Myths about the first hard disk drive

Googling *the first HDD weighed over a ton and stored 5 MB of data* can produce over 100,000 hits. Both values are incorrect.

As a participant in the storage industry for almost fifty years I am aware of a number of apocryphal disk drive stories but in 2008 while vetting the storage section of the Computer History Museum’s Revolution Exhibit I encountered and resolved one of many myths about the first hard disk drive (“HDD”); the Golden Gate Bridge Paint Myth – the coating of disks of the first HDD was not similar to the paint on the Golden Gate Bridge.

In modern times HDDs are shipped on or about their announcement dates, with a nominal published price that is volume dependent and declines over time. In the 1950s when the first HDD was announced, it was industry practice to pre-announce products and to provide long lead times from the official announcement to actual production shipment. Pricing rarely changed over the life of the product. As a consequence of these 1950s practices a number of myths exist about the first HDD resulting in misstatements of progress to date. This article is my attempt to clarify some original parameters of the first HDD. I then use the clarified parameters to measure the progress of the HDD industry.

**Names and dates**

What’s in a name – the first HDD has had a number of labels, the unambiguous name is the IBM 350 disk storage (“IBM 350” hereinafter). It is frequently called the RAMAC disk drive, but that term has at least two meanings.

Two HDDs were announced at an IBM press conference on September 14, 1956, one as the IBM 350 Random Access Memory, a bundled component of the IBM 305 RAMAC system and the second as the IBM 355 Random Access Memory, a required component of the IBM 650 RAMAC system. At announcement five engineering prototypes of the IBM 350 had been installed commencing with a June 1956 installation at Zellerbach Paper, SF CA. As many as fourteen engineering prototypes, labelled IBM 350A disk storage units shipped before production unit shipments began in November 1957. The production units were labeled IBM 350 “disk storage” or “file” in product literature and sales manuals. The “A” units were all replaced by production units and for the most part scrapped.

The second model, subsequently labeled the IBM 355 disk storage unit was simultaneously announced; its first delivery date is uncertain and was likely well after production IBM 350 units, its volume appears low and because of its multiple actuators its price was substantially higher than that of the IBM 350. There are no known test models prior to production shipment. Accordingly, the IBM 355 is not considered to be the first HDD.

As was the practice, there were a number of pre-announcements of the IBM 350 beginning as early as May 1955. On September 4, 1956, IBM internally announced the drives and the systems and this
date is sometimes cited as an announcement date; IBM’s archives going so far as giving September 4 for the drive announcement, but inconsistently stating September 14 as the announcement date for the system that used the drive.

There is no example of IBM officially designating any disk drive as RAMAC. The term was used by IBM San Jose employees interchangeably to refer to both the IBM 350 disk storage (i.e. “RAMAC drive”) and the associated 305 RAMAC system (i.e., “RAMAC system”) and the usage for the HDD apparently spread through the industry as engineers left IBM San Jose for other firms in the HDD industry. In 1984 the American Society of Mechanical Engineers designated its 16th International Historic Landmark as the “IBM 350 RAMAC Disk File,” a “technology that spawned an industry.” The term RAMAC is ambiguous so this article uses IBM 350 for the first HDD, but it should be understood that many observers use the term RAMAC as a synonym.

Disk drive progress is tracked both in the laboratory and by production shipments. Accordingly, progress measurements hereinafter are based upon a November 1957 first production shipment of an IBM 350. Most observers set their measurement baseline in 1956 inappositely applying data of the June 1956 engineering prototype to the measurement of later production units.

**Weight**

The IBM 350 weighed 1730 lbs. (785 kg) slightly less than one ton (2,000 lbs., 907 kg). The “more than one ton” belief may include the 441 lb. (200 kg) weight of the separate air compressor unit.

**Capacity**

Multiple sources have the IBM 350 with an equivalent capacity of 4.4 MB (megabyte) or 5.0 MB; the most appropriate value is 3.75 MB.

There is no dispute that the IBM 350 stored five million “characters” but what were those characters in terms of yesterday’s megabytes? The IBM 350 recorded its characters using an 8 bit recording code, of which 6 bits represented data plus a Parity bit and a Space bit. It is common in storage recording for there to be more bits recorded on the medium than there are data bits but the storage capacity is stated in terms of data bits, resulting in an equivalent IBM 350 capacity of 5 million \( \times \frac{6}{8} = 3.750 \text{ MB} \).

The IBM maintenance documents disclose a nominal recording frequency of 97.6 kbit/sec and a 1200 RPM rotational speed from which an unformatted capacity of 6.09 MB can be calculated. Today HDDs are specified only in formatted capacity but there was a time when drive manufacturers specified both the unformatted capacity and the formatted capacity with one or more approved formats, e.g. drives offered as compatible with the Seagate ST412 in the 1980s were specified by Seagate as 10.0 MB (formatted) and 12.6 MB (unformatted). IBM never published directly its unformatted capacity as indeed it is no longer published for modern drives; however, from this analysis the derived specification for the IBM 350 is 3.75 MB (formatted) or 6.09 MB (unformatted).
A capacity of 4.4 MB\textsuperscript{25} (actually 4.375 MB) can be calculated using a 7 bit character\textsuperscript{26} (5 * 7/8 = 4.375). There is no teaching of a 7-bit character in any of the contemporaneous IBM 350 documentation. No proponent of 4.4 MB has ever stated a valid basis for using which 7 of the 8 recorded bits. Some 7 bit proponents concatenate the 6 data bits plus the Parity bit to arrive at 7 potential data bits without justifying just the elimination of the Space bit. It is well established that magnetic recording has a soft error rate, requiring as a minimum some form error detection, e.g., parity, in order to assure the integrity of the played back data. It is certainly possible to record using the parity bit as a data bit; the playback would not be reliable and IBM would certainly have voided the warranty of any user so foolish.

An even more widespread myth is a 5.0 MB capacity as demonstrated by the Google search cited at the beginning of this article. This may simply be the lack of understanding of the difference between the 5 million 6-bit characters of the IBM 350 and 5 million 8-bit characters of a modern HDD. One justification, so far unpublished, would involve a one-for-one mapping of the 8 recorded bits of into an 8 bit byte for some form of emulating an IBM 350 with a modern drive. Such emulation would take 5.0 MB of a modern HDD but since two of the eight bits contain no data, it is not valid to say that it is equivalent to 5.0 MB of data in today’s HDD.

**Price**

HDDs then were large complex machines with many field replaceable parts and adjustments, so much so that a large installation could demand and get an on-site customer (service) engineer for maintenance purposes. Detailed maintenance documents revealed much about the feeds and speeds of the IBM 350\textsuperscript{27} but little has been published about its price.

In 2006 a speaker at the 50th anniversary celebration of the “RAMAC” (IBM 350) stated its purchase price as $50,000\textsuperscript{28} which has become a repeated myth. Research reveals that actual purchase price was $34,500 in 1957, about $300,000 today.

One reason for the price confusion is that at the time of announcement IBM did not necessarily offer products either separately or for purchase\textsuperscript{29}. The IBM 350 (later IBM 350 Model 1) was announced as a bundled component of the IBM 305 RAMAC system – the system offered at a monthly rental of $3,200. Noted historian Emerson Pugh, reported that the IBM 350 represented $650/month of this total\textsuperscript{30}. IBM’s archives were unable to locate any pricing information on this first disk drive; however, they provided a copy of the May 1958 announcement of a “second disk storage for the 305 RAMAC” system, the IBM 350 Model 2, with a rental price of $700/month and a purchase price of $36,400. The original IBM 350 was designated the Model 1. IBM was quite rigorous in establishing its purchase price to rental ratio so the $700/month rental of the 350 Model 2 strongly indicates the purchase price of the Model 1 should have been less than $36,400 and not nearly $50,000.

In 1961 the US Army’s Ballistic Research Laboratory Third Survey\textsuperscript{31} disclosed six prices\textsuperscript{32} for IBM 350 disk storage, four with a rental price of $650/month (presumably Model 1’s), one at $700/month (presumably a Model 2) and one at Boeing Wichita, with a purchase price of $34,500. Considering the

© T. Gardner, 2014, 2018 all rights reserved
strong agreement with known IBM rental prices, I conclude the IBM 350 Model 1 purchase price was $34,500.

This gives one bases of performance as $9,200/MB in 1957 dollars or about $80,000/MB in 2017 dollars33.

**Today’s manufacturing cost**

Western Digital one of the three remaining HDD manufacturers with about a 50% market share reported an average manufacturing cost at year end 2014 of $43 per HDD.34 Due to the highly competitive market today Western Digital is a reasonable surrogate from a cost perspective for IBM, 1957’s only HDD supplier. IBM has not willingly published costs on the IBM 350; however, with a reasonable assumption that IBM’s gross margin on the IBM 350 was at least 60%,35 it is possible to estimate that the manufacturing cost per box has declined from about $14,000 to $43, about 320-to-one in current dollars or 2800-to-one in constant dollars while at the same time the capacity per box has increased by about 300-million-to-one.

**Today’s price per unit capacity**

The market for HDDs is now very different than in 1957 when there was one product, sold by a direct sales force to the government and enterprises. Today most HDDs reach the end user bundled in a laptop or desktop without out any visible price. Retail prices are posted on the internet and in printed publications for all sorts of HDDs but such prices are very volatile depending more on inventory conditions and cash needs in the channels than actual costs. Selling below cost is known to happen. Thus, it is difficult to aggregate the many such prices and capacities so this information is not a particularly good measure of the progress since 1957.

Backblaze36 a large buyer of lowest cost per gigabyte HDDs for cloud storage is a reasonable surrogate for IBM’s 1957 customers in terms of sales channel and need for low price. Figure 1 depicts actual price per GB paid by Blackblaze second quarter 2017 and includes for reference the IBM 350 price per MB extended assuming this price declined at a Moore’s Law rate, that is, decreasing by a factor of two every 24 months. The break in the Backblaze curve is the 2011 floods in Thailand.

The graph shows the current price per GB actually paid by Backblaze is substantially less than that which one would have paid had the HDD industry only improved this parameter at a

© T. Gardner, 2014, 2018 all rights reserved
Moore’s Law rate. Backblaze’s year end 2014 price per GB of 3.75 ¢/GB is about a 2-billion-to-one improvement in price/performance in constant dollars, representing a compound annual improvement of 45.4% per year, a rate well in excess of a 41.4% Moore’s Law rate. Price per gigabyte has continued to fall albeit at a somewhat lower rate.

**Recording density today**

Much of the progress of HDDs has been driven by the exponential growth in the density of the information recording. The IBM 350’s 3.75 MB was stored at a density of 2 million bits per square inch and used both surfaces of fifty 24-inch (610 mm) diameter disks. As of 2018 the highest known HDD density announced to ship was in 2016 with an areal density of 1.3 Terabits per square inch, a 650 million-to-one improvement - the entire capacity of the IBM 350 could be stored on a small fraction of one surface of today’s 2½-inch HDD.

Figure 2 depicts the areal density improvement this century and depicts the IBM 350 areal density extended assuming it increased at a Moore’s Law rate.

Over nearly 60 years areal density of real products has progressed at varying rates, sometime slow and sometimes fast but looking back to the first production shipment area density improved thru 2016 at a rate that slightly exceeded a Moore’s Law. However the current rate of improvement is well below a Moore’s Law rate and is projected to remain so. Thus it is likely that the next announced areal density milestone will fall below the Moore’s Law trend line.

**Summary**

The HDD industry has improved technical and price performance at rates exceeding a Moore’s Law rate for fifty-nine years, a period longer than Moore’s Law has applied to semiconductor complexity. HDD technology, like semiconductor complexity, has lately improved at lower rates so it is unlikely these measures of HDD performance will continue at such a high rate. But the fifty-nine year progress has been remarkable.
Acknowledgements I’d like to thank the staff at IBM’s corporate archives that have been extremely cooperative in finding and providing copies of contemporaneous documents. Backblaze publishing its purchase price history is a service to our industry and I greatly appreciate their providing the underlying numerical data. I also appreciate the documentation rigor demanded by my colleagues at the Computer History Museum and by the several somewhat anonymous editors on Wikipedia.

About the author: Tom Gardner is a consultant to the storage industry having had assignments with storage companies such as Auspex, Memorex, Network Appliance, Seagate, SyQuest, Toshiba and Western Digital. He is a volunteer at the Computer History Museum and with the IEEE Silicon Valley Technical History Committee. He holds a BSEE from Brooklyn Polytechnic, an MS, Engineering Science from Rensselaer Polytechnic and an MS, Management from the Stanford Graduate School of Business. All opinions expressed herein are his own and may not reflect the opinions of any organization with which he is affiliated.

---

1 The Golden Gate Paint then was a mixture of Linseed Oil as binder with Lead Oxide as pigment while the IBM 350 disks used various other binders, ultimately an epoxy, Iron Oxide as the pigment and other additives, see: [http://chmhdd.wikifoundry.com/page/Golden+Gate+Bridge+Paint+Myth](http://chmhdd.wikifoundry.com/page/Golden+Gate+Bridge+Paint+Myth)

2 “305 RAMAC and 650 RAMAC, two electronic data processing machines using IBM's random access memory, a stack of disks that stores millions of facts and figures less than a second from management's reach.” IBM Press Release, September 14, 1956. See: [http://www-03.ibm.com/ibm/history/exhibits/650/650_pr2.html](http://www-03.ibm.com/ibm/history/exhibits/650/650_pr2.html)

3 “… the huge electronic data processing machine equipped with the “juke box” memory file, known as RAMAC for Random Access Memory Accounting, was announced by press, radio and television September 14.” IBM San Jose News, IBM Corp., September 1956, p. 1

4 “The heart of the [305] system is the IBM 350 Random Access Memory [later disk storage] with 50,000 addressable locations storing a total of 5,000,000 alphameric characters.” IBM Letter 256-114, September 4, 1956.

5 Originally announced as the 650 RAM system, the name was soon thereafter changed to 650 RAMAC to avoid copyright issues.


7 Ibid. Before September 13, 1956 one unit had been installed at the US Navy, Norfolk VA

8 Associated Press news article distributed circa November 1, 1957, as e.g. “RAMAC Joins Big Airline,” The Indiana Gazette, Indiana Pennsylvania, November 4, 1957, p.22. See also, “305, 608 NOW IN OPERATION,” Datamation January-February 1958. “First production models of two new IBM computers, the 305 RAMAC and the 608 have been delivered to customers.”

9 For example, the U.S. Library of Congress Copyright Office Catalogs of Copyright Entries, from 1957 to 1960, have four listings for “IBM 350 disk storage unit;” none of the many IBM RAMAC listings are associated with the IBM 350.

© T. Gardner, 2014, 2018 all rights reserved
“The latest addition to the 650 [system] is the IBM 355 random access memory [later disk storage], a storage medium in which any group of data may be reached quickly and directly, without search.” IBM 650 press fact sheet, September 14, 1956. See http://www-03.ibm.com/ibm/history/exhibits/storage/storage_355.html

Op cit. IBM Press Release, September 14, 1956; production of both the 350 and the 355 was promised for “mid-1957” thus there were at least two and perhaps as many as fourteen 350’s shipped before any 355 shipped. The 350 did slip from mid-1957 to late 1957.

IBM states over 1,000 305 systems were shipped thru 1961, each one having at least one 350; some having 2. According to Phister, Data Processing Technology and Economics, Second Edition, © 1979, the number of first generation HDDs installed on all IBM systems in 1961 was 900 units leaving very little room for 355s on 650s. A low number of disks on IBM 650 systems is supported by a 1962 survey of educational computing that identified only 2 RAMAC units out of 38 installed IBM 650s ["Computing In The University," American Mathematical Society and the Ohio State Research Center, Datamation, May 1962]

Recent discussions with development and production personnel who worked at IBM San Jose in the late 1950s.

IBM Press Conference, May 6, 1955. IBM announced that it was bringing out a new product “that takes information from a stored program using a multi-million character random access memory that makes it possible for the new system to do a whole job automatically without using batch processing.”


The term RAMAC has been used officially by IBM to refer to the 305 RAMAC system, the 650 RAMAC system and to a series of RAMAC Virtual Arrays. To this author’s knowledge it has never been directly applied to the IBM 350 disk storage in any product or sales literature. Informally many in the disk drive industry use RAMAC as a synonym for the IBM 350.

See for example, “Emerging Trends in Data Storage on Magnetic Hard Disk Drives,” Ed Grochowski, Datatech (September 1988), ICG Publishing, pp. 11-16. Figure 2, Areal density projection, separately displays trend lines of availability dates for “demo” units and for production units.


Op cit. IBM CE Manual, pg 7 and Figure 86 describe the 8 bit positions in each character as comprising S, X, 0, 1, 2, 4, 8, and R bits, with S being the Space bit merely providing a space and bit R having no numeric or alphabetic value, but is added to certain characters so that every character will have an odd number of bits, that is, odd parity. The six data bits then are X, 0, 1, 2, 4 and 8.

A simple example is the CD-ROM which encodes 8 data bits into 14 channel bits plus 2 separation bits and then further embeds a block of 2048 data bytes into a 2352 byte block containing control and correction bytes. There are thus (2352/2048) * (16/8) = 2.4 times as many bytes recorded on a CD-ROM as is its stated capacity in bytes.

© T. Gardner, 2014, 2018 all rights reserved
The published CD-ROM capacity is always given in terms of data bytes and not the more than two times as many bytes actually recorded on the medium.

22 The IBM 355 is described as having a 6 million “character” capacity but these are digit characters (0-9) that can be represented by 4 bits rather than six bit alphameric characters so the capacity of the IBM 355 in modern terms is 3.00 MB (6 megadigits * 4/8).

23 Op cit. IBM CE Manual, each sector is 1/5 a revolution with a gap of 1800 µsec and 8000 character bits
Bit time = (sector time – gap time) / (bits/sector)
= (10000 µsec – 1800 µsec)/8000 = 10.25 µsec/bit or 97.6 kbit/sec

24 Op cit. IBM CE Manual, 10000 tracks * 50 msec/track * 97.6/8 kbyte/sec = 6.09 MB

25 See e.g., Computer History Museum, “RAMAC actuator and disk stack,” Memory Size 4.4 MB
http://www.computerhistory.org/revolution/memory-storage/8/233/923

26 e.g., Op cit., ASME RAMAC, “... a total capacity of 5 million binary decimal encoded characters (7 bits per character) ...


29 Its 1956 Consent Decree required IBM to offer its products for purchase not later than January 25, 1957.


32 Ibid. Fourteen prices were disclosed but eight applied to the later double capacity versions of the IBM 350, Models 11 thru 13.


34 WDC Quarterly Fact Sheet – Q4FY14, Western Digital Corp.

35 IBM’s gross margin on equipment sales from 1958 to 1960 was in the range of 50% to 55% of equipment revenue [Phister, p.49]. According to multiple IBM sources, HDDs were generally among the most profitable of IBM’s equipment product lines well into the 1980s, at times approaching 90% gross margin. Thus a 60% assumption for the average IBM 350 gross margin is conservative.

36 “Farming hard drives: 2 years and $1M later,” Backblaze Blog, November 26, 2013
Backblaze is a cloud backup service company located in Silicon Valley, CA. At the end of 2013 they had 27,134 consumer grade drives installed. See:

© T. Gardner, 2014, 2018 all rights reserved
Price and price/megabyte, of course have declined so the compounded rate is actually negative but is expressed herein as a positive number, corresponding to megabyte/price, for comparison to the well-known Moore’s Law rate. Moore’s Law restated and applied to HDDs would be price/megabyte decreasing by a factor of two every two years, a -28.7% compound annual rate with the HDD performance being a somewhat faster rate of -31.3% over the period.

Seagate 5TB 2½-inch ST5000, announced for availability in October 2015. Areal density from Seagate manual 100804767, Rev. F