DNA Data Storage and Cryptography

by Anna Vinnedge

United States Military Academy, West Point, NY

Department of Mathematical Sciences

Abstract

In this poster we are giving an overview of DNA cryptography, a new area of research in data storage and security. We will provide definitions, descriptions, and examples related to the topic.

Motivation

- The typical storage systems that currently hold the worlds data in the form of 0's and 1's will not be able to keep up. For this reason, finding ways to store data efficiently has become an increasingly relevant problem.
- One potential alternative for data storage is DNA-based data storage.
- With the use of a DNA-based data storage system, it is essential that we develop means of encryption and decryption that can work under this new medium in order to keep data secure.

Introduction

Recent research has opened up the cyber world to the possibility of harnessing DNA as a medium for very large scale computations, data storage, and encryption and decryption techniques. DNA, made up of two inter-woven sub strands composed of two molecules is routinely being sequenced and accurately copied. It is also very stable making it a promising medium for the safe storage of data. A single gram of DNA has the potential to store a single zettabyte (One billion terabytes) and also has the potential to be very useful in the encryption and decryption of data. Allowing the nucleotides to represent binary values allows theoretical algorithms to be derived for encrypting and decrypting data.

DNA Data Storage

DNA Structure

DNA, or deoxyribonucleic acid, is the genetic material in humans and all other living things. DNA is composed of two polynucleotide chains that coil around each other in a double helix formation. Each polynucleotide is composed of smaller units each made with one of four nitrogen containing nucleobases: cytosine (C), guanine (G), adenine (A), or thymine (T)

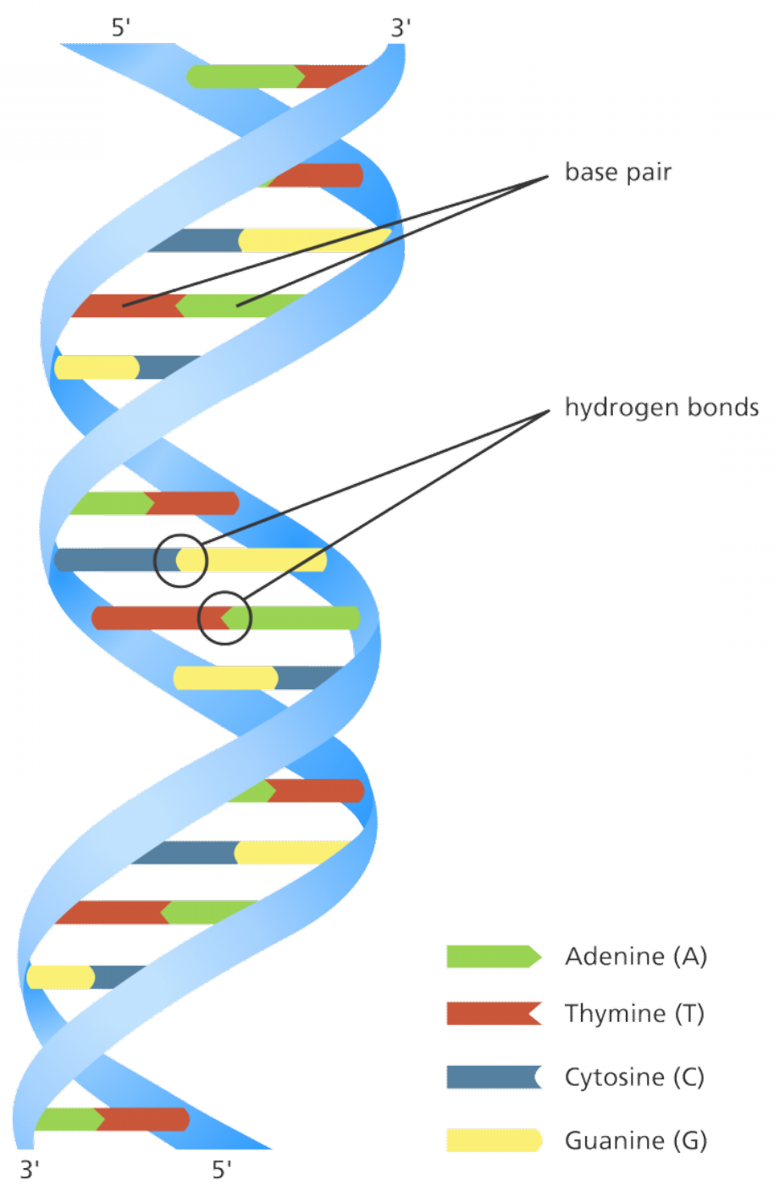


Figure 1: Structure of DNA

Binary Structure of DNA Data Storage

Researchers are harnessing these pattern based strands to look into ways to store data. By allowing each nucleobase to represent a two element binary value, DNA can become representative of an XOR table to be utilized as a bitwise arithmetic operator. For the purposes of DNA storage and DNA cryptography, the following XOR digits are represented by the following nucleotides:

00=A
01=T
10=C
11=G

In order to read data stored in DNA, it is sequenced in the same way a human genome is sequenced and each nucleotide is then converted to its binary representation. Encoding DNA to represent data is done through DNA synthesizing.

DNA Cryptography

Cryptography Concept

Cryptography is the the area of research that focuses on protecting information and communication using codes by encryption and decryption. Encryption uses codes to secure a message and decryption uses codes to reveal an original message.

Most cryptographic algorithms employ either a *public key* (symmetric) or a *public key* (assymetric). With symmetric encryption, the key for encryption and decryption is the same and that key must be kept secret. With assymetric cryptography, the key used for encryption is different from the key used for decryption which allows for the keys to stay public.

DNA as a Cryptographic Medium

DNA cryptography is a new proposed technique of encryption and decryption in which data can be hid within a DNA sequence. This technique is very promising due to the powerful potential of DNA computing and data storage. Simply hiding data within DNA is not enough to keep it secret, but applying known as well as new ciphers to data already sequenced into DNA shows great promise

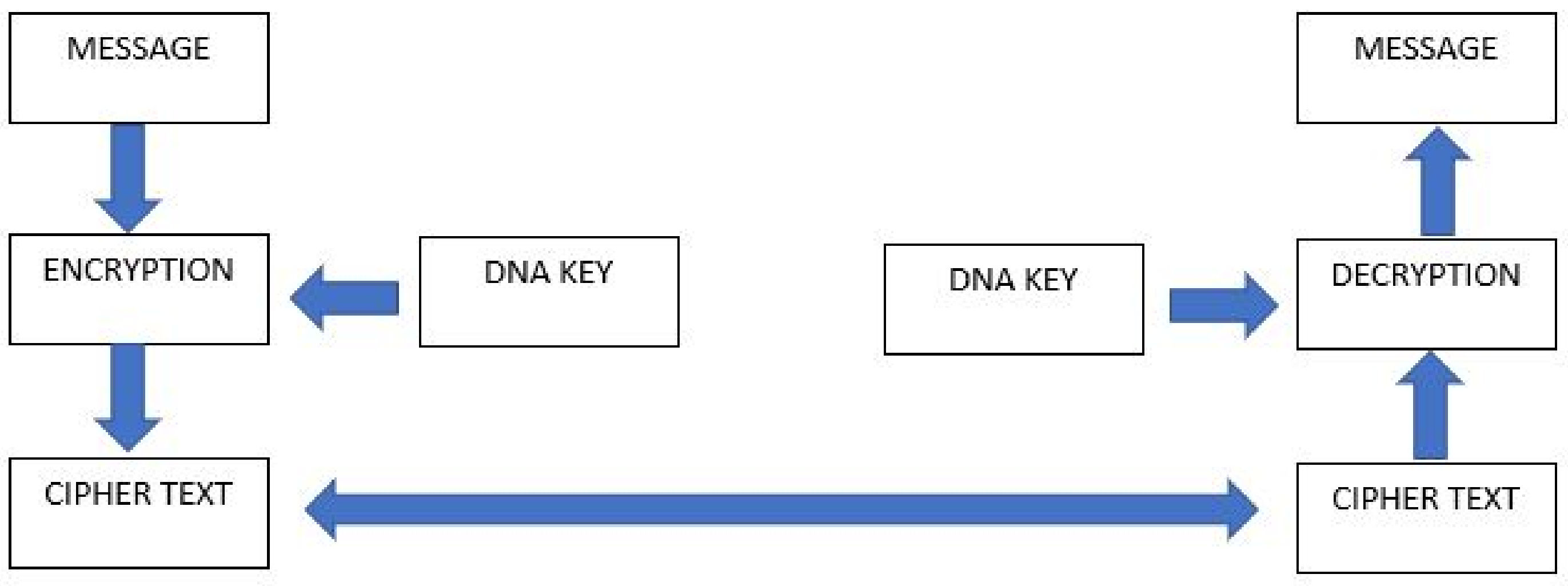


Figure 2: Basic flow of DNA cryptography

Simple Algorithms

One of the simplest and earliest encryption methods is the Caesar cipher, which is a shift cipher in mod 26 (to represent each lower case letter of the alphabet).

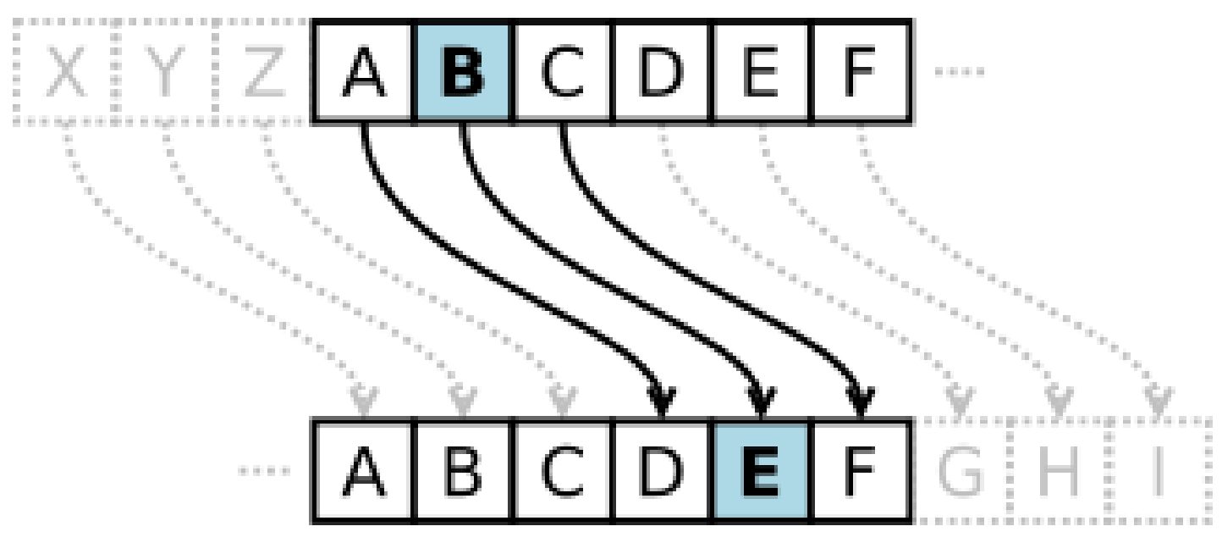


Figure 3: Alphabetic Shift Cipher

In order to make this algorithm applicable to DNA, the shift would occur when the message was still in alphabetic form but that shifted message would then be converted to binary representations in ASCII format. After converted to ASCII, DNA would be synthesized, representative of the message using the nucleotide to binary pairing described earlier.

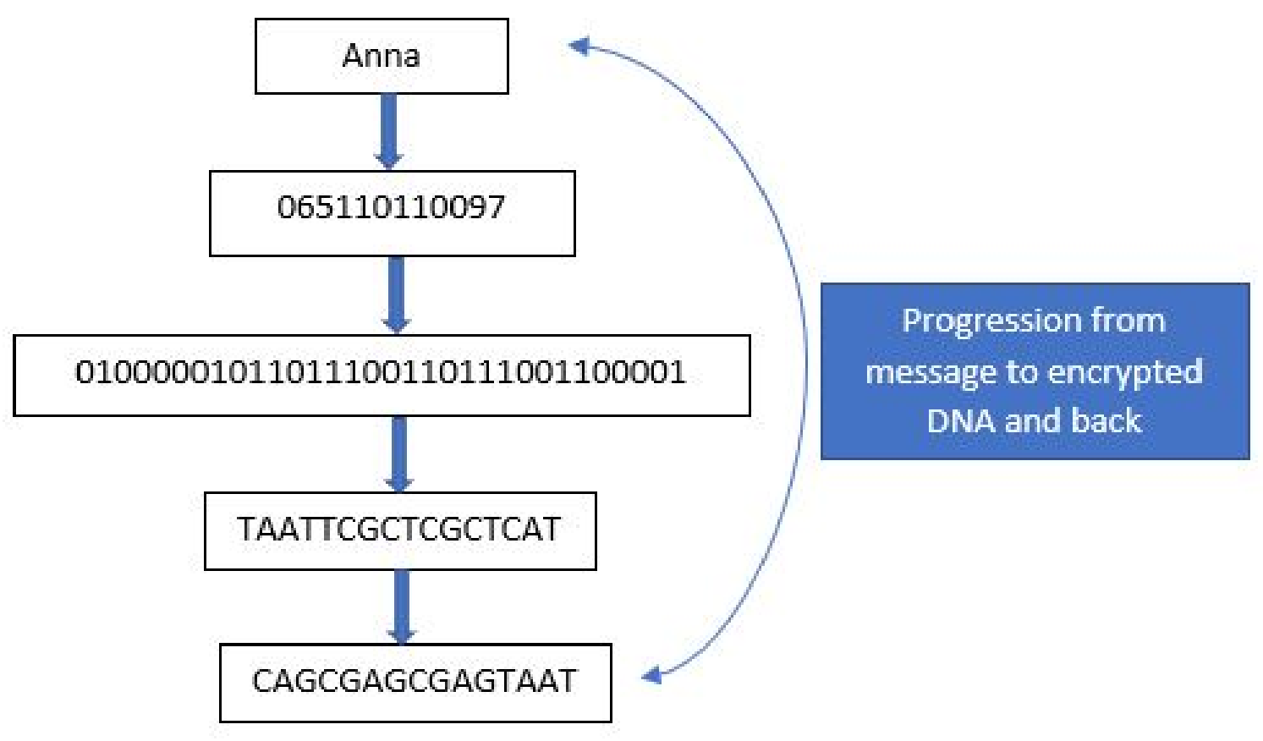


Figure 4: Example of DNA Shift Cipher

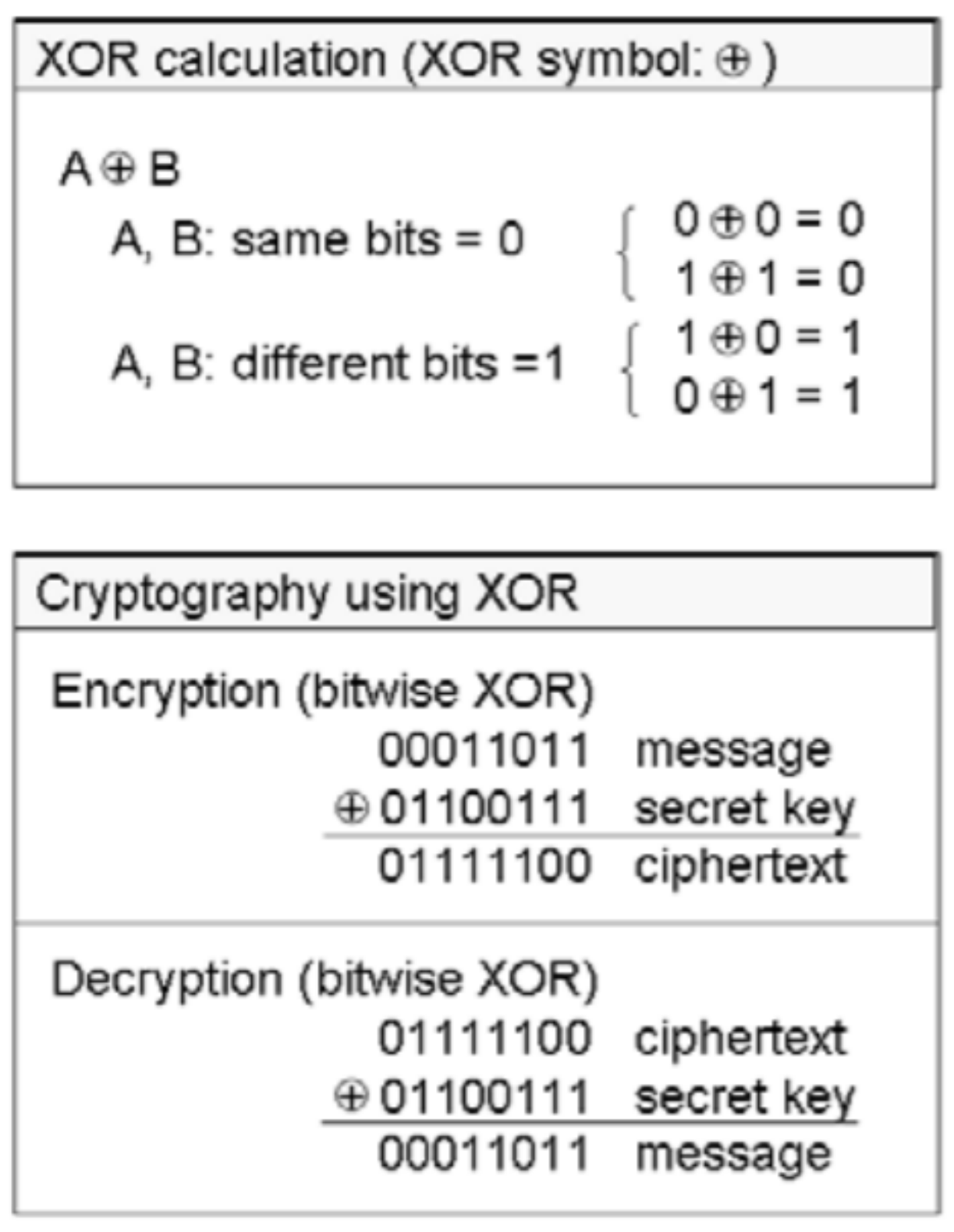
One Time Pad

OTP is the only mathematically unbreakable encryption. It works using a key that is as long as the message it is encrypting and the key is truly random.

To apply this to DNA cryptoraphy, let p represent the plaintext, k represent the key, and c represent the ciphertext. So, we have that

$$p \text{ XOR } k = c$$
$$c \text{ XOR } k = p$$

To implement this process using DNA, the process of DNA self assembling takes place to generate a random key.



XOR calculation (XOR symbol: ⊕)

A ⊕ B

A, B: same bits = 0 { 0 ⊕ 0 = 0, 1 ⊕ 1 = 0

A, B: different bits =1 { 1 ⊕ 0 = 1, 0 ⊕ 1 = 1

Cryptography using XOR

Encryption (bitwise XOR)

```
  00011011  message
⊕ 01100111  secret key
  01111100  ciphertext
```

Decryption (bitwise XOR)

```
  01111100  ciphertext
⊕ 01100111  secret key
  00011011  message
```

Figure 5: XOR Operation for OTP

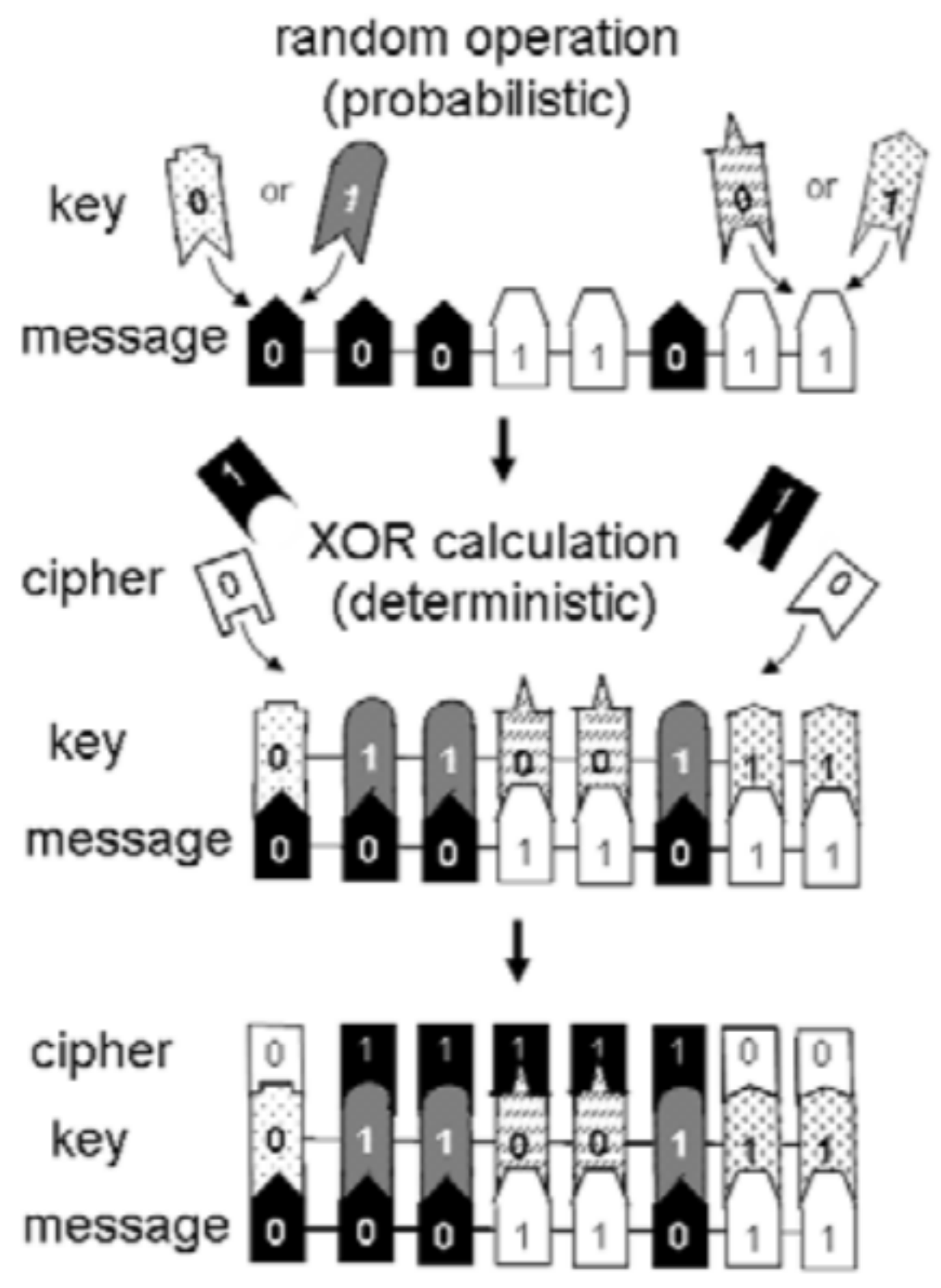


Figure 6: Secret Key Generation Using Physical Random Process of DNA Assembly

Prior to the discussion of DNA data storage and encryption, OTP was not a probable cryptosystem due to the massive amount of data and computing power it would take to use. One proposed method of OTP using DNA is to use a DNA database as a codebook which can construct a mapping table with unique characteristics for inverse mapping. Essentially, several DNA sequences will be made up of evenly distributed random bit sequences, which, using logical XOR operations and biological properties of the nucleotide strands, can be constructed.

References

[1] Miki Hirabayashi, Kazuhiro Oiwa, and Hiroaki Kojima. Design of true random one-time pads in dna xor cryptosystems. *Proceedings in Information and Communications Technology*, 2.

[2] Mandrita Mondal and Kumar S. Ray. Review on dna cryptography. *Electronics and Communication Science Unit*, 48.

[3] Rechael Rettner. Dna: Definition, structure discovery. *LiveScience*.

[4] Margaret Rouse. *SearchSecurity*.

[5] Federico Tavella, Alberto Giaretta, Triona Marie Dooley-Cullinane, Mauro Conti, Lee Coffey, and Sasitharan Balasubramaniam. Dna molecular storage system: Transferring digitally encoded information through bacterial nanonetworks. *IEEE Transactions on Emerging Topics in Computing*.

[6] Sang Yup. Dna data storage is closer than you think. *Scientific American*.