System Integration of Elliptic Curve Cryptography on an OMAP Platform

Sergey Morozov  
ECE Department  
Virginia Tech  
Blacksburg, VA 24061  
Email: morozovs@vt.edu

Christian Tergino  
ECE Department  
Virginia Tech  
Garden City Park, NY 11040  
Email: etergino@vt.edu

Patrick Schaumont  
ECE Department  
Virginia Tech  
Blacksburg, VA 24061  
Email: schaum@vt.edu

Abstract—Elliptic Curve Cryptography (ECC) is popular for digital signatures and other public-key crypto-applications in embedded contexts. However, ECC is computationally intensive, and in particular the performance of the underlying modular arithmetic remains a concern. We investigate the design space of ECC on TI’s OMAP 3530 platform, with a focus on using OMAP’s DSP core to accelerate ECC computations for the ARM Cortex A8 core. We examine the opportunities of the heterogeneous platform for efficient ECC, including the efficient implementation of the underlying field multiplication on the DSP, and the design partitioning to minimize the communications overhead between ARM and DSP. By migrating the computations to the DSP, we demonstrate a significant speedup for the underlying modular arithmetic with up to 9.24x reduction in execution time, compared to the implementation executing on the ARM Cortex processor. Prototype measurements show an energy reduction of up to 5.3 times. We conclude that a heterogeneous platform offers substantial improvements in performance and energy, but we also point out that the cost of inter-processor communication cannot be ignored.

I. INTRODUCTION

Modern mobile devices and embedded systems are designed to provide an ever increasing number of services for the end-user. Secure communication in particular is in demand for mobile phones, cars, remote operation and sensing, and more. A vital component of a secure communications software stack includes signatures and key-exchange, requiring public-key cryptographic algorithms. In contrast to symmetric-key cryptography, the computational costs associated with public key cryptography are high - typically three orders of magnitude above symmetric-key cryptography. Public-key algorithms like RSA, DSA, and elliptic curve cryptography (ECC) all deal with operands hundreds to thousands of bits in size and use modular arithmetic for the underlying computations. In particular, modular multiplications are a critical underlying part for public-key cryptography, and it can be prohibitively expensive on smaller embedded processors.

While algorithmic-level improvements are continually proposed, efficient mapping in hardware or software is always important. In an embedded context, where the application has a strong influence on the system architecture, dedicated hardware solutions are appealing. This includes for example, expanding the capabilities of an existing processor with specialized instructions or designing completely new co-processors.

Researchers often approach the problem by coming up with an HDL description of the new co-processor or datapath, and conclude the research with a simulation or an FPGA prototype of the design.

We approached this problem from another angle. We have observed that recent trends in mobile and embedded computing have pushed increasingly complex System-on-Chip designs into heterogeneous platforms such as TI’s OMAP and DaVinci chips. These SoCs now frequently include multiple heterogeneous cores: a standard RISC core could be augmented with DSP, SIMD extensions, and a graphics processing unit. Our research specifically addresses the use of a VLIW DSP processor for cryptography. Previous research has shown that such a processor could efficiently implement modular arithmetic [1], [2]. In addition, the side-channel characteristics of ECC on a DSP have been investigated [3]. Our objective was to study the integration costs that arose from ECC in a heterogeneous, DSP-enabled SoC.

The contributions of this paper are two-fold. First, we present an original implementation of binary-field modular arithmetic on a VLIW DSP processor. Our solution implements multi-precision arithmetic that utilizes the DSP’s XOR-multiply instructions, and also exploits multiple concurrent datapaths to achieve a significant speed-up of binary field multiplication relative to the ARM RISC core on the same chip.

Second, we examine the design space of elliptic curve cryptography systems on a typical heterogeneous multicore platform, in this case the TI OMAP 3530. This chip contains an ARM Cortex A8 and a C64x+ DSP core. We present a solution for high speed ECC cryptography services under varying ECC cryptographic parameters. We show that the multicore implementation demonstrates a significant speed-up over existing production-grade ECC libraries for a general-purpose processor. We then make a strong effort to create a fair comparison between our heterogeneous ARM+DSP implementation, and OpenSSL ARM-only implementation. The comparison covers performance as well as energy cost of the resulting design.

The remainder of this paper is presented as follows. We first review the basis of elliptic-curve cryptography, enumerating the computational requirements. We then discuss an imple-
mentation of binary-field modular arithmetic on a VLIW DSP. In Section 4, we integrate the resulting library in a system-library for elliptic-curve cryptography on the OMAP platform. We analyze the results in Section 5 and conclude.

II. COMPUTATIONAL COST OF ELLIPTIC CURVE CRYPTOGRAPHY

The basics of elliptic curve cryptography are extensively discussed in literature; see for example Hankerson et al. [4]. In this section, we will only highlight the computational characteristics of ECC.

ECC is defined using Elliptic Curves over prime or binary finite fields. Prime fields $GF(p)$ use integers less than a p-bit prime; binary fields $GF(2^m)$ use polynomials with binary coefficients of degree less than an m-degree irreducible polynomial. The elliptic curve domain parameters, as well as the field definitions, are standardized, such as by SEC [5] or by NIST [6]. Our experiments were performed using prime curves on 160 and 224 bit fields, and using binary curves on 163, 283, and 409 bit fields. These curves are called secp160r1, secp224r1, and secp163r1, secp283r1, secp409r1 for prime-field and binary-field SEC curves respectively. The large field-sizes used in ECC imply that the fundamental operations have to be implemented using multi-precision arithmetic. For example, the prime-field secp160r1 curve requires 160-bit operands, which each take 5 words on a 32-bit processor. Each operation in 160-bit arithmetic needs to be expressed in terms of a 32-bit datapath.

The central operation in ECC is the point multiplication, which multiplies a point on a curve with a scalar number k. Cryptographic protocols are constructed based on the point multiplication [6]; for example, generating a digital signature requires one point multiplication, while verifying it requires two point multiplications. The point multiplication is computationally expensive. A straightforward implementation of a point multiplication on the secp160r1 curve requires 1600 modular multiplications and 960 modular squarings, each of them in a 160-bit field. Hence, the crypto-engineering community makes a significant research effort to optimize the elliptic-curve algorithms. For example, the Explicit-Formulas Database [7] is an excellent source to find the latest optimizations to the ECC algorithms. Regardless of this optimization effort, however, it is clear that optimized mapping of the existing algorithms to processors is important. In this paper, we focus on this aspect, starting with efficient mapping of binary-field modular multiplication.

III. EFFICIENT MODULAR MULTIPLICATION ON A VLIW DSP

In this section, we describe an efficient implementation of modular arithmetic on a VLIW DSP processor, using the TI C64x+ in the OMAP as an example. We focus on the case of binary-field arithmetic; a discussion on efficient modular arithmetic on a VLIW DSP for prime field may be found in [1].

$GF(2^m)$ relies on binary polynomial multiplications. A large majority of processors do not have support for this arithmetic. For this reason, software implementations of ECC often stick to prime field ($GF(p)$). However, the C64x+ does have support for binary polynomial multiplication, an often-used operation in the domain of error control coding [8]. In our binary-field operations, we make use of this feature. In addition, we target an implementation that makes optimal use of the instruction-level parallelism in the VLIW DSP.

A. Binary Field Arithmetic

Binary field numbers are within $GF(2^m)$ and are represented in the form

$$A(x) = a_0x^0 + a_1x^1 + \ldots + a_{m-1}x^{m-1},$$

where each coefficient $a_i = \{0,1\}$. A binary field also needs an irreducible polynomial $f(x)$:

$$f(x) = T(x) + x^m$$

$$T(x) = t_0x^0 + t_1x^1 + t_2x^2 + \ldots + t_{m-1}x^{m-1}$$

where every coefficient, $t_i \in \{0,1\}$. All operations in $GF(2^m)$ are performed modulo $f(x)$. Addition and subtraction are equivalent to performing a bitwise Exclusive-OR between the two operands. To multiply two polynomials, one will first perform a polynomial multiplication, and next reduce the result using the irreducible polynomial that defines the field. When multiplying two polynomials $A(x)$ and $B(x)$ of degree $m-1$, the result is a polynomial $C(x)$ of degree $2m-2$.

$$C(x) = c_0x^0 + c_1x^1 + \ldots + c_{2m-2}x^{2m-2},$$

with

$$c_0 = a_0b_0,$$

$$c_1 = a_1b_0 + a_0b_1,$$

$$c_2 = a_2b_0 + a_1b_1 + a_0b_2$$

$$\ldots$$

$$c_{2m-2} = a_{m-1}b_{m-1}$$

The result of the modular multiplication of $A(x)$ and $B(x)$ is $C(x) \mod f(x)$. We next describe an efficient implementation of the $\mod f(x)$ operation.

B. Unbalanced Exponent Modular Reduction

The binary polynomials in the fields we used (secp163r1, secp283r1, secp409r1) are unbalanced: this means that $T(x)$ in (2) is of low degree and much smaller than $m$. An efficient
method for reduction over binary fields, which exploits this unbalanced character, was introduced in [9]. To reduce $C(x)$, it is first divided into two parts $C_h(x)$ and $C_l(x)$, both of which have a degree lower than $m$. Then, the reduction can be formulated as follows.

Since $C(x) = C_h(x) + C_l(x)x^m$
and $T(x) = x^m \mod f(x)$

then $C(x) \mod f(x) = C_h(x) + C_l(x)T(x) \mod f(x)$

In case the product $C_h(x)T(x)$ still has coefficients with a degree higher than $m$, the reduction step in Equation (4) is applied repeatedly. For the polynomials used in our experiments, no more than two iterations are needed for a complete reduction. We can now map the efficient multiplication and reduction on the OMAP’s DSP processor as follows.

C. Binary-Field Multiplication on TI C64x+ DSP

The C64x+ is a VLIW processor. It contains two identical data paths, each with four functional units. Each instruction is 32-bits and the processor can execute 8 instructions every cycle. The processor contains two register files, one for each data path. Each register file has 32 32-bit registers. Of particular importance for this paper is the XORMPY instruction on the C64x+. This instruction is identical to a normal multiplication except that the partial products are XORed together instead of added. This is ideal for binary polynomial multiplication. The operand sizes for XORMPY are limited to 32-bits and 9-bits and the product size is limited to 32-bits.

When implementing long-word arithmetic such as for $GF(2^m)$ multiplication, we need to use all output bits of XORMPY, in order to minimize the total number of XORMPY per modular multiplication. Therefore, we limit the input operand sizes to 24-bit and 8-bit respectively. Figure 1 shows an operand, “A”, being multiplied by an operand, “B.” Operand “A” must be shifted into 24-bit subwords. Each subword can then be multiplied by each byte of the “B” operand. This produces partial products which can then be XORed together to form a complete product, which still needs to be reduced.

Creation of the partial products, shifting, and accumulating them, is the most computationally intensive part of the $GF(2^m)$ modular multiplication. Given the C64x+’s parallelized architecture, this series of instructions can be computed quickly. Every cycle, the DSP can perform two XORMPYs, two shifts along with four XOR operations. Figure 2 shows the resulting execution diagram. In this figure, the rows represent cycles and the columns show concurrent operations. While only two XORMPYs can be dispatched every clock cycle, the datapath is internally pipelined, so that up to 8 XORMPYs can concurrently execute. Our $GF(2^m)$ fully exploits the parallelism offered by the DSP. The final step is the modular reduction, illustrated next.

D. Modular Reduction on the TI C64x+ DSP

After the full polynomial product is formed, it must be reduced into $GF(2^m)$. For the standard curves we studied, the unbalanced exponent modular reduction is applicable. However, depending on the specific form of $T(x)$, we use two different methods for implementing the unbalanced exponent reduction. Figure 3 illustrates both cases. If three of four or more terms in $T(x)$ are within 9 bits of each other, then it is beneficial to use XORMPYs, shifts and XORs to reduce the product. Otherwise, it is more efficient to shift and add for reduction. If most of terms in $T(x)$ are within a 9-bit window, then these terms can fit in the 9-bit operand of XORMPY. Therefore, XORMPY can be used by multiplying $T(x)$ by $C_h(x)$ using 24-bits at a time. For example, in the case of $GF(2^{163})$, the irreducible polynomial is: $f(x) = x^{163} + x^7 + x^6 + x^3 + x^0$. So $T(x) = x^7 + x^6 + x^3 + x^0$ or 0xC9. Performing the first iteration of reduction requires $T(x)$ be multiplied with seven 24-bit words. The next and final iteration of the reduction requires $T(x)$ be multiplied with one 24-bit word. An example that illustrates the other case is...
reduction in $GF(2^{233})$. In this case, $f(x) = x^{233} + x^{74} + x^0$. A product in this case is 465 bits long. To reduce, we shift all bits above bit 232 to the right 233, then XOR them with the bits occupying bits 232-0. We XOR that result with $C_h(x)$ shifted to the right 233-74 = 159 bits. Since the result extends out over 233 bits, these steps must be repeated once more.

E. Performance Evaluation

We implemented the binary field modular multiplication described above for five different field sizes on an OMAP processor. We mapped the proposed algorithms in C. We compiled this code with the TI C6x C compiler under high optimization (-O3). By inspecting the assembly code generated by the compiler, and carefully rewriting the C code, we were able to achieve a parallel schedule exactly as shown in Figure 2. Our prototyping platform is a Beagle Board (http://beagleboard.org); the DSP processor executes at 360 MHz while the ARM processor executes at 500 MHz.

We compare the results of the DSP to those achievable with the optimized openSSL cryptographic library for ARM. Table II shows the execution times and cycle counts, achieved for the DSP and the ARM. The table shows both binary field and prime field implementations. The DSP implementation for the prime field follows the design explained above, while the DSP implementation for the prime field used the library developed by Yan et al in [1].

The DSP implementations outperform the ARM implementations significantly, even though the DSP runs at a lower clock frequency. This is clearly the result of instruction-level parallelism. The binary-field implementations show a higher speedup than the prime-field implementations. This is because the ARM does not have an XORMPY instruction, and hence has to emulate binary-field multiplication with shifts and XORs.

Our strategy to implement unbalanced exponent reductions is demonstrated in Table III. Using the shift-and-add method works better for reductions in $GF(2^{233})$ and $GF(2^{409})$, while XORMPY works better for reductions in $GF(2^{163}$ and $GF(2^{283})$. This supports the claim that if most set bits in the irreducible polynomial fit in a 9-bit window, it is more effective to reduce using XORMPYs.

![Fig. 3. Unbalanced Exponent Modular Reduction. Left, using XORMPY. Right, using shift-XOR.](image-url)

### Table II

<table>
<thead>
<tr>
<th>Field Size</th>
<th>ARM A8 500MHz Time (µs)</th>
<th>C64x+ DSP 360MHz Time (µs)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>$GF(2^{233})$</td>
<td>17.4 8700 2.35 846 7.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GF(2^{283})$</td>
<td>30.4 15200 3.29 1184 9.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GF(P160)$ [1]</td>
<td>3.34 1670 0.83 299 4.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GF(P224)$ [1]</td>
<td>7.81 3905 1.57 565 4.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table III

<table>
<thead>
<tr>
<th>Field Size</th>
<th>Irreducible Polynomial</th>
<th>Shift-Add Method (Cycles)</th>
<th>XORMPY Method (Cycles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$GF(2^{163})$</td>
<td>$x^{163} + x^7 + x^0 + x^7 + x^0$</td>
<td>28 19</td>
<td></td>
</tr>
<tr>
<td>$GF(2^{233})$</td>
<td>$x^{233} + x^{74} + x^0$</td>
<td>33 42</td>
<td></td>
</tr>
<tr>
<td>$GF(2^{283})$</td>
<td>$x^{283} + x^{12} + x^2 + x^5 + x^0$</td>
<td>90 64</td>
<td></td>
</tr>
<tr>
<td>$GF(2^{409})$</td>
<td>$x^{409} + x^{37} + x^0$</td>
<td>52 66</td>
<td></td>
</tr>
</tbody>
</table>

### IV. Elliptic-Curve System Integration

We next discuss how to integrate this fast modular arithmetic into Elliptic Curve Cryptography, executing on a heterogeneous platform that includes the ARM as well as the DSP. We first introduce the target platform, and then discuss the proposed partitioning.

A. TI OMAP

We use the TI OMAP processor as a driving example, but we note that many similar architectures exist. OMAP is popular among cell-phone manufactures such as Nokia and Samsung, and has also found its way into internet tablets and smartphones. Though many variations of the device exist, most versions combine an ARM CPU with a DSP, packaged together with a myriad of other peripherals [10].

Figure 4 shows the organization of the relevant subsystems. The central part of OMAP3530 is the microprocessor with the ARM Cortex-A8 core. This subsystem itself includes multiple submodules: processor’s level 1 and level 2 caches, interrupt controller, and bridge modules to the main bus referred to as L3 interconnect. The C64x+ DSP is located in a separate subsystem, with its own caches, interrupt controller, and bus bridge/memory management unit. The L3 Interconnect bus is implemented using Open-core protocol (OCP), and connects the ARM, DSP, SDRAM controller and other subsystems together. There is also an additional bus referred to as the L4 Interconnect to connect additional peripherals.

To facilitate interprocessor communication the OMAP3530 chip includes an interprocessor communication (IPC) module attached to the L4 Interconnect. The IPC consists of a memory-mapped mailbox which contains small hardware FIFOs and connections to the interrupt controllers of both ARM and DSP subsystems. These FIFOs store 4 messages of 32-bits in size, and are typically used to exchange memory addresses. The actual sharing of data is done through a shared memory, typically a predefined section on the RAM on the L3 bus.
The above implies that interprocessor communication between the ARM and DSP carries significant overhead. A typical sequence requires that the sending processor prepares the shared memory region, resolves cache coherence issues, transmits a message to the IPC queue, and triggers an interrupt for the reading processor. The reading processor has to either be polling the IPC queues or process the interrupt, translate the transmitted addresses to its own memory mapping, and read the data as necessary. Table I shows the overhead as measured using two communication libraries provided by TI. The fourth column shows the number of ARM processor cycles for the round-trip transfer of a byte. We note that there is a strong dependency on the message size, and that communication overhead is over a hundred clock cycles per byte, even when optimized libraries are used. This observation will strongly influence the partitioning of ECC functionality over the ARM and the DSP.

B. ECC partitioning

Since the ARM in the OMAP has a central control function, it is natural to implement ECC as a service in terms of the system control software on this ARM. The ARM will thus initiate the ECC point multiplication, and spawn off computations to the DSP as needed.

In ECC, a distinction must be made between the low-level modular arithmetic in terms of binary-field and prime-field operations, and the point operations, which implement the scalar multiplication of EC points. A straightforward partitioning would map the modular arithmetic onto the DSP, and keep the point operations on the ARM. However, given the communication-overhead numbers cited earlier, this can be an unfavorable choice. Even for the largest field size, a field multiplication is only of the order of 30 microseconds, while the round-trip time to the DSP is at least 150 microseconds. Hence, the speedup provided by the DSP in executing modular arithmetic is completely lost in interprocessor communication.

We therefore migrated the entire point multiplication to the DSP. The ARM only handles the EC domain parameters, communicates the EC point and the scalar to the DSP, and retrieves the result of the point multiplication. An additional argument in favor of this partitioning is that the ARM is free for other tasks while the DSP is executing the point multiplication.

In the following section, we discuss the performance of this design as compared to a design that only uses the ARM.

V. RESULTS

We implemented all of the designs discussed so far on TI’s Beagle Board. This board provides accurate performance evaluation, and also enabled us to obtain power- and energy-measurements. We first discuss the characteristics of the select EC point multiplications, then present performance evaluation, and finally provide analysis of power and energy consumption.

A. Selected Elliptic Curve Designs

Table V lists our implementation choices for binary-field and prime-field ECC implementation. Projective coordinates are used to reduce the amount of modular arithmetic required during point-multiplication. As mentioned earlier, the prime-field ECC implementation is ported from the work reported in [1].

An important objective of this work was to provide a fair comparison for elliptic-curve cryptography between a single-core ARM design, and a DSP-accelerated design. We use the implementation used in the popular OpenSSL crypto library as the baseline ARM implementation. By default, the algorithms in the library differ slightly from the variants listed in Table V. However, we made sure that the point multiplication on the ARM (OpenSSL) and the ARM+DSP (this paper) was identical in terms of computational complexity.

B. Performance

Table VI shows the resulting execution time of a point multiplication for different field sizes for a design using only the ARM versus a design using the ARM with the DSP. In the case of the DSP accelerated design, we measure the time from just prior to sending the operands to DSP, to right after the ARM receives the reply containing the computed point. In the case of OpenSSL ARM implementation, we measure the time for the point multiplication function to return.

The results indicate that our ARM-DSP implementation is able to preserve a significant part of the speed-up gained through accelerated field arithmetic, although it is less than
the ideal speedup achievable for field multiplications on the DSP. This is explained as follows. When a complex EC Point multiplication executes on a DSP, there is a certain amount of utility tasks, during which the DSP datapath cannot be fully utilized. Such additional tasks include fetching operands from memory, and creating and zeroing out temporary variables. Hence, the utilization of the DSP over the entire EC point multiplication is lower than when doing only modular multiplications. This observation is supported by the fact that a larger performance drop is seen for larger field sizes.

### C. Power Efficiency

The Beagle Board’s PCB design includes a test point for power measurement. Two exposed pins enable measurement of the voltage drop across a 0.3 ohm resistor. The resistor is located right between the board’s power supply and the primary power regulator chip TPS65950. We measured the power by bringing out a trigger signal to one of the input/output pins on the PCB, and using this trigger to capture the voltage drop across the resistor with an oscilloscope. By computing the current through the resistor and multiplying that current by the power supply voltage we were able to get a precise estimate of the power draw by the entire board. We motivate the use of the voltage drop across a 0.3 ohm resistor as the power measurement. Two exposed pins enable measurement of the power draw by the entire board. We motivate the use of the Beagle Board as a test platform rather than the OMAP 3530 chip.

### VI. Conclusion

Our objective in this research was to study heterogeneous multicore ECC design on a current-generation SoC platform. Our results show that there are clear and tangible performance and power efficiency benefits to heterogeneous design even after all the overhead and tradeoffs are taken into account. Our research shows that it is worthwhile exploring complex SoC platforms for cryptographic acceleration.

### ACKNOWLEDGMENT

We thank the Security and Architecture lab at University of Connecticut for their input on the implementation of prime-field arithmetic on the OMAP 3530.

### REFERENCES


