Functional Verification of the HP PA 8000 Processor

The advanced microarchitecture of the HP PA 8000 CPU has many features that presented significant new verification challenges. These include out-of-order instruction execution, register renaming, speculative execution, four-way superscalar operation, decoupled instruction fetch, concurrent system bus interface, and PA-RISC 2.0 architecture enhancements. Enhanced functional verification tools and processes were required to address this microarchitectural complexity.

by Steven T. Mangelsdorf, Raymond P. Gratias, Richard M. Blumberg, and Rohit Bhatia

Computer system performance has been improving recently at a rate of 40 to 60 percent per year. This growth rate has been fueled by several factors. Advancements in integrated circuit technology have made higher microprocessor clock rates and larger caches possible. There have been contributions from system software as well, such as compilers that emit more efficient machine code to realize a given function. The PA-RISC instruction set architecture has evolved to keep pace with changes in technology and customer workloads.

These factors alone, however, would not have been sufficient to satisfy customer demand for increased performance in a very competitive industry. The balance has been made up by innovations in microarchitecture that increase the amount of useful work that a microprocessor performs in a clock cycle. This has increased the complexity of the design and thus the effort required for successful functional verification.

Many of our previous microprocessor projects have reused existing cores (although generally with significant modifications and enhancements). In contrast, the HP PA 8000 CPU has a new microarchitecture that borrows little from previous projects. Some of the features in its microarchitecture presented significant new verification challenges:

- Out-of-order execution. A 56-entry queue of pending instructions is maintained by an ***instruction reorder buffer*** (IRB). The queue hardware selects instructions for execution that have their operands available irrespective of program order.
- Register Renaming. Write-after-write and write-after-read ordering dependencies are eliminated by remapping references from an architectured register to a temporary register.
- Speculative Execution. The PA 8000 predicts whether a branch is taken and can tentatively execute instructions down the predicted path. The side effects of all such instructions must be canceled if the prediction turns out to be incorrect.
- Four-Way Superscalar Operation. The PA 8000 has ten functional units and can sustain an execution rate of four instructions per cycle.
- Decoupled Instruction Fetch. Instructions are fetched and inserted into the queue by an autonomous ***instruction fetch unit*** (IFU). The IFU performs branch prediction and caches the target addresses of recently taken branches in a ***branch target address cache*** (BTAC).
- Concurrent System Bus Interface. Memory requests can be issued out of order, and data returns can be accommodated out of order. Up to 16 requests can be outstanding at a time.
- PA-RISC 2.0 Architecture Enhancements. These provided important new capabilities, such as 64-bit addressing and computation, but they necessitated tool rework and limited reuse of existing test cases.

This paper describes the enhanced functional verification tools and processes that were required to address the daunting microarchitectural complexity of the PA 8000.

Verification Overview

The purpose of functional verification is to identify defects in the design of a microprocessor that cause its behavior to deviate from what is permitted by the specification. The specification is the PA-RISC instruction set architecture and the bus protocols established by industry standards or negotiated with the designers of other system components. Performance specifications, such as instruction scheduling guidelines committed to the compiler developers, may also be considered.

Although it is not possible to prove the correctness of a microprocessor design absolutely through exhaustive simulation or existing formal verification techniques, a functional verification effort must achieve two things to be considered successful. First and foremost, it must provide high confidence that our products will meet the quality expectations of our customers. At the same time, it must identify defects early enough in the design cycle to avoid impacting the product's time to market.

A typical defect caught early in the design cycle might cost only one engineering day to debug and correct in the RTL (Register Transfer Language). Close to tape release, it might take five to ten days to modify transistor-level schematics and layout, modify interblock routing, and repeat timing analysis. Therefore, tape release can be delayed if the defect rate is not driven down quickly.

After tape release, lost calendar time is the primary cost of defects because the time required to fabricate a new revision of the design is at best a few weeks and at worst a few months. Defects that are so severe that they block a software partner's development, tuning, or testing efforts can put them on the critical schedule path of the product. The worst-case scenario is a masking defect that blocks further testing efforts for a certain functional area of the design, and this delays the discovery of additional defects by the time required to fabricate a new revision. One or more masking defects in series can quickly devastate the product schedule.

The PA 8000 verification effort consisted of a presilicon phase and a postsilicon phase. The purpose of the presilicon phase was to find defects concurrently with the design, when the cost of correcting them was small, and to drive up the quality level at first tape release so that the first prototypes would be useful to our software partners. This was done using three tactics: RTL simulation, accelerated simulation, and switch-level simulation. The postsilicon effort consisted of aggressive characterization of hardware prototypes to complete verification before systems were shipped to customers. Also, performance verification was done at various stages in the project.

RTL Simulation

Most previous PA-RISC microprocessor projects have built their functional verification efforts around an internally developed RTL simulator that compiles RTL descriptions of blocks into C code which are then compiled with HP's C compiler. Block execution is scheduled dynamically using an event-driven algorithm. This simulation technology achieves modest performance (about 0.5 Hz running on a typical workstation), but it does provide capabilities for rapid prototyping such as the ability to simulate very high-level RTL and quick model builds. Therefore, our RTL simulator became the cornerstone of our verification effort early in the design.

Fig. 1 shows the verification environment used for RTL simulation. There are four basic components in the environment:

- The RTL model for the PA 8000.
- Bus emulators, which can apply interesting stimulus to the input buses of the PA 8000 including responses to its transactions. We included emulators for all components sharing the system bus including the memory system, I/O adapter, and third-party processors.
- Checking software, which monitors the behavior of the PA 8000 and verifies that it complies with the specifications. This also helps speed debugging by flagging behavioral violations as soon as they occur.
- A variety of test case sources and tools that can compile the test cases into an initial state for the PA 8000 model and configure the bus emulators.

Checking Software

The most important check is a thorough comparison between instructions retiring in the PA 8000 model and instructions retiring in the PA-RISC architectural simulator. ***Retiring*** means exiting the instruction reorder buffer, or IRB (see ***Article 1***). A tool called the ***depiper*** captures information about each instruction retiring in the PA 8000 model, including what resources (such as destination registers) are being modified and the new values. The synchronizer compares this with similar information obtained from the PA-RISC architectural simulator which is also running the same test case. This provides very high confidence that the PA 8000 complies with the basic PA-RISC instruction set architecture. A final-state comparison of all processor and memory state information is also done at the end of each test case.

The depiper also provides the synchronizer with information about architecturally transparent events such as cache misses. Using this information, the synchronizer can perform strong checks in the areas of cache coherency, memory access ordering consistency, and memory-to-cache transfers. In addition, a number of checkers were developed for other areas:

- A checker for the instruction queues, including whether the order in which instructions are sent to functional units complies with data dependencies
- A checker for protocol violations on the system bus
- A checker for the bus interface block, discussed in more detail below
- A checker that detects unknown (X) values on internal nodes.

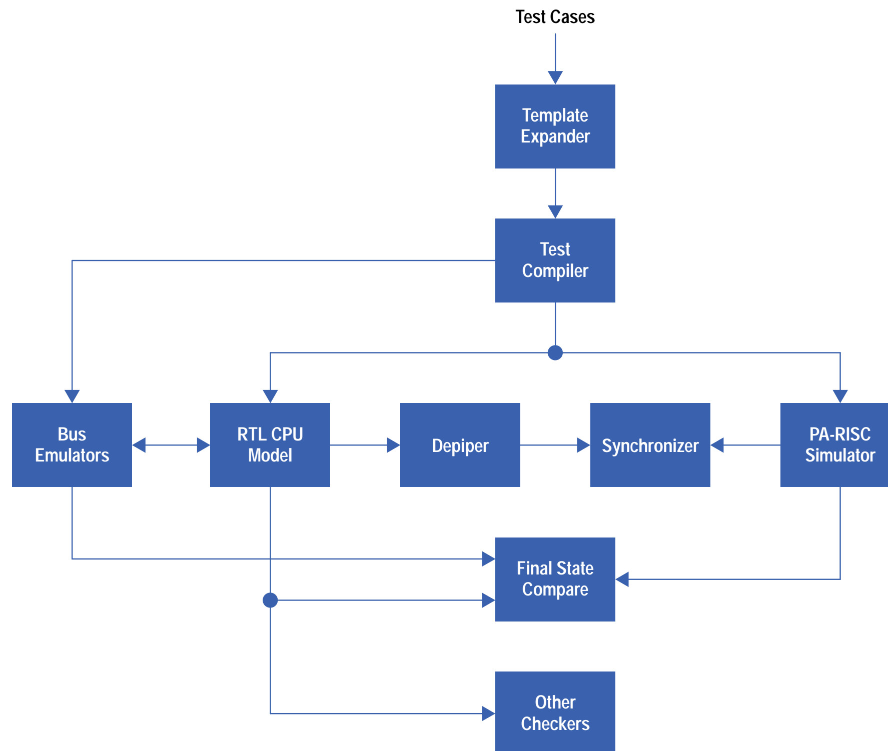


Fig. 1. *Block diagram of the presilicon RTL (Register Transfer Language) verification environment.*

Test Case Sources

A test case is essentially a test program to be run through the RTL model of the processor to stress a particular area of functionality. These are generally written in a format similar to PA-RISC assembly language, with annotations to help specify initial cache and TLB contents. In addition, a control file can be attached to a test case to specify the behavior of the bus emulators. The emulators have useful default behavior, but if desired the control files can precisely control transaction timing.

A test case is compiled using a collection of tools that includes the PA-RISC assembler. The result of the compilation is a set of state initializations for the RTL model. These include the processor registers, caches, TLB, and memory. In addition, the bus emulators are initialized with the commands they will use during execution of the test case.

Previous PA-RISC microprocessor projects had built up a library of test cases and architectural verification programs (AVPs). Although we did run these, it was clear from the beginning that a large source of new cases would be required. The existing cases were very short, so their ability to provide even accidental coverage for a machine with a 56-entry IRB was questionable. Moreover, we needed cases that targeted the unique microarchitectural features of the PA 8000.

We developed a test case template expander to improve our productivity in generating the large number of cases required. An engineer could write a test template specifying a fundamental interaction, and the tool would expand this into a family of test cases. Some of the features of this tool included:

- The ability to sweep a parameter value. This was often used to vary the distance between two interacting instructions.
- The ability to fill in an unspecified parameter with a random value.
- An if construct, so that a choice between two alternatives could be conditional on parameters already chosen.
- Instruction groups, so that an instruction could be specified that had certain characteristics without specifying the exact instruction.

We also used the pseudorandom code generator and test coverage measurement techniques discussed below in the RTL simulation environment. To improve our coverage of multiprocessor functionality, we configured our bus emulators to generate random (but interacting) bus traffic.

Structural Verification

A block can be described by a single large RTL procedure or by a schematic that shows the interconnection of several smaller blocks, each of which is described by RTL. At the beginning of the project, RTL tends to be written at a high level because it can simulate faster and is easier to write, debug, and maintain when the design is evolving rapidly. Block

designers, however, have a need to create schematics for their blocks, so there is a risk that these will diverge from the RTL reference.

We considered three strategies to verify that the two representations of the block were equivalent. The first, formal verification, was not pursued because the required tools were not yet available from external vendors. The second strategy was to rely on the switch-level verification effort. This was unattractive because defects would be found too late in the design cycle, and the planned number of vectors to be run might not have provided enough coverage. The strategy selected was to retire the higher-level RTL description and replace it in the RTL model with the lower-level representation. The more timely and thorough verification that this provided compensated for some disadvantages, including slower simulation and more difficulty in making changes. We also used this strategy selectively, relying on switch-level simulation to cover regular blocks such as data paths with little risk.

Divide and Conquer

In any large design effort, one faces a choice of whether to verify components individually, together, or both. Verifying a component separately has several potential advantages. Simulation time is greatly reduced. Input buses can be directly controlled, so effort need not be expended manipulating the larger model to provide interesting stimulus. Finally, dependencies between subprojects are eliminated.

For separate verification to succeed, the interfaces to other components must be very well-specified and clearly documented. Investments must be made in a test jig to provide stimulus to the component and in checking software to verify its outputs. In addition, some portion of the verification must be repeated with all components integrated to guard against errors in the specifications or different interpretations of them.

The PA 8000's bus interface block was particularly well-suited to separate verification. The block had clean external interfaces but contained a lot of complexity, including the hardware to manage multiple pending memory accesses. A software checking tool was written to monitor the block's interfaces and verify its operation. Checking that a request on one bus ultimately results in a transaction on the other bus is a simple example of numerous checks performed by this tool. A very low defect rate demonstrated the success of the divide-and-conquer strategy for this block.

Most of our remaining verification effort was focused on the complete PA 8000. As a final check, a system-level RTL model was built that included several processors, the memory controller, the I/O adapter and other components. Although throughput was very low, basic interactions between the components were verified using this model.

Accelerated Simulation

The speed of the RTL simulator was adequate to provide quick feedback on changes and for basic regression testing, but we lacked confidence that on a design as complex as the PA 8000 it would be sufficient to deliver an adequate quality level. We saw a strong need for a simulation capability that was several orders of magnitude faster so that we could run enough test cases to ferret out more subtle defects. We considered two technologies to provide this: cycle-based simulation and in-circuit emulation.

Cycle-based simulation provides a much faster software simulation of the design. With an event-driven simulator such as our RTL simulator, a signal transition causes all blocks that the signal drives to be reexecuted, and any transitions on the outputs of these blocks are similarly propagated until all signals are stable. The overhead to process every signal transition, or event, is fairly high. Cycle-based simulators greatly improve performance by eliminating this overhead. The design is compiled into a long sequence of Boolean operations on signal values (AND, OR, etc.), and execution of this sequence simulates the operation of the logic in a clock cycle. The name cycle-based simulator comes from the fact that the signal state is only computed at the ends of clock cycles, with no attempt to simulate intermediate timing information. Our investigation revealed that speedups of 500 times were possible, so a simulation farm of 100 machines could have a throughput on the order of 25,000 Hz. The biggest drawback of this strategy was that cycle-based simulators were not yet available from external vendors.

With in-circuit emulation, the gates in a Boolean representation of the design are mapped onto a reconfigurable array of field-programmable gate arrays (FPGAs). The design is essentially built using FPGAs, and the emulated processor is connected to the processor socket in an actual system. The clock rate of the emulation system is on the order of 300,000 Hz, so very high test throughput is possible. It is even possible to boot the operating system. Unfortunately, there were many issues involved in using in-circuit emulation successfully:

- Custom printed circuit boards would have to be designed for the caches, large register files, and any other regular structures that consume too much emulation capacity. Changes in the design would be difficult to accommodate.
- A system was needed to exercise the emulated processor, including a memory controller and I/O devices. Firmware and hardware tinkering would have been needed to make this system functional at the slow clock rates required by the emulation system.

- Productivity was reduced by long compile times and limited observability of internal signals. Only one engineer at a time could use the system for debugging.
- The strategy was difficult to extend to multiprocessor testing. It was prohibitively expensive to emulate multiple processors. We planned to use a software emulator to create third-party bus traffic and verify the processor's responses, but there was a risk that the software's performance would throttle the emulation system's clock rate.
- The emulation system was a very large capital investment.

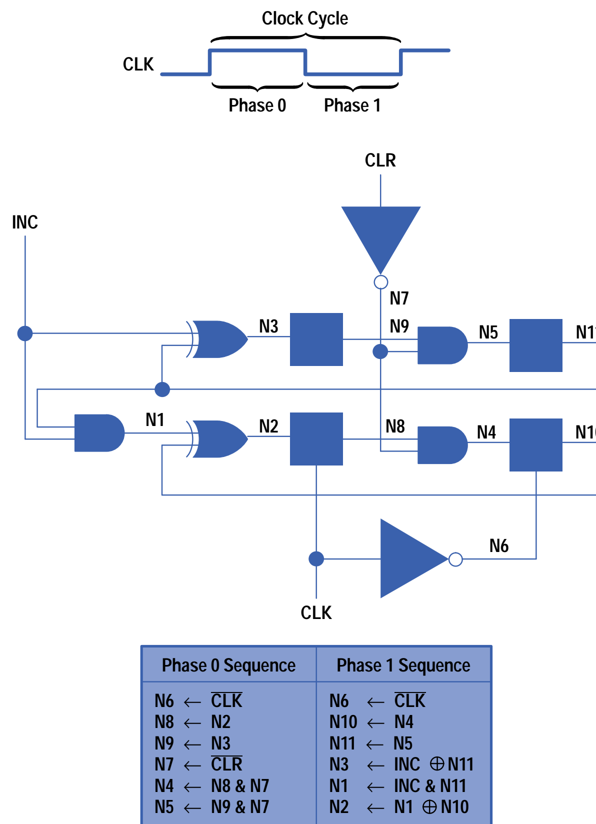
We were quite wary of in-circuit emulation since its use on a previous project had failed to make a significant contribution to functional verification. We were also willing to give up the performance advantage of in-circuit emulation to avoid tackling the ease-of-use issues. The decision to use cycle-based simulation would have been simple except that it meant that we would have to develop the simulator ourselves. R&D organizations in HP are challenged to focus on areas of core competency and look to external vendors to fulfill needs such as design tools that are common in the industry. We did select cycle-based simulation because we were confident that its lower risk and higher productivity would translate into a competitive advantage.

We were careful to reuse components wherever possible and to limit the scope of the project to providing the tool functionality required to verify the PA 8000. We did not attempt to create a simulation product useful to other groups within HP. This turned out to be a good decision because comparable tools have recently started to become available from external vendors.

Cycle-Based Simulation Compiler

The cycle-based simulation compiler operates only on simple gate-level primitives such as logic gates and latches, so higher-level RTL must first be synthesized into a gate-level equivalent. We had to develop our own translator for this because the RTL language used by our RTL simulator was defined before the industry standardization of such languages. Another simplification is that signal values are limited to 0 and 1, with no attempt to model an unknown (X) state.

Fig. 2 shows a simple example circuit, a two-bit counter, that we will use to illustrate the compilation process. The user must describe to the compiler information about the circuit's clocks. The clock cycle is broken down into two or more phases, with the state of the clocks fixed during each phase. This circuit has a clock cycle of two phases, and the clock (CLK) is low during the first phase and high during the second phase.



Phase 0 Sequence	Phase 1 Sequence
N6 ← $\overline{CLK}$	N6 ← $\overline{CLK}$
N8 ← N2	N10 ← N4
N9 ← N3	N11 ← N5
N7 ← $\overline{CLR}$	N3 ← INC ⊕ N11
N4 ← N8 & N7	N1 ← INC & N11
N5 ← N9 & N7	N2 ← N1 ⊕ N10

Fig. 2. Cycle-based simulation compilation example.

The compiler uses this information to determine which gates need to be evaluated during each phase. This is done in two steps. First, for each phase, the compiler propagates the clock values into the circuit. This uses simple rules of Boolean logic, such as the fact that the output of an AND gate with a zero input must be zero. The goal is to identify latches with a zero control, which are therefore provably opaque during that phase. Next, again for each phase, the compiler finds all gates that can be reached from a clock or other input through a path that does not contain an opaque latch.

Next, a sequence of Boolean operations is emitted corresponding to the gates in each phase. Because we used PA-RISC machines for simulation, the sequences were actually output in PA-RISC assembly language. The sequences totaled more than two million instructions for the PA 8000 design. The gates are ordered in sequence so that a gate is not emitted until its inputs have been computed. Cycles, or loops, in the circuit are handled by looping through the gates in the cycle until all circuit nodes are stable.

Numerous optimizations are done on the output assembly language sequences:

- Clock signals have known values during each phase, which can be propagated into the circuit. These constant values can simplify or eliminate some of the Boolean operations.
- The 32 PA-RISC registers are used to minimize loads and stores to memory. Boolean operation scheduling and victim register selection are employed to minimize the number of loads and stores.
- The compiler can determine which circuit nodes carry information from one phase to the next. The remaining nodes are temporaries whose values need not be flushed to memory after their final use within a phase.
- To eliminate NOT operations corresponding to inverting gates, the compiler can represent nodes in inverted form and perform Demorgan transformations of Boolean operations (e.g., NOT-AND is equivalent to OR-NOT).
- Aliasing of circuit nodes is done to eliminate code for simple buffers and inverters.

Any one of the Boolean operations in the output assembly language sequence operates on all 32 bits of the PA-RISC data path, as shown in Fig. 3. We make use of this parallelism to run 32 independent test cases in parallel. This is possible because the simulator always executes exactly the same sequence of assembly language instructions regardless of the test case (assuming the circuit being simulated is the same). This does not reduce the time to solution for a given test case, but it does increase the effective throughput of the simulator by 32 times. This was still very useful because our verification test suites are divided into a vast number of fairly short test cases.

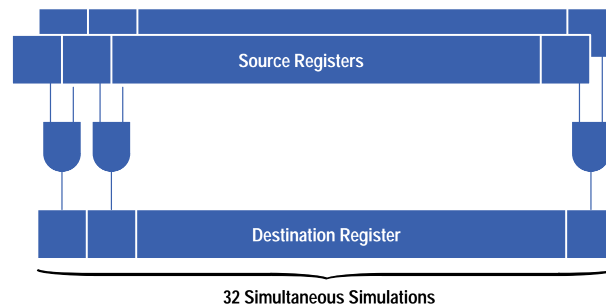


Fig. 3. *Multislot cycle-based simulation.*

The compiler allows the user to write C++ behavioral descriptions of blocks such as memories and register files that are not efficient to represent using gate-level primitives. The compiler automatically schedules the calls to this C++ code, and an API (application programming interface) gives the code access to the block's ports.

Pseudorandom Testing

We had learned from previous projects that the type of defects likely to escape the RTL simulation effort would involve subtle interactions among pending instructions and external bus events. With up to 56 instructions pending inside the processor and a highly concurrent system bus with multiprocessing support, it is not possible to count—much less fully test—all of the interactions that might occur. We believed that the value of handwritten test cases and test cases randomly expanded from templates was reaching diminishing returns, even with the low simulation throughput achievable with the RTL simulator.

We had also learned that pseudorandom code generators were a very effective means of finding these kinds of defects. Such a program generates a pseudorandom sequence of instructions that use pseudorandom memory addresses and pseudo-random data patterns. However, it is important that the program make pseudorandom selections in a manner that considers the microarchitecture of the processor and the kinds of interaction defects that are likely to occur.

Selecting memory addresses is a good example. Memory addresses are 64 bits wide. If they were selected truly randomly, reusing the same address within a test case would be an impossibly rare event. This would fail to stress important aspects of

the machine, such as the logic that detects that a load is dependent on a preceding store with the same address. There are hundreds of selections that a generator makes in which the microarchitecture must be carefully considered.

We chose to target the cycle-based simulation environment for a new pseudorandom code generator. Our pseudorandom code generator was carefully tuned for the microarchitecture of the PA 8000 and included support for the new PA-RISC 2.0 instruction set. Hundreds of event probabilities could be specified by a control file to provide engineering control over the types of cases being generated.

We also chose not to port the rich set of checking software from the RTL simulation environment to the cycle-based simulation environment because of the effort involved and risk that performance would be reduced. Generators such as our pseudorandom code generator predict the final register and memory state of the processor, and defects will generally manifest themselves as mismatches between the simulated and predicted final state. It is possible that an error in state will be overwritten before the end of a test case, but a defect won't be missed unless this happens in every test case that hits it, which is extremely unlikely statistically. Our experience with hardware prototype testing, in which internal signals are unavailable and all checking must be done through final state, also made us confident in this strategy.

Cycle-Based Simulation Environment

Fig. 4 shows the cycle-based simulation environment, which will be described by following the life cycle of a typical test case. The job controller controls the 32 independent simulations that are running in the data path positions, or slots, of the cycle-based simulation model. It starts and ends test cases in the 32 slots independently. It is controlled by a UNIX® shell, which is driven either by a script or interactively for debug activities.

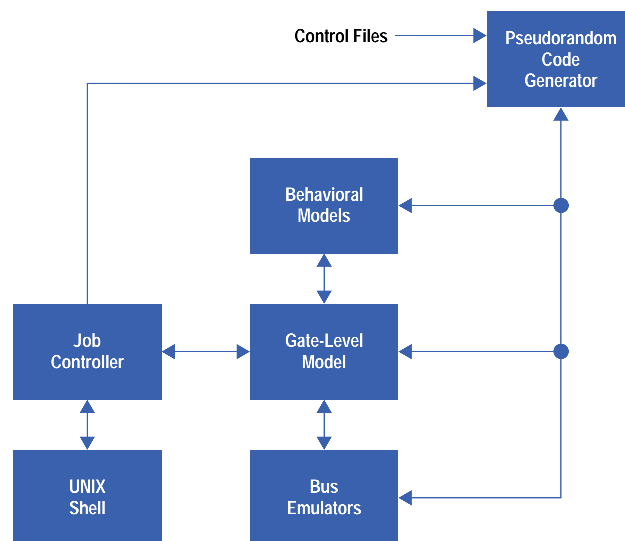


Fig. 4. *Block diagram of the cycle-based simulation environment.*

When a slot becomes available, the controller commands the pseudorandom code generator to generate a new test case, occasionally first reading a new control file. The test case is specified by the initial state of memory and the processor's registers, and the pseudorandom code generator specifies the initial state of the caches as well to prevent an initial flurry of misses.

The pseudorandom code generator downloads the initial state of the simulation into various components of the simulated model. These include the gate-level model, behavioral models representing caches, register files, and other regular structures, and emulators representing bus devices such as the memory system, I/O adapter, and third-party processors. The model is then stepped for numerous clock cycles until a breakpoint trigger fires to indicate the end of the test case. The pseudorandom code generator is then commanded to extract the relevant final state from the simulated model and compare it with the final state that it predicted to determine whether the test case passed.

We used a simulation farm of up to 100 desktop workstations and servers for cycle-based simulation. Jobs were dispatched to these machines under the control of HP Task Broker.[1] Each job ran several thousand test cases that were generated using a specific pseudorandom code generator control file.

Multiprocessor Testing

Multiprocessor testing was a key focus area. We wrote emulators for additional processors and the I/O adapter, which share the memory bus. It was only necessary to emulate the functionality required to initiate and respond to bus transactions, but the emulators were accurate enough that defects in the processor related to cache coherency would manifest themselves as mismatches in the final memory state.

We established linkages with our pseudorandom code generator so that the emulators would be more effective. When a test case started, the pseudorandom code generator downloaded control file information so that parameters such as transaction density and reply times could be easily varied. The pseudorandom code generator also downloaded the memory addresses used by the test case so that the emulators could initiate transactions that were likely to cause interactions.

Coverage Improvement

Improving the test coverage of our pseudorandom code generator was an ongoing activity. The pseudorandom code generator has hundreds of adjustable values, or ***knobs***, in its control file, which can be varied to focus the generated test cases. We found that the defect rate quickly fell off when all knobs were left at their default settings.

We used two tactics to create more effective control files. First, we handcrafted files to stress particular functional areas. Second, we generated files using pseudorandom techniques from templates, each template specifying a particular random distribution for each knob. We found with both strategies that it was important to monitor the quality of the files generated.

We did this in two ways. First, our pseudorandom code generator itself reported statistics on the test cases generated with a given control file. A good example is the frequency of traps. Traps cause a large-scale reset inside the processor, including flushing the instruction queues, so having too many traps in a case effectively shortens it and reduces its value. We made use of instrumentation like this to steer the generation of control files.

Feedback is often needed based on events occurring within the processor, which our pseudorandom code generator cannot generally predict. For example, an engineer might need to know how often the maximum number of cache misses are pending to be confident that a certain area of logic has been well-tested. Test case coverage analysis was accomplished by an add-on tool in the simulation environment. This tool included a basic language that allowed engineers to describe events of interest using Boolean equations and timing delays. The list of events could include those that were expected to occur regularly or even those that a designer never expected to occur. Both ends of this spectrum could provide useful information.

Once the events were defined, the add-on tool provided monitoring capabilities during the simulation. As test cases were run, the tool would generate output every time it detected a defined event. This output was then postprocessed and assembled into an event database. The event database could contain results of thousands of test case runs. Event activity reports were then generated from this event database. These reports included statistics such as frequency of events, duration of events, the average, maximum, and minimum distance between two occurrences of a event, and so on.

The event activity reports were then analyzed by engineers to identify weak spots in coverage and provide feedback to the generation of control files. This methodology provided one other benefit as well. For many functional defects, especially ones that were hard to hit, the conditions required to manifest the defect were coded and defined as an event. Then this add-on tool was used with a model that contained a fix for the defect to prove that the conditions required for the defect were being generated.

Switch-Level Simulation

In typical ASIC design methodologies, an RTL description is the source code for the design, and tools are used to synthesize transistor-level schematics and IC layout mechanically from the RTL. Verifying the equivalence of the synthesized design and the RTL is largely a formality to guard against occasional tool defects. In the full-custom methodology used on the PA 8000, however, designers handcraft transistor-level schematics to optimize clock rate, die area, and power dissipation. Therefore, we needed a methodology to prove equivalence of the handcrafted schematics and the RTL.

At the time the project was undertaken, formal verification tools to prove this equivalence were not available. Instead, we turned to an internally developed switch-level simulator. Although much slower than the RTL simulator, the switch-level simulator included essential features such as the ability to model bidirectional transistors, variable drive strength, and variable charge ratios. Thanks to this careful effort in switch-level verification on the PA 8000, not a single defect was found on silicon that was related to a difference between the transistor-level schematics and the RTL.

Verification was performed by proving that a block behaved the same when running a test case in the RTL simulator and in the switch-level simulator. First, a full-chip RTL simulation of a test case was done with the ports of a block monitored. These vectors were then turned into stimulus and assertions for a switch-level simulation of the block. Initializing the state of the block identically in the two environments was a challenge, especially since the hierarchies and signal names of the RTL and schematic representations can differ.

Initially, this strategy was used to turn on the switch-level simulator models of individual blocks on the chip. This helped to distribute the debug effort and quickly bring all blocks up to a reasonable quality level. Afterward, the focus shifted to full-chip switch-level simulator verification. In addition to collecting vectors at the ports of the chip, thousands of internal signals were monitored in the RTL simulation and transformed into assertions for the switch-level simulation. These were valuable for debugging and raising our confidence that there were no subtle behavioral differences between the two models.

The RTL simulation effort was a plentiful source of test cases, but they were targeted at functional defects rather than implementation errors, and the slower speed of the switch-level simulator allowed only a portion of them to be run. To improve coverage, the process shown in Fig. 5 was used at the block level. The RTL description for the block was converted into an equivalent gate-level model using tools developed for cycle-based simulation. Automated test generation tools, normally used later in the project for manufacturing, were then used to create test vectors for the gate-level model. If the switch-level simulation using these vectors failed, then the two representations were known to differ. While the automated test generation tools do not generate perfect test vectors, this process still proved to be a valuable source of additional coverage.

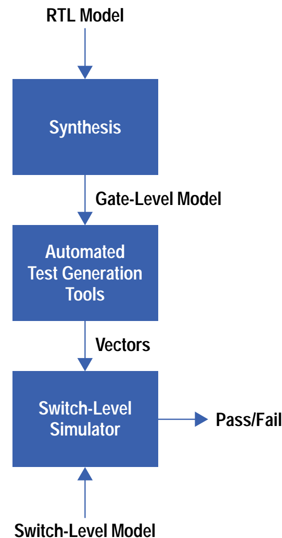


Fig. 5. *Process used at the block level to improve test coverage.*

The switch-level simulator also supports several different kinds of quality checks. These include dynamic decay checking to detect undriven nodes, drive fight checking to detect when multiple gates are driving the same node, and a race checking methodology. This was implemented by altering how the clock generator circuits were modeled to create overlap between the different clocks on the chip. Failures that arose from overlapped clocks pointed to paths requiring detailed SPICE simulations to prove that the race could not occur in the real circuits. Reset simulations were done from random initial states to ensure that the chip would power up properly. Finally, a switch-level simulator model was built from artwork netlists to prove that there were no mismatches between the artwork and the schematics that were missed by other tools.

Postsilicon Verification

Presilicon verification techniques are adequate to find nearly all of the defects in a design and bring the level of quality up to the point where the first prototypes are useful to software partners. However, defects can be found in postsilicon verification that eluded presilicon verification for many reasons.

First, test code can be run at hardware speeds, several orders of magnitude faster than even the fastest simulators. Errors in the simulation models or limitations in the simulators themselves can cause the behavior of the silicon to differ from that predicted by the simulators. Finally, most simulation is targeted at system components, such as the PA 8000 itself, rather than the entire system. Errors in the specifications for interfaces between components, or different interpretations of these specifications, can become apparent when all components are integrated and the system exercised.

Overlapped Test Coverage

Previous projects had established the value of running test code from as many sources as possible. Each test effort had its own focus and unique value, but each also had its own blind spots. This is even true for pseudorandom code generators. In the design of these very complex programs, many decisions are made that affect the character and style of the generated code. There can be code defects as well that cause coverage holes. The large overlap in coverage between efforts proved to be an invaluable safety net against the limitations and blind spots of individual tools.

The PA 8000 verification team focused its effort on pseudorandom code testing. Experience showed that this would be the primary source of subtle defects and would allow us to find most defects before our software partners. We ran several tools including our pseudorandom code generator and generators used in the development of the PA 7200 and PA 7300LC processors and the HP 9000 Model 725 workstation. Several tools were capable of generating true multiprocessing test cases that included data sharing between random sequences running on different processors. Data sharing with DMA processes was implemented as well.

We developed a common test environment for most of these random code generators. The HP-UX* operating system is not a suitable environment because its protection checks do not permit many of the processor resources to be easily manipulated. Our test environment allowed random testing of privileged operations and also included many features to improve repeatability and facilitate debugging. For example, it performed careful initialization before each test case so that, aided by logic analyzer traces, we could move a failing test case to the RTL simulator for easy debugging (in hardware, there is no access to internal signals). We established a plug-and-play API so that the investment in the environment could be leveraged across several generators.

In parallel with our pseudorandom testing, our software partners pursued their own testing efforts. While primarily targeted at their own software, this provided stress for the processor as well. The test efforts included the HP-UX and MPE/XL operating system kernels, I/O and network drivers, commands, libraries, and compilers. Performance testing also provided coverage of benchmarks and key applications. Finally, although it did not find any defects in the PA 8000, HP's Early Access Program made available preproduction units to customers and external application developers.

Ongoing Improvement

When defects were found, we used the process shown in Fig. 6 to learn as much as possible about why the defect was missed previously and how coverage could be improved to find additional related defects. After the root cause of the defect was determined, actions were taken in the areas of design, presilicon verification, and postsilicon verification.

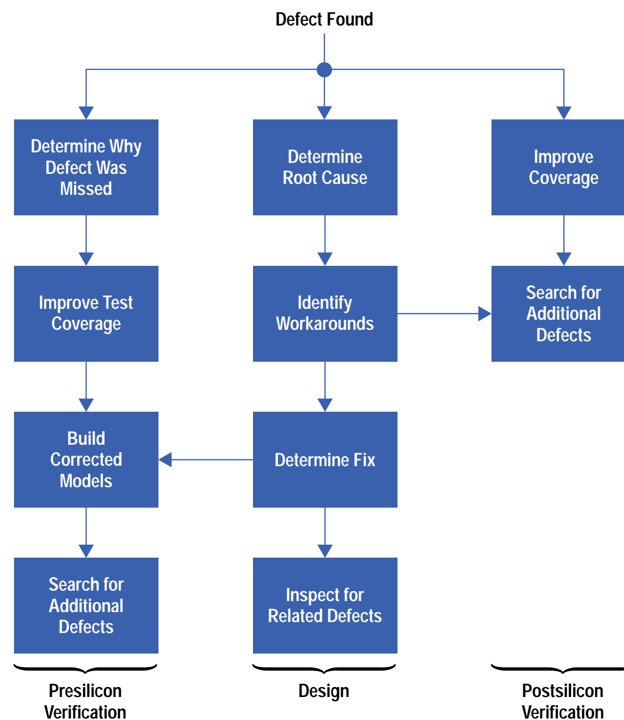


Fig. 6. *Postsilicon quality improvement process.*

The designers would identify workarounds for the defect and communicate these to our software partners, at the same time seeking their input to evaluate the urgency for a tape release to fix the defect. The design fix was also determined, and inspections were done to validate the fix and brainstorm for similar defects.

In the area of presilicon verification, reasons why the defect was missed would be assessed. This usually turned out to be a test case coverage problem or a blind spot in the checking software. Models would then be built with the design fix and other corrections. Test coverage would be enhanced in the area of the defect, and simulations were done to prove the fix and search for any related defects. Cycle-based simulation played a big role here by finding several introduced defects and incomplete fixes.

The postsilicon verification activities were similar. Coverage for the tool that found the defect would be enhanced, either by focusing it with control files or by tool improvements. Spurred by a healthy rivalry, engineers who owned other tools would frequently improve them to show that they could hit the defect as well. All of this contributed to finding related defects.

Performance Verification

At the time the PA 8000 was introduced in products, it was the world's fastest available microprocessor. Careful microarchitectural optimization and verification of the design against performance specifications were factors in achieving this leadership performance.

In a microarchitectural design as complex as the PA 8000, seemingly obscure definition decisions and deviations of the design from the specification can cause a significant loss of performance when system-level effects and a variety of workloads are considered. A good example is a design defect that was found and corrected in the PA 8000. When a cache miss occurred under certain circumstances, a dirty cache line being evicted from the cache would be written out on the system bus before the read request for the missing line was issued. Since the addresses of the two lines have similar low-order bits, both mapped to the same bank of main memory. The memory controller would begin processing the write as soon as it was visible on the bus, busying the memory bank and delaying the processing of the more critical read.

A detailed microarchitectural performance simulator was written early in the project to help guard against such issues. It was used to project performance and generate a statistical profile for a variety of benchmarks and applications. Workloads with surprising results or anomalous statistics were targeted for more detailed analysis, and through this process opportunities were identified to improve the microarchitecture. Particularly valuable feedback came from the compiler development team, who used the simulator to evaluate the performance of compiler prototypes. The concurrent development of tuned compilers with close cooperation between hardware and software teams was a key contributor to the PA 8000's performance leadership.

The microarchitectural performance simulator was written at a somewhat abstract level, so it could not provide feedback on whether the detailed design met the performance specifications. Comparing the performance of the RTL simulator against the microarchitectural simulator was the obvious way to address this, but the RTL simulator was far too slow. As a compromise, we performed this comparison on key performance kernels that were tractable enough for the RTL simulator. We also developed a path by which a workload could be run up to a critical point using the microarchitectural simulator, at which point the state of the memory, caches, and processor registers could be transferred into the RTL simulator for detailed simulation.

Performance verification continued in the postsilicon phase of the project. The PA 8000 incorporated several performance counters that could be configured to count numerous events. These were used to help identify workloads or segments of workloads needing closer analysis. The PA 8000's external pins and debug port provided sufficient information to determine when instructions were fetched, issued for execution, and retired. Isolation of specific performance issues was aided by a software tool called the depiper which presented a visual picture of instruction execution. Through these efforts, several performance-related hardware defects were identified and corrected before production.

Results

Achieving a defect-free design at first tape release is not a realistic expectation for a design as complex as the PA 8000. Nevertheless, we were extremely satisfied with the quality we achieved at first tape release. The first prototypes were capable of booting the operating system and running virtually any application correctly. In fact, only one defect was ever hit by an application, although a few defects were encountered in stress testing of system software.

Fig. 7 shows the sources of defects found and corrected after first tape release. Surprisingly, about a third of the total defects were found by continued use of the presilicon verification tools (mostly the cycle-based simulation environment) for a few months following tape release. This indicates that despite the outstanding performance of cycle-based simulation, the project would have benefited from even more throughput, or perhaps use of the tool earlier. A third of the defects were also found by one of the pseudorandom code generators running on hardware prototypes. Inspections were a significant source of defects. The remaining defects were split between turn-on work, performance analysis work, and partner software testing. Since very few defects were discovered by partners, we could generally communicate workarounds ahead of time and take other steps to minimize the impact.

Fig. 8 shows the impact of the defects found after first tape release on our software partners. A large majority were never seen outside the environment in which they were found and had no significant impact. About half of these involved functional areas, such as debugging hardware, that are not even visible to applications or system software. Most of the remaining defects had only a moderate impact. Examples are defects that were found by a partner at the expense of their testing resources, defects that required a workaround in system software, and defects that required certain performance-related features in the processor to be disabled. Only a handful of defects were severe enough to temporarily block or significantly disrupt a partner's development and testing efforts. All but one of these were early multiprocessing defects that slightly delayed bringing up the multiprocessing operating system.

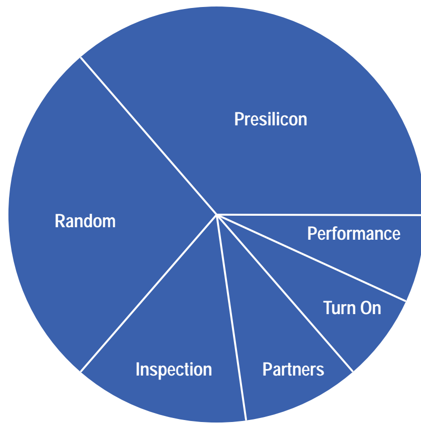


Fig. 7. *Sources of defects found and corrected after first tape release.*

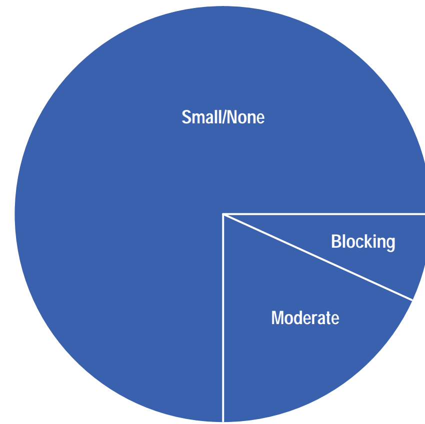


Fig. 8. *Partner impact.*

Cycle-Based Simulation Results

The cycle-based simulation effort made an essential contribution to the verification of the PA 8000. Fig. 9 shows the sources of defects that eluded our RTL simulation effort (which incorporated existing best practices). If we had not made the investment in cycle-based simulation, the number of defects that would have had to be found by postsilicon techniques would have been three times higher. It was much less expensive to fix the defects caught by cycle-based simulation as the design progressed than it would have been to fix them in later revisions.

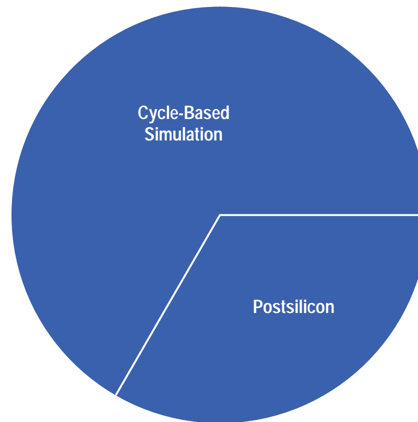


Fig. 9. *Defects escaping traditional presilicon verification.*

Also, because cycle-based simulation tended to find the most severe defects early, no masking defects were present, and the number of serious blocking defects that we had to manage after the first tape release was reduced by three to six times. If our software partners had been exposed to this level of severe defects, it is probable that the product's time to market would have been impacted.

Finally, cycle-based simulation provided a high-confidence regression test before each tape release. Several incomplete bug fixes and new defects that had been introduced in the design were found in time to be corrected before a tape release.

Conclusions

Continuous innovation in functional verification tools and processes is required to keep pace with the increasing microarchitectural complexity of today's CPUs. This paper has described the methodologies used to verify the PA 8000. These met our most important goal of improving the quality of the PA 8000 to the high level demanded by our customers. By finding defects early, they also helped us conserve our engineering resources and quickly deliver the industry-leading performance of the PA 8000.

Acknowledgments

The authors would like to thank all of the different teams who contributed to the successful functional verification of the PA 8000. Special thanks to all of the individuals in the Engineering Systems Laboratory in Ft. Collins, Colorado and the General Systems Laboratory in Roseville, California who were involved in the design and verification effort. Many thanks

also to our partners: the Cupertino Open Systems Laboratory for testing efforts and flexibility in working around defects, the Systems Performance Laboratory and Cupertino Language Laboratory for performance verification work, and the Integrated Circuit Business Division for supporting our simulation tools.

Reference

1. T.P. Graf, et al, "HP Task Broker: A Tool for Distributing Computational Tasks," ***Hewlett-Packard Journal***, Vol. 44, no. 4, August 1993, pp. 15-22.

HP-UX 9.* and 10.0 for HP 9000 Series 700 and 800 computers are X/Open Company UNIX 93 branded products.

UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company Limited.

X/Open is a registered trademark and the X device is a trademark of X/Open Company Limited in the UK and other countries.

▶ Go to Next Article

▶ Go to Journal Home Page