Implementation of a pipelined modular multiplier architecture for GF(p) elliptic curve cryptography computation

ADNAN ABDUL-AZIZ GUTUB*, ABDUL-RAHMAN M. EL-SHAFEI** AND MOHAMMED A. AABED**

*Center of Research Excellence in Hajj and Omrah, Umm Al-Qura University, Makkah, Saudi Arabia. Associate Researcher, Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh, Saudi Arabia. Email: aagutub@uqu.edu.sa

**Computer Engineering Department, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia

Abstract

This paper reflects the achievement of improved results in terms of cost for Pipelined Crypto Modular Multiplier Architecture when compared with its earlier versions of Parallel Crypto Architecture. The improved pipelined modular multiplier is implemented on Field Programmable Gate Array (FPGA), designed in four stages to be properly suitable for elliptic curve crypto computation. To avoid inversion complexity, the elliptic computations arithmetic utilizes projective coordinates instead of the normal affine coordinates. We adjusted the elliptic curve crypto addition operation with efficient scheduling for this pipelining.

The proposed hardware is compared to the previous parallel (non-pipelined) models that were similarly designed. All considered architectures have been synthesized for 160-bits operations showing interesting features. Detailed results showed that this work gave overall efficiency in the cost, which shows a promising direction for further research.

Keywords: Cryptography hardware (Crypto Hardware); Elliptic curve cryptography (ECC); Parallel crypto architecture (Parallel Crypto Arch.); Pipelined crypto architecture (Pipelined Crypto Arch.); Projective coordinate cryptosystems (Proj. Crypto Sys.)

Introduction

Encryption used to be considered a very mysterious subject. It seemed to have no important practical applications outside the government/military world and was
considered a harmless mathematical diversion if anyone else should pursue it. However, in today's world where everything we do is recorded on computers and all computers are linked to the Internet (and thus to each other), we find that encryption is of vital importance in many areas, such as the protection of privacy and confidentiality, transmission of secure information (e.g. credit card details), and to provide authentication of the sender of a message or even authenticate the time that message was sent. At the same time, law enforcement authorities worry that if everyone routinely kept their personal records encrypted, evidence held under a search warrant would be unusable. For this reason, they have tried to limit the ability of the general public to use strong (unbreakable) encryption. Therefore, the study of information security and cryptosystems has become an urgent need and a rich field of research and contribution. To cope with these challenges, one of the key areas has been the focus of research in the last two decades and is addressed in this work, Elliptic Curve Cryptography (ECC).

The study of elliptic curves by algebraists, algebraic geometers and number theorists dates back to the middle of the nineteenth century. The existence of an extensive literature describes the beautiful and elegant properties of these marvelous objects. In 1984, Hendrik Lenstra described an ingenious algorithm for factoring integers that relies on properties of elliptic curves (ECs). This discovery prompted researchers to investigate other applications of ECs in cryptography and computational number theory.

Public-key cryptography was conceived in 1976 by Diffie and Hellman. The first practical realization followed in 1977 when Rivest, Shamir and Adleman proposed their now well-known RSA cryptosystem, in which security is based on the intractability of the integer factorization problem. ECC was discovered for crypto usage in 1985 by Neal Koblitz and Victor Miller. Elliptic curve cryptographic schemes are public-key mechanisms that provide the same functionality as RSA schemes but with less computation. Over the years, researchers have developed techniques for designing and proving the security of RSA, Discrete Logarithm (DL) and EC protocols under reasonable assumptions. The fundamental security issue that remains is the hardness of the underlying mathematical problem that is necessary for the security of all protocols in a public-key family: the integer factorization problem for RSA systems, the discrete logarithm problem for DL systems, and the elliptic curve discrete logarithm problem for
ECC systems. The perceived difficulty of these problems directly impacts performance, since it dictates the sizes of the domain and key parameters. That in turn affects the performance of the underlying arithmetic operations. The security of these schemes is based on the difficulty of a different problem, namely the elliptic curve discrete logarithm problem (ECDLP). Currently the best algorithms known to solve the ECDLP have fully exponential running time, in contrast to the sub-exponential time algorithms known for the integer factorization problem. This means that a desired security level can be attained with significantly smaller keys in elliptic curve systems than is possible with their RSA counterparts. For example, it is generally accepted that a 160-bit elliptic curve key provides the same level of security as a 1024-bit RSA key. The advantages that can be gained from smaller key sizes include speed and efficient use of power, bandwidth, and storage.

ECC computations are performed in hardware as well as software, where hardware ECC computations have been proved to be faster and more efficient (Gutub 2006). This work proposes improving ECC hardware by pipelining its modular multiplier that is implemented on FPGA. We tune the used modular multiplier as a 4-stage pipeline suitable for elliptic curve crypto computation. To avoid inversion complexity, the elliptic computations arithmetic utilizes projective coordinates instead of the normal affine coordinates as proved by Gutub (2006), i.e. the proposed architecture considers representing the elliptic curve points as projective coordinate points in order to reduce the number of all inversion operations to one. We also adjusted the elliptic curve crypto addition operation with efficient scheduling for correct pipelining. Our proposed hardware is compared to parallel (non-pipelined) models designed very similarly to ensure fair comparisons and conclusions.

The flow of the paper will be as follows: first we provide some background behind the security of ECC and elliptic curve crypto computations. This section will also provide an example of simple procedure for using ECC for encryption and decryption. Several available techniques for implementing ECC at high-speed are explored in the section to follow. All ECC design schemes are to sustain the high throughput required by applications, where hardware-based designs are shown as the solution reaching acceptable performance-cost ratio. Section 4 describes the ECC scalar multiplication
concept and algorithm, which is detailed and reconsidered within this work as coordinate systems (Section 5). Section 6 covers designing the hardware components used with the modules data flow that is improved for this architecture. The proposed pipelined design is detailed in Section 7. Section 8 elaborates in the pipelining stages derivation proofing the accuracy of the proposed design. The hardware implementation and simulation results are given in Section 9 followed by the comparison and analysis remarks (Section 10). Finally, Section 11 concludes the work.

**Way to ECC**

This section provides background behind the security of elliptic curve crypto computations. It also provides an example of a simple procedure for using ECC for encryption and decryption.

The cryptographic protocols generally serve very similar objectives and are based on almost same principles (Blake *et al.*, 1999). They contain a function which, by means of a parameter called the encryption key, can be easily computed. The inverse of this function is hard to compute unless a trapdoor function (a second key corresponding to the former one) is known. A general assumption made during the analysis of the security of a system is that all information about the system except the trapdoor key are known by the adversary. The previously mentioned group of public and private key systems is based on the way these keys are generated and kept.

In a private key system, encryption and decryption are performed using the same key which should be kept secret, otherwise the system is broken. In public key cryptosystems, encryption and decryption are done using two different keys. One of the keys is published and the other is kept secret. When one party is going to sign a message, the encryption key is kept secret but the key to verify the signature will be published. On the other hand, when a secret message is to be sent, the encryption key will be published while the key to open the message will be kept secret by the owner. For example, when messages sent to a user are encrypted by his public key, he is the only person who has access to the corresponding private key and can decrypt the message. There are several types of public key cryptosystems. A major group of these systems is based on the
difficulty of solving the discrete logarithm problem, which is the basic security feature behind Elliptic Curve Cryptography (ECC).

ECC bases its security on the difficulty in solving the discrete logarithm problem (DLP). ECC’s main operation can be simplified to the sum of elliptic curve points, i.e. the ECC operation is to compute point $Q$ given point $P$, where $Q=kP$ and this $kP$ is interpreted as the sum of points $P+P+.....+P+P$ ($k$-times). Here, the points $P$ and $Q$ should satisfy an elliptic curve equation defined on a finite field (a finite set of elements that satisfies some axioms). The logarithm discrete problem is now defined: Given the points $P$ and $Q$, find the integer $k$ such that $Q=kP$. This point summation operation ($kP$) is called scalar multiplication. The difficulty in solving this conjectured problem increases as the number of elements (valid points) increase.

There are many ways to apply elliptic curves for encryption/decryption purposes. In its most basic form, users randomly chose a base point $(x,y)$ lying on the elliptic curve $E$. The plaintext (the original message to be encrypted) is coded into an elliptic curve point $(x_m,y_m)$. Each user selects a private key '$n$' and computes his public key $P=n(x,y)$. For example, user A’s private key is $n_A$ and his public key is $P_A=n_A(x,y)$. For anyone to encrypt and send the message point $(x_m,y_m)$ to user A, he/she needs to choose a random integer $k$ and generate the ciphertext $C_m=\{k(x,y),(x_m,y_m)+kP_A\}$. The ciphertext pair of points uses A’s public key, wherein only user A can decrypt the plaintext using his private key. To decrypt the ciphertext $C_m$, the first point in the pair $C_m\ k(x,y)$ is multiplied by A’s private key to get the point: $n_A(k(x,y))$. When this point is subtracted from the second point of $C_m$, the result will be the plaintext point $(x_m,y_m)$. The complete decryption operations are $[(x_m,y_m)+kP_A]-A[k(x,y)]= (x_m,y_m)+k[n_A(x,y)]- n_A[k(x,y)]= (x_m,y_m)$.

The most time consuming operation in the encryption and decryption procedure is finding the multiples of the base point $(x,y)$. The approach used to implement this is discussed in the Section 4, after covering the literature review of Section 3.

**Literature review**

Elliptic Curve Cryptography (ECC) is gaining increased research acceptance and has lately been the subject of several standards (Tawalbeh et al., 2010). This interest is mainly due to the high level of security with relatively small keys provided by ECC. To
sustain the high throughput required by applications such as network servers, high speed implementations of public-key cryptosystems are needed. For that purpose, hardware-based processors or accelerators are often the optimum solution for reaching an acceptable performance-cost ratio. The fundamental question that arises is how to choose the appropriate efficiency–flexibility tradeoff. In this section, techniques for implementing ECC at a high speed are explored.

Some studies deal with architectures using only one field size and one hardwired irreducible polynomial (or modulo). This polynomial is of course inconsistent through reconfiguration of reconfigurable platforms. A fully hardwired asynchronous Applicaion-Specific Integrated Circuit (ASIC) design is presented by K. Ja'rvinen et al. (2004). It is based on a special Optimal Normal Basis (ONB) multiplier. Another similar ASIC implementation, based on synchronous polynomial basis multiplier, can be found in (Sozzani et al., 2005). Designs for curves defined over GF(p) finite fields are explained in Batina et al., (2004), McIvor et al. (2004), Orlando & Paar, (2001) and Ors et al.(2003), where all use Montgomery multiplications. Three architectures (Bantina et al., 2004; Orlando & Paar, 2001; Ors et al., 2003) use systolic array Montgomery modular multipliers while the design in McIvor et al., (2004) uses Montgomery multiplication based on parallel schoolbook multipliers. A systolic array Montgomery modular multiplier is also modeled by N. Mentens et al. (2004) but for GF(2^m) curves. Other research work using reconfigurable hardware can be found in Bajracharya et al. (2004) and Nguyen et al., (2003). They focus on an efficient hardware/software partition for the scalar multiplication. Reconfigurable architectures based on digit-serial polynomial basis multiplier can be found in Ansari & Hasan (2006), Lutz & Hasan (2004), Orlando & Paar (2000) and Shu et al., (2005). The research by Ansari & Hasan (2006) also mentions ASIC results performing the fastest implementation. Saqib (2004) presented utilizing Hessian elliptic curves from Smart (2001) but they were used in order to extract more parallelism. The same design using the DLP is presented by Saqib et al. (2004). Nevertheless, the low frequency of this design shows that large multipliers must be handled with special care. More about Karatsuba multipliers can be found in Dyka & Langendoerfer (2005), Gathen et al., (2005) and Rodríguez-Henríquez & Koç (2003). Considering several designs, the hardware shown by Ansari & Hasan (2006) seems to be
the best representative of a high-speed architecture using a hardwired irreducible polynomial, a single field size, and an unknown base point $P$. It is a pipelined and parallel structure based on a large Most Significant Digit (MSD)-first multiplier and an accurate scheduling of the DLP algorithm.

The Sakiyama et al. (2007) work presents a reconfigurable curve-based crypto-processor that accelerates scalar multiplication of ECC and Hyper Elliptic Curve Cryptography (HECC) of genus 2 over $\text{GF}(2^m)$. By allocating several copies (say $x$) of processing cores that embed reconfigurable Modular Arithmetic Logic Units (MALUs) over $\text{GF}(2^m)$, scalar multiplication of ECC/HECC can be accelerated by exploiting Instruction-Level Parallelism (ILP). The supported field size can be arbitrary up to a limit related to the number of cores, i.e. up to $x(n+1)-1$. The super-scaling feature is facilitated by defining a single instruction that can be used for all field operations and point/divisor operations. Chen et al. (2007) presented a high-performance elliptic curve cryptographic processor for GF(p) general curves which featured a systolic arithmetic unit. The work proposed a new unified systolic array that efficiently implements addition, subtraction, multiplication and division. At the system level, the control dependencies in the operation sequence and the mismatched communication between the systolic array and the separate storage would stall the pipeline in the systolic array. The design avoided these pipeline stalls successfully using optimization methods. The processor is synthesized in 0.13-micron standard-cell technology. It required 1.01-ms to compute a 256-bit scalar multiplication for general curves over GF(p).

Kumar et al. (2006) showed different architectural enhancements in a Least Significant Digit (LSD) multiplier for binary fields $\text{GF}(2^m)$. They proposed two different architectures: the Double Accumulator Multiplier (DAM), and N-Accumulator Multiplier (NAM), which are faster than traditional LSD multipliers. The evaluation of the multipliers for different digit sizes gave best choices and showed that currently used digit sizes are not efficient. The paper suggested that one should always use the NAM architecture to get the best timings. Considering the time area product, DAM or NAM gave the best performance depending on the digit size.

A GF(p) processor suitable for RSA, DSA and ECDSA is presented by K. Itoh et al. (1999). It is based on a Montgomery multiplier with a 16-bit digit size, implemented
in a TMS320C6201 DSP (digital signal processor) hardware. Another processor using MAC2424 (24x24-bit) for GF(p$^m$) OEF is presented by Chung et al. (2000). It includes an interesting analysis of the pre-computation needed by the signed window scalar multiplication method. More about GF(p$^m$) arithmetic can be found in Bertoni et al. (2003) and Bajard et al. (2003). Two 64-bit dual-field processors can be found in Eberle et al. (2004) and Satoh & Takano (2003). They also use Montgomery multipliers and are both able to compute integer and binary polynomial arithmetic (for GF(p) and GF(2$^m$)).

In particular, RSA, DSA, ECDSA and DH are supported. The architecture proposed in Satoh & Takano (2003) is an ASIC, while Eberle et al. (2004) is implemented on FPGA with ASIC extrapolations. The work in Eberle et al. described an architecture based on a 64-bit datapath with its core based on a 64x64-bit multiplier able to process both integer and binary polynomial arithmetic. The processor currently implements the main public-key protocols like RSA, DSA, ECDSA and DH with arbitrary key sizes and curves.

**Scalar multiplication**

The ECC algorithm used for calculating $nP$ from $P$ is based on the binary method. The algorithm used for scalar multiplication is based on the binary method (Gutub, 2006; Gutub & Ibrahim, 2003), since it is efficient for hardware implementation. The binary method algorithm is shown below:

---

**Inputs:** $k$: a constant, $P$: point on the elliptic curve  
**Output:** $Q$: another point on the elliptic curve, $Q = k \cdot P$

Define: $w$: number of bits in $k$, where $k_i$ is the $i^{th}$ bit in $k$

If $k_{w-1}=1$ then $Q := P$ else $Q := 0$;

for $i := w-2$ down to 0 do

$Q := Q + Q$; \hspace{1em} Point Doubling

If $k_{w-1}=1$ then $Q := Q + P$; \hspace{1em} Point Addition

Return $Q$;

---

Basically, the binary method algorithm scans the bits of $n$ and doubles the point $Q$ $k$-times. Whenever a particular bit of $n$ is found to be one, an extra operation is needed. This extra operation is $Q+P$. As a result, adding and doubling elliptic curve points are the most basic operations in each iteration. The point addition and doubling requires
performing inversion operations. It is known that these inversion operations are the most expensive operations since their cost is equivalent to $9\sim30$ modular multiplications (Gutub & Ibrahim, 2003); i.e., inversion is extremely slow, which made researchers generally try to avoid it if possible. The use of coordinate systems other than the Affine coordinate system (which will be illustrated later) greatly reduces the number of inversions required in the operations of the scalar multiplication on the expense of extra multiplications.

ECC effectively uses point doubling and addition operations in the arithmetic execution. The optimized formulae available for these operations eliminate the costly field inversion from the main loop of the scalar multiplication; i.e. fast operations are achieved by using projective coordinates (Gutub, 2006). However, the operation in projective coordinates involves more scalar multiplication than in affine coordinate and ECC on projective coordinates, and will be efficient only when the implementation of scalar multiplication is much faster than a multiplicative inverse operation. Therefore, transfer is needed from one coordinate to another for avoiding the inversion process cost. The following section is dedicated to illustration of the coordinate systems structure used for these purposes.

**Coordinate systems**

An elliptic curve can be represented by different coordinate systems. Following are descriptions of two coordinates, i.e. affine coordinate and standard projective coordinate procedures (Miyaji, 1992). This standard projective coordinate system is found appropriate for parallel implementation, as detailed in Gutub & Ibrahim (2003), which led to our choice for this proposed pipelined hardware.

**Affine coordinate:**

Let $E$ be an elliptic curve over $\text{GF}(p)$, has the following equation:

$$E: y^2 = x^3 + ax + b \pmod{p},$$

where $a$ and $b$ are constants satisfying $4a^3 + 27b^2 \neq 0 \pmod{p}$.

Let $P=(x_1,y_1)$, $Q=(x_2,y_2)$, and $P+Q=(x_3,y_3)$, be points of $E(\text{GF}(p))$.

1. **Addition formula:** $x_3 = \lambda^2 - x_1 - x_2$, $y_3 = \lambda(x_1 - x_3) - y_1$, where $\lambda = (y_2 - y_1)/(x_1 - x_2)$

2. **Doubling formula:** $x_3 = \lambda^2 - 2x_1$, $y_3 = \lambda(x_1 - x_3) - y_1$, where $\lambda = (3x_1^2 + a)/(2y_1)$
Addition time = 1 Inversion + 1 Squarings + 2 Multiplications + 6 Subtractions
Doubling time = 1 Inversion + 2 Squarings + 2 Multiplications + 1 Addition + 3 Subtractions

Standard projective coordinate:

For standard projective coordinates, we set \( x = X/Y \) and \( y = Y/Z \), giving the equation:

\[
E: Y^2Z = X^3 + aXZ^2 + bZ^3 \pmod{p},
\]

Let \( P = (X_1, Y_1, Z_1) \), \( Q = (X_2, Y_2, Z_2) \) and \( P + Q = (X_3, Y_3, Z_3) \) be points of \( E(\text{GF}(p)) \),

1. Addition formula: \( X_3 = vA \), \( Y_3 = u(v^2X_1Z_2 - A) - v^3Y_1Z_2 \), \( Z_3 = v^3Z_1Z_2 \)
where, \( u = Y_2Z_1 - Y_1Z_2 \), \( v = X_2Z_1 - X_1Z_2 \), \( A = u^2Z_1Z_2 - v^3 - 2v^2Y_1Z_2 \)

2. Doubling formula: \( X_3 = 2hs \), \( Y_3 = w(4B-h) - 8s^2Y_1^2 \), \( Z_3 = 8s^3 \)
where, \( w = aZ_1^2 + 3X_1^2 \), \( s = Y_1Z_1 \), \( B = X_1Y_1s \), \( h = w^2 - 8B \)

Addition time = 12 Multiplications + 4 Subtractions
Doubling time = 7 Multiplications + 1 Addition + 1 Subtraction

The work in Gutub & Ibrahim (2003) proposed running these ECC point operations with projective coordinates on hardware with parallel multipliers. It was found that four parallel multipliers would give the maximum parallelization with the least number of multiplication steps as proven in Tawalbeh et al. (2010). We, in this work, are concentrating on the implementation of the ECC point addition operation for comparison purposes. Figure 1 shows a rearranged ECC point addition data flow graph assuming 4 parallel multipliers. The computation through this data flow graph (Figure 1) requires 4 steps of modular multipliers and 4 steps of modular adders.
Related hardware components

This section introduces the designs and algorithms considered for studying our pipelined hardware. We give a brief idea about modular addition first, followed by modular multiplication. Figure 2 shows the design implemented for modular addition operations in this model, extracted from the previous work by Gutub (2006).

![Diagram of modular addition](image)

**Fig. 2:** Non-pipelined modular adder.
To obtain a pipelined implementation of this adder, it has been divided into two stages, as shown in Figure 3. The Ripple Carry Adders (RCAs) are split into two modules, each of which is a separate stage. The two RCAs in Figure 3 are divided into two stages using latches as shown in Figure 4. Note that this block diagram is a simplified example of a 6-bit operation.

As was mentioned earlier, the multiplication process is the most sophisticated and time consuming process in the ECC systems. Thus, optimizing the multiplier design and delay is a fundamental requirement for system efficiency. The straightforward approach to compute modular multiplication is by performing multiplication followed by reduction (Gutub, 2007). The multiplication can be computed through several addition operations. Then, the reduction is performed through several subtractions, by subtracting the
modulus several times, until the result is less than the modulus. This approach is inefficient and suffers from very low speed. It can, however, be improved by merging modulo subtraction with the multiplication-add operations (Gutub, 2007), as in the algorithm below.

Define \( k \): number of bits in \( x \); \( x_i \): the \( i^{th} \) bit of \( x \)

Input: \( x, y, \) and \( n; \) where \( x,y < n; \)

Output: \( P = xy \mod n \)

1. \( P := 0; \)
2. For \( i = k - 1 \) down to 0;
3. \{ 
4. \( P := 2P; \)
5. \( \) If \( P \geq n \) Then \( P := P - n; \)
6. \( \) If \( x_i = 1 \) Then
7. \( \{ P := P + y; \)
8. \( \) If \( P \geq n \) Then \( P := P - n \}; \)
9. \}
10. \} End;

The algorithm above is for GF(p) modulo multiplication and found to be very appropriate for hardware implementation (Gutub, 2007). It has a bounding ‘for’ loop, which includes iterative modulo multiplication reduction operations. The bounding loop can be designed in hardware as a controller that will control the number and processes of the iterations. The modulo multiplication reduction is implemented in hardware with three adders and three multiplexers connected, as shown in Figure 5. There are no registers in the hardware design; the small boxes shown are symbols to clarify the mapping of bit-flow. The adder can function as a subtractor by inverting one of its inputs. The complete process of \( x.y \mod n \) will need \( k \) clock cycles, if each modulo reduction iteration is performed in one clock cycle. The multiplication of \( P \) by two (as in step 4 of the algorithm above) is performed by a shift to the bits of \( P \) toward the left. The multiplexers Mux-1 and Mux-3 are controlled by the subtractor’s output-carry-bit. Therefore, the complete subtractions are to be made for the Mux to give the output. The reader is referred to Gutub (2007) for more details.
Proposed pipelined multiplier design

We propose improving the previous ECC design in Figure 6, detailed by Gutub & Ibrahim (2003), into a new 4-stage pipelined modular multiplication approach as shown in Figure 7. The design is similar in principle to our previous pipelined design in Gutub (2006); however, its architecture components are designed based on pipelining the standard projective coordinates which makes it an improved hardware with multipliers of four new stages. Each stage contains a different modular multiplication operation. Each multiplication operation loops through all stages $k$-times.

Fig. 5: Modular non-pipelined multiplier.

Fig. 6: Previous parallel architecture.
The modular adder consists of four \( k/2 \) digit carry-propagate adder and three \((k/2+1)\) digits carry-propagate adder. The modular multiplication can accommodate a maximum of 4 different multiplication stages (Figure 8), where each multiplication can be processed independently from other multiplications timings. However, new modular multiplication does not start looping through the pipeline until stage 1 is free. The output is delivered from stage 4 after \( 4k \) cycles, as shows in Figure 8.

It should be noted that the control box is used to select when a new modular multiplication operation will be inserted. It also controls the looping of existing operations to complete \( 4k \) cycles. The control box will allow a new multiplication operation if no current multiplication operation occupies this stage. If all stages are busy,
the inputs will be stalled. Figure 9 shows the design implemented for the pipelined modular multiplication operations in our proposed model.

![Modular pipelined multiplier](image)

Figure 9 shows the design implemented for the pipelined modular multiplication operations in our proposed model.

**Figure 9**: Modular pipelined multiplier.

Figure 9 shows the four stages and where each stage needs to finalize its outputs for the next stage to start its operation accurately. Stage 1 processes the inputs at the first cycle; it contains two adders connected through multiplexers to direct the data correctly, i.e. either from as new inputs or as fed back from Stage 4 as an intermediate result. Note that Stages 2, 3, and 4 are also made up of adders that work together as a pipeline; they are following the k-times cycle shown in Figure 8 until the output is ready and to be taken from connection Z at Stage 4.

**Derivation of pipelined point addition framework**

The procedure for standard projective coordinates point addition can be defined corresponding to the registers as detailed in the following register sequence:

\[
\begin{align*}
R_1 &= Y_1 Z_2 \\
R_2 &= Y_2 Z_1 \\
R_3 &= X_1 Z_2 \\
R_4 &= Z_1 X_2 \\
R_5 &= R_2 - R_1 \\
R_6 &= R_3 + R_4 \\
R_7 &= R_4 - R_3 \\
R_8 &= R_7 Z_2 \\
R_9 &= R_5 R_3 \\
R_{10} &= R_5 R_7 \\
R_{11} &= Z_1 Z_2 \\
R_{12} &= R_4 R_{10} \\
R_{13} &= R_3 R_{10} \\
R_{14} &= R_1 R_9 \\
R_{15} &= R_2 R_{10} \\
R_{16} &= R_3 - R_{15} \\
R_{17} &= R_{12} - R_{16} \\
R_{18} &= Y_1 R_{13} \\
R_{19} &= Z_1 R_{13} \\
R_{20} &= R_5 R_{17} \\
R_{21} &= X_3 = R_10 R_7 \\
R_{22} &= Y_3 = R_{20} - R_{18}
\end{align*}
\]
The above procedure is rescheduled to avoid any dependency. It considers proper arrangement for trying to fully utilize all of the 4-stage modular pipelined multiplier (Figure 9). This rescheduling is shown in Table 1, which notes the number of clock cycles in an accumulation manner.

<table>
<thead>
<tr>
<th>Clock Cycle Accumulation</th>
<th>Modular Multiplication</th>
<th>Modular Adder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>0</td>
<td>R₁</td>
<td></td>
</tr>
<tr>
<td>1N</td>
<td>R₂</td>
<td>R₁</td>
</tr>
<tr>
<td>2N</td>
<td>R₃</td>
<td>R₂</td>
</tr>
<tr>
<td>3N</td>
<td>R₄</td>
<td>R₃</td>
</tr>
<tr>
<td>4N</td>
<td>R₁₁</td>
<td>R₄</td>
</tr>
<tr>
<td>4N+1</td>
<td>R₄</td>
<td>R₃</td>
</tr>
<tr>
<td>4N+2</td>
<td>R₁₁</td>
<td>R₄</td>
</tr>
<tr>
<td>4N+3</td>
<td>R₉</td>
<td>R₁₁</td>
</tr>
<tr>
<td>4N+4</td>
<td>R₁₁</td>
<td>R₉</td>
</tr>
<tr>
<td>4N+5</td>
<td>R₁₀</td>
<td>R₁₁</td>
</tr>
<tr>
<td>8N</td>
<td>R₉</td>
<td>R₁₀</td>
</tr>
<tr>
<td>8N+3</td>
<td>R₁₄</td>
<td>R₉</td>
</tr>
<tr>
<td>8N+5</td>
<td>R₁₅</td>
<td>R₁₄</td>
</tr>
<tr>
<td>8N+6</td>
<td>R₁₂</td>
<td>R₁₅</td>
</tr>
<tr>
<td>8N+7</td>
<td>R₁₄</td>
<td>R₁₂</td>
</tr>
<tr>
<td>8N+8</td>
<td>R₁₃</td>
<td>R₁₄</td>
</tr>
<tr>
<td>12N+3</td>
<td>R₁₂</td>
<td>R₁₅</td>
</tr>
<tr>
<td>12N+5</td>
<td>R₁₃</td>
<td>R₁₂</td>
</tr>
<tr>
<td>12N+6</td>
<td>R₁₃</td>
<td>R₁₂</td>
</tr>
<tr>
<td>12N+7</td>
<td>R₁₃</td>
<td>R₁₂</td>
</tr>
<tr>
<td>12N+8</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
<tr>
<td>12N+9</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
<tr>
<td>12N+10</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
<tr>
<td>16N+7</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
<tr>
<td>16N+8</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
<tr>
<td>16N+9</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
<tr>
<td>16N+10</td>
<td>R₁₉</td>
<td>R₁₂</td>
</tr>
</tbody>
</table>

In fact, this gives the clear estimation of the total number of clock cycles needed for an ECC point addition operation as \(16N+10\), where \(N\) is the size of the modulus in bits.

**Hardware implementations and simulations**

The purpose of the hardware implementation is to give some common platform and fair comparison between our proposed pipelined architecture and similar previous designs.
The focus in this study is not targeted toward industrial purposes. It does not give the
details of the architecture implementation; instead, the aim is to extract the hardware time
and area parameters of the main blocks to build a fair comparison study between the
designs. Therefore, our implementation exploration here is going to be limited to the
level needed to serve this comparison goal. We will implement the basic blocks of
hardware that are commonly used to build all studied designs, i.e. our model here as well
as similar previous architectures. The major common components needed by all designs
are the modular multiplier and modular adder.

For simulation, we have used Silos Compiler (free Verilog compiler) to generate the
results. A sample of the output results are shown for each design implemented in the
project. For synthesis, we used the Virtex-4 XC4VSX35 FPGA library of Xilinx IES
software. Unfortunately, this available FPGA platform does not create an Area report,
unlike synthesis using ASIC libraries. Note that the ASIC area report advantage gives an
accurate estimate of the space requirements in micrometer\(^2\) and number of gates needed.
However, as mentioned before, this FPGA implementation cannot be precise for
industrial usages; it is mainly a practical tool for fair comparison and academic study. We
used it for comparing pipelined and non-pipelined versions of the ECC operations
hardware. In this section, we have included a brief summary of the synthesis results
(Table 2) followed by some output report-briefing subsections for each design.

<table>
<thead>
<tr>
<th></th>
<th>Modular Adder</th>
<th>Modular Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parallel</td>
<td>Pipelined</td>
</tr>
<tr>
<td>Time Path delay</td>
<td>8.36ns</td>
<td>7.4ns</td>
</tr>
<tr>
<td>Frequency</td>
<td>119.6Mhz</td>
<td>135.2MHz</td>
</tr>
<tr>
<td>FPGA Hardware</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slices</td>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td>Flip Flops</td>
<td>-</td>
<td>28</td>
</tr>
<tr>
<td>LUTs</td>
<td>32</td>
<td>48</td>
</tr>
<tr>
<td>IOBs</td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td>This study Hardware</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area Estimation Figure</td>
<td>82</td>
<td>141</td>
</tr>
</tbody>
</table>

*Parallel Design (Non-pipelined) Modular Adder*

- **Synthesis Results**
(a) Time:
The maximum combinational path delay is 8.359\textit{ns}. This is equivalent to having a maximum frequency of 119.6\textit{MHz}.

(b) FPGA Hardware:
- Number of Slices: 18 out of 15360 (0%).
- Number of 4 input LUTs: 32 out of 30720 (0%).
- Number of bonded IOBs: 32 out of 448 (7%).

Fig. 10: Non-Pipelined Modular Adder Block Diagram.

- **Macro Statistics**
  - Adders/Subtractors: 2
  - 8-bit adder: 1
  - 9-bit adder: 1
  - XORs: 8
  - 1-bit XOR3: 8

*Pipelined Modular Adder*

- **Synthesis Results**

(a) Time:

The minimum input arrival time before clock is $4.190\text{ns}$. The maximum output required time after clock is $7.399\text{ns}$. The maximum combinational path delay was not found. This is equivalent to having a maximum frequency of $135.2\text{MHz}$.

(b) FPGA Hardware:

- Number of Slices: 29 out of 15360 (0%)
- Number of Slice Flip Flops: 28 out of 30720 (0%)
- Number of 4 input LUTs: 48 out of 30720 (0%)
- Number of bonded IOBs: 36 out of 448 (8%)

Fig. 11: Pipelined Modular Adder Block Diagram.

- **Macro Statistics**
Adders/Subtractors: 4
- 4-bit adder carry in: 1
- 4-bit adder carry out: 2
- 5-bit adder carry in: 1
- Registers: 29
- Flip-Flops: 29
- XORs: 8
- 1-bit XOR3: 8

Parallel Design (Non-pipelined) Modular Multiplier

Fig. 12: Non-Pipelined Modular Multiplier Block Diagram.

- Synthesis Results
(a) Time:
The minimum period is 6.052\textit{ns} (Maximum Frequency: \textit{165.238MHz}). The minimum input arrival time before clock is 6.633\textit{ns}. The maximum output required time after clock is 9.784\textit{ns}. The maximum combinational path delay is 10.365\textit{ns}.

(b) FPGA Hardware:

- Number of Slices: 59 out of 15360 (0%)
- Number of Slice Flip Flops: 33 out of 30720 (0%)
- Number of 4 input LUTs: 105 out of 30720 (0%)
- Number of bonded IOBs: 36 out of 448 (8%)

\begin{itemize}
  \item \textbf{Macro Statistics}
  \begin{itemize}
    \item Adders/Subtractors: 4
    \item 5-bit subtractor: 1
    \item 8-bit adder: 1
    \item 9-bit adder: 2
    \item Counters: 1
    \item 5-bit up counter: 1
    \item Registers: 28
    \item Flip-Flops: 28
    \item Comparators: 1
    \item 5-bit comparator equal: 1
    \item Multiplexers: 1
    \item 8-bit 4-to-1 multiplexer: 1
    \item XORs: 8
    \item 1-bit XOR3: 8
  \end{itemize}
\end{itemize}

\textit{Pipelined Modular Multiplier}

\begin{itemize}
  \item \textbf{Synthesis Results}
  \begin{itemize}
    \item Time:
      The minimum period: 3.201\textit{ns} (Maximum Frequency is 312.402\textit{MHz}). The minimum input arrival time before clock is 3.904\textit{ns}. The maximum output
required time after clock is $6.909\text{ns}$. The maximum combinational path delay is $7.629\text{ns}$

(b) FPGA Hardware:

- Number of Slices: 105 out of 15360 (0%)
- Number of Slice Flip Flops: 139 out of 30720 (0%)
- Number of 4 input LUTs: 141 out of 30720 (0%)
- Number of bonded IOBs: 36 out of 448 (8%)

Fig. 13. Pipelined Modular Multiplier Block Diagram.

- Macro Statistics
  - Adders/Subtractors: 8
  - 4-bit adder carry out: 4
  - 5-bit adder carry in/out: 3
Comparisons and analysis remarks

Based on the results obtained from the 160-bit synthesis, the pipelined Homogenous ECC point addition and the parallel Homogenous ECC point addition have been compared. For pipelined point–addition, the longest path in the implementation will be the path of a modular addition. This is because modular addition operations are given their own designated clock cycles separate than the multiplications. Note that the critical path of the modular multiplication is less than the modular addition. Therefore, the period used for pipelined point addition will depend on the longer path, i.e. the modular addition. The reason that modular addition takes a longer period is that the critical path depends on the full-adder of the carry-save adder along with the \( k/2+1 \) carry-propagate adder. In contrast, the critical path of the modular multiplication depends only on 2-to-1 Mux along with a \( k/2 \) carry-propagate adder. Therefore, the implementation total period is \( 7.399ns \) with total number of clock cycles: \( 16N+10 \), as discussed earlier. Hence, the total time needed for a point addition using 160-bit operands is

\[
T=7.399n*(16*160 + 10)=19.01\text{microsecond}.
\]

For a parallel implementation of the point-addition, the longest period includes a modular multiplication and 2 modular additions. This is because the modular additions are implemented in combinational logic.

The total period is: \( 6.052n + 2*8.359n = 22.77ns \).

The total number of clock cycles is \( 4N \). Therefore, the total time needed for a point addition using 160-bit operands is

\[
T=(4*160)*22.77n= 14.57\text{microseconds}.
\]

Using the implemented multiplication and addition units, we compared the proposed design (Figure 7) with previous parallel design shown in Figure 6, both studied in relation to their area and time. Since the basic components are the same implemented in FPGA,
the comparison is believed to be fair and very practical. The study considered the area and timing estimations. To make our study consistent with the previous study in Gutub & Ibrahim (2003), we assume the basic hardware unit as the multiplier. All other units are quantified relative to this multiplier unit, as summarized in Table 3.

**Conclusion**

In this paper, we redesigned multiplier hardware as pipelined for Elliptic Curve Cryptography (ECC) computations. The design adopted projective coordinates ECC arithmetic to reduce the inversion complexity. The pipelined architecture is implemented and synthesized through a Xilinx Virtex-4 FPGA platform for 160-bits. Based on the synthesis results provided, we concluded that the parallel implementation of the point addition is faster than our pipelined approach; however, the pipelined approach is more advanced in term of chip area. Combing the speed and area as a figure of merit cost-values showed that this work gave overall efficiency in its area time cost which made it very attractive showing a promising research direction for researchers to work on.

**Table 3. Cost comparison.**

<table>
<thead>
<tr>
<th>Cost approximation based on similar Area estimate considered in previous designs [33]</th>
<th>Pipelined</th>
<th>Parallel</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (figure relating to size of hardware)</td>
<td>2</td>
<td>4</td>
<td>~50% efficient area</td>
</tr>
<tr>
<td>(AT) (Area (\times) Time)</td>
<td>38</td>
<td>58</td>
<td>~34% efficient cost</td>
</tr>
<tr>
<td>(AT^2) (Area (\times) Time (\times) Time)</td>
<td>722</td>
<td>845</td>
<td>~15% efficient cost</td>
</tr>
</tbody>
</table>

**Cost approximation based on this work implementation**

<table>
<thead>
<tr>
<th>Area (figure relating to size of hardware)</th>
<th>Pipelined</th>
<th>Parallel</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (\times) Time</td>
<td>10,678</td>
<td>14,804</td>
<td>~28% efficient cost</td>
</tr>
<tr>
<td>(AT^2) (Area (\times) Time (\times) Time)</td>
<td>202,882</td>
<td>216,144</td>
<td>~6% efficient cost</td>
</tr>
<tr>
<td>(A^2T) (Area (\times) Area (\times) Time)</td>
<td>6,001,036</td>
<td>15,011,256</td>
<td>~60% efficient cost</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

We would like to thank King Fahd University of Petroleum and Minerals (KFUPM) for partially hosting this research. Special appreciation to the students of the course COE 509, Applied Cryptosystems - Techniques & Architectures, for their valuable initiatives and positive cooperation. Thanks to both research centers: The Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh, and The Center of Research Excellence in Hajj and Omrah, Umm Al-Qura University (UQU), Makkah, for collaborative moral support toward the achievements in this work.

REFERENCES


