"As We May Think"

Vannevar Bush

CS294-6 *Reading The Classics* 9/21/04

Notes by: Eylon Caspi eylon@eecs.berkeley.edu
AJ Shankar aj@eecs.berkeley.edu
Jingtao Wang jingtaow@eecs.berkeley.edu

In a 1945 paper [1] for the popular press, Vannevar Bush proposed that science be put to use in organizing the vast record of human knowledge. Inspired by his previous work in microfilm mass storage, Bush envisioned an information workstation—the *memex*—capable of storing, navigating, and annotating an entire library's worth of information. His idea of push-button linking between documents is commonly held to be the forefather of modern hypertext. However, Bush's vision lacked several key innovations present in today's information workstations, including searchable, digital content and rapid information sharing on a network. Bush tells a masterful story of technology's state of the art in 1945 and of taking its trends to their logical conclusion, revealing how his vision was both guided by and limited by that technology.

This document shall explore and summarize Bush the man, the contributions and limitations of his paper "As We May Think," and our class discussion thereof. We begin with (1) a biography of Bush and (2) a discussion of his vision for organizing the human record. We discuss in turn Bush's predictions on technology for (3) information acquisition, mass storage, (4) automation, and (5) information retrieval, and the culmination of those technologies in (6) the memex. We then discuss (7) the limitations of the memex and of Bush's vision at large.

1. Bush Biography

We give a brief introduction of Bush's life[1], so as to establish his personal background motivating "As We May Think."

1.1 Bush's Early Life

Bush was born on March 11, 1890, in Chelsea, Massachusetts[2], to father Richard Perry Bush[3] (a universalist minister) and mother Emma Linwood Bush. He had two sisters. He studied engineering at Tufts College, earning in 1913 both a bachelor's degree and a master's degree in the time usually taken to earn a bachelor's degree. His master's thesis included the invention of the Profile Tracer, which can be used in surveying work to measure distances over uneven ground.

---

[1] Most of the information in this section is from [4] [7] [8] [9].

[2] A different source (http://www.nuclearfiles.org/rebios/vannevarbush.htm) claimed that Bush was born in Everett, Massachusetts.

[3] We could find no evidence implying any relationship between the family of Vannevar Bush and the family of presidents George H. Bush and George W. Bush.

After graduation, Bush worked shortly as a test engineer for General Electric but was laid off after a fire broke out in his plant. He then took a position teaching math at Tufts' sister college, Clark University in Massachusetts, in 1914. On September 5, 1916, he married Phoebe Davis [7]. He undertook higher education at MIT, earning a doctorate of engineering in less than a year (1917), then returned to Tufts as an assistant professor [8].

1.2 Academic Work

In 1919, Bush left Tufts and joined MIT's electrical engineering department. He would stay at MIT for 25 years. At first, he worked on electrical power transmission and on submarine detection, then in the 1930's, on analog and mechanical computers. In 1931, Bush completed the first *differential analyzer*, a mechanical computer for solving differential equations. It would be used during WWII to calculate ballistics tables. Interestingly enough, Claude Shannon—inventor of information theory, which became the foundation of digital communication and digital computers—was one of Bush's students at MIT.

During the 1930s, Bush also worked on technology for document retrieval and information organization. He made extensive use of microfilm, which was growing in popularity as a storage medium, especially among librarians. In 1938, together with John H. Howard, Bush designed and built the *rapid selector*, a machine for high-speed referencing of information stored on microfilm. This technology would play a key role in his landmark article, "As We May Think."

In 1937, Bush was elected president of the Carnegie Institution. In this prestigious role, he influenced the direction of research in the U.S. and informally advised the government on scientific matters.

1.3 Military Advisory, WWII

In the 1930s, US military research was relatively small and disorganized. It was performed primarily by military staff and often duplicated between different branches of the military. On June 12, 1940, Bush met with President Roosevelt to detail a plan for changing the organization of military research. Bush proposed a new organization, called the National Defense Research Committee (NDRC), which would bring together government, military, business, and scientific leaders to coordinate military research. Roosevelt agreed, and the NDRC was created with Bush as its chairman. In 1941, the new Office of Scientific Research and Development (OSRD) subsumed the NDRC, and Bush became its director. Bush thus came into a position of tremendous influence, including supervising the Manhattan Project which developed the first atomic bomb. Bush's work not only helped the Allies win the war, but it changed the way scientific research was done in the U.S. Bush demonstrated that technology was the key to winning a war, and in turn earned a new respect for scientists.

By late 1944, Bush believed Allied victory was inevitable, and he began to consider the future role of science. In reply to President Roosevelt's request for post-war direction, he published [1] and [2]. He argued that, after the war, the nation would still need permanent support for research, and he outlined in [2] a plan of continued support for research. He wrote, "It is my judgment that the national interest in scientific research and scientific education can best be promoted by the creation of a National Research Foundation." The National Science Foundation (NSF) was created in 1950 to answer that call. Though not as powerful as Bush's proposal called for, the NSF nevertheless secured the marriage between science and government.

1.4 After WWII

After WWII, Bush worked as a military consultant and research advisor. He served as director of AT&T and of Merck and Co. He oversaw the creation of the NSF and served on its Advisory Committee for several years. He served as chairman of the MIT Corporation and remained involved for many years as its honorary chairman. In these roles, Bush continued to push for analog computers, even as digital computers rose to prominence.

Vannevar Bush died on June 30, 1974, years before the rise of the Internet and World Wide Web, which would be deeply influenced by "As We May Think". In 1990, the NSF created the Vannevar Bush Award in memorial of Bush's contribution to basic scientific research in the U.S.

2. Bush's Vision: Science and the Record

In "As We May Think," Bush turned his thoughts to the future of scientific research. It was clear to him that science was an essential tool for interacting with the world, as well as for organizing and disseminating knowledge. Bush believed that the sheer volume of information becoming available to scientists would overwhelm traditional methods of acquisition, storage, and analysis, and so new methods were needed. He argued that science itself, *i.e.* technology, could help in the task of organizing knowledge. And so in the rest of the paper, Bush proposed novel solutions to the various problems of maintaining the *record*, the universal body of knowledge acquired by mankind.

3. Acquisition and Storage

Bush first turned to the task of extending the record: acquiring new data. He understood that images would comprise a substantial part of that record and saw a need to make photography instant and cheap. He proposed the use of a head-mounted camera and "dry photography" (non-wet developing) to visually document a scientist's workday. Dry photography was not quite a realized technology at the time, and so Bush suggested extending two existing technologies: the facsimile and the television. Whatever the method, however, there is still the matter of storing the staggering number of pictures that would no doubt be generated by a shutter-happy scientist. Here Bush counted on advances in microfilm that would allow pictures to be stored in a miniscule space.

Microfilm already existed with 20x linear reduction. Bush noted that, with 100x linear reduction, the entire Encyclopedia Britannica could be reproduced in matchbox size for a nickel. Such an inexpensive storage medium would then lead to ubiquitous distribution of information.

For extending the written record, Bush suggested dictation technology. He envisioned connecting a Vocoder[4] (an electronic speech analyzer) to a stenotype (a phonetic typewriter), so as to translate human speech to a typed sequence of phonemes. The subsequent problem of converting the phonemes into proper written language remained unsolved. Bush suggested a rather extreme approach, which is to devise a natural language suited particularly for analysis and parsing by machine: "All [a scientist] needs to do is to take advantage of existing mechanisms and to alter his language."

It is interesting to note that Bush did not discuss digital content storage. He discussed digital representations for calculation and indexed retrieval (more on this below) but not for content storage. Digital mass storage would be commercialized in the coming decade with the introduction of tape drives (*e.g.* UNIVAC 1951, IBM 1953) and hard drives (*e.g.* IBM 1956), but Bush did not see that far.

4. Calculation and Automation

Bush was quite optimistic (and rightly so) about the power of machines to manipulate and analyze data once it is in the system *in a suitable representation.* Calculation could already be automated by mechanical and electro-mechanical computers—including Bush's own Babbage-esque *Differential Analyzer* from his MIT days in the 1930s—given input by keyboard or punch-card. Bush noted that large systems would require the chaining of machines using compatible, intermediate representations. Bush also predicted blazing fast electrical machines capable of, "performing at 100 times present speeds, or more." He went well beyond the staid prediction that computers would crunch numbers like nobody's business (even to the point where data entry was the limiting factor), and made the startlingly accurate (so far, at least) prediction that computers would eventually perform all laborious computations—even calculus, "advanced mathematics", and in fact any kind of logical (*i.e.* formal) thought: "put a set of premises into such a device and turn the crank, and it will readily pass out conclusion after conclusion, all in accordance with logical law." The difficulty would be in getting the premises or formulae *into* the computer in a suitable representation.

5. Retrieval

So far, Bush's acquisition and storage mechanisms seemed reasonable, but they had a major shortcoming for retrieval of information: images could not be searched and

[4] The Vocoder [3] was an electronic speech coder designed by Homer Dudley at Bell Labs in the 1930's. A spectral analyzer component converted sounds into codes, and a subsequent synthesizer component used filters excited by noise to reproduce the original sound. Dudley's work forms the basis of modern, low-bitrate sound compression, *e.g.* in telephony and RealAudio. The Voder, which was essentially the Vocoder's speech synthesizer driven by a keyboard, was demonstrated at the 1939 World's Fair.

retrieved by machine. Bush recognized a need for digital indexing of documents, that is, associating a machine-readable index code with each document, so that a machine could identify and retrieve a document from mass store automatically. He termed this retrieval *selection*. Punch card selectors had existed for decades, but Bush extended the idea to larger, faster systems. In the paper, he gives the example of a department store accounting system, where each transaction combines punch cards from the customer, salesperson, merchandise, and general accounts. The general accounts card(s) would have to be selected and updated by machine in a back room, *e.g.* matching customer codes to update billing, *e.g.* matching merchandise codes to update inventory.

Bush proposed several mechanisms for machine readable representation to speed up selection. Whereas punch card readers were traditionally mechanical, Bush proposed reading their dots using photocells, and further, storing those dots on microfilm. Microfilm was not only faster, but it had the ability to store both a document and its index code(s) together. This unified, searchable storage would be a key component of Bush's *memex* (described below), and Bush had already developed the technology at MIT in the *rapid selector* project[5]. Bush further proposed that dots could be stored magnetically on a sheet of metal, but the mention was only in passing and was devoid of detail. Nevertheless, the prediction would be realized in 1956 with the invention of the IBM hard drive, which quickly became the dominant technology for rapid document retrieval.

6. The Memex

Having described a groundwork of enabling technologies for information acquisition, storage, and retrieval, Bush then put them together in the *memex*, a hypothetical, desk-size information work-station. It contained a massive amount of microfilm, perhaps a library's worth, including purchased material as well as personal documents that would be scanned or penned in. Documents could be recalled by document code and page number, as in the *rapid selector*, then displayed on one of two screens. Bush's key observation was that in addition to rapid selection by number, an essential part of the system was *associative* annotation and selection: arbitrary pages could be linked together by the user, forming a "trail" of documents that could be later recalled. This is probably the first description of hypertext, which clearly is a good idea—*c.f.* the World Wide Web. The technical mechanism for Bush's trails was simply that each page of microfilm included a list of index codes of other pages, in the usual form of positional dots, and the user would push a button to record in that list a link between the two documents presently being displayed.

---

[5] The *rapid selector* was a device to automate microfilm document retrieval, using photocells to search for a particular pattern of dots on microfilm. It was prototyped by Bush and his graduate students at MIT, 1938-40. The technology was rumored to have been used for cryptoanalysis during WWII and was later redirected for libraries at the request of the US Dept. of Commerce (being built by ERA, a company employing Bush's former students). Though the device was never widely used, it was tested and publicized by librarian Ralph Shaw. Photocell-based microfilm selection dates back at least to Emanuel Goldberg's 1927 patent in Germany and its use in several Zeiss-Ikon machines in the 1930's, but Goldberg's name was buried by the Nazis, and Bush was apparently unaware of his work. [4]

Bush recognized the tremendous value of trails. He envisioned entire texts being distributed with built-in trails, and scientists sharing custom-made trails by copying the associated pages and links. The trail concept is quite compelling, and we could not adequately determine why it has not been embraced over the last half-century. Even on the World Wide Web, which is a trail-based system, only recently have there appeared tools for user trails and annotations (*e.g.* A9 [5]), and they have yet to gain widespread popularity, despite the fact that pretty much everyone agrees that user trails can be extremely valuable.

7. Limitations and Mispredictions

While the bulk of "As We May Think" was firmly grounded in the technology of the day, Bush appended a coda that made some far-reaching predictions about information acquisition: since humans operate by electrical impulses, whether in the brain or in signals to muscles, it might be possible to feed these impulses directly into machines, rather than having to endure a tedious and lossy translation to and from a kinetic (mechanical) version first. This is a pretty cool idea, and scientists are just now beginning to explore the opportunities it presents, *e.g.* in prostheses.

Bush made an impressive number of accurate predictions about the future of science: the beginning of the information age, the power of computers to complete non-creative tasks, association as a recall mechanism, and so on. However, he did fall short in a few significant areas.

Bush's notion of information content (images) was not digital or symbolic. That is, he did not envision content in a format that a machine could search or analyze at anything but the coarsest of levels. Consequently, he had to rely on manual indexing of documents: all trails and index codes had to be painstakingly constructed by hand. There was no way, for instance, to find all pages in the record containing the word "orangutan." The closest method would be for an author to construct a viewable, hyperlinked keyword index (as in the back of a textbook) for a particular document or collection. The lack of keyword searching might be a critical reason why the memex never saw widespread use, though there are clearly more mundane technical problems stemming from the use of microfilm (*e.g.* the lack of microfilm dry photography, hampering user scanned content). Bush's reliance on microfilm was necessitated by the lack, at the time, of any technology for digital mass storage. Interestingly, it is his student, Claude Shannon, who would complete the digital picture, in defining the nature of digital information, digital compression, and the use of Boolean algebra for digital computation.

Bush's version of hypertext corresponds quite closely to the idea of the World Wide Web without search engines—useful, but crippled in a significant way. Given this similarity, and the popularity of the web even before the advent of powerful search engines, one might wonder why the memex idea was not adopted in some form or another in the first several decades after its design. A plausible explanation stems from another key observation missed by Bush: the concept of networking.

In Bush's vision, entire libraries were replicated and stored locally for each memex user, and trails were constructed independently, to be shared only by hand. Not only was this economically wasteful, it was also intellectually wasteful, as the same trail might have to be independently created several times over. It is clear now, with the Internet, that networking greatly increases the efficiency of knowledge dissemination. Each new piece of knowledge is published once and shared on demand, rather than published hundreds or thousands of times, per user. Furthermore, a fast and searchable network gives the user immediate access to the newest knowledge of the *entire* (online) record, rather than waiting for acquaintances and publishers to provide copies. The ease of information sharing in turn produces a more comprehensive and fertile public record.

Bush also seemed to underestimate the value of science and the record for the layman. He accurately predicted the use of the record in laboratory research, business accounting, and law. Yet he neglected to see that it could affect the greatest number of people through entertainment and recreation, as in television, gaming, product information, and event/travel planning.

8. Conclusion

Vannevar Bush's contributions took two disparate sides. On the one hand, he spent many years devoted to the application of science to the military, including supervising the Manhattan project and later directing and advising companies in the military industry. On the other hand, he firmly believed that science could better the life of man in times of peace, as evidenced by his academic work at MIT and his paper "As We May Think." That dichotomy reveals a balance and an optimism towards both science and the military, seemingly typical of America in the 1940s and rarer today, predating a growing cynicism borne of such things as Vietnam and the exploding national debt.

In "As We May Think," Bush envisioned a world where the record of human knowledge were accessible at the touch of a finger. His vision was a kernel for the information age, and his idea for hypertext would be a foundation for the World Wide Web. While his vision was far reaching, it was clearly limited by his understanding of the technology of the day. He discounted the promise of digital information and of rapid information sharing by network. As such, Bush fell victim to a common flaw of technologists, to overestimate the impact of a technology in the near term and greatly underestimate its impact in the long term. The *Jargon File* [6] (*New Hacker's Dictionary*) lists Bush's eponym, *vannevar* as a, "bogus technological prediction… *esp*. one that fails by implicitly assuming that technologies develop linearly." Nevertheless his ideas from "As We May Think" had a lasting impact, felt on every computer desktop today.

9. References

[1] Vannevar Bush, "As We May Think," *The Atlantic Monthly*, July 1945.

[2] Vannevar Bush, "Science The Endless Frontier", United States Government Printing Office, Washington, July 1945. Internet: http://www1.umn.edu/scitech/VBush1945.html

[3] “Homer Dudley’s Speech Synthesisers, ‘The Vocoder’ (1940) & ‘Voder’(1939).” Internet: http://www.obsolete.com/120_years/machines/vocoder/
[4] Michael K. Buckland, “Emanuel Goldberg, Electronic Document Retrieval, And Vannevar Bush's Memex,” *Journal of the American Society for Information Science*, vol. 43, no. 4 (May 1992): 284-294. Preprint available on internet: http://www.sims.berkeley.edu/~buckland/goldbush.html
[5] Internet: http://www.a9.com
[6] *The Jargon File*, Internet: http://catb.org/esr/jargon/html/V/vannevar.html
[7] “Events in the Life of Vannevar Bush.” Internet: http://www.cs.brown.edu/research/graphics/html/info/timeline.html
[8] “Internet Pioneers Vannevar Bush.” Internet: http://www.ibiblio.org/pioneers/bush.html
[9] “Vannevar Bush: the Man Behind the Scenes of the Atomic Bomb.” Internet: http://www.doug-long.com/bush.htm