Simulated Annealing

A Basic Introduction

Prof. Olivier de Weck

Massachusetts Institute of Technology

Dr. Cyrus Jilla

Outline

- Heuristics
- Background in Statistical Mechanics
  - Atom Configuration Problem
  - Metropolis Algorithm
- Simulated Annealing Algorithm
- Sample Problems and Applications
- Summary

Learning Objectives

- Review background in Statistical Mechanics: configuration, ensemble, entropy, heat capacity
- Understand the basic assumptions and steps in Simulated Annealing
- Be able to transform design problems into a combinatorial optimization problem suitable to SA
- Understand strengths and weaknesses of SA

Heuristics

What is a Heuristic?

- A Heuristic is simply a rule of thumb that hopefully will find a good answer.
- Why use a Heuristic?
  - Heuristics are typically used to solve complex (large, nonlinear, nonconvex (ie. contain many local minima)) multivariate combinatorial optimization problems that are difficult to solve to optimality.
- Unlike gradient-based methods in a convex design space, heuristics are NOT guaranteed to find the true global optimal solution in a single objective problem, but should find many good solutions (the mathematician's answer vs. the engineer's answer)
- Heuristics are good at dealing with local optima without getting stuck in them while searching for the global optimum.

Types of Heuristics

- Heuristics Often Incorporate Randomization

- **2 Special Cases of Heuristics**
  - Construction Methods
    - Must first find a feasible solution and then improve it.
  - Improvement Methods
    - Start with a feasible solution and just try to improve it.

- **3 Most Common Heuristic Techniques**
  - Genetic Algorithms
  - Simulated Annealing
  - Tabu Search
  - New Methods: Particle Swarm Optimization, etc…

Origin of Simulated Annealing (SA)

- **Definition:** A heuristic technique that mathematically mirrors the cooling of a set of atoms to a state of minimum energy.
- **Origin:** Applying the field of Statistical Mechanics to the field of Combinatorial Optimization (1983)
- Draws an **analogy** between the cooling of a material (search for minimum energy state) and the solving of an optimization problem.
- **Original Paper** Introducing the Concept
  - Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., "Optimization by Simulated Annealing," *Science*, Volume 220, Number 4598, 13 May 1983, pp. 671-680.

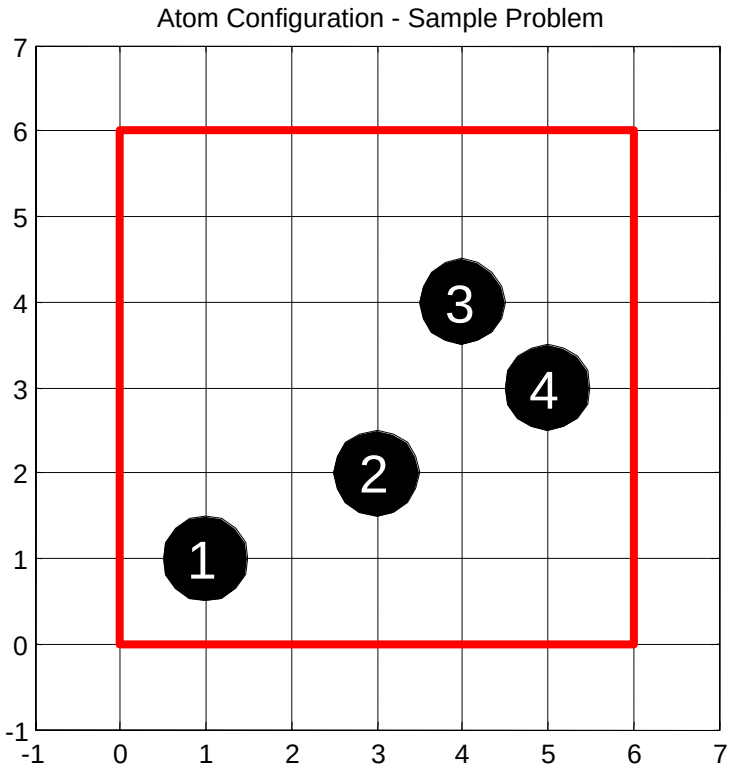
Statistical Mechanics

The Analogy

- **Statistical Mechanics:** The behavior of systems with many degrees of freedom in thermal equilibrium at a finite temperature.
- **Combinatorial Optimization:** Finding the minimum of a given function depending on many variables.
- **Analogy:** If a liquid material cools and anneals too quickly, then the material will solidify into a sub-optimal configuration. If the liquid material cools slowly, the crystals within the material will solidify optimally into a state of minimum energy (i.e. ground state).
  - This ground state corresponds to the minimum of the cost function in an optimization problem.

Sample Atom Configuration

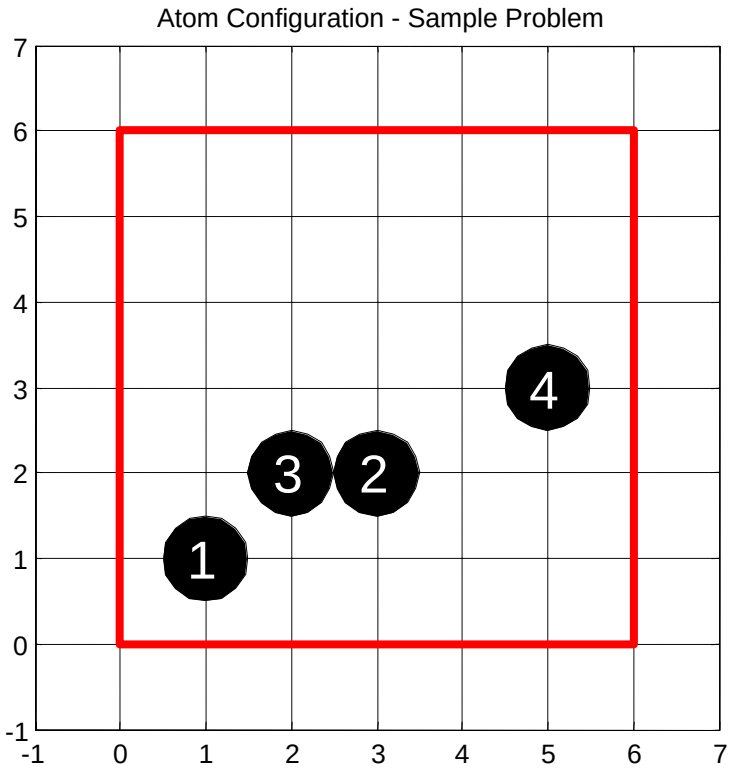
Original Configuration



E=133.67

Energy of original (configuration)

Perturbed Configuration



E=109.04

Perturbing = move a random atom to a new random (unoccupied) slot

Configurations

- Mathematically describe a configuration
  - Specify coordinates of each “atom”

$$\{r_i\} \text{ with i=1,..,4} \quad r_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, r_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, r_3 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, r_4 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

  - Specify a slot for each atom

$$\{r_i\} \text{ with i=1,..,4} \quad R = \begin{bmatrix} 1 & 12 & 19 & 23 \end{bmatrix}$$

Energy of a state

- Each state (configuration) has an energy level associated with it

$$H(q,\dot{q},t) = \sum_i \dot{q}_i p_i - L(q,\dot{q},t)$$

Hamiltonian

$$H = T + V = E_{tot}$$

kinetic potential

Energy

Energy of configuration
=
Objective function value of design

Energy sample problem

- Define energy function for “atom” sample problem

$$E_i = \underbrace{mgy_i}_{\text{potential energy}} + \underbrace{\sum_{j \neq i}^{N} \left(\left(x_i - x_j \right)^2 + \left(y_i - y_j \right)^2 \right)^{\frac{1}{2}}}_{\text{kinetic energy}}$$

$$E(R) = \sum_{i=1}^{N} E_i(r_i)$$

Absolute and relative position of each atom contributes to Energy

Compute Energy of Config. A

- Energy of initial configuration

$$E_1 = 1 \cdot 10 \cdot 1 + \sqrt{5} + \sqrt{18} + \sqrt{20} = 20.95$$

$$E_2 = 1 \cdot 10 \cdot 2 + \sqrt{5} + \sqrt{5} + \sqrt{5} = 26.71$$

$$E_3 = 1 \cdot 10 \cdot 4 + \sqrt{18} + \sqrt{5} + \sqrt{2} = 47.89$$

$$E_4 = 1 \cdot 10 \cdot 3 + \sqrt{20} + \sqrt{5} + \sqrt{2} = 38.12$$

Total Energy Configuration A: $E(\{r_A\}) = 133.67$

Boltzmann Probability

$$N_R = \frac{P!}{(P-N)!}$$

Number of configurations

P=# of slots=25
N=# of atoms =4

N_R=6,375,600

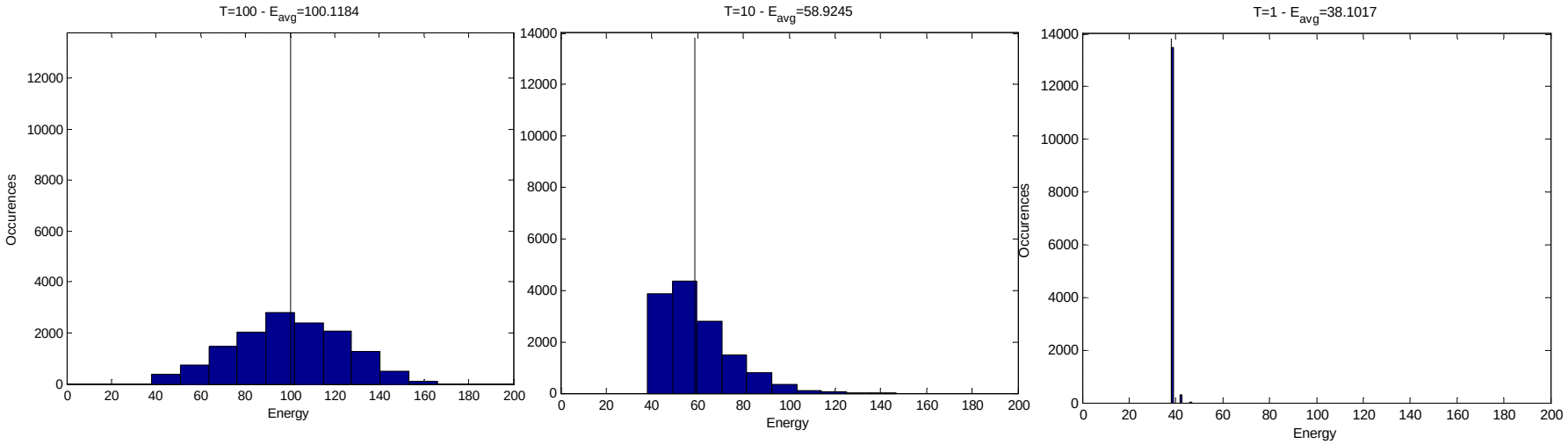
What is the likelihood that a particular configuration will exist in a large ensemble of configurations?

$$P(\{r\}) = \exp\left[\frac{-E(\{r\})}{k_B T}\right]$$

Boltzmann probability depends on energy and temperature

Boltzmann Collapse at low T

T=100 T=10 T=1



T high → T low

Boltzmann Distribution collapses to the lowest energy state(s) in the limit of low temperature

⇨ Basis of search by Simulated Annealing

Partition Function *Z*

- Ensemble of configurations can be described statistically
- Partition function, Z, generates the ensemble average

$$Z = \mathrm{Tr}\exp\left(\frac{-E_i}{k_B T}\right) = \sum_{i=1}^{N_R} \exp\left(\frac{-E_i}{k_B T}\right)$$

Initially (at T>>0) equal to the number of possible configurations

Free Energy

$$E_{avg} = \overline{E}(T) = \sum_{i=1}^{N_R} E_i(T)$$ ← Average Energy of all Configurations in an ensemble

$$F(T) = -k_B T \ln Z = \overline{E}(T) - TS$$

$$E_{avg} = \overline{E}(T) = \frac{\sum_{i=1}^{N_R} \exp\left(-\frac{E_i(T)}{k_B T}\right) E_i(T)}{Z} = \frac{-d \ln Z}{d\left(1/k_B T\right)}$$

Relates average Energy at T with Entropy S

Specific Heat and Entropy

- Specific Heat

$$C(T) = \frac{d\bar{E}(T)}{dT} = \frac{\left[\frac{1}{N_R}\sum_{i=1}^{N_R} E_i^2(T) - \bar{E}(T)^2\right]}{k_B T^2} = \frac{\overline{E^2(T)} - \bar{E}(T)^2}{k_B T^2}$$

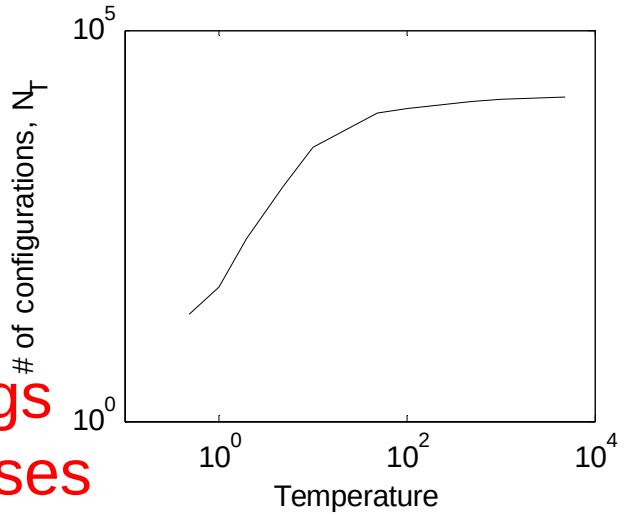
- Entropy

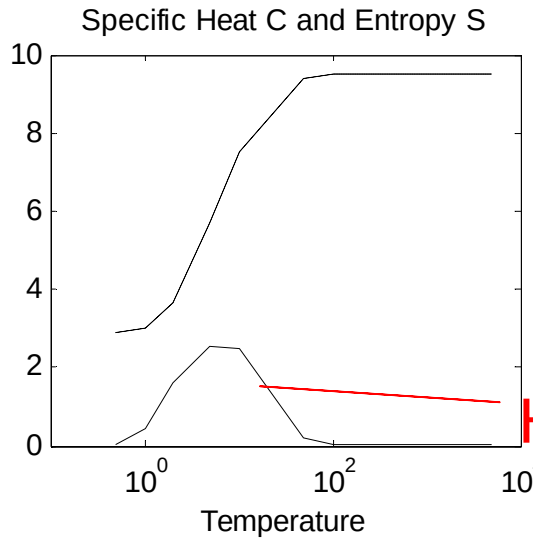
$$S(T) = S(T_1) - \int_T^{T_1} \frac{C(T)}{T} dT \qquad \frac{dS(T)}{dT} = \frac{C(T)}{T}$$

Entropy is ~ equal to ln(# of unique configurations)

Low Temperature Statistics

configs decreases

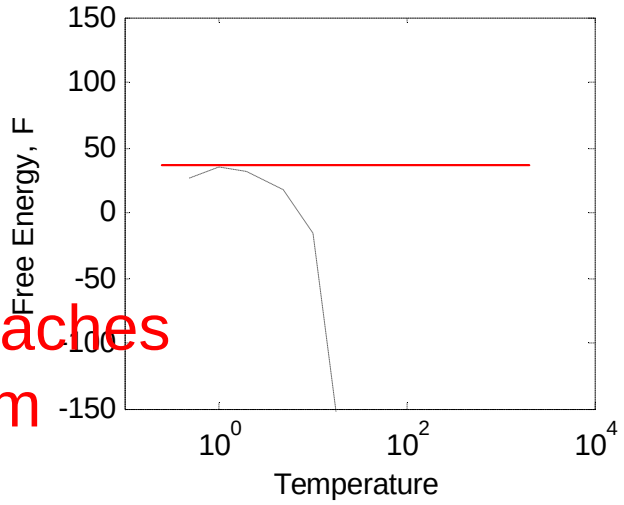


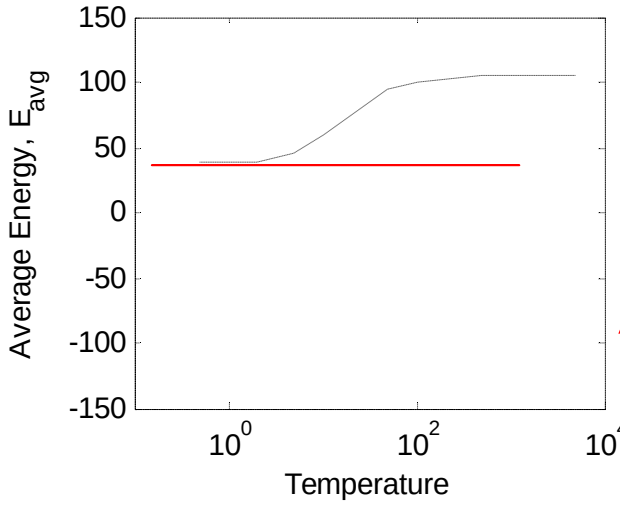


Entropy S decreases with lower T

Ht. Capacity

Approaches E* from below



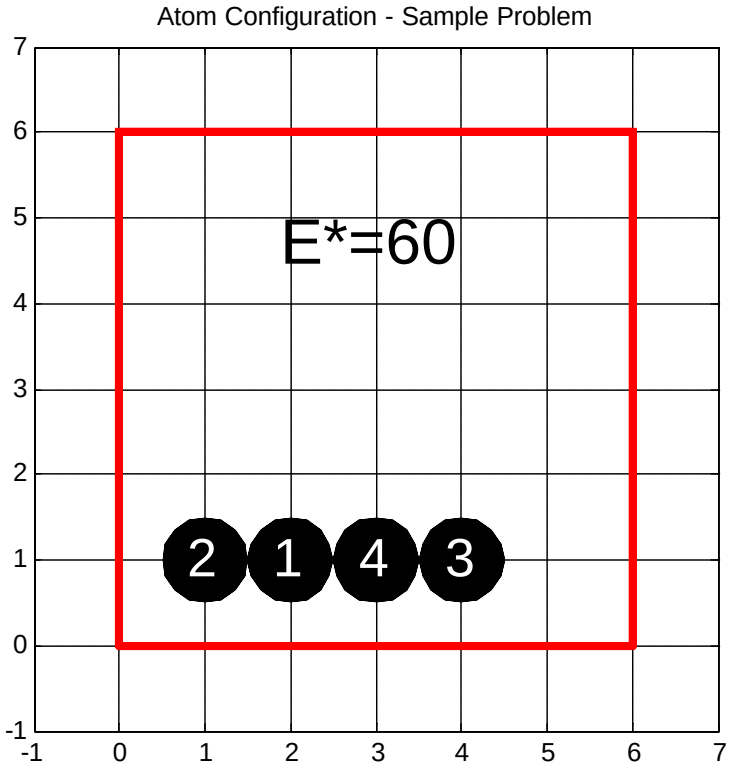


Approaches E* from above

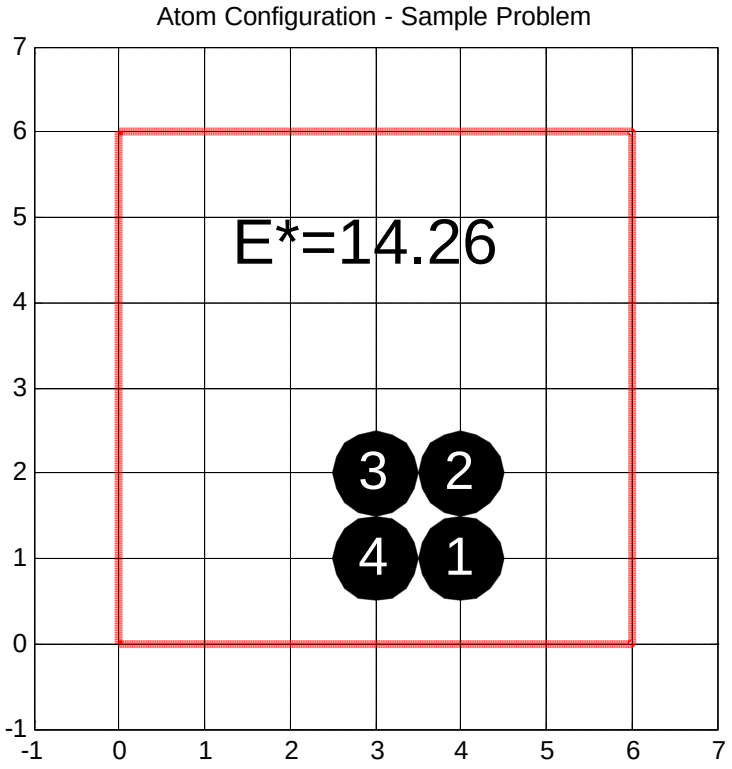
Minimum Energy Configurations

- Sample Atom Placement Problem

Uniqueness?



Strong gravity g=10



Low gravity g=0.1

Simulated Annealing

Dilemma

- Cannot compute energy of all configurations !
  - Design space often too large
  - Computation time for a single function evaluation can be large
- Use Metropolis Algorithm, at successively lower temperatures to find low energy states
  - Metropolis: Simulate behavior of a set of atoms in thermal equilibrium (1953)
  - Probability of a configuration existing at T → Boltzmann Probability $P(r,T)=\exp(-E(r)/T)$

The SA Algorithm

- **Terminology:**
  - X (or R or Γ) = Design Vector (i.e. Design, Architecture, Configuration)
  - E = System Energy (i.e. Objective Function Value)
  - T = System Temperature
  - Δ = Difference in System Energy Between Two Design Vectors

- **The Simulated Annealing Algorithm**

  1) Choose a random X_i, select the initial system temperature, and specify the cooling (i.e. annealing) schedule

  2) Evaluate $E(X_i)$ using a simulation model

  3) Perturb X_i to obtain a neighboring Design Vector (X_{i+1})

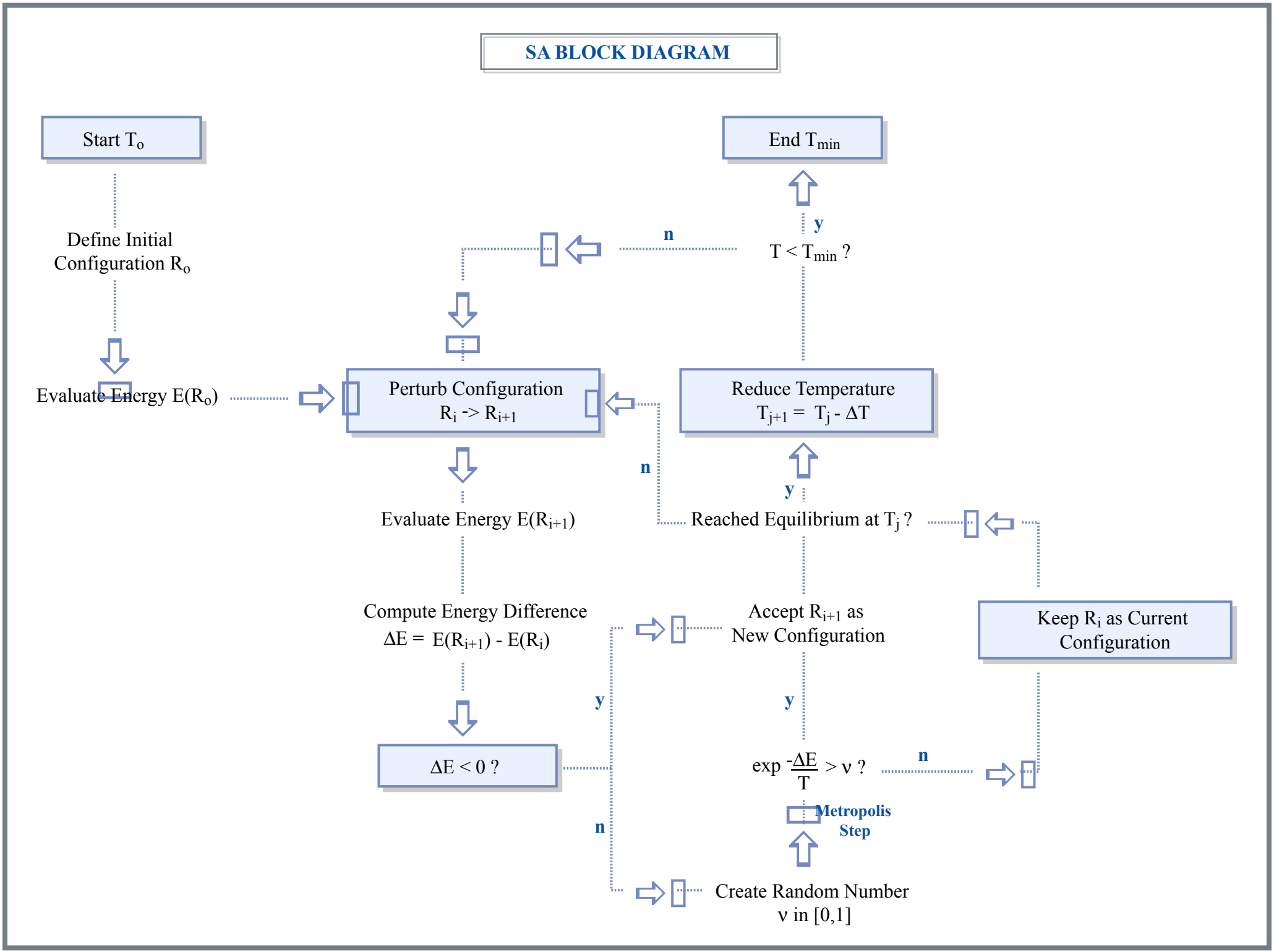
  4) Evaluate $E(X_{i+1})$ using a simulation model

  5) If $E(X_{i+1}) < E(X_i)$, X_{i+1} is the new current solution

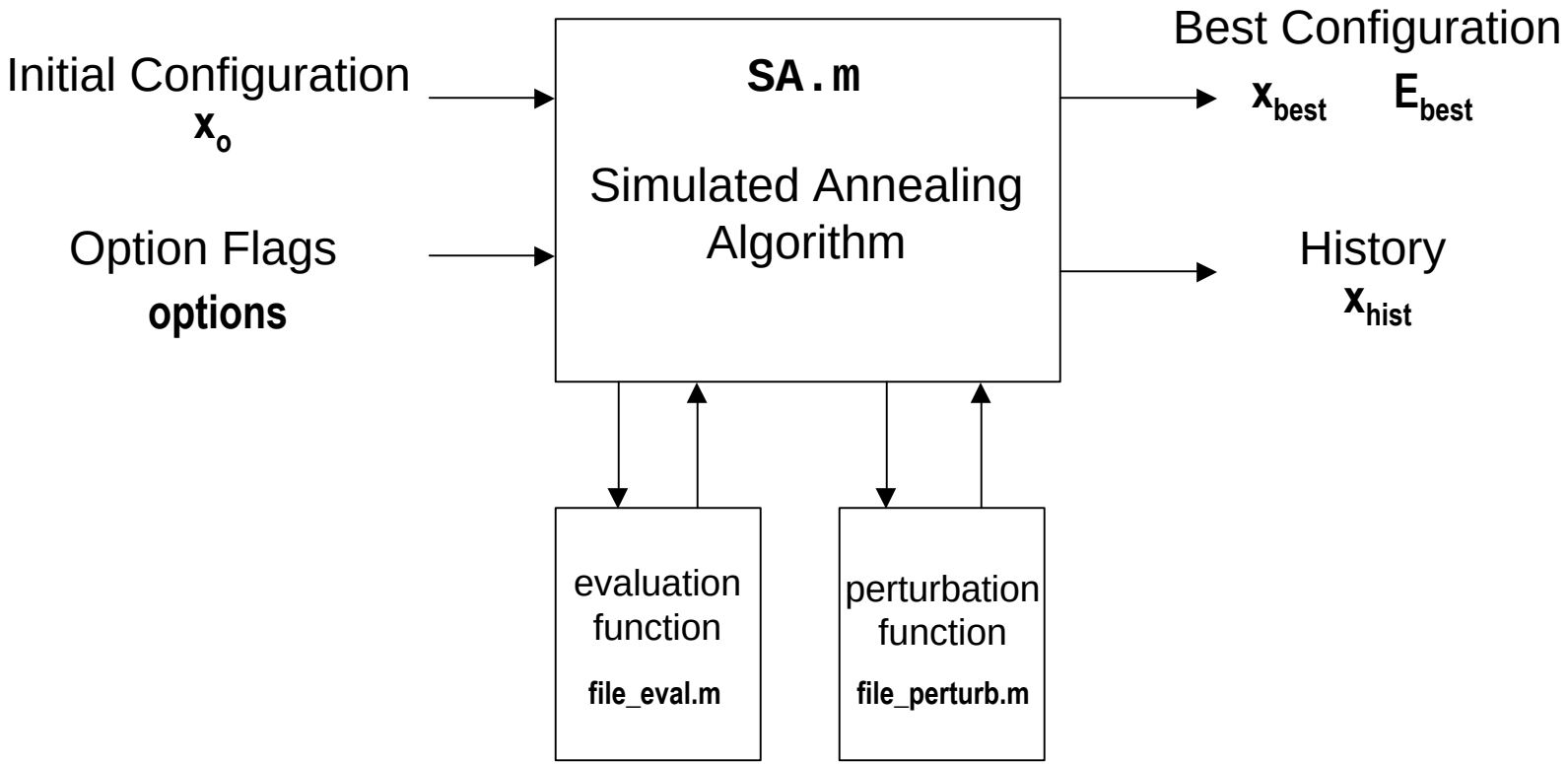
  6) If $E(X_{i+1}) > E(X_i)$, then accept X_{i+1} as the new current solution with a probability $e^{(-\Delta/T)}$ where $\Delta = E(X_{i+1}) - E(X_i)$.

  7) Reduce the system temperature according to the cooling schedule.

  8) Terminate the algorithm.


SA BLOCK DIAGRAM
Start T_o
Define Initial Configuration R_o
Evaluate Energy $E(R_o)$
Perturb Configuration R_i -> R_{i+1}
Evaluate Energy $E(R_{i+1})$
Compute Energy Difference $\Delta E = E(R_{i+1}) - E(R_i)$
$\Delta E < 0$?
y
n
Create Random Number ν in [0,1]
exp $\frac{-\Delta E}{T}$ > ν ?
n
y
Metropolis Step
Accept R_{i+1} as New Configuration
y
Keep R_i as Current Configuration
Reached Equilibrium at T_j ?
n
y
Reduce Temperature $T_{j+1} = T_j - \Delta T$
$T < T_{min}$?
n
y
End T_{min}

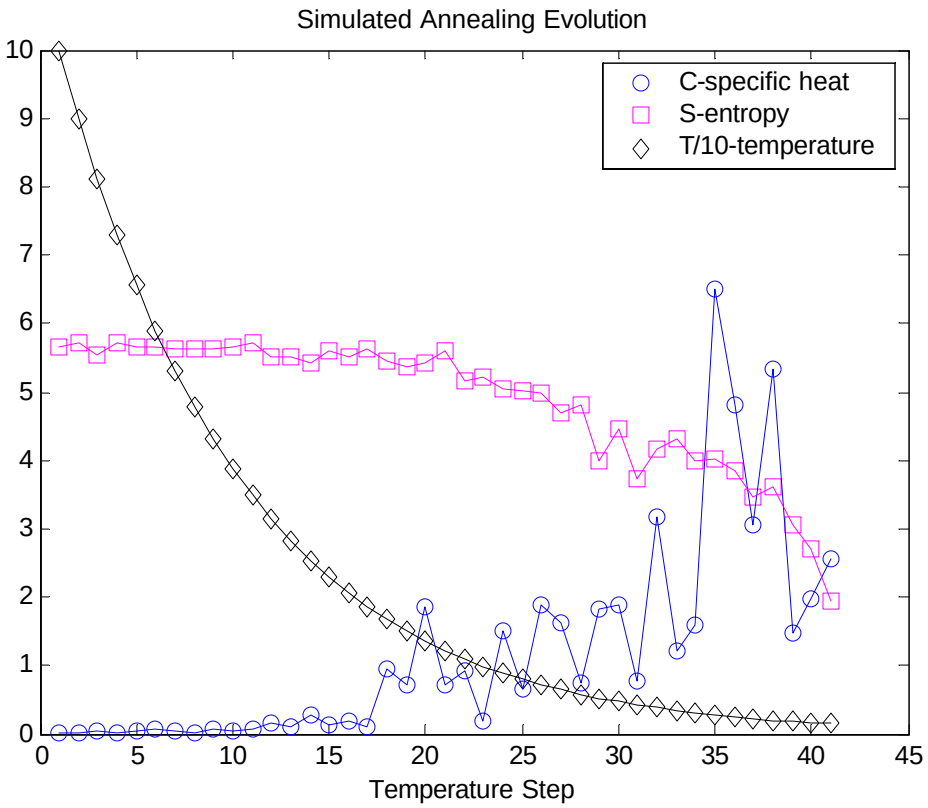
Matlab Function: **SA.m**



Exponential Cooling

- Typically (T_1/T_o)~0.7-0.9

$$T_{k+1} = \left(\frac{T_1}{T_o} \right)^k T_k$$



Can also do linear or step-wise cooling

…

But exponential cooling often works best.

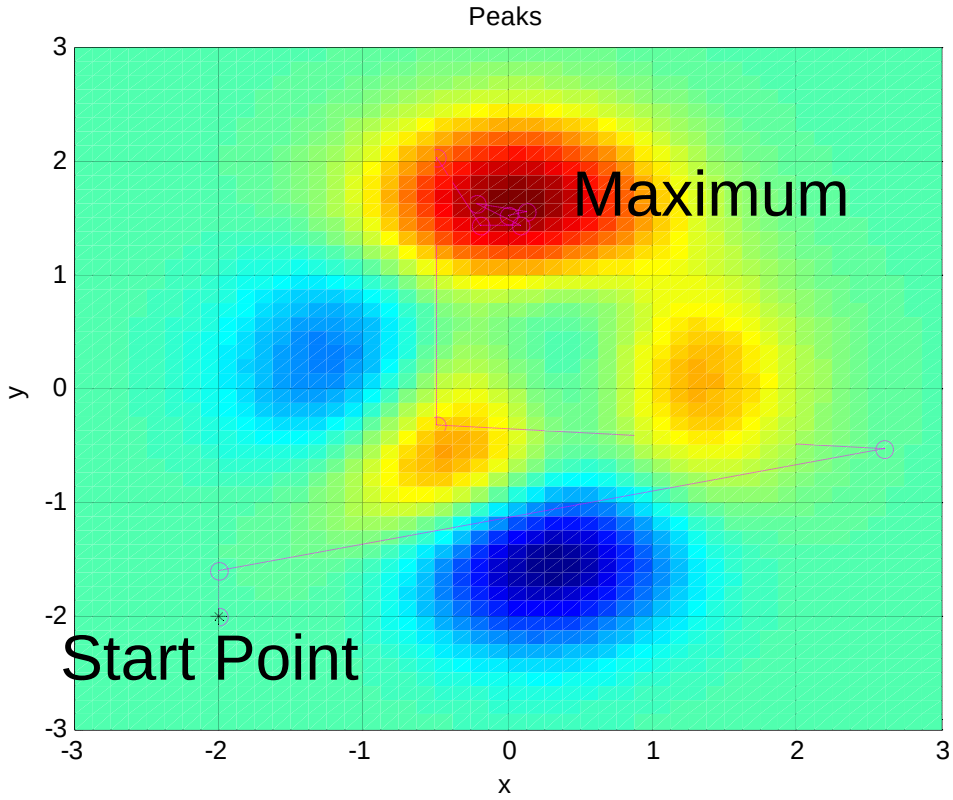
Key Ingredients for SA

- A concise description of a configuration (architecture, design, topology) of the system (Design Vector).
- A random generator of rearrangements of the elements in a configuration (Neighborhoods). This generator encapsulates rules so as to generate only valid configurations.
- A quantitative objective function containing the trade-offs that have to be made (Simulation Model and Output Metric(s)). Surrogate for system energy.
- An annealing schedule of the temperatures and/or the length of times for which the system is to be evolved.

Sample Problems

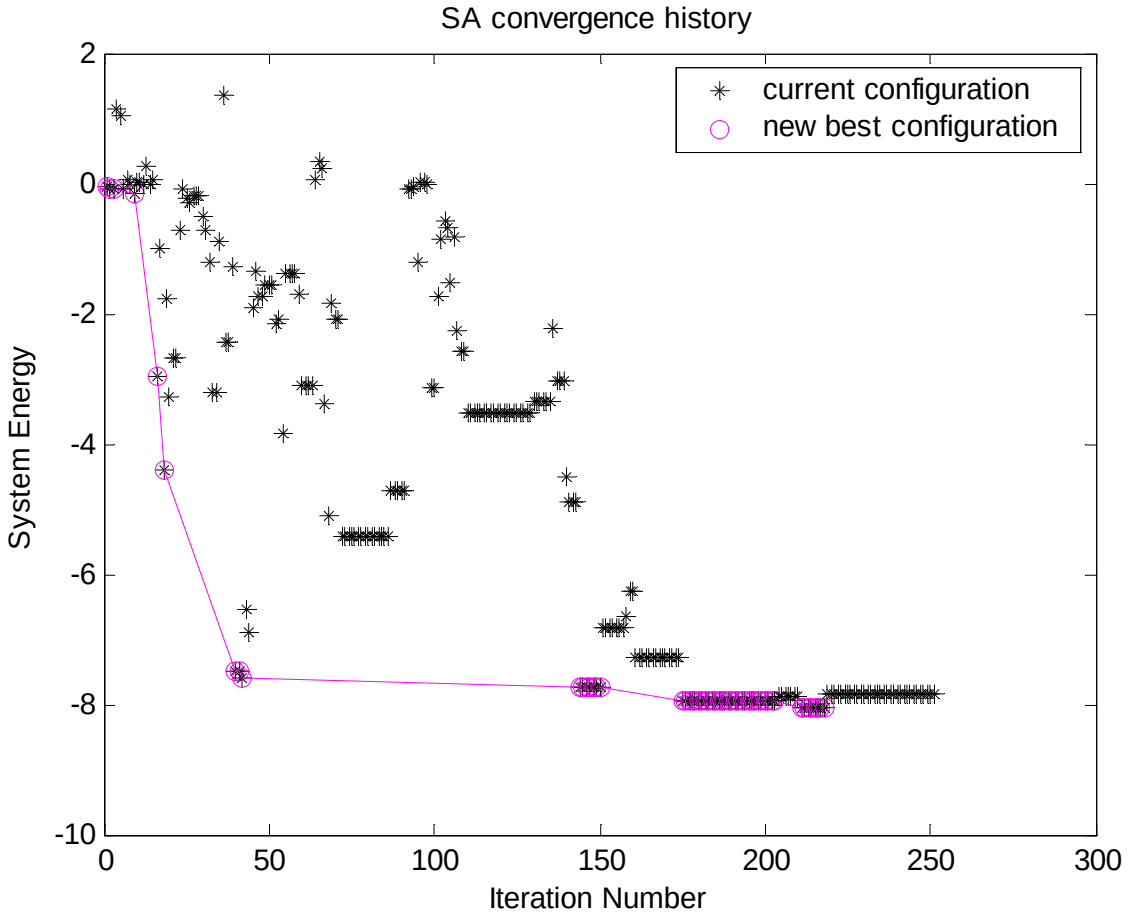
Matlab “peaks” function

- Difficult due to plateau at z=0, local maxima
- SAdemo1
- x_o=[-2,-2]
- Optimum at
- x*=[0.012, 1.524]
- z*=8.0484



“peaks” convergence

- Initially ~ nearly random search
- Later ~ gradient search



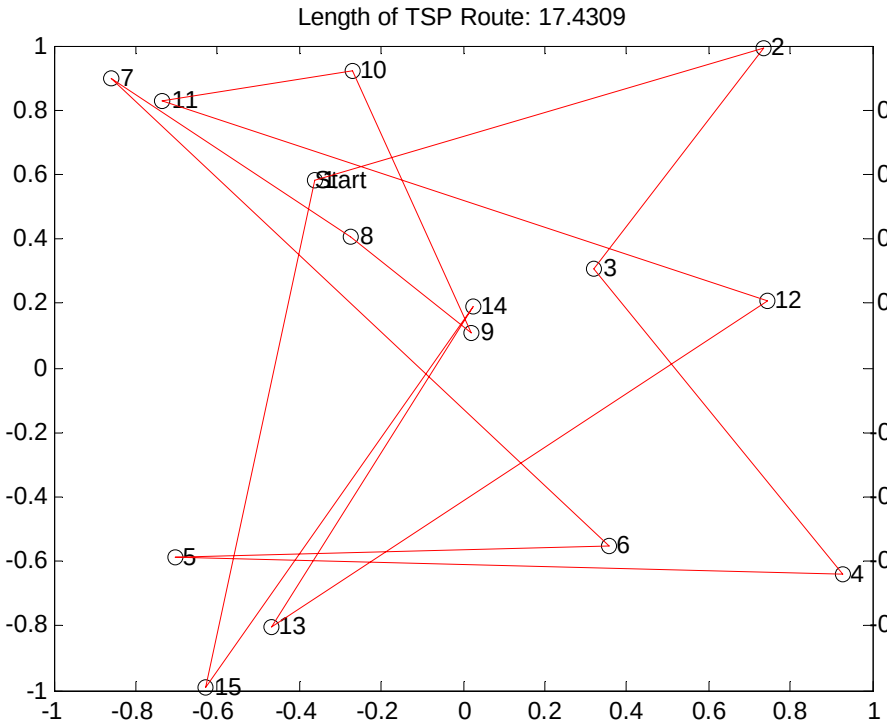
Traveling Salesman Problem

- *N* cities arranged randomly on [-1,1]
- Choose N=15
- SAdemo1
- Minimize “cost” of route (length, time,…)
- Visit each city once, return to start city

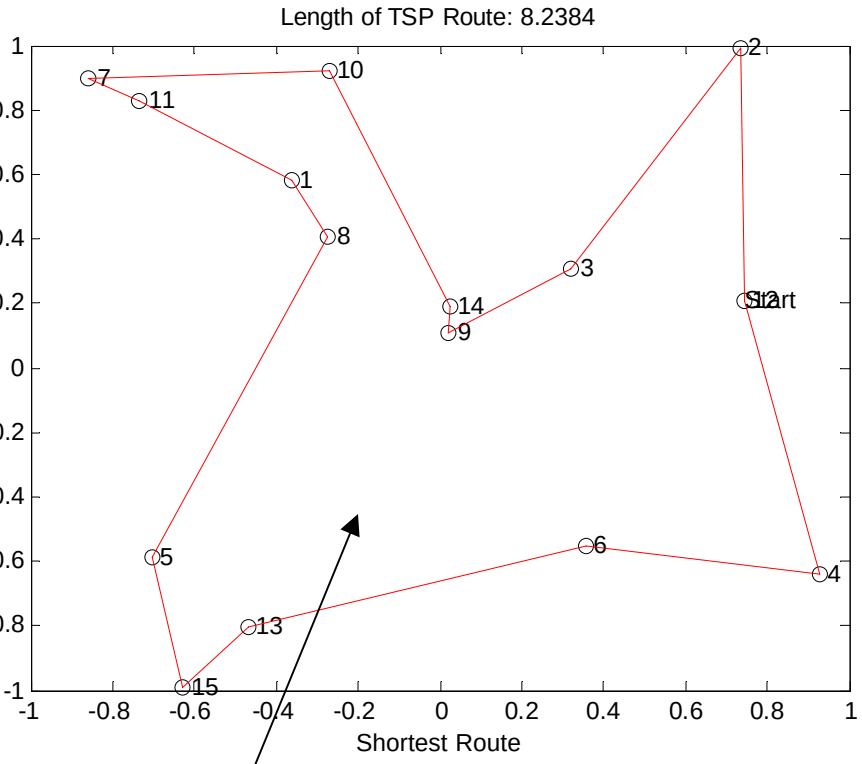
$$l(R) = \underbrace{\sum_{i=1}^{N} \sqrt{\sum_{j=1}^{2} \left(x_j(\ _{i+1}) - x_j(\ _i) \right)^2}}_{\text{visit N cities}} + \underbrace{\sqrt{\sum_{j=1}^{2} \left(x_j(R_1) - x_j(R_N) \right)^2}}_{\text{return home to first city}}$$

TSP Problem (II)

Initial (Random) Route
Length: 17.43



Final (Optimized) Route
Length: 8.24



Result with SA

Structural Optimization

- Define:
  - Design Domain
  - Boundary Conditions
  - Loads
  - Mass constraint
- Subdivide domain
  - N x M design “cells”
  - Cell density ρ=1 or ρ=0
- Where to put material to minimize compliance?

$$\text{find} \quad r_i \qquad i = 1,\ldots,N$$

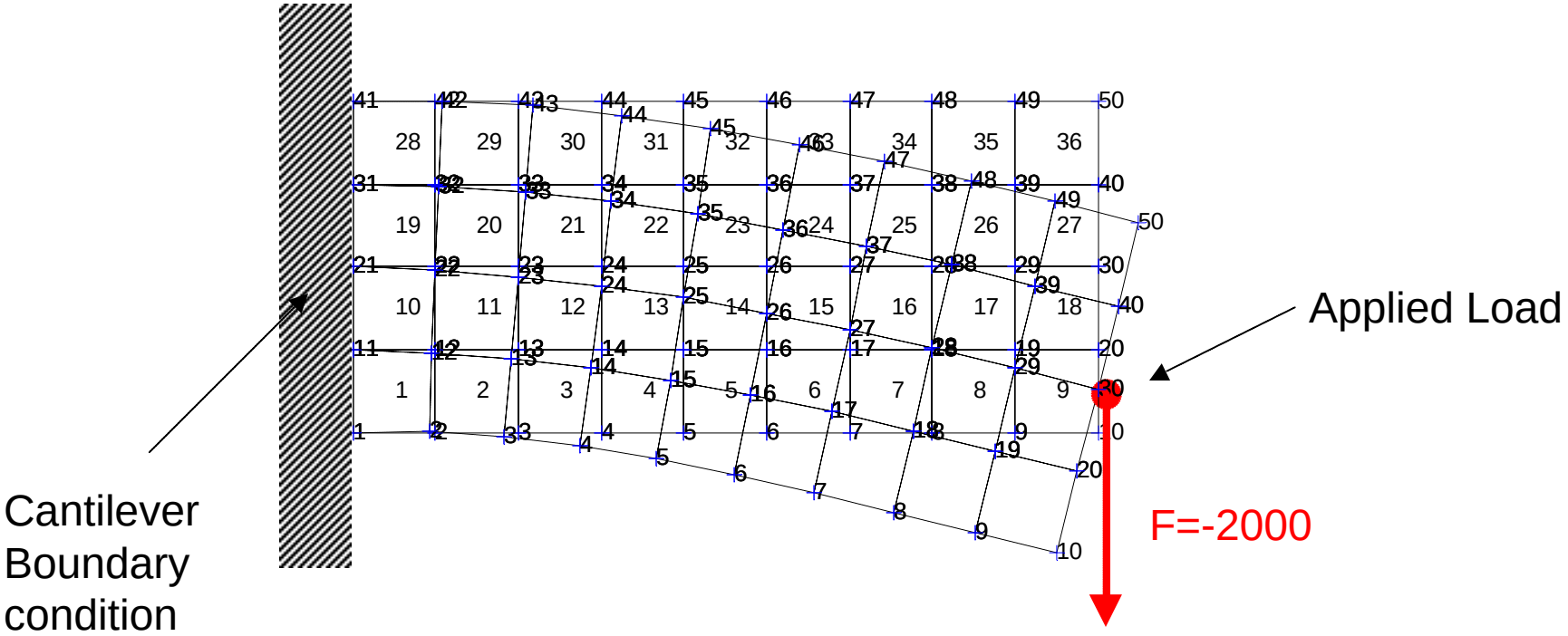
$$\min C = f^T u(r_i)$$

$$\text{s.t.} \quad u = K^{-1} f$$

$$\text{s.t.} \quad \sum_{i=1}^{N} V_i r_i \le m_{\max}$$

Structural Topology Optimization (II)

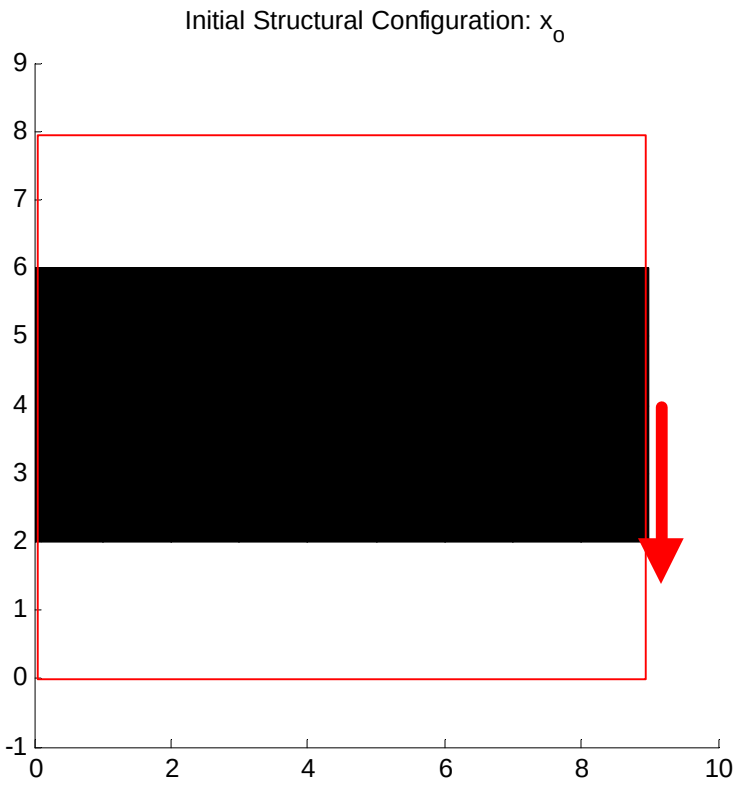
"Energy" = strain energy = compliance
Computed via Finite Element Analysis



Deformation not drawn to scale

Structural Toplogy Optimization

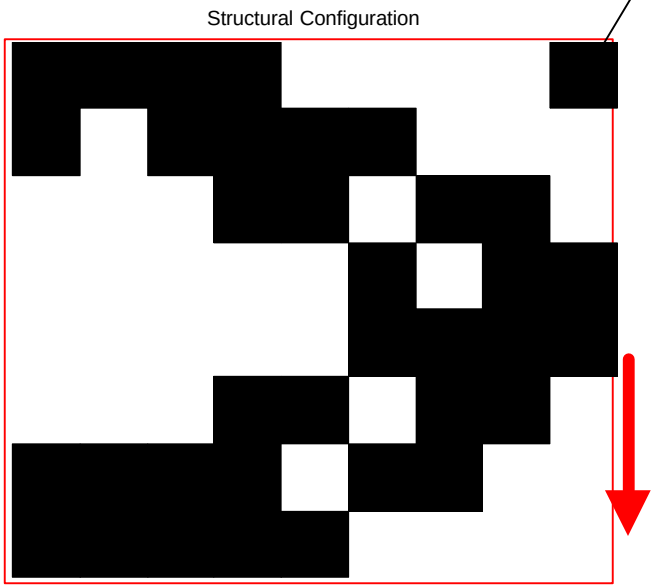
Initial Structural Toplogy



Compliance C=593.0

“Final” Toplogy

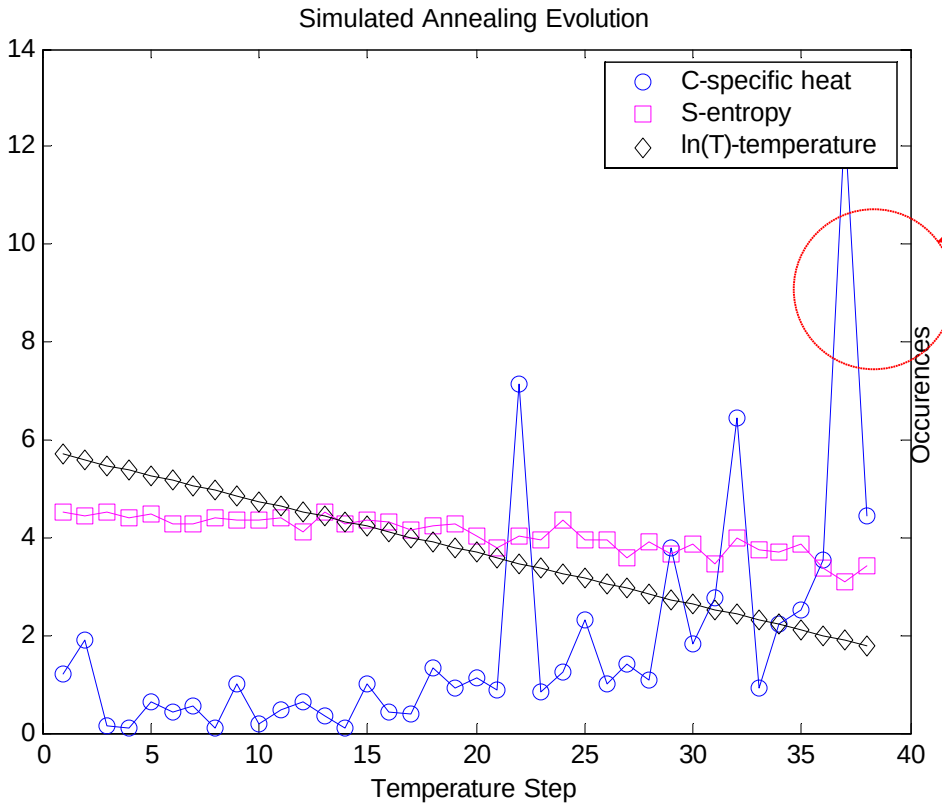
“island”

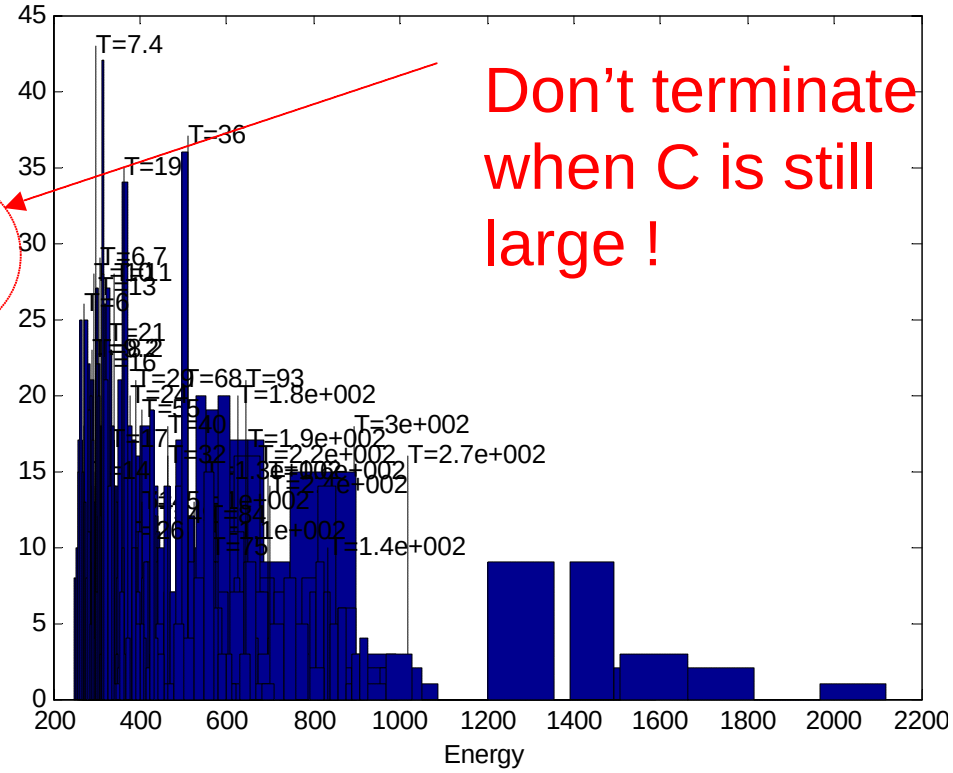


Compliance C=236.68

Not “optimal” – often need post-processing

Structural Optimization – Convergence Analysis





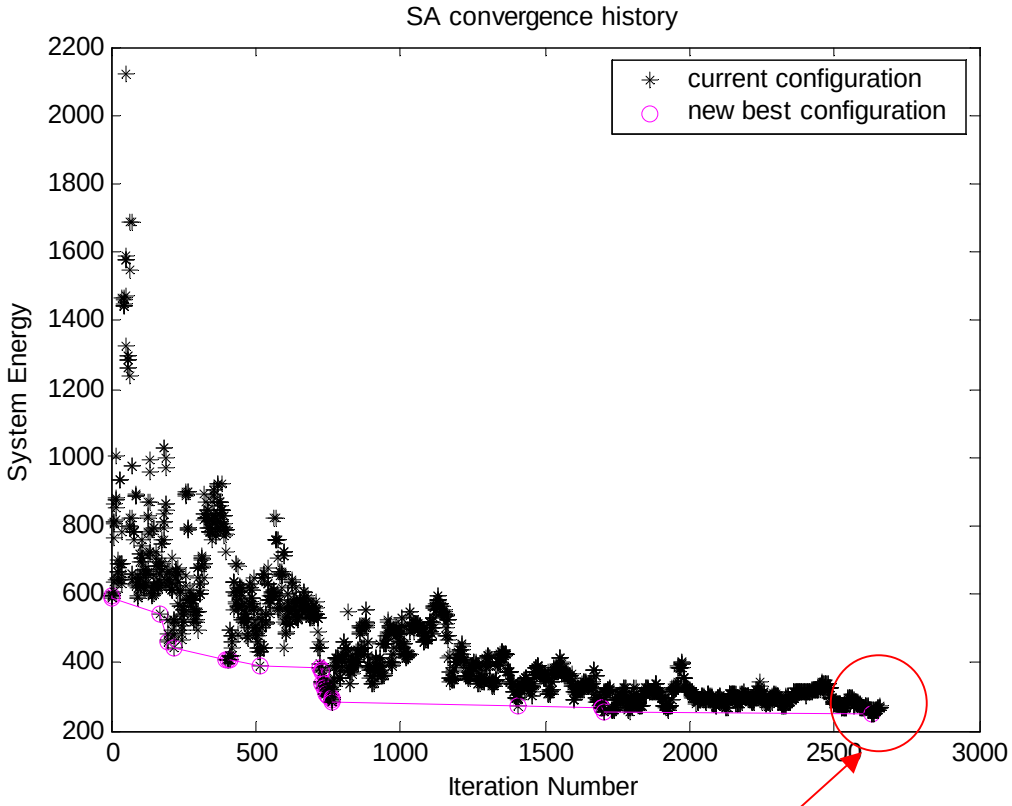
Don't terminate when C is still large !

Boltzmann Distributions

Evolution of

- Entropy, Temperature, Specific Heat

Premature termination

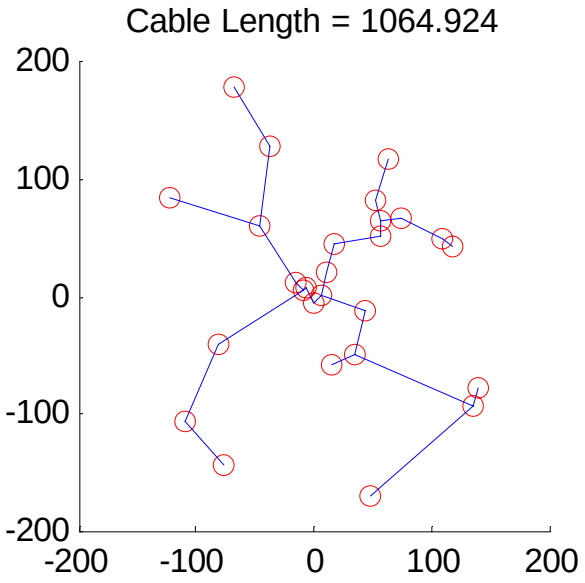


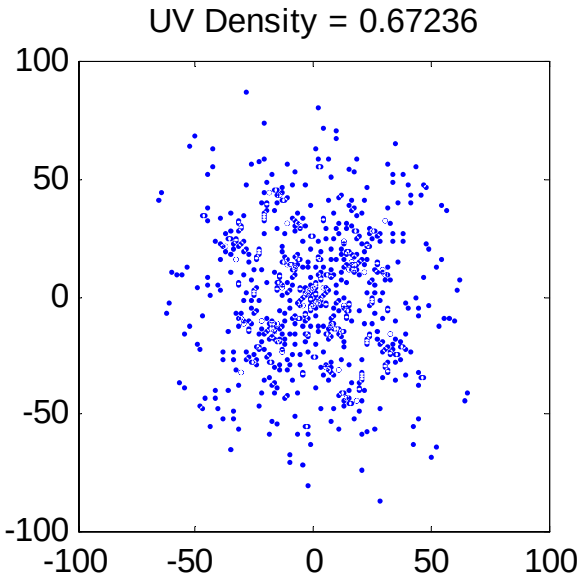
Indicator: Best Configuration found only shortly before Simulated Annealing terminated.

Final Example: Telescope Array Optimization

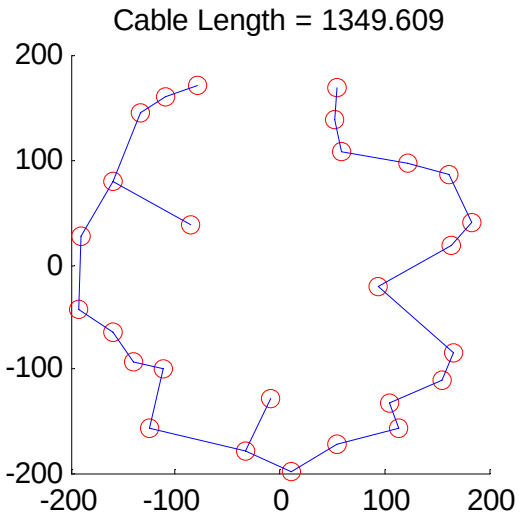
- Place *N*=27 stations in xy within a 200 km radius
- Minimize UV density metric
- Ideally also minimize cable length

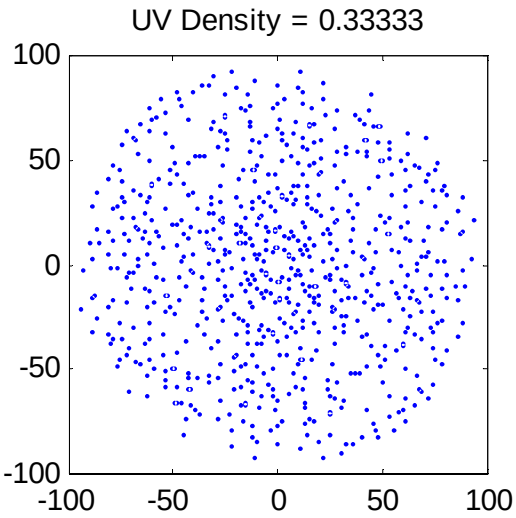
Initial Configuration





Optimized Solution





Simulated Annealing Improved UV density from 0.67 to 0.33

- Simulated Annealing transforms the array:
  - Hub-and-Spoke → Circle-with-arms

Reference available

Summary

Summary: Steps of SA

- **The Simulated Annealing Algorithm**

  1) Choose a random X_i, select the initial system temperature, and outline the cooling (ie. annealing) schedule

  2) Evaluate $E(X_i)$ using a simulation model

  3) Perturb X_i to obtain a neighboring Design Vector (X_{i+1})

  4) Evaluate $E(X_{i+1})$ using a simulation model

  5) If $E(X_{i+1}) < E(X_i)$, X_{i+1} is the new current solution

  6) If $E(X_{i+1}) > E(X_i)$, then accept X_{i+1} as the new current solution with a probability $e^{(-\Delta/T)}$ where $\Delta = E(X_{i+1}) - E(X_i)$.

  7) Reduce the system temperature according to the cooling schedule.

  8) Terminate the algorithm.

Research in SA

- Alternative Cooling Schedules and Termination criteria
- Adaptive Simulated Annealing (ASA) – determines its own cooling schedule
- Hybridization with other Heuristic Search Methods (GA, Tabu Search …)
- Multiobjective Optimization with SA

References

- Cerny, V., "Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm", *J. Opt. Theory Appl.*, 45, 1, 41-51, 1985 .

- de Weck O. “System Optimization with Simulated Annealing (SA), Memorandum

- Cohanim B., Hewitt J, and de Weck O. L. “The Design of Radio Telescope Array Configurations using Multiobjective Optimization: Imaging Performance versus Cable Length”, The Astrophysical Journal, (in press), 2004

- Jilla, C.D., and Miller, D.W., “Assessing the Performance of a Heuristic Simulated Annealing Algorithm for the Design of Distributed Satellite Systems,” *Acta Astronautica*, Vol. 48, No. 5-12, 2001, pp. 529-543.

- **Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., “Optimization by Simulated Annealing,” *Science*, Volume 220, Number 4598, 13 May 1983, pp. 671-680.**

- Metropolis,N., A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, "Equation of State Calculations by Fast Computing Machines", *J. Chem. Phys.*,21, 6, 1087-1092, 1953.