Design of a High-speed and large-capacity NAND Flash storage system based on Fiber Acquisition

Qing Li, Shanqing Hu*
School of Information and Electronic
Beijing Institute of Technology
Beijing, China
liqing02278@126.com, hushanqing@bit.edu.cn

Teng Long, Feng Liu
School of Information and Electronic
Beijing Institute of Technology
Beijing, China
longteng@bit.edu.cn, liufeng@bit.edu.cn

***Abstract*—Now, the quality of higher speed and larger capacity are required to the real-time data acquisition and playback storage system. This paper designs a large-scale storage system which uses the optical interface board, synchronous board, and multi storage boards. This system puts forward an advanced storage structure which has several NAND Flashes with multi-buses, forming a parallel pipeline design. By using the key technologies of bad block management and the ECC algorithm, which greatly avoids the influence of the invalid block to the storage system and reduces the probability of error data as well. It cannot only improve the storage bandwidth and capacity substantially, but also ensure the reliability of the storage system effectively. As a result, the storage system achieves the capacity of 30TB and the bandwidth of 25GBps. Also, this system uses fiber interfaces to link to the external network. The number and rate of the fiber channel can be dynamically changed in real time based on the dynamic reconfiguration technology by MIF files, which makes the storage system standard, universal, and scalable.**

Keywords—nand flash; high speed; large capacity; ecc; dynamic reconfiguration

I. INTRODUCTION

In the field of signal processing, the mainstream adoption of high speed and large capacity data storage technology is divided into two aspects, the high speed disk array storage technology and high speed solid-state storage technology [1]. The traditional disk array storage technology is mature and stable, but its disadvantage is that it has poor resilience with the harsh environment. However, solid-state data storage system has advantages of low power, impact resistance, high reliability etc. Especially, the NAND Flash for its non-volatile, low price, high efficiency becomes an ideal device in the mass storage system.

In recent years, the different main control chips as the brain of SSD(Solid State Disk), have big difference in the running speed, including Marvell, Samsung, Intel, etc. [2]. One of its functions is to allocate data in the flash memory in a reasonable way, and the second is to transfer the data, between the flash chips with the external SATA interface [3, 4]. So, the algorithm that used in the master control chip, can definitely lead to the performance of the SSD directly [5]. Thus, the NAND Flash-based storage management technology and the ECC (Error Checking and Correcting) algorithm [6] have become a research hotspot [7]. In [8], the MANF storage structure of multi groups of NAND Flash was proposed, which improved the bandwidth of real-time storage system to 600MBps, and the capacity was increased to 384GB. However, this structure can't adapt to the bandwidth of high-speed serial interfaces.

Aiming at the shortcomings of the above storage system, in this real-time data acquisition and playback storage system, this paper builds a SOPC storage system based on FPGA, which also introduces the NiosII soft core embedded in FPGA. It realizes the management of bad block table, and the monitoring of entire system. Storage area puts forward a multi-chips and multi-groups storage array. Making use of the inherent accessing time of the NAND Flash chips, it also proposes a parallel pipeline architecture, which not only greatly expands the storage capacity to 30TB, but also increases the speed of the memory bandwidth up to 25GBps. Thanks to the technology of dynamic reconfiguration based on MIF files, the mode of system's fiber interfaces can be changed in real time, which can greatly adapt to the flexibility of today's high-speed interconnect interfaces.

II. REAL TIME STORAGE SYSTEM DESIGN

In the field of signal processing, the quality of higher speed and larger capacity are required to the real-time storage system, this paper designs a real-time data acquisition and playback storage system, which is based on multi NAND Flashes. This system is composed of M blocks of NANF Flash boards, a control board, a synchronous board and M/2 blocks of fiber interface boards. Each fiber interface board provides N ways of fiber channel, so it can totally have MN/2 numbers of fiber channel for the acquisition and playback of data. The NAND Flash storage board uses multi pieces of NAND Flash with multi-buses forming a parallel pipeline structure and uses the key technologies of bad block management and the ECC algorithm controlling the maintenance and management of the large capacity of data, whose capacity is more than 10TB. The fiber interface board and synchronous board have PCI interfaces, and are connected to the main-control board, as the system control. Each fiber interface board provides a fiber pathway for unloading. And the data can via fiber channel to

978-1-4799-6233-4/14/$31.00 ©2014 IEEE

the end-point unloading server. The structure of system is shown in Fig. 1.

This acquisition and playback storage system can work in the following four modes: data acquisition mode, playback mode, unloading mode and read-back mode. These different modes make this storage system flexible and efficient.

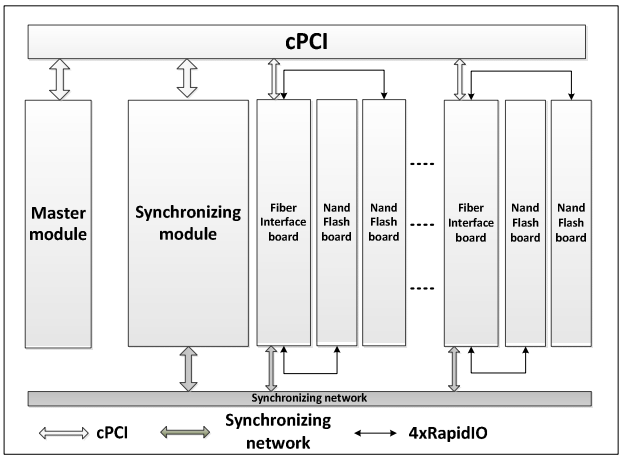

Fig.1. Structure of High-speed storage system

A. Acquisiton Mode

In fiber acquisition mode, the user can choose to receive how many number of fiber channels and choose how fast the data to transmit. And the fiber support 8 kinds of rates (1.25Gbps / 1.5Gbps / 2.0Gbps / 2.5Gbps / 3.125Gbps / 4.8Gbps / 5Gbps/6.25Gbps). When the data is transferring, synchronous board starts all optical interface board to collect the data, ensuring that all fiber data recording at the same time. The FPGA of fiber interface board receives data through the external fiber module, and transmit the data to the two NAND Flash storage boards by the 4xSRIO. In the storage board, there are two FPGAs; each of them corresponds to a set of NAND Flash array, storing the data of one set of 4xSRIO. The data flow diagram is shown in Fig. 2.

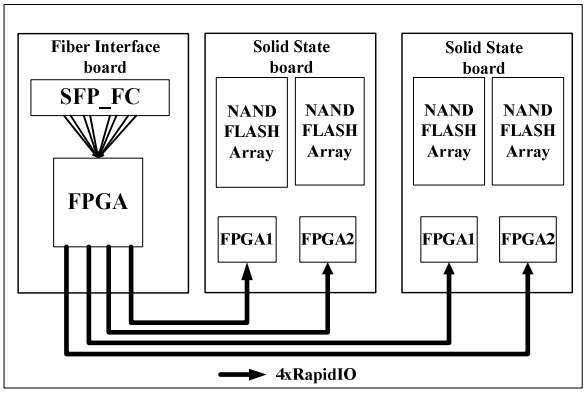

Fig.2. Data flow of fiber acquisition mode

B. Replay Mode

In fiber playback mode, after the user selects the playback file, NAND Flash storage boards read data from the NAND Flash array and pass the data to the fiber interface board. Then, fiber interface board playbacks data to the user signal processors in a specific way. Playback mode is the reverse process of fiber acquisition mode, during playback mode, the FPGA of fiber interface board can make use of the technology of dynamic reconfiguration, let the data send from different fiber modules at optional rates.

C. Unloading Mode

In fiber unloading mode, after the user selects the unloading files, the unloading operation is started by the unloading server. First, the FPGA of NAND Flash storage board reads the data from the NAND Flash array. Then, the FPGA sends the data to the fiber interface board by 4xSRIO. At last, the FPGA of fiber interface board transmits the data to the PCIE board and the FPGA of PCIE board uploads the data to the host computer, which can form the final documents. During the interaction of commands, the fiber board and storage board communicate by the 4xSRIO, and the PCIE board and fiber board communicate by the PCIE. The data flow diagram is shown in Fig. 3.

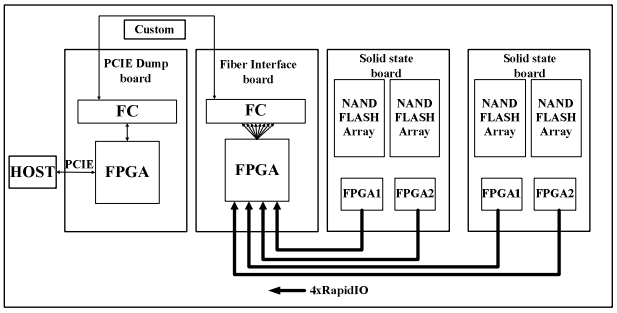

Fig.3. Data flow of fiber unloading mode

D. Read-back Mode

The fiber read-back mode is the reverse process of the unloading mode. After the user selects the read-back files, the data from the host computer can be transmitted to the PCIE board by PCIE. Then, the data can be sent to the fiber board. At last, the FPGA of fiber board reads each channel of 4xSRIO to make the data store in the NAND Flash array. Each storage board has two sets of NAND Flash array.

E. Dynamic Reconfiguration of Fiber channel

During the acquisition of different size of files in the large-scale storage system, in order to achieve the qualities of high-bandwidth and high-speed of this system, the number and rate of each set of fiber channels in each fiber interface board are designed to be dynamically adjusted in real time, thanks to the technology of dynamic reconfiguration.

The technology of dynamic reconfiguration based on MIF file can reconfigure the rate or bit wide of GXB module in a real time by modifying the MIF file in the FPGA. Such not only meets the requirements of changeable rate of fiber, but also improve the flexibility of the applications of FPGA.

The dynamic configuration requires a MIF file, which is used to initialize the RAM. So, the different working mode of the different configuration parameters of GXB must produce

different MIF file. During the dynamic reconfiguration, the GXB must be in the state of reset. Then, the new MIF file with new parameters can be unloaded to the GXB through the ALTGX_RECONFIG module. The bit wide of the MIF file is 16, and the size is no more than 60. So, a RAM which is 16bit×64 depth can be used to save the MIF file. If it needs to reconfigure multi sets of channel, it can improve the depth of the RAM. The configuration architecture is shown in Fig. 4.

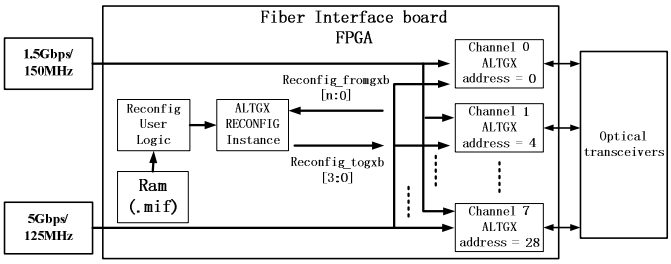

Fig.4. Dynamic reconfiguration based on MIF file

III. NAND FLASH STORAGE SYSTEM BASED ON FPGA

In order to improve the bandwidth and the capacity of the acquisition and playback storage system, this paper designs a parallel pipeline structure with multi NAND Flashes in the storage board.

The hardware design of the storage board includes one chip of FPGA, multiple chips of NAND Flash. The system's structure is shown in Fig. 5. This design is a SOPC system controlled by FPGA, which can realize all kinds of operations including the pipeline control of NAND Flash, the management of bad block table, and the control of the data interface. The design has $M \times N$ groups of NAND Flash, each of group have K chips. And each of the N group shares a data bus. And there are M buses totally, which realize the pipeline monitoring of the FPGA. Using high speed interfaces 4xSRIO to achieve the interconnection for the command or data between the memory arrays of NAND Flash with the front end devices, which speed is up to 1280MBps.

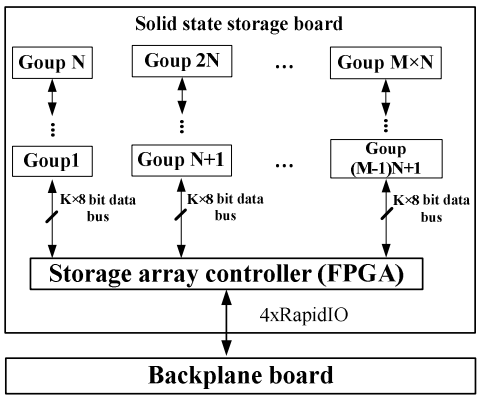

Fig.5. Structure of NAND Flash Control Module

A. A Parallel Pipeline Array Design of NAND Flash

In the design of the storage circuit of NAND Flash, it makes multiple NAND Flashes in a parallel combination, which can greatly improve the system's storage capacity and the storage bandwidth. For each chip of Flash, for example, the write operation can be divided into three phases, namely the loading time T_{TA}, the transferring time T_{TRANS} and the page programming time T_{TPROG}. So equation (1) is the total time of the processing of write operation.

$$T_{WR} = T_{CA} + T_{TRANS} + T_{PROG} \tag{1}$$

We can assume the amount of data is Q, the number of pages of write operation is p, and the capacity of each page of Flash is S. Then equation (2) is the whole time it takes for the write operation.

$$T_{WR} = \lceil p \rceil T_{CA} + pT_{TRANS} + \lceil p \rceil T_{PROG} \tag{2}$$

Where: $p=Q/S$.

When writing a group of NAND Flash, the chip number of each group is K, then we can exchange $p=Q/KS$ in the formula above. So, the number of page written to each NAND Flash is only $1/K$ to the primary number, which reduces the time of writing process down to K times.

Assuming the number of the buses is M, the number of the groups on each bus is N, there will be as much as $M \times N$ times for the writing operation of data. Thus, equation (3) and equation (4) is the writing time it takes for the whole array.

$$T_{WR} = \lceil p \rceil T_{CA} + pT_{TRANS} + \lambda\left(\lceil p/N \rceil - 1\right) + T_{PROG} \tag{3}$$

$$\lambda = MAX\left\{T_{PROG} - (N-1)\left(T_{CA} + T_{TRANS}\right), 0\right\} \tag{4}$$

Where: $p=Q/KMS$.

So, the total accessing time is reduced to M×N times. The read operation is in the same way.

In this paper, the system we designed uses 12 groups of NAND Flash chips to form a pipeline storage structure. Every 3 groups of the Flash chips share one set of data bus, which has 4 buses and each of them has a 64bit data bandwidth. At 80MHz, according to the chip characteristics, equation (5) is the bandwidth of this board.

$$B_{WR} = 8bit \times K \times M \times (80MHz/2) \tag{5}$$

So, in this paper the bandwidth of the parallel and pipeline design of NAND Flash is 1280MBps.

B. Software Design of Master Control

The NiosII as the soft core embedded in FPGA is responsible for the implementation of reading and writing process of parallel pipeline operation, and the global monitoring of the working state of the whole system. It makes the management of the logic layer of FPGA abstracted, greatly reduces the difficulty of the FPGA development. In this paper, it makes use of the commands sending from the front device, which is given to the soft core of FPGA NiosII to judge whether the operation is to erase or record or read. The operation flow chart is shown in Fig. 6. After finishing the

initialization of NAND Flash, the operation of NiosII gets into the waiting process. Then, the software of FPGA can go to the corresponding process by receiving the command send by the front system.

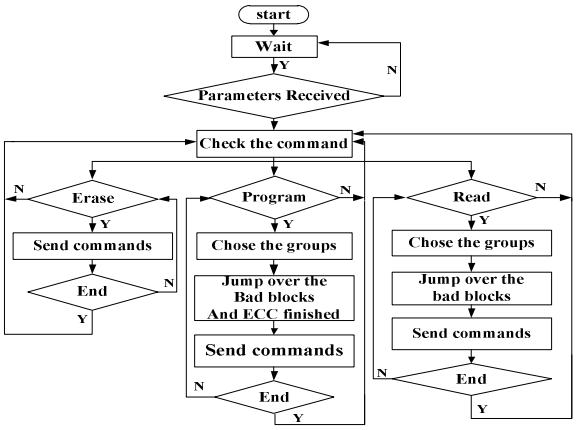

Fig.6. NiosII software Flow of NAND Flash

C. Bad Block Management Design

Due to its technology reason of NAND Flash, there will be some block address cannot be accessed, which is referred to the invalid block. One of the key technologies of the design is to manage the invalid block reasonable, so as to achieve the reliability of the data storage. It requires checking the invalid block of the chip before the accessing of data. It also needs to establish a bad block table for each group. Also, the invalid block may be divided into the initial bad blocks and the used bad blocks. Fig. 7 shows the specific processes in the detection of bad block and the creation of bad block table about the two types of invalid block. For initial bad block, before the operation of NAND Flash, it needs a traversal of effective flag of each block of NAND Flash. If this flag is non-0xFF, it is considered a bad block. For the used bad blocks, after the wiping operation of NAND Flash, the state of the lowest bit of I/O bus shows whether the block is valid. The invalid block leads to a discontinuous physical address of operation. The control program of FPGA achieve a transparent physical address of NAND Flash chips, the users can use the continuous logical address to accessing the data.

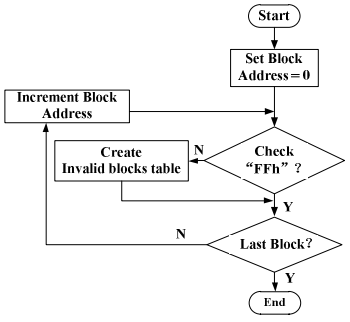

Fig.7. Building Process of Bad Block Table

D. ECC Parallel Design

Since the NAND Flash chips often appears the phenomenon of bit inversion, which leads to the mistakes of store data. Therefore it must use ECC algorithm to solve the above problems. This design uses the algorithm of 1024Bytes with 24bit error corrected [9], which achieves high reliability with zero error bits. The ECC algorithm uses the BCH code, which can be expressed as *BCH (n, k, t)*, where n is the BCH code length, k is the length of the data, t is the error correcting bits [10]. For the NAND Flash used in this paper, the minimum demand of ECC is 1080Bytes data with 24bit error corrected. So it needs to design a software module to satisfy the demand of error correction. Since the page size of each chip is 8192 bytes, it forms the type of code is *BCH (8528, 8192, 24)*. During the encoding operation, the data to be encoded through the BCH encoder in order and the output data is the check code [11, 12]. The equation (6) is the generator polynomial of Binary BCH code.

$$g(x) = m_1(x)m_3(x)\cdots m_{2t-1}(x) \tag{6}$$

Where: if α, α^3,…, α^{2t-1} are the polynomial roots of this BCH codes, then $m_i(x)$ is the minimal polynomial of $\alpha^i (1 \le i \le 2t)$.

In decoding operation, the data to be decoded is sent to the ECC decoder. When the ECC decoder finds an error, it can calculate the error location, and then the Chien Search Circuit can use the IBM (non-inverse Berlekamp Massey) algorithm to find and correct the error.

IV. CONCLUTION

Real-time data acquisition and playback storage system is a powerful tool for radio system and a helpful debugger for signal processor [13]. In this paper, the FPGA is in type of Stratix IV from the Altera Company, and NAND Flash is in type of MT29F128G [14] from the Micron Company.

Each storage subsystem uses one fiber interface board and four NAND Flash storage boards. Make use of the method of parallel pipeline control, which is based on the key technology of ECC error detection and correction. The capacity of the storage subsystem has achieved to 6TB, and bandwidth has achieved to 5GBps, which is possible to recording various data at a high speed continuously.

In real-time data acquisition and playback storage system, 20 NAND Flash storage boards are used to record data. Such a system is able to keep recording to 400 minutes with bandwidth up to 25GBps. The acquainted real time data can be replayed dynamically by the fiber channel. And the measured data also can be used to the debugging or the verification of algorithm in the future work.

REFERENCES

[1] Bo.Ying, "Extendible VXS bus solid state recording system[D]", Nanjing: Nanjing University of Science and Technology, 2008.

[2] F. Douglis, R. Caceres, F. Kaashoek, K. Li, B. Marsh and J.A. Tauber, "Storage Alternatives for Mobile Computers," Proc. First Symp. Operating Systems Design and Implementation (OSDI), pp. 25-37, 1994.

[3] C.Wang, W.Liu, “The high speed data recording system based on file system[J],” Journal of Beijing institute of technology university, 2010.

[4] Lansun.Shen, “Principle and application of high-speed data acquisition system [M],” Beijing: post and telecom press, 1995.

[5] E. Gal and S. Toledo, "Algorithms and data structures for flash memories," ACM Computing Surveys, Vol. 37, No. 2, pp.138-163, Jun. 2005.

[6] C. Yang, Y. Emre and C. Chakrabarti, "Product Code Schemes for Error Correction in MLC NAND Flash Memories," IEEE Trans. on VLSI Systems, vol.20, no.12, pp.2302-2314, 2012.

[7] Guangnian.Yang, Rongliang.Guo, Guozheng.Zhang, “The development of storage technology [J],” Computer and digital engineering, 2008.

[8] Lifeng.Wang, Shanqing.Hu, Feng.Liu, Teng.Long, “A high speed solid state data recording system based on serial rapidIO,” Computer engineering Publising, 2011.

[9] Xinmei.Wang, Guozhen.Xiao, “Error correcting code: Principle and Method,” Xi’an Electronic and Engineering University, 2001.

[10] H. Choi, W. Liu, and W. Sung, "VLSI Implementation of BCH Error Correction for Multilevel Cell NAND Flash Memory," IEEE Trans. on VLSI Systems, vol. 18, no. 5, pp.843-847, May 2010.

[11] STMicroelectronics corporation, Inc, “Error correction code in single level cell NAND flash memories[Z],” 2004.

[12] Samsung corporation, Inc, “ NAND Flash ECC Algorithm[Z],” 2003.

[13] T.E. Curtis and A.B. Webb, "High-Performance Signal Acquisition-Systems for Sonar Applications," International Conference on Analogue to Digital and Digital to Analogue Conversion, vol. 343, pp. 87-94, 1991.

[14] Information on http://www.micron.com/products/nand-flash/mlc-nand