????:
http://www.design-reuse.com/articles/15952/raid6-accelerator-in-a-powerpc-iop-soc.html

?? :OSSLab vxr, thx ??:
??/Abstract-- IO Processors (IOP) are key elements for storage application to attach to a Host the maximum number of IO's such as RAID disks, capable for high speed data movement between the Host and these IO's and able to perform function such as parity computation for RAID (Redundant Array of Independent/ Inexpensive Disks) applications. A second generation [1] 800MHz PowerPC SOC working with 800MHz DDR2 SDRAM memory and on-chip L2 cache, executes up to 1600 DMIPS is described in this paper. It is designed around an high speed central bus (12.8 GBytes/sec PLB) with crossbar switch capability, integrating 3 PCI express ports, plus one PCI-X DDR interface and RAID 5 & 6 hardware accelerator. The RAID 6 algorithm implement 2 disk parities, with the second Q parity generation based on the finite Galois Field (GF) Primitive Polynomial functions. The SOC has been implemented in a 0.13 um, 1.5 V nominal-supply, bulk CMOS process. Active power consumption is 8W typical when all the IP run simultaneously.
IO Processors (IOP)??????(Host)???IO???RAID??????????????, ???Host???IO??????????????????RAID?????, ??????. ???????????? 800MHz SoC, ??800MHz DDR2 SDRAM??????????, ??????1600DMIPS(Dhrystone MIPS: ?????????????, ???MFLOPS????????). ??SoC??????????????(crossbar switch capability)????????(PLB??????? 12.8GB/s), ????PCI express??, ?????PCI-X DDR???RAID 5&6???????. RAID 6?????2??????, ??????Galois Field (GF)??????????Q??????(second Q parity). ??SoC???0.13??, ??1.5V??, CMOS?????. ????IP(IP core: Intellectual Property core=>??????ASIC??FPGA??????, ??????ASIC???????????????)???,??????????8w.

II-??/INTRODUCTION
This paper describes a PowerPC system-on-a-chip (SOC) which is intended to address the high-performance RAID market segment. The SOC uses IBM's Core-Connect technology [2] to integrate a rich set of features including a DDRII-800 SDRAM controller, three 2.5Gb/s PCI-Express interfaces[3], hardware accelerated XOR for RAID 5 and RAID 6, I2O messaging, three DMA controllers, a 1Gb Ethernet port, a parallel peripheral Bus, three UARTs, general purpose IO, general purpose timers, and two IIC buses.
?????????????RAID?????Power SoC. SoC??IBM?Core-Connect?????????, ????DDRII-800 SDRAM??????, ??2.5Gb/s PCI-Express??, RAID5/6????XOR??, I2O????(Intelligent I/O), ??DMA???, ??GbE??????, ????????? (parallel peripheral Bus), ??UART, GPI/O??, ?????(GPT)???IIC???.

II- ????/SYSTEM OVERVIEW
This SOC design consists of a high performance 32-bit RISC processor core, which is fully compliant with the PowerPC specification. The PowerPC architecture is well known for its low power dissipation coupled with high performance, and it is a natural choice for embedded applications The processor core for this design is based upon an existing, fixed voltage PowerPC 440 core [2]. The core includes a hardware multiply accumulate unit, static branch prediction support, and a 64-entry, fully-associative translation look aside buffer. The CPU pipeline is seven processor stages deep. Single cycle access, 64-way set associative, 32-KByte instruction and data caches are connected to the processor core.
???SoC??????????PowerPC??????32??RISC?????. PowerPC????????????, ??????????????. ?????????????, ?????PowerPC 440????. PowerPC 440???????MAC(??/??)??, ?????????64-entry????TLB(?Cache?????, ??????????) . CPU??????7???. ??????????????(single cycle access), 64????(64-way set associative), 32KBytes?????????(?PLB ???128bits; ????????????). ??????????IP(Intellectual Property)??, ???RAID 5? RAID 6?????????DMA????.

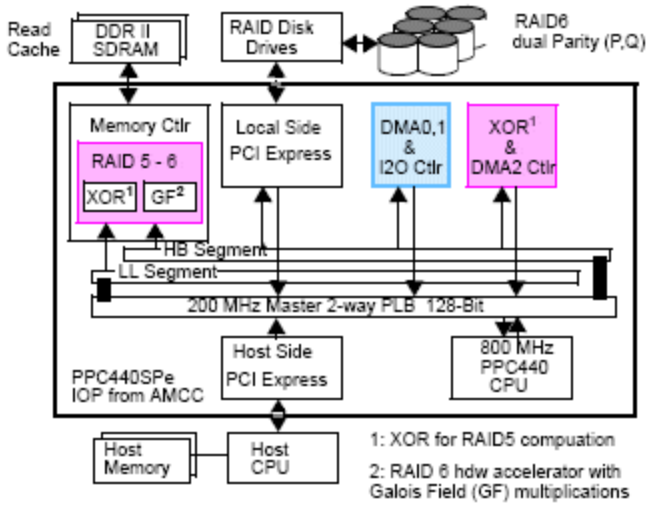

Figure 1 RAID 5?RAID6?IOP?????? / RAID 5 & 6 IOP processor block diagram

A second level (L2) cache of 256 KB is integrated to improve processor performance. Applications that do not require L2 may optionally use the L2 as on chip SRAM. The L2 memory arrays include redundant bits for parity and spares that can be connected after test and configured with on chip fuses.

???256KB???????????????. ??????L2?????, ???????L2????????SRAM. ????????(clip fises)????????, ????L2??? ??????????????(parity and spares)???????(redundant bits).

III - ????????????/ON CHIP HIGH SPEED CROSSBAR BUS:

The key element of this SOC for high speed data transfer is the central 128b wide 200 MHz crossbar PLB (Processor Local Bus) [1]. Two out of eleven masters can simultaneously access one of the two PLB slave buses: one specialized in High Bandwidth(HB) data transfer and a second one with Low Latency (LL). The same physical memory in the SDRAM can be accessed either on the HB or the LL slave bus through two aliased address ranges. By convention (but not required) the LL bus segment is used by the PowerPC to achieve low latency access to memory while the HB bus segment is used for large data moves by the DMA engines.

??SoC??????????128????, 200MHz??????PLB(Processor Local Bus). 11?master??2?????????PLB slave???????: ??????????????(HB:High Bandwidth data transfer)?????????(LL: Low Latency)??. ??????????,????HB??LL slave?????????SDRAM?????????. ????(???), LL???????Power PC?????????????, ?HB???????DMA??(DMA engine)????????.

??????????, ??HB??LL slave?????????SDRAM???????????. ????(???), LL?????????PowerPC???????????, ?HB??? ??????DMA??????????.

The Crossbar architecture separates the 64b address, 128b read data, and the 128b write data busses allowing simultaneously duplex operations per master with two independent masters resulting in a peak theoretical bandwidth of 10 Gbytes/sec.

?????(Crossbar architecture)?? 64????, 128???????128?????????, ???????master??????????????10GB????.

While the Crossbar arbiter supports 64 bit addressing, the PowerPC440 CPU is a 32 bit processor that can address up to 4 GB of physical address, the 64 entry TLB transforms this address to a real 36 bit PLB address (upper 28 bits are 0s) for 64GB access of the total address space.

????????(Crossbar arbiter)??64????, ??PowerPC440 CPU????????4GB??????32?????. 64-entry TLB??????????? ?????64GB?????36??PLB??, . Raid 5 ?Raid 6 ???

The RAID 5 have been developed to provide data recovery in the case of a single drive failure by adding a parity disk that is used with the remaining disks to rebuilt the failing data.

RAID 5????????????(failure)???????????, ????????????????????????(failing data)???.
Notice that the Error on a disk drive must be detected by an another error detection circuit such as CRC checking.
???????????(Error)??????????(error detection circuit)??CRC???????.
??Raid 6?????????
AMCC PowerPC 440SP/440SPe: ????Raid 5 ?Raid 6???
Raid 5??????????? P ,P??????
P = D0 Xor D1 Xor D2
????HDD ,Raid 5????????P?(????)

	硬碟0	硬碟1	硬碟2	硬碟3
條帶0	D0	D1	D2	**P0**
條帶1	D3	D4	**P1**	D5
條帶2	D6	**P2**	D7	D8
條帶3	**P3**	D9	D10	D11
條帶4	D12	D13	D14	**P4**

In RAID 6, ??? P,Q with GF coefficients is needed:

P = D0 Xor D1 Xor D2

Q = (A0×a)?(A1×b) ?(A2×c)

The RAID 6 algorithm for Q parity generation is based on the Galois Field (GF) Primitive Polynomial functions. With the PPC440SPe it is possible to use several different values of the GF polynomial, including the values 0x11D and 0x14D which corresponds to the equations:
??Q??????RAID6??????Galois Field (GF)????????(Primitive Polynomial functions). ??PPC440SPe???????GF????????, ????????(equation)??-0x11D??-0x14D:

0x11D: $X^8 + X^4 + X^3 + X^2 + 1$

0x14D: $X^8 + X^6 + X^3 + X^2 + 1$

V - RAID??????/RAID Hardware assist options

There are two options for the RAID hardware assist.

???RAID hardware assist??.

The fist one is to attach directly to the PLB on chip bus the RAID assist as an independent unit with its own DMA controller and perform the dual P,Q parity between the different operands through the control of the DMA engine.

??????????DMA?????????????PLB??RAID assist???????, ????DMA????, ????????????P,Q????.

The second option is to have this RAID hardware assist directly in the memory controller and enable it if the address on the PLB on chip system bus falls in one predefined address range. In this case some predefined function and parameters must be included in the reserved bits of the 64- bits PLB address.

??????????????????RAID??????(RAID hardware assist)?????, ???????-PLB??????????????????. ??????, ????????????????64?? PLB??????????.

VI - PLB??RAID 5????(Hardware assist)/RAID 5 Hardware assist on PLB

The block diagram figure 2 shows that a XOR and Not XOR function is attached directly to the PLB and computes the parity in each cycle when a new operand is entered in the unit.

?2??????????????????????, PLB??????XOR/NOT XOR???????????????.

The Hardware XOR engine computes a bit-wise XOR on up to 16 data streams with the results stored in a designated target. The XOR engine is driven by a linked list Command Block structure specifying control information, source operands, target operand, status information, and next link. Source and target can reside anywhere in PLB and/or PCI address space.

??XOR?????????????, ???16?????(16??XOR??)????????-XOR??(1 xor 2 xor 3 xor 4 xor .... xor n | n=16). XOR????Command Block??? ??????, ?????, ?????, ?????????????. ??????????PLB?(??)PCI??????????.

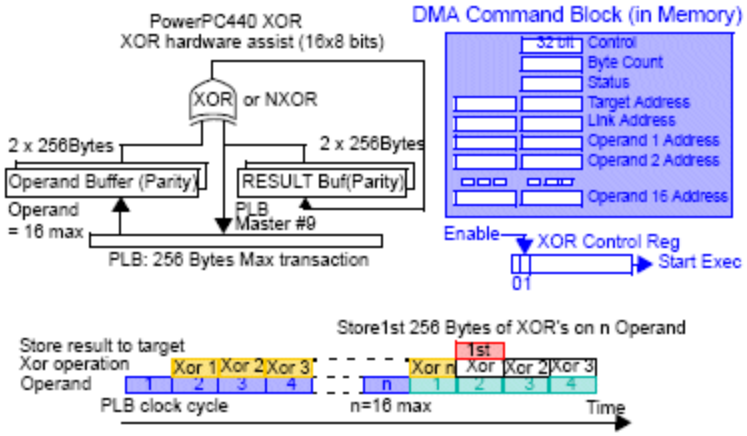

Fig 2: DMA aware XOR unit operations

VII - ??????(Memory Queue)??RAID????/RAID Hardware assist in Memory Queue

The block diagram figure 3 show that a XOR and Galois field computation for Q parity is integrated in the Memory Queue of the SDRAM Memory controller. Two circuits are implemented; one working on Writing to the Memory acting like a Read Modify Write, and a second one with a Multiple Read of Data as shown on figure 4.
?3????????SDRAM????????????(Memory Queue)???Q??????XOR?Galios????. ?????????; ???????????-?-?(Read Modify Write)???????, ???????????????(Multiple Read of Data)???, ??4??.

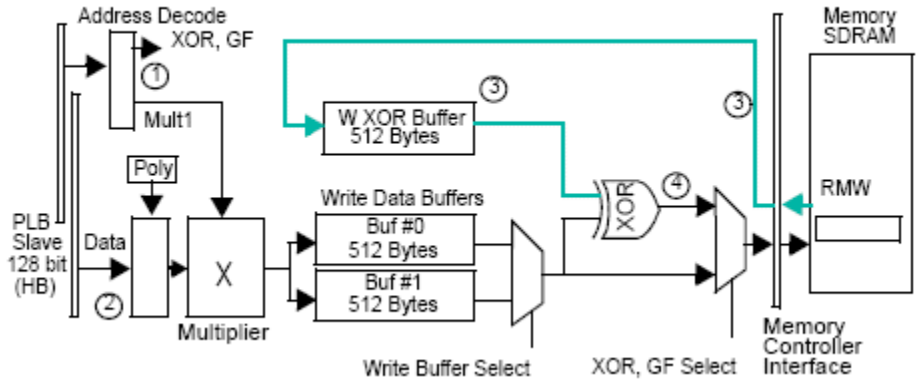

Fig 3: ?RAID 6??????(Memory Queue)???XOR??/Write XOR function in Memory Queue for RAID 6

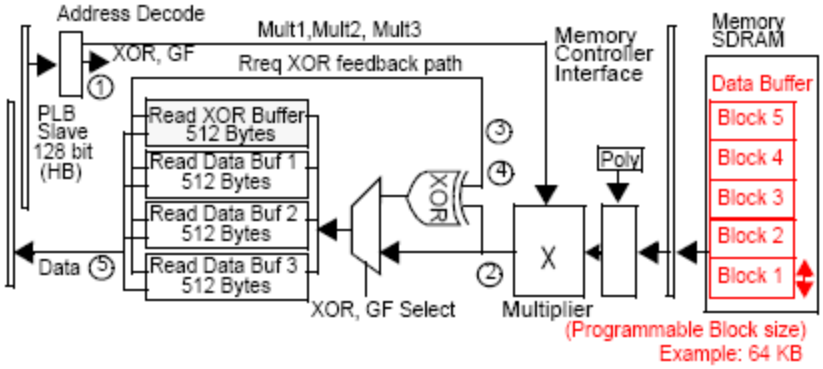

Fig 4: ?RAID 6??????(Memory Queue)???XOR??/Read XOR function in Memory Queue for RAID 6

The following figures 5 and 6 explain how the address on the PLB is decoded to activate the RAID 5/6 hardware assist in the Memory queue if this address falls in a predefined window.

???5??6??PLB??????????????????(Memory queue)??RAID 5/6 ????(RAID 5/6 hardware assist), ??????????????(window)???.

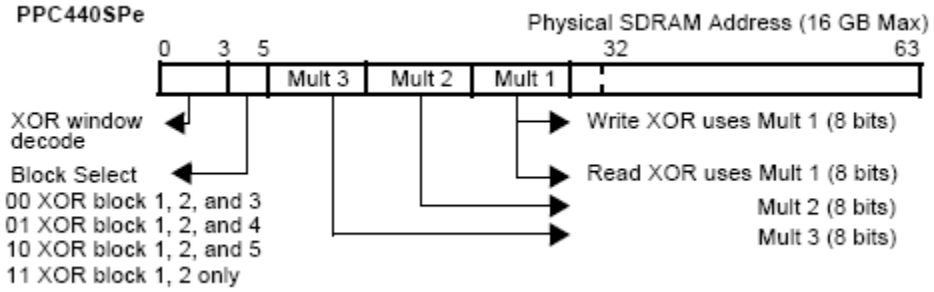

Fig 5: PLB???????????(Memory Queue)??RAID 6??/PLB Address to access RAID 6 function in Memory Queue

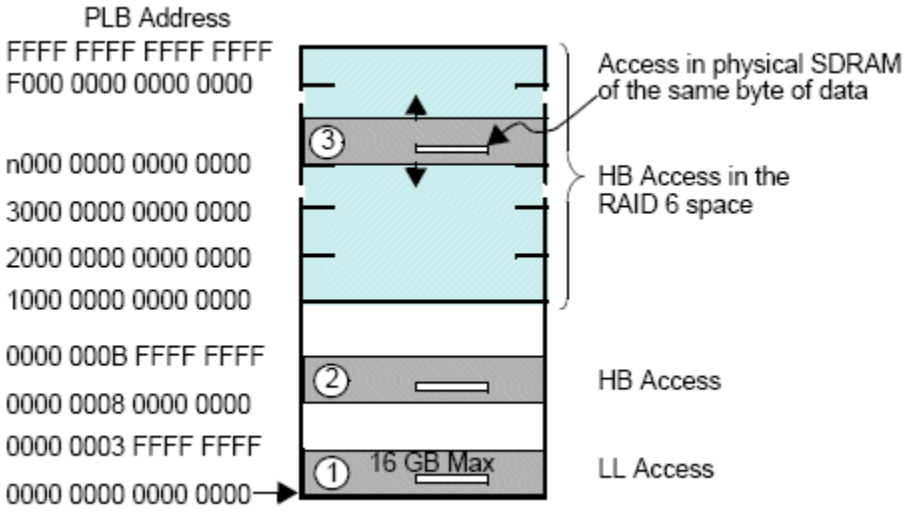

Fig 6: ???????/Memory System Address

V - ??????/Disk STRIP operation

One of the important operation in RAID 5 or 6 is the update of a strip of 64 KB, for example, as illustrated by the following figure 7. The Host that wants to update a strip in the RAID disks has to provide the new data to the IOP in a temporary memory space. Then it is needed to recompute the P and the Q parity. The first step consist in computing a new parity P,Q with the XOR of the current parity and the Strip that will be replaced. Then the P and Q are computed a second time with the new data and the last parity. After all these operations, the new P and Q as the new Data can be send to the Disk controller.

?RAID 5?6?, ???????64KB???????, ????????7??. Host????RAID??????????????????????????IOP. ??????????P?Q?????(P and the Q parity). ??????????????XOR??????P, Q???????????????. ??P?Q????????????????????????. ?????????, ??P?Q?????????????????.

These operations are illustrated on the following figure.

?????????????.

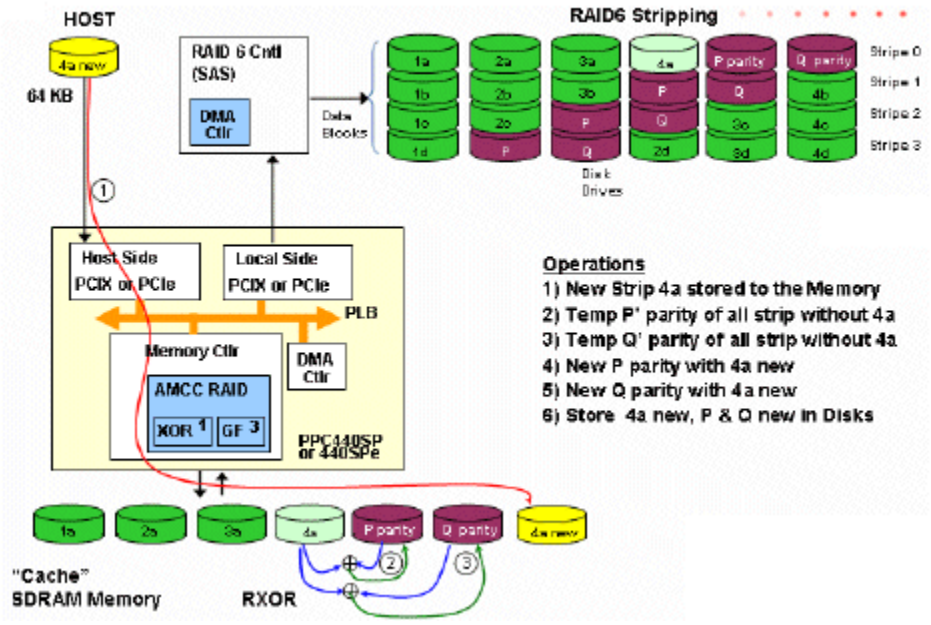

Figure 7: ??RAID?????/Strip update with RAID 6

XII- ????/TEST RESULTS

The performance throughput depends on the number of disk drives. The following curve shows one example of throughput for full stripe RAID 6 that has been measured with the PPC440SPe integrating RAID 6 hardware assist.

??????????????. ??????????????RAID6????PPC440SPe??????RAID 6?????.

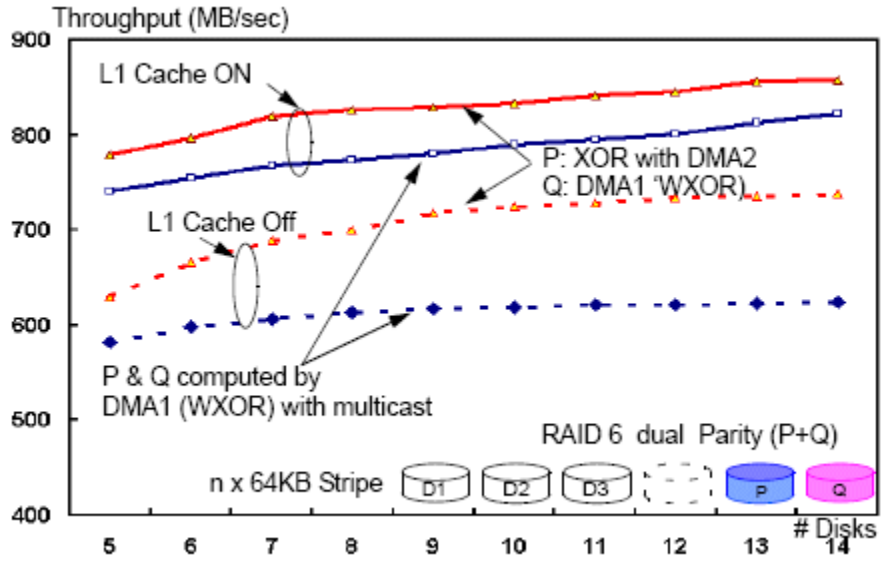

Figure 8: ????RAID 6???/Full Stripe RAID 6 throughput

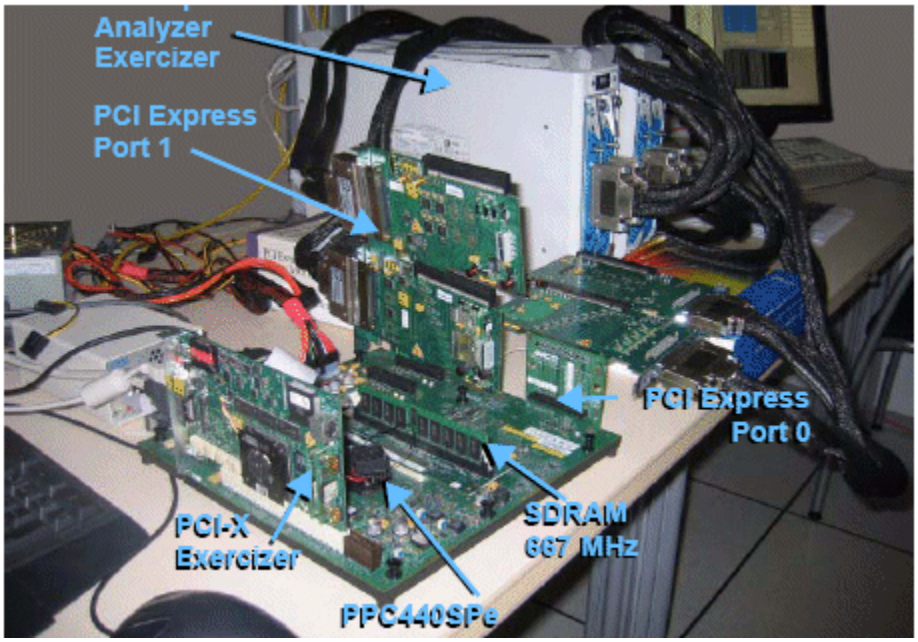

Figure 9: ??RAID??????????/Board for RAID performance benchmarking

A special board with modular approach for PCI-Express, PCI-X, SDRAM DDR2 DIMMS, and peripheral attachments has been developed. It permits the debug of the SOC device with DDR1 and DDR2 SDRAM as well as PCI Express and PCI-X DDR connectors. Debug was done with the IBM Riscwatch debugger through the JTAG serial link I/O.

????PCI-Express, PCI-X, SDRAM DDR2 DIMMS??????????????????. ?????DDR1?DDR2 SDRAM??PCI Express?PCI-X DDR????SoC????.
??JTAG????I/O, ??IBM?Riscwatch?????????.

XI- SOC IMPLEMENTATION

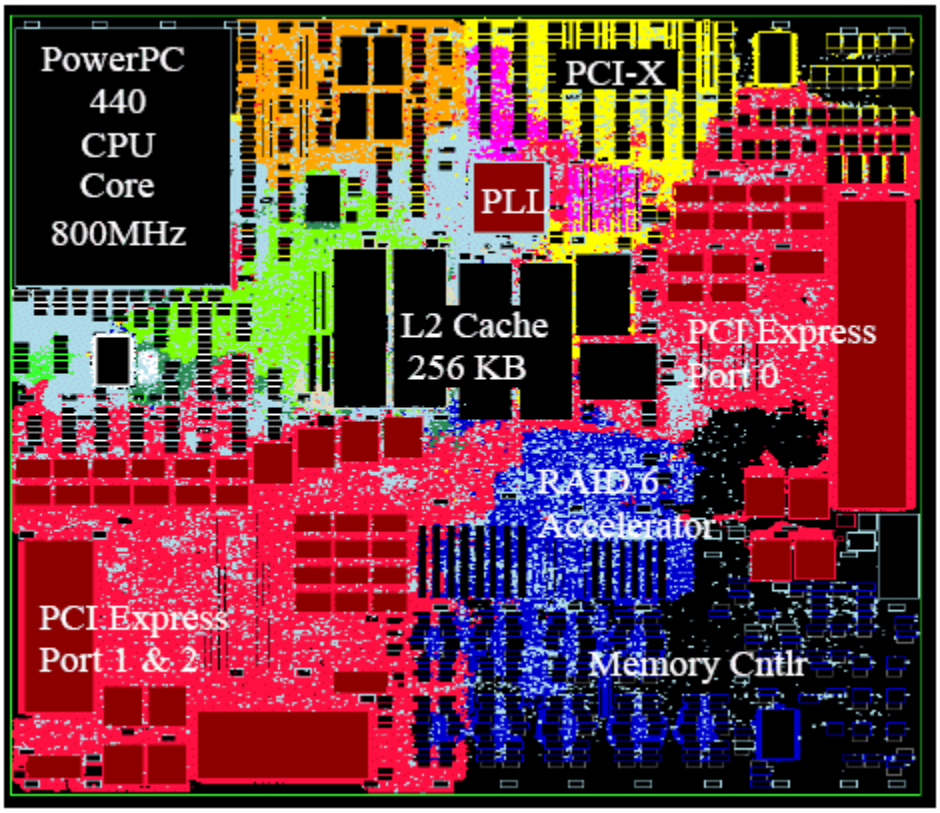

Figure 10: PowerPC IOP????/PowerPC IOP Chip layout

????/Main Features

800MHz PowerPC CPU

256 KB L2 Cache

DDR800 SDRAM Memory Controller

200MHz Internal PLB bus - 12.8GB/s

3 PCI Express Ports (8-4-4 lanes) 2.5Gb/sec per lane

1 PCI-X DDR 266 interface

DMA, I2O, RAID 5 & 6 Hdwr assist, etc.

??/Technology

CMOS 0.13 um Copper ??

7 levels of Metal

11.757 million gates

Gate area = 3x12 channels of 0.4um

??/Packaging

27mm FC-PBGA (Flip chip Plastic Ball Grid Array)

1mm pitch

495 Signal I/Os

675 Pads

Due to the large number of I/O (495) needed to integrate all the peripherals, the I/Os are placed in an area array across the die. A peripheral approach for IO implementation was possible with a staggered structure; however, it would have resulted in a larger die size, and a more noise sensitive part because of large simultaneous switching.

?????I/O??(495?)????????, I/O?????????????(die; ??(wafer)???????). ????IO????????????????(Staggered Structure; ???????); ??, ??? ?????????,??????????, ??????????????(Simultaneous Switching: ??????????????, ??????SSN).

The device is based on an ASIC with integrated synthesizable cores - also named IP's - with the exception of the PowerPC CPU core which is a precharacterized hard core with optimized timing analysis and tuned clock distribution to achieve 800MHz.

Logic is described in Verilog and synthesis done with Synopsys synthesis tool. The physical design including floorplaning,

placement and wiring was done with IBM's proprietary Chip Bench tool. Special care was taken in physical implementation for minimization of noise induced by coupling and simultaneous switching on top of the conventional signal integrity verification.

Extensive simulation of each core with simulation after SOC integration has resulted in a first pass good product.

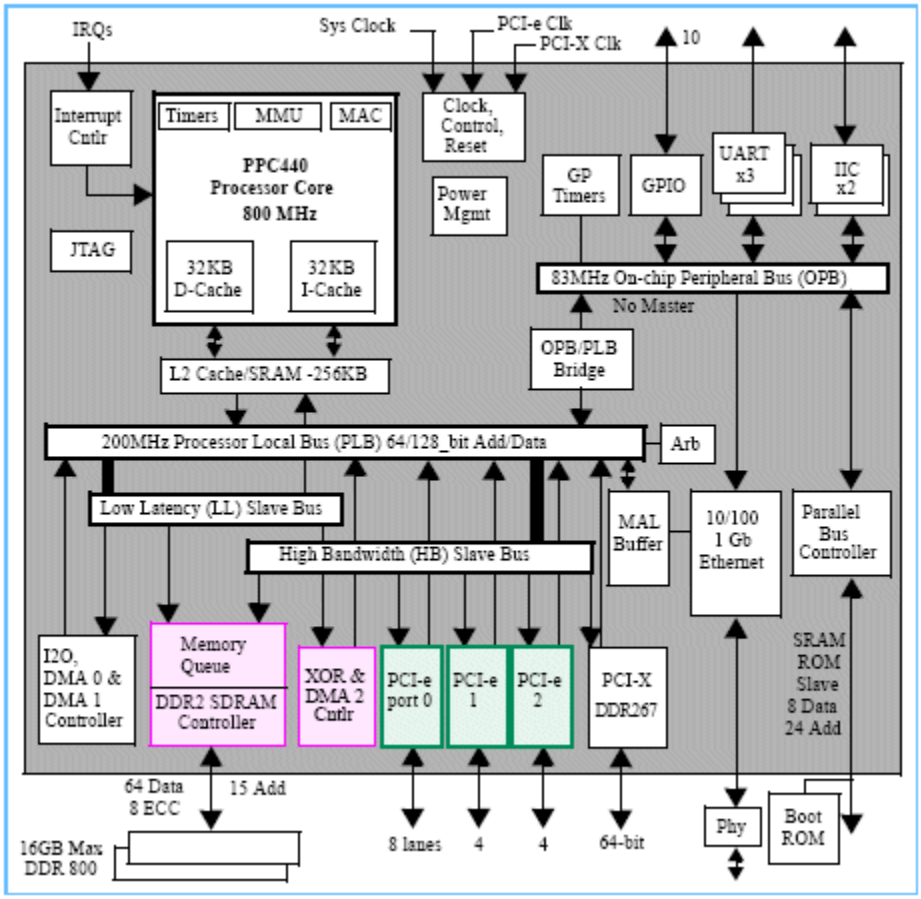

Figure 11: PowerPC IOP???/PowerPC IOP block diagram

??/CONCLUSION

An high performance SOC based on a PowerPC CPU core for RAID storage application, have been tested at 800MHz. with main interfaces such as DDRII SDRAM at 800MHz and three PCI-Express ports. Two RAID hardware accelerators, permits to achieve data throughput in the range of 1 GBytes per second.

??????, ????PowerPC CPU??????SoC, ????????DDRII 800MHz???PCI-Express????????800MHz. ??RAID?????, ???????1GBytes??????? ??.

[1] A PowerPC SOC IO Processor for RAID applications, G.Boudon & al. IPSOC04

[2] IBM Corp. (1999) Coreconnect Bus Architecture

[3] "Integrating PCI Express IP in a SoC" Ilya Granovsky, Elchanan Perlin, IP/SOC06