A New Look at Survey Propagation and its Generalizations

Elitza Maneva* Elchanan Mossel† Martin J. Wainwright‡

Abstract

We study the survey propagation algorithm [19, 5, 4], which is an iterative technique that appears to be very effective in solving random k-SAT problems even with densities close to threshold. We first describe how any SAT formula can be associated with a novel family of Markov random fields (MRFs), parameterized by a real number ρ. We then show that applying belief propagation—a well-known "message-passing" technique—to this family of MRFs recovers various algorithms, ranging from pure survey propagation at one extreme ($\rho = 1$) to standard belief propagation on the uniform distribution over SAT assignments at the other extreme ($\rho = 0$). Configurations in these MRFs have a natural interpretation as generalized satisfiability assignments, on which a partial order can be defined. We isolate cores as minimal elements in this partial ordering, and prove that any core is a fixed point of survey propagation. We investigate the associated lattice structure, and prove a weight-preserving identity that shows how any MRF with $\rho > 0$ can be viewed as a "smoothed" version of the naive factor graph representation of the k-SAT problem ($\rho = 0$). Our experimental results show that message-passing on our family of MRFs is most effective for values of $\rho \neq 1$ (i.e., distinct from survey propagation); moreover, they suggest that random formulas may not typically possess non-trivial cores. Finally, we isolate properties of Gibbs sampling and message-passing algorithms that are typical for an ensemble of k-SAT problems. We prove that the space of cores for random formulas is highly disconnected, and show that for values of ρ sufficiently close to one, either the associated MRF is either highly concentrated around the all-star assignment, or it has exponentially small conductance. Similarly, we prove that for ρ sufficiently close to one, the all-star assignment is attractive for message-passing when analyzed in the density-evolution setting.

*Division of Computer Science, UC Berkeley, CA. Email: elitza@cs.berkeley.edu. Supported by NSF grant grant CCR-0121555

†Department of Statistics, UC Berkeley, CA. Email: mossel@stat.berkeley.edu. Supported by a Miller Fellowship in Computer Science and Statistics.

‡Department of Electrical Engineering and Computer Science and Department of Statistics, UC Berkeley, CA Email: wainwrig@eecs.berkeley.edu. Supported by NSF grant IIS-9988642, and a grant from Intel Corporation.

1 Introduction

The *survey-propagation* algorithm [19, 5] is an iterative "message-passing" technique designed to solve high-density random k-SAT problems. Non-rigorous arguments based on the replica method as well as experimental results both suggest that it may be effective even very close to the satisfiability threshold. Nonetheless, the reasons underlying this remarkable performance are not yet fully understood.

In this paper, we provide a new perspective on survey propagation, as well as a broader class of related algorithms. We begin by introducing a new family of Markov random fields (MRFs), parameterized by a pair (ω_o, ω_*) of non-negative real numbers, that can be associated with any k-SAT problem. Of especial interest is the line given by $\rho = \omega_*$ and $\omega_o = 1 - \rho$. In particular, we demonstrate that a range of algorithms—including survey propagation (SP) as a special case—can all be recovered as the well-known belief propagation algorithm [27] as applied to suitably restricted MRFs with parameters along the line segment $\rho \in [0, 1]$. This equivalence is significant, because belief propagation is a widely used method for computing approximations to marginal distributions in general Markov random fields.

Based on this equivalence, we turn to examining the combinatorial and analytical properties of our extended Markov structures. Configurations in these MRFs have a natural interpretation as generalized satisfiability assignments, on which a partial ordering and hence a lattice structure can be defined. We refer to the minimal element in any such lattice as a *core*, and we prove that any core is a fixed point of the pure form of survey propagation ($\rho = 1$). However, our experimental results suggest that non-trivial cores typically do not exist for random formulae. This observation motivates considering the broader family of Markov random fields for the range $0 < \rho < 1$, as well as the associated belief propagation algorithms, which we denote by $\mathrm{SP}(\rho)$. By exploiting the lattice structure of generalized assignments, we establish a new combinatorial identity involving generating functions on lattices. This identify reveals how the distribution for $\rho \in (0, 1)$ can be viewed as a "smoothed" version of the MRF with $\rho = 0$. The latter MRF is simply the uniform distribution over

(ordinary) satisfying assignments, which is conjectured to be disconnected for high formula densities [18, 19, 5]

Our experimental results on the SP(ρ) algorithms indicate that they are most effective for values of ρ close to but different from one. One intriguing possibility is that the effectiveness of pure survey propagation (i.e., SP(1)) may be a by-product of the fact that SP(ρ) is most effective for values of ρ less than but close to 1. In addition, we consider alternative sampling-based methods (e.g., Gibbs sampling) for computing marginals for the extended MRFs. Success of such alternative methods provides independent confirmation of the significance of the extended MRF representation in the success of survey propagation. Finally, we study properties of both message-passing and Gibbs sampling that are typical over a random ensemble of k-SAT problems. We establish results that link the typical behavior of Gibbs sampling and message-passing algorithms under suitable initialization, and when applied to the extended family of MRFs with ρ sufficiently close to one.

The fact that the pure form of survey propagation (i.e., SP(1) in our notation) is a form of belief propagation was first conjectured by Braunstein et al. [5], and established independently of our work by Braunstein and Zecchina [6], and Aurell et al. [2]. However, both of the latter papers treat only the special case $\rho = 1$, and do not provide a natural combinatorial interpretation. Our result is a generalization, in that it applies to the full range of $\rho \in [0, 1]$. Moreover, the combinatorial structures intrinsic to our Markov random fields—viz. generalized assignments, cores and lattices—highlight the importance of values $\rho \neq 1$.

The remainder of this paper is organized as follows. In §2, we introduce the background and notation necessary to set up the problem. Statements and discussion of our main results can be found in §3. Due to space constraints, most proofs are omitted from this extended abstract, but can be found in the technical report [16].

2 Background

2.1 The k-SAT problem and graphical representation: We begin with notation and terminology necessary to describe the k-SAT problem. Let C and V represent index sets for the clauses and variables, respectively, where $|V| = n$ variables and $|C| = m$. We denote elements of V using the letters i, j, k etc., and members of C with the letters a, b, c etc. We use x_S to denote the subset of variables $\{x_i : i \in S\}$.

In the k-SAT problem, the clause indexed by $a \in C$ is specified by the pair $(V(a), J_a)$, where $V(a) \subset V$ consists of k elements, and $J_a := (J_{a,i} : i \in V(a))$ is a k-tuple of $\{0, 1\}$-valued weights. The clause indexed by a is *satisfied* by the assignment x if and only if $x_{V(a)} \neq J_a$. Equivalently, letting $\delta(y, z)$ denote an indicator function for the event $\{y = z\}$, if we define the function $\psi_{J_a}(x) := 1 - \prod_{i \in V(a)} \delta(J_{a,i}, x_i)$, then the clause a is satisfied by x if and only if $\psi_{J_a}(x) = 1$. The overall formula consists of the AND of all the individual clauses, and is satisfied by x if and only if $\prod_{a \in C} \psi_{J_a}(x) = 1$. For later use, we define the sets

$$\begin{aligned} C(i) &:= \{a \in C : i \in V(a)\}, \\ C^+(i) &:= \{a \in C(i) : J_{a,i} = 0\}, \end{aligned}$$

with $C^-(i) := C(i) \backslash C^+(i)$. Note that $C(i) = C^+(i) \cup C^-(i)$ is the set of clauses including the variable x_i. For each pair $(a, i) \in E$, the set $C(i) \backslash \{a\}$ of clauses that include neighbor i (where we exclude a) can be divided into two (disjoint) subsets, depending on whether their preferred assignment of x_i *agrees* (in which case $b \in C_a^s(i)$) or *disagrees* (in which case $b \in C_a^u(i)$) with the assignment of x_i preferred by clause a. More formally, we define

$$\begin{aligned} C_a^s(i) &:= \{b \in C(i) \backslash \{a\} : J_{a,i} = J_{b,i}\}, \\ C_a^u(i) &:= \{b \in C(i) \backslash \{a\} : J_{a,i} \neq J_{b,i}\}. \end{aligned}$$

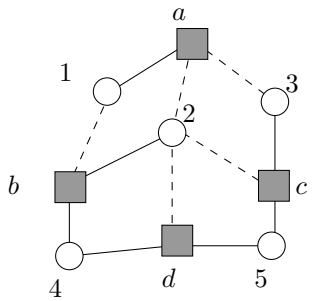

Figure 1. Factor graph representation of a 3-SAT problem on $n = 5$ variables with $m = 4$ clauses, in which circular and square nodes correspond to variables and clauses respectively. Solid and dotted edges (a, i), respectively, correspond to the weightings $J_{a,i} = 0$ and $J_{a,i} = 1$ respectively. The clause a is defined by the neighborhood set $V(a) = \{1, 2, 3\}$ and weights $J_a = (0, 1, 1)$. In traditional notation, this corresponds to the formula $(x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\bar{x}_1 \vee x_2 \vee x_4) \wedge (\bar{x}_2 \vee x_3 \vee x_5) \wedge (\bar{x}_2 \vee x_4 \vee x_5)$.

As illustrated in Figure 1, any instance of the k-SAT problem can be associated with a particular bipartite graph on the variables (denoted by circular nodes) and clause (denoted by square nodes), where the edge (a, i) between the clause $a \in C$ and variable $i \in V$ is included in E if and only if $i \in V(a)$. Following Braunstein et al. [5], it is convenient to introduce two labellings of any given edge—namely, solid or dotted, corresponding

to whether $J_{a,i}$ is equal to 0 or 1 respectively. We denote by

$$p(x) \propto \prod_{a \in C} \psi_{J_a}(x) \tag{2.1}$$

the uniform distribution on satisfying assignments of a formula. This distribution is well defined only if the formula is satisfiable.

2.2 Random instances and thresholds: The k-SAT problem for $k \geq 3$ is a classical NP complete problem [8]. This fact does not rule out the existence of efficient algorithms for deciding if random formulae are satisfiable, or for finding satisfying assignments for random formulae when they are satisfiable. Accordingly, of interest to us are random instances of the k-SAT problem, where given a density parameter α, we choose $m = \alpha n$ clauses uniformly and without replacement from the set of all k-clauses on n variables. In terms of the factor graph representation, this procedure samples a random (n, m)-bipartite graph, in which each clause $a \in C$ has degree k.

Clearly, a random formula becomes increasingly difficult to satisfy as the clause density α increases. Friedgut [12] showed that the probability that a formula is satisfiable exhibits a sharp threshold at a value $\alpha_c(n)$. It is widely believed that $\alpha_c(n)$ is independent of n. Rigorous bounds on α_c can be found in various papers [13, 7, 9, 10, 14, 1], whereas Monasson and Zecchina [20] derive approximations based on"replica method" calculations that yield $\alpha_c \approx 4.267$. For even values of k, these replica calculations have been shown rigorously [11, 21] to yield bounds on the threshold.

In several papers in the statistical physics literature [18, 19, 5], it is argued that in addition to the threshold α_c, there is another threshold $\alpha_d < \alpha_c$ (for $k = 3$, $\alpha_d \approx 3.921$), which marks the transition between a phase in which the satisfying assignments form a single cluster, and a phase where they form an exponential number of disconnected clusters. The clustering is in terms of a neighborhood structure in which assignments that are close in Hamming distance are considered neighbors. It is conjectured, moreover, that one manifestation of this phase transition is in the complexity of finding a solution. In particular, polynomial time algorithms that use only local information are expected to fail with high probability for random k-SAT instances with $\alpha > \alpha_d$. This conjecture is consistent with previous results [24] on local algorithms applied to the uniform distribution over SAT assignments (2.1), defined by the usual factor graph representation of k-SAT given in Figure 1.

2.3 Survey propagation: We now provide an explicit description of the SP(ρ) family of algorithms. For any given $\rho \in [0, 1]$, the algorithm involves updating messages from clauses to variables, as well as from variables to clauses. Each clause $a \in C$ passes a real number $\eta_{a \to i} \in [0, 1]$ to each of its variable neighbors $i \in V(a)$. In the other direction, each variable $i \in V$ passes a triplet of real numbers $\Pi_{i \to a} = (\Pi^u_{i \to a}, \Pi^s_{i \to a}, \Pi^*_{i \to a})$ to each of its clause neighbors $a \in C(i)$.

SP(ρ) updates:

Message from clause a to variable i:

$$\eta_{a \to i} = \prod_{j \in V(a) \setminus \{i\}} \Big[\frac{\Pi^u_{j \to a}}{\Pi^u_{j \to a} + \Pi^s_{j \to a} + \Pi^*_{j \to a}} \Big].$$

Message from variable i to clause a:

$$\Pi^u_{i \to a} = \Big[1 - \rho \prod_{b \in C^u_a(i)} (1 - \eta_{b \to i}) \Big] \prod_{b \in C^s_a(i)} (1 - \eta_{b \to i})$$

$$\Pi^s_{i \to a} = \Big[1 - \prod_{b \in C^s_a(i)} (1 - \eta_{b \to i}) \Big] \prod_{b \in C^u_a(i)} (1 - \eta_{b \to i})$$

$$\Pi^*_{i \to a} = \prod_{b \in C^s_a(i)} (1 - \eta_{b \to i}) \prod_{b \in C^u_a(i)} (1 - \eta_{b \to i}).$$

Remarks:

1. Although we have omitted the time step index for simplicity, the message passing equations should be interpreted as defining a recursion on (η, Π). The initial values for η are chosen randomly in the interval $(0, 1)$.

2. The idea of the ρ parameter is to provide a smooth transition from the original naive belief propagation algorithm to the survey propagation algorithm. As shown in [5], setting $\rho = 0$ yields the belief propagation updates applied to the probability distribution (2.1), whereas setting $\rho = 1$ yields the pure version of survey propagation.

Supposing that the messages converge, the overall conviction of a value at a given variable are computed from the incoming set of equilibrium messages as

$$\mu_i(1) \propto \Big[1 - \rho \prod_{b \in C^+(j)} (1 - \eta_{b \to j}) \Big] \prod_{b \in C^-(j)} (1 - \eta_{b \to j}), \tag{2.3}$$

with similar formula for $\mu_i(0)$ and $\mu_i(*)$. To be consistent with their interpretation as (approximate)

marginals, the triplet $\{\mu_i(0), \mu_i(*), \mu_i(1)\}$ at each node $i \in V$ is normalized to sum to one. The bias of variable x_i is defined as $|\mu_i(0) - \mu_i(1)|$. The decimation algorithm based on survey propagation [5, 6] consists of the following steps:

1. Run SP(1) on the SAT problem. Extract the fraction β of variables with the largest biases, and set them to their preferred values.

2. Simplify the SAT formula, and return to Step 1.

Once the maximum bias over all variables falls below a pre-specified tolerance, the Walk-SAT algorithm is applied to the formula to find the remainder of the assignment (if possible).

3 Our contributions

This section contains a high-level overview of our main contributions; detailed statements and proofs can be found in the technical report [16]. We begin by defining the novel family of MRFs, and then establish that survey propagation is equivalent to belief propagation as applied to suitably restricted forms of these MRFs. We then examine the combinatorial properties of these MRFs, as well as the empirical performance of Gibbs sampling compared to message-passing. Last, we provide some theoretical results on the behavior of Gibbs sampling and message-passing algorithms on random ensembles of formulae.

3.1 Distributions over generalized assignments:

The first step to forming our extended MRFs is to allow the variables $x = \{x_1, \ldots, x_n\}$ to take values in $\{0, 1, *\}^n$, to which we refer as a *generalized assignment*. It will be convenient, when discussing the assignment of a variable x_i with respect to a particular clause a, to use the notation $s_{a,i} := 1 - J_{a,i}$ and $u_{a,i} := J_{a,i}$ to indicate, respectively, the values that are *satisfying* and *unsatisfying* for the clause a.

DEFINITION 3.1. *A generalized assignment x is* invalid *for a clause a if either*

(a) all variables are unsatisfying (i.e., $x_i = u_{a,i}$ for all $i \in V(a)$), or

(b) all variables are unsatisfying except for exactly one index $j \in V(a)$, for which $x_j = *$.

Otherwise, the generalized assignment x is valid for clause a, and we denote this event by $\mathrm{VAL}_a(x_{V(a)})$. *We say that a generalized assignment is* valid *for a formula if it is valid for all of its clauses.*

We say that a variable x_i is the *unique satisfying variable* for a clause if it is assigned $s_{a,i}$ whereas all other variables in the clause (i.e., the variables $\{x_j : j \in V(a)\backslash\{i\}\}$) are assigned $u_{a,j}$. A variable x_i is *constrained* by clause a if it is the unique satisfying variable. We let $\mathrm{CON}_{i,a}(x_{V(a)})$ denote the indicator function for the event that x_i is the unique satisfying variable in the generalized assignment $x_{V(a)}$ for clause a. We say that a variable is constrained in an assignment if it is constrained by at least one clause.

The motivation behind defining case (b) as invalid is that the variable x_j is effectively forced to $s_{a,i}$, and so cannot be assigned the $*$ symbol. We define $n_*(x)$, $n_c(x)$ and $n_o(x)$ to be respectively the number of $*$-variables, the number of unconstrained variables, and the number of constrained variables in assignment x. (Note that by definition, it holds that $n_*(x) + n_c(x) + n_o(x) = n$.)

Let ω_o and ω_* be parameters in the interval $[0, 1]$, corresponding respectively to the "weights" of unconstrained variables and stars. We define weights of generalized assignments in the following way: invalid assignments x have weight $W(x) = 0$, and for any valid assignment x, we set

$$W(x) := (\omega_o)^{n_o(x)} \times (\omega_*)^{n_*(x)}.$$

Our primary interest is the probability distribution given by $p_W(x) \propto W(x)$. Note that if $\omega_o = 1$ and $\omega_* = 0$ then the distribution $p_W(x)$ is the uniform distribution (2.1) over satisfying assignments. Note however that if $\omega_* \neq 0$, then—in contrast to the earlier distribution (2.1)–this definition is valid for any SAT problem whether or not it is satisfiable, since the all-$*$ vector is always a valid generalized assignment.

Definition of the MRF: Next we show how the distribution p_W can be represented by a Markov random field for any choices of $\omega_o, \omega_* \in [0, 1]$. Doing so requires the addition of another dimension to our state space, which allows us to assess whether a given variable is constrained or unconstrained. We define the *parent set* of a given variable x_i, denoted by P_i, to be the set of clauses for which x_i is the unique satisfying variable. For each $i \in V$, let $\mathcal{P}(i)$ be the set of all possible parent sets of clause i. Due to the restrictions imposed by our definition, P_i must be contained in either $C^+(i)$ or $C^-(i)$ but not both. Therefore, the cardinality[1] of $\mathcal{P}(i)$ is $2^{|C^-(i)|} + 2^{|C^+(i)|} - 1$.

Our extended Markov random field is defined on the Cartesian product space $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, where $\mathcal{X}_i := \{0, 1, *\} \times \mathcal{P}(i)$. The distribution factorizes as a product of compatibility functions at the variable and clause nodes of the factor graph, which are defined as

[1] Note that it is necessary to subtract one so as not to count the empty set twice.

follows. Each variable node $i \in V$ has an associated compatibility function of the form:

$$\Psi_i(x_i, P_i) \quad := \quad \begin{cases} \omega_o & : \quad P_i = \emptyset, x_i \neq * \\ \omega_* & : \quad P_i = \emptyset, x_i = * \\ 1 & : \quad \text{for any other valid } (P_i, x_i) \end{cases}$$

The role of these functions is to assign weight to the generalized assignments according to the number of unconstrained and star variables, as in the weighted distribution p_W.

The compatibility functions at the clause nodes serve to ensure that only valid assignments have non-zero probability, and that the parent sets $P_{V(a)} := \{P_i \ : \ i \in V(a)\}$ are consistent with the assignments

$$x_{V(a)} := \{x_i \ ; \ i \in V(a)\}$$

in the neighborhood of a. More precisely, we require that the generalized assignment $x_{V(a)}$ is valid for a (i.e., $\mathrm{VAL}_a(x_{V(a)}) = 1$) and that for each $i \in V(a)$, exactly one of the two following conditions holds:

(a) $a \in P_i$ and x_i is constrained by a or

(b) $a \notin P_i$ and x_i is not constrained by a.

The following compatibility function corresponds to an indicator function for the intersection of these events:

$$\Psi_a\big(x_{V(a)}, P_{V(a)}\big) := \\ \mathrm{VAL}_a(x_{V(a)}) \times \prod_{i \in V(a)} \delta\big(\mathrm{Ind}[a \in P_i], \ \ \mathrm{CON}_{a,i}(x_{V(a)})\big).$$

We now form a Markov random field over generalized assignments and parent sets by taking the product of variable and clause compatibility functions

$$p_{gen}(x, P) \propto \prod_{i \in V} \Psi_i(x_i, P_i) \quad \prod_{a \in C} \Psi_a\big(x_{V_a}, P_{V(a)}\big). \tag{3.4}$$

With these definitions, it is straightforward to verify that $p_{gen} = p_W$, so that we have our desired Markov representation of p_W.

3.2 A unified class of algorithms: Having defined the MRF p_{gen}, it is possible to write out the associated belief propagation updates. Although (at least in principle) this step is straightforward, the calculations involved are non-trivial; complete details can be found in the technical report [16]. At a high level, two key facts are that the clause-to-variable messages can be parameterized by only three numbers $\{M^u_{a\to i}, M^s_{a\to i}, M^*_{a\to i}\}$, and that the variable-to-clause message requires only six values.

With this set-up, we have the following:

THEOREM 3.1. *For all $\omega_* \in [0, 1]$, the BP updates on the extended (ω_*, ω_o)-MRF are equivalent to the* $\mathrm{SP}(\omega_*)$ *family of algorithms under the following restrictions:*

(a) the constraint $\omega_o + \omega_* = 1$ is imposed, and

(b) all messages are initialized such that $M^u_{a\to i} = M^*_{a\to i}$ for every edge (a, i).

Remarks: The essence of Theorem 3.1 is that the pure survey propagation algorithm, as well as all the ρ-variants thereof, are all equivalent to belief propagation on our extended MRF with suitable parameter choices. It thus generalizes independent work by other groups [6, 2] on the special case SP(1). This broader equivalence is important for a number of reasons:

1. Belief propagation is a widely-used algorithm for computing approximations to marginal distributions in general Markov random fields. It also has a variational interpretation as an iterative method for attempting solve a non-convex optimization problem based on the Bethe approximation [27, 26]. Among other consequences, this variational interpretation leads to other algorithms that also solve the Bethe problem, but unlike belief propagation, are guaranteed to converge.

2. Given the link between SP and extended MRFs, it is natural to study combinatorial and probabilistic properties of the latter objects. In §3.3, we show how so-called "cores" arise as fixed points of SP(1), and we prove a weight-preserving identity that shows how the extended MRF for general ρ is a "smoothed" version of the naive MRF.

3. Finally, since BP (and hence SP) is computing approximate marginals for the MRF, it is natural to study other ways of computing marginals and examine if these lead to an effective way for solving random k-SAT problems. We begin this study in §3.4.

3.3 Combinatorial properties: This section is devoted to investigation of the combinatorial properties associated with our extended Markov random field. Consider a directed graph G, in which the vertex set consists of all valid generalized assignments. For a given pair of valid generalized assignments x and y, the graph includes a directed edge from x to y if there exists an index $i \in V$ such that $x_j = y_j$ for all $j \neq i$, and $x_i \neq y_i = *$. We label this edge with i. Notice that the out-degree of a valid assignment x is exactly equal to its number of unconstrained variables $n_o(x)$. We separate the valid assignments into $n + 1$ levels, so that assignment x is in level $n_*(x)$. Thus, every edge

is from an assignment in level l to one in $l+1$, where $l \in \{0, 1, \ldots, n-1\}$. Since G is acyclic, we can use its structure to define a partial ordering; in particular, we write $y < x$ if there is a directed path in G from x to y. Notice that all directed paths from x to y are labeled by indices in the set $T = \{i \in V : x_i \neq y_i = *\}$, and only the order in which they appear is different.

Cores: As a particular case of Theorem 3.1, setting $\omega_* = 1$ and $\omega_o = 0$ yields the extended MRF that underlies the SP(1) algorithm. In this case, the only valid assignments with positive weight are those without any unconstrained variables. More formally, we define a *core assignment* to be a valid generalized assignment $x \in \{0, 1, *\}^n$ such that for any $i \in V$ such that $x_i \neq *$, the variable x_i is constrained by at least one clause of the formula. Thus, the distribution p_W for $(\omega_o, \omega_*) = (0, 1)$ is simply uniform over the core assignments.

For a valid assignment x and a subset $S \subseteq V$, let $\gamma_S(x)$ be the minimal $y \leq x$, such that the path from x to y is labeled only by indices in S. It is easy to show that there exists a unique minimal y, that can be found by sequentially converting unconstrained variables from the set S into $*$-assignments. The following result connects fixed points of SP(1) to these minimal generalized assignments:

PROPOSITION 3.1. *For a valid assignment x, let* SP(1) *be initialized by:*

$$\Pi^u_{i\to a} = \delta(x_i, u_{a,i}), \quad \Pi^s_{i\to a} = \delta(x_i, s_{a,i}), \quad \Pi^*_{i\to a} = 0.$$

Then within a finite number of steps, the algorithm converges and the output fields are

$$\mu_i(b) = \delta(y_i, b),$$

where $y = \gamma_V(x)$ and $b \in \{0, 1, *\}$.

Thus, SP(1), when suitably initialized, simply strips the valid assignment x down to its core $\gamma_V(x)$. Moreover, Proposition 3.1, in conjunction with Theorem 3.1, leads to viewing the pure form of survey propagation SP(1) as performing an approximate marginalization over cores. This view raises the key question: do cores exist for random formulae? While it is easy to construct formulae with or without cores, to date we have been unable to prove (or disprove) the existence of cores for high-density random formula. We have, however, performed experiments to address this question; these results and the case $k = 2$ lead to a natural conjecture, which is described below.

Experiments on cores: We have performed a large number of the following experiments: starting with a satisfying assignment x, choose one its unconstrained variables uniformly at random, change it to $*$, and repeat until there are no unconstrained variables. This procedure is equivalent to taking a random path from x in G, by choosing at each step a random outgoing edge. Any such path terminates at the core $\gamma_V(x)$. We refer to this procedure as "peeling". It is interesting to examine at each step of this process the number of unconstrained variables (or outgoing edges). If it becomes 0 at any point, the process stops at a core assignment. Figure 2 shows the results of such an experiment for $n = 1000$, using different values of α, and $k = 3$. The plotted curves are the evolution of the number of unconstrained variables as the number of $*$'s increases. For small n, for example $n = 100$, and α close to threshold, satisfying assignments often correspond to core assignments; a similar observation was also made by Braunstein and Zecchina [6]. In contrast, for larger n, this correspondence is rarely the case, as for example Figure 2 demonstrates. The generated curves suggest that $\gamma_V(x)$ is almost always the all-$*$ assignment, and moreover that for high density α, there is a critical level in G where the out-degrees are very low. Increasing α results in failure to find a satisfying assignment, rather than in the formation of real core assignments.

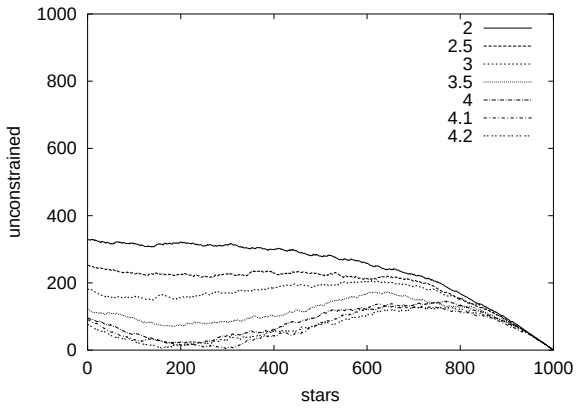

Figure 2. Evolution of the number of unconstrained variables in the following process: start with a satisfying assignment, change a random unconstrained variable to $*$ and repeat. Plotted is the result for $n = 1000$, for random formulae with $k = 3$ and $\alpha = \{2, 2.5, 3, 3.5, 4, 4.1, 4.2\}$. In particular, core assignments fall on the x-axis, and satisfying assignments fall on the y-axis.

Cores for $k = 2$: For $k = 2$, the event that there is a path in G from a satisfying assignment to the all-$*$ assignment has a very natural interpretation in terms of the event that the *pure-literal rule* succeeds

in finding an assignment. The pure-literal rule [23] is an algorithm consisting of the following steps: assign 1 to a variable if it only appears positive, and 0 if it only appears negative, reduce the formula, and repeat. It is straightforward to check that the sequence of variables given by the labels on any path from the all-$*$ assignment to a satisfying assignment can be identified with a sequence of steps of the pure-literal type. Furthermore, it is known [23] that there is a phase transition for the event that the pure-literal rule succeeds at $\alpha = 1$.

It is natural to conjecture that an analogous property holds for $k \geq 3$: in particular, to postulate that for all $\alpha < \alpha_c$, then with high probability there exists a satisfying assignment x and a sequence of variables, such that there exists a path in G from x to the all-$*$ assignment labeled by this sequence. Our experimental results suggest that this may even be true for almost all satisfying assignments.

If (as suggested by our results) non-trivial cores typically do *not* exist and therefore cannot explain the success of pure survey propagation, an alternative explanation is required. Accordingly, we propose studying the behavior of SP(ρ) for $\rho \in (0, 1)$. Our experimental results, consistent with similar reports from Kirkpatrick [15], show that SP(ρ) tends to be most effective in solving k-SAT for values of $\rho < 1$. If so, the good behavior of SP(1) may well follow from the similarity of SP(1) updates to SP(ρ) updates for $\rho \approx 1$. To further explore this issue, the effects of varying the weight distribution (ω_o, ω_*), and consequently the parameter ρ, are discussed in the following section.

Weight distribution: One of the benefits of our analysis is that it suggests a large pool of algorithms to be investigated. One option is to vary the values of ω_o and ω_*. A "good" setting of these parameters should place significant weight on precisely those valid assignments that can be extended to satisfying assignments. At the same time, the parameter setting clearly affects the level of connectivity in the space of valid assignments. Connectivity most likely affects the performance of belief propagation, as well as any other algorithm that we may apply to compute marginals or sample from the distribution.

Figure 3 shows the performance of belief propagation on the extended MRF for different values of (ω_o, ω_*), and applied to particular random formula with $n = 10000$, $k = 3$ and $\alpha = 4.2$. For weights satisfying $\omega_o + \omega_* > 1$, the behavior is very predictable: although the algorithm converges, the choices that it makes in the decimation steps lead to a contradiction. Note that these results show a sharp transition in algorithm behavior as the weights cross the line $\omega_o + \omega_* = 1$, which is representative of the more general behavior.

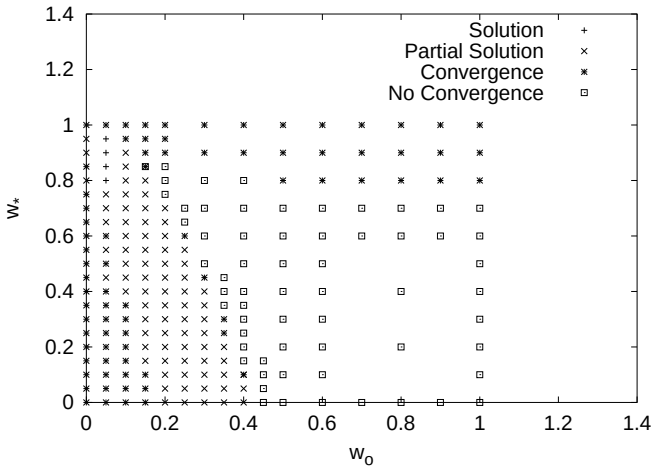

Figure 3. Performance of BP for different settings of ω_o, and ω_* for a particular random formula with $n = 10000$, $k = 3$, $\alpha = 4.2$. We distinguish between four cases: (i) BP converges and the decimation steps yields a complete solution, (ii) BP converges and the decimation steps yield a partial solution, completed by using Walk-SAT, (iii) BP converges but the decimation steps don't lead to a solution, and (iv) BP does not converge.

The following result, which we prove in [16], provides some justification for the excellent performance in the regime $\omega_o + \omega_* \leq 1$.

THEOREM 3.2. *If $\omega_o + \omega_* = 1$, then $\sum_{y \leq x} W(y) = (\omega_*)^{n_*(x)}$ for any valid assignment x. If $\omega_o + \omega_* < 1$, then $\sum_{y \leq x} W(y) \geq (\omega_*)^{n_*(x)}$ for any valid assignment x.*

In particular, if x is a satisfying assignment, then it holds that $n_*(x) = 0$, so that (for $\omega_o + \omega_* = 1$) the total weight of assignments below x in the partial order is exactly 1. Thus, Theorem 3.2 has a very natural interpretation in terms of a "smoothing" operation, in which the uniform weight assigned to each satisfying assignment is spread over the lattice associated with it.[2] Consequently, the (ω_o, ω_*)-MRF may be regarded as a smoothed version of the uniform distribution over satisfying assignments. Theorem 3.2 also provides intuition for the necessity of decimation. Indeed, although the measure p_W is closely related to the measure p, it is clearly not the case that marginals computed in one

[2] Note, however, that any generalized assignment that belongs to two or more lattices is assigned a weight only once. Otherwise, the transformation would be a convolution operation in a strict sense.

measure should be equal to the marginals computed in the other. Instead, the two marginals should agree on which variables are strongly biased towards 0 or 1. It is thus natural to set variables with large p_W-marginal probabilities for 0 or 1 to their preferred values.

3.4 Gibbs sampling: Based on our experiments, the algorithm SP(ρ) is very effective for appropriate choices of the parameter ρ. The link provided by Theorem 3.2 suggests that the distribution p_W, for which SP(ρ)—as an instantiation of belief propagation on the extended MRF—is computing approximate marginals, is likely to possess good "smoothness" properties. One expected consequence of such "smoothness" is that algorithms other than BP should also be effective in computing approximate marginals. Interestingly, rigorous conditions that imply (rapid) convergence of BP [25]—namely, uniqueness of Gibbs measures on the computation tree—are quite similar to conditions implying rapid convergence of Gibbs samplers, which are often expressed in terms of "uniqueness", "strong spatial mixing", and "extremality" (see, for example [17, 3]).

It is interesting to explore the application of sampling methods to the extended MRF as a means of computing unbiased stochastic approximations to the marginal distributions, and hence biases at each variable. Accordingly, we implemented a Gibbs sampler for the family of extended MRFs developed in Section 3.1, and used it to estimate marginal probabilities on the variables. (As a caveat, we cannot guarantee a priori that the dyanamics were run sufficiently long to reach the stationary distribution.) We compare the approximate marginals computed by the SP(β) family of algorithms (to which we refer as *pseudomarginals*) to the (stochastic) estimates computed by the Gibbs sampler. Given the manner in which the SP pseudomarginals are used in the decimation procedure, the most natural comparison is between the biases $\mu_i(0) - \mu_i(1)$ provided by the $SP(\beta)$ algorithm, and the biases $\tau_i(0) - \tau_i(1)$ associated with the Gibbs sampler (where τ_i are the approximate marginals obtained from Gibbs sampling on the extended MRF with parameter β). The results of such comparisons for the SP parameter $\rho \in \{0.95, 0.9, 0.7, 0.5\}$ and the Gibbs sampling parameter $\beta \in \{0.4, 0.5, 0.7, 0.9\}$ are shown in Figure 4. Comparisons are made for each pair (ρ, β) in these sets, and over a range of clause densities $\alpha \in \{4.2, 4.1, 4.0.3.8, 3.6, 3.4\}$. For fairly dense formulae (e.g., $\alpha \geq 4.0$), the general trend is that the SP(ρ) biases with larger ρ agree most closely with the Gibbs biases with β relatively smaller (i.e., $\beta < \rho$). For lower clause densities (e.g., $\alpha = 3.4$), the agreement between the SP(ρ) and Gibbs biases on $MRF(\beta)$ when $\beta = \rho$ is substantially closer. This pat-

SAT α	Gibbs β			
	0.4	0.5	0.7	0.9
4.2	**0.0493**	0.1401	0.3143	0.4255
4.1	**0.0297**	0.1142	0.3015	0.4046
4.0	0.0874	**0.0416**	0.2765	0.3873
3.8	0.4230	0.4554	0.1767	**0.0737**
3.6	0.4032	0.4149	0.1993	**0.0582**
3.4	0.4090	0.4010	0.2234	**0.0821**

(a) Comparison to SP(0.95)

SAT α	Gibbs β			
	0.4	0.5	0.7	0.9
4.2	**0.0440**	0.1462	0.3166	0.4304
4.1	0.0632	**0.0373**	0.2896	0.4119
4.0	**0.0404**	0.0666	0.2755	0.3984
3.8	0.1073	**0.0651**	0.2172	0.3576
3.6	0.1014	**0.0922**	0.1620	0.3087
3.4	0.3716	0.3629	0.1948	**0.0220**

(b) Comparison to SP(0.9)

SAT α	Gibbs β			
	0.4	0.5	0.7	0.9
4.2	SP fails	SP fails	SP fails	SP fails
4.1	**0.0230**	0.0985	0.3236	0.4341
4.0	0.0493	**0.0079**	0.3273	0.4309
3.8	0.0531	**0.0194**	0.2860	0.4104
3.6	0.0980	**0.0445**	0.2412	0.3887
3.4	0.0365	**0.0356**	0.1301	0.3869

(c) Comparison to SP(0.7)

SAT α	Gibbs β			
	0.4	0.5	0.7	0.9
4.2	SP fails	SP fails	SP fails	SP fails
4.1	**0.1925**	0.2873	0.3989	0.4665
4.0	**0.0483**	0.1092	0.2986	0.4179
3.8	0.0924	**0.0372**	0.3235	0.4323
3.6	**0.0184**	0.0304	0.2192	0.4009
3.4	0.0323	**0.0255**	0.0718	0.3613

(d) Comparison to SP(0.5)

Figure 4. Comparison of $SP(\rho)$ pseudomarginals for $\rho \in \{0.95, 0.9, 0.7, 0.5\}$ to marginals estimated by Gibbs sampling on weighted MRFs with $\beta \in \{0.4, 0.5, 0.7, 0.9\}$ for the range of SAT problems $\alpha \in \{4.2, 4.1, 4.0.3.8, 3.6, 3.4\}$. Each entry in each table shows the average ℓ_1 error between the biases computed from the $SP(\rho)$ pseudomarginals compared to the biases computed from Gibbs sampling applied to $MRF(\beta)$. Calculations were based on top 50 most biased nodes on a problem of size $n = 1000$. The bold entry within each row (corresponding to a fixed α) indicates the $MRF(\beta)$ that yields the smallest ℓ_1 error in comparison to the SP biases.

ter of results is consistent with the view that belief propagation would overcount due to cycles in the factor graph.

3.5 Expansion arguments for random formulae: In this section, we describe the use of simple random graph arguments in order to obtain typical properties of cores, as well as the behavior of Gibbs sampling or message-passing algorithms applied to the MRF associated with a randomly chosen formula. Throughout this section, we use p^{ϕ}_{W} to denote the MRF distribution for a fixed formula ϕ.

We begin with a result that establishes that cores, if they exist, are typically at least a certain linear fraction $c(\alpha, k)$ of the total number n of variables.

PROPOSITION 3.2. *Let ϕ be a random k-sat formula with $m = \alpha n$ clauses where $k \geq 3$. Then for all positive integers C the probability that ϕ has a core with C clauses is bounded by $\left(\frac{e^2 \alpha C^{k-2}}{n^{k-2}}\right)^C$. Consequently, if we define $c(\alpha, k) := (\alpha e^2)^{-1/(k-2)}$, then with probability tending to one as $n \to +\infty$, there are no cores of size strictly less than $c(\alpha, k)\, n$.*

This result shows that the size of cores (when they exist) is typically linear in n. Moreover, by definition, the extended MRF for $\rho = 1$ assigns positive mass to the all-$*$ vector. It follows that the state space of the MRF for $\rho = 1$ typically satisfies one of the following properties:

(a) Either the state space is trivial, meaning that it contains only the all $*$ state, or

(b) The state space is disconnected with respect to all random walks based on updating a small linear fraction of the coordinates in each step.

Our next goal is to establish that a similar phenomenon persists when ρ is close to 1 (i.e., when $1 - \rho$ is small).

First, we need to introduce some notions from the analysis of the mixing properties of Markov chains. Let T be a reversible chain with respect to a measure p on a state space Ω. For sets $A, B \subset \Omega$, write

$$q_T(A, B) = \sum_{x \in A, y \in B} p(x) T_{x \to y} = \sum_{x \in A, y \in B} p(y) T_{y \to x}.$$

The *conductance* of the chain T is defined as

$$c(T) = \inf_{S \subset \Omega} \{ \frac{q_T(S, S^c)}{p(S)(1 - p(S))} \}.$$

It is well-known that $c(T)/2$ is an upper bound on the spectral gap of the chain T and that $2/c(T)$ is a lower bound on the mixing time of the chain. We note moreover that the definition of T implies that for every two sets A, B it holds that $q_T(A, B) \leq \min\{p(A), p(B)\}$.

DEFINITION 3.2. *Consider a probability measure p on a space Ω of strings of length n. Let T be a Markov chain on Ω. The radius of T denoted by $r(T)$ is defined by*

$$r(T) := \sup\{d_H(x, y) : T_{x,y} > 0\},$$

where d_H is the Hamming distance. We let the radius r-conductance of p denote by $c(r, p)$ be $\sup\{c(T) :$ T is reversible w.r.t. p and $r(T) \leq r\}$.

Now returning to the random k-SAT problem, we write p_ρ for the measure $p_W = p^{\phi}_{W}$ with $\omega_* = \rho$ and $\omega_o = 1 - \rho$.

PROPOSITION 3.3. *Consider a randomly chosen k-SAT formula with density α, and the following two events for a fixed ρ:*

(a) A_n is the event that $p^{\phi}_{\rho}[n - n_*(x) \leq 2\sqrt{(1-\rho)}\, n] \geq 1 - \exp(-\Omega(n))$.

(b) B_n is the event that the measure p^{ϕ}_{ρ} satisfies $c(\sqrt{(1-\rho)}\, n, p_\rho) \leq \exp(-\Omega(n))$.

Then there exists a $\rho_0 \in (0, 1)$ such that if $\rho > \rho_0$ then $\mathrm{Prob}[\phi \in A_n \cup B_n] \to 1$ as $n \to +\infty$.

Proposition 3.3 shows that for values of ρ sufficiently close to 1, any random sampling technique based on local moves (e.g., Gibbs sampling), if started at the all $*$ assignment, will take exponentially long to get to an assignment with more than a negligible fraction of non-$*$. We now establish an analogous claim for belief propagation update on the extended Markov random fields. More precisely, we prove that if ρ is sufficiently close to 1, then running belief propagation with initial messages that place most of their mass on the all-$*$ assignment will produce messages that also place most of their mass on the all-$*$ assignment.

This result is proved in the "density-evolution" setting [22], by which we mean that the number of iterations is taken to be less than the girth of the graph, so that cycles have no effect. More formally, we establish the following:

THEOREM 3.3. *For every formula density $\alpha > 0$, arbitrary scalars $\epsilon'' > 0$ and $\delta > 0$, there exists $\rho' < 1$, $\epsilon' \in (0, \epsilon'')$ and $\gamma > 0$ such that for all $\rho \in (\rho', 1]$ and $\epsilon \in (0, \epsilon')$, the algorithm $SP(\rho)$ satisfies the following condition.*

Consider a random formula ϕ, a random clause b and a random variable i that belongs to the clause b. Then with probability at least $1-\delta$, if $SP(\rho)$ is initialized with all messages $\eta^0_{a \to j} < \epsilon$, then the inequality $\eta^t_{b \to i} < \epsilon'$ holds for all iterations $t = 0, 1, \ldots, \gamma \log n$.

4 Conclusion

We introduced a novel family of extended Markov random fields, parameterized by $\rho \in [0, 1]$, that can be associated with any k-SAT problem. We showed that applying belief propagation to these fields recovers various algorithms, ranging from survey propagation ($\rho = 1$) to belief propagation on the naive MRF representation of k-SAT ($\rho = 0$). The combinatorial properties of the extended MRFs provide insight into the performance of survey propagation (as well as broader class of related algorithms), including the role of cores and lattices, the importance of considering $\rho \neq 1$, the significance of decimation, and the smoothness properties of fields with $\rho > 0$ relative to the naive representation ($\rho = 0$).

References

[1] D. Achioptas and Y. Peres. The threshold for random k-SAT is $2^k 2 \log 2 - o(k)$. In *Proceedings of FOCS*, pages 223–231, 2003.

[2] E. Aurell, U. Gordon, and S. Kirkpatrick. Comparing beliefs, surveys, and random walks. Technical report, 2004. URL:http://lanl.arXiv.org/cond-mat/0406217.

[3] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. To appear, 2004.

[4] A. Braunstein, M. Mezard, M. Weigt, and R. Zecchina. Constraint satisfaction by survey propagation. Technical report, 2003. Preprint at URL:http://lanl.arXiv.org/cond-mat/0212451.

[5] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: an algorithm for satisfiability. Technical report, 2003. Preprint at URL:http://lanl.arXiv.org/cs.CC/0212002.

[6] A. Braunstein and R. Zecchina. Survey propagation as local equilibrium equations. Technical report, 2004. Preprint at URL:http://lanl.arXiv.org/cond-mat/0312483.

[7] V. Chvatal and B. Reed. Mick gets some (the odds are on his side). In *Proceedings of 33rd FOCS*, Pittsburgh Pennsylvania, 1992.

[8] S. Cook. The complexity of theorem-proving procedures. In *Proceeding of 3rd STOC*, page 151, 1971.

[9] W. F. de la Vega. On random 2-sat. Unpublished manuscript., 1992.

[10] O. Dubois, Y. Boufkhad, and J. Mandler. Typical random 3-sat formulae and the satisfiability threshold. In *Proceedings of 11'th SODA*, pages 126–127, 2000.

[11] S. Franz and M. Leone. Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.*, 111:535, 2003. URL:http://lanl.arXiv.org/cond-mat/0208280.

[12] E. Friedgut. Neccesary and sufficient conditions for sharp threhsolds of graph properties and the k-problem. *J. Amer. Math. Soc.*, 12:1017–1054, 1999.

[13] A. Goerdt. A remark on random 2-sat. *J. Computer System and Sciences*, 53:469–486, 1996.

[14] A. Kaporis, L. M. Kirousis, and E. G. Lalas. The probabilistic analysis of a greedy satisfiability algorithm. In *Proceedings of 10'th Annual European Symposium on Algorithm*, pages 574–585, 2000.

[15] S. Kirkpatrick. On survey propagation. Personal communication, 2004.

[16] E. Maneva, E. Mossel, and M. J. Wainwright. A new look at survey propagation and its generalizations. Technical report, Department of Statistics, UC Berkeley, September 2004. URL: https://lanl.arXiv.org/abs/cs.CC/0409012.

[17] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

[18] M. Mezard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297, 812, 2002. (Scienceexpress published online 27-June-2002; 10.1126/science.1073287).

[19] M. Mezard and R. Zecchina. Random k-satisfiability: from an analytic solution to an efficient algorithm. *Phys. Rev. E*, 66, 2002.

[20] R. Monasson and R. Zecchina. Statistical mechanics of the random k-satisfiability model. *Phys. Rev. E.*, 3:1357–1370, 1997.

[21] D. Panchenko and M. Talagrand. Bounds for diluted mean field spin glass models. 2004. Preprint at URL:http://lanl.arXiv.org/math.PR/0405357.

[22] T. Richardson and R. Urbanke. The capacity of low-density parity check codes under message-passing decoding. *IEEE Trans. Info. Theory*, 47:599–618, February 2001.

[23] J. Rosenthal, J. Plotkin, and J. Franco. The probability of pure literals. *Journal of Computational Logic*, 9:501–513, 1999.

[24] P. Svenson and M. G. Nordahl. Relaxation in graph coloring and satisfiability problems. *Phys. Rev. E.*, 59:3983–3999, 1999.

[25] S. Tatikonda and M. I. Jordan. Loopy belief propagation and Gibbs measures. In *Uncertainty in Artificial Intelligence (UAI), Proceedings of the Eighteenth Conference*, 2002.

[26] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational methods. Technical report, UC Berkeley, Department of Statistics, No. 649, 2003. Preprint at URL:http://www.stat.berkeley.edu/tech-reports/index.html.

[27] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In *Exploring Artificial Intelligence in the New Millennium*, chapter 8. Science and Technology books, 2003.