GYÖRFI LÁSZLÓ

Az információtechnológia természettörvényei

Györfi László
matematikus
az MTA rendes tagja

1947-ben született Hercegfalván. 1970-ben diplomázott az ELTE Természettudományi Karának matematika–fizika szakán. 1978-tól a matematikatudomány kandidátusa, 1988-tól akadémiai doktora lett; 1995-től az MTA levelező, majd 2001-től rendes tagja.

Pályáját a Távközlési Kutató Intézetben kezdte, 1975–1990 között az MTA Informatikai és Elektronikai Kutatócsoportban dolgozott. 1990 óta a BME Számítástudományi és Információelméleti Tanszékének egyetemi tanára. 1995 óta vezeti az MTA Informatikai és Elektronikai Kutatócsoportját.

Meghatározó szerepe volt a Budapesti Műszaki Egyetemen a műszaki informatika és az alkalmazott matematika szak tantervének kidolgozásában, valamint a szakok alapításában és indításában. Kifejlesztette és bevezette a tömegkiszolgálás, az információelmélet, a kódelmélet és a matematikai statisztika tárgyakat. Számos hallgató témavezetője volt, közülük három már nemzetközileg elismert egyetemi tanár. Tagja az MTA Távközlési Rendszerek Bizottságának.

Fő kutatási területe: nemparaméteres statisztika, a többszörös hozzáférésű csatornák kódolása.

Az információtechnológia alapvető feladata az információ tömörítése és védelme az információ átvitele, tárolása során. A tömörítés lehet veszteségmentes, amikor az üzenetsorozatot úgy kódolják, hogy az üzenet egyértelműen reprodukálható legyen. Veszteséges tömörítés esetén nem követeljük meg a tökéletes reprodukciót. Az információ védelme jelentheti a sérülés elleni védelmet, továbbá az adatvédelmet – vagyis a titkosítást –, a hozzáférés-védelmet, illetve a hitelesítést – vagyis a manapság oly sokat emlegetett digitális aláírást. Az előadás az információelmélet egyik meglepő és fontos természettörvényét, a hibajavító kódolás elvi határait mutatja be.

Információtechnológiai feladatok

Az információelmélet bizonyos információtechnológiai feladatok gazdaságos megoldásának elvi határait és az ezeket a határokat közelítő kódolási eljárásokat foglalja egységbe. E feladatok közé tartozik az információ tömörítése és védelme az információ átvitele, illetve tárolása során. Az információval, adattal lehet mást is csinálni, például adatkezelést, információfeldolgo-

Adattömörítés:
feladata, hogy egy üzenetsorozatot úgy kódoljon, hogy egyrészt a kódolt sorozat minél rövidebb legyen, másrészt a kódsorozatból az üzenetsorozat egyértelműen reprodukálható legyen.

Veszteséges tömörítés:
alkalmazásakor megengedünk bizonyos torzítást, miközben célunk a gazdaságos, tömör reprezentáció.

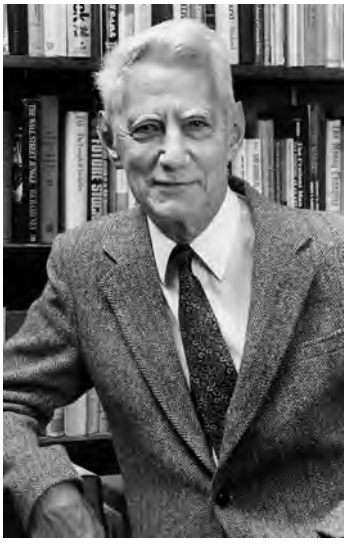
zást stb., mely feladatok – az előbbiekkel együtt – a tág értelemben vett informatika témái.

Az információ tömörítésének, a *forráskódolás*nak két típusát különböztetjük meg. Az egyik a *veszteségmentes* – ezt **adattömörítés**nek is hívjuk –, a másik a **veszteséges tömörítés**, amely megenged torzítást is a reprodukció során.

Az adattömörítés feladata, hogy egy üzenetsorozatot gazdaságosan reprezentáljon, vagyis kódoljon úgy, hogy egyrészt a kódolt sorozat minél rövidebb legyen, másrészt a kódsorozatból az üzenetsorozat egyértelműen reprodukálható legyen. Ilyen problémával találkozunk, ha például könyvet, programot, adatsorozatot kell tömöríteni.

Képzeljük el, hogy a magyar szépirodalmat szeretnénk CD-re vinni. Nem közömbös, hogy hány CD-n fér el, tehát érdemes tömöríteni. Egyáltalán nem nehéz 1:10-es tömörítési arányt elérni, amikor is tömörítéssel tízszer kevesebb CD kell, mint tömörítés nélkül. Egy másik példa, ha mobiltelefonon szeretnénk szöveget átküldeni: ilyenkor a kis adatsebességű mobilon akkor tudjuk gyorsan átküldeni a szöveget, ha átküldés előtt tömörítjük; 1:10-es tömörítéssel például tizedannyi idő alatt tudjuk átküldeni a tömörített üzenetet.

Az első tömörítő eljárás a *Morse-kód* volt, amely az ábécé gyakran előforduló betűihez rövid, a ritkábban előfordulókhoz hosszabb ti-tá (mai szóhasználattal bináris) kódszavakat rendelt.

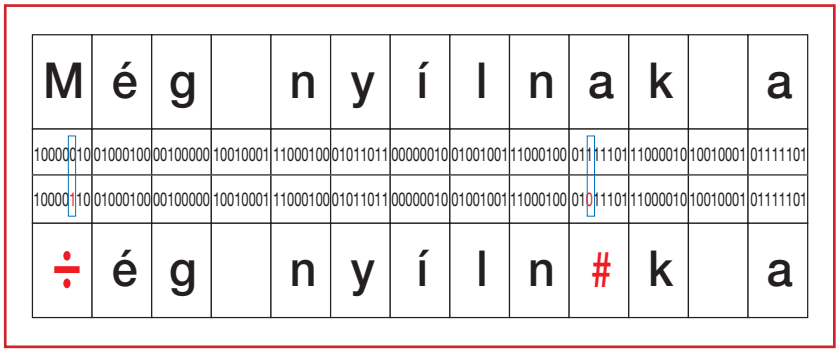


Claude Shanon (1916–2001)

A tömörítés minőségét a *tömörítési aránny*al jellemezhetjük, ami a tömörített hossznak és az eredeti adatsorozat hosszának az aránya. Mindenki számára világos, hogy a tömörítési aránynak, a tömöríthetőségnek van határa. Az adattömörítés természettörvényét Claude Shannon fedezte fel, amikor kiszámította a tömörítési arány elvi alsó határát, a *forrásentrópiát*, és megadott olyan kódolási eljárásokat, amelyek ezt az elvi alsó határt elérik. A mindennapi gyakorlatban is alkalmazunk ilyen tömörítő eljárásokat, amikor különböző tömörítő programokat használunk.

A veszteséges forráskódolás esetén nem cél a tökéletes reprodukció, vagyis megengedünk torzítást, de a cél továbbra is a gazdaságos, tömör reprezentáció. Mindennapi alkalmazásai a beszéd, a zene, a kép és a videó tömörítése. Kép tömörítése esetén például nyilván felesleges megkövetelni, hogy a reprodukált kép képpontról képpontra egyezzen meg az eredeti képpel, csupán azt szeretnénk, hogy szemmel ne érzékeljünk romlást. Ebben a feladatban két célfüggvényünk van. Az egyikkel mérjük a tömörítést, a másikkal a torzítást, vagyis azt, hogy a tömörítés utáni reprodukció mennyire hasonlít az eredetire. Ha két, egymásnak ellentmondó célunk van, nevezetesen alacsony értéken tartani mind a tömörítési arányt, mind a torzítást, akkor a probléma úgy kezelhető, ha az egyiket – például a torzítást – egy előírt értéken rögzítjük, és emellett minimalizáljuk a tömörítési arányt. Az elvi határ ekkor is tisztázható, de az elvi határt közelítő kódok ma még nem ismertek. Ugyanakkor léteznek a gyakorlatban hatékony veszteséges tömörítő eljárások, amelyeket sikerrel alkalmaznak a mobiltelefonban és a kép, a videó és a zene kódolására.

Az információ védelme jelentheti az információ sérülése elleni védelmet



1. ábra. *A küldött bitek meghibásodásának hatása egy szövegben*

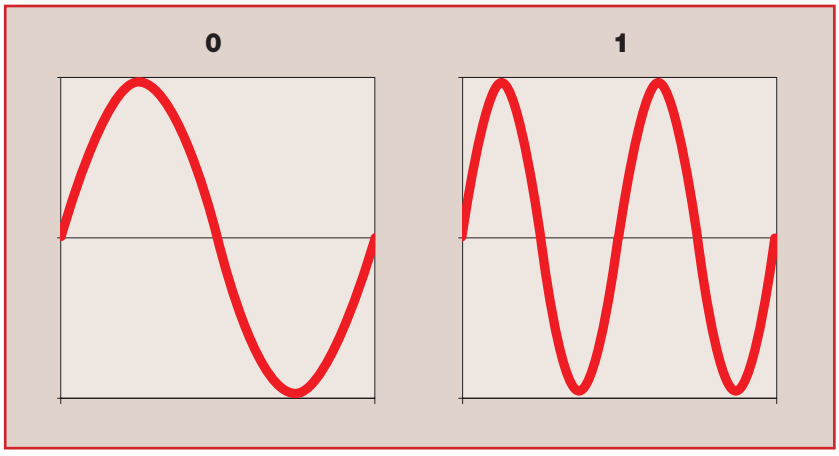
(csatornakódolás), vagy az adatvédelmet (titkosítás), vagy a hozzáférésvédelmet, illetve a hitelesítést (digitális aláírás). Ha például interneten szeretnék egy banki tranzakciót lebonyolítani, akkor nyilván elvárom, hogy a megadott adatok pontosan legyenek továbbítva (hibajavító kódolás), más személy ne tudja meg ezeket az adatokat még akkor sem, ha az információtovábbítás nyilvános hálózaton, például mobileszközön történik (titkosítás), a bank számára pedig bizonyított legyen, hogy valóban én kezdeményeztem a tranzakciót (digitális aláírás).

A védelmi feladatok közül nézzük részletesen a **csatornakódolás**t, más néven **hibajavító kódolás**t, mégpedig először néhány hibajavító elvet és technikát. A közeg zavarai miatt az adóban a modem bemenete és a vevőben a modem kimenete különbözhet (1. ábra). Az adótól a vevőbe kell eljuttatni az üzenetet egy fizikai közegen (vezeték, rádiós frekvenciasáv stb.) keresztül. A távközlő mérnök is ezzel a feladattal foglalkozik. Nevezetesen az adóba és a vevőbe olyan áramköröket, modemeket tervez, amelyek az adóban a bitekhez a közeghez illeszkedő jelalakokat rendelnek, illetve a vevőben a torzított jelalakokból következtetnek a lehetséges bitekre (2. és 3. ábra).

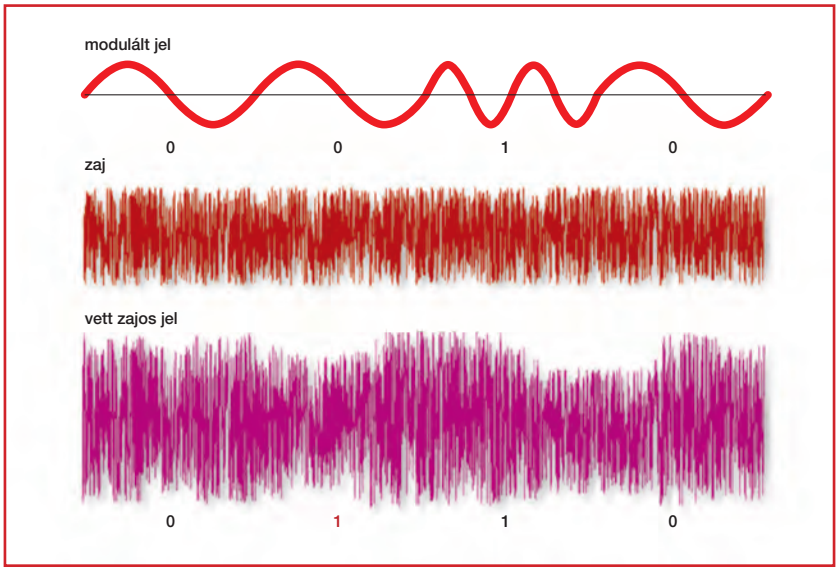
A távközlő mérnök feladata az, hogy ennek a hibázásnak a valószínűségét alacsony értéken tartsa. Itt kezdődik az **információelmélet** feladata, amikor a távközlő mérnök eredményét adottságként tekintjük, amelyen vagy nem tudunk, vagy nem akarunk javítani. Tudomásul vesszük, hogy adott egy többé-kevésbé megbízhatatlan eszköz, ezt nevezzük csatornának, és ennek segítségével akarunk megbízható átvitelt biztosítani.

Csatornakódolás, hibajavító kódolás:
célja, hogy a hibásan vett kódszóból vissza lehessen állítani az eredeti üzenetet.

Információelmélet:
információtechnológiai feladatok (információ tömörítése és védelme) gazdaságos megoldásának elvi határaival és az ezeket a határokat közelítő kódolási eljárásokkal foglalkozó tudomány.



2. ábra. *Példa jelalakokra*

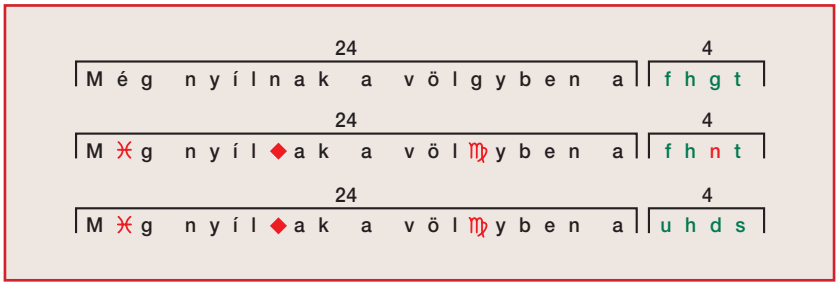


3. ábra. *A moduláció és a zajos vétel folyamata*

A csatornakódolásnak két típusa van. Az első a **hibajelző kódolás**, amely még napjainkban is döntően jellemzi az adatátvitelt. Az adó az üzenetsorozatot blokkokra osztja, és minden blokkot ellát úgynevezett **hibajelző (paritás-ellenőrző) karakterek**kel. Ezt hívjuk *redundanciá*nak is. Az üzenetet és a paritás-ellenőrző karaktereket együtt *kódszó*nak nevezzük. A vevő a vett blokkból kiszámolja a hibajelző karaktereket, és ha egyezést talál, akkor ezt nyugtázza az adónak, egyébként újraküldést kér. Ebben az esetben rendelkezésre áll egy visszairányú csatorna a nyugták számára. A modem is ezt az elvet követi. Vannak olyan kódok, például a Reed–Solomon-kódok, amelyeknél *m* darab paritás-ellenőrző karakter esetén bármely, legfeljebb *m* darab karakter meghibásodását lehetséges jelezni. A 4. ábra példájában egy 24 betű hosszú üzenetből a Reed–Solomon-kódot használva kiszámolunk 4 hibajelző betűt (első sor). A második sorban szerepel a vett 28 betű, ahol piros színnel jelöltük meg a 4 hibásan vett betűt. A vevő a vett sorozat első 24 betűje alapján kiszámolja a 4 hibajelző betűt (harmadik sor), és mivel a második és a harmadik sor utolsó 4 betűje különbözik, ezért észreveszi a hibát.

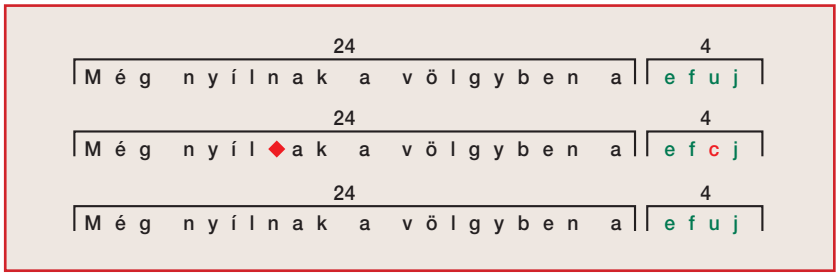
Hibajelző kódolás:
célja, hogy észre lehessen venni, ha a vételben hiba történt.

Hibajelző (paritás-ellenőrző) karakterek:
a védendő üzenetet az üzenettől függő karakterekkel látjuk el, amelyek segítségével az esetleges hibákat tudjuk jelezni vagy javítani.



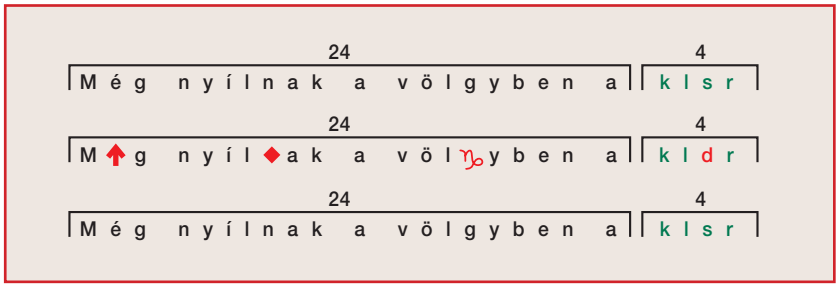
4. ábra. *Hibajelző kódolás*

A hibajavító kódolás akkor is használható, ha ilyen visszairányú csatorna nincs. Erre példa lehet az űrszonda problémája, ahol ráadásul a nagy távolság miatt a jelszint jóval kisebb, mint a zajszint, tehát gyakori a hibázás. Az 5. ábra szemlélteti, hogy egy 4 hibajelző betűt használó Reed–Solomon-kód képes 2 hibát kijavítani.



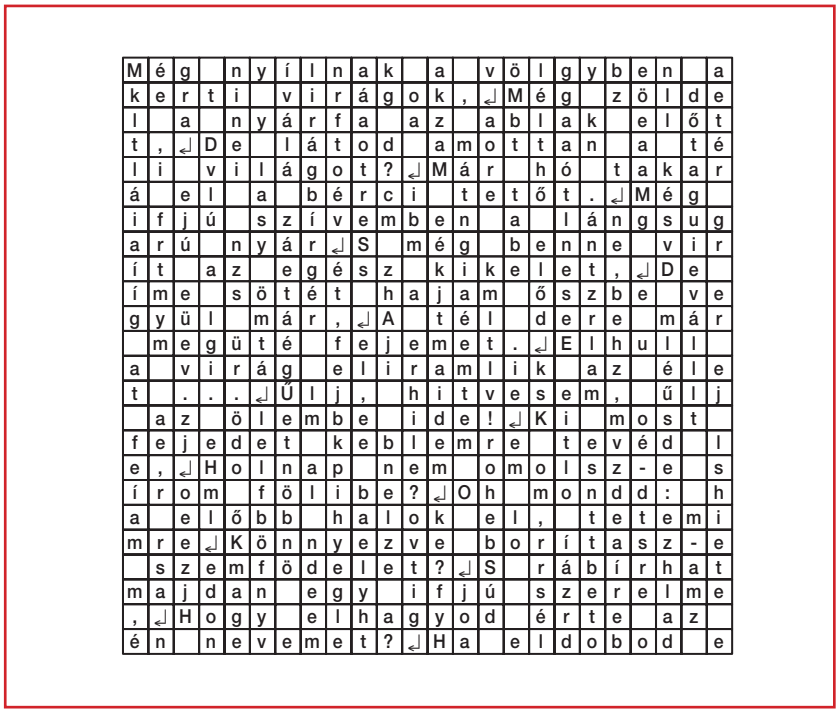
5. ábra. *Hibajavító kódolás*

Ha t darab hiba történt, akkor $2t$ ismeretlenünk van, a t hiba helye és a t megsérült karakter. Lényegében ez az oka annak, hogy az előbb említett, m darab paritás-ellenőrző karaktert használó Reed–Solomon-kód képes megtalálni m ismeretlent, tehát bármely legfeljebb $m/2$ darab hibát kijavítani.



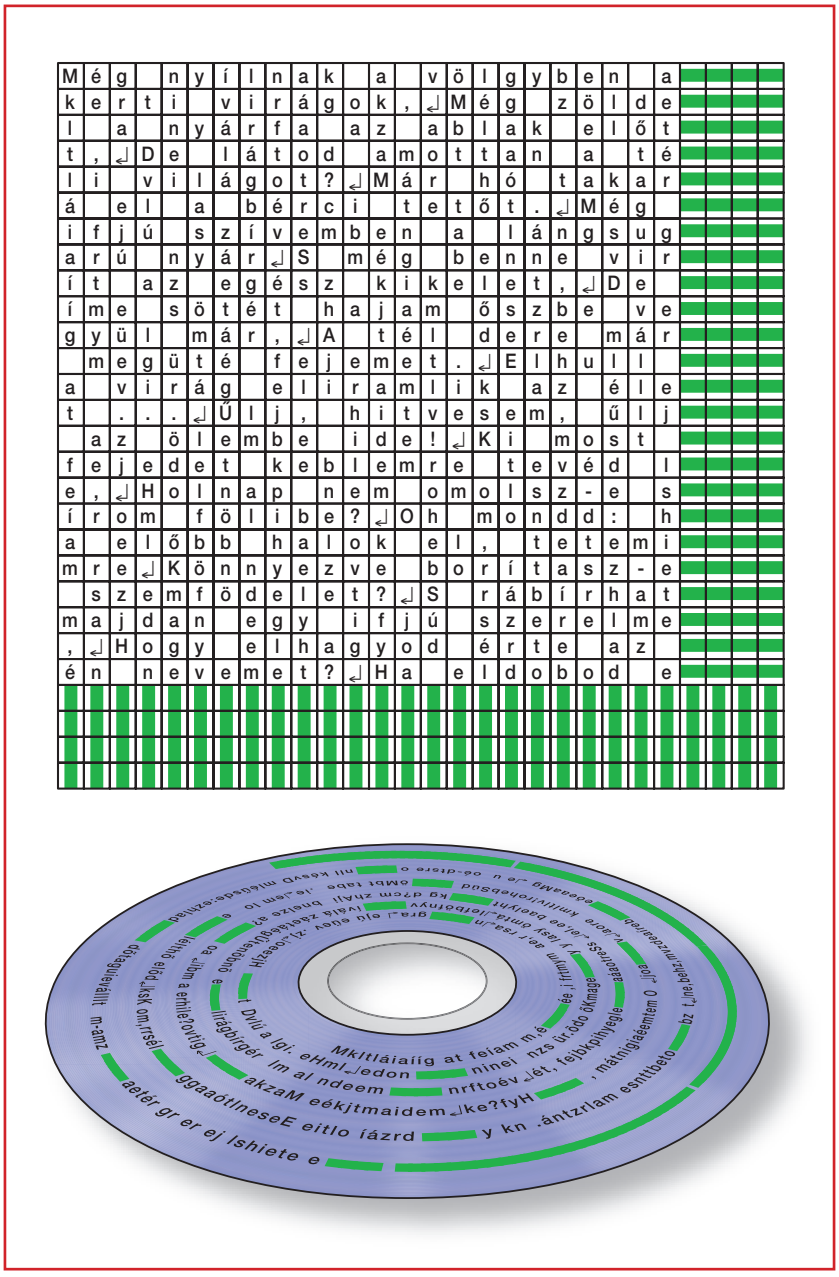
6. ábra. *Törléses hiba javítása*

Érdemes egy speciális hibázási mechanizmusról beszélni, amikor a hibás karakterek helyét ismerjük, ezt hívjuk *törléses hibá*nak. Ha t darab törléses hiba történt, akkor csak t ismeretlenünk van, a t meghibásodott karakter. Ennek megfelelően az előbb említett, m darab paritás-ellenőrző karaktert használó Reed–Solomon-kód képes bármely, legfeljebb m darab törléses hibát kijavítani (6. ábra).

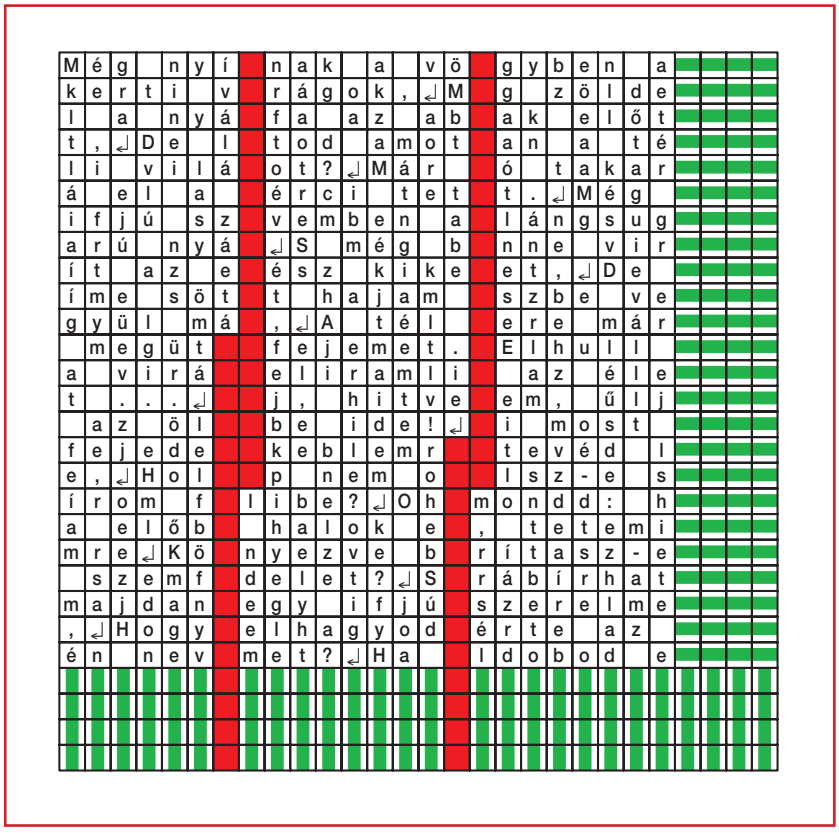


7. ábra. *A szöveget egy 24 x 24-es táblázatba írjuk*

A visszairányú csatorna hiányára egy másik példa a CD, ahol a vett hibás betűsorozat esetén nem kérhetek ismételt küldést. Itt ráadásul a hibázás mechanizmusa kellemetlen, mert a hibák csomókban fordulnak elő. Bor Zsolt előadásából is tudhatjuk (ME 1. köt., 307–321. p.), hogy a CD-lemezen a digitális információt a spirálpályák mentén elhelyezkedő negyedhullámhossz mélységű gödröcskék hossza és a gödröcskék távolsága tartalmazza. Ha a CD a winchesterhez hasonlóan az olvasó optikával és mechanikával együtt egy zárt dobozban lenne, akkor gyakorlatilag nem fordulna elő hiba, viszont ekkor éppen a CD fő előnyei tűnnének el. A lemez felületének esetleges sérülésekor vagy a lencse szennyeződésekor azonban egész karaktersorozatok sérülnek meg, ezek a *csomós hibák*. A csomós hibák ellen védekezik az *átfűzési (interleaving) technika*. Az üzeneteket (hangmintákat)



8. ábra. *Minden oszlopot és minden sort ellátunk 4 hibajelző betűvel (zöld csíkok), és a táblázatot oszlopfolyamatosan írjuk a CD-re*



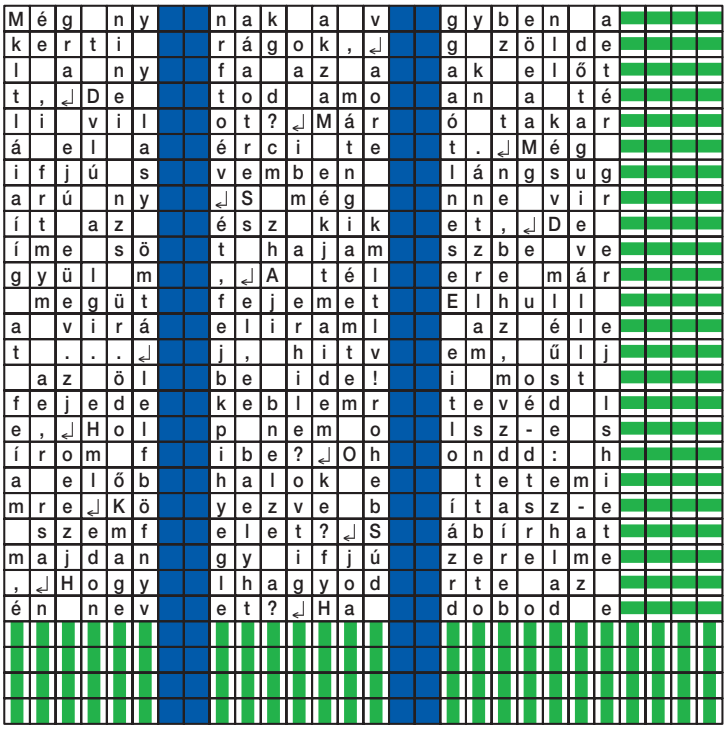
9. ábra. *A CD-n a hibák oszlopfolyamatosan történnek, amelyeket a piros csíkok szemléltetnek*

egy 24×24-es táblázatba írjuk (7. ábra), és minden sort és minden oszlopot 4 paritás-ellenőrző karakterrel kódolunk (8. ábra). Az így kapott 28×28-as táblázatot oszlopfolytonosan tároljuk a lemezen. A csomós hibák az oszlopok mentén fordulnak elő (9. ábra), és ezeket a hibás oszlopokat hibajelzéssel detektáljuk; a hibás oszlopok sorszámai a sorokra elvégzett hibajavításszámra a hibahelyek, azaz mesterségesen törléses hibákat generáltunk, tehát legfeljebb 4 hibás oszlop kijavítható (10. ábra).

Természetszerűen vetődik fel egy hibajavító kódolás minőségének a kérdése. Jellegzetesen egy kódot két számmal jellemzünk, az egyik a kihasználtság, az üzenethossznak és a kódszóhossznak az aránya, a másik a hibavalószínűség, vagyis annak a valószínűsége, hogy a dekódolt üzenet nem egyezik az eredeti üzenettel. A csatornakódolásnak az az alapproblémája, hogy milyen kihasználtságot érhetünk el, ha nagyra törően a hibavalószínűséget kis értékre akarjuk leszorítani.

A véletlen törvényei

Az eddig tárgyalt feladatokban az információ legfontosabb tulajdonsága az volt, hogy véletlen. Ha a tömörítendő adat nem lenne véletlen, azaz adott lenne, akkor nem kellene tömöríteni. Ha a hibázó csatorna nem lenne véletlen, akkor a javítás is triviális lenne, következésképp az információelmélet törvényei főleg a véletlen törvényeit használják fel, illetve fejlesztik tovább.



10. ábra. *Az oszlopok szerinti hibajelző betűk segítségével jelezhetjük a hibás oszlopokat, amelyeket lilával szinezünk. A sorok menti hibajavítás számára ezek ismert helyű hibák, azaz törléses hibák. 4 hibás oszlop esetében a sorok szerinti hibák kijavíthatók*

A véletlennel kapcsolatban a legtöbb ember gyanakszik, hiszen az egyrészt jelenthet szerencsét, ami elkerüli, másrészt jelenthet bajt, katasztrófát, ami viszont megtalálja. A *valószínűség-számítás* a véletlen tömegjelenségek törvényeit tárja fel, ugyanakkor egy szuverén egyén nem szereti, ha a tömeg egy jelentéktelen pontjaként kezelik, tehát elsőre úgy tűnik, hogy számára a valószínűségszámítás érdektelen. Ennek az ellenkezőjéről szeretnék mindenkit meggyőzni.

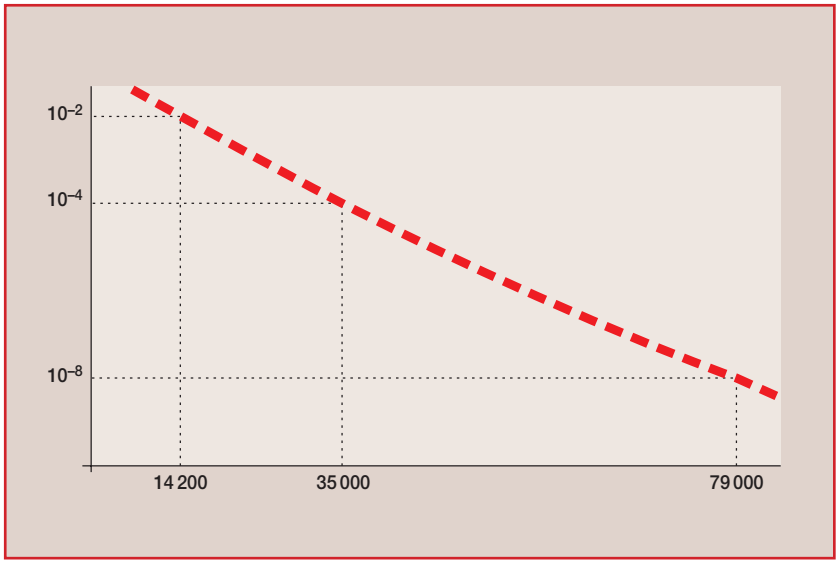
A klasszikus valószínűség-számítás főleg a szerencsejátékok, illetve a matematikai statisztika bizonyos problémáival foglalkozott. Ez utóbbi esetén általában kevés adatból próbáltak törvényszerűséget levezetni, azaz jellegzetesen olyan megállapításokat, amelyek nagy, körülbelül 95 százalékos biztonsággal igazak. Kérdés az, hogy ez a 95 százalék tényleg nagy-e az egyén szempontjából, aki ezt a törvényszerűséget fel akarja használni. Ha nyáridőben a kedvenc meteorológusom reggel azt mondja, hogy a zápor valószínűsége 5 százalék, akkor ez számomra csak annyit jelent, hogy vagy esik, vagy nem, hiszen ha bőrig áztam, akkor nem vigasztal, hogy ennek pici volt a valószínűsége. A valószínűség-számítás jelentősége ott kezdődik, amikor a törvényszerűség helyett törvény van, vagyis a valószínűből majdnem biztos – pestiesen szólva: tuti – lesz. Mindenkinek van egy tapasztalati fogalma a tutiról. Az, hogy nem lesz hármas találatom a lottón, az valószínű. (A hármas találat valószínűsége körülbelül 0,0008.) Teljesen szubjektív, hogy az a kijelentés, hogy nem lesz négyes találatom, tuti-e vagy ezt csak az ötös találatra mondom. (A négyes találat valószínűsége körülbelül 10^{-5}, az ötösé 10^{-8}.) Törvény alatt a későbbiekben a tutit értem, vagyis amikor a véletlen tömegjelenséggel kapcsolatban ilyen értelemben eltűnik a véletlen.

A valószínűség-számítás legfontosabb törvénye a *nagy számok törvénye,* amely szerint, ha egy véletlen esemény bekövetkezésére „sok” kísérletet végzünk, és kiszámítjuk a bekövetkezések számának és a teljes kísérlethossznak az arányát, akkor ez az arány „közel” lesz egy számhoz, mégpedig a véletlen esemény valószínűségéhez. Kérdés, hogy mit jelent a „sok”, és mit jelent a „közel”. Lássunk erre egy példát!

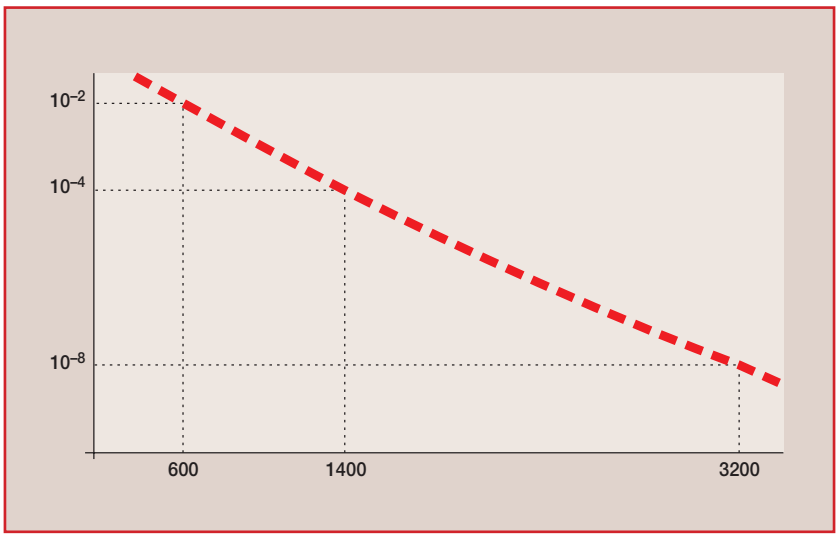
Egy képzeletbeli ország parlamenti választásának az estéjén a két nagy párt elnöke egy exit poll-felmérés alapján már az urnazáráskor szeretné tudni, hogy mi a listás szavazás eredménye. Tételezzük fel, hogy az erőviszonyok eléggé kiegyenlítettek; például mindkét elnök legfeljebb 49 százalékos eredmény esetén is szeretné ezt tutira tudni este 7-kor. A felmérés akkor hibás, ha legalább 50 százalékos, mivel ekkor egyikük túl korán suttogja szemlesütve világgá, hogy „győztünk”. Megfordítva, ha legalább 51 százalékos eredményt ér el, de a felmérés legfeljebb 50 százalékos, akkor is hibázunk, hiszen ekkor az elnök feleslegesen gratulál az ellenfelének. Ilyen kiélezett helyzetben tehát a tűrés 1 százalék. Kérdés, hogy egy exit poll-felmérés során hány szavazót kell megkérdezni ahhoz, hogy 1 százalék tűréssel tuti eredményt kapjunk.

Bizonyítható, hogy adott tűrés mellett a téves következtetés valószínűsége, a hibavalószínűség a mintanagyságnak exponenciálisan gyorsan csökkenő függvénye, ami azt jelenti, hogy a mintanagyság megduplázásával a hibavalószínűség a négyzetére csökken (11. ábra).

Ezek elborzasztó mintanagyságok: 10^{-4} -es hibavalószínűséghez 35 ezer szavazót kell megkérdezni, mégpedig szigorúan véletlenszerűen, azaz a választói névjegyzékből 35 ezer nevet kisorsolni, megkeresni a szavazókörzetét, és abból a szavazókörzetből valakit megkérdezni. Ellenérdekű felek együttműködésére jó példa lehet az, hogy ha mindkét párt elnöke megrendel egy felmérést, és mindegyiknek a költségvetése csak 17 500-as felmérésre futja, akkor kicserélik az adataikat, és rögtön van tuti eredményük. Valaki persze joggal vetheti fel, hogy a közvélemény-kutatások általában csak ezres mintaszámmal dolgoznak. Ez akkor indokolt, ha a helyzet nem any-



11. ábra. *A hibavalószínűség függése a mintanagyságtól 1 százalékos tűrés esetén*

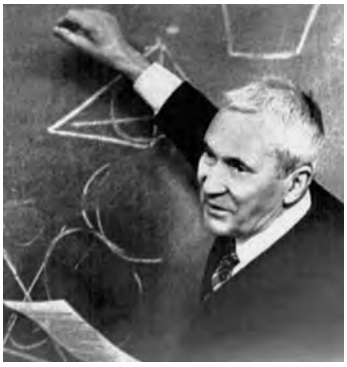


12. ábra. *A hibavalószínűség függése a mintanagyságtól 5 százalékos tűrés esetén*

nyira kiélezett. Ha például 5 százalék tűrés elég, akkor lényegesen kisebb minta szükséges (12. ábra).

Az is megmutatható, hogy ezek az adatok nem függnek attól, hogy hányan vesznek részt a szavazásban, tehát adott tűrés és hibavalószínűség esetén ugyanannyi minta kell az Amerikai Egyesült Államok elnökválasztási eredményének előrejelzésekor, mint a magyar parlamenti választáskor, ezért exit poll-felmérést csak listás szavazás esetén érdemes készíteni.

A véletlen törvényeinek jelentős alkalmazási területe a kvantumfizika. Itt, a Mindentudás Egyetemén is több ilyen témájú előadást hallhattunk, amikor a fizikus az elemi részecskék véletlen viselkedését, kölcsönhatását egy egyszerű modellel jellemzi, és valószínűség-számítási technikával – többnyire egy rafinált nagy számok törvényével – levezeti a makroszkopikus viselkedést. Ha ez a levezetett viselkedés összhangban van a mérésekkel, akkor határtalan örömmel állapítja meg, hogy felfedezett egy új részecskét. (Lásd Horváth Zalán ME 3. köt. 155–171. p.; Mihály György ME 2. köt. 241–257. p.; Sólyom Jenő ME 2. köt. 273–288. p. és Vicsek Tamás ME 1. köt. 223–234. p. előadását.)



Kolmogorov, Andrej Nyikolajevics (1903–1987)

A modern valószínűség-számítást Andrej Nyikolajevics Kolmogorov alapozta meg egy 1933-ban publikált cikkében. Magyarországon e diszciplína úttörője Rényi Alfréd volt.

A hibajavító kódolás törvénye

Térjünk vissza a hibajavító kódolás problémájára! Emlékeztetnék arra, hogy egy kódot két számmal jellemzünk, az egyik a kihasználtság, az üzenethossznak és a kódszóhossznak az aránya, a másik a hibavalószínűség, vagyis annak a valószínűsége, hogy a dekódolt üzenet nem egyezik az eredeti üzenettel. Mindenki számára természetes, hogy a csatorna kihasználtsága növelhető a hibavalószínűség növelésével. Példaként tekintsük a bináris szimmetrikus csatorna esetét, vagyis amikor a csatorna bemenete és kime-

nete is 0 vagy 1 értékű, és p annak a valószínűsége, hogy a bemenet és a kimenet különbözik. Legyen $p = 0{,}1$, vagyis egy elég rossz csatornánk van, hiszen átlagosan minden tizedik bit elromlik, tehát átlagosan minden három karakterből kettő elromlik.

Épeszű ember számára egy ilyen csatorna fabatkát sem ér. Megmutatom, hogy egy ilyen vacak csatorna is lehet értékes. Legyen az a feladatunk, hogy egy hosszú, például 1000 soros programot akarunk átvinni úgy, hogy igényesek vagyunk: azt kérjük, hogy a teljes átvitel meghibásodásának a valószínűsége legyen mondjuk 10^{-6}.

Nézzünk először egy mindenki számára természetesen adódó technikát, az *ismétléses kód*ot! Ha csak egyetlen bit átvitele lenne a feladatunk, akkor alkalmazhatjuk ezt az egyszerű eljárást. A 0-t például három 0 küldésével, azaz 000-val, az 1-et három 1 küldésével kíséreljük meg, és a vevőben arra szavazunk, amelyik többségben van (13. ábra).

Rényi Alfréd (1921–1970)

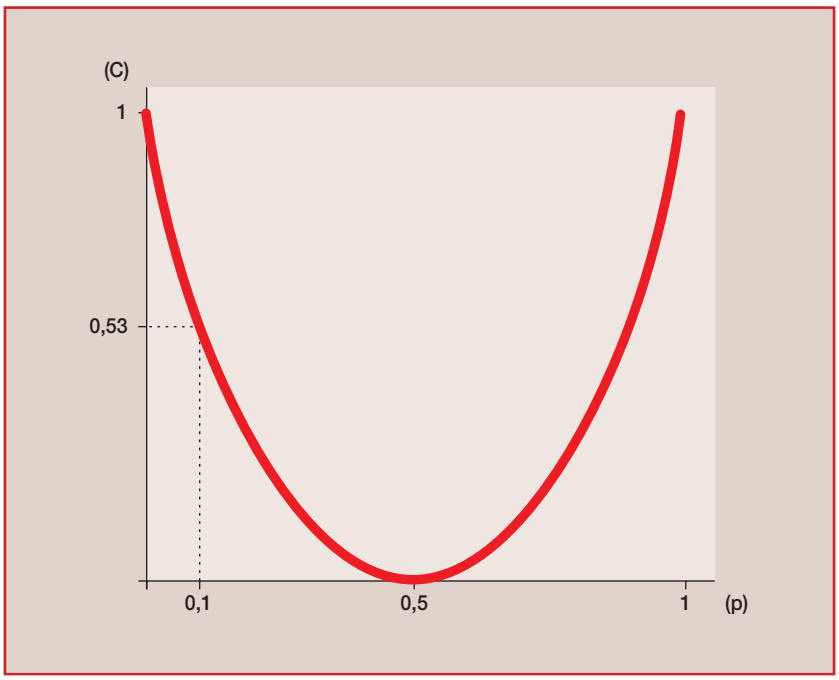
13. ábra. *Ismétléses kód*

Ellenőrizhető, hogy 19 hosszú ismétlés esetén az átvitel hibavalószínűsége már 10^{-6}, de pazaroltunk, mivel a csatornát 1/19-es, azaz körülbelül 5 százalékos kihasználtsággal üzemeltettük.

Ha a blokk-kódolási elvet alkalmazzuk, vagyis nem egy bitet, hanem egy k hosszú üzenetblokkot kódolunk n hosszú kódszóba, akkor nyilván rögzített k/n csatornakihasználtság mellett érdekel bennünket a dekódolás hibavalószínűsége, és mindenki azt várja, hogy kis hibavalószínűséget csak kis kihasználtság árán érhetünk el.

Érdekes módon ez nem így van. A fentebb már emlegetett Claude Shannon 1948-ban publikált cikkében harminckét évesen nemcsak az adattömörítés, hanem a csatornakódolás elvi határát – a „fénysebességet” – is felfedezte, és ő bizonyította elsőként, hogy létezik tökéletes titkosító.

Shannon – véleményem szerint – a csatornakódolás esetén volt a legmerészebb, a legzseniálisabb. Felfedezte, hogy az elvi határ szempontjából nem feltétlenül kell a kihasználtság csökkentésével fizetni a hibavalószínűség csökkentéséért, nem kell ilyen földhöz ragadt módon gondolkodni. Felfedezte, hogy létezik a kihasználtságnak egy szintje, ezt nevezzük



14. ábra. *Kapacitásgörbe*

csatornakapacitásnak (C), úgy, hogy ha a rögzített kihasználtságot C alatt tartjuk, akkor az üzenethossz növelésével található olyan kód, amely segít abban, hogy a dekódolás hibavalószínűsége tetszőlegesen kicsi legyen (14. ábra).

Csatornakapacitás:
a kihasználtságnak az a maximális értéke, amely alatt az üzenethossz növelésével található olyan kód, hogy a dekódolás hibavalószínűsége tetszőlegesen kicsi lehet.

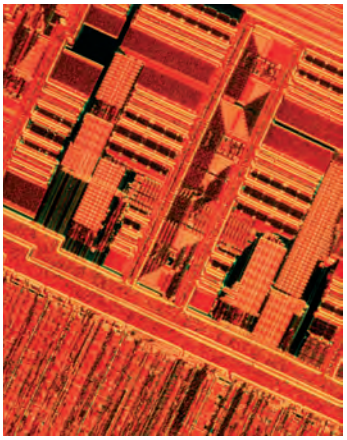
A fenti példában $p = 0{,}1$ esetén $C = 0{,}53$, tehát a csatorna 50 százalékos kihasználtságával elérhető, hogy annak a valószínűsége, hogy egy hosszú programnak legalább egy karaktere elromoljon az átvitel során, legyen kisebb, mint 10^{-6}, és csak a program méretével azonos hosszúságú redundanciát kell hozzáadnunk a kódolás során. Nyilvánvaló, hogy léteznek az ismétléses kódnál hatékonyabb eljárások, de a csatornakódolási tétel minden józan elvárást felülmúl.

Képzeljük el, hogy egy 10 bites, tehát igen rövid üzenetet szeretnénk 50 százalékos kihasználtsággal, azaz 20 bit hosszú kódszavakkal átvinni. Bár a legkisebb hibavalószínűségű kódot nem tudjuk megtalálni, de magát a legkisebb hibavalószínűséget jól tudjuk becsülni. Az eredmény: ez a hibavalószínűség túl nagy, és ettől elcsüggedünk. Azt mondja erre Shannon, hogy ne bánkódjunk, ha egy egyszerű feladatot nem tudunk megoldani, akkor próbálkozzunk egy nehézzel, egy jóval nehezebbel, nevezetesen ne 10 bites, hanem 1000 bites üzenetet küldjünk át 50 százalékos kihasználtsággal, azaz 2000 bit hosszú kódszavakkal. Itt jön az igazi meglepetés: ekkor a minimális hibavalószínűség már mindenki számára elfogadhatóan kicsi lesz. Nyilván történelmietlen dolog eljátszani azzal a gondolattal, hogyan alakult volna ez a diszciplína, ha Shannon meg sem születik. Meggyőződésem, hogy a csatornakapacitást máig sem találták volna fel, hiába az eddig összegyűlt tapasztalat a digitális távközlés területén.

Az üzenethossz – és ezzel a kódszóhossz – növelésével egy tömegjelenséget konstruálunk úgy, hogy az eredmény, a biztonságos átvitel tuti lesz az egyén, a távközlési szolgáltatás felhasználója számára, és ehhez a szolgáltató-

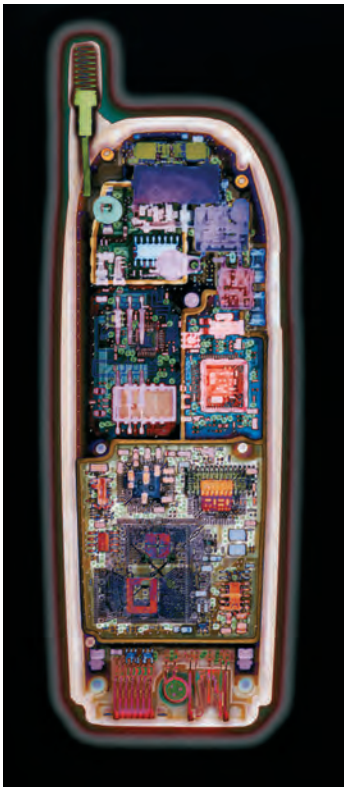
nak nem kell pazarlóan bánni a jellegzetesen igen drága távközlési erőforrással. Ha egy csatorna értékét, árát csak a kapacitása határozná meg, akkor a fenti csatorna feleannyit érne, mint egy nem hibázó csatorna – azzal is indítottam a példát, hogy ez egy mit sem érő, vacak csatorna. Hangsúlyozni kell azonban, hogy a kapacitás a hasznosítható kihasználtságok elvi határa, elvi maximuma, és a zajos csatornák zöménél ezt ma még igen nehéz megközelíteni. A GSM-ben például csúcsidőben is csak a kapacitásnak körülbelül 10 százaléka a kihasználtság.

Visszatérve a csatornakapacitásra, joggal vetődik fel a kérdés, hogy miért nem működik a valamit valamiért elv, a hibavalószínűség leszorításához miért nem kell a kihasználtságot lerontani. Shannon itt a véletlent többszörösen is munkára fogta. Egyrészt a kódolás bevezetésével egy ügyes kísérletet tervezett, ahol a véletlenszerűen hibázó csatorna a kísérlet egy komponense, másrészt a jó kód létezését egy ravasz véletlen kódválasztással bizonyította.



Szilikonchipek

Számomra bámulatos Shannon képzelőereje és absztrakciós készsége. A nagy tudományos felfedezésekhez többnyire egy új, az addigi elméletekkel ütköző tapasztalat vezetett, márpedig 1948-ban egyetlenegy példa létezett digitális kommunikációra: a távíró, amelynél viszont nem volt szigorú előírás a hibavalószínűségre. A 20. század tudománytörténete minőségileg más, új gondolkodási technikákat eredményezett. Gondoljunk arra például, hogy egészen Descartes-ig úgy vélték, hogy az egyenletes mozgás fenntartásához is erőre van szükség, ugyanis még nem tudtak olyan pontosan sebességet mérni, hogy ennek a kiinduló feltételnek, hipotézisnek a hibája kiderüljön. Ezek után viszont könnyű dolga volt Newtonnak, hiszen csak a differenciálszámítást kellett kidolgoznia, majd kimennie az almáskertbe. Ugyanakkor még a 20. századi elméleti fizika nagyszerű eredményei között is csak elvétve akad olyan törvény, amely addig nem tapasztalt jelenségről szólt, azaz egy elméleti modell alapján először prognosztizálták a jelenséget, és csak utána „mérték ki" laboratóriumban. Örömmel tapasztaltam, hogy ilyen eredmények Magyarországon is vannak, mégpedig fiatal fizikusoké, ugyanis a 2004-es Talentum-díj egyik kitüntetettje, Domokos Péter alacsony hőmérsékletek területén két jelenséget is megjósolt, amelyek létezését később Párizsban, illetve Stanfordban laboratóriumi kísérlettel bizonyították.



Shannon az információtechnológia természettörvényeit akkor fedezte fel, amikor még nem is létezett digitális távközlés. 1948-ban ugyanezen a kutatóhelyen, a Bell Laboratóriumban találták fel a tranzisztort, de a kódolási, dekódolási eljárásokat hardverben, digitális céláramkörökben lehetett csak megvalósítani még harminc évig, ezért csupán katonai hírközlési és űrkutatási feladatokban használták fel az információelmélet eredményeit. A mikroprocesszor megjelenésével a dekódolási algoritmusokat már olcsón, szoftverben implementálták, és így megnyílt az út a tömeges digitális távközlési szolgáltatások előtt. De 1948-ban a szóban forgó jelenségeket Shannonnak még „fejben" kellett lejátszania.

Ajánlott irodalom

Bertsekas, Dimitri – Gallager, Robert: Data Networks. New Yersey: Prentice Hall, 1992.

Blahut, Richard E.: Theory and Practice of Error Control Codes. Massachusetts: Addison-Wesley, 1983.

Bor Zsolt: A mindentudó fénysugár. In: Mindentudás Egyeteme 1. köt., Bp.: Kossuth K., 2004: 307–321.

Buttyán Levente – Vajda István: Kriptográfia és alkalmazásai. Bp.: Typotex K., 2004.

Buttyán Levente – Györfi László – Vajda István: Adatbiztonság: titkosítás, hitelesítés, digitális aláírás. *Magyar Tudomány,* 2005. 5. sz

Cover, Thomas M. – Thomas, Joy A.: Elements of Information Theory. New York: Wiley, 1991.

Csibi Sándor (szerk.): Információ közlése és feldolgozása. Bp.: Tankönyvkiadó, 1986.

Csiszár Imre – Körner János: Information Theory: Coding Theorems for Discrete Memoryless Systems. Bp.: Akadémiai K., 1981.

Gallager, Robert G.: Information Theory and Reliable Communication. New York: Wiley, 1968.

Györfi László – Győri Sándor – Vajda István: Információ- és kódelmélet. Bp.: Typotex K., 2000.

Györfi László: Claude E. Shannon (1926–2001). *Magyar Tudomány,* 46.=108. évf. (2001) 5. sz.: 614–618.

Horváth Zalán: Mikrokozmosz – világunk építőköveinek kutatása. In: Mindentudás Egyeteme 3. köt., Bp.: Kossuth K., 2004: 155–171.

Linder Tamás – Lugosi Gábor: Bevezetés az információelméletbe. Bp.: Tankönyvkiadó, 1990.

Mihály György: Mire jó a kvantumfizika? In: Mindentudás Egyeteme 2. köt., Bp.: Kossuth K., 2004: 241–257.

Nemetz Tibor – Vajda István: Algoritmusos adatvédelem. Bp.: Akadémiai K., 1991.

Rényi Alfréd: Valószínűségszámítás. Bp.: Tankönyvkiadó, 1966.

Simonyi Károly: A fizika kultúrtörténete a kezdetektől 1990-ig. 4., átdolg. kiad. Bp.: Akadémiai K., 1998.

Sólyom Jenő: Az alacsony hőmérsékletek titkai. In: Mindentudás Egyeteme 2. köt., Bp.: Kossuth K., 2004: 273–288.

Szász Domokos: Kolmogorov, a kozmikus matematikus. *Magyar Tudomány,* 48.=110. évf. (2003) 4. sz.: 499–503.

Takács Ferenc: Hangstúdiótechnika. Bp.: Műegyetemi K., 2004.

Tanenbaum, Anrew S.: Számítógéphálózatok. Bp.: Panem – [London]: Prentice-Hall, 1999.

Vetier András: Szemléletes mérték- és valószínűségelmélet. Bp.: Tankönyvkiadó, 1991.