Maxwell-Boltzmann and Bose-Einstein Distributions for the SAT Problem

Claudio Angione[1], Annalisa Occhipinti[2], and Giuseppe Nicosia[2]

[1]Computer Laboratory, University of Cambridge
[2]Department of Mathematics and Computer Science, University of Catania

Abstract. Recent studies in theoretical computer science have exploited new algorithms and methodologies based on statistical physics for investigating the structure and the properties of the Satisfiability problem. We propose a characterization of the SAT problem as a physical system, using both quantum and classical statistical-physical models. We associate a graph to a SAT instance and we prove that a Bose-Einstein condensation occurs in the instance with higher probability if the quantum distribution is adopted in the generation of the graph. Our method allows a comprehensive analysis of the SAT problem based on a new definition of entropy of an instance, without requiring the computation of its truth assignments. Finally, we develop four new SAT solvers based on quantum and classical statistical distributions, and we test them on both random and real-life SAT instances. As a result, our method can be readily used to develop fast and efficient solvers of large-scale computational problems, namely AI planning, model checking, and hardware and software verification.

1 Introduction

Nowadays, the SAT problem is regularly used for solving large-scale computational problems, such as AI planning, protein structure prediction, haplotype inference, circuit-level prediction of crosstalk noise, model checking, and hardware and software verification [1]. As a result, it has received significant research attention [2, 3], and numerous solver algorithms have been proposed and improved (e.g., by analyzing the structure of each instance [4]). Since evaluating all these techniques requires the generation of hard satisfiable instances, several methods for generating random instances have been presented to test the solvers performance [5, 6].

Although many SAT solvers are based on local search techniques [7, 8] and on Conflict Directed Clause Learning algorithms (CDCL) [9], new methods for the SAT problem analysis consist of translating each instance into a graph (e.g., by using the planar graph approach [10] or the Bose-Einstein distribution and the S2G algorithm [11]).

Over the last few years, several research fields have witnessed a remarkable expansion due to the collaboration between physicists and computer scientists [12–14]. Therefore, studying the k-SAT problem through models studied in statistical physics represents an unprecedented opportunity to find a new characterization for the satisfiability problem.

In this work, we prove that the S2G algorithm is unlikely to generate Bose-Einstein condensed graphs by using the Maxwell-Boltzmann distribution (i.e., classical physics

$G = (V, E)$	k-SAT	Statistical physics
node	clause	degeneration state of the energy level of the node
edge	link between two clauses	one particle for each degeneration state involved
node weight	fitness of a clause	value of the energy level
edge weight	probability of being established	weight on particles
out-degree update	parameter θ	temperature of the system

Table 1. *Dictionary translating the graph (left) into the k-SAT problem (centre) and statistical physics language (right).*

instead of quantum physics). To this aim, we propose a S2G algorithm based on the Maxwell-Boltzmann distribution and produce its phase diagrams, which we compare to the Bose-Einstein case. We also define the entropy of a k-SAT instance and we introduce its temperature, in order to translate each instance into a complete physical system (summarized in Table 1). Finally, we propose four SAT solvers inspired by CHAINSAT [15] and augmented with classical and quantum statistical-physical approaches to classify an instance. The analysis of the physical parameters of our algorithm allows us to establish a priori which of the four SAT solvers is the best for each instance.

2 A characterization of SAT through quantum physics

By translating a SAT formula into a graph, the S2G algorithm [11] shows evidence of a process equivalent to the Bose-Einstein condensation in quantum physics. A vertex v_i is a clause C_i of the k-SAT formula $F = C_1 \wedge C_2 \wedge ... \wedge C_m$. The graph is built by defining global and local fitness functions (named f^G and f^L respectively) for evaluating literals and clauses. Then, a metric is defined to compute how many literals are not in common between two clauses. This metric is proved to be related to the Hamming distance [16].

2.1 The Bose-Einstein distribution

Let us consider an isolated system of N identical and indistinguishable bosons confined to a space of volume V and sharing a given energy E. Let us assume that these bosons can be distributed into a set of energy levels, where each level E_i is characterized by an *energy* ϵ_i and a *degeneration* g_i, representing the number of different physical states that can be found at that level. The N identical and indistinguishable particles are distributed among the energy levels, and each level E_i contains n_i particles, to be accommodated among its g_i quantum states. One can readily check that n_i particles may be put on the level E_i (consisting of g_i states) in $[n_i + (g_i - 1)]!$ different ways. Since bosons are indistinguishable and the physical states are equivalent, the number of possible assignments of n_i bosons on E_i is $w_i = \frac{(n_i+g_i-1)!}{n_i!(g_i-1)!} = \binom{n_i+g_i-1}{n_i}$. By iterating for all the energy levels E_i, a distribution $\{n_i\}$ (i.e., a distribution with n_i particles on the level E_i, $\forall i$) can be obtained in $W = \prod_i w_i$ different ways. Specifically, w_i is the number of distinct microstates associated with the ith level of the spectrum, while W is the number of distinct microstates associated with the whole distribution set $\{n_i\}$.

The distribution corresponding to the statistical equilibrium is the most probable one, thus it is the one that may be reached in the largest number of possible ways.

Hence, we compute the maximum W subject to the conservation of the number of particles $\sum_i n_i = N$ and to the preservation of the energy of the system $\sum_i \epsilon_i n_i = E$. Using the method of Lagrange's multipliers (applied to $\log W$), we have the following definition of *Bose-Einstein distribution*:

$$n_i = \frac{g_i}{e^{\alpha+\beta\epsilon_i} - 1}, \tag{1}$$

where $\alpha = -\frac{\mu_C}{k_B T}$ and $\beta = \frac{1}{k_B T}$ are inversely proportional (by means of Boltzmann's constant k_B) to the absolute temperature T of the system at the equilibrium, and μ_C represents the chemical potential [17].

Given an ideal Bose-Einstein gas in equilibrium below its transition temperature, the *Bose-Einstein Condensation* (BEC) is the property that a finite fraction of particles occupies the lowest energy level. Below a critical temperature T_{BEC} near to 0 K, all the particles become absolutely identical, with no possible measurement that can tell them apart. The gas shows a very unusual state of aggregation of particles, called *Bose-Einstein condensate*, also referred to as "*the fifth state of matter*". According to Penrose and Onsager [18], we can provide a criterion of BEC for an ideal gas in equilibrium: BEC $\iff$ $\frac{\langle n_0 \rangle}{N} = e^{O(1)}$, No BEC $\iff$ $\frac{\langle n_0 \rangle}{N} = o(1)$, where $\langle n_0 \rangle$ is the average number of particles that occupy the lowest energy level E_0.

2.2 Condensation phenomena in a SAT formula

The graph construction is thought of as a dynamical process. Given the graph $G = (V, E)$ at the $(i-1)$th iteration, the S2G algorithm adds a clause C_{t_i} to G as a node $v(C_{t_i})$, by estimating the probability of being connected to each node $v(C_{t_j})$ already in the graph as $\Pi_{t_j} = \frac{k_{t_j} \cdot f^L(C_{t_j})}{\sum_{\nu=1}^{|V|} k_{t_\nu} \cdot f^L(C_{t_\nu})}$, where $k_{t_j} = degree(v(C_{t_j}))$ is the connectivity of C_{t_j} (i.e., the number of links shared by $v(C_{t_j})$), and $f^L(C_{t_j})$ is the fitness of the clause C_{t_j}. This probability distribution ensures that a new vertex is likely linked to an existing one with high fitness value or/and high connectivity [19]. A node $v(C_{t_i})$ entering the graph is assigned the energy $\epsilon_{t_i} = -T \cdot \log f_r^L(C_{t_i})$, where $T = 1/\beta$, and β is a parameter used to model the temperature of the system.

In the mapping of S2G to quantum physics, every clause of the graph is associated with the degeneration state of the energy level of that clause. For each link established between two clauses, the S2G algorithm assigns a particle to each of the two degeneration states of the two clauses involved. The S2G-PA version of the algorithm also includes the concept of preferential attachment. In particular, at each iteration i, the node that joins the graph is forced to connect to a number of nodes between 1 and a fixed upper bound ρ. Furthermore, an outgoing link is rewarded $\theta < 1$, whereas an incoming link is rewarded 1, and therefore the graph is regarded as a directed graph. The condensation of the SAT formula over its fittest clause is eventually mapped to the emergence of a star-like topology in the graph. This phenomenon is associated with the BEC of bosons on the lowest energy level available.

3 The Maxwell-Boltzmann S2G

An approach to construct graph from k-SAT instances may also involve the classical physics, namely the Maxwell-Boltzmann (MB) distribution. Here we address the key question whether the S2G algorithm [11] could generate Bose-Einstein (BE) condensed graphs by using the MB distribution. In particular, we prove that the BEC in the satisfiability problem is more likely to occur if we use the BE distribution in the construction of the graph. To this end, we develop a new version of the S2G algorithm to generate graphs based on the MB distribution and analyze their phase diagrams, comparing the behavior in the MB and BE cases.

3.1 The Maxwell-Boltzmann distribution

From the classical-physics standpoint, particles must be regarded as distinguishable entities. The N particles are distributed among the energy levels, and each level E_i contains n_i particles. Note that the set $\{n_i\}$ of the occupation numbers does not fully describe the system, since particle are distinguishable. Therefore, we need to distinguish between which particles are occupying each energy level. In other words, we need to count not merely the possible sets $\{\{n_i\}_i\}$ of occupation numbers, but also the possible microstates in each set.

The number of possible sets of occupation numbers is counted by taking into account that particles are distinguishable. Therefore, the selection of which particle is accommodated on which energy level must be ordered. At this stage, we do not investigate the order inside energy levels because we will do so when counting the microstates. As a result, the distribution set n_i is obtainable in $\frac{N!}{\prod_i n_i!}$ different ways. Nevertheless, due to the Gibbs correction factor (see [17]) or, equivalently, using Stirling's approximations for the factorial $N!$, we obtain the correct counting $\frac{1}{\prod_i n_i!}$.

As regards microstates, since particle are distinguishable, when accommodating n_i particles on the energy level E_i, any of its n_i particles may be put into any of its g_i quantum states independently from one another, and all the resulting microstates are considered as distinct. In other words, each particle must be assigned one of the g_i quantum states, and this can be done in g_i different ways. By iterating this assignment independently from one particle to another, n_i particles may be put on the level E_i (consisting of g_i states) in $g_i^{n_i}$ different ways.

Combining the two results, we finally obtain $W = \prod_i \frac{g_i^{n_i}}{n_i!}$, which is the total number of possible distinct microstates of the system. Using the same method as in the BE case, this expression leads to the following definition of *Maxwell-Boltzmann distribution*:

$$n_i = \frac{g_i}{e^{\alpha+\beta\epsilon_i}}. \tag{2}$$

3.2 A S2G version based on classical physics

Since particles in the MB distribution are distinguishable, the idea underlying the MB version of the S2G algorithm is that also the links between clauses must be distinguishable. Therefore, when we run the S2G algorithm using the MB distribution, we

(a) BE case (b) MB case

Fig. 1. *Different graphs with the same* $FW = 1$. In (a), a BEC has taken place (i.e., one node has a huge fraction of edges and the remaining fraction is shared among all the other nodes), while (b) does not show any condensation.

(i) switch off the preferential attachment (PA) used in [11], and (ii) ensure that if a new node i will be linked to a node j with probability p_{ij}, then $degree(i) = degree(i) + p_{ij}$ and $degree(j) = degree(j) + p_{ij}$. This new pattern could provide graphs with many connected components. Hence, the original definition of Fraction Winner $FW = w_{links}/t_{links}$, where w_{links} = number of links shared by the winner, and t_{links} = number of links of the whole graph [11], leads to graphs with the same FW, but with different topological structure (Figure 1).

Our goal is to study whether the graph built by using the MB distribution can show cases of BEC. For this reason, we introduce a new definition of FW, which takes into account the presence of connected components. Let $G(V, E)$ be a graph with $|V|$ vertices and $|E|$ edges. We compute the number of connected components (CCs) by running the Depth First Search (DFS) Algorithm [20]. From the CCs, we obtain the number of nodes shared by these components (CNs), namely all the nodes of G with nonzero degree. Then, we define the Generalized Fraction Winner as $GFW = FW \cdot CNs/|V|$. It is important to underline that this definition is a generalization of the previous definition of FW. Indeed, in the BE construction, the PA ensures that CNs $= |V|$. In Figure 2 we show the comparison between the GFW in the BE and MB cases. Remarkably, the MB values are always lower than BE values, indicating that the probability of Bose-Einstein condensation is higher if the BE distribution is adopted in the S2G algorithm.

4 Experimental results

4.1 The entropy of a *k*–SAT instance

The aim of this section is to characterize the k-SAT problem as a physical system, in order to analyze the complexity of certifying the satisfiability of a random k-SAT instance and eventually to find a satisfying assignment. Let us introduce a definition of entropy for a k-SAT instance, based on a discretized form of the von Neumann entropy [22]:

$$VN_d(x, p, q) = -\frac{1}{n}\sum_{i=0}^{n} p(x_n)\log\left[q(x_n)p(x_n)\right], \tag{3}$$

where p and q are respectively a true distribution and an approximated distribution of a random discretized variable x with n possible values. In this work, we set $n =$ number

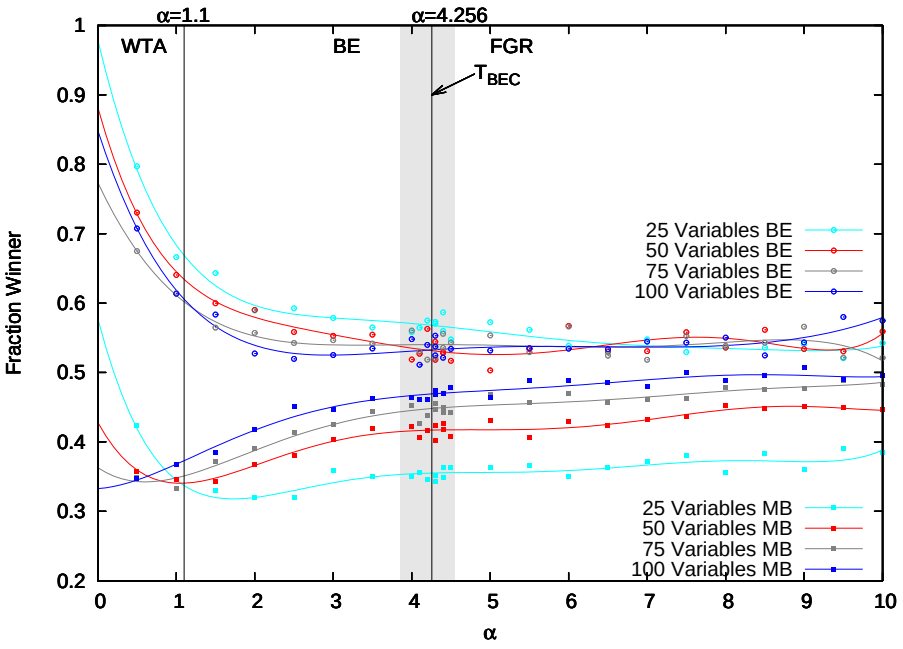

Fig. 2. *Generalized Fraction Winner BE-MB.* This plot shows the comparison between the GFW in the MB and BE cases. The MB plots always shows lower values than BE plots. Hence, BEC takes place under BE conditions with more probability than under MB conditions. The experimental data points have been fitted with a sixth-order regression curve using the least-squares approach. The gray box shows the region of the phase diagram where the phase transition between satisfiable and unsatisfiable instances of 3-SAT is located [21].

of energy levels, and we compute entropy every time that a new clause is added to the graph. The final entropy is computed after that the last clause of the k-SAT instance is added to the final graph. In a BE system, when a new node ν is added to the graph using the S2G algorithm, we define:

$$\begin{aligned} q_{BE}(\nu) &= \text{number of different physical states } g_i \text{ on the level} \\ &\quad E_i/\text{number of nodes already added to the graph;} \\ p_{BE}(\nu) &= \begin{cases} 1 \text{ if } \nu \text{ has been added to the level } E_i \\ 0 \text{ otherwise.} \end{cases} \end{aligned}$$

Conversely, in a MB system we set

$$\begin{aligned} q_{MB}(\nu) &= \text{number of different physical states } g_i \text{ on the level} \\ &\quad E_i/\text{number of nodes already added to the graph;} \\ p_{MB}(\nu) &= \text{probability } pi, \text{ computed by S2G, that a new node} \\ &\quad \text{is added to the level } E_i. \end{aligned}$$

In the characterization, the number of nodes of the graph equals the number of available degeneration states, q is the expected distribution followed by the new node when linking to an existing one, and p is the actual distribution, outcome of the actual link established.

In Figure 3 we plot the von Neumann final entropy as function of the clauses-to-variables ratio α. The BE systems always have lower entropy than the MB ones, and the entropy behavior is in keeping with the distribution of BEC shown in [11]. Hence, the larger is the probability that BEC occurs, the larger is the entropy. Moreover, the plot shows that when α decreases the entropy increases. This process is associated with an increment of the spatial (geometric) disorder of the particles [23]. It follows that the disorder of the particles is related to the satisfiability of the instance (which depends on the value of α). In this way, we are able to characterize the complexity of certifying the satisfiability of a k-SAT instance by computing its entropy (see Table 2). We have also performed further analyses, omitted due to lack of space, using the Rényi entopy [24] and the information gain [25] to confirm our results.

Entropy	Particle status	Information	SAT-instance structure	SAT/UNSAT	Probability of BEC
high	disorder	high	many free variables available to satisfy the few clauses of the formula	SAT	high
low	order	low	few free variables available to satisfy the many clauses of the formula	UNSAT	low

Table 2. *Dictionary translating entropy status of the SAT formula into its structure*. When available, a free variable can be set to TRUE or FALSE so as to satisfy the instance. An instance in the SAT or UNSAT phase is satisfiable or unsatisfiable (respectively) with high probability.

4.2 The role of the non-integer out-degree θ

In [11], the parameter $\theta \in]0; 1]$ represents the *out-degree* increment of the node $v(C_{t_i})$ linked to the existing node $v(C_{t_j})$ during the generation of the graph. In particular,

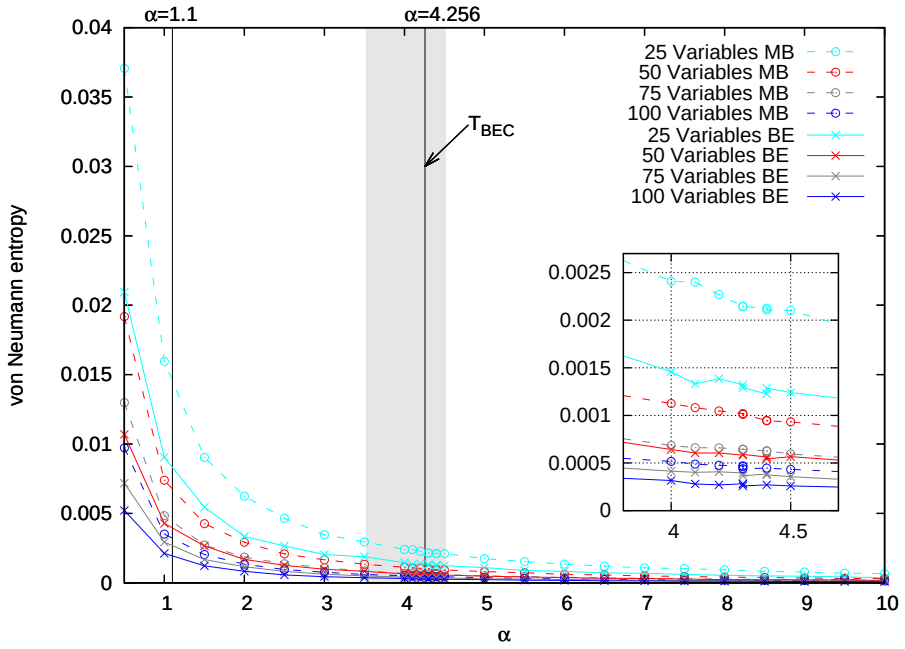

Fig. 3. *von Neumann entropy of* 3*-SAT*. When α decreases the entropy increases. Since the entropy represents the spatial disorder of the particles [23], in the k-SAT problem the physical concept of disorder is related to the complexity of certifying the satisfiability of the instance. The plot shows that BE systems always have lower entropy than MB ones, in agreement with the physical distribution of particles. Indeed, BE systems are often characterized by particles established on the same energy level, which in case of BEC is the lowest available.

the same edge between the new node $v(C_{t_i})$ and $v(C_{t_j})$ increases their connectivities k_{t_i} and k_{t_j} as $k_{t_i} = k_{t_i} + \theta$ and $k_{t_j} = k_{t_j} + 1$. Making use of the relation $k_i = \theta \cdot od(v(C_i)) + id(v(C_i))$, where od and id are the *out-degree* and *in-degree* of $v(C_{t_i})$ respectively, nodes aim to connect to a particular node, which gets richer and richer. Indeed, as incoming links are rewarded more than outgoing links (1 and $\theta < 1$ respectively), the connectivity of the node that acquires links increases much more than the connectivity of the nodes linking to it. In this work, we aim to show the connection between θ and the temperature of the system under investigation, namely the k-SAT instance.

It is largely known that BEC takes place only in a dilute gas of bosons cooled to temperatures near to the absolute zero [26]. Therefore, we analyze the system focusing on values of θ near to zero, in order to prove that for low values of θ the BEC takes place with more probability than for high values of θ. In order to understand the role played by θ, we analyze 41 different values of $\theta \in]0, 1]$. Figure 4 shows that the probability of BEC increases as θ decreases. Hence, if θ approaches zero, the number of BECs increases, suggesting that the parameter θ in the S2G algorithm is related to the absolute temperature in the corresponding physical system.

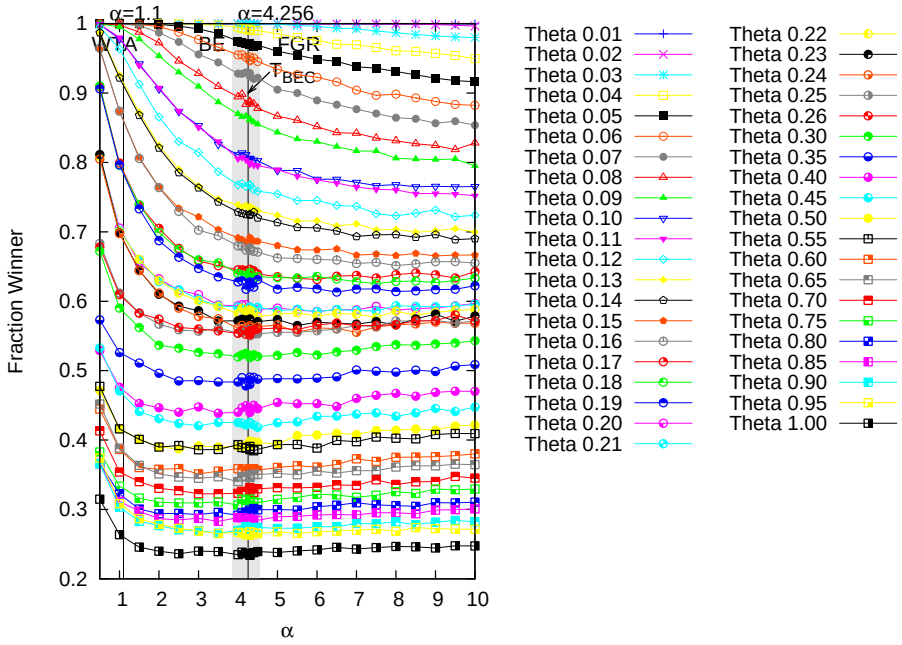

Fig. 4. *Fraction Winner of* 3*-SAT*. We report the number of links shared by the winner node against α. The FW is almost linear for $\theta \in [0.25, 1]$. For $\theta \in [0.05, 0.20]$, we observe a nonlinear variation, meaning that if θ approaches 0, the number of BEC cases increases in a nonlinear way. This behavior of θ is strongly related to the physical property of having a large number of BECs when the absolute temperature approaches 0.

In order to investigate thoroughly the nonlinearity shown by the FW as function of θ, we define $g(\alpha)$ as the third-order polynomial fitting computed with the least-squares approach for each curve plotted in Figure 4. We define the following measure of nonlinearity over an interval $[a, b]$ as $\int_a^b |g''(\alpha)|$. Using this formula we can compute the amount of variation in the first derivative of g, thus quantifying the nonlinearity of g in its domain. Notably, the results shown in Figure 5 allow to identify some clusters of fraction winner plots showing similar nonlinear behavior.

5 The Maxwell-Boltzmann CHAINSAT

In section 3 we focused on the fraction winner related to the MB and the BE distributions. The different results obtained by running the S2G algorithm with these two distributions lead to the creation of the final graphs with different topological structures. That is, the energy assigned by S2G to each clause (i.e., the weight of the clause) changes depending on the distribution used. In order to study the changes that each distribution involves, here we run the LC version of CHAINSAT [11] using the weight provided by the S2G algorithm run with MB and BE distributions. The SAT instances are cre-

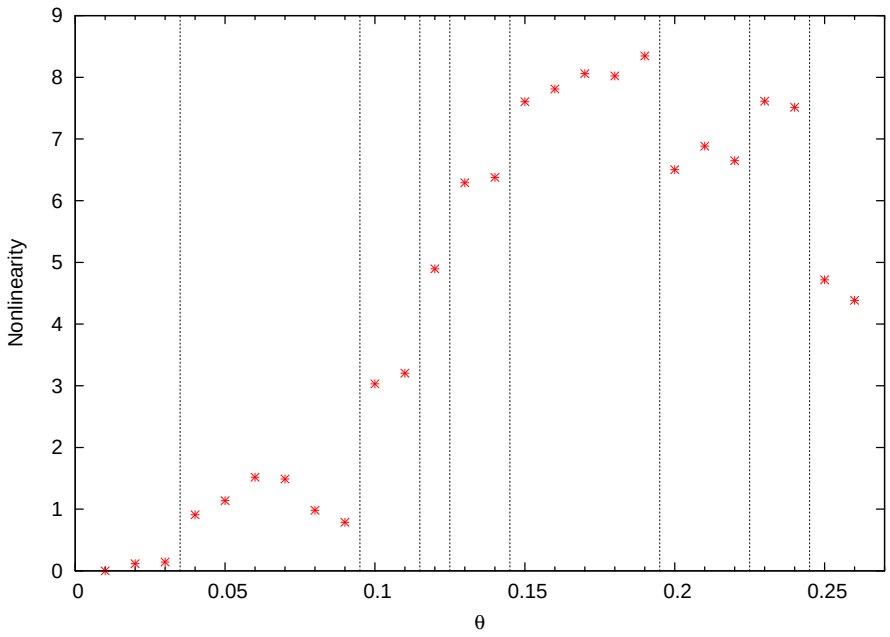

Fig. 5. *Nonlinearity of the Fraction Winner as function of θ*. The plots of the FW have been clustered according to their nonlinearity, with $\theta \in]0, 0.26]$. When θ approaches 0, the FW exhibits a linear behavior, while the nonlinearity starts from $\theta = 0.04$. Interestingly, in $[0.09, 0.14]$ there is a large increment of nonlinearity.

ated through A. van Gelder's generator MKCNF.C[1]. We investigate how the number of satisfied clauses and the number of flips change in each modified version of CHAINSAT (named MB-CHAINSAT and BE-CHAINSAT). Surprisingly, as shown Figure 9, by comparing the percentage of satisfied clauses we can speculate that MB-ChainSAT satisfies the same number of instances as BE-ChainSAT.

We investigate this comparison more thoroughly by analyzing the difference of satisfied clauses for each value of n (number of variables). In particular, we observe that MB-CHAINSAT and BE-CHAINSAT satisfy approximately the same number of clauses (Table 3). We also study the number of flips performed by the two algorithms, that is their computational effort. Figure 6 displays the number of flips (normalized to 1) obtained by running BE-CHAINSAT and MB-CHAINSAT. MB-CHAINSAT requires a higher computational effort than BE-CHAINSAT, except for $n = 25$. A detailed analysis of the number of flips is shown in Table 3.

The average number of flips used for solving an instance with MB-CHAINSAT is 1223655.71 against 1211622.34 with BE-CHAINSAT. Therefore, on average, BE-CHAINSAT needs a lower computational effort than MB-CHAINSAT. This is due to

[1] ftp://dimacs.rutgers.edu/pub/challenge/satisfiability/contributed/UCSC/instances.
We set the generator to obtain both satisfiable and unsatisfiable formulas, so as to obtain a purely uniform random k-SAT distribution.

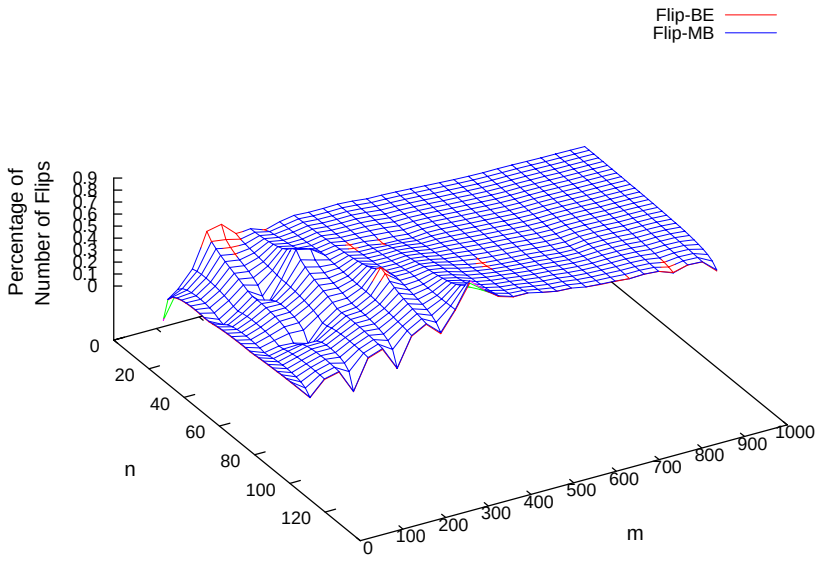

Fig. 6. *Computational Effort of* LC-CHAINSAT. We plot the number of flips (normalized to 1) performed by BE-CHAINSAT and MB-CHAINSAT, using the LC version. BE-CHAINSAT employs less flips than MB-CHAINSAT, therefore requiring a lower computational effort. This highlights the importance of the topological structure of a graph. Since with the BE distribution the graph has a FW higher than the graph built with the MB distribution, there exists a node that, sharing a large number of links, allows to find a solution performing less flips.

the structure of the BE-graphs, with a fraction winner higher than MB-graphs (as shown in Section 3): a node that shares more links enables to find a truth assignment for the instance, performing less flips. Therefore, if we consider graphs with $n > 25$ and our LC version of CHAINSAT, BE-CHAINSAT performs better than MB-CHAINSAT, as it solves the same number of instances performing less flips.

Conversely, when using the NLC version of CHAINSAT [11], Figure 7 shows that MB-CHAINSAT generally outperforms BE-CHAINSAT in terms of satisfied clauses, but when m approximates 400, BE-CHAINSAT satisfies more clauses than MB-CHAINSAT. We investigate this comparison more thoroughly by analyzing the difference of satisfied clauses for each value of n (number of variables). The results are shown in Table 4. Finally, in Figure 8 we show that in the NLC case MB-CHAINSAT is computationally less expensive than BE-CHAINSAT.

In conclusion, for the LC version, the BE distribution (namely, BE-CHAINSAT) always obtains better results than MB-CHAINSAT and CHAINSAT, as it solves the same percentage of clauses with a lower number of flips. For the NLC version, the MB distribution (namely, MB-CHAINSAT), outperforms the other algorithms for each value of n.

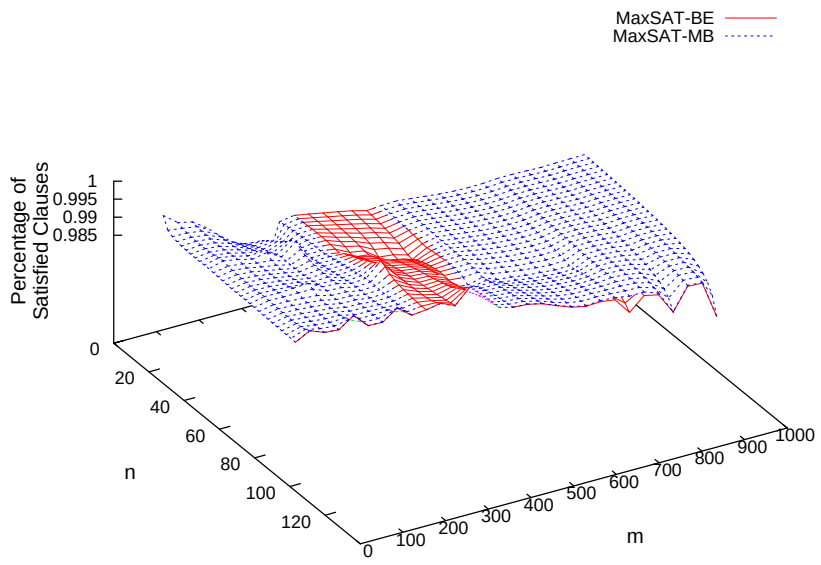

Fig. 7. *MaxSAT for* NLC-CHAINSAT. We plot the percentage of clauses satisfied by BE-CHAINSAT and MB-CHAINSAT as function of the number of clauses m and variables n, by using the Not-Linked-Clauses versions (NLC). MB-CHAINSAT seems to solve a higher percentage of clauses than BE-CHAINSAT. Actually, MB-CHAINSAT outperforms BE-CHAINSAT only for few values of α. The two algorithms solve exactly the same number of instances for $n = 50$, $n = 75$ and $n = 100$. (A more detailed analysis is shown in Table 4.)

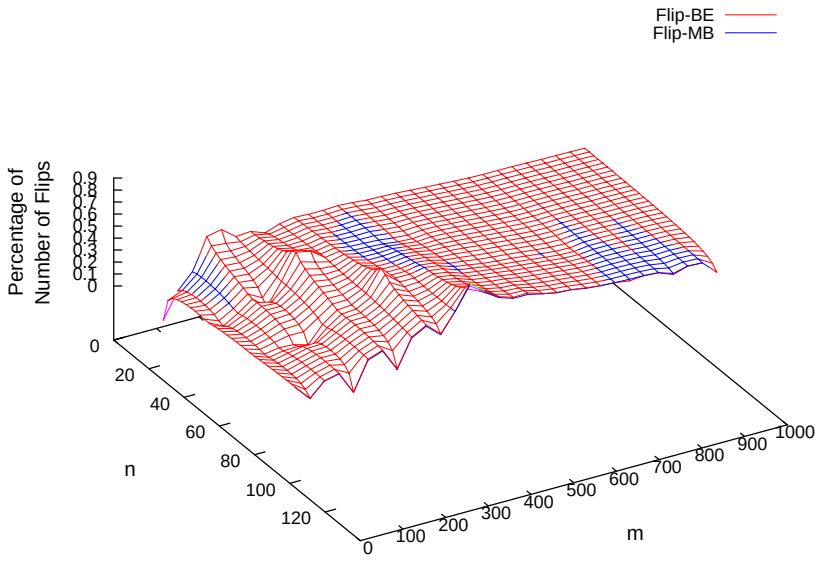

Fig. 8. *Computational Effort of* NLC-CHAINSAT. We plot the number of flips (normalized to 1) performed by the two algorithms by using the NLC version. BE-CHAINSAT employs more flips than MB-CHAINSAT, thus requiring a higher computational effort.

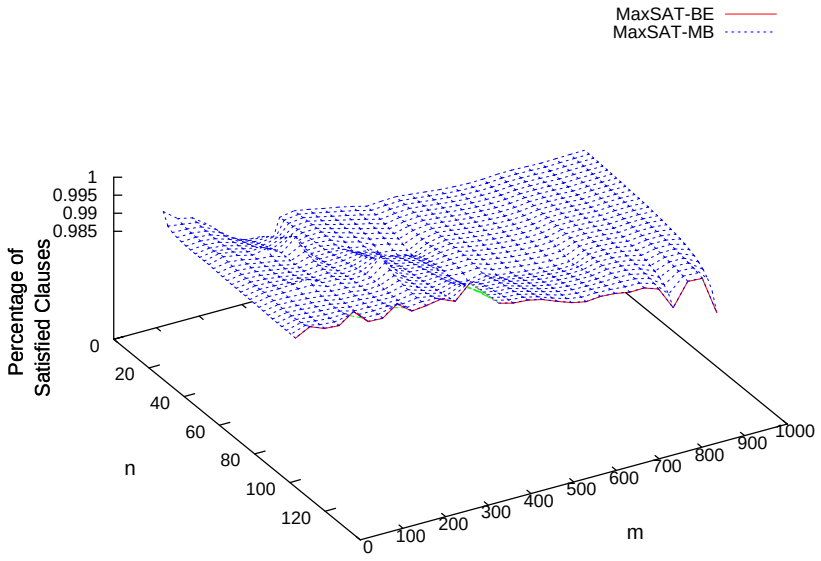

Fig. 9. *MaxSAT for LC-ChainSAT.* We plot the percentage of clauses satisfied by BE-ChainSAT and MB-ChainSAT as function of the number of clauses m and variables n, by using the Linked-Clauses version (LC). Although MB-ChainSAT seems to solve a higher percentage of clauses than BE-ChainSAT, the detailed analysis in Table 3 highlights that they solve the same number of instances for each value of n.

	LC-Maxwell Boltzman		LC-Bose Einstein		ChainSAT	
n	MaxSAT	Flips	MaxSAT	Flips	MaxSAT	Flips
25	0.8567	1185219.00	0.8567	1187645.11	0.8567	**1176755.92**
50	0.8580	1200317.11	0.8580	**1159281.68**	0.8580	1165854.05
75	0.8587	1228486.09	0.8587	**1219520.08**	0.8587	1231789.81
100	0.8590	1280600.65	0.8590	**1280042.75**	0.8590	1296716.23
Average	0.8581	1223655.71	0.8581	**1211622.34**	0.8581	1217779.00

Table 3. *Summary of* LC *SAT solvers performance*. For each value of n (number of variables), we report the percentage of satisfied clauses and the number of flips performed by MB-CHAINSAT, BE-CHAINSAT and CHAINSAT, using the LC version. All the algorithms ensure the same percentage of satisfied clauses. In terms of flips, CHAINSAT outperforms MB-CHAINSAT and BE-CHAINSAT only for $n = 25$. When $n > 25$, BE-CHAINSAT outperforms MB-CHAINSAT and CHAINSAT, as it solves the same percentage of clauses with a lower number of flips.

	NLC-Maxwell Boltzman		NLC-Bose Einstein		ChainSAT	
n	MaxSAT	Flips	MaxSAT	Flips	MaxSAT	Flips
25	0.8567	**1175219.03**	0.8563	1177645.11	0.8567	1176755.92
50	0.8577	**1162949.63**	0.8580	1166531.26	0.8580	1165854.05
75	0.8587	**1211373.56**	0.8587	1229056.62	0.8587	1231789.81
100	0.8590	**1269268.24**	0.8587	1275695.48	0.8590	1296716.23
Average	0.8580	**1204702.62**	0.8579	1212232.12	0.8581	1217779.00

Table 4. *Summary of* NLC *SAT solvers performance*. For each value of n (number of variables), we report the percentage of clauses satisfied and the number of flips performed by MB-CHAINSAT, BE-CHAINSAT and CHAINSAT by using the NLC version. As regards flips, MB-CHAINSAT outperforms the other algorithms for each value of n.

6 Discussion

In this research work, we have proposed a statistical-physical characterization for the Satisfiability problem. Starting from the S2G algorithm, we have developed a new algorithm in order to translate a SAT instance into a graph by using the BE (quantum) or the MB (classical) statistical distributions. The phase diagram of the graph provided by the algorithm shows evidence of condensation as the clauses-to-variables ratio decreases. Furthermore, we have carried out a systematic study to employ the characterization in the well-known CHAINSAT solver [15], without requiring a priori investigation of its solutions. The fitness-based sorting provided by our algorithm allows to enhance CHAINSAT.

In order to investigate the role of the quantum physics in our algorithm, we have cross-compared the behavior of our algorithm when using classical or quantum physics. From the phase diagrams, it is evident that graphs are more likely to undergo BEC if they are generated through the algorithm based on quantum physics. Moreover, when using our quantum and classical algorithms to drive the CHAINSAT solver, BE-CHAINSAT is computationally less or more expensive than MB-CHAINSAT, in terms of flips performed, depending on the CHAINSAT version used.

Notably, we have also defined the entropy of a SAT instance according to the information gained during the generation of its graph through S2G. From a statistical-physical standpoint, we have shown that the temperature of the SAT problem is mapped by the parameter representing the out-degree assigned dynamically to the nodes of the

graph associated with the instance. These findings highlight the emergence of a comprehensive characterization of the k-SAT problem using the classical and quantum particle distributions of statistical physics.

References

1. E. Clarke, A. Gupta, H. Jain, and H. Veith. Model checking: Back and forth between hardware and software. *Verified Software: Theories, Tools, Experiments*, pages 251–255, 2008.
2. U. Hustadt and R.A. Schmidt. *On evaluating decision procedures for modal logic*. MPI Informatik, Bibliothek & Dokumentation, 1997.
3. F. Lardeux, F. Saubion, and J.K. Hao. Three truth values for the SAT and MAX-SAT problems. In *International Joint Conference on Artificial Intelligence*, volume 19, page 187. Lawrence Erlbaum associates LTD, 2005.
4. Y. Hamadi, S. Jabbour, and L. Sais. Control-based clause sharing in parallel SAT solving. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 499–504. Morgan Kaufmann Publishers Inc., 2009.
5. T. Castell and M. Cayrol. Hidden Gold in Random Generation of SAT Satisfiable Instances. In *International Joint Conference on Artificial Intelligence (1)*, pages 372–377. Morgan Kaufmann, 1997.
6. SAT-competition. http://www.satcompetition.org/. 2012.
7. D.N. Pham, J. Thornton, and A. Sattar. Building structure into local search for SAT. In *20th International Joint Conference on Artificial Intelligence*, pages 2359–2364, 2007.
8. A. Belov, M. Järvisalo, and Z. Stachniak. Depth-driven circuit-level stochastic local search for SAT. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence-Volume One*, pages 504–509. AAAI Press, 2011.
9. G. Audemard and L. Simon. Predicting learnt clauses quality in modern SAT solvers. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 399–404, 2009.
10. H. Kautz and B. Selman. Unifying SAT-based and graph-based planning. In *International Joint Conference on Artificial Intelligence*, volume 16, pages 318–325. Lawrence Erlbaum associates LTD, 1999.
11. C. Angione, A. Occhipinti, G. Stracquadanio, and G. Nicosia. Bose-Einstein Condensation in Satisfiability Problems. *European Journal of Operational Research*, 227(1):44–54, 2013.
12. D. Mitchell, B. Selman, and H. Levesque. Hard and easy distributions of SAT problems. In *Proceedings of the National Conference on Artificial Intelligence*, number July, pages 459–465. AAAI Press/MIT Press, 1992.
13. M. Mézard and A. Montanari. Constraint satisfaction networks in physics and computation. *Clarendon Press, Oxford*, 1(9):11, 2007.
14. M. Mézard and T. Mora. Constraint satisfaction problems and neural networks: a statistical physics perspective. *Journal of Physiology-Paris*, 103(1):107–113, 2009.
15. M. Alava, J. Ardelius, E. Aurell, P. Kaski, S. Krishnamurthy, P. Orponen, and S. Seitz. Circumspect descent prevails in solving random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 105(40):15253, 2008.
16. T. Richardson and R. Urbanke. *Modern coding theory*. Cambridge University Press, 2008.
17. R.K. Pathria. *Statistical mechanics*. Butterworth-Heinemann, 1996.
18. O. Penrose and L. Onsager. Bose-Einstein condensation and liquid helium. *Physical Review*, 104(3):576–584, 1956.
19. G. Bianconi and A.L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86(24):5632–5635, 2001.

20. T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2001.
21. D. Achlioptas and F. Ricci-Tersenghi. Random formulas have frozen variables. *SIAM Journal on Computing*, 39(1):260–280, 2009.
22. A. Kopp, X. Jia, and S. Chakravarty. Replacing energy by von Neumann entropy in quantum phase transitions. *Annals of Physics*, 322(6):1466–1476, 2007.
23. M. D. Bal'makov. Entropy and disorder. *Glass physics and chemistry*, 27(4):287–297, 2001.
24. A. Rényi. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.
25. R.K. Niven. Exact Maxwell–Boltzmann, Bose–Einstein and Fermi–Dirac statistics. *Physics Letters A*, 342(4):286–293, 2005.
26. C.J. Pethick and H. Smith. *Bose-Einstein condensation in dilute gases*. Cambridge university press, 2001.