Phase transition for random quantified Boolean Formulas

N. Creignou\textsuperscript{1} H. Daudé\textsuperscript{2} Uwe Egly\textsuperscript{3} R. Rossignol\textsuperscript{4}

\textsuperscript{1}LIF, Marseille
\textsuperscript{2}LATP, Marseille
\textsuperscript{3}TU Wien
\textsuperscript{4}Orsay

Alea, 2008
Outline

1. Threshold phenomena for SAT problems
2. Quantified satisfiability problems
   - Probabilistic model: (a,e)-QSAT
3. Threshold phenomena for (a,e)-QXOR-SAT
   - The case $e \geq 3$
   - The case $e = 2$
4. Threshold phenomenon for (1,2)-QSAT
   - Complexity of (1,2)-QSAT
   - Experimental results
   - Representation of (1,2)-QCNF-formulas as labeled digraphs
   - Lower and upper bounds for the threshold
Threshold phenomena for SAT problems

Quantified satisfiability problems

Probabilistic model : (a,e)-QSAT

Threshold phenomena for (a,e)-QXOR-SAT

The case \( e \geq 3 \)

The case \( e = 2 \)

Threshold phenomenon for (1,2)-QSAT

Complexity of (1,2)-QSAT

Experimental results

Representation of (1,2)-QCNF-formulas as labeled digraphs

Lower and upper bounds for the threshold
A threshold phenomenon for 3-SAT

\[ \Pr(SAT_{n,L}) = \text{probability that a 3-CNF formula over } n \text{ variables with } L = c \cdot n \text{ clauses is satisfiable.} \]

\[ \Pr(SAT_{n,c \cdot n}) \to 1 \text{ for } c \leq 3.52 \]

(Kaporis, Kirousis, Lalas, 2003)

\[ \Pr(SAT_{n,c \cdot n}) \to 0 \text{ for } c \geq 4.506 \]

(Dubois, Boufkhad, Mandler, 2000)

The critical ratio is estimated at around 4.25
Given a constraint satisfaction problem, depending on the size of the scaling window the transition SAT/UNSAT is either *sharp* or *coarse*.

- The transition for 3-SAT is sharp (Friedgut, 1998).
- The transition for 2-SAT is sharp, the critical ratio is 1 (Chvatal, Reed, Goerdt, 1992).
An example of a sharp transition: 3-XOR-SAT
An example of a coarse transition: 2-XOR-SAT

\[ f_1(x) = \exp \left( \frac{x}{2} \right) \left( 1 - 2x \right)^0.25 \]
Random instances are useful to evaluate the performance of SAT solvers: hard instances are at the threshold.
Outline

1. Threshold phenomena for SAT problems
2. Quantified satisfiability problems
   - Probabilistic model: (a,e)-QSAT
3. Threshold phenomena for (a,e)-QXOR-SAT
   - The case $e \geq 3$
   - The case $e = 2$
4. Threshold phenomenon for (1,2)-QSAT
   - Complexity of (1,2)-QSAT
   - Experimental results
   - Representation of (1,2)-QCNF-formulas as labeled digraphs
   - Lower and upper bounds for the threshold
Input: $\Phi$ a closed formula of the form $Q_1 x_1 Q_2 x_2 \ldots Q_n x_n \varphi$, where $Q_1, \ldots, Q_n$ are arbitrary quantifiers and $\varphi$ is a CNF-formula.

Question: Is $\Phi$ true?

- QSAT is PSPACE-complete (ranges over the full polynomial hierarchy depending on the number of quantifiers alternation).
- QSAT allows the modelling of various problems (games, model checking, verification, etc.).
- QSAT is a monotone property.
Is there a phase transition for QSAT?
What is a "good" probabilistic model?
Is there an easy-hard-easy pattern for random instances?

Cadoli et al. 97, Gent and Walsh 99, Chen and Interian 05
(a,e)-QSAT

We will first restrict our attention to formulas with only one alternation of quantifiers

An \textit{(a,e)-QCNF-formula} is a closed quantified formula of the following type

$$\forall X \exists Y \varphi(X, Y),$$

- \(X\) and \(Y\) denote distinct set of variables,
- \(\varphi(X, Y)\) is an \((a + e)\)-CNF-formula such that each clause contains exactly \(a\) variables from \(X\) and exactly \(e\) variables from \(Y\).
- (a,e)-QSAT the property for an (a,e)-QCNF-formula of being true.
Threshold phenomena for SAT problems
Quantified satisfiability problems
Threshold phenomena for (a,e)-QXOR-SAT
Threshold phenomenon for (1,2)-QSAT

Probabilistic model : (a,e)-QSAT

(a,e)-QCNF((m,n),L)-formulas

- \( m \) the number of universal variables, \( \{x_1, \ldots, x_m\} \)
- \( n \) for the number of existential variables, \( \{y_1, \ldots, y_n\} \)
- Random formulas \( \forall X \exists Y \varphi(X, Y) \) obtained by choosing uniformly independently and with replacement \( L \) clauses among the \( N = 2^{a+e} \binom{m}{a} \binom{n}{e} \) the possible clauses.
We are interested in the probability that a randomly chosen 
\((a, e)\)-QCNF\((m, n, L)\)-formula is true.

\[
\Pr_{(m,n,L)}((a,e)\text{-QSAT})
\]

Non-quantified case: \(e\)-SAT, i.e., \(a=0\),

\[
\Pr_{n,L}(e\text{-SAT})
\]
(a,e)-QXOR-SAT a natural variant to initiate an investigation

- QXOR-SAT is in P (linear algebra framework)
- Well studied in the non-quantified version: random graph tools.
Outline

1. Threshold phenomena for SAT problems
2. Quantified satisfiability problems
   - Probabilistic model: (a,e)-QSAT
3. Threshold phenomena for (a,e)-QXOR-SAT
   - The case $e \geq 3$
   - The case $e = 2$
4. Threshold phenomenon for (1,2)-QSAT
   - Complexity of (1,2)-QSAT
   - Experimental results
   - Representation of (1,2)-QCNF-formulas as labeled digraphs
   - Lower and upper bounds for the threshold
Threshold phenomena for SAT problems
Quantified satisfiability problems
Threshold phenomena for (a,e)-QXOR-SAT
Threshold phenomenon for (1,2)-QSAT

The case $e \geq 3$

The case $e = 2$

A sharp threshold

\[
\Pr_{n,L}(e\text{-Max-rank}) \leq \Pr_{(m,n,L)}((a,e)\text{-QXOR-SAT}) \leq \Pr_{n,L}(e\text{-XOR-SAT})
\]

\[
\frac{2\Pr_{n,L}(e\text{-XOR-SAT}) - 1}{2} \leq \Pr_{n,L}((a,e)\text{-QXOR-SAT}) \leq \Pr_{n,L}(e\text{-XOR-SAT})
\]

e-XOR-SAT exhibits a sharp threshold when $L$ is $\Theta(n)$ and so does (a,e)-QXOR-SAT, with a critical value $c_3 \approx 0.918$ for $e = 3$. 
A sharp threshold

\[
\Pr_{n,L}(e-\text{Max-rank}) \leq \Pr_{(m,n,L)}((a,e)-\text{QXOR-SAT}) \leq \Pr_{n,L}(e-\text{XOR-SAT})
\]

\[
\frac{2\Pr_{n,L}(e-\text{XOR-SAT}) - 1}{2} \leq \Pr_{n,L}((a,e)-\text{QXOR-SAT}) \leq \Pr_{n,L}(e-\text{XOR-SAT})
\]

e-XOR-SAT exhibits a sharp threshold when \( L \) is \( \Theta(n) \) and so does \((a,e)-\text{QXOR-SAT}\), with a critical value \( c_3 \approx 0.918 \) for \( e = 3 \).
A sharp threshold

\[ \Pr_{n,L}(e\text{-Max-rank}) \leq \Pr_{(m,n,L)}((a,e)\text{-QXOR-SAT}) \leq \Pr_{n,L}(e\text{-XOR-SAT}) \]

\[ \frac{2\Pr_{n,L}(e\text{-XOR-SAT}) - 1}{2} \leq \Pr_{n,L}((a,e)\text{-QXOR-SAT}) \leq \Pr_{n,L}(e\text{-XOR-SAT}) \]

e-XOR-SAT exhibits a sharp threshold when \( L \) is \( \Theta(n) \) and so does \((a,e)\)-QXOR-SAT, with a critical value \( c_3 \approx 0.918 \) for \( e = 3 \).
A sharp threshold

\[ \Pr_{n,L}(e-\text{Max-rank}) \leq \Pr_{(m,n,L)}((a,e)-\text{QXOR-SAT}) \leq \Pr_{n,L}(e-\text{XOR-SAT}) \]

\[ \frac{2\Pr_{n,L}(e-\text{XOR-SAT}) - 1}{2} \leq \Pr_{n,L}((a,e)-\text{QXOR-SAT}) \leq \Pr_{n,L}(e-\text{XOR-SAT}) \]

e-XOR-SAT exhibits a sharp threshold when \( L \) is \( \Theta(n) \) and so does \((a,e)-\text{QXOR-SAT}\), with a critical value \( c_3 \approx 0.918 \) for \( e = 3 \).
Representation of (a,2)-QXOR-formulas as labeled graphs

Let $X = \{x_1, x_2, x_3\}$ and let $Y = \{y_1, \ldots, y_7\}$. The formula

$\forall X \exists Y \ \varphi(X, Y)$

with $\varphi(X, Y)$ being a conjunction of the following equations

\begin{align*}
y_1 \oplus y_2 &= x_1 & y_1 \oplus y_7 &= x_2 \\
y_2 \oplus y_3 &= x_3 & y_2 \oplus y_6 &= x_2 \oplus 1 \\
y_3 \oplus y_4 &= x_2 \oplus 1 & y_3 \oplus y_5 &= x_3 \\
y_4 \oplus y_5 &= x_3 \oplus 1 & y_6 \oplus y_7 &= x_1 \oplus 1
\end{align*}
Bad cycles

Cycle is **bad** if it has a nonzero weight, and **good** otherwise.
Bad cycles and \((a,2)\)-QXOR-SAT

\[
\Pr((a,2)\text{-QXOR-SAT}) = \Pr(G_a(s) \text{ has no bad cycle})
\]
Threshold phenomena for SAT problems
Quantified satisfiability problems
Threshold phenomena for (a,e)-QXOR-SAT
Threshold phenomenon for (1,2)-QSAT

The case $e \geq 3$

The case $e = 2$

The distribution function for $(a,2)$-QXOR-SAT

$(a,2)$-QXOR-SAT has a coarse threshold.

$$\Pr_{(m,n,cn)}((a,2)\text{-QXOR-SAT}) \xrightarrow{n \to +\infty} H_m(c).$$

The distribution function $H_m$ depends on the number of universal variables.
The distribution functions $H_0$, $H_a$ and $H_\infty$

\begin{align*}
H_a(x) &= e^x \sqrt{1 - 2x} (1 - 4x^2)^{-1/8} \\
H_0(x) &= e^{x/2} (1 - 2x)^{0.25} \\
H_\infty(x) &= e^x \sqrt{1 - 2x}
\end{align*}
Threshold phenomena for SAT problems
Quantified satisfiability problems
Threshold phenomena for (a,e)-QXOR-SAT
Threshold phenomenon for (1,2)-QSAT

1. Threshold phenomena for SAT problems
2. Quantified satisfiability problems
   - Probabilistic model: (a,e)-QSAT
3. Threshold phenomena for (a,e)-QXOR-SAT
   - The case $e \geq 3$
   - The case $e = 2$
4. Threshold phenomenon for (1,2)-QSAT
   - Complexity of (1,2)-QSAT
   - Experimental results
   - Representation of (1,2)-QCNF-formulas as labeled digraphs
   - Lower and upper bounds for the threshold
(1,2)-QSAT

- $m$ the number of universal variables, $\{x_1, \ldots, x_m\}$
- $n$ for the number of existential variables, $\{y_1, \ldots, y_n\}$
- Random formulas $\forall X \exists Y \varphi(X, Y)$ obtained by choosing uniformly independently and with replacement $L = cn$ clauses among the $N = 2^3 \binom{m}{1} \binom{n}{2}$ possible 3-clauses having 1 universal and 2 existential literals.

We are interested in the probability that such a randomly chosen formula is true.

$\mathbb{P}_{m,c}(n)$
If $m$ is constant, $(1,2)$-QSAT is solvable in linear time.

If $m = \alpha \lceil \log n \rceil$, $(1,2)$-QSAT is solvable in polynomial time.

If $m = n$, then $(1,2)$-QSAT is $\text{coNP}$-complete.
When $m(n) = n$: the threshold occurs at $c = 1$
When $m(n) = 2$: the threshold occurs at $c = 2$
Intermediate regime when $m = \alpha \lceil \log n \rceil$
Let $\phi: \forall x_1 \exists y_1 y_2 (x_1 \lor y_1 \lor y_2) \land (\overline{x_1} \lor y_1 \lor \overline{y_2})$.

The quantified formula is satisfiable if and only if there is no pure path $y \rightsquigarrow \overline{y} \rightsquigarrow y$ for any existential variable $y$. 

Threshold phenomena for (a,e)-QXOR-SA T

Complexity of (1,2)-QSAT

Labeled digraphs

Representation of (1,2)-QCNF-formulas as labeled digraphs
A necessary condition for unsatisfiability

Every unsatisfiable (1,2)-QCNF formula contains a pure bicycle.

Let $B$ be the number of pure bicycles in a (1,2)-QCNF formula.

$$1 - \mathbb{P}_{m,c}(n) \leq \Pr(B \geq 1) \leq \mathbb{E}(B).$$
A sufficient condition for unsatisfiability

Every (1,2)-QCNF formula that contains some simple snake is unsatisfiable.

Let $X$ be the number of simple snakes of size $s + 1 = 2t$ in a (1,2)-QCNF formula.

$$1 - \mathbb{P}_{m,c}(n) \geq \Pr(X \geq 1) \geq \frac{(\mathbb{E}(X))^2}{\mathbb{E}(X^2)}$$
Mean of the number of pure bicycles

Let $p$ be such that $N \cdot p \sim c \cdot n$.

$$
\mathbb{E}(B) = \sum_{s=2}^{n} (n)_s 2^s [(2s)^2 - 1] c(m, s + 1) p^{s+1} ,
$$

where

$$
c(m, s + 1) = \sum_{k=1}^{\min(m, s + 1)} \binom{m}{k} \cdot 2^k \cdot S(s + 1, k) \cdot k!
$$

with $S(m, k)$ denoting the Stirling number of the second kind.
A lower bound for the threshold

**Theorem**

When $1 < c < 2$, and $m = \lceil \alpha \ln n \rceil$ with $\alpha > \frac{1}{\ln(2)}$,

$$\mathbb{E}(B) \leq C (\ln n)^{9/2} \cdot n^{\alpha H(c) - 1} + o(1)$$

where $C$ is a finite constant depending only on $\alpha$ and $c$, and

$$H(c) = \ln(c) + \left(\frac{2}{c} - 1\right) \ln(2 - c).$$

Let $a(\alpha)$ be the solution of the equation $\alpha \cdot H(c) = 1$, then for $c < a(\alpha)$ the above result shows that $\mathbb{E}(B) = o(1)$. 
Main result

Let $m = \lceil \alpha \ln n \rceil$ where $\alpha > 0$. There exist $1 < a(\alpha) \leq b(\alpha) \leq 2$ such that the following holds:

- if $c < a(\alpha)$, then $\mathbb{P}_{m,c}(n) \xrightarrow{n \to +\infty} 1$,
- if $c > b(\alpha)$, then $\mathbb{P}_{m,c}(n) \xrightarrow{n \to +\infty} 0$.

Moreover:

1. if $\alpha \leq \frac{1}{\ln 2}$, then $a(\alpha) = b(\alpha) = 2$,
2. if $\frac{1}{\ln 2} < \alpha \leq \frac{2}{\ln 2 - 1/2}$, then $a(\alpha) < b(\alpha) = 2$ and $a$ is strictly decreasing,
3. if $\alpha > \frac{2}{\ln 2 - 1/2}$, then $a(\alpha) < b(\alpha) < 2$, $a$ and $b$ are strictly decreasing and $\lim_{\alpha \to +\infty} a(\alpha) = \lim_{\alpha \to +\infty} b(\alpha) = 1$. 
The lower and upper bounds

\[ F_{IG} = a(\alpha) \quad \text{and} \quad b(\alpha) \]

**Fig.:** \( a(\alpha) \) and \( b(\alpha) \)
Conclusion

- Validation of the probabilistic models proposed for the study of the QSAT transition.
- For (a,e)-QXOR-SAT, introduction of quantifiers has an effect at the level of the distribution function.
- For (1,2)-QSAT, introduction of quantifiers influences the location of the threshold, the number of universal variables plays a crucial role.