SA1

SIMULATED ANNEALING

- Origins : Statistical Physics
- Maxwell / Boltzmann : Kinetic Theory of Gases

* # of particles ~ 10^{23} atoms/cm^3

* Only thermal equilibrium at temperature T may be considered / observed

* Gas molecules in a volume V in thermal equilibrium at temperature T

* n_i : # of molecules with kinetic energy E_i

$$\boxed{\frac{n_1}{n_2} = \frac{e^{-\frac{E_1}{kT}}}{e^{-\frac{E_2}{kT}}} = e^{-\frac{(E_1-E_2)}{kT}}}$$

k: Boltzmann constant

* molecules with their combination / energies form a "thermal state" x_i

QUESTION: At thermal equilibrium at temperature T, which states are more likely observable?

* state → x_i its "energy" → E_i

$$\boxed{P(x_i) = C\, e^{-\frac{E_i}{kT}}}$$

$\sum_i p(x_i) = 1$ ⇒ normalization constant C.

SA 2

- At high temperatures, almost all states can be observed
- At low temperatures, only states with small energy can be observed
- At absolute zero temperature (T=0) only the zero energy level is observable

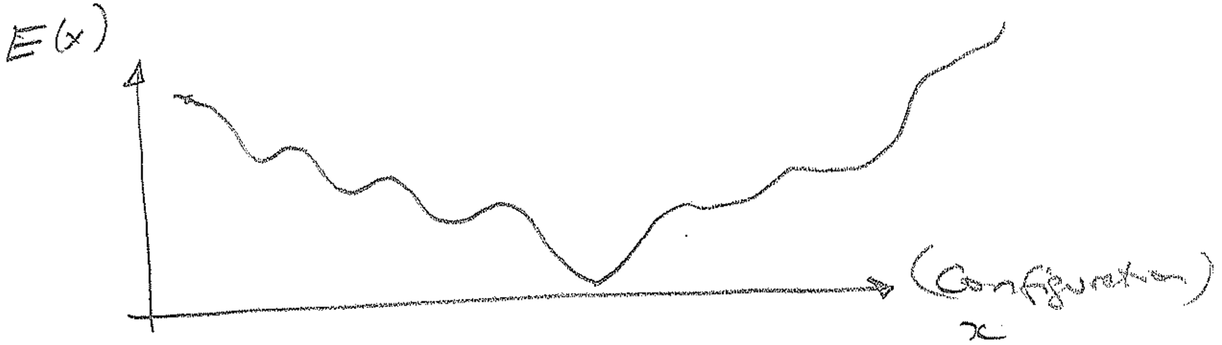
- This idea could be utilized in "optimization"

ANNEALING

- Changing temperature ⇒ change the "state"
- Heat a metal above "melting"
- Cool down slowly at a temperature
- If everything goes right ⇒ a desired "state" may be achieved
  - SiO_2 → melt → cool → glass
  - Iron +·· → steel.
  - During "cooling", at each temperature different possible "states" may be observed
  - Should "cool" intelligently
  - Requires "intelligent" cooling schemes.

- Possible problem during cooling:

* At a certain temperature (molecule) configuration may reach to a "thermal equilibrium" which is not the desired one ⇒ Local minimum of the energy.



- Solution: sometimes a "reheating" may be necessary.
- Forging swords, forming glasses, etc
- Same idea could be used in optimization
- Metropolis - Hastings in 1953
- Generalization by Kirkpatrick-Gelatt-Vecchi in 1983

Main idea: go downhill (descent), but allow with some probability going uphill sometimes

- Cool down (reduce the temperature)
- Stop at a reasonable point.

* descent / ascent ⇒ annealing cooling / reheating
⇒ escape from local minima
probability of ascent ⇒ Boltzmann distribution

* Cool at absolute zero ⇒ to achieve zero energy level according to Boltzmann law

SA4

Problem:

$$\min_{x} E(x)$$ (find GLOBAL minimum)

- In gradient descent ⇒ problem of local minimum.

SIMULATED ANNEALING

- Choose initial temperature T
- Start with x_0
- Evaluate $E(x_0)$
- Choose $x_1 = x_0 + h$ (Neighborhood!)
- Evaluate $E(x_1)$
- If $E(x_1) < E(x_0)$, accept x_1 as new point.
- If $E(x_1) > E(x_0)$ accept x_1 with PROBABILITY p

  $p(x_1) = e^{-\frac{(E(x_1) - E(x_0))}{T}}$ (Botzmann!)

  (Cooling!)
- Lower the temperature
- Repeat.

* Stop at a reasonable point
(i.e. when E is below a treshold
or temperature is too low)

SA5

- To implement the Boltzmann Rule, a RANDOM NUMBER GENERATOR is required
  - Evaluate $p(x_i) = \bar{e}^{\frac{E(x_i)-E(x_{i-1})}{T}} < 1$
  - $p(x_i) \in (0,1)$
  - pick a RANDOM NUMBER $r \in (0,1)$
  - if $r < p(x_i) \Rightarrow$ accept x_i
  - otherwise remain at x_{i-1}.

- If temperature T is too high → any move is almost acceptable
- As T gets lower only "downhill" directions can be accepted

(i.e. for T↑ → random search
for T↓ → gradient descent)

- Cooling scheme is important
- Should not cool TOO FAST (⇒ gradient descent)
  ⇒ may be trapped to a local minimum

- Possible cooling schemes
  - $T_{k+1} = \alpha T_k$ $\quad 0<\alpha<1 \Rightarrow$ Geometric cooling
  - $T_{k+1} = \bar{e}^{\alpha T_k}$ $\quad \alpha>0 \Rightarrow$ exponential cooling
  - $T_{k+1} = T_k - \beta k$ $\quad \beta>0 \Rightarrow$ linear cooling
  - $T_{k+1} = \dfrac{c}{\ln k}$ $\quad c>0 \Rightarrow$ logarithmic cooling

Convergence: Under certain conditions, convergence to GLOBAL MINIMUM is guaranteed with PROBABILITY 1 as $k \to \infty$

Conditions: Cooling should not be TOO FAST

$(\sum T_k \to \infty)$

STOCHASTIC GRADIENT DESCENT

— In most of the cases, same as SHUFFLING

- training set $\{x_1, x_2, \dots, x_N\}$
- Start with index set $J_0 = \{1, 2, \dots, N\}$
- SHUFFLE the index set RANDOMLY
  $J_s = \{sh(1), sh(2), \dots, sh(N)\}$
  — Do the incremental gradient descent
  — Note that $sh(i) = random(1, N) \neq sh(j) \quad j \le i$
- Sometimes ADDING NOISE to gradient can also be used

$$w \leftarrow w - \eta \left(\frac{\partial E}{\partial w} + n(t) \right)$$

$n(t)$: a random but small term

Silicon dioxide

From Wikipedia, the free encyclopedia

The chemical compound **silicon dioxide**, also known as **silica** (from the Latin *silex*), is an oxide of silicon with the chemical formula **SiO_2**. It has been known for its hardness since antiquity. Silica is most commonly found in nature as sand or quartz, as well as in the cell walls of diatoms.[1][2]

Silica is manufactured in several forms including fused quartz, crystal, fumed silica (or pyrogenic silica, trademarked Aerosil or Cab-O-Sil), colloidal silica, silica gel, and aerogel.

Silica is used primarily in the production of glass for windows, drinking glasses, beverage bottles, and many other uses. The majority of optical fibers for telecommunications are also made from silica. It is a primary raw material for many whiteware ceramics such as earthenware, stoneware, porcelain, as well as industrial Portland cement.

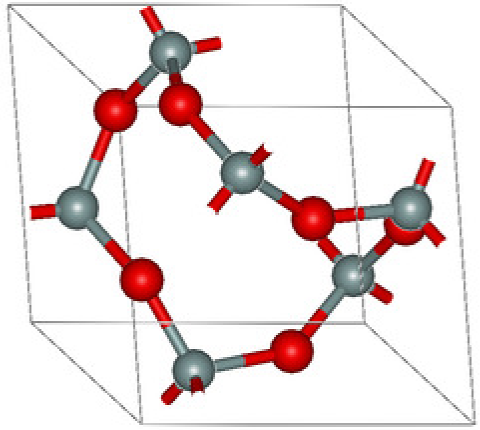
Silica is a common additive in the production of foods, where it is used primarily as a flow agent in powdered foods, or to absorb water in hygroscopic applications. It is the primary component of diatomaceous earth which has many uses ranging from filtration to insect control. It is also the primary component of rice husk ash which is used, for example, in filtration and cement manufacturing.

Thin films of silica grown on silicon wafers via thermal oxidation methods can be quite beneficial in microelectronics, where they act as electric insulators with high chemical stability. In electrical applications, it can protect the silicon, store charge, block current, and even act as a controlled pathway to limit current flow.[3]

A silica-based aerogel was used in the Stardust spacecraft to collect extraterrestrial particles. Silica is also used in the extraction of DNA and RNA due to its ability to bind to the nucleic acids under the presence of chaotropes. As hydrophobic silica it is used as a defoamer component. In hydrated form, it is used in toothpaste as a hard abrasive to remove tooth plaque.

In its capacity as a refractory, it is useful in fiber form as a high-temperature thermal protection fabric. In cosmetics, it is useful for its light-diffusing properties and natural absorbency. Colloidal silica is used as a wine and juice fining agent. In pharmaceutical products, silica aids powder flow when tablets are formed. Finally, it is used as a thermal enhancement compound in ground source heat pump industry.

Silicon dioxide



IUPAC name

Silicon dioxide

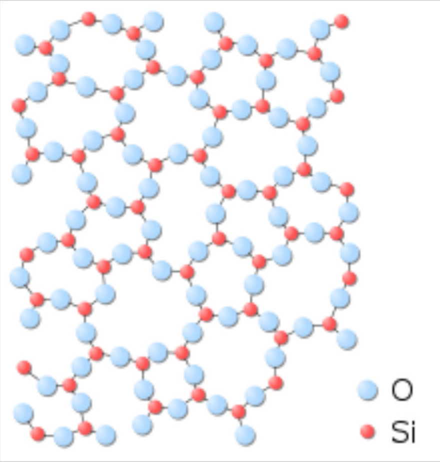
Other names

Quartz

Silica
Silicic oxide

Silicon(IV) oxide

For example, in the unit cell of α-quartz, the central tetrahedron shares all 4 of its corner O atoms, the 2 face-centered tetrahedra share 2 of their corner O atoms, and the 4 edge-centered terahedra share just one of their O atoms with other SiO_4 tetrahedra. This leaves a net average of 12 out of 24 total vertices for that portion of the 7 SiO_4 tetrahedra which are considered to be a part of the unit cell for silica (see 3-D Unit Cell (http://www.mindat.org/min-3337.html)).



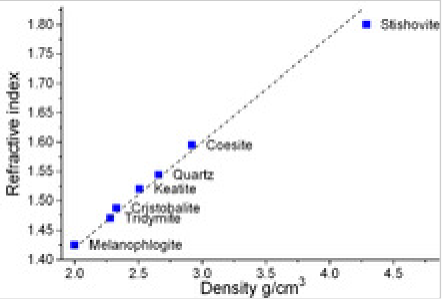
The amorphous structure of glassy silica (SiO_2) in two-dimensions. No long-range order is present; however there is local ordering with respect to the tetrahedral arrangement of oxygen (O) atoms around the silicon (Si) atoms. Note that a fourth oxygen atom is bonded to each silicon atom, either behind the plane of the screen or in front of it; these atoms are omitted for clarity.

SiO_2 has a number of distinct crystalline forms (polymorphs) in addition to amorphous forms. With the exception of stishovite and fibrous silica, all of the crystalline forms involve tetrahedral SiO_4 units linked together by shared vertices in different arrangements. Silicon-oxygen bond lengths vary between the different crystal forms, for example in α-quartz the bond length is 161 pm, whereas in α-tridymite it is in the range 154–171 pm. The Si-O-Si angle also varies between a low value of 140° in α-tridymite, up to 180° in β-tridymite. In α-quartz the Si-O-Si angle is 144°.[5]

Fibrous silica has a structure similar to that of SiS_2 with chains of edge-sharing SiO_4 tetrahedra. Stishovite, the higher pressure form, in contrast has a rutile like structure where silicon is 6 coordinate. The density of stishovite is 4.287 g/cm^3, which compares to α-quartz, the densest of the low pressure forms, which has a density of 2.648 g/cm^3.[6] The difference in density can be ascribed to the increase in coordination as the six shortest Si-O bond lengths in stishovite (four Si-O bond lengths of 176 pm and two others of 181 pm) are greater than the Si-O bond length (161 pm) in α-quartz.[7] The change in the coordination increases the ionicity of the Si-O bond.[8] But more important is the observation that any deviations from these standard parameters constitute microstructural differences or variations which represent an approach to an amorphous, vitreous or glassy solid.

Note that the only stable form under normal conditions is α-quartz and this is the form in which crystalline silicon dioxide is usually encountered. In nature impurities in crystalline α-quartz can give rise to colors (see list).

Note also that both high temperature minerals, cristobalite and tridymite, have both a lower density and index of refraction than quartz. Since the composition is identical, the reason for the discrepancies must be in the increased spacing in the high temperature minerals. As is common with many substances, the higher the temperature the farther apart the atoms due to the increased vibration energy.



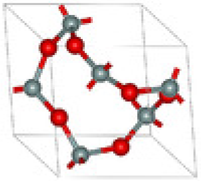
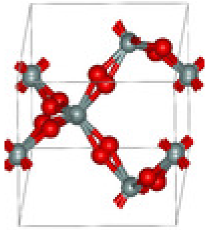
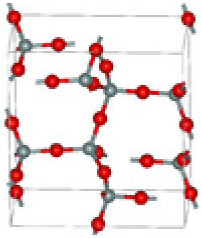
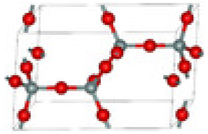
Relation between refractive index and density for some SiO_2 forms.[4]

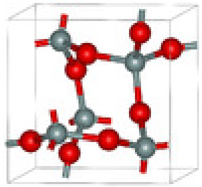
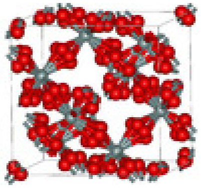
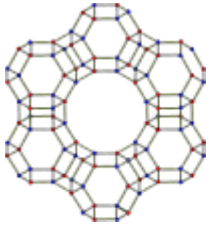
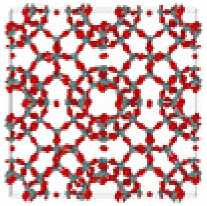
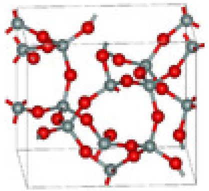
The high pressure minerals, seifertite, stishovite, and coesite, on the other hand, have a higher density and index of refraction when compared to quartz. This is probably due to the intense compression of the atoms that must occur during their formation, resulting in a more condensed structure.

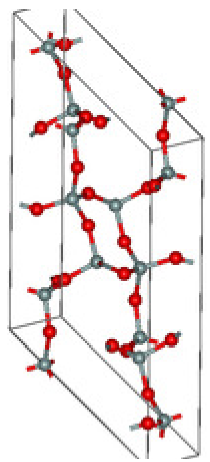
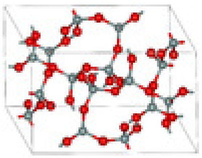
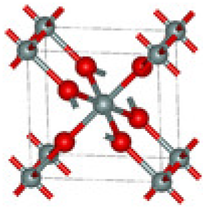
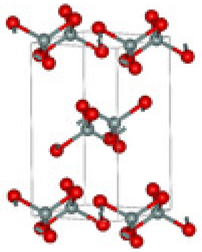
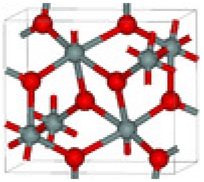
Faujasite silica is another form of crystalline silica. It is obtained by dealumination of a low-sodium, ultra-stable Y zeolite with a combined acid and thermal treatment. The resulting product contains over 99% silica, has high crystallinity and high surface area (over 800 m^2/g). Faujasite-silica has very high thermal and acid stability. For example, it maintains a high degree of long-range molecular order (or crystallinity) even after boiling in concentrated hydrochloric acid.[9]

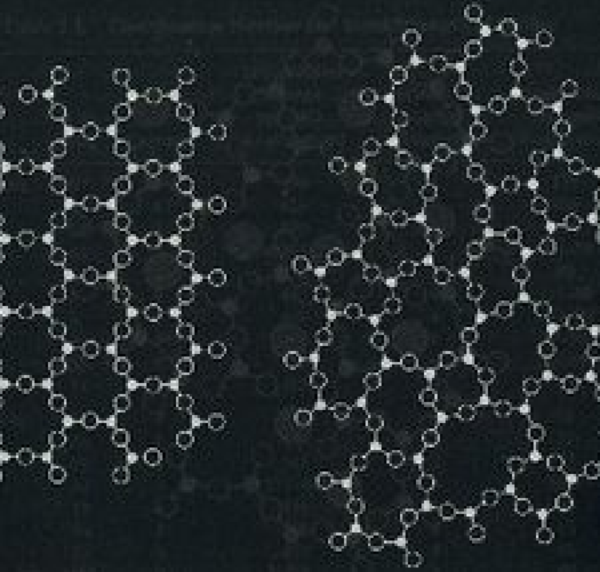
Molten silica exhibits several peculiar physical characteristics that are similar to the ones observed in liquid water: negative temperature expansion, density maximum, and a heat capacity minimum.[10] When molecular silicon monoxide, SiO, is condensed in an argon matrix cooled with helium along with oxygen atoms generated by microwave discharge, molecular SiO_2 is produced which has a linear structure. Dimeric silicon dioxide, $(SiO_2)_2$ has been prepared by reacting O_2 with matrix isolated dimeric silicon monoxide, (Si_2O_2). In dimeric silicon dioxide there are two oxygen atoms bridging between the silicon atoms with an Si-O-Si angle of 94° and bond length of 164.6 pm and the terminal Si-O bond length is 150.2 pm. The Si-O bond length is 148.3 pm which compares with the length of 161 pm in α-quartz. The bond energy is estimated at 621.7 kJ/mol.[11]

Crystalline forms of SiO_2[5]

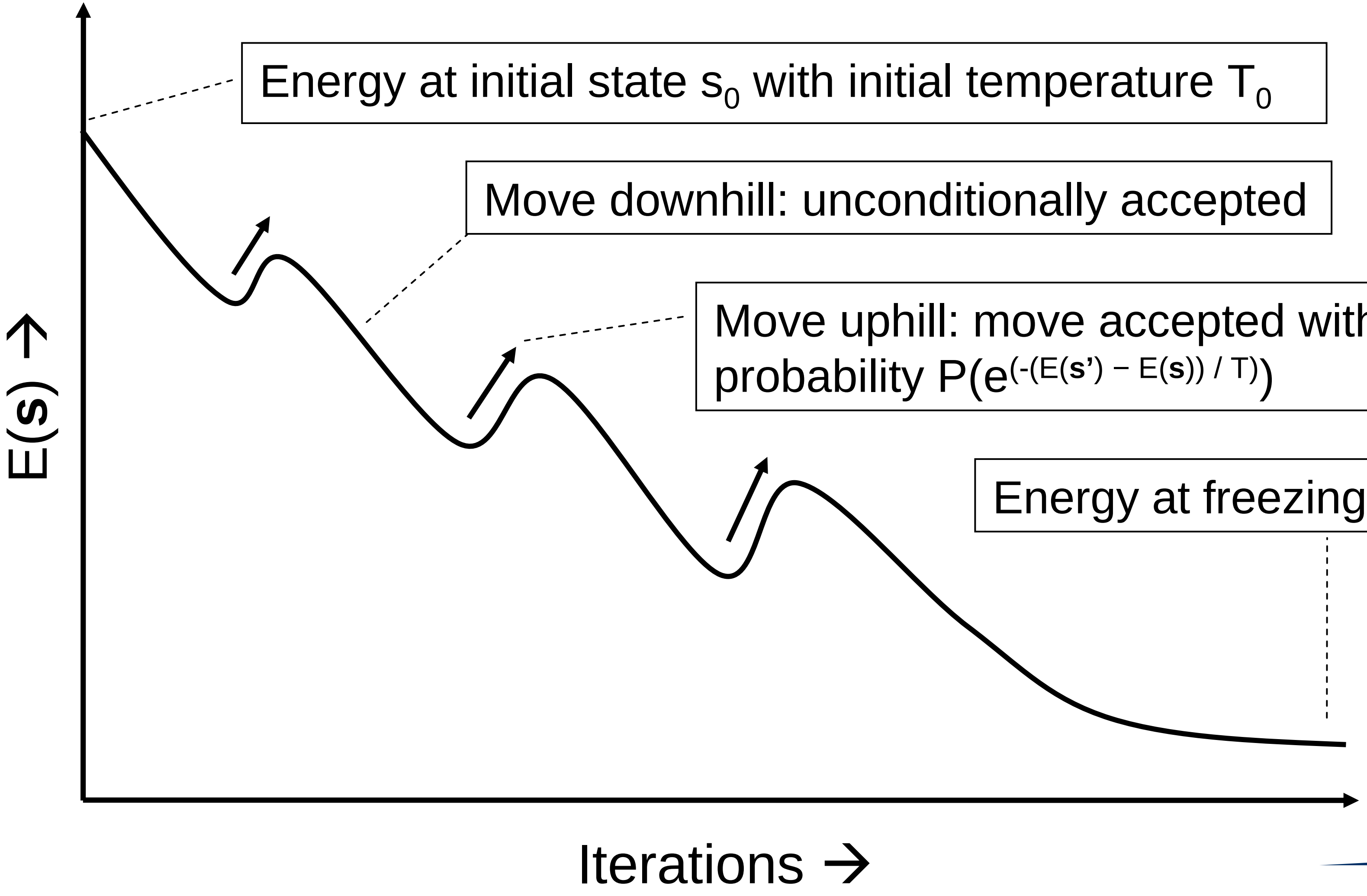
Form	Crystal symmetry Pearson symbol, group No.	Notes	Structure
α-quartz	rhombohedral (trigonal) hP9, $P3_121$ No.152[12]	Helical chains making individual single crystals optically active; α-quartz converts to β-quartz at 846 K	
β-quartz	hexagonal hP18, $P6_222$, No.180[13]	closely related to α-quartz (with an Si-O-Si angle of 155°) and optically active; β-quartz converts to β-tridymite at 1140 K	
α-tridymite	orthorhombic oS24, $C222_1$, No.20[14]	metastable form under normal pressure	
β-tridymite	hexagonal hP12, $P6_3/mmc$, No. 194[14]	closely related to α-tridymite; β-tridymite converts to β-cristobalite at 2010 K	

α-cristobalite	tetragonal tP12, $P4_12_12$, No. 92[15]	metastable form under normal pressure	
β-cristobalite	cubic cF104, $Fd\bar{3}m$, No.227[16]	closely related to α-cristobalite; melts at 1978 K	
faujasite	cubic cF576, $Fd\bar{3}m$, No.227[17]	sodalite cages connected by hexagonal prisms; 12-membered ring pore opening; faujasite structure.[9]	
melanophlogite	cubic (cP*, $P4_232$, No.208)[4] or tetragonal ($P4_2/nbc$)[18]	Si_5O_{10}, Si_6O_{12} rings; mineral always found with hydrocarbons in interstitial spaces-a clathrasil[19]	
keatite	tetragonal tP36, $P4_12_12$, No. 92[20]	Si_5O_{10}, Si_4O_{14}, Si_8O_{16} rings; synthesised from glassy silica and alkali at 600–900K and 40–400 MPa	

moganite	monoclinic mS46, C2/c, No.15[21]	Si_4O_8 and Si_6O_{12} rings	
coesite	monoclinic mS48, C2/c, No.15[22]	Si_4O_8 and Si_8O_{16} rings; 900 K and 3–3.5 GPa	
stishovite	Tetragonal tP6, P4_2/mnm, No.136[23]	One of the densest (together with seifertite) polymorphs of silica; rutile-like with 6-fold coordinated Si; 7.5–8.5 GPa	
fibrous	orthorhombic oI12, Ibam, No.72[24]	like SiS_2 consisting of edge sharing chains	
seifertite	orthorhombic oP, Pbcn[25]	One of the densest (together with stishovite) polymorphs of silica; is produced at pressures above 40 GPa.[26]	



Convergence of SA



Energy at initial state s_0 with initial temperature T_0

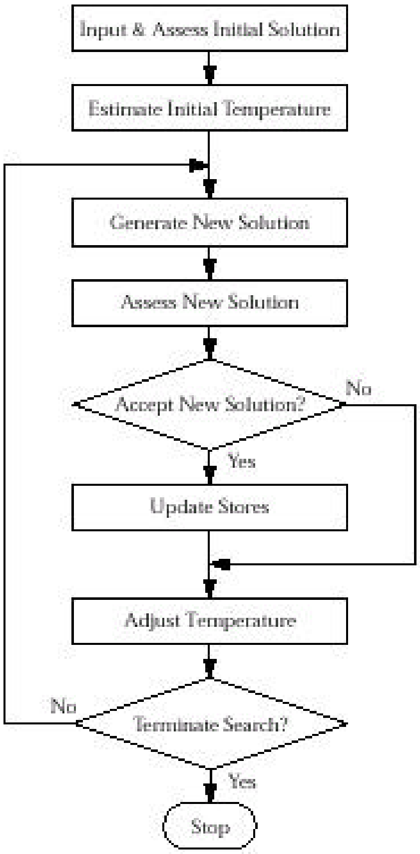
Move downhill: unconditionally accepted

Move uphill: move accepted with probability $P(e^{(-(E(\mathbf{s}') - E(\mathbf{s})) / T)})$

Energy at freezing point

Statistical Thermodynamics and Computing

Physics	Simulated Annealing
State	Feasible solution to the problem
Energy	Value returned by eval()
Equilibrium State	Local Optimum
Ground State	Global Optimum
Temperature	Control Parameter
Annealing	Search by reducing T
Boltzmann-Gibbs distribution	Probability of selecting a new point



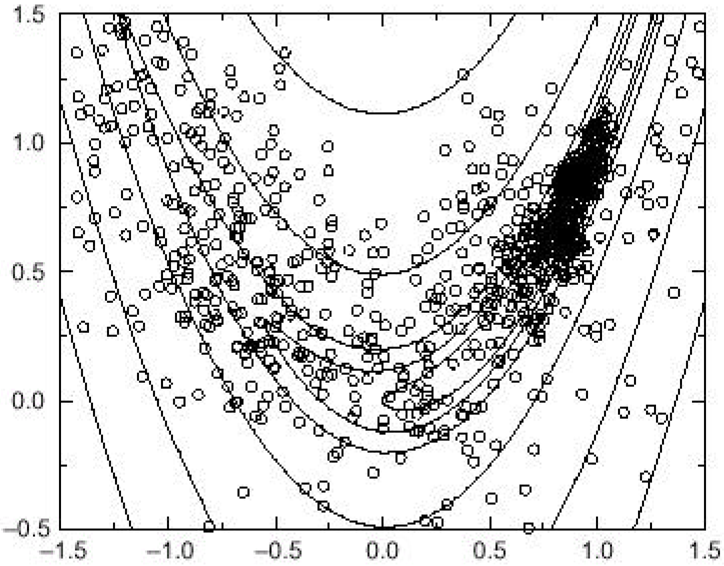
The structure of the simulated annealing algorithm

The following elements must be provided:

- a representation of possible solutions
- a generator of random changes in solutions
- a means of evaluating the problem functions and
- an *annealing schedule* - an initial temperature and rules for lowering it as the search progresses.

These steps are discussed in section 5.

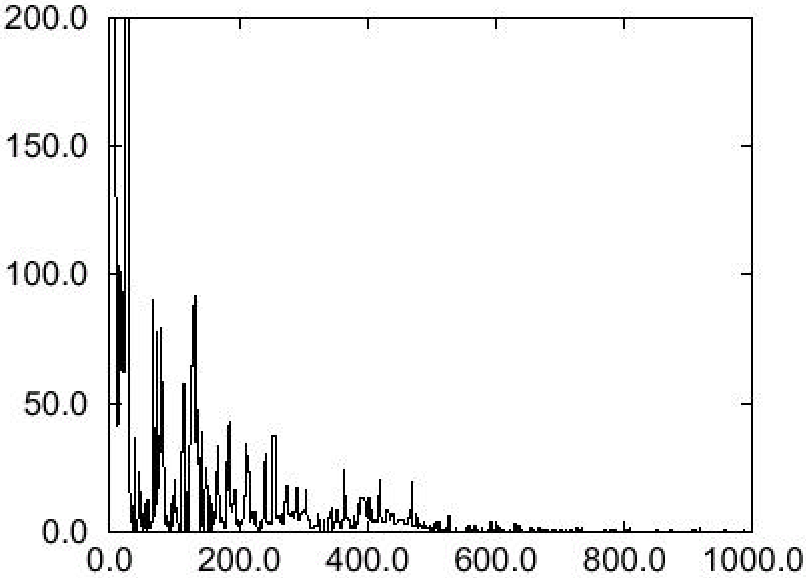
The following figure shows the progress of a SA search on the two-dimensional Rosenbrock function, $f = [(1-x_1)^2 + 100\ (x_2 - x_1^2)]^2$:



Search pattern

Although this function is amenable to solution by more efficient methods, it is useful for purposes of comparison. Each of the solutions accepted in a 1000-trial search is shown (marked by symbols). The algorithm employed the adaptive step size selection scheme of equations (5) and (6) shown later in section 5.1. It is clear that the search is wide-ranging but ultimately concentrates in the neighbourhood of the optimum.

The following figure shows the progress in reducing the objective function for the same search:



Objective function reduction

Initially, when the annealing temperature is high, some large increases in f are accepted and some areas far from the optimum are explored. As execution continues and T falls, fewer uphill excursions are