



# Bandgap optimization in combinatorial graphs with tailored ground states: application in quantum annealing

Siddhartha Srivastava<sup>1</sup> · Veera Sundararaghavan<sup>1</sup>

Received: 30 May 2022 / Revised: 16 July 2022 / Accepted: 17 July 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

A mixed-integer linear programming (MILP) formulation is presented for parameter estimation of the Potts model. Two algorithms are developed; the first method estimates the parameters such that the set of ground states replicate the user-prescribed data set; the second method allows the user to prescribe the ground states multiplicity. In both instances, the optimization process ensures that the bandgap is maximized. Consequently, the model parameter efficiently describes the user data for a broad range of temperatures. This is useful in the development of energy-based graph models to be simulated on Quantum annealing hardware where the exact simulation temperature is unknown. Computationally, the memory requirement in this method grows exponentially with the graph size. Therefore, this method can only be practically applied to small graphs. Such applications include learning of small generative classifiers and spin-lattice model with energy described by Ising hamiltonian. Learning large data sets poses no extra cost to this method; however, applications involving the learning of high dimensional data are out of scope.

**Keywords** Potts model · Ising model · Parameter estimation · Mixed integer linear programming · Boltzmann machines

## List of symbols

$\theta$	Set of parameters for potts model
$S$	A state of the graph
$\Delta E$	Band gap
$\eta$	Negative Log-likelihood

---

✉ Siddhartha Srivastava  
sidsriva@umich.edu

Veera Sundararaghavan  
veeras@umich.edu

<sup>1</sup> Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA

$C$	Set of graph's connection
$S$	Set of all possible states
$S_D$	Set of the data states
$S_E^\theta$	Set of all Excited states
$S_G^\theta$	Set of all Ground states
$V$	Set of graph's vertices
$E$	Potts Energy
$E_0$	Energy of ground state
$E_1$	Energy of 1 <sup>st</sup> excited state
$G$	Simple undirected weighted Graph
$N_C$	Number of graph connections
$N_{DS}$	Number of data states
$N_{ES}$	Number of excited states
$N_{GS}$	Number of ground states
$N_L$	Number of labels
$N_{TS}$	Number of total states
$N_V$	Number of graph vertices
$p(\mathbf{S})$	Probability of a state, $\mathbf{S}$
$s_i$	Label of Vertex with index $i$
$v_i$	Vertex with index $i$

## 1 Introduction

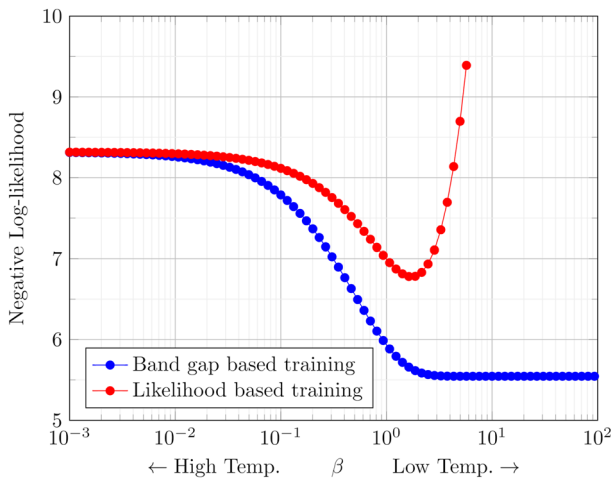
Potts energy model was initially developed to describe interacting spins on a crystalline lattice. Since then, it has become an archetypal model in other fields involving operations research, network theory, and physics of phase transition. The motion of biological cells was described by Graner and Glazier (1992) using a large-Q Potts model. A similar approach was used in Miodownik (2007) to study grain boundary motion in polycrystalline microstructures during thermally induced grain growth and recrystallization process. In such studies, the system's dynamics is represented as a transition probability governed by the model's energy description. These problems can be simulated using Monte Carlo based simulations. On the other hand, there are problems where the equilibrium solutions are required, for instance, in computer vision, Potts model is often used to describe the cut energy of a segmentation problem (c.f. Boykov et al. 2001). These problems are usually solved using the Graph-cut method. The computation of this process becomes exceedingly challenging as more generality is introduced. Bagon's thesis (Bagon 2012) provides an excellent review of these generalities and suggests practical algorithms.

Traditionally, these models are trained by considering them as Markov Random Fields (MRFs) and using gradient-based approaches to maximize the likelihood (Descombes et al. 1999). However, analytical estimates of the gradients are hard to compute. Among approximate techniques, Hinton's contrastive divergence method (Hinton 2002) provides an efficient way to approximate the gradients in the parameter optimization problem successively. An excellent review of this subject is

presented in Fischer and Igel (2012). Recently, the advent of Quantum annealing technology has made it easier to sample states from the model's probability distribution (Adachi and Henderson 2015). This development has significantly eased the approximation of the required gradients. However, these methods have a critical drawback. These techniques only work for finite temperature probability distribution. Consequently, the model trained using these techniques is often temperature-dependent and shows disagreement with the data as the temperature is lowered (Srivastava and Sundararaghavan 2020). As an example, the negative log-likelihood of a model trained using this technique is presented in Fig. 1. It can be seen that the minimum is close to the training  $\beta$  (inverse temperature), which was chosen as  $\beta = 1$ . A possible reason for this problem is that the training results in a locally optimal solution. Using quantum annealers adds another layer of complication because the simulation temperature is not known and depends on the graph size (Srivastava and Sundararaghavan 2020).

In contrast, this work is based on the band gap's maximization, while the ground states are chosen as the data states. This approach guarantees that the states' probability distribution gets closer to that of the data set as the temperature is reduced. Moreover, it ensures that the model adequately represents the data set for a broad range of temperatures. However, the downside of this approach is that there is no guarantee of the existence of parameters for every data set. This fact can be easily motivated by noticing that the number of ground states can be more than the number of model parameters and may result in an over-constrained optimization problem. Such problems do not exist at a non-zero temperature as all the states appear with non-zero probability.

In this paper, a Mixed Integer Linear Programming (MILP) formulation is presented to estimate Potts model parameters. Two variations of the algorithm are presented.



**Fig. 1** Comparative analysis of likelihoods of models trained using the Likelihood maximization and band gap maximization. The predicted models are presented in Appendix B.1. A lower value of Negative Log likelihood signifies a better trained model

The first algorithm assigns a prescribed data set as the model's ground states while maximizing the bandgap. The second algorithm identifies a set of ground states with a prescribed multiplicity while maximizing bandgap. It should be noted that the computational complexity of both the algorithms grows exponentially with the size of the problem. Therefore, these methods are only suited for small graph structures. These problems arise in designing energies of smaller motifs in a lattice structure.

The paper is organized as follows: The formulation for the Potts energy is reviewed in sect. 2. Concepts like the ground state, bandgap, and probability of a state are also reviewed. A theorem is presented to estimate the efficiency of the developed algorithms quantifiably. The problem statement is summarized in sect. 3. The developed algorithms are presented in sect. 4. A case study for the Ising model is presented in Sect. 5. Few details on the computational complexity are also outlined. Section 6 provides a summary of the paper.

## 2 Mathematical formulation

Potts model is a type of a discrete pairwise energy model on an undirected simple graph. In lieu of introducing some useful terms, following definition for graph is used:

*Graph:* A graph,  $G$ , is a pair of sets  $(\mathcal{V}, \mathcal{C})$ , where  $\mathcal{V}$  is the set of vertices and  $\mathcal{C}$  is the set of edges/connections. For each element  $e \in \mathcal{C}$  there is a corresponding ordered pair  $(x, y); x, y \in \mathcal{V}$  i.e.  $\mathcal{C} \subseteq \mathcal{V} \times \mathcal{V}$ . A Graph,  $G = (\mathcal{V}, \mathcal{C})$  is undirected if an edge does not have any directionality i.e  $(x, y) \equiv (y, x)$ . A graph is simple if  $(x, x) \notin \mathcal{C}$  for all  $x \in \mathcal{V}$ .

Also, this work requires the graph to be finite, i.e., the number of vertices is finite. Next, the definition of Potts energy is introduced.

### 2.1 Potts model

Consider a finite undirected simple graph  $G(\mathcal{V}, \mathcal{C})$ . The number of vertices are denoted by  $N_V = |\mathcal{V}|$  and the number of edges are denoted by  $N_C = |\mathcal{C}|$ . The indices of connections and vertices are related using the maps,  $\pi_1$  and  $\pi_2$  such that for a connection with index,  $k \in \{1, \dots, N_C\}$ , the index of the corresponding vertices are  $\pi_1(k)$  and  $\pi_2(k)$  with  $1 \leq \pi_1(k) < \pi_2(k) \leq N_V$ . This essentially means  $e_k \equiv (v_{\pi_1(k)}, v_{\pi_2(k)})$ . Each vertex,  $v_i \in V$  is assigned a state  $s_i \in \{1, 2, \dots, N_L\}$  for all  $i \in 1, \dots, N_V$ . This determines the complete state of the graph as an ordered tuple  $\mathbf{S} = (s_1, \dots, s_i, \dots, s_n) \in \{1, \dots, N_L\}^{N_V}$ . The set of all possible states is referred to as  $\mathcal{S} = \{1, \dots, N_L\}^{N_V}$  with the total number of states denoted by  $N_{TS} = |\mathcal{S}| = N_L^{N_V}$ . The Potts energy for a particular state can be evaluated as follows:

$$E(\mathbf{S}) = \sum_{i=1}^{N_V} H_i U(s_i) + \sum_{k=1}^{N_C} J_k V(s_{\pi(k,1)}, s_{\pi(k,2)}) \quad (1)$$

where,  $U(s)$  is the energy of labeling a vertex with label  $s$ , and  $V(s_i, s_j)$  is the energy of labeling two connected vertices as  $s_i$  and  $s_j$ . The parameters  $H_i$  and  $J_k$  are referred to as the Field strength and Interaction strength, respectively.

Since the graph is undirected, following symmetry is imposed:

$$V(s_i, s_j) = V(s_j, s_i)$$

The parameter set is represented as a vector,  $\theta = [\theta_1, \dots, \theta_{N_v+N_c}]^T$ . In this work, it is specialized to following form:

$$\theta = [H_1, \dots, H_{N_v}, J_1, \dots, J_{N_c}]^T$$

This notation allows to describe energy as a matrix-product evaluated as  $E(\mathcal{S}|\theta) = \epsilon(\mathcal{S})\theta$  where  $\epsilon(\mathcal{S})$

$$\epsilon(\mathcal{S}) = \left[ U(s_1), \dots, U(s_{N_v}), V(s_{\pi_1(1)}, s_{\pi_2(1)}), \dots, V(s_{\pi_1(N_c)}, s_{\pi_2(N_c)}) \right]$$

### 2.1.1 Ground states and band gap

For a given set of parameters,  $\theta$ , the set of ground states ( $\mathcal{S}_G(\theta) \subseteq \mathcal{S}$ ) is the set of states with minimum energy,  $E_0(\theta)$ , i.e.

$$\mathcal{S}_G(\theta) = \operatorname{argmin}_{\mathcal{S} \in \mathcal{S}} E(\mathcal{S}|\theta), \quad E_0(\theta) = \min_{\mathcal{S} \in \mathcal{S}} E(\mathcal{S}|\theta)$$

In contrast, all the non-minimal states are referred to as excited states. The set of all excited states, denoted by  $\mathcal{S}_E(\theta)$ , can be evaluated as:

$$\mathcal{S}_E(\theta) = \mathcal{S} - \mathcal{S}_G(\theta)$$

The cardinalities of the set of ground states ( $\mathcal{S}_G$ ) and excited states ( $\mathcal{S}_E$ ) are denoted by  $N_{GS}$  and  $N_{ES}$ , respectively. All excited states may or may not have the same energy. However, the minimum excited energy referred to as the ‘first excited energy’ is used in defining the band gap and is evaluated as:

$$E_1(\theta) = \min_{\mathcal{S} \in \mathcal{S}_E(\theta)} E(\mathcal{S}|\theta)$$

It should be noted that no assumption is made on the multiplicity of states with energy  $E_1(\theta)$ . The band gap (a positive quantity) defines the energy gap between  $\mathcal{S}_G$  and  $\mathcal{S}_E$ . It is estimated as:

$$\Delta E(\theta) = E_1(\theta) - E_0(\theta)$$

### 2.1.2 Probability distribution

At any given temperature,  $T$ , the probability of occurrence of a state,  $\mathcal{S}$  is described by the Boltzmann distribution as:

$$p(\mathcal{S}|\theta, \beta) = \frac{1}{Z} e^{-\beta E(\mathcal{S})} \tag{2}$$

where  $\beta = 1/k_B T$  is the inverse thermodynamic temperature,  $k_B$  is the Boltzmann constant and  $Z$  denotes the partition function which is estimated as

$$Z = \sum_{S \in \mathcal{S}} e^{-\beta E(S)}$$

## 2.2 Parameter estimation

Given a data set,  $\mathcal{S}_D \subseteq \mathcal{S}$ , the parameters set,  $\theta$ , is optimized such that the states in  $\mathcal{S}_D$  have higher probability of occurrence at a prescribed  $\beta$  value. Mathematically, this procedure entails minimization of negative log-likelihood as defined below:

$$\eta(\theta, \beta) = - \sum_{S \in \mathcal{S}_D} \log p(S|\theta, \beta) \quad (3)$$

It can be observed that at high temperatures i.e.  $\beta \rightarrow 0$ , all states occur with equal likelihood and therefore

$$\eta_0 = \lim_{\beta \rightarrow 0}(\theta, \beta) = N_{DS} \log(N_{TS})$$

where  $N_{DS} = |\mathcal{S}_D|$ . On the other hand, at low temperatures i.e.  $\beta \rightarrow \infty$ , only ground states occur with equal probability and occurrence of any other state has probability 0. Consequently, the value of  $\eta$  in this limit is finite only when  $\mathcal{S}_D \subseteq \mathcal{S}_G$ . It is evaluated as:

$$\eta_\infty(\theta) = \lim_{\beta \rightarrow \infty}(\theta, \beta) = N_{DS} \log(N_{GS})$$

It is desirable to estimate parameters such that the ground state replicates the data set, and the bandgap is maximized. The reason will be apparent after the next theorem (proof in Appendix A).

**Theorem** For a given set of parameters,  $\theta_D$ , such that (i)  $\mathcal{S}_G(\theta_D) = \mathcal{S}_D$  (ii)  $\Delta E > 0$ , following statements hold true:

(a)  $\eta(\theta_D, \beta)$  monotonically decreases with  $\beta$  and the low temperature limit

$$\eta_\infty(\theta_D) = \lim_{\beta \rightarrow \infty} \eta(\theta_D, \beta) = N_{GS} \log(N_{GS}) \quad (4)$$

(b)  $\eta(\theta_D, \beta)$  is bounded as:

$$N_{GS} \log(N_{GS}) < \eta(\theta_D, \beta) \leq N_{GS} \log(N_{GS} + N_{ES} e^{-\beta \Delta E}) \quad (5)$$

(c) For any  $\epsilon > 0$ , there exists a  $\beta^*$  such that for all  $\beta > \beta^*$ ,  $\eta(\theta_D, \beta) - \eta_\infty(\theta_D, \beta) < \epsilon$  where  $\beta^*$  is estimated as:

$$\beta^* = \frac{1}{\Delta E} \left( \log \frac{N_{ES}}{N_{GS}} - \log (e^{\epsilon/N_{GS}} - 1) \right) \tag{6}$$

The consequence of this theorem is that it guarantees that if the parameters are chosen appropriately,  $\eta$  will approach to its global minimum in the low temperature (high  $\beta$ ) limit. Moreover, at a finite  $\beta$ ,  $\eta$  is bounded from above by a decreasing function. It can be seen in Fig. 2a, that the bound gets tighter for higher values of  $\Delta E$ . It is also shown that the trained model is efficient in the range of  $\beta$  determined by  $[\beta^*, \infty)$ . Figure 2a shows that a higher bandgap allows a broader range of temperatures.

### 3 Problem statement

Given a finite undirected simple graph  $G(\mathcal{V}, \mathcal{C})$ , find parameters,  $\theta$  that maximizes the band gap in following two situations:

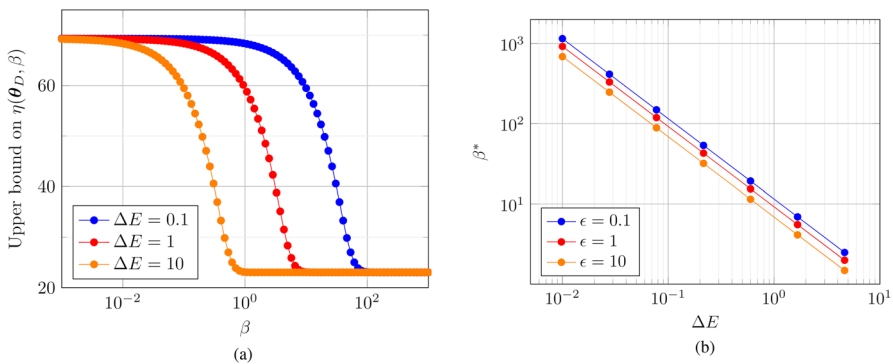
Case 1:  $\mathcal{S}_D$  is prescribed and  $\mathcal{S}_G(\theta_D) = \mathcal{S}_D$ .

Case 2: Ground state multiplicity,  $N_{GS}$ , is prescribed.

To make this optimization problem well posed, it is additionally imposed that  $H_i^{\min} \leq H_i \leq H_i^{\max}$  and  $J_k^{\min} \leq J_k \leq J_k^{\max}$ . Moreover, the functions  $U(s)$  and  $V(s_i, s_j)$  are predetermined and not calibrated in the optimization process.

### 4 Methods

A Mixed Integer Linear Programming (MILP) problem is formalized for parameter estimation of Potts model. A brief overview of the MILP formulation is presented below:



**Fig. 2** An illustration of bounds for a trained Potts model with  $N_{GS} = 10$  and  $N_V = 10$ . **a** The upper bound on  $\eta$  with respect to  $\beta$  for various values of energy gap **b**  $\beta^*$  as a function of band gap for various bounds on  $\eta$

*Mixed integer linear programming (MILP)*: An optimization problem is considered to be of MILP type when the objective function is linear in the decision variables and some of the decision variables are integer. A typical setup of MILP problem is given in Eq. (7) where  $\mathbf{x}$  is the decision variable of size  $N$ ,  $I$  is the set of indices of  $\mathbf{x}$  which are integers and the matrices  $\mathbf{A}_{eq}$ ,  $\mathbf{b}_{eq}$ ,  $\mathbf{A}$  and  $\mathbf{B}$  are used to define linear constraints.

$$\begin{aligned}
 \text{Optimize:} \quad & \min_{\mathbf{x}} \mathbf{c}\mathbf{x} \\
 \text{Inequality constraints:} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\
 \text{Equality constraints:} \quad & \mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq} \\
 \text{Bounds:} \quad & \mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub} \\
 \text{Integer variables:} \quad & \mathbf{x}_I \in \mathbb{Z}
 \end{aligned} \tag{7}$$

The MILP formulation for the two cases is presented next. In both cases, the decision variables include the parameters,  $\theta$ , and some auxiliary variables. These variables are introduced along with the algorithm description. Moreover, the algorithms do not enforce that  $\Delta E > 0$ . Therefore, the results are accepted only if this condition is met.

#### 4.1 Algorithm 1: Parameter estimation for potts model with Data set (PEPDAS)

The energies of individual states can be evaluated as a matrix product operation (shown in Section 2) which works well with linear programming framework. However, the calculation of band gap requires calculation of a minimum of energy over  $S_E$ . This operation introduces a non-linearity. Thus, following auxiliary variables are introduced to pose this optimization as a linear programming problem:

- $E_1$  (real valued scalar): It represents the energy of the 1<sup>st</sup> excited state.
- $\mathbf{m} = [m_1, \dots, m_{N_{ES}}]$  (binary valued vector of size  $N_E$ ): It is defined such that its value is 1 on exactly one index and 0 everywhere else. The index with value 1 must correspond to one of the 1<sup>st</sup> excited state.
- $M$  (real valued scalar): It represents a large positive number. For computational purposes it can be evaluated as:

$$M = \left( \max_s |U(s)| \right) \sum_{i=1}^{N_V} (|H_i^{\max}| + |H_i^{\min}|) + \left( \max_{s_1, s_2} |V(s_1, s_2)| \right) \sum_{k=1}^{N_C} (|J_k^{\max}| + |J_k^{\min}|) \tag{8}$$

The decision variable in this formulation are given as:

$$\mathbf{x} = [\theta, E_1, \mathbf{m}]^T$$

Consider a data set,  $\underline{S}_D = \{\bar{S}_1, \dots, \bar{S}_{N_{DS}}\}$ . The optimization cost ( $-\Delta E$ ) is estimated by substituting the  $E(\bar{S}_1)$  as that of ground state and  $E_1$  for the 1st excited state energy. Thus the cost is evaluated as:



$$\text{Cost} = E(\bar{S}_1) - E_1$$

The energy of all data states are explicitly equated as follows:

$$E(\bar{S}_1) - E(\bar{S}_i) = 0, \quad \forall i \in \{2, \dots, N_{DS}\}$$

The 1<sup>st</sup> excited energy,  $E_1$  is estimated by bounding it from above by energies of all the excited states. It is bounded from below by the energy of state corresponding to the index at which  $m_i = 1$ . The upper bound on  $E_1$  insures that if  $m_i = 1$ , then  $E_1(\theta) = E(S_i)$ . Here the states,  $S_i$  (written without a bar) denote all states that are not present in  $S_D$ . These conditions can be imposed using following set of equations and inequality:

$$\begin{aligned} E(S_i) - E_1 + Mm_i &\leq M, & \forall i \in \{1, \dots, N_{ES}\} \\ -E(S_i) + E_1 &\leq 0, & \forall i \in \{1, \dots, N_{ES}\} \\ \sum_{i=1}^{N_{ES}} m_i &= 1 \end{aligned}$$

Most computing software only allows integer valued variables. In such a case, the binary value of variable  $m$  can be explicitly enforced by setting following bounds on integer valued  $m$ :

$$0 \leq m_i \leq 1, \quad \forall i \in \{1, \dots, N_{ES}\}$$

This formulation is presented in Box 1 in the matrix format.

Optimization cost:

$$c = [\varepsilon(\bar{S}_1) \quad -1 \quad \mathbf{0}_{1 \times N_{ES}}]$$

Inequality constraints

$$A = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \varepsilon(S_i) & -1 & [0, \dots, 0, \underbrace{M}_{i^{th} \text{ index}}, 0, \dots, 0]_{1 \times N_{ES}} \\ -\varepsilon(S_i) & 1 & \mathbf{0}_{1 \times N_{ES}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \vdots \\ M \\ 0 \\ \vdots \end{bmatrix}$$

Equality constraints:

$$A_{eq} = \begin{bmatrix} \mathbf{0}_{1 \times N_V} & 0 & \mathbf{1}_{1 \times N_{ES}} \\ \varepsilon(\bar{S}_2) - \varepsilon(\bar{S}_1) & 0 & \mathbf{0}_{1 \times N_{ES}} \\ \vdots & \vdots & \vdots \\ \varepsilon(\bar{S}_{N_{DS}}) - \varepsilon(\bar{S}_1) & 0 & \mathbf{0}_{1 \times N_{ES}} \end{bmatrix}, \mathbf{b}_{eq} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Bounds:

$$\begin{aligned} lb &= [H_1^{\min}, \dots, H_{N_V}^{\min}, J_1^{\min}, \dots, J_{N_C}^{\min}, -M, \mathbf{0}_{1 \times N_{ES}}] \\ ub &= [H_1^{\max}, \dots, H_{N_V}^{\max}, J_1^{\max}, \dots, J_{N_C}^{\max}, M, \mathbf{1}_{1 \times N_{ES}}] \end{aligned}$$

Box 1: MILP formulation for PEPDAS method

## 4.2 Algorithm 2: Parameter estimation for potts model with ground state multiplicity (PEPGSM)

In this formulation only the variable  $N_{GS}$  is provided by the user instead of  $\mathcal{S}_{Data}$ . This condition adds the complexity of locating the ground states and evaluating the ground state energy,  $E_0(\theta)$ . This problem is resolved by including following auxiliary variables:

- $E_0$  (real valued scalar): It represents the ground state energy.
- $\mathbf{l} = [l_1, \dots, l_{N_{TS}}]$  (binary valued vector of size  $N_{TS}$ ): It is defined such that it's value is 1 on exactly  $N_{GS}$  indices and 0 everywhere else. The index has value 1 if and only if it corresponds to the ground state.
- $E_1$  and  $M$  as defined in algorithm 1
- $\mathbf{m} = [m_1, \dots, m_{N_{TS}}]$  (binary valued vector of size  $N_{TS}$ ): It is same as algorithm 1, except that the index are now enumerated based on the set  $\mathcal{S}$

The decision variable in this formulation are given as:

$$\mathbf{x} = [\theta \ E_0 \ E_1 \ \mathbf{l} \ \mathbf{m}]$$

The optimization cost is given as:

$$\text{Cost} = E_0 - E_1$$

The estimation of  $E_0$  is done using the same idea of bounding  $E_0$  from above and below. The bound is tight only for indices where  $l_i = 1$ .

$$\begin{aligned} -E(\mathcal{S}_i) + E_0 &\leq 0, & \forall i \in \{1, \dots, N_{TS}\} \\ E(\mathcal{S}_i) - E_0 + Ml_i &\leq M, & \forall i \in \{1, \dots, N_{TS}\} \\ \sum_{i=1}^{N_{TS}} l_i &= N_{GS} \end{aligned}$$

For the estimation of  $E_1$ , the upper bound is lifted on indices corresponding to ground states. This allows to estimate minimum over non-optimal states. Moreover, index of 1<sup>st</sup> excited state cannot coincide with ground state i.e.  $l_i = 1$  and  $m_i = 1$  cannot occur simultaneously. These conditions are imposed using following inequalities and equations:

$$\begin{aligned} -E(\mathcal{S}_i) + E_1 - Ml_i &\leq 0, & \forall i \in \{1, \dots, N_{TS}\} \\ E(\mathcal{S}_i) - E_1 + Mm_i &\leq M, & \forall i \in \{1, \dots, N_{TS}\} \\ l_i + m_i &\leq 1, & \forall i \in \{1, \dots, N_{TS}\} \\ \sum_{i=1}^{N_{TS}} l_i &= 1 \end{aligned}$$

The condition of binary valued variables is imposed on integer variables as follows:

$$0 \leq l_i, m_i \leq 1, \quad \forall i \in \{1, \dots, N_{ES}\}$$

This formulation is presented in Box 2 in the matrix format.

Optimization cost:

$$c = [\mathbf{0}_{1 \times (N_V + N_C)} \quad 1 \quad -1 \quad \mathbf{0}_{1 \times N_{TS}} \quad \mathbf{0}_{1 \times N_{TS}}]$$

Inequality constraints

$$A = \begin{bmatrix} \mathbf{0}_{1 \times (N_V + N_C)} & 0 & 0 & \mathbf{1}_{1 \times N_{TS}} & \mathbf{1}_{1 \times N_{TS}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\varepsilon(S_i) & 1 & 0 & \mathbf{0}_{1 \times N_{TS}} & \mathbf{0}_{1 \times N_{TS}} \\ \varepsilon(S_i) & -1 & 0 & [0, \dots, 0, \underbrace{M}_{i^{th} \text{ index}}, 0, \dots, 0]_{1 \times N_{TS}} & \mathbf{0}_{1 \times N_{TS}} \\ -\varepsilon(S_i) & 0 & 1 & [0, \dots, 0, \underbrace{-M}_{i^{th} \text{ index}}, 0, \dots, 0]_{1 \times N_{TS}} & \mathbf{0}_{1 \times N_{TS}} \\ \varepsilon(S_i) & 0 & -1 & \mathbf{0}_{1 \times N_{TS}} & [0, \dots, 0, \underbrace{M}_{i^{th} \text{ index}}, 0, \dots, 0]_{1 \times N_{TS}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ M \\ 0 \\ M \\ \vdots \end{bmatrix}$$

Equality constraints:

$$A_{eq} = \begin{bmatrix} \mathbf{0}_{1 \times (N_V + N_C)} & 0 & 0 & \mathbf{1}_{1 \times N_{TS}} & \mathbf{0}_{1 \times N_{TS}} \\ \mathbf{0}_{1 \times (N_V + N_C)} & 0 & 0 & \mathbf{0}_{1 \times N_{TS}} & \mathbf{1}_{1 \times N_{TS}} \end{bmatrix}, \mathbf{b}_{eq} = \begin{bmatrix} N_{GS} \\ 1 \end{bmatrix}$$

Bounds:

$$lb = [H_1^{\min}, \dots, H_{N_V}^{\min}, J_1^{\min}, \dots, J_{N_C}^{\min}, -M, -M, \mathbf{0}_{1 \times N_{TS}}, \mathbf{0}_{1 \times N_{TS}}]$$

$$ub = [H_1^{\max}, \dots, H_{N_V}^{\max}, J_1^{\max}, \dots, J_{N_C}^{\max}, M, M, \mathbf{1}_{1 \times N_{TS}}, \mathbf{0}_{1 \times N_{TS}}]$$

Box 2: MILP formulation for PEPGSM method

## 5 Results and discussions

In this section, an example is presented to show the efficiency of both the algorithms. It is shown by example that the predicted  $\eta$  decays and is bounded. Moreover, the PEPGSM method can predict ground states that provide higher bandgap compared to randomly picked ground states. Next, the computational cost of this method is discussed.

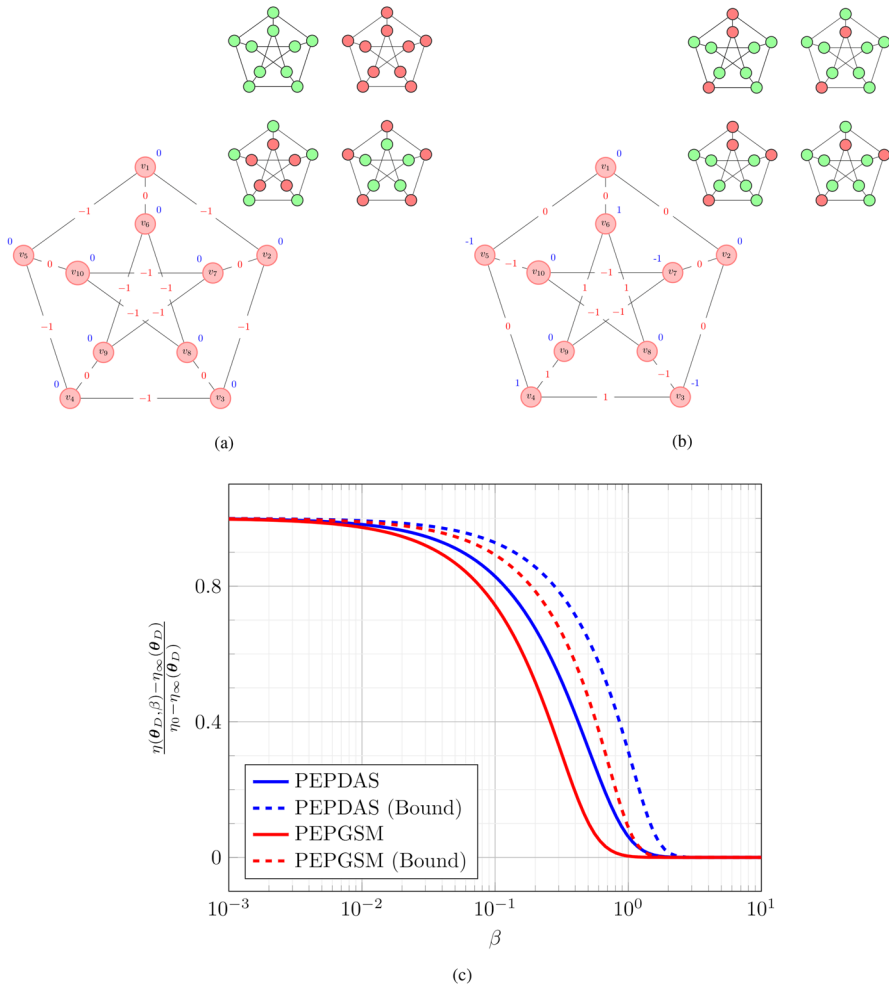
### 5.1 Examples

The parametric estimation of Ising model is presented as an application of this method. In this model, the states take a binary form i.e.  $N_L = 2$ . Traditionally the labels are denoted as  $\{+1, -1\}$  and the corresponding energy functions are defined as:

$$U(+1) = +1, \quad U(-1) = -1$$

$$V(+1, +1) = V(-1, -1) = 1, \quad V(+1, -1) = -1$$

Therefore, the energy can be effectively written as:



**Fig. 3** Optimal Ising parameters of a Peterson graph with 4 ground states found using **a** PEPDAS method, and **b** PEPGSM method. The ground states are presented as the colored graph in the top right corner of each image. A green label denotes the ‘+1’ state, and the red label denotes the ‘-1’ state. **c** The normalized Negative log-likelihood of the optimized graphs

$$E(S) = \sum_{i=1}^{N_V} H_i S_i + \sum_{k=1}^{N_C} J_k S_{\pi(k,1)} S_{\pi(k,2)} \tag{9}$$

This model is applied on a 10-noded Peterson graph with  $|H| \leq 1$  and  $|J| \leq 1$ . First, the graph is trained by prescribing up to 4 data states using the PEPDAS method. Next, the graph is trained by prescribing the number of states from 1 to 4 using the PEPGSM method. The predicted band gaps are shown in Table 1. It can be observed that the PEPGSM method predicts the same bandgap as the PEPDAS method for data sets with a size up to 3. However, for 4 data points, the PEPGSM method can

**Table 1** Predicted maximum band gap for Peterson graph

Algorithm	$N_{GS} = 1$	$N_{GS} = 2$	$N_{GS} = 3$	$N_{GS} = 4$
PEPDAS	8.0	6.0	4.0	6.0
PEPGSM	8.0	6.0	4.0	4.0

identify ground states that provide higher bandgap. The predicted parameters for a graph with four ground states are shown in Fig. 3. Likelihood estimates are not well defined in the case of PEPGSM method as it is not trained using the data. However, for comparison,  $\eta$  is estimated using the set of ground states in place of the data set. The results for negative log-likelihood of the PEPDAS predicted model and PEPGSM predicted model are shown in Fig. 3c. As expected, PEPGSM predicted model performs better than PEPDAS predicted model in terms of the range of  $\beta$  for which they can be used. The details of the other three models are presented in Appendix B.2.

The models for  $N_{GS} = 4$ , are further used in sampling the ground states from the D’Wave emulator (Quantum annealing device/emulator) (Johnson et al. 2011). A single instance of the simulation with 10000 samples gave a distribution of  $p = [0.2543, 0.2480, 0.2526, 0.2451]$  for PEPDAS and  $p = 0.2481, 0.2531, 0.2426, 0.2562]$  for PEPGSM models. In both cases, no state other than the prescribed ground state appeared in the samples. This effect can be attributed to these emulators’ typically low simulation temperature (approx.  $\beta \sim 10$ ). Moreover, due to the degeneracy of the ground state, the sampling probability for each ground state is very close to  $1/N_{GS}^{-1}$

### 5.2 Computation size

One of the limiting features of these algorithms is that it grows exponentially with the graph size. An exact number of variables and equations is provided in Table 2. It should be noted that the number of states,  $N_{TS} = N_L^{N_V}$  and is the reason for the large size of the decision variable. The system of equations and inequalities in both algorithms have large sparse blocks which provide some computational easing. It should also be noted that the sparsity of graph,  $G$ , does not give considerable advantage in the algorithm as the size of the problem is mainly dictated by the number of labels,  $N_T$ , and the number of vertices,  $N_V$ .

**Table 2** Variable size for Algorithms PEPDAS and PEPGSM

Quantity	PEPDAS	PEPGSM
Total variables	$N_V + N_C + 1 + N_{ES}$	$N_V + N_C + 2 + 2N_{TS}$
Integer (Binary) variables	$N_{ES}$	$2N_{TS}$
Inequality conditions	$2N_{ES}$	$4N_{TS} + 1$
Equality conditions	$N_{DS}$	2

## 6 Conclusions

Two algorithms were developed and analyzed for estimating parameters of Potts model. The functionality of each method is as follows:

1. PEPDAS method estimates the parameters to exactly replicate the ground states as the prescribed data set.
2. PEPGSM method estimates the parameters to identify ground states based on their prescribed quantity.

Both algorithms maximize the band gap between the ground and excited states of the model. It was shown that models optimized in this manner have a higher probability of being in the ground state for a broader range of temperatures. The upper bounds on the optimized model's performance are also estimated. This efficiency is measured in terms of the range of temperature for which ground states' likelihood remains in the desired range. The examples included in the paper show promising practical results on small graphs. As suggested in the main body of the paper, these methods do not scale well with the graph size, and their usage should be restricted to small problems. A promising extension of this work to address larger graphs is via successively optimizing parameters of randomly chosen subgraphs using the proposed PEPDAS and PEPGSM methods. This extension, therefore, promises a heuristic method for parameter estimation of large graphs.

## Appendices

The codes are available at <https://github.com/sidsriva/PEP>.

### Appendix 1: Proof of theorem

(a) Since  $\mathcal{S}_G(\theta, \beta) = \mathcal{S}_D$ , the Negative Log Likelihood,  $\eta(\theta_D, \beta)$ , is estimated as:

$$\eta(\theta_D, \beta) = N_{GS}\beta E_0 + N_{GS} \log Z$$

The derivative is estimated as::

$$\frac{d\eta}{d\beta} = N_{GS}((E_0 - \mathbb{E}(E))) \quad (10)$$

where

$$\mathbb{E}(E) = \sum_{S \in \mathcal{S}} E(S)p(S|\theta_D, \beta)$$

Since  $\Delta E > 0$ , the expected energy is strictly bounded below as  $\mathbb{E}(E) > E_0$ . Consequently:

$$\frac{d\eta}{d\beta} < 0$$

In the low temperature limit, Eq. (2) estimates that the probability of all excited states approaches 0 while all ground states are equally likely with probability  $(N_{GS})^{-1}$ . Therefore, the value of  $\eta$  in this limit is estimated as Eq. (4).

(b) Let  $S_G \in \mathcal{S}_G$  and  $P = p(S_G|\theta_D, \beta)$  so that  $\eta(\theta_D, \beta) = -N_{GS} \log P$ . The probability of occurrence of a ground state is given by  $N_{GS}P$  and occurrence of a excited state is given as  $(1 - N_{GS}P)$ . Moreover, for any finite value of  $\beta$  both of these probabilities are finite. Therefore, the expectation of energy,  $\mathbb{E}$ , can be bounded as

$$\mathbb{E} = N_{GS}PE_0 + \sum_{S \in \mathcal{S}_E} E(S)p(S|\theta_D, \beta) \leq N_{GS}PE_0 + (1 - N_{GS}P)E_1$$

Substituting in Eq. (10),

$$\frac{d\eta}{d\beta} = E_0 - \mathbb{E}(E) \leq (N_{GS}P - 1)N_{GS}\Delta E$$

Substituting  $P = e^{-\eta/N_{GS}}$  gives the following differential inequality

$$\frac{d\eta}{d\beta} \leq (N_{GS}e^{-\eta/N_{GS}} - 1)N_{GS}\Delta E \tag{11}$$

Consider the differential equation for  $\beta \in [0, \infty)$ ,

$$\frac{d\xi}{d\beta} = (N_{GS}e^{-\xi/N_{GS}} - 1)N_{GS}\Delta E \tag{12}$$

with initial condition  $\xi(\theta_D, 0) = \eta(\theta_D, 0) = N_{GS} \log N_{TS}$ . Noting that  $N_{GS}e^{-\xi/N_{GS}} - 1 = N_{GS}P - 1 > 0$ , this ODE is integrated to give the following solution:

$$\xi(\theta_D, \beta) = N_{GS} \log (N_{GS} + N_{ES}e^{-\beta\Delta E}) \tag{13}$$

Using Comparison Lemma (Khalil 2002), for all  $0 < \beta < \infty$ ,

$$\eta(\theta_D, \beta) \leq \xi(\theta_D, \beta) \tag{14}$$

This proves the upper bound. The lower bound is a direct consequence from monotonicity proved in part 1.

(c) For any  $\beta < \infty$

$$\eta(\theta_D, \beta) - N_{GS} \log N_{GS} \leq N_{GS} \log \left( 1 + \frac{N_{ES}}{N_{GS}} e^{-\beta\Delta E} \right)$$

For any  $\epsilon > 0$ , choose a  $\beta > \beta^*(\epsilon)$  using Eq. (6) and observe that,

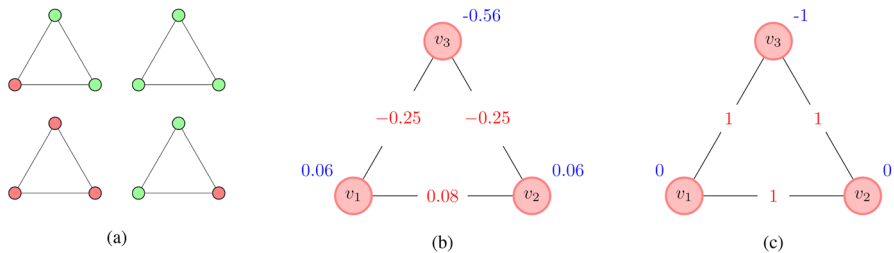
$$N_{GS} \log \left( 1 + \frac{N_{ES}}{N_{GS}} e^{-\beta \Delta E} \right) < \epsilon$$

This proves the third statement.

## Appendix 2: Optimized graphs

### Appendix 2.1: K-3 graph

A fully connected 3-noded graph is optimized for 4 data states. The energy of the graph is modeled using Ising model Eq. (9) with  $|H| \leq 1$  and  $|J| \leq 1$ . The optimized parameters using the (1) Minimization of Negative Log-likelihood, and (2) PEPDAS method are presented in Fig. 4

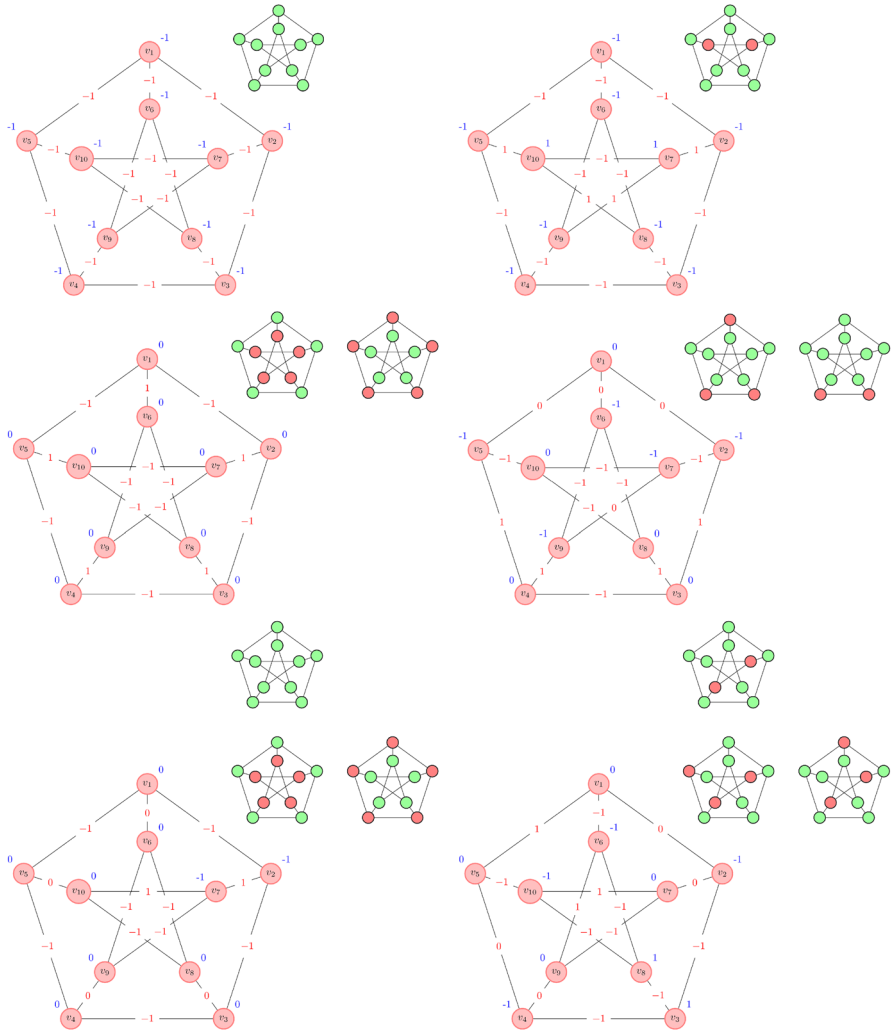


**Fig. 4** **a** Training data set of states with green representing a '+1' state and red representing a '-1' state. **b** Optimized graph using minimization of Negative Log-likelihood at  $\beta = 1$  **c** Optimized graph using PEPDAS method. The field terms are mentioned in blue color and interaction terms are mentioned in red color

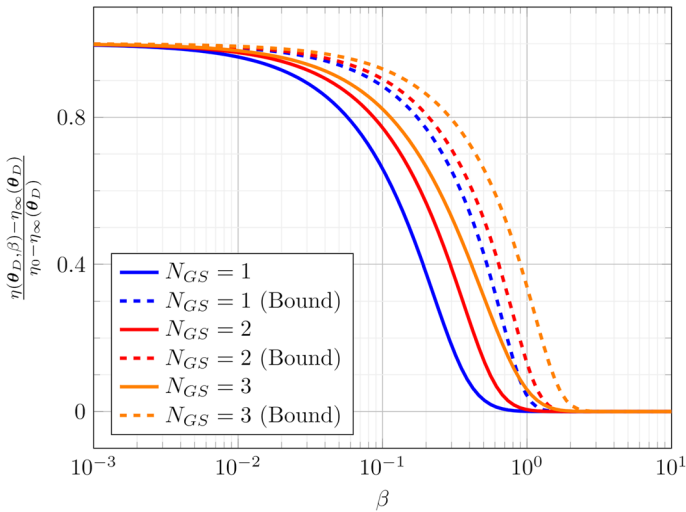
### Appendix 2.2: Peterson graph

A Peterson graph is first optimized for upto 3 user prescribed data states using PEPDAS method. Then it is optimized for 3 ground states using PEPGSM method. The energy of the graph is modeled using Ising model Eq. (9) with  $|H| < 1$  and  $|J| < 1$ . The optimized graphs are presented in Fig. 5 and their respective Negative log likelihood is presented in Fig. 6.





**Fig. 5** Optimal Ising parameters of a Peterson graph found using PEPDAS method (left) and PEPGSM method (right). The ground states are presented as the colored graph in the top right corner of each image. A green label denotes the '+1' state and the red label denotes the '-1' state



**Fig. 6** Normalized Negative log likelihood and their respective bounds for Peterson graphs trained using PEPDAS method

## References

- Graner François, Glazier James A (1992) Simulation of biological cell sorting using a two-dimensional extended potts model. *Phys Rev Lett* 69(13):2013
- Miodownik Mark (2007) Monte carlo potts model. In: Janssens Koenraad G.F., Raabe Dierk, Kozeschnik Ernst, Miodownik Mark A, Nestler Britta (eds) *Computational materials engineering*. Academic Press, Burlington, pp 47–108
- Boykov Yuri, Veksler Olga, Zabih Ramin (2001) Fast approximate energy minimization via graph cuts. *IEEE Trans Pattern Anal Mach Intell* 23(11):1222–1239
- Shai Bagon (2012) Discrete energy minimization, beyond submodularity: Applications and approximations. arXiv preprint [arXiv:1210.7362](https://arxiv.org/abs/1210.7362)
- Descombes Xavier, Morris Robin D, Zerubia Josiane, Berthod Marc (1999) Estimation of Markov random field prior parameters using Markov chain monte Carlo maximum likelihood. *IEEE Trans Image Process* 8(7):954–963
- Hinton Geoffrey E (2002) Training products of experts by minimizing contrastive divergence. *Neural Comput* 14(8):1771–1800
- Asja Fischer and Christian Igel (2012) An introduction to restricted boltzmann machines. In: *Iberoamerican Congress on Pattern Recognition*, Springer, pp. 14–36
- Steven H Adachi and Maxwell P Henderson (2015) Application of quantum annealing to training of deep neural networks. arXiv preprint [arXiv:1510.06356](https://arxiv.org/abs/1510.06356)
- Siddhartha Srivastava, Veera Sundararaghavan (2020) Machine learning in quantum computers via general boltzmann machines: Generative and discriminative training through annealing. arXiv preprint [arXiv:2002.00792](https://arxiv.org/abs/2002.00792)
- Johnson Mark W, Amin Mohammad HS, Gildert Suzanne, Lanting Trevor, Hamze Firas, Dickson Neil, Harris Richard, Berkley Andrew J, Johansson Jan, Bunyk Paul et al (2011) Quantum annealing with manufactured spins. *Nature* 473(7346):194–198
- Hassan K. Khalil (2002) Fundamental properties. In: *Nonlinear systems*, Prentice Hall, pp. 87 – 110.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.