Error Correction Schemes for DNA Storage with Nanopore Sequencing

Laura Conde-Canencia, Belaid Hamoum, Dominique Lavenier, Emeline Roux

Error Correction Schemes for DNA Storage with Nanopore Sequencing

Laura CONDE-CANENCIA[1], Belaid HAMOUM[1] and Dominique LAVENIER[2]

[1] Lab-STICC, CNRS UMR 6285, Université Bretagne Sud, 56100, Lorient, France

[2] GenScale, INRIA, Campus Beaulieu, 35042, Rennes, France

Corresponding Author: laura.conde-canencia@univ-ubs.fr

1. Introduction

DNA storage is an emerging technology that uses DNA molecules to store data. This type of storage system is much more compact than any other due to the data density of the DNA. Moreover the capability for longevity and for resistance to obsolescence of DNA is undeniable: DNA is a universal and fundamental data storage mechanism in biology. For these and other reasons, DNA used as a memory-storage material in nucleic acid memory products promises a viable and compelling alternative to electronic memory. The exponential decrease in DNA synthesis costs should make the technology cost-effective for long-term data storage within about ten years. Several DNA-based storage systems have reported since 2012 [CHU12] [YON13] [GRA15] [ZHI16] [BOR16]. Companies such as Microsoft are leading research on this topic [MIC15] and have already announced their plan to use DNA storage in their data centers by 2020 [MIC17].

2. Motivation and goals of our work

The objective of our work is to design coding schemes allowing information to be efficiently encoded on DNA molecules, and to be read back using very low cost sequencing devices based on nanopore technology. The first step of our work develops coding techniques targeting the nanopore constraints in order to reduce the error performance of the global storage system. The second part demonstrates the feasibility of the approach by (1) synthetizing DNA molecules encoded with the proposed coding schemes; (2) reading the information by sequencing the DNA molecules with a nanopore device; (3) applying error detection and error correction techniques to the output signal to retrieve the initial information.

3. The MinION technology

We consider the Oxford Nanopore MinION device as it is currently the portable solution that offers ultra-long reads and a very reasonable cost. The MinION weighs under 100 g and can be plugged into a PC or a laptop using a high-speed USB 3.0 cable. It uses biomolecular nanopores from which electrical signals are detected when DNA molecules go through them. By interpreting these signals, DNA molecules can be deciphered. However, the nanopore technology presents high error rates (~10-15%), and so does the MinION. The challenge of our project is to provide efficient base-caller tools for correcting errors inherent to that technology. Our approach takes advantage of the DNA storage principle that allows for coding techniques to ensure robust decoding.

Acknowledgements

The authors would like to thank the Labex Cominlabs for funding this research project.

References

1. [CHU12] Church, G. M.; Gao, Y.; Kosuri, S. (2012), "Next-Generation Digital Information Storage in DNA" Science. 337 (6102): 1628.
2. [YON13] Yong, E. (2013). Synthetic double-helix faithfully stores Shakespeare's sonnets. Nature.
3. [GRA15] Grass, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. (2015). "Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes". Angewandte Chemie International Edition. 54 (8): 2552.
4. [ZHI16] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church and W. L. Hughes, “Nucleic acid memory”, NATURE MATERIALS | VOL 15 | APRIL 2016 Online: http://arep.med.harvard.edu/pdf/Zhirnov_NAM_2016.pdf
5. [BOR16] James Bornholt, (2016). A DNA-Based Archival Storage System. ASPLOS '16 Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems.