Probabilistic Analysis of a Generalization of the Unit Clause Literal Selection Heuristic for the k-Satisfiability Problem

by

Ming-Te Chao
Case Western Reserve University
Department of Computer Engineering and Science
Cleveland, OH 44106

and

John Franco
Indiana University
Department of Computer Science
Bloomington, IN 47405

TECHNICAL REPORT NO. 165

Probabilistic Analysis of a Generalization of the Unit Clause Literal Selection Heuristic for the k-Satisfiability Problem

by

Ming-Te Chao, Case Western Reserve University

and John Franco, Indiana University

January, 1985

This material is based on work supported by the U.S. Air Force under grant number AFOSR-84-0372.

To appear in Information Sciences.

ABSTRACT

Two algorithms for the k-Satisfiability problem are presented and a probabilistic analysis is performed. The analysis is based on an instance distribution which is parameterized to simulate a variety of sample characteristics. The algorithms assign values to literals appearing in a given instance of k-Satisfiability, one at a time, until a solution is found or it is discovered that further assignments cannot lead to finding a solution. One algorithm chooses the next literal from a unit clause if one exists and randomly from the set of remaining literals otherwise. The other algorithm uses a generalization of the Unit-Clause rule as a heuristic for selecting the next literal: at each step a literal is chosen randomly from a clause containing the least number of literals. The algorithms run in polynomial time and it is shown that they find a solution to a random instance of k-Satisfiability with probability bounded from below by a constant greater than zero for two different ranges of parameter values. It is also shown that the second algorithm mentioned finds a solution with probability approaching one for a wide range of parameter values.

1. Introduction

This paper is concerned with the probabilistic performance of heuristics for the k-Satisfiability problem (k-SAT). k-SAT is the problem of determining whether all of a collection of k-literal disjunctions (clauses) of Boolean variables are *true* for some truth assignment to the variables. This problem is NP-complete so there is no known polynomial time algorithm for solving it. k-SAT is a special case of the Satisfiability problem (SAT) which is the problem of determining whether all of a collection of disjunctions of Boolean variables are *true* for some truth assignment to the variables.

The analysis is based on an equally likely instance distribution which has been used in other studies of algorithms for this problem. This model has two parameters in addition to k: n, the number of disjunctions, and r, the number of variables from which disjunctions are composed. The model (which we refer to as $M(n,r,k)$) is described in greater detail in the next section. In [8] it was shown that, under $M(n,r,k)$, if $\lim_{n,r\to\infty} \frac{n}{r} > 2^k \ln 2$ then random instances have no solution with probability approaching 1. In [1] it was shown that Backtracking solves k-SAT in exponential average time for all limiting ratios of n to r which are constant. In [8] it was shown that a variant of the Davis-Putnam Procedure [4] which searches for all solutions to a given instance requires exponential time in probability under $M(n,r,k)$ for all limiting ratios of n to r which are constant. But, in [6] it was shown that, for all $k \geq 3$, the Pure-Literal heuristic can be used to solve random instances of k-SAT in polynomial time with probability approaching 1 when $\lim_{n,r\to\infty} \frac{n}{r} < 1$. In [3] it was shown that the Unit-Clause and maximum occurring literal heuristics can be used to solve instances of 3-SAT generated according to $M(n,r,3)$ in polynomial time with probability bounded from below by a constant when $\lim_{n,r\to\infty} \frac{n}{r} < 2.9$. In this paper it is shown that a generalization of the Unit-Clause rule can be used as a literal selection heuristic to solve random instances of k-SAT, for $4 \leq k \leq 40$, in polynomial time with probability bounded from below by a constant when $\lim_{n,r\to\infty} \frac{n}{r}$ is less than $\frac{3.09*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$ and with probability approaching 1 when $\lim_{n,r\to\infty} \frac{n}{r}$ is less than $\frac{1.845*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$. A similar analysis shows that, for all $k \geq 3$, the Unit-Clause heuristic alone solves random instances in polynomial time with bounded probability when $\lim_{n,r\to\infty} \frac{n}{r} < \frac{2^{k-1}}{k}\left(\frac{k-1}{k-2}\right)^{k-2}$. These results are useful because they indicate the effectiveness of the two heuristics when used in a Backtrack algorithm for k-SAT.

There are a number of papers which investigate the probabilistic performance of SAT. These papers present results which are closely related to the results obtained for k-SAT but are based on the constant-density model for SAT: construct each of n clauses independently by placing each of r variables independently in a clause with probability p and complementing those variables in each clause with probability 1/2. Average case results using the constant-density model or a variation are in [1], [9], [10], [11] and [12]. Probabilistic results using the constant-density model are in [7]. According to the results in [7], when the average number of literals in a clause is fixed at k, random instances of SAT are nearly always proven to have no solutions in polynomial time.

2. k-Satisfiability and The Probabilistic Model

The following terms are used to describe k-SAT. Let $V = \{v_1, v_2 ... v_r\}$ be a set of r boolean variables. Associated with each variable v_i is a positive literal, denoted by v_i, and a negative literal, denoted by $\bar{v}_i$ and literal v_i has value *true* iff the variable v_i has value *true* and literal $\bar{v}_i$ has value *true* iff the variable v_i has value *false*. The literals v_i and $\bar{v}_i$ are said to be complementary. Denote by L the set $\{v_1, v_2 ... v_r, \bar{v}_1, \bar{v}_2 ... \bar{v}_r\}$ of literals associated with the variables of V. If l is a literal then $comp(l)$ is the literal which is complementary to l. A clause is a subset of L such that no two literals in the subset are complementary. A truth assignment to V is an assignment of truth values to every variable in V. A clause c is satisfied by truth assignment t if at least one literal in c has value *true* under t. Let $A_i(V)$ denote the set of i-literal clauses that can be composed of literals from L. An instance I of k-SAT is a collection of clauses chosen from $A_k(V)$ and the problem is to find a truth assignment to V which satisfies all clauses in I, if one exists, and to verify that no such truth assignment exists otherwise. A truth assignment which satisfies all clauses in I is said to be a solution to I.

The probabilistic model used for analysis is presented by describing the method used to construct random instances. A random instance of k-SAT contains n clauses chosen uniformly, independently and with replacement from $A_k(V)$. The distribution associated with this construction is referred to as $M(n, r, k)$.

3. The Algorithms *UC* and *GUC*

The algorithms we consider, called *UC* and *GUC*, take as input a collection of clauses I and output "a solution exists" or "cannot determine whether a solution exists". Both algorithms contain a single loop; at each iteration of the loop a literal is chosen and some clauses and literals are removed from I. Let $C_i^{I,\sigma}(j)$, for all $1 \le i \le k$, denote the collection of clauses in I containing exactly i literals at the end of the j^{th} iteration where σ denotes the sequence of chosen literals. We shorten $C_i^{I,\sigma}(j)$ to $C_i(j)$. Then $C_i(0) = \phi$ for all $1 \le i \le k-1$ and $|C_k(0)| = n$. If the $j+1^{st}$ chosen literal is l then the lines

> Remove from I all clauses containing l
> Remove from I all occurrences of $comp(l)$

have the following effect

$$\forall\, 1 \le i \le k-1 \;\; C_i(j+1) = \{c : c \in C_i(j) \textit{ and } l \notin C_i(j) \textit{ and } comp(l) \notin C_i(j) \\ \textit{or } c \cup \{comp(l)\} \in C_{i+1}(j)\}$$
$$C_k(j+1) = \{c : c \in C_k(j) \textit{ and } l \notin C_k(j) \textit{ and } comp(l) \notin C_k(j)\}.$$

The algorithm *UC* is based on the Unit-Clause heuristic [4]. The Unit-Clause heuristic requires choosing the $j+1^{st}$ literal from $C_1(j)$ if $C_1(j) \neq \phi$. The algorithm *GUC* employs a heuristic for choosing literals which is a generalization of the Unit-Clause heuristic. The generalization requires choosing the $j+1^{st}$ literal from a smallest clause in I: that is, the $j+1^{st}$ literal is chosen randomly from $C_m(j)$ where $m = \min\{i : C_i(j) \neq \phi\}$.

UC(I):

> $j \leftarrow 0$
> Repeat
>> If $C_1(j) \neq \phi$ Then choose l randomly from $C_1(j)$
>> Else choose l randomly from L
>> $L \leftarrow L - \{l, comp(l)\}$
>> Remove from I all clauses containing l
>> Remove from I all occurrences of $comp(l)$
>> $j \leftarrow j + 1$
>
> Until I is empty or there exist two complementary unit clauses in I
> If I is empty Then Output("a solution exists")
>> Else Output("cannot determine whether a solution exists")

GUC(I):

> $j \leftarrow 0$
> Repeat
>> Let $m = \min\{i : C_i(j) \neq \phi\}$
>> Choose l randomly from $C_m(j)$
>> Remove from I all clauses containing l
>> Remove from I all occurrences of $comp(l)$
>> $j \leftarrow j + 1$
>
> Until I is empty or there exist two complementary unit clauses in I
> If I is empty Then Output("a solution exists")
>> Else Output("cannot determine whether a solution exists")

UC and *GUC* run in less than $O(r^2 n)$ time since I must be empty after r iterations of the loop and the remove operations need look at no more than $r * n$ literals. An instance I of SAT has a solution if *UC* or *GUC* run on I outputs "a solution exists": one solution to I may be found by assigning the value *true* to the variables whose positive literals were chosen and the value *false* to all other variables.

4. Analysis of *UC* and *GUC*

In this section it is shown that if instances are generated according to $M(n,r,k)$ and $\lim_{n,r\to\infty} \frac{n}{r} < \frac{2^{k-1}}{k}\left(\frac{k-1}{k-2}\right)^{k-2}$ then for some $\epsilon > 0$, the probability that UC outputs "a solution exists" is greater than ϵ in the limit. It is also shown that the probability that GUC outputs "a solution exists" is bounded from below by a positive constant ϵ when $\lim_{n,r\to\infty} \frac{n}{r}$ is less than $\frac{3.09*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$ and ϵ approaches 1 in the limit when $\lim_{n,r\to\infty} \frac{n}{r}$ is less than $\frac{1.845*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$ under $M(n,r,k)$. To simplify the analysis it is assumed that both algorithms continue selecting literals and removing clauses and literals from I even if complementary unit clauses appear in I. Also, we will assume that the ratio of n to r is a function of k only and that k is fixed.

The following theorem will be used to show how the collections of clauses in $C_i(j)$ are distributed for both algorithms.

Theorem 1:

Let V_{r-j} be the subset of variables associated with unchosen literals after j literals have been chosen. Suppose for all $1 \le i \le k$ the clauses in $C_i(j)$ are independent and are equally likely to be any clause in $A_i(V_{r-j})$. Then for all $1 \le i \le k$ the clauses in $C_i(j+1)$ are independent and equally likely to be any clause in $A_i(V_{r-j-1})$.

Proof:

Let c_1 be a clause in $C_i(j+1)$ and let $\hat{c}_1$ be the clause in $C_i(j)$ from which c_1 was derived after the $j+1^{st}$ literal was chosen. Let x and y be members of $A_i(V_{r-j-1})$ and let X, X_1, Y and Y_1 be collections of clauses obtained from $A_i(V_{r-j-1})$. Let $|X| = |Y|$. Define the operator $\sqcup$ as follows: if X is as above and l is the chosen literal then $X \sqcup \{l\}$ is a collection of clauses obtained from $A_{i+1}(V_{r-j})$ such that the m^{th} clause of $X \sqcup \{l\}$ contains all the literals of the m^{th} clause of X in addition to the chosen literal l. Denote by $C_i^i(j+1)$ the sub-collection of $C_i(j+1)$ which came from $C_i(j)$ and denote by $C_i^{i+1}(j+1)$ the sub-collection of $C_i(j+1)$ which came from $C_{i+1}(j)$. Denote by $\bar{C}_i(j,l)$ the collection of clauses in $C_i(j)$ not containing literal l or $comp(l)$ and denote by $C_{i+1}(j,l)$ the collection of clauses in $C_{i+1}(j)$ containing literal l. If A and B are collections of clauses and each clause in A can be paired with an identical clause in B then we say that $A \sqsubseteq B$. Then

$pr(c_1 = x) =$

$$pr(\hat{c}_1 = x \textit{ or } \hat{c}_1 = x \cup \{v\} \textit{ and } \bar{v} \text{ was chosen } \textit{or } \hat{c}_1 = x \cup \{\bar{v}\} \text{ and } v \text{ was chosen})$$

$$= pr(\hat{c}_1 = y \textit{ or } \hat{c}_1 = y \cup \{v\} \textit{ and } \bar{v} \text{ was chosen } \textit{or } \hat{c}_1 = y \cup \{\bar{v}\} \text{ and } v \text{ was chosen})$$

$$= pr(c_1 = y).$$

Also, $pr(C_i(j+1) = X) =$

$$\sum_{X_1 \sqsubseteq X} pr(C_i^i(j+1) = X_1 \text{ and } C_i^{i+1}(j+1) = X - X_1)$$

$$= \sum_{X_1 \sqsubseteq X} pr(\bar{C}_i(j,l) = X_1) * pr(C_{i+1}(j,l) = X \sqcup \{l\} - X_1 \sqcup \{l\})$$

$$= \sum_{Y_1 \sqsubseteq Y} pr(\bar{C}_i(j,l) = Y_1) * pr(C_{i+1}(j,l) = Y \sqcup \{l\} - Y_1 \sqcup \{l\})$$

$$= pr(C_i(j+1) = Y).$$

where we have made use of the hypothesis that all clauses in $C_i(j)$ and $C_{i+1}(j)$ are independent and uniform over $A_i(V_{r-j})$ and $A_{i+1}(V_{r-j})$ and the fact that this hypothesis remains true over $A_i(V_{r-j-1})$ and $A_{i+1}(V_{r-j-1})$ for all clauses other than the one from which the chosen literal was taken since the literal was chosen randomly from the set of smallest clauses.

Corollary 1:

For all $0 \le j \le r$ and $1 \le i \le k$ all clauses in $C_i(j)$ are independent and equally likely to be any clause in $A_i(V_{r-j})$.

Proof:

By induction on j. The basis step holds because of the assumed distribution on given instances. The induction step holds because of theorem 1.

Because of corollary 1 a system of differential equations for finding the expected number of clauses in $C_i(j)$ for all $2 \leq i \leq k$ may be obtained. Let $n_i(j)$ denote the number of clauses in $C_i(j)$, let $w_i(j)$ denote the number of i – literal clauses added to $C_i(j)$ as a result of choosing the j^{th} variable and let $z_i(j)$ denote the number of clauses eliminated from $C_i(j)$ as a result of choosing the j^{th} variable. These three terms depend on I and σ but this dependence is omitted from the terms for the sake of simplicity. The $w_i(j)$ term may be thought of as representing the "rate of flow" of clauses into $C_i(j)$ when the j^{th} variable is chosen and the $z_i(j)$ term may be thought of as representing the "rate of flow" of clauses out of $C_i(j)$ when the j^{th} variable is chosen. If the average rate of flow into $C_1(j)$ is always less than 1 the number of clauses in $C_1(j)$ will not, in probability, grow very large since at least one clause is removed from $C_1(j)$ whenever $C_1(j) \neq \phi$. In this case the probability that a complementary pair of clauses exists in $C_1(j)$ for some j is small. However, if the average rate of flow into $C_1(j)$ rises above 1 for a constant fraction of the values of $\frac{j}{r}$ then the number of clauses in $C_1(j)$ gets large for a fraction of the values of $\frac{j}{r}$ since the flow out of $C_1(j)$ is asymptotically no more than one unless $|C_1(j)|$ is large. In this case the probability that there is a complementary pair of clauses in $C_1(j)$ for some j is near 1. Since, as will be seen from the analysis below, if the expected flow into $C_1(j)$ goes above $1+\delta$ for any $\delta > 0$ then it stays above 1 for a constant fraction of values of $\frac{j}{r}$, the conditions under which $E\{w_1(j)\} < 1$ for all $1 \leq j < r$ are the conditions under which UC and GUC find a solution with bounded probability. Furthermore, in the case of GUC, if the average "rate of flow" of clauses into $C_2(j)$ is always less than one then the average number of clauses in $C_1(j)$ is $o(1)$ and the probability that a complementary pair of unit clauses is encountered on any iteration tends to zero. Thus, the conditions under which $E\{w_2(j)\} < 1$ for all $1 \leq j < r$ are the conditions under which GUC finds a solution with probability tending to 1.

We now develop the differential equations for finding $E\{w_1(j)\}$ and $E\{w_2(j)\}$, solve them and find the conditions on $\frac{n}{r}$ which cause $E\{w_1(j)\} < 1$ and $E\{w_2(j)\} < 1$ for all $1 \leq j < r$ for GUC and which cause $E\{w_1(j)\} < 1$ for all $1 \leq j < r$ for UC. Later, it will be shown that these results imply the desired probabilistic results.

Consider UC first. Clearly, for $1 \leq i \leq k$

$$n_i(j+1) = n_i(j) + w_i(j+1) - z_i(j+1).$$

Taking expectations gives

$$E\{n_i(j+1)\} = E\{n_i(j)\} + E\{w_i(j+1)\} - E\{z_i(j+1)\}$$

which can be written

$$E\{n_i(j+1)\} - E\{n_i(j)\} = E\{w_i(j+1)\} - E\{z_i(j+1)\}. \qquad (1)$$

For large r we can approximate (1) by

$$\frac{dE\{n_i(j)\}}{dj} = E\{w_i(j+1)\} - E\{z_i(j+1)\}. \tag{2}$$

But, for all $2 \leq i \leq k$

$$E\{z_i(j+1)\} = E\{E\{z_i(j+1)/n_i(j)\}\}$$

$$= E\{\frac{i * n_i(j)}{r-j}\} = \frac{i * E\{n_i(j)\}}{r-j} \tag{3a}$$

because of corollary 1. Also, for all $1 \leq i < k$

$$E\{w_i(j+1)\} = E\{E\{w_i(j+1)/n_{i+1}(j)\}\}$$

$$= E\{\frac{(i+1) * n_{i+1}(j)}{2(r-j)}\} = \frac{(i+1) * E\{n_{i+1}(j)\}}{2(r-j)} \tag{3b}$$

and

$$E\{w_k(j+1)\} = 0.$$

Therefore (2), for $2 \leq i \leq k$ can be written

$$\frac{dE\{n_i(j)\}}{dj} = \frac{(i+1) * E\{n_{i+1}(j)\}}{2(r-j)} - \frac{i * E\{n_i(j)\}}{r-j} \tag{4a}$$

and

$$\frac{dE\{n_k(j)\}}{dj} = -\frac{k * E\{n_k(j)\}}{r-j}. \tag{4b}$$

The solution to these differential equations under the assumption that $E\{n_k(0)\} = n$ and $E\{n_i(0)\} = 0$ for all $1 \leq i < k$ is

Theorem 2:

For all $2 \leq i \leq k$

$$E\{n_i(j)\} = \frac{1}{2^{k-i}} \binom{k}{i} \left(1 - \frac{j}{r}\right)^i \left(\frac{j}{r}\right)^{k-i} n.$$

Proof:

Straightforward solution to (4a) and (4b).

Theorem 2 can be proved another way: the number of clauses in I containing $i > 1$ literals at iteration j is binomially distributed with parameters n and $p = 2^i\binom{j}{k-i}\binom{r-j}{i}/(2^k\binom{r}{k})$. The mean is np which tends to the expression in Theorem 2 when $r \to \infty$. We have used differential equations to prove Theorem 2 to prepare for the analysis of *GUC*.

Using Theorem 2 we can find the conditions under which $E\{w_1(j)\} < 1$ for all $1 \le j < r$.

Theorem 3:

Given that inputs to *UC* are distributed according to $M(n, r, k)$,

if $k \ge 3$ and $\lim_{n,r\to\infty} \frac{n}{r} < \frac{2^{k-1}}{k}\left(\frac{k-1}{k-2}\right)^{k-2}$ then $\forall 1 \le j < r,\ E\{w_1(j)\} < 1$.

Proof:

From (3b) and Theorem 2

$$E\{w_1(j+1)\} = \frac{E\{n_2(j)\}}{r-j} = \frac{1}{2^{k-2}}\binom{k}{2}\left(1-\frac{j}{r}\right)\left(\frac{j}{r}\right)^{k-2}\frac{n}{r}. \tag{5}$$

Taking the derivative gives

$$\frac{dE\{w_1(j)\}}{dj} = \frac{1}{2^{k-2}}\binom{k}{2}\frac{n}{r}\left[\frac{(k-2)}{r}\left(\frac{j}{r}\right)^{k-3}\left(1-\frac{j}{r}\right) - \frac{1}{r}\left(\frac{j}{r}\right)^{k-2}\right].$$

This is zero when $j = \frac{k-2}{k-1}r$. Substituting $\frac{k-2}{k-1}r$ for j in (5) and setting $E\{w_1\left(\frac{k-1}{k-2}r+1\right)\} < 1$ gives, for large r,

$$\frac{n}{r} < \frac{2^{k-1}}{k}\left(\frac{k-1}{k-2}\right)^{k-2}.$$

Now consider *GUC*. We can use (2) to model the accumulation of clauses in $C_i(j)$ for all $2 \le i \le k$ but the flow equations of (3a) and (3b) must be changed. Equations (3a) and (3b) may still be used for $E\{z_i(j+1)\}$ and $E\{w_{i-1}(j+1)\}$ if the $j+1^{st}$ literal (l) is not chosen from $C_i(j)$. Thus the expectations conditioned on the event $l \notin C_i(j)$ are

$$E\{z_i(j+1) | l \notin C_i(j)\} = \frac{i * E\{n_i(j) | l \notin C_i(j)\}}{r-j} \tag{6a}$$

and

$$E\{w_{i-1}(j+1)|l \notin C_i(j)\} = \frac{i * E\{n_i(j)|l \notin C_i(j)\}}{2(r-j)}. \tag{6b}$$

If the $j+1^{st}$ literal is chosen from $C_i(j)$ then one of the clauses of $C_i(j)$ is surely removed from I so (3a) and (3b) become

$$E\{z_i(j+1)|l \in C_i(j)\} = \frac{i * (E\{n_i(j)|l \in C_i(j)\} - 1)}{r-j} + 1 \tag{7a}$$

and

$$E\{w_{i-1}(j+1)|l \in C_i(j)\} = \frac{i * (E\{n_i(j)|l \in C_i(j)\} - 1)}{2(r-j)}. \tag{7b}$$

The probability that l is chosen from $C_i(j)$, denoted $p_i(j)$, and (6a), (6b), (7a) and (7b) can be used to compute $E\{z_i(j+1)\}$ and $E\{w_{i-1}(j+1)\}$. We have

$$E\{z_i(j+1)\} =$$

$$E\{z_i(j+1)|l \notin C_i(j)\} * (1 - p_i(j)) + E\{z_i(j+1)|l \in C_i(j)\} * p_i(j)$$

$$= \frac{i * E\{n_i(j)\}}{r-j} + p_i(j) * \left(1 - \frac{i}{r-j}\right) \tag{8a}$$

and

$$E\{w_{i-1}(j+1)\} =$$

$$E\{w_{i-1}(j+1)|l \notin C_i(j)\} * (1 - p_i(j)) + E\{w_{i-1}(j+1)|l \in C_i(j)\} * p_i(j)$$

$$= \frac{i * E\{n_i(j)\}}{2(r-j)} - p_i(j) * \frac{i}{2(r-j)}. \tag{8b}$$

The differential equations are:

$$\frac{dE\{n_k(j)\}}{dj} = -\frac{k * E\{n_k(j)\}}{r-j} - p_k(j) * \left(1 - \frac{k}{r-j}\right) \tag{9a}$$

and for all $2 \le i \le k-1$

$$\frac{dE\{n_i(j)\}}{dj} = \frac{(i+1) * E\{n_{i+1}(j)\}}{2(r-j)} - \frac{i * E\{n_i(j)\}}{r-j}$$

$$-p_{i+1}(j) * \frac{i+1}{2(r-j)} - p_i(j) * \left(1 - \frac{i}{r-j}\right). \tag{9b}$$

Let j_m, if it exists, denote the minimum j for which $E\{w_{k-m}(j)\} \geq 1$ and let $j_{\max}(\nu)$ be the value of j for which $E\{w_{k-\nu}(j)\}$ is maximum where ν is the maximum i such that j_i exists. We will use $j_{\max}$ when the precise value of ν is unimportant. As for UC we will find the conditions that guarantee
$E\{w_1(j_{\max}(k-1))\} < 1$ in the case of GUC. Also, we will find the conditions that guarantee
$E\{w_2(j_{\max}(k-2))\} < 1$ for these will also be the conditions under which GUC solves random instances of $k-SAT$ in polynomial time with probability approaching 1 as $r \to \infty$.

Lemma 1:

If $j_{\max} = \alpha r$ where α is a constant between 0 and 1 then

a) $\forall\, 2 \leq i \leq k-1,\ \forall\, 1 \leq j \leq j_{k-i}\ \ E\{w_{i-1}(j)\} = O(1/r)$.

b) $\forall\, 1 \leq m \leq k-2\ \ j_m \leq j_{m+1} \leq j_{\max}$ if j_m and j_{m+1} exist.

c) $\forall\, 1 \leq i \leq k-2,\ \forall\, 1 \leq j \leq j_{k-i-1}\ \ p_i(j) = O(1/r)$.

d) $\forall\, 2 \leq i \leq k-1,\ \forall\, 1 \leq j \leq j_{k-i}\ \ \frac{i * E\{n_i(j)\}}{r-j} = O(1/r)$.

e) $\forall\, 2 \leq i \leq k,\ \forall\, 1 \leq j \leq j_{max}\ \ p_i(j) * \frac{i}{r-j} = O(1/r)$.

f) $j_1 = 1$ if it exists.

Proof:

a) Since GUC always chooses the next literal from a smallest clause, $E\{n_i(0)\} = 0$ for all $1 \leq i \leq k-1$ and $E\{w_i(j)\} < 1$ for $1 \leq j \leq j_{k-i}$, $E\{n_i(j)\}$ is less than a constant for all $1 \leq j \leq j_{k-i}$. Therefore $E\{w_{i-1}(j)\} = O(1/r)$ for $1 \leq j \leq j_{k-i}$.

b) Follows from a).

c) From a) $E\{w_i(j)\} = O(1/r)$ for all $1 \leq j \leq j_{k-i-1}$. But $p_i(j)$ can be no greater than the maximum rate at which clauses enter $C_i(j)$ which is the maximum $E\{w_i(j)\}$ over $1 \leq j \leq j_{k-i-1}$.

d) From the proof of a) $E\{n_i(j)\}$ is less than a constant for all $1 \leq j \leq j_{k-i}$. The desired result follows.

e) Obvious.

f) Since $E\{n_k(0)\} = n$, $E\{w_{k-1}(1)\}$ is approximately kn/r. This is greater than 1 if j_1 exists because, from (9a) and (8b), $E\{w_{k-1}(j)\}$ is nonincreasing as j increases.

Lemma 2:

a) $\forall\ 1 \le i \le k-2$, if $\forall\ j_{k-i-1} < j \le j_{k-i}$, $E\{w_{i+1}(j)\} \ge 1$ and $E\{w_i(j)\} < 1$ then $\forall\ j_{k-i-1} < j \le j_{k-i}$, $p_i(j) = E\{w_i(j)\} - o(1)$.

b) $\forall\ 1 \le i \le k-1$, if $\forall\ j_{k-i} < j \le j_{k-i+1}$, $E\{w_i(j)\} \ge 1$ and $E\{w_{i-1}(j)\} < 1$ then $\forall\ j_{k-i} < j \le j_{k-i+1}$, $p_i(j) = 1 - E\{w_{i-1}(j)\} + o(1)$.

c) $\forall\ 1 \le i \le k$, if $\forall\ j_{k-i+1} < j \le j_{\max}$ $\exists\ i'(j) < i$ s.t. $E\{w_{i'(j)}(j)\} \ge 1$ and $p_i(j_{k-i+1}) = o(1)$ then $\forall j_{k-i+1} < j \le j_{\max}$ $p_i(j) = o(1)$.

Proof:

a) Since the ratio of n to r is assumed to be a function only of k, $E\{n_{i+1}(j)\}$ rises no faster than linearly with j. Therefore $E\{w_i(j)\}$ rises no faster than O(1/r) per iteration of *GUC*. Thus, as r approaches infinity $p_i(j)$ approaches the steady state equilibrium probability that there is at least one clause in $C_i(j)$. Since $E\{w_i(j)\} < 1$, $E\{w_{i-1}(j)\} = O(1/r)$ and *GUC* always chooses the next literal from a smallest clause, the equilibrium probability is, in the limit, $E\{w_i(j)\}$.

c) The $j+1^{st}$ chosen literal will be chosen from $C_{i''(j)}(j)$ where $i''(j) \le i'(j)$ for all $j_{k-i+1} < j \le j_{max}$ in the limit.

b) Follows from a) and c) of Lemma 2 and c) of Lemma 1.

The following differential equations result after using the results of Lemma 1 and Lemma 2 in (9a) and (9b) provided the conditions of Lemma 2 are satisfied: for all $1 \le j \le j_{max}$

$$\frac{dE\{n_k(j)\}}{dj} = -\frac{k * E\{n_k(j)\}}{r-j} - o(1), \tag{10a}$$

for all $1 \le j \le j_{k-i}$ and for all $2 \le i \le k-1$

$$\frac{dE\{n_i(j)\}}{dj} = o(1), \tag{10b}$$

for all $j_{k-i} < j \le j_{k-i+1}$ and for all $2 \le i \le k-1$

$$\frac{dE\{n_i(j)\}}{dj} = \frac{(i+1) * E\{n_{i+1}(j)\}}{2(r-j)} - \frac{i * E\{n_i(j)\}}{2(r-j)} - 1 + o(1), \tag{10c}$$

and for all $j_{k-i+1} < j \le j_{\max}$ and for all $2 \le i \le k-1$

$$\frac{dE\{n_i(j)\}}{dj} = \frac{(i+1) * E\{n_{i+1}(j)\}}{2(r-j)} - \frac{i * E\{n_i(j)\}}{r-j} + o(1). \qquad (10d)$$

The solutions to (10a)-(10d) follow. These solutions were obtained first for $j_1 \le j < j_2$ then for $j_2 \le j < j_3$ and so on. It is easy to check that the conditions of Lemma 1 and Lemma 2 are satisfied for each interval before solving the appropriate differential equations for that interval. The solution to (10a) for all $1 \le j \le j_{\max}$ under the boundary condition $E\{n_k(0)\} = n$ is

$$E\{n_k(j)\} = \left(1 - \frac{j}{r}\right)^k n - o(r). \qquad (11a)$$

The solution to (10b) for all $1 \le j \le j_{k-i}$ under the boundary condition $E\{n_i(0)\} = 0$ for all $1 \le i \le k-1$ is

$$E\{n_i(j)\} = o(r). \qquad (11b)$$

The solution to (10d) for all $j_{k-i+1} \le j \le j_{\max}$ under the boundary condition

$$E\{w_{i-1}(j_{k-i+1})\} = \frac{i * E\{n_i(j_{k-i+1})\}}{2(r - j_{k-i+1})} - p_i(j_{k-i+1}) * \frac{1}{2(r - j_{k-i+1})} = 1 \qquad (11d'')$$

is

$$E\{n_i(j)\} = \sum_{m=0}^{k-i} \frac{1}{2^m} \binom{m+i}{m} \left(\frac{j}{r}\right)^m \left(1 - \frac{j}{r}\right)^i \mathrm{const}(j_{k-i-m+1}) r + o(r) \qquad (11d)$$

where, for all $3 \le i \le k-1$

$$\mathrm{const}(j_{k-i+1}) = \frac{E\{n_i(j_{k-i+1})\}}{r\left(1 - \frac{j_{k-i+1}}{r}\right)^i} - o(1)$$

$$- \sum_{m=1}^{k-i} \frac{1}{2^m} \binom{m+i}{m} \left(\frac{j}{r}\right)^m \mathrm{const}(j_{k-i-m+1}) \qquad (11d')$$

and $\mathrm{const}(j_1) = n/r$. The solution to (10c) for all $3 \le i \le k-1$ and $j_{k-i} < j \le j_{k-i+1}$ is

$$E\{n_i(j)\} = \mathrm{dots}(j_{k-i}) \left(1 - \frac{j}{r}\right)^{\frac{i}{2}} r - \frac{2\left(1 - \frac{j}{r}\right) r}{i-2} + o(r)$$

$$-\sum_{m=0}^{k-i-1} \frac{m+1}{2^{m+1}} \binom{m+i+1}{m+1} \left(\frac{j}{r}\right)^m \left(1-\frac{j}{r}\right)^{i+1} \text{const}(j_{k-i-m}) r \times$$

$$\sum_{l=0}^{m} \frac{m! \left(\frac{i}{2}\right)! \left(\frac{j}{r}\right)^{-l} \left(1-\frac{j}{r}\right)^{l}}{(m-l)! \left(\frac{i}{2}+l+1\right)!} \qquad (11c)$$

where

$$\text{dots}(j_{k-i}) = \frac{2}{i-2} \left(1-\frac{j_{k-i}}{r}\right)^{1-\frac{i}{2}} - o(1)$$

$$+\sum_{m=0}^{k-i-1} \frac{m+1}{2^{m+1}} \binom{m+i+1}{m+1} \left(\frac{j_{k-i}}{r}\right)^m \left(1-\frac{j_{k-i}}{r}\right)^{\frac{i}{2}+1} \text{const}(j_{k-i-m}) \times$$

$$\sum_{l=0}^{m} \frac{m! \left(\frac{i}{2}\right)! \left(\frac{j_{k-i}}{r}\right)^{-l} \left(1-\frac{j_{k-i}}{r}\right)^{l}}{(m-l)! \left(\frac{i}{2}+l+1\right)!} \qquad (11c')$$

and for all $j_{k-2} < j \leq j_{\max}$

$$E\{n_2(j)\} = r\left(1-\frac{j}{r}\right) \ln\left(1-\frac{j}{r}\right) + r\left(1-\frac{j}{r}\right) \text{dots}(j_{k-2}) + o(r)$$

$$-\sum_{m=0}^{k-3} \frac{m+1}{2^{m+1}} \binom{m+3}{m+1} \left(\frac{j}{r}\right)^m \left(1-\frac{j}{r}\right)^3 \text{const}(j_{k-2-m}) r \sum_{l=0}^{m} \frac{m! \left(\frac{j}{r}\right)^{-l} \left(1-\frac{j}{r}\right)^{l}}{(m-l)!(l+2)!} \qquad (12a)$$

where

$$\text{dots}(j_{k-2}) = -\ln\left(1-\frac{j_{k-2}}{r}\right) - o(1)$$

$$+\sum_{m=0}^{k-3} \frac{m+1}{2^{m+1}} \binom{m+3}{m+1} \left(\frac{j_{k-2}}{r}\right)^m \left(1-\frac{j_{k-2}}{r}\right)^2 \text{const}(j_{k-2-m}) \times$$

$$\sum_{l=0}^{m} \frac{m! \left(\frac{j_{k-2}}{r}\right)^{-l} \left(1-\frac{j_{k-2}}{r}\right)^{l}}{(m-l)!(l+2)!} \qquad (12b)$$

The points j_i which exist are the points at which

$$E\{w_{k-i}(j)\} = \frac{(k-i+1) * E\{n_{k-i+1}(j)\}}{2(r-j)} + o(1) = 1.$$

Thus the following expression locates, in the limit, the points j_i for $2 \le i \le k-2$:

$$1 = -(k-i+1)\sum_{m=0}^{i-2}\frac{m+1}{2^{m+2}}\binom{k-i+m+2}{m+1}\left(\frac{j_i}{r}\right)^m\left(1-\frac{j_i}{r}\right)^{k-i+1}\mathrm{const}(j_{i-m-1})\times$$

$$\sum_{l=0}^{m}\frac{m!\,(\frac{k-i+1}{2})!\left(\frac{j_i}{r}\right)^{-l}\left(1-\frac{j_i}{r}\right)^{l}}{(m-l)!\,(\frac{k-i+1}{2}+l+1)!}-\frac{k-i+1}{k-i-1}+\frac{\mathrm{dots}(j_{i-1})}{2}\left(1-\frac{j_i}{r}\right)^{\frac{k-i-1}{2}}(k-i+1). \tag{13}$$

The following expression gives the limiting value of $E\{w_1(j)\}$ for $j_{k-2} < j \le j_{\max}(k-1)$:

$$-\sum_{m=0}^{k-3}\frac{m+1}{2^{m+1}}\binom{m+3}{m+1}\left(\frac{j}{r}\right)^m\left(1-\frac{j}{r}\right)^2\mathrm{const}(j_{k-m-2})\sum_{l=0}^{m}\frac{m!\left(\frac{j}{r}\right)^{-l}\left(1-\frac{j}{r}\right)^{l}}{(m-l)!(l+2)!}$$

$$+\ln\left(1-\frac{j}{r}\right)+\mathrm{dots}(j_{k-2}). \tag{14}$$

The maximum of $E\{w_1(j)\}$ in the limit can be computed numerically for a given ratio of n to r and for $k \ge 4$ as follows:

```
FINDMAXFLOW(n, r, k) :
    const(j_1) := n/r
    dots(j_1) := k/(k + 1) * const(j_1) + 2/(k - 3)
    for i := 2 to k - 2 do begin
        Locate j_i using (13)
        Compute const(j_i) using (11d') and (11d")
        Compute dots(j_i) using (11c') and (12b)
    End
    i := k - 1
    Locate j such that E{w_1(j)} is maximum using (14)
    Return E{w_1(j)}.
```

Note that for $1 \le i \le k-2$, $\frac{j_i}{r}$ is a function only of k since $\frac{n}{r}$ is assumed to be a function only of k. Also, note that $\frac{j}{r}$ is a function only of k where j is located by the next to last line of *FINDMAXFLOW*.

Using *FINDMAXFLOW* the maximum ratio of n to r giving $E\{w_1(j_{\max}(k-1))\} < 1$ may be found. The following table lists these results for values of k from 4 to 40. Inspection of the table reveals that $E\{w_1(j_{\max}(k-1))\} < 1$ if n/r is less than $\frac{3.09*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$ for $4 \le k \le 40$.

k	n/r	$\frac{3.09*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$
4	5.6	5.56
6	17.8	17.24
8	57.4	55.4
10	191	184
12	647	630
14	2255	2203
16	7980	7822
18	28500	28113
20	103000	102078
30	7.14×10^7	7.14×10^7
40	5.57×10^{10}	5.54×10^{10}

FINDMAXFLOW can be modified to find the maximum ratio of n to r giving $E\{w_2(j)\} < 1$ for all $1 \le j < r$: replace the line "for $i := 2$ to $k-2$..." with the line "for $i := 2$ to $k-3$...", the line "Return $E\{w_1(j)\}$" with the line Return $E\{w_2(j)\}$ and the line "Locate j such that $E\{w_1(j)\}$..." with the line "Locate j such that $E\{w_2(j)\}$ is maximum" where $E\{w_2(j)\}$ is obtained by dividing (11c) by $.666 * (r-j)$. The following table lists the results for values of k from 4 to 40.

k	n/r	$\frac{1.845*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$
4	2.5	3.27
6	9.6	10.27
8	32.6	33
10	110	110
12	380	376
14	1330	1315
16	4711	4669
18	16980	16785
20	61700	60947
30	4.3×10^7	4.26×10^7
40	3.35×10^{10}	3.32×10^{10}

We now prove the main results. In the theorems below k is regarded to be a constant.

Theorem 4:

UC verifies that a solution exists for satisfiable instances generated according to $M(n, r, 3)$ with probability greater than ϵ for some $\epsilon > 0$ when $\lim_{n,r\to\infty} \frac{n}{r} < \frac{2^{k-1}}{k}\left(\frac{k-1}{k-2}\right)^{k-2}$.

Proof:

From theorem 3 $E\{w_1(j)\} < 1$ for all $1 \leq j < r$ when $\lim_{n,r\to\infty} \frac{n}{r} < \frac{2^{k-1}}{k}\left(\frac{k-1}{k-2}\right)^{k-2}$. From corollary 1 the clauses entering $C_1(j+1)$ from $C_2(j)$ are statistically independent. Suppose all clauses entering $C_1(j+1)$ are regarded as entering $C_1(j+1)$ in some order which is decided arbitrarily. Then the probability that the q^{th} clause entering $C_1(j+1)$ is complementary to no clause in $C_1(j+1)$ is

$$\left(1 - \frac{1}{2(r-j)}\right)^{n_1(j)+q-1}.$$

Therefore, the probability that none of the clauses entering $C_1(j+1)$ is complementary to any clause in $C_1(j+1)$ is

$$\left(1 - \frac{1}{2(r-j)}\right)^{n_1(j)*w_1(j)+w_1(j)*(w_1(j)-1)/2}$$

so the probability that no complementary pair is encountered during a run of UC is

$$\sum\prod_{j=0}^{r-1}\left(1 - \frac{1}{2(r-j)}\right)^{n_1(j)*w_1(j)+w_1(j)*(w_1(j)-1)/2} pr(\ldots n_1(j), w_1(j)\ldots)$$

$$> \sum\prod_{j=0}^{r-1}\left(1 - \frac{1}{2r}\right)^{\frac{2r}{r-j}(n_1(j)*w_1(j)+w_1(j)*(w_1(j)-1)/2)} pr(\ldots n_1(j), w_1(j)\ldots)$$

$$= \sum\left(1 - \frac{1}{2r}\right)^{2r\sum_{j=0}^{r-1}\frac{2n_1(j)*w_1(j)+w_1(j)*(w_1(j)-1)}{2(r-j)}} pr(\ldots n_1(j), w_1(j)\ldots). \quad (15)$$

If the sum in the exponent of (15) is less than $\frac{\kappa n}{r}$ (where κ is a constant) with probability bounded from below by 2/3 then (15) is bounded from below by $\frac{2}{3}(1-\frac{1}{2r})^{2n\kappa}$ which approaches a constant as r approaches infinity if the limiting ratio of n to r is constant. To show that the sum in the exponent of (15) is less than $\frac{\kappa n}{r}$ with probability greater than 2/3 we show that the expectation of the sum is bounded from above by $\frac{\kappa n}{3r}$ and apply markov's inequality.

To show that the expectation of the sum in the exponent of (15) is less than $\frac{\kappa n}{3r}$ we need only show that the expectation of each term in the sum is less than $\frac{\kappa n}{3r^2}$. Denote by $s_1(j)$ the j^{th} term in the sum. Then

$$E\{s_1(j)\}$$
$$\leq \frac{1}{2(r-j)}\left(E\{w_1^2(j)\} + \sum_{s=0}^{n}\sum_{t=0}^{n} 2*s*t*pr(n_1(j)=t, w_1(j)=s)\right). \qquad (16)$$

The second term within parentheses is bounded by $\gamma_1(1-\frac{j}{r})\frac{n}{r}$ for $j < r - r^{8/9}$ and by $\gamma_2(1-\frac{j}{r})\frac{n}{r}$ for $j \geq r - r^{8/9}$ where γ_1 and γ_2 are constants greater than zero. Consider the first case, $1 \leq j < r - r^{8/9}$. Define $n_l = E\{n_2(j)\} - n^{3/4}$ and $n_u = E\{n_2(j)\} + n^{3/4}$. Since $n_2(j)$ is binomially distributed with mean $E\{n_2(j)\}$ proportional to $\frac{j}{r}(1-\frac{j}{r})^2 n$, the probability that $n_l < n_2(j) < n_u$ is greater than $1 - 2e^{-n^{3/2}/E\{n_2(j)\}}$ from [5] and this is greater than $1 - 2e^{-\sqrt{n}}$ since $E\{n_2(j)\} < n$. The double sum of (16) can be split into three parts:

$$\sum_{s=0}^{n}\sum_{t=0}^{n}\sum_{u=0}^{n_l} 2*s*t*pr(n_1(j)=t, n_2(j)=u, w_1(j)=s)$$
$$+\sum_{s=0}^{n}\sum_{t=0}^{n}\sum_{u=n_l}^{n_u} 2*s*t*pr(n_1(j)=t, n_2(j)=u, w_1(j)=s)$$
$$+\sum_{s=0}^{n}\sum_{t=0}^{n}\sum_{u=n_u}^{n} 2*s*t*pr(n_1(j)=t, n_2(j)=u, w_1(j)=s)$$
$$< \frac{8n^2}{e\sqrt{n}} + 2*E\{w_1(j)\}*E\{n_1(j)\} \qquad (17)$$

in the limit since $E\{w_1(j)\} < 1$ for all $1 \leq j < r$ and $|n_u - n_l| \to 0$. But $E\{w_1(j)\}$ is proportional to $(1-\frac{j}{r})\frac{n}{r}$ from (3b) and Theorem 2. Also, $E\{n_1(j)\}$ is bounded by a constant for all $1 \leq j \leq r$ since $E\{w_1(j)\} < 1$ and at least one clause is removed from $C_1(j)$ if $C_1(j) \neq \phi$. So (17) is less than $\gamma_1(1-\frac{j}{r})\frac{n}{r}$ where γ_1 is a constant greater than zero.

Now consider the case $r - r^{8/9} \le j < r$. In this range $pr(w_1(j) = s) \le (O((1-\frac{j}{r})\frac{n}{r}))^s/s!$ since $w_1(j)$ is binomially distributed. Furthermore, $(1-\frac{j}{r})^{37} \le (1-\frac{j}{r})\frac{1}{r^4}$ in this range. Therefore the double sum of (16) is bounded from above by

$$\sum_{s=0}^{37}\sum_{t=0}^{n} 74 * t * pr(n_1(j) = t) + \sum_{s=38}^{n}\sum_{t=0}^{n} 2 * s * t \left(1 - \frac{j}{r}\right)\frac{1}{r^4}$$

$$\le 74 * E\{n_1(j)\} + 2\left(1 - \frac{j}{r}\right)\frac{n^4}{r^4}. \qquad (18)$$

Since $w_1(j)$ is binomially distributed with mean $E\{w_1(j)\} \to 0$ and $E\{w_1(j+1)\} - E\{w_1(j)\} = O(1/r)$ and since UC always chooses the next literal from a unit clause if one exists, $E\{n_1(j)\}$ is bounded from above by a constant times $E\{w_1^2(j)\}$. It will be shown later that $E\{w_1^2(j)\}$ is proportional to $E\{w_1(j)\}$ hence proportional to $(1-\frac{j}{r})\frac{n}{r}$ in the interval $j > r - r^{8/9}$. This and the fact that $\frac{n}{r}$ is assumed to be a constant less than $2^k \ln(2)$ results in the following inequality based on (18):

$$\sum_{s=0}^{n}\sum_{t=0}^{n} 2 * s * t * pr(n_1(j) = t, w_1(j) = s) \le \gamma_2(1 - \frac{j}{r})\frac{n}{r}$$

as $r \to \infty$.

We now need to find a bound on $E\{w_1^2(j)\}$. But $w_1(j)$ is distributed binomially hence $E\{w_1^2(j)\} = \sigma^2(w_1(j)) + (E\{w_1(j)\})^2 < E\{w_1(j)\} + (E\{w_1(j)\})^2$ and

$$E\{w_1^2(j)\} < \gamma_3(1 - \frac{j}{r})\frac{n}{r}.$$

Let $\gamma = \max\{\gamma_1, \gamma_2\}$. Substituting $\gamma(1-\frac{j}{r})\frac{n}{r}$ for the double sum in (16) and then $\gamma_3 * (1-\frac{j}{r})\frac{n}{r}$ for $E\{w_1^2(j)\}$ in the resulting inequality gives

$$E\{s_1(j)\} \le \left(\frac{\gamma_3 + \gamma}{2}\right)\frac{n}{r^2} = \frac{\kappa}{3} * \frac{n}{r^2}.$$

From this the expectation of the sum in the exponent of (15) is less than $\frac{\kappa n}{3r}$. By markov's inequality the probability that the sum is greater than $\frac{\kappa n}{r}$ is less than 1/3. Therefore, the probability that the sum is less than $\frac{\kappa n}{r}$ is greater than 2/3. Thus (15) is greater than $\frac{2}{3}(1-\frac{1}{2r})^{2n\kappa}$ which approaches $\frac{2}{3}e^{-\frac{n\kappa}{r}}$ as r approaches infinity. Let $\epsilon = \frac{2}{3}(1-\frac{1}{2r})^{2n\kappa}$.

Theorem 5:

GUC verifies that a solution exists for random instances generated according to $M(n,r,k)$ with probability greater than ϵ for some $\epsilon > 0$ when $\lim_{n,r\to\infty} \frac{n}{r} < \frac{3.09*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2}$ and $4 \le k \le 40$.

Proof:

We first introduce a new algorithm UC' which runs concurrently with *GUC*: initially, UC' is given a copy I' of the input I to *GUC* and each time a literal l is chosen by *GUC* all occurrences of $comp(l)$ removed from I are removed by UC' from I' and all clauses removed by *GUC* from I are removed by UC' from I' except the clause containing the chosen literal - the chosen literal in this clause is replaced with a literal chosen randomly from the set of all pairs of literals and their complements neither of which have yet been chosen unioned with $\{l, comp(l)\}$ and the clause is removed if it contains l, $comp(l)$ is removed if it appears in the clause and nothing happens to the clause otherwise. Let $\hat{n}_2(j)$ and $\hat{w}_1(j)$ have the same meaning as $n_2(j)$ and $w_1(j)$ except applied to UC'.

The proof follows the proof of Theorem 4 up to (16); all the random variables in (16) now represent flows and class sizes under *GUC*. Again we consider the region $j < r - r^{8/9}$. The double sum of (16) can then be bounded from above by the same expression in (17) except with $\hat{w}_1(j)$ substituted for $w_1(j)$ and $\hat{n}_2(j)$ substituted for $n_2(j)$. Let $n_l = E\{\hat{n}_2(j)\} - n^{3/4}$ and $n_u = E\{\hat{n}_2(j)\} + n^{3/4}$. This bound may be written

$$\sum_{s=0}^{n}\sum_{t=0}^{n}\sum_{u=0}^{n_l} 2*s*t*pr(n_1(j)=t, \hat{n}_2(j)=u, \hat{w}_1(j)=s)$$

$$+\sum_{s=0}^{n}\sum_{t=0}^{n}\sum_{u=n_l}^{n_u} 2*s*t*pr(n_1(j)=t, \hat{n}_2(j)=u, \hat{w}_1(j)=s)$$

$$+\sum_{s=0}^{n}\sum_{t=0}^{n}\sum_{u=n_u}^{n} 2*s*t*pr(n_1(j)=t, \hat{n}_2(j)=u, \hat{w}_1(j)=s)$$

$$< \frac{8n^2}{e\sqrt{n}} + 2*E\{\hat{w}_1(j)\}*E\{n_1(j)\} \qquad (19)$$

in the limit since $\hat{n}_2(j)$ is binomially distributed with mean equal to the mean of $n_2(j)$ under UC. But $E\{\hat{w}_1(j)\}$ is proportional to $(1-\frac{j}{r})\frac{n}{r}$ since this mean is equal to the mean of $w_1(j)$ under UC. Also, $E\{n_1(j)\}$ is bounded by a constant since $E\{w_1(j)\} < 1$ for all $j < r - r^{8/9}$ and GUC always chooses the next literal from the set of unit clauses when a unit clause exists in I. Therefore, (19) is less than $\gamma_1(1-\frac{j}{r})\frac{n}{r}$ where γ_1 is a constant greater than zero. The double sum in (16) is bounded by $\gamma_2(1-\frac{j}{r})\frac{n}{r}$ for $j \geq r - r^{8/9}$ by using the same argument as in Theorem 4 except with $\hat{w}_1(j)$ replacing $w_1(j)$. Also,

$$E\{w_1^2(j)\} \leq E\{\hat{w}_1^2(j)\} < \gamma_3(1-\frac{j}{r})\frac{n}{r}.$$

The remainder of the proof follows the proof of Theorem 4.

Theorem 6:

GUC verifies that a solution exists for random instances generated according to $M(n,r,k)$ with probability approaching 1 as $r \to \infty$ when
$\lim_{n,r\to\infty} \frac{n}{r} < \frac{1.845*2^{k-2}}{k+1}\left(\frac{k-1}{k-2}\right)^{k-2} - 1$ for $4 \leq k \leq 40$.

Proof:

First we show that the terms represented by (16) are $O(r^{-25/18})$ for $j < r - r^{8/9}$. Consider the double sum in (16) first. From the proof of Theorem 5 the double sum is bounded from above by (19). Since $E\{w_2(j)\} < 1$ for all $j < r - r^{8/9}$, $E\{n_2(j)\}$ is bounded from above by a constant for all $j < r - r^{8/9}$ and this implies $E\{w_1(j)\} \leq \frac{K'}{r-j}$ where K' is a constant. By Little's law and the fact that GUC always chooses literals from $C_1(j)$ if $C_1(j) \neq \phi$, $E\{n_1(j)\}$ is less than a constant times the maximum of $E\{w_1(j)\}$ in the interval $j < r - r^{8/9}$. Thus $E\{n_1(j)\} = O(r^{-8/9})$ for $j < r - r^{8/9}$ and (19) approaches $\gamma_4(1-\frac{j}{r})\frac{n}{r} * O(r^{-8/9})$ where γ_4 is a constant greater than zero.

Now consider the $E\{w_1^2(j)\}$ term in (16). Since $E\{w_1(j)\} \leq \frac{K'}{r-j}$, $pr(w_1(j) = s) \leq \frac{K'}{r-j}$. But

$$E\{w_1^2(j)\} = \sum_{s=0}^{r^{1/8}} s^2 * pr(w_1(j) = s) + \sum_{s=r^{1/8}+1}^{n} s^2 * pr(w_1(j) = s)$$

$$\leq \sum_{s=0}^{r^{1/8}} \frac{s^2 * K'}{r-j} + \sum_{s=r^{1/8}+1}^{n} s^2 * pr(\hat{w}_1(j) = s)$$

$$\le \frac{1}{r^{1/2}} + \frac{2}{e^{\beta r^{1/4}}} \quad \text{for all } j < r - r^{8/9}$$

where β is a constant. Thus, for $j < r - r^{8/9}$ the terms of (16) are all $O(r^{-25/18})$ where it has been assumed that $\frac{n}{r}$ is a constant. For $j > r - r^{8/9}$ the terms of (16) are $O(1/r)$ from the proof of Theorem 4. Thus, in (15), the mean of the sum in the exponent is $O(r^{-2/18})$. By Markov's inequality, this sum is less than $O(r^{-1/18})$ and the entire exponent is less than $O(r^{17/18})$ with probability greater than $1 - O(r^{-1/18})$. But $(1 - \frac{1}{2r})^{O(r^{17/18})} \to 1$ as $r \to \infty$. Hence (15) tends to 1 as $r \to \infty$.

6. Conclusions

We have presented algorithms for $k - SAT$ based on the Unit-Clause heuristic and a generalization of the Unit-Clause heuristic and have shown the conditions under which these algorithms find a solution to a random instance of $k - SAT$ in polynomial time with probability bounded from below by a constant under $M(n, r, k)$. We have also shown the condition under which the latter algorithm finds a solution with probability approaching 1 under $M(n, r, k)$. All the conditions are of the form $\lim_{n,r\to\infty} \frac{n}{r} < F_X(k)$ where X indicates the algorithm and the kind of probabilistic result and $F_X(k) = O(2^k/k)$ for all X. This is a dramatic improvement over the result in [6] which showed that the Pure-Literal heuristic finds a solution to a random instance of $k - SAT$ in polynomial time with probability approaching 1 as $r \to \infty$ under $M(n, r, k)$ if $\lim_{n,r\to\infty} \frac{n}{r} < 1$. However, it is likely that a random instance has a solution with probability approaching 1 if $\lim_{n,r\to\infty} \frac{n}{r} < O(2^k)$.

The method used to obtain the result is general enough to be used to analyze the performance of similar algorithms applied to other NP-complete problems; two obvious candidates are the Chromatic Number problem and the Set Cover problem. The method is also general enough to be used on a variety of input distributions: these distributions should have the property that all components (clauses in the case of $k - SAT$) of the same size are independently and uniformly chosen from the set of all possible components of that size consisting of variables given in the problem instance. For example, this method can be used to get performance results for SAT under the constant-density model. Finally, if flow equations are too difficult to solve, a numerical solution usually is possible. For example, an algorithm for which flow equations can be written but are hard to solve is UC modified so that a chosen literal always occurs more often than its complement (the variable is chosen randomly and then the maximum occurring literal associated with that variable is the chosen literal).

References

[1] Brown, C.A. and Purdom, P.W., "An average time analysis of backtracking," *SIAM J. Comput.* **10** (1981), pp 583-593.

[2] Chao, M.T., "Probabilistic analysis and performance measurement of algorithms for the Satisfiability problem," Ph.D. Dissertation, Case Western Reserve University (1984).

[3] Chao, M.T. and Franco, J., "Probabilistic analysis of the Unit Clause and Maximum Occurring Literal selection heuristics for the 3-Satisfiability problem," Indiana University Tech. Report (1984).

[4] Davis, M. and Putnam, H., "A computing procedure for quantification theory," *J.ACM* **7** (1960), pp 201-215.

[5] Erdos, P. and Spencer, J., *Probabilistic Methods in Combinatorics*, Academic Press, 1974.

[6] Franco, J., "Probabilistic analysis of the pure literal heuristic for the satisfiability problem," *Annals of Operations Research* **1** (1984), pp 273-289.

[7] Franco, J., "Sensitivity of probabilistic results on algorithms for NP-complete problems to input distributions," Case Western Reserve University (1984).

[8] Franco, J. and Paull, M., "Probabilistic analysis of the Davis Putnam Procedure for solving the satisfiability problem," *Discrete Applied Mathematics* **5** (1983), pp 77-87.

[9] Goldberg, A., "Average case complexity of the satisfiability problem," proc. 4^{th} Workshop on Automated Deduction (1979), pp 1-6.

[10] Goldberg, A., Purdom,P.W. and Brown, C.A., "Average time analysis of simplified Davis-Putnam procedures," *Information Processing Letters* **15** (1982), pp 72-75.

[11] Purdom, P.W., "Search rearrangement backtracking and polynomial average time," *Artificial Intelligence* **21** (1983), pp 117-133.

[12] Purdom, P.W. and Brown, C.A., "The pure literal rule and polynomial average time," to appear in *SIAM J. Comput.*.