Architecting Datacenters for Sustainability: Greener Data Storage using Synthetic DNA

Bichlien H. Nguyen*[1], Julie Sinistore[2], Jake. A Smith[1], Praneet S. Arshi[2], Lauren M. Johnson[2], Tim Kidman[2], T.J. DiCaprio[1], Doug Carmean[1], Karin Strauss*[1]

[1] Microsoft, Redmond, USA
[2] WSP Inc., USA
* Corresponding Authors, bnguy@microsoft.com and kstrauss@microsoft.com

Abstract

Global digital data generation has been growing at a breakneck pace. Although not all generated data needs to be stored, a non-trivial portion does. Synthetic deoxyribonucleotide acid (DNA) is an attractive medium for digital information storage. If kept under appropriate conditions, DNA can reliably store information for thousands of years [1]. It also has a practical estimated density of 1 Exabyte per cubic inch, which is much higher than commercial data storage media.

Buildings, infrastructure, electronic computing, storage, and networking equipment, and other physical resources all contribute to the environmental impacts, particularly, emissions, energy and water consumption, and waste generation of digital data storage. DNA data storage has the potential to limit these impacts by drastically reducing the resources required to maintain very large volumes of data.

In this paper, we describe how to store digital information in synthetic DNA, present a cradle-to-grave life cycle assessment (LCA) of archival DNA data storage, and compare the resulting environmental impacts with those of traditional hard disk drives (HDDs) and tape storage based on greenhouse gas (GHG) emissions, energy usage, and blue water consumption (BWC). We conclude that DNA shows promise when compared to HDDs and tape, and we follow that conclusion with a discussion of how further innovation in biotechnology could be used to improve the sustainability of future datacenters.

1 Introduction

The rate of digital information generation far outpaces increases in our capacity to store it. According to IDC, the "Global DataSphere" (all digital data generated globally) is expected to grow from 44 Zettabytes (10^{21} bytes) in 2020 to 175 Zettabytes in 2025 (approximately 32% per year) [2]. IDC predicts that about 10% of all data will be stored, and 49% of that will be in public clouds. This will result in demand for over 8 Zettabytes of storage. Currently, magnetic media such as HDDs and tape are used for a large percentage of long-term cloud data storage. However, these technologies may be unsuitable for the world's increasing storage requirements.

Two notable data storage evaluation metrics are data density and durability. Figure 1 compares storage technologies: each bar represents the data density of each storage technology broken into recent volumetric data density information and projected volumetric data density (based on limitations in scaling practical storage systems). The bottom text shows the typical durability for each technology. Tape storage, currently the densest commercial storage medium at a demonstrated density of over 37 Gigabits/mm^3 [3], also has the best durability. Unsurprisingly, tape is typically used for archival storage. However, data stored on tape still needs to be rewritten onto new media every few years, which can take days. Millions of cartridges would be needed for the 8 Zettabytes of data that IDC predicts will be stored in public clouds by 2025. Other commercial storage technologies could be used to store this data, but that may only aggravate the problem because they are either less dense or durable than tape.

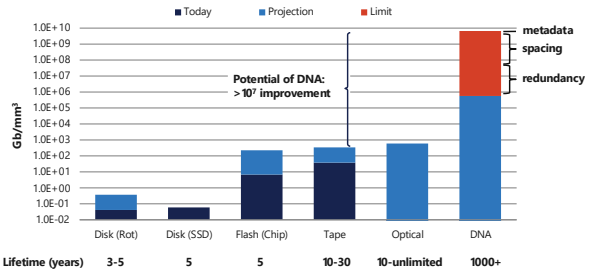

Figure 1: Comparison of data capacity and durability across data storage technologies.

DNA has been investigated as an alternative archival data storage medium [4]–[7]. As shown in Figure 1, DNA has a theoretical density of over 1 Exabyte/mm^3 (i.e., 1,000,000,000 Gigabyte/mm^3) and durability on the order of hundreds to thousands of years — both of which make it quite attractive for this application. A practical implementation requires overheads such as metadata, including object identifiers, addresses, and logical redundancy for error correction, physical spacing and physical redundancy. Fortunately, even at this lower effective density, DNA still offers a clear advantage over other commercial media at an estimated density of over 1 Exabyte/in^3.

The main environmental impacts of data storage libraries result from the physical buildings, storage equipment, infrastructure, data access, and environmental control necessary for their operations. Each of these elements produces GHG emissions, energy and water consumption, and waste disposal burdens. The significant gains in density and durability from DNA data storage should thus result in a lower environmental burden.

Past DNA data storage research has covered the implementation of improved system architectures [5], [8], preservation techniques [9], [10], and automation [11], [12]. However, DNA's sustainability aspects have remained unexplored until now. In this work, we performed a cradle-to-grave life cycle analysis (LCA) of a hypothetical full production DNA data storage system (using multiple chemistry options) and compared it with other storage technologies on three metrics: GHG emissions, energy use, and water consumption. Our findings suggest that DNA data storage could have lower environmental impacts than both HDD and tape storage.

2 DNA Data Storage Basics

DNA 101

A deoxyribonucleotide acid (DNA) strand (also known as an oligonucleotide) is a linear polymer composed of sequences of four natural nucleotides (A, C, G, T) that are commonly referred to as bases. In nature, two complementary strands are typically paired into DNA's famous double-helix structure. It is possible to predict interactions between DNA strands in a double helix based on their sequences due to base-pairing interactions (A binds to T, and G to C). From an information storage perspective, since the interactions are known, the information on one strand of the double helix is redundant with its complementary strand.

Storing data in DNA

DNA data storage is a method for storing digital information in synthetic DNA strands. Storing information in DNA starts with converting a sequence of bits into a sequence of nucleotides. Figure 2a shows a simple example mapping between bits and bases — every two bits in a sequence are translated into one of the four nucleotide types. For example, in Figure 2b, the binary string 01101100 maps to the DNA sequence CGTA. Although appropriate for illustration, such a simple mapping is rarely used because synthetic DNA is prone to errors (base deletions, insertions, and substitutions, and missing sequences). In practice, encodings are more sophisticated, and error-correcting algorithms are often employed to improve the system's robustness [1], [6]. Once the sequences that represent the bits to be stored are determined, the next step is to create the molecules that represent the sequences through a process called DNA synthesis.

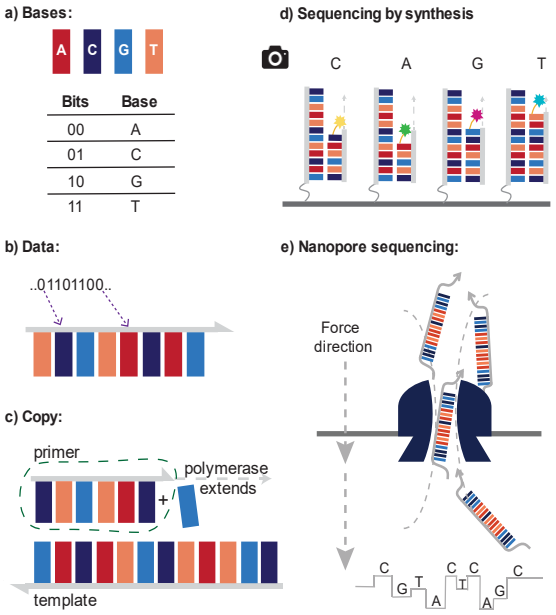

Figure 2: DNA basic overview. Fig. 2a shows the four nucleotide bases (A, C, G, T) and a simple mapping of bits to bases. Fig. 2b provides an example of encoding data in DNA. Fig. 2c highlights the ability to copy the DNA using polymerase chain reaction (PCR). Fig. 2d provides an overview of how DNA is read using sequencing by synthesis, where incorporated bases are fluorescently labelled. Fig. 2e shows a schematic of how DNA can be read using a nanopore sequencer.

DNA synthesis

Synthesis of *de novo* oligonucleotides has traditionally been performed through a process called standard synthesis or phosphoramidite synthesis. This process occurs in cycles composed of four complex chemical steps for the addition of a single nucleotide [13]. These steps are: (1) deblocking, which enables the next base to attach, (2) addition, which adds a blocked base (i.e., a chemically modified base that prevents additional

bases from attaching), (3) oxidation, which strengthens the newly formed bond, and (4) capping, which adds a group that prevents further strand growth where addition has not happened in the current cycle. Standard chemistry relies heavily on organic reagents, such as acetonitrile, which can be volatile, flammable, and toxic. Despite these challenging handling issues, standard synthesis has been used in the biotechnology industry for the past 40 years due to its maturity as a process [14].

Enzymatic synthesis is a nascent alternative approach based on using an enzyme called terminal deoxynucleotide transferase (TdT) to add bases to an existing DNA strand. Although much less mature, in recent years this type of synthesis has generated interest as a promising method for *de novo* DNA synthesis [15]–[17]. Enzymatic DNA synthesis is expected to require fewer steps, and it is performed predominantly in aqueous salt buffers that mimic biological pH — an easy-to-handle solvent.

DNA replication

DNA can be easily replicated through a process called polymerase chain reaction (PCR), as shown in Figure 2c [18]. In this reaction, a short priming oligonucleotide (a short strand of DNA, about 20 nucleotides in size) binds to the template DNA strand to be replicated. A polymerase (i.e., an enzyme that "completes" nucleotides missing in the double helix) then extends that priming sequence and "fills in the blanks" by incorporating complementary bases as it zips along the template to form double-stranded DNA. This reaction can be repeated many times to generate the desired number of molecular replicas.

DNA sequencing

The two most common techniques used for reading DNA are sequencing-by-synthesis (e.g., Illumina sequencing instruments) and nanopore devices (e.g., Oxford Nanopore Technologies sequencing instruments). Both use aqueous buffers, so their handling is easier than in standard synthesis.

In sequencing-by-synthesis, illustrated in Figure 2d, the DNA strand of interest serves as the template for a polymerase to create a complementary strand using fluorescent base types and blocking groups to ensure one base addition per cycle. Each fluorescent base emits at different wavelengths and can be monitored and distinguished optically during that cycle [19]. The fluorescent group is cleaved at the end of the cycle, allowing sequencing to proceed to the next cycle.

In nanopore sequencing, illustrated in Figure 2e, the strand of interest is pulled through a voltage-gated nanochannel. As the strand is translocated through the channel, bases perturb the current differently to generate a unique current signature that can be inferred to map back into those specific bases [20].

Putting it all together: DNA data storage system

A DNA data storage system uses the multistep process outlined in Figure 3. First, since there are limitations on the length of synthetic DNA sequences (about 150 to 300 bases in length is typical today), data to be stored in DNA is partitioned into smaller pieces (about 15 to 30 bytes) before being mapped into DNA sequences, given a sequence number, or index, to identify their position in the original file, and augmented with additional error correction information. These bit sequences are then mapped to sequences of the four DNA nucleotides.

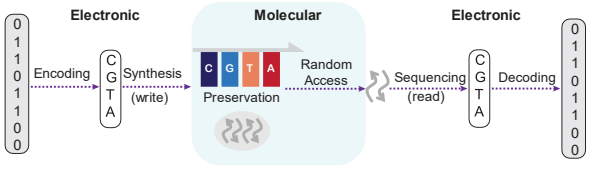

Figure 3: Overview of DNA data storage system.

Once data has been encoded into DNA bases, the sequences are written into physical DNA oligonucleotide strands through standard or enzymatic synthesis, typically using a 2D array platform that creates multiple unique DNA sequences in parallel in a single synthesis run. Array synthesis may employ fluidic deposition, photolithographic, and electrochemical synthesis techniques [13].

After synthesis, the oligonucleotides on the array are removed from the surface and pooled to create a complex mixture of DNA strands. Each DNA pool may contain multiple files and inherently does not provide spatial isolation of the data. The pools are deposited in a DNA library, which is then spatially organized and addressed, so multiple pools can be stored on the same substrate.

To retrieve a file stored in DNA, the pool is physically retrieved from its library, and the file is accessed using PCR, which selectively copies the DNA oligonucleotides that encode the data sequences to be recovered. PCR's selectivity is accomplished by assigning different primer sites to each of the files stored in a pool and later using complementary primers associated to the file to be read in the PCR reaction [8]. The molecules are then sampled and sequenced, and the data is error corrected and decoded back into the original file.

3 Life Cycle Assessment

In this section, we compare the environmental impacts of traditional archival data storage media (HDD and tape) with DNA-based storage media. We conducted

this analysis through a screening-level LCA to identify potential areas of concern. We also performed a cradle-to-grave analysis using GaBi LCA software and quantified GHG emissions, energy, and BWC for each storage media.

3.1 Model assumptions

Functional unit

To accurately compare the LCA results, we defined a common functional unit across the storage media. Because all the products store data over time, we defined our unit of comparison, or functional unit, as 1 TB of data stored for a year with a read rate of 10% (100 GB) for that year (note that this comparison is for archival storage, so read performance was not a factor). We included a 2% data sensitivity case to test how variations in data access patterns affect the results.

Geographical boundary

Infrastructure for energy, waste treatment, and other processes vary across regions and can significantly impact LCA results, so we standardized our model using United States data, including for grid energy and material production.

System boundary

The cradle-to-grave assessment covered three main stages for each storage media: (1) production, (2) use, and (3) end-of-life. Figure 4 outlines process-flow diagrams for each storage type. For production, we inventoried the raw materials required to fabricate each type and its associated manufacturing inputs (such as heat, power, water, and chemicals). For use, we modeled the energy necessary for writing, maintaining, and accessing a 1TB functional unit of archival storage. We assumed that the datacenter building infrastructure would be equivalent across all storage methods and limited it to what was directly required for storage. This was a conservative assumption: datacenter physical infrastructure would likely be lower for DNA data storage due to its higher data density. For end-of-life treatment and destruction, we used shredding and incineration as the disposal methods for HDDs and tape, and incineration as the disposal method for DNA. Since water is used as an input for the DNA synthesis process, we also included wastewater treatment for DNA data storage.

Impact categories

For this LCA, we quantified GHG emissions, BWC, and energy as the impact categories. We quantified GHG emissions using the IPCC AR5 characterization factor for GWP100, excluding biogenic carbon ($kgCO_2eq$). Blue water refers only to surface and groundwater and excludes rainwater. Water consumption is the portion of water use that is not returned to its original water source and cannot be reused after withdrawal (e.g., water lost via evaporation or water incorporated into a product or plant). We used the GaBi BWC characterization method to quantify blue water (liters). Primary energy was quantified by net calorific value (MJ) otherwise known as low heating value depending on dataset availability.

Data sources

Primary data sources for DNA production, storage, reading, and end-of-life were projections of future production systems extrapolated from current systems. We sourced primary HDD and tape data from literature and primary product information from online resources as detailed below. We sourced secondary data inputs, or life cycle inventory (LCI) data, from GaBi professional and electronics databases (service pack 39) [21] and the Ecoinvent version 3.5 database (with temporal coverage of 2018-2019) [22].

3.2 HDD and tape storage

We modeled HDD and tape storage from cradle-to-grave to match the system boundary of the DNA archival storage system (Figure 4a) and derived HDD production impacts from an existing manufacturer's study that included materials used, distribution, manufacturing, and packaging [23]. LCI data for data storage tape does not exist, so we modeled tape production impacts by tearing down a commercially available tape product (LTO Ultrium 8 Data Cartridge with a compressed data storage capacity of 30TB) and manually identifying and weighing its components. We determined the HDD-use phase assumptions for writing the data and accessing 10% of it annually using the product's energy specs (with the assumption that the disks are not spun down). The tape use phase included the energy to write the tape and access 10% annually using the energy and the product's performance specs and an assumption that, once written, the tape would need to be rewound 48 times per year. For both, we also considered additional energy consumption for servers and networking equipment and media power usage effectiveness.

3.3 DNA data storage scenarios

Given that DNA data storage has not been implemented in a datacenter and its full deployment requirements cannot be anticipated, we approached the LCA by pinpointing the most significant components of the DNA data storage process (shown in Figure 3). We identified synthesis and sequencing as the major contributors to production impacts (Figure 4b).

DNA synthesis volumes

Total volumes and types of chemical reagents required to store 1 TB of data in DNA vary depending on the exact manufacturing processes (acetonitrile versus

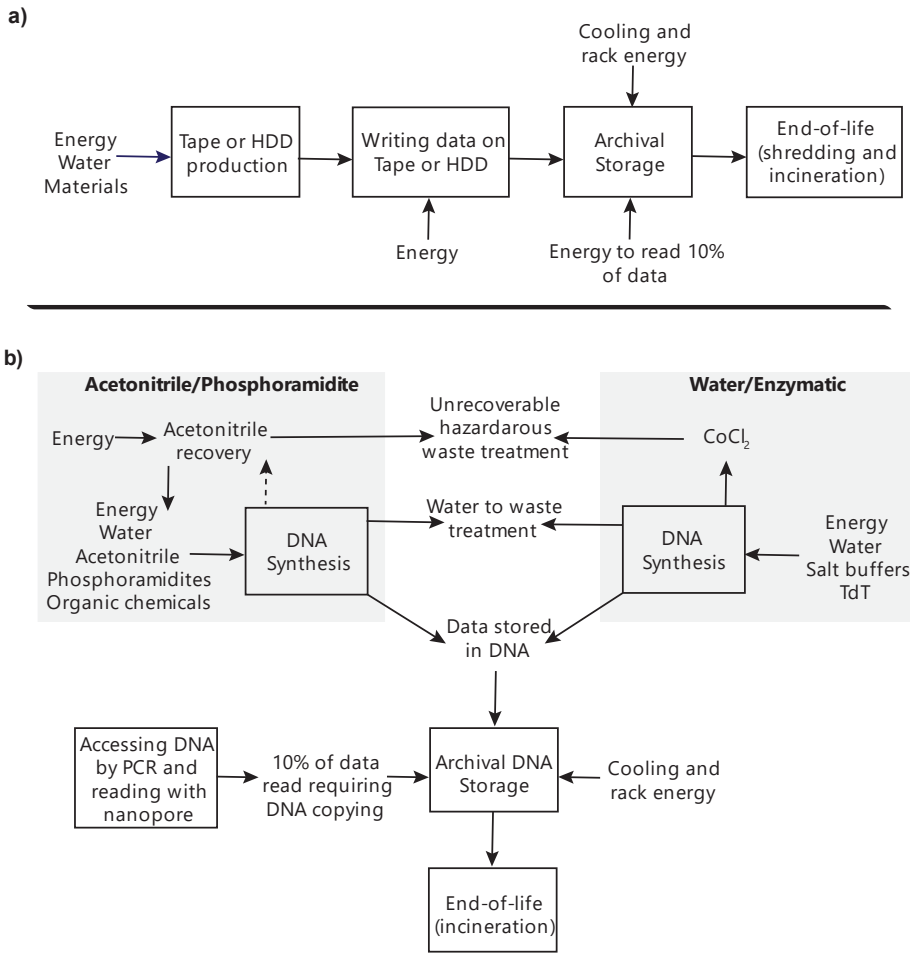

Figure 4. Process flow diagrams for the cradle-to-grave LCA of the different storage media. Boxed text represents. Unboxed text represents inputs. Boxed text represents life-cycle phases. Fig. 4a shows inputs and outputs for storing 1 TB of data and reading 100 GB a year with tape or HDD. Fig. 4b delineates the process of storing 1 TB of data and reading 100 GB a year with DNA. The flow diagram begins with the inputs and outputs for either phosphoramidite/acetonitrile-based DNA synthesis or TdT/enzymatic DNA synthesis and converges once the data has been written into DNA.

enzymatic synthesis) and assumed parameters, so we modeled worst- and best-case scenarios for each DNA synthesis method to account for that. We modeled the worst-case scenario as a synthesis process in which the reagents for each chemical input are used once and then discarded. In the best-case scenario, 90% of reagents could be reused throughout the synthesis run.

For standard DNA synthesis, we incorporated two acetonitrile production methods into the LCA for comparison. We considered acetonitrile derived from propane found in fossil fuels (conventional acetonitrile) and bio-acetonitrile derived from ethanol (bio-acetonitrile). We assumed enzymatic DNA synthesis to be an aqueous system.

Retrieving data in DNA

We assumed the use of the PCR method to selectively amplify the DNA strands encoding the 10% of the data to be read from the 1 TB DNA pool. Once retrieved from the pool, we assumed the strands would be sequenced with a nanopore device.

3.4 LCA results

A tale of two acetonitriles in standard synthesis

Figure 5a shows the cradle-to-grave LCA results from comparing the two types of acetonitrile used for synthesis against the GHG emissions, energy, and water consumption dimensions. The LCA comparison between conventional acetonitrile and bio-acetonitrile production (not shown) did not result in a clear best. We expected bio-acetonitrile would be significantly more sustainable than conventional acetonitrile; however, while bio-acetonitrile production did reduce GHG emissions and energy consumption by 79% and 85% respectively (compared to conventional acetonitrile production), it increased BWC by 84%.

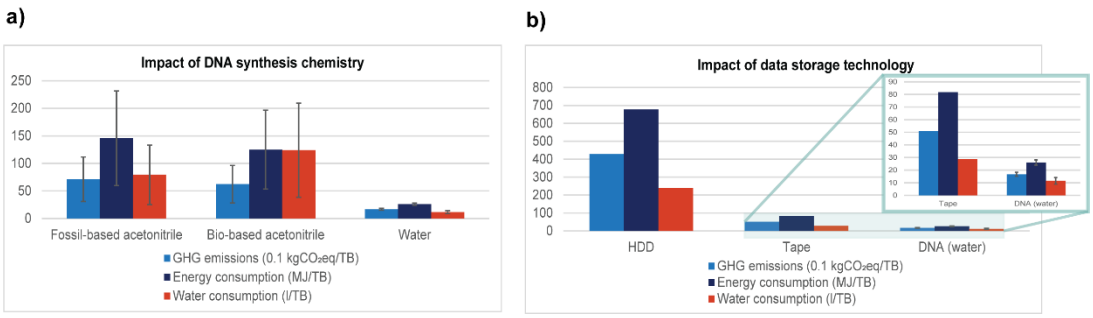

Figure 5. Cradle-to-grave LCA results estimating GHG, energy utilization, and water consumption. Fig. 5a summarizes the impacts of three different DNA synthesis techniques for DNA data storage. Fig. 5b provides an overall comparison of HDD, tape, and DNA data storage (water/enzymatic synthesis).

Unsurprisingly, the primary contributor to water consumption in the LCA for bio-acetonitrile DNA synthesis was bio-acetonitrile production. The water consumption increase was a direct result of higher water inputs for ethanol production from biomass. For conventional acetonitrile, water consumption resulted only from the production of miscellaneous input chemicals to the synthesis process.

For energy consumption, the production of the four base phosphoramidites used to create the oligonucleotides was the main contributor to both conventional and bio-acetonitrile DNA synthesis.

Standard vs. enzymatic synthesis

Enzymatic DNA synthesis utilizes fewer chemicals than phosphoramidite synthesis, and it is performed primarily in aqueous neutral buffered solutions (i.e., saltwater). Since it does not use acetonitrile, enzymatic synthesis has lower GHG emissions, energy consumption, and water usage than either conventional or bio-acetonitrile DNA synthesis.

As expected, direct water consumption during the DNA synthesis process drove the water metric. Breaking down the drivers for GHG emissions and energy consumption, we found the production of the salts (e.g., tris-acetate) used in the buffered water to be the primary contributor.

Overall HDD, Tape, and DNA comparison

Of the storage media evaluated in this study, HDD storage has the highest GHG emissions, energy consumption, and water usage, as shown in Figure 5b. The main driver of the environmental impacts was the use-phase energy demands of storing 1 TB of data with a 10% access rate. This includes the power required for spinning the disks, accessing and reading the data, and cooling the system. Tape, by comparison, has a significantly lower impact than HDDs even though the main contributor to environmental impacts was also use-phase energy demands.

The environmental impacts of DNA data storage are less clear-cut and depend heavily on the method used to manufacture the DNA. DNA data storage using standard synthesis may be more sustainable than existing HDD and tape storage when closer to its best-case scenario. In its worst-case scenario, it falls short of expectations. In contrast, DNA data storage with enzymatic synthesis appears to significantly reduce environmental impacts across all storage types and metrics, regardless of best- or worst-case scenario assumptions.

3.5 Limitations in analysis

While both HDD and tape are commercially available storage technologies, they lack readily available inputs and existing datasets; therefore, we modeled them using literature, product teardowns, and assumptions. GaBi databases lacked a significant number of DNA data storage components and processes. Mini-LCAs had to be completed for each lifecycle phase, including phosphoramidite DNA synthesis and enzymatic DNA synthesis and sequencing. Within these sub-LCAs, many of the inputs lacked LCI data and had to be modeled with proxies serving as functional counterparts or manufactured in similar processes. An example is the use of amylase production as a proxy for TdT in enzymatic synthesis. Both are enzymes, and large-scale enzyme manufacturing generally involves the fermentation of microorganisms engineered to produce the desired enzyme [24]. Since amylase manufacturing is done at scale, we selected the results from an amylase LCA as the most appropriate approximation for TdT in the quantities needed for DNA data storage.

Using data proxies and assumptions limits the results' applicability to future at-scale rollouts of these technologies. Future work will include the modeling of reagents and other important DNA synthesis inputs and consider archival-type HDD storage applications. At the time of this paper's writing, we are actively creating LCAs for each input proxy to incorporate into future assessments.

4 Discussion and Conclusion

We expect that the next few decades will see a great deal of innovation and improvement in datacenter sustainability. As we transition from the digital revolution to the fourth industrial revolution (a fusion of physical, digital, and biological technologies), biotechnology advances promise to have great impact on datacenters.

DNA data storage could both increase the density and durability of archival storage systems and be more environmentally sustainable than existing storage media. Though the future impacts of DNA data storage will largely depend on improvements in DNA synthesis and sequencing technologies, in this paper we demonstrated the potential of DNA storage to improve GHG emissions, energy use, and water consumption.

We foresee that optimizing phosphoramidite chemical reagent consumption for standard DNA synthesis or reaching a mature enzymatic DNA synthesis process will reduce DNA data storage's environmental footprint. The fact that DNA is inherently biological also enable a new end-of-life disposal method: the DNA could be biodegraded.

Beyond DNA data storage, we believe biotechnology has the potential to address other datacenter sustainability needs. New concrete mixtures can reduce the carbon emissions associated with datacenter construction — biomaterials, sand and bacteria compositions, and graphene reinforcement are all solutions under development [25], [26].

Another area of interest is identifying greener methods to power datacenters, such as clean biofuels derived from biomass. The renewable nature of these fuels may help prevent the release of previously sequestered carbon and eliminate the emission of unhealthy, volatile organic compounds and sulfur compounds typical of fossil fuels [27], [28].

With datacenters shifting toward higher levels of circularity, we expect electronic components to be harvested from their original boards for redirection to their highest possible value use. However, the remaining printed circuit board substrates with custom metal tracks are still likely to become physical waste. To address this, we foresee a future in which such boards are biodegraded. In this scenario, biodegradable materials are combined with biological or synthetic biology-based technologies so the metals present in the degraded boards can be easily scavenged. E-waste composting could become a reality.

These examples illustrate that biotechnology may help address the multiple sustainability challenges currently faced by datacenters. Though some of these technologies are not yet fully mature (or have even surpassed proof of concept), it is important to discuss them now so that their development can be nurtured and expedited. Early success in these areas can unlock significant paradigm shifts in datacenter sustainability.

5 Acknowledgements

We thank Alessandra Pistoia, Elizabeth Willmott, and other members of the Microsoft Sustainability Team for their support and guidance on sustainability topics. We thank Winston Sanders, Anand Narasimhan, Nicholas Keehn, Yuan-Jyue Chen, and Luis Ceze for enlightening discussions and their valuable feedback on this work. We thank Marc Steuben for proofreading this manuscript.

6 References

[1] R. N. Grass, R. Heckel, M. Puddu, D. Paunescu, and W. J. Stark, "Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes," *Angewandte Chemie International Edition*, vol. 54, no. 8, pp. 2552–2555, 2015, doi: 10.1002/anie.201411378.

[2] "Where in the World Is Storage: Byte Density Across the Globe." IDC, 2013, [Online]. Available: http://www.idc.com/downloads/where_is_storage_ infographic_243338.pdf.

[3] "Sony Develops Magnetic Tape Technology with the World's Highest Recording Density." Sony, 2014, [Online]. Available: http://www.sony.net/SonyInfo/News/Press/201404/14-044E/.

[4] G. M. Church, Y. Gao, and S. Kosuri, "Next-Generation Digital Information Storage in DNA," *Science*, vol. 337, no. 6102, pp. 1628–1628, Sep. 2012, doi: 10.1126/science.1226355.

[5] N. Goldman *et al.*, "Towards Practical, High-capacity, Low-maintenance Information Storage in Synthesized DNA," *Nature*, vol. 494, no. 7435, pp. 77–80, Feb. 2013, doi: 10.1038/nature11875.

[6] Y. Erlich and D. Zielinski, "DNA Fountain Enables a Robust and Efficient Storage Architecture," *Science*, vol. 355, no. 6328, pp. 950–954, Mar. 2017, doi: 10.1126/science.aaj2038.

[7] L. Organick *et al.*, "Random Access in Large-scale DNA Data Storage," *Nat Biotechnol*, vol. 36, no. 3, pp. 242–248, Mar. 2018, doi: 10.1038/nbt.4079.

[8] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, "A DNA-Based Archival Storage System," *SIGPLAN Not.*, vol. 51, no. 4, pp. 637–649, Jun. 2016, doi: 10.1145/2954679.2872397.

[9] W. D. Chen *et al.*, "Combining Data Longevity with High Storage Capacity—Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles," *Adv. Funct. Mater.*, vol. 29, no. 28, p. 1901672, Jul. 2019, doi: 10.1002/adfm.201901672.

[10] A. X. Kohll *et al.*, "Stabilizing Synthetic DNA for Long-term Data Storage with Earth Alkaline Salts,"

Chem. Commun., vol. 56, no. 25, pp. 3613–3616, 2020, doi: 10.1039/D0CC00222D.

[11] S. Newman *et al.*, "High Density DNA Data Storage Library via Dehydration with Digital Microfluidic Retrieval," *Nat Commun*, vol. 10, no. 1, p. 1706, Dec. 2019, doi: 10.1038/s41467-019-09517-y.

[12] C. N. Takahashi, B. H. Nguyen, K. Strauss, and L. Ceze, "Demonstration of End-to-End Automation of DNA Data Storage," *Scientific Reports*, vol. 9, no. 1, p. 4998, Mar. 2019, doi: 10.1038/s41598-019-41228-8.

[13] S. Kosuri and G. M. Church, "Large-scale de novo DNA Synthesis: Technologies and Applications," *Nat Methods*, vol. 11, no. 5, pp. 499–507, May 2014, doi: 10.1038/nmeth.2918.

[14] S. L. Beaucage and M. H. Caruthers, "Deoxynucleoside Phosphoramidites—A New Class of Key Intermediates for Deoxypolynucleotide Synthesis," *Tetrahedron Letters*, vol. 22, no. 20, pp. 1859–1862, Jan. 1981, doi: 10.1016/S0040-4039(01)90461-7.

[15] S. Palluk *et al.*, "De novo DNA Synthesis Using Polymerase-nucleotide Conjugates," *Nat Biotechnol*, vol. 36, no. 7, pp. 645–650, Aug. 2018, doi: 10.1038/nbt.4173.

[16] H. H. Lee, R. Kalhor, N. Goela, J. Bolot, and G. M. Church, "Terminator-free Template-independent Enzymatic DNA Synthesis for Digital Information Storage," *Nat Commun*, vol. 10, no. 1, p. 2383, Dec. 2019, doi: 10.1038/s41467-019-10258-1.

[17] S. Barthel, S. Palluk, N. J. Hillson, J. D. Keasling, and D. H. Arlow, "Enhancing Terminal Deoxynucleotidyl Transferase Activity on Substrates with 3′ Terminal Structures for Enzymatic De Novo DNA Synthesis," *Genes*, vol. 11, no. 1, p. 102, Jan. 2020, doi: 10.3390/genes11010102.

[18] L. Garibyan and N. Avashia, "Polymerase Chain Reaction," *Journal of Investigative Dermatology*, vol. 133, no. 3, pp. 1–4, Mar. 2013, doi: 10.1038/jid.2013.1.

[19] "An Introduction to Next-Generation Sequencing Technology." Illumina, [Online]. Available: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.

[20] D. Deamer, M. Akeson, and D. Branton, "Three Decades of Nanopore Sequencing," *Nat Biotechnol*, vol. 34, no. 5, pp. 518–524, May 2016, doi: 10.1038/nbt.3423.

[21] *GaBi Software System and Database for Life Cycle Engineering 1992-2020*. .

[22] G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, and B. Weidema, "The Ecoinvent Database Version 3 (Part I): Overview and Methodology," *Int J Life Cycle Assess*, vol. 21, no. 9, pp. 1218–1230, Sep. 2016, doi: 10.1007/s11367-016-1087-8.

[23] "Exos X12 Sustainability Report." Seagate, [Online]. Available: https://www.seagate.com/global-citizenship/product-sustainability/exos-x12-sustainability-report/.

[24] R. Carlson, "On DNA and Transistors," 2016. http://www.synthesis.cc/synthesis/2016/03/on_dna_and_transistors.

[25] R. Singh, M. Kumar, A. Mittal, and P. K. Mehta, "Microbial Enzymes: Industrial Progress in 21st Century," *3 Biotech*, vol. 6, no. 2, p. 174, Dec. 2016, doi: 10.1007/s13205-016-0485-8.

[26] M. Seifan, A. K. Samani, and A. Berenjian, "Bioconcrete: Next Generation of Self-healing Concrete," *Appl Microbiol Biotechnol*, vol. 100, no. 6, pp. 2591–2602, Mar. 2016, doi: 10.1007/s00253-016-7316-z.

[27] H. M. Jonkers, A. Thijssen, G. Muyzer, O. Copuroglu, and E. Schlangen, "Application of Bacteria as Self-healing Agent for the Development of Sustainable Concrete," *Ecological Engineering*, vol. 36, no. 2, pp. 230–235, Feb. 2010, doi: 10.1016/j.ecoleng.2008.12.036.

[28] J. Sheehan, V. Camobreco, J. Duffield, M. Graboski, and H. Shapouri, "An Overview of Biodiesel and Petroleum Diesel Life Cycles." National Renweable Energy Laboratory, 1998, [Online]. Available: https://www.nrel.gov/docs/legosti/fy98/24772.pdf.

[29] A. Kumar, D.-S. Kim, H. Omidvarborna, and S. K. Kuppili, "Combustion Chemistry of Biodiesel for Use in Urban Transport Buses: Experiment and Modeling." Mineta National Transit Research Consortium, 2014, [Online]. Available: https://transweb.sjsu.edu/sites/default/files/1146-biodiesel-bus-fuel-combustion-chemistry.pdf.