

Proseminar Speicher- und Dateisysteme

- Hausarbeit -

„Holografischer und DNA-basierter Speicher“

Betreuer: Jakob Lüttgau

Abgabetermin: 22.02.2019



Fabian Böhm

B.Sc. Wirtschaftsinformatik

Gliederung:

1. Einleitung.....	Seite 3
2. Holografischer Speicher.....	Seite 3
2.1. Technik.....	Seite 3
2.2. Probleme & Zukunft.....	Seite 3-4
3. DNA-basierter Speicher	
3.1. Aufbau der DNA.....	Seite 4
3.2. DNA-Sequenzierungsverfahren.....	Seite 4-5
3.3. Nanopore Technologie.....	Seite 5
3.4. Datenspeicherung in der DNA.....	Seite 5-6
3.5. Warum DNA-basierter Speicher?	Seite 7
3.6. Geschichte.....	Seite 7-8
3.7. Chancen und Probleme.....	Seite 8
3.8. Aktuelles.....	Seite 9
4. Zusammenfassung.....	Seite 9
5. Quellen- & Abbildungsverzeichnis.....	Seite 10

Einleitung

In der folgenden Ausarbeitung werde ich kurz auf den holografischen Speicher beziehen und diesen kurz erläutern, indem ich etwas über die Technik erzähle, die dahinter steckt und welche Probleme dort bisher mit einhergehen und welche Zukunftsmöglichkeiten diese Technologie mit sich bringt. Das Thema DNA-basierter Speicher werde ich etwas genauer erläutern, wobei ich zunächst auf den biologischen Aufbau der DNA gehe und wie man dort überhaupt Daten heraus lesen kann. Des Weiteren gehe ich auf die verschiedenen Sequenzierungsmethoden ein und erläutere die Möglichkeiten der Technologie anhand von Beispielen und Versuchen aus bisherigen Experimenten. Zu guter Letzt werde ich noch aktuelle Chancen und Probleme darlegen.

Holografischer Speicher

Die Entwicklung des holografischen Speichers begann Mitte der 2000er Jahre, als die Firma InPhase einen ersten Prototypen entwickelte, der die doppelte Größe eines handelsüblichen Toasters hatte.

Die Idee dahinter war, dass Daten dort auf einer 3mm dicken transparenten Plastikscheibe gespeichert werden sollten. Außerdem muss man sich diese Speicherung „räumlich“ vorstellen, da erstmals das gesamte Volumen zur Speicherung genutzt wurde, im Gegensatz zu herkömmlichen Speichertypen, wie DVD oder CD, wo nur die Oberfläche beschrieben wird.

Der Prototyp hatte, laut Angaben der Firma in einer Pressemeldung im Jahre 2005, eine Speicherkapazität von 300GB, wobei bis zu 1,6TB angestrebt werden sollte. Um dies in Relation zu setzen: Das wäre die ungefähre Speicherkapazität von 340 DVDs. 2005 eine enorme Revolution, aber heute geht man sogar davon aus, dass 1 Petabyte (1.000TB) möglich ist. Des Weiteren konnte der Prototyp die Daten 10-mal schneller auslesen als eine CD.

Die Anwendung holografischer Speicher sahen die Entwickler vor allem in dem Bereich der Massenspeicher, wie Bibliotheken, Archive oder auch Fernsehanstalten.

Technik

Wie man in der Abbildung 1 erkennen kann schreiben 2 Laserstrahlen aus der selben Quelle Punkt für Punkt holografische 3D-Bilder in das transparente Speichermedium. So entsteht ein Schachbrettmuster aus hell-dunkel mit je ca. 1Mio. Bits.

Der sog. Referenzstrahl ist direkt auf das Medium gerichtet, wobei der Signalstrahl auf einen Modulator fällt, der durch Bitmuster verändert wird. Sobald sich beide Strahlen treffen, entsteht ein Interferenzmuster und das Hologramm wird dauerhaft sichtbar sein.

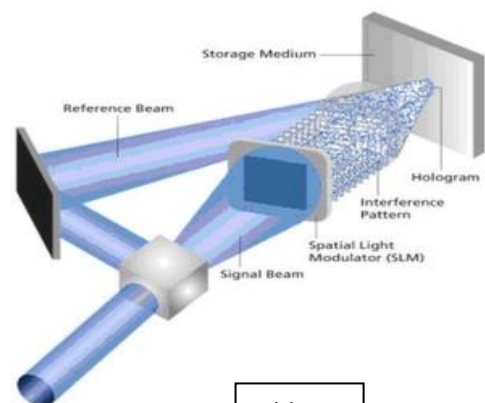


Abb. 1

Probleme und Zukunft

Der 1. Prototyp besaß eine Transferrate von 160MB/s (später bis zu 960MB/s) und eine Übertragungsrate von einer zehntausendstel Sekunde. Dass es aber nur ein einmal beschreibbarer Lesespeicher ist, ist nur ein Problem. Zudem kommen noch weitere, wie beispielsweise die hohen Kosten, die einhergehen. Die Firma InPhase hatte 100Mio. €

investiert, ohne auch nur eine einzige marktfähige Holodisk auf den Markt gebracht zu haben, geschweige denn eine verkauft zu haben. Der Grund dafür, waren die hohen Kosten für die Entwicklung des Produktes, wobei hauptsächlich die Materialfrage für die Langzeitstabilität das Problem war.

Auch wenn mittlerweile mehr als 13 Jahre vergangen sind, seit der erste Prototyp präsentiert worden ist, sehen einige Experten trotzdem noch eine Zukunft in den holografischen Speichermedien, da das Potenzial als Massenspeichermedium eingesetzt zu werden enorm hoch ist, zur Zeit die Firmen aber kein Profit aus der Herstellung ziehen würden.

DNA-basierter Speicher

Aufbau der DNA

1953 stellten Watson & Crick ein Strukturmodell über das gesamte bis dato gesammelte Wissen der DNA auf. Sie erkannten auf Röntgenbilder erstmals die Doppelhelix. In dieser Doppelhelix zeigen die Basen immer nach innen und sind komplementär (gleiche Paarung).

- ➔ Adenin + Thymin & Cytosin + Guanin
---- „Leiter“ ----
- ➔ Phosphat-Desoxyribose-Rückgrat
---- „Holm“ ----

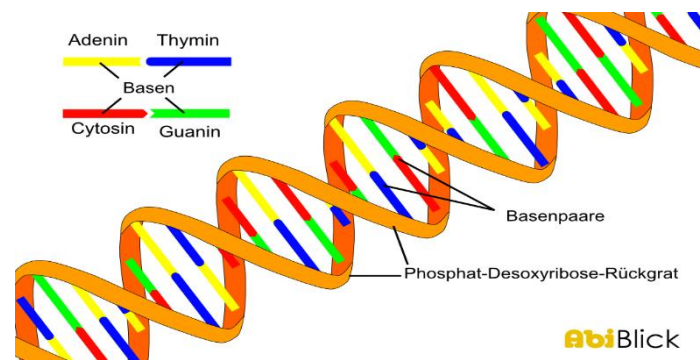


Abb. 2: Aufbau der DNA

Durch Hydrolyse kann man die einzelnen Bestandteile voneinander trennen.

Die DNA ist also eine Abfolge aus Nukleotidbasen¹. 3 zusammenhängende Basen bilden ein sog. Triplet, welches bestimmte Aminosäuren codiert, welche wiederum Proteine und Enzyme bildet. Dies macht aufgrund von verschiedenen Ausprägungen und Anzahlen die unterschiedlichen Merkmale für die Lebewesen aus.

Zusammenfassend kann man sagen, dass Basen durch ihre Anordnung unsere Erbinformationen übersetzen.

Insgesamt gibt es 20 verschiedene Aminosäuren, wodurch direkt sichtbar wird, dass ein Triplet auch unbedingt notwendig ist. ($4^3 = 64$). Das wiederum bedeutet auch, dass gleiche Aminosäuren aus verschiedenen Basenpaarungen codiert werden können. Demnach ist es dann wichtig die Infos des Basensequenzen zu entschlüsseln.

DNA-Sequenzierungsverfahren

Die DNA-Sequenzierung dient der Bestimmung der Basenabfolge innerhalb eines Gens. Es gibt mittlerweile viele und auch moderne Varianten diese zu bestimmen, wobei ich im folgenden die etwas ältere, aber verständlichere Variante nach Sanger genauer erläutern und nur kurz auf eine der moderneren Methoden eingehen werde.

1.Schritt

¹ Nukleotid: Verbindung aus Phosphat, Zucker und Basenbestandteil

Jeder Strang der DNA-Doppelhelix wird chemisch markiert und unter Hitzeeinwirkung werden die Wasserstoffbrücken, die die Basenpaare zusammenhalten, getrennt. So entstehen zwei Einzelstränge

2. Schritt

Ein sog. Primer² setzt sich an den Einzelstrang und die DNA-Polymerase³ setzt sich an der Primer. Der Baustoff sind die beigefügten Nukleotide mit einem zusätzlichen Abbruch-Nukleotid. Wenn das Abbruch-Nukleotid „verbaut“ wurde, dann ist Schluss. So entstehen verschieden lange Fragmente und da man weiß, welches das Abbruch-Nukleotid ist, weiß man auch das gesuchte Nukleotid (Primer-Hybridisierung).

3. Schritt

Auftrennung der Länge nach mit Hilfe der Gelelektrophorese. Dies ist ein Verfahren um Moleküle zu trennen und sichtbar zu machen.

Aufbau: - Gelmatrix (elektrisch geladen)
 - Anode (+)
 - Kathode (-)

Ablauf:

Unterschiedliche Moleküle (Größe = Stärke der Ladung) wandern weiter bzw. kürzer durch das Gel in Richtung Anode, da die DNA durch das Phosphat von Natur aus negativ geladen ist. Gleiche Moleküle (gleiche Größe/Ladung) bilden Banden.

4. Schritt

Betrachtung der Ansätze aller chemischen Gene. Die Abfolge ist rekonstruierbar und man kann an der jeweiligen Markierung den DNA-Strang ableiten. Dieser muss dann nur noch in sein komplementäres Abbild umgewandelt werden.

Nanopore Technologie

Eine der moderneren Varianten zur DNA-Sequenzierung ist die Nanopore Technologie.



Abb. 3: Nanopore Technologie

Der Name entstammt der nanometergroßen Öffnung (kaum größer als DNA-Strang) in einer Membran. Bei dieser Variante fädelt man einen DNA-Strang durch diese Membran und zieht ihn bis zum Ende durch. Jedes Mal, wenn eine Base (Adenin, Cytosin, Guanin, Thymin) die Engstelle passiert, misst ein eingebauter Messmechanismus die Daten in Echtzeit und der Code wird entschlüsselt. Somit ist kein rechenintensives Zusammenpuzzeln von Mio. Bruchstücken mehr nötig. Vor allem Feldforscher, die Daten in Echtzeit benötigen, ist diese Technologie von Vorteil, da die Geräte nur die Größe eines USB-Sticks benötigen.

² Primer: radioaktiver (beim Röntgen sichtbar) Startpunkt bei der DNA Replikation)

³ DNA-Polymerase: Enzyme, welche die Synthese aus der DNA katalysieren

Datenspeicherung in der DNA

Wie werden nun die Daten in der DNA gespeichert bzw. spezieller, wie kommt ein Foto auf DNA?

Beispiel: ein digitales Foto besteht aus einer Sequenz aus 1 und 0. Nun muss man von Binärcodierung auf Code der DNA übersetzen.

Zunächst wird erst einmal künstliche DNA hergestellt werden. Ein Molekül besteht ja aus einer Abfolge der Basen (A,C,G,T) und diese kann man nun selbst wählen.

$2^7 2^6 2^5 2^4 2^3 2^2 2^1 2^0$ Zitat von Goldman: „Jede digitale Info beruht auf 0 und 1, ebenso gut kann sie auch in langen Reihen aus A,C,G und T umgewandelt werden.“
 0 1 1 0 1 0 0 1 Nick Goldman war einer der ersten, dir eine komplette Nachricht in DNA gespeichert hat.
 $4^3 4^2 4^1 4^0$ Allerdings muss man sich das Ganze ein wenig anders vorstellen. Der Schreibprozess ist hierbei der Herstellungsprozess (DNA-Synthese) und der Leseprozess ist die DNA-Sequenzierung.

Beispiel für die Umwandlung von Binärcode in Quartärcode:

→ Bsp.: Zahl 105
 binär = 01101001
 quartär = 1221
 → A=0, C=1, G=2 und T=3
 DNA = CGGC

	vorheriger Buchstabe			
	A	C	G	T
0	C	G	T	A
1	G	T	A	C
2	T	A	C	G

211012012 als
 TCTAGCGAT

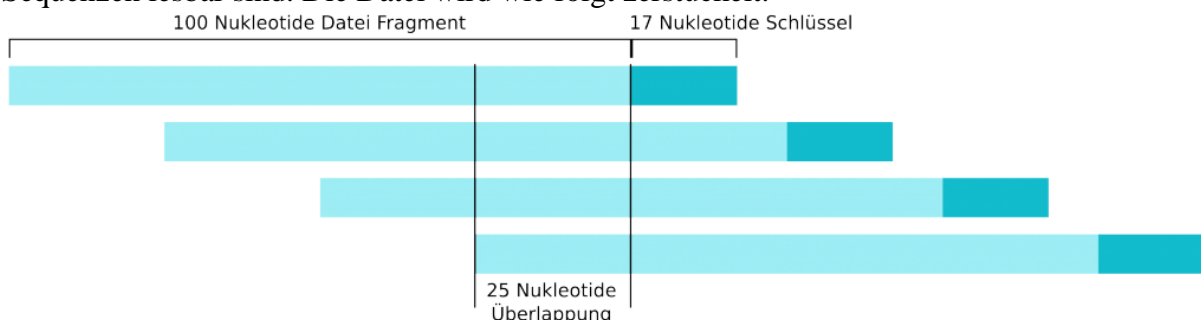
Heutige Sequenzierungsmethoden weisen allerdings Schwächen auf.

Ungenauigkeiten beim Lesen der DNA, wenn gleiche Buchstaben aufeinander folgen, ist ein Problem. Man kann sich das ungefähr so vorstellen, wenn man versucht vielen Nullen ohne Tausender-Separator zu lesen.

Nick Goldman und sein Kollege Birney überlegten sich einen Trick, indem sie den Ternärcode statt Quartärcode nutzten.

- Bsp.:
 wurde gerade ein „T“ geschrieben, dann verbleiben nur noch die Werte 0,1 und 2, die man den Nukleotiden A, C und G zuordnen kann

Nun muss man aber wissen, dass die Datei bei der DNA-Sequenzierung nicht als kompletter DNA-Strang gespeichert wird, da, auf die Sanger-Methode bezogen, nur 500 Nukleotid lange Sequenzen lesbar sind. Die Datei wird wie folgt zerstückelt.



Wie man sehen kann, wird die Datei in 117 Nukleotid-lange Sequenzen aufgesplittet. 100

Nukleotide davon enthalten die eigentliche Datensequenz und weitere 17 dienen als Schlüssel, um die Daten am Ende wieder zusammenzufügen.

Außerdem werden die Fragmente in 4 Teile geteilt, die sich gegenseitig in 25 Nukleotiden überlappen. Dies ist ein Trick, um später fehlerfreier lesen zu können.

Zum Schluss werden die Fragmente zur DNA-Synthese ins Labor geschickt, die jedes Fragment in millionenfacher Kopie herstellt. (Bei späterer Auswertung wurden erstaunlicherweise 100% der Daten wieder hergestellt, inkl. Der Kopien).

Warum DNA-basierter Speicher?

Der Hintergedanke beim DNA-basierten Speicher ist zum einen, dass der weltweit benötigte Speicher zwar wächst, aber auch die Daten. Um Vergleiche zu ziehen, eine Diskette hat eine Speicherkapazität von 3,25MB, magnetische Speicher schon 185GB und optische Speicher sogar bis zu 1 Petabyte. Eine Studie der EMC Corporation, ein US-Speicherhersteller, sagt voraus dass bereits 2022 die verfügbare Speichermenge nur noch für 15% aller Daten reichen würde.

Auch wenn viele Daten flüchtig sind, gibt es auch Daten, die vor allem in den Naturwissenschaften heutzutage nur noch nicht gebraucht oder verwertet werden können. Wenn man diese Daten für spätere Berechnungen speichern könnte, wäre das ein großer Vorteil.

Ein weiterer Punkt ist auch, dass heutige Speichermedien eine viel zu geringe Lebensdauer haben. Eine DVD zum Beispiel 10 Jahre und Flash 30 Jahre. Und mal ehrlich, wer kann heutzutage noch eine Diskette lesen?

Allerdings ist nicht nur die Datengröße ein Problem, sondern auch die räumliche Größe. Man kann sagen, dass 1mm³ gebraucht werden, um 100GB zu speichern. Bis zum Jahr 2022 soll das weltweite Datenvolumen auf 44 Zettabyte (44 Billion GB) anwachsen. Umgerechnet wären das 440.000l Speicherplatz. Besser wären doch, wenn man genau die gleiche Menge an Daten auf 44ml speichern könnte. Gesucht ist 1 Exabyte/mm³ (1 Milliarde GB) und die DNA bietet genau das. Um es besser zu verdeutlichen: auf 1 Gramm DNA passen Daten von 1 Million CDs.

Wie bereits indirekt erwähnt ist der Anwendungsbereich des DNA-basierten Speichers die Langzeitspeicherung von Daten. DNA kann, wenn richtig gekühlt, mehrere Jahrtausende halten und Daten speichern.

Die DNA ist die Grundlage des Lebens. Das Speichermedium an sich – die DNA – wird auch in tausenden Jahren gleich sein und lesbar bleiben.

Geschichte

1988 wurde erstmals die Idee geäußert digitale Daten in Biomoleküle zu lagern. Allerdings erst 6 Jahre später kam Adleman auf die Idee des DNA-Computers. Er lieferte einen Beweis, dass man mit DNA programmieren kann, indem er einige kleinere mathematische Gleichungen mit Hilfe des Computers löste. 1997 merkte man dann aber, dass die DNA-Computer zu der Zeit falsch bzw. nicht effizient genutzt wurde und man entschied sich auf boolesche Funktionen umzusteigen, da die Forscher dort den optimalen nutzen sahen. 2002 wurde dann der erste programmierbare molekulare Computer in Israel am Weizmann Institut entwickelt. Dieser bestand auch Enzymen und DNA-Molekülen statt aus Siliziumchips, die ansonsten in herkömmlichen PCs angewendet werden. 2 Jahre später haben

dann Forscher des Weizmann Instituts den DNA-Computer mit einem In- Outputmodul⁴ gekoppelt. Dieser war in der Lage Krebsaktivitäten in Zellen aufzuspüren und bei Auftreten dieser ein Medikament zu geben.

2013 war das Jahr als Goldman und Birney ihre Methode, codierte DNA zu synthetisieren und als Datenträger zu verwenden, im Magazin „Nature“.

Ihre erste Nachricht, die „nur“ 739KB groß war, enthielt unter anderem folgende Daten:

- 154 Sonette von Shakespeare
- JPEG Foto der Forscher neben einem Baum vor dem Institut
- Publikation von Watson & Crick als PDF-Datei
- MP3-Auszug aus der berühmten Rede von Martin Luther Kings „I have a dream“
- Kodierungsanleitung

Die Forscher schickten die Nachricht an die Firma Agent Technologies, um alles Daten zu entschlüsseln (Dauer = 2 Wochen). Anschließend schickte die Firma den Forschern ein Paket mit winzigen, gefriergetrockneten Partikeln, kleiner als ein Sandkorn, in Proberöhrchen zurück. 100% des Daten plus aller Millionen Kopien konnten wieder entschlüsselt werden, wobei beispielsweise 1 Sonett Shakespeares 0,3 Pikogramm (0,0000000000003g) wog. Bis heute gab es noch viele weitere erfolgreiche Versuche, wie zum Beispiel gelang Forschern eine Filmsequenz im Erbgut lebender Bakterien zu speichern.

Chancen & Probleme

Drei große Kernargumente stützen die These, dass DNA-basierte Computer eine vielversprechende Zukunft haben.

Das erste Argument ist die Datenarchivierung, die bei DNA-basiertem Speicher bei mehreren tausend Jahren liegt und somit herkömmliche Speichermedien deutlich aussticht. Außerdem geht es bei der Datenarchivierung irgendwann auch um die räumliche Größe, wobei DNA-basierter Speicher nur einen winzig kleinen, fast schon unbedeutend großen Bereich benötigen würde, wie bereits schon zuvor erwähnt.

Ein weiteres Argument ist die enorm hohe Speicherkapazität. Am besten kann man die an einem Beispiel deutlich machen:

1 Gramm DNA = 1 Million CDs = 215 Petabyte = 215.000TB

➔ Computer mit 1l Flüssigkeit und 6 Gramm DNA = 3072 Exabyte = 3,1 Mio. Petabyte = 3,1 Mrd. TB

Das dritte gute Argument das für eine Zukunft von DNA-basiertem Speicher spricht ist die Parallelisierung. Zwar ist ein DNA-Computer nicht in der Lage Daten schnell auszulesen, dafür können aber viele komplexe Aufgaben gleichzeitig gelöst werden. Heutige DNA-Computer können bis zu 1.000 Peta-Operationen pro Sekunde gleichzeitig lösen, wohingegen die besten Supercomputer nur einige 10 Peta-Operationen lösen können.

Natürlich gibt es aber auch einige Probleme die damit einhergehen, wie zum Einen die zurzeit noch unvorstellbar hohen Kosten. Hier kann auch ein Beispiel den besten Überblick geben:

DNA-Synthese = 12.400\$ / MB

DNA Sequenzierung = 220\$ / MB

➔ Um nun Daten auf 1 Gramm DNA zu speichern, die ja bekanntlich bis zu 215 Petabyte speichern kann, kommt man auf 3 Trillion \$

➔ Anders ausgedrückt: Auf eine Zelle, die 2 Nanogramm wiegt, passen 43MB Daten, wobei Kosten von 542.660\$ entstehen.

⁴ In- Outputmodul: Kommunikation/Interaktion eines Informationssystems mit der Außenwelt bzw. mit den Benutzern oder anderen Informationssystemen

Ein weiteres Problem ist die technische Umsetzung. Bei der Methode nach Sanger besteht die Möglichkeit bis zu 1000 Basenpaare aufwendig zu analysieren und später wieder zusammenzusetzen. Das kleinste menschliche Chromosom, das Y-Chromosom aber alleine besitzt schon 50 Mio. Basenpaare.

Zu guter Letzt kommen wir zu dem Punkt der langen Zugriffszeiten, wo der DNA-basierte Speicher mehrere Stunden benötigt und andere herkömmliche Speichermedien nur wenige Millisekunden.

Allerdings ist wie bereits erwähnt der zukünftige Anwendungsbereich in der Datenarchivierung und liegt nicht darin Daten enorm schnell auslesen zu können.

	Strang 1		Strang 2		Strang 3	
Aktuelles	0	XOR	0	→	0	
	1	XOR	1	→	0	
Anfang 2016 hatte sich eingeschaltet und hat sich zum Goldmann und Birney zu	1	XOR	0	→	1	Microsoft in das Projekt mit Ziel gemacht die Methode von verbessern. Sie wollten
gezielter auf die Daten zugreifen, damit man dann nicht mehr den kompletten DNA-Speicher auslesen muss. Somit entstehen eine hohe Kosten- und Zeitersparnis.	0	XOR	1	→	1	
Zudem wollten die eine elegantere Lösung bezüglich der Redundanz erschaffen, indem sie mittels einer XOR-Verknüpfung den ersten oder zweiten Strang mit Hilfe eines dritten Strangs wieder rekonstruieren können. Somit kann man wichtige Dateien mehrmals verknüpfen und besonders sicher speichern.						

Zusammenfassung

Abschließend kann man sagen, dass die DNA die Grundlage des Lebens ist und somit auch immer ein Teil des Menschen bleiben wird, welche sich nicht verändert und somit für allezeit lesbar sein wird.

Zwar ist der DNA-Speicher kein Speicher für Endverbraucher, da diese meist die Intention haben schnell auf Daten zugreifen zu wollen. Hinzu kommen noch enorme Kosten und die Notwendigkeit einen DNA-Computer ständig kühlen zu müssen.

Allerdings ist sie meiner Meinung nach der Speicher der Zukunft, da heutige Speichermodule keine lange Lebensdauer besitzen und man somit Daten häufig umschreiben muss. In der Datenarchivierung wäre das von enormen Vorteil.

Quellenverzeichnis

Abitur-Wissen Biologie – Genetik von Albert Kollmann (Stark-Verlag, Ausgabe 94/703)

<http://www.scinexx.de/wissen-aktuell-21220-2017-03-06.html>

<https://www.dna-sequenzierung.com/sanger-sequenzierung/>

<https://de.wikipedia.org/wiki/DNA-Sequenzierung>

https://de.wikipedia.org/wiki/Holografischer_Speicher

<https://www.frustfrei-lernen.de/biologie/dna-dns-aufbau-struktur.html>

<http://scienceblogs.de/bioinfowelten/2016/07/27/die-zweckentfremdung-der-dna-teil-1-speicher-der-zukunft/2/>

<https://lab.dessimoz.org/media/DNA-storage/Die%20Zeit.pdf>

<https://www.simplyscience.ch/definitionen/articles/nukleotid.html>

<https://www.sueddeutsche.de/wissen/dna-forscher-speichern-historischen-film-im-erbgut-lebender-bakterien-1.3588165>

https://de.wikipedia.org/wiki/Eingabe_und_Ausgabe

https://www.deutschlandfunk.de/tolle-idee-was-wurde-daraus-holografische-speichermedien.676.de.html?dram:article_id=402409

<https://www.storage-insider.de/holografie-speichern-mit-licht-a-88545/>

https://de.wikipedia.org/wiki/Optischer_Datenspeicher

<http://www.biologie-schule.de/gelektrophorese.php>

<https://de.wikipedia.org/wiki/DNA-Computer>

<https://de.wikipedia.org/wiki/Primer>

Abbildungsverzeichnis

Seite 3: <https://www.storage-insider.de/holografie-speichern-mit-licht-a-88545/>

Seite 4: <https://www.abiblick.de/biologie/dna/>

Seite 5: <https://www.spektrum.de/news/ein-dna-lesegeraet-fuer-die-hosentasche/1143028>

Seite 6 (alle): <http://scienceblogs.de/bioinfowelten/2016/07/27/die-zweckentfremdung-der-dna-teil-1-speicher-der-zukunft/2/>

Seite 9: <http://scienceblogs.de/bioinfowelten/2016/07/27/die-zweckentfremdung-der-dna-teil-1-speicher-der-zukunft/2/>