Genomic Digital Signal Processing

Parameswaran Ramachandran

and

Andreas Antoniou

Department of Electrical Engineering,
University of Victoria, BC, Canada.

Signal Processing and Genomics

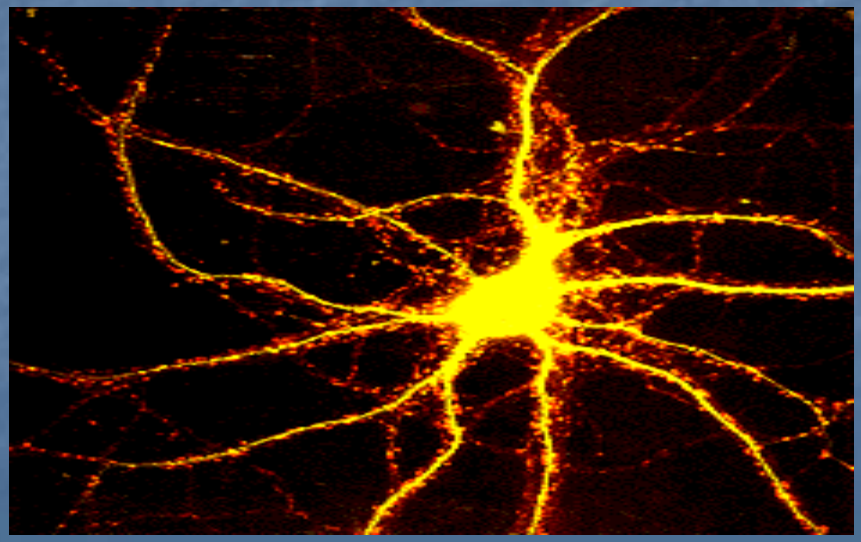
- Most signals and processes in nature are continuous. However, genomic information occurs in the form of discrete sequences.
- As will be shown, **DNA (deoxyribonucleic acid)** molecules as well as **proteins** can be represented by numerical sequences.
- **Digital signal processing (DSP)** evolved to process numerical sequences. Therefore, it provides numerous powerful and efficient tools that can be used for the analysis of genomic data.
- **Open access** to raw genomic data has spawned strong interest in exploring the application of DSP to genomics.

Basics of Molecular Biology

(see Alberts et al. [1])

- The **cell** is the fundamental unit of life. All living organisms are made of cells.
- About 75–100 trillion cells in the human body!
- Ability to replicate independently makes cells ***living***.

Red Blood Cells



A Nerve cell (neuron). The orange dots, called synapses, are the junctions through which a neuron communicates with other neurons.

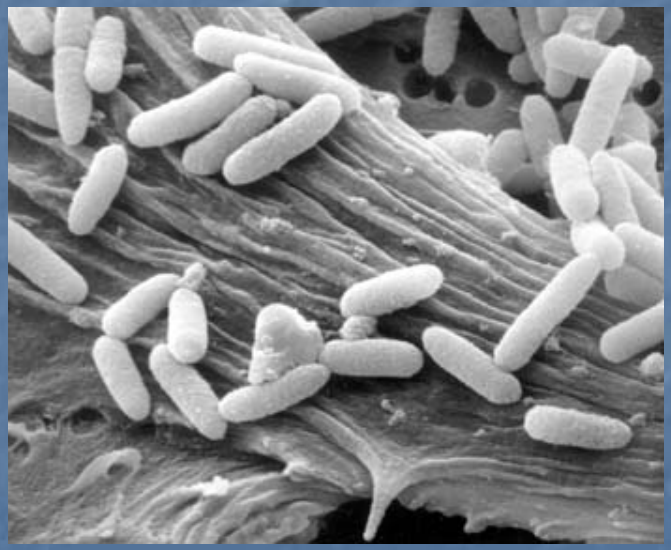
Classification of Living Organisms

- **Procaryotes**
  - **Absence** of a distinct membrane-bound **nucleus**
  - DNA is **NOT** organized into **chromosomes.**
  - Bacteria and cyanobacteria are procaryotes.

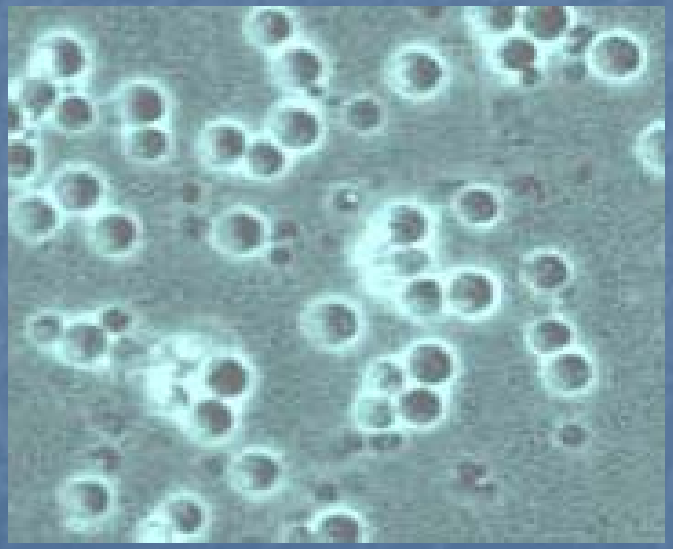
- **Eucaryotes**
  - **Presence** of a distinct membrane-bound **nucleus**
  - DNA is organized into **chromosomes.**
  - Can be **single-celled** (yeasts, amoebas) or **multicellular** (plants, animals, people)

Model Organisms

- **Very large** diversity of living organisms
- However, biologists focus their attention on a small number of representative organisms.



E. Coli Bacteria



Baker's yeast (S. Cerevisiae)

Model Organisms (cont'd)

the plant Arabidopsis

the fly Drosophila (fruit fly)

C. Elegans (roundworm)

Mouse

Human beings

Harry Nyquist (1889–1976)

The Genome

- An organism's **genome** is the blueprint for making and maintaining itself.
- Nearly all cells of an organism contain the genome. An exception is the **red blood cell** which lacks DNA.
- In procaryotes, the genome is found as a single circular piece. Eucaryotic genomes are often very long and are divided into packets called **chromosomes.**
- Different eucaryotes have different numbers of chromosomes:

Human: 46 **Mouse:** 40 **Dog:** 78 **Pepper:** 24

Coffee: 88 **Apple:** 34 **Horse:** 64 **Drosophila:** 8

The DNA

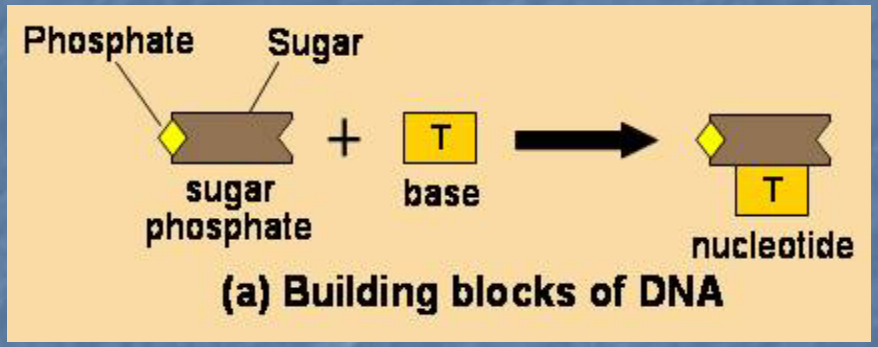
- Genomic information is encoded in the form of DNA inside the nuclei of cells.
- A DNA molecule is a long linear polymeric chain, composed of **four** types of subunits. Each subunit is called a **base.**
- The four bases in DNA are adenine **(A),** thymine **(T),** guanine **(G),** and cytosine **(C).**
- DNA occurs as a **pair** of strands. Bases pair up across the two strands. A always pairs with T and G always pairs with C. Hence, the two strands are called **complementary.**

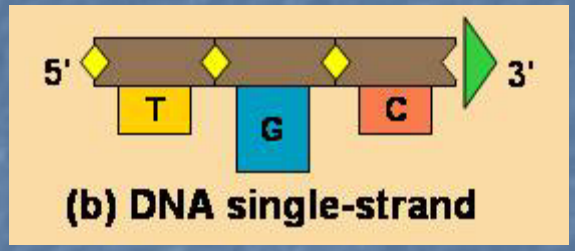
Fun fact:

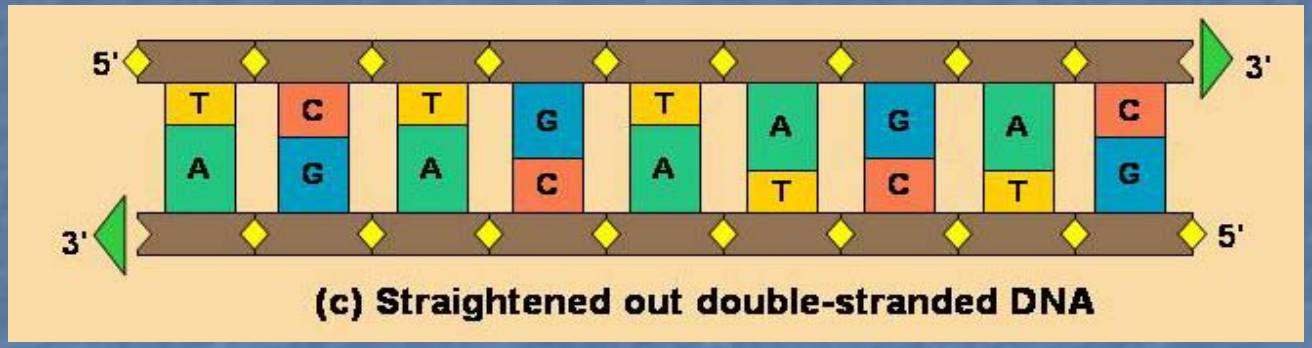
If our cells were enlarged to the size of an aspirin pill, our DNA would be about 10 kms. long!

DNA (cont'd)

DNA and its building blocks

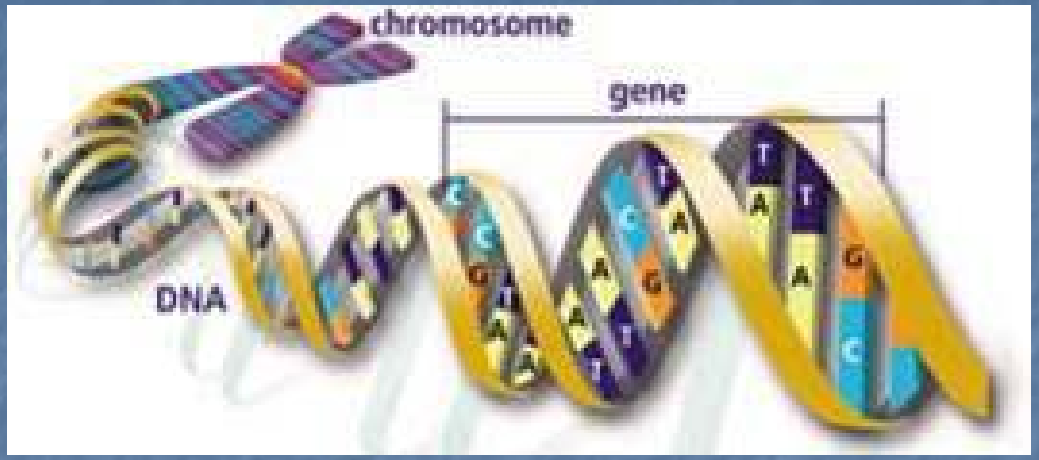






- The sugar in DNA is called **deoxyribose.**

DNA (cont'd)

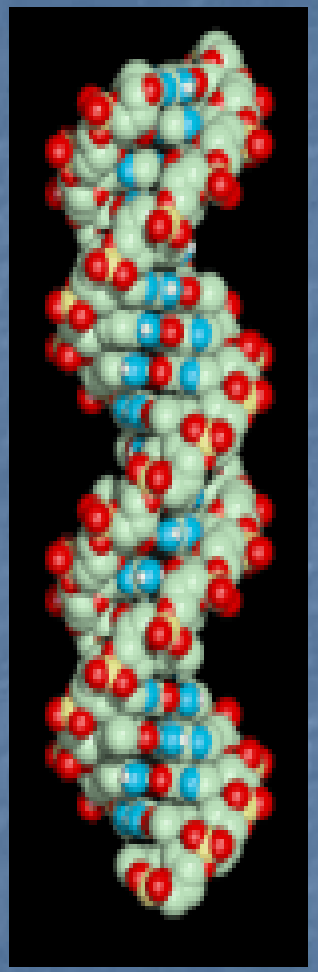


The DNA double helix structure

Image Credit: U.S. Department of Energy Human Genome Program. http://www.ornl.gov/hgmis

- The DNA double helix is **right handed.** This is the same as the thread in regular **bolts** and **screws.**

The right handed thread



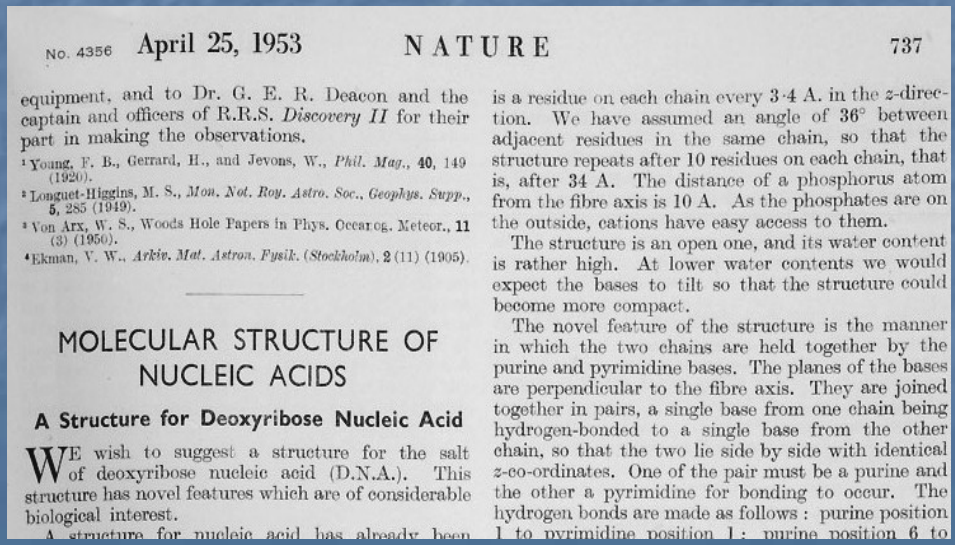
The double helix, animated!

Discovery of the DNA Double Helix

- In 1953, **James D. Watson** and **Francis Crick** proposed the double helical structure of DNA through their landmark paper in the British journal **Nature.**
- For this discovery, they shared the **1962 Nobel prize** for Physiology and Medicine with **Maurice Wilkins** who, with **Rosalind Franklin,** provided the data on which the structure was based.

Watson **Crick**

Cambridge, 1953.



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been

is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to

The landmark paper – Nature 171, 737-738 (1953).

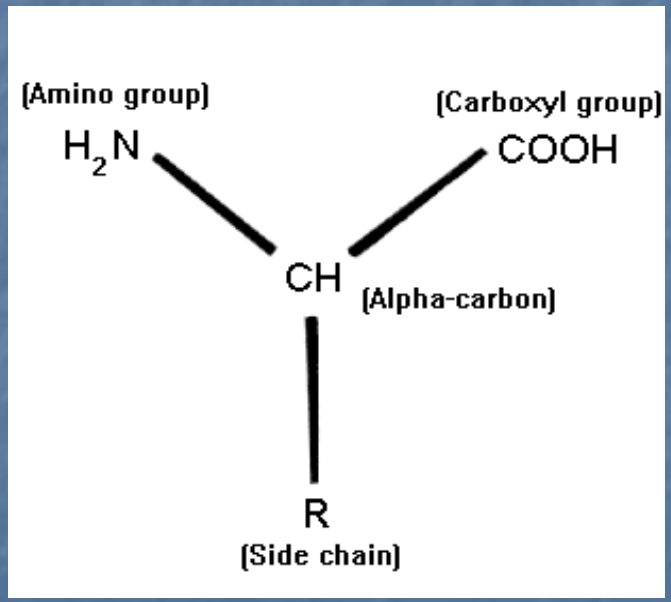
Proteins

- Proteins are the building blocks of cells. Most of the dry mass of a cell is composed of proteins.
- Proteins
  - form the structural components (e.g., skin proteins)
  - catalyze chemical reactions (e.g., enzymes)
  - transport and store materials (e.g., hemoglobin)
  - regulate cell processes (e.g., hormones)
  - protect the organism from foreign invasion (e.g., antibodies)
- Proteins are long polymers of subunits called **amino acids.**

Proteins (cont'd)

- Amino acids are molecules with a central carbon atom (called α-Carbon) attached to a carboxylic acid group, an amino group, a hydrogen atom, and a variable side chain. Only the side chains vary between amino acids.

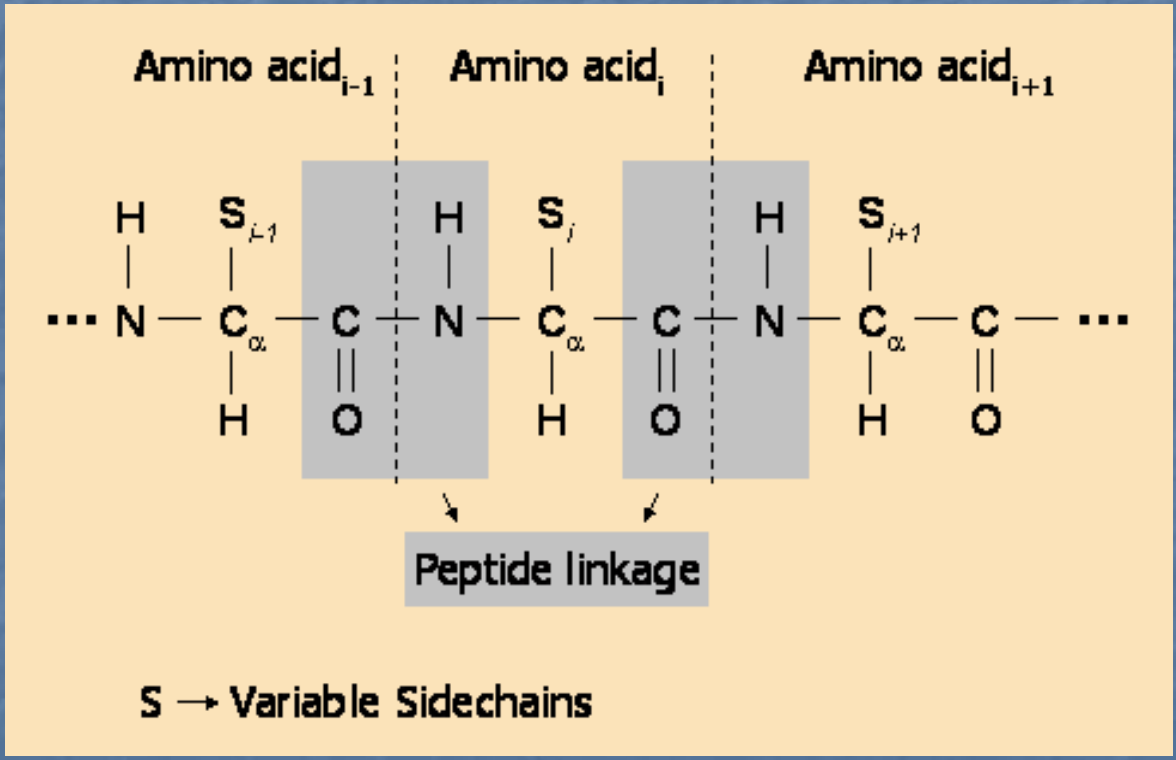
- There are 20 different amino acids that form proteins. Out of these, we need to take 9 amino acids through our diet as we cannot synthesize them. These are called **essential** amino acids.



Amino acid

Proteins (cont'd)

- Individual amino acids are linked by covalent linkages called **peptide bonds.**



Formation of a protein chain

Proteins (cont'd)

- Proteins perform their functions by folding into unique **three-dimensional (3-D)** structures.

Image Credit: Catholic University of Brussels, Biotechnology

Ribosomal subunit

Credit: Argonne Photo Gallery

Genes

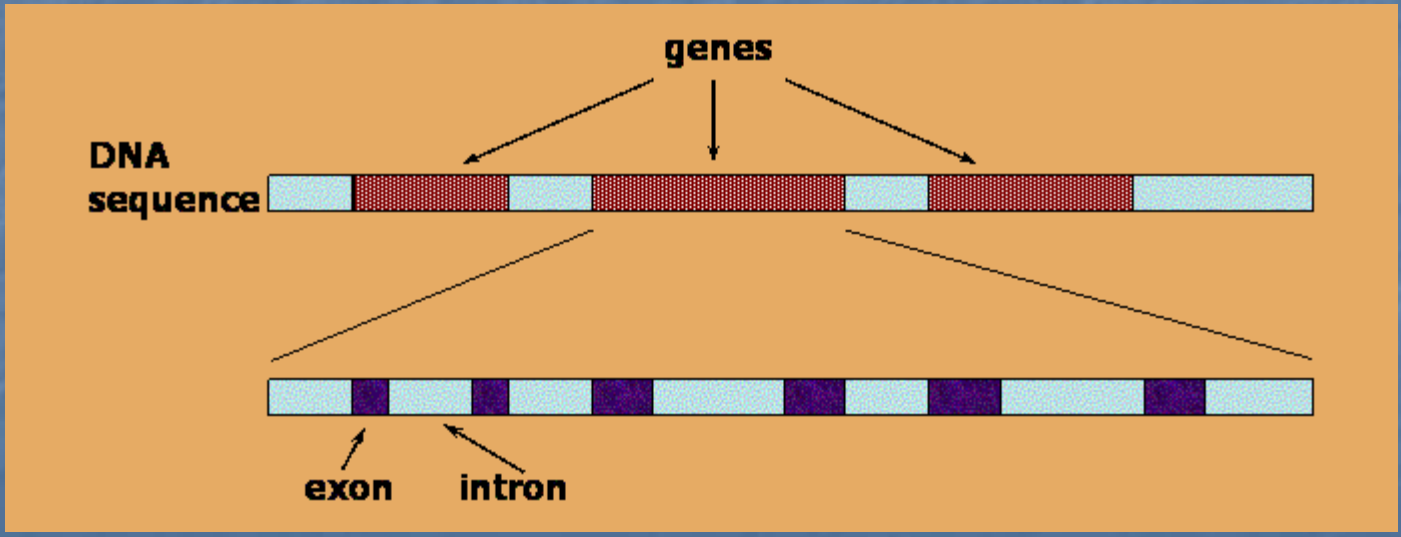
- Genes are regions in the genome that carry instructions to make proteins.
- Genes are inherited from parent to offspring, and thus are preserved across generations.
- Genes determine the traits of an organism. It is only due to genes that a **dog** always gives birth to a **dog,** and a **cat** always gives birth to a **cat!**

dog ⇨ **dog**

cat ⇨ **cat**

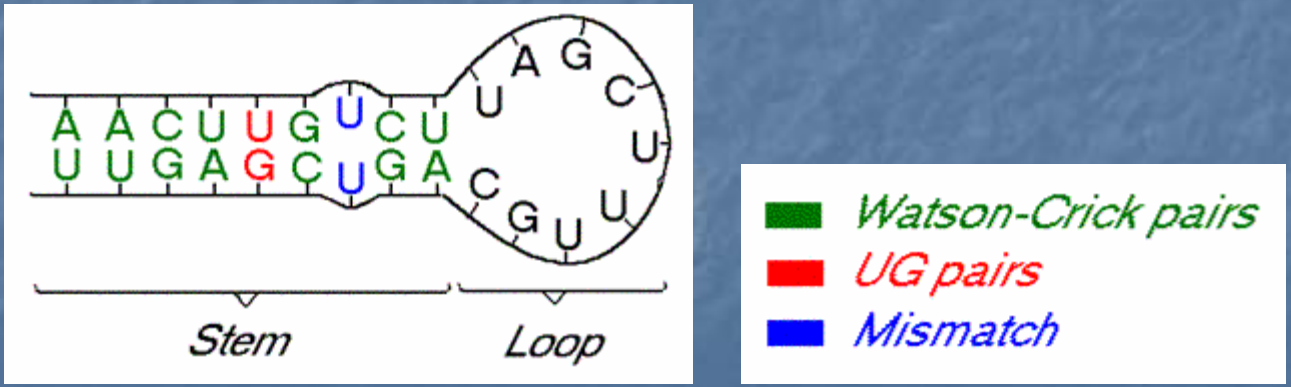
Genes (cont'd)

- Procaryotic genes mostly occur as uninterrupted stretches of DNA.
- Eucaryotic genes are mostly divided into many fragments called **exons.** The exons are separated by noncoding regions called **introns.**



The RNA

- Ribonucleic acid **(RNA)** is a chemical similar to a single strand of DNA.
- Unlike in the DNA, in RNA the sugar **ribose** occurs instead of deoxyribose and the base uracil **(U)** occurs instead of thymine.
- RNA delivers DNA's genetic message to the cytoplasm of a cell where proteins are made.



An example RNA chain. Notice the folding and looping.

The Central Dogma of Molecular Biology

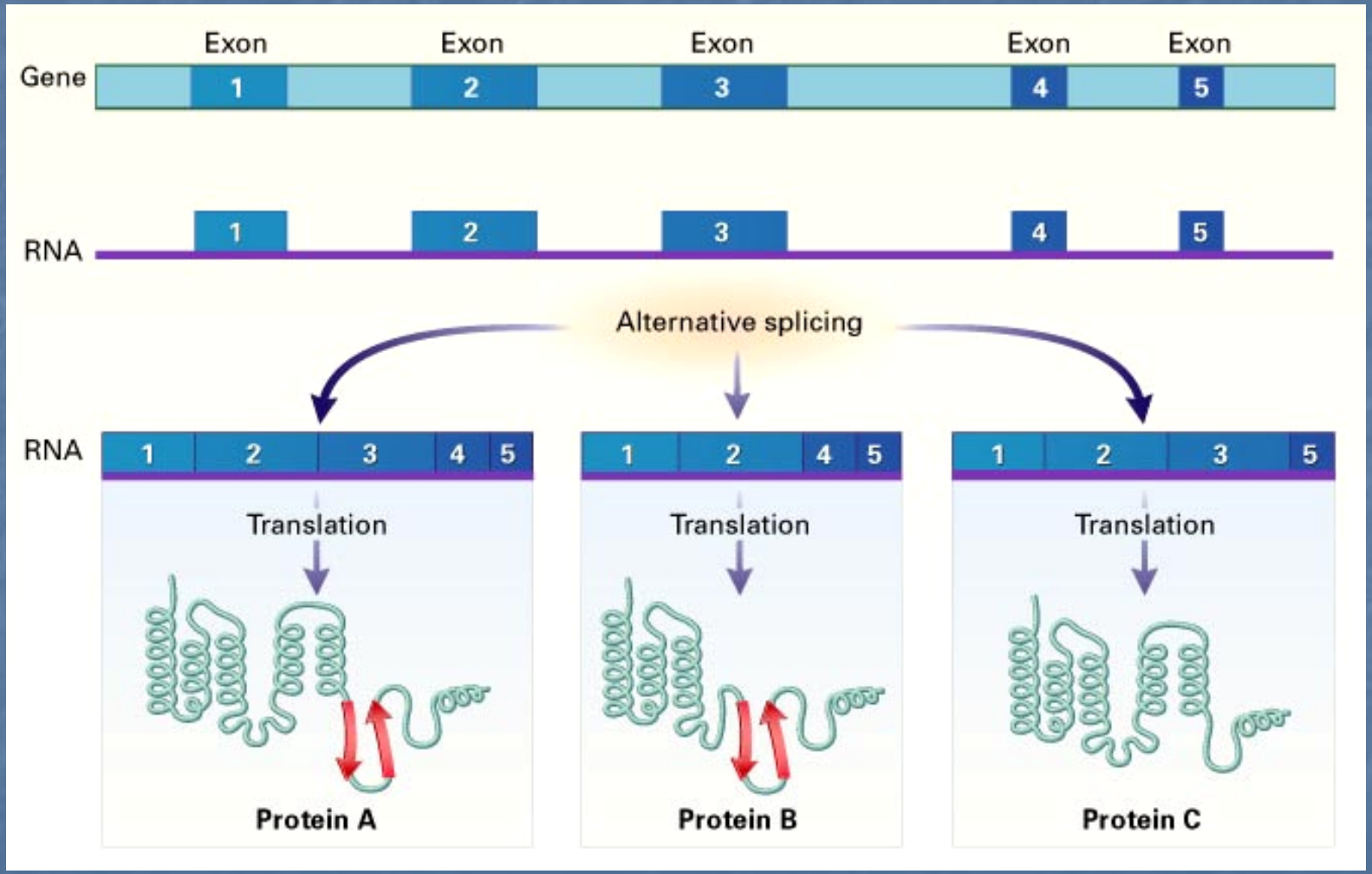
- **Transcription:** Genes are copied into RNA called messenger RNA **(mRNA).**
- **Translation:** Proteins are then made from the mRNA transcripts.
- The above two steps are fundamental to all life and hence are together called the **central dogma of molecular biology.**

The Central Dogma.

Splicing and Alternative Splicing in Eucaryotes

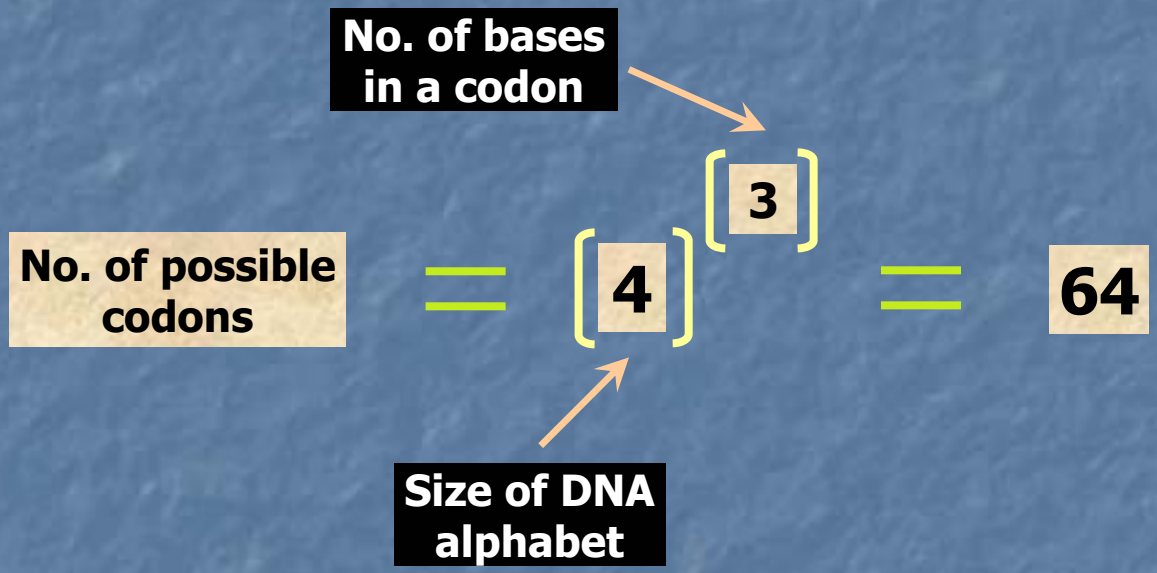
- When DNA is copied into mRNA during transcription, the introns are eliminated by a process called **splicing.**
- The same gene can code for different proteins. This happens by joining the exons of a gene in different ways. This is called **alternative splicing.**
- Alternative splicing seems to be one of the main purposes for which the genes in eucaryotes are split into exons.
- The mRNA obtained after splicing is uninterrupted and is used for making proteins.

Alternative Splicing (cont'd)



The Genetic Code

- Rule by which genes code for proteins
- Groups of three bases, **called codons,** code for the individual amino acids.



- Since the number of codons is greater than the number of amino acids, more than one codon can code for an amino acid. The genetic code is hence said to be **degenerate.**

Genetic Code (cont'd)

FIRST POSITION	SECOND POSITION OF CODON: T	C	A	G	THIRD POSITION
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)	T
	TTC Phe (F)	TCC Ser (S)	TAC Tyr (Y)	TGC Cys (C)	C
	TTA Leu (L)	TCA Ser (S)	**TAA (STOP)**	**TGA (STOP)**	A
	TTG Leu (L)	TCG Ser (S)	**TAG (STOP)**	TGG Trp (W)	G
C	CTT Leu (L)	CCT Pro (P)	CCT Pro (P)	CGT Arg (R)	T
	CTC Leu (L)	CCC Pro (P)	CCC Pro (P)	CGC Arg (R)	C
	CTA Leu (L)	CCA Pro (P)	CCA Pro (P)	CGA Arg (R)	A
	CTG Leu (L)	CCG Pro (P)	CCG Pro (P)	CGG Arg (R)	G
A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)	T
	ATC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)	C
	ATA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)	A
	ATG Met (M) **(START)**	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)	G
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)	T
	GTC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)	C
	GTA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)	A
	GTG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)	G

- ATG acts as the only **START** codon (also codes for Methionine).
- TAA, TAG, and TGA act as alternative 

codons.

Reading Frames

- A DNA strand is always read for codons in the **5'–to–3'** direction (This has to do with the asymmetrical molecular structure of the sugar molecules that make up the nucleotides, i.e., 5'-carbons at one end and 3'-carbons at the other).
- Each of the two strands can be read in three different ways depending on the starting point.
- Thus, there are six different ways in total. Each one is called a **reading frame.**

```
CGT  AGC  TTA  CTG  . . .
. CG  TAG  CTT  ACT  G . .
. . C  GTA  GCT  TAC  TG .
```

Three ways of reading a DNA strand

Open Access Databases

- **Entrez** database of the National Center for Biotechnology Information **(NCBI)** – Provides free access to whole genome sequences, protein sequences, protein 3-D structures, etc.

  Web Link: http://www.ncbi.nih.gov/Entrez

- Protein Data Bank (PDB) – A worldwide repository for the processing and distribution of 3-D structure data of proteins

  Web Link: http://www.rcsb.org/pdb

Conversion of DNA Character Strings to Numbers

- To apply DSP techniques, the DNA should be represented by numerical sequences.
- Two choices:
  - Create four binary sequences, one for each character (base), which specify whether a character is present (1) or absent (0) at a specific location

    These are known as indicator sequences (Tiwari et al. [4]).
  - Assign meaningful real or complex numbers to the four characters A, T, G, and C

    In this way, a single numerical sequence representing the entire character string is obtained.

Binary Indicator Sequences

DNA Sequence	···	A	T	T	G	C	A	C	C	G	T	G	A	···
Indicator seq. for A	···	1	0	0	0	0	1	0	0	0	0	0	1	···
Indicator seq. for T	···	0	1	1	0	0	0	0	0	0	1	0	0	···
Indicator seq. for G	···	0	0	0	1	0	0	0	0	1	0	1	0	···
Indicator seq. for C	···	0	0	0	0	1	0	1	1	0	0	0	0	···

- On adding all the four indicator sequences, we get a sequence of all 1s since a location must have one of the 4 possible characters.
- Hence, any three of the four indicator sequences completely characterize the full DNA character string.
- Indicator sequences can be analyzed to identify patterns in the structure of a DNA string.

The Discrete Fourier Transform (DFT)

- For a finite-length sequence $x[n]$ of length N, a corresponding periodic sequence $\tilde{x}[n]$ with period N can be formed as

$$\tilde{x}[n] = \sum_{r=-\infty}^{\infty} x[n + rN]$$

where $n = 0, 1, 2, \ldots, N\text{-}1$ and r is an integer.

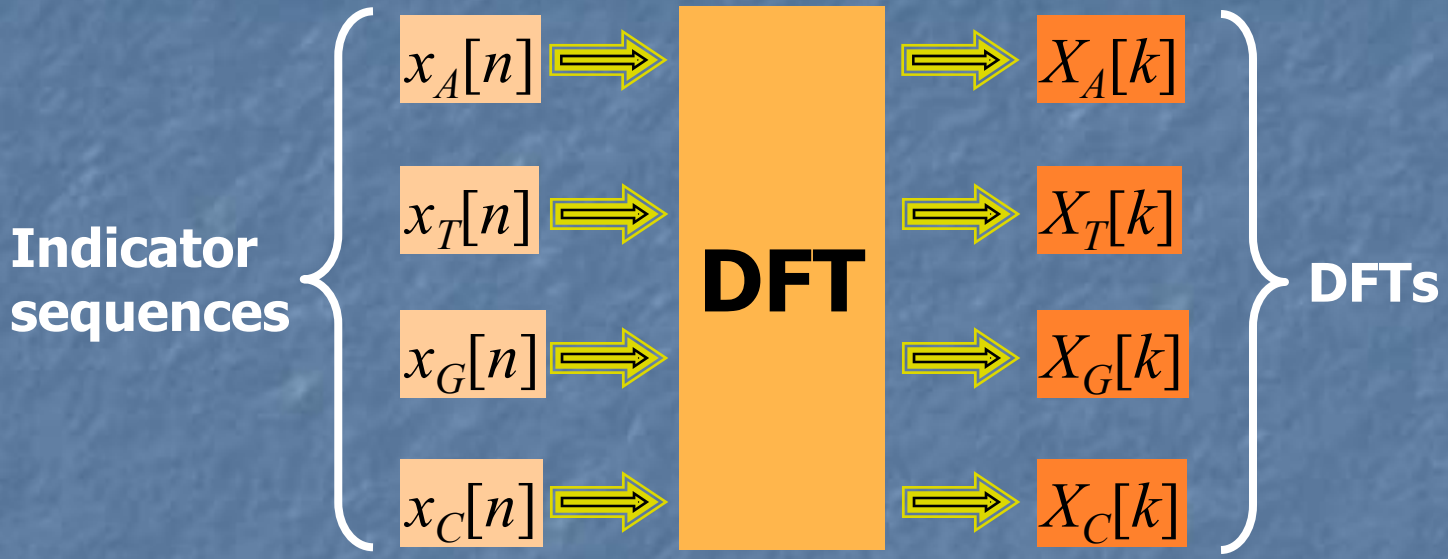
- The DFT of $\tilde{x}[n]$ is given by

$$\tilde{X}[k] = \sum_{n=0}^{N-1} \tilde{x}[n] W_N^{kn} \quad \text{where} \quad W_N = e^{-j2\pi/N}$$

Period-3 Property of Protein-Coding Regions

- Protein-coding regions of DNA have been found to have a peak at frequency $2\pi/3$ in their Fourier spectra. This is called the period-3 property (see Tiwari et al. [4]).
- The period-3 property is related to the different statistical distributions of codons between protein-coding and noncoding DNA sections.
- The period-3 property can be used as a basis for identifying the coding and non-coding regions in a DNA sequence, as will be shown later.

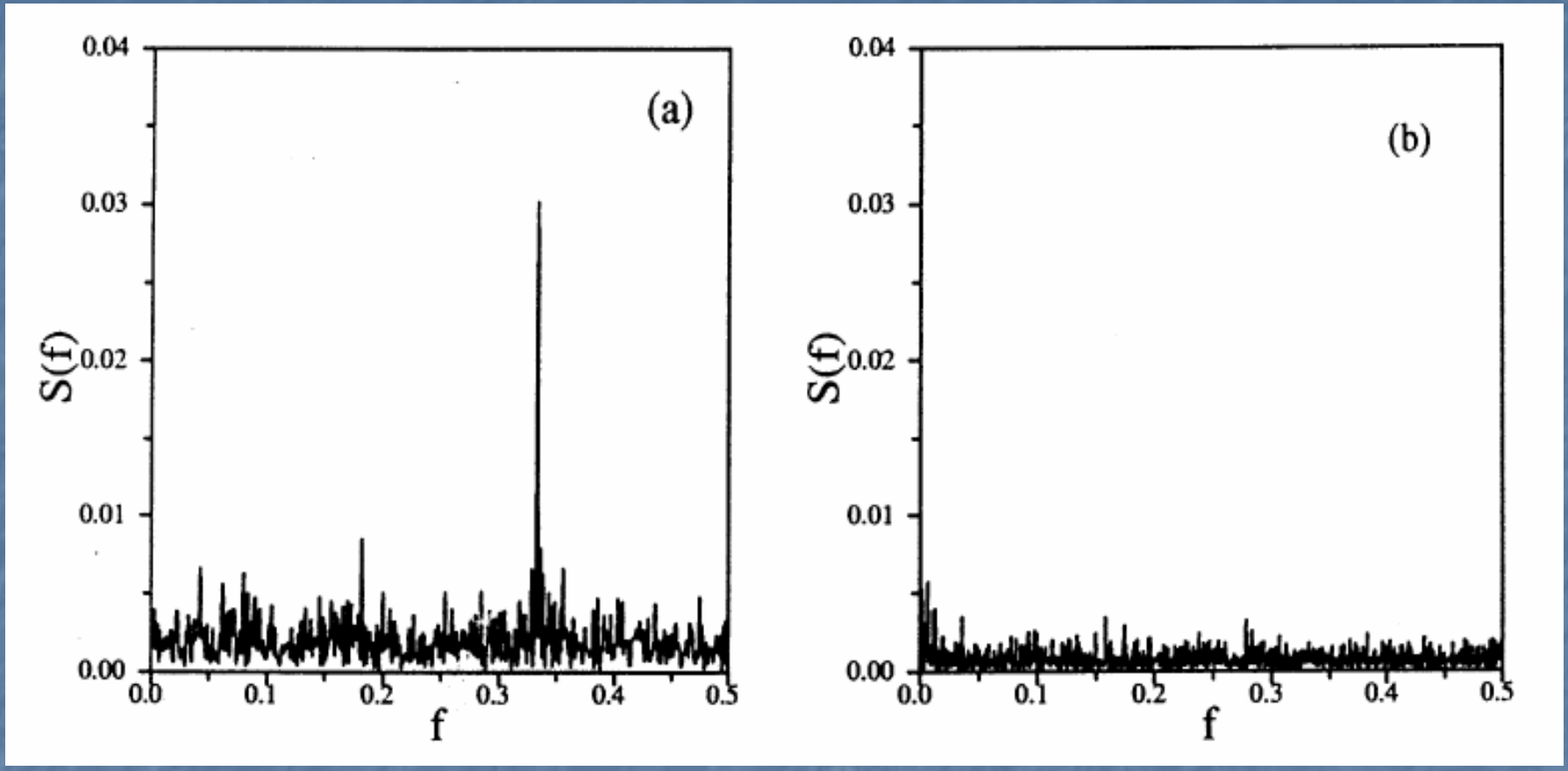
Period-3 Property (cont'd)



$$PSD = |X_A[k]|^2 + |X_T[k]|^2 + |X_G[k]|^2 + |X_C[k]|^2$$

- PSD of protein-coding regions show a peak at $2\pi/3$.

Period-3 Property (cont'd)



PSD of a protein-coding region

PSD of a non-coding region

Tiwari et al., *CABIOS*, vol. 13, no. 3, 1997.

Geometric Representations

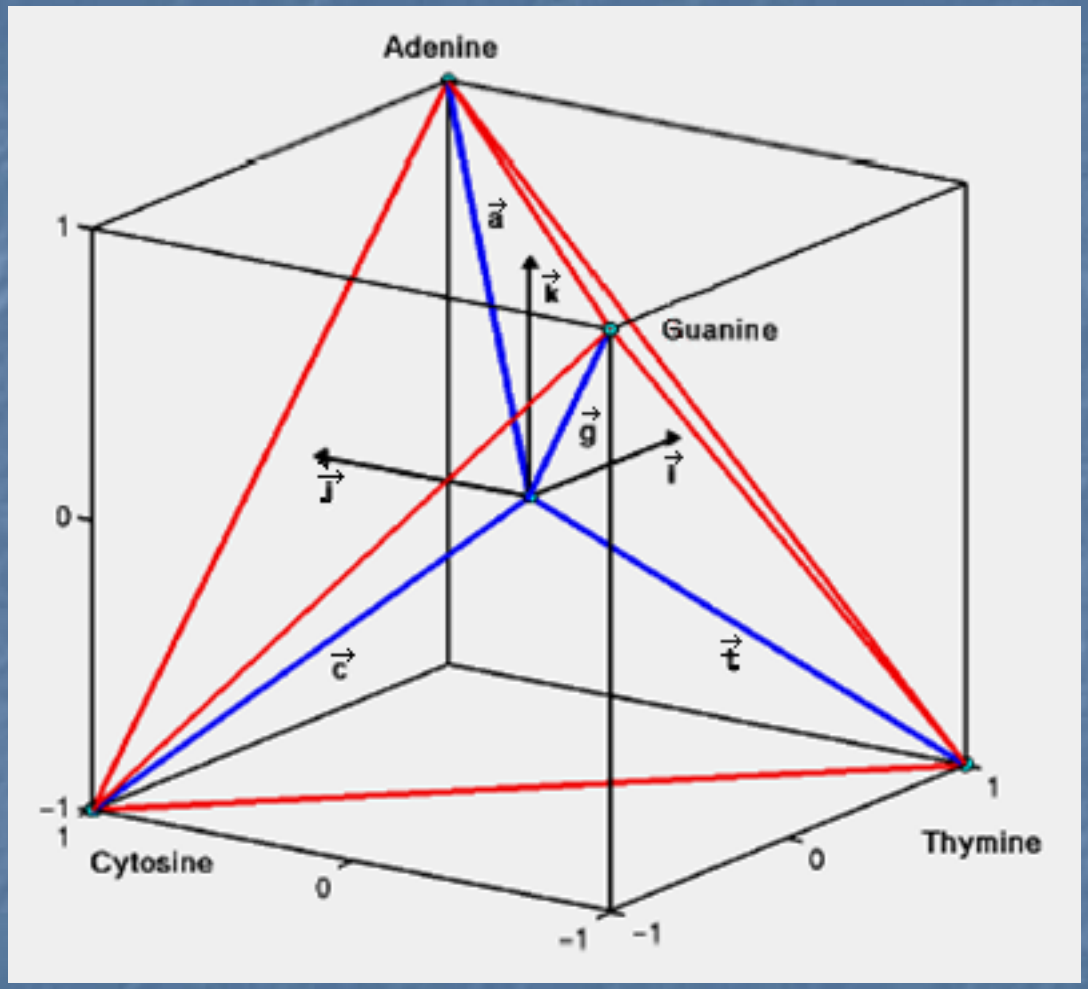
- Numbers can be assigned to the DNA bases through geometric representations.
- One possibility (see Cristea [3]) is to assign to the 4 bases the 4 vectors from the center to the vertices of a **regular tetrahedron**.
- The vertices of a regular tetrahedron form a subset of the vertices of a **cube.** Hence, the 4 vectors point towards alternate cube vertices.
- **Assumptions:**

  Cube side length: 2 units. **Origin:** Cube center.

  Hence, the co-ordinates of the vertices of the cube are

  $(\pm 1, \pm 1, \pm 1)$

Geometric Representations (cont'd)



The tetrahedral representation

The base vectors are

$$\vec{a} = \vec{i} + \vec{j} + \vec{k}$$

$$\vec{t} = \vec{i} - \vec{j} - \vec{k}$$

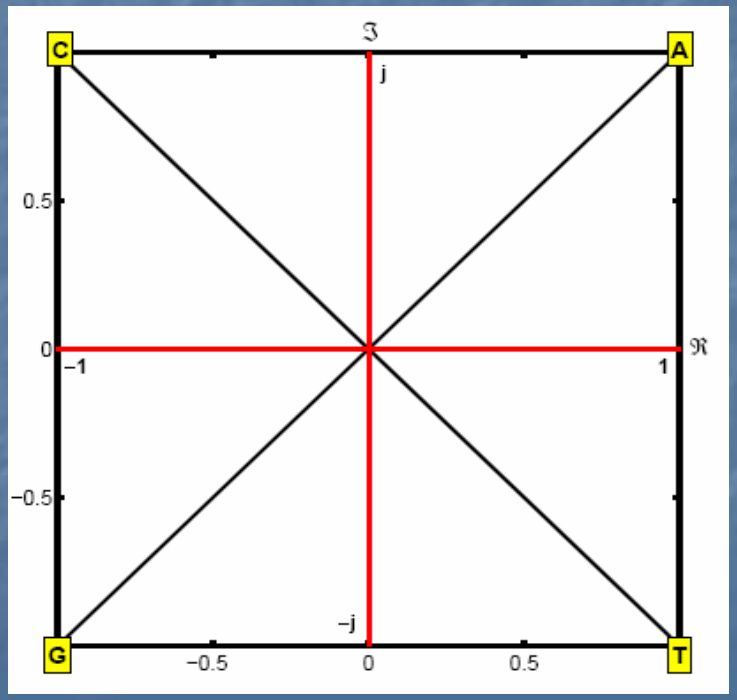
$$\vec{g} = -\vec{i} - \vec{j} + \vec{k}$$

$$\vec{c} = -\vec{i} + \vec{j} - \vec{k}$$

Figure reproduced from P. Cristea, *ELSEVIER Signal Processing*, vol. 83, no. 4, 2003.

Geometric Representations (cont'd)

- The dimensionality of the tetrahedral representation can be reduced to two by projecting the tetrahedron onto a suitable plane.
- Many projection planes can be obtained. The simplest choice is defined by a pair of co-ordinate axes.



Projection onto the $\vec{i} - \vec{j}$ plane

Mapping the $\vec{i}$ axis to the Real axis and the $\vec{j}$ axis to the Imaginary axis, we get

$$a = 1 + j$$
$$t = 1 - j$$

Complex Conjugates

$$g = -1 - j$$
$$c = -1 + j$$

Complex Conjugates

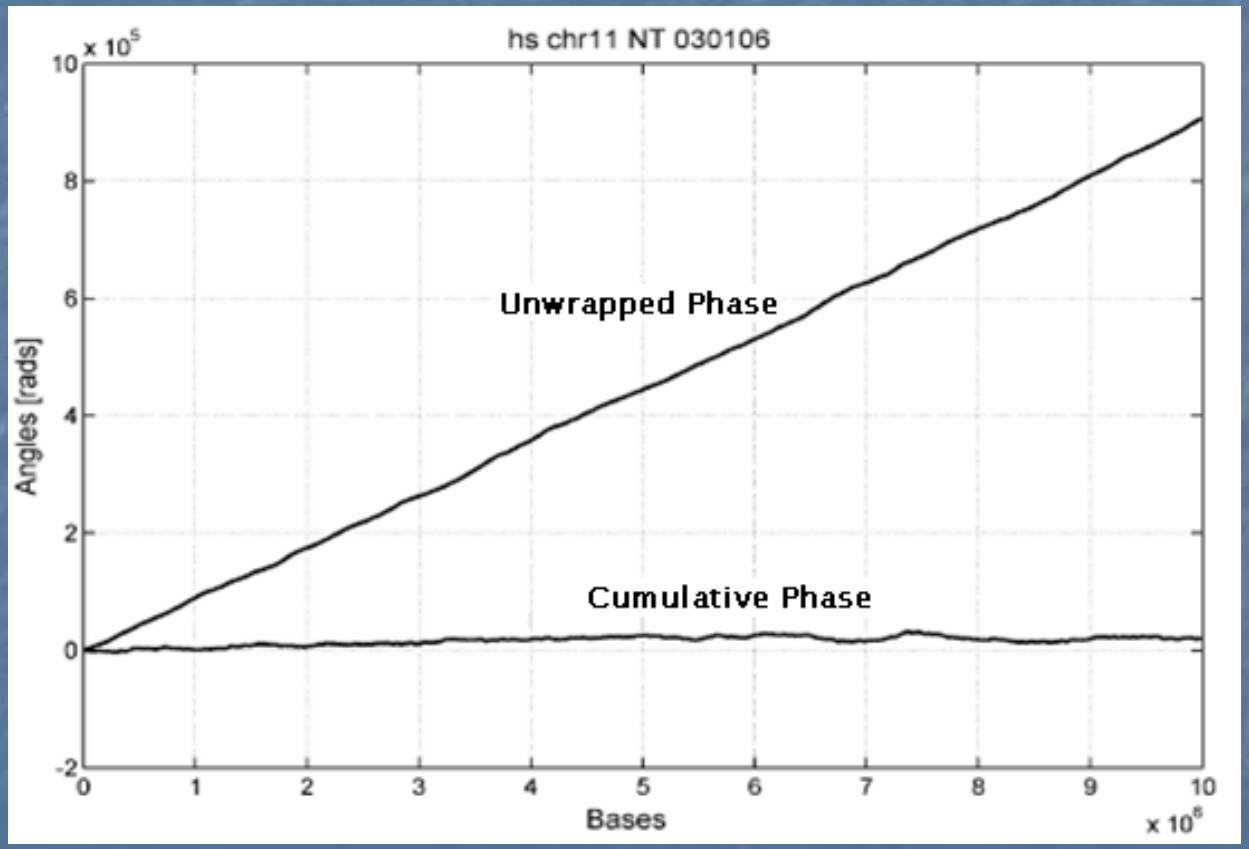
Figure reproduced from P. Cristea, *ELSEVIER Signal Processing*, vol. 83, no. 4, 2003.

Geometric Representations (cont'd)

- **Cumulative Phase:** It is the sum of the phase values of the complex numbers in a sequence starting from the first element up to the current element.
- **Unwrapped Phase:** It is the cumulative phase with the discontinuities produced by crossings of the negative real axis of the complex plane removed. It is obtained as follows:
  - If the complex number moves from the 2^{nd} to the 3^{rd} quadrant, add 2π to the cumulative phase.
  - If the complex number moves from the 3^{rd} to the 2^{nd} quadrant, subtract 2π from the cumulative phase.

Geometric Representations (cont'd)

- The complex representations can be used to observe long range phase characteristics of genomes.



Nearly linear phase characteristic of human chr. 11.

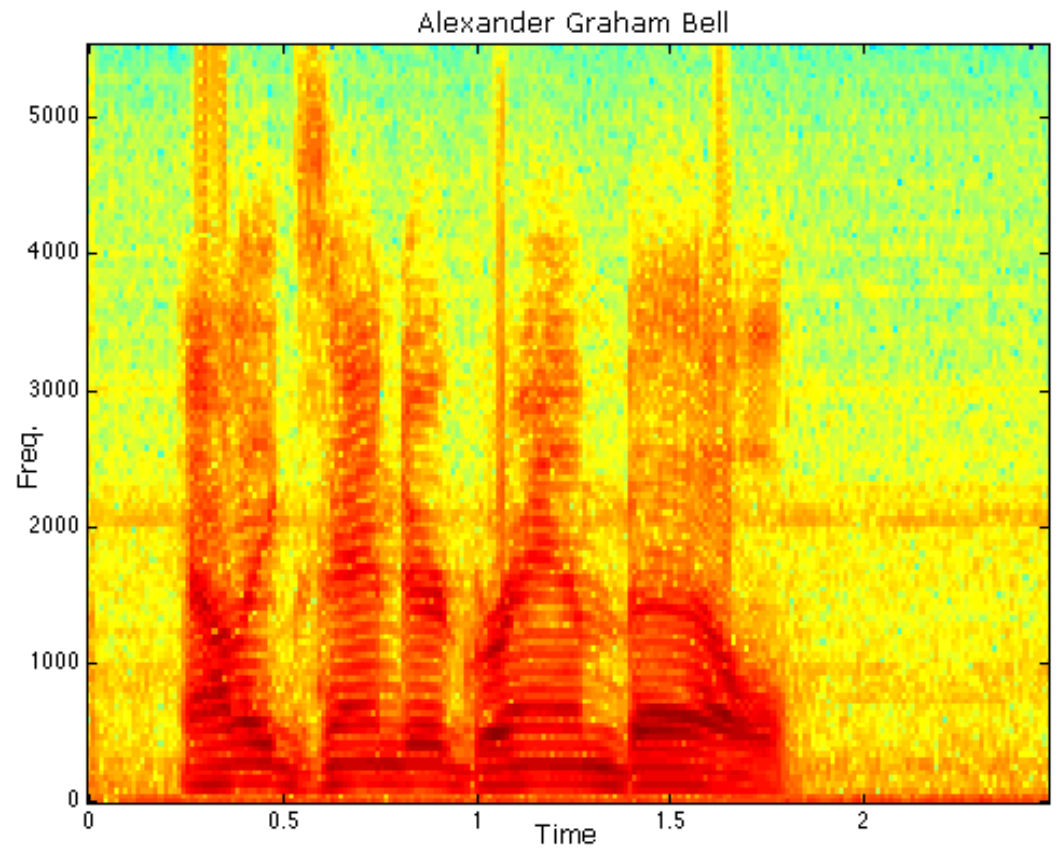
P. Cristea, *ELSEVIER Signal Processing*, vol. 83, no. 4, 2003.

The nearly linear unwrapped phase characteristic shows a long-range correlation in the occurrence of bases. Thus, the intergenic regions are NOT completely random!

Short-Time Fourier Transform (STFT)

- Provides a localized measure of the frequency content of a long sequence
- Example application: Used to analyze speech signals for their time-varying frequency content
- Two steps:
  - Apply a sliding window to the long sequence in order to divide it into short sections
  - Take the DFT of each individual section
- The individual DFTs form the columns of the STFT matrix. A plot of the magnitude of the STFT values is called a spectrogram.

STFT (cont'd)



A spectrogram of the utterance of the phrase "Alexander Graham Bell". The greater the redness the higher the energy level.

http://www.owlnet.rice.edu/~engi202/matlab.html

- Trained observers can identify the words in a speech signal just by looking at the spectrograms!

Color Spectrograms of DNA

- Very useful visualization tools, providing information about the local nature of DNA sequences
- A way of obtaining DNA spectrograms (see Anastassiou [2]) is by using the indicator sequences.
- Two steps:
  - Reduce the number of sequences from four to three so as to have three STFT matrices, one for each of the three primary colors **Red (R)**, **Green (G)**, and **Blue (B)**
  - Superimpose the three STFT matrices to obtain a single spectrogram

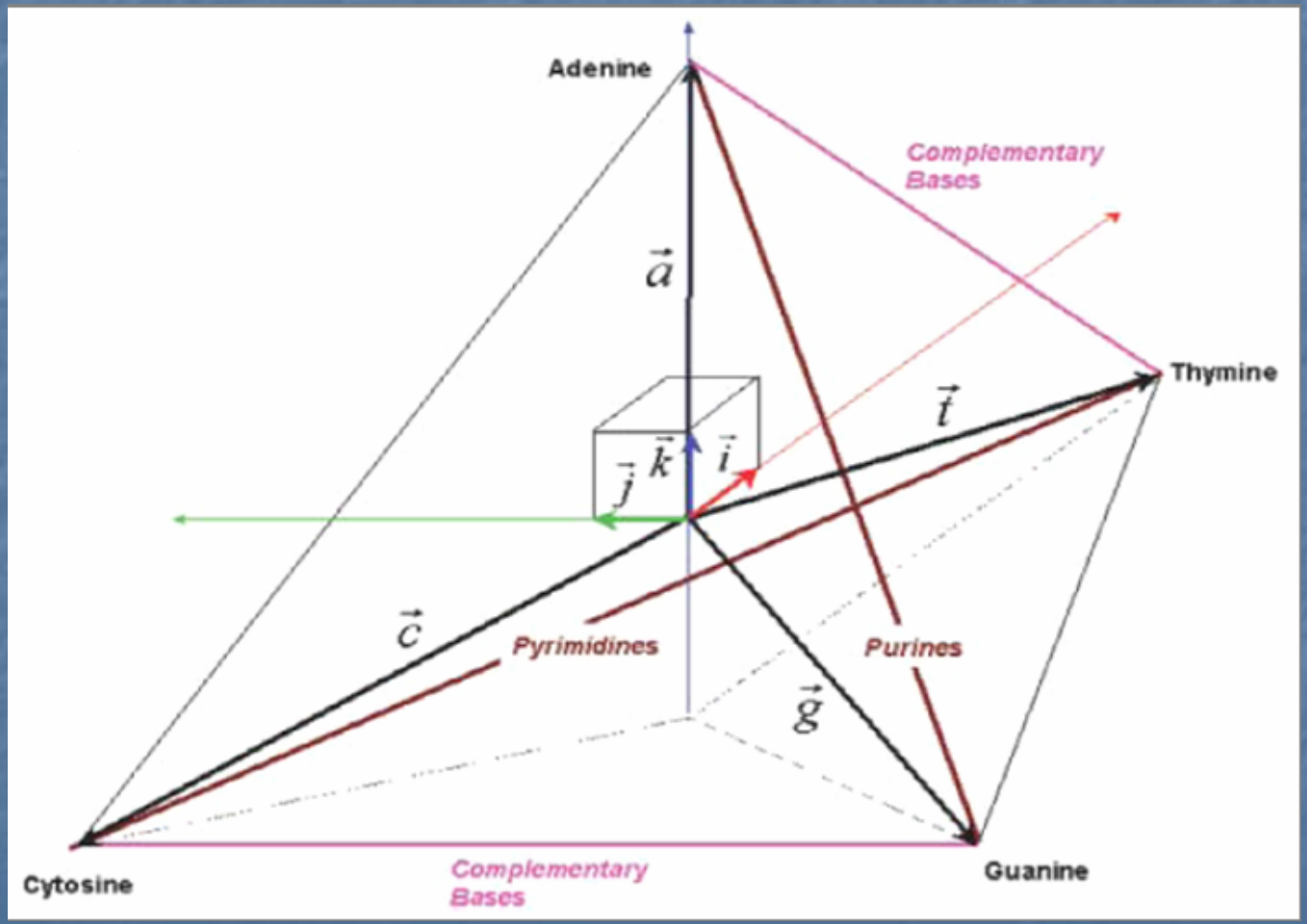
Color Spectrograms (cont'd)

Reducing the number of sequences

- Three steps:
  - Represent the 4 sequences by 4 3-dimensional vectors pointing from the center to the vertices of a regular tetrahedron
  - Resolve the 4 vectors along 3 mutually perpendicular directions namely $\vec{r}$, $\vec{g}$, and $\vec{b}$
  - Form 3 4-dimensional vectors, i.e., the 4 component-values along the $\vec{r}$ direction will form a vector, and so on
- In effect, transform 4 3-dimensional vectors into 3 4-dimensional vectors.

Color Spectrograms (cont'd)

Reducing the number of sequences (cont'd)



Rotated Geometrical Representation

P. Cristea, *J. Cell. Mol. Med.*, vol. 6, no. 2, 2002.

Color Spectrograms (cont'd)

Reducing the number of sequences (cont'd)

- On resolving the four three-dimensional vectors, we get

$$(a_r, a_g, a_b) = (0,0,1) \qquad (t_r, t_g, t_b) = \left(\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3}\right)$$

$$(g_r, g_g, g_b) = \left(-\frac{\sqrt{2}}{3}, -\frac{\sqrt{6}}{3}, -\frac{1}{3}\right) \qquad (c_r, c_g, c_b) = \left(-\frac{\sqrt{2}}{3}, \frac{\sqrt{6}}{3}, -\frac{1}{3}\right)$$

which give rise to the three numerical sequences

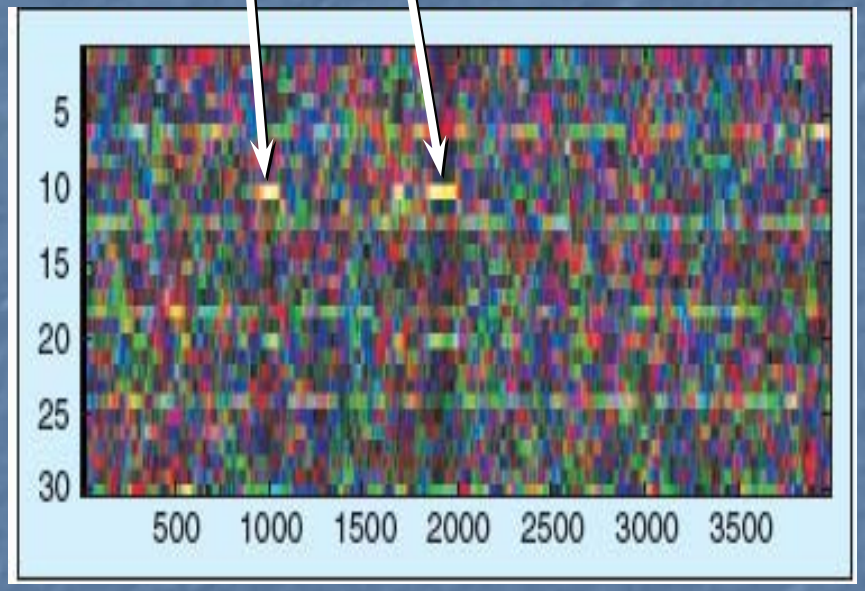
$$x_r[n] = \frac{\sqrt{2}}{3}(2x_T[n] - x_C[n] - x_G[n]), \qquad x_g[n] = \frac{\sqrt{6}}{3}(x_C[n] - x_G[n])$$

and $$x_b[n] = \frac{1}{3}(3x_A[n] - x_T[n] - x_C[n] - x_G[n]).$$

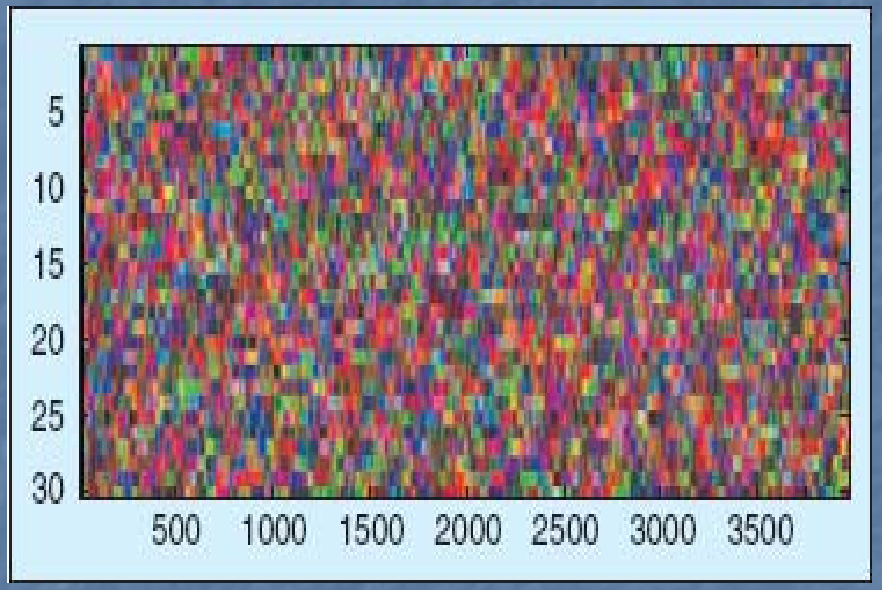
Color Spectrograms (cont'd)

- Color spectrograms can be used to locate repeating DNA sections.

Regions containing repeats



Real DNA section

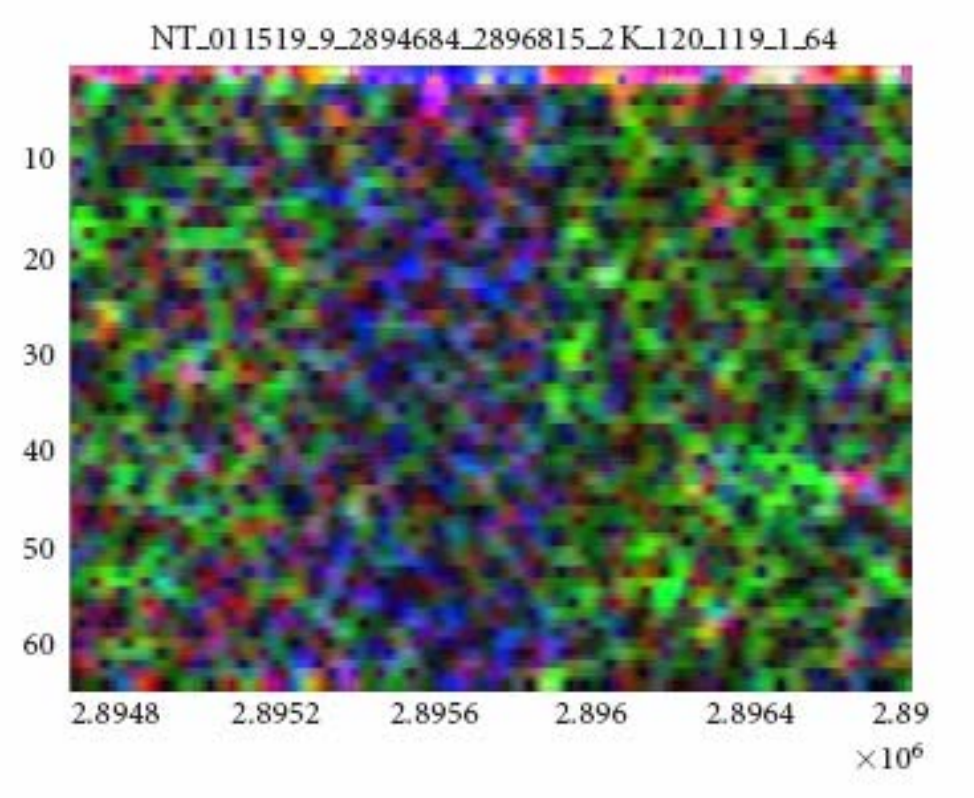


Artificial DNA section where each of the 4 bases occurs with probability ¼

D. Anastassiou, *IEEE Signal Processing Magazine*, Jul. 2001.

Color Spectrograms (cont'd)

- Spectrograms can be used to locate CG rich regions in DNA called CpG islands. The `p' in CpG simply denotes that C and G are linked by a phosphodiester bond.



CpG islands (green) separated by regions rich in A (blue).

D. Sussillo et al., *EURASIP Journal on Applied Signal Processing*, 2004.

Digital Filters for Identifying Protein-Coding Regions

- As was mentioned, protein-coding regions in DNA exhibit the period-3 property, i.e., there is a peak at frequency $2\pi/3$ in their Fourier spectra.
- This appears to be a fairly consistent property of coding regions. Hence, researchers have regarded it as a good indicator of coding regions.
- By locating the period-3 property coding sections can be identified.

  This can be done with digital filters (Vaidyanathan et al. [5]).

Application of Digital Filters (cont'd)

- The digital filter method consists of the following steps:
  - Design a narrowband bandpass digital filter with its passband centered at $\omega_0 = 2\pi/3$.
  - Process the 4 indicator sequences $x_A[n]$, $x_T[n]$, $x_G[n]$, and $x_C[n]$, one by one, with the digital filter where, n denotes base location. Let the corresponding output sequences be $y_A[n]$, $y_T[n]$, $y_G[n]$, and $y_C[n]$.
  - Define

$$Y[n] = |y_A[n]|^2 + |y_T[n]|^2 + |y_G[n]|^2 + |y_C[n]|^2$$

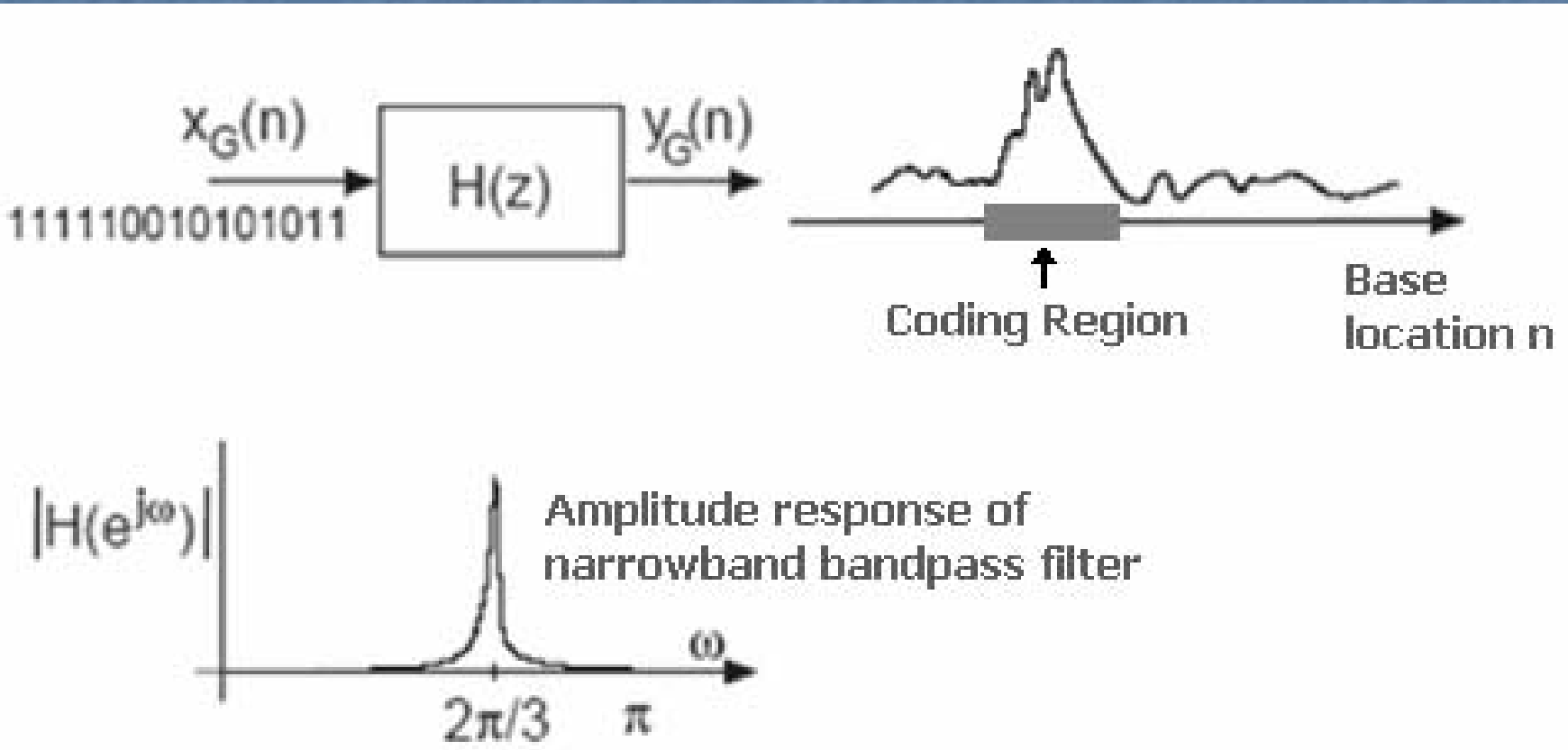
- A plot of $Y[n]$ can be used to indicate coding regions.

Application of Digital Filters (cont'd)

Intuitive Explanation

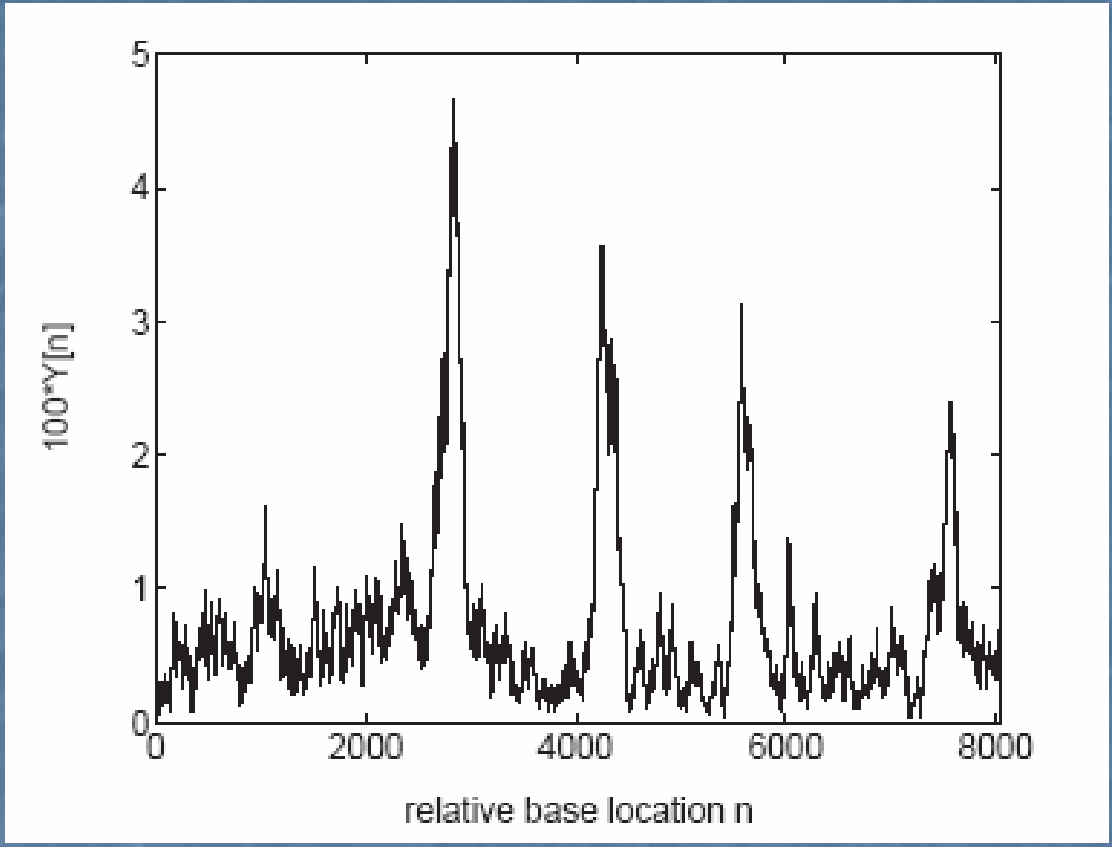
- Suppose we wish to analyze a very long DNA sequence using the digital filter method.
- Until we reach the protein-coding region, we will get a very low-level noise-like signal that is approximately constant.
- Once the protein-coding region begins to be processed, $Y[n]$ would go up by several decibels. This would mark the beginning of the protein-coding part.
- When the output level goes down to the noise level again, that would mark the end of the protein-coding part.

Application of Digital Filters (cont'd)



P. P. Vaidyanathan et al., *Journal of the Franklin Institute*, (341), 2004.

Application of Digital Filters (cont'd)



Output $Y[n]$ of the bandpass filter method, showing the coding regions (exons) as peaks in a *C. Elegans* gene.

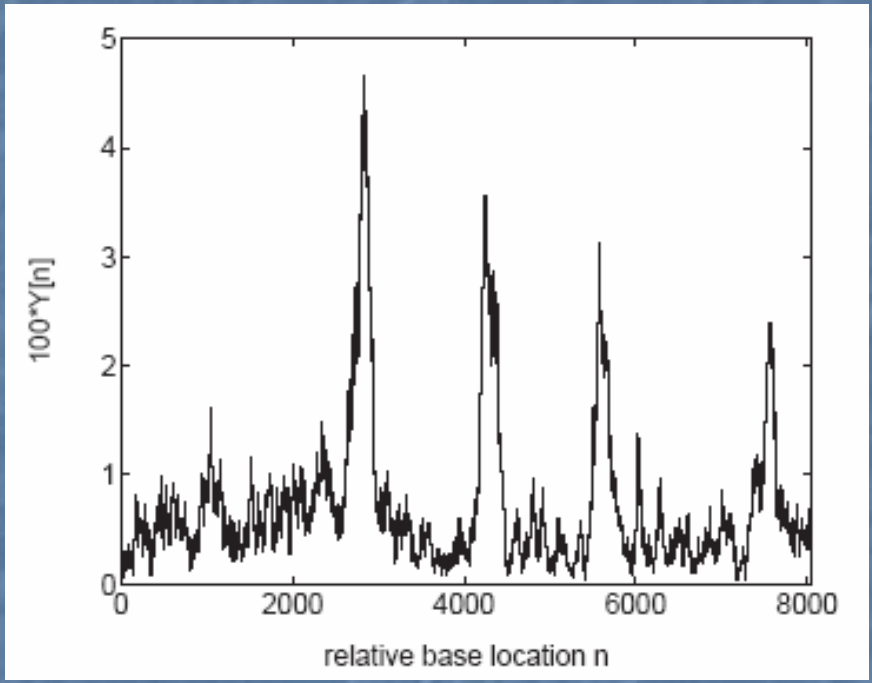
P. P. Vaidyanathan et al., *Journal of the Franklin Institute*, (341), 2004.

Application of Digital Filters (cont'd)

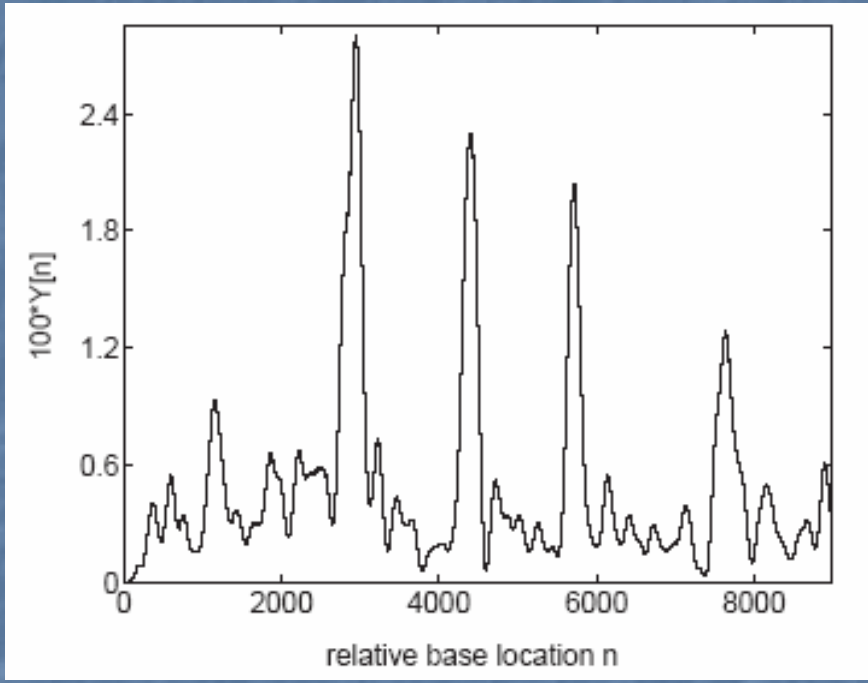
- A second-order, highly selective, narrowband bandpass filter can identify the period-3 property but it cannot remove much of the 1/f noise (background noise present in DNA sequences due to long-range correlation between base pairs).
- To also eliminate the 1/f noise as well as identify the period-3 property, one would need to use a high-order, highly selective, narrowband bandpass filter with larger minimum stopband attenuation.

P. P. Vaidyanathan et al., *Journal of the Franklin Institute*, (341), 2004.

Application of Digital Filters (cont'd)



Bandpass filter method for *C. Elegans* gene (background noise present)



Multistage filter method for the same gene (background noise removed)

P. P. Vaidyanathan et al., *Journal of the Franklin Institute*, (341), 2004.

Conclusions

- The application of DSP methods to genomic data have begun to make important contributions to genomic research.
- Open access to raw genomic data makes it easy for DSP experts to get involved in genomic research.
- With the huge number powerful DSP techniques developed over the years being applied to genomics, we can hope to see rapid advances in specialized areas such as customized drug design and genetic remedies, which will greatly benefit humankind.

References

1. B. Alberts et al., *Essential Cell Biology,* Garland Publishing, New York, 1998.
2. D. Anastassiou, "Genomic signal processing," *IEEE Signal Processing Magazine,* Jul. 2001.
3. P. D. Cristea, "Large-scale features in DNA genomic signals," *ELSEVIER Signal Processing,* vol. 83, no. 4, Apr. 2003.
4. S. Tiwari et al., "Prediction of probable genes by Fourier analysis of genomic sequences," *CABIOS,* vol. 13, no. 3, 1997.
5. P. P. Vaidyanathan et al., "The role of signal-processing concepts in genomics and proteomics," *Journal of the Franklin Institute,* vol. 341, 2004.

To download the slides for the talk

go to

www.ece.uvic.ca/~andreas

and click on

Genomic Digital Signal Processing

under Recent Lectures.