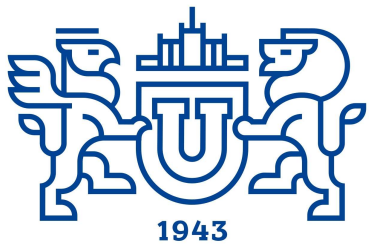




Russian Supercomputing Days

Семинар «Квантовые вычисления»

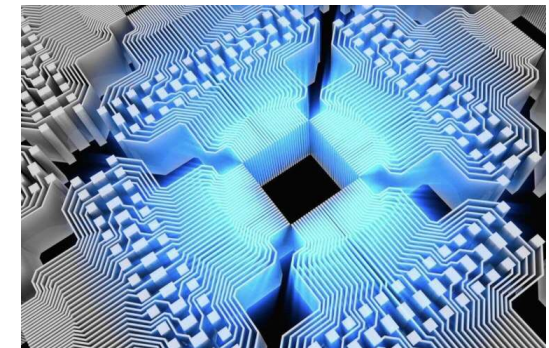
Применение квантовых аннилеров для решения научно-практических и бизнес-задач многопараметрической оптимизации: текущее состояние и перспективы



ЮУрГУ

Николай Владимирович Малетин,
Лаборатория «Квантовая инженерия света»
Южно-Уральского Государственного Университета
MaletinNV@my.msu.ru, Telegram: @MaletinNV

26 сентября 2023 года
Москва

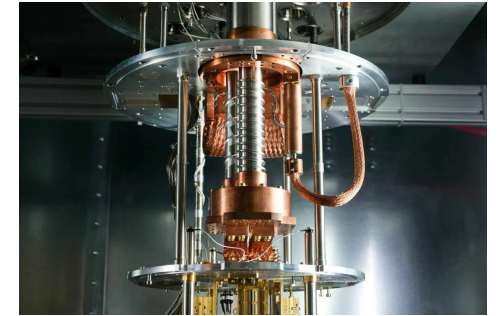


Применение квантовых аннилеров для решения научно-практических и бизнес-задач многопараметрической оптимизации: текущее состояние и перспективы

Обсуждаемые вопросы:

1. В чем преимущества квантовых аннилеров перед квантовыми компьютерами вентиляного типа с точки зрения решения задач оптимизации.
2. Текущие возможности современных квантовых аннилеров: примеры практического применения в бизнеса и оценка вычислительной мощности.
3. Основные математические особенности разработки алгоритмов решения многопараметрических задач оптимизации для квантовых аннилеров.
4. Примеры разработки квантовых алгоритмов для решения некоторых научно-практических задач в лаборатории «Квантовая инженерия света» ЮУрГУ.
5. Прогнозы и перспективы.

$$H(q_1, q_2, \dots, q_N) = \sum_{i=1}^N \sum_{j \geq i}^N \alpha_{ij} q_i q_j$$



Практические возможности квантовых вычислений сегодня = Квантовые вычислители с шумами + Алгоритмы, устойчивые к вычислительным ошибкам/ шумам

UQC – Universal Quantum Computer

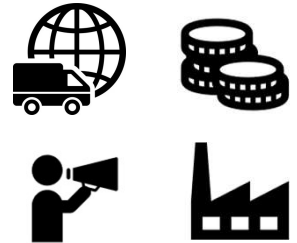
NISQ – Noisy Intermediate Scale Quantum devices

шумы => вычислительные ошибки => ~~UQC~~

UQC	VS	NISQ
Реализация		
на горизонте 20+ лет*		уже есть и активно развиваются
Количество «алгоритмических» кубитов		
> 1000		от 100 до 1000
Уровень шума		
предельно низкий		сравнительно высокий
Возможно использовать		
для реализации любых квантовых алгоритмы		для реализации квантовых алгоритмов, устойчивых к вычислительным ошибкам/ шумам !

Задачи оптимизации для решения на NISQ-устройствах, обсуждаемые в литературе

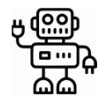
Комбинаторная оптимизация и задачи оптимального управления:
логистика, финансы, маркетинг, планирование и управление сложными производственными процессами.



Моделирование и прогнозирование сложных систем (природных, экономических, технологических):

- глобальный прогноз погоды,
- прогноз дефолтов,
- сейсморазведка,
- томография.



Обучение систем AI/ML 

*Источник: BSG <https://www.bcg.com/publications/2019/quantum-computers-create-value-when>

Квантовый отжиг – лучший на текущий момент квантовый метод оптимизации

для компьютеров
адиабатического типа:
Quantum Annealing (QA)



Базовые
квантовые
методы
оптимизации



для компьютеров вентиляного типа
(GTQC - gate-type quantum computers):
Quantum Approximate Optimization Algorithm (QAOA),
Variational Quantum Eigensolver (VQE)

① QA D-Wave – QC №1 for business – первый и пока единственный квантовый компьютер с помощью которого уже решаются практические задачи оптимизации.

D:wave
The Quantum Computing Company™



5760 кубитов D-Wave vs 433 кубитов IBM
QA на порядок опережают GTQC по количеству кубитов.

③ При отсутствии доступа к квантовым аннилерам для приближенного решения можно использовать компьютерные эмуляторы систем Изинга (<https://opg.optica.org/oe/fulltext.cfm?uri=oe-27-7-10288&id=408024>), достаточно мощные реализации которых есть в России и распараллеливаемые алгоритмы приближенного решения задач QUBO.

② На сегодняшний день
Качество решений QA > Качества решений QAOA

при использовании сопоставимых вычислительных мощностей (экспериментальное исследование: <https://arxiv.org/abs/2301.00520>).

По мнению некоторых экспертов в решении задач оптимизации QA всегда будут превосходить GTQC по качеству решений при использовании сопоставимых вычислительных мощностей (<https://phys.org/news/2023-05-team-quantum-advantage-optimization-problems.html>).

QA D-Wave уже используется при решении достаточно масштабных практических и квази-практических задач оптимизации в бизнесе

2021г., компании Groovenauts и Mitsubishi Estate: оптимизации сбора отходов в районе Токио Маруноути

Параметры	До	После	Эффект
Маршрут сбора отходов, км	2296	1004	-56%
Количество транспортных средств	75	31	-59%
Выбросы CO2			-57%

2022 г., компания SavantX: оптимизация логистики в порту Лос-Анджелеса

Параметры	До	После	Эффект
Загрузка крана, %	45	72	60%
Время возврата грузовика, мин.	66	58	-12%
Среднее расстояние между кранами, м.	8900	6200	-30%

Источник: D-Wave Systems,
<https://www.dwavesys.com/learn/customer-success-stories/>

2022 год., компания Multiverse Computing: оптимизация портфеля

Моделирование управления портфелем банка BBVA на исторических данных с учетом транзакционных издержек.

- размер портфеля – 56 активов;
- моделируемый период – 8 лет;
- полученный коэффициент Шарпа – 12,16;
- время расчета на D-Wave – 171 секунда;
- время расчета на обычном ПК – 32,5 часа.

Результаты обнадеживают,
но это еще не квантовое преимущество, т.к.

- 1) на самых мощных суперкомпьютерах расчет будет быстрее,
- 2) используемые в D-Wave алгоритмы гибридные.

Крупнейшие клиенты D-Wave:

Бизнес-корпорации:
Google, Mastercard, GE, Volkswagen, Toyota, Mitsubishi.

Научные учреждения:
NASA, LANL, FZJ.



«Заходите к нам на «ОГОНЁК»,
ОТТОЖЖЁМ КВАНТОВО!»
Алан Баратс,
генеральный директор
«D-Wave Systems»,

где-то в параллельной вселенной

Математика квантового отжига. Основное

алгоритм
градиентного
спуска



За счет эффекта квантового туннелирования вероятность достижения глубоких точек минимума при использовании QA больше, чем при использовании классических алгоритмов.

Квантовый аннилер позволяет решать лишь только один тип задач – задачи QUBO, но к ним могут быть сведены многие задачи оптимизации

Задача QUBO (Quadratic Unconstrained Binary Optimization – квадратичной неограниченной бинарной оптимизации): **Найти координаты глобального минимума функции**

$$H(q_1, q_2, \dots, q_N) = \sum_{i=1}^N \sum_{j \geq i}^N \alpha_{ij} q_i q_j, \quad (1)$$

где q_1, q_2, \dots, q_N – бинарные переменные, принимающие значение 0 или 1, α_{ij} – постоянные коэффициенты, кодирующие условие задачи.

Бинарная дискретизация произвольной переменной $x \in [x_{min}, x_{max}]$: $x = x_{min} + \frac{x_{max} - x_{min}}{2^N - 1} \sum_{k=1}^N 2^{k-1} q_k. \quad (2)$

① Метод QUBO

достаточно **универсален**, т.к. к задаче QUBO можно свести многие типы задач оптимизации в частности задачу оптимизации любой функции с линейными ограничениями.

② Существует

универсальный метод сведения к задаче QUBO задачи оптимизации любой функции с линейными ограничениями с помощью введения дополнительных бинарных переменных.

③ Есть стандартный набор

универсальных алгоритмов приближенного решения больших задач QUBO посредством понижения их размерности/декомпозиции на задачи меньших размеров.

④ Метод инвариантен к

типу компьютера, т.к. задачи QUBO можно решать и на **вентильном квантовом компьютере** с помощью алгоритма QAOA, но QA решает их лучше.

Аппаратные ограничения + недостатки универсальных алгоритмов...

Полносвязная задача QUBO – задача QUBO, все коэффициенты α_{ij} в гамильтониане $H(q_1, q_2, \dots, q_N)$ которой, кроме, возможно, α_{00} , не равны нулю. Большая часть практически интересных задач являются либо полносвязными, либо близкими к полносвязным

1 Квантовая вычислительная мощность D-Wave ограничена

5760 физических кубитов + 15 связей на кубит \approx 124 алгоритмических кубита в смысле решения полносвязных задач QUBO

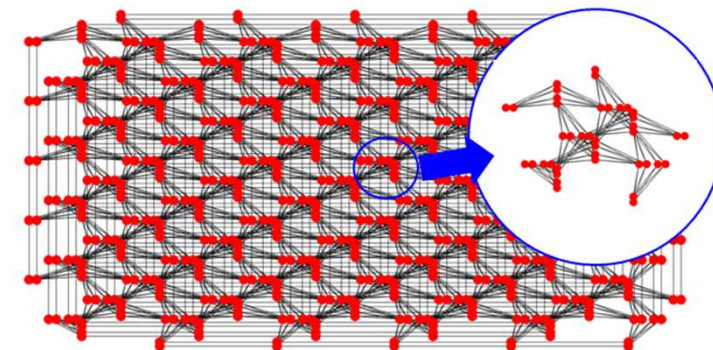
1 бинарная переменная \approx 1 «алгоритмический» кубит

2 Аппаратное ограничение на диапазон α_{ij}

$$\frac{\max_{i,j} |\alpha_{ij}|}{\min_{i,j} |\alpha_{ij}|} < 10^3$$

3 Универсальный метод сведения к задаче QUBO задачи оптимизации любой функции с линейными ограничениями **неприменим для сложных задач**, т.к. уже для не очень сложных функций количество дополнительных бинарных переменных быстро и нелинейно растет с ростом количества основных бинарных переменных.

4 Универсальные алгоритмы приближенного решения больших задач QUBO посредством понижения их размерности/ декомпозиции на задачи меньших размеров **плохо применимы для больших задач**, т.к. точность получаемых решений ухудшается с ростом масштаба задачи.



Топология Pegasus QA D-Wave Advantage

... превращают разработку алгоритма в серьезную математическую задачу

Основные этапы математической разработки QA-алгоритма решения масштабной задачи оптимизации:

1. Декомпозиция исходной задачи на более простые подзадачи и/или подзадачи меньшей размерности.
2. Перформулировка подзадач в формат QUBO с минимально возможным количеством бинарных переменных.
3. Разработка декомпозиционно-итерационных алгоритмов для решения задач QUBO, полученных на предыдущем шаге, с учетом внутренних математических симметрий исходных подзадач.

Применение QS в задачах вычислительного материаловедения

1 Симулирование квантовых систем на GTQC

Изучение квантовой системы посредством ее моделирования на другой квантовой системе (GTQC).
Пионерская идея Ричарда Фейнмана, предложенная им в 1982 г.

2 Параметризация модельных потенциалов межмолекулярного взаимодействия используя QA

Расчет инженерно-физических характеристик материалов

(предел текучести, пластичность, прочность на разрыв и пр.)
требует **моделирования систем**, состоящих из $\sim 10^9$ частиц.

Шаг 1. Численно решаем уравнение

Шредингера методами DFT для системы из 10^2 - 10^3 частиц на интервалах 10^2 - 10^3 пикосекунд. Получаем энергию системы E_m^{obs} в $\sim 10^1$ - 10^2 различных конфигурациях m .

Шаг 2. Параметризуем модельный полуэмпирический потенциал взаимодействия $U(\vec{\theta}, r)$,

минимизируя функцию невязки

$$Err^2(\vec{\theta}) = \sum_{m=1}^M (E_m^{mod}(\vec{\theta}) - E_m^{obs})^2. \quad (3.1)$$

где $E_m^{mod}(\vec{\theta}) = \sum_{i=1}^K \sum_{j>i}^K U(r_{ij}^{(m)})$, (3.2)

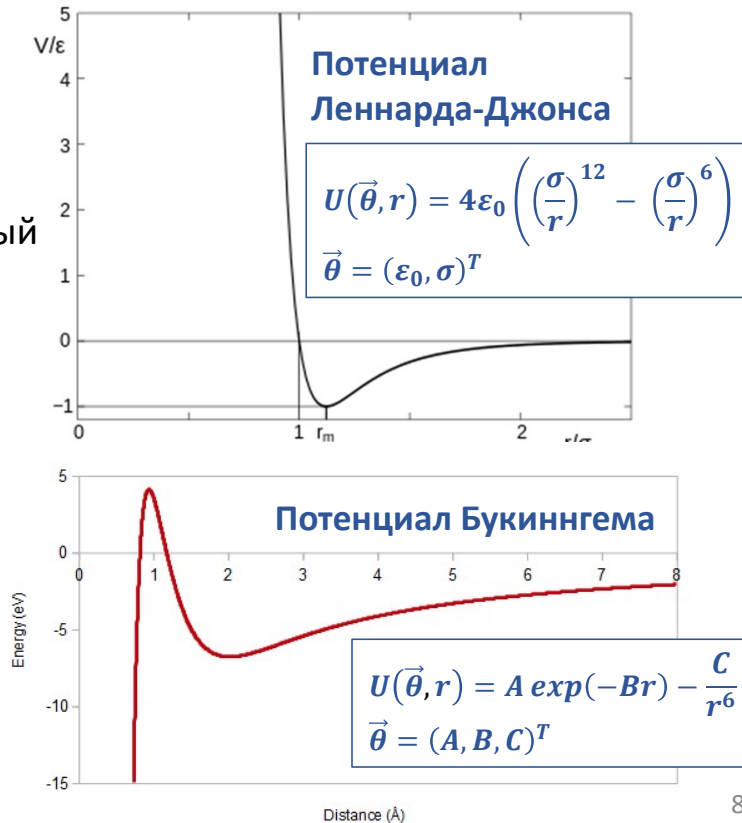
$\vec{\theta}$ – вектор параметров потенциала θ_i

Шаг 3. Моделируем систему

из $\sim 10^9$ частиц, используя полученный на Шаге 2 потенциал $U(\vec{\theta}, r)$.
Получаем инженерно-физические характеристики.

сложная задача оптимизации

Количество параметров θ_i в однокомпонентной системе от 2 до ~ 10 ,
в многокомпонентной системе ~ 10 - 10^2 ,



Математика квантового отжига. Стандартные методы

Потенциал Леннарда-Джонса

$$U(r) = 4\varepsilon_0 \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right), \quad (4)$$

$$E_m^{mod}(\varepsilon_0, \sigma) = \sum_{i=1}^K \sum_{j>i}^K U(r_{ij}^{(m)}), \quad (5)$$

$$Err^2(\varepsilon_0, \sigma) = \sum_{m=1}^M (E_m^{mod}(\varepsilon_0, \sigma) - E_m^{obs})^2 \rightarrow \min. \quad (6)$$

Шаг 1. Замена переменных: $a_m = \sum_{i=1}^K \sum_{j>i}^K \frac{1}{(r_{ij}^{(m)})^{12}}$,

$$b_m = \sum_{i=1}^K \sum_{j>i}^K \frac{1}{(r_{ij}^{(m)})^6}, \quad l = 4\varepsilon_0\sigma^{12}, \quad s = -4\varepsilon_0\sigma^6. \quad (7)$$

Шаг 2. Бинарная дискретизация по формуле

$$x = x_{min} + \frac{x_{max} - x_{min}}{2^{N-1}} \sum_{k=1}^N 2^{k-1} q_k, \quad (2):$$

$$l = \sum_{k=0}^{N_b^{(l)}} \beta_k^{(l)} q_k^{(l)}, \quad s = \sum_{k=0}^{N_b^{(s)}} \beta_k^{(s)} q_k^{(s)}. \quad (8)$$

В итоге:

$$Err^2(\vec{q}^{(s)}, \vec{q}^{(l)}) = \sum_{m=1}^M (a_m \sum_{k=0}^{N_b^{(l)}} \beta_k^{(l)} q_k^{(l)} + b_m \sum_{k=0}^{N_b^{(s)}} \beta_k^{(s)} q_k^{(s)} - E_m^{obs})^2, \quad (9)$$

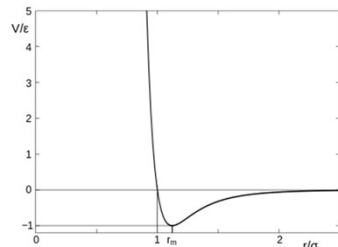
Функция (9) является гамильтонианом QUBO, т.к.

$$q^2 \equiv q$$

Масштабируемость алгоритма логарифмическая

Зависимость кол-ва бинарных переменных N_b от линейных размеров сетки дискретизации $(N_d^{(l)}, N_d^{(s)})$ – количества узлов сетки для переменных l и s :

$$N_b(N_d^{(l)}, N_d^{(s)}) = \log_2 N_d^{(l)} + \log_2 N_d^{(s)}. \quad (10)$$



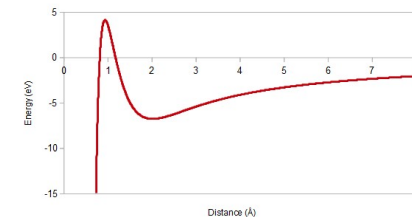
Потенциал Букиннгема

$$U(r) = A \exp(-Br) - \frac{C}{r^6}, \quad (11)$$

$$Err^2(A, B, C) = \sum_{m=1}^M (E_m^{mod} - E_m^{obs})^2 = \sum_{m=1}^M (A f_m(B) - C d_m - E_m^{obs})^2 \rightarrow \min, \quad (12)$$

где $f_m(B) = \sum_{i=1}^K \sum_{j>i}^K \exp(-Br_{ij}^{(m)})$,

$$d_m = \sum_{i=1}^K \sum_{j>i}^K \frac{1}{(r_{ij}^{(m)})^6}, \quad (13)$$



Шаг 1. Бинарная дискретизацию по формуле (2) преобразует функцию $Err^2(A, B, C)$ в **трансцендентную функцию от бинарных переменных**.

Шаг 2. Теорема 1. Любая функцию от n бинарных переменных $f(q_1, q_2, \dots, q_n)$ может быть представлена в виде полинома от q_1, q_2, \dots, q_n степени не большей, чем n .

Шаг 3. Теорема 2. Задача оптимизации произвольного полинома от 3-х бинарных переменных может быть сведена к задаче QUBO от 5-ти бинарных переменных.

Масштабируемость алгоритма

линейно-логарифмическая

$$N_b(N_d^{(A)}, N_d^{(B)}, N_d^{(C)}) \approx 4N_d^{(B)} \log_2 N_d^{(A)} + \log_2 N_d^{(C)}. \quad (14)$$

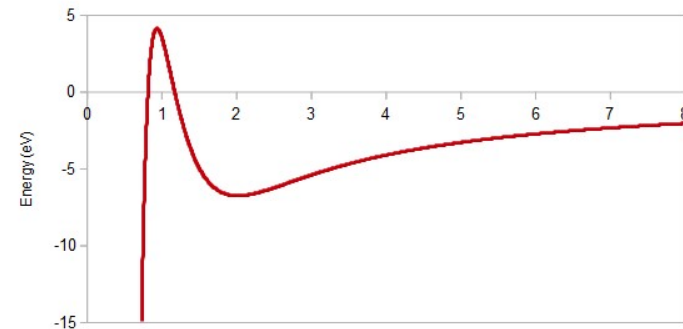
Математика квантового отжига. Аппроксимационно-декомпозиционный метод сведение к QUBO с использованием рядов Тейлора

$U(r) = A \exp(-Br) - \frac{C}{r^6}$, (11) – потенциал Букиннгема

$$Err^2(A, B, C) = \sum_{m=1}^M (E_m^{mod} - E_m^{obs})^2 = \sum_{m=1}^M (A f_m(B) - C d_m - E_m^{obs})^2 \rightarrow \min, (12)$$

где $f_m(B) = \sum_{i=1}^K \sum_{j>i}^K \exp(-B r_{ij}^{(m)})$,

$$d_m = \sum_{i=1}^K \sum_{j>i}^K \frac{1}{(r_{ij}^{(m)})^6}, (13)$$



Основная идея метода: «вытащить» параметр B из экспоненты $\exp(-Br)$, используя разложение в ряд Тейлора.

Первый этап: численно рассчитываем значения первых V коэффициентов a_v ряда Тейлора по r функции $A \exp(-Br)$.

Масштабируемость алгоритма логарифмическая.

Второй этап: Приравняв полученные значения a_v к их формульному выражению, получаем систему уравнений для A и B , которая сводится к задаче минимизации функции

$$Err^2(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_V) = \left(\sum_{v=0}^V (-1)^v a_v \varepsilon_v r_{max}^v \right)^2 (15.1)$$

при наличии ограничений

$$A(B)^v = a_v v! (1 - \varepsilon_v) \text{ для всех целых } v \text{ от } 0 \text{ до } V. (15.2)$$

Задача (15) сводится к задаче QUBO. **Масштабируемость алгоритма линейная.**

Третий этап: Если на втором этапе полученные значения ε_v оказываются достаточно небольшими, т.е. наши данные хорошо описываются потенциалом Букиннгема, то можно использовать еще одну аппроксимацию и получить **алгоритм с логарифмической масштабируемостью.**

Применение QA к задаче инверсии сейсмических данных

Задача инверсии сейсмических данных – определения структуры и состава геологических пород посредством анализа данных об отраженных акустических волнах.

Модельная задача: акустическое 1D приближение для горизонтально-слоистой среды с априорной информацией о параметрах слоев, выраженной произвольными функциями.

Математическая модель. Минимизируемый функционал:

$$Err_b^2(\mathbf{AI}, \mathbf{t}) = \sum_{k=1}^N \left[a_k^{out} - \sum_{i=1}^M \frac{(AI_{i+1} - AI_i)}{(AI_{i+1} + AI_i)} (a_k^{in} \cos(k\omega_0 t_i) - b_k^{in} \sin(k\omega_0 t_i)) \right]^2 + \sum_{k=1}^N \left[b_k^{out} - \sum_{i=1}^M \frac{(AI_{i+1} - AI_i)}{(AI_{i+1} + AI_i)} (a_k^{in} \cos(k\omega_0 t_i) + b_k^{in} \sin(k\omega_0 t_i)) \right]^2 - \sum_{i=1}^{M+1} \ln[\phi_i^{AI}(\mathbf{AI}_i)] - \sum_{i=1}^M \ln[\phi_i^t(t_i)].$$

Шаг 1. Декомпозиция исходной задачи на более простые подзадачи

на уравнение от одной неизвестной + функционал более простого вида

$$Err_b^2(\mathbf{q}, \mathbf{t}) = \sum_{k=1}^N \left[a_k^{out} - \sum_{i=1}^M q_i (a_k^{in} \cos(k\omega_0 t_i) - b_k^{in} \sin(k\omega_0 t_i)) \right]^2 + \sum_{k=1}^N \left[b_k^{out} - \sum_{i=1}^M q_i (a_k^{in} \cos(k\omega_0 t_i) + b_k^{in} \sin(k\omega_0 t_i)) \right]^2 - \sum_{i=1}^M \ln[\phi_i^r(q_i)] - \sum_{i=1}^M \ln[\phi_i^t(t_i)].$$

Шаг 2. Сведение к задачи QUBO с минимальным количеством бинарных переменных

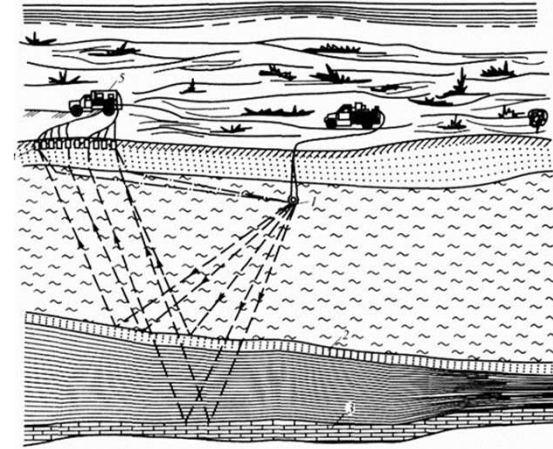
$$H(\mathbf{q}) = \sum_{k=1}^N \left\{ \left[a_k - \sum_{i=1}^M \sum_{j=0}^{n-1} \sum_{l=0}^{n-1} A_{ki}^{jl} q_i^{jl} \right]^2 + \left[b_k - \sum_{i=1}^M \sum_{j=0}^{n-1} \sum_{l=0}^{n-1} B_{ki}^{jl} q_i^{jl} \right]^2 \right\} + \sum_{i=1}^M \sum_{j=0}^{n-1} \sum_{l=0}^{n-1} C_i^{jl} q_i^{jl} + \lambda \sum_{i=1}^M (\sum_{j=0}^{n-1} \sum_{l=0}^{n-1} q_i^{jl} - 1)^2.$$

Шаг 3. Разработка декомпозиционно-итерационного алгоритма для решения задачи

QUBO, учитывающего внутренние математические симметрии исходной задачи

Модернизация алгоритма Energy-Impact Decomposing (EIP).

Стандартный алгоритм EIP – один из базовых универсальных алгоритмов приближенного решения задач QUBO.



Размер задачи QUBO $K = M/\epsilon^2$ бинарных переменных, где M – количество слоев, ϵ – точность дискретизации. При $M = 100$, $\epsilon = 0,1$ получаем $K = 10\,000$.

Оценка времени расчета на D-Wave среднего месторождения (10 x 10 км, сетка с шагом 25 м) ~ 1,5 мес.

Малетин Н.В., «О возможности решения масштабных одномерных задач инверсии сейсмических данных на современных квантовых отжигателях», *Геофизика*, 2023, 2, 102, DOI: [10.34926/geo.2023.59.58.012](https://doi.org/10.34926/geo.2023.59.58.012)

Прогнозы и перспективы

1 Оборудование

QA D-Wave:

Экстраполяция на 2033 год ≈ 26 тыс. физических кубитов

Связанность (кол-во фактических связей к кол-ву возможных) ≈ 0,25%-0,30%

Количество алгоритмических кубитов ≈ 2%-3% от количества физических

Экстраполяция на 2033 год

≈ 500-700 алгоритмических кубитов (≈ 26 тыс. * (2 – 3 %))

Экстраполяция на 2028 год

≈ 300-450 алгоритмических кубитов

GTQC IBM:

Экстраполяция на 2033 год ≈ 50 тыс. физических кубитов

Амбиция IBM на 2033 год ≈ 100 тыс. физических кубитов

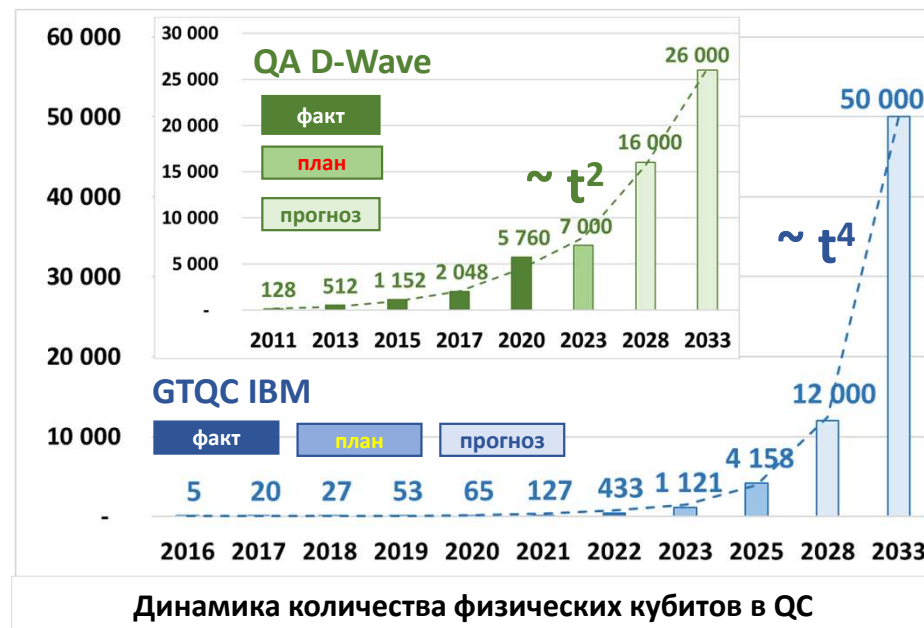
(источник: IBM <https://research.ibm.com/blog/100k-qubit-supercomputer>)

2 Математика и алгоритмы

Накопление базы эвристик и разработка на их основе эффективных **распараллеливаемых** алгоритмов приближенного решения больших задач QUBO для разных типов задач.

3 Задачи

Поиск **практических задач/ классов задач, где методология QUBO + QA** оказывается **эффективней** применяемых сейчас методов решения, например, задач оптимизации функций с очень сложным ландшафтом.



4 На пороге квантового преимущества?

Прогноз для QAOA от 420 до ~ 1000 физических кубитов

(<https://doi.org/10.22331/q-2020-05-11-264>, <https://doi.org/10.1038/s41598-019-43176-9>)

Прогноз для QA ~ 4400 - 6600 физических кубитов

5 Кто победит – QA или оптимизация на GTQC?

Показатель	QA	GTQC
Количество компаний всего/ в т.ч. преодолевших 50 физических кубитов	2 / 1	~20 / 10
Количество физических платформ всего / в т.ч. основных	1 / 1	10 / 5

QUANTUM BITS
IAM
QUANTUM SATIS

СПАСИБО ЗА ВНИМАНИЕ!