INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

Optimization with the Hopfield network based on correlated noises: an empirical approach

Jacek Mańdziuk*

TR-97-019

May 1997

Abstract

This paper presents two simple optimization techniques based on combining the Langevin Equation with the Hopfield Model. Proposed models - referred as Stochastic Model (SM) and Pulsed Noise Model (PNM) - can be viewed as straightforward stochastic extensions of the Hopfield optimization network.

Optimization with SM, unlike in previous related models, in which δ-correlated Gaussian noises were considered, is based on Gaussian noises with positive autocorrelation times. This is a reasonable assumption from a hardware implementation point of view.

In the other model - PNM, Gaussian noises are injected to the system only at certain time instances, as opposite to continuously maintained δ-correlated noises used in the previous related works.

In both models (SM and PNM), intensities of noises added to the model are *independent* of neurons' potentials. Moreover, instead of impractically long inverse logarithmic cooling schedules, *linear* cooling is tested.

With the above strong simplifications neither SM nor PNM is expected to rigorously maintain Thermal Equilibrium (TE). However, approximate numerical tests based on the canonical Gibbs-Boltzmann distribution show, that differences between rigorous and estimated values of TE parameters are relatively low (within a few percent). In this sense both models are said to perform Quasi Thermal Equilibrium.

*Senior Fulbright Scholar visiting ICSI and EECS Dept. UCBerkeley. On leave from Institute of Mathematics, Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, Poland, e-mail: mandziuk@alpha.im.pw.edu.pl

1 Introduction

In this paper two, simple optimization techniques based on combining the Langevin Equation - based optimization with the Hopfield Model are introduced. Proposed models - referred as Stochastic Model (SM) and Pulsed Noise Model (PNM) - can, in short, be viewed as straightforward stochastic extensions of the Hopfield optimization circuit. Instead of classical Hopfield differential equations, which for a given starting point deterministically describe a trajectory in the search space, new models are defined by stochastic differential equations, obtained by adding a noise term to the Hopfield Model. Similarly to the simulated annealing method [12], a noise term is multiplied by the coefficient (temperature), which decreases in time.

Both models follow the idea of Stochastic Neural Network [13] and Diffusion Machine [21].

Optimization with Stochastic Model, unlike in referred works, in which δ-correlated Gaussian noises were considered, is based on Gaussian noises with positive autocorrelation times. This is a reasonable assumption from the hardware implementation point of view. Unfortunately, theoretical work on SM performance seems to be difficult, since transitions of the stochastic process describing the model are not Markovian.

The paper is mainly focused on comparison between experimental results obtained for three types of Gaussian noises that were tested in computer simulations. Distinctions between noises are based on the relation between the length of the noise autocorrelation time τ, and the RC time constant, which governs the relaxation time of the Hopfield electrical circuit.

In the other model - Pulsed Noise Model, Gaussian noises are injected to the system only at certain time instances, as opposite to continuously maintained δ-correlated noises used in the previous related approaches.

In both models (SM and PNM), intensities of noises added to the system are *independent* of neurons' potentials. Finally, instead of impractically long inverse logarithmic cooling schedules, *linear* cooling is tested.

With the above strong simplifications neither SM nor PNM is expected to rigorously maintain Thermal Equilibrium (TE). However, approximate numerical tests based on the canonical Gibbs-Boltzmann distribution show, that differences between the rigorous and estimated values of TE parameters are relatively low (within a few percent). In this sense both models are said to perform Quasi Thermal Equilibrium.

Both models are tested on the 10-city Travelling Salesman Problem (TSP). Numerical results show that both models solve the small-size TSP efficiently. It should be noted that no effort has been devoted neither to selecting the suitable energy function nor to finding the optimal or sub-optimal set of energy coefficients. Moreover, in both models, improvement is expected with slower cooling schedules.

Results presented in this paper are also presented in part in [15] and [16] for PNM and SM, respectively.

The paper is organized as follows: the next three subsections briefly introduce the background of this work: the Hopfield Model, the Travelling Salesman Problem, and the Langevin (Diffusion) Equation - all of them in the context of solving $NP-Hard$ optimization problems. In Section 2 previous related works are presented and their main conclusions are discussed. The next Section describes Stochastic Model, and presents numerical results for solving TSP and for TE tests. Section 4 covers the description and simulation results for Pulsed Noise Model. Final remarks and conclusions are placed in the last Section.

Notation remark: usually, the term *Stochastic Model* (*SM* in short) will address the *idea* of the optimization method proposed, whereas the plural term *Stochastic Models* (or *SMs*) will represent *various* realizations based on white, moderate or quasi-static noise. The distinction will also be clearly indicated by the context.

1.1 The Hopfield Model

In 1982 Hopfield [9] introduced a neural network model of a Content Addressable Memory (CAM) composed of many, highly-interconnected two-state, McCulloch-Pitts neurons [18]. The subsequent papers described the continuous version of the model, which was composed of the collection of continuous (graded) response neurons. The application domain of the continuous model was either the construction of a CAM [10] or solving combinatorial optimization problems [11].

Since after the classical paper of Hopfield and Tank [11] lots of theoretical and experimental research has been published on this subject, the description of the Hopfield Model (HM) presented in this Section is confined to the minimum indispensable for introducing the notation.

The continuous Hopfield Model composed of N neuronal units, in terms of an electrical circuit *Fig. 1*, is described by the set of differential equations (1):

$$C_i \frac{du_i}{dt} = \sum_{j=0}^{N-1} t_{ij} v_j + I_i - \frac{u_i}{r_i}, \tag{1}$$

where, u_i, v_i denote the input potential and the output potential of the i-th neuron, respectively, $(i = 0, 1, \ldots, N-1)$, I_i is the external input current to the i-th neuron, $t_{ij} = \frac{1}{R_{ij}}$ is a weight from output of neuron j to input of neuron i, $t_{ij} = t_{ji}$, and

$$\frac{1}{r_i} = \frac{1}{R_i} + \sum_{j=0}^{N-1} t_{ij} \tag{2}$$

The input-output relation of the $i-th$ neuron, $v_i = g(u_i)$ is a sigmoidal amplifier's response function:

$$g(x) = \frac{1}{2}(\, 1 + tanh(\alpha x) \,) \tag{3}$$

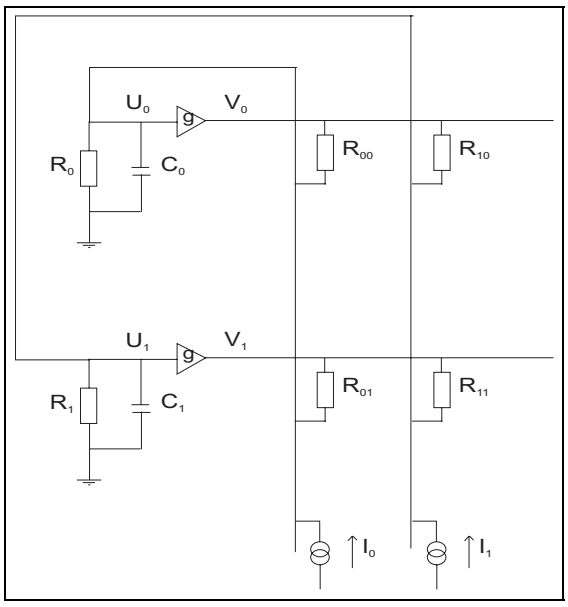


Figure 1: Description of the Hopfield Model as an electrical circuit. For the sake of clarity, network composed of two neurons is presented.

In case $C_i = C_j,\ (i,j = 0,\ldots,N-1)$, if $R_i, R_{ij},\ (i,j = 0,\ldots,N-1)$ are chosen so as $r_i = r_j,\ (i,j = 0,\ldots,N-1)$, after redefining: $t_{ij} := \frac{t_{ij}}{C_i},\ I_i := \frac{I_i}{C_i},\ r_i := R$, $C_i := C,\ (i,j = 0,\ldots,N-1)$ set of equations (1) finally simplifies to:

$$\frac{du_i}{dt} = \sum_{j=0}^{N-1} t_{ij} v_j + I_i - \frac{u_i}{RC}, \tag{4}$$

where RC is the relaxation time of the system.

Solving the given optimization task with the Hopfield Model is based on minimization of the energy function (Lyapunov function for (4)), which is of the following generic form:

$$E = -\frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} t_{ij} v_i v_j - \sum_{i=0}^{N-1} I_i v_i + \frac{1}{RC} \sum_{i=0}^{N-1} \int_0^{v_i} g^{-1}(x) dx \tag{5}$$

For sufficiently high gain α in (3) and in the regime $\frac{du_i}{dt} \longrightarrow 0$, [10, 20] energy (5) reduces to

$$E = -\frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} t_{ij} v_i v_j - \sum_{i=0}^{N-1} I_i v_i \tag{6}$$

Therefore, while solving optimization problem with the Hopfield network one is to choose weights t_{ij} and external inputs I_i so as (global) minima of (6) correspond to (optimal) solutions of the given problem at hand.

Certainly, since

$$\frac{du_i}{dt} = -\frac{\partial E}{\partial v_i} \tag{7}$$

that is the minimization is based on the deterministic gradient descent method, the convergence to the global minimum is not guaranteed. Moreover, the final stable state (the obtained solution) depend on the initial conditions of (4). However, despite some obvious limitations, the Hopfield continuous model (4), as well as two-state model can be successfully used for solving combinatorial optimization problems, especially in situations when finding a good (not optimal) solution is sufficient or for the problems with relatively many global minima, e.g. the N-Queens Problem [14, 17].

1.2 The Travelling Salesman Problem

The Travelling Salesman Problem is one of the standard benchmark problems for evaluation of the optimization methods. There are two main reasons for that: first, the existence of many local minima in the solution space makes the problem really hard, and second: the problem has practical meaning. In case of the HM there are two more important aspects: first, this was the problem originally considered by Hopfield and Tank and by many others afterwards. Secondly, the TSP is particularly bad suited for the gradient-descent method incorporated by the HM.

In this paper a version of the TSP with a full graph K_n on a plane for the problem of size n and symmetric weight matrix T is considered. This means, that there exist a direct connection between any two different cities, and that the length of this connection is independent of the direction of traversing it.

TSP is mapped onto the Hopfield network by a square $n \times n$ matrix V of "nearly binary" elements. Expression "nearly binary" means that gain α in the neuron's activation function is big enough to eventually drive the network towards one of the corners of the $[0,1]^{n\times n}$ hypercube. The solution of the problem is read out from V after binarization of its elements. The same generic form of the energy function for the TSP as the one presented in [11] is used, with only different coefficients: $A = 5, B = 5, C = 10, D = 5, n\_ = n$ and $\alpha = 10$. The set of coefficients is chosen so as to be "reasonable" and no effort has been spent on optimizing coefficients' values.

Namely, the following energy function is used:

$$E = \frac{A}{2}\sum_{x=0}^{n-1}\sum_{i=0}^{n-1}\sum_{j=0,j\neq i}^{n-1} v_{xi}v_{xj} + \frac{B}{2}\sum_{i=0}^{n-1}\sum_{x=0}^{n-1}\sum_{y=0,y\neq x}^{n-1} v_{xi}v_{yi} +$$

$$\frac{C}{2}\Big(\sum_{x=0}^{n-1}\sum_{i=0}^{n-1} v_{xi} - n\Big)^2 + \frac{D}{2}\sum_{x=0}^{n-1}\sum_{y=0,y\neq x}^{n-1}\sum_{i=0}^{n-1} d_{xy}v_{xi}(v_{y,i+1} + v_{y,i-1}), \tag{8}$$

where d_{xy} denotes the distance between city x and city y, and all indices are taken modulo n.

A detailed explanation of the above choice of the energy function as well as a description of a mapping of the problem to matrix V are presented in [11].

Many results published afterwards support the following remarks:

- the choice of the energy form (8) - regardless of coefficient values - is not optimal. More efficient energy forms can be found in [2, 3, 4, 6], to cite only a few papers. Improvement is based on the observation that the global inhibition term in (8), that is the term multiplied by C, tends to average the influence of particular neurons, and in turn forces the system back towards the center of the hypercube $[0,1]^{n \times n}$. The solution for this harmful effect is incorporating the global inhibition term into the inhibition terms for rows and columns, e.g. in the following way

$$\frac{A}{2}\sum_{x=0}^{n-1}(\sum_{i=0}^{n-1}\sum_{j=0,j\neq i}^{n-1} v_{xi}v_{xj} - 1) + \frac{B}{2}\sum_{i=0}^{n-1}(\sum_{x=0}^{n-1}\sum_{y=0,y\neq x}^{n-1} v_{xi}v_{yi} - 1) \tag{9}$$

  and/or adding the term that would push v_{xi} values out from the center of the hypercube, e.g. by adding the following term:

$$\frac{(A+B)}{2}\sum_{x=0}^{n-1}\sum_{i=0}^{n-1} v_{xi}(1 - v_{xi}) \tag{10}$$

  to the energy equation,

- even for the form of energy (8) much improvement can be achieved by a judicious choice of parameters A, B, C and D, which clearly depend on the city-set. Especially the tradeoff between the syntactical constraints (valid tours) and the efficiency of solutions (tour lengths), that is the ratio $\frac{C}{D}$ plays a crucial role in the overall performance of the Hopfield Model.

The network used for our simulations was not optimized in any of the above aspects.

1.3 The Langevin Equation

The Boltzmann Machine [1] combined with the *simulated annealing* technique [12] is a well known model for stochastic optimization over binary variables. The model can be extended to the case of optimization over continuous variables based on the Langevin (Diffusion) Equation:

$$dX(t) = -\nabla E(X(t))dt + \sqrt{2T}dW(t), \tag{11}$$

where $E(\cdot)$ is the function being minimized, $X \in \mathcal{R}^N$, $W(t)$ is the N- dimensional standard Brownian (Wiener) process, and T is the temperature.

Under some conditions on E, $X(t)$ converges weakly to equilibrium with probability density $\pi_T(X)$ given by the Gibbs distribution:

$$\pi_T(X) = \frac{1}{Z(T)} exp\Big(-\frac{E(X)}{T}\Big), \tag{12}$$

where $Z(T)$ is the normalizing constant. For temperatures T near *zero* π_T is concentrated on the global minima of E.

The advantage of the fact that steady-state probability density of (11) is given by the Gibbs-Boltzmann distribution is taken in the optimization algorithm proposed by Geman and Hwang [8]. In [8] it is showed that under some restrictive conditions for E and X, when the minimization is confined to the hypercube $[0,1]^N$ by so-called "reflecting boundaries", and with the inverse logarithmic temperature annealing (cooling) schedule

$$T(t) = \frac{c}{ln(2+t)}, \tag{13}$$

and for "sufficiently large" c, system (11), regardless of starting point $X(0)$, converges weakly with the probability measure to the Gibbs distribution (12). In other words, while maintaining the Thermal Equilibrium, system (11) gradually (as T decreases) converges with probability to the global minimum of E.

This result was extended in [7] to the case of minimization over $\mathcal{R}^N$.

2 Previous related work

Mathematical proof for the ability of the Langevin algorithm to eventually converge to the global minimum forms a strong basis for the research on practical implementation of the Langevin Equation-based minimization methods. One of the enticing possibilities is combining circuit implementation simplicity of the HM with the minimization power of the Langevin Equation - based algorithm. However, the obstacle in a simple combining of the two methods is that the dynamics of the $i-th$ neuron in the HM is governed by $-\frac{\partial E}{\partial v_i}$, whereas in the Langevin Equation is governed by $-\frac{\partial E}{\partial u_i}$, where v_i is a non-linear transformation of u_i.

Thus, as reported in [13], in order to keep up with the Boltzmann law (the Gibbs distribution) the noise has to be injected to the Hopfield Model in the following way:

$$\frac{du_i(t)}{dt} = \sum_{j=0}^{N-1} t_{ij}v_j + I_i - \frac{u_i}{RC} + 2\sqrt{\frac{T}{\alpha(t)}}cosh\Big(\ \alpha(t)u_i(t)\ \Big)\gamma_i(t), \tag{14}$$

where $\gamma_i(t), \gamma_j(t); (i,j = 0,\ldots,N-1, i \neq j)$ are pairwise independent δ-correlated Gaussian noises with intensity (15):

$$Cov[\gamma_i(t), \gamma_i(s)] = \frac{dg^{-1}}{dv_i}\delta(t-s) \tag{15}$$

In such a case, it can be shown that the probability of observing a configuration $V \in [0,1]^N$ at temperature T is given by

$$p_{V,T} = \frac{1}{Z(T)}exp\Big\{-\frac{E(V)}{T}\Big\}, \tag{16}$$

where

$$Z(T) = \int_{[0,1]^N} exp\Big\{-\frac{E(V)}{T}\Big\}\, dV \tag{17}$$

The above combination of the HM and the Langevin algorithm results in, what authors of [13] called, Stochastic Neural Network (SNN), which given enough time converges with probability to the global minimum. The proof of the ability of SNN to maintain TE with the inverse logarithmic cooling (13) presented in [13] is a significant theoretical result. Unfortunately, there are some practical limitations that prevented SNN from straightforward implementation in hardware as well as in computer simulations [5]. Mainly the two of them:

- an extremely slow, inverse logarithmic cooling schedule (13) is required, which makes the implementation ineffective. On the other hand, one may think of application of faster annealing schedules, however in such a case the convergence may not be mathematically guaranteed,

- a suitable scheme for changing α in (14) in time is also required. Coordination of both temperature and gain schedules makes solving (14) a complicated and time-consuming task. This also seems to be a real obstacle in computer simulations [5] and in analogue implementation. In fact, in [13] it is suggested that increase of α in (14) should be done much slower than temperature decrease. This is a necessary condition since *cosh* function should be restricted from a rapid growth.

An idea of slow increase of α was also proposed in [11], in the context of the classical (not stochastic) HM.

The last remark concerning SNN is about the nature of noises in (14). Certainly, δ-correlated noises used for the mathematical proof does not exist in practical situations. Although there is no problem with computer simulations, in any hardware realization the length of the noise autocorrelation time must be positive. This remark does not apply exclusively to SNN, but has a more general meaning - in case of δ-correlated noises, theoretical proofs does not fit tightly to reality. Obviously in practice, noises with less "sharp" characteristics can be successfully used. However in that case, stochastic processes (14) are not Markovian, since the transitions at time t are not independent of the past.

Similar mathematical result concerning the convergence of the combination of HM and LE optimization was independently proved in [21]. In the proof presented in [21] there is no requirement for changing of gain α. This, as mentioned above, significantly simplifies the method. The model presented in [21] is described by the set of equations (18):

$$du_i = -\frac{\partial E}{\partial v_i}dt + \sqrt{\frac{2T}{g'(u_i)}}dW_i \tag{18}$$

and is called Diffusion Machine. The particular choice of g as in (3), leads to (14).

The work presented in this paper was also inspired by ref. [19], where an optoelectronic system performing video-rate simulated annealing is presented and accuracy of its hardware implementation versus computational simulation results is analyzed.

3 Stochastic Model

In order to avoid some implementation limitations of Stochastic Neural Network and Diffusion Machine mentioned in the previous section and to simplify the computational model, SM is defined based on four postulates:

- the noise added to each neuron has positive autocorrelation time τ,
- amplifier's gain α is kept constant (in a high limit),
- the annealing schedule is reasonably fast,
- the noise intensity is independent of the potential of the neuron, that is the noise is injected in the simplest possible way.

In order to address the above postulates, SM is described by the following set of differential equations:

$$\frac{du_i(t)}{dt} = \sum_{j=0}^{N-1} t_{ij} v_j(t) + I_i - \frac{u_i(t)}{RC} + \gamma_i(t) T(t), \tag{19}$$

where $\gamma_i(t)$ denotes a one-dimensional Gaussian noise, and $\gamma_i(t), \gamma_j(t), (i \neq j, i, j = 0, \ldots, N-1)$ are pairwise independent.

Temperature schedule is linear, with a fixed stepsize equal to the inverted number of iterations (denoted by β) in the cooling process,

$$T(t) = T(0)(1 - \frac{t}{\beta}), \tag{20}$$

Instead of δ-correlated noises with intensity (15), Gaussian noises with a non-zero autocorrelation time τ, are proposed. More precisely, at time $t = 0$ and at each time $t = p\tau, (p = 1, 2, \ldots)$, for each neuron $i, (i = 0, \ldots, N-1)$ a noise value $\gamma_i(t)$ is generated (independently from other neurons) according to the Gaussian $N(0, 1)$ distribution. Intensity of the i-th noise $\gamma_i(t)$ during $[(p-1)\tau, p\tau)$ period (in $\frac{\tau}{h}$ iterations, where h is the stepsize of a numerical method used for solving (19)), are set according to the linear change with a step equal to $[\gamma_i(p\tau) - \gamma_i((p-1)\tau)]\frac{h}{\tau}$, that is

$$\gamma_i\Big((p-1)\tau + kh\Big) = \gamma_i\Big((p-1)\tau\Big)(1 - \frac{kh}{\tau}) \; + \; \gamma_i(p\tau)\frac{kh}{\tau}$$

$$k = 0, \ldots, \frac{\tau}{h}, \qquad p = 1, 2, 3, \ldots \tag{21}$$

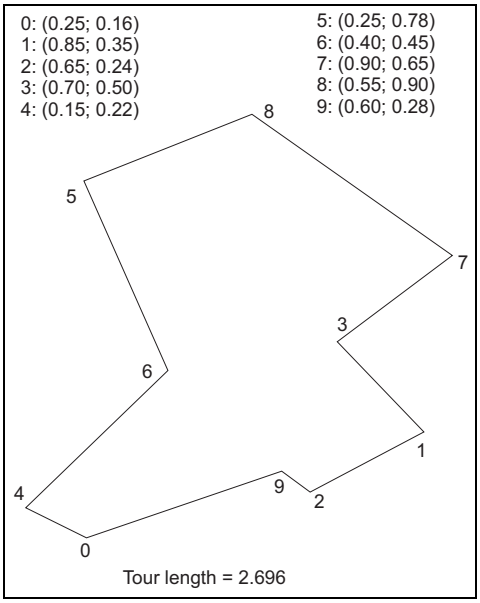


Figure 2: The shortest tour in the exemplar 10-city set. The first, second and third shortest tour lengths are equal to: $2.696, 2.765, 2.767$, respectively

---

Based on relations between time constants of the system, three cases are considered:

- $\tau < RC$ - white noise,
- $\tau = RC$ - moderate noise,
- $\tau > RC$ - quasi-static noise,

where RC is the relaxation time of the model.

3.1 Numerical results

Numerical results are presented in two aspects: quality of tours found (Section 3.1.1) and ability of the system to perform Thermal Equilibrium-like behavior (Section 3.1.2). Computer simulations of SM are presented for the city set depicted in *Fig.* 2. The number of iterations β in (20) is equal to $5 * 10^6$. Parameter β is set based on some number of preliminary tests, however its value cannot be treated as the optimal or suboptimal choice. Obviously, the greater β, the longer the cooling. Hence, value of β used in simulations is set based on a compromise between the time required to perform a test and an average quality of resulting tours.

Unless otherwise stated, in all simulations of Stochastic Model the starting temperature $T(0)$ is set to 100.

3.1.1 Minimization performance

Results of simulations performed for the Hopfield Model and Stochastic Model with white, moderate and quasi-static noise are presented in *Table* 1.

noise type	parameters	best	mean	worst	%failures
no noise		2.69	3.35	3.75	64
white	$\tau = 0.1$	2.69	2.85	3.14	0
moderate	$\tau = 1$	2.69	2.92	3.27	0
quasi-static	$\tau = 10$	2.76	3.25	3.57	0

Table 1. Results of computer simulations for the city-set from *Fig. 2* - Hopfield Model (the first row) and Stochastic Models.

The main quantitative observations are as follows:

- best results were obtained for the white noise case, with the average tour length being 6.07% greater than the shortest tour. Slightly worse tours were obtained for the moderate noise case - 8.69%, and significantly worse for the quasi-static noise case and Hopfield Model - 21.09% and 24.53%, respectively,
- all SMs achieved 100% convergence to valid tours, comparing to 36% of the HM,
- one of the first three shortest tours was found in roughly $35\%, 25\%, 5\%, 5\%$ of trials for Stochastic Model with white, moderate, quasi-static noise, and for the Hopfield Model, respectively.

Certainly, results are preliminary and obtained for the size of data, which is relatively small and does not permit final conclusions. However, in author's opinion, the numerical evidence supports the following general observations:

- all three SMs significantly outperform HM in the number of successful trials and the white and moderate models also in the quality of obtained tours,
- among SMs, the case of white noise and moderate one are significantly better than the quasi-static noise case,
- in both (Stochastic and Hopfield) models improvement is expected by more suitable choice of parameters or more efficient form of the energy function. However, due to a gradient-descent minimization scheme incorporated in HM, its performance cannot be improved significantly,
- experimental results for the Thermal Equilibrium testing presented in the next Section indicate the possibility of a further improvement of SM performance, especially when slower cooling schedules are applied.

3.1.2 Quasi Thermal Equilibrium

The main objective of this work is an experimental analysis of properties of proposed optimization models rather than solving a given TSP problem at hand. Certainly, due to a very straightforward idea of noise injection as well as "inadmissibly" fast cooling schedules, SM is unlikely to rigorously achieve TE.

A theoretical analysis of TE properties of SM is extremely difficult, since due to the positive autocorrelation time of noises, the Markov property does not hold. Hence, the method of examination whether SM exhibits TE - like properties is based on numerical tests. However, in the experimental approach, the huge number of potential binary configurations V, $V \in \{0,1\}^N$, makes direct observation of the Gibbs-Boltzmann distribution infeasible.

Therefore, the approximate method, based on the canonical Gibbs-Boltzmann distribution (22) is used (cf. [19]). In the canonical Gibbs-Boltzmann distribution,

$$p_T = \frac{1}{Z'(T)}\Omega(E)exp\Big\{-\frac{\mu E}{T}\Big\}, \tag{22}$$

$\Omega(E)$ denotes the number of binary configurations V with energies between E and $E + \delta E$, μ is the scaling coefficient, and $Z'(T)$ is the normalizing constant.

For any two temperatures T_i and T_j, $(T_i > T_j)$, the logarithm (23) of the ratio of distributions (22) calculated at T_i and T_j,

$$ln\frac{p_{T_i}}{p_{T_j}} = a(T_i, T_j)\mu E + b(T_i, T_j), \tag{23}$$

where

$$a(T_i, T_j) = \frac{T_i - T_j}{T_i T_j} \tag{24}$$

does not incorporate the unknown distribution $\Omega(E)$. Hence, having distributions collected for various temperature pairs, if the system rigorously holds the canonical Gibbs-Boltzmann distribution, one will expect:

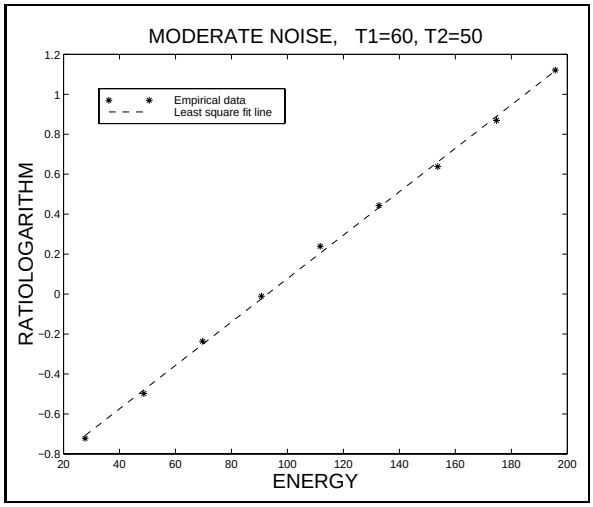
- a linear relation between ratiologarithm $ln\frac{p_{T_i}}{p_{T_j}}$ and energy E according to (23)
- a constant, independent of T_i and T_j, value of μ in (23)

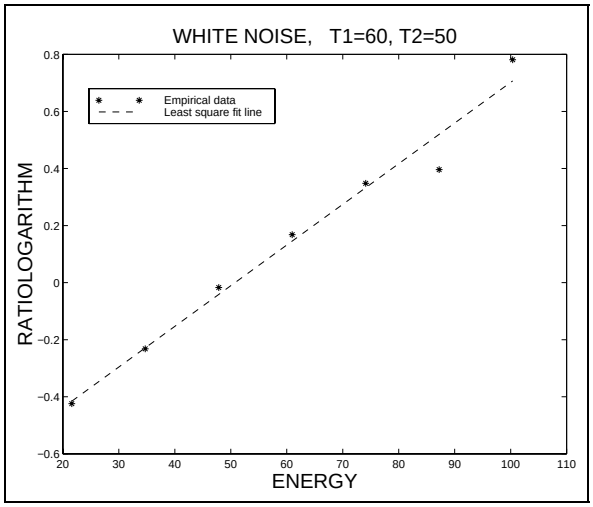
In order to examine the above properties, for various temperatures T, distributions of configurations are collected in the following way:

$(A1)$ system (19) is cooled to T according to (20), and T is fixed,

$(A2)$ $M = 16\ 000$ sample configurations $v \in [0,1]^{n \times n}$ are collected, one at each $p\tau - 1$ time, $(p = 1, \ldots, 16\ 000)$, i.e. right before a noise injection,

(a)





(b)

Figure 3: Quasi-linear relationship between energy intervals and ratiologarithms for temperature pair $(60, 50)$. Stochastic Model with moderate and white noise.

$(A3)$ configurations are binarized (with a threshold equal to 0.5) and for binary configurations $v \in \{0,1\}^{n \times n}$ energy values are calculated from (6).

Having collected, at various temperatures, the data according to $(A1) - (A3)$, for several pairs of "close" temperatures $(T_i, T_j), T_i > T_j$,

$(B1)$ histograms of energy distributions in temperatures T_i and T_j are plotted,

$(B2)$ in the overlapping area of the histograms, the ratiologarithm $ln\frac{p_{T_i}}{p_{T_j}}$ versus E is plotted.

The value of μ, for a given pair (T_i, T_j), is determined in the following way:

$(C1)$ a few last points in $(B1)$, i.e. points representing the highest energy values, are omitted in calculations. This is because of a small number of data in these high-energy categories - only a few samples per category. In such a case the ratio and the logarithm of the ratio are unreliable,

$(C2)$ based on the least-square-fit method the experimental slope $S(T_i, T_j)$ in $(B2)$ is calculated,

$(C3)$ the value of μ for the pair (T_i, T_j), denoted by μ_{T_i,T_j} is calculated as:

$$\mu_{T_i,T_j} = \frac{S(T_i, T_j)}{a(T_i, T_j)} \tag{25}$$

Ideally, if the data were collected under the Boltzmann law, plot in $(B2)$ would be a straight line with a slope equal to $a(T_i, T_j)\mu$ given by (23)-(24). Certainly, "simple" SMs does not rigorously hold TE, however in plots $(B2)$ a quasi-linear relation is observed for all tested temperature pairs. Examples of such relations are presented in $Fig.$ 3. Experimental points in $Fig.$ 3 are denoted by stars. The dashed line represents the least square fit approximation of the experimental data.

Estimated μ_{T_i,T_j} values for the set from $Fig.$ 2, in case of white and moderate noise are presented in $Table$ 2. Results for the quasi-static noise are not reported, because of relatively poorer minimization performance of the model in this case in comparison with the two other cases.

noise type	(T_i, T_j)	μ_{T_i,T_j}	(T_i, T_j)	μ_{T_i,T_j}	(T_i, T_j)	μ_{T_i,T_j}	(T_i, T_j)	μ_{T_i,T_j}
white	$(100, 90)$	4.48	$(90, 80)$	3.75	$(80, 70)$	3.97	$(70, 60)$	3.61
	$(60, 50)$	4.27	$(50, 40)$	3.91	$(40, 30)$	4.15	$(30, 20)$	5.10
moderate	$(100, 90)$	3.33	$(90, 80)$	3.62	$(80, 70)$	3.47	$(70, 60)$	3.33
	$(60, 50)$	3.26	$(50, 40)$	3.29	$(40, 30)$	3.09	$(30, 20)$	3.37

Table 2 Results of testing Thermal Equilibrium properties of Stochastic Model. Estimated μ_{T_i,T_j} values for various pairs (T_i, T_j) in case of white and moderate noise.

Again, if distributions were collected under the Boltzmann law, μ_{T_i,T_j} should have a constant value independent of temperatures of sampling T_i and T_j. Although it is clear from $Table$ 2, that μ is not constant, statistical parameters of sets of μ_{T_i,T_j} values obtained for both noises, presented in $Table$ 3, show that relative deviations of these values are small. In particular, for moderate noise, the relative deviation is less than **5%**. In this sense we conclude that Stochastic Model with white or moderate noise provide the Quasi Thermal Equilibrium.

noise type	$\langle\mu\rangle$	σ_μ	$\sigma_\mu/\langle\mu\rangle$
white	4.16	0.44	10.65%
moderate	3.34	0.14	4.19%

Table 3. Mean value $\langle\mu\rangle$ and standard deviation σ_μ of sets of μ_{T_i,T_j} presented in $Table$ 2.

3.1.3 Technical details

The following comments explain some aspects of the TE testing method presented in the previous Section in more detail.

Straight lines obtained in $(C2)$ based on the least-square-fit method, for all μ_{T_i,T_j} for both noises fit the experimental data very closely.

The coefficient of determination from regression analysis, that is the difference between 1 and the ratio of the residual sum of squares and the total sum of squares, denoted by r^2, has been computed for each choice of (T_i, T_j) for both noises. By definition $r^2 \in [0,1]$, and the greater r^2 the closer linear dependence of the data. In particular, $r^2 = 1$ means perfect match between the experimental data and the estimated straight line.

In the experiment coefficient r^2, *for all tested pairs* (T_i, T_j), exceeds 0.97 and 0.91 for moderate and white noise, respectively (cf. *Fig.* 3).

A number of points omitted in calculations of $\mu_{T_i|T_j}$ in $(C1)$ depends on two factors: the size of a common area of the histograms calculated for temperatures T_i and T_j, or in some sense how close the temperatures are to one another, and on the distribution of samples collected in the highest energy intervals. In any case, in each test the number of samples left out is smaller than 3.4%, and is mainly caused by the relative shift of the histograms.

Besides the results presented in the previous section, some experiments have been performed on the same set of cities but with other $T(0)$. The general conclusion is that unless sampling temperatures are relatively too high, results remain qualitatively the same.

For example, in moderate noise case, for $T(0) = 500$ and temperature pairs $(500, 440)$, $(440, 380)$, $(380, 320)$, $(320, 260)$ similar results were obtained as for pairs $(100, 90)$, $(90, 80)$, ..., $(40, 30)$, $(30, 20)$. Namely, $\langle\mu\rangle = 3.07$, $\sigma_\mu = 0.1$ and $\sigma_\mu/\langle\mu\rangle = 3.26\%$.

In another test, sampling was performed at temperatures $100, 90, \ldots, 20$, but with $T(0) = 500$, that is initial cooling phase was longer. Again, results remained qualitatively the same.

In all tests, β was scaled, so as the temperature decrement remained constant (the higher $T(0)$, the greater β).

4 Pulsed Noise Model

The other stochastic extension of the Hopfield Model considered in this paper is Pulsed Noise Model. PNM is defined by the same set of differential equations as SM, namely (19), where as previously $T(t)$ is decreasing linearly according to (20). The difference between PNM and SM is in the characteristics of noises added to the system. In PNM, at each time $t = p\tau$, $(p = 0, 1, 2, \ldots)$, for each neuron i, a noise value $\gamma_i(t)$ is generated independently from other neurons, according to Gaussian $N(0,1)$ distribution. In the period between $p\tau$ and $(p+1)\tau$ intensity of noise $\gamma_i(t)$ is equal to *zero* - there is no noise in the system. Consequently, between successive $p\tau$ times, $(p = 0, 1, 2, \ldots)$, system (19) performs a gradient descent relaxation.

The motivation behind PNM is similar to the one for SM:

- straightforward noise injection,
- fast cooling,
- fixed amplifier's gain.

Moreover,

- higher stability of the model, because of the lesser amount of noise injected,
- cheaper implementation, since the noise is injected only at certain time instances, as opposite to continuously maintained noise.

Simulation results for PNM, for the set from *Fig.* 2 are presented in *Table* 4. Parameter β in (20) is equal to 10^7. Value of β is chosen based on some number of preliminary tests, and - as in SM case - represents a tradeoff between the average duration of a test and the average quality of results. Starting temperature $T(0)$ is equal to 10 000.

Parameters	Best	Mean	Worst	%Failures
$\tau = RC = 1$	2.69	2.95	3.10	0

Table 4. Results of numerical simulations of Pulsed Noise Model for the city set from *Fig.* 2.

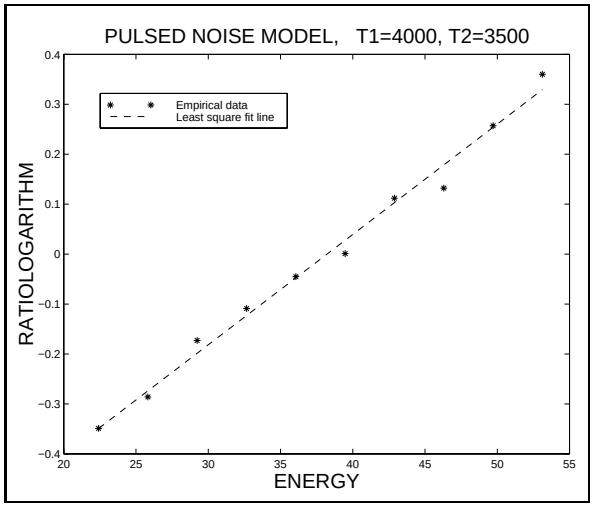
Results presented in *Table* 4 show that even with linear cooling PNM is efficient in solving small-size TSP. Valid salesman tours are obtained in all tests. Average tour length is about 9.6% greater than the optimal one.

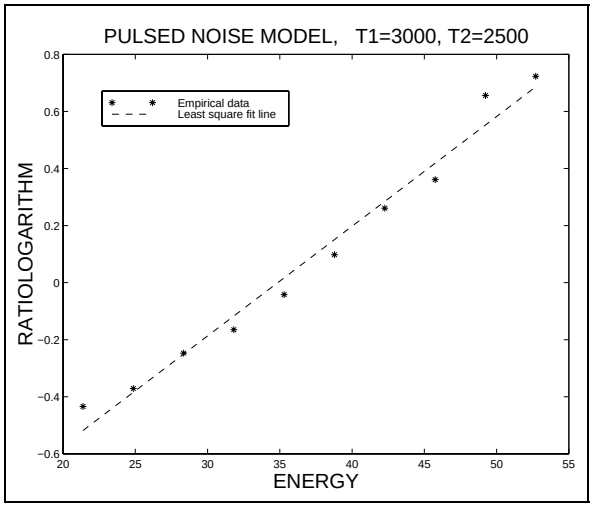
Similarly to SM case, improvement is expected with slower cooling schedules, optimized energy form and/or problem-dependent tuning of energy coefficients.

PNM at first glance seems to not differ from a sequence of trials of HM, in which the final state of the previous test serves as the starting point of the next one, and additional noise $\gamma_i(t)$ is injected to the starting point at each trial. The difference is based on the fact that PNM, after each noise injection, performs exactly $\frac{\tau}{h}$ relaxation steps - which for high temperatures does not guarantee reaching the local minimum.

In PNM, as temperature decreases, the "chain" of configurations v gradually approaches a neighborhood of a deep minimum in the energy surface. As temperature decreases PNM becomes stack in this "good" neighborhood. Along with further decreasing of temperature, PNM becomes more and more similar to the multiple-run HM, and finally reaches the minimum.

(a)





(b)

Figure 4: Pulsed Noise Model. Quasi-linear relationship between energy intervals and ratiologarithms for temperature pairs $(4000, 3500)$ and $(3000, 2500)$.

4.1 Experimental testing of Thermal Equilibrium properties

The same method as the one described in Section 3.1.2 is used to check the ability of PNM to maintain Quasi Thermal Equilibrium.

First, distributions are collected at various temperatures according to $(A1)-(A3)$. Then, ratiologarithm $(B1)-(B2)$ is plotted. Finally, experimental μ values are calculated based on $(C1)-(C3)$.

Sampling temperatures are equal to $5000, 4500, 4000, \ldots, 2000$ combined into pairs $(5000, 4500)$, $(4500, 4000), \ldots, (2500, 2000)$.

For all tested temperature pairs, in plots $(B2)$ a quasi-linear relation is observed. Actually, for all tested pairs, the estimated line in $(C2)$ fits the experimental data $(B2)$ very closely. The coefficient of determination r^2 described in Section 3.1.3 is greater than 0.96 *for all temperature pairs.*

Exemplar plots of ratiologarithm versus energy for $(T_1, T_2) = (4000, 3500)$ and $(T_1, T_2) = (3000, 2500)$ are presented in *Fig.* 4.

Experimental μ_{T_i,T_j} values calculated in the experiment are presented in *Table* 5. The mean value $\langle\mu\rangle$ and the standard deviation σ_μ are equal to 584.89 and 27.60, resp. Therefore, the relative standard deviation $\frac{\sigma_\mu}{\langle\mu\rangle}$ is equal to 4.71%.

On the same basis as for SM we conclude that PNM performs Quasi Thermal Equilibrium.

T_i, T_j	μ_{T_i,T_j}	T_i, T_j	μ_{T_i,T_j}	T_i, T_j	μ_{T_i,T_j}
5000, 4500	593.7	4500, 4000	601.0	4000, 3500	618.7
3500, 3000	588.9	3000, 2500	576.6	2500, 2000	530.1

Table 5. Estimated μ_{T_i,T_j} values for various (T_i, T_j).

4.2 Technical details

Thermal Equilibrium tests for SM and PNM were performed in different temperature ranges. The reason for that is the different way of adding noise to both models. In PNM noise is not maintained continuously, and therefore its intensity is higher than in SM case.

Certainly, temperature has a relative meaning and the proper range of temperature values also depends on the choice of the annealing schedule.

The other problem associated with a proper choice of temperature ranges is that sampling temperature should be neither too high nor too low. At high temperature nothing interesting happens in the system. Distributions for "close" temperatures are very similar (almost identical). On the other hand, for very low temperatures the system is already stuck in the neighborhood of a minimum (final solution).

The crucial temperature range is somewhere in between. The choice of temperature ranges used in simulations was guided by the two above extreme possibilities.

5 Final remarks

Stochastic Model and Pulsed Noise Model presented in this paper are simple stochastic modifications of the deterministic Hopfield Model. The main advantage of these approaches is their simplicity and implementation feasibility. Unlike in the previous related works regarding Stochastic Neural Networks [13] and Diffusion Machine [21], in SM and PNM, intensities of Gaussian noises injected to the system are independent of neurons' potentials.

Moreover, instead of impractically long inverse logarithmic cooling schedules, the linear cooling is tested.

Additionally, Stochastic Model is based on a reasonable assumption that the length of autocorrelation time of the noise must be positive.

Finally, the advantage of Pulsed Noise Model is that noise is injected only at certain time instances, which makes implementation cheaper in terms of the amount of energy required to maintain the noise.

Definitely, with the above strong simplifications neither SM nor PNM is expected to rigorously maintain Thermal Equilibrium. However, approximate numerical tests based on the canonical Gibbs-Boltzmann distribution show, that differences between

the rigorous and estimated TE parameters are relatively low (within a few percent). In this sense both models are said to perform Quasi Thermal Equilibrium.

Simulation results for the 10-city TSP show that, for the size of problem considered, the above inaccuracies are negligible, and both models perform very well even with "inadmissibly" fast, linear cooling.

Current work is focused on supporting experimental results with theoretical analysis and on numerical tests for bigger-size TSP.

Acknowledgments

This work was supported by the Tempus grant no. IMG-95-PL-1014 and the Fulbright grant no. 20895.

Special thanks are due to prof. Eugene Wong from UC Berkeley for invaluable discussions and comments about this work and about the nature of stochastic optimization processes, and to dr. Philippe Lalanne from IOTA for inspiration for this research. I would also like to thank prof. Bernard C. Levy from UC Davis, and James Beck and Brian E. D. Kingsbury from ICSI for helpful comments.

References

[1] D. H. Ackley, G. W. Hinton and T. J. Sejnowski, "A learning algorithm for Boltzmann Machines", Cognitive Sci., 1985, 9, pp. 147-169

[2] Y. Akiyama, A. Yamashira, M. Kajiura, Y. Anzai and H. Aiso, "The Gaussian Machine: A Stochastic Neural Network for Solving Assignment Problems", Journal of Neural Network Computing, 1991, Winter, pp. 43-51

[3] A. R. Bizzarri, "Convergence properties of a modified Hopfield-Tank model", Biological Cybernetics, 1991, 64, pp. 293-300

[4] R. D. Brandt, Y. Wang, A. J. Laub and S. K. Mitra, "Alternative Networks for Solving the Travelling Salesman Problem and the List-Matching Problem", Intern. Conf. Neural Networks, San Diego, 1988, pp. 333-340

[5] J. H. Cervantes and R. R. Hildebrant, "Comparison of three neuron- based computation schemes", IEEE Conf. on Neural Networks, San Diego, USA, 1987, III-657:671

[6] L. Chen and K. Aihara. Chaotic Simulated Annealing by a Neural Network Model with Transient Chaos, *Neural Networks*, 8(6):915-930, 1995

[7] T-S. Chiang, C-R. Hwang and S-J. Sheu, "Diffusion for global optimization in $\mathcal{R}^n$ ", SIAM J. Control and Optimization 1987, 25(3) pp. 737-753

[8] S. Geman and C-R. Hwang, "Diffusions for global optimization", SIAM J. Control and Optimization 1986, 24(4) pp. 1031-1043

[9] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", Proc. Natl. Acad. Sci. USA, 1982, 79, pp. 2554-2558

[10] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons", Proc. Natl. Acad. Sci. USA, 1984, 81, pp. 3088-3092

[11] J. J. Hopfield and D. W. Tank, "Neural Computation of Decisions in Optimization Problems", Biol. Cyber., 1985, 52, pp. 141-152

[12] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by Simulated Annealing", Science, 1983, 220, pp. 671-680

[13] B. C. Levy and M. B. Adams, "Global optimization with stochastic neural networks", IEEE Conf. on Neural Networks, San Diego, USA, 1987, III-681:689

[14] J. Mańdziuk, "Solving the N-Queens Problem with a binary Hopfield-type network. Synchronous and asynchronous model", Biological Cybernetics, 1995, 72, pp. 439-446

[15] J. Mańdziuk, "Pulsed noise - based stochastic optimization with the Hopfield model", Proceedings of the IEEE International Conference on Neural Networks (ICNN'97), Houston, Texas, USA, June 1997, *to appear*

[16] J. Mańdziuk, "Non-delta-correlated Gaussian Noises and the Hopfield Optimization Circuit: an Empirical Study", Proceedings of the 2nd International Conference on Computational Intelligence and Neuroscience (ICCIN'97), Research Triangle Park, North Carolina, USA, March 1997, vol. 2: 110-113

[17] J. Mańdziuk and B. Macukow, "A neural network designed to solve the N-Queens Problem", Biological Cybernetics, 1992, 66, pp. 375-379

[18] W. A. McCulloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity", Bulletin of Mathematical Biophysics, 1943, 5, pp. 115-133

[19] D. Prevost, Ph. Lalanne, J. C. Rodier, P. Chavel, A. Dupret E. Belhaire and P. Garda, "Video-rate Simulated Annealing for stochastic artificial retinas", Optics Communication, 1996, 132, pp. 427-431

[20] M. J. S. Smith and C. L. Protman, "Practical design and analysis of a simple "neural" optimization circuit", IEEE Trans. Circiuts and Systems, 1989, 36, pp. 42-50

[21] E. Wong, "Stochastic neural networks", Algorithmica, 1991, 6, pp. 466-478