# Further Investigations into Regular XORSAT *

**Matti Järvisalo**
Laboratory for Theoretical Computer Science
P.O. Box 5400, FI-02015 Helsinki University of Technology, Finland
matti.jarvisalo@tkk.fi
Accompanying web page: `http://www.tcs.hut.fi/~mjj/benchmarks/aaai06/`

## Introduction

Recent years have witnessed rapid progress both in the foundations of and in applying state-of-art solvers for the propositional satisfiability problem (SAT). The study of sources for hard SAT instances is motivated by the need for interesting benchmarks for solver development and on the other hand by theoretical analysis of different proof systems.

In this respect *satisfiable* instance families are especially interesting. In contrast to unsatisfiable instance families, there are few theoretical results for satisfiable formulas (Alekhnovich, Hirsch, & Itsykson); for the successful DPLL method, restricted heuristics need to be considered.

While real-world problems serve as best benchmark instances in many sense, such instances are typically very large and unavailable in abundance. More "artificial" empirically hard satisfiable CNF families include (see references therein for more) regular random $k$-SAT (Boufkhad *et al.*), encodings of quasi-group completion (Achlioptas *et al.* 2000), XORSAT models inspired by statistical physics (Ricci-Tersenghi, Weight, & Zecchina 2001; Jia, Moore, & Selman 2005), and the *regular XORSAT model* (Haanpää *et al.* 2006) motivated by expansion properties of random regular bipartite graphs.

Experimental comparison with other available generators for notably hard satisfiable 3-CNF formulas shows that the regular XORSAT model gives extremely hard instances for state-of-the art clausal SAT solvers (Haanpää *et al.* 2006). In this paper we generalize the regular XORSAT model for $k > 3$, and investigate how this relates to the hardness of the instances. By increasing the degree of the underlying regular constraint graphs, we observe a sharp increase in problem difficulty with respect to the number of variables, motivating further analysis of regular XORSAT.

## Preliminaries

Let $X$ be a set of Boolean variables. For each $x \in X$ there are two *literals*, $x$ (positive) and $\bar{x}$ (negative). A *clause* of length $k$ (a $k$-clause) is a disjunction of $k$ distinct literals. A propositional formula in $k$*-conjunctive normal form* (a $k$-CNF formula) is a conjunction of $k$-clauses. A *truth assignment* $\tau$ associates a truth value $\tau(x) \in \{0, 1\}$ with each variable $x \in X$. A truth assignment *satisfies* a CNF formula if it satisfies every clause in it. A clause is satisfied if it contains at least one satisfied literal, where a literal $x$ (respectively, $\bar{x}$) is satisfied if $\tau(x) = 1$ ($\tau(x) = 0$). The $k$*-satisfiability problem* is to decide whether a given $k$-CNF formula admits a satisfying truth assignment.

Let $G = (V, E)$ be an undirected graph. The *degree* $d$ of a vertex is the number of vertices adjacent to it. A graph is $d$*-regular* if all of its vertices have degree $d$. Graph $G$ is *bipartite* if there exist $X, Y \subseteq V$ such that $X \cup Y = V$, $X \cap Y = \emptyset$, and every edge is incident to one vertex in $X$ and one in $Y$. Such a pair $(X, Y)$ is a *bipartition* of $G$.

## The Regular $d$-XORSAT Model

We now describe the regular $d$-XORSAT model for generating satisfiable $d$-CNF formulas. The regular (3-)XORSAT model was originally introduced in (Haanpää *et al.* 2006).

A regular $d$-XORSAT instance with $n$ variables is constructed as follows. Let $X = \{x_i\}_{i=1}^n$ be an associated set of $n$ Boolean variables and let $Y = \{y_i\}_{i=1}^n$ be a set of $n$ elements, each corresponding to an equation in a system of $n$ linear equations over $X$. A *constraint graph* $G = (V, E)$ with bipartition $(X, Y)$ characterizes the occurrences of the variables in the equations, that is, $\{x_j, y_i\}$ is an edge of $G$ if and only if the variable $x_j \in X$ occurs in the equation $y_i \in Y$. In the regular $d$-XORSAT model a constraint graph $G$ is selected uniformly at random from the set of all $d$-regular graphs with bipartition $(X, Y)$. Once a constraint graph $G$ has been selected, construct a system of linear equations based on $G$ as follows. Let $A = (a_{ij})$ be the $n \times n$ matrix whose entries are defined for all $i, j = 1, \ldots, n$ by

$$a_{ij} = \begin{cases} 1 & \text{if } \{x_j, y_i\} \in E, \\ 0 & \text{if } \{x_j, y_i\} \notin E. \end{cases}$$

Select uniformly at random a $\vec{z} \in \{0, 1\}^n$ and let $\vec{b} \in \{0, 1\}^n$ so that $\vec{b} \equiv A\vec{z} \pmod 2$. The system of linear equations is now $A\vec{x} \equiv \vec{b} \pmod 2$, where $\vec{x} = (x_1, \ldots, x_n)$ is a column vector of variables. By construction $A\vec{z} \equiv \vec{b}$

*Work supported by the Academy of Finland (project #211025), Helsinki Graduate School in Computer Science and Engineering, and the Emil Aaltonen Foundation. The author is grateful for insightful discussions with Petteri Kaski and Ilkka Niemelä.

$\pmod 2$ and thus the system always has at least one solution. Finally, transform the system $A\vec{x} \equiv \vec{b} \pmod 2$ into a $k$-CNF formula by introducing for every equation $\sum_{k=1}^{d} x_{j_k} \equiv b_i \pmod 2$ a set of $2^{d-1}$ clauses that forbid the combinations of truth values that violate the equation, i.e., all clauses on the variables $x_{j_k}$ with an even (respectively, odd) number of positive literals if $b_i = 1$ ($b_i = 0$).

## The Generator

The process described in (Haanpää *et al.* 2006) for generating $d$-regular bipartite graphs uniformly at random (u.a.r.) is unfortunately inefficient for $d > 3$. Thus we use the following (although not u.a.r.) approach to generate $d$-regular bipartite graphs with $2n$ vertices.

1. Let $G = K_{n,n}$ (the complete $2n$-vertex bipartite graph) and $H$ be the $2n$-vertex graph with no edges.
2. Repeat $d$ times:

2a) Obtain $G'$ by relabeling the vertices of $G = (V, E)$ with a permutation $\pi : 2n \to 2n$ selected uniformly at random.

2b) Let $M'$ be a maximum matching of $G'$ and $M = \pi^{-1}(M')$ (restore to the original vertex labeling).

2c) Add the edges of $M$ to $H$.

2d) Remove the edges of $M$ from $G$.

3. Return $H$.

An $\mathcal{O}(d|E|)$ algorithm (Schrijver 1999) is used for finding maximum matchings in $d$-regular bipartite graphs. See the accompanying web page for implementations of the original regular XORSAT generator and the described variation.

## Experiments

We investigate the hardness of regular $d$-XORSAT for various $d$ with the complete DPLL-based SAT solver zChaff (Fig. 1) (http://www.princeton.edu/˜chaff/zchaff.html) and the local search solver WalkSAT (Fig. 2) (http://www.cs.washington.edu/homes/kautz/walksat/), plotting the median number of decisions/flips over 15 instances. A comparison with satisfiable random $k$-XORSAT (Ricci-Tersenghi, Weight, & Zecchina 2001) with clauses-to-variables ratio 1 is presented in Fig. 3 using the DPLL-solver Satz (http://www.laria.u-picardie.fr/˜cli/EnglishPage.html). See the accompanying web page for comparisons running other solvers.

## Conclusions

By increasing the degree of the underlying regular constraint graphs, we observe a sharp increase in problem difficulty with respect to the number of variables. Regular $d$-XORSAT gives instances with only 50 variables on which state-of-the-art SAT solvers make in the order of $10^6$ decisions, the instances being harder than satisfiable random XORSAT.

Interesting further work includes theoretical analysis of the behavior of DPLL (with learning) and local search methods on regular $d$-XORSAT. A related research problem is to find a satisfiable $k$-CNF family on which the lower bound for DPLL converges to $\Omega(2^n)$ as $k \to \infty$.

## References


Achlioptas, D.; Gomes, C.; Kautz, H.; and Selman, B. 2000. Generating satisfiable problem instances. In *AAAI'00*, 256–261. AAAI Press.

Alekhnovich, M.; Hirsch, E.; and Itsykson, D. Exponential lower bounds for the running time of DPLL algorithms on satisfiable formulas. *J. Automated Reasoning*. To appear.

Boufkhad, Y.; Dubois, O.; Interian, Y.; and Selman, B. Regular random $k$-SAT: Properties of balanced formulas. *J. Automated Reasoning*. To appear.

Haanpää, H.; Järvisalo, M.; Kaski, P.; and Niemelä, I. 2006. Hard satisfiable clause sets for benchmarking equivalence reasoning techniques. *J. Satisfiability, Boolean Modeling and Computation* 2(1):27–46.

Jia, H.; Moore, C.; and Selman, B. 2005. From spin glasses to hard satisfiable formulas. In *SAT'05*, volume 3542 of *LNCS*, 199–210. Springer.

Ricci-Tersenghi, F.; Weight, M.; and Zecchina, R. 2001. Simplest random $K$-satisfiability problem. *Physical Review E* 63:026702.

Schrijver, A. 1999. Bipartite edge-colouring in $\mathcal{O}(\Delta m)$ time. *SIAM J. Computing* 28(3):841–846.


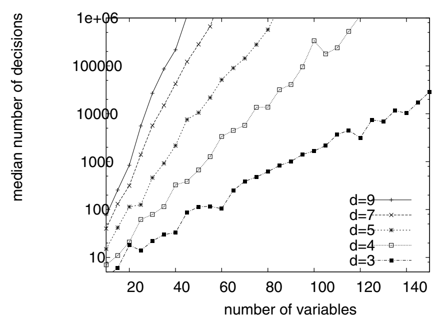


Figure 1: zChaff on regular $d$-XORSAT

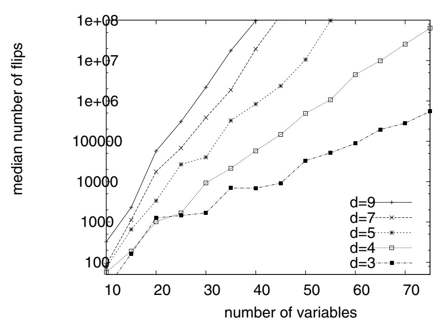


Figure 2: WalkSAT on regular $d$-XORSAT

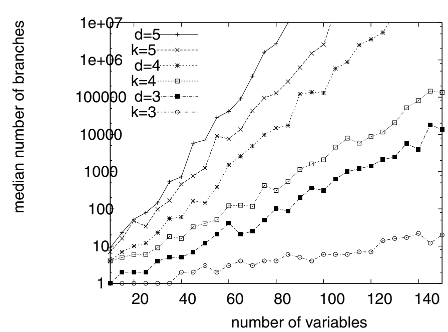


Figure 3: Satz: Regular $d$-XORSAT *v* random $k$-XORSAT