

A NOVEL TECHNIQUE FOR THE CONVERSION  
OF DIGITAL DATA INTO DNA SEQUENCE

# DATA TO DNA

---



GIRIK MALIK



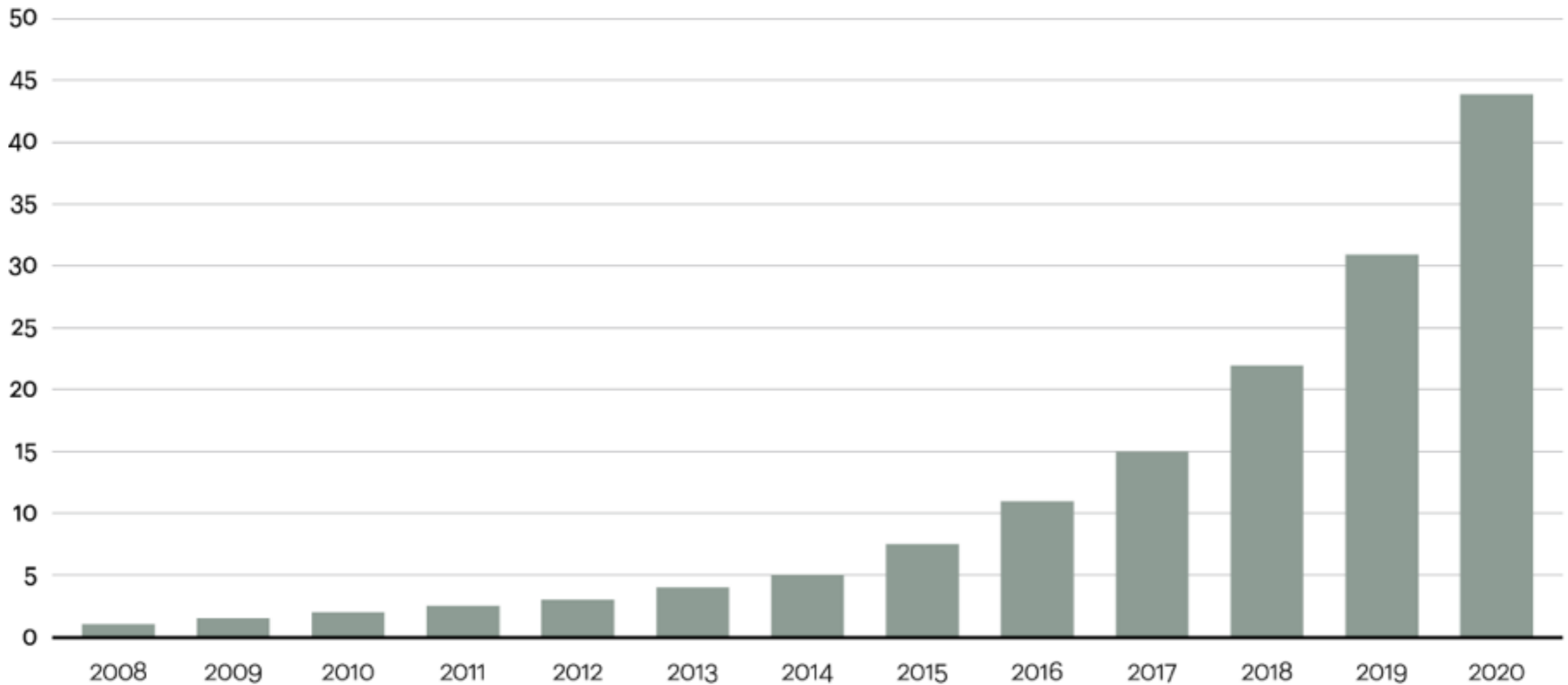
THE OHIO STATE UNIVERSITY  
COLLEGE OF MEDICINE

# MOTIVATION

Figure 1

**Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020**

## Data in zettabytes (ZB)



Source: Oracle, 2012

## PROBLEM

- ▶ We are generating data at an astronomical rate !
- ▶ Seagate alone has shipped > 2 billion hard drives
- ▶ > 30 billion devices connected to the Internet
- ▶ The amount of storage per unit area is reaching limits
- ▶ Hard drives store 1 million times than at their inception
- ▶ Squeeze more data in same space

## WHY DNA ?

- ▶ DNA can be stored for a longer period of time
- ▶ No maintenance cost
- ▶ Secure and effective
- ▶ Store more in less space
- ▶ Can store passwords, classified documents and information

## NIBBLE – OUR TECHNIQUE

- ▶ Nibble : *half-a-byte*
- ▶ Stores everything in 4 bases of DNA
- ▶ Combination of 4 bases give 256 unique codes
- ▶ Map ASCII table

## POINTER APPROACH

- ▶ Stores data in sequentially and random manner
- ▶ Cuts down synthesis cost
- ▶ Maps data to less than 25% of organisms DNA (demonstrated on E.Coli)

## WORLD DATA CONVERTED TO DNA

Estimated World data by 2020	Number of Characters in world data	Number of bases of DNA covering world data	Weight of DNA bases covering world data (Daltons)	Weight of DNA bases covering world data (grams)
45 ZB = $4.5 \times 10^{22}$ bytes	$4.5 \times 10^{22}$	$1.8 \times 10^{23}$	$5.94 \times 10^{25}$	98.63601185

## CLAIMS

- ▶ Ideal for long-term data storage/archival
- ▶ Can be used for data security
- ▶ Can store entire world data by 2020 in less than 99g DNA
- ▶ Storage density 467 PB/g
- ▶ Can eliminate the cost of synthesis, retrieving data by mere sequencing
- ▶ Robust - retrieve entire data even if the DNA strand is lost



## LIMITATIONS

- ▶ Slow sequencing rate
- ▶ Synthesis too expensive to work for routine activities
- ▶ Storing world data based on a statistical probability
- ▶ Dependent on faster and cheaper technology
- ▶ Too futuristic !



## AFTERMATH



OTC #: \_\_\_\_\_

**Intellectual Property Disclosure Form****(Instructions are provided on the last page)**

A.	Title of Intellectual Property:	<p><b>DNA based Data Storage: A step towards more robust data storage</b></p> <p><b>Biological Molecule based Data Storage: A step towards more robust data storage</b></p>
B.	Suggested Keywords:	<p><b>DNA, Data Storage, Password, Cryptography, Big Data, Data Warehouse, Data Archival, Pointer based Storage, Addressing based Storage, DNA disk, Protein based Storage, DNA to protein, Data to Protein, Translation</b></p>
C.	Description of Intellectual Property	<p><b>General purpose</b> <i>Is the Intellectual Property a new process, composition of matter, a device, or one or more products? A new use for, or an improvement to, an existing product or process?</i></p> <p><b>Yes the Intellectual Property is a new process solving the problem of Massive and Big Data Storage</b></p> <p><b>Technical description</b> <i>A detailed description that will be a primary source of information for the patent attorney as an application is being prepared. (Please attach a separate sheet - ATTACHMENT A)</i></p> <p><b>Problem Solved</b> <i>What are the advantages and improvements over existing methods, devices or materials? What features are believed to be new?</i> <b>Both 'long term' and 'big data' storage demand novel storage solutions. Current storage platforms and approaches are not scalable due to immense demands on the space, cost and energy involved in maintaining big data servers. The pointer based Data storage is something new which provides even more robust data storage with an option of recovering back all the data based on just a pointer file, even if the mapping sequence is lost.</b></p>
D.	<p>Attach photocopies of original crucial documentation of ideas and data, e.g. lab notes, diagrams, notes from phone conversations, etc and maintain all original documents. (ATTACHMENT B)</p>	

# PATENT

Google Patents



Sign in

SEARCH TERMS

SEARCH FIELDS



MORE ^








## A biomolecule based data storage system

### Abstract

The present invention describes a biomolecule based storage system for converting, storing the data in DNA coded form and retrieving data using pointer file approach. User input data is converted into 4base DNA sequence, called Nibble, which is further mapped onto the DNA sequence of an organism. The first position of each converted nibble is then obtained and stored in a pointer file. By mapping the positions of pointer file onto the DNA sequence of the organism, the data can be retrieved.

### Classifications

**G06N3/12** Computer systems based on biological models using genetic models

[View 1 more classifications](#)

### Description

#### FIELD OF INVENTION

[0001] The present invention relates to data storage system, particularly storing data in a naturally occurring or synthetically created biomolecule such as but not limited to Deoxyribonucleic acid (DNA), Ribonucleic acid (RNA), proteins, primary metabolites, secondary metabolites, their complexes and other combinations.

BACKGROUND OF THE INVENTION [0002] Computer data is continuously growing in

WO2016059610A1

WO Application

Find Prior Art

**Application number:** PCT/IB2015/057964

**Other languages:** [French](#)

**Inventor:** [Girik MALIK](#), [Pawan K. DHAR](#)

**Original Assignee:** [Malik Girik](#), [Dhar Pawan K](#)

**Priority date:** 2014-10-18

**Filing date:** 2015-10-16

**Publication date:** 2016-04-21

**Info:** [Non-patent citations \(3\)](#), [Legal events](#), [Similar documents](#)

**External links:** [Espacenet](#), [Global Dossier](#), [PatentScope](#), [Discuss](#)

### Claims (1)

1. WE CLAIM:

1) A biomolecule based data storage system, comprising:

an E.coli Master DNA file, said file containing physical DNA sequence of E.coli;

an ASCII map having 256 characters and 256 combinations of 4-base DNA sequence, said 4-base combination is called a Nibble;

# PATENT

▶ Application published on WIPO

▶ Application No. WO2016059610

▶ International Application No. PCT/IB2015/057964

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(10) International Publication Number  
**WO 2016/059610 A1**

(43) International Publication Date  
21 April 2016 (21.04.2016)

(51) International Patent Classification:  
G06F 19/10 (2011.01) G06V 3/72 (2006.01)

(21) International Application Number:  
PCT/IB2015/057964

(22) International Filing Date:  
16 October 2015 (16.10.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
2975/DEL/2014 18 October 2014 (18.10.2014) IN

(71) Inventors; and  
Applicants : MALIK, Girik [IN/IN]; RZ-112, Manas Kunj, Mangal Bazaar, Uttam Nagar, New Delhi 110059 (IN). DHAR, Pawan K. [IN/IN]; 301, Tower 10, Purvanchal Heights, Uttar Pradesh, Sector Zeta 1, Greater Noida, Greater Noida 201306 (IN).

(74) Agent: RANA, Vikrant; S. S. Rana & Co., Advocates, Patent & Trademark Attorneys, 317, Lawyers' Chambers, High Court of Delhi, New Delhi 110003 (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:  
— as to the identity of the inventor (Rule 4.17(i))

[Continued on next page]

(54) Title: A BIOMOLECULE BASED DATA STORAGE SYSTEM



(57) Abstract: The present invention describes a biomolecule based storage system for converting, storing the data in DNA coded form and retrieving the data using pointer file approach. User input data is converted into 4base DNA sequence, called Nibble, which is further mapped onto the DNA sequence of an organism. The first position of each converted nibble is then obtained and stored in a pointer file. By mapping the positions of pointer file onto the DNA sequence of the organism, the data can be retrieved.

Fig. 1: Process of conversion of Data to DNA and pointer

WO 2016/059610 A1

**THANK YOU VERY MUCH!**