ELSEVIER

Artificial Intelligence 92 (1997) 317-329

Artificial Intelligence

Research Note

# The maximum length of prime implicates for instances of 3-SAT

Paul E. Dunne *, Trevor J.M. Bench-Capon

*Department of Computer Science, University of Liverpool, Liverpool, L69 7ZF, UK*

Received August 1996; revised February 1997

**Abstract**

Schrag and Crawford (1996) present strong experimental evidence that the occurrence of prime implicates of varying lengths in random instances of 3-SAT exhibits behaviour similar to the well-known phase transition phenomenon associated with satisfiability. Thus, as the ratio of number of clauses ($m$) to number of propositional variables ($n$) increases, random instances of 3-SAT progress from formulae which are generally satisfiable through to formulae which are generally not satisfiable, with an apparent sharp threshold being crossed when $m/n \sim 4.2$. For instances of 3-SAT, Schrag and Crawford (1996) examine with what probability the longest prime implicate has length $k$ (for $k \geqslant 0$)—unsatisfiable formulae correspond to those having only a prime implicate of length 0—demonstrating that similar behaviour arises. It is observed by Schrag and Crawford (1996) that experiments failed to identify any instance of 3-SAT over nine propositional variables having a prime implicate of length 7 or greater, and it is conjectured that no such instances are possible. In this note we present a combinatorial argument establishing that no 3-SAT instance on $n$ variables can have a prime implicate whose length exceeds $\max\{\lceil n/2\rceil + 1, \lfloor 2n/3\rfloor\}$, validating this conjecture for the case $n = 9$. We further show that these bounds are the best possible. An easy corollary of the latter constructions is that for all $k > 3$, instances of $k$-SAT on $n$ variables can be formed, that have prime implicates of length $n - o(n)$. © 1997 Elsevier Science B.V.

*Keywords:* 3-CNF formulae; Prime implicates; Satisfiability

* Corresponding author. E-mail: ped@csc.liv.ac.uk.


*PII* S0004-3702(97)00018-0

## 1. Definitions and notations

$X_n = \langle x_1, x_2, \ldots, x_n \rangle$ denotes a set of $n$ propositional variables. A *literal* is either a variable $x$ or its negation $\bar{x}$. A clause $C$ is a disjunction of literals; $C$ is said to be *trivial* if it contains both the literal $x$ and its negation, and is *nontrivial* otherwise. A *CNF formula* $\phi$ over $X_n$ is a conjunction of (nontrivial) clauses $\{C_1, C_2, \ldots, C_m\}$. For any CNF formula $\phi$, $f_\phi$ denotes the $n$-variable propositional logic function represented by $\phi$. The decision problem *satisfiability* (SAT) asks whether a given CNF formula $\phi$ is such that there exists any instantiation $\alpha$ of the propositional variables $X_n$ of $\phi$, for which $f_\phi(\alpha) = 1$. A $k$-CNF formula is a CNF formula in which every clause has exactly $k$ literals. The decision problem $k$-SAT is the satisfiability problem restricted to $k$-CNF formulae. Similarly, for integers $1 \leqslant k_1 < k_2 < \cdots < k_r$, a $(k_1, \ldots, k_r)$-CNF is a CNF formula in which the length of any clause is one of $\{k_1, \ldots, k_r\}$.

A CNF formula $\phi(X_n)$ is a *maximal unsatisfiable* formula if $\phi$ is unsatisfiable and $\forall C \in \phi$, the CNF formed by removing $C$ from $\phi$ is satisfiable.

If $f(X_n)$ is a propositional logic function over the variables $X_n$, then a *0-point* of $f$ is an instantiation $\alpha \in \langle 0, 1 \rangle^n$ of the variables such that $f(\alpha) = 0$. A clause $C$ is an *implicate* of $f(X_n)$ if for all instantiations $\alpha$ that yield $C(\alpha) = 0$, such an instantiation renders $f(\alpha) = 0$. A clause $C$ is a *prime implicate* of $f$ if it is an implicate of $f$ and no proper subset of the literals forming $C$ defines an implicate of $f$.

$$k\text{-SAT}(n) \stackrel{\text{def}}{=} \{\phi\text{: } \phi \text{ is an instance of } k\text{-SAT over } X_n\},$$

i.e., $k$-SAT$(n)$ is the set of $k$-CNF formulae with $n$ propositional variables.

$$\mathit{rank}(f) \stackrel{\text{def}}{=} \max\{|C|\text{: } C \text{ is a prime implicate of } f\}.$$

$$r(n,k) \stackrel{\text{def}}{=} \max\{\mathit{rank}(f_\phi)\text{: } \phi \in k\text{-SAT}(n)\}.$$

For $\phi \in (k_1, \ldots, k_r)$-SAT$(n)$, $\phi_i$ denotes the set of clauses $C$ in $\phi$ that contain exactly $i$ literals.

$\lceil x \rceil$ denotes the smallest integer $y$ such that $y \geqslant x$.

$\lfloor x \rfloor$ denotes the largest integer $y$ such that $y \leqslant x$.

## 2. Main result

### 2.1. Preliminaries

In order to obtain the result, we proceed via three main stages. First we show that the value of $r(n,k)$ is exactly determined by a measure defined on maximal unsatisfiable instances of $(1, 2, \ldots, k)$-SAT: this measure is denoted $\mu(n,k)$ when introduced and used subsequently. The remaining parts then deal with proving upper and lower bounds for the specific case of $\mu(n,3)$. Our main result concerning $r(n,3)$ then follows as an easy corollary.

**Definition 2.1.** $\forall n$, $\forall k \geqslant 2$, the measure $\mu(n,k)$ is defined as

$$\max\left\{\sum_{i=1}^{k-1}(k-i)|\phi_i|:\ \phi \in (1,2,\ldots,k)\text{-SAT}(n) \text{ and } \phi \text{ is maximally unsatisfiable}\right\}. \quad (1)$$

The motivation underlying the definition of $\mu(n,k)$ is the following. Consider *any* $\phi \in (1,2,\ldots,k)$-SAT$(n)$ that is maximally unsatisfiable, i.e., $\phi$ need not maximise the sum given in Definition 2.1. We can form a $k$-CNF formula $\psi$ from $\phi$ by introducing new variables into each clause of $\phi$ that contains fewer than $k$ literals. Now, the maximum number of new variables that could be added is $\sum_{i=1}^{k-1}(k-i)|\phi_i|$, since we can use $k-1$ new variables for each clause of length 1 in $\phi$, $k-2$ new variables for each clause of length 2 in $\phi$, etc. Suppose that $\langle y_1, y_2, \ldots, y_s\rangle$ is the set of new literals added to create the $k$-CNF $\psi$ from $\phi$. We can observe two facts about $\psi$: firstly, $\psi$ is an instance of $k$-SAT$(n+s)$; secondly, the clause $y_1 \vee y_2 \vee \cdots \vee y_s$ is a *prime* implicate of $f_\psi$ (a formal proof of the second assertion is given in Lemma 2.2). Thus, the definition of $\mu(n,k)$ can be interpreted as capturing the maximum value $s$, such that a $k$-CNF of $n+s$ variables having a prime implicate of length $s$ can be formed from a $(1,2,\ldots,k)$-CNF of $n$ variables. It follows from these observations that if it is not possible to form a $\phi \in (1,2,\ldots,k)$-SAT$(n)$ which is *both* maximally unsatisfiable *and* "has room" for $s$ variables to be added to make a $k$-CNF formula, i.e., $\sum_{i=1}^{k-1}(k-i)|\phi_i| < s$, then no $k$-CNF of $n+s$ variables can have a prime implicate of length $s$. A formal justification of these claims is given in the following lemma.

**Lemma 2.2.** $\forall n \geqslant k \geqslant 2$, $r(n,k) = \max\{t:\ t \leqslant \mu(n-t,k)\}$.

**Proof.** We first establish that $r(n,k) \geqslant \max\{t:\ t \leqslant \mu(n-t,k)\}$. Let $m$ denote $\max\{t:\ t \leqslant \mu(n-t,k)\}$ and the $n$ propositional variables be partitioned into two sets $y_1,\ldots,y_m$ and $x_1,\ldots,x_{n-m}$. We show how to construct a $k$-SAT instance $\phi(n)$ for which $(y_1 \vee \cdots \vee y_m)$ is a prime implicate. From our definition of $m$, there exists a $(1,2,\ldots,k)$-SAT instance $\psi$ on $n-m$ variables such that $\sum_{i=1}^{k-1}(k-i)|\psi_i| \geqslant m$ and $\psi$ is maximally unsatisfiable. $\phi(n)$ is formed from $\psi$ by adding $k-i$ literals from $\{y_1,\ldots,y_m\}$ to each clause in $\psi_i$. Since $m \leqslant \mu(n-m,k)$ this can be accomplished using all of the $m$ literals $y_i$. The resulting $k$-SAT instance $\phi$ has $y_1 \vee \cdots \vee y_m$ as an implicate, since setting $y_i := 0$, $1 \leqslant i \leqslant m$ yields a formula that is equivalent to $\psi$ and hence unsatisfiable, i.e., equivalent to 0. Furthermore $y_1 \vee \cdots \vee y_m$ must be a *prime* implicate of $\phi(n)$, for if a single $y_i$ is set to 1 with the remainder set to 0, then $\phi(n)$ reduces to a CNF formula whose clauses form a strict subset of the clauses of $\psi$. Such a formula must be satisfiable since $\psi$ is a maximal unsatisfiable instance of SAT.

It remains to show that $r(n,k) \leqslant \max\{t:\ t \leqslant \mu(n-t,k)\}$. Let $\phi \in k$-SAT$(n)$ such that $rank(f_\phi) = r(n,k) = m$. Without loss of generality we can (by relabelling literals and variables) assume that $(x_1 \vee x_2 \vee \cdots \vee x_m)$ is a maximum length prime implicate of $f_\phi$ and that no $\chi \in k\text{-SAT}(n)$ has $rank(f_\chi) = r(n,k)$ and $|\chi| < |\phi|$. Note that a consequence of the latter property is that no clause of $\phi$ contains the literal $\bar{x}_i$, for any $1 \leqslant i \leqslant m$. Let $\delta_m$ denote the partial instantiation of the $n$ variables, given by $\langle x_i := 0\colon\ 1 \leqslant i \leqslant m\rangle$. Consider the formula $\psi \in (1,2,\ldots,k)\text{-SAT}(n-m)$ that results by applying $\delta_m$ to $\phi$, i.e., reducing the number of variables in a clause containing an instance of the literal $x_i$ $(1 \leqslant i \leqslant m)$. The resulting formula $\psi$ must be unsatisfiable since $x_1 \vee \cdots \vee x_m$ is an implicate of $f_\phi$. Furthermore, $\psi$ must be a maximal unsatisfiable instance. For suppose this were not so, and that some clause $C \in \psi$ could be deleted without rendering $\psi - \{C\}$ satisfiable. If $C \in \phi$, i.e., $C$ did not contain any of the literals $x_i$ $(1 \leqslant i \leqslant m)$, then $\chi = \phi - \{C\}$ would still have $rank(f_\chi) = m$ contradicting our assumption that $\phi$ had a minimal number of clauses. If $C \notin \phi$ then $(Q \vee C) \in \phi$ for some subset $Q$ of the literals $\{x_1,\ldots,x_m\}$. Without loss of generality, suppose that $Q = \{x_1,\ldots,x_r\}$, where $r < k$. If, for some $x_i$, $x_1$ say, it is the case that $x_1 \notin C_j$, $\forall C_j \in \phi - \{(Q \vee C)\}$, then

$$f_\phi(\langle x_1 := 1, x_2 := 0, \ldots, x_m := 0\rangle) = f_{\psi-\{C\}} = 0$$

(notice that here we use the property of no clause of $\phi$ containing the literal $\bar{x}_i$ for any $1 \leqslant i \leqslant m$) contradicting the assumption that $\bigvee_{i=1}^{m} x_i$ is a prime implicate. Thus, we may assume that $\forall x_i \in Q$, $\exists C_i \in \phi - \{(Q \vee C)\}$ such that $x_i \in C_i$. Now we get

$$f_{\phi-\{(Q\vee C)\}}(\delta_m) = f_{\psi-\{C\}} = 0$$

which contradicts the choice of $\phi$ having a minimal number of clauses.

So, having established that $\psi$ is a maximal unsatisfiable instance of $(1,2,\ldots,k)$-SAT, it follows that $\sum_{i=1}^{k-1}(k-i)|\psi_i| \leqslant \mu(n-m,k)$. It is also the case, however, that $m \leqslant \sum_{i=1}^{k-1}(k-i)|\psi_i|$, since each clause in $\psi_i$ can account for at most $k-i$ literals in the prime implicate $x_1 \vee \cdots \vee x_m$. We thus have the inequality $m \leqslant \mu(n-m,k)$, and hence $r(n,k) \leqslant \max\{t\colon\ t \leqslant \mu(n-t,k)\}$ as claimed. □

We note, in passing, that the conjecture in [3]—$r(9,3) = 6$—follows quite easily from the characterisation given by Lemma 2.2: suppose that $r(9,3) \geqslant 7$. Lemma 2.2 then implies that either $\mu(2,3) \geqslant 7$ or $\mu(1,3) \geqslant 8$. The latter is clearly impossible, since the only maximal unsatisfiable formula of a single variable is $x \wedge \bar{x}$. If the former were true, then there would be a maximal unsatisfiable $\psi$ say, in $(1,2)\text{-SAT}(2)$ such that $2|\psi_1| + |\psi_2| \geqslant 7$. Now it must be the case that $|\psi_1| \leqslant 2$, otherwise $\psi_1$ contains a literal and its negation and $\psi$ is not maximal. Without loss of generality, suppose that $\psi_1 = x_1 \wedge x_2$. Since $2|\psi_1| + |\psi_2| \geqslant 7$, so $|\psi_2| \geqslant 3$. But if we choose three distinct clauses of length 2 over two variables then at least one must be of the form $(x_1 \vee y)$ or $(x_2 \vee y)$. Again the maximality of $\psi$ is contradicted since we have the two clauses $x_1 \wedge (x_1 \vee y) = x_1$ or $x_2 \wedge (x_2 \vee y) = x_2$. This leaves only the case $|\psi_1| \leqslant 1$, but such would require $|\psi_2| \geqslant 5$: this is impossible since there are only four distinct length 2 clauses of two variables.

**Lemma 2.3.** $\mu(n,3) \geqslant \max\{n+3, 2n\}$.

**Proof.** For $n = 1$ and $n = 2$ the lemma may be verified directly from the $(1,2,3)$-CNF formulae

$$x_1 \wedge \bar{x}_1;\ x_1 \wedge x_2 \wedge (\bar{x}_1 \vee \bar{x}_2).$$

For $n \geqslant 3$, let $\psi(n)$ denote the 2-SAT instance

$$\psi(n) \stackrel{\text{def}}{=} (x_1 \vee x_n) \wedge (\bar{x}_1 \vee \bar{x}_n) \wedge \bigwedge_{i=1}^{n-1} (x_i \vee \bar{x}_{i+1}) \wedge \bigwedge_{i=1}^{n-1} (\bar{x}_i \vee x_{i+1}).$$

We show that $\psi(n)$ is a maximal unsatisfiable instance of 2-SAT$(n)$. First observe that $\psi(n)$ *is* unsatisfiable. For consider any instantiation $\alpha \in \{0,1\}^n$ of its variables. If every variable has the same value instantiation then either the clause $(x_1 \vee x_n)$ or the clause $(\bar{x}_1 \vee \bar{x}_n)$ is false, rendering $\psi(n)$ false. On the other hand, for any instantiation in which some variables take the value 0 and some the value 1, there must be some index $i$, such that the value of $x_i$ differs from the value of $x_{i+1}$. In this case one of the clauses $(x_i \vee \bar{x}_{i+1})$ or $(\bar{x}_i \vee x_{i+1})$ must be 0, again rendering $\psi(n)$ equal to 0. To see that $\psi(n)$ is maximal, consider any clause $C \in \psi(n)$ and the 2-SAT instance $\chi(n,C) = \psi(n) - \{C\}$ obtained by removing $C$ from $\psi(n)$. If $C = (x_1 \vee x_n)$ then the instantiation $x_i := 0$, $\forall 1 \leqslant i \leqslant n$, satisfies $\chi$ since all of the remaining clauses contain at least one instance of a negated literal. A similar argument holds for the case of $C = (\bar{x}_1 \vee \bar{x}_n)$, using the instantiation $x_i := 1$. If $C = (x_i \vee \bar{x}_{i+1})$, then by considering the instantiation $x_j := 0$ $(1 \leqslant j \leqslant i)$, $x_j := 1$ $(i+1 \leqslant j \leqslant n)$, it is easy to see that this satisfies $\chi(n,C)$. Similarly, for $C = (\bar{x}_i \vee x_{i+1})$ the instantiation $x_j := 1$ $(1 \leqslant j \leqslant i)$, $x_j := 0$, $i+1 \leqslant j \leqslant n$ produces a satisfying assignment of $\chi(n,C)$. □

### *2.2. Formula graphs for 2-SAT instances*

By virtue of Lemma 2.2, we wish to establish an upper bound on $2|\phi_1| + |\phi_2|$ for any maximally unsatisfiable $\phi \in (1,2,3)$-SAT$(n)$. In order to do this, we consider two different possibilities for $\phi$: the case when $\phi$ has at least one clause containing a single literal; and the case when every clause of $\phi$ contains at least two literals. The former case yields to a relatively straightforward inductive argument (Case 1 in the proof of Lemma 2.10 below). The second case, however, turns out to be rather more complicated and, in order to complete the upper bound proof, we need to bound the number of clauses of length 2 in maximally unsatisfiable instances of $(2,3)$-SAT$(n)$. This we do in two stages: first considering the case were there are no clauses of length 3, i.e., maximally unsatisfiable instances of 2-SAT$(n)$; and, having dealt with these, maximally unsatisfiable instances of $(2,3)$-SAT$(n)$ containing at least one clause of length 3. Lemma 2.7 and Case 1 of Lemma 2.9 deal with the former instances; Lemma 2.8 and Case 2 of Lemma 2.9 are directed towards the latter class of formulae.

To assist in deriving the bounds needed for each of the $(2,3)$-SAT$(n)$ cases, we use the well-known concept of *formula graphs*.

**Definition 2.4.** Let $\phi \in$ 2-SAT$(n)$. The *formula graph* of $\phi$—$G_\phi(V,E)$—is a directed graph with $2n$ vertices corresponding to the $2n$ possible literals. There is a directed edge $(x,y)$ in $E$ if and only if the clause $(\bar{x} \vee y) \in \phi$. Thus, each clause $(x \vee y) \in \phi$ generates exactly two edges in $G_\phi(V,E)$: $(\bar{x},y)$ and $(\bar{y},x)$. If $Q$ is a directed path of literals in $G_\phi$ then $\bar{Q}$ denotes the path in which each literal in $Q$ is negated and the direction of each edge in $Q$ is reversed. The definition of formula graph implies that the path $Q$ exists if and only if the path $\bar{Q}$ exists.

Formula graphs were introduced in [1], where they are used as the basis of a linear time algorithm for 2-SAT. Further use is made of these graphs in [2], where a sharp satisfiability threshold is exhibited for 2-SAT instances, by analysing combinatorial properties of random formula graphs rather than random instances of 2-SAT. We rely on some basic facts and constructions from [1] in our subsequent development.

**Definition 2.5.** Let $G_\phi(V,E)$ be a formula graph.

A *strongly connected component* of $G_\phi$ is a subgraph $S(W,F)$ of $G_\phi$ with the following properties: $W \subseteq V$; $F = (W \times W) \cap E$; $\forall x,y \in W$ $(x \neq y)$, there are directed paths from $x$ to $y$ and from $y$ to $x$ in $G_\phi$; $\forall x \in W$, $\forall z \notin W$, either there is no directed path from $x$ to $z$ or there is no directed path from $z$ to $x$ in $G_\phi$. For any strongly connected component $S(W,F)$ of $G_\phi$ there exists a *complementary component* $\bar{S}(\bar{W},\bar{F})$ in $G_\phi$, formed by negating the literals in $W$ and reversing the direction of each edge in $F$. $S$ and $\bar{S}$ may not be distinct.

A *contradictory cycle* is a directed cycle in $G_\phi$ such that both the literals $x_i$ and $\bar{x}_i$ appear on the cycle, for some variable $x_i$. An instantiation $\alpha \in \{0,1\}^n$ of the variables of $\phi$ is *inconsistent with* $G_\phi$ if there is a directed path from a literal whose value is 1 under $\alpha$ to a literal whose value is 0 under $\alpha$.

The strongly connected components of $G_\phi$—$S_1,\ldots,S_k$—induce a partition of the $2n$ literals of $\phi$ into $k$ sets. If these components are regarded as "super-vertices"—$V_1,\ldots,V_k$—then the edges of $G_\phi$ that do not belong to any $S_j$, i.e., connect a literal in one component to a literal in another, define a directed acyclic graph over these super-vertices.

**Lemma 2.6.** $\forall \phi \in$ 2-SAT$(n)$:

(a) *$\alpha$ is inconsistent with $G_\phi$ if and only if $f_\phi(\alpha) = 0$.*

(b) *$\phi$ is unsatisfiable if and only if $G_\phi$ has a contradictory cycle.*

**Proof.** Both parts are implicit in the analysis of formula graphs from [1]. □

We need a slightly stronger form of Lemma 2.6(b) for our purposes.

**Lemma 2.7.** *If $G_\phi$ has a contradictory cycle then it has a simple contradictory cycle, i.e., one on which any literal occurs at most once.*

**Proof** (*Outline* [1]). Let $P$ be a contradictory cycle in $G_\phi$, such that no other contradictory cycle contains fewer edges than $P$ and let $x$, $\bar{x}$ be complementary literals on $P$. Since $P$ is assumed to contain as few edges as possible, it follows that the paths $x \rightarrow \cdots \rightarrow \bar{x}$ and $\bar{x} \rightarrow \cdots \rightarrow x$ on $P$ are simple paths. Let $V = \{v_1, v_2, \ldots, v_{k-1}, v_k\}$ be the literals that are visited (exactly) twice on $P$; let $Y$ be the literals visited on the simple path from $x$ to $\bar{x}$ but not on the path from $\bar{x}$ to $x$; and $Z$ be the literals visited on the simple path from $\bar{x}$ to $x$ but not on the path from $x$ to $\bar{x}$. The cycle $P$ may be expressed as:

$$P = x \rightarrow R \rightarrow \bar{x} \rightarrow S \rightarrow x$$

where

$$R = \{x \rightarrow\} Y_1 \rightarrow v_1 \rightarrow Y_2 \rightarrow v_2 \rightarrow \cdots \rightarrow Y_{k-1} \rightarrow v_{k-1} \rightarrow Y_k \rightarrow v_k \rightarrow Y_{k+1} \{\rightarrow \bar{x}\},$$

$$S = \{\bar{x} \rightarrow\} Z_{k+1} \rightarrow w_k \rightarrow Z_k \rightarrow w_{k-1} \rightarrow Z_{k-1} \rightarrow \cdots$$
$$\rightarrow w_2 \rightarrow Z_2 \rightarrow w_1 \rightarrow Z_1 \{\rightarrow x\},$$

where $\{w_1, \ldots, w_k\} = \{v_1, \ldots, v_k\}$; $\{Y_i\colon\ 1 \leqslant i \leqslant k+1\}$ is a partition of $Y$ into $k+1$ subsets (some of which may be empty); and, similarly, $\{Z_j\colon\ 1 \leqslant j \leqslant k+1\}$ is a partition of $Z$ into $k+1$ subsets. Now, since $G_\phi$ is a formula graph, the existence of a path $u \rightarrow Q \rightarrow t$ implies the existence of a path $\bar{t} \rightarrow \bar{Q} \rightarrow \bar{u}$. It follows, from the analysis of $P$ that $G_\phi$ contains paths:

$$\bar{R} = \{x \rightarrow\} \bar{Y}_{k+1} \rightarrow \bar{v}_k \rightarrow \bar{Y}_k \rightarrow \bar{v}_{k-1} \rightarrow \cdots \rightarrow \bar{Y}_2 \rightarrow \bar{v}_1 \rightarrow \bar{Y}_1 \{\rightarrow \bar{x}\},$$

$$\bar{S} = \{\bar{x} \rightarrow\} \bar{Z}_1 \rightarrow \bar{w}_1 \rightarrow \bar{Z}_2 \rightarrow \bar{w}_2 \rightarrow \cdots \rightarrow \bar{Z}_k \rightarrow \bar{w}_k \rightarrow \bar{Z}_{k+1} \{\rightarrow x\},$$

cf. Fig. 1, where for simplicity we have assumed that $v_i = w_i$, and reduced each $Y_j$, $Z_j$ to a single literal in each subset.

Now we can construct a cycle

$$x \rightarrow \bar{Y}_{k+1} \rightarrow \bar{v}_k \rightarrow \bar{Y}_k \rightarrow \bar{v}_{k-1} \rightarrow \cdots \rightarrow \bar{v}_1 \rightarrow \bar{Y}_1 \rightarrow \bar{x}$$
$$\rightarrow Z_{k+1} \rightarrow w_k \rightarrow Z_k \rightarrow w_{k-1} \rightarrow \cdots \rightarrow w_2 \rightarrow Z_2 \rightarrow w_1 \rightarrow Z_1 \rightarrow x$$

(cf. the corresponding cycle in Fig. 1). This cycle contains the same number of edges as $P$. If it is not a simple cycle, then one of the new complementary literals introduced must be the same as one of the literals in $P$. This, however, would imply that a contradictory cycle containing fewer edges than $P$ could be constructed, contradicting our initial choice. □

**Lemma 2.8.** *Let $\phi = \psi \wedge \chi \in (2,3)$-SAT$(n)$ be a maximal unsatisfiable formula, where $\phi_2 = \psi$, $\phi_3 = \chi$ and $|\phi_3| > 0$. Let $S(W, F)$ be any strongly connected component of $G_\psi$. Then $|F| \leqslant 2(|W| - 1)$.*

[1] This result is not given in [1] (where it is not needed for the algorithm presented); the "normal form" presented in [2] appears to assume its correctness, but gives no explicit proof to that effect.

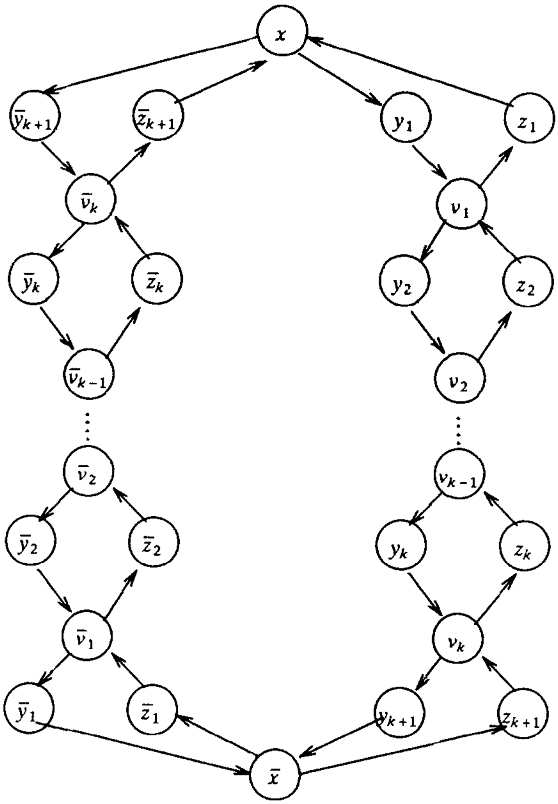


Fig. 1. Construction of a simple contradictory cycle.

**Proof.** We may assume that $|W| > 1$, otherwise the result is immediate. Consider any $x \in W$. Using $x$ as the root we can build a directed tree that describes simple paths from $x$ to any other literal in $W$, e.g. form a breadth-first spanning tree whose first level consists of those literals $y$, such that $(x, y)$ is an edge of $S$, and whose remaining levels are formed by expanding each first level literal in the same way, the process continuing until all literals in $S$ have been accounted for: call this tree $T^+(x)$. Similarly, we can build, starting from edges directed into $x$, a directed tree that describes simple paths to $x$ from any other literal in $W$: call this tree $T^-(x)$. Since $T^+$ and $T^-$ are trees on $|W|$ vertices, in total they contain at most $2(|W| - 1)$ edges. Note that some edges of $S(W, F)$ may appear in both $T^+(x)$ and $T^-(x)$ so that this bound is not necessarily exact. Furthermore, for every distinct pair of literals $y$, $z$ in $W$ we have paths $y \Rightarrow x$, $z \Rightarrow x$ in $T^-$ and paths $x \Rightarrow y$, $x \Rightarrow z$ in $T^+$. Hence the edges of these trees preserve the strongly connected structure of $S$, i.e., we have both a directed path from $y$ to $z$ and a directed path from $z$ to $y$. Let $\bar{T}^+(\bar{x})$ and $\bar{T}^-(\bar{x})$ be the corresponding structures in the complementary component $\bar{S}$ of $S$. (Note that $S$ and $\bar{S}$ are disjoint from the assumptions that $\phi$ is maximally unsatisfiable and $|\phi_3| > 0$.) Suppose that there is some edge $(u, v)$ in $S$ that is not contained in either $T^+$ or $T^-$, and hence the edge $(\bar{v}, \bar{u})$ is not contained in $\bar{T}^+$ or $\bar{T}^-$. Such an edge corresponds to a clause $C = (\bar{u} \vee v) \in \psi$. We claim that in this case the (2,3)-SAT instance $\phi - C$ would be unsatisfiable. To see this consider any $\alpha \in \{0,1\}^n$: since $\phi$ is unsatisfiable it follows that either $f_\chi(\alpha) = 0$ or $f_\psi(\alpha) = 0$. In the former case, $f_{\phi - C}(\alpha) = 0$, and so we can assume that $f_\chi(\alpha) = 1$, $f_\psi(\alpha) = 0$. From Lemma 2.6(a) this means that $\alpha$ is inconsistent with $G_\psi$. But then $\alpha$ must also be inconsistent with $G_{\psi - C}$ since our construction of $T$ and $\bar{T}$ means that there is a directed path from $x$ to $y$ in $G_\psi$ if and only if there is a directed path from $x$ to $y$ in $G_{\psi - C}$. It follows that every strongly connected component $S(W, F)$ contains no more than $2(|W| - 1)$ edges as claimed. □

**Lemma 2.9.** *$\forall n \geqslant 2$, if $\phi \in$ (2,3)-SAT$(n)$ is maximally unsatisfiable then $|\phi_2| \leqslant 2n$.*

**Proof.** Let $\phi \in$ (2,3)-SAT$(n)$ that is maximally unsatisfiable.

*Case* 1: $|\phi_3| = 0$. So, $\phi \in$ 2-SAT$(n)$, and since $\phi$ is unsatisfiable it follows, from Lemma 2.7, that $G_\phi$ contains a simple contradictory cycle. Such a cycle can contain at most $2n$ literals, thus be formed from at most $2n$ different edges (i.e., clauses of $\phi$). So if $|\phi| > 2n$, we could remove some clause of $\phi$ without affecting the contradictory cycle within $G_\phi$. This, however, would mean that $\phi$ was not maximal.

*Case* 2: $|\phi_3| > 0$. Consider the formula graph of $\phi_2$, $G_{\phi_2}(V, E)$. Suppose that this has $k$ strongly connected components $S_i(W_i, F_i)$, $1 \leqslant i \leqslant k$. We know, from the definition of formula graph, that $|\phi_2| = |E|/2$. However,

$$
\begin{aligned}
|E| &= \sum_{i=1}^{k} |F_i| + \left| \left\{ (v,w) \in E\colon \ (v,w) \notin \bigcup_{i=1}^{k} F_i \right\} \right| \\
&\leqslant \sum_{i=1}^{k} 2(|W_i| - 1) + \left| \left\{ (v,w) \in E\colon \ (v,w) \notin \bigcup_{i=1}^{k} F_i \right\} \right| \\
&= 2\sum_{i=1}^{k} |W_i| - 2k + \left| \left\{ (v,w) \in E\colon \ (v,w) \notin \bigcup_{i=1}^{k} F_i \right\} \right| \\
&= 4n - 2k + \left| \left\{ (v,w) \in E\colon \ (v,w) \notin \bigcup_{i=1}^{k} F_i \right\} \right|.
\end{aligned}
$$

Without loss of generality, we can assume that $\phi = \phi_2 \wedge (x_1 \vee x_2 \vee x_3) \wedge \psi$, where $\phi_3 = (x_1 \vee x_2 \vee x_3) \wedge \psi$. Since $\phi$ is maximal, it follows that $\phi_2 \wedge \psi$ is satisfiable, and furthermore, every satisfying assignment must have $(x_1 \vee x_2 \vee x_3) = 0$. So, consider any assignment $\alpha \in \{0,1\}^n$, and the "*cross-component*" edges—$\{(v,w) \in E\colon \ (v,w) \notin \bigcup_{i=1}^{k} F_i\}$. If $f_\psi(\alpha) = 1$ then either $\alpha$ must be inconsistent with $G_{\phi_2}$, or $f_{(x_1 \vee x_2 \vee x_3)}(\alpha) = 0$. It follows that cross-component edges which do not essentially contribute to forming inconsistent paths from the literals $x_1$, $x_2$, $x_3$ may be deleted from $G_{\phi_2}$ and the corresponding clauses from $\phi_2$. Recalling that the cross-component edges define a directed acyclic graph over the components, a straightforward but lengthy analysis of the differing possible relationships between the host components of $x_1$, $x_2$, and $x_3$, establishes that if there are more than $2k$ such edges, then some are redundant. [2] In consequence, we have that $|E| \leqslant 4n$ and hence $|\phi_2| \leqslant 2n$. □

### 2.3. Combining the results

**Lemma 2.10.** $\forall n \geqslant 1$, $\mu(n,3) \leqslant \max\{n+3, 2n\}$.

**Proof.** By induction on $n$. The base case, $n = 1$, is immediate from the fact that the only unsatisfiable CNF formula of a single variable is the CNF $x \wedge \bar{x}$.

Inductively assume that the lemma holds for all values $1 \leqslant m < n$, and let $\phi \in (1,2,3)\text{-SAT}(n)$ such that $\phi$ is maximally unsatisfiable. We first observe that for each of the $n$ variables $x_i$, there must be a clause of $\phi$ containing the literal $x_i$ and a clause of $\phi$ containing the literal $\bar{x}_i$.

(For suppose this were not so: then either no clause of $\phi$ involves the variable $x_i$; or exactly one of the literals $x_i/\bar{x}_i$ occurs in $\phi$. In the former instance, $\phi \in (1,2,3)\text{-SAT}(n-1)$ and so from the inductive hypothesis

$$2|\phi_1| + |\phi_2| \leqslant \max\{n+2, 2n-2\} < \max\{n+3, 2n\}$$

as required. In the latter case, suppose without loss of generality, that the variable $x_1$ occurs only in positive form in $\phi$. We may express $\phi$ as

$$\left(x_1 \vee \phi^{(1)}\right) \wedge \phi^{(2)}$$

where $\phi^{(1)} \in (1,2)\text{-SAT}(n-1)$ and $\phi^{(2)} \in (1,2,3)\text{-SAT}(n-1)$, i.e., $\phi^{(1)}$ consists of those clauses of $\phi$ that depend on $x_1$ and $\phi^{(2)}$ is formed from the remaining clauses of $\phi$. Since $\phi$ is maximally unsatisfiable and $\phi^{(2)} \subset \phi$ it follows that there is an instantiation $\beta \in \{0,1\}^{n-1}$ of the variables of $\phi^{(2)}$ for which $f_{\phi^{(2)}}(\beta) = 1$. But now we obtain the contradiction that the instantiation $\langle x_1 := 1, \beta\rangle$ satisfies $\phi$.)

*Case* 1: $|\phi_1| \geqslant 1$. Without loss of generality, let $x_1 \in \phi_1$. We may express $\phi$ as

$$x_1 \wedge \left(\bar{x}_1 \vee \phi^{(1)}\right) \wedge \phi^{(2)}$$

where $\phi^{(1)} \in (1,2)\text{-SAT}(n-1)$ is formed from those clauses of $\phi$ in which the literal $\bar{x}_1$ appears, and $\phi^{(2)} \in (1,2,3)\text{-SAT}(n-1)$ consists of those clauses of $\phi$ in which the variable $x_1$ does not appear. We claim that $\psi \in (1,2,3)\text{-SAT}(n-1)$ defined by $\phi^{(1)} \wedge \phi^{(2)}$ is maximally unsatisfiable. Certainly, this formula is unsatisfiable, for if $\beta \in \{0,1\}^{n-1}$ satisfied $\psi$ then $\langle x_1 := 1, \beta\rangle$ would satisfy $\phi$. If $\psi$ is not maximal then there is a clause $C$ of $\psi$ such that the instance $\psi - C$ is unsatisfiable. If $C \in \phi^{(1)}$ then the clause $(\bar{x}_1 \vee C) \in \phi$. Consider the CNF $\phi - (\bar{x}_1 \vee C)$. This is satisfiable

[2] The analysis considers a spanning forest of the component, formed using only the cross-component edges (which contains at most $k-1$ edges). Such a forest accounts for all of the inconsistent paths that can arise from a single literal, $x_1$. To account for the inconsistent paths from $x_2$ and $x_3$ an inductive argument over the number of components ($k$) can be used to show that these can only take a small number of additional edges.

since $\phi$ is maximal, and any satisfying assignment has the form $\langle x_1 := 1, \beta\rangle$ for some $\beta \in \{0,1\}^{n-1}$. But now we obtain the contradiction that

$$f_{\phi-(\bar{x}_1 \vee C)}(x_1 := 1, \beta) = 1,$$
$$f_{\phi-(\bar{x}_1 \vee C)}(x_1 := 1, \beta) = f_{\psi-C}(\beta) = 0.$$

The case where $C \in \phi^{(2)}$ results in a similar contradiction.

Thus, since $\psi = \phi^{(1)} \wedge \phi^{(2)}$ is a maximal unsatisfiable instance in $(1,2,3)$-SAT$(n-1)$, from the inductive hypothesis it follows that

$$2|\psi_1| + |\psi_2| \leqslant \max\{n+2, 2n-2\}.$$

However,

$$|\psi_1| = |\phi_1| - 1 + |\{C \in \phi_2\colon\ C = (\bar{x}_1 \vee y)\}|,$$
$$|\psi_2| = |\phi_2| - |\{C \in \phi_2\colon\ C = (\bar{x}_1 \vee y)\}| + |\{C \in \phi_3\colon\ C = (\bar{x}_1 \vee y \vee z)\}|.$$

Hence,

$$2|\psi_1| + |\psi_2| = 2|\phi_1| + |\phi_2| + |\{C \in \phi\colon\ \bar{x}_1 \text{ occurs in } C\}| - 2.$$

Rearranging this and using the upper bound on $2|\psi_1| + |\psi_2|$ yields

$$2|\phi_1| + |\phi_2| \leqslant \max\{n+4, 2n\} - |\{C \in \phi\colon\ \bar{x}_1 \text{ occurs in } C\}|.$$

Since there must be at least one clause of $\phi$ containing the literal $\bar{x}_1$, we have the upper bound required for this case of the inductive argument.

*Case* 2: $|\phi_1| = 0$. In this case, we have $\phi \in (2,3)$-SAT$(n)$, and $\phi$ is maximally unsatisfiable. From Lemma 2.9, $|\phi_2| \leqslant 2n$. If $n = 2$, then any $\phi \in (1,2)$-SAT$(2)$, for which $2|\phi_1| + |\phi_2| > 4$, must either contain a clause of length 1—i.e., Case 1 applies—or is not maximally unsatisfiable. □

**Theorem 2.11.** $\forall n \geqslant 3$, $r(n,3) = \max\{\lceil n/2\rceil + 1, \lfloor 2n/3\rfloor\}$.

**Proof.** From Lemmas 2.2, 2.3, and 2.10 we have

$$r(n,3) = \max\{t\colon\ t \leqslant \max\{n-t+3, 2(n-t)\}\}.$$

The inequality $t \leqslant n-t+3$ yields $t \leqslant \lceil n/2\rceil + 1$; similarly, the inequality $t \leqslant 2(n-t)$ gives $t \leqslant \lfloor 2n/3\rfloor$. Combining these proves the theorem. □

## 3. Prime implicates in $k$-SAT instances for $k > 3$

For completeness, we present in this section bounds on $r(n,k)$ for fixed $k \geqslant 4$. These are obtained by bounding $\mu(n,k)$.

**Theorem 3.1.** $\forall k \geqslant 4$, $\mu(n,k) = \Omega(n^{k-2})$; $\mu(n,k) = O(n^{k-1})$.

**Proof.** For the lower bound, we use a recursive construction, building from the formula $\psi(n)$, described in the proof of Lemma 2.3. We denote by $\phi^{(k)}$ the formula for which the lower bound claimed holds. At every stage of the construction $\phi^{(k)} \in (k-1)\text{-SAT}(3^{k-3}n)$. Since we are assuming that $k$ is constant, the coefficient $3^{k-3}$ is not significant in the analysis.

*Base*: If $k = 3$, the formula $\psi(n)$ of Lemma 2.3 is used.

*Recursive step*: If $k > 3$, let $X$, $Y$, $Z$ be disjoint sets of $N = 3^{k-4}n$ variables, and $\phi^{(k-1)}(Y)$ and $\phi^{(k-1)}(Z)$ be instances of $\phi^{(k-1)}$. Partition the clauses of $\phi^{(k-1)}(Z)$ into $N$ (non-empty) sets $C_i$. $\phi^{(k)}(X,Y,Z)$ is the $k$-CNF formula:

$$\bigwedge_{i=1}^{N} \bigwedge_{C \in C_i} (x_i \vee C) \wedge \bigwedge_{i=1}^{N} \bigwedge_{C \in \phi^{(k-1)}(Y)} (\bar{x}_i \vee C).$$

$\phi^{(k)}(X,Y,Z)$ is unsatisfiable: consider any $\alpha \in \{0,1\}^{3N}$: if any $x_i$ is set to 1 by $\alpha$ then $\phi^{(k)}$ reduces to the unsatisfiable formula $\phi^{(k-1)}(Y)$. If every $x_i$ has the value 0 then $\phi^{(k)}$ reduces to the unsatisfiable formula $\phi^{(k-1)}(Z)$. To see that $\phi^{(k)}$ is maximal, consider the effect of removing any clause $D$. If $D$ is of the form $(\bar{x}_i \vee C)$ then setting $x_j := 0$ (for $j \neq i$) and $x_i := 1$, reduces $\phi^{(k)}$ to $\psi^{(k-1)}(Z) \wedge (\phi_{(k-1)}(Y) - \{C\})$, where $\psi^{(k-1)}$ is a strict subset of the clauses of $\phi^{(k-1)}(Z)$. Since $\phi^{(k-1)}$ is maximal, there are assignments to $Y$ and $Z$ that satisfy $\psi^{(k-1)}$ and $\phi^{(k-1)} - \{C\}$. If $D$ has the form $(x_i \vee C)$ then the assignment $x_i := 0$, $1 \leqslant i \leqslant N$, reduces $\phi^{(k)}$ to the formula $\phi^{(k-1)}(Z) - \{C\}$ which is satisfiable.

Let $\sigma(N,k)$ denote the number of clauses in $\phi^{(k)}$. We then have the recurrence relations: $\sigma(n,3) = 2n$; $\sigma(3N,k+1) = (N+1)\sigma(N,k)$. These yield $\mu(n,k) = \Omega(n^{k-2})$ as claimed.

For the upper bound let $\phi \in (1,2,3,\ldots,k)\text{-SAT}(n)$ be a formula that maximises the value of $\mu(n,k)$. By introducing at most $k-1$ new variables, we may express $\phi$ as $\psi$, an instance of $(k-1,k)\text{-SAT}(n+k-1)$ in which the number of clauses of length $k-1$ is no larger than $2^{k-1}\mu(n,k)$. If

$$|\psi_{k-1}| > (2^{k-1}-1)\binom{n+k-1}{k-1}$$

then $\psi$ must be unsatisfiable, since for any $\alpha \in \{0,1\}^{n+k-1}$ there are exactly $\binom{n+k-1}{k-1}$ clauses of length $k-1$ that are not satisfied by $\alpha$. The total number of possible length $k-1$ clauses is exactly $2^{k-1}\binom{n+k-1}{k-1}$, so if $|\psi_{k-1}|$ exceeds the bound given, then every instantiation $\alpha$ fails to satisfy at least one clause of $\psi$. It is straightforward to convert this bound to give $\mu(n,k) = O(n^{k-1})$. □

**Corollary 3.2.** *For all fixed* $k \geqslant 4$: $r(n,k) \geqslant n - O(n^{1/(k-2)})$; $r(n,k) \leqslant n - \Omega(n^{1/(k-1)})$.

**Proof.** Immediate from Lemma 2.2 and Theorem 3.1. □

## 4. Conclusions

In this note the question of how many variables are contained in the longest prime implicate of an instance of $k$-SAT$(n)$ has been addressed. For the case of 3-SAT instances, exact bounds on this length have been derived, establishing that for $n \geqslant 3$, no instance can have a prime implicate of length exceeding $\max\{\lceil n/2 \rceil + 1, \lfloor 2n/3 \rfloor\}$ and, furthermore, explicit constructions of instances achieving these bounds have been given. The results confirm a conjecture put forward in [3]. For general $k \geqslant 4$, it has been shown that instances with prime implicates of length $n - f(n)$ can be formed where $\lim_{n \to \infty} f(n)/n = 0$. In these cases, a small gap remains between the upper and lower bound results.

## References

[1] B. Aspvall, M.F. Plass and R.E. Tarjan, A linear-time algorithm for testing the truth of certain quantified boolean formulas, *Inform. Process. Lett.* **8** (1979) 121–123.

[2] A. Goerdt, A threshold for unsatisfiability, in: I. Havel and V. Krbec, eds., *Proceedings 17th Symposium on Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science **629** (Springer, Berlin, 1992) 264–274.

[3] R. Schrag and J.M. Crawford, Implicates and prime implicates in random 3-SAT, *Artif. Intell.* **81** (1996) 199–222.