# Extending an Automated Testing Framework to Support Agile eXtreme Programming Development Concepts in an Embedded Real-Time Environment

Militina Gorobets Department of Electrical and Computer Engineering, University of Calgary Calgary, Canada

Michael Smith (Senior Member IEEE) Department of Electrical and Computer Engineering, University of Calgary Calgary, Canada Mike.Smith @ucalgary.ca Albert Tran Department of Electrical and Computer Engineering, University of Calgary Calgary, Canada

James Miller Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada jumm@ualberta.ca

Abstract—Automated testing frameworks provide a useful tool whether employing test-last or Agile eXtreme Programming test driven development test-first approaches. Additional consideration of how the testing framework's performance may generate false positive or negative test results becomes important when an embedded system's real-time performance must be taken into account. In this paper, we consider the advantages and disadvantages of two tools to improve the framework's performance by co-opting, and extending, on-chip hardware performance monitors to enhance Embedded-Agile development. We discuss providing faster full code test coverage analysis and lowoverhead live black-box and white-box testing using a FPGAbased test insertion co-processor. The co-processor development was used as a test bench for exploring FPGA-Agile development

Keywords—Embedded-Agile Development, FPGA-Agile Development, Automated testing frameworks, real time systems.)

#### I. INTRODUCTION

Grenning [1] was an early proponent of adapting the business (Enterprise) world's Agile processes [2] for embedded system development. It was reasoned that Agile's barely sufficient processes would be advantageous for embedded engineers who recognized the need for reliable products and had extensive domain knowledge, but were not eager to implement process improvement edicts from above.

The eXtreme Programming Inspired (XPI) embedded system lifecycle [3] has multiple test driven development (TDD) stages to support a hardware-software co-design process for embedded devices requiring an interaction between real-time digital signal processing (DSP) algorithms and external hardware. Early XPI stages use TDD for identification of the customer's (research team) requirements. Existing customer tests and new unit tests are used to support initial development using research-oriented development in a high level language (e.g. MATLAB). Later XPI stages extend the tests again during C++ simulation and movement onto a system-under-test (SUT) to handle embedded peripheral inter-action.

Enterprise, or business-oriented, Agile's TDD processes require refocusing for the specialized Embedded Agile environment. For example, the critical need to meet expected strict embedded system real-time requirements would suggest an early evaluation of non-functional code aspects (time and memory requirements) that might be considered inappropriate by some advocates who adhere more strictly to the standard Agile concepts. Thus the standard concept of "Refactoring for code maintainability" from Enterprise-Agile will need to become subservient to Embedded-Agile approaches of "Refactoring code for speed or power consumption" to take advantage of specific processor architectural characteristics [3].

In this paper, we consider two tools to providing additional support for the automated testing framework that *a priori* is needed for Embedded-Agile development, and would prove useful in a more conventional test-last development environment. We first discuss the concept of co-opting existing on-chip hardware performance monitors to provide low-overhead test coverage analysis tool, *E-RECOVER*. This tool is intended to reduce the introduction of potential false positive or negative results as a result of impacting system performance during testing. The theoretical and actual performance impact of this tool is discussed. The second part of the paper discusses the concepts of a prototype FPGA Agile test support, ATS, co-

processor designed to provide low-overhead, live, black-box and white-box testing.

#### II. TOOL 1 : LOW OVERHEAD TEST COVERAGE ANALYSIS

Developers and testers use code coverage tools to gauge which statements have (or have not) been executed during the testing process. Statements that have not been executed (i.e. untested code) have a potential for leaving latent defects in the code. Incorporating coverage measurements into the software reliability analysis has shown improvements in system reliability estimations, providing the developer with insight on where to focus testing activities [4]. Collecting and analyzing coverage report data throughout the project may shed light on the efficiency of the performed code coverage testing as product evolves. This may further help developers effectively write tests to target areas of code that would potentially be more error-prone.

Code coverage analysis is as important on embedded systems as on desktop systems; however, the embedded developer is operating under more stringent memory constraints. Realtime performance issues means that code coverage analysis must not distort the performance characteristics of the system under test (SUT) [5]. To solve such conflicts, embedded developers do have advantages over their desktop counterparts; in particular, easier access to a wider and more sophisticated range of hardware performance monitoring capabilities.

#### A) Software based instrumentation

Traditional methods of providing code coverage rely on a software approach to code instrumentation. This process inserts software probes into source- or object-code, which record whether particular segments of a program are executed [6]. As the code is executed, the inserted code probes cause software interrupts (SWI) to be activated whenever a given branch node is reached. As is shown schematically in Figure 1, the SWI stops the current code execution and invokes a software interrupt service routine (ISR) that identifies the code location and records the code coverage information. Code instrumentation can deliver high coverage rates but significantly changes the execution characteristics of a program from its normal flow, as well as impacting the code size. Tikir and Hollingsworth [6] indicate that the PURECOV tool, based on this approach, slows down the code execution of the SPEC95 benchmark suite from 1.8 (Tomcatv) to 19.8 times (Perl).

To reduce such overhead, they suggest incrementally inserting the necessary instrumentation code the first time a function is called in order to decrease the instrumentation costs and using more sophisticated analysis of possible execution paths. They indicate that the cost of determining when to invocate dynamic deletion of "used" instrumentation code to improve analysis of long running programs may negate any time savings. On an embedded system, there is a further problem. Unnecessary instrumentation has the potential of forcing the developer to load the program under test into slow off-processor memory incurring a 2 to 6 times performance penalty.



Fig. 1 Source code instrumentation causes code bloat, as well as high overheads introduced when jumping into/out of software probe code (based on [17]).



Fig. 2 Block diagram of a hardware assisted code coverage tool using the branch vector TRACE hardware

# B) Hardware Assisted Code Coverage

Direct use of the internal hardware present on some processors to assist the code coverage analysis can overcome software instrumentation issues. VTune and OProfile tools periodically sample the processor's program counter (PC) while the program is running, with addresses then matched to the original source files. However, Anderson et al. [7] found that the sparseness of the PC samples provided only a fraction of full software instrumentation information, necessitating repeated executions and negating any gains in elapsed developer time

To overcome this sparse coverage issue we can make use of the special hardware performance monitors (HPM) designed into modern processors We have investigated making use of branch vector management provided by hardware TRACE Units, such as those of the Analog Devices family of Blackfin processors. Here, a hardware stack automatically captures program code flow vectoring information. If this information is cross referenced to the original source code during post processing, there is no need to record sparse general PC information.

The Blackfin family is designed for real-time DSP operations and includes features that can be co-opted to further reduce monitoring overhead. The "Loop Compression" option prevents the logging of recent duplicate entries onto the hardware stack, reducing the amount of branch vector data logging that is generated for (i) loops calling leaf-functions, or (ii) the logging generated for the double nested loops common in algorithms used in embedded systems for DSP-oriented applications. In addition, (iii) the TRACE unit also does not log any program counter discontinuities associated with zero-overhead (hardware) loops which an optimizing compiler will introduce to replace standard software loop operations. On the Blackfin, the combination of the two available zero- overhead hardware loop units with 2-level loop compression offers the potential for considerable performance gains.

Fig. 2 shows how this TRACE unit can be used within a tool that provides the same detailed coverage information as found with software instrumentation but with minimal performance degradation. However, for the tool to be useful within a test driven development (TDD) process, reports must be generated from the extensive full coverage details in an expedient manner. Again, extending the ideas found to improve the speed of software instrumentation, we propose adding control features (1) to identify, in real time, the code (functions) currently under test; (2) to generate code coverage reports for only those functions that are currently in the scope of interest of the developer; and (3) to lower the performance impact of the tool by automatically activating the HCM for only those functions currently of interest

#### III. HARDWARE-BASED CODE COVERAGE IMPLEMENTATION

The new Rapid Efficient Code COVERage tool, E-RECCOVER, was added to the Source-Forge UnitTest++ testing framework initially targeting the Blackfin processor for its distinct feature of a hardware TRACE unit which stores the program instruction branched to and the memory address of the branch instruction which caused the discontinuity. Figure 2 provides the block diagram of a hardware assisted code coverage tool using the branch vector TRACE hardware. The original subroutines and functions are left unmodified, except for the system main() where the TRACE buffer hardware is configured. When the TRACE hardware buffer gets full, the processor will generate a TRACE exception handled by an exception service routine (ESR). The contents of the TRACE buffer will be moved to a software holding array. Otherwise, the ESR will pass the unhandled exception to the generic system exception handler. Because the custom ESR handler is written in optimized Blackfin assembly, it occupies only a few hundred bytes of code space, and can handle transferring the TRACE buffer contents very efficiently.

Several components are required to generate meaningful output from the list of branch vectors. Because the physical code memory location may be moved around on each compilation, a code mapping script must be run every time the project is compiled. This source mapping table (SMT) was originally generated via an exposed VisualDSP++ IDE Automation Interface based on Microsoft's Component Object Model (COM) technology, accessed through a COM connection between the EmbeddedUnit GUI and VDSP. However, this approach took roughly 60 seconds for every 10k lines of code (LOC), too slow for repeatedly running the tests required for a TDD approach. This bottleneck was overcome by using intermediate debug output files generated during code compilation which were passed through an elfdump tool to create two intermediate DWARF-format files source code/memory address mapping and the program symbol table, which EmbeddedUnit used to generate the SMT in less than one second for 10 KLOC.

## A. Expected performance analysis

As with other instrumentation approaches, the performance impact of the proposed hardware assisted test coverage tool can be expected to be very algorithm dependent. However, an estimate of the expected average performance impact of the tool can be determined by making a number of simple assumptions. (i) L lines of high-level code are executed, each requiring N processor clock cycles taking a time T = N L cycles to execute; (ii) a fraction k of the code is instrumented for hardware assisted code coverage; (iii) a branch occurs every B lines of the high level language; (iv) the TRACE unit activates an exception service routine only when the branch vector hardware stack of H entries is full; (v) the disruption of a processor pipeline of depth P due to entry into and exit from the exception service routine results in a 2P cycle overhead; (vi) the exception routine to read the branch vector pairs (source and destination) from the TRACE unit and store them to local memory for later recovery requires 2 \* H \* M cycles where M is number of cycles to perform a memory access; and (vii) R registers must be saved and recovered during the exception service routine for a performance cost of 2 \* R \* M cycles. This leads to a "conservative" estimate of the performance impact of using hardware assisted code coverage is

# PerformanceImpact = InstrumentationOverhead/T= kL(2P + 2(H + R)M)/(BHNL)

A worse case practical estimate of the overhead can be obtained by assuming 100% code coverage requirement (k = 1), with each high-language statement requiring 5 to 30 cycles to execute (average N = 8) and branches occurring on average every B = 5 high-level statements. For the ADSP-BF5XX family, we would expect an instruction pipeline of depth P =10, and R = 7 registers saved. The hardware TRACE stack of depth H = 16 entries requires an average of 2 \* M = 8 cycles for each branch vector pair stored to internal memory for later analysis. Thus the expected worse-case theoretical performance impact of the proposed fast full code coverage tool is approximately 30%. In practice, we would expect a significantly lower performance impact as we have not taken the HCM loop compression capabilities into account during the analysis, leading to an overhead of (2P + 2 (H + R) M) / H or approximately 10 - 12 cycles for each branch recorded, lower overhead than what is achievable for software instrumentation approaches.

# B. Actual performance impact

In a test, the execution time (cycle count) of an uninstrumented function DemoFunction() was determined. This consisted of a quadruple nested loop with the inner loop conditionally calling a leaf function WriteAudioPort() that writes to a device (audio peripheral) along an external processor bus. This function models characteristics found in many DSP algorithms. Given the large number of branches occurring in this over simplified function, the overall performance impact of code coverage monitoring will be exaggerated; but the true performance impact / branch vector generated ratio becomes straight forward to determine.

Our EmbeddedUnit testing framework allows determination of the non-instrumented time for the *DemoFunction()*. Extensions to the framework allowed activating and deactivating the code coverage hardware via CODE\_TRACE() macros, with extended features to permit the evaluation using standard TRACE hardware features and then activating the special loop compression activated features.

The *DemoFunction()* had an inner loop causing three jumps (conditional evaluation, jump into, and out of *WriteAudioPort())*. With trip counts for the 4 loops of 14, 16, 18, and 20, the expected number of branch vectors generated was around 242,000. However, the results from running the program showed only 167,000 branches occurring. Detailed inspection of the machine code showed that the optimizing compiler was able to reduce the C++ inner loop conditional statements into single conditionally-executed machine level instructions (if-condition-do-assignment). These statements reduce to no-operation instructions if the condition is not met; avoiding any pipeline disruptions and associated performance degradation – a typical example of the algorithm and processor dependence associated with program flow.

With a baseline non-instrumented code execution of 4.95 million cycles, the overhead associated with recording each branch is about 17 cycles per branch vector, compared to the 10 - 12 cycles per vector predicted in the previous section. Activating the loop compression option caused a function call to *WriteAudioPort()* in the inner (zero-overhead hardware-controlled) loops to be logged only once for every iteration of the outer loop. With the 2 Blackfin hardware loop registers handling two inner loops and the two outer loops handled by software, the branch count was calculated from the assembly code generated by the compiler as 14 \* 16 + 18 \* 20 = 808; which was confirmed by the EmbeddedUnit measurements.

The loop compression option reduces the number of branch vectors recorded, but not the overhead for documenting an individual branch vector; a cost that remains constant. With this example involving essentially continual branching, the performance impact for full code coverage is 56% when using the un-optimized E-RECCOVER tool settings. This reduces to an overhead of 0.25% after activating the TRACE unit's loop compression option to cause the removal of over 99% of the logging information recognized as being redundant. This performance hit is an order of magnitude smaller than reported by Shye et al. [14] sampling the Itanium BVB every 10 million

clock cycles while retaining the 100% code coverage available from software instrumentation.

# IV. TOOL 2: FPGA AGILE TEST SUPPORT, ATS, CO-PROCESSOR

Combining in-house and third party threads may be needed in an embedded application to meet changing market requirements. In an Enterprise environment, software instrumentation of the threads is one approach used to determine whether the correct locks (mutexes) are held when accessing shared memory. However, positives and negatives into data race analysis and may require more memory than readily available in a real-time embedded development environment.

Many modern embedded processors and micro-controllers provide low-level developer debugging support through breakpoints activated by on-chip hardware bus-watch units that monitor (snoop) bus activity. Huang et al. [9] proposed co-opting these units to produce a hardware-assisted, lowoverhead and low-memory requirement E-RACE tool capable of inserting data-race tests upon recognizing accesses to shared memory. They indicated that such hardware-supported data-race analysis approaches would be particularly useful when the original (legacy or third party) thread source code was not readily available for implementing software instrumented shared memory access detection. The authors further suggested that such a low-overhead tool would also prove useful when developing new threads via an Embedded-Agile TDD approach.

Huang et al. [9] evaluated the E-RACE tool concept using the Analog Devices family of Blackfin (ADSP-BF5XX) processors. The BF5XX debug instruction-bus watch unit could be programmed to throw exception events on recognizing specific instruction-bus activity. The associated exception service routine (eXSR) was then co-opted to insert a low overhead hardware-supported test to determine whether a monitored thread unblocked in time to meet desired hardware or software hard time constraints. However the BF5XX data-bus watch unit threw emulation events upon recognizing shared data memory accesses. Portions of the associated emulation service routine (ESR) run on the development workstation making this approach 10,000+ times less time efficient in inserting tests than via an eXSR. This voided many of the potential advantages of using a hardware-assisted tool to support an Acceptance Test Driven Development (A-TDD) Embedded-Agile approach to developing threads without hidden shared memory data races. Smith et al. [10] suggested that problems associated with the different capabilities of the existing datawatch and instruction-watch units could be overcome by the development of an external FPGA-based ATS coprocessor. This approach, shown schematically in Fig. 3, is suitable for any embedded system, and would use an interrupt to inform the SUT that the co-processor had recognized the bus activity associated with a particular thread accessing shared memory. There would be no overhead when the ATS co-processor was simply monitoring bus activity, unlike a software instrumentation approach.



Fig. 3. Schematic of the operation of an FPGA-based ATS co-processor. Tests are needed for new and legacy code to ensure that the shared memory accesses only occur when threads own the proper MUTEX values. Without the code re-compilation needed to support software instrumentation, the ATS co-processor's watch units (WUs) perform low-overhead snoop operations on the processor's data buses to identify shared memory access and enabling the insertion of tests to recognize potential race conditions.

# A. Identification of ATS Requirements

Development of the ATS coprocessor was undertaken with two results in mind. The first was to meet the identified need to extend an automated testing framework to better support. Embedded-Agile development. The second was to provide a test bench to identify potential Agile development approaches to develop a coprocessor. The assumption here was that the final real time performance of a system using software-based DSP algorithms was insufficient for customer needs. This would necessitate that existing parts of the algorithm be moved over to an FPGA-based DSP coprocessor to enhance performance as shown schematically in Fig. 4.

We classify this as an FPGA-Agile development approach as tests for the existing software already exist, and are therefore available to monitor the development, and final performance of the FPGA DSP co-processor. To explore potential generalizable approaches to Embedded-Agile XP development of FPGA products we deliberately started development of the Test Support (ATS) co-processor using the Agile technique of using customer wish list items (WLx) as an approach to expressing requirements that identify the desired hardware-based ATS FPGA co-processor characteristics. We then take a customer identified high priority item from the wish-list and then attempt to meet the spirit of a number of Agile extreme programming (XP) development concepts. Each development stage (Sprint) should start with the definition of tests that must be met to satisfy the identified wish list requirement. Just enough code should then be written to satisfy those tests. The Sprint end result should be a minimal viable product (MVP) with capabilities that satisfy the tests derived from the requirements. The minimum viable product concept delineates a difference between Agile XP and standard Embedded or FPGA development approaches. The MVP is not a prototype whose development is considered wasteful in both time and other resources in an Agile concept. The MVP is a fullyfunctional unit of the final product written to a high standard

with the intention that it can be retained and released onto the market without the cost of modification.

--- WL1: The ATS will contain watch unit blocks capable of recognizing specific address and data bus activities/

— WL2: Each watch unit would contain a (watchdog) timer which, paired with an instruction bus, could watch activities to evaluate hard-time constraints of a particular thread.

— WL3: A pair of bus-watch unit registers would provide the capability to efficiently watch a range of memory locations shared across new and legacy threads.

— WL4: On identifying a watch condition match, the ATS issues an interrupt to the processor to cause a test insertion.

— WL5: A separate communication channel exists between the processor and the ATS co-processor to allow indication of which watch units. The ATS must be temporarily deactivated to allow the interrupted watched condition to complete.

— WL6: A specified number of particular bus activities must occur before the ATS again communicates the recognition of the watched activity

## B. Planned Co-design Embedded-Agile Migration Stages

We propose that there be several migration stages (MS) to provide the maximum opportunities to meet the Agile XP requirement of maintaining a minimum viable product (MVP) during Co-design Embedded-Agile development stages.

**MS1:** Previous tests have been used to validate the existing DSP blocks in Fig. 4, and indicate that software components DSP4 and DSP5 need to be parallelized, migrated to a coprocessor, to meet the embedded system's real-time requirements. **MS2:** The functionality of each proposed FPGA component block, M-DSPx, required by a DSPx software component is mocked in software and validated using the existing tests. Mocking will suggest how each DSPx block might be configured inside the FPGA co-processor. If the mocked code is reconfigured as RISC-like operations (ROPs) then the number of parallel ROPs possible within a "reasonable" FPGA architecture provides an early proxy to whether the required speed improvement will be achieved following migration.

**MS3/MS4:** The functionality of the SPI-supported test encoder and interpreter required for the specific application is identified with the mocking validated with the existing tests. In MS3 the encoder and Master SPI interface together with the interpreter and SPI slave interface are all mocked in software. Then, MS4, the mocked encoder and decoder are interfaced with the processor's actual SPI hardware working in echo mode, i.e. the SPI Master-out slave-in (MOSI) line tied back to the Master-in slave-out (MISO) line. These combined software and hardware mocking stages are validated using the existing tests.**MS5:** By following an Agile TDD process,

the coding of the mocked SPI unit and interpreter will have been preceded by the development of additional unit tests. These tests can be used to validate the result of migrating. the mocked SPI slave unit, and then the mocked test interpreter, onto the actual FPGA **MS6**: The DSP algorithm blocks identified in MS1 are then migrated. Black box unit and acceptance testing will be possible through the SPI slave unit and test interpreter now functioning on the FPGA



an FPGA co-processor. The Hardware-Software refactoring process is supported at unit- and acceptance-test levels through tests developed during earlier XPI stages. The DSP4 block has already been migrated to the FPGA as M-DSP4, while DSP5 has been planned for migration as PM-DSP5. The SPI-supported test encoder and interpreter allow separation of the validation of the co-processor functionality from the processor dependent stage of integrating the co-processor onto the embedded system busses.

**MS7:** With the functionality of the FPGA co-processor established through the SPImocked address, data and control bus operations, the co-processor can then be connected to the actual embedded system buses. Tests for this stage can still conveniently involve use of the SPI test interpreter when appropriate. Details of this final processor dependent stage is beyond the scope of the current article but has been outlined for the Analog Devices BF5XX family.

# V. RESULTS FROM ATS COPROCESSOR DEVELOPMENT

Agile processes have been widely used in software development, and recently migrated to embedded systems. We explored the adaptation and adoption of Agile processes to address the aspect of reliable development of FPGA-based coprocessors. We explored an initial Hardware-Software Co-Design Embedded-Agile XP process based on the assumption that an existing C++ program must be migrated onto an FPGA-based co-processor to improve real time performance, or onto an FPGA-based soft processor to meet cost or power.

These processes were demonstrated through the development of an Agile Test Support (ATS) co-processor. A Test-Encoder-Interpreter and a Result-Encoder-Interpreter interface was used to achieve an automated testing framework. Although the ATS development was successful using this approach, our experiments indicated that it was inefficient to attempt to directly validate the existing C++ DSP algorithms ported on a one-to-one basis to a FPGA coprocessor from the CPU side. A far better approach would be to migrate the existing tests into a FPGA Agile testing framework. The original embedded system tests can then be used as acceptance tests for the integrated co-processor.

#### VI. CONCLUSION

In this paper, we having considered two tools to improve a testing framework's support for Embedded-Agile development by co-opting, and extending, on-chip hardware performance monitors to support faster full code test coverage analysis and low-overhead live black-box and white-box testing using a FPGA-based test insertion co-processor.

#### ACKNOWLEDGMENT

Financial support was provided by Analog Devices and Natural Sciences and Engineering Council of Canada (NSERC) through a Collaborative Research and Development grant (CRD 299423-03), NSERC Discovery Grants (RGPIN/03652-2007 and 2016) and the University of Calgary.

#### REFERENCES

- J. Greening, "Extreme programming and embedded system development, Embedded Systems Conference, 2002.
- [2] K. Beck. "Test Driven Development: By Example". Addison-Wesley Professional, 2002.
- [3] M. Smith, J. Miller, L. Huang, and A. Tran., "A more agile approach to embedded system development", IEEE Software, Vol. 26#3, 50–57, 2009.
- [4] M. Chen, M. Lyu and W. Wong, "Effect of Code Coverage on Software Reliability Measurement", IEEE Transactions on Reliability, Vol. 50, 165 – 170, 2001.
- [5] C-Y. Huang, C-H. Chiu, C-H. Lin and H-W. Tzeng, "Code Coverage Measurement for Android Dynamic Analysis Tools", IEEE International Conference on Mobile Services (MS2015), 209 – 216, 2015.
- [6] N. Kumar, B. R. Childers et al.,"Low overhead program monitoring and profiling." Proc. of the 6th ACM SIGPLAN-SIGSOFT workshop on Program analysis for software tools and engineering. 28-34, Lisbon, Portugal, ACM. 2005.
- [7] M. Tikir and J. Hollingsworth, "Efficient Instrumentation for Code Coverage Testing", Proceedings of International Symposium on Software Testing and Analysis, 86-96, 2002,
- [8] J. Anderson, L. Berc, J. Dean, S. Ghemawat et al., "Continuous Profiling: Where have all the cycles gone?, ACM Trans. on Computer Systems, 15:4, 357 – 390, 1997.
- [9] L. Huang, M Smith, A. Tran, J. Miller, "E-RACE, A Hardware-Assisted Approach to Lockset-Based Data Race Detection for Embedded Products", IEEE 19th International Symposium on Software Reliability Engineering, ISSRE Seattle, USA, 2008.
- [10] M. Smith, D. Deng, S. Islam, and J. Miller. "Enhancing Hardware Assisted Test Insertion Capabilities on Embedded Processors using an FPGA-based Agile Test Support Co-processor" J. Sig. Processing Systems, 1–14, 2013.