Storing the information age: 50 years of tape storage innovation

By Wayne Glynne, Storage Business Manager, IBM Australia and New Zealand

The setting - June 1949

Recorded music was on 78 rpm platters and "wire recorders" were the mainstay of radio. Magnetic tape had only recently emerged from the laboratory for use in a few radio broadcasts. A room full of vacuum tubes and miles of wire accomplished what a hand-held calculator would be able to do faster by the 1980s. Information storage meant books, filing cabinets, or -- to those at the leading edge of data processing technology -- paper punch cards. Reels of tape, tape cartridges and programmable computers were the stuff of science fiction.

All that was about to change.

The challenge

A more compact means of storing data permanently *had* to be developed. For example, the US social security administration needed a way to solve its data storage problem -- acres of filing cabinets filled with punch cards containing social security data on every working American.

The limitations of punch cards were becoming obvious. There were 80 characters per card and a good speed was 100 cards per minute. That's 133 characters per second. The first marketed tape drive, the IBM 726, operated at 7,500 characters per second, 56 times faster than the punch card rate.

Developers tried more dense paper cards for data storage, but the improvement was only a few times better. Punched paper tape offered no great advantage either. Magnetic tape had just come into its own in audio equipment and offered the most promise. Eventually, it would be shown that even at the first working density of 100 bits per inch of a half-inch wide tape, a 10.5 inch diameter reel of tape could hold the equivalent of more than 35,000 punched cards. Today, a high-end LTO Tape Cartridge is 1,000 times faster and stores more information in one inch of tape than the 726 Tape drive was capable of storing on over 500 feet of tape ... without data compression.

It was by no means clear that wisdom lay in gambling millions of dollars, the careers of some of the brightest engineers in America, to say nothing of the future of the then 35-year-old IBM Corporation, on the infant technology of magnetic tape that was based on gluing bits of rusted iron onto strips of plastic.

The suggestion that such a vaguely understood technology as 'magnetics' could supplant the familiar and highly profitable punched card, met with strong and vocal opposition within the company. Even some of the most highly respected and positioned figures within IBM were convinced it was foolish.

Tom Watson Jr, by then executive vice president of IBM under his father T.J. Watson Sr, immediately grasped the promise of magnetic recording as a result of his experience with electronics in the Army Air Corps during World War II. He and a number of the company's top engineers and scientists decided that the gamble of moving to magnetic tape *had* to be taken. According to one IBM Fellow, Watson Jr *"went all around the room asking people if it was the right thing to do or not..and some people said "yes" and some people said other things. Then he told all those people who said other things that maybe it was time they work on other problems."*

The solution

With support from the senior executives, a handful of IBM's population in Poughkeepsie NY, set to work. No one single person came up with the idea of magnetic tape. It grew out of a number of meetings and discussions.

Magnetic recording was an unexplored topic in 1949. As in most creative efforts, improvisation would play a role in the development of the magnetic tape drive. Very fast starts and stops were desirable in order to minimise the wasted tape and read-write delays. With tape speeds anticipated to be 100 to 200 inches-per-second, it was impractical to accelerate bulky tape storage reels rapidly enough. Storage loops would be necessary in the tape path for gradual acceleration of the reels. The vacuum column provided this loop of tape and allowed the early prototype to mimic punch card records.

Once the idea of the vacuum column was hit upon and the vacuum switch conceived, developers needed some very thin, flexible material in order to fabricate a sensitive pressure-sensing diaphragm. With nothing suitable to hand, the quickest solution that occurred to the engineers, was to run to the nearest drug store for a pair of rubber baby pants. They worked.

The vacuum pump, necessary to produce the vacuum in the columns, provided another challenge. Throughout development, the motor from an old General Electric vacuum cleaner was used. When the tape drive was released for production, that same motor was included because the prototype drives had been optimised for it. It turned out that this was an *obsolete* vacuum cleaner. A GE sales rep searched warehouses all over the country to get us under way.

It was a time of some confusion. At one point a shipment of tape was delivered to the receiving dock. The man in charge came over and said, 'We just got a shipment of tape from 3M, but we're going to have to send it all back' Why? 'It doesn't have any glue on it!' "

IBM's first magnetic tape unit, the IBM 726, was announced in 1952 - the year Thomas J. Watson Jr became IBM president and the employee population passed 40,000. It was announced with the IBM 701 Defense Calculator, the company's first commercially marketed electronic computer. A development model was called the Tape Processing Machine

by the engineers who designed and built it – an indication of the pivotal role played by magnetic tape in launching the data processing industry.

Tape technology

Throughout the 1950s and 60s, magnetic tape units offered successive improvements to data storage. By the 1970s, the world had begun moving beyond an industrial paradigm into an information age.

To improve the operation of tape and move some of its control function out of the central processing unit (CPU), some of the instructions were first moved to the tape drive. This was the beginning of tape microcode, or embedded instructions for the control of tape. Drives introduced in the early 1970s increased that function considerably and also debuted false-symptom codes for real time diagnosis of problems and failures, a function currently taken for granted on all storage devices.

Devices for processing and storing information became increasingly specialised and sophisticated, offloading tasks to specialised peripherals instead of relying on the CPU. These devices continued to add function and hardware capability so rapidly that within a few generations, what was once considered a mainframe resided in a peripheral device. Powerful personal computers would soon be readily available to the average consumer. Day-to-day transactions were becoming increasingly computerised and contributing to the explosion of data being generated ... and needing to be stored. At the same time, customer storage requirements and expectations grew. Tape storage and management needed to improve in reliability and performance and to accommodate a growing number of operating systems and applications. Enter automation.

IBM produced the first automated tape storage device for information in 1974, bringing with it several innovations that in turn, led to further enhancements to the management and architecture of data storage. The first cache control device was developed in 1984 as a cache controller for disks, and created the foundation for improved storage controllers. This initial cache control unit led to advances in data storage control function that optimised the use of both DASD and tape storage.

Up to this time, the industry standard had been a 10.5-inch round tape reel that stored 180 MB of data and had an uncompressed data rate of 1.25 MB per second. A breakthrough was the introduction of a 5.5-inch square cartridge, storing up to 200 MB data, and executing at the rate of 3 MB/sec, all the while requiring *less than half* the floor space of an equivalent installation of magnetic tape units from just a decade earlier.

Enhancements

A series of enhancements and firsts for tape storage quickly followed. Improved Data Recording Capability (IDRC), introduced in 1986, provided a significant improvement in data compression for tape. The ESCON (Enterprise Systems Connection Architecture) channel interface delivered in 1987, enabled tape storage across several kilometers - a distance previously unheard of.

In 1991, tape storage saw the introduction of the 36-track drive and a new extended-length chromium dioxide tape, providing 800 MB of storage. With the Improved Recording Capability (IDRC), capacity was expanded to more than 2.4 GB, the highest data capacity available at that time.

A significant increase in the embedded control function (microcode) enabled optimisation of the operation, reliability, performance and maintenance of tape storage. Tape-control microcode had become a critical component of high speed, high reliability tape as well as disk storage. Lessons learned with one, helped improve the other.

Changes

Direct Access Storage Devices (DASD) or hard disk drives, introduced in 1956, had undergone dramatic increases in capacity and areal density. At the same time cost-per-gigabyte had dropped steeply. New storage devices using optical recording technologies were being developed. New tape formats, arrays of disks, optical devices and other technologies entered the marketplace and threatened to relegate tape storage to nothing more than backup or archival repositories. Similar to earlier predictions of the end of the mainframe era, forecasts of the demise of tape storage were shortsighted.

Management of multiple tape formats and architectures became an increasing problem for large and midrange customers throughout the 1990s. At the same time, e-business was becoming a requirement for a company's success and competitive advantage. The complexity of servers, extended and shared databases, security issues and connectivity across multiple locations were common components of the IT industry. A seamless, non-disruptive and efficient solution needed to be developed for data storage management.

In 1993, new tape drives were introduced with the advanced connectivity of ESCON and then FIBRE support. More efficient and higher-capacity tape library dataservers followed shortly thereafter. Improved management of storage for disk and tape along with innovative cache control systems and software accompanied these devices.

Once again, the CPU and by this time *networks* of CPUs, were freed from managing storage. Control of tape volumes migrated to increasingly powerful library managers that were, essentially, "computers" residing in the library itself for the express purpose of managing the flow of data in and out. By the 1990s, the amount of data in a tape library could be hundreds of terabytes. By the end of the decade ... petabytes.

Advancements

Significant performance and connectivity advancement soon followed with tape storage devices that, for the first time, "virtualised" a tape image to the host and significantly improved the utilisation of tape resources. This optimised the filling of tapes while presenting a single image to the host

computer and operational software, dramatically improving and optimising tape storage and migration of large datasets.

From 1995 to 1998, as the network world was unfolding, tape developers introduced a series of products that uniquely improved customers' abilities to access and retain critical business information, and positioned them for the network storage explosion to come.

A revolutionary mid-point load, fast-time-to-data tape architecture was unveiled in 1996 with the first small-form-factor, midpoint-load tape device. Capacity and performance was extended to 7.5 GB with the enhanced capacity cartridge in 1997.

As the 20th Century drew to a close, Linear Tape Open (LTO) was announced and a new technology roadmap begun for tape. LTO was an aggressive effort established between IBM, Hewlett Packard and Seagate to develop a truly open systems tape solution. In less than two years, this technology went from laboratory to marketplace. The IBM Ultrium LTO drive brought with it new automation solutions and re-emphasised the cost advantage of tape storage. It also provided new fuel for the future.

Transition to a new century of magnetic tape storage

Innovations in tape storage that began mid-20th century, particularly those relating to density and performance, were being achieved at an accelerating rate. The IT world braced for the highly publicised 'Y2K disaster' as the 1900s made room for the 2000s. While clocks struck and then passed midnight without incident, tape storage provided much of the confidence that if things *had* gone awry, data recovery would have been successful.

The unfolding of a brand new century found 100 GB of native capacity already available on a single LTO cartridge. In enterprise storage, capacity was doubled by the introduction of Extended Length media. And, the performance and reliability of both the initial 10 GB media and the new extended-length media were attained through an intensive co-development with leaders in the magnetic media industry.

Still, the long-term relevance of magnetic tape was again questioned by those watching disk storage capacity increase and cost decrease. Remarkable advances were made in disk storage technology; and both disk and optical storage were challenging tape's cost advantage. Tape needed its own imaginative advances to exploit its cost benefits.

Advances in storage in the early 2000s included: the ability to partition physical libraries into a number of logical libraries, FIBRE channel support, peer-to-peer duplication of data, support of Linux and other Open Systems, new connectivity options across an ever increasing number of storage platforms. At the same time, the reliability, serviceability and overall cost of tape subsystems improved, while tape technology in general was called upon to hold increasingly greater amounts of the world's data.

Even though technological advances in storage were coming quickly, they often resulted in incompatibility with existing infrastructure and architecture. A more unified view of information storage had to be developed. The time had come for a new paradigm for data storage.

Storage in the 21st century and beyond

The LTO head was the first tape drive recording head to break away from the older, modular, closed and contoured head designs that have carried the burden of linear tape drive recording for so long. Armed with the proven capabilities of these disk head technologies, significant increases in tape track density were possible.

This improved head process allowed the mass of the head to be reduced while achieving synergy with disk head technology and process advances. A timing-based servo provides very precise position information to the drive so that interleaved bands of data can be written in a serpentine manner, eight tracks at a time, to construct 384 tracks capable of storing 100 GB of uncompressed data in the first generation LTO cartridge. Coupled with improved media, actuator and format innovations, significant increases in the reliable placement of very narrow tracks on a thin tape were demonstrated with the first LTO tape product.

The LTO family of tape drives is expected to deliver increased storage capacity and performance for several generations. They will provide a 200GB capacity cartridge capability in 2002 with a significant improvement in data from 15 MB per second to 30 MB per second. Tape storage currently provides over 500 TB of uncompressed capacity in a fully automated IBM Tape Library, in the space equivalent to eight, five-drawer file cabinets. These automated tape solutions are also able to support multiple tape formats in the same physical library separated over distances of over 100 kilometers.

Looking ahead: the 1-Terabyte tape cartridge

In May 2002, on the 50th anniversary of IBM's introduction of tape storage as a crucial part of computer processing and storage of information, IBM announced that one terabyte (TB) of uncompressed information was successfully written and more importantly, read back successfully, in a *single* half inch tape cartridge equivalent in size to the current LTO tape cartridge.

As enterprise customers move to consolidate their storage environments, the ability to store large volumes of data in a small footprint will grow in importance. The compression of one TB of data into a four-inch wide by five-inch long by one-inch thick cartridge exceeds a density of 1 gigabit per square inch on the recorded media. This was achieved using advanced particulate tape technology coupled with improved high density track placement made possible by utilising novel track following, timing based servo invented by IBM and the most advanced recording head technology used in linear tape drives.

With proven hard disk technologies and inventions unique to IBM tape development expertise, 1 terabyte is only a milestone...not a barrier. The future looks very good for not only improved capacity, but also significant increases in data rate, reliability and management of tape storage for the next 50 years. As long as tape offers a cost effective, reliable and high capacity storage capability, it will be an important part of the future of storage solutions.