A Survey of Recent Developments in Testability, Safety and Security of RISC-V Processors

Original
A Survey of Recent Developments in Testability, Safety and Security of RISC-V Processors / Anders, Jens; Andreu, Pablo; Becker, Bernd; Becker, Steffen; Cantoro, Riccardo; Deligiannis, Nikolaos; Elhamawy, Nourhan; Faller, Tobias; Hernandez, Carles; Mentens, Nele; Namazi Rizi, Mahnaz; Polian, Ilia; Sajadi, Abolfazl; Sauer, Mathias; Schwachhofer, Denis; SONZA REORDA, Matteo; Stefanov, Todor; Tuzov, Ilya; Wagner, Stefan; Zidaric, Nuša. - (2023), pp. 1-10. (Intervento presentato al convegno 2023 IEEE European Test Symposium (ETS) tenutosi a Venice (Italy) nel 22-26 May 2023) [10.1109/ETS56758.2023.10174099].

Availability:
This version is available at: 11583/2978944 since: 2023-05-30T20:36:24Z

Publisher:
IEEE

Published
DOI:10.1109/ETS56758.2023.10174099

Terms of use:
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)
A Survey of Recent Developments in Testability, Safety and Security of RISC-V Processors

Jens Anders∗, Pablo Andreu‡, Bernd Becker∗, Steffen Becker†, Riccardo Cantoro‡, NikoIas I. Deligiannis∗, Nourhan Elhamawy‡, Tobias Faller∗, Carles Hernandez‡, Nele Mentens§, Mahnaz Namazi Rizi‡, Ilia Polian‡, Abolfazl Sajadi‡, Mathias Sauer††, Denis Schwachhofer†, Matteo Sonza Reorda‡, Ilya Tuzov‡, Stefan Wagner‡, Nuša Zidarič‡

Abstract—With the continued success of the open RISC-V architecture, practical deployment of RISC-V processors necessitates an in-depth consideration of their testability, security and safety aspects. This survey provides an overview of recent developments in this quickly-evolving field. We start with discussing the application of state-of-the-art functional and system-level test solutions to RISC-V processors. Then, we discuss the use of RISC-V processors for safety-related applications; to this end, we outline ISA extensions, RISC-V allows custom extensions to be designed, consisting of alternative base parts with added optional space variants and “mandates” a convenient, modular processor design, as shown in Table I. As a result, the PiCoRV32 core implements the base integer ISA of 32-bits (RV32I) and the base integer ISA of 32-bits, using 16 registers (RV32E) and supports the integer multiplication and division (RV32M) and the base integer ISA of 32-bits, using 16 registers (RV32I).

Table I: Selected RISC-V cores: academic and open source

In Table I we list selected RISC-V cores, with focus on academic and open source designs. In the first two columns we list the core and supported base ISA and possible extensions. Then we list number of pipeline stages and some of the features. The latter is of importance for System-on-Chips (SoCs), ISA extensions and coprocessor designs. Finally we provide technology with integrated RoCC, Tilelink Crossbar, RoCC, and available interfaces; the latter is of interest for System-on-Chips (SoCs), ISA extensions and coprocessor designs. Finally we provide technology with

In Tables I we list selected RISC-V cores, with focus on academic and open source designs. In the first two columns we list the core and supported base ISA and possible extensions. Then we list number of pipeline stages and some of the implemented core features, e.g., branch prediction or out-of-order execution, and available interfaces; the latter is of importance for System-on-Chips (SoCs), ISA extensions and coprocessor designs. Finally we provide technology with √ denoting fabricated ASIC chips.

Recently, several surveys have emerged to cover the current state-of-the-art in RISC-V developments and features. For a comparative survey of selected open-source cores listing
benchmark and maximum frequency, core area, and power consumption for both FPGA and ASIC technologies see [2]. The survey in [19] is focusing on open-source hardware integration of machine learning applications (ML). Authors of [1] classified the RISC-V software ecosystem into four categories: application fields (space systems, IIoT, AI-based heterogeneous systems), RISC-V implementations (soft-cores, System-on-Chip (SoC), emulators and simulators), software architecture (bare-metal, OS, development tools), and deployment features (security, reliability, low-power). The survey in [20] presents ISA extensions designed by the RISC-V community and the progress of official RISC-V ISA extension specifications. Abella et al. [21] provided an overview of why RISC-V is successful and some contributions on RISC-V and security against side-channel attacks, a dual core LockStep system with diverse redundancy, and System-Level Test.

In this paper, we survey the recent developments around RISC-V with focus on the following aspects: test, safety, and security. More specifically, in the area of functional test in Section II-A we focus on test generation methods for Software-Based Self-Test, Burn-In and System-Level Test and the first steps in the development of a model of non-functional properties of hardware. Then, in Section III we describe the potential and existing solutions of RISC-V for safety-related applications, with focus on functional correctness and timing verification. Finally, in Section IV we provide an overview of RISC-V security from the cryptographic, hardware and ISA perspective, focusing on the granularity of added hardware and added custom ISA extensions in the context of hardware/software co-design, and conclude the section with RISC-V system security.

II. RISC-V AND FUNCTIONAL TEST

In this section we present how RISC-V is able to support research on generating functional tests. In Section II-A we start with the challenges of generating Software-Based Self-Test (SBST) and Burn-In (BI) programs and how the RISC-V specification and ecosystem can help there.

Afterwards, we move from SBST to System-Level Test (SLT) in Section II-B where we show how the open-source nature of RISC-V and the tools it spawned support research in automated test program generation. Here we explain why it is difficult to use traditional testing methods for SLT and why a greybox approach to generating tests becomes necessary. Then, we go into detail on how an open-source SoC generation framework supports the development of a high-level model of non-functional properties of hardware.

A. RISC-V and Functional Test Development

Software-Based Self-Test (SBST) [22] allows at-speed, native in-field testing of processors with respect to permanent faults by running software programs on the processor core, requiring no Design for Testability (DfT) infrastructure and averts over-testing of manufactured chips. It is well known that the manual development of SBST programs (also known as Self Test Libraries, or STLs) is an arduous task that requires in-depth knowledge of the underlying architecture and expensive project time - often months - to achieve a satisfying test coverage. In many instances, specific test details have to be derived from the hardware architecture implementation. Depending on the test, these details span from register initialization sequences up to reasoning about fault propagation paths.

RISC-V offers a modular and fully customizable ISA specification that features a variety of pre-defined extensions. Combining the base ISA with a set of extensions enables designs ranging from power-efficient embedded microcontrollers to high-performance chip designs. This modular structure of the RISC-V ISA leads to an incremental processor design. Consequently, the development of SBSTs is led alongside the incremental implementation of each ISA extension. As the ISA extensions form defined boundaries of SBST modules the reuse of those software modules for other RISC-V cores is encouraged.

For example, both R15CY and Zero-R15CY cores (see Table [1]) implement the RV32I base specification and the M extension. Hence, it is reasonable to assume that some parts and test structures of the same STLs can be re-used for both cores as shown in Figure [1]. Though, when re-using the tests for a different implementation a loss in test coverage is to be expected. Nevertheless, with reduced effort a test engineer can extend and port the existing tests and merely develop additional tests for the remaining, not yet-covered C and F extensions supported by the R15CY core. Although porting is possible, the reader should note that even though the two systems implement the same ISA extensions, hard-to-test faults are architecture-dependent [23].

The architecture dependency of hard-to-test faults implies a high complexity for test generation. For each new architecture, the development effort for testing hard-to-test faults has to be repeated. On top of that, the non-intuitive fault behavior and propagation make manual development unfeasible. Therefore, tool-aided, automatic SBST development is required for targeting these faults.

In the context of test generation, formal methods like bounded model checking (BMC) provide a facility to find test sequences for hard-to-test faults under functional constraints. The test engineer formalizes these functional constraints in a test specification. This test specification models the set of valid test behaviors that are allowed to be generated by the BMC solver. This test specification, among other things,
includes the set of allowed instructions as specified by the ISA, allowed Control and Status Registers (CSRs) accesses and configurations, and the memory model.

Except for the most recently developed extensions, all parts of the official RISC-V ISA exist as formal specifications and build a ground truth for the RISC-V ecosystem of tools. According to the use case, these specifications are combined and inserted into various processes exploiting their reusability. Similarly, these formal specifications are reusable for test generation and allow for the automatic derivation of large parts of the test specification. Hence, RISC-V’s open-source and openly available ISA specifications lay the groundwork for automated SBST generation per extension module.

It has been shown that SBST generation embedding parts of the formal RISC-V specification allows for rapid tool-aided SBST development targeting hard-to-test faults in scalar, single-issue pipelined processor core modules [24]. Depending on the supported ISA extensions of the processor core, a test specification is derived in a modular fashion. By transforming the required specification parts into a test specification and combining them with comparatively few additional constraints, a test procedure can be developed incrementally. The additional constraints are SoC architecture-specific and include parameters like memory regions and exception handling. Additionally, the test procedure type is specified depending on the environment the SBST is run. This environment can be a small-scale test after tape-in during development where the SoC is embedded into a custom test bench and the SoC’s outputs are accessible; or an at-speed end-of-manufacturing test under harsh conditions to test for infant mortality of chips by artificially aging them (Burn-In test); or a timed self-test of chips in-field during operation to detect degradation due to electromigration and fatigue caused by external stress.

Burn-In test (BI) [25], is omnipresent in the safety-critical domain. More specifically, dynamic BI methodologies require functional stress-inducing stimuli. For the case where the devices under test (DUTs) are RISC-V processors [26], the stimuli generation process can be greatly aided and guided by the incremental ISA structure. For instance, such a process requires the identification of the appropriate vectors (i.e., assembly instructions) in order to compose a typically short and stress-effective program for the DUT. That is, a program that is able to maximize the number of logical switches in short periods of time. By having a clear reference of the implemented sub-set of instructions then we can rely on strategies (e.g., SAT-solvers [27], evolutionary algorithms [28]) to identify such sequences and an efficient stress program for the DUT.

Overall, RISC-V’s formal specifications are a cornerstone for functional test constraint derivation e.g., the ISA itself, CSR behavior, and the memory model, directly providing the transition from a specification to functional test generation. RISC-V eases the complexity of STL development by providing an open, modular, well-organized ISA as well as an ecosystem of tools (e.g., compiler, instruction set simulator, virtual prototype) that can be used alongside electronic design automation (EDA) software to develop and validate RISC-V functional test programs.

### B. How RISC-V supports research in System-Level Test

#### System-Level Test (SLT)

System-Level Test (SLT) has emerged as an additional test step in the last decade [30]. It is used to improve the quality assurance of complex SoCs, as traditional testing is not sufficient anymore. Figure 2 shows that SLT is executed as the very last step after Final Test.

SLT tries to approximate the end-user environment of the DUT as closely as possible. For example, a smartphone SoC might be placed into a board that resembles an actual smartphone, plus some monitors and sensors to observe the non-functional properties of the SoC. A test engineer or a system then boots an operating system, in the above example, it might be Android or iOS, and checks for unexpected behavior. If that is not the case, more applications that are typical in the DUT’s mission mode are executed, e.g., a video is streamed onto the device or a browser is started. The DUT is considered defective if, at any point, unusual events such as a crash or unexpected errors occur. During SLT, the DUT is considered a blackbox. This is due to the complexity of these SoCs but also due to the fact that some Intellectual Properties are locked, and only I/O interfaces and behavioral descriptions are provided.

There are some issues with this approach. In particular, these test suites are currently manually composed by test engineers. Furthermore, the quality of those cannot be determined beforehand, as the only available metric is a simple pass/fail. So the defect level and diagnosis of returned ICs are used to determine the quality of the suite after the fact. Finally, typical testing times for SLT are ranging from 15 min to up to 2 h and SLT programs are also very large compared to structural tests or SBST.

With the above issues, it is unfeasible to use traditional methods such as fault simulation to gauge the quality of the test suite. Instead, different metrics and approaches for generating shorter SLT programs are developed [29]. One important factor in generating these tests is that we want to control non-functional properties, such as temperature or power consumption, of our DUT because some defects that are assumed to be detected exclusively by SLT are marginal defects [31]. The detection of this kind of defects depends on certain requirements, such as having a component at a specific
temperature or a temperature gradient between components or a specific sequence of interactions that can only happen in the mission mode of the DUT.

Writing SLT programs that control non-functional properties by hand requires a lot of time and knowledge about the architecture of the DUT, which makes it infeasible in practical use. Instead, there is ongoing research on methods to automatically generate program snippets that are able to reliably control these properties. To develop these methods, it is necessary to analyze multiple different architectures. This is where an open-source SoC generator framework called Chipyard [5] can help.

Furthermore, in Section II-A we already showed how we can easily derive constraints for methods such as evolutionary algorithms from the freely-available specification. This is useful for SLT program generation as well. For example, we can use greybox-based methods that are guided by feedback to generate program snippets that optimize a specific non-functional property of our DUT [12]. These methods include genetic programming [33] or mutation-based greybox fuzzing [34].

Chipyard has the distinct advantage that it allows us to freely configure SoCs based on either an in-order (Rocket [4], CVA6 [11], Ibex [8] and an educational core called Sodor) or out-of-order (BOOM [15]) core or even mix both. For example, we can add or remove caches and change their size or modify aspects of the cores themselves, such as the number of issues in the BOOM core, the size of the return order buffer, and much more. We can also freely enable or disable extensions and choose between a 32-bit and 64-bit architecture. The flexible configurability allows us to examine multiple configurations for real-world different scenarios, e.g., embedded processors or high-end smartphone SoCs, and identify common and architecture-specific features to derive a generic model of non-functional properties to use for SLT program generation.

Additionally, there is a wide range of well-tested periphery available that we can utilize as well to test interactions between the core and the periphery. And, due to the open-source nature of Chipyard, we are able to add our own periphery and make changes to existing ones, such as, for example, bus monitors, to observe activity on all the buses in the DUT. We can use this as feedback for a fuzzer to, for example, maximize the contention on each bus or cover all the possible transactions with all possible sources and destinations.

Chipyard already supports many of the designs listed in Table I. As several of these designs are also compatible with traditional ASIC design flows, it is possible to extract more information such as power simulations, map them to an FPGA to speed-up simulation time and get hardware measurements, and extract additional insight by adding output of bus monitors as feedback for the aforementioned greybox monitors.

Moreover, the RISC-V tools and cores are easy to set up with other commercial tools for further development and investigation. This allows us to work on new, more elaborate metrics to evaluate the performance of the different SLT snippets. Not just for non-functional properties but also for standard fault models (e.g., stuck-at or transition fault delays). Thus, we have some insights into possible causes for marginal defects and can figure out a way to imitate the conditions in which they manifest themselves. Causes of these marginal defects are still unknown, but by having such insights we might be able to find solutions to mitigate them and generalize the behavior of these.

On top of that, by having this transparency of the architecture, we can develop the SLT snippets to target specific peripherals or specific areas in the core that we believe to be problematic. Therefore, the SLT snippets can be developed as compact as possible, reducing the test times but maintaining the fault coverage or even achieving a higher fault coverage.

We also had a chance to compare a RISC-V design with and without DfT structures. We draw the conclusion that DfT structures are needed to facilitate the traditional structural test insertions as well as the SLT test insertion by speeding up the simulation process, having full controllability and observability of the DUT and resulting in much higher fault coverages than without DfT structures. This visibility is what helps to identify more faults and gives the insight on possible causes of a defect.

The freely available RISC-V specification and the rich ecosystem of open-source cores and tools allow researchers to work on functional program generation easily by, on the one hand, enabling them to easily derive constraints to automatically generate valid assembly snippets, and, on the other hand, allow to analyze and use architectures for different application scenarios to extract common and architecture-specific features to guide test program generation.

III. RISC-V for Safety-related Applications

In the context of safety-critical systems several different ISAs are widely used in different applications domains. Some illustrative examples are the case of the SPARC [36] for space, the TriCore [37] in the automotive domain, and the PowerPC in avionics and aerospace. Recently, ARM processors [38] have also increased their presence in safety-critical domain applications allowing safety-critical application developers enjoying from the wide ARM software ecosystem.

In this context, RISC-V arises as a great opportunity to break the existing safety-critical system fragmentation, and many industrial players in the safety-critical domain have shown their interest in RISC-V. The modularity, openness, and flexibility of RISC-V are the key features for its potential success in safety-critical systems. However, the adoption of RISC-V is also subject to having processor architectures that meet the stringent needs for certification of functional safety products. Currently, the RISC-V functional safety special interest group is working on a white-paper covering the best practices for the design of RISC-V processors safety-related applications. The following two subsections about functional correctness (Section III-A) and timing verification (Section III-B) elaborate on the impact of the functional safety in the design of RISC-V based processors. Finally, in the last Section III-C we review some of the existing RISC-V processors targeting safety-critical applications.
A. Functional Correctness

The functional safety measures at hardware level commonly include fault tolerance, error detection and correction mechanisms. First of all, the processor internal memory structures, such as register files and L1/L2 caches, must be protected with error correction codes (ECC) [39]. Single-error correction and double-error detection (SECDED) Hamming codes are commonly considered in this context because of their moderate resource and performance overheads [40]. Likewise, the internal interconnect and external communication buses are subject to ECC protection.

The replication of entire on-chip components is another widely-used technique to detect and correct hardware faults. The basic fault detection capabilities are achieved by dual modular redundancy. In multicore processors it is typically implemented in the form of dual-core lockstep (DCLS) [41], a mechanism that executes the same program on two CPU cores in parallel, compares cores outputs at each clock cycle and signals any discrepancies. Some lockstep processors allow flexible usage of the redundant (slave) cores, providing a lock mode for reliable execution of critical applications, and split mode for high-performance execution of non-critical applications [42].

The triple-modular redundancy (TMR) or higher multiplicity replication schemes, for instance triple-core lockstep [43], enable real-time error correction.

The diverse execution of the replicated cores plays an important role in any replication scheme, whenever it is supposed to prevent common-mode failures. Diversity can be introduced at ISA and/or microarchitectural levels, resulting in heterogeneous fault-tolerant CPU architectures [44]. These solutions, however, are often considered impractical due to their increased cost and complexity of software running on top of heterogeneous cores. A more affordable (low-overhead) approach for diverse redundancy is time staggering mechanism, which introduces controlled inter-core delay to ensure that replicated (homogeneous) CPU cores execute different instructions at any given time instant.

Additional fault detection mechanisms, commonly provisioned in dependable CPU architectures, include watchdog timers, performance monitoring (profiling), trace and debug support units.

The RISC-V implementation with integrated fault tolerance and error detection mechanisms can be verified by means of fault injection (FI) experiments. FI is a well-known testing technique [45], used to verify the behaviour of critical system in presence of faults, evaluate the efficiency of integrated safety mechanisms, identify safety vulnerabilities and error propagation paths. In fact, the usage of FI testing during the system design is recommended by ISO 26262 standard [46]. At the early design stages, the simulation-based FI allows to verify safety mechanisms implemented in high-level HDL model of the system, e.g. ECC, lockstep execution, etc. Whereas at the late design stages the FPGA-based FI and physical FI are useful to evaluate safety features with all implementation details taken into account, including synthesis-place-route optimizations and impact of implementation technology.

B. Timing Verification

Functional safety certification processes impose several requirements to the software development and validation processes [47]. Several of these requirements are related to the timing behaviour of applications and are tightly coupled with the characteristics of the processor in which the software is executed. In general, and regardless of the functional safety application domain, the execution time of software needs to be bounded at the unit level and to enable a reliable integration of such elements the hardware/software architecture must ensure freedom from interference at the scheduling level. These software requirements are usually translated into the need of deriving the worst-case execution time (WCET) for each software unit and the need for temporal and spatial isolation between the different software items.

WCET derivation is challenged by the presence of hardware features that introduce execution time variability such as caches and shared resources. Thus, processor designs targeting safety-critical applications need to implement specific features targeted to control, reduce or completely remove such execution time variability and achieve a predictable behaviour. For instance, execution time predictability can be attained by replacing caches by scratchpads or by implementing restrictive policies for resource sharing such as time-division multiplexing [48]. However, in general, achieving constant execution time in specific resources comes at the expense of computing performance and/or flexibility which is not acceptable in many application domains. Thus, the current trend towards attaining predictability in complex hardware platform relies on the use of timing monitoring and quota enforcement policies [49].

Resource partitioning can be implemented with software only means and with both hardware and software support. However, fine-grain and performance efficient partitioning requires specific hardware support. In that respect, the RISC-V ISA has defined the H extension for virtualization. The H extension enables the execution of unmodified OS as hypervisor guests and thus, the co-existence of applications with heterogeneous needs in terms of performance and criticality. Unfortunately, the partitioning implemented at the H-extension level does not prevent the execution time interference originated at micro-architectural features like caches, queues and interconnects. For that, additional hardware support such as cache partitioning, time-predictable arbitration policies, and dedicated hardware for fine-grain isolation is needed [50].

C. Existing RISC-V Solutions for Safety-related Applications

Several RISC-V processors have been adapted or designed to meet functional safety requirements. The NOEL-V [51] from FrontGrade Gaisler is a processor for the space domain that implements the RV64GCH ISA configuration. However, the GPL version of the NOEL-V core does not implement fault-tolerant features. The NOEL-V has also been used to build the SELENE platform [2]. The SELENE platform is a multicore SoC that implements hardware-based diverse redundancy in
the form of non-intrusive light-lockstep [53], the safeSU [54] hardware monitors for timing verification and enforcement, and provides virtualization support for partitioning hypervisors such as Jailhouse [55] and XNG, both already supporting RISC-V. RISC-V based systems from the PULP project [56] have also been adapted to ease its adoption in safety-related applications. For instance, a RISC-V cluster based on the PULP project [57] implement hybrid modular redundancy to select between dual and triple redundancy with lockstep execution in a flexible manner.

Other RISC-V processors targeting safety-related applications that are commercially available (not open-source) are the SiFive P550 Application class processor with configurable dual-core lockstep, RAS capabilities, and ECC in interconnect and memories, and the Fraunhofer EMSA5-FS which has been certified as ASIL-D ready according to ISO 26262:2018 for functional safety in vehicles. A RISC-V processor tailored for high-energy physics applications (e.g., particles detectors at CERN) based on Fraunhofer AIRISC core [58] combines radiation-hardened fabrication technology with several architectural protection measures, including fine-grain TMR and extensive memory scrubbing.

IV. RISC-V FOR SECURITY APPLICATIONS

In this section, we discuss recent developments concerning RISC-V cores with respect to security. We begin with a conceptual overview of these developments from cryptographic, hardware, and ISA perspectives. These perspectives are presented in Sections IV-A, IV-B, and IV-C, respectively. In Section IV-D, we briefly cover specific RISC-V security topics from a system point of view.

A. Conceptual Overview

In this section, we present RISC-V features, supporting secure communication, from three perspectives. We begin the discussion by categorizing the security features based on the type of cryptographic computation they support. This is presented as the cryptographic perspective (Figure 3—right). The pyramid shows examples of basic building blocks, primitives, and protocols required to provide information security. They are placed in a pyramid to express the hierarchical dependency among them, i.e., the basic building blocks are used to construct (composed) primitives that are used to construct security related protocols. The hardware perspective (Figure 3—middle) shows hardware design decisions distinguishing between dedicated hardware added to the main RISC-V core and a coprocessor interfaced with the main core. Finally, the ISA perspective distinguishes between Base ISA, Standard Extensions, and Custom Extensions (Figure 3—left).

1) The cryptographic perspective: The pyramid in Figure 3 shows, in a top-down fashion, some main building components needed to assure secure communications. Cryptographic protocols providing secure communication channels (for example the DTLS protocol) are designed as challenge-response handshakes with appropriate steps to assure confidentiality, data integrity, authentication and non-repudiation, and furthermore to prevent known attacks such as replay, man-in-the-middle, etc. These protocols require several cryptographic primitives, a.o. digital signatures (like ECDSA) for authentication. Such cryptographic primitives are composed of basic building blocks. For example, the round function of the block cipher AES is composed of basic primitives such as SubBytes, ShiftRows, MixColumns, AddRoundKey, and relies on finite field arithmetic.

2) The hardware perspective: In the past decades, it became clear that security is a necessity for many embedded software applications. However, introducing security-related features in a system induces a communication and computational overhead. Such overhead can be reduced by adding dedicated hardware to a general-purpose embedded processor in order to accelerate the security-related computations in an application. In the middle of Figure 3 we show the hardware perspective of security-related features existing in RISC-V based cores and distinguish between dedicated co-processors tightly/loosely coupled with the main core and dedicated hardware additions to the main core. For example, attaching a cryptographic co-processor, such as the DTLS Cryptographic Engine [59], is a common solution. Alternatively, dedicated hardware, such as the AES functional unit [63], can be added to the datapath of the main core. The work presented in [73] analyzes the building blocks of the Lightweight Cryptography (LWC) finalists to find the most suitable acceleration granularity, e.g., their smallest hardware addition is for the Xoodyak parity plane manipulation (xorrol) with only a 1.155x increase in the area of the main core.

3) The ISA perspective: The RISC-V ISA is designed to be modular including a relatively small stable base ISA and multiple standard extensions while it also allows the design of custom instruction extensions. The distinction based on the presence of custom extensions is shown on the left of Figure 3. For example, the aforementioned DTLS engine interacts with the main core via a memory-mapped interface.
Table II: Cryptographic applications and granularity of added hardware and added custom ISA extensions in the context of hardware/software co-design

<table>
<thead>
<tr>
<th>RISC-V ISA</th>
<th>RISC-V Main Core</th>
<th>HW perspective</th>
<th>Application and granularity (Figure 1)</th>
<th>Implementation comments</th>
<th>Technology CAD tools</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV32 I</td>
<td>BlueSpec RISC-V</td>
<td>✓</td>
<td>memory mapped interrupts (wfi) FSM in coproc.</td>
<td>entire DTLS as coprocessor: ECDHE &amp; ECDSA AES-128-GCM, SHA2-256</td>
<td>SPA-hardened configurable prime field ECC accelerator, stand-alone crypto possible</td>
<td>65nm LP CMOS, fabricated</td>
</tr>
<tr>
<td>RV32 IM</td>
<td>BlueSpec RISC-V</td>
<td>✓</td>
<td>memory mapped interrupts (wfi), Sapphire inn. decode-control</td>
<td>entire LWE-PQC as Sapphire coprocessor: Frodo, NewHope qTESLA, CRYSTALS-Kyber CRYSTALS-Dilithium</td>
<td>modular ALU with config. prime, SRAM-based NTT butterfly unit, Keccak core distribution sampler</td>
<td>40nm LP CMOS, fabricated</td>
</tr>
<tr>
<td>RV32 I</td>
<td>Rocket</td>
<td>✓</td>
<td>RoCC</td>
<td>entire crypto coprocessor: AES, ECC, SHA-256</td>
<td>4-stage pipelined coprocessor, finite field multiplier</td>
<td>28nm CMOS, post-layout, Synopsys VCS and PrimeTime PX</td>
</tr>
<tr>
<td>RV32 IM + Custom</td>
<td>Ibex</td>
<td>✓</td>
<td>AES IP connected to CSRs</td>
<td>AES coprocessor with one custom instruction</td>
<td>AES IP in decode stage mode as insn. parameter (ECB, CBC, CFB, OFB)</td>
<td>Nexys Artix-7, Xilinx Vivado 2018.1 post-P&amp;R simulation</td>
</tr>
<tr>
<td>RV32/64 IM + Custom</td>
<td>SCARV 32-bit</td>
<td>✓</td>
<td>main core decoder</td>
<td>AES-GCM with custom insns for: SubBytes, MixColumns, and SubBytes + MixColumns</td>
<td>AES-FUS for 4 32-bit extension sets and 1 64-bit set, carry-less mult. for GCM</td>
<td>Yosys post-synthesis RISC-V Cryptography Extension K proposals</td>
</tr>
<tr>
<td>RV32 IMCF</td>
<td>CV32E40P (RISCY)</td>
<td>✓</td>
<td>Accelerator connected to FPR</td>
<td>finite field accelerator, used for IBC (PQC): SIKE</td>
<td>Montgomery Multiplier, Modular Adder and Subtractor</td>
<td>Xilinx ZedBoard, Xilinx Vivado 2017.4</td>
</tr>
<tr>
<td>RV32 I</td>
<td>VexRiscv</td>
<td>✓</td>
<td>ABP decoder connected to the main core</td>
<td>prime field coprocessor, used for IBC (PQC): SIKE</td>
<td>Modular arithmetic adder, dual Montgomery multiplier for all SIKE primes</td>
<td>Xilinx Virtex-7, post-P&amp;R</td>
</tr>
<tr>
<td>RV32 IM</td>
<td>CV32E40P (RISCY)</td>
<td>✓</td>
<td>32-bit AHB Intercconnect</td>
<td>NTT accelerator, hash accelerator, used for LBC (PQC): NewHope</td>
<td>DMR hardened NTT and NTT-LUT core, hash core with Keccak</td>
<td>Xilinx ZedBoard Zynq-7000, FI simulation</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>VexRiscv</td>
<td>✓</td>
<td>Flexible pipeline: plugins</td>
<td>prime field coprocessor, w/ 4 custom insn, used for LBC (PQC): Kyber, NewHope</td>
<td>2 variants: synthesized with a fixed prime, and flexible prime as parameter</td>
<td>Artix-35T, Xilinx Vivado 2019.1, iCE40 UltraPlus, kcestorm</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>CV32E40P (RISCY)</td>
<td>✓</td>
<td>main core decoder</td>
<td>finite field accelerator, w/ 4 custom insns used for LBC (PQC): LAC</td>
<td>PQ-ALL, ternary multiplier, Chien module, modular reduction , SHA256</td>
<td>Xilinx ZCU102 Zynq UltraScale+</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>SCR1</td>
<td>✓</td>
<td>32-bit AHB-Lite Intercconnect</td>
<td>vector coprocessor, w/ 12 custom insns used for LBC (PQC): Kyber, NewHope, LAC</td>
<td>NTT core, Keccak core, binomial sampler, Random Number Generator</td>
<td>TSMC 28nm HPC, fabricated</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>Ibex</td>
<td>✓</td>
<td>Vector Instruction Interface</td>
<td>vector coprocessor, w/ 16 custom insns used for LBC (PQC): CRYSTALS-Kyber</td>
<td>NTT, CWM, modular arithmetic for each exLane, register pooling, automatic index generation</td>
<td>Xilinx Alveo U250 Vivado 2019.2</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>SweRV-EL2</td>
<td>✓</td>
<td>main core decoder</td>
<td>finite field accelerator, w/ 4 custom insns used for: AES, Reed. Solomon codes</td>
<td>carry-less multiplication, polynomial reduction, parameterized degree, irre. poly.</td>
<td>Nexys Artix-7, Verilator v4.032</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>CV32E40P (RISCY)</td>
<td>✓</td>
<td>main core decoder</td>
<td>various accelerators, w/ 29 custom insn., used for LBC (PQC): NewHope, Kyber, Lightsaber, Firesaber, Saber</td>
<td>NTT, modular arithmetic, Keccak in decode stage bimodal sampling, Karatsuba, Toom-Cook multipliers in execute stage</td>
<td>Xilinx Zynq-7000, UMC 65nm, Cadence Incisive Enterprise Simulator, Joudes</td>
</tr>
<tr>
<td>RV32 GC + Custom</td>
<td>Rocket</td>
<td>✓</td>
<td>main core decoder</td>
<td>LWC finalists: separate extensions for each candidate by benchmarking, e.g., Ascon sigma0, sigma1 inn., e.g., XOODYAK xorrol inn.</td>
<td>main core, Zbhku FU, ans ISE-specific LWC FU for each candidate, also implementing AES-GCM for reference</td>
<td>SASEBO-GH1 board Kintex-7 target, Xilinx Vivado 2019.1</td>
</tr>
<tr>
<td>RV32 IM CF</td>
<td>CV32E40P (RISCY)</td>
<td>✓</td>
<td>main core decoder</td>
<td>Ascon-p with a custom insn: permutation with number of rounds as parameter</td>
<td>Ascon-p in Decode stage, Internal state in GPR Ascon, Ascon-hash, ISAP</td>
<td>65nm LL CMOS Cadence Encounter RTL Compiler, NanoRoute</td>
</tr>
</tbody>
</table>

PQC = Post-Quantum Crypto, IBC = Isogeny Based Crypto, LBC = Lattice-Based Crypto, LWE = Learning With Errors, ECC = Elliptic Curve Crypto. LWC = Light-Weight Crypto., SPA = Simple Power Analysis, FI = Fault Injection, DMR = Dual Modular Redundancy CSR = Control and Status Registers, FPR = Floating Point Registers, GPR = General Purpose Registers

and does not require any (custom) additions to the ISA [59]. In contrast, the work in [62, 63] presents a design of custom AES instructions for RISC-V. In general, with custom ISA extensions, ISA and hardware designers can choose a suitable level of granularity of the instructions and dedicated hardware added to the RISC-V based architecture. For example, the
authors of [62] implemented the entire AES primitive with different modes, including ECB, as a custom instruction, while [63] presents separate RISC-V ISA instructions for utilizing the SubBytes and MixColumns basic blocks as well as a single instruction for utilizing both SubBytes + MixColumns.

Finally, there exists an ongoing effort to propose and standardize RISC-V cryptography ISA extensions containing two orthogonal sets of specific instructions: first, scalar & entropy source instructions [75], and second, vector cryptographic instructions [76]. The scalar & entropy source instructions are cryptographic instruction proposals for smaller cores that do not implement vector extensions. They include a subset of "constant-time" instructions with data-independent execution time to prevent timing side-channels. In contrast, vector cryptographic instructions are proposed to be used in large and high performance cores that have large vector registers that are reused. Some examples are AES and SHA-2, with instructions on different granularity, for example single-round (building block) and all-round (primitive). It is also possible to vectorize SHA-3 using a number of general-purpose instructions, allowing for simultaneous processing of multiple pieces of data to increase the performance of SHA-3 [76].

B. Selected RISC-V Cores with Security-related Features

In this section, we present selected state-of-the-art proposals for RISC-V implementations with security support. These proposals are shown in Table I where the order of the rows roughly follows the top-down cryptographic perspective illustrated with the pyramid in Figure 3. The columns in Table I are organized starting with the ISA perspective and the RISC-V main core, in the first two columns, followed by three columns summarizing the hardware perspective, with checkmarks for dedicated hardware added to the main core and co-processor designs. When custom extensions are present, indicated in the ISA column, the instruction decoder in the main core has to be modified but we do not consider this as an addition to the main core in Column 3. Column 5 lists some implementation details to show the relationship between the dedicated hardware and the main core. For example, the Sapphire co-processor in Row 3 has its own instructions decoder and controller, and could be reused with any main core.

Column 6 provides some application details from the cryptographic perspective, illustrated with the pyramid in Figure 3, together with the granularity of the ISA extensions and corresponding hardware additions. Column 7 provides very brief implementation comments, such as expensive cryptographic computations, i.e. sub-modules illustrated with the pyramid in Figure 3 and architectural decisions. Finally, in Column 8, we provide some information on the technology and CAD tools used to design the RISC-V cores, focusing mostly on fabricated ASIC and configured FPGA designs.

C. Towards Side-Channel Attack Resilience for RISC-V

During cryptographic operations, computations are performed using sensitive data such as cryptographic keys. It has been shown that such sensitive data can be revealed by exploiting side-channel leakage, such as power consumption or electromagnetic emanations. Thus, in the past two decades, we have seen the rise of side-channel analysis (SCA) for cryptographic circuits and systems. More recently, such SCA trend could be witnessed for RISC-V based systems featuring support for security-related applications. We briefly discuss some side-channel hardened RISC-V solutions, aiming to prevent, or at least to hinder, the extraction of sensitive data from RISC-V based systems.

As mentioned earlier, the DTLS Cryptographic Engine [59] is often attached as a co-processor to a RISC-V core. This engine implements countermeasures to prevent simple power analysis (SPA) from distinguishing point addition and point doubling. Similarly, the Sapphire co-processor implements constant-time SPA-hardened sub-modules [60]. The designers of Sapphire also conducted simple power analysis on the fabricated chip to verify the constant-time execution. The Sapphire co-processor has no differential power analysis (DPA) countermeasures, but its designers have discussed how to include such countermeasures using the Sapphire’s programmability.

The authors of [77] propose a secure RISC-V ISA and architecture by including an additional secure pipeline which is completely separated from the unprotected (original) pipeline. The instruction decode stage disables the unprotected pipeline during execution of protected instructions to prevent leakage. The secure pipeline executes a set of instructions protected by Domain-Oriented Masking (DOM) with multi-cycle nonlinear operations, e.g., 4-cycle 1-bit ADD instruction with two fresh random numbers per share. The idea of custom ISA instructions to mitigate SCA has been extended in [78] to include DOM and other masking approaches. The authors have presented 22 custom instructions such as conversion of operands (Boolean to/from Arithmetic masking), Boolean masking operations, arithmetic masking operations, and masked finite field arithmetic for $GF(2^8)$. They place a masked ALU containing dedicated hardware, including random bit generators, in parallel to the unprotected ALU in the SCARV core, with a 1.45× area increase for the ASIC implementation. Additional randomness and re-masking are used to compensate for not duplicating the datapath. The work in [78] also provides preliminary SCA results on FPGA implementations of the SCARV core by running and comparing unmasked AES using only the base ISA and masked AES using the aforementioned custom extensions.

Finally, the authors of [79] have used formal verification methods to design a secure Ibex core with additional features, preventing leakage, such as secure register file, AND-gated computation unit, and clear for the hidden LSU buffer. The security features increase the total Ibex core area by only 9.9%.

D. RISC-V System Security

This section briefly discusses selected RISC-V security topics, such as Root-of-Trust (RoT), trusted boot, memory protection and isolation, and trusted execution environment (TEE), from the perspective of added (cryptographic) hardware. Some of these topics are also covered in [1], [20], [80].
A secure bootloader is considered to be the smallest RoT, on top of which the chain of trust and trusted code base are built. For example, Sanctum [81] uses trusted on-chip ROM to hold the bootloader. In addition, the system includes a hardware TRNG and PUF, and software SHA-3. ECDSA and AES implementations [82] to provide strong isolation using enclaves and to protect against software attacks, such as cache timing attacks and passive address translation attacks. Sanctum has been designed for Rocket RISC-V with minimal hardware extensions and implemented on Xilinx Zynq 7000 FPGA.

Keystone [83] is an open-source framework for building customized TEEs with only basic primitives implemented in hardware. The framework is comprised of a device-specific secret key available only to the trusted boot process, a hardware source of randomness, and a trusted boot process. The Keystone RoT can be either a hardware crypto engine or tamper-proof software. It uses physical memory protection (PMP) and a security monitor running in machine mode to enforce memory isolation. For example, a SoC with two Rocket cores using Keystone implements two hardware accelerators, SHA-3 and ECDSA, connected via Tilelink [84]. Another example is the ITUS RISC-V based secure SoC [85] which includes two Rocket cores and integrates Keystone-enclaves for TEE. The SoC, implemented on Xilinx Kintex 705, also includes a hardware key management unit with one-time programmable memory, TRNG and PUF, a hardware signature verification unit with boot sequencer, SHA3, the XMS and ECDSA digital signature schemes, and a memory protection unit with AES-GCM. In addition, [85] also considers countermeasures for side-channel attacks. Keystone has been also modified to run on the CVA6 core and on a GPU-scale accelerator with 4096 cores, divided into clusters of 8 [86].

V. Conclusion

The RISC-V ISA and processors implementing it are receiving an ever-increasing attention from both the scientific community and the industry. The test, safety and security aspects discussed in this survey are key prerequisites for RISC-V processor being useful for actual products. While many basic concepts and approaches previously developed for different architectures can be reused with minor modifications, the open nature of RISC-V ISA calls for new and more universal solutions for some of the problems mentioned. Therefore, this survey can only provide a snapshot of recently obtained results, and the scientific work will continue in the foreseeable future. In order to obtain the best solutions, it is important to establish and maintain connection between RISC-V architects and specialists in test methods, safety and hardware-oriented security.

Acknowledgment

This work was supported in part by the German Federal Ministry of Education and Research (BMBF) within the project Scale4Edge under contract no. 16ME0132 and by the Italian ICSC National Research Centre for High Performance Computing, Big Data and Quantum Computing within the NextGenerationEU program. UPV researchers have received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 877056, Agencia Estatal de Investigación from Spain under grant agreement no. PCI2020-112092 within the NextGenerationEU program. Furthermore, this research was supported by Advantest as part of the Graduate School ”Intelligent Methods for Test and Reliability” (GS-IMTR) at the University of Stuttgart. For this work, Leiden University received funding from the Dutch Research Council (NWO) through the PROACT project (NWA.1215.18.014).

References
