Rádiót 450 kHz-en sugározzák), míg UHF FM rádiózására a CCIR sávot használják 87,5 MHz és 108 MHz között.⁴

Először tekintsük át a középhullámú AM rádiózás jellegzetességeit. A középhullámú adás legnagyobb előnye, hogy nagy távolságokra is elér.⁵ Ahogy már említettük, Európában a 531 kHz és 1611 kHz közötti sávon folyik AM rádióadás, és az egyes csatornák 9 kHz távolságban helyezkednek el egymástól. Nem véletlen tehát, hogy középhullámon AM-DSB modulációt használunk, hiszen az FM jelek sugárzásához jóval nagyobb sávzélességre van szükség. A jellemző középfrekvencia $f_{KF} = 455$ kHz, így a szuperheterodin vevő RF szűrőjének a 910 kHz-cel nagyobb tükrőfrekvenciát már el kell nyomnia: ez viszonylag könnyen teljesíthető követelmény, még egy hangolható szűrő esetében is. A 9 kHz távolságban lévő, szomszédos adók elnyomásáról a KF szűrő gondoskodik. A moduláció jellegéből adódik, hogy a lehető legmagasabb átvihető frekvencia 4500 Hz, tehát az emberi beszéd maradéktalanul érthető a vételi oldalon. Az AM adás egyik lényeges tulajdonsága, melyre még hivatkozni fogunk, hogy *mono* az adás.⁶

³Industrial, Scientific and Medical radio bands

⁴Régebben az OIRT sáv volt kijelölve UHF FM rádiózására a 64 és 73 MHz közötti tartományon, a szocialista blokk országaiban, azonban manapság már ezekben az országokban is a CCIR sávot használják.

⁵A Kossuth Rádiót tiszta időben akár Afrikában és Oroszország távoli részein is fogni lehet. Érdekes, hogy a Kossuth Rádiót sugárzó Solti adótorony Európa legnagyobb teljesítményű rádióállomása, 2 MW teljesítménnyel képes adni.

⁶Természetesen létezhet sztereó AM rádióadás, azonban ez soha nem terjedt el, így gyakorlatilag nem is nagyon használják.

Irodalomjegyzék

- [1] Jiří Adámek. *Foundations of Coding: Theory and Applications of Error-Correcting Codes with an Introduction to Cryptography and Information Theory*. John Wiley & Sons, Inc., 1991.
- [2] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Son Inc., 2006.
- [3] Fodor György. *Hálózatok és Rendszerek*. Műegyetemi Kiadó, 2008.
- [4] Sallai Gyula and Abos Imre. A távközlés, információ- és médiatechnológia konvergenciája. *Magyar Tudomány*, pages 844 – 851, 2007.
- [5] Simon Haykin. *Communication Systems*. John Wiley & Sons Inc., 2001.
- [6] Géher Károly. *Híradástechnika*. 1993.
- [7] B. P. Lathi and Zhi Ding. *Modern Digital and Analog Communication Systems*. Oxford Press, 2009.
- [8] Balogh László, Kollár István, Németh József, Péceli Gábor, and Sujbert László. *Digitális Jelfeldolgozás*. 2008. Segédlet a „Digitális jelfeldolgozás” (BMEVIMM4084) tárgyhoz.
- [9] Györfi László, Györi Sándor, and Vajda István. *Információ- és kódelmélet*. 2002.
- [10] Alan V. Oppenheim, Alan S. Willsky, and S. Hamid Nawab. *Signals and Systems*. Prentice Hall Inc., 1997.
- [11] John G. Proakis and Masoud Salehi. *Fundamentals of Communication Systems*. Pearson Education Inc., 2014.