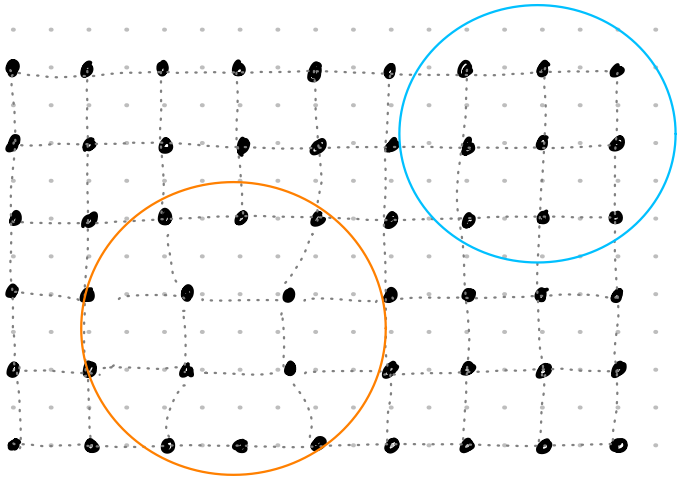
Simulated Annealing Methods

SA is a global optimization heuristic method inspired by the actual metallurgical process of annealing used to obtain metal crystallization in low energy states.

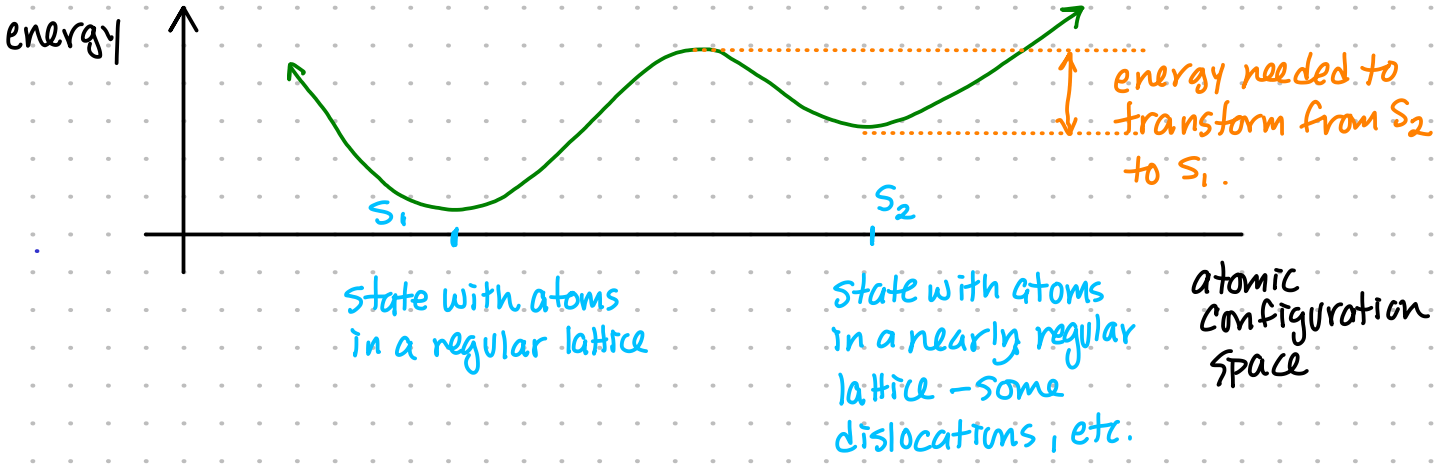
typical metals show many non uniform features in the structure



region of low stress
low energy

region of high stress
high energy

The metal cannot easily relieve the stress because it takes extra energy to move atoms from one configuration to another.
Trying to convey this idea in a one-dimensional plot:

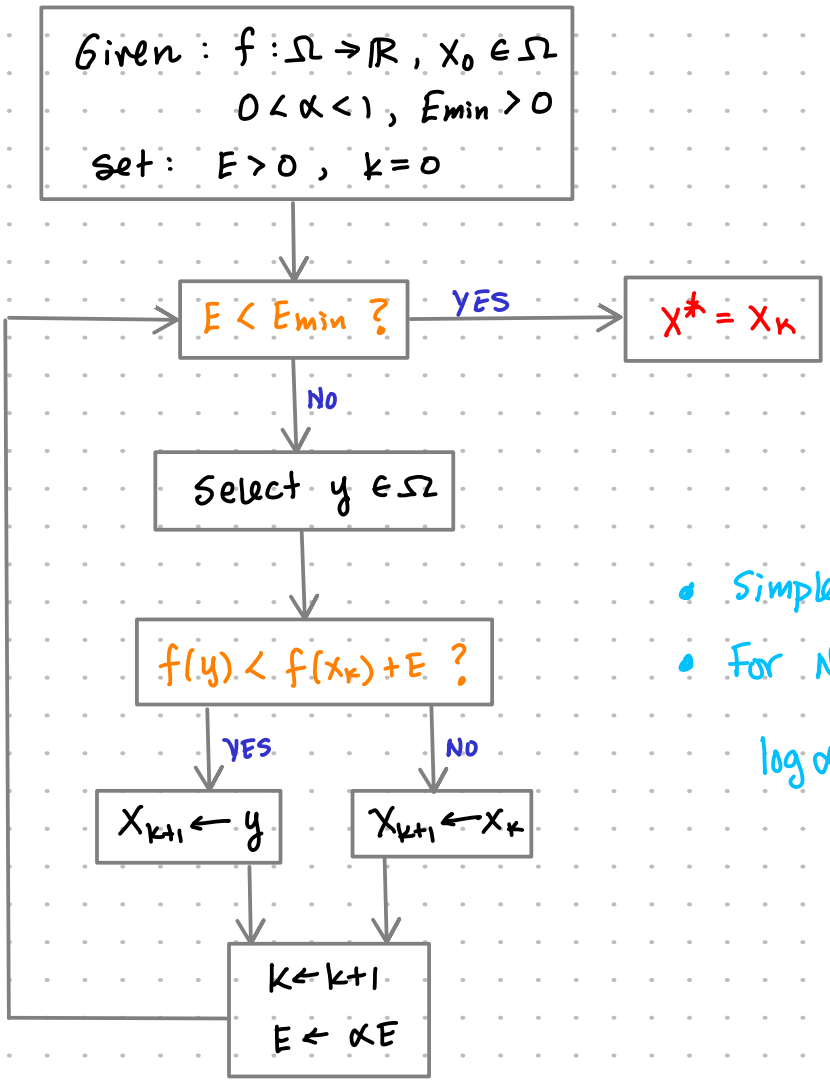


Adding energy ΔE to a system in state S_2 allows the atomic configuration to change to any neighboring state with energy up to $E(S_2) + \Delta E$. If ΔE is large enough, then the metal can transition "over the enery hill" into the basin of state S_1.

How can we use this concept as a global optimization strategy?

Using an added energy idea, we can allow for a non-decreasing iterate sequence $\{x_0, x_1, \dots\}$ not necessarily satisfying $f(x_k) > f(x_{k+1})$. Instead we require $f(x_k) + E_k > f(x_{k+1})$ with added energy E_k. If $E_k \to 0$ then in the limit we approach a monotone decreasing sequence (local optimal). But the non-monotonicity allows for the possibility of a global basin search.

A first Idea:



- Simple threshold
- For N iterations:

$$\log \alpha = \frac{1}{N} \log \frac{E_{min}}{E}$$

How to choose a new candidate y

y should (a) be a neighbor of x_k

(b) allow exploration of new regions

(1) $y = x_k + \varepsilon \qquad \varepsilon \in \mathcal{N}(0,\sigma)^n$

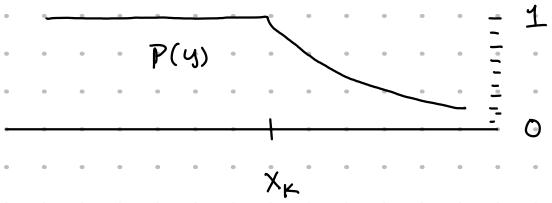
on average $y = x_k$ but arbitrarily long steps are possible.

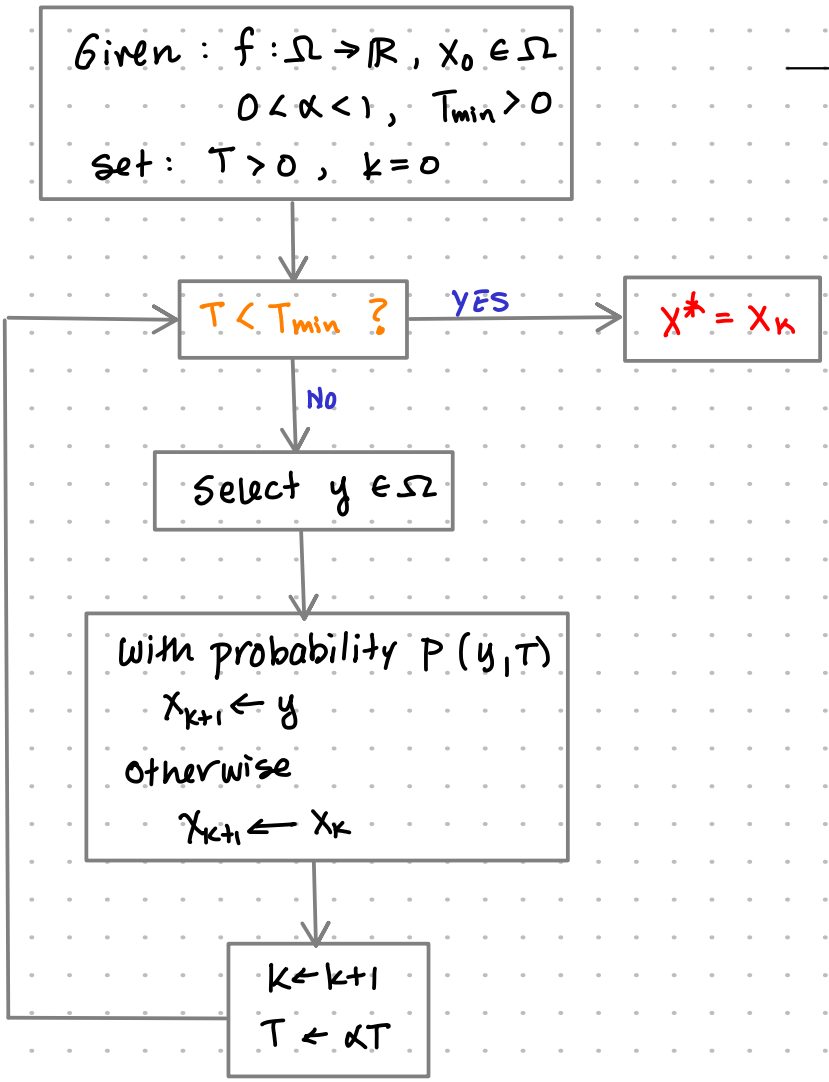
(2) $y = x_k + \varepsilon \qquad \|\varepsilon\| \leq \Delta$

limited step length but combinations of accepted steps can be arbitrarily long.

However...

A second Approach: let $P(y,T) := \min\left\{1, \exp\left[\frac{f(x_k)-f(y)}{T}\right]\right\}$





Some Implementation Notes

(1) When choosing $y \in \Omega$, it must be a neighbor of x_k

(a) $y = x_k + u$, u is random gaussian vector in $\mathbb{R}^n$

(b) y randomly chosen within box constraints around x_k

(2) Shrink factor α is typically close to one so that many iterations are used.

(3) Also keeping the best known solution is a good strategy.

(4) For $\alpha = 1$, $E \gg 1$, SA is much like random search.

For $\alpha = 0$, $E_{min} = 0$, SA is random search requiring descent

For $E_{min} = 0$ and y chosen over all of Ω, $x_k \to x^*$.

(5) The probability function should satisfy

- $P(y) = 1$ when $f(y) \leq f(x_k)$
- $P(y) < 1$ when $f(y) > f(x_k)$
- $\lim_{T \to 0} P(y,T) = \begin{cases} 1 & f(y) \leq f(x_k) \\ 0 & f(y) > f(x_k) \end{cases}$ (or $E \to 0$)