Quantum Implementation of ASCON Linear Layer

Soham Roy\textsuperscript{1}, Anubhab Baksi\textsuperscript{2}, and Anupam Chattopadhyay\textsuperscript{2}

\textsuperscript{1} Indian Institute of Technology, Madras, India
\textsuperscript{2} Nanyang Technological University, Singapore
sroy@smail.iitm.ac.in, anubhab001@e.ntu.edu.sg, anupam@ntu.edu.sg

Abstract. In this paper, we show an in-place implementation of the ASCON linear layer. An in-place implementation is important in the context of quantum computing, we expect our work will be useful in quantum implementation of ASCON. In order to get the implementation, we first write the ASCON linear layer as a binary matrix; then apply two legacy algorithms (Gauss-Jordan elimination and PLU factorization) as well as our modified version of Xiang et al.’s algorithm/source-code (published in ToSC/FSE’20). Our in-place implementation takes 1595 CNOT gates and 119 quantum depth; and this is the first in-place implementation of the ASCON linear layer, to the best of our knowledge.

Keywords: ASCON · Quantum Implementation · Linear Layer · In-place Implementation

1 Introduction

ASCON [DEMS19] is a lightweight cryptographic primitive that provides confidentiality, integrity, and authenticity to data transmission. It is recently selected as the winner of the LWC project by NIST\textsuperscript{1}.

In recent years, quantum computing has become an active research area for improving the performance and security of cryptographic primitives. The potential benefits of quantum computing include increased processing speeds and the ability to perform certain operations more efficiently than classical computers. Therefore, exploring the potential of quantum computing to improve the performance of ASCON is of significant interest to the cryptography community. With the increasing popularity of quantum computing, an efficient implementation of ciphers on quantum logic is becoming increasingly important. Therefore, it comes as no surprise that the researchers have been working on finding improved quantum implementation for the ciphers (see, e.g., [BJS+21,JBB+22,JBK+22a,JSK+22]).

Contribution

In this work, we present possibly the first-ever in-place implementation of the ASCON linear layer. The related implementations are available as an open-source project\textsuperscript{2}. The linear layer is described in terms of rotation and XOR of five 64-bit registers. Thus, it can be equivalently expressed as a $320 \times 320$ binary matrix.

To the best of our finding, the only work to deal with the quantum implementation of ASCON is done in [LJS+22]. However, this work does not implement the linear layer explicitly (implements the state update function as one module). Internally, the naïve quantum implementation of the linear layer (Section 3) is invoked, which results in a out-of-place implementation with doubled qubit count.

When it comes to in-place implementation, we know about the PLU factorization, Gauss-Jordan elimination and a recent development by XZLBZ [XZL+20]. All the methods have been used in some capacity in the literature (Section 3). In total, we show 3 in-place implementations of the ASCON linear layer.

In this process, we would like to note that the authors of [XZL+20] explicitly state that their algorithm/source-code works efficiently up to $32 \times 32$ binary matrix. Notably, the task of finding good in-place implementation of $64 \times 64$ binary matrices and beyond is mostly kept out-of-scope. We contribute to the algorithm as well as the source-code of XZLBZ [XZL+20] so that now it can work even with the ASCON binary matrix (which is of dimension $320 \times 320$).

As noted in Section 3, the naïve implementation in quantum doubles the number of qubits and requires a total of (number of qubits + number of XOR operation) CNOT gates in general. When the qubit count $\times$ CNOT count metric is considered, the implementation reported by us is 16.93\% cheaper. The benchmarks corresponding to these implementations, along with two more in-place implementations (obtained by applying the Gauss-Jordan elimination and PLU factorization) are given in Section 4.

\textsuperscript{1}https://csrc.nist.gov/News/2023/lightweight-cryptography-nist-selects-ascon
\textsuperscript{2}https://github.com/sohamroy19/ascon-linear-layer/
2 ASCON Linear Layer

As defined in [DEMS19], the linear layer $p_L$ provides diffusion within each 64-bit register word $x_i$. It applies a linear function $\Sigma_i(x_i)$ to each word $x_i$, defined as follows (‘$\gg$’ indicates right rotation):

\[
\begin{align*}
x_0 & \leftarrow \Sigma_0(x_0) = x_0 \oplus (x_0 \gg 19) \oplus (x_0 \gg 28) \\
x_1 & \leftarrow \Sigma_1(x_1) = x_1 \oplus (x_1 \gg 61) \oplus (x_1 \gg 39) \\
x_2 & \leftarrow \Sigma_2(x_2) = x_2 \oplus (x_2 \gg 1) \oplus (x_2 \gg 6) \\
x_3 & \leftarrow \Sigma_3(x_3) = x_3 \oplus (x_3 \gg 10) \oplus (x_3 \gg 17) \\
x_4 & \leftarrow \Sigma_4(x_4) = x_4 \oplus (x_4 \gg 7) \oplus (x_4 \gg 41)
\end{align*}
\]

When simplified, this can be written as a binary non-singular matrix of dimension $320 \times 320$. The Hamming weights of each row and also column of this matrix are both 3 (i.e., the differential and linear branch numbers of the matrix are 3), and the multiplicative order of the matrix is 64. A graphical view is given in Figure 1 (blue indicates 1 and yellow indicates 0).

![Figure 1: ASCON linear layer (320 \times 320 binary matrix in graphical form)](image)

3 Quantum Implementation of Linear Layer

The problem of optimization of a binary non-singular matrix while implementing for a classical device is well-studied [BP10, LSL19, Köll19, Max19, ZXL+20, BF19, BDK+21, LWF+22]. However, the research works dedicated to find quantum quantum implementation/optimization are few and far between. Here we attempt to
collate the major research works. We also cover the basic theory which is relevant for linear layer implementation (for the implementation of the non-linear component, one may refer to, e.g., [DBSC19,CBC23]).

3.1 Metrics

**Qubit** Analogous to the concept of bit in the classical computing, we use quantum bits (qubits for short). It is customary to write the qubits using the Dirac’s ket notation, e.g., $|0\rangle$ or $|1\rangle$.

**CNOT and SWAP Gates** The Controlled NOT (CNOT for short) and SWAP gates form the basis of linear operations. The CNOT gate basically works like an in-place XOR gate. The circuit diagrams for the gates are as shown in Figure 2.

![Figure 2: Basic quantum gates](image)

A SWAP gate can be implemented using 3 CNOT gates (see, e.g., [DBSC19, Figure 3]). It may not be possible to swap two qubits if those are not adjacent, see [Ser17, Page 6] or [WFH11].

**Quantum Circuit and Depth** The quantum circuits are composed of the quantum gates and those operate on the qubits. For instance, the quantum circuit corresponding to the implementation stated in Example 4 is given in Figure 3.

![Figure 3: A quantum circuit (with CNOT and SWAP gates)](image)

The depth of a logical circuit can be defined as the number of combinational logic gates along the longest path of the circuit. Unlike the classical depth, the quantum depth is generally not simple to calculated (noted in [JBK22]), due to the fact fact a quantum gate can have only 1 fan-out. More relevant discussion on depth can be found in [ZH23]; and apparently this is the only paper that attempts to optimize for quantum depth by using a greedy algorithm.

3.2 Naïve Quantum Implementation

The so-called d-XOR implementation is proposed to indicate the naïve implementation in a classical circuit [KPPY14]. The d-XOR count for the ASCON linear layer is 640, as each row has two XOR operations. However, due to the fan-out restriction and the in-place nature of a quantum circuit, this implementation becomes incompatible.

For the naïve quantum implementation of a $n \times n$ binary matrix, we first need to introduce (up to) $n$ ancilla qubits (effectively doubling the qubit count) initialized at $|0\rangle$. Then the ancilla qubits are updated with the initial values of the first $n$ qubits. This helps retain the original matrix while the $n$ qubits are overwritten.

For the lack of a better terminology, we call this the naïve quantum implementation or direct-CNOT (d-CNOT for short) implementation. An example can be seen from Example 1. Here and in the subsequent Examples, we use the variable $x$ with subscript starting from 0 to indicate the matrix; if needed the variable $y$ with subscript starting from 0 is also used to indicate each row of the matrix (thus, $x$ variables act as the input and $y$ variables act as the output for the implementation).
Example 1 (Naïve quantum). Consider the binary matrix $M^{4\times4}$:

$$
\begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1
\end{pmatrix}
$$

Our original qubits are $\{x_0, x_1, x_2, x_3\}$. In the first step, 4 ancilla qubits are created and initialized to 0:

$$
x_4 \leftarrow |0\rangle \\
x_5 \leftarrow |0\rangle \\
x_6 \leftarrow |0\rangle \\
x_7 \leftarrow |0\rangle
$$

After this, the ancilla qubits are updated in-place, so that each copies one original qubit:

$$
x_4 \leftarrow x_4 \oplus x_0 \\
x_5 \leftarrow x_5 \oplus x_1 \\
x_6 \leftarrow x_6 \oplus x_2 \\
x_7 \leftarrow x_7 \oplus x_3
$$

Equivalently, $M$ is augmented with the $4 \times 4$ identity matrix to create this $8 \times 4$ matrix, which we name $M'$:

$$
M = \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1
\end{pmatrix} \quad \rightarrow \quad \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1
\end{pmatrix} = M'
$$

Finally, we update the original qubits in-place (in $M'$) with the help of the ancilla qubits:

$$
x_0 \leftarrow x_0 \oplus x_5 \\
x_1 \leftarrow x_1 \oplus x_6 \\
x_1 \leftarrow x_1 \oplus x_7 \\
x_2 \leftarrow x_2 \oplus x_4 \\
x_2 \leftarrow x_2 \oplus x_5 \\
x_2 \leftarrow x_2 \oplus x_7 \\
x_3 \leftarrow x_3 \oplus x_4
$$

Therefore, the naïve quantum implementation takes 8 qubits and 11 CNOT operations in total, with 3 quantum depth.

Remark 1. Generally, the ancilla qubits are reset to $|0\rangle$ thereafter (by appplying CNOT gate with its current content), so that those can be used later. If we consider this, then the CNOT gate count would increase.

It may not always be necessary to double the number of qubits in the naïve quantum implementation. This can manifest in two ways (see Examples 2 and 3):

1. Not all original qubits are updated (happens when the $i^{th}$ row of the matrix is identical to the $i^{th}$ row of an identity matrix).
2. Not all original qubits are needed to update other qubits (happens when the $i^{th}$ column of the matrix is identical to the $i^{th}$ column of an identity matrix).
Example 2 (Not all qubits are updated). The naïve quantum implementation of the following matrix is possible without copying $x_3$ to an ancilla qubit, as the 3rd row has not been modified:

$$\begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}$$

Example 3 (Not all original qubits are needed). The naïve quantum implementation of the following matrix is possible without copying $x_0$ and $x_2$ to ancilla qubits, as the 0th and 2nd columns have not been modified:

$$\begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}$$

Thus, it can be implemented using only 1 ancilla qubit to store a copy of the original $x_1$.

There is also another notion that captures the essence of in-place implementation. This is the so-called $s_\epsilon$-XOR implementation [BDK+21, BKD+21b]. In particular, when $\epsilon = 1$, this implementation reduces to factorization of a binary non-singular matrix into matrices for which the quantum implementation is apparent.

### 3.3 In-place Implementations

**Legacy Algorithms** Two legacy algorithms are known to have a quantum-friendly outcome:

1. Gauss-Jordan elimination (mentioned in [Köl19]);
2. PLU factorization (used in some form in [ASR12, GLS16, ASAM18, vH19, ZWS+20, JNRV20, PWL+23]).

As these two algorithms are well-known, we briefly describe those in Examples 4 and 5 for the sake of completeness (similar to Example 1, we use the variable $x$ with subscript starting from 0). For the PLU factorization, we use the Sage implementation.

**Example 4 (Gauss-Jordan elimination).** Consider the binary matrix $M^{4 \times 4}$:

$$\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1
\end{pmatrix}$$

By applying the Gauss-Jordan elimination we obtain the following factorization of the matrix:

$$\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}$$

In this factorization each matrix is necessarily an elementary matrix, corresponding to an elementary operation. Thus, the sequence of gates given by the above factorization is as follows:

$$x_2 \leftarrow x_2 \oplus x_3$$
$$x_1 \leftarrow x_1 \oplus x_3$$
$$x_2, x_3 \leftarrow x_3, x_2$$
$$x_2 \leftarrow x_2 \oplus x_1$$

This implementation incurs 3 CNOT and 1 SWAP gates with 4 quantum depth (can be seen from Figure 3).
Example 5 (PLU factorization). Consider the binary matrix $M^{4 \times 4}$:

\[
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}
\]

By applying the PLU factorization, we obtain the following:

\[
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}
= \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

These three matrices are respectively a permutation matrix, a lower triangular matrix and an upper triangular matrix; and can be implemented following [vH19, Algorithm 1]. The implementation is given as follows (that incurs 5 CNOT and 1 SWAP gates, with 4 quantum depth):

\[
x_0 \leftarrow x_0 \oplus x_2 \\
x_1 \leftarrow x_1 \oplus x_3 \\
x_2 \leftarrow x_2 \oplus x_3 \\
x_3 \leftarrow x_3 \oplus x_1 \\
x_2 \leftarrow x_2 \oplus x_0 \\
x_2, x_3 \leftarrow x_3, x_2
\]

\[
XZLBZ and Its Modification
\]

A new idea for in-place implementation of binary matrices is proposed by XZLBZ [XZL+20]. This is subsequently used by various research works such as [JBK+22b, HS22, YJBS23].

The details of the algorithm used by XZLBZ [XZL+20] is omitted here for brevity though, we present a summarized view for the sake of completeness. The algorithm runs it two phases.

1. In the first phase, an initial factorization is done. In this process, the first priority is to use $A^*$ search. However, if there is no resolution from the $A^*$ search, then the algorithm falls back to Gauss-Jordan elimination.

2. In the second phase, the algorithm tries to find an alternate representation by replacing gap number of CNOT operations with a reduced number of CNOT operations. The gap variable is iterated over, initially being set to the length of the sequence. The program terminates if the gap hits a certain lower bound (which is hard-coded as 4 in their source-code).

In this work, we take a closer look and make some in-house modifications to XZLBZ. The main objective for such change is to accommodate the algorithm/source-code run with larger (i.e., $> 32 \times 32$) binary matrices.

In another direction, the work of Zhu-Huang [ZH23] deals with taking the output from XZLBZ, then post-process it to find a lower quantum depth. A greedy algorithm is used for this purpose. This is among the first, if not the first, work to purposefully optimize for quantum depth. One notable feature in the output is that, unlike XZLBZ [XZL+20], there is no final relabel in [ZH23]. In other words, in a typical XZLBZ output, there are assignments to the $y$ variables. In contrast, Zhu-Huang’s implementation does not need any special $y$-variable assignment (as a $y$-variable corresponds to its $x$-variables). We also make changes to XZLBZ so that our modified version optionally returns with/without final relabel\footnote{Our implementations (given in the repository) follow the without final relabel format, where the initial permutation is declared at the top.}. See Example 6 for an implementation with final relabel and another implementation without final relabel of the same matrix.

Example 6 (Final relabel). Consider the following matrix:

\[
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1
\end{pmatrix}
\]
The matrix is implemented with final relabel by XZLBZ [XZL+20] as follows (notice that the y-variables do not directly match the corresponding x-variables at the LHS):

\[
\begin{align*}
    x_3 &\leftarrow x_3 \oplus x_0 \\
    (x_0 &\leftarrow x_0 \oplus x_1) \rightarrow y_3 \\
    (x_1 &\leftarrow x_1 \oplus x_2) \rightarrow y_0 \\
    (x_2 &\leftarrow x_2 \oplus x_3) \rightarrow y_2 \\
    (x_3 &\leftarrow x_3 \oplus x_4) \rightarrow y_1 \\
    (x_4 &\leftarrow x_4 \oplus x_0) \rightarrow y_4
\end{align*}
\]

The matrix can also be implemented without final relabel, by using 2 SWAP gates at the beginning (this is similar to [ZH23]). In this implementation, the last assignment of an x-variable acts as the corresponding y-variable, and is given as follows:

\[
\begin{align*}
    x_0, x_1 &\leftarrow x_1, x_0 \\
    x_1, x_3 &\leftarrow x_3, x_1 \\
    x_1 &\leftarrow x_1 \oplus x_3 \\
    (x_3 &\leftarrow x_3 \oplus x_0) \rightarrow y_3 \\
    (x_0 &\leftarrow x_0 \oplus x_2) \rightarrow y_0 \\
    (x_2 &\leftarrow x_2 \oplus x_1) \rightarrow y_2 \\
    (x_1 &\leftarrow x_1 \oplus x_4) \rightarrow y_1 \\
    (x_4 &\leftarrow x_4 \oplus x_3) \rightarrow y_4
\end{align*}
\]

The implementation with final relabel can be described by following permutation matrix; and this is termed as the reduced matrix by XZLBZ [XZL+20]:

\[
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

SMT/MILP Model To the best of our finding, the only other idea for in-place implementation of the linear layer is reported in [BKD+21b, BKD21a]. The idea here is to convert the factorization into an SMT or MILP instance. However, it becomes mostly impractical to deal with > 8 × 8 binary matrices, due to the solver taking long time.

4 Quantum Benchmark

With the background covered in Section 3, the following quantum benchmarks are reported in Table 1:

1. Naïve corresponds to using ancillary qubits to enable computing the new state of a qubit while maintaining the previous one (in a back-up qubit), which is required for the calculation of new state of two other qubits.
2. Gauss-Jordan elimination factorizes any binary matrix using elementary operations (which correspond to CNOT or SWAP gates).
3. PLU factorization outputs a given binary matrix into a permutations matrix, a lower triangular matrix and an upper triangular matrix.
4. Modified XZLBZ is our in-house modification of what is introduced in [XZL+20].

In the naïve quantum implementation, the qubit count × CNOT count = 640 × 960; but the same metric for the result obtained using modified XZLBZ is 320 × 1595. Thus, we save about about 16.93%.

\[\square\]
Table 1: Quantum benchmarks of ASCON linear layer

<table>
<thead>
<tr>
<th>Method</th>
<th>Qubit count</th>
<th>CNOT count</th>
<th>Quantum depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>640</td>
<td>960</td>
<td>26</td>
</tr>
<tr>
<td>Gauss-Jordan</td>
<td>320</td>
<td>2413</td>
<td>358</td>
</tr>
<tr>
<td>PLU</td>
<td>2413</td>
<td>288</td>
<td></td>
</tr>
<tr>
<td>Modified XZLBZ</td>
<td>1595</td>
<td>119</td>
<td></td>
</tr>
</tbody>
</table>

5 Conclusion

In this work, we extract the ASCON linear layer (which is a $320 \times 320$ binary non-singular matrix) and show its quantum benchmarks. To the best of our knowledge, this is the first-of-its-kind analysis on ASCON. We show 4 in-place implementations of the matrix, which is useful in finding a suitable quantum implementation (as in-place implementations do not require any ancilla/garbage qubit).

The best in-place implementation (in terms of either CNOT count or quantum depth) that we obtain is from a modified version of the XZLBZ [XZL+20]. We contribute in respective parts of the algorithm and the source-code of XZLBZ [XZL+20], so that now our version of XZLBZ works well even with the ASCON linear layer.

The naïve quantum implementation doubles the qubit count, whereas the in-place implementations maintain the same qubit count. Thus, one may be interested in a trade-off where the resulting circuit incurs relatively low CNOT count and/or quantum depth, but not all qubits are duplicated. Put in other words, some of the qubits are updated in-place, but the rest are duplicated and updated out-of-place.

References


ZWS

JBB

JNRV20. Samuel Jaques, Michael Naehrig, Martin Roetteler, and Fernando Virdia. Implementing grover oracles

YJBS23. Yujin Yang, Kyungbae Jang, Anubhab Baksi, and Hwajeong Seo. Optimized implementation and analysis

XZL

WFH11. David S. Wang, Austin G. Fowler, and Lloyd C. L. Hollenberg. Surface code quantum computing with error

PWL

LJS

JSK

JBK

GLRS16. Markus Grassl, Brandon Langenberg, Martin Roetteler, and Rainer Steinwandt. Applying Grover’s algorithm to

ZH23. Chengkai Zhu and Zhenyu Huang. Optimizing the depth of quantum implementations of linear layers.

K¨ ol19. Lukas K¨ olsch. Xor-counts and lightweight multiplication with fixed elements in binary finite fields. In

DEMS19. Christoph Dobrauig, Maria Eichlseder, Florian Mendel, and Martin Schl¨ affer. Ascon v1.2. Submission
round-2/spec-doc-round2/ascon-spec-round2.pdf. 1, 2


LSL+19. Shun Li, Siwei Sun, Chaoyun Li, Zihao Wei, and Lei Hu. Constructing low-latency involutory mds matrices with lightweight circuits. IACR Transactions on Symmetric Cryptology, pages 84–117, 2019. 2


YJBS23. Yujin Yang, Kyungb ae Jang, Anubhab Baksi, and Hwajeong Seo. Optimized implementation and optimization of cha in quantum computing. Applied Sciences, 13(8), 2023. 6
