Abstract
Typical lockstep systems are complicated to design, and typically require a large pin count of logic to be added to the system. The IBM PowerPC 750GX contains a lockstep facility that allows lockstep operation without complicated and bulky external logic, because the lockstep facility is located in the chip.

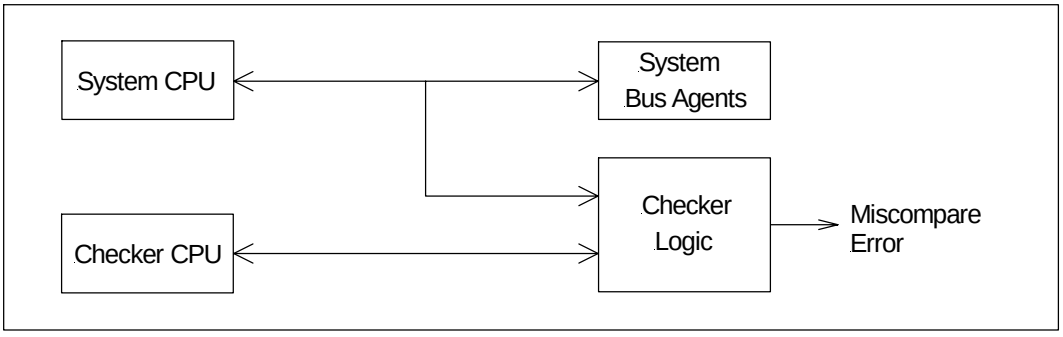
1. Introduction
Information in this note concerning the IBM PowerPC 750GX dd1.2 processor is intended to supplement that found in the ***IBM PowerPC® 750GX RISC Microprocessor User's Manual, Version 1.0***, and it applies only to revision dd1.2 (and any succeeding revisions) of the PowerPC 750GX processor. That document is referred to in this application note as the User's Manual. Since information about the processor is subject to change, designers should check the IBM Microelectronics Technical Library on the web for the most recent version of all PowerPC documentation.

2. The Lockstep Method
The lockstep processor technique is a way of achieving high reliability in a microprocessor system, by adding a second identical processor (the Checker) to a system to monitor and verify the operation of the system processor. The two processors are initialized to the same state during system startup, and they receive the same inputs (code, bus operations, and asynchronous events), so during normal operation, the state of the two processors is identical from clock to clock. They are said to be operating in lockstep.

The lockstep technique assumes that an error in either processor will cause a difference between the states of the two processors, which will eventually be manifested as a difference in the outputs, so the lockstep system monitors the outputs of the two processors and flags an error if there is a discrepancy.

In the conceptual lockstep system diagram of Figure 1, the system CPU connects normally to the system CPU bus agents. The CPU bus also connects to the Checker logic. The Checker logic steers the required inputs to the Checker CPU, and compares the outputs of the two processors. If there is any difference in the outputs of the two processors, the Checker logic asserts an error signal.

Figure 1. Lockstep System Concept



2.1 Typical Lockstep System
Figure 2 shows the block diagram of a typical lockstep system. In order to more clearly illustrate the lockstep concept, Figure 2 has been greatly simplified in two respects. First, all bussed signals are treated as either inputs or outputs instead of I/Os, and second, all outputs are shown as valid on the same clocks. Note the following features of a typical lockstep system:

1) CPU inputs from the system are connected to both processors.
2) Outputs from the system CPU are connected to the system bus agents, and to the Checker logic.
3) Outputs from the checker CPU are connected only to the Checker logic. The Checker CPU does not drive the system CPU bus at any time.
4) The Checker logic compares the outputs of the two processors on the clocks that they are valid.

Figure 2. Typical Lockstep System

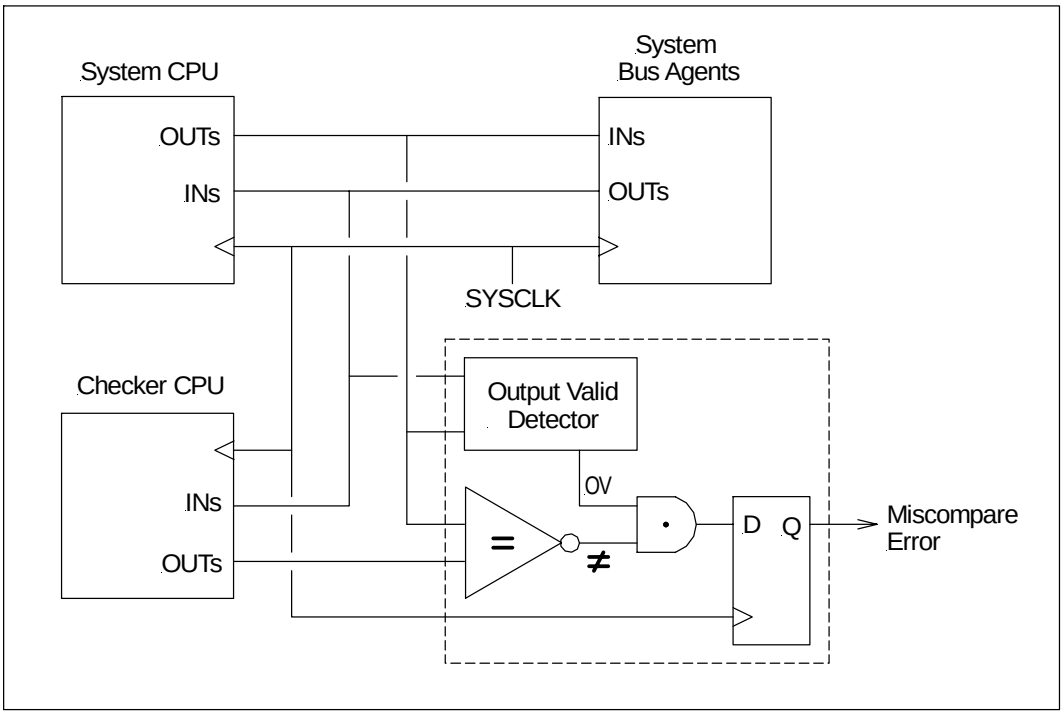


Table 1 shows an estimate of the number of signal I/Os required to implement the steering logic and control logic required in the typical lockstep system of Figure 2. The interface to the CPU bus might require 126 I/Os, the interface to the Checker processor might require 108 I/Os, and another 16 are budgeted for miscellaneous control pins, for a total estimate of 250 I/Os. With a bare minimum of 40 power and ground pins, at least a 292 pin package is required.

Table 1. Pd as a Function of Fcore & Vdd, Standard-Power Parts, 105C			
CPU Bus Interface	Data	64	
	Address	32	
	Address Attributes	12	
	Control Lines	18	126
Checker Processor Interface	Data	64	
	Address	32	
	Address Attributes	10	
	BiDi Control Lines	2	108
Miscellaneous Control I/Os	Misc	16	16

PowerPC 750 family systems generally use a “bridge”: a memory controller and/or PCI interface. If designing the bus and lockstep logic from scratch, a design team can combine the functions in one package and reduce the number of pins that are uniquely required for the lockstep function.

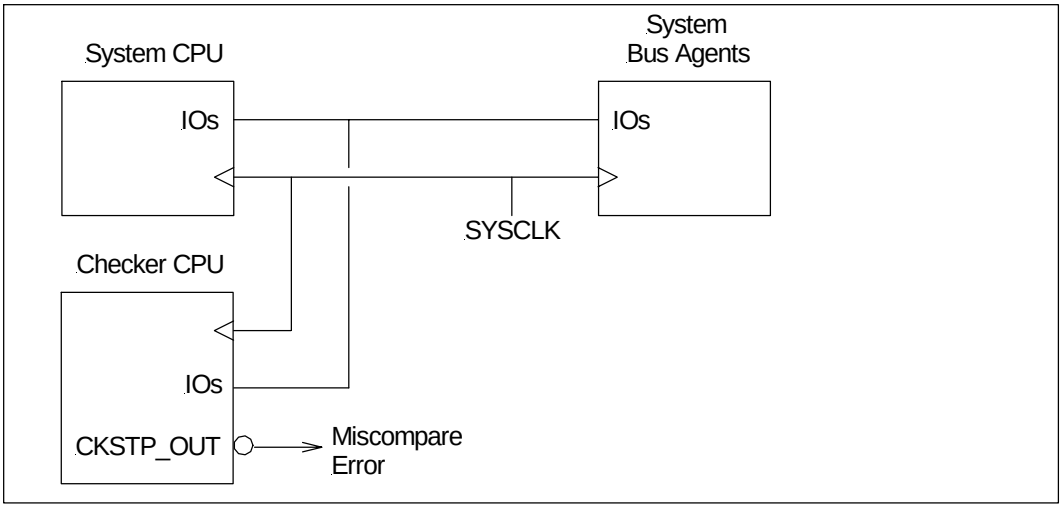
2.2 Features of the PowerPC 750GX Lockstep Facility

The IBM PowerPC 750GX processor contains an integrated lockstep facility (LSF). The LSF contains all of the functions required of a lockstep system, including the required data steering and state checking. Miscompares are signaled by the checkstop out pin of the Checker processor, allowing miscompare diagnosis, and giving the system designer a simple miscompare signal that can be used by the system processor or system logic. As shown in Figure 3, designing a PowerPC 750GX lockstep system is as

simple as connecting two PowerPC 750GX processors together, pin-strapping one as the lockstep processor, and deciding where to go for lunch.

The LSF has the following advantages:
1) The LSF offers designers the high reliability of a lockstep architecture,
2) No external buffers or state-checking logic has to be designed,
3) No external buffers or state-checking logic has to be added to the system,
4) There is no impact on bus speed, timing, or signal integrity,
5) The Checker processor checkstops, allowing miscompare diagnosis,
6) Miscompares are signaled by a simple output signal, allowing the designer flexibility in responding to the miscompare.

Figure 3. PowerPC 750GX Lockstep System Block Diagram



3. The PowerPC 750GX Lockstep Facility

A 750GX lockstep system is composed of two 750GX processors, a Checker and a Master, along with the usual bus logic for a single-processor system. The Master operates normally, and is unaffected by the inclusion of the Checker processor in the system. The Checker is a silent partner of the Master, and only affects the system if it detects a lockstep failure.

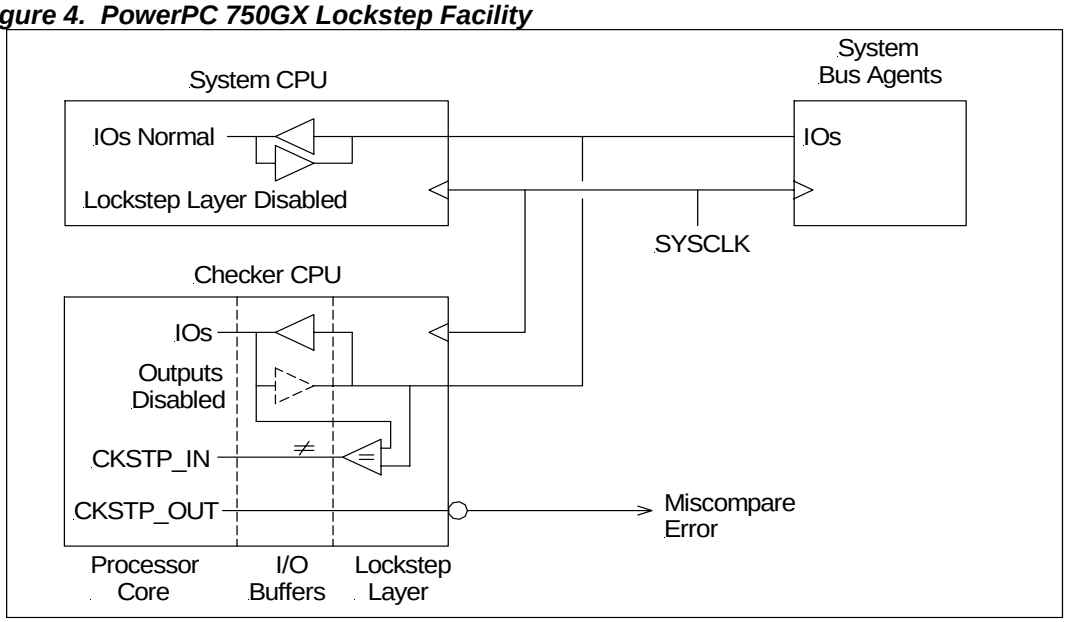
During hardware reset, the Checker and Master are initialized to the same state, and thereafter receive identical inputs. Thus while operation is normal, the internal state of the processors will be identical, and the outputs of the processors (if enabled) will be identical. If an error occurs in either processor, then the outputs of the two processors will miscompare.

The lockstep Checker functions are embedded in the Checker processor, eliminating the need for complicated logic and buffers with high pin counts.

3.1 The Master

As shown in Figure 4, the lockstep facility can be thought of as a logical layer that is added between the processor and the bus. While in normal (Master) mode, the lockstep circuitry is disabled. Thus the Master processor in a 750GX lockstep system is a 750GX that has not been placed in Checker mode. The operation of the Master is identical to the operation of the 750GX in a single processor system.

Figure 4. PowerPC 750GX Lockstep Facility



3.2 The Checker

As shown in Figure 4, the Checker is a 750GX that is operating with the lockstep facility enabled. In describing lockstep operation, it is useful to think of the Checker in terms of the processor core and the lockstep layer. The operation of the processor core in lockstep mode is identical to its operation in normal (non-lockstep) mode; it is the lockstep layer that provides the lockstep functionality.

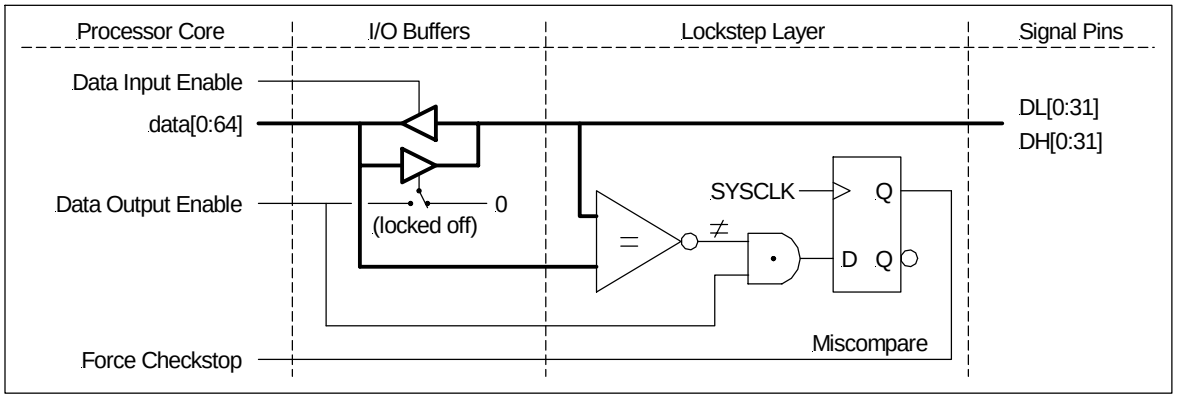
During lockstep operation, lockstep layer of the Checker has the following general characteristics:

1) Passes all inputs through to the processor core to maintain the state of the Checker core equal to that of the Master,
2) Compares the bussed outputs of the Checker core to the outputs of the Master while the outputs are valid, and
3) Disables the bus drivers of the Checker, so that it does not drive the bus.

Figure 5 shows a more detailed view of the 750GX in Checker mode.

- The operation of the processor core layer is unchanged from normal mode.
- The I/O buffers are locked off so that the Checker does not drive the bus.
- The lockstep layer compares the outputs of the Master processor to the outputs of the processor core of the Checker processor on the clocks that the outputs are defined as valid. A difference causes a Miscompare to be reported. The Miscompare signal forces the Checker to Checkstop.

Figure 5. Checker Mode Overview



3.3 System Electrical Considerations

Putting the 750GX into Checker mode has no effect on the electrical or timing characteristics. So system layout, timing, and signal quality issues are identical to a system using two 750GX processors in a multiprocessor configuration. See the IBM Application Note *Toward a 200MHz Bus Dual PowerPC 750GX System* for details, and the IBM PowerPC750GX Evaluation Kit for examples. Most manufacturers of commercially available bridges also make available reference designs or evaluation kits that feature dual 750GX processors. All of the topological characteristics and requirements of a dual 750GX multiprocessor system apply to a lockstep system, and no additional requirements apply.

Since lockstep mode does not change any of the IO characteristics, 200MHz bus operation is supported using the same layout considerations as apply to a 200MHz multiprocessor system.

3.4 System Thermal Considerations

The power dissipation of the 750GX is the same in normal and lockstep modes. Since the core speed, core power supply, and code load is the same for the Master and Checker, the maximum power dissipation is identical. Actual power dissipation will of course vary from module to module.

3.5 Detailed Connectivity

Table 2 represents a CPU bus populated by a Master 750GX, a Checker 750GX, and system logic.

- The IO column indicates the direction of the signal net at the processors.
- The Master 750GX, Checker 750GX, and System Logic columns show the pins of those bus agents.
- A "----" indicates that there is normally a connection. A "--x--" indicates that there is normally no connection. For example, BR# is connected between the Master and the Checker, and between the Checker and the system logic. Also see the notes to the Table.
- All connectivity requirements specified in the 750GX Datasheet and 750GX User's Manual, must be followed. The requirements in this section are in addition to those found in the Datasheet and the User's Manual.

Table 2. Connectivity

IO	Master		Checker		System Logic	Notes
O	BR#	----	BR#	----	BR#	2
I	BG#	----	BG#	----	BG#	1
IO	ABB#	----	ABB#	----	ABB#	2,3
IO	TS#	----	TS#	----	TS#	2
IO	A[0:31]	----	A[0:31]	----	A[0:31]	2
IO	AP[0:3]	----	AP[0:7]	----	AP[0:7]	3
IO	TT[0:4]	----	TT[0:4]	----	TT[0:4]	2
O	TBST#	----	TBST#	----	TBST#	2
O	TSIZ[0:2]	----	TSIZ[0:2]	----	TSIZ[0:2]	2
IO	GBL#	----	GBL#	----	GBL#	2
O	WT#	----	WT#	----	WT#	2
O	CI#	----	CI#	----	CI#	2
I	AACK#	----	AACK#	----	AACK#	1
IO	ARTRY#	----	ARTRY#	----	ARTRY#	3
I	DBG#	----	DBG#	----	DBG#	1
I	DBWO#	----	DBWO#	----	DBWO#	1,4
IO	DBB#	----	DBB#	----	DBB#	3
IO	D[0:63]	----	D[0:63]	----	D[0:63]	2
IO	DP[0:7]	----	DP[0:7]	----	DP[0:7]	3

Table 2. Connectivity						
IO	**Master**		**Checker**		**System Logic**	**Notes**
I	DBDIS#	----	DBDIS#	----	DBDIS#	1,3
I	TA#	----	TA#	----	TA#	1
I	DRTRY#	----	DRTRY#	----	DRTRY#	1
I	TEA#	----	TEA#	----	TEA#	1
I	INT#	----	INT#	----	INT#	1,5
I	SMI#	----	SMI#	----	SMI#	1,5
I	MCP#	----	MCP#	----	MCP#	1,5
I	SRESET#	----	SRESET#	----	SRESET#	1,5
I	HRESET#	----	HRESET#	----	HRESET#	1,5
O	RSRV#	----	RSRV#	----	RSRV#	6
I	TBEN	----	TBEN	----	TBEN	1
I	TLBISYNC#	----	TLBISYNC#	----	TLBISYNC#	1
O	QREQ#	----	QREQ#	----	QREQ#	6
I	QACK#	----	QACK#	----	QACK#	1
I	CKSTP_IN#	----	CKSTP_IN#	----	CKSTP_IN#	8
O	CKSTP_OUT#	----	CKSTP_OUT#	----	CKSTP_OUT#	8
	SYSCLK		SYSCLK		SYSCLK	6
I	PLL_CFG[0:4]	----	PLL_CFG[0:4]	----	PLL_CFG[0:4]	1
I	PLL_RNG[0:1]	----	PLL_RNG[0:1]	----	PLL_RNG[0:1]	1
O	CLK_OUT	-x--	CLK_OUT	-x--	CLK_OUT	6
I	TCK	----	TCK	---	TCK	7
I	TMS	----	TMS	---	TMS	7
I	TDI		TDI		TDI	7
O	TDO		TDO		TDO	7
I	TRST#	----	TRST#	----	TRST#	7
I	LSSD_MODE	----	LSSD_MODE	----	LSSD_MODE	1
I	L1_TSTCLK	----	L1_TSTCLK	----	L1_TSTCLK	1
I	L2_TSTCLK	----	L2_TSTCLK	----	L2_TSTCLK	1
I	BVSEL	----	BVSEL	----	BVSEL	1

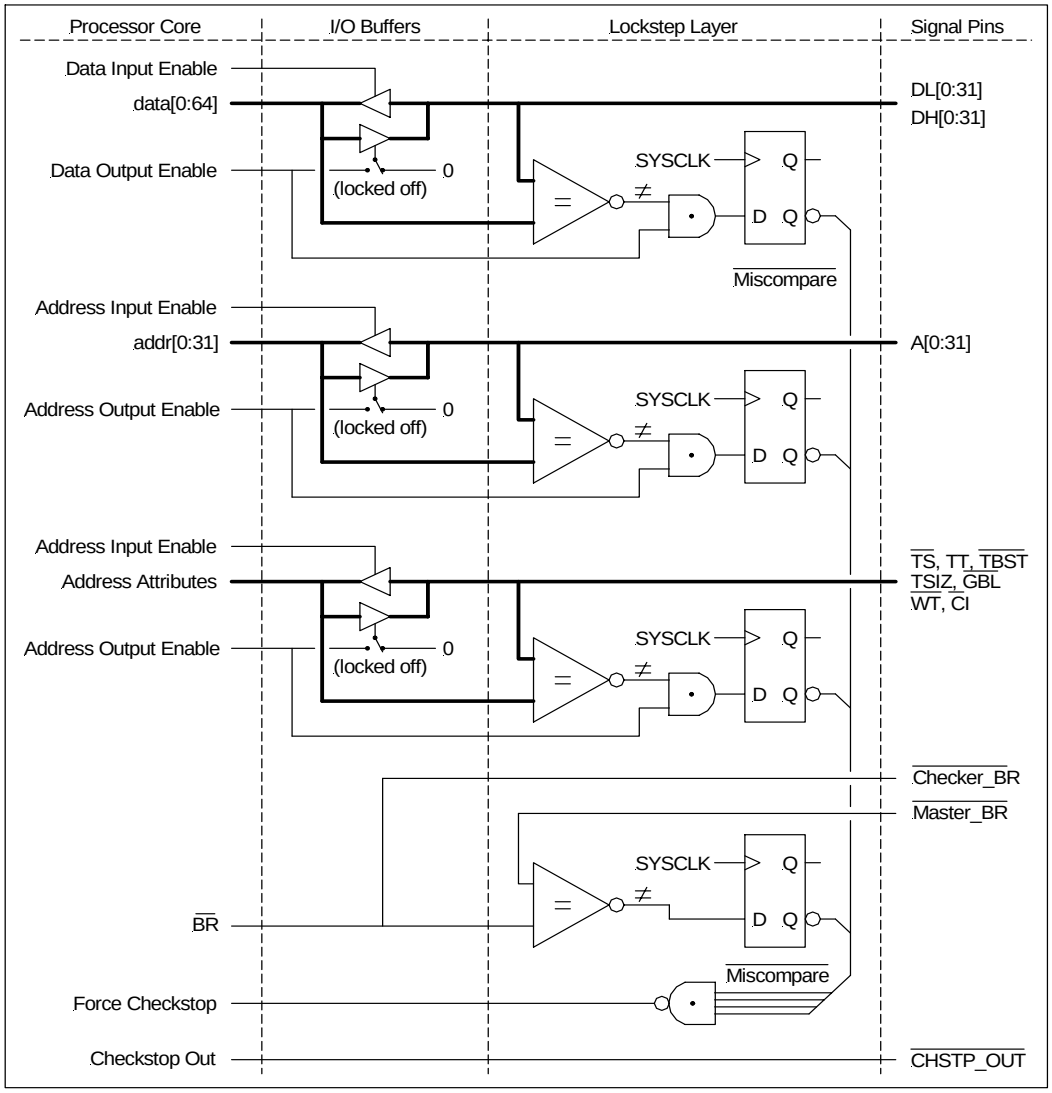
Notes to Table 2:

1. Input-only signals are not compared by the Checker. Inputs must be either connected between the Master and the Checker, or handled in some other way so that that they are identical when sampled by both processors (ex: pulled up).
2. Input-Output and Output-only signals are compared on clock cycles that they are valid outputs. BR# and TS# are compared on any cycle that they could be valid.
3. ABB#, AP[0:3], ARTRY#, DBB#, DBDIS#, and DP[0:7] are not compared by the Checker. These signals must be either connected between the Master and the Checker, or handled in some other way so that that they are identical when sampled by both processors (ex: pulled up). The Checker will never drive these signals.
4. DBWO# is used to select normal vs Checker mode. Whenever DBWO# is sampled asserted on the last clock that HRESET# is sampled asserted (for example, by connecting DBWO# to HRESET#), lockstep mode is selected. In systems that do not use DBWO# in normal operation, connect DBWO# to HRESET# for the Checker, and pull DBWO# high for the Master.

5. Inputs that are normally effectively asynchronous (INT#, SMI#, MCP#, SRESET#, HRESET#) must be received by both processors on the same clock. Therefore, the same ac, dc, timing, and edge rate conditions apply to these signals as to the address and data signals.
6. SYSCLK, CLK_OUT, RSRV#, and QREQ# are not compared by the Checker.
7. TCK, TMS, TDI, TDO, TRST#, LSSD_MODE, L1_TSTCLK, L2_TSTLCK, & BVSEL are not compared by the Checker. Connect the JTAG signals as normal for your debugger hardware in a multiprocessor system (MUX or daisy-chain). See Section 9 for more information.
8. The connectivity of CKSTP_IN# and CKSTP_OUT# depends on the system design. See Section 5 for more information.

Figure 6 shows a conceptual view of the Checker, showing the groupings of the various systems, and the conceptual logic in the lockstep layer. Note that only outputs are compared, and only on the clock cycles that the signal could be valid, given the current state of the bus. BR# is compared on every clock.

Figure 6. Checker Mode Conceptual Diagram



In response to a miscompare, the lockstep logic forces a checkstop, even if the machine check exception is enabled. There is no way to disable the checkstop or to otherwise alter the operation of lockstep mode. To disable the lockstep checkstop, disable lockstep mode. In response to the checkstop state, the Checker asserts the CHKSTP_OUT# output.

4. Initialization

In order to initialize the system for lockstep operation, it is necessary to configure both processors for deterministic operation, and to put the Checker into lockstep mode.

4.1 Array Reset and Deterministic Operation

In non-lockstep systems, as long as conformance to the architecture is maintained, it is not necessary for the processor to behave in a completely deterministic manner at all times. For example, as long as the processor maintains coherency, performance, and other operational specifications, it does not matter if a particular cache block is stored in way 1 or way 3 of an L1 cache. Equally, it does not matter if that behavior is determined by a semi-random process during processor reset.

However, in a lockstep system, both processors must appear to the system to be in an identical state on each clock. Previous 750 designs required code sequences to synchronize the arrays to prevent lockstep failures. Now much of this work is done by the array reset provided by the 750GX. Using array reset, it is simple to set the 750GX into a deterministic state following hreset. Array reset must be activated on both processors whenever the 750GX is used in lockstep mode.

DBDIS# is used as a typical 750 strapping pin to initiate an array reset. Whenever DBDIS# is sampled asserted on the last clock that HRESET# is sampled asserted (for example, by connecting DBDIS# to HRESET#), an array reset will be started.

It takes 32k core clock cycles after the deassertion of HRESET# for the 750GX to initialize the arrays. When array initialization has completed, the 750GX will issue a bus request for the reset vector.

Array reset initializes all of the 750GX arrays except the GPRs, the FPRs, and the BATs. As usual, general purpose registers and floating point registers must be initialized to a known and appropriate value before being stored to memory. The BAT registers must also be initialized as usual in order to avoid errors during address translation.

4.2 Selecting Lockstep Mode

The Checker processor must be put into lockstep mode during the hreset sequence. DBWO# is used as a typical 750 strapping pin to select lockstep mode. Whenever DBWO# is sampled asserted on the last clock that HRESET# is sampled asserted (for example, by connecting DBWO# to HRESET#), lockstep mode is selected. Systems that use DBWO# during normal operation must also assert DBWO# while HRESET# is asserted.

5. System Actions on Miscompare

The elegance of the 750GX lockstep facility makes the typical application simple: connect the two processors and the bridge as shown in Table 2, and then choose an action to take if a miscompare is detected. Several options are possible.

The information available from the system depends on which processor checkstopped:

- If only the Checker checkstops, then a miscompare occurred, and the system does not know which processor malfunctioned.
- If both processors checkstop, then the system threw a checkstop error (ex: double memory fault).
- If only the Master checkstops, then there is a catastrophic error (possibly both processors malfunctioned, or there is a bus hang), since anything that checkstopped the Master should also checkstop the Checker.

5.1 Tie all checkstop inputs and outputs together

Connect the CKSTP_IN# and CKSTP_OUT# pins of both processors together. This causes both processors to checkstop if either processor checkstops. A miscompare or a checkstop of the Master processor will checkstop both processors.

- This option is simple, and freezes the processors in place for possible troubleshooting by analysis of the internal arrays via the RISCWatch debugger.
- The downside of this option is that the system is locked up after a miscompare, and no recovery is possible.

5.2 Use logic to hreset both processors

This option uses minimal external logic to assert HRESET# to both processors if the Checker asserts CKSTP_OUT#.

- This option resets both processors, thus recovering the system hardware from the miscompare. The hardware is restarted as soon as possible after the miscompare.
- This option does not seek to recover data or retain any information about the miscompare.

5.3 Use logic to achieve an orderly shutdown

This option uses state machines in an ASIC or service processor to attempt to gather & store information on the fail, and to bring the system to a recoverable state, such as a system image snapshot from a previous checkpoint. The service logic then resets both processors and restarts the board, which resumes operation from the checkpoint.

- This option offers the highest capability of error recovery, since a miscompare causes a roll-back to a previous checkpoint, and then a restart.
- This option loses the internal state of the Checker & Master at the checkstop, but if a processor checkstops, it would presumably be swapped out relatively quickly, and in the lab the service logic could be set to checkstop both processors on a miscompare, so that the internal arrays can be examined for the source of the problem.

5.4 Use logic to automatically isolate & fix the problem

- This option is not usually feasible.
- A timely replacement of the board is usually more cost-effective

6. Alternate Processor Bus Modes

Alternate bus modes that are supported by the 750GX are supported in lockstep mode. These alternate modes include 32-bit Data Bus mode, DRTRY# mode, and no-DRTRY# mode.

Contact IBM PowerPC Applications Engineering (ppcsupp@us.ibm.com) before designing a lockstep system using any of these alternate bus modes

7. Multiprocessor System Architecture

The lockstep facility is designed to work in a multiprocessor environment, but there are issues concerning bus loading and signal quality and timing that should be considered carefully. We recommend that designers check with IBM PowerPC Applications Engineering (ppcsupp@us.ibm.com) before designing a multiprocessor system using lockstep.

8. A Voting Processor System Architecture

It is possible to build a simple quasi-voting processor architecture using one Master and two Checker processors. This system is able to differentiate between a Master fault and a Checker fault, and resume operation on a Checker fault.

This system uses a single Master and two Checker processors, and service logic to manage miscompares.

If a single Checker detects a miscompare, then the error is in that Checker. Three recovery methods are feasible:

1. The system can take itself offline, record the error, and restart all three processors, and restore the image from the latest checkpoint (or whatever recovery method is preferred), or
2. Since no corruption has occurred, because there was no error in the computing unit, the system can save its state, restart all three processors, reload the saved state, and resume operation, or
3. The system can take the Checker offline until the card is replaced. The board would revert to single-Checker operation until replaced.

If both Checkers detect a miscompare, then the error is in the Master. The system would take itself offline, record the error, restart all three processors, and restore the image from the latest checkpoint.

If all three processors checkstop simultaneously, then there was a non-recoverable system error, such as a memory access double-fault.

If the Master signals any error without concurrence from both Checkers, then there is a serious hardware error, and the card must be taken offline & replaced. It is possible that data has been corrupted, since the lockstep algorithm was not in operation.

This system architecture has better error coverage than the single-Checker system, in that it can differentiate between an error in the Master vs an error in one or both of the Checkers. The 750GX lockstep facility also makes this design relatively inexpensive and simple to design, because the most challenging error detection work is done by the processor.

System operation at bus speeds up to 200MHz is possible using commercially available bridges.

9. RISCWatch Considerations

RISCWatch works in the normal fashion with the 750GX in both normal and lockstep modes. While the system is not operating in lockstep mode, RISCWatch can be used normally. All JTAG signals must conform to the timing specifications in the 750GX Datasheet.

While the system is operating in lockstep mode, TCK and TRST# must meet the same setup and hold times to SYSCLK as are specified for CPU bus signals. Since this is generally not possible, we recommend that RISCWatch (and other JTAG debuggers) only be used when the system is not in lockstep mode. This is not typically an issue, since code development is independent of lockstep mode operation.

If a miscompare occurs and the processors checkstop, the Master and Checker can normally be accessed individually via RISCWatch, without having to synchronize TCK and TRST# to SYSCLK.

IBM Systems and Technology Group
2070 Route 52, Bldg. 330
Hopewell Junction,
NY 12533-6351

The IBM home page can be found at
http://www.ibm.com

The IBM Microelectronics home page
can be found at http://www.ibm.com/chips

November 17, 2005