

Teorija informacija u biosistemima

Dragana Bajić

Departman za Energetiku, elektroniku i telekomunikacije

Fakultet tehnički nauka

Srbija

dragana.bajic@gmail.com

Sažetak— Teorija informacija predstavlja teorijski osnov prenosa i zapisa prevashodno diskretnih podataka. Njen naziv i sadržaj su formalizovani sredinom XX veka. Međutim, eonima pre toga, priroda je implementirala koncepte teorije informacija za prenos i zapis podataka u prostoru i vremenu. Za prenos podataka unutar organizma priroda koristi digitalnu frekvencijsku modulaciju, a sami signali su komprimovani. Za prenos i zapis na makro planu, priroda koristi četvoronivoski zapis, uz implementaciju zaštitnog kodovanja, komprimovanja, ispravljanja grešaka, kriptografije i sinhronizacije. Ovaj rad daje pregled tehnika koje priroda koristi, sa posebnim osvrtom na sinhronizaciju.

Ključne riječi- teorija informacija, DNK, nervni akcioni potencijal (information theory, DNK, nerve action potential)

I. UVOD

Teorija informacija kao teorijski osnov prenosa i zapisa podataka utemeljena je objavljivanjem klasičnih radova Kloda E. Šenona. Prvi, iz 1948, uveo je matematičke koncepte teorije informacija [1], a sam termin "teorija informacija" (TI) doživeo je svoju premijeru godinu dana kasnije [2]. Svoj proboj TI zahvaljuje digitalizaciji i diskretnim signalima, zasnovanim na još jednom Šenonovom radu - teoremi o odmeravanju (odabiranju, semplovanju) (1949, [3]). Doduše, teorema je bivala izvođena i ranije. Između ostalog, formalno prvenstvo priznato je Vladimiru Kotelnjnikovu 1999. godine, dodelom nagrade "for the first theoretically exact formulation of the sampling theorem" za rad [4], objavljen 16 godina pre Šenonovog.

Međutim, isključivo Šenonovi radovi su inicirali svetsku digitalizaciju, a u teoriju informacija uključili segmente zaštitnog kodovanja [5], kriptografije [6] i komprimovanja [1], popularizovanog Hafmenovim kodom [7]. Digitalizacija je postala sveprisutna a njene mogućnosti neizmerne – od aplikacija u multimedijalnim podacima koji prožimaju govor, muziku, zvukove, slike, filmove i animiranu višedimenzionalnu stvarnost, do sofisticirane obrade medicinskih signala i slika koja omogućavaju preciznu dijagnostiku.

Digitalna era implicira da je prethodna bila analogna. I zaista, telefonija, televizija, radio, fotografski material, gramofonske ploče, video i audio magnetne trake i kasete samo su neki od prethodnih - analognih - sistema za prenos za zapis. Međutim, analogni sistemi nisu postojali oduvek. Prvi

fotografski negativ je iz 1839, telefonski razgovor 1876, EKG je zapisan 1887 a iste godine je osmišljen i "fonograf" (preteča gramofona), rentgenska slika 1895, prva govorna radio poruka 1897, sve ostalo je imalo svoj početak (i kraj) u XX veku.

Analogna era prenosa i zapisa trajala je manje od dva veka i predstavlja singularitet. Naime, i pre Šenona postojali su diskretni prenosi i zapisi, od binarnog prenosa Morzeovim simbolima i bljeskanja ogledalom sa brda na brdo, do zapisa simbolima koji pripadaju ograničenom skupu, a koja uobičajeno nazivamo "slova". Čak su i piktografski zapisi diskretni, samo je broj simbola nešto veći. Od ostalih segmenata, kriptografija je postojala i pre nego što je Šenon uklopio klasične šifre u okvire teorije informacija. Komprimovanje izbacivanjem samoglasnika, praktično zasnovano na (Šenonovoj) entropiji, koristili su narodi u vremena kada su medijumi za zapis bili skupi. Morzeov kod je koristi iste principe kao i Hafmenov, samo što je realizovan intuitivno a ne po fiksnim pravilima, i nije prefiksna jer postoji granica između (binarnih) simbola.

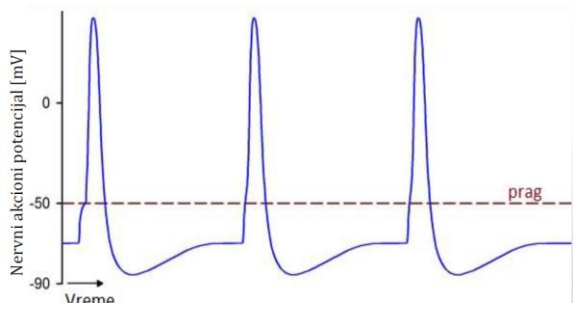
Diskretni sistemi prenosa i zapisa iz vremena pre Šenona imaju zajedničku osobinu: osmislili su ih ljudi. Međutim, eonima pre formalizovanja teorije informacija, dok je život na Zemlji nastajao, priroda je koncepte TI ugradila u svoje sisteme za prenos podataka u prostoru i vremenu.

Tema ovog rada je pregled metoda teorije informacija koje imaju svoje analogije u prenosu ili zapisu informacija DNK strukturama i neuralnim vlaknima. Neke od tih analogija u nisu uočene u dostupnoj literaturi. Posebna pažnja posvećena je sinhronizaciji, to jest procedurama koje omogućavaju da se u nizu diskretnih informacija pronađe specifična struktura.

II. OŽIČENI DIGITALNI PRENOSNI SISTEM

Mehanicistički pristup funkcionalnosti ljudskog tela svodi ga na mrežu senzora i aktuatora kontrolisanu centralnim procesorom. Senzori su čula, aktuatori mišići, centralni procesor je mozak, a nervna vlakna prenosni medijum. To je ekstremno pojednostavljen pristup, jer informacije se prenose i renalnim sistemom, a i hormoni imaju svoju ulogu. Međutim, dominantni prenosni put su neuroni a nosioci informacija su impulsi nervnog akcionog potencijala (NAP, Slika 1). Intenzitet informacije nije srazmeran amplitudi, već broju simbola u jedinici vremena, odnosno njihovoj učestanosti. Jači signal – više impulsa, slabiji signal – manje impulsa, što znači da je za prenos informacija

nervnim vlaknima odgovorna impulsna modulacija, i to frekvencijska.



Slika 1. Niz impulsâ nervnog akcionog potencijala

Digitalna frekvencijska modulacija nije jedina analogija. Priroda je prenos poboljšala primenom destruktivnog komprimovanja. Veza između pobude i frekvencije impulsâ nije linearna, već logaritamska, odnosno veza između pobude S i učestanosti impulsâ f iskazana je Weber–Fechnerovim zakonom po kojem je $f \approx a \cdot \log(S) + b$ [8]. Logaritamska transformacija smanjuje amplitudski opseg i omogućava prenos u realnom vremenu. Naime, interval između impulsa je fiziološki ograničen (kod čoveka oko 5ms) a najjača pobuda kojoj takav interval odgovara tipično je 10^6 puta veća od najmanje. Da je veza linearna, interval koji odgovara najmanjoj pobudi bio bi 10^6 puta duži, pa bi vreme potrebno da se oseti na primer pero koje padne na ruku iznosilo više od jednog sata. Cena koja se plaća za efikasniji prenos je niska rezolucija: za razlikovanje dve različite pobude potrebna je veća promena njihovog intenziteta.

III. ZAPIS I PRENOS U PROSTORU I VREMENU

Malo je poznato da je Šenonova doktorska disertacija posvećena sasvim drugoj temi: genetici [9]. Bilo je to nekoliko godina pre otkrića DNK, dok je još dominirala klasična Mendelova teorija o nasleđivanju. Šenon je izveo opštu formulu za raspodelu nekoliko povezanih posmatranih karakterističnih osobina jedinki posle određenog broja generacija, pod pretpostavkom da je sparivanje jedinki slučajno. Formula, iako originalna i inovativna u vreme kada je branio doktorat (1940. godine), prošla je nezapaženo, možda zato što nikada nije objavljena, a Šenon je promenio oblast interesovanja.

Otkriće strukture DNK 1953. i dešifrovanje genetskog koda 1961-1969 donelo je nove, izazovne signale. DNK molekul predstavlja nosioca informacije o genetskom zapisu organizma – praktično, memorijski blok. Radi se o veoma dugačkom nizu sastavljenom od više miliona konstitutivnih elemenata – nukleotida. Postoje samo četiri različita nukleotida, označena sa A, G, T, C. [10] Sa tačke gledišta teorije informacija, radi se o četvorosimbolnom alafabetu. Informacija se prenosi redosledom nukleotida a kompleksnost postiže njihovim velikim brojem. DNK prenosi informacije u prostoru, kretanjem nosioca, i u vremenu, nasleđivanjem. DNK zapisi su trajni – pročitan je DNK zapis sibirskog mamuta koji je živio pre 1 200 000 godina.

A. Signali zasnovani na DNK

DNK molekul sastoji se od dva lanca nukleotida (dva stranda). Paralelni nukleotidi međusobno su povezani po šablona A-T i G-C tako da su strandovi komplementarni. U ljudskom telu ima 10^9 nukleotida (diskretnih nosilaca informacija), a ukupna dužina DNK molekula ravna je 400 rastojanja od Zemlje do Sunca.

Tripleti nukleotida nazivaju se kodoni. Od 64 moguća tripleta preslikavaju se u 20 amino-kiselina prema Tabeli I. Niz amino-kiselina čini protein, ključni faktor svih procesa u živim bićima. Informacije o svakom proteinu zapisane su u segmentima DNK molekula. Ti segmenti predstavljaju gen za dati protein. [10].

Tabela 1. Preslikavanje tripleta nukleotida (korona) u amino-kiseline. Označeni su start i stop kodoni.

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	STOP	UGA	STOP
UUG	Leu	UCG	Ser	UAG	STOP	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

B. Kriptografija

Šifre sa asimetričnim (javnim) ključem [11] smatraju se za jedan on najznačajnijih prodora u istoriji svetske kriptografije. [12]. Zasnovane su na jednosmernim funkcijama, odnosno funkcijama koje se lako računaju u jednom, a teško u drugom smeru. Množenje dva broja je lako (jednostavan primer 4×6), ali na osnovu rezultata (24) ne može se jednoznačno utvrditi koji je faktor prvi, a koji drugi jer ima 8 mogućnosti.

Priroda na istom principu štiti informacije o proteinima kao ključnim faktorima svojih procesa. Preslikavanje kodona u amino-kiselinu je lako, ali preslikavanje unazad nije, jer se ista amino-kiselina preslikava iz različitih kodona. Budući da prosečan protein može da sadrži mnogo više od 100 amino-kiselina, a svaka se generiše na dva do (najčešće) četiri načina, u proseku postoji više od 10^8 kombinacija kodona koji generišu protein. Iako je takva zaštita za današnje računare trivijalna (osmocifrena lozinka), treba imati u vidu da priroda pre više miliona godina nije računala na kompjuterizovane neprijatelje.

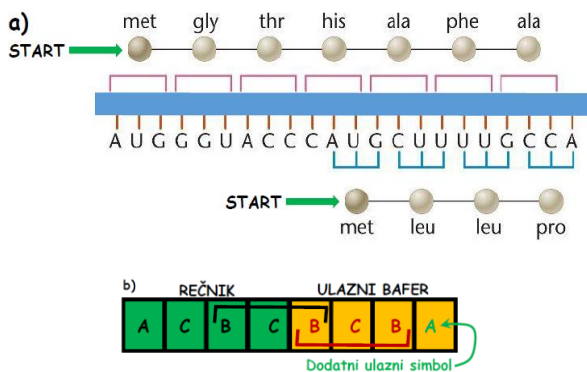
Ovaj rad je delimično finansiran grantom 451-03-68/2023-14/200156 (exTR32040) Ministarstva prosvete, nauke i tehnološki razvoj Republike Srbije.

C. Komprimovanje

Iako je broj kvazi-binarnih nukleotida koji definišu genom imponzantan (za ljude, kao što je rečeno, 10^9), priroda je primenila nedestruktivno komprimovanje tako što je pojedine gene "preklopila", odnosno iste nukleotide DNK lanca iskoristila za kodovanje različitih gena. To je karakteristika pre svega virusa jer je potrebno spakovati mnogo podataka u malom genomu, ali i u ljudskom genomu je pronađeno 774 para preklapajućih gena [10].

Na Slici 2a prikazana su dva početka dva različita gena u istom DNK lancu [10]. Iako isti nukleotidi ulaze u sastav oba gena, kodoni nisu isti. Kodon se sastoji od tri nukleotida tako da ima tri mogućnosti za njegov početak, pa je prvi nukleotid kodona u "donjem" genu pomećen za jedan nukleotid u odnosu na kodone u "gornjem" genu.

Isti model predstavlja jedan od principa primenjenih u LZ77 kodu [13]: a primer je prikazan na Slici 2b. Nekodovana informacija BCB iz ulaznog bafera koduje se na osnovu informacija koje su kodovane i nalaze se u rečniku, i na osnovu nekodovanog sadržaja ulaznog bafera, to jest sebe same. Drugim rečima, kodovani i nekodovani sadržaj se preklapaju i tako omogućavaju veću uštedu (Slika 2b).



Slika 2. Nedestruktivno kodovanje. a) ušteda na osnovu preklapanja dva gena. b) Ušteda LZ77 postupka preklapanjem kodovanog i nekodovanog sadržaja (dodatni ulazni simbol nije od značaja za preklapanje).

D. Kontrola grešaka

Prvi nivo zaštite od grešaka u DNK zapisu je prosto ponavljanje poruka: postoje dva stranda sa komplementarnim zapisima, što omogućava da se detektuju nepravilnosti.

Međutim, prave greške nastaju tokom procesa ćelijske deobe, što podrazumeva repliciranje svih DNK zapisa koji se u ćeliji nalaze. U procesu replikacije DNK strandovi se razdvoje, a nad svakim počinje da deluje, nukleotid po nukleotid, encim DNK polimeraza i svakom nukleotidu generiše njegov komplemetn. Tada su moguće greške supstitucije, kada se generiše pogrešan nukleotid, kao i greške ubacivanja i brisanja, što je u telekomunikacijama ekvivalentno pogrešno primljenom bitu i "cycle-slip-u" u telekomunikacijama. DNK polimeraza pogreši jednom u 100000 nukleotida. Znajući da jedna ćelija može da sadrži i do 6 milijardi parova nukleotida, tokom samo jedne ćelijske deobe generiše se i do 120000 grešaka [10].

Greške ispravlja sama DNK polimeraza u proceduri koja se naziva "proofreading" a koja greške prepoznaje i otklanja na osnovu informacione suvišnosti u hemijskim spojevima. Tako se ukloni oko 99% grešaka, a zaostale greške podležu drugom nivou ispravljanja grešaka, "mismatch repair", gde više različitih enzima sagledava celokupnu DNK strukturu i zamenjuje pogrešni nukleotid ispravnim. Zaostala greška nakon čitave procedure je 10^{-9} [10].

Za poređenje, verovatnoća greške od 10^{-5} u SDH transportnim sistemima se smatra za prekid veze [14].

Pokušaji da se u proceduri prepoznaju zaštitni kodovi iz teorije informacija, izuzev prostog ponavljanja, za sada su bezuspešni.

Na nivou amino-kiselina, dodatnu zaštitu od grešaka pruža jednosmerna funkcija prevođenja kodona u amino-kiseline. Na primer, amino-kiselina alanin se koduje i sa GCU, i sa GCC, i sa GCA, i sa GCG. Greška (konkretno pod nazivom "tautomerska") će prevesti A u C, tj. GCA će postati GCC. Međutim, to ništa ne menja jer amino-kiselina ostaje ista.

E. Sinhronizacija

Problem sistema koji prenose diskretne informacije je nalaženje početka.

Kao što je već rečeno, segmentacija DNK lanca na kodone može da se radi na tri načina, jer kodon sadrži tri nukleotida.

Proteini imaju oznake za početak i kraj, slične start-stop sekvencama pri asinhronom prenosu na sloju podataka OSI, odnosno TCP/IP sistema. Iz Tabele 1 se vidi da postoje tri stop kodona čija je to jedina funkcija, ali da ulogu start sekvence igra amino-kiselina "methionine".

Informacije o proteinu su zapisane u DNK lancu, i ključno pitanje je kako u čitavom dugačkom lancu naći početak upravo onog gena koji koduje dati protein, a da se pri tome ne troši energija na razmotavanje i potonje uparivanje čitavog dugačkog lanca.

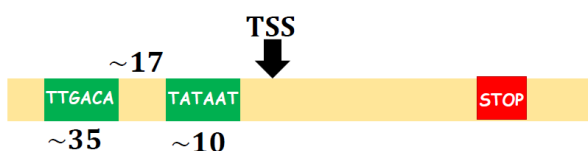
Odgovor je u procesu "transkripcije" i encimu RNK polimerazi koji "klizi" duž lanca slično procesu akvizicije u sinhronim transportnim sistemima. Kada encim dođe na određenu lokaciju na strandu, čvrsto se veže i na osnovu DNK šablona (gena) uz koji se vezao generiše mRNK (messenger RNK). mRNK direktno upravlja sintezom lanca proteina u procesu koji se naziva translacija.

Ključno pitanje procesa transkripcije svodi se na "Kako otkriti početak", odnosno kako RNK polimeraza "zna" gde treba da se veže. Odgovor je svojevrsna sinhronizacija po genu, uz pomoć stop kodona.

Zapis gena na lancu DNK sastoji od segmenta za kodovanje – to je šablon na osnovu koga se formira mRNK i koji, po konvenciji, počinje na poziciji +1 ("transcription start site" – TTS), – i od segmenta koji se ne koristi za kodovanje. Takođe po konvenciji, prvi segment se označava pozitivnim brojevima počev od +1 a drugi negativnim. Jedan od zadataka ovog drugog segmenta koji se naziva "promoter" region jeste da pruži informaciju o početku, to jest o TTS [10].

Ovaj rad je delimično finansiran grantom 451-03-68/2023-14/200156 (exTR32040) Ministarstva prosvete, nauke i tehnološki razvoj Republike Srbije.

Deo RNK polimeraze ('sigma' deo) nalazi karakteristične delove u promotor regionu i na osnovu njih pronalazi početak. Promoteri kod prokariota nisu međusobno identični, ali imaju nekoliko karakterističnih osobina. Radi se o dve šestosimbolne sekvence (Slika 3) označene sa -10 i -35. U idealnom slučaju, sekvenca -10 ima oblik TATAAT a sekvenca -35 TTGACA. Rastojanje između njih je 17 nukleotida ali može da varira. Prema [15], u 92% slučajeva rastojanje je 17 ± 1 . Rastojanje je dovoljno veliko i sekvence mogu da se razmatraju svaka za sebe, odnosno nisu deo jedne distribuirane sekvence [16]. TTS se u tri četvrtine slučajeva nalazi na rastojanju 7 ± 1 od -10 sekvence. Realno, u proseku tri od šest simbola mogu da budu izmenjeni - radi se o 'koncenzus' sekvencama, gde ne moraju svi simboli da budu ispravni da bi se sekvenca prepoznala. U Tabeli 2 izlistane su verovatnoće pojavljivanja simbola pojedinih sekvenci. I karakteristične su za bakteriju *Escherichia coli* na kojoj je čitava priča osinhronizaciji i započeta [17,18].



Slika 3. Sinhronizacija kod prokariota: "start" sekvence (promoteri) i "stop" sekvence

Tabela 2. Verovatnoća pojavljivanja nukleotida promotor regiona [%]

T	A	T	A	A	T
77%	76%	60%	61%	56%	82%
T	T	G	A	C	A
69%	79%	61%	56%	54%	54%

Ponekad se javlja i treća sekvenca, zvana UP element, AAAWWTWTTTNNNAANNN gde W označava A ili T a N bilo koji nukleotid, a koja je centrirana oko pozicije -50.

Završetak šablona se označava stop sekvencom iz Tabele 1. Ti kodoni se po pravilu ne pojavljuju u regionu za kodovanje (u tu svrhu se u ramovima na drugom OSI nivou koriste bit stuffing i byte stuffing procedure).

Iz prethodnog primera može se zaključiti da jednostavno svođenje sinhronizacije u procesu ekspresije gena na telekomunikacioni problem nije u potpunosti opravdano. Postoji niz nedoumica u vezi sa verovatnoćama iz Tabele 2, kao i u vezi sa samim procesom sinhronizacije. Analiza sekvenci ne daje odgovor kako RNK polimeraza 'zna' kada treba da započne transkripciju, ako analize kažu da transkripcija započinje u 75% slučajeva na rastojanju 6, 7 ili 8 od sinhronišuće sekvence u slučaju *E. coli* bakterije koja predstavlja najjednostavniji primer.

Sa druge strane, postavlja se pitanje zašto se sinhronizacionim pravilima koja su značajno relaksiranija u

poređenju sa rigidnim ITU-T preporukama već milionima godina uspešno transkribuje protein iz njihovih gena. Bifiks i kros-bifiks fazi analiza [19, 20] koja se pokazala uspešnom u telekomunikacionim sinhronizacijama, u kontekstu RNK polimeraze nije uspela da pruži nikakav odgovor.

F. Entropija

Entropija je veličina koja predstavlja osnovu teorije informacija [1]. i inspirisala je prvo povezivanje teorije informacija sa genetikom. Eksperti za kodovanje analizirali su mapiranje četvorosimbolne DNK strukture u aminokiseline (Tabela 1), pokušavajući da otkriju entropiju i eventualne zaštitne kodove. Nisu otkrili ništa i dve discipline su se privremeno razile.

Jedan od prvih 'povratnika' bio je bard evropskog klasičnog zaštitnog kodovanja Joakim Hagenauer. Pored već pomenutih analiza sinhro-sekvenci *E. coli* bakterije [17, 18] koristio je uzajamnu informaciju – razliku združene entropije i obe uzajamne entropije – kao meru rastojanja pri analizi klasifikacije genoma, kao i za mapiranje gena pojedinih bolesti.

IV. ZAKLJUČAK

Koncepti teorije informacija su ugrađeni u živi svet. Za prenos informacija u telu koristi se digitalna frekventijska modulacija, a efikasnost povećava primenom destruktivnog logaritamskog komprimovanja. Može da se povuče analogija između šifrovanja sa asimetričnim ključem i preslikavanja kodona u amino-kiseline, komprimovanja preklapanjem kodona u genima i LZ77 koda, start i stop kodona i formiranja rama na sloju podataka OSI referentnog nivoa.

Neka rešenja su slična rešenjima teorije informacija, ali još uvek nisu do kraja protumačena – na primer, procedura ispravljanja grešaka i proces sinhronizacije. Naročito je interesatna činjenica da je, čak i u najjednostavnijim organizmima, proces sinhronizacije vrlo neodređeno definisan uz veliki broj stepena slobode a ipak funkcioniše, potpuno suprotno rešenjima sinhronizacije i u sinhronim i u asihronim sistemima. Ova rešenja prirode bi mogla da inspirišu nova istraživanja i dovedu do novih inženjerskih rešenja.

ZAHVALNICA

Istraživanje je izvršeno pod okriljem EU COST akcije CA20120 – Intelligence-Enabling Radio Communications for Seamless Inclusive Interactions (INTERACT) - VT-IoT-Health.

LITERATURA

- [1] Shannon, C.E. A mathematical theory of communication. Bell Syst. Tech. J. 1948, 27, 379–423.
- [2] Shannon, C.E. Some Topics on Information Theory. In Shannon: Collected Papers; Sloane, N.J.A., Wyner, A.D., Eds.; Wiley: Hoboken, NJ, USA, 1993; pp. 458–459.

- [3] Shannon, C.E. Communications in the presence of noise. Proc. IRE 1949, 37, 10–21.
- [4]. Kotelnikov, V.A. On the transmission capacity of "ether" and wire in electrocommunications. Izd. Red. Upr. Svyazzi RKKA (1933),
- [5]. Shannon, C.E. General treatment of the problem of coding. Trans. IRE Prof. Group Inf. Theory 1953, 1, 102–104.
- [6] Shannon, C.E. Communication theory of secrecy systems. Bell Syst. Tech. J. 1949, 28, 656–715.
- [7] Huffman, D. A Method for the Construction of Minimum-Redundancy Codes Proceedings of the IRE. 40 (9): 1098–1101. (1952).
- [8] Brown, B.H. Smallwood, R.H. Barber, D.C. Lawfor P.V. Medical Physics and Biomedical Engineering, IoP, London, 1998.
- [9] Shannon, C.E. An algebra for theoretical genetics, Ph.D. thesis, Massachusetts Institute of Technology, Dept. of Mathematics (1940).
- [10] Zvelebil, M. Baum, J.O. Understanding Bioinformatics, Garland Science, New York (2008).
- [11] Hellman, M. Merkle, R. and Diffie, W. Public key system, Stanford University, 1976.
- [12] Tanenbaum, A Computer networks, Prentice Hall, New Jersey 2003
- [13] Ziv, J. Lempel, A. A Universal Algorithm for Sequential Data Compression. IEEE Transactions on Information Theory. 23 (3): 337–343. 1977.
- [14] ITU T Recommendation G.704.
- [15] Harlez, C.B. Reynolds, R.P. Analysis of E. coli promoter sequences, Nucleic Acids Res. March 11; 15(5), pp 2343–2361 (1987).
- [16] de Lind van Wijngaarden, A. J. Willink, T. J. Frame Synchronization Using Distributed Sequences", IEEE Transactions on Communications, Vol. 48, No. 12, pp 2127–2138, (2000).
- [17] P Hanus, B. Goebel, J. Dingel, J. Weindl, J. Zech, Z. Dawy, J. Hagenauer, J. C. Mueller: Information and Communication Theory in Molecular Biology, Electrical Engineering, pp 10001-10012, (2007).
- [18] Weindl, J. Hagenauer, J. Applying Techniques from Frame Synchronization for Biological Sequence Analysis, ICC2007, (2007).
- [19] Bajić, D. Search, Sequences, Synchronization and States: a Different Approach, University of Novi Sad, 2006
- [20] Bajić, D. Stefanović, Č. Krosbifiks analiza sinhronizacije rama DNK sekvenci DOGS, Str. 132, Kelebjaja; 2008

ABSTRACT

Information theory represents the theoretical basis of the transmission and recording of primarily discrete data. Its name and content were formalized in the middle of the 20th century. However, eons before that, nature implemented the concepts of information theory to transmit and record data in space and time. For data transmission within the organism, nature uses digital frequency modulation, and the signals themselves are compressed. For transmission and recording on a macro level, nature uses quadruple recording, with the implementation of protective coding, compression, error correction, cryptography, and synchronization. This paper provides an overview of the techniques used by nature, with special reference to synchronization.

INFORMATION THEORY IN BIOSYSTEMS

Dragana Bajić