COMPUTING USING DNA SYSTEMS

DEEPAK SHARMA

DEPARTMENT OF CHEMICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2019

COMPUTING USING DNA SYSTEMS

by

DEEPAK SHARMA

Department of Chemical Engineering

Submitted

in partial fulfillment of the requirements of the degree of ***Doctor of Philosophy***

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2019

I would like to dedicate this thesis to my grandfather (Late) **Rameshwar Das** . . .

Certificate

This is to certify that the thesis titled **"Computing Using DNA systems"** being submitted by **Mr. Deepak Sharma** in the Department of Chemical Engineering, Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy**, in Chemical Engineering, is a record of the original, bona-fide research work carried out by him under my guidance and supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted for the award of any other degree, associateship or similar title of any university or institution.

Dr. Manojkumar Ramteke

Department of Chemical Engineering

Indian Institute of Technology Delhi

Acknowledgements

I owe a high debt to all those people who have directly or indirectly made this dissertation possible.

My journey to Ph.D. began with my advisor **Prof. Manojkumar Ramteke** at Indian Institute of Technology Delhi having an endless source of determined motivation and widespread knowledge. His visionary supervision and never-ending thoughts are the primary source of inspiration for me which helped to overcome the technical difficulties throughout my Ph. D. One of my motivation to work with him is his excitement towards solving new research problems (δx progress). He always shows me a big picture of my research, which helps me understand the specific aspect of my work. He did tremendous help throughout my thesis. There are lots of things which I had learnt from him like his positive attitude, work commitment, quality work, gentle, caring behavior. I always feel honored to present myself as his student.

I sincerely express my gratitude to my SRC committee, **Prof. Anurag Rathore, Prof. Shalini Gupta**, and **Prof. Anup Singh**. The expertise of the committee in biotechnology, biosensors, was always valuable. They always ready and give me their valuable feedbacks related to research, career and also on the presentation of my work in each semester. I have benefitted significantly from their suggestions and pieces of advice from time to time.

I would like to extend my thanks to past and present members of our group: **Dr. Vibhu Trivedi, Dr. Shiv Prakash, Dr. Nitish Ghune, Debashish Panda, Amrish Kumar, Feleke**

Bayu, Anubha for excellent discussion inside and outside the lab. I would also like to thanks **Salil, Shivendra, Tanmoy, Nilotpal, Sefali** from M. Tech and dual for fruitful discussion.

I would like to give a special thanks to my friends **Joyjit, Nalin, Pranjalya**, and **Sugeet** for technical discussion. I am much thankful to **Shabina, Abhijeet**, and **Kaushal** for supporting me in my course work. I am thankful to **Prateek** and **Khyati** for keeping me healthy by doing the workout on the ground and regular cycling trips. I am indebtedly thankful to everyone for anytime chai break including **Uma, Devendra, Neelam, Kamalika, Pooja, Gaurav, Anu, Sanchayan, Vandit, Rahul, Devshish**, and **Pawan** who have significantly improved my Ph.D experience. My special thanks to **Dr. Jogender Singh, Dr, Manish Jain, Dr. Dinesh Attarde, Dr. Dinesh Kumar**, and **Dr. Prasanta Kalita** for their guidance and regular support as elder brother. I would also love to thanks **Abhishek** and **Sneha** for sending me recent updates and research papers and **Dipender** for his consistent support.

I dedicate this thesis to my grandpa (Late) **Rameshwar Das** who always support and motivate me for study. His dream was to see me as a doctorate. My deepest gratitude to my family for continually encouraging me and believing in my capabilities. I am thankful to my parents **Roop Chand** and **Usha Rani**, for their persistent confidence in me. Without their love, care, patience, faith, and praise through the time of the thesis, this work would not have been possible. Special thanks to my brother **Abhishek**, sister-in-law **Poonam** for their love and encouragement and **Ankur** for his love, encouragement and proofreading. I am very thankful to exclusive members of my family **Pulkit, Kartik, Maysha** and **Dhanmay** for their cute smile. I am feeling enormous pleasure in expressing my gratitude to my father-in-law **Rajesh** and mother-in-law **Maya**. Finally, I show my special appreciation to my wife, **Reetu**, for her love and constant support, for all the late nights and early mornings tea, for proofreading, and for keeping me sane over the past few months. But most of all, thank you for being my best friend. I owe you everything.

Without you, I would not be the person I am today. I appreciatively acknowledge Indian Institute of Technology Delhi for providing me the facilities and an excellent academic environment to carry out my research work.

Abstract

Adleman's illustration of molecular computing using DNA paved a way towards entirely a new direction of computing. In DNA computing, the given combinatorial search space is represented in the form of DNA space. The uniqueness of Watson-Crick base pairing of DNA sequences leads to the formation of the long dsDNA molecules representing different model solutions for the given combinatorial problem. Since millions of DNA molecules are present in a small volume of DNA solution used, checking of all possible solutions is assured. The optimal solution is then extracted from the pool of DNA solutions using the biochemical processes. Thus, the exponential time complex combinatorial problem on a traditional computer essentially turns out to be a separation problem involving a polynomial number of steps in DNA computing experiments. Despite a promising concept, the implementations of existing DNA computing procedures were restricted only to smaller size formulations.

In the present study, a DNA computer aided with nearest neighbor heuristics and iterative implementation is developed to solve three short-term scheduling problems of multi-grade polymer plant involving multiple feasible solutions. Subsequently, a new DNA computing method is developed to obtain a significantly improved biochemical separation of correct dsDNA from the pool of dsDNA. For this, a multi-stage circular structure formation and digestion of dsDNA molecules is used. The new methodology is used to solve a bigger size Hamiltonian cycle problem involving 18 vertices. Further, the study is extended to design universal logic gates (AND, OR and NOT) using DNA. A simple Boolean logic circuit

involving all universal logic gates is solved using the developed DNA based logic gates which opens the possibility of solving large size computation using DNA based logic gates.

सारांश

डीएनए का उपयोग करके मॉलिक्यूलर कंप्यूटिंग के एडलमैन के चित्रण ने कंप्यूटिंग की पूरी तरह से एक नई दिशा का एक मार्ग प्रशस्त किया। डीएनए कंप्यूटिंग में, दिए गए कॉम्बिनेटरियल सर्च स्पेस डीएनए स्पेस के रूप में दर्शाया जाता है। डीएनए सीक्वेंसेस की वाटसन-क्रिक बेस पैरिंग की युनिकेनेस्स डी एस डीएनए मोलेक्युल्स का निर्माण करती है जो दिए गए संयोजन समस्या के लिए मॉडल समाधान है। चूंकि लाखों डीएनए अणु उपयोग किए गए डीएनए सोलूशन्स की एक छोटी मात्रा में मौजूद हैं, जो सभी संभावित समाधानों की जांच का आश्वासन देते हैं। ऑप्टीमल समाधान फिर बायोकैमिकल प्रोसेसेस (प्रक्रियाओं) का उपयोग करके डीएनए सोलूशन्स के पूल से निकाला जाता है। इस प्रकार, एक पारंपरिक कंप्यूटर पर एक्सपोनेंशियल टाइम काम्प्लेक्स कबिनटोरिअल प्रॉब्लम अनिवार्य रूप से एक सेपरेशन प्रॉब्लम बन जाती है जिसमे डीएनए कंप्यूटिंग एक्सपेरिमेंट्स पोलीनोमिअल नंबर ऑफ़ स्टेप्स है। एक आशाजनक अवधारणा के बावजूद, मौजूदा डीएनए कंप्यूटिंग प्रक्रियाओं का इम्प्लीमेंटेशन्स केवल छोटे आकार के योगों तक सीमित रखा गया था।

वर्तमान अध्ययन में, एक डीएनए कंप्यूटर नेअरेस्ट नेबर हेयरिस्टिक्स और इट्रेटिव इम्प्लीमेंटेशन की मदद से विकसित किया गया है जिससे मल्टी-ग्रेड पॉलीमर प्लांट की तीन शार्ट-टर्म स्केडुलिंग प्रोब्लेम्स का समाधान किया गया हैं जिसके कई संभव समाधान हैं। इसके बाद, एक नई डीएनए कंप्यूटिंग विधि विकसित की है जिसमे डी एस डीएनए को सिग्नीफिकांतली इम्प्रूव्ड बायोकैमिकल सेपरेशन विधि से डी एस डीएनए के पूल से

निकला गया हैं। इसके लिए, एक बहु-चरण सर्कुलर स्ट्रक्चर फार्मेशन और डी एस डीएनए अणुओं के डाईजेशन का उपयोग किया जाता है। नई पद्धति का उपयोग से बड़े आकार की 18 वर्टेक्स की हैमिल्टनियन साइकिल की समस्या को हल किया गया हैं। इसके अलावा, डीएनए का उपयोग करते हुए अध्ययन को लॉजिक गेट्स (एंड, और एवं नॉट) डिजाइन करने के लिए बढ़ाया गया है। एक साधारण बूलियन लॉजिक सर्किट जिसमें सभी लॉजिक गेट शामिल हैं को विकसित डीएनए आधारित लॉजिक गेट्स का उपयोग करके हल किया गया है जो डीएनए आधारित लॉजिक गेट्स का उपयोग करके बड़े आकार के कम्प्यूटेशन को हल करने की संभावना को खोलता है।

Table of contents

List of figures

List of tables

Nomenclature

$\% \ w/v$	Percent Weight/Volume
$^{\circ}C$	degree Celsius
$\wedge$	AND
$\neg$	NOT
$\overline{PO_i}$	reverse primer sequence corresponds to i^{th} order
$\overline{x_i}$	i^{th} variable of SAT problem represented by 0 (i = 1, 2, 3, . . . , n)
$\sigma(j)$	sequence of orders produced in unit j
$\vee$	OR
A	Adenine
a_i	i^{th} vertex of SAT problem (i = 1, 2, 3, . . . , n+1)
bp	Base pair
C	Cytosine
c	binary variable [c $\in$ 0, 1]

C_i i^{th} clause of SAT problem (i = 1, 2, 3, . . . , m)

$C_{i \to j}$ Edge, i = *current vertex, j = next vertex*

DNA P input DNA used to solve logic gate

DNA Q input DNA used to solve logic gate

DNA R DNA used as input to the Boolean circuit

DNA S DNA used as input to the Boolean circuit

DNA T DNA used as input to the Boolean circuit

DNA T_1 template DNA 1 used to solve logic gate

DNA T_2 template DNA 2 used to solve logic gate

DNA T_3 template DNA 3 used to solve logic gate

DNA T_4 template DNA 4 used to solve logic gate

DNA U DNA used as input to the Boolean circuit

DNA V Intermediate output of the Boolean circuit

DNA W Intermediate output of the Boolean circuit

DNA X Intermediate output of the Boolean circuit

DNA Y Intermediate output of the Boolean circuit

DNA Z Final output of the Boolean circuit

DNA Deoxyriiiibose nucleic acid

dNTP deoxyribose Nucleoside triphosphate

$dsDNA$ double stranded DNA

DT Due Time

$E(t,i,z)$ Extraction operator where t represents the sequences of tube, an i^{th} bit of variable string $x_1\ x_2\ x_3 \ldots\ x_n$ is equal to $z\ \in\ \{0,1\}$

E Edge

$EDTA$ Ethylenediaminetetraacetic acid

$EtBr$ ethidium bromide

F Forbidden sequence

G Guanine

$GC\ content$ Amount Guanine and Cytosine present in the DNA sequence

HPP Hamiltonian Path Problem

I initiator DNA sequence for solving maze

I_1^p Makespan

$l\ and\ \bar{l}$ loop and its complementary sequence for solving maze

$L\ and\ L_V$ length of DNA sequence

MFE minimum free energy

min time in minutes

$min(I_1)$ minimize the makespan

N number of vertices

N_E ssDNA sequences of all edges

N_V ssDNA sequences of all vertices

$n_{\sigma(j)}$ total number of orders to be produced in unit j with the sequence $\sigma(j)$

NP Non polynomial

O_i Order, $i = 1, 2, 3 \ldots N$

P_i position of the vertex (i = 1, 2, 3, . . . , n)

PCR Polymerase Chain Reaction

PO_i forward primer sequence corresponds to i^{th} order

PTU_{1-4} Processing Time of extruder 1 to extruder 4

RNA Ribonucleic acid

RT Release Time

s_t *and* $\overline{s_t}$ stem and its complementary sequence for solving maze

SAT Satisfiability Problem

sec time in seconds

$ssDNA$ single stranded DNA

T Thymine

t_h *and* $\overline{t_h}$ toehold and its complementary sequence for solving maze

$t_h^{'}$ *and* $\overline{t_h^{'}}$ initiator and its complementary sequence for solving maze

$t_{i,j}^{c}$ sequence-dependent changeover time

$t_{i,j}^{p}$ Processing time of order, *i*, in processing unit j

t_{i}^{r} Start release time of i^{th} order

t_{i}^{s} Start release time of i^{th} order

$t_{i}^{'}$ $t_{i-1} - t_{i}$ (i = 1, 2, 3, . . . , n)

TBE Tris Borate EDTA

TSP Travelling Salesman Problem

U_i Extruder, $i = 1, 2, 3 \ldots N$

V Vertex

v variable DNA [v ∈ A, T, G, C]

V_f Final (f^{th}) vertex

V_i Starting (i^{th}) vertex

V_i^c i^{th} vertex having value c [(i = 1, 2, 3, . . . , n) and c ∈ 0, 1]

x_i i^{th} variable of SAT problem represented by 1 (i = 1, 2, 3, . . . , n)

Y Hairpin DNA corresponds to vertex for solving maze

Y_{En} Entry DNA sequence for solving maze

Y_{Ex} Exit DNA sequence for solving maze

Z Hairpin DNA corresponds to edge for solving maze