

A new approach to organize, edit and add data for DNA-based data storage

Jorge Eduardo Guerrero^{1,2}, Afsaneh Sadremomtaz^{1,2} and Reza Zadegan^{1,2,}*

1. Department of Nanoengineering, NC A&T State University, Greensboro, NC, USA

2. Joint School of Nanoscience and Nanoengineering, Greensboro, NC, USA

* Corresponding author email: rzadegan@ncat.edu

With the expanding data storage capacity needs, DNA as an alternative to the archival storage medium offers potential advantages, including higher density and data retention for information storage^{1,2}. However, the majority of DNA-based memory systems are write-once and read-only, although few studies have suggested overwriting digital data on the existing DNA using chemical modifications of bases³. Using those strategies requires constantly updating the entire data coding and iteratively synthesizing the DNA pool. Therefore, considering the complexity and cost, those methods needed some amendments to become industrially scalable. Inspired by magnetic tapes⁴ and multisession-CD⁵, in this work, we created a DNA storage system coined the Molecular File System (MolFS), to organize, store, and edit digital information in a DNA pool. MolFS uses DNA pools that consist of multiple sessions, where each session contains data block and unique index sections to store and edit the files. We used indexes to describe the file system hierarchy, locate files along with the blocks, recognize the sessions, and identify the file versions. This approach reduces the editing cost compared to the state-of-the-art methods, and editing or adding data requires only synthesizing a new DNA pool containing the DNA session of the differential file. As proof of concept, we encoded 2.3 Kbytes of graphic and text data into 2 DNA pools. To edit the existing DNA pool, we added 8 new differential data blocks to existing pools, reaching 13.8 Kbytes of data stored from sessions 1 to 5. We performed nanopore sequencing and recovered the data from the MolFS sessions accurately and precisely.

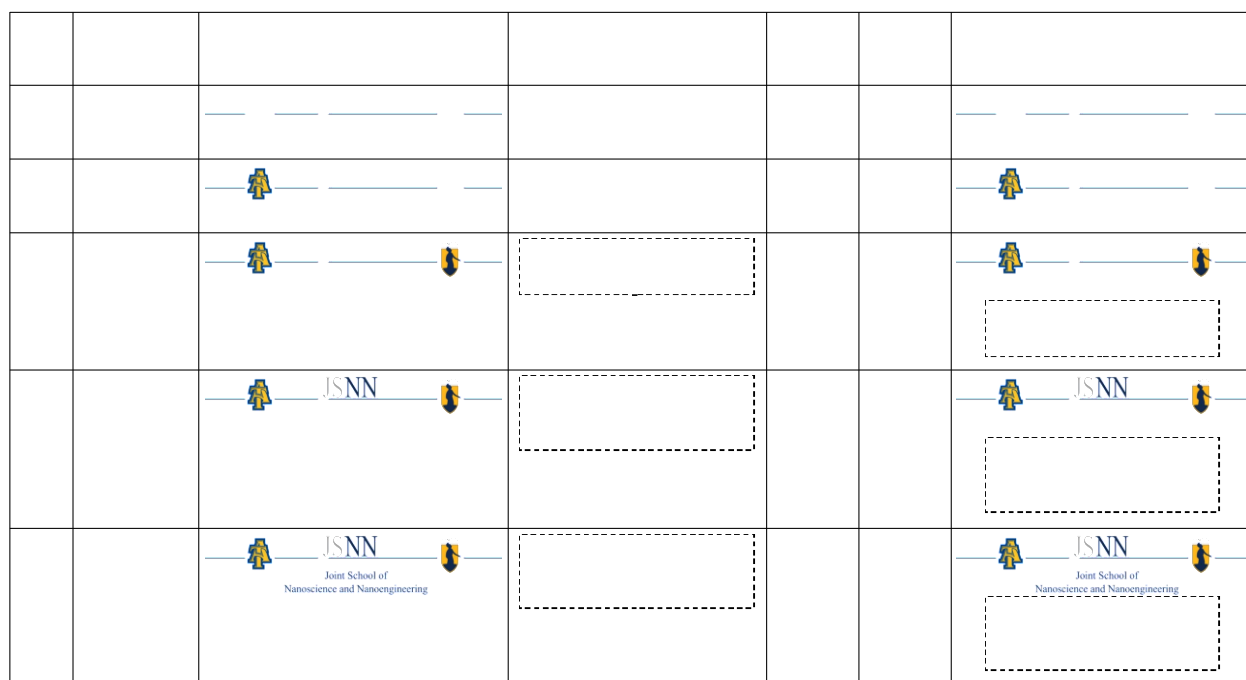


Figure 1. Molecular File System. We created a Molecular File System containing multiple sessions in the DNA pools. For each session we synthesized DNA pools to encode the index and data blocks, and through five sessions the MolFS enabled us to create folders and insert and edit files.

This work is supported through the National Science Foundation award # MCB 2027738.

References

1. Zhirnov, V., Zadegan, R. M., Sandhu, G. S., Church, G. M. & Hughes, W. L. Nucleic acid memory. *Nature Materials* vol. 15 366–370 (2016).
2. Gantz, J. & Reinsel, D. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. *Idc* (2012).
3. Tabatabaei Yazdi, S. M. H., Yuan, Y., Ma, J., Zhao, H. & Milenkovic, O. A Rewritable, Random-Access DNA-Based Storage System. *Sci. Rep.* 5, 14138 (2015).
4. Pease, D. *et al.* The linear tape file system. *2010 IEEE 26th Symp. Mass Storage Syst. Technol. MSST2010* 4, (2010).
5. Optical Storage Technology Association. Universal Disk Format Specification. (1995).