TEMPERATURE OF DATA IN STORAGE DEVICES AND DATA STORAGE SOLUTIONS FOR THE NEXT DECADE

Kai-Zhong GAO[1, 2], *Sr. Member IEEE*
[1]International Business and Technology Services, Kaizhong.Gao@intlbt.com
[2]Argonne National Laboratory, kgao@anl.gov

Magnetic data storage technology, such as HDD has tremendous growth over the past 60 years, and only until recently, the areal density capability (ADC) has slowing down [1-3]. It was until last couple of years, 3D NAND based SSD starts to take over some of the HDD markets [4]. In our recent study, a standard data "temperature" definition is determined and we show that this temperature of data is an intrinsic property of digital data for storage, which links both data storage capacity and access speed of a device together [5]. We show that this data temperature definition is consisted with "hot" and "cold" data concept utilized in the field of data storage system field.

Using this standard data temperature definition, we are able to analyze all digital data properties in a given device, such as in HDD. Fig. 2 shows an analysis of a given HDD used under different conditions, the results show that the maximum average data temperature capability for a given drive, regardless how to use it, depend only on the total number of files on that particular device. As shown in Fig. 3, the HDD drive capacity is increased by several orders of magnitudes over the past few decades. The average data temperature capability in the HDD is actually reduced, as illustrated in Fig. 4. The peak temperature of the drive does not change much, but the average data temperature is monotonically reduced as the drive capacity is increased. Further analysis shows that this trend is independent to the choice of applications, and only depend on the technology choice. In other words, for all HDD based technology, it follows a universal line for a given application. Further study also reveals how 3D NAND based SSD technology replaced HDD enterprise market over the past few years.

Based on this analysis, we further show that the digital universe evolution, to certain aspect, can be mapped to the physical world. As the size of digital universe increases, the amount of low temperature data also increase rapidly. This leads to a new opportunity for "cold" data storage, or data archive storage technology. Currently magnetic tape recording and Blu-ray based data archive systems are widely utilized for data achieve applications, as part of cloud based data centers, such as the one used in Amazon, Facebook etc. A comparison of these technologies shows its potential and limitations for the coming years, as the digital universe continue to grow. Using the same analysis, we will also show how STT-MRAM, as an emerging technology will join the field, and in what market it is likely to replace existing technology.

For data archive technology, one of the limitations is the time to first data, i.e. its peak data temperature capability. The storage archive system based on Blu-ray or magnetic tape recording has its intrinsic limitations which prevent many applications. Another limitation is associated with the method on how to achieve data and how to remove obsolete data. The technologies that enable magnetic tape recording and Blu-ray for data achieve system have certain limitations which raise additional concerns beyond technologic and economic perspectives. Based on the analysis, we show that being able to remove unwanted data, is the desired feature, even for data achieve systems. For a grand scale, this also has profound impact for the data center operations. Proper technology associated with new generations of data achieve system is desired, which enable opportunities in the coming decades.

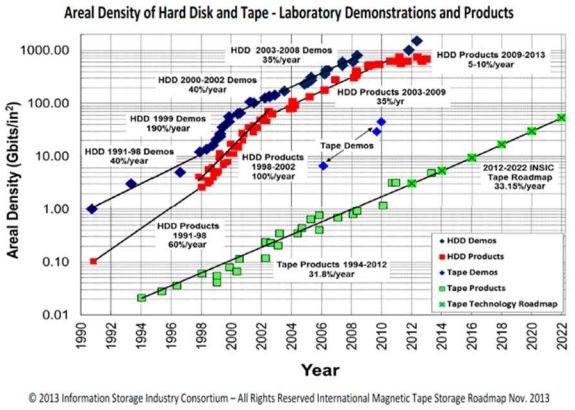
With large capacity HDDs in the market with reduced cost, HDD based data archive system also plays a role in data archive system, despite its higher cost than Blu-ray and tape systems. Beyond the energy assisted technologies, recent enablers for drive capacity growth is mainly based on system architecture change, such as: Helium Drive with more platters, SMR, Dual Readers, IMR and BMR etc. One of the advantages of HDD based data archive system is its peak data temperature capability. We show that for a carefully designed hardware system, it is possible to further extend its capacity for a storage system, sometime even based on a given, fixed drive configurations. In addition, an optimized solution enables a lower power consumption for HDD storage system, which can be significantly more efficient than any of the existing systems, such as the one used in Facebook and google cloud data centers.

Kaizhong Gao
kgao@anl.gov

Based on the same theory, we study future HDD technology potential. We show that although more data will be saved on HDDs, the overall HDD potential is limited. On the other hand, the HDD based data archive system, if fully optimized, has its potential in the following decades. As the HDD industry fully explore assisted technology and trying to integrate them into product. An interesting situation may occur where the decades long development of Heat Assisted Magnetic Recording may eventually be suitable for "Cold" (data archive) storage system.

REFERENCES

1) Information Storage Industry Consortium (INSIC) Tape and HDD Technology Roadmap 2012-22 (Link), 2015-25 (Link).
2) ASTC Technology Roadmap (Link)
3) K. Z. Gao, Write Architectures and Their Impact on Hard Disk Drive Capacity, Intermag, 2017 DE03.
4) L. Mearian, Flash memory's density surpasses hard drives for first time, Feb. 2016, Computer World (Link)
5) K. Z. Gao, Temperature of Data, News and Views, International Business and Technology Service 2016 (Link)



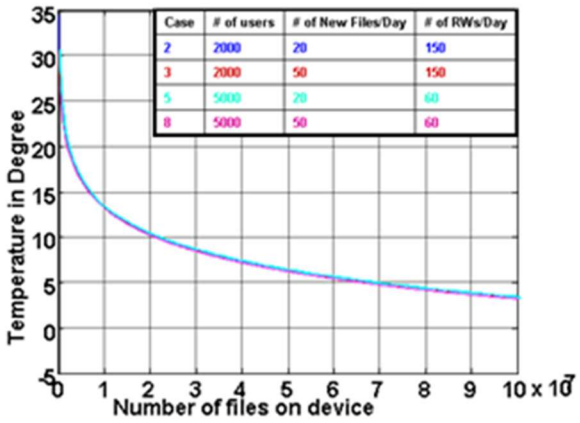
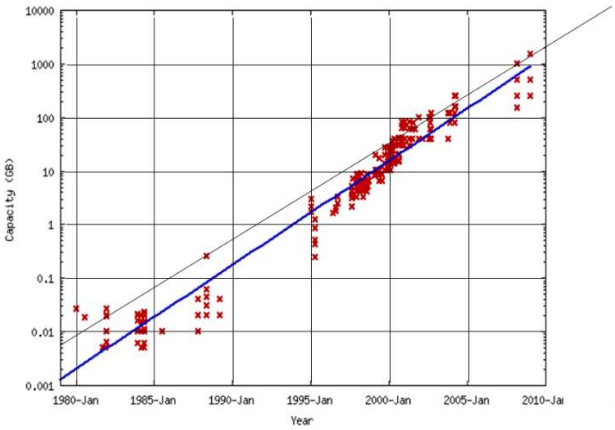


Figure 1 (left). Technology and product roadmap of HDD and Tape based on INSIC released study.

Figure 2 (right). Data temperature vs. number of files on a given device, for drive have different number of read and write attempt per day.



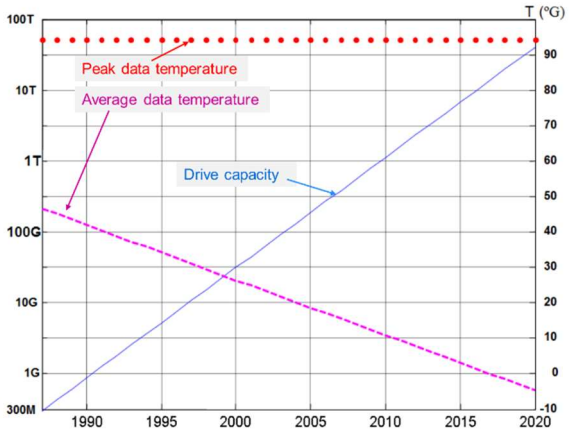


Figure 3 (left). HDD product capacity vs. time as listed on Wikipedia, with average 43% annual capacity growth rate over the past 3 decades.

Figure 4 (right). Peak data temperature, average data temperature capability and drive capacity vs. time. Data temperature capability for a given product is reduced with capacity increase.