

# NP on Logarithmic Space

Frank Vega

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 30, 2023

## NP on Logarithmic Space

Frank Vega $^{\ast 1}$ 

#### Abstract

P versus NP is considered as one of the most important open problems in computer science. This consists in knowing the answer of the following question: Is P equal to NP? It was essentially mentioned in 1955 from a letter written by John Nash to the United States National Security Agency. However, a precise statement of the P versus NPproblem was introduced independently by Stephen Cook and Leonid Levin. Since that date, all efforts to find a proof for this problem have failed. Another major complexity classes are L and NL. Whether L = NL is another fundamental question that it is as important as it is unresolved. We prove that  $NP \subseteq NSPACE(\log^2 n)$  just using 1L-reductions.

2020 MSC: MSC 68Q15, MSC 68Q17

KEYWORDS: Computational Algorithm, Complexity Classes, Completeness, Polynomial Time, Reduction, Logarithmic Space

## 1 Introduction

In 1936, Turing developed his theoretical computational model [10]. The deterministic and nondeterministic Turing machines have become in two of the most important definitions related to this theoretical model for computation [10]. A deterministic Turing machine has only one next action for each step defined in its program or transition function [10]. A nondeterministic Turing machine could contain more than one action defined for each step of its program, where this one is no longer a function, but a relation [10].

Let  $\Sigma$  be a finite alphabet with at least two elements, and let  $\Sigma^*$  be the set of finite strings over  $\Sigma$  [2]. A Turing machine M has an associated input alphabet  $\Sigma$  [2]. For each string w in  $\Sigma^*$  there is a computation associated with M on input w [2]. We say that M accepts w if this computation terminates in the accepting state, that is M(w) = "yes" [2]. Note that, Mfails to accept w either if this computation ends in the rejecting state, that is M(w) = "no", or if the computation fails to terminate, or the computation ends in the halting state with some output, that is M(w) = y (when M outputs the string y on the input w) [2].

Another relevant advance in the last century has been the definition of a complexity class. A language over an alphabet is any set of strings made up of symbols from that alphabet [4]. A complexity class is a set of problems, which are represented as a language, grouped by measures such as the running time, memory, etc [4]. The language accepted by a Turing machine M, denoted L(M), has an associated alphabet  $\Sigma$  and is defined by:

$$L(M) = \{ w \in \Sigma^* : M(w) = "yes" \}.$$

Moreover, L(M) is decided by M, when  $w \notin L(M)$  if and only if M(w) ="no" [4]. We denote by  $t_M(w)$  the number of steps in the computation of M on input w [2]. For  $n \in \mathbb{N}$  we denote by  $T_M(n)$  the worst case run time of M; that is:

$$T_M(n) = max\{t_M(w) : w \in \Sigma^n\}$$

where  $\Sigma^n$  is the set of all strings over  $\Sigma$  of length n [2]. We say that M runs in polynomial time if there is a constant k such that for all  $n, T_M(n) \leq n^k + k$ [2]. In other words, this means the language L(M) can be decided by the Turing machine M in polynomial time. Therefore, P is the complexity class of languages that can be decided by deterministic Turing machines in polynomial time [4]. A verifier for a language  $L_1$  is a deterministic Turing machine M, where:

$$L_1 = \{w : M(w, u) = "yes" \text{ for some string } u\}.$$

We measure the time of a verifier only in terms of the length of w, so a polynomial time verifier runs in polynomial time in the length of w [2]. A verifier uses additional information, represented by the string u, to verify that a string w is a member of  $L_1$ . This information is called certificate. NP is the complexity class of languages defined by polynomial time verifiers [8].

It is fully expected that  $P \neq NP$  [8]. Indeed, if P = NP then there are stunning practical consequences [8]. For that reason, P = NP is considered as a very unlikely event [8]. Certainly, P versus NP is one of the greatest open problems in science and a correct solution for this incognita will have a great impact not only in computer science, but for many other fields as well [3]. Whether P = NP or not is still a controversial and unsolved problem [1]. We provide an important step forward for this outstanding problem using the logarithmic space complexity.

#### 1.1 The Hypothesis

A function  $f: \Sigma^* \to \Sigma^*$  is a polynomial time computable function if some deterministic Turing machine M, on every input w, halts in polynomial

time with just f(w) on its tape [10]. Let  $\{0,1\}^*$  be the infinite set of binary strings, we say that a language  $L_1 \subseteq \{0,1\}^*$  is polynomial time reducible to a language  $L_2 \subseteq \{0,1\}^*$ , written  $L_1 \leq_p L_2$ , if there is a polynomial time computable function  $f: \{0,1\}^* \to \{0,1\}^*$  such that for all  $x \in \{0,1\}^*$ :

 $x \in L_1$  if and only if  $f(x) \in L_2$ .

An important complexity class is NP-complete [5]. If  $L_1$  is a language such that  $L' \leq_p L_1$  for some  $L' \in NP$ -complete, then  $L_1$  is NP-hard [4]. Moreover, if  $L_1 \in NP$ , then  $L_1 \in NP$ -complete [4]. A principal NP-complete problem is SAT [5]. An instance of SAT is a Boolean formula  $\phi$  which is composed of:

- 1. Boolean variables:  $x_1, x_2, \ldots, x_n$ ;
- 2. Boolean connectives: Any Boolean function with one or two inputs and one output, such as  $\wedge(AND)$ ,  $\vee(OR)$ ,  $\neg(NOT)$ ,  $\Rightarrow(implication)$ ,  $\Leftrightarrow(if and only if);$
- 3. and parentheses.

A truth assignment for a Boolean formula  $\phi$  is a set of values for the variables in  $\phi$ . A satisfying truth assignment is a truth assignment that causes  $\phi$  to be evaluated as true. A Boolean formula with a satisfying truth assignment is satisfiable. The problem *SAT* asks whether a given Boolean formula is satisfiable [5]. We define a *CNF* Boolean formula using the following terms:

A literal in a Boolean formula is an occurrence of a variable or its negation [4]. A Boolean formula is in conjunctive normal form, or CNF, if it is expressed as an AND of clauses, each of which is the OR of one or more literals [4]. A Boolean formula is in 3-conjunctive normal form or 3CNF, if each clause has exactly three distinct literals [4]. For example, the Boolean formula:

$$(x_1 \lor \neg x_1 \lor \neg x_2) \land (x_3 \lor x_2 \lor x_4) \land (\neg x_1 \lor \neg x_3 \lor \neg x_4)$$

is in 3CNF. The first of its three clauses is  $(x_1 \lor \neg x_1 \lor \neg x_2)$ , which contains the three literals  $x_1, \neg x_1$ , and  $\neg x_2$ .

A logarithmic space Turing machine has a read-only input tape, a writeonly output tape, and read/write work tapes [10]. The work tapes may contain at most  $O(\log n)$  symbols [10]. In computational complexity theory, L is the complexity class containing those decision problems that can be decided by a deterministic logarithmic space Turing machine [8]. NL is the complexity class containing the decision problems that can be decided by a nondeterministic logarithmic space Turing machine [8]. In general, DSPACE(S(n)) and NSPACE(S(n)) are complexity classes that are used to measure the amount of space used by a Turing machine to decide a language, where S(n) is a space-constructible function that maps the input size n to a non-negative integer [9]. The complexity class DSPACE(S(n)) is the set of languages that can be decided by a deterministic Turing machine that uses O(S(n)) space [9]. The complexity class NSPACE(S(n)) is the set of languages that can be decided by a nondeterministic Turing machine that uses O(S(n)) space [9].

A function  $f: \Sigma^* \to \Sigma^*$  is a logarithmic space computable function if some deterministic Turing machine M, on every input w, halts using logarithmic space in its work tapes with just f(w) on its output tape [10]. Let  $\{0,1\}^*$  be the infinite set of binary strings, we say that a language  $L_1 \subseteq \{0,1\}^*$  is logarithmic space reducible to a language  $L_2 \subseteq \{0,1\}^*$ , written  $L_1 \leq_l L_2$ , if there is a logarithmic space computable function f: $\{0,1\}^* \to \{0,1\}^*$  such that for all  $x \in \{0,1\}^*$ :

$$x \in L_1$$
 if and only if  $f(x) \in L_2$ 

The logarithmic space reduction is used for the completeness of the complexity classes L, NL and P among others.

The two-way Turing machines may move their head on the input tape into two-way (left and right directions) while the one-way Turing machines are not allowed to move the input head on the input tape to the left [6]. Hartmanis and Mahaney have investigated the classes 1L and 1NL of languages recognizable by deterministic one-way logarithmic space Turing machine and nondeterministic one-way logarithmic space Turing machine, respectively [6]. They have shown that  $1L \neq 1NL$  (by looking at a uniform variant of the string non-equality problem from communication complexity theory) and have defined a natural complete problem for 1NL under deterministic one-way logarithmic space reductions [6]. Furthermore, they have proven that  $1NL \subseteq L$  if and only if L = NL [6].

We can give a certificate-based definition for NL [2]. The certificatebased definition of NL assumes that a logarithmic space Turing machine has another separated read-only tape, that is called "read-once", where the head never moves to the left on that special tape [2].

**Definition 1.1.** A language  $L_1$  is in NL if there exists a deterministic logarithmic space Turing machine M with an additional special read-once input tape polynomial  $p : \mathbb{N} \to \mathbb{N}$  such that for every  $x \in \{0, 1\}^*$ :

$$x \in L_1 \Leftrightarrow \exists u \in \{0,1\}^{p(|x|)}$$
 then  $M(x,u) = "yes"$ 

where by M(x, u) we denote the computation of M, x is placed on its input tape, the certificate string u is placed on its special read-once tape, and M uses at most  $O(\log |x|)$  space on its read/write tapes for every input x where  $|\ldots|$  is the bit-length function. The Turing machine M is called a logarithmic space verifier.

An oracle Turing Machine M has an additional tape, the oracle tape, and three states  $q_?$ ,  $q_{yes}$  and  $q_{no}$  [7]. When M enters  $q_?$  (M is said to query the oracle), then M goes to the state  $q_{yes}$  or the state  $q_{no}$  according to whether the string written in the oracle tape belongs or does not belong to a set called the oracle [7]. A language accepted by an oracle Turing Machine M with oracle A is denoted by  $L^A(M)$  [7]. The class of languages accepted by deterministic and nondeterministic oracle Turing Machine Mworking in space S(n), with oracle A, is denoted by  $DSPACE^A(S(n))$  and  $NSPACE^A(S(n))$ , respectively [7]. We state the following Hypothesis:

**Hypothesis 1.2.** There is a nonempty language  $L_2 \in 1L$ , such that there is another language  $L_3$  which is closed under logarithm space reductions in NP-complete with a deterministic **square logarithmic** space Turing machine M using an additional special read-once input tape polynomial  $p: \mathbb{N} \to \mathbb{N}$ , where:

 $L_3 = \{w : M(w, u) = y, \exists u \in \{0, 1\}^{p(|w|)} \text{ such that } y \in L_2\}$ 

when by M(w, u) we denote the computation of M, w is placed on its input tape, and the certificate string u is placed on the special read-once tape of M. In this way, there is a NP-complete language defined by a **square logarithmic** space verifier M such that when the input is an element of the language, then there exists a certificate u such that M outputs a string which belongs to a single language in 1L.

We show the principal consequences of this Hypothesis:

#### **Theorem 1.3.** If the Hypothesis 1.2 is true, then $NP \subseteq NSPACE(\log^2 n)$ .

Proof. We can simulate the computation M(w, u) = y in the Hypothesis 1.2 by a nondeterministic square logarithmic space oracle Turing machine N such that the string y is written in the oracle tape in the computation of N(w), since we can read the certificate string u within the read-once tape by a work tape in a nondeterministic logarithmic space generation of symbols contained in u [8]. Certainly, we can simulate the reading of one symbol from the string u into the read-once tape just nondeterministically generating the same symbol in the work tapes using a logarithmic space [8]. We could remove each symbol or a logarithmic amount of symbols generated in the work tapes, when we try to generate the next symbol contiguous to the right on the string u. In this way, the generation will always be in logarithmic space. This proves that  $L_3$  is in  $NSPACE(\log^2 n)^{1L}$  since the string y written in the oracle tape is queried whether  $y \in L_2$  or not. That is equivalent to say that  $L_3$  is in  $NSPACE(\log^2 n)$  when the Hypothesis 1.2 is true, since  $NSPACE(\log^2 n)^{1L} = NSPACE(\log^2 n)$  as a consequence of that  $NSPACE(\log^2 n)$  is closed under 1L-reductions [7]. Due to  $L_3$  is closed under logarithm space reductions in NP-complete, then every NP problem is logarithmic space reduced to  $L_3$ . This implies that  $NP \subseteq NSPACE(\log^2 n)$  since  $NSPACE(\log^2 n)$  is closed under logarithm space reductions as well.

#### 1.2 The Problems

Now, we define the problems that we are going to use.

#### Definition 1.4. ONE-IN-THREE 3-TIMES 3SAT (3T-3SAT)

INSTANCE: A Boolean formula  $\phi$  in CNF such that each variable appears exactly thrice with two positive and one negated literal occurrences.

QUESTION: Is there a truth assignment such that each clause has at least one true literal?

REMARKS: Every single certificate must contain all the variables of  $\phi$ and if we eliminate one single variable from that truth assignment, then this stops of being a certificate for the instance.

#### Definition 1.5. 1SUM

INSTANCE: An unary string  $0^K$  and a collection B of positive integers such that every element in B has the same bit-length |K| where  $|\ldots|$  means the bit-length function.

QUESTION: Is the sum of all elements of the collection B equal to K?

### 2 Results

#### **Theorem 2.1.** 3T- $3SAT \in NP$ -complete.

*Proof.* Consider the problem *ONE-IN-THREE 3SAT* with no negated literals. That is almost 3T-3SAT but the instances in *ONE-IN-THREE 3SAT* are Boolean formulas in 3CNF. We know that *ONE-IN-THREE 3SAT*  $\in$  *NP-complete* [5]. Consider an instance  $\psi$  of *ONE-IN-THREE 3SAT* in which the variable x appears k times. So, we replace the first occurrence of x by  $x_1$ , the second by  $x_2$  and so on, where  $x_1, x_2, \ldots, x_k$  are k new variables. Next, we add the expression

$$(\neg x_1 \lor x_2) \land (\neg x_2 \lor x_3) \land \ldots \land (\neg x_k \lor x_1)$$

which is equivalent to

$$x_1 \Rightarrow x_2 \Rightarrow \ldots \Rightarrow x_1.$$

Note that, each clause above has fewer than 3 literals. The final result satisfies the condition on the selected variable x. Suppose we are given in  $\psi$  the expression

$$\dots (u \lor v \lor x) \dots \land \dots (x \lor y \lor z) \dots$$

and therefore, the transformed expression into another Boolean formula  $\phi$  would be

$$\dots (u \lor v \lor x_1) \dots \land \dots (x_2 \lor y \lor z) \dots \land \dots (\neg x_1 \lor x_2) \land (\neg x_2 \lor x_3) \land \dots$$

where the variable  $x_2$  appears exactly thrice since  $x_2$  appears twice and  $\neg x_2$  appears once. Consider an instance  $\psi$  of *ONE-IN-THREE 3SAT* in which the variable x appears once. In this case, we add the another new clause  $(x \lor \neg x)$  into  $\phi$ . In this way, the variable x appears exactly thrice since x appears twice and  $\neg x$  appears once. Hence, we proceed with this subroutine for each variable in  $\psi$  from *ONE-IN-THREE 3SAT* to finally obtain the equivalent instance  $\phi$  in 3T-3SAT.

#### Theorem 2.2. $1SUM \in 1L$ .

*Proof.* Given an collection of positive integers B, we can read its elements from left to right, check that every element in B has the same bit-length |K|, sum them one by one into a single value and compare whether this calculated value is equal to K. We can make all this computation in a deterministic one-way using logarithmic space. Certainly, we can count and store the number of bits of each element of the collection that we read from the input and check whether they are all equal to the unique bit-length |K|. Indeed, we never need to read to the left on the input for the acceptance of the elements in 1SUM in a deterministic logarithmic space.

**Theorem 2.3.** There is a deterministic square logarithmic space Turing machine M, where:

 $3T-3SAT = \{w : M(w, u) = y, \exists u \text{ such that } y \in 1SUM\}$ 

when by M(w, u) we denote the computation of M, w is placed on its input tape, u is placed on the special read-once tape of M, and u is polynomially bounded by w.

*Proof.* The input could be a Boolean formula  $\phi$  in CNF such that each variable appears exactly thrice with two positive and one negated literal occurrences. The Boolean formula  $\phi$  contain n variables and m clauses. We can create a certificate array A which contains indexes values that represents the position of exactly one literal per clause. We read at once the indexes values of the array A and we reject when this index is out of range in relation to the clause in the  $i^{th}$  position. Besides, we check that the array contains exactly m element: one index per clause. While we read the indexes values of the array A using every position i, we check those constraints in A and output the number that represent the positive literals concatenated  $\lceil \log_2 n \rceil$  times its binary representation (the function R(x, y) concatenates the binary string x

Algorithm 1 Square logarithmic space verifier with output

```
1: /*A valid instance for 3T-3SAT with its certificate*/
 2: procedure VERIFIER(\phi, A)
        /*Initialize the number of variables*/
 3:
 4:
        n \leftarrow number-of-variables(\phi)
        /*Output the value of 0^{K*}/
 5:
        output 0^{2 \cdot \sum_{i=1}^{n} R(i, \lceil \log_2 n \rceil)}
 6:
        /*Initialize the number of clauses*/
 7:
        m \leftarrow number-of-clauses(\phi)
 8:
        /*Output the open square bracket of collection B^*/
 9:
        output,
10:
        /*Iterate for the elements of the certificate array A^*/
11:
12:
        for i \leftarrow 1 to m do
            /*Assign the current index*/
13:
            j \leftarrow A[i]
14:
           if j = null then
15:
                return "no"
16:
            else if j > number-of-literals(c_i) \lor j < 1 then
17:
                return "no"
18:
            else if i = m \wedge A[i+1] \neq null then
19:
                return "no"
20:
            else if c_i[j] < 0 then
21:
                /*We fill it by zeroes until the bit-length of K^*/
22:
                output fill-by-zeroes(2 \cdot R(-c_i[j], \lceil \log_2 n \rceil))
23:
            else
24:
                /*We fill it by zeroes until the bit-length of K^*/
25:
                output fill-by-zeroes(R(c_i[j], \lceil \log_2 n \rceil))
26:
            end if
27:
            if i < m then
28:
                output,
29:
30:
            else
31:
                /*Output the close square bracket of collection B^*/
                output ]
32:
            end if
33:
        end for
34:
35: end procedure
```

just y times) and the absolute value for negated literals concatenated  $\lceil \log_2 n \rceil$  times its binary representation with another 0 bit behind (which means the concatenation result is multiplied by 2) just assuming the literals are defined as integer according to the DIMACS files representation as input (http://www.satcompetition.org/2009/format-benchmarks2009.html).

By Theorems 2.1 and 2.2, we obtain that for all:

$$\phi \in 3T\text{-}3SAT \Leftrightarrow \exists A \text{ such that } (0^K, B) \in 1SUM$$

with  $K = 2 \cdot \sum_{i=1}^{n} R(i, \lceil \log_2 n \rceil)$  such that when  $(0^K, B) \in 1SUM$ , then we guarantee do not output the positive and negated literal of a single variable, we indeed do output all variables and output always one single value for each clause. To sum up, we can create this verifier that only uses a **square logarithmic** space in the work tapes such that the array A is placed on the special read-once tape, because we read at once the indexes values in the array A. Hence, we only need to iterate from the cells of the array A to verify whether the array is an appropriated certificate according to the described constraints and check that every index j is correct.

This square logarithmic space verifier with output will be the Algorithm 1. We introduce some constraints in the Algorithm 1 in order to guarantee the algorithmic procedure. For example, we assume that a value does not exist in the array A into a cell of some position i when A[i] = null. In addition, we immediately reject when the mentioned comparisons between the indexes values j and the size of the clause do not hold at least into one single binary digit. That means the machine enters into the rejecting state when the certificate is not valid. Remember that, we assume the variables are between 1 and n due to the DIMACS files representation as input.

## **Theorem 2.4.** $NP \subseteq NSPACE(\log^2 n)$ .

*Proof.* This is a directed consequence of Theorems 1.3 and 2.3 because of the Hypothesis 1.2 is true. Certainly, 3T-3SAT is closed under logarithm space reductions in NP-complete. Indeed, we can reduced SAT to 3T-3SAT in logarithmic space and every NP problem could be logarithmic space reduced to SAT by the Cook's Theorem Algorithm [5].

## 3 Conclusions

The significance of NP being a subset of  $NSPACE(\log^2 n)$  would be immense, as it would imply that all problems in NP can be solved using a poly-logarithmic amount of space. This would have important implications for the study of complexity theory and the development of algorithms for practical applications. This would have important implications for a wide range of fields, including computer science, mathematics, physics, and engineering.

## Acknowledgments

The author thanks Emmanuel (CEO of NataSquad) for the financial support.

## References

- Scott Aaronson. P <sup>?</sup>= NP. Electronic Colloquium on Computational Complexity, Report No. 4, 2017.
- [2] Sanjeev Arora and Boaz Barak. Computational complexity: a modern approach. Cambridge University Press, USA, 2009.
- [3] Stephen Arthur Cook. The P versus NP Problem. https:// www.claymath.org/wp-content/uploads/2022/06/pvsnp.pdf, April 2000. Clay Mathematics Institute. Accessed 9 January 2023.
- [4] Thomas Cormen, Charles Eric Leiserson, Ronald Linn Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, USA, 3rd edition, 2009.
- [5] Michael Randolph Gare and David Stifler Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco: W. H. Freeman and Company, USA, 1 edition, 1979.
- [6] Juris Hartmanis and Stephen Ross Mahaney. Languages Simultaneously Complete for One-Way and Two-Way Log-Tape automata. SIAM Journal on Computing, 10(2):383–390, 1981. doi:10.1137/0210027.
- [7] Pascal Michel. A survey of space complexity. *Theoretical computer science*, 101(1):99–132, 1992. doi:10.1016/0304-3975(92)90151-5.
- [8] Christos Harilaos Papadimitriou. Computational complexity. Addison-Wesley, USA, 1994.
- [9] Walter John Savitch. Relationships between nondeterministic and deterministic tape complexities. Journal of computer and system sciences, 4(2):177-192, 1970. doi:10.1016/S0022-0000(70)80006-X.
- [10] Michael Sipser. Introduction to the Theory of Computation, volume 2. Thomson Course Technology Boston, USA, 2006.

NATASQUAD, 10 RUE DE LA PAIX 75002 PARIS, FRANCE \*1 Email(s): vega.frank@gmail.com \*1 (corresponding author)