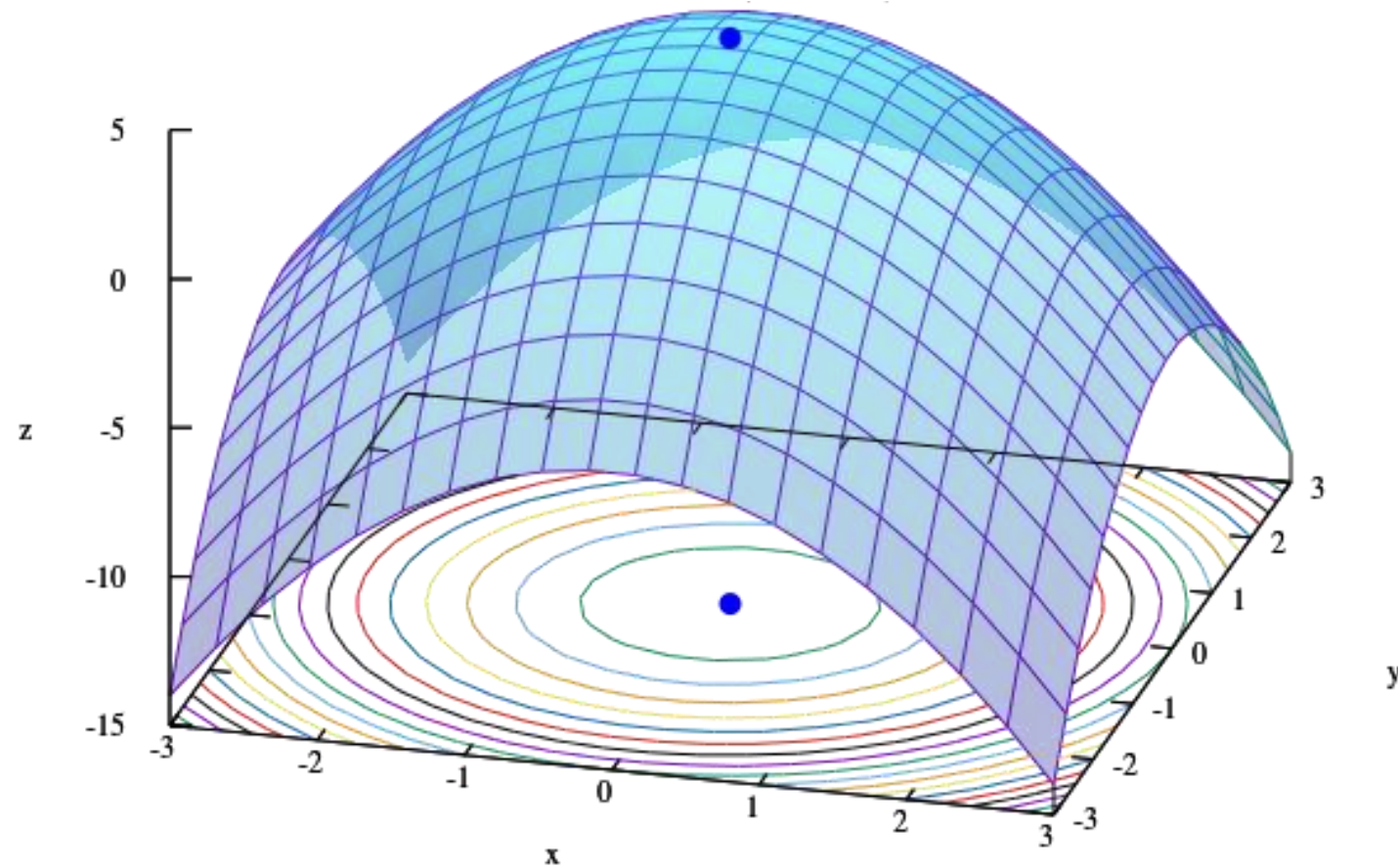


# Стохастическая оптимизация

Для дифференцируемых функций – градиентный спуск

Для недифференцируемых или неизвестных функций – локальный/глобальный поиск



# Глобальный поиск vs локальный поиск

Глобальный поиск – пытаемся посмотреть на все пространство состояний и найти лучшее (Монте-Карло, случайное блуждание).

Локальный поиск – используем текущее состояние системы и движемся от него.

Гибридные алгоритмы – гибриды локального и глобального поиска.

# Кросс-энтропийный метод

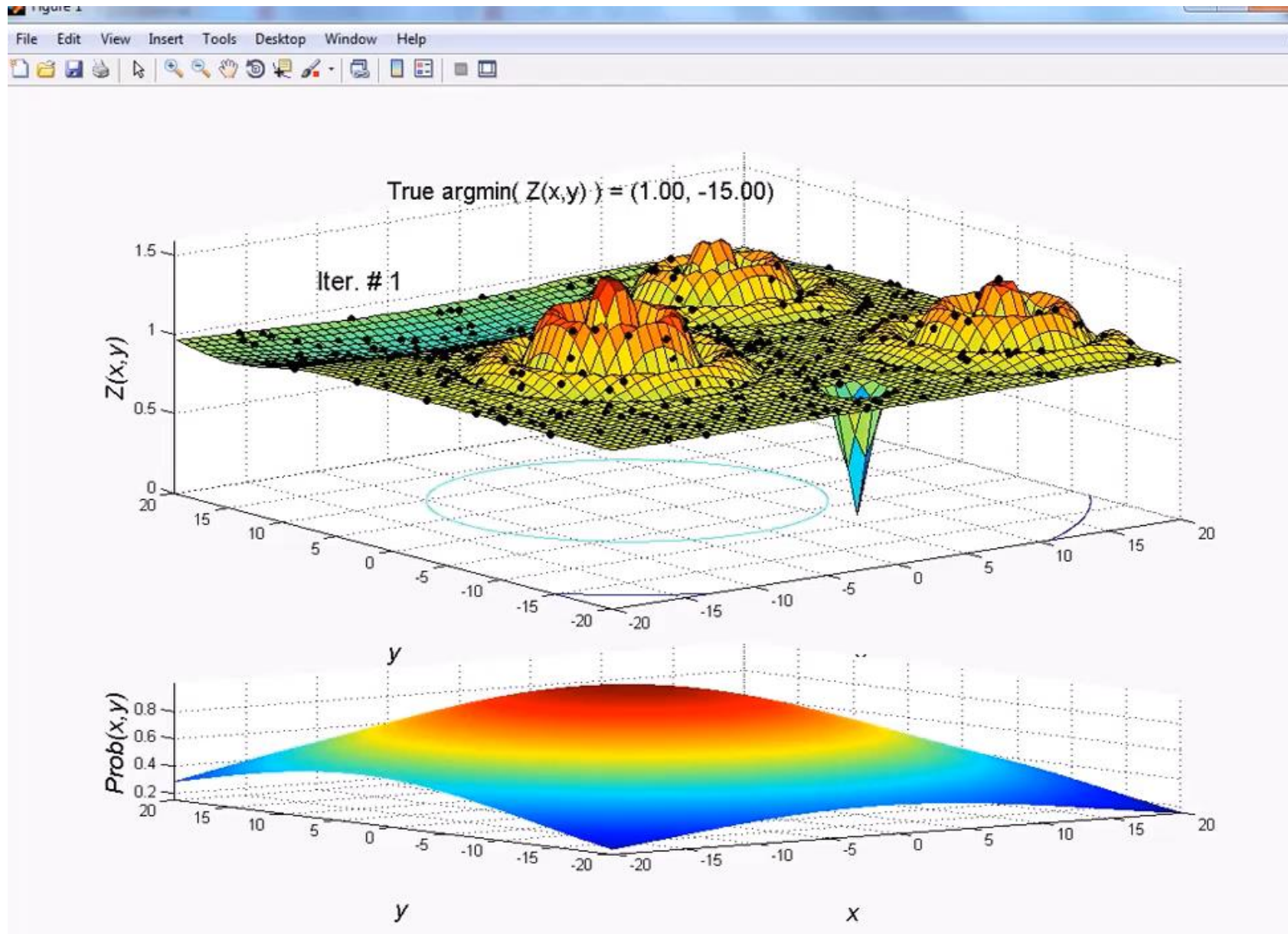
1. На шаге  $t = 1$  выбираем начальный вектор параметров распределения  $v_0$ .
2. Сгенерируем случайную выборку  $x_1 \dots x_N$  по распределению  $f(x; v_{t-1})$ .
3. Посчитаем новый вектор параметров основываясь на  $k$  лучших (по критерию  $H$ ) примерах из выборки:

$$v_t = \arg \max_u \frac{1}{k} \sum_{x_i \in \text{best } k} H(x_i) \frac{f(x_i; u)}{f(x_i; v_{t-1})} \log f(x_i; v_{t-1})$$

4. Если сошлось или достигли условия останавливаемся, иначе  $\rightarrow$  2.

\*для некоторых распределений  $\arg \max_u$  можно получить аналитически.

# Кросс энтропийный поиск



# Hill Climbing

Аналог **Gradient Descent**.

Считаем оптимизируемую функцию по нескольким возможным смещениям (например – изменения одной координаты/параметра) и идем в сторону наибольшего подъема.

Модификация – **Stochastic Hill Climbing**:

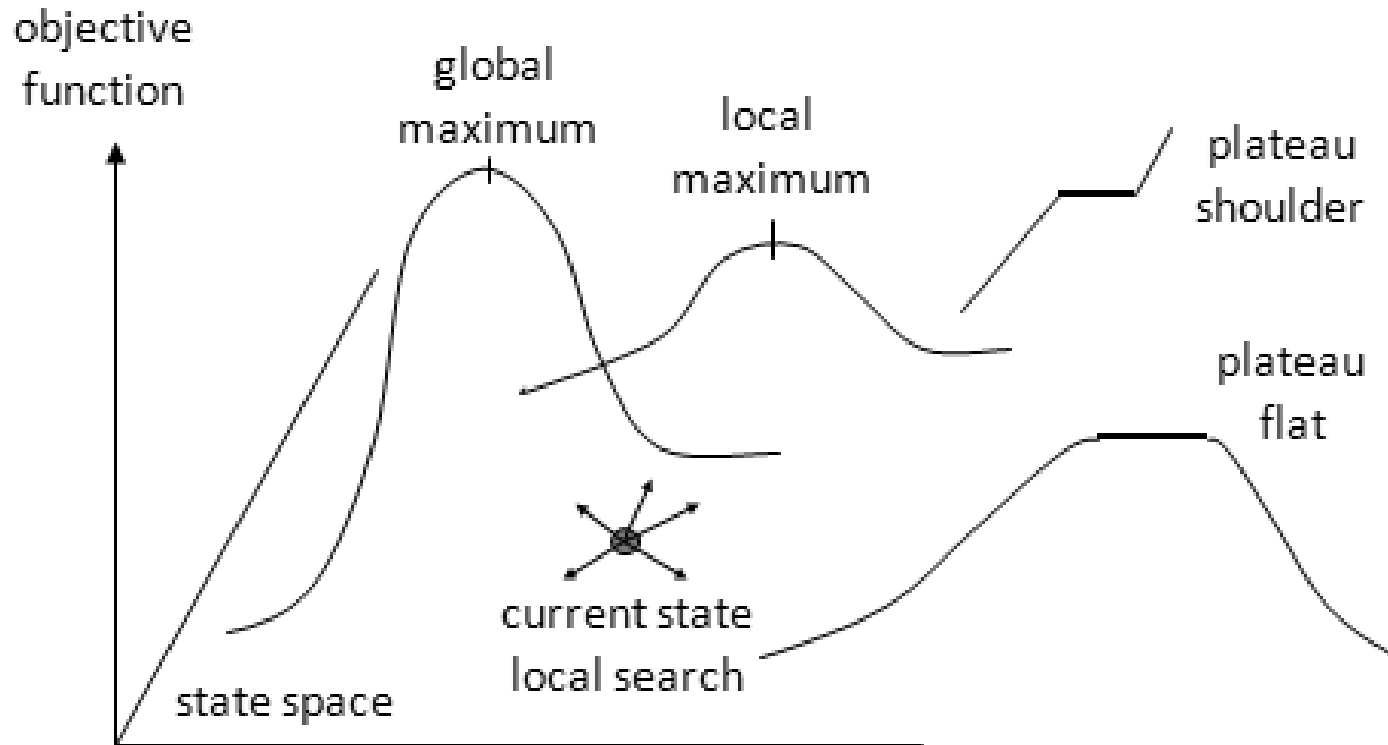
Идем с вероятностью, зависящей от значения функции (например – по softmax).

Модификация – **Tabu Search**:

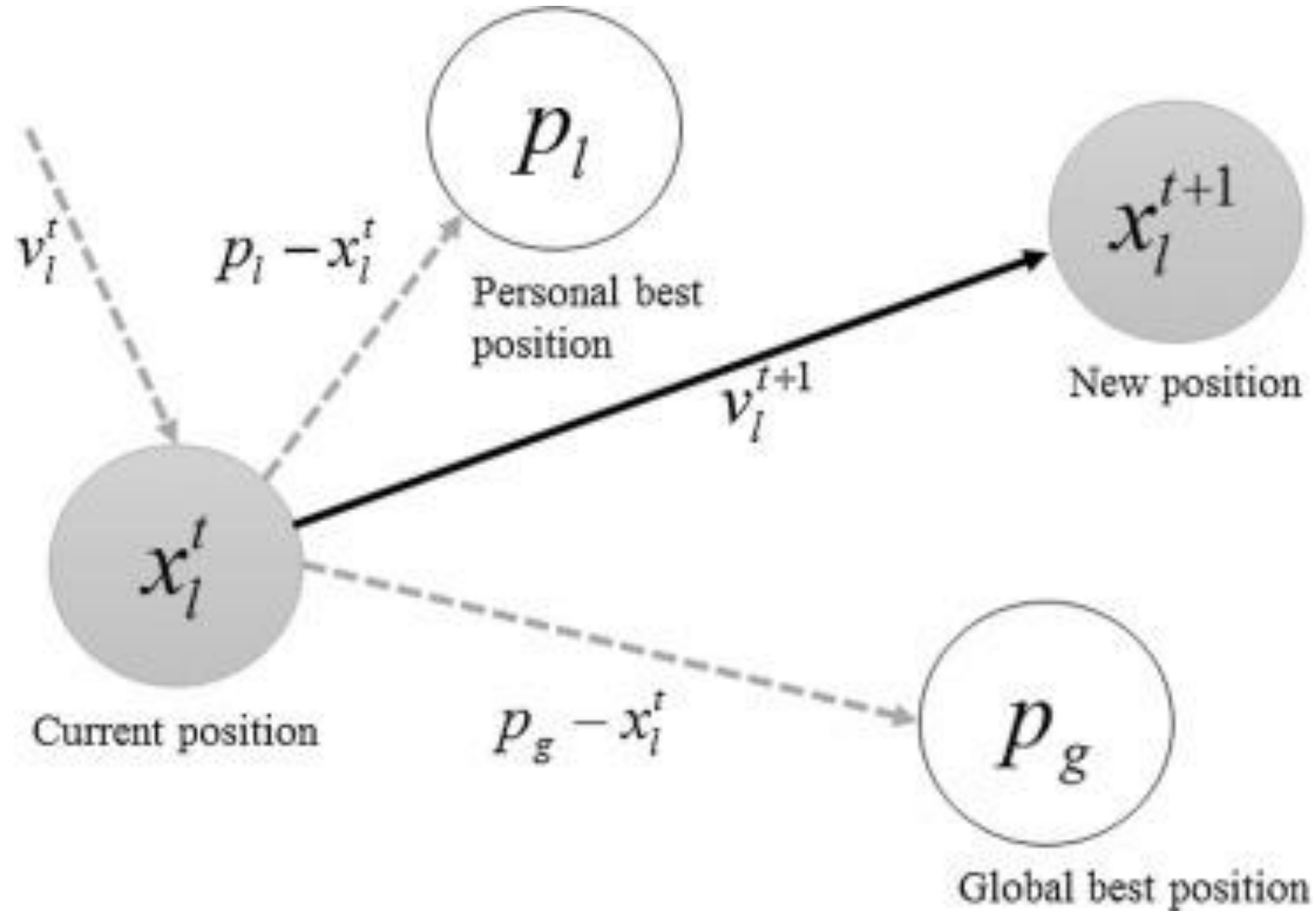
Запоминаем последние несколько позиций и в них не возвращаемся.

Модификация – **Particle swarm optimization**:

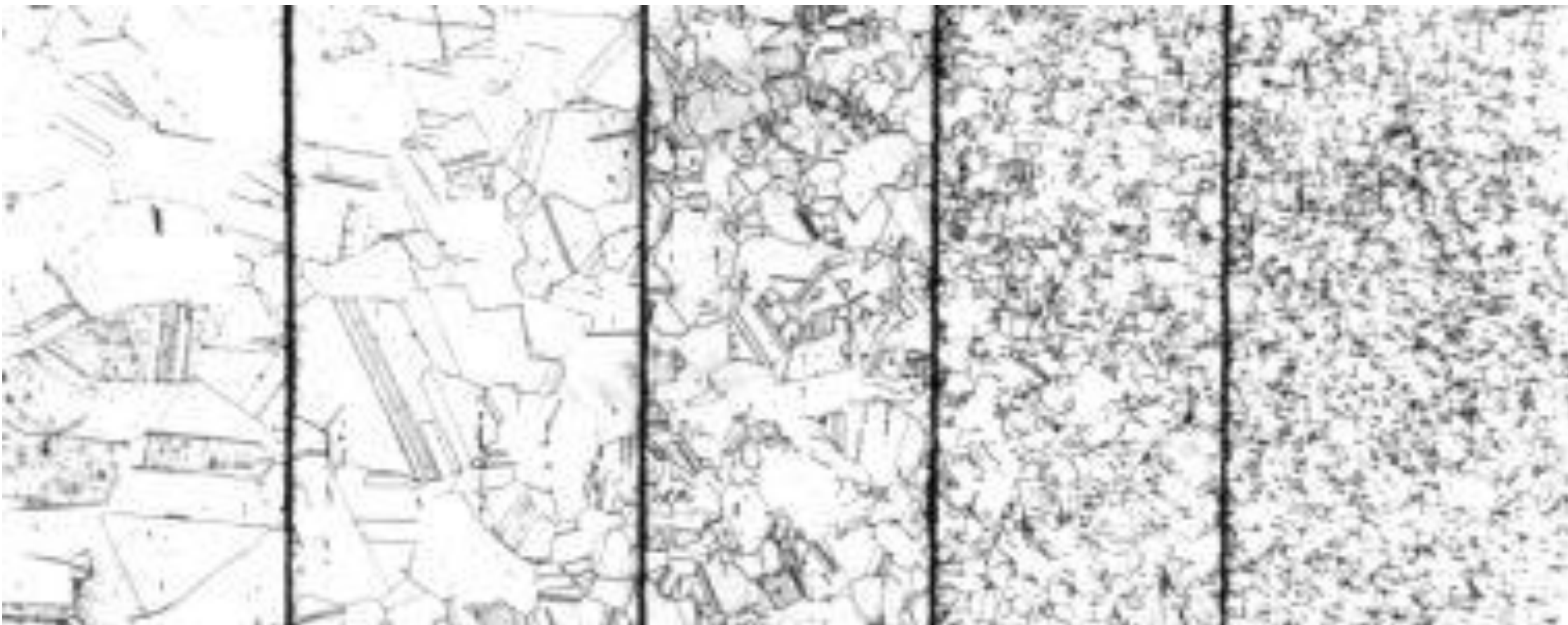
Запускаем много агентов, которые могут обмениваться информацией.



# Particle swarm optimization



# Отжиг (Annealing)



Структура метала с отжигом и без

# Отжиг (Simulated Annealing)

Введем в систему «температуру»:

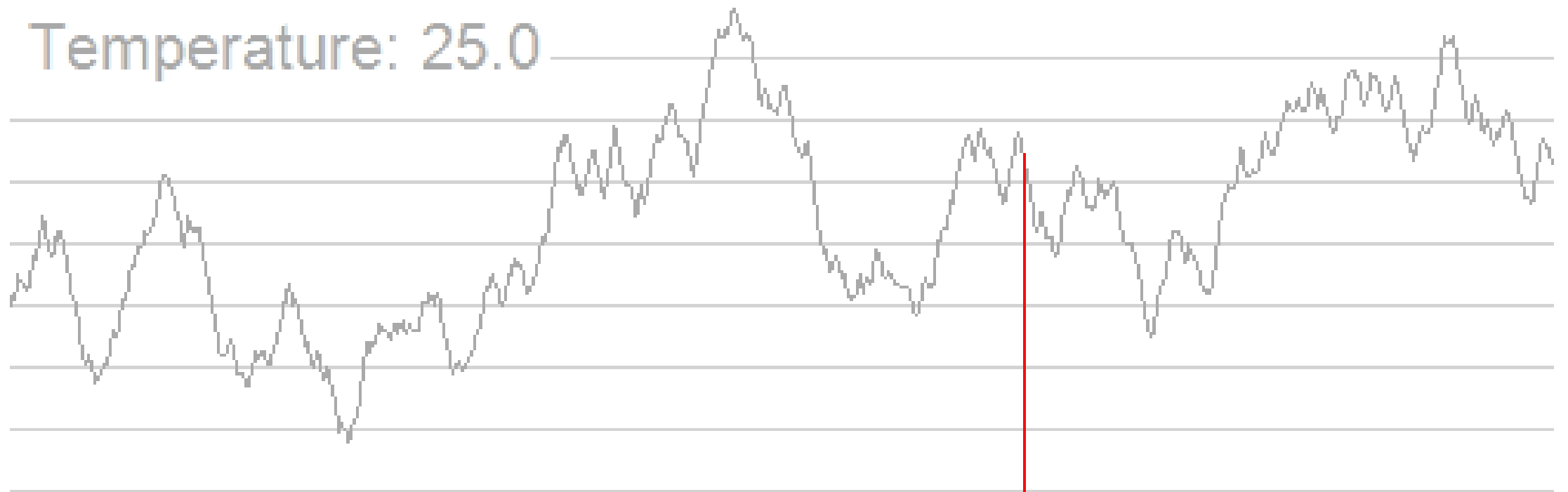
Например, для Softmax:

$$P(s_i) = \frac{e^{\frac{\Delta E(s_i)}{T}}}{\sum e^{\frac{\Delta E(s_j)}{T}}}$$

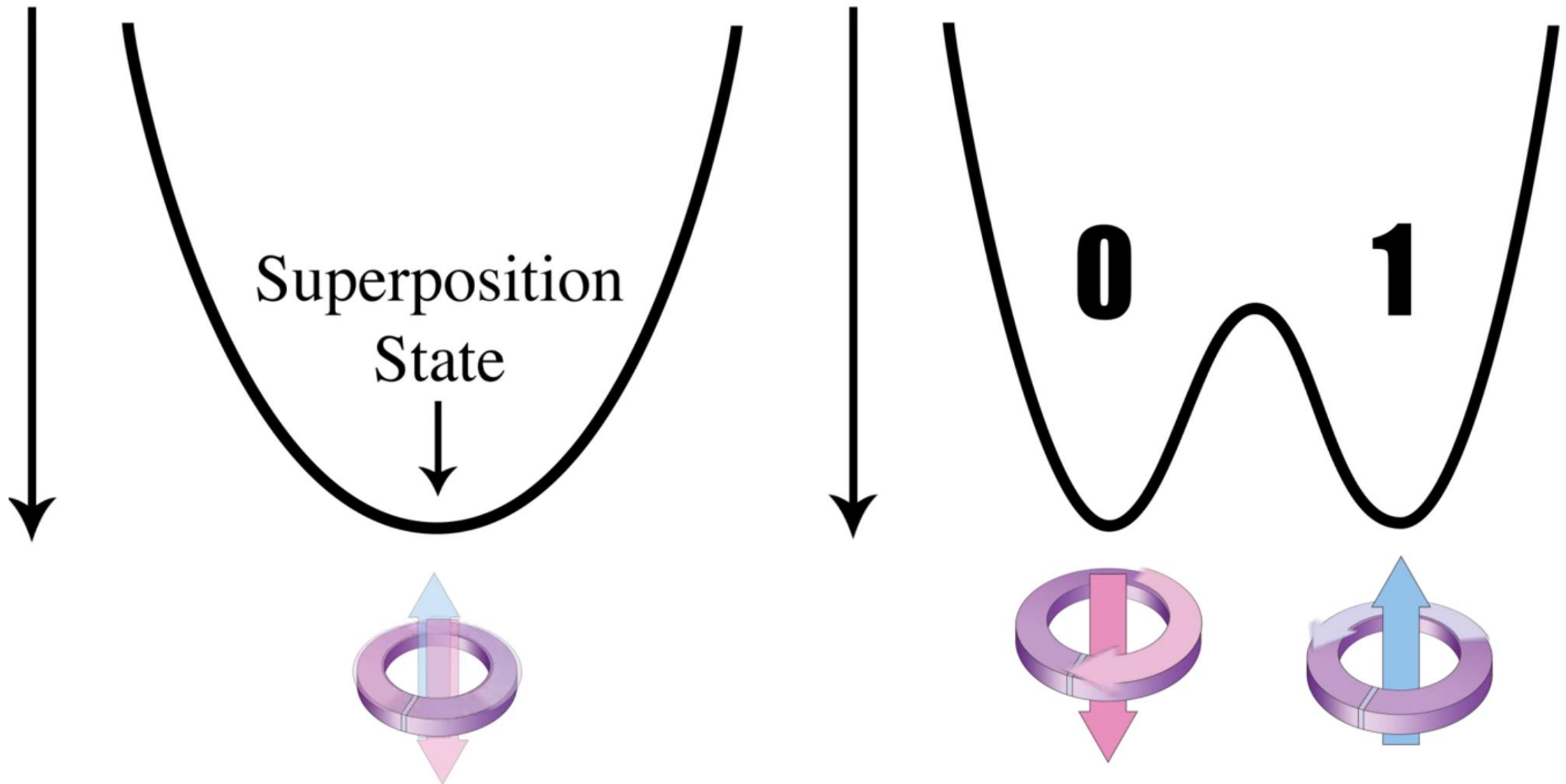
	T - Температура			
	10000	10	1	0.1
$\Delta E$	P( $\Delta E$ )	P( $\Delta E$ )	P( $\Delta E$ )	P( $\Delta E$ )
10	0.25016	0.43944	0.99325	1
5	0.25004	0.26653	0.00669	1.9287E-22
0	0.24991	0.16166	4.51E-05	3.7200E-44
-2	0.24986	0.13235	6.1E-06	7.6676E-53



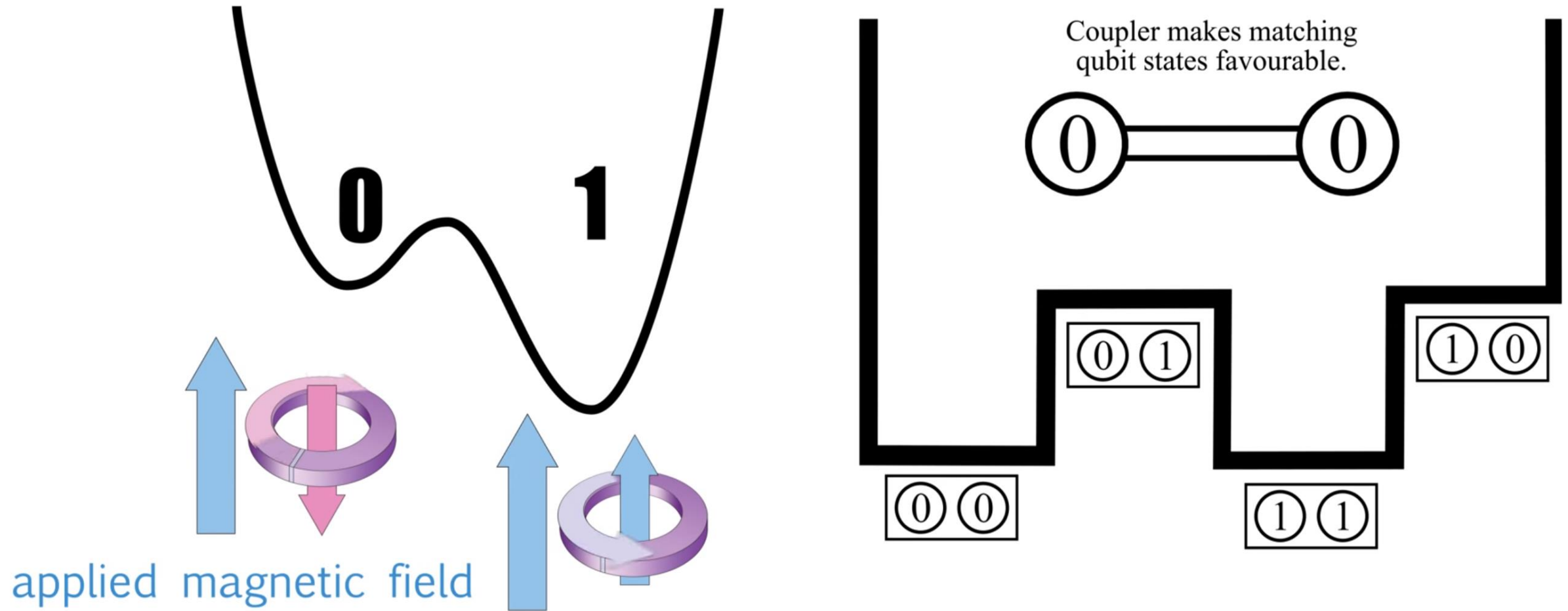
# Отжиг (Simulated Annealing)



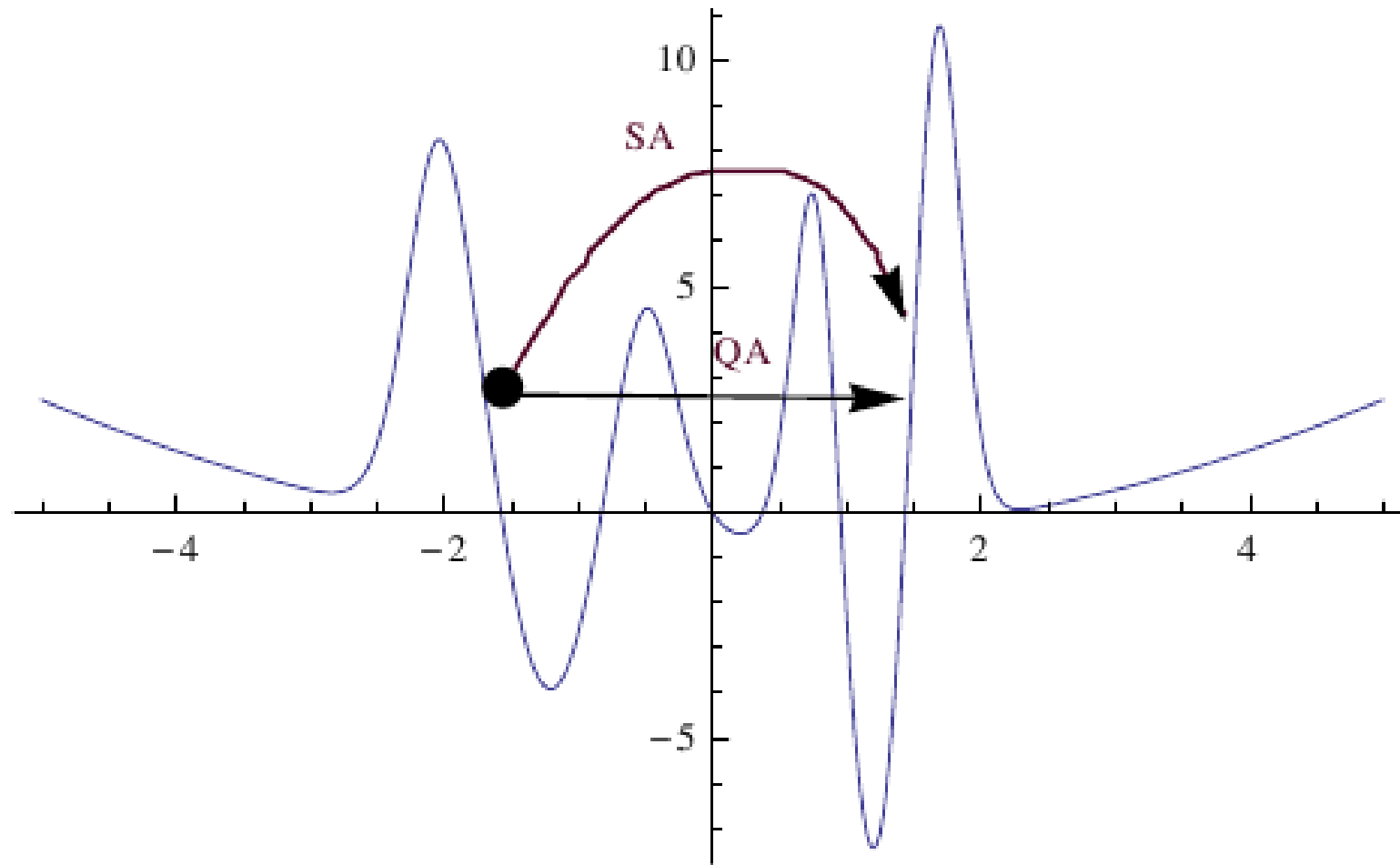
# Квантовый отжиг



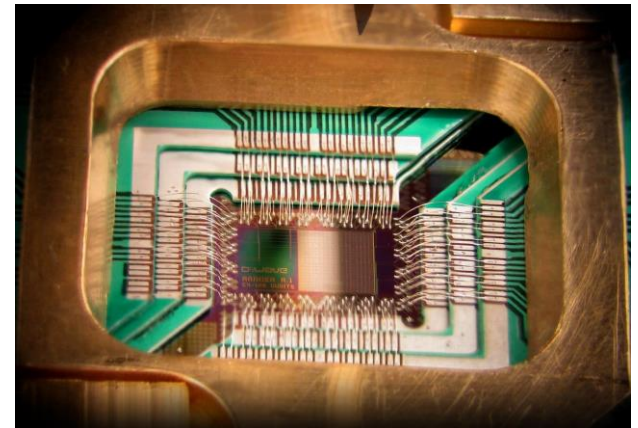
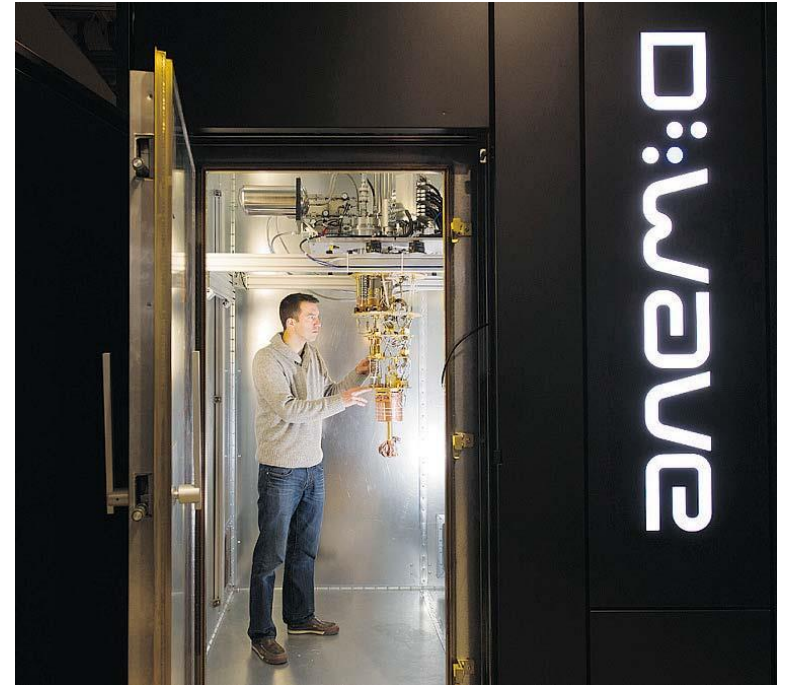
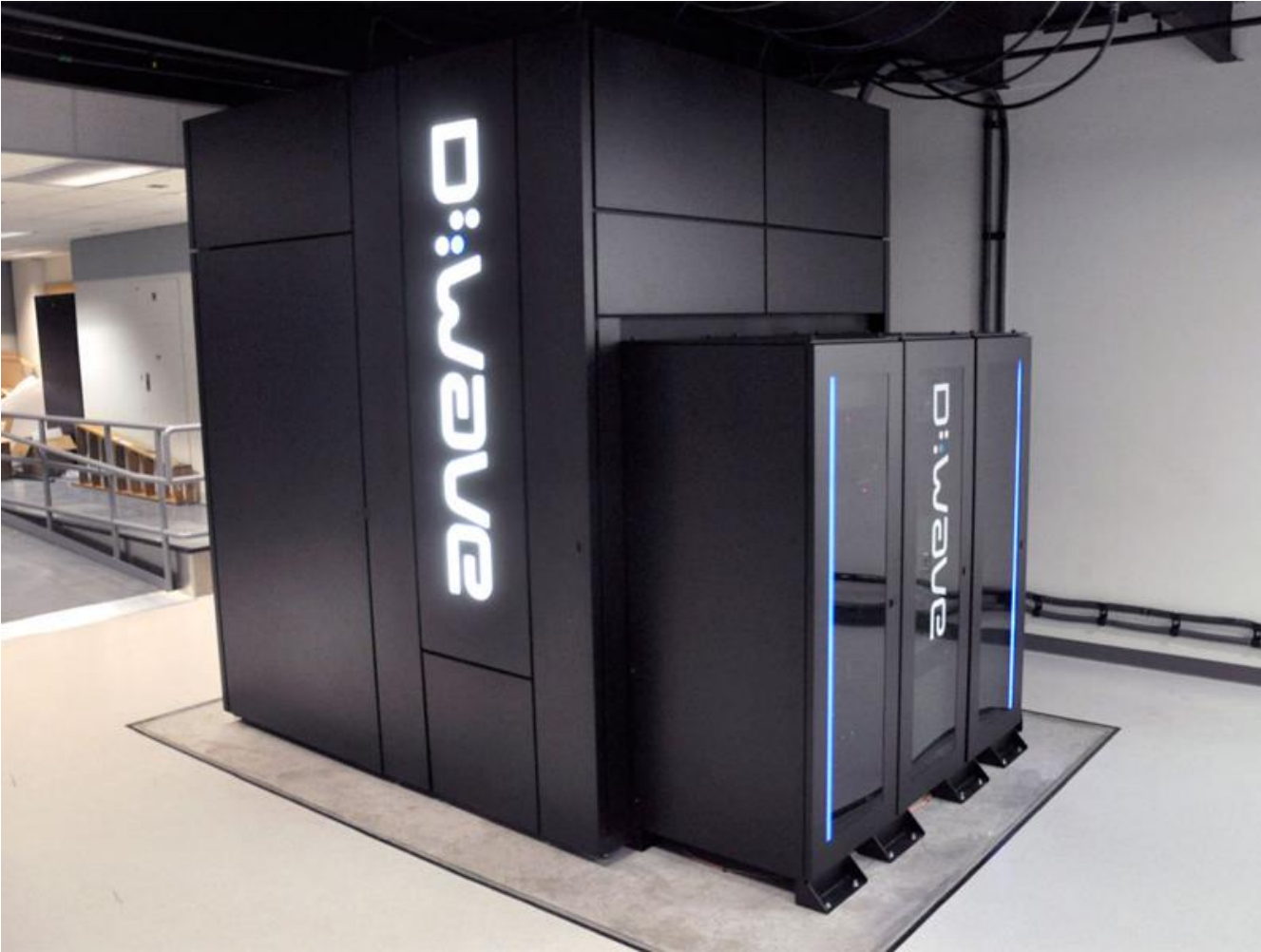
# Magnets and couplers



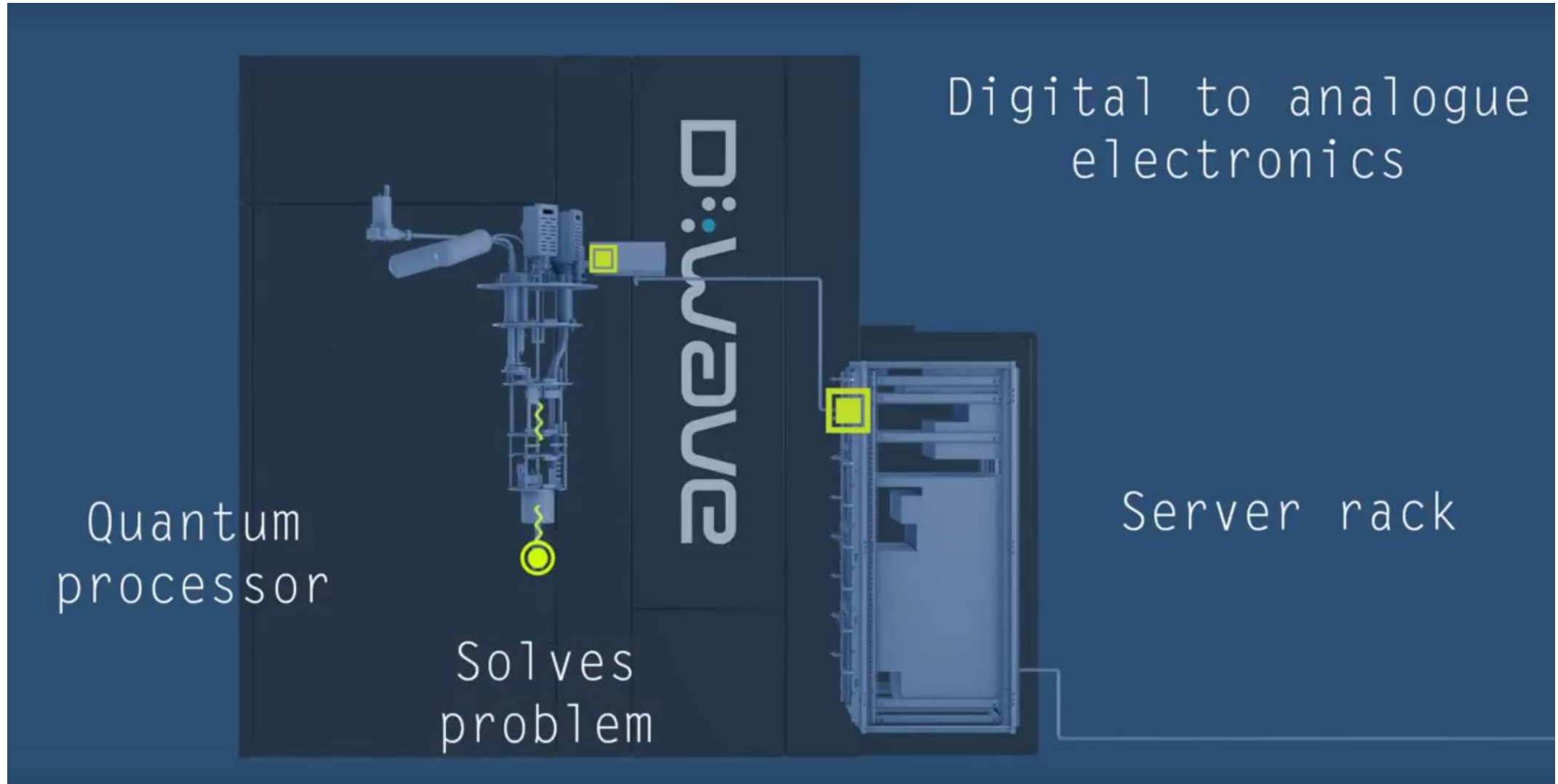
# Туннельный эффект



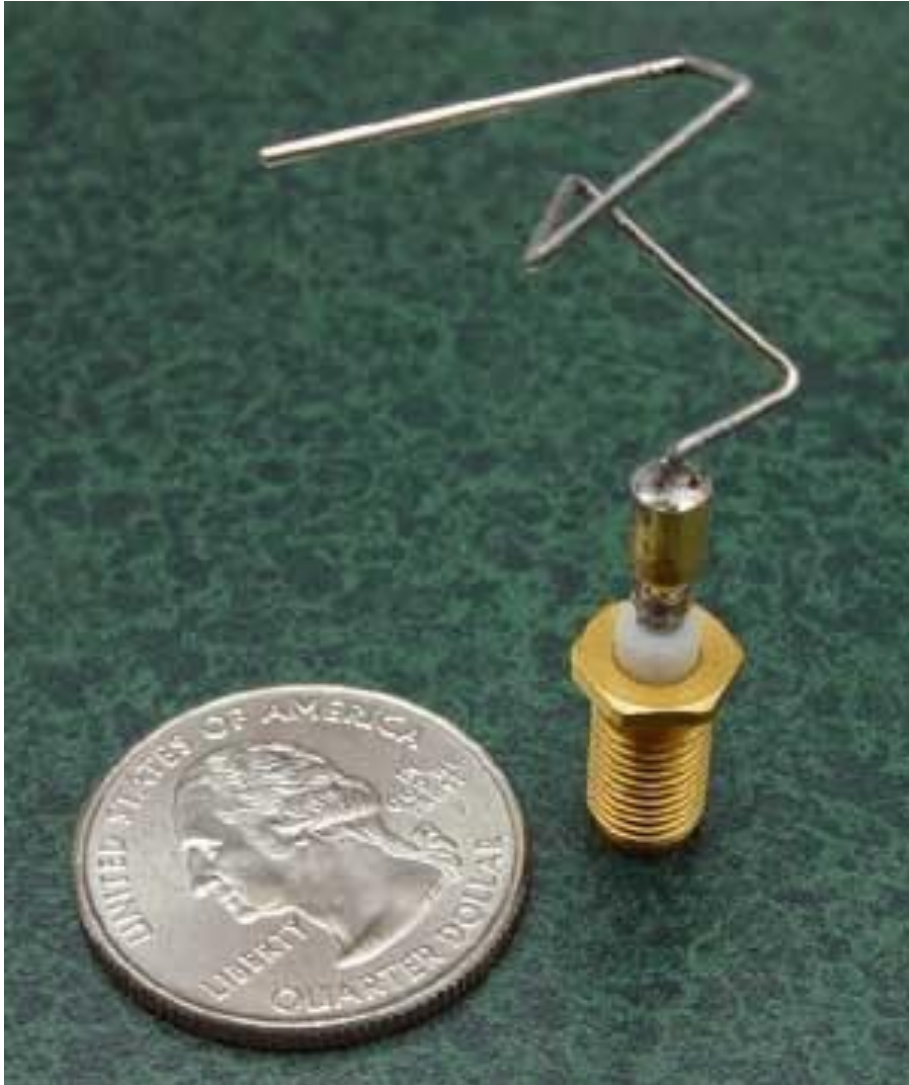
# D-Wave



# D-Wave



# Генетический алгоритм

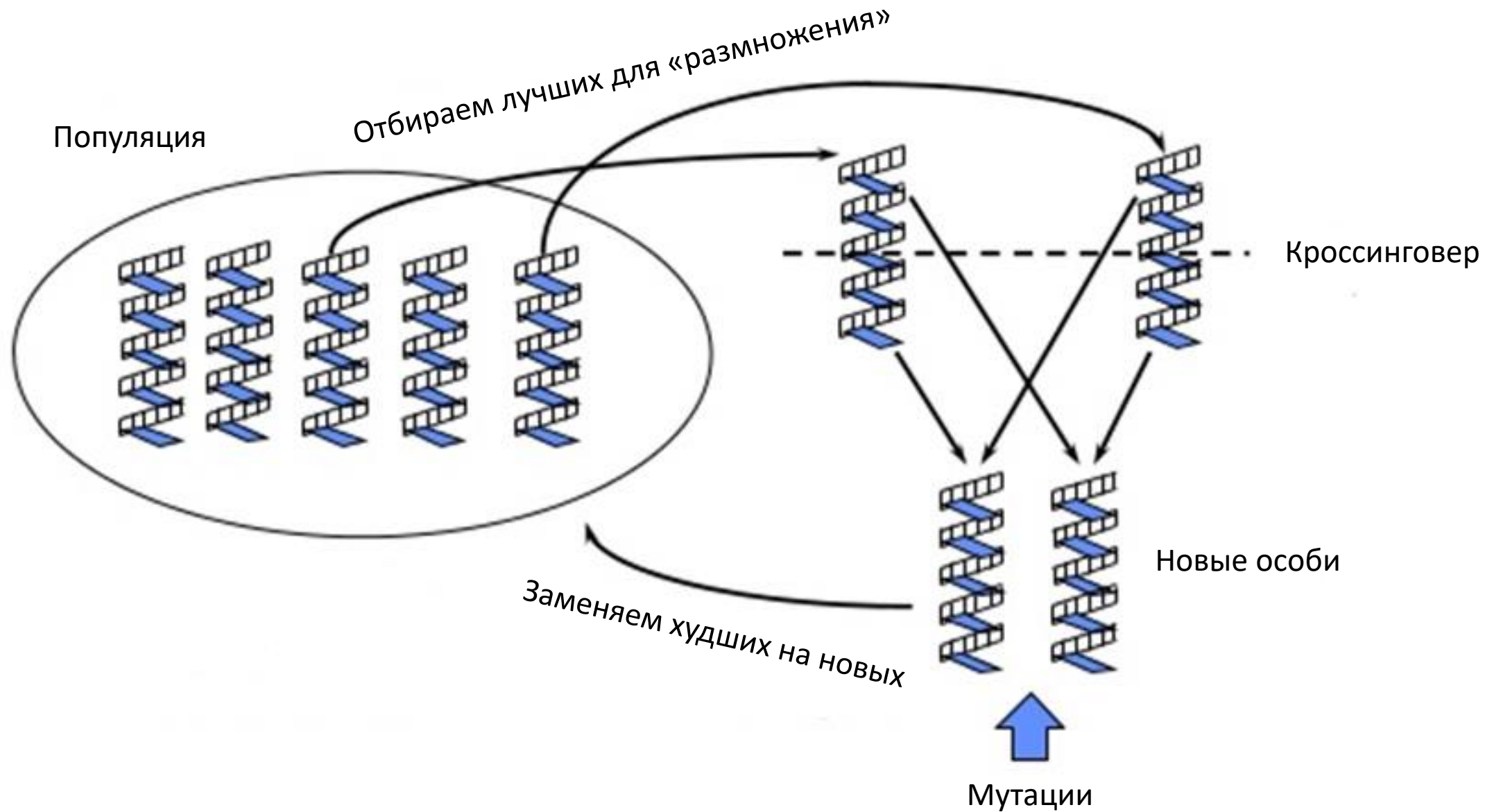


## **“Evolved antenna”:**

полученная с помощью симуляции —  
генетического алгоритма —  
антенна, использующаяся на спутниках,  
измеряющих магнитосферу земли,  
запущенных в 2006 году — миссия Space Technology 5.



# Генетический алгоритм





# No Free Lunch Theorem

$$\sum_f P(d_m^y | f, m, a_1) = \sum_f P(d_m^y | f, m, a_2)$$

$d_m^y \{y_1, y_2 \dots y_m\}$  - последовательность полученных значений оптимизируемой функции.

$$y = f(x)$$

$a_1, a_2$  — алгоритмы оптимизации.