

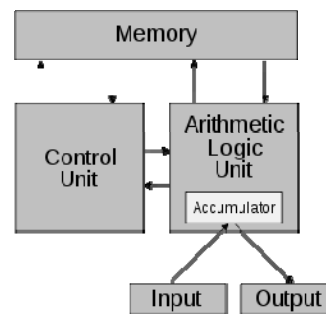
Computing with DNA & Review and Study Suggestions

1

Wednesday, April 29, 2009

von Neumann Architecture

- Refers to the existing computer architectures consisting of
 - a processing unit
 - a single separate storage structure to hold both instructions and data
- Implements a Turing Machine



2

DNA Computing

- Existing genetic engineering tools can slice, create, and rebuild DNA sequences:
 - Custom sequences can be made “to order”.
 - Duplication, replication, and selection operations exist for DNA sequences.
- Many ($\approx 10^{15}$) different DNA sequences can be operated on at the same time in a single test tube.
- DNA acts as memory; computation is carried out by complementary bases

3

Wednesday, April 29, 2009

How can DNA manipulations lead to computation?

In 1993, Leonard Adleman (USC) proposed DNA computing

- Storage
 - DNA and binary strings encode information
- Alphabet
 - DNA has a 4-letter alphabet: A, C, G, T
 - Computers have a 2-letter alphabet: 0, 1
- Operations
 - We can synthesize any strand of DNA
 - We can generate any binary sequence

4

Wednesday, April 29, 2009

How can DNA strings carry out computation?

DNA strings attach to each other (twist into a double helix) if they have complementary elements in corresponding positions.

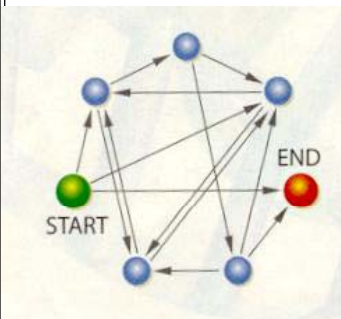
- A and T are complements
- C and G are complements

Adleman devised a DNA manipulation technique to solve the Hamiltonian Path (HP) problem.

5

Wednesday, April 29, 2009

- DNA sequences were created
- They allowed the execution of trillions of operations in parallel



From Adleman's '98 Scientific American paper

6

Wednesday, April 29, 2009

Adleman's first DNA computation operated on a 7-node, 14-edge directed graph.

The algorithm is probabilistic.

- With high probability it will generate a Hamiltonian path, if one exists.

It is a hard-wired, not a programmed solution.

- The setup is for one graph.
- It needs to be repeated from scratch for another graph.

Demonstrated feasibility and started a new area:

DNA Computing

7

Wednesday, April 29, 2009

Overview of the HP DNA algorithm

Given is a directed graph G and nodes s and t

1. Generate DNA sequences to represent nodes and edges.
2. Generate DNA strands representing paths.
3. From among the paths, select a path that:
 - begins at s and ends at t
 - has length $n-1$
 - is a simple path

8

Wednesday, April 29, 2009

Encoding of nodes and edges

For every node v_i :

generate a random DNA strand R_i of length 20

For every edge (v_i, v_j) :

create the string $S(i,j)$, which consists of the first half of R_i and the second half of R_j .

Example:

v_2 as $R_2 = \text{ATCGAACGTTTAAACGTAGT}$

v_3 as $R_3 = \text{TCGAATTACGTAGAACGTTT}$

edge (v_2, v_3) as **TAAACGTAGT**TCGAATTACG

9

Wednesday, April 29, 2009

Initial conditions

Start node s has strand R_s and end node t has strand R_t

For every edge (s, v_i) , create the edge strand consisting of R_s and the first half of R_i

For every edge (v_i, t) , create the edge strand consisting of the second half of R_i followed by R_t

Consider a path $\langle s, v_1, v_2, \dots \rangle$

$R_s = \text{TTCATTTCGCCTTAAAGGACT}$

$R_1 = \text{GTAGTACGTTTCGGCGTAGA}$

$R_2 = \text{ATCGAACGTATTAACGTAGT}$

TTCATTTCGCCTTAAAGGACTGTAGTACGTTTCGGCGTAGA**ATCGAACGTATTA**

...

10

Wednesday, April 29, 2009

10

What goes on in the test tube?

- Complement of the node strands – lots of them
- Edge strands – lots of them

DNA strands attach to form double strands if they have complementary elements in the corresponding positions.

11

Wednesday, April 29, 2009

11

Generate

$R_s = \text{TTCATTTCGCCTTAAAGGACT}$

$R_1 = \text{GTAGTACGTTTCGGCGTAGA}$

$R_2 = \text{ATCGAACGTATTAACGTAGT}$

What goes into the tube

$c(R_s) = \text{AAGTAAGCGGAATTCCTGA}$

$c(R_1) = \text{CATCATGCAAAGCCGCAACA}$

$c(R_2) = \text{TAGCTTGCATAATTGCATCA}$

edge (s, v_1) : $\text{TTCATTTCGCCTTAAAGGACTGTAGTACGTT}$

edge (v_1, v_2) : $\text{TCGGCGTAGAATCGAACGTA}$

Reaction

$\text{TTCATTTCGCCTTAAAGGACTGTAGTACGTTTCGGCGTAGAATCGAACGTA}$

$\text{AAGTAAGCGGAATTCCTGACATCATGCAAAGCCGCAACATAGCTTGCATAATTGCATC}$

12

Putting it all together

1. Generate DNA sequences representing nodes and edges.
2. Mix to generate the paths.
3. From among the paths, select a path that
 - begins at s and ends at t
 - has length $n-1$
 - is a simple path

Step 3 involves operations like merge, amplify, test-if-empty, separate, separate-by-length, separate-by-positions.

13

Wednesday, April 29, 2009

SCIENCE CLASSICS

BY LARRY GONICK

THE SOLUTION

AS COMPUTER COMPONENTS SHRINK YEAR BY YEAR, SCIENTISTS DREAM OF THEIR ULTIMATE GOAL: A CHEMICAL COMPUTER, WHOSE WORKING PARTS WOULD BE INDIVIDUAL MOLECULES.

BUT THIS HAS REMAINED ONLY A DREAM—UNTIL NOW. LEONARDY ADLEMAN OF THE UNIVERSITY OF SOUTHERN CALIFORNIA HAS JUST SHOWN HOW TO DO COMPUTATION USING DNA.

IF YOU HAVE PROBLEMS, DISSOLVE THEM.

ADLEMAN, A COMPUTER SCIENTIST, CHOSE A TASK THAT REPRESENTS A WHOLE CLASS OF HARD-TO-SOLVE PROBLEMS. COMPUTER GUYS CALL IT THE TRAVELING SALESMAN PROBLEM.

IN THIS VERSION, THE MARKETING REP HAS A MAP OF SEVERAL CITIES WITH ONE-WAY STREETS BETWEEN SOME OF THEM. THE PROBLEM IS TO FIND A ROUTE (IF IT EXISTS) THAT PASSES THROUGH EACH CITY EXACTLY ONCE, WITH A DESIGNATED BEGINNING AND END.

COULDN'T YOU CALL IT SOME-THING A LITTLE LESS GENDER BIASED, A LITTLE MORE... NOW?

HOW ABOUT THE MOBILE MARKETING REP PROBLEM?

YOU CONFUSURE! HOW ABOUT THE TRAVELING SALESMAN PATH PROBLEM?

WHEN THE NUMBER OF CITIES IS LARGE—SAY MORE THAN 100—THIS PROBLEM IS TOO MUCH FOR EVEN THE FASTEST COMPUTER.

FOR HIS DNA COMPUTATION, ADLEMAN CHOSE THIS SIMPLE ARRANGEMENT OF 7 CITIES AND 18 STREETS.

HE REPRESENTED EACH CITY CHEMICALLY BY A SINGLE STRAND OF DNA 20 BASES LONG. ITS SEQUENCE CHOSEN AT RANDOM.

1 AAATGACTAGGACTCCCA
2 CTGGAGGGCTTAAGGAAT
3 ATTTCCTTATCCG
4 CCATCCCTC

THE ACTUAL SEQUENCES DON'T MATTER?

A STREET BETWEEN TWO CITIES IS THE COMPLEMENTARY 20-BASE STRAND THAT OVERLAPS EACH CITY'S STRAND HALFWAY. THIS STREET LITERALLY JOINS THE TWO CITIES.

PATH 1 → 2

1 2

IN DNA, C ALWAYS PAIRS WITH G, AND T ALWAYS PAIRS WITH A, SO IN CLOSE-UP, IT LOOKS LIKE THIS:

G T A G C T A G
C A T G C A T G

A MULTICITY TOUR BECOMES A PIECE OF DOUBLE-STRANDED DNA, WITH THE CITIES LINKED IN SOME ORDER BY THE STREETS.

NOTE: SOME CITIES MAY BE VISITED MORE THAN ONCE.

14

<http://www.msri.org/external/larryg/pages/11.htm>

Progress since 1993

Other problems were solved using Adleman's approach (formula satisfiability, factoring numbers, combinatorial problems with a large solution space).

It was shown that DNA computation does not provide additional computational power beyond that of a Turing machine.

Between 2002-04, researchers from the Weizmann Institute of Science (Israel) developed a programmable molecular computing machine composed of enzymes (hardware) and DNA molecules (data).

15

Wednesday, April 29, 2009

15

What about quantum computing?

- A quantum computer uses quantum mechanical phenomena - superposition and entanglement - to perform operations on data.
 - Quantum properties are used to represent data and perform operations.
- Quantum computing has the potential to support a new kind on computation
 - A system with n qubits can perform 2^n calculations at one
- The technology supporting quantum computation has not yet been developed

16

Quantum Computing

- Quantum computers would be different from traditional computers as well as DNA computers.
- A quantum computer could not solve unsolvable problems (like the Halting Problem)
- But, exponential time algorithms may have feasible solutions.
 - Fast integer factorization would break cryptographic systems
 - Security of existing methods for public key encryption (RSA systems) would no longer be guaranteed

17

Wednesday, April 29, 2009

Review and study suggestions

- Exam will have some programming questions, some explain questions
- Sample questions will be posted on Blackboard this week
- Programming questions
 - Write one line statements achieving what is specified
 - Write small programs (like in exam 2)
 - Know how to generate plots/visuals (not Cytoscape)
- Explain Questions
 - Give short answers addressing the point
 - Don't reproduce material from sources

18

Wednesday, April 29, 2009

Review and study suggestions

- Work through lab material and lab problems
 - Understand what the focus of the lab was
 - Read posted solutions
 - Re-solve problems
- Understand code posted in association with lectures
 - Change it, break it ...
- Make sure to review ...
 - Recursion
 - Graph operations
 - Dictionaries
 - Be able to read class definitions and methods

19

Review and study suggestions

- **Chapters covered from Zelle**
 - 1, 2, 3, 4
 - 6, 7, 8, 11
 - 13 (except 13.3)
- **Software packages (basic uses)**
 - Matplotlib, VPython, NetworkX, NumPy
- **Learning Python Book**
 - Mainly for looking up material and seeing different examples
 - There are a number of Python features we did not cover

20

Wednesday, April 29, 2009