FPC3 Xilinx Xi5
Design and notes

Mark Hayter and Richard Black

10th March 1994

The version 3 Fairisle Port Controller uses a xilinx 3090 as its main controller. This is substantially larger than the previous versions. Advantage has been taken of this to include many experimental and debugging aids in the device. The new design, the Xi5 bits, is based on the earlier Xi3 version with additions. The occasion was also exploited to make support of standard confirmant headers easier - in particular to present, and remap on contiguous VPI bits. The FPC3 also includes the transmission system on the main PCB, the improved txrx control chip forms the taxi5 design. This document describes the interface to these chips, and should be read along with the Xi3 and txrx chip descriptions.

1 Main xilinx changes

The Xi5 xilinx on the FPC3 is based on extensions to the Xi3 version on the FPC2. The queue management is identical, but the interupt and status details have changed. Externally the main changes are that the xilinx gets to see all of the SRAM data bus, allowing increases in sizes of timers and status information, and connections are made to the txrx chip to allow the CPU to monitor the state of the line.

Internal changes:

- **Byte Swapping.** When reading the header of an incomming cell, the Xi5 design swaps over the first two octets. In other words they are written in a big-endian format even though the rest of the cell is written in little-endian format. This ensures that the VPI bits in the standard header are contiguous in that word. Also it is these VPI bits that are used for referring to the cell buffers when performing remap operations.

- **Source stamping** The source information is now written to the same data bits as are used for the PT field in standard conformant headers. This means that a cell for Wanda (where the stamp is written to both the first two words) can be identified by checking the PT field. Hence no VPIs need be reserved for this purpose.

- **Fifos.** The FPC3 uses synchronous fifos and some changes must be made to enable the use these. In addition the xilinx will become responsible for resetting the fifos, both to ensure timing constraints are met and to allow separate resetting of them.

- **Status.** The status register will be increased in size to pass the extra information from the txrx chip. The report of line violations will be done using a count of the number of violations sine the last status clear. The counter will stop at all bits 1 to indicate overflow. There will also be an indication of the current violation state. If any of the latched status bits is set then an IRQ (maskable in IOC) is raised, and there is a separate IRQ for violations. If the receive fifo should become full through error this is also reported.

- **Free running clock (FRC).** The free running clock is increased to 16 bits, incremented every frame pulse. This count will continue to be readable by the processor and put in as the cell timestamp.

- **Timer clock (TC).** New to Xi5 this clock may be read and written by the processor. It will decrement every frame pulse and cause a FIQ on reaching zero. The FIQ is removed when TC is written with a new value - the IOC mask can be used to disable it at other times.

- **Timestamping.** The new xilinx can check the incomming VCI cells. The FDL specification has 16 bits in the VCI field, but the xilinx remap system only supports 11 bits. Three of the spare bits will, if non-zero, enable automatic timestamping into the **data** portion of the cell. The *inbound* VCI (which is used as the offset into the remap vector for cells going into the fabric) is used for determining the offset to place the timestamp in both incomming and outgoing cells. Table 1 indicates the mapping used - for convenience both the offset in the xilinx cell buffer and the offset in the data portion of the cell are shown. Clearly there is no synchronisation between port controllers, so the only meaningful thing to do is to compare measurements made on the same port controllers - though port controllers on the same fabric will increase at the same time. This function can be enabled and disabled by writing a command to the xilinx.

VCIin	In from Rx		Out to fabric	
bits	Buffer	Data	Buffer	Data
13..15	word	word	word	word
0	none	none	none	none
1	3	1	2	0
2	5	3	4	2
3	7	5	6	4
4	9	7	8	6
5	11	9	10	8
6	13	11	12	10
7	none	none	none	none

Table 1: Timestamp offsets

- **Interrupts.** The xilinx uses three IOC FIQ lines and two (edge sensitive) IRQ lines. All masking of interupts must be done by IOC.

- **Local cells.** The "local" bit in the transmit buffer address has been removed. All cells have their VCIs remapped going in to the fabric.

- **Pass-on multicast.** The portc route byte has been expanded by one bit. If set then a cell from the fabric is put into both transmit and loopback fifos (the cell is rejected if either is more than half full). This allows (probably with some VCI trickery) serial multicast by passing the cell on to another output. The portc route byte becomes:

spare 7-3	Select 2-1	Active 0	Use
X	XX	0	No cell
X	00	1	Tx fifo
X	01	1	Loop fifo
X	10	1	Both fifos
X	11	1	Loop fifo

Table 2: FPC3/Xi5 Port Controller Route byte

2 Clocks and Timestamping

The changes made to the Xi5 bits for assisting experimental work are the addition of clocks and timestamping. The design of these was driven by the requests made by experimenters who wished to measure the performance of the switch and investigate different cell scheduling mechanisms. The timer clock **TC** was added to enable controlled generation of CPU interrupts in particular to aid with rate based scheduling disciplines. Timestamping in the data portion of cells was added to enable delay measurements.

The width of the clocks was chosen to be 16 bits because when a stream of cells is passing through the port controller from a particular connection, the clock(s) will respectively be used to measure (timestamp) and schedule the cells for transmission according to a negotiated rate. For a given cell the time spent at the port controller is measured by the difference (modulo the clock size) between the time of its arrival and the time of its departure on the FRC. If the clock is insufficiently wide then it may roll over completely whilst the cell is at the port controller, and the measurement of the cell's residence time at the port controller will be incorrect by one full clock period. With a clock of width 16 bits, and a frame pulse every $3\mu s$, cells from a periodic (CBR) source of greater than 1.9 Kb/s, the clock will never roll over completely within the interarrival time of two cells. It will therefore be sufficient to use the FRC for delay measurements and the TC for scheduling to ensure that all streams are accurately measured.

A stream of cells transmitted through the network with the timestamping VCI bits set up appropriately will permit the measurement of jitter and delay at each node (switch) traversed by the stream. This will be particularly valuable for cell level scheduling experiments, where different scheduling algorithms will be compared. These algorithms will aim to ensure that each class of traffic meets its QOS requirements, and that the overall performance requirements of the ATM multiplexer at a port controller are met. Implementing the delay stamping in hardware saves approximately 10 instructions per cell in the fiq for experiments in which cell delays must be measured. It has the additional advantage that it can be selectively enabled for a particular connection and for a particular port controller, by setting up the delay stamping VCIs appropriately.

3 TXRX Xilinx (Taxi5)

The transmission system xilinx will be extended to provide more status information as described above. In addition the following changes will be made:

- **Line speed.** In order to allow the testing using various line speeds the transmission system can be slowed. The line will still use 100Mbit data rate (125 MHz clock), and will insert "sync" characters to slow the useful data rate. An 8 bit value **n** will be given as the maximum value for a counter. The txrx will enable transmission 1 cycle in **n**, or 16 cycles in **n** (not uniformly spread) depending on a flag. This gives a reasonable coverage of the range 0.4 Mbps - 100 Mbps. Note at very slow speeds this could cause problems for the timeout mechanism.

- **Links**. The links on the existing card (HEC CRC and handshake) are be replaced by software control.

4 Xilinx interface.

The interface and registers will be very similar to the "xi3" design. The basic outline is shown in the Xi5/Xi6 Info sheet. The "TimeStamp" generated by the xilinx consists of the status register in the top 16 bits and the FRC in the low 16 bits. Thus it includes source and time information.

A read operation from a register (Rm set) uses the selR field to select the register to be read in the low 16 bits, and the selS to select the status register to be read on the top 16 bits. In addition there are two flag bits. nRL is only used if ArmP is being read and should be set to prevent automatic reload. clrV is only used if the status register is read, and causes the violation and HECfail counters to be cleared.

A write to memory may cause the transmit buffer counter to be incremented by setting the Inc bit in the address. In this case if the data written has the enW bit set then the top 16 bits of the data will be written to TC as well as to memory.

In a write to registers address bits are used to select the register(s) to be written. There is nothing to stop data being written to several registers in the same operation. Again the Inc bit may be set to cause the transmit counter to increment. TC will be written if bit enW is set in the data and either the Tc bit or the Inc bit is set in the address.

The command register interprets the bit pattern given and peforms zero or more operations. Bits should be set to enable the operation.

- **GoTx.** Start the transmission system after a reset or NACK. It is the responsibility of the software to ensure all the pointers and links are in a sane state before this command is issued.

- **setD.** Set the internal delay-stamp enable flag from EdF. This allows the automatic timestamping of cells based on their VCIin to be enabled and disabled.

- **RqW.** Requests a buffer be moved from the free queue to the recieved queue. This operation will be performed as soon as the receive system is searching for a new start of cell. Only one request will be held pending, until the buffer has been unlinked from the free queue additional requests are ignored. The buffer will have the TimeStamp written in both words 0 and 1 so software may detect it by the top VCI bit being set.

- **rRx,rTx,rLp.** These three bits enable updating of the reset state of the three fifos. For any enabled lines the EdF value is copied into the reset register. If the reset register bit is zero (which it will be after the xilinx is reset) then the fifo is disabled and reset.

- **Wtr.** If this bit is set then the TxRx chip parameters are loaded. The txCount value is used as the reload value for a down counter which reloads on reaching zero. The counter decrements at the byte transmit rate. If the inN bit is clear then data will only be transmitted when the count is zero, if inN is set transmission is enabled whenever the count < 16. The nHan and nCrc bits replace the links on the old transmission card, and if set disable handshake and HEC CRC generation.

The interupts generated by the xilinx are shown in table 3. The IR and IF lines are edge triggered, so will only generate an interupt when the status appears. The current value of those signals may be read via the IOC control register.

Type	Name	bit	Use
FIQ	FH[0]	fiq[0]	Rx cell, NACK or FreeEmpty
FIQ	FH[1]	fiq[1]	Tx Availible
FIQ	FLbar	fiq[2]	TC timed out
IRQ	IR	irqA[2]	latched status
IRQ	IF	irqA[3]	violation or HECfail
IRQ	IL4	irqA[4]	receive fifo went full

Table 3: Xilinx interrupts

5 Software

Although the FIQ code for Xi5 is almost identical to Xi3, it was realised early in the design that a significant amount of work would be needed on the Wanda management code to use Xi5 correctly, and that that overhaul was already overdue. Unfortunately that work has not even begun. The current software is a set of horrible hacks that try and grunge the Xi5 device under the existing management code.

One problem is the fact that VCIs have a different value when transmitted and received (due to the byte swap), including the Meta Signalling VCI! The other is the lack of the local bit. This hasn't even been fixed for local connections. Instead the software circulates round half of the remap entries which are not used by normal remapping (because Wanda still only supports 2^{10} connections not 2^{11}).

Amazingly enough this holds together well enough for some experiments to have taken place. From the performance point of view the FPC3 has been observed to be capable of the full line rate when FIFO queueing. Although there is never enough CPU left at this rate to run the operating system, the FIQ was observed to empty its work queue from time to time, indicating that it is not fully saturated.

More complex experiments have also taken place, but they will be detailed elsewhere.

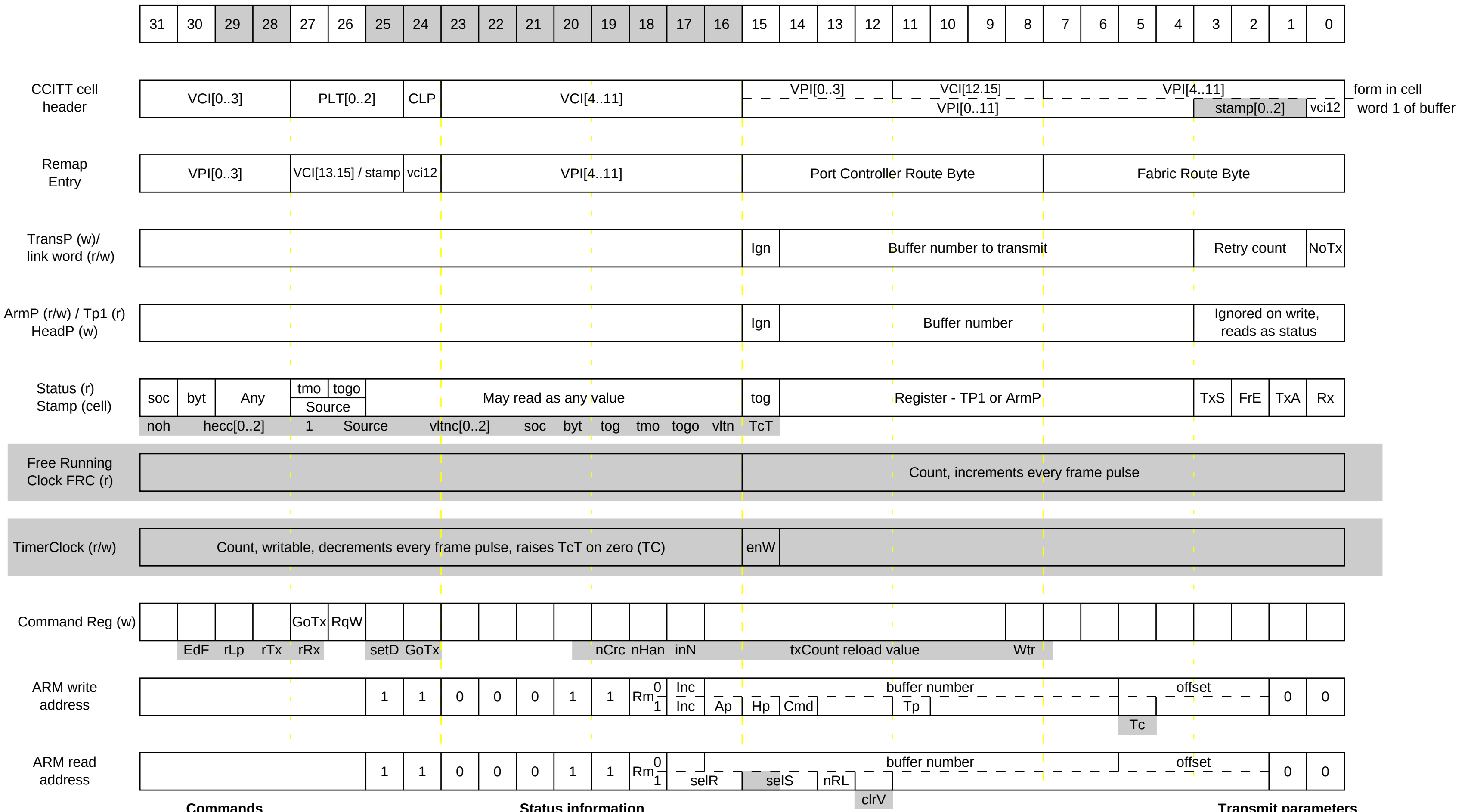
6 The Xi6 design

The Xi6 design is a retro fit of the byte swapping and payload type code into the Xi3 design for FPC2 systems. FPC2 specific features are as usual. The data and address interface to the xilinx chip is a subset of that specified in this document which is indicated in the diagram. The FRC register is actually present, but always returns zeros.

7 Approved Bits

The only approved bits for Xi5 are ThUlachne13HACK.lca. The only approved bits for Xi6 are W3Emaera1.lca. Use of all other bits (which are known to contain bugs) is prohibited.

FPC2 / Xi6 and FPC3 / Xi5 Data and Address -- Anything on gray is in Xi5 but not Xi6



Commands
(unused bits should be set to zero)

GoTx	- Start the transmission system
setD	- Set to copy EdF to delaystamp enable
RqW	- Request a buffer for Wanda
rRx	- Set RX system reset to EdF
rTx	- Set TX system reset to EdF
rLp	- Set LP system reset to EdF
EdF	- Enable/disable(reset) flag, set to enable
Wtr	- Set to write txrx information
txCount	- Max value (N) for slow tx
inN	- Set for 16 in N tx, clear for 1 in N
nCrc	- Set to disable HEC CRC generation
nHan	- Set to disable Handshake generation

Status information

Source: 00 - LpFifo, 01 RxFifo, 1x 11 WandaRq

togo	- There are cells to be transmitted
tmo	- (latched) A cell was timed out on rx
tog	- (latched) A cell was lost since no buffers
byt	- (latched) Stray bytes were seen
soc	- (latched) Extra SOCs were seen
vltnc	- Count of violations
hecc	- Count of HEC failures
vltn	- Current violation state
noh	- Rx cells have no HEC
TxS	- Transmir stopped, ie RESET or NACK
FrE	- Free Queue is empty
TxA	- Transmit buffers availible
Rx	- Receive buffers to collect
TcT	- Timer Clock timer out (passed 0)

Cell Header

VPI	- Virtual Path ID
VCI	- Virtual Channel ID
CLP	- Cell Loss Priority
PLT	- Payload Type (set to 11x for Wanda cells)
stamp	- Word to timestamp in (=VCI[12..14])

ARM write address

Rm	- 1 for register 0 for SRAM
Inc	- set to increment txCNT, writes TC if enW set
Cmd	- set to write command
Tp	- set to write TransP
Ap	- set to write ArmP
Tc	- set to write TC, enW must also be 1
Hp	- set to write HeadP

Transmit parameters

Ign - Ignored on write, reads as status
Retry Count - Number of times to retry on NACK
NoTx - Don't transmit, just put on free queue

NB: ALL CELLS HAVE THE VCI REMAPPED
Inbound VPI[10..0] (word1[14..4]) -> remap entry

ARM read address

	selR	selS
00	Armp	status
01	TP1	status&clr
10	FRC	TC
11	FRC	TC

nRL - set for no reload on ArmP read
clrV - set to clear vltnc & hecc on status read