Quantum Circuits for Toom-Cook Multiplication

Srijit Dutta, ¹ Debjyoti Bhattacharjee, ² and Anupam Chattopadhyay ²

¹ Department of Computer Science and Engineering, IIT Bombay, India
² School of Computer Science and Engineering, Nanyang Technological University, Singapore

(Dated: June 26, 2018)

In this paper, we report efficient quantum circuits for integer multiplication using Toom-Cook algorithm. By analyzing the recursive tree structure of the algorithm, we obtained a bound on the count of Toffoli gates and qubits. These bounds are further improved by employing reversible pebble games through uncomputing the intermediate results. The asymptotic bounds for different performance metrics of the proposed quantum circuit are superior to the prior implementations of multiplier circuits using schoolbook and Karatsuba algorithms.

I. INTRODUCTION

Quantum computing has gathered significant attention by solving certain problems much faster than any known classical algorithm. In contrast to Boolean logic, quantum bits (qubits) not only represent the classical 0 and 1 states but also any complex combination or superposition of both, leading to a significant speed-up in computing. The Deutsch-Jozsa algorithm ¹ and Shor’s factorization algorithm ² are well-known examples demonstrating the power of quantum computing. This capability gives rise to the bounded-error quantum polynomial time (BQP) complexity class with an open quest among computer scientists and mathematicians to establish the exact relation between BQP and other complexity classes. In order to accelerate scientific computing using the capabilities of a quantum computer, efficient quantum circuits for basic mathematical functions are needed. The efficiency of a quantum circuit is measured by lower computational space (number of qubits) and lower computational time (logical depth). For fault-tolerant, error-protected quantum circuits to implement the quantum algorithms, it is projected that a large number of physical qubits are required for every logical qubit ³. Naturally, potential solutions to reduce the number of logical qubits contribute to the overall efficiency of the quantum circuit.

Multiplication is one of the elementary mathematical operations of arithmetic. Fast long integer arithmetic is at the very core of many computer algebra systems. In quantum computing, apart from being used as a block in itself, integer multiplication is used as a subroutine in many applications such as Shor’s integer factorization algorithm and in Newton iterations for calculating many functions like the inverse ⁴.

In this paper, we present quantum implementation of the Toom-Cook multiplication algorithm ⁵ ⁶, which can attain better asymptotic complexity than simple schoolbook multiplication and the Karatsuba based integer multiplication ⁷. We further improve these bounds by analyzing pebble games on complete trees.

II. PRIOR WORKS

The problem of multiplication in the quantum domain has been explored previously. For small numbers, the naive schoolbook multiplication works best, with a runtime complexity \(O(n^2)\) that also translates to the logical depth in a quantum circuit realization. Karatsuba multiplication, implemented in quantum circuits ⁷, is usually faster when the multiplicands are longer than 320 - 640 bits, which also provides asymptotic improvement in terms of Toffoli cost and Toffoli depth over the schoolbook multiplication. However, the number of qubits required for Karatsuba-based quantum implementation is higher than the schoolbook multiplication. In the realm of quantum circuits, so far, the Schönhage-Strassen method (using Fast Fourier transform) and Toom-Cook multiplication algorithm are not reported, even though, it is known from the classical implementations that these algorithms result in better run time, when the operand size is much larger. We primarily focus on the Toom-Cook multiplication in this work. As reported in our results, this leads to significant savings of all the performance metrics for an efficient quantum circuit, clearly outperforming the prior works.

The family of Toom-Cook methods is an infinite set of polynomial algorithms (Toom-2,5, Toom-3, Toom-4, etc) ⁸. Instead of using the more common Toom-3 implementation, we present the work with Toom-2.5 to avoid a division by 3 required by the former. This leads to reduction in overall circuit costs, as quantum division is costlier in terms of Toffoli count and Toffoli depth than simple addition or shift operations. Most of the higher Toom implementations require a similar division by constants that are not multiples of 2. The implementation of such divisions incur higher quantum costs and therefore we avoid them.

When moving to quantum domains, gate sets need to go beyond classical to create the superposition effect of the inputs. The standard universal quantum gate library that efficiently implements fault-tolerant quantum error correction codes is the Clifford+T library ⁹ ¹⁰. In this library, the cost of implementing a \(T\)-gate is sufficiently high to customarily neglect the cost of other Clifford group gates, while determining the total cost of the
quantum circuit. Therefore, the number of T-gates is a metric to judge the cost of a quantum circuit. Also, the number of qubits used in a quantum circuit is another important standard, since the current quantum technologies still struggle to achieve error free computation for large count of qubits. A study of the space-time trade off can be performed [11] using these two metrics. Another metric of importance is the T-depth. T-depth is defined as the number of T-stages is a quantum circuit where each such stage consists of one or more T or T† gates performed concurrently on separate qubits. It is important to note that an input circuit with continuous parameter gates (e.g. z-rotation gate $R_z(\theta)$) is decomposed using a set of discrete, basis gates, typically from the Clifford+$T$ library. The exact number of Clifford+$T$ gates needed for such a continuous parameter gate depends on the desired accuracy, and the discrete gate set provides only an approximation. In the context of the current work, we consider T-count and T-depth of the circuit to be proportional to the Toffoli-count and Toffoli-depth respectively, by following the Toffoli decomposition proposed in [12].

III. TOOM-COOK MULTIPLIER

Given two large integers $n_1$ and $n_2$, the Toom-Cook algorithm splits them into $k$ smaller parts of length $l$. The multiplication sub-operations are then computed recursively using Toom-Cook multiplication again, till we are able to apply another algorithm on it for the last stage of recursion, or until the desired multiplier is reached. The input numbers are divided into limbs of a given size, each in the form of polynomial, and the limb size is used as radix. Instead of multiplying the obtained polynomials directly, they are evaluated at a set of points and the values multiplied together at those points. Based on the products obtained at those points, the product polynomial is computed by interpolation. The final result is then obtained by substituting the radix.

In general, Toom-$k$ runs in $\Theta(c(k)n^e)$, where $n$ denotes input size, $k$ is the number of parts that the input operand is decomposed into and $e = \log_2 (2k - 1)$. $c(k)$ is the time spent on auxiliary additions and multiplications by small constants. The Karatsuba algorithm [13] is a special case of Toom-Cook multiplication (Toom-2), where the input operand is split into two smaller ones. It reduces 4 multiplications to 3 and so operates at $\Theta(n^\log_2 3)$. In general, Toom-$k$ reduces $k^2$ multiplications to $2k - 1$ ordinary long multiplication (equivalent to Toom-1) with complexity $\Theta(n^2)$.

A. Implementation Details

Let $x$ and $y$ be two $n$ bit numbers. To proceed with Toom-2.5 algorithm, we first decompose $x$ and $y$ into two and three parts respectively. Express $x = x_12^i + x_0$ and $y = y_22^{2i} + y_12^i + y_0$ with $i \geq 1$. Typically $i$ is chosen as

$$\max\left\{ \left\lfloor \frac{\log_2 x}{k} \right\rfloor, \left\lfloor \frac{\log_2 y}{k} \right\rfloor \right\},$$

where $k = 2.5$ in our case. We define the following four product terms:

$$P = x_0y_0,$$  \hspace{1cm} (1)
$$Q = (x_0 + x_1)(y_0 + y_1 + y_2),$$  \hspace{1cm} (2)
$$R = (x_0 - x_1)(y_0 - y_1 + y_2),$$  \hspace{1cm} (3)
$$S = x_1y_2$$  \hspace{1cm} (4)

Then, the product $xy$ is evaluated as :-

$$xy = A2^{3i} + B2^{2i} + C2^i + D$$ (5)
$$A = S$$ (6)
$$B = -P + \frac{1}{2}Q + \frac{1}{2}R$$ (7)
$$C = \frac{1}{2}Q - \frac{1}{2}R - S$$ (8)
$$D = P$$ (9)

Note that only 4 multiplications are required for computation of the product. Also, each of these multiplications consists of numbers of size smaller than the original problem size (bit-width). Each smaller multiplication is between one number of bit-width $n/2$ and another of bit-width $n/3$. Since this method is applied in recurrence the second time for our analysis, we consider that the smaller limbs formed from the number which was split into 3 parts originally is now split into 2 parts and vice-versa. So after two steps, we get 16 smaller problems of size $n/6$ each. Thus, we obtain the basic recurrence for the number of steps $T(n)$.

$$T(n) = 16T\left(\frac{n}{6}\right)$$ (10)

All additions (the intermediate ones as well as the final ones) are performed by separate adders which have bounded cost.

B. Gate count Analysis

For gate count analysis, we consider only the Toffoli count required by the quantum circuit or sub-circuit. This is because the other used gates (NOTs, CNOTs) do not contribute to the T-count of the circuit, considering the Clifford+$T$ library. The designed circuit maps $(x,y,0,0) \mapsto (x,y,g,xy)$, where $g$ denotes some garbage output resulting as computation of $A, B, C$ and $D$. The product is copied after the computation and the circuit is then run backwards (uncomputed) to set the garbage outputs back to 0.

In our circuit implementation, the Cuccaro adder is used [14]. For addition of two $n$ bit numbers, Cuccaro adder requires $2n - 1$ Toffoli gates. It is also established that the cost $A_n$, for an in-place adder adding two $n$ bit numbers, is bounded by $2n$ Toffoli gates.

Let $T_{n,n}$ denote the multiplication call to Toom-2.5 circuit for calculating product to two $n$ bit numbers and
The recursive nature of the problem gives rise to an inherent tree structure as shown in Fig. 1. The size of a node is representative of the problem size at that level. For example, the root level denotes the complete problem (n bit multiplication). According to the recursion presented in Equation (10), each node will have 16 children nodes denoting a smaller problem (n/6 bit multiplication). For the Toom-2.5 circuit with an input of size n at any level x in the tree, there are 16^x nodes of size n^6−x each for a total cost of n(16/6)^x at level x (level numbering starting at 0 from root). So, the space cost Q_{orig} of the complete tree is

\[ Q_{orig} = n \sum_{0}^{N-1} \left( \frac{16}{6} \right)^x \]

\[ = n \frac{(16/6)^{\log_6 n} - 1}{(16/6) - 1} \]

\[ = O(n(8/3)^{\log_6 n}) = O(n^{1+\log_6 (8/3)}) \]

\[ \approx O(n^{1.547}) \]

where tree height, \( N = \log_6 n \).

The reversible pebble game is a combinatorial game played on rooted Directed Acyclic Graphs (DAGs). Each pebble represents some amount of space. The rules are similar to those used in the pebble game to model irreversible computation except that we simply cannot remove pebbles by reversibility constraint. There is a reverse computation for each corresponding computation performed, implying that during the game, the pebbles may still be removed but it is subject to the same conditions as applied during placing the pebbles. We use this reversible game to obtain better asymptotic bounds in the number of qubits (space) to implement the Toom 2.5 algorithm.

We want to find a level in the recursion tree such that the size of each node’s sub-tree is approximately equal to the sum of the size of all nodes at that level chosen and above. Once all the nodes in the chosen level have been computed, we uncompute all the sub-trees below it. This is performed to minimize space — the size of these sub-trees is chosen to be approximately equal to the remaining size of the tree above them. Let the required height be k from the leaves of the tree. The cost of all height k sub-trees is

\[ n \sum_{N-k}^{N-1} \left( \frac{16}{6} \right)^x \]

Therefore, cost of a single height k sub-tree is

\[ \frac{n}{16^{N-k}} \sum_{N-k}^{N-1} \left( \frac{16}{6} \right)^x = \frac{n}{6^{N-k}} \sum_{0}^{k-1} \left( \frac{16}{6} \right)^x \]

We want this to equal the cost of all nodes above the
the number of sub-trees $ST_k$ and corresponding depth $D_k$ is defined as follows.

$$ST_k = 16 \left(1 - \frac{16}{\log_2 3 \log_6 n} \right) \log_6 n$$

$$D_k = \frac{\log_2 n}{6 \left(1 - \frac{16}{\log_2 3 \log_6 n} \right) \log_6 n}$$

$$ST_k * D_k = n \left(\frac{8}{3} \left(1 - \frac{16}{\log_2 3 \log_6 n} \right) \log_6 n \right) \approx n^{1.43}$$

The product $ST_k * D_k$ gives an overall depth for computing the entire $k^{th}$ level of the recursion tree.

The method proposed above is most efficient if both the numbers to be multiplied are approximately of the same bit-width. In case one of them is much bigger than the other, it is better if the bigger number is repeatedly divided into 3 parts in each turn, until the smaller parts of both the numbers are roughly the same size. Following this method, the asymptotic computational complexity can be shown to be more efficient than that of the alternating Toom-2.5 method adopted.

The circuit of the described implementation is shown in Fig. 2. It describes the circuit for $T_{n,n}$ that multiplies $x$ (decomposed into $x_0, x_1$) and $y$ (decomposed into $y_0, y_1, y_2$). All symbols and variables mentioned hold the same meanings as described in the analysis above. The adder, subtractor and shifting blocks are represented as ‘Adder’, ‘Sub’ and ‘Shift’ respectively. The $T_{2,2}$ blocks denote Toom-2.5 sub-circuits of smaller bit-width.

TABLE I: Asymptotic performance analysis of the quantum implementation of various multiplication methods.

| Method          | QC  | TC  | TD              |
|-----------------|-----|-----|-----------------|
| Naive [7]       | $O(n)$ | $O(n^2)$ | $O(n^2)$ |
| Naive Improved. [16] | $O(n)$ | $O(n^2)$ | $O(n \log n)$ |
| Karatsuba [17]  | $O(n^{1.427})$ | $O(n^{1.063})$ | $O(n^{1.158})$ |
| Toom-2.5        | $O(n^{1.404})$ | $O(n^{1.068})$ | $O(n^{1.148})$ |
| Const. Mult. [17] | $O(n)$ | $O(n^2)$ | $O(n)$ |

| QC | TC | TD |
|----|----|----|
| Qubit count | Toffoli count | Toffoli depth |

TABLE II: Cost of quantum implementation of multiplication.

| Method          | QC  | TC  | TD              |
|-----------------|-----|-----|-----------------|
| Naive [7]       | $4n + 1$ | $4n^2 - 3n$ | $4n^2 - 4n + 1$ |
| Karatsuba [17]  | $n \left(\frac{8}{3} \right) \frac{\log_2 n}{\log_6 n}$ | $42n \log_2 n$ | $n \left(\frac{8}{3} \right) \left(1 - \frac{1}{\log_2 3} \right) \log_2 n$ |
| Toom-2.5        | $n \left(\frac{8}{3} \right) \frac{\log_2 n}{\log_6 n}$ | $49n \log_6 n$ | $n \left(\frac{8}{3} \right) \left(1 - \frac{1}{\log_2 3} \right) \log_6 n$ |
| Const. Mult. [17] | $3n + 1$ | $4n(n + 1)$ | $8n$ |
IV. RESULTS AND DISCUSSIONS

Table I presents the asymptotic results of implementation of various multiplication methods while Table II provides the exact constants involved. The naïve multiplication method suggested in [7] allows implementation with the lowest number of qubits asymptotically but fares badly in terms of Toffoli count and depth.

In [16], the implementation of logarithmic depth adders have been provided. The naïve (shift-add) multiplier can be improved in depth by using the logarithmic depth adder as a submodule. The n-bit adder has a depth of order $O(n \log n)$, thus the multiplier shall have a depth of $O(n \log n)$. However, for both ‘in place’ and ‘out of place’ adders described in [16], extra ancilla are required for intermediate computation. Also, the Toffoli count is greater compared to the Cuccaro Adder. Thus, the multiplier developed by extension though optimized in depth, will have greater asymptotic Toffoli and qubit count, equal to $O(n^2)$. In Table I, we provide the asymptotic complexity of such a multiplier. However, in the absence of an explicit design, we are unable to provide the exact constants involved in the cost metrics and hence Naïve Improved method mentioned in Table I is excluded from Table II. Toom-2.5 requires less number of qubits than Karatsuba [7]. Toom-2.5 outperforms both the naïve and Karatsuba methods in terms of Toffoli count as well as Toffoli depth, highlighting the efficiency of the proposed method.

Pavlidis et al. [17] presented a depth optimized multiplier, for multiplication by a constant only. Therefore, it is unfair to be directly compared with our implementation and the Karatsuba multiplication implementations presented in [7]. It has a Toffoli depth of $8n$, a cost of $4n(n + 1)$ and qubit count of $3n + 1$.

The Clifford+T quantum gate library has garnered much interest in the implementation of fault-tolerant quantum circuits [18]. As mentioned in [19], the cost of Toffoli gate is higher compared to the NOT and CNOT gates. The Toffoli gate may be decomposed using Clifford+T-gates, which makes cost metrics associated with Toffoli gates important. Therefore, Toffoli count and Toffoli depth are used as the performance metrics to begin with. The cost of mapping a Toffoli gate to the Clifford+T fault tolerant library is upper bounded by $7 \times$ Toffoli count and $3 \times$ Toffoli depth [12]. Therefore, fault tolerant implementation of the proposed multiplication method would have at most $7 \times$ Toffoli count and $3 \times$ Toffoli depth of the values mentioned in Table I. It is further possible to improve these values by optimization techniques proposed in [11][12].

Fig. 3a presents a comparison of the Toffoli count required by the various methods for variation in the bit-width of the inputs. The naïve multiplication method performs better in terms of total Toffoli cost at smaller input sizes (< 300 bits), but is outperformed by the Karatsuba and Toom algorithms at higher bit-widths.

Fig. 3c shows the variation in the qubit requirements by the different implementations across a range of input sizes. In this case the shift and add method (naïve) outperforms both the recursive algorithms as it increases linearly. However, this low space requirement leads to a higher depth as demonstrated, in Fig. 3d in a logarithmic scale. Both Toom-2.5 and the Karatsuba implementations perform much better in this respect.

We also present a bound on the CNOT counts of the considered implementations. In the proposed Toom-2.5 circuit shown in Fig. 2 CNOT gates are present in the Cuccaro adders and copy blocks. It can be seen from [14] that the number of CNOT gates in a n bit adder can be bounded by $5n$. Proceeding similarly as the Toffoli count analysis, we get an exactly similar recurrence relation as presented in Gate Count Analysis in Section III. Let $CC_n$ denote the number of CNOT gates in $T_{n,n}$. Also, let $Ac_n$ denote the number of CNOT for an in-place n bit adder.

$$CC_n = 16CC_{n/6} + 40Ac_{n/6} + 22Ac_{n/3} + 4Ac_{n/2} + 4Ac_{n/6}$$

$$= 16^{\log_6 n} CC_1 + 40(Ac_1 + 16Ac_6 + \ldots)$$

$$+ 22(Ac_3 + 16Ac_6 + \ldots) + 4(Ac_2 + 16Ac_6 + \ldots)$$

$$+ 4(Ac_6 + 16Ac_{12} + \ldots) + COPY_{cnot}$$

where $COPY_{cnot}$ denotes the number of CNOTs used in the two copy blocks. However, the number of CNOT gates arising out of the Copy blocks are of the order $O(n)$ and is dominated by the terms of order $n^{\log_6 16}$. $CC_1 = 0$ because 1-bit multiplier just consists of 1 Toffoli gate.

$$CC_n \approx 2(58n[(\frac{16}{6})^{\log_6 n} - 1])$$

$$= 116n(n^{\log_6 (16/6)} - 1) \leq 116n^{\log_6 16}$$

By similar analysis, the CNOT count of the Karatsuba multiplier can be bounded by $100n^{\log_2 3}$. For the naïve method, controlled adders are considered as described in [7]. Each such adder has 2n CNOTs and the multiplier uses $n - 1$ such adders. Thus, the total CNOT count is $2n^2 - 2n$. For the Constant multiplier in [17] 2n CNOT gates are employed. These observations are summarized in Fig. 4.

In [20], it has been established that the T-gate is at least 6 times costlier compared to the CNOT gate, which emphasizes the importance of T-count and T-depth. However, with increasing circuit size, the cost of CNOT may take a dominant role if we follow the analysis in terms of upper/lower bounds [19]. From that perspective, the study of overall cost is important. As we found for the case of multipliers, the CNOT count of Toom-2.5 Multiplier grows at slightly lower rate compared to that of the Karatsuba Multiplier with increasing input size. Considering the fact that Toffoli count of Toom-2.5 Multiplier already outperforms Karatsuba at large input sizes, the proposed design is clearly more efficient.
FIG. 3: Comparison of the quantum multiplier implementations based on: (a) Toffoli Count, (b) Toffoli Depth and (c) #Qubits.

FIG. 4: Variation in CNOT counts across different implementations with increasing input size.

V. CONCLUSION

Designing an efficient quantum circuit with low resource requirements and faster run time is an important challenge with significant repercussions across several domains, such as, scientific computing and security. In this work, we reported an efficient quantum circuit for integer multiplication based on Toom-Cook algorithm. We provide design results, and techniques for lowering the resource requirements. In terms of asymptotic complexity, the presented implementation outperforms the state-of-the-art results for multiple performance metrics.

[1] D. Deutsch and R. Jozsa, in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 439 (The Royal Society, 1992) pp. 553–558.
[2] P. W. Shor, in Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on (Ieee, 1994) pp. 124–134.
[3] E. T. Campbell, B. M. Terhal, and C. Vuillot, Nature 549, 172 (2017).
[4] M. Soeken, M. Roetteler, N. Wiebe, and G. De Micheli, in 2017 Design, Automation & Test in Europe Conference & Exhibition (DATE) (IEEE, 2017) pp. 470–475.
[5] A. L. Toom, in Soviet Mathematics Doklady, Vol. 3 (1963) pp. 714–716.
[6] S. A. Cook and S. O. Aanderaa, Transactions of the American Mathematical Society 142, 291 (1969).
[7] A. Parent, M. Roetteler, and M. Mosca, arXiv preprint arXiv:1706.03419 (2017).
[8] D. E. Knuth, The art of computer programming, Vol. 2 (Pearson Education, 1997).
[9] M. Amy, D. Maslov, and M. Mosca, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 33, 1476 (2014).
[10] A. G. Fowler, A. M. Stephens, and P. Groszkowski, Physical Review A 80, 052312 (2009).
[11] R. Wille, M. Soeken, D. M. Miller, and R. Drechsler, Integration, the VLSI Journal 47, 284 (2014).
[12] N. Abdessaied, M. Amy, M. Soeken, and R. Drechsler, in Multiple-Valued Logic (ISMVL), 2016 IEEE 46th International Symposium on (IEEE, 2016) pp. 150–155.
[13] A. Karatsuba, in Sov. Phys. Dokl., Vol. 7 (1963) pp. 595–596.
[14] S. A. Cuccaro, T. G. Draper, S. A. Kutin, and D. P. Moulton, arXiv preprint quant-ph/0410184 (2004).
[15] C. H. Bennett, SIAM Journal on Computing 18, 766 (1989).
[16] T. G. Draper, S. A. Kutin, E. M. Rains, and K. M. Svore, arXiv preprint quant-ph/0406142 (2004).
[17] A. Pavlidis and D. Gizopoulos, arXiv preprint arXiv:1207.0511 (2012).
[18] Y. S. Weinstein, Physical Review A 87, 032320 (2013).
[19] D. Maslov, arXiv preprint arXiv:1602.02627 (2016).
[20] V. V. Shende and I. L. Markov, arXiv preprint arXiv:0803.2316 (2008).