nature materials

SUPPLEMENTARY INFORMATION

DOI: 10.1038/NMAT4594

Nucleic acid memory

Victor Zhirnov, Reza M. Zadegan, Gurtej S. Sandhu, George M. Church and William L. Hughes

Victor Zhirnov is at the Semiconductor Research Corporation, 1101 Slater Road, Durham, North Carolina 27703, USA. Reza M. Zadegan and William L. Hughes are in the Department of Materials Science and Engineering, Boise State University, 1910 University Drive, Boise, Idaho 83725-2090, USA. Gurtej S. Sandhu is at Micron Technology, Inc., PO Box 6, 8000 South Federal Way, Boise, Idaho 83707-0006, USA. George M. Church is in the Department of Genetics, Harvard University, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. e-mail: willhughes@boisestate.edu; gmc@harvard.edu

Table of contents

DOI: 10.1038/NMAT4594

S1. Storage Needs in 2040

Global trends in information storage are based on research by Hilbert and Lopez,[1] whose research provided a detailed inventory of all storage media and its use. A summary of their findings is shown in Figure 1a, from which several observations can be made: (A) the majority of data was analog before 2002, (B) analog data reached a maximum around 2000 and steadily decreased afterwards, (C) the amounts of stored analog and digital data became equal around 2002, and (D) after 2007, the vast majority of information became digital—a trend that continues today.

Extrapolation of the digital line in Figure 1a provides projections of global data storage. An important caveat is that the growth rate of digital storage (red line) is considerably higher than the total growth rate (blue line). This reflects the fact that analog data has dominated total storage capacity for the majority of the measurement period. Extrapolation of the digital line in Figure 1a is an upper bound, while the extrapolation of the total storage capacity is a conservative estimate. Figure 1b shows global memory demand projections — both a conservative estimate and an upper bound. These extrapolations are compared to independent estimates of the data storage in 2010–2014 (solid red dots) and projected storage needs in 2020 and 2030 (open red dots) in Figure 1b. All estimages and projectsions are within the defined boundaries (formed by the "conservative" and "upper bound" lines).

As indicated in Figure 1b, future information and communication technologies are expected to generate enormous amounts of data, far surpassing today's data flows. As a reference, about 10^{22} bits of data were stored worldwide in 2014. In addition, global memory demand is estimated to exceed 3×10^{24} bits by 2040 (Figure 1b). By this conservative estimate, the total mass of silicon wafers required for flash memory would exceed silicon supply (see section S2), and would be equal to 3×10^{24} bits × 1.63×10^{-12} g/bit = 2.4×10^{9} kg (see section S4).

S2. Global Silicon Wafer Supply Trend

The data for global silicon wafer supply was acquired from reference [2] and shows a linear growth between 1990 and 2015 (Figure S1). Figure S2 is an extrapolation of this data into the future, which projects that silicon wafer production will be ~2.4×10^{7} kg in 2040.

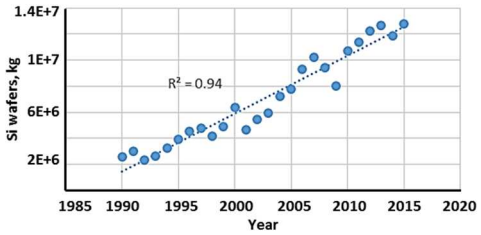

Figure S1 | Global Si wafer supply between 1990 and 2015.

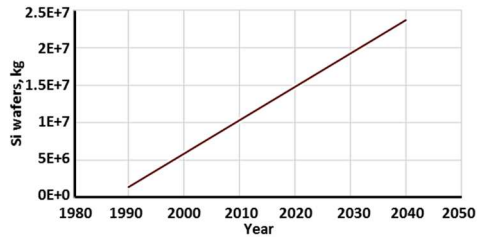

Figure S2 | Projected global Si wafer supply extrapolated from Figure S1.

Remarks on sources of high-purity quartz for wafer grade silicon

The process to prepare a useable single-crystal silicon material for semiconductor applications begins by mining for relatively pure quartz (silicon dioxide). The quartz mineral used to grow the semiconductor grade silicon single-crystals has to be extremely pure. For example the silicon oxide in the form of beach sand is unsuitable for semiconductors, because the time and energy expenditures of its purification would be prohibitive. Very few deposits are classified as high-purity quartz. Most of the quartz used in the semiconductor-grade silicon processes is shipped from select places, notably USA (North Carolina), Australia, Norway, and a few other locations.[3]

S3. Energetics and Speed of Flash and DNA Memory Operations

For NAND flash memory, if only one isolated floating gate cell is considered, its energy of operation is remarkably low. For example, the write energy, E_{cell}, can be estimated as the sum of energy of barrier deformation and the energy needed to "pump" N electrons through the barrier into the storage node. For the ~15 nm flash cell, E_{cell} *is* ~10^{-16} J. Next, in a flash memory system, the memory cells are connected to form an array, and, in many instances the properties of interconnecting array wires determine the operational characteristics of the memory system, e.g., its energy of operation. A given flash cell in an array is selected (e.g., for write operation) by applying appropriate voltages to both interconnect lines, thus charging them. The line charging energy in a typical flash array is ~10^{-11}-10^{-12} J, which is orders of magnitude larger than the "nominal" cell energy. In addition, in practical flash memory systems, large amounts of energy are consumed in peripheral circuitry, such as in multiplexers and voltage pumps, I/Os, controllers, peripheral logic, etc. Computation-intensive algorithms are required for auxiliary operations such as garbage collection, wear leveling, and error correction. All these overheads raise the system-level write energy to 10^{-9}-10^{-10} J/bit. The energetics of flash memory operations is summarized in Table S3. Hence, it is important to note that the total *system-level* energy of flash memory is often dominated by the peripheral circuitry, which requires energy many orders of magnitude above that of the individual floating gate cells.

Table S1 | Energetics of flash memory

Cell level	10^{-16} J/bit
Array level	10^{-11}-10^{-12} J/bit
System level	10^{-10}-10^{-9} J/bit

If one applies a similar system-level analysis to DNA memory in living cells, the DNA molecule serves as the storage medium, while molecular elements used to retrieve the information from DNA and carry this information to certain locations in the cell can be regarded as part of the peripheral circuitry. For example, the mRNA molecules are counted as the memory interface, and the ribosome can be regarded as the output interface of the system.

For a most conservative estimate of the DNA memory performance *in cells*, the entire DNA information content is read and written at least once during one cell division cycle (e.g., for the DNA replicating itself). In *E.coli*, for example, the replication time for a cell is nominally 2400 s, and the total memory size is 9.6 Mbit. Therefore, a characteristic access time per bit, assuming only two accesses (one read and one write per division cycle), is:

$$t_{bit} \sim \frac{2400}{2 \cdot 9.6 \cdot 10^{6}} \sim 100 \mu s \qquad \text{(SE1)}$$

The total power consumption of an *E.coli* cell is about is 1.4×10^{-13} W. Assuming that all energy is used only for memory access, the *system-level* energy per accessed bit (system-level) for an *E.coli* cell is:

$$E_{bit} \sim \frac{1.4 \cdot 10^{-13} \cdot 2400}{2 \cdot 9.6 \cdot 10^{6}} \sim 10^{-17} \frac{J}{bit} \qquad \text{(SE2)}$$

The above estimate assumes that all energy of the cell is consumed to support DNA operations, and is thus highly conservative. If only the energy of nucleotide synthesis is considered (which is imprecisely equivalent to the cell-level operation of flash), in prokaryotic cells such as *E. coli*, the cost of DNA replication accounts for only 2% of the energy budget of the cell.[4] Taking 2% of the above number (SE2) results in the DNA-access energy of 2×10^{-19} J/bit. A similar number can be obtained considering the thermodynamics of dNTP/ATP hydrolysis[5] and taking into account the cost of synthesis of a polymerase molecule.

Table S2 | Energetics of DNA memory in *E. coli* cell

Nucleotide level	2×10^{-19} J/bit
System level	10^{-17} J/bit

Of course, the above estimates of the energetics of DNA memory operations in cells do not directly translate into the envisioned DNA-based storage systems referred to here as Nucleic Acid Memory (NAM). Additional energy may be needed for additional steps required for converting DNA information into a digitally readable format. The estimate for this component will eventually depend on the technique adopted.

DOI: 10.1038/NMAT4594

S4. Weight of a Flash Bit

Flash memory is organized into dense arrays of cells that are fabricated on a silicon wafer. The array capacity is constrained, in part, by the mass of the silicon. The wafer thickness is diameter-dependent (Table S3) because of the mechanical properties of silicon. For example, the initial wafer must be thick enough to avoid cracking during high-volume and speed manufacturing.

Table S3 | Diameter and thickness of standard Si wafers

Wafer diameter (mm)	Wafer thickness (mm)
100	0.525
150	0.625
200	0.725
300	0.775
450*	0.925

* Preproduction wafer diameter under development

The weight of silicon needed for one flash memory bit, m_{flash}, is determined by the thickness of the wafer, h, and the area of the flash cell, A_{cell}, where $\rho_{Si} = 2.33$ g/cm^3 is the density of silicon:

$$m_{flash} = \rho_{Si} \cdot A_{cell} \cdot h \quad \text{(SE3)}$$

The area of one NAND flash cell is $A_{cell} = 4F^2$, where F is its smallest dimension.[6] Based on the physics of operation of flash memory (Supplement S6), F is 10-15 nm. For practical scaling reasons, $F = 15$ nm has been chosen for this analysis, yielding an $A_{cell} = 9 \times 10^{-12}$ cm^2. Assuming $h = 0.775$ mm for a 300 mm wafer, Eq. SE3 results in an $m_{flash} = 1.63 \times 10^{-12}$ g/bit. It should be noted that the thickness of an actual silicon die in a flash package is considerably smaller that the thickness of the initial wafer because the die thickness is reduced to 20–50 µm by back grinding and polishing. However, the weight (SE3) is representative of the total amount of silicon required to build a flash bit. Flash memory can also be organized in 3-dimensional (3D) stacks of several flash memory dies in one package. The number of stacks and the 3D density therefore depends on the die thickness (see above). If a hypothetical ultimate 3D stacking of the flash is considered, where the die thickness is equal to the smallest cell spatial dimension, F, the 3D packing density of the memory cells is $n_{3D} = 1/32F^2$. For $F = 15$nm, a total of 10^{16} bit of flash could fit into the volume of a 1 cm cube.[7]

S5. Weight of a DNA Bit

The DNA memory size (e.g., in a genome) is typically measured in grams (g), picograms (pg), number of nucleotides (nt), bases (b), or base-pairs (bp) with unit conversions of: 1 pg = 978 Mbase or 1 Mbase = 1.02×10^{-15} g for double-stranded DNA (ds-DNA) or 1 pg = ~1.9 Gbase for single-stranded DNA (ss-DNA) or RNA.[8] For conversion to binary units (bits), the information in megabases is multiplied by a factor of 2 (quaternary to binary conversion), e.g., 1 Mbase = 2 Mbits. Each of the 2 bits of equivalent binary information is approximately 0.34 nm of length along a DNA "tape." The resulting weight of a DNA bit is $m_{DNA} = 5.1 \times 10^{-22}$ g/bit, which is 10 orders of magnitude less than the weight of a flash bit.

Table S4 | DNA storage capacity (genome size) for several representative cellular organisms

Classification	Cell Type	Volume (µm^3)	Genome Size (1 Mb = 2 Mbit)
Cyanobacteria	P. marinus	0.1	1.75 Mb = 3.5 Mbit
Bacteria	E.coli	1	4.6 Mb = 9.2 Mbit
Human cells	Average parameters	~1000	~3,000 Mb = 6 Gbit

It is also instructive to consider operational characteristics of DNA memory in a living cell, in which case all possible overheads of a practical memory system are included. For example, an *E.coli* cell contains 4.6 Mbp = 9.2 Mbit of ds-DNA memory, all confined within a 1µm^3 volume (~10^{-12} g liquid weight), and this results in a 3-dimensional memory density of ~10^{19} bit/cm^3 and the weight of a "full" DNA bit of ~1.09×10^{-19} g/bit – still five orders of magnitude less than flash. Theoretically, the same amount of ds-DNA could be stored with a memory density of ~10^{21} bit/cm^3 when stored as a crystal. However, since other components are needed during information reading and writing processes, it can be speculated that the practical density of DNA memory would be close to that of living cells. Additionally, each storage node needs to be connected to other components to be functional, reducing the practical density of DNA memory. Considering the aforementioned facts, the realistic storage density of double-stranded DNA memory is ~10 Mbit/µm^3 or 10^{19} bit/cm^3, which is much denser than the density limits for

flash memory. If such density could be achieved in synthetic environments, the lower bound of predicted storage needs of year 2040 (~10^{24} bit) could be placed in a 100×100×10 cm^3 box.

S6. Scaling Limit of Flash Memory

The current baseline memory technologies (DRAM, SRAM, and flash) are based on storing electron charge in a storage node. Two distinguishable states 0 and 1 are created by the presence (e.g., state 1) or absence (e.g., state 0) of electrons in a specific location (the charge storage node). In order to prevent losses of the stored charge, the storage node is defined by energy barriers of sufficient height E_b to retain a charge (as shown in Figure S3). The properties of the barrier, i.e., barrier height, E_b, and barrier width, a, determine the retention time of a memory cell.

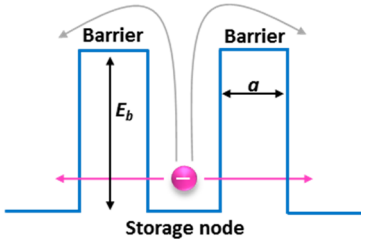

Figure S3 | A generic electron charge-based memory element

There are two fundamental mechanisms for the losses of the stored charge. The first is the thermal over-barrier transitions (thermionic emission), which is related to the Boltzmann probability as discussed in the main paper and in Supplement S8. Also, the "through-barrier" tunneling transition can occur, which is a quantum mechanical effect that can be derived from the famous *Heisenberg Uncertainty Principle*:

$$\Delta x \Delta p \geq \frac{\hbar}{2} \tag{SE4}$$

(SE4) implies that by initially setting the particle on one side of the barrier one can find the particle on either side if Δx is larger than the barrier width a. If a barrier is present within the position uncertainty interval Δx, and the barrier width a is less than Δx, the particle does not "feel" the presence of the barrier. Eq. SE4 can be used to estimate the conditions for significant quantum mechanical tunneling through a barrier of finite width a: assuming the uncertainty in momentum $\Delta p = \sqrt{2mE_b}$ from (SE4) one obtains:

$$\Delta x \sim a_{\min} = \frac{\hbar}{2\sqrt{2mE_b}} \tag{SE5}$$

For $a < a_{min}$, the tunneling probability is significant, and particle localization is therefore not possible. An explicit formula for the tunneling probability can be derived from (SE5) using elementary algebraic transformations (derivation details are given in reference [9]):

$$p_T \approx \exp\left(-\frac{2\sqrt{2m}}{\hbar} \cdot a \cdot \sqrt{E_b}\right) \tag{SE6}$$

Equations (SE5) and (SE6) demonstrate that a heavier mass of the information-bearing particle allows for smaller separation between distinguishable states and leads therefore to more states per unit volume or area. Detailed calculations using (SE6) yield a > 5nm for a 10-year electron retention. The corresponding practical minimum size of the floating gate cell is ~10 nm.[9]

Devices smaller than 5 nm utilize information bearing particles whose mass exceeds that of the electron to suppress tunneling. It has been theoretically shown that "atomic switches" based on moving atoms as information carriers can offer superb memory characteristics relative to electron-based devices in deep nanometer domain.[10] Nanoionic devices have already attracted the attention of many research groups, who have demonstrated several promising experimental memory devices.[9, 11]

Similarly, DNA memory uses molecular fragments (nucleotides) as information carriers — each consisting of more than 10 atoms. The molecular information carriers are densely packed in a linear array with a distance between nucleotides of only 0.34 nm, i.e., more than ten times that of electronic memory.

DOI: 10.1038/NMAT4594

S7. Stability of DNA in Nature

S7a. Stability of DNA

A multitude of physical, chemical, and biological factors can affect the integrity of information in DNA molecules. These factors include pH, ionizing radiation and ultraviolet light exposure, molecular tensions, mechanical forces, salt concentration of the medium, heavy metals, air pollutants, chemotherapeutics, nucleases, inflammation, free radicals, and hydrolysis (Figure S4).[12, 13, 14, 15] pH has a drastic effect on DNA stability, where low pH causes depurination of DNA strands and thus disrupts the information stability. Moreover, it has been suggested that degradation of ancient DNA has been accelerated in museum storage where possibly both temperature and oxygen accessibility to the sample has increased.[16]

S7b. Errors

Errors represent a serious challenge for practical organic chemical synthesis of DNA, requiring error correction procedures.[17] Typical error levels are 1 error in ~100-1000 bits.[18, 19, 20] Researchers have proposed different error reduction methods. One elegant solution is using the MutS protein, which is a part of the DNA repair system in many living cells.[21] The MutS protein binds to DNA mismatches, deletions, or insertions. The MutS affinity to DNA erroneous locations allows for removal of damaged DNA molecules from a mixed pool of good and defective copies. This method yields DNA with one error per 10^4 bases. However, further improvements of the fidelity of DNA constructs are needed. As a benchmark, the error reduction mechanisms in living cells allow for error rates of 1 error per 10^8–10^{10} bases.[22, 23, 24]

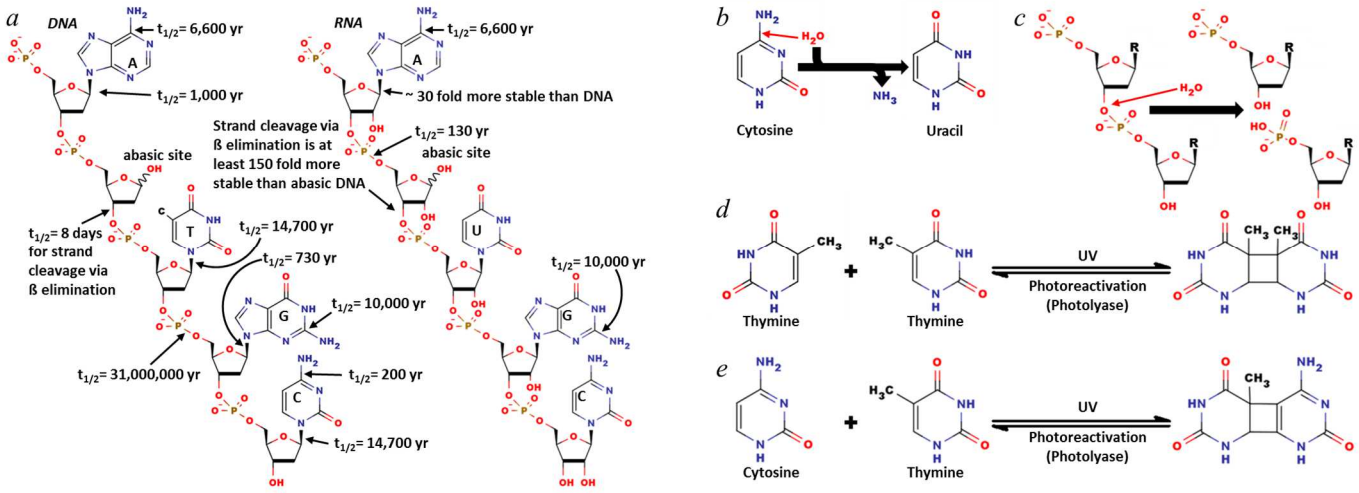

Figure S4 | (a) half-life of DNA and RNA molecules through different damage mechanisms. (b) and (c) are examples of hydrolysis of nucleic acids, deamination of cytosine (b) produces uracil that will cause the point mutation of G>A / C>T, whereas hydrolysis of phosphodieste bond (c) causes strand break damage. (d) and (e) are mechanisms of pyrimidine dimer formations, from two thymine monomers (d) or, a monomer of thymine and cytosine (e). These damages are reversible in cells through a photoreactivation process.

There has also been tremendous progress in new technologies for DNA sequencing — another key component for DNA memory development. There are promising demonstrations of micro-manufactured DNA devices such as a "DNA transistor" for sequencing.[25, 26] The DNA transistor is a microfabricated nanopore structure with embedded electrodes, which is used for both controlling the translocation of DNA through the pore and also for measuring the transversal tunneling current during the translocation. Researchers have theoretically and experimentally shown that the four DNA bases possess distinguishable electronic signatures.[27, 28, 29] Theoretical analysis of an ideal case of the DNA readout suggests that the entire human genome (6 Gbit) might be read in ~7h without parallelization with a 0.1% error rate.[27] The corresponding equivalent access time per bit is ~10 μs/bit. This result, while optimistic is not inconsistent with the estimates of characteristic access time per bit of DNA memory in the cell, performed above: ~100 μs/bit with error rate < 10^{-6} %. Recent experimental studies showed that slower translocation rates are advantageous for reliable readout,[25, 26] suggesting trade-offs between high throughput (fast translocation) and low read errors (slow translocation). Further studies are needed to fully understand these trade-offs. Note that the DNA transistor uses nondestructive electrical readout and is therefore compatible with semiconductor electronics.

S7c. Cells, information storage, replication (copy-paste) and repair (error-correction) strategies

When amassed, cellular DNA encounters 10^4 - 10^5 damages per day[30] even though eukaryotes and prokaryotes both exhibit repair mechanisms that substantially minimize the number of defects. For example, the unrepaired damages are 10^{-4} to 10^{-6} per gamete in human cells[31, 32] and 10^{-7} to 10^{-8} for each cell division in *Escherichia coli* (*E. coli*).[33] Topologically, the information integrity of chromosomal DNA is location dependent – suggesting that microenvironments may impact DNA damage and/or repair rates.[34] Moreover, the mutation rate for mitochondrial DNA

(mtDNA) is 10x that of nuclear DNA (nDNA), which indicates histones may increase both the density and stability of nDNA during packaging of chromosomes.[35] Alternatively, researchers have reported that mtDNA accumulates more errors because its polymerase has lower fidelity and its microenvironment contains harmful factors such as free radicals.[36] Recent studies suggest that mtDNA may also benefit from protein protection via the following three mechanisms: (1) creating multiple copies of itself, (2) base excision repair and double-strand break repair, and (3) complete digestion if repair is not viable; for example ethanol, lipopolysaccharide, or H_2O_2 can induce mtDNA loss in select cells.[35, 36] In comparison, nDNA in human cells is maintained through checkpoint mechanisms that occur throughout the cell cycle such as Gap 1 (G1), Gap 2 (G2), or metaphase checkpoints.[37] There are a number of mechanisms that ensure the integrity of the cellular DNA, and some are briefly described in the following paragraphs.

UV damage contributes to pyrimidine dimer formation of two transverse adjacent TT or TC bases (Fig. S4 d,e).[38] Exogenous chemical agents such as benzopyrene or cisplatin and reactive oxygen species also cause base pairing errors via 8-oxo-guanines and cyclopurines. Regardless of the source of pyrimidine dimerization, cellular DNA repair occurs via nucleotide excision repair (NER), which exploits proteins that sequentially detect and cut damaged DNA and then insert and ligate repaired DNA.[39] Bacteria can utilize three proteins to perform these tasks including *UvrA*, *UvrB*, and *UvrC*. Alternatively, UV damage is repaired via photoreactivation. During photoreactivation, chromophores convert light into chemical work capable of binding the pyrimidine dimer and reverting it to the unchanged form via a photolysis enzyme.[40]

Spontaneous DNA damage, caused by free radicals or other reactive species, are often repaired through base excision repair (BER).[38] For example, an attack of reactive oxygen and water species can cause deamination of cytidine, adenine, or guanine that produces uracil, inosine, and xanthosine. Damaged bases are sequentially cut and removed through cleavage of the covalent bonds between the bases and sugar-phosphate backbone via DNA glycosylase and AP endonuclease. Post cleavage, the gap is filled using polymerase and sealed using a ligase enzyme.[41]

Single strand breaks (SSBR) are another type of DNA damage that are direct products of reactive oxygen species reaction or an intermediate product of enzymes such as Type I topoisomerase that removes supercoiling during replication and transcription. A few enzymes are responsible for repairing these types of damages that are present in the cells and prepare the 3′ and 5′ ends for sealing the strand breaks.[39]

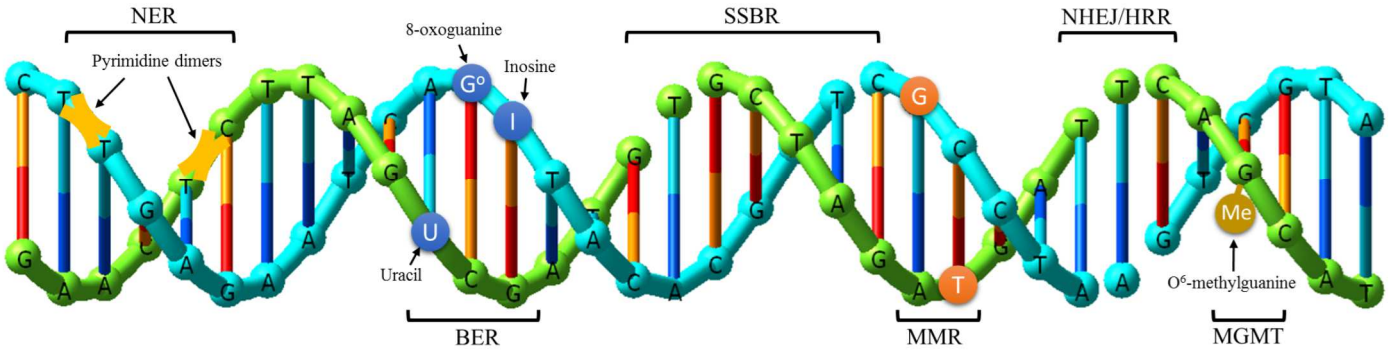

Figure S5 | A summary of the common DNA repair mechanisms of a cell

Base-base mismatches and *insertion-deletion loops* usually occur during DNA replication. The mechanism of DNA mismatch repair (MMR) is conserved from bacteria to mammals and is conducted on the newly synthesized strand that contains the mistake. The newly synthesized DNA is recognized through its unmethylated state in the bacteria, and Okazaki fragments are the markers for the newly synthesized strands in eukaryotic cells. After the newly synthesized strand has been detected and incision has been performed, mismatch is removed from the nascent strand by 5′to 3′ exonuclease activity of the associated exonuclease, and, finally, the gap will be resynthesized by a specific polymerase.[39, 41]

Double-strand breaks usually occur when DNA content is exposed to ionizing radiation. These types of lesions are associated with suppression of replication and transcription or chromosomal rearrangement and are usually deleterious and lead to disruption of functions. Two mechanisms are responsible for repairing these types of lesions: (1) nonhomologous end joining (NHEJ), which utilizes enzyme ligase IV to find the lesions and use the overhang pieces as a template, and (2) homologous recombination repair (HRR) that uses the homologues chromosome as a template for lesion repair.[42] In mammalian cells, alkylating agents damage the DNA and O^6-methylguanine-DNA methyltransferase (MGMT) is an enzyme that removes O^6-alkylation adducts from the O^6 position of the damaged guanine and thus reverts it to the undamaged guanosine.[43] In addition to the aforementioned active cellular mechanisms, there are a few environmental conditions that affect the integrity of DNA information and are the major concern of this perspective.

DOI: 10.1038/NMAT4594

S8. DNA Degradation — Water Attack on DNA

In this study, we performed NAM retention analysis using a generic barrier model for memory devices made from DNA (Figure 3a). Since in NAM, the mass of atoms that form nucleotides is large, tunneling is suppressed and only over-barrier transitions contribute to the memory losses. The rate, r, at which the memory state is lost depends on the barrier height, E_b, and number of collisions with the barrier per unit time, often referred to as thermal attempt frequency, f_0:

$$r = f_0 \exp\left(-\frac{E_b}{k_B T}\right) \tag{SE7}$$

where k_B is Boltzmann's constant and T is temperature, The pre-factor in (SE7) is determined by an average time τ between barrier strikes, which is a function of their average travel distance to the barrier, Δx, and velocity, v:

$$\tau = \frac{\Delta x}{v} \tag{SE8}$$

For a many-particle system the total attempt frequency can be estimated as:

$$f_0 \sim \frac{N_{2D}}{\tau} \sim \frac{v}{\Delta x} N_{2D} \tag{SE9}$$

where N_{2D} is the number of particles in a two-dimensional plane in the proximity to the barrier wall and moving towards the barrier. In many cases, simple estimates of Δx and N_{2D} can be made using basic geometric relations in matter. For example, if a particle system is characterized by the volumetric (3D) density, $n_{3D}=n$ (e.g., number of atoms per cm^3), then the number of atoms per unit of length in an arbitrary direction (e.g., per cm) is calculated as:

$$n_{1D} \sim n^{\frac{1}{3}} \tag{SE10a}$$

Nearest-neighbor distance between the particles is:

$$\Delta x = \frac{1}{n_{1D}} \sim n^{-\frac{1}{3}} \tag{SE10b}$$

or

$$n \sim \frac{1}{(\Delta x)^3} \tag{SE10c}$$

Moreover, the number of atoms per unit of area in an arbitrary cross-section (e.g., per cm^2) is:

$$n_s = n_{2D} \sim n^{\frac{2}{3}} \sim \frac{1}{(\Delta x)^2} \tag{SE10d}$$

Using (SE10a-d), the number of particles hitting the barrier, N_{2D} in (E4), can be estimated as:

$$N_{2D} = \frac{1}{4\pi} \cdot n_{2D} \cdot A = \frac{1}{4\pi} \cdot n^{\frac{2}{3}} \cdot A = \frac{1}{4\pi} \cdot \frac{1}{(\Delta x)^2} \tag{SE11}$$

where A is the cross-sectional area of the barrier structure and the factor 1/4π is used to select particles going in one specified direction. Substituting (SE11) into (SE9) and using (SE10c) results in:

$$f_0 \sim A \cdot \frac{v}{\Delta x} \cdot \frac{1}{4\pi} \cdot \frac{1}{(\Delta x)^2} = \frac{Avn}{4\pi} \tag{SE12}$$

At the absence of external stimuli, the particles gain a velocity v only due to thermal excitations. Using the thermal energy of $\varepsilon = \frac{1}{2}k_B T$ can be found from the kinetic energy relation:

$$\frac{k_B T}{2} = \frac{mv^2}{2} \tag{SE13a}$$

$$v \sim \sqrt{\frac{k_B T}{m}} \tag{SE13b}$$

where m is the mass of the particle. By putting (SE13b) into (SE12) we calculate f_0 as:

$$f_0 \sim \frac{An}{4\pi}\sqrt{\frac{k_B T}{m}} \tag{SE14}$$

The formula (SE14) was derived using a simplified approach and can be refined in a number of ways. Nevertheless, it gives a reasonably accurate estimate for the attempt frequency in different physical systems such as electrons in solids, e.g., in electron-based memories,[44] or atoms/ions in nanoionics resistive memory or RRAM.[10] In the following, it will also be extended to explore DNA as a memory element.

The dominant source of DNA damage is cleavage of chemical bonds via hydrolysis, and it includes the following mechanisms: (a) depurination, (b) backbone cleavage, and (c) deamination. In all these mechanisms, the water molecules strike different sites of the DNA, so the system in question is "reciprocal" to the previously stated, i.e., the external particles properties determine the pre-factor in (SE8).

The molecular density of water can be expressed through its gravimetric density, ρ, the molar mass, M, and the Avogadro's number, N_A:

$$n = \frac{\rho N_A}{M} = \frac{1\left(\frac{g}{cm^3}\right)\cdot 6.022\cdot 10^{23}\left(\frac{molecules}{mole}\right)}{18\left(\frac{g}{mole}\right)} = 3.34\cdot 10^{22}\left(\frac{molecules}{cm^3}\right) \tag{SE15}$$

The corresponding average distance between water molecules $\Delta x \sim n^{1/3} = 0.31\text{nm}$. The area of collision, A, depends on the specific mechanism of the DNA degradation.

Depurination – elimination of A and G bases

For depurination (Figure S6), the area of collision A_{dp} is determined by the nucleotide size, typically $s = 0.34\text{nm}$. Assuming that a nucleotide is represented by a sphere of radius $R = s/2 = 0.17\text{nm}$, the resulting area is:

$$A_{dp} = 4\pi R^2 = 4\pi\left(\frac{s}{2}\right)^2 = \pi s^2 = 0.36nm^2 \tag{SE16}$$

Putting (SE15) and (SE16) into (SE12) yields an attempt frequency for DNA depurination equal to $3.85\times10^{11}\ s^{-1}$ at T = 343 K. The experimentally derived number (the pre-factor in the Arrhenius equation) is $f_0 = 3.10\times10^{11}\ s^{-1}$.[45]

Input data[45]: E_a=31.2 kcal/mol=2.17×10^{-19} J=1.35 eV (experiment, dsDNA)

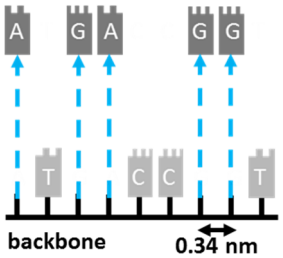

Figure S6 | DNA depurination occurs via protonation of the N7 atom (not shown)

DOI: 10.1038/NMAT4594

Backbone cleavage

The backbone cleavage occurs at depurinated sites, and the corresponding area of collision A_{bc} can be assumed to be related to the surface area of the bare backbone cylindrical fragment of radius r_{bb} (it is assumed that $r_{bb} \approx s$):

$$A_{bc} = 2\pi r_{bb} \cdot s = 2\pi s^2 = 0.73nm^2 \quad \text{(SE17)}$$

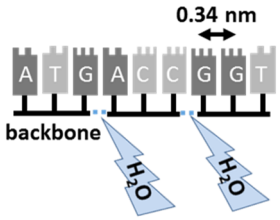

Figure S7 | DNA backbone cleavage

Putting (SE17) into (SE12) yields an attempt frequency of the DNA backbone cleavage equal to 3.8×10^{11} s^{-1} per site at T = 343 K. The experimentally derived number is $f_0 = 2.4\times10^{11}$ s^{-1}.[46]

Input data[46]: E_a = 24.5 kcal/mol =1.7×10^{-19} J = 1.06 eV (experiment, dsDNA)

E_a = 25 kcal/mol =1.74×10^{-19} J = 1.09 eV (experiment, ssDNA)

Since backbone cleavage occurs at depurinated sites, it can be regarded as secondary to depurination. Figure S8 displays the theoretical retention time as a function of temperature for the DNA backbone cleavage mechanism calculated using both Eq. E5b in the main manuscript, and the parameters derived in this section.

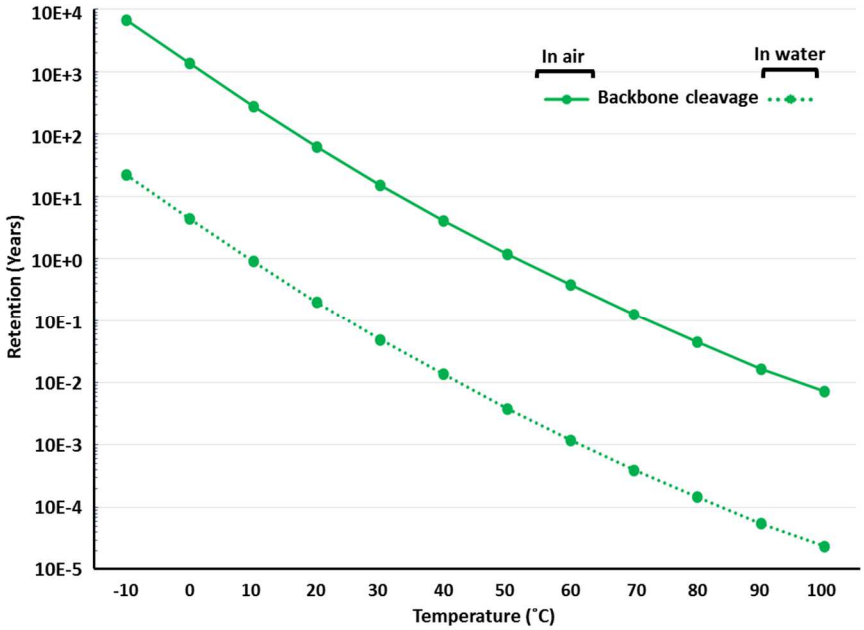

Figure S8 | DNA backbone cleavage: Theoretical retention time as a function of temperature

Deamination

For deamination, the area of collision A_{da} is small and is related to the characteristic covalent radius of the nitrogen atom in the amine radical (Figure S9), which is $r_N \sim 71$ pm. The corresponding reaction area can be assumed to roughly be $A_{da} \sim r_N^2 \approx 0.005 nm^2$, which when put into (SE14) results in $f_0 \sim 5.3\times10^9$ s^{-1}, not far from the experimental attempt frequency of DNA deamination which is equal to 3.7×10^9 s^{-1} at T = 343 K.

Input data[47]: E_a=29kcal/mol=2.01×10^{-19} J=1.26 eV (experiment, ssDNA)

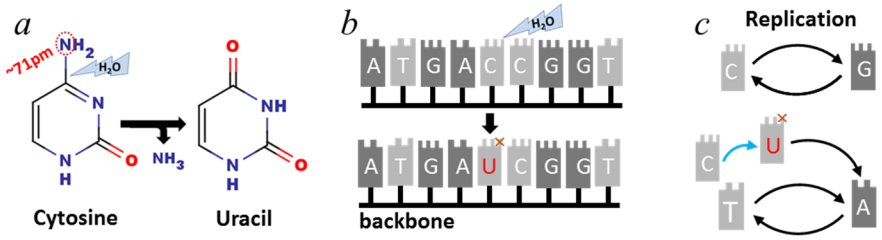

Figure S9 | a) mechanism of DNA deamination. b) an example of information alteration by deamination o cytosine and formation of uracil that leads to G>A and C>T mutation during replication process (c)

References

1. Hilbert, M., Lopez, P. *Science* **332,** 60-65 (2011).
2. Winegarner, R. M. Silicon Industry; 2011.
3. Hughes, E. *Industrial Materials* (2013).
4. Lane, N., Martin, W. *Nature* **467,** 929-934 (2010).
5. Adleman, L. M. *Science* **266,** 1021-1024 (1994).
6. Zhirnov, V. V., Marinella, M. J. Memory Technologies: Status and Perspectives. *Emerging Nanoelectronic Devices*. Wiley, 2015.
7. Zhirnov, V. V., Cavin, R. K. *Microsystems for Bioelectronics: Scaling and Performance Limits*. Elsevier, 2014.
8. Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., *et al. Nucleic Acids Res* **35,** D332-338 (2007).
9. Waser, R., Dittmann, R., Staikov, G., Szot, K. *Advanced Materials* **21,** 2632-2663 (2009).
10. Zhirnov, V. V., Meade, R., Cavin, R. K., Sandhu, G. *Nanotechnology* **22,** 254027 (2011).
11. Hasegawa, T., Terabe, K., Sakamoto, T., Aono, M. *MRS Bulletin* **34,** 929-934 (2009).
12. Branzei, D., Foiani, M. *Nat Rev Mol Cell Biol* **9,** 297-308 (2008).
13. Friedberg, E. C. *Nature* **421,** 436-440 (2003).
14. Campos, P. F., Craig, O. E., Turner-Walker, G., Peacock, E., Willerslev, E., *et al. Ann Anat* **194,** 7-16 (2012).
15. Zadegan, R. M., Norton, M. L. *International Journal of Molecular Sciences* **13,** 7149-7162 (2012).
16. Pruvost, M., Schwarz, R., Correia, V. B., Champlot, S., Braguier, S., *et al. Proc Natl Acad Sci U S A* **104,** 739-744 (2007).
17. Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E. M., *et al. Nature* **494,** 77-80 (2013).
18. Hecker, K. H., Rill, R. L. *Biotechniques* **24,** 256-260 (1998).
19. Matzas, M., Stahler, P. F., Kefer, N., Siebelt, N., Boisguerin, V., *et al. Nat Biotech* **28,** 1291-1294 (2010).
20. Technologies, I. D.
21. Carr, P. A., Park, J. S., Lee, Y. J., Yu, T., Zhang, S., *et al. Nucleic Acids Res* **32,** e162 (2004).
22. McCulloch, S. D., Kunkel, T. A. *Cell Res* **18,** 148-161 (2008).
23. Preston, B. D., Albertson, T. M., Herr, A. J. *Semin Cancer Biol* **20,** 281-293 (2010).
24. Kunkel, T. A. *Cold Spring Harb Symp Quant Biol* **74,** 91-101 (2009).
25. Luan, B., Stolovitzky, G., Martyna, G. *Nanoscale* **4,** 1068-1077 (2012).
26. Luan, B. Sequencing DNA using a Solid-State Nanopore. *SRC Workshop on Semiconductor Concepts from Synthetic Biology - SemiSynBio*. Cambridge, MA, (February 22-23, 2013).
27. Lagerqvist, J., Zwolak, M., Di Ventra, M. *Nano Lett* **6,** 779-782 (2006).
28. Krems, M., Zwolak, M., Pershin, Y. V., Di Ventra, M. *Biophys J* **97,** 1990-1996 (2009).
29. Huang, S., He, J., Chang, S., Zhang, P., Liang, F., *et al. Nat Nano* **5,** 868-873 (2010).
30. Scharer, O. D. *Angew Chem Int Ed Engl* **42,** 2946-2974 (2003).
31. CROWE, F. W., SCHULL, W. J., NEEL. Springfield, J. V. *Journal of the American Medical Association* **162,** 615-616 (1956).
32. De Bont, R., van Larebeke, N. *Mutagenesis* **19,** 169-185 (2004).
33. Clancy, S. *Nature Education* **1,** 103 (2008).
34. Wang, G., Vasquez, K. M. *DNA repair* **19,** 143-151 (2014).
35. Yakes, F. M., Van Houten, B. *Proceedings of the National Academy of Sciences* **94,** 514-519 (1997).
36. Alexeyev, M., Shokolenko, I., Wilson, G., LeDoux, S. *Cold Spring Harb Perspect Biol* **5,** a012641 (2013).
37. Odell, I. D., Wallace, S. S., Pederson, D. S. *J Cell Physiol* **228,** 258-266 (2013).
38. Slupphaug, G., Kavli, B., Krokan, H. E. *Mutat Res* **531,** 231-251 (2003).
39. Kulkarni, A., Wilson, D. M., 3rd *Am J Hum Genet* **82,** 539-566 (2008).
40. Croteau, D. L., DellaVecchia, M. J., Perera, L., Van Houten, B. *DNA repair* **7,** 392-404 (2008).
41. Sirbu, B. M., Cortez, D. *Cold Spring Harb Perspect Biol* **5,** a012724 (2013).
42. Williams, G. J., Hammel, M., Radhakrishnan, S. K., Ramsden, D., Lees-Miller, S. P., *et al. DNA repair* **17,** 110-120 (2014).
43. Xiao, W., Fontanie, T. *Mutation Research/DNA Repair* **336,** 133-142 (1995).
44. Zhirnov, V. V., Cavin, R. K. C. *Microsystems for Bioelectronics: Scaling and Performance Limits*. Elsevier 2015.
45. Lindahl, T., Nyberg, B. *Biochemistry* **11,** 3610-3618 (1972).
46. Lindahl, T., Andersson, A. *Biochemistry* **11,** 3618-3623 (1972).
47. Lindahl, T., Nyberg, B. *Biochemistry* **13,** 3405-3410 (1974).