Combining the k-CNF and XOR Phase-Transitions *

Jeffrey M. Dudek
Rice University

Kuldeep S. Meel
Rice University

Moshe Y. Vardi
Rice University

Abstract

The runtime performance of modern SAT solvers on random k-CNF formulas is deeply connected with the 'phase-transition' phenomenon seen empirically in the satisfiability of random k-CNF formulas. Recent universal hashing-based approaches to sampling and counting crucially depend on the runtime performance of SAT solvers on formulas expressed as the conjunction of both k-CNF and XOR constraints (known as k-CNF-XOR formulas), but the behavior of random k-CNF-XOR formulas is unexplored in prior work. In this paper, we present the first study of the satisfiability of random k-CNF-XOR formulas. We show empirical evidence of a surprising phase-transition that follows a linear trade-off between k-CNF and XOR constraints. Furthermore, we prove that a phase-transition for k-CNF-XOR formulas exists for $k = 2$ and (when the number of k-CNF constraints is small) for $k > 2$.

1 Introduction

The Constraint-Satisfaction Problem (CSP) is one of the most fundamental problem in computer science, with a wide range of applications arising from diverse areas such as artificial intelligence, programming languages, biology and the like [Apt, 2003; Dechter, 2003]. The problem is, in general, NP-complete, and the study of run-time behavior of CSP techniques is a topic of major interest in AI, cf. [Dechter and Meiri, 1994]. Of specific interest is the behavior of CSP solvers on random problems [Cheeseman *et al.*, 1991]. Specifically, a deep connection was discovered between the density (ratio of clauses to variables) of random propositional CNF fixed-width (fixed number of literals per clause) formulas and the runtime behavior of SAT solvers on such formulas [Mitchell *et al.*, 1992; Crawford and Auton, 1993; Kirkpatrick and Selman, 1994]. The key experimental findings are: (1) as the density of random CNF instances increases, the probability of satisfiability decreases with a precipitous drop, believed to be a phase-transition, around the point where the satisfiability probability is 0.5, and (2) instances at the phase-transition point are particularly challenging for DPLL-based SAT solvers. Indeed, phase-transition instances serve as a source of challenging benchmark problems in SAT competitions [Belov *et al.*, 2014]. The connection between runtime performance and the satisfiability phase-transition has propelled the study of such phase-transition phenomena over the past two decades [Achlioptas, 2009], including detailed studies of how SAT solvers scale at different densities [Coarfa *et al.*, 2003; Mu and Hoos, 2015].

For random k-CNF formulas, where every clause contains exactly k literals, experiments suggest a specific phase-transition density, for example, density 4.26 for random 3-SAT, but establishing this analytically has been highly challenging [Coja-Oghlan and Panagiotou, 2013], and it has been established only for for $k = 2$ [Chvátal and Reed, 1992; Goerdt, 1996] and all large enough k [Ding *et al.*, 2015]. A phase-transition phenomenon has also been identified in random XOR formulas (conjunctions of XOR constraints). Creignou and Daudé [1999] proved a phase-transition at density 1 for variable-width random XOR formulas. Creignou and Daudé [2003] also proved the existence of a phase transition for random ℓ-XOR formulas (where each XOR-clause contains exactly ℓ literals), for $\ell \geq 1$, without specifying an exact location for the phase-transition. Dubois and Mandler [2002] independently identified the location of a phase transition for random 3-XOR formulas. More recently, Pittel and Sorkin [2015] identified the location of the phase-transition for ℓ-XOR formulas for $\ell > 3$.

Despite the abundance of prior work on the phase-transition phenomenon in the satisfiability of random k-CNF formulas and XOR formulas, no prior work considers the satisfiability of random formulas with *both* k-clauses and variable-width XOR-clauses together, henceforth referred as k-CNF-XOR formulas. Recently, successful hashing-based approaches to the fundamental problems of constrained sampling and counting employ SAT solvers to solve k-CNF-XOR formulas [Chakraborty *et al.*, 2013a; 2013b; 2014b; 2016; Meel *et al.*, 2016]. Unlike previous approaches to sampling and counting, hashing-based approaches provide strong theoretical guarantees and scale to real-world instances involving formulas with *hundreds of thousands* of variables. The scalability of these approaches crucially depends on the runtime performance of SAT solvers in handling k-CNF-XOR

*The author list has been sorted alphabetically by last name; this should not be used to determine the extent of authors' contributions.

formulas [Ivrii *et al.*, 2015]. Moreover, since the phase-transition behavior of k-CNF constraints have been analyzed to explain runtime behavior of SAT solvers [Achlioptas and Coja-Oghlan, 2008], we believe that analysis of the phase-transition phenomenon for k-CNF-XOR formula is the first step towards demystifying the runtime behavior of CNF-XOR solvers such as CryptoMiniSAT [Soos *et al.*, 2009] and thus explain the runtime behavior of hashing-based algorithms.

The primary contribution of this work is the first study of phase-transition phenomenon in the satisfiability of random k-CNF-XOR formulas, henceforth referred to as the k-CNF-XOR phase-transition. In particular:

1. We present (in Section 3) experimental evidence for a k-CNF-XOR phase-transition that follows a linear trade-off between k-CNF clauses and XOR clauses.
2. We prove (in Section 4) that the k-CNF-XOR phase-transition exists when the ratio of k-CNF clauses to variables is small. This fully characterizes the phase-transition when $k = 2$.
3. We prove (in Section 4) upper and lower bounds on the location of the k-CNF-XOR phase-transition region.
4. We conjecture (in Section 5) that the exact location of a phase-transition for $k \geq 3$ follows the linear trade-off between k-CNF and XOR clauses seen experimentally.

2 Notations and Preliminaries

Let $X = \{X_1, \cdots, X_n\}$ be a set of propositional variables and let F be a formula defined over X. A *satisfying assignment* or *witness* of F is an assignment of truth values to the variables in X such that F evaluates to true. Let $\#F$ denote the number of satisfying assignments of F. We say that F is *satisfiable* (or *sat.*) if $\#F > 0$ and that F is *unsatisfiable* (or *unsat.*) if $\#F = 0$.

We use $\mathsf{Pr}(X)$ to denote the probability of event X. We say that an infinite sequence of random events $E_1, E_2, \cdots$ occurs *with high probability* (denoted, w.h.p.) if $\lim_{n\to\infty} \mathsf{Pr}(E_n) = 1$. We use $\mathsf{E}[Y]$ and $\mathsf{Var}[Y]$ to denote respectively the expected value and variance of a random variable Y. We use $\mathsf{Cov}[Y, Z]$ to denote the covariance of random variables Y and Z. We use $o_k(1)$ to denote a term which converges to 0 as $k \to \infty$.

A k-*clause* is the disjunction of k literals out of $\{X_1, \cdots, X_n\}$, with each variable possibly negated. For fixed positive integers k and n and a nonnegative real number r, let the random variable $F_k(n, rn)$ denote the formula consisting of the conjunction of $\lceil rn \rceil$ k-clauses, with each clause chosen uniformly and independently from all $\binom{n}{k}2^k$ possible k-clauses over n variables.

The early experiments on $F_k(n, rn)$ [Mitchell *et al.*, 1992; Crawford and Auton, 1993; Kirkpatrick and Selman, 1994] led to the following conjecture:

Conjecture 1 (Satisfiability Phase-Transition Conjecture)**.** *For every integer $k \geq 2$, there is a critical ratio r_k such that:*

1. *If $r < r_k$, then $F_k(n, rn)$ is satisfiable w.h.p.*
2. *If $r > r_k$, then $F_k(n, rn)$ is unsatisfiable w.h.p.*

The Conjecture was quickly proved for $k = 2$, where $r_2 = 1$ [Chvátal and Reed, 1992; Goerdt, 1996]. In recent work, Ding, Sly, and Sun established the Satisfiability Phase Transition Conjecture for all sufficiently large k [Ding *et al.*, 2015]. The Conjecture has remained elusive for small values of $k \geq 3$, although values for these critical ratios r_k can be estimated experimentally, e.g., r_3 seems to be near 4.26.

An XOR-clause over n variables is the 'exclusive or' of either 0 or 1 together with a subset of the variables $X_1, \cdots, X_n$. An XOR-clause including 0 (respectively, 1) evaluates to true if and only if an odd (respectively, even) number of the included variables evaluate to true. Note that all k-clauses contain *exactly* k variables, whereas the number of variables in an XOR-clause is not fixed; a uniformly chosen XOR-clause over n variables contains $\frac{n}{2}$ variables in expectation.

For a fixed positive integer n and a nonnegative real number s, let the random variable $Q(n, sn)$ denote the formula consisting of the conjunction of $\lceil sn \rceil$ XOR-clauses, with each clause chosen uniformly and independently from all 2^{n+1} XOR-clauses over n variables. Creignou and Daude [1999; 2003] proved a phase-transition in the satisfiability of $Q(n, sn)$: if $s < 1$ then $Q(n, sn)$ is satisfiable w.h.p., while if $s > 1$ then $Q(n, sn)$ is unsatisfiable w.h.p.

A k-CNF-XOR formula is the conjunction of some number of k-clauses and XOR-clauses. For fixed positive integers k and n and fixed nonnegative real numbers r and s, let the random variable $\psi_k(n, rn, sn)$ denote the formula consisting of the conjunction of $\lceil rn \rceil$ k-clauses and $\lceil sn \rceil$ XOR-clauses, with each clause chosen uniformly and independently from all possible k-clauses and XOR-clauses over n variables. (The motivation for using fixed-width clauses and variable-width XOR-clauses comes from the hashing-based approaches to constrained sampling and counting discussed in Section 1.) Although random k-CNF formulas and XOR formulas have been well studied separately, no prior work considers the satisfiability of random mixed formulas arising from conjunctions of k-clauses and XOR-clauses.

3 Experimental Results

To explore empirically the behavior of the satisfiability of k-CNF-XOR formulas, we built a prototype implementation in Python that employs the CryptoMiniSAT[1] [Soos *et al.*, 2009] solver to check satisfiability of k-CNF-XOR formulas. We chose CryptoMiniSAT due to its ability to handle the combination of k-clauses and XOR-clauses efficiently [Chakraborty *et al.*, 2014b; 2014a]. The objective of the experimental setup was to empirically determine the behavior of $\mathsf{Pr}(\psi_k(n, rn, sn)$ is sat) with respect to r and s, the k-clause and XOR-clause densities respectively, for fixed k and n.

3.1 Experimental Setup

We ran 11 experiments with various values of k and n. For $k = 2$, we ran experiments for $n \in \{25, 50, 100, 150\}$. For $k = 3$, we ran experiments for $n \in \{25, 50, 100\}$. For $k = 4$ and $k = 5$, we ran experiments for $n \in \{25, 50\}$. We were not able to run experiments for values of n significantly larger

[1] http://www.msoos.org/cryptominisat4/

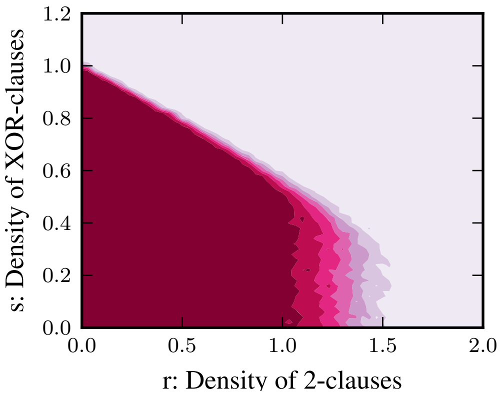

Figure 1: Phase transition for 2-CNF-XOR formulas

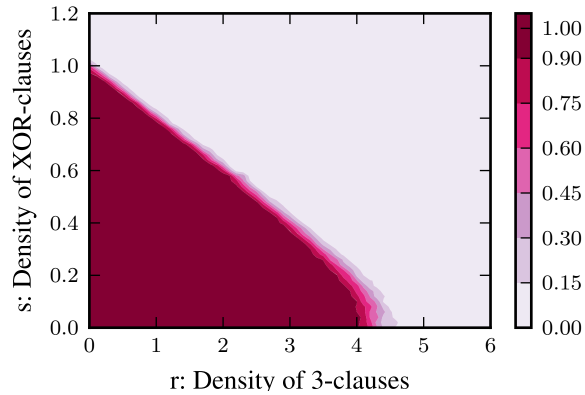

Figure 2: Phase transition for 3-CNF-XOR formulas

than those listed above: at some k-clause and XOR-clause densities, the run-time of CryptoMiniSAT scaled far beyond our computational capabilities.

In each experiment, the XOR-clause density s ranged from 0 to 1.2 in increments of 0.02. Since the location of phase-transition for $k-$CNF depends on k, the range of k-clause density r also depends on k. For $k = 3$, r ranged from from 0 to 6 in increments of 0.04; for $k = 5$, r ranged from 0 to 26 in increments of 0.43, and the like.

To uniformly choose a k-clause we uniformly selected without replacement k out of the variables $\{X_1, \cdots, X_n\}$. For each selected variable X_i, we include exactly one of the literals X_i or $\neg X_i$ in the k-clause, each with probability $1/2$. The disjunction of these k literals is a uniformly chosen k-clause. To uniformly choose an XOR-clause, we include each variable of $\{X_1, \cdots, X_n\}$ with probability $1/2$ in a set A of variables. Additionally we include in A exactly one of 0 or 1, each with probability $\frac{1}{2}$. The 'exclusive-or' of all elements of A is a uniformly chosen XOR-clause. For each assignment of values to k, n, r, and s, we evaluated satisfiability, using CryptoMiniSAT, of 100 uniformly generated formulas of $\psi_k(n, rn, sn)$ by constructing the conjunction of $\lceil rn \rceil$ k-clauses and $\lceil sn \rceil$ XOR-clauses, with each clause chosen uniformly and independently as described above. The percentage of satisfiable formulas gives us an empirical estimate of $\mathsf{Pr}(\psi_k(n, rn, sn)$ is satisfiable).

Each experiment was run on a node within a high-performance computer cluster. These nodes contain 12-processor cores at 2.83 GHz each with 48 GB of RAM per node. Each formula was given a timeout of 1000 seconds.

3.2 Results

We present scatter plots demonstrating the behavior of satisfiability of k-CNF-XOR formulas. For lack of space, we present results only for three experiments[2]. The plots for $k = 2, 3$ and 5 are shown in Figure 1, 2, and 3 respectively. The value of n is set to 150, 100, and 50 respectively for the three experiments above.

[2]The data from all experiments is available at `http://www.cs.rice.edu/CS/Verification/Projects/CUSP/`

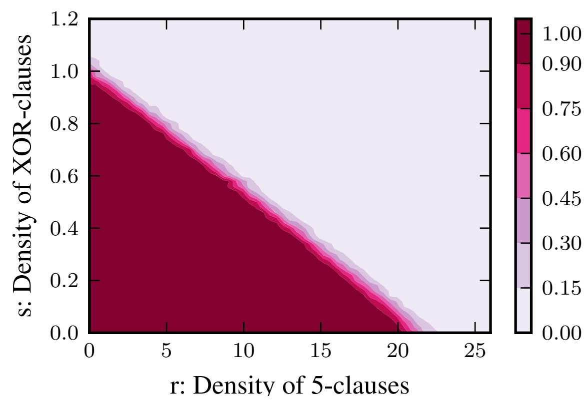

Figure 3: Phase transition for 5-CNF-XOR formulas

Each figure is a 2D plot, representing the observed probability that $\psi_k(n, rn, sn)$ is satisfiable as the density of k-clauses r and the density of XOR-clauses s varies. The x-axis indicates the density of k-clauses r. The y-axis indicates the density of XOR-clauses s. The dark (respectively, light) regions represent clause densities where almost all (respectively, no) sampled formulas were satisfiable.

Note that $\psi_k(n, rn, sn)$ consists only of XOR clauses when $r = 0$. Examining the figures along the line $r = 0$ the phase-transition location is around ($r = 0$, $s = 1$), which matches previous theoretical results on the phase-transition for XOR formulas [Creignou and Daudé, 1999]. Likewise, $\psi_k(n, rn, 0) = F_k(n, rn)$ and, by examining the figures along the line $s = 0$, we observe phase-transition locations that match previous studies on the phase-transition for k-CNF formulas for $k = 2, 3,$ and 5 [Achlioptas, 2009]. Note that the phase-transition we observe for 2-CNF formulas is slightly above the true location at $s = 1$ [Chvátal and Reed, 1992; Goerdt, 1996]; the correct phase-transition point for 2-CNF formulas is observed only when the number of variables is above 4096 [Wilson, 2000].

In all the plots, we observe a large triangular region where the probability that $\psi_k(n, rn, sn)$ is satisfiable is nearly 1. We likewise observe a separate region where the observed probability that $\psi_k(n, rn, sn)$ is satisfiable is nearly 0. More surprisingly, the shared boundary between the two regions for large areas of the plots seems to be a constant-slope line. A closer examination of this line at the bottom-right corners of the figures for $k = 2$ and $k = 3$, where the k-clause density is large, reveals that the line appears to "kink" and abruptly change slope. We discuss this further in Section 5.

4 Establishing a Phase-Transition

The experimental results presented in Section 3 empirically demonstrate the existence of a k-CNF-XOR phase-transition. Theorem 1 shows that the k-CNF-XOR phase-transition exists when the density of k-clauses is small. In particular, the function $\phi_k(r)$ (defined in Lemma 3) gives the location of a phase-transition between a region of satisfiability and a region of unsatisfiability in random k-CNF-XOR formulas.

Theorem 1. *Let $k \geq 2$. There is a function $\phi_k(r)$, a constant $\alpha_k \geq 1$, and a countable set of real numbers $\mathcal{C}_k$ (all defined in Lemma 3) such that for all $r \in [0, \alpha_k) \backslash \mathcal{C}_k$ and $s \geq 0$:*

(a). If $s < \phi_k(r)$, then w.h.p. $\psi_k(n, rn, sn)$ is satisfiable.

(b). If $s > \phi_k(r)$, then w.h.p. $\psi_k(n, rn, sn)$ is unsatisfiable.

Proof. Part (a) follows directly from Lemma 9. Part (b) follows directly from Lemma 14. The proofs of these lemmas are presented in Sections 4.1 and 4.2 respectively. □

$\phi_k(r)$ is the *free-entropy density* of k-CNF, drawing on concepts from spin-glass theory [Gogioso, 2014]. From the expression for $\phi_k(r)$ in Lemma 3, it is easily verified that $\phi_k(0) = 1$ and that $\phi_k(r)$ is a monotonically decreasing function of r. Thus when the k-clause density (r) is 0, Theorem 1 says that an XOR-clause density of 1 is a phase-transition for XOR-formulas, matching previously known results [Creignou and Daudé, 1999]. As the k-clause density increases, $\phi(r)$ is decreasing and so the XOR-clause density required to reach the phase-transition decreases.

Theorem 1 fully characterizes the random satisfiability of $\psi_k(n, rn, sn)$ when $r < 1$. In the case $k = 2$, prior results on random 2-CNF satisfiability characterize the rest of the region. If $r > 1$, then $F_2(n, rn)$ is unsatisfiable w.h.p. [Chvátal and Reed, 1992; Goerdt, 1996] and so the 2-clauses within $\psi_2(n, rn, sn)$ are unsatisfiable w.h.p. without considering the XOR-clauses. Therefore $\psi_2(n, rn, sn)$ is unsatisfiable w.h.p. if $r > 1$. This, together with Theorem 1, proves that $\phi_2(r)$ is the complete location of the 2-CNF-XOR phase-transition.

Moreover, Lemma 4 shows that $\alpha_k \geq (1 - o_k(1)) \cdot 2^k \ln(k)/k$ (where $o_k(1)$ denotes a term that converges to 0 as $k \to \infty$) and so Theorem 1 shows that a phase-transition exists until near $r = 2^k \ln(k)/k$ for sufficiently large k.

For small $k \geq 3$, the region $r < 1$ characterized by Theorem 1 is only a small portion of the region where the subset of k-clauses remains satisfiable. Moreover, the location of the phase-transition $\phi_k(r)$ given by Theorem 1 is difficult to compute directly. Theorem 2 gives explicit lower and upper bounds on the location of a phase-transition region.

Theorem 2. *Let $k \geq 3$. There is a function $\Lambda_b(k, r)$ (defined in Lemma 5) such that for all $s \geq 0$ and $r \geq 0$:*

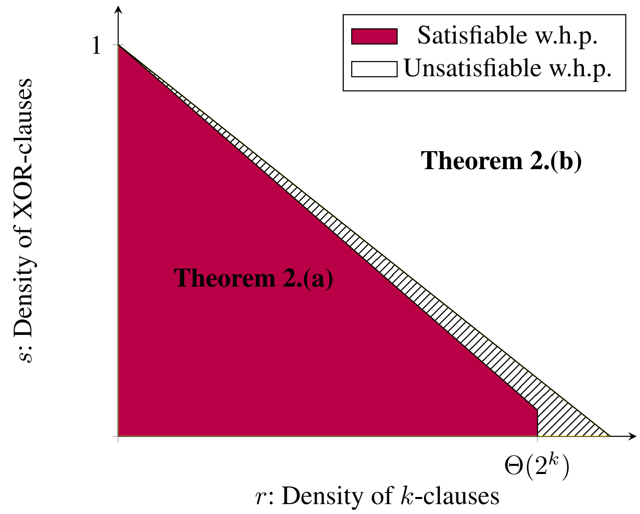

Figure 4: Satisfiability of $\psi_k(n, rn, sn)$ as $n \to \infty$

(a). If $s < \frac{1}{2}\log_2(\Lambda_b(k, r))$ and $r < 2^k \ln(2) - \frac{1}{2}((k + 1)\ln(2) + 3)$, then w.h.p. $\psi_k(n, rn, sn)$ is satisfiable.

(b). If $s > r\log_2(1 - 2^{-k}) + 1$, then w.h.p. $\psi_k(n, rn, sn)$ is unsatisfiable.

Proof. Part (a) follows directly from Lemma 10. Part (b) follows directly from Lemma 15. The proofs of these lemmas are presented in Sections 4.1 and 4.2 respectively. □

Both the upper bound $r\log_2(1 - 2^{-k}) + 1$ and (using the expression for $\Lambda_b(k, r)$ in Lemma 5) the lower bound $\frac{1}{2}\log_2(\Lambda_b(k, r))$ are linear in r. When the k-clause density r is 0, Theorem 2 agrees with Theorem 1. As the k-clause density increases past $\Theta(2^k)$, Theorem 2 no longer gives a lower bound on the location of a possible phase-transition.

4.1 A Proof of the Lower Bound

We now establish Theorem 1.(a) and Theorem 2.(a), which follow directly from Lemma 9 and Lemma 10 respectively.

The key idea in the proof of these lemmas is to decompose $\psi_k(n, rn, sn)$ into independently generated k-CNF and XOR formulas, so that $\psi_k(n, rn, sn) = F_k(n, rn) \wedge Q(n, sn)$. We can then bound the number of solutions to $F_k(n, rn)$ from below with high probability and bound from below the probability that $F_k(n, rn)$ becomes unsatisfiable after including XOR-clauses on top of $F_k(n, rn)$.

The following three lemmas achieve the first of the two tasks. The first, Lemma 3, gives a tight bound on $\#F_k(n, rn)$ for small k-clause densities.

Lemma 3. *Let $k \geq 2$ and let α_k be the supremum of $\{r : \exists\delta > 0$ s.t. $\Pr(F_k(n, rn)$ is unsat.$) \leq O(1/(\log n)^{1+\delta})\}$. Then $\alpha_k \geq 1$. Furthermore, there exists a countable set of real numbers $\mathcal{C}_k$ such that for all $r \in [0, \alpha_k) \backslash \mathcal{C}_k$:*

(a). The sequence $\frac{1}{n}\mathsf{E}\left[\log_2(\#F_k(n, rn)) \mid F_k(n, rn) \text{ is sat.}\right]$ converges to a limit as $n \to \infty$. Let $\phi_k(r)$ be this limit.

(b). For all $\epsilon > 0$, w.h.p. $(2^{\phi_k(r)-\epsilon})^n \leq \#F_k(n, rn)$.

(c). For all $\epsilon > 0$, w.h.p. $(2^{\phi_k(r)+\epsilon})^n \geq \#F_k(n, rn)$.

Proof. These proofs are given in [Abbe and Montanari, 2014]. $\alpha_k \geq 1$ is given as **Remark 2**. Part (a) is given as **Theorem 3**. Parts (b) and (c) are given as **Theorem 1**. □

We abuse notation to let $\phi_k(r)$ denote the limit of the sequence in Lemma 3.(a) for all $r > 0$, although a priori this sequence may not converge for $r \geq \alpha_k$. Later work refined the value of α_k in Lemma 3 for sufficiently large k and so extended the tight bound on $\#F_k(n, rn)$. In particular, Lemma 4 implies that $\alpha_k \geq (1 - o_k(1)) \cdot 2^k \ln(k)/k$.

Lemma 4. *Let* $k \geq 2$*. For all* $r \geq 0$*, if* $r \leq (1 - o_k(1)) \cdot 2^k \ln(k)/k$ *then* $\Pr(F_k(n, rn)$ *is sat.*$) \geq 1 - O(1/n)$*.*

Proof. The proof of this is given as **Theorem 1.3** of [Coja-Oghlan and Reichman, 2013]. □

It is difficult to compute $\phi_k(r)$ directly. Instead, Lemma 5 provides a weaker but explicit lower bound on $\#F_k(n, rn)$.

Lemma 5. *Let* $k \geq 3$*,* $\epsilon > 0$*, and* $r \geq 0$*. Let* β_k *be the smallest positive solution to* $\beta_k(2 - \beta_k)^{k-1} = 1$ *and define* $\Lambda_b(k, r) = 4(((1 - \beta_k/2)^k - 2^{-k})^2/(1 - \beta_k)^k)^r$*.*

If $r < 2^k \ln(2) - \frac{1}{2}((k + 1)\ln(2) + 3)$*, then w.h.p.* $\frac{1}{2}(\Lambda_b(k, r) - \epsilon)^{n/2} \leq \#F_k(n, rn)$*.*

Proof. The proof of this is given on page 264 of [Achlioptas *et al.*, 2011] within Section 6 (**Proof of Theorem 6**); the definition of $\Lambda_b(k, r)$ is given as equation (20). □

Before we analyze how the solution space of $F_k(n, rn)$ interacts with the solution space of $Q(n, sn)$, we must characterize the solution space of $Q(n, sn)$. The following lemma shows that the solutions of $Q(n, sn)$ are pairwise independent, meaning that a single satisfying assignment of $Q(n, sn)$ gives no information on other satisfying assignments.

Lemma 6. *Let* $n \geq 0$ *and* $s \geq 0$*. If* σ *and* σ' *are distinct assignments of truth values to the variables* $\{X_1, \cdots, X_n\}$*:*

(a). $\Pr(\sigma$ *satisfies* $Q(n, sn)) = 2^{-\lceil sn \rceil}$

(b). $\Pr(\sigma$ *satisfies* $Q(n, sn) \mid \sigma'$ *satisfies* $Q(n, sn)) = 2^{-\lceil sn \rceil}$

Proof. The proof of this is given in the proof of Lemma 1 of [Gomes *et al.*, 2007]. □

The following lemma bounds from below the probability that a formula H (in Lemma 8 we take $H = F_k(n, rn)$) remains satisfiable after including XOR-clauses on top of H. This result and proof is similar to Corollary 3 from [Gomes *et al.*, 2006].

Lemma 7. *Let* $\alpha \geq 1$*,* $s \geq 0$*,* $n \geq 0$*, and let* H *be a formula defined over* $\{X_1, \cdots, X_n\}$*. Then* $\Pr(H \wedge Q(n, sn)$ *is satisfiable* $\mid \#H \geq 2^{\lceil sn \rceil + \alpha}) \geq 1 - 2^{-\alpha}$*.*

Proof. Let R be the set of all truth assignments to the variables in X that satisfy H; there are $\#H$ such truth assignments. For every truth assignment $\sigma \in R$, let Y_σ be a 0-1 random variable that is 1 if σ satisfies $H \wedge Q(n, sn)$ and 0 otherwise. Note that $\mathsf{Var}[Y_\sigma] = \mathsf{E}[Y_\sigma^2] - \mathsf{E}[Y_\sigma]^2 \leq \mathsf{E}[Y_\sigma^2]$. Since Y_σ is a 0-1 random variable, $Y_\sigma^2 = Y_\sigma$ and thus $\mathsf{Var}[Y_\sigma] \leq \mathsf{E}[Y_\sigma]$.

Let σ and σ' be distinct truth assignments in R. By Lemma 6, $\mathsf{E}[Y_\sigma Y_{\sigma'}] = \mathsf{E}[Y_\sigma]\mathsf{E}[Y_{\sigma'}] = 2^{-\lceil sn \rceil} \cdot 2^{-\lceil sn \rceil}$. Thus $\mathsf{Cov}[Y_\sigma, Y_{\sigma'}] = \mathsf{E}[Y_\sigma Y_{\sigma'}] - \mathsf{E}[Y_\sigma]\mathsf{E}[Y_{\sigma'}] = 0$.

Let the random variable Y be the number of solutions to $H \wedge Q(n, sn)$, so $Y = \#(H \wedge Q(n, sn)) = \sum_\sigma Y_\sigma$. Thus $\mathsf{Var}[Y] = \mathsf{Var}[\sum_\sigma Y_\sigma] = \sum_\sigma \mathsf{Var}[Y_\sigma] + \sum_{\sigma \neq \sigma'} \mathsf{Cov}[Y_\sigma, Y_{\sigma'}]$. Since the covariance of Y_σ and $Y_{\sigma'}$ is 0 for all pairs of distinct truth assignments σ and σ' in R, we get that $\mathsf{Var}[Y] = \sum_\sigma \mathsf{Var}[Y_\sigma]$. Since $\mathsf{Var}[Y_\sigma] \leq \mathsf{E}[Y_\sigma]$ for all truth assignments σ in R, we get that $\mathsf{Var}[Y] \leq \sum_\sigma \mathsf{E}[Y_\sigma]$. Since $\mathsf{E}[Y] = \mathsf{E}[\sum_\sigma Y_\sigma] = \sum_\sigma \mathsf{E}[Y_\sigma]$, we conclude that $\mathsf{Var}[Y] \leq \mathsf{E}[Y]$. Moreover, since $\mathsf{E}[Y_\sigma] = \Pr(\sigma \text{ satisfies } Q(n, sn)) = 2^{-\lceil sn \rceil}$ we get $\mathsf{E}[Y] = \#H \cdot 2^{-\lceil sn \rceil}$.

Let the event $\neg E_n$ denote that $H \wedge Q(n, sn)$ is unsatisfiable. Thus if $\neg E_n$ occurs then then $Y = 0$ and so $|Y - \mathsf{E}[Y]| \geq \mathsf{E}[Y]$. This implies that $\Pr(\neg E_n) \leq \Pr(|Y - \mathsf{E}[Y]| \geq \mathsf{E}[Y])$. Chebyshev's inequality says that $\Pr(|Y - \mathsf{E}[Y]| \geq \mathsf{E}[Y]) \leq \mathsf{Var}[Y] \ / \ \mathsf{E}[Y]^2$. It follows that $\Pr(\neg E_n) \leq \mathsf{Var}[Y] \ / \ \mathsf{E}[Y]^2$. Since $\mathsf{Var}[Y] \leq \mathsf{E}[Y]$, we get that $\Pr(\neg E_n) \leq \mathsf{E}[Y]^{-1}$. Therefore by plugging in the value for $\mathsf{E}[Y]$ we get $\Pr(\neg E_n) \leq (\#H)^{-1} \cdot 2^{\lceil sn \rceil}$.

Finally, if we assume that $\#H \geq 2^{\lceil sn \rceil + \alpha}$ then $(\#H)^{-1} \cdot 2^{\lceil sn \rceil} \leq 2^{-\alpha}$. Therefore $\Pr(\neg E_n \mid \#H \geq 2^{\lceil sn \rceil + \alpha}) \leq 2^{-\alpha}$. □

Using the key behavior of XOR-clauses described in Lemma 7, we can transform lower bounds (w.h.p.) on the number of solutions to $F_k(n, rn)$ into lower bounds on the location of a possible k-CNF-XOR phase-transition.

Lemma 8. *Let* $k \geq 2$*,* $s \geq 0$*, and* $r \geq 0$*. Let* $B_1, B_2, \cdots$ *be an infinite convergent sequence of positive real numbers such that* $B_i^n \leq \#F_k(n, rn)$ *occurs w.h.p. for all* $i \geq 1$*. If* $s < \log_2(\lim_{i \to \infty} B_i)$*, then w.h.p.* $\psi_k(n, rn, sn)$ *is satisfiable.*

Proof. For all integers $n \geq 0$, let the event E_n denote the event when $\psi_k(n, rn, sn)$ is satisfiable. We would like to show that $\Pr(E_n)$ converges to 1 as $n \to \infty$.

The general idea of the proof follows. We first decompose $\psi_k(n, rn, sn)$ as $\psi_k(n, rn, sn) = F_k(n, rn) \wedge Q(n, sn)$. Let the event L_n denote the event when the number of solutions of $F_k(n, rn)$ is bounded from below (by a lower bound to be specified later). We show that L_n occurs w.h.p.. Next, we use Lemma 7 to bound from below the probability that $F_k(n, rn) \wedge Q(n, sn)$ remains satisfiable given that $F_k(n, rn)$ has enough solutions; we use this to show that $\Pr(E_n \mid L_n)$ converges to 1 as $n \to \infty$. Finally, we combine these results to prove that $\Pr(E_n)$ converges to 1.

Since $2^s < \lim_{i \to \infty} B_i$, there is some integer $i \geq 1$ such that $2^s < B_i$. Define the event L_n as the event when $\#F_k(n, rn) \geq B_i^n$. Then L_n occurs w.h.p. by hypothesis.

Next, we show that $\Pr(E_n \mid L_n)$ converges to 1. Choose $\delta > 0$ and $N > 0$ such that $2^{s + \delta + 1/N} < B_i$; we can always find sufficiently small δ and sufficiently large N such that this holds. Since we are concerned only with the behavior of $\Pr(E_n \mid L_n)$ in the limit, we can restrict our attention only to large enough n. In particular, consider $n > 2N$. Then we get that $2^{sn + \delta n + 2} < B_i^n$ and so $2^{\lceil sn \rceil + \delta n + 1} < B_i^n$. Let $\alpha =$

$\delta n + 1$, so that $2^{\lceil sn \rceil + \alpha} \leq B_i^n$. Then Lemma 7 says that $\Pr(E_n \mid L_n) \geq 1 - 2^{-\delta n - 1}$. Since $1 - 2^{-\delta n - 1}$ converges to 1 as $n \to \infty$, $\Pr(E_n \mid L_n)$ must also converge to 1.

Thus both $\Pr(E_n \mid L_n)$ and $\Pr(L_n)$ converge to 1 as $n \to \infty$. Since $\Pr(E_n \cap L_n) = \Pr(E_n \mid L_n) \cdot \Pr(L_n)$, this implies that $\Pr(E_n \cap L_n)$ also converges to 1. Since $\Pr(E_n \cap L_n) \leq \Pr(E_n) \leq 1$, this implies that $\Pr(E_n)$ converges to 1. ☐

Finally, it remains only to use Lemma 8 to obtain bounds on the k-CNF-XOR phase-transition. The tight lower bound on $\#F_k(n, rn)$ from Lemma 3.(b) corresponds to a tight lower bound on the location of the phase-transition.

Lemma 9. *Let $k \geq 2$, and let α_k, $\mathcal{C}_k$, and $\phi_k(r)$ be as defined in Lemma 3. For all $r \in [0, \alpha_k) \backslash \mathcal{C}_k$ and $s \in [0, \phi_k(r))$, $\psi_k(n, rn, sn)$ is satisfiable w.h.p..*

Proof. Let $B_i = 2^{\phi_k(r) - 1/i}$. By Lemma 3.(b), $B_i^n \leq \#F_k(n, rn)$ w.h.p. for all $i \geq 1$. Furthermore, $\lim_{i \to \infty} B_i = 2^{\phi_k(r)}$ and so $s < \log_2(\lim_{i \to \infty} B_i)$. Thus $\psi_k(n, rn, sn)$ is satisfiable w.h.p. by Lemma 8. ☐

The weaker lower bound on $\#F_k(n, rn)$ from Lemma 5 corresponds to a weaker lower bound on the location of the phase-transition.

Lemma 10. *Let $k \geq 3$, $s \geq 0$, and $r \geq 0$. If $r < 2^k \ln(2) - \frac{1}{2}(k+1)\ln(2) + \frac{3}{2}$ and $s < \frac{1}{2}\log_2(\Lambda_b(k, r))$, then $\psi_k(n, rn, sn)$ is satisfiable w.h.p..*

Proof. Let $B_i = (\Lambda_b(k, r) - 1/i)^{1/2}$. This is an increasing sequence in i, so $\log_2(B_{i+1}/B_i)$ is positive for all $i \geq 1$. Consider one such $i \geq 1$ and define $N_i = 1/\log_2(B_{i+1}/B_i)$. Then for all $n > N_i$ it follows that $2^{1/n} < B_{i+1}/B_i$ and so $B_i^n < \frac{1}{2} B_{i+1}^n$. By Lemma 5, $\frac{1}{2} B_{i+1}^n \leq \#F_k(n, rn)$ w.h.p. and therefore $B_i^n < \frac{1}{2} B_{i+1}^n \leq \#F_k(n, rn)$ w.h.p. as well.

Furthermore, $\lim_{i \to \infty} B_i = \Lambda_b(k, r)^{1/2}$ and so $s < \log_2(\lim_{i \to \infty} B_i)$. Thus $\psi_k(n, rn, sn)$ is satisfiable w.h.p. by Lemma 8. ☐

4.2 A Proof of the Upper Bound

We now establish Theorem 1.(b) and Theorem 2.(b), which follow directly from Lemma 14 and Lemma 15 respectively.

Similar to Section 4.1, the key idea in the proof of these lemmas is to decompose $\psi_k(n, rn, sn)$ into independently generated k-CNF and XOR formulas, so that $\psi_k(n, rn, sn) = F_k(n, rn) \wedge Q(n, sn)$. We can then bound the number of solutions to $F_k(n, rn)$ from above with high probability and bound from below the probability that $F_k(n, rn)$ becomes unsatisfiable after including XOR-clauses on top of $F_k(n, rn)$.

The first of these two tasks is accomplished through Lemma 3.(c), which gives a tight upper bound on $\#F_k(n, rn)$ for small k-clause densities, and by Lemma 11, which gives a weaker explicit upper bound on $\#F_k(n, rn)$.

Lemma 11. *For all $\epsilon > 1$, $k \geq 2$, and $r \geq 0$, w.h.p. $\#F_k(n, rn) < (2\epsilon \cdot (1 - 2^{-k})^r)^n$.*

Proof. Let $X = \#F_k(n, rn)$. For a random assignment on n variables σ, note that $\Pr(\sigma \text{ satisfies } F_k(n, 1)) = (1 - 2^{-k})$. Since the $\lceil rn \rceil$ k-clauses of $F_k(n, rn)$ were chosen independently, this implies that $\mathsf{E}[X] = 2^n (1 - 2^{-k})^{\lceil rn \rceil}$.

By Markov's inequality, we get $\Pr(X \geq \epsilon^n \mathsf{E}[X]) \leq \mathsf{E}[X] / (\epsilon^n \mathsf{E}[X]) = \epsilon^{-n}$. Since $1 - 2^{-k} < 1$ and so $2^n \epsilon^n (1 - 2^{-k})^{\lceil rn \rceil} \leq 2^n \epsilon^n (1 - 2^{-k})^{rn}$, it follows that $\Pr(X \geq \epsilon^n 2^n (1 - 2^{-k})^{rn}) \leq \Pr(X \geq \epsilon^n \mathsf{E}[X]) \leq \epsilon^{-n}$. Thus $\lim_{n \to \infty} \Pr(X < \epsilon^n 2^n (1 - 2^{-k})^{rn}) = 1$. ☐

The following lemma bounds from below the probability that a formula H (in Lemma 13 we take $H = F_k(n, rn)$) remains satisfiable after including XOR-clauses on top of H. This result and proof is similar to Corollary 1 from [Gomes *et al.*, 2006].

Lemma 12. *Let $\alpha \geq 1$, $s \geq 0$, $n \geq 0$, and let H be a formula defined over $X = \{X_1, \cdots, X_n\}$. Then $\Pr(H \wedge Q(n, sn) \text{ is unsatisfiable} \mid \#H \leq 2^{\lceil sn \rceil - \alpha}) \geq 1 - 2^{-\alpha}$.*

Proof. Let the random variable Y denote $\#(H \wedge Q(n, sn))$ as in Lemma 7. Markov's inequality implies that $\Pr(Y \geq 1) \leq \mathsf{E}[Y]$. Recall from Lemma 7 that $\mathsf{E}[Y] = \#H \cdot 2^{-\lceil sn \rceil}$, so $\Pr(Y \geq 1) \leq \#H \cdot 2^{-\lceil sn \rceil}$. If $\#H \leq 2^{\lceil sn \rceil - \alpha}$, then $\#H \cdot 2^{-\lceil sn \rceil} \leq 2^{-\alpha}$. Thus $\Pr(Y \geq 1 \mid \#H \leq 2^{\lceil sn \rceil - \alpha}) \leq 2^{-\alpha}$. Since $H \wedge Q(n, sn)$ is unsatisfiable exactly when $Y = 0$, we conclude $\Pr(H \wedge Q(n, sn) \text{ is unsatisfiable}) \geq 1 - 2^{-\alpha}$. ☐

Using the key behavior of XOR-clauses described in Lemma 12, we can transform upper bounds (w.h.p.) on the number of solutions to $F_k(n, rn)$ into upper bounds on the location of a possible k-CNF-XOR phase-transition.

Lemma 13. *Let $k \geq 2$, $s \geq 0$, and $r \geq 0$. Let B_1, B_2, $\cdots$ be an infinite convergent sequence of positive real numbers such that $\#F_k(n, rn) \leq B_i^n$ occurs w.h.p. for all $i \geq 1$. If $s > \log_2(\lim_{i \to \infty} B_i)$, then w.h.p. $\psi_k(n, rn, sn)$ is unsatisfiable.*

Proof. For all integers $n \geq 0$, let the event $\neg E_n$ denote the event when $\psi_k(n, rn, sn)$ is unsatisfiable. We would like to show that $\Pr(\neg E_n)$ converges to 1 as $n \to \infty$.

The general idea of the proof follows. Note that $\psi_k(n, rn, sn) = F_k(n, rn) \wedge Q(n, sn)$ as in Lemma 8. Let the event U_n denote the event when the number of solutions of $F_k(n, rn)$ is bounded from above (by an upper bound to be specified later). We show that U_n occurs w.h.p.. Next, we use Lemma 12 to bound from below the probability that $F_k(n, rn) \wedge Q(n, sn)$ becomes unsatisfiable given that $F_k(n, rn)$ has few solutions; we use this to show that $\Pr(\neg E_n \mid U_n)$ converges to 1 as $n \to \infty$. Finally, we combine these results to prove that $\Pr(\neg E_n)$ converges to 1.

Since $2^s > \lim_{i \to \infty} B_i$, there is some integer $i \geq 1$ such that $2^s > B_i$. Define the event U_n as the event when $\#F_k(n, rn) \leq B_i^n$. Then U_n occurs w.h.p. by hypothesis.

Next, we show that $\Pr(\neg E_n \mid U_n)$ converges to 1. Choose $\delta > 0$ and $N > 0$ such that $2^{s - \delta - 1/N} > B_i$. As in Lemma 8 we are concerned only with the behavior of $\Pr(\neg E_n \mid U_n)$ in the limit so we can restrict our attention only to large enough n. In particular, consider $n > N$. Then we get that $2^{\lceil sn \rceil - \delta n - 1} > 2^{sn - \delta n - n/N} > B_i^n$. Let $\alpha = \delta n + 1$, so that $2^{\lceil sn \rceil - \alpha} \geq B_i^n$. Then Lemma 12 says that $\Pr(\neg E_n \mid U_n) \geq$

$1 - 2^{-\delta n - 1}$. Since $1 - 2^{-\delta n-1}$ converges to 1 as $n \to \infty$, $\mathsf{Pr}(\neg E_n \mid U_n)$ must also converge to 1.

Thus both $\mathsf{Pr}(\neg E_n \mid U_n)$ and $\mathsf{Pr}(U_n)$ converge to 1 as $n \to \infty$. Since $\mathsf{Pr}(\neg E_n \cap U_n) = \mathsf{Pr}(\neg E_n \mid U_n) \cdot \mathsf{Pr}(U_n)$, this implies that $\mathsf{Pr}(\neg E_n \cap U_n)$ also converges to 1. Since $\mathsf{Pr}(\neg E_n \cap U_n) \leq \mathsf{Pr}(\neg E_n) \leq 1$, this implies that $\mathsf{Pr}(\neg E_n)$ converges to 1. □

Finally, it remains only to use Lemma 13 to obtain bounds on the k-CNF-XOR phase-transition. The tight upper bound on $\#F_k(n, rn)$ from Lemma 3.(c) corresponds to a tight upper bound on the location of the phase-transition.

Lemma 14. *Let $k \geq 2$, and let α_k, $\mathcal{C}_k$, and $\phi_k(r)$ be as defined in Lemma 3. Then for all $r \in [0, \alpha_k) \backslash \mathcal{C}_k$ and $s > \phi_k(r)$, $\psi_k(n, rn, sn)$ is unsatisfiable w.h.p..*

Proof. Let $B_i = 2^{\phi_k(r)+1/i}$. By Lemma 3.(c), $B_i^n \geq \#F_k(n, rn)$ w.h.p. for all $i \geq 1$. Furthermore, $\lim_{i\to\infty} B_i = 2^{\phi_k(r)}$ and so $s > \log_2(\lim_{i\to\infty} B_i)$. Thus $\psi_k(n, rn, sn)$ is unsatisfiable w.h.p. by Lemma 13. □

The weaker upper bound on $\#F_k(n, rn)$ from Lemma 11 corresponds to a weaker upper bound on the phase-transition.

Lemma 15. *Let $k \geq 2$, $s \geq 0$, and $r \geq 0$. If $s > 1 + r \log_2(1 - 2^{-k})$, then $\psi_k(n, rn, sn)$ is unsatisfiable w.h.p..*

Proof. Let $B_i = ((1 + 1/i) \cdot 2(1 - 2^{-k})^r)$. By Lemma 11, $B_i^n \geq \#F_k(n, rn)$ w.h.p. for all $i \geq 1$. Furthermore, $\lim_{i\to\infty} B_i = 2(1 - 2^{-k})^r$ and so $s > \log_2(\lim_{i\to\infty} B_i)$. Thus $\psi_k(n, rn, sn)$ is unsatisfiable w.h.p. by Lemma 13. □

5 Extending the Phase-Transition Region

Section 4 proved that a phase-transition exists for k-CNF-XOR formulas when the k-clause density is small. Our empirical observations in Section 3 suggest that a phase-transition exists for higher k-clause densities as well. In this section, we conjecture two possible extensions to our theoretical results.

The first extension follows from Theorem 1, which implies that $s = \phi_k(r)$ gives the location of the phase-transition for small k-clause densities. It is thus natural to conjecture that $\phi_k(r)$ gives the location of the k-CNF-XOR phase-transition for all (except perhaps countably many) $r > 0$. This would follow from a conjecture of [Abbe and Montanari, 2014].

The second extension follows from the experimental results in Section 3, which suggest that the location of the phase-transition follows a linear trade-off between k-clauses and XOR-clauses. This leads to the following conjecture:

Conjecture 2 (k-CNF-XOR Linear Phase-Transition Conjecture)**.** *Let $k \geq 2$. Then there exists a slope $L_k < 0$ and a constant $\alpha_k^* > 0$ such that for all $r \in [0, \alpha_k^*)$ and $s \geq 0$:*

(a). If $s < rL_k + 1$, then w.h.p. $\psi_k(n, rn, sn)$ is satisfiable.

(b). If $s > rL_k + 1$, then w.h.p. $\psi_k(n, rn, sn)$ is unsatisfiable.

Theorem 2 bounds the possible values for L_k. Moreover, if the Linear k-CNF-XOR Phase-Transition Conjecture holds, then Theorem 1 implies that $\phi_k(r)$ is linear for all $r < \alpha_k$ and $r < \alpha_k^*$. Explicit computations of $\phi_k(r)$ (or sufficiently tight bounds) would resolve this conjecture.

Note that this conjecture does not necessarily describe the entire k-CNF-XOR phase-transition; a phase-transition may exist when $r > \alpha_k^*$ as well. The experimental results in Section 3 for $k = 2$ and $k = 3$ suggest that the location of the phase-transition may "kink" and become non-linear for large enough k-clause densities. We leave the full characterization of the k-CNF-XOR phase-transition for future work, noting that a full characterization would resolve the Satisfiability Phase-Transition Conjecture.

6 Conclusion

We presented the first study of phase-transition phenomenon in the satisfiability of k-CNF-XOR random formulas. We showed that the free-entropy density $\phi_k(r)$ of k-CNF formulas gives the location of the phase-transition for k-CNF-XOR formulas when the density of the k-CNF clauses is small. We conjectured in the k-CNF-XOR Linear Phase-Transition Conjecture that this phase-transition is linear. We leave further analysis and proof of this conjecture for future work.

Pittel and Sorkin [Pittel and Sorkin, 2015] recently identified the location of the phase-transition for random ℓ-XOR formulas, where each clause contains exactly ℓ literals. This suggests that a phase-transition may also exist in formulas that mix k-CNF clauses together with ℓ-XOR clauses.

In this work we did not explore the runtime of SAT solvers over the space of k-CNF-XOR formulas. Historically, other phase-transition phenomena have been closely connected empirically to solver runtime. Developing this connection in the case of k-CNF-XOR formulas is an exciting direction for future research and may lead to practical improvements to hashing-based sampling and counting algorithms.

Acknowledgments

The authors would like to thank Dimitris Achlioptas for helpful discussions in the early stages of this project.

Work supported in part by NSF grants CCF-1319459 and IIS-1527668, by NSF Expeditions in Computing project "ExCAPE: Expeditions in Computer Augmented Program Engineering", by BSF grant 9800096, by the Ken Kennedy Institute Computer Science & Engineering Enhancement Fellowship funded by the Rice Oil & Gas HPC Conference, and Data Analysis and Visualization Cyberinfrastructure funded by NSF under grant OCI-0959097. Part of this work was done during a visit to the Israeli Institute for Advanced Studies.

References

[Abbe and Montanari, 2014] E. Abbe and A. Montanari. On the concentration of the number of solutions of random satisfiability formulas. *Random Structures & Algorithms*, 45(3):362–382, 2014.

[Achlioptas and Coja-Oghlan, 2008] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proc. of FoCS*, pages 793–802, 2008.

[Achlioptas *et al.*, 2011] D. Achlioptas, A. Coja-Oghlan, and F. Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. *Random Structures & Algorithms*, 38(3):251–268, 2011.

[Achlioptas, 2009] D. Achlioptas. Random satisfiability. In *Handbook of Satisfiability*, pages 245–270. IOS Press, 2009.

[Apt, 2003] K. Apt. *Principles of Constraint Programming*. Cambridge Univ. Press, 2003.

[Belov *et al.*, 2014] A. Belov, D. Diepold, M. J. Heule, and M. Järvisalo. Sat competition 2014. 2014.

[Chakraborty *et al.*, 2013a] S. Chakraborty, K. S. Meel, and M. Y. Vardi. A scalable and nearly uniform generator of SAT witnesses. In *Proc. of CAV*, pages 608–623, 2013.

[Chakraborty *et al.*, 2013b] S. Chakraborty, K. S. Meel, and M. Y. Vardi. A scalable approximate model counter. In *Proc. of CP*, pages 200–216, 2013.

[Chakraborty *et al.*, 2014a] S. Chakraborty, D. J. Fremont, K. S. Meel, S. A. Seshia, and M. Y. Vardi. Distribution-aware sampling and weighted model counting for SAT. In *Proc. of AAAI*, pages 1722–1730, 2014.

[Chakraborty *et al.*, 2014b] S. Chakraborty, K. S. Meel, and M. Y. Vardi. Balancing scalability and uniformity in SAT witness generator. In *Proc. of DAC*, pages 1–6, 2014.

[Chakraborty *et al.*, 2016] S. Chakraborty, K. S. Meel, and M. Y. Vardi. Algorithmic improvements in approximate counting for probabilistic inference: From linear to logarithmic SAT calls. In *Proc. of IJCAI*, 2016.

[Cheeseman *et al.*, 1991] P. Cheeseman, B. Kanefsky, and W. M. Taylor. Where the really hard problems are. In *Proc. of IJCAI*, pages 331–340, 1991.

[Chvátal and Reed, 1992] V. Chvátal and B. Reed. Mick gets some (the odds are on his side). In *Proc. of FoCS*, pages 620–627, 1992.

[Coarfa *et al.*, 2003] C. Coarfa, D. D. Demopoulos, A. S. M. Aguirre, D. Subramanian, and M. Y. Vardi. Random 3-sat: The plot thickens. *Constraints*, 8(3):243–261, 2003.

[Coja-Oghlan and Panagiotou, 2013] A. Coja-Oghlan and K. Panagiotou. Going after the k-sat threshold. In *Proc. of SToC*, pages 705–714, 2013.

[Coja-Oghlan and Reichman, 2013] A. Coja-Oghlan and D. Reichman. Sharp thresholds and the partition function. In *Journal of Physics: Conference Series*, volume 473, page 012015, 2013.

[Crawford and Auton, 1993] J. M. Crawford and L. D. Auton. Experimental results on the crossover point in satisfiability problems. In *Proc. of AAAI*, pages 21–27, 1993.

[Creignou and Daudé, 1999] N. Creignou and H. Daudé. Satisfiability threshold for random xor-cnf formulas. *Discrete Applied Mathematics*, 9697:41 – 53, 1999.

[Creignou and Daudé, 2003] N. Creignou and H. Daudé. Smooth and sharp thresholds for random k-xor-cnf satisfiability. *RAIRO:ITA*, 37(2):127–147, 2003.

[Dechter and Meiri, 1994] R. Dechter and I. Meiri. Experimental evaluation of preprocessing algorithms for constraint satisfaction problems. *Artificial Intelligence*, 68:211–241, 1994.

[Dechter, 2003] R. Dechter. *Constraint Processing*. Morgan Kaufmman, 2003.

[Ding *et al.*, 2015] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. In *Proc. of SToC*, pages 59–68. ACM, 2015.

[Dubois and Mandler, 2002] O. Dubois and J. Mandler. The 3-xorsat threshold. *Comptes Rendus Mathematique*, 335(11):963 – 966, 2002.

[Goerdt, 1996] A. Goerdt. A threshold for unsatisfiability. *Journal of Computer and System Sciences*, 53(3):469 – 486, 1996.

[Gogioso, 2014] S. Gogioso. Aspects of statistical physics in computational complexity. *arXiv:1405.3558*, 2014.

[Gomes *et al.*, 2006] C. P. Gomes, A. Sabharwal, and B. Selman. Model counting: A new strategy for obtaining good bounds. In *Proc. of AAAI*, volume 21, pages 54–61, 2006.

[Gomes *et al.*, 2007] C. P. Gomes, A. Sabharwal, and B. Selman. Near-uniform sampling of combinatorial spaces using XOR constraints. In *Proc. of NIPS*, pages 670–676, 2007.

[Ivrii *et al.*, 2015] A. Ivrii, S. Malik, K. Meel, and M. Vardi. On computing minimal independent support and its applications to sampling and counting. *Constraints*, pages 1–18, 2015.

[Kirkpatrick and Selman, 1994] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random boolean expressions. *Science*, 264(5163):1297–1301, 1994.

[Meel *et al.*, 2016] K. S. Meel, M. Vardi, S. Chakraborty, D. J. Fremont, S. A. Seshia, D. Fried, A. Ivrii, and S. Malik. Constrained sampling and counting: Universal hashing meets sat solving. In *Proc. of Beyond NP Workshop*, 2016.

[Mitchell *et al.*, 1992] D. Mitchell, B. Selman, and H. Levesque. Hard and easy distributions of sat problems. In *Proc. of AAAI*, pages 459–465, 1992.

[Mu and Hoos, 2015] Z. Mu and H. H. Hoos. On the empirical time complexity of random 3-sat at the phase transition. In *Proc. of IJCAI*, pages 367–373, 2015.

[Pittel and Sorkin, 2015] B. Pittel and G. B. Sorkin. The satisfiability threshold for k-xorsat. *Combinatorics, Probability and Computing*, FirstView:1–33, 10 2015.

[Soos *et al.*, 2009] M. Soos, K. Nohl, and C. Castelluccia. Extending SAT Solvers to Cryptographic Problems. In *Proc. of SAT*. Springer-Verlag, 2009.

[Wilson, 2000] D. B. Wilson. Random 2-sat data. http://dbwilson.com/2sat-data/, 2000.