A

Seminar

Report

On

DNA COMPUTING

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS IN DEGREE COURSE IN COMPUTER ENGINEERING, BACHELOR OF ENGINEERING

BY

Harshal Dalvi

Pranav Firake

Aditya Gaikar

Taral Lokhande

Under the guidance of:

Mrs. Archana Gulati

DEPARTMENT OF COMPUTER ENGINEERING

DATTA MEGHE COLLEGE OF ENGINEERING

AIROLI, NAVI MUMBAI – 400708

Academic Year 2012 – 2013

DEPARTMENT OF COMPUTER ENGINEERING

DATTA MEGHE COLLEGE OF ENGINEERING

UNIVERSITY OF MUMBAI

CERTIFICATE OF APPROVAL FOR SEMINAR REPORT

This is to certify that

Harshal Dalvi

Pranav Firake

Aditya Gaikar

Taral Lokhande

have submitted their seminar on "DNA COMPUTING" in partial fulfillment of undergraduate program in Computer Engineering in the academic year 2012-2013.

Mrs. Archana Gulati	Prof. A. P. Pande	Dr. S. D. Sawarkar
(Internal Guide)	(HoD)	(Principal)
	Computer Department	

Academic Year 2012 – 2013

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards the Computer Technology Department of D.M.C.E.

It gives us the pleasure to present this seminar on DNA COMPUTING and would like to take this opportunity to thank everyone who has been a part of making this seminar.

Our gratitude and sincere thanks to Principal, Dr. S D Sawarkar and Prof. A. P. Pande (HoD of Computer Department) for providing all the necessary system facilities and privileges that enabled us to complete the seminar project in time. I would also like to thank Mrs. Archana Gulati for her guidance, genuine suggestions and constant encouragement, without which this seminar project would not have been a success.

Finally, I would like to thank my colleagues and friends who helped me in completing this seminar successfully.

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The twentieth century will be remembered for three major achievements – The evolution of computers, decoding of the human genome and evolution from Newtonian physics to quantum physics. Since the beginning of time man has performed computations or calculation. The method and nature of these computations has however changed from manual in the stone ages to mechanical in the medieval ages to electronic in the new computer age.

Computers, by definition, are machines which receive input, manipulate and store the input, and produce an output. They've quickly grown in the size and processing power. Computers are commonly known to consist of integrated circuits mainly constructed of silicon; however, a computer is never considered to be "alive".

What is going to be the future of computing systems? Can we look beyond silicon to embrace other mediums for computing? Computers inspired by biological or physical systems are possible alternatives. Microprocessors made of silicon will eventually reach their limits of speed and miniaturization. Chip makers need a new material to produce faster computing speeds.

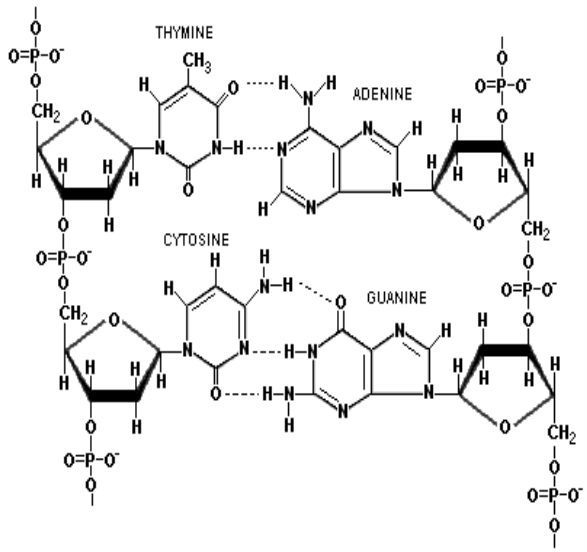
You won't believe where scientists have found the new material they need to build the next generation of microprocessors. Millions of natural supercomputers exist inside living organisms, including your body. DNA (Deoxyribo Nucleic Acid) molecules, the material our genes are made of, have the potential to perform calculations many times faster than the world's most powerful human-built computers. Technological advances however could use these building blocks of our genome in creating computer processors and data storage, and catapult processing speeds to incomprehensible levels not possible by today's standards.

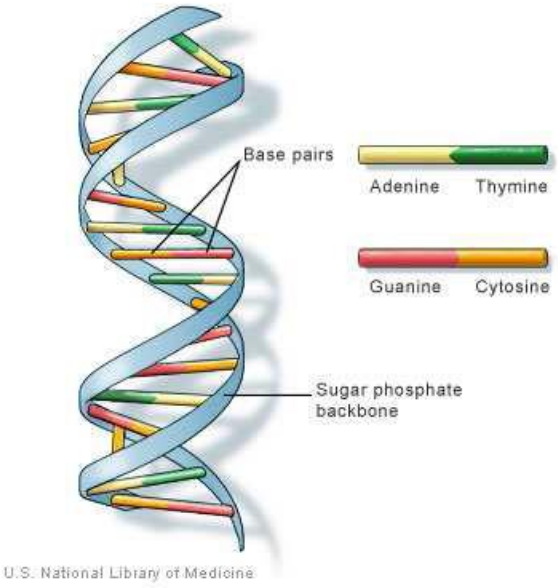
DNA computing is an alternative to the way computers work today. While this technology is not readily available, or being mass produced, the theory behind it is quite old and the development is ongoing and catching more speed. Companies like IBM are attempting to use DNA to produce the next generation of processors. Before discussing how DNA can be used in computers, it's important to first understand the basic structure of a DNA molecule.

1.1 What is DNA?

Deoxyribonucleic acid (DNA) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. Along with RNA and proteins, DNA is one of the three major macromolecules essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information.

These two strands run in opposite directions to each other and are therefore anti-parallel, one backbone being 3' (three prime) and the other 5' (five prime). This refers to the direction the 3rd and 5th carbon on the sugar molecule is facing. Attached to each sugar is one of four types of molecules called nucleobases (informally, bases). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.





1.2 Operations on DNA

Mixing: combine the contents of two test tubes into a third one to achieve union.

Melting: break apart a double-stranded DNA into its single-stranded complementary components by heating the solution. Melting in vitro is also known under the name of denaturation.

Annealing: bond together two single-stranded complementary DNA sequences by cooling the solution. Annealing in vitro is known as hybridization.

Amplifying (copying): make copies of DNA strands by using the Polymerase Chain Reaction PCR. The DNA polymerase DNA single-strand called template, and a shorter oligonucleotide called a primer, that is annealed to it.

Separating the strands by length using a technique called gel electrophoresis that makes possible the separation of strands by length.

Extracting those strands that contain a given pattern as a substring by using affinity purification.

Cutting DNA double-strands at specific sites by using commercially available restriction enzymes.

Ligating: paste DNA strands with compatible sticky ends by using DNA ligases. Indeed, another enzyme called DNA ligase, will bond together, or ``ligate", the end of a DNA strand to another strand.

Substituting: substitute, insert or delete DNA sequences by using PCR site-specific oligonucleotide mutagenesis.

Marking single strands by hybridization: complementary sequences are attached to the strands, making them double-stranded. The reverse operation is unmarking of the double-strands by denaturing, that is, by detaching the complementary strands. The marked sequences will be double-stranded while the unmarked ones will be single-stranded.

Destroying the marked strands by using exonucleases, or by cutting all the marked strands with a restriction enzyme and removing all the intact strands by gel electrophoresis.

Detecting and Reading: given the contents of a tube, say ``yes" if it contains at least one DNA strand, and ``no" otherwise. PCR may be used to amplify the result and then a process called sequencing is used to actually read the solution.

2. THE ADLEMAN EXPERIMENT

The concept of DNA computing was born in 1993, when Professor Leonard Adleman, a mathematician specializing in computer science and cryptography at the Laboratory of Molecular Science, Department of Computer Science, University of Southern California accidentally stumbled upon the similarities between conventional computers and DNA while reading the book "Molecular Biology of the Gene," written by James Watson, who co-discovered the structure of DNA in 1953. Adleman came to the conclusion that DNA had computational potential to solve complex mathematical problems.

In 1994, Leonard Adleman introduced the idea of using DNA to solve complex mathematical problems. In fact, DNA is very similar to a computer hard drive in how it stores permanent information about your genes.

2.1 The Experiment

Adleman is often called the inventor of DNA computers. Also he is one of the creators of one of the strongest encryption algorithms called RSA. In 1994, Leonard Adleman surprised the scientific community by using the tools of molecular biology to solve a different computational problem. His article in a 1994 issue of the journal Science outlined how to use DNA to solve a well-known mathematical problem, called the directed Hamilton Path problem, also known as the "traveling salesman" problem. The goal of the problem is to find the shortest route between a number of cities, going through each city only once. As you add more cities to the problem, the problem becomes more difficult. Adleman chose to find the shortest route between seven cities.

This computer solved the traveling salesman problem. There was nothing remarkable about the problem itself, which dealt with finding the shortest route through a series of points. Nor was there anything special about how long it took Adleman to solve it — seven days — substantially greater than the few minutes it would take an average person to find a solution. What was exciting about Adleman's achievement was that he had solved the problem using nothing but deoxyribonucleic acid (DNA) and molecular chemistry. You could probably draw this problem out on paper and come to a solution faster than Adleman did using his DNA test-tube computer. Here are the steps taken in the Adleman DNA computer experiment:

1. Strands of DNA represent the seven cities. In genes, genetic coding is represented by the letters A, T, C and G. Some sequence of these four letters represented each city and possible flight path.
2. These molecules are then mixed in a test tube, with some of these DNA strands sticking together. A chain of these strands represents a possible answer.
3. Within a few seconds, all of the possible combinations of DNA strands, which represent answers, are created in the test tube.
4. Adleman eliminates the wrong molecules through chemical reactions, which leaves behind only the flight paths that connect all seven cities.

Specifically, the method based on Adleman's experiment would be as follows:

1. Generate all possible routes.
2. Select itineraries that start with the proper city and end with the final city.
3. Select itineraries with the correct number of cities.
4. Select itineraries that contain each city only once.

All of the above steps can be accomplished with standard molecular biology techniques.

2.2 Reactions Required in the Experiment

Synthesis

A desired strand of DNA can be synthesized in lab. This is possible for strands up to a certain length. Longer 'random' strands are available. They consist of DNA sequences that have been cloned from many different organisms. The synthesizer is supplied with the four nucleotide bases in solution, which are combined according to a sequence entered by the user.

Denaturing, annealing and ligation

Double-stranded DNA may be dissolved into single strands (or denatured) by heating the solution to a temperature determined by the composition of the strand. Heating breaks the hydrogen bonds between complementary strands. Since the hydrogen bonds between strands are much weaker than the covalent bonds within strands, the strands remain undamaged by this process. Since a G-C pair is joined by three hydrogen bonds, the temperature required to break it is slightly higher than that for an A-T pair, joined by only two hydrogen bonds.

Hybridization separation

Separation by hybridization is an operation often used in DNA computation, and involves the extraction from a test tube of any single strands containing a specific short sequence (e.g., extract all

strands containing the sequence TAGACT). If we want to extract single strands containing the sequence X, we first create many copies of its complement. We attach to these oligonucleotides a biotin molecule which binds in turn to a fixed matrix.

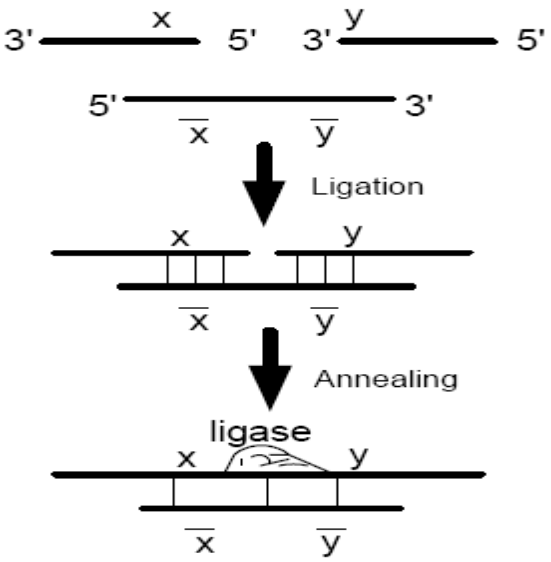
Gel-Electrophoresis

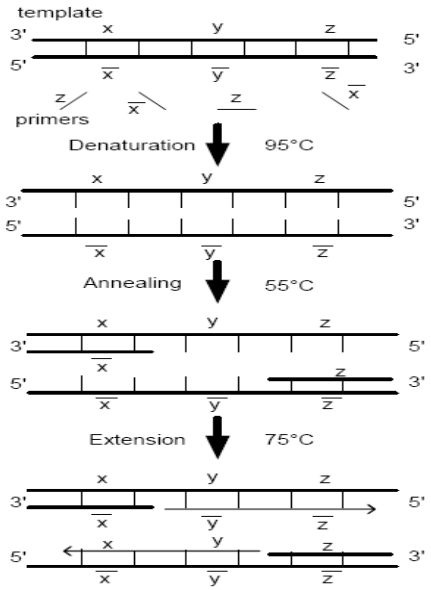
Gel electrophoresis is an important technique for sorting DNA strands by size. Electrophoresis is the movement of charged molecules in an electric field. Since DNA molecules carry negative charge, when placed in an electrical field they tend to migrate towards the positive pole. The rate of migration of a molecule in an aqueous solution depends on its shape and electrical charge. Since DNA molecules have the same charge per unit length, they all migrate at the same speed in an aqueous solution. However, if electrophoresis is carried out in a gel (usually made of agarose, polyacrylamide or a combination of the two) the migration rate of a molecule is also affected by its size.

Primer extension and PCR

The DNA polymerases perform several functions, including the repair and duplication of DNA. Given a short primer oligonucleotide, p, in the presence of nucleotide triphosphates, the polymerase extends p (always in the 50/30 direction) if and only if p is bound to a longer template oligonucleotide, t.

Another useful method of manipulating DNA is the PolymeraseChainReaction, or PCR. PCR is a process that quickly amplifies the amount of a specific molecule of DNA in a given solution using primer extension by polymerase. Each cycle of the reaction doubles the quantity of this molecule, giving an exponential growth in the number of strands.





3. DNA COMPUTER

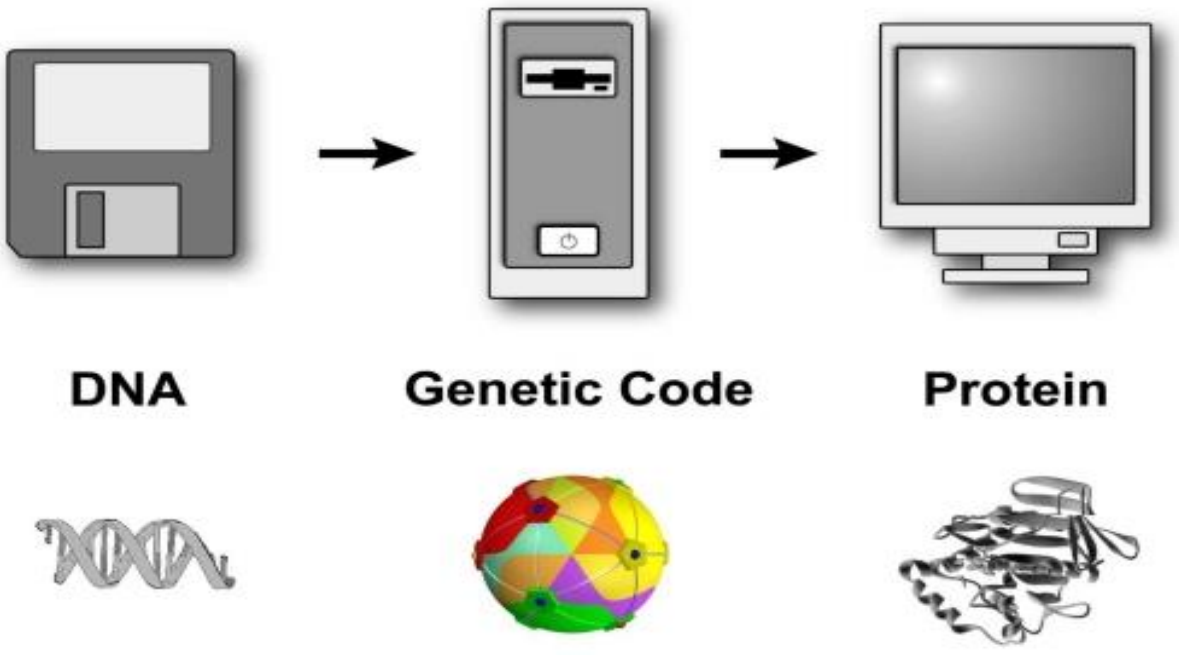
A DNA computer is basically a collection of specially selected DNA strands whose combinations will result in the solution to some problem, depending on the problem at hand. Technology is currently available both to select the initial strands and to filter the final solution. The promise of DNA computing is massive parallelism: with a given setup and enough DNA, one can potentially solve huge problems by parallel search. This can be much faster than a conventional computer, for which massive parallelism would require large amounts of hardware, not simply more DNA. Since Adleman's original experiment researchers have developed several different models to solve other mathematical and computational problems using molecular techniques.

There are three reasons for using DNA computing to solve computational problems:

(1) The information density of DNA is much greater than that of silicon: 1 bit can be stored in approximately one cubic nanometer. Other storage media, such as videotapes, can store 1 bit in 1,000,000,000,000 cubic nanometer.

(2) Operations on DNA are massively parallel: a test tube can contain trillions of strands. Each operation on a test tube of DNA is carried out on all strands in the tube in parallel.

(3) DNA computing is an interdisciplinary field where: biologists, computer scientists, physics, mathematicians, chemists, etc. find a lot of interesting problems which can be applied to both theoretical and practical areas of DNA computing.



3.1 Applications of DNA Computers

As far as applications are concerned, this can be quite useful in figuring out how to route telephone calls, plane trips, and basically any problem that can be turned into a Hamiltonian problem. It is also been claimed that DNA can be used to solve optimization problems involving business management. This would involve optimizing the routing of raw materials. It is even said that DNA can be used in devising the wiring schematics for circuits.

1. Applications making use of ”classic” DNA computing schemes where the use of massive parallelism holds an advantage over traditional computing schemes, including potential polynomial time solutions to hard computational problems.

2. Applications making use of the ‘natural’ capabilities of DNA, including those that make use of informational storage abilities and those that interact with existing and emerging biotechnology.

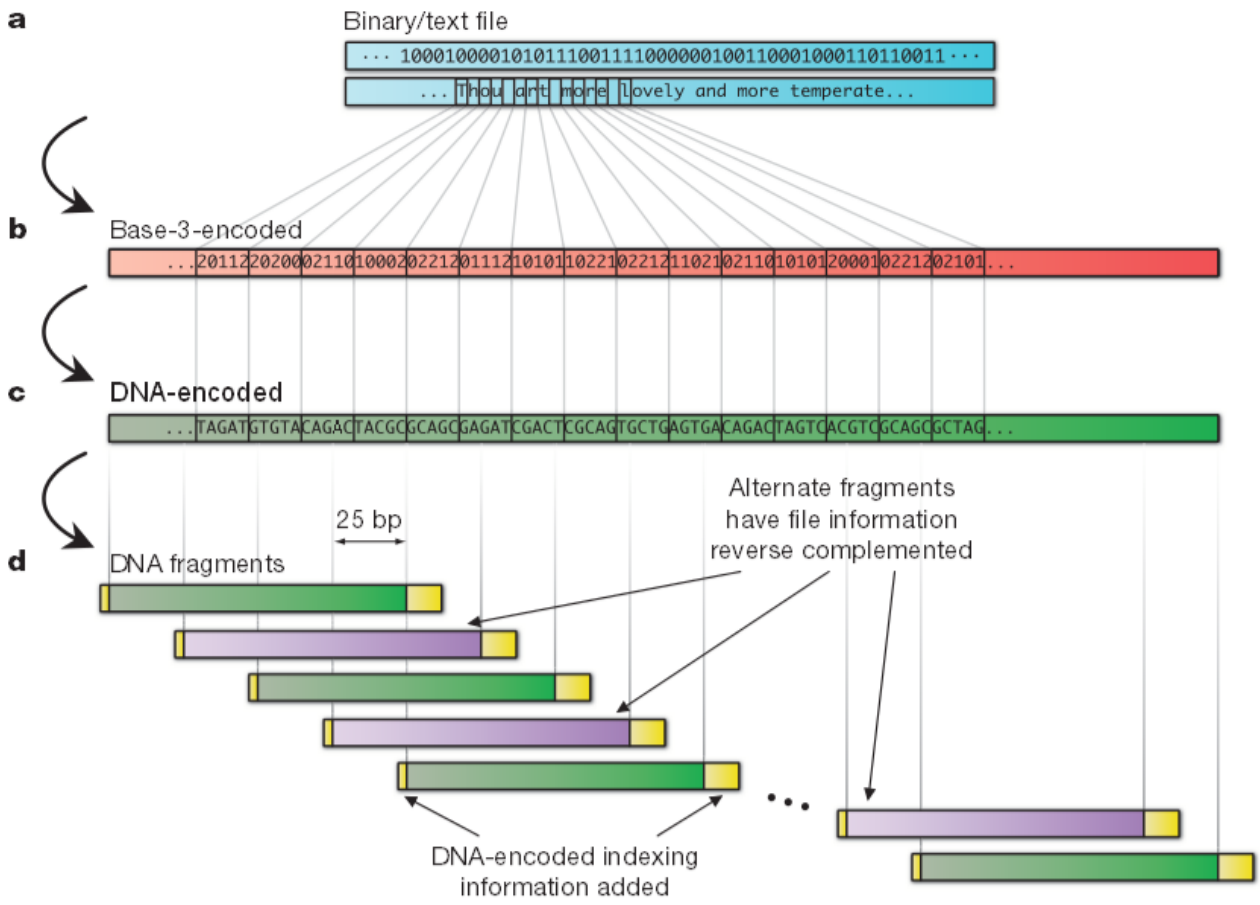
3. Contributions to fundamental research within both computer science and the physical sciences, especially concerning exploring the limitations of computability and to understanding and manipulating bimolecular chemistry.

4. Classical DNA computing techniques have already been theoretically applied to a real life problem: breaking the Data Encryption Standard (DES). Although this problem has already been solved using conventional techniques in a much shorter time than proposed by the DNA methods, the DNA models are much more flexible, potent, and cost effective. The brief description about DES as follows.

4. DATA STORAGE USING DNA

When it comes to storing information, hard drives don't hold a candle to DNA. Our genetic code packs billions of gigabytes into a single gram. A mere milligram of the molecule could encode the complete text of every book in the Library of Congress and have plenty of room to spare. All of this has been mostly theoretical — until now. In a new study, researchers stored an entire genetics textbook in less than a picogram of DNA — one trillionth of a gram — an advance that could revolutionize our ability to save data.

To encode a digital file, researchers divide it into tiny blocks of data and convert these data not into the 1s and 0s of typical digital storage media, but rather into DNA's four-letter alphabet of As, Cs, Gs and Ts. Each DNA fragment also contains a digital "barcode" that records its location in the original file. Reading the data requires a DNA sequencer and a computer to reassemble all of the fragments in order and convert them back into digital format. The computer also corrects for errors; each block of data is replicated thousands of times so that any chance glitch can be identified and fixed by comparing it to the other copies.



4.1 Writing Data to DNA

One Bit per Base Encoding

If each base in the sequence represents one bit of encoded value, there are 16 possible encodings, representing the truth tables for all Boolean functions of two variables. Encodings 0 ("always zero") and 15 ("always one") are devoid of information and may be ignored. The remaining 14 encodings represent all possible ways each base can specify one bit of information independent of context.

Encoding	A	C	G	T	Encoding	A	C	G	T
0	0	0	0	0	8	0	0	0	1
1	1	0	0	0	9	1	0	0	1
2	0	1	0	0	10	0	1	0	1
3	1	1	0	0	11	1	1	0	1
4	0	0	1	0	12	0	0	1	1
5	1	0	1	0	13	1	0	1	1
6	0	1	1	0	14	0	1	1	1
7	1	1	1	0	15	1	1	1	1

Why encode only one bit per base? The three dimensional structure of the DNA molecule, often characterised as a regular double helix, in fact depends upon the sequence of bases making it up. Encoding only one bit per base, thereby sacrificing 50% of the potential information density, creates redundancy which permits encoding of arbitrary information within the chemical constraints of the DNA molecule. It's worth keeping in mind that the genetic code exhibits better than three-to-one redundancy (64 possible codons, 21 amino acids plus a stop code encoded), quite probably due to the same constraints.

Two Bit per Base Encoding

Encoding two bits per base pair achieves the maximum information density attainable. As discussed above, however, it sacrifices the freedom to chose among multiple redundant encodings to obtain desired chemical properties and reliable replication. Still, it's worth looking at how one can encode two bits per base, because there may be circumstances in which "double density DNA" encoding is feasible.

There are twenty-four possible ways to encode two bits of information in each base pair. Encoding 0 assigns values from binary 00 through 11 to the bases in alphabetical (arbitrary)

Encoding	A	C	G	T	Encoding	A	C	G	T
0	00	01	10	11	12	10	00	01	11
1	00	01	11	10	13	10	00	11	01
2	00	10	01	11	14	10	01	00	11
3	00	10	11	01	15	10	01	11	00
4	00	11	01	10	16	10	11	00	01
5	00	11	10	01	17	10	11	10	00
6	01	00	10	11	18	11	00	01	10
7	01	00	11	10	19	11	00	10	01
8	01	10	00	11	20	11	01	00	10
9	01	10	11	00	21	11	01	10	00
10	01	11	00	10	22	11	10	00	01
11	01	11	10	00	23	11	10	01	00

order; the other encodings represent all permutations of the list of four two bit binary values. Since the number of permutations of n items is $n!$, there are a total of 4! or 24 two bit encodings.

4.2 An Image in the Genome

To show how information of any kind can be encoded in DNA, let's work through a concrete example in excruciating detail. The information we'll encode is an image of the Swiss flag. Examining the image we're going to encode reveals two key properties: it contains only two colours, red and white, and the regularities in the image allow it to be compressed to an image only 18 pixels square. We're interested in the pattern, not the colour.

From Image to Bits

Having reduced the initial image into an 18×18 monochrome image, this can then be rewritten as the following string of binary digits, with zeroes representing the black pixels and ones the white (this choice is arbitrary--the person who decodes the message may choose the opposite encoding, but will still discover the pattern).

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Simple data stream would look like:

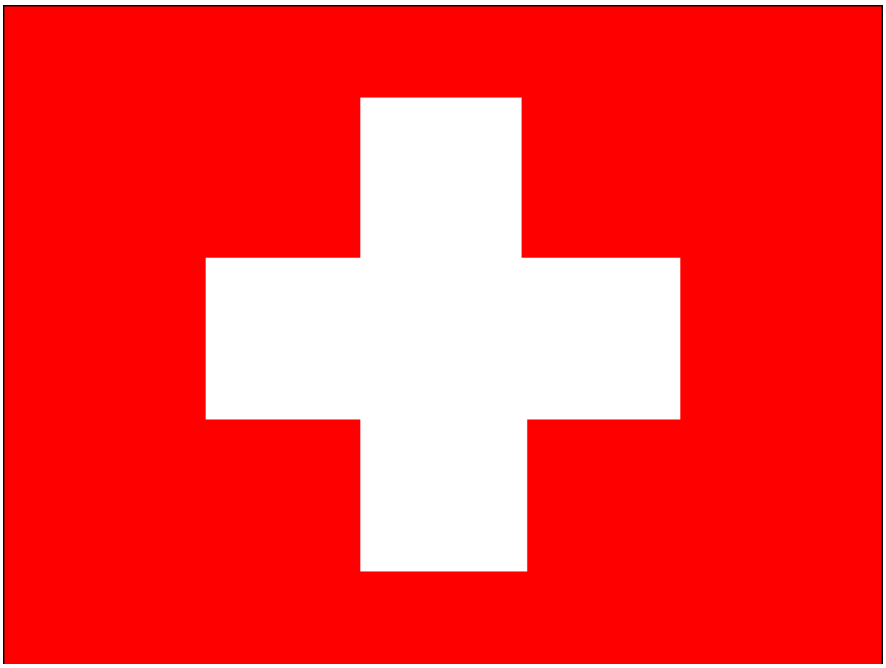
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1
1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

From Bits to DNA

Now we're ready to encode the bits representing the messages as a sequence of nucleotides in a strand of DNA. First, we choose one of the forms of encoding described above. For this example, I'll encode one bit per base and use Encoding 6 as given in the table above. Encoding 6 prescribes that A or T nucleotides encode zeroes, and C or G encode ones. The choice of which to use for a given one or zero is arbitrary. Since complementary pairs are always A-T or C-G, we can think of Encoding 6 as coding the data in the choice of complementary pair, while ignoring which member of the pair is on which strand of the double helix. Chemically, this is an attractive way to encode information since it allows us to avoid a purine-rich sequence which could trigger transcription; to do so we simply make sure to include sufficient pyrimidines (thymine or cytosine) to be safe; since the information is encoded in the choice or complementary pair, their order is irrelevant to the coding and can be flipped where necessary to meet this constraint.

The result of our encoding, then, would be something like the following:

```
TTTTATATATATAAATATAAAATATAAAAATTATTAAATATAAGGCCTTTAAAAATATATACCG
CAAATAATATTTATAGCCCTTTTTTTTATTTTTCCGCAATAAATAATTATAGCCCAATATTTTT
CGCGCGGGCGGCGCTATTGCGGCCGGCGCCCGTAAACGGCGCCCGCGCCCATTTGCCCGGGCCC
CGCCTTTTTAATTGGGCTTTTTAAATTAATTCGGCATTATTTAATATAACCGGAATAATTTTTA
TAAGGCCTTATTAAAATTTAACCGGTATTAATTAAAATTTAATTAATTTAAATTTAAATTTAAT
ATTA
```



CONCLUSION

Before you trash your silicon-based computer and start trying to process words with DNA, remember that it'll be a while before the wet computers show up in showrooms. DNA computers can't be found at your local electronics store yet. The technology is still in development.

Bio molecular computers, made of DNA and other biological molecules, only exist today in a few specialized labs, remote from the regular computer user. DNA computer components -- logic gates and biochips -- will take years to develop into a practical, workable DNA computer. If such a computer is ever built, scientists say that it will be more compact, accurate and efficient than conventional computers.

The current applications of DNA chips are restricted to the field of medicine. Affymetrix Inc. pioneered the research in the field of DNA medicine. However now many companies such as Motorola and Corning and the Hewlett-Packard spinoff Agilent Technologies have joined this rapidly growing technology. Each of these challengers is applying its industrial expertise to making its own DNA microarrays or chips. DNA chips or arrays have been used to solve many problems in the field of medicine.

These DNA computers can be used in fluids, such as a sample of blood or in the body, and make decisions at the level of a single cell. DNA computers could conceivably be implanted in the body to both diagnose and kill cancer cells or monitor and treat diabetes by dispensing insulin when needed.

The first DNA computers are unlikely to feature word processing, e-mailing and solitaire programs. Instead, their powerful computing power will be used by national governments for cracking secret codes, or by airlines wanting to map more efficient routes. Studying DNA computers may also lead us to a better understanding of a more complex computer -- the human brain. In the future, some speculate, there may be hybrid machines that use traditional silicon for normal processing tasks but have DNA co-processors that can take over specific tasks they would be more suitable for.

WEBLIOGRAPHY

- http://www.howstuffworks.com
- http://www.rsa.com
- http://www.features.techworld.com
- http://www.networkworld.com
- http://www.ezinearticles.com
- http://www.nature.com
- http://www.latestinfo.com
- http://computerworld.com
- http://en.wikipedia.org
- http://webopedia.com