High-throughput Video Data Access Control of Striping with SSDs for 8K Super Hi-Vision

Takeshi Kajiyama[†] (member), Kodai Kikuchi[†] (member) and Eiichi Miyashita[†] (member)

Abstract We have developed a high-throughput recording and playback method for full-specification 8K SHV video whose data rate is up to 2400 MB/s. The method fully exploits hardware with a specific design for handling large amounts of data, sequential transfer, and frame unit random reads including fast-forward, rewind and jog-shuttle. The data transfer speed is more than 3000 MB/s when using SSD striping. Steady recording and playback, including frame unit random reads, were demonstrated on an 8K SHV recorder and player.

Keywords: Super Hi-Vision, 8K SHV, video recorder, striping, RAID-0, SSD.

1. Introduction

Full-specification 8K Super Hi-Vision (8K SHV) is the highest level specification of 8K SHV^{(1)–(5)}, which consists of ultra-high resolution video and three-dimensional audio providing an extremely strong sense of reality to viewers. The uncompressed video data rate of full-specification 8K SHV at a frame rate of 120 fps reaches 18 GB/s (144 Gb/s). Video compression would facilitate the recording of such high-data-rate video. The authors have developed an 8K SHV real-time compression recorder/player and a method for controlling the compressed video data rate on compact circuits for practical use⁽⁶⁾. This recorder/player targets high-quality video for studio production use, i.e., compressed full-specification 8K SHV video with data rates of up to 2400 MB/s and a compression ratio of 1:8.

Although NAND flash memory⁽⁷⁾ is a nonvolatile memory enabling high-data-rate recording, the ordinary forms of NAND storage, called solid state drives⁽⁸⁾ (SSDs), are still not fast enough for recording 8K SHV; their data transfer speeds are less than 600 MB/s and are limited by the interface performance of the serial ATA⁽⁹⁾ (SATA). A simple solution to this problem is a parallelizing technique, called striping (RAID-0), that uses multiple SSDs operating simultaneously to increase data transfer speeds. While this technique has no redundancy unlike RAID⁽¹⁰⁾, the theoretical top speed of

striping with N SSDs is expected to be about N times the transfer speed of the slowest SSD. However, RAID-0 controllers have a trade-off between random data access performance and sequential data transfer speed, and conventional RAID-0 controllers tend to assign first priority to random access, while video data requires sequential data transfer. Here, we propose a method of controlling video data access that is specialized for large amounts of sequential data.

This paper describes a high-throughput sequential data transfer technique and a frame data unit random read for full-specification 8K SHV video at data transfer speeds of more than 2400 MB/s. These methods work by considering both the compressed 8K SHV frame data (up to 20 MB per frame) and the data access patterns in multiple playback modes, e.g., fast-forward, rewind, and jog-shuttle.

2. Problem and improvement of parallel read/write speed of SSD striping

In this section, we give an overview of the striping technique and describe an evaluation of a conventional RAID-0 controller. After that, we describe the problems with the conventional RAID-0 controller and the basic idea on how to improve throughput.

2.1 Overview of striping technique

Striping techniques are classified into hardware striping and software striping. The former uses a RAID-0 control board that consists of a host interface called a PCI Express⁽¹¹⁾ (PCIe), a storage interface (usually SATA or serial attached SCSI), and a RAID-0 controller. An example of hardware striping is shown in Fig. 1. The

Received March 21, 2016; Revised May 29, 2016; Accepted June 3, 2016

[†]Science & Technology Research Laboratories, Japan Broadcasting Corporation
(Tokyo, Japan)

Fig.1 Example of hardware striping

Fig.2 Example of software striping

latter is implemented as software inside a host, but an interface conversion board that enlarges the number of SSD interfaces via PCIe may also be installed in the host. Fig. 2 shows an example of software striping.

2.2 Evaluation of conventional RAID-0 controllers

The data transfer speed of ordinary RAID-0 controllers levels out owing to multiple factors in spite of increasing N . To investigate these factors, we evaluated the speed of sequential read/writes between a host PC and three kinds of striped SSD (SSD A, B and C) in hardware striping (Fig. 1). The interface between the host PC and RAID board was a PCIe Gen2 x8 Lane (4000 MB/s max.), and the interface between the RAID board and each SSD was a SATA Gen3 (600 MB/s max.). The maximum striping number was eight. Fig. 3 and Fig. 4 show the relationship between the striping number N and sequential write and read speeds. The software striping illustrated in Fig. 2 was similarly evaluated. In this case, the interfaces to the host PC and SSD were the same as in hardware striping, but the maximum striping number was sixteen. Fig. 5 and Fig. 6 show the relationship between N and the sequential write and read speeds. In all of these figures, solid lines indicate measured values, and dashed lines indicate the expected speed calculated as N times the single SSD speed. These data were benchmarked with CrystalDiskMark¹²⁾ of benchmark software. The file size to use for the test was more than 1 GB and other setting parameters, including a size of direct memory access (DMA), were default value in benchmark software, board driver software, and operating system. Table 1

Fig. 3 Write speed of conventional hardware striping

Fig. 4 Read speed of conventional hardware striping

lists the specifications of the devices.

In the case of hardware striping, the write speed is considerably lower than the expected speed at $N=8$ in Fig. 3. Similarly, the read speed is lower at $N=8$ in Fig. 4. The maximum speed at $N=8$ is under 2000 MB/s in both Fig. 3 and Fig. 4; it did not reach even half the theoretical speed of 4000 MB/s in PCIe. Moreover, the transfer speed at $N=8$ is lower than at $N=4$ with SSD B.

In the case of software striping, the write and read speeds are less than the expected speed at $N=8$ and considerably lower at $N=16$ (Fig. 5 and Fig. 6.) The maximum write speeds are under 2200 MB/s at $N=16$ in Fig. 5, and read speeds are under 2000 MB/s at $N=16$ in Fig. 6; they are considerably lower than the PCIe speed. These results indicate that the data transfer speeds of striping using more than one SSD are limited by a

Fig. 5 Write speed of conventional software striping

Fig. 6 Read speed of conventional software striping

Table 1 Hardware Specifications

Host PC	CPU: Core i7 (Intel) M/B : P8Z68 (ASUS) OS : Windows7 (Microsoft)
RAID board	ARC-1882i (areca)
Interface conversion board	SAS 9201-16i (AVAGO Technologies)
SSD A	SSD 840 PRO (Samsung)
SSD B	PX-256M2P (Plextor)
SSD C	PX-128M5M(Plextor)

number of bottlenecks.

2.3 Factors influencing transfer speed

Two factors can be considered as bottlenecks of the transfer speed. One is the operating speed limit of the

Fig. 7 Overview of test system and board

RAID-0 controller; the other is the overhead including the latency time on the PCIe and SATA interfaces. While the former is a fixed value reflecting only the hardware performance, the latter may be affected by reducing the overhead so that throughputs up to the theoretical limit of the hardware may be possible. Above all, high-speed video data such as an 8K SHV require the full operational capacity of the hardware.

3. Analysis of data transfer properties and video data access control method

This section shows the relationship between the amount of data transferred and throughput on the PCIe and SATA interface, and it explains the video data access control method to reduce the overhead. We prepared a test board, as shown in Fig. 7, to evaluate the transfer speed versus data size; the test board had three FPGAs connected via Low-voltage differential signaling (LVDS) links with a data transfer bandwidth of up to 4000 MB/s in total. The host interface had PCIe Gen2 x 8 lanes and a bandwidth of up to 4000 MB/s. The SSD interface also had SATA Gen2 (300 MB/s max.) that were slower than SATA Gen3 interfaces because of the hardware limitations. The maximum striping number was sixteen.

3.1 Analysis of data transfer speed on PCIe

Two kinds of commands were prepared in the PCIe protocol. One was the programmed input/output (PIO), which has a small latency and low throughput owing to its overheads with small data packets. The other was DMA, which has large latency time and high throughput with large data packets. DMA is most effective at transferring large amounts of data like video streams, but its latency needs to be shorter for high-throughput transfers. Fig. 8 is an overview of the time to transfer video frame data in a DMA. In this figure, R is the DMA repetition count for a video frame, T_{start} is the latency at the start of the DMA data transfer, T_{hd} is the time for

Fig. 8 Overhead of DMA transfer in PCIe

Fig. 9 DMA transfer speed versus DMA size in PCIe

transferring a packet header of PCIe, and T_{data} is the time for transferring packet data. The figure indicates that a longer T_{start} may decrease throughput, so reducing T_{start} should be able to improve throughput; however, reducing latency is difficult because it depends on the hardware operation speed. Instead, reducing the DMA repetition counts R by enlarging the size of the DMA transfers for each video frame may be a better way to improve throughput.

Fig. 9 shows the relationship between the DMA transfer size and transfer speed in continuous transfer, which were measured on a test board between the PCIe controller of the host PC and that of the test board shown in Fig. 7. The horizontal axis indicates the data size of a DMA transfer, and the vertical axis indicates the transfer speed. As shown in Fig. 9, the throughput is more than 3400 MB/s when the transfer size is 20 MB / frame. On the other hand, the throughput is less than 3000 MB/s when the size is 2 MB per transfer. This result shows that larger transfers improve DMA throughput.

3.2 Analysis of data transfer speed in SSD via SATA

SSDs typically have a logical sector size of 512B or 4kB. The SATA interface can set the data size in an ATA

Fig. 10 Sequential transfer speed in SSD via SATA versus sector count

transfer command in terms of the number of sectors. The throughput likely increases when a transfer command has a higher sector count. Fig. 10 shows the sequential read/write speeds for sector count in an ATA command between the SATA controller on the test board and an SSD, which was measured with the same type of SSD (C) in Figs. 3 to 6. Note that SSD C has 512B of logical sector size and the SATA interface was a SATA Gen2. The horizontal axis indicates the sector count during a transfer, and the vertical axis indicates the throughput in continuous transfer. As shown in Fig. 10, the transfer speed increases as the sector count increases. Both the read and write speeds reach a high level when the sector count is 1024 and slightly increase with the sector count after that. The maximum speeds are 262 MB/s for read and 215 MB/s for write at 3072 sectors. These results show that the throughput can be improved with a sector count over 1024.

3.3 Access control method for video data

The above analysis confirms that a transfer with a larger data size improves the throughput of the PCIe and SATA interface. Meanwhile, reading storage by the frame unit is preferable because there are several playback modes that require the ability of randomly reading from frame data, such as fast-forward, rewind, and jog-shuttle. Therefore, we designed a method for controlling data access by the frame unit. Here, let us define the following data mapping parameters. D_F is the size of the compressed frame data. D_S is the data size to be read/written for each SSD divided by the striping number N , i.e., $D_S = D_F / N$. The unit used to exchange data between an SSD and the RAID-0 controller is called a unit block, the size of which is equal to D_S . The unit used to exchange data between the host and RAID-0 controller is a unit block set, the size of which is equal to D_F . Fig. 11 shows how the unit block is related to the

Fig. 11 Data mapping on SSDs

Table 2 Example of page and sector count

D_F [MB]	N	D_s [kB]	S (512B/sector)
20	16	1280	2560
	8	2560	5120
10	16	640	1280
	8	1280	2560

unit block set. D_P is the internal read/write unit in the SSDs; usually it is 8 or 16 kB. The logical sector count S is the number corresponding to D_s . In the case of a 512B sector size, S is calculated as $S = D_s / 512B$. Table 2 shows examples of the various relationships between the parameters. Thus, the frame unit data access is controlled by arranging the size of the data that is exchanged between the host PC (or recorder/player) and storage. This video data control method achieves high-speed sequential data transfer with less overhead by enabling the size of the DMA transfer to be enlarged up to D_F and the sector count at an ATA command to be up to S .

3.4 Analysis of random read speed

In contrast with write operations, which require only sequential data transfer, read operations need the random access. In general, random accesses to SSDs tend to be slower than sequential accesses. To confirm the effect of the proposed method on frame-unit random reads, we evaluated the random read speeds for various D_s . Fig. 12 shows the results of random reads speeds in a SSD of the same type as in Fig. 10. The horizontal axis indicates random read data size, and the vertical axis indicates continuous random read speed. The random read speed increased rapidly until 1024 kB and leveled out after that. Therefore, the random read speed can be

Fig. 12 Random read speed versus data size in SSD via SATA

maximized by setting N under the condition that D_s is not less than 1024 kB.

4. High-throughput read/write test

This section describes the performance of the video data access control method with 16 SSDs (SSD C) on the test board shown in Fig. 7. Fig. 13 shows the measured write speed versus striping number N with $D_F = 20$ MB. The transfer speed increased proportionally with N and reached 3211 MB/s at $N=16$. Fig. 14 shows the measured read speeds. The transfer speed increased with N and reached 3118 MB/s at $N=16$.

The write speed fell slightly from expected speed with an increase in N , and read speed fell moderately from expected. These are most likely caused by an increase of slower SSDs and slow read/write operations with N because waiting time to read/write an unit block set is decided by the slowest SSD. This tendency is noticeable at read because ordinary SSDs have no large cache memory for read operation in contrast to write operation.

From those results, we can see that the proposed method freed up the bottlenecks and striping using 16 SSDs achieved read/write speeds of more than 3000 MB/s.

5. Implementation in 8K SHV recorder and player

To test the sustainability of the data transfer speed and practical performance of random read operations, we implemented our method in an 8K SHV recorder and player and examined the recording and playback operations. The video data stream had a constant 1200 MB/s data rate, half that of compressed full-specification 8K SHV, because of the limitations of the recorder and player, i.e., reduced information on the chroma sub-sampling of 4:2:0. The storage consisted of eight SSDs

Fig. 13 Write speed of proposed method ($D_F = 20$ MB)

Fig. 14 Read speed of proposed method ($D_F = 20$ MB)

(SSD C), having a capacity of 1547 MB/s for write and 1622 MB/s for read with $D_F = 10$ MB and $D_s = 1280$ kB (see Table 2). Recording and normal playback, requiring sustainability in sequential data transfers, were successfully demonstrated for 14 minutes at recording, which is upper limit of storage capacity, and more than 8 hours at playback. Fast-forward, rewind, and jog-shuttle operations, requiring random accesses to frame data, were also demonstrated.

6. Conclusions

We developed video data access control method that achieves high-speed sequential writes and frame unit random reads at the same time. This method increases the transfer speed of striping by reducing overhead in the PCIe and SATA interface. Using our method, the transfer speed increased with the number of SSDs used in parallel up to 16 whereas the transfer speed of ordinary RAID-0 controllers leveled off at around 8 parallel SSDs. A read/write speed of over 3000 MB/s,

which is required for recording and playback of full-specification 8K SHV, was demonstrated with 16 SSDs. We also tested this method with a 1200 MB/s 8K SHV compressed video stream and 8 SSDs on an 8K SHV recorder and player. Steady operation was demonstrated for both sequential transfers and frame unit random reads.

References

- 1) M. Sugawara, M. Kanazawa, K. Mitani, H. Yamashita and F. Okano: "Ultrahigh-Definition Video System with 4000 Scanning Lines," *SMPTE Mot. Imag. J.*, 112 (10&11) pp.339-346 (2003)
- 2) T. Yamashita, K. Masaoka, K. Ohmura, M. Emoto, Y. Nishida and M. Sugawara: "Super Hi-Vision: Video Parameters for Next-Generation Television," *SMPTE Mot. Imag. J.*, 121(4), pp.63-68 (2012)
- 3) K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama and A. Ando, "A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)," *SMPTE Mot. Imag. J.*, 117(3), pp.40-49 (2008)
- 4) SMPTE ST 2036-1 "Ultra-High Definition Television - Image Parameter Values for Program Production," (2013)
- 5) ITU-R Recommendation BT.2020, "Parameter Values for Ultra-High Definition Television Systems for Production and International Program Exchange," (2012)
- 6) K. Kikuchi, T. Kajiyama, E. Miyashita: "8K Super Hi-Vision Compression Recorder Using Compact Bit-rate Control Method," *International Workshop on Advanced Image Technology*, 1C-6 (2016)
- 7) R. Micheloni, L. Crippa and A. Marelli: "Inside NAND Flash Memories," Springer Science & Business Media, (2010)
- 8) R. Micheloni, A. Marelli and K. Eshghi: "Inside Solid State Drives (SSDs)," Springer Science & Business Media, (2012)
- 9) Serial ATA International Organization: Serial ATA Revision 3.0, (June 2, 2009)
- 10) D.A. Patterson, G. Gibson, R.H. Katz: "A Case for Redundant Arrays of Inexpensive Disks (RAID)," *Proceedings of the 1988 ACM SIGMOD international conference on Management of data*, pp.109-116 (1988)
- 11) PCI-SIG: PCI Express Base Specification Revision 3.1a, (Dec. 7, 2015)
- 12) <http://crystalmark.info/>

Takeshi Kajiyama received his B.E. and M.E. degrees from the University of Electro-Communications in 2001 and 2003. He joined NHK in 2003. Since 2008, he has been researching and developing video recorder systems at the Science and Technology Research Laboratories of NHK.

Kodai Kikuchi received his B.E. and M.E. degrees from Chiba University in 2009 and 2011. He joined NHK in 2011. Since 2013, he has been researching and developing video recorder systems at the Science and Technology Research Laboratories of NHK.

Eiichi Miyashita holds a Ph.D. degree in electronics engineering from Kyushu University. He joined NHK in 1987. Since 1990, he has been involved in research on perpendicular magnetic recording at the Science and Technology Research Laboratories of NHK. He is now a senior research Engineer of the Advanced Television Systems Research Division.