ICON
A Symbol of Excellence

Implementation of optimal experimental design algorithms on a quantum computer

Sergei Leonov
(ICON Innovation Center)
Joint work with Valerii Fedorov

April 30, 2018

Outline

- Motivation 1: collaboration with Lockheed Martin on quantum computing
- Motivation 2: current landscape in oncology studies as a driver for better modeling
- Examples of designs: combinatorial and model-based
- Quantum computing: some examples

Motivation 1: Quantum computing

- About 2 years ago: start of a joint project between ICON Innovation Center and Lockheed Martin
- Exploring statistical problems to be solved on a quantum computer D-Wave: *https://www.dwavesys.com/quantum-computing*
- D-Wave: a quantum annealer, designed to solve quadratic unconstrained binary optimization problems (QUBO)

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left[\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{h}^T \mathbf{x}\right], \qquad (6)$$

elements of $n \times 1$ vector $\mathbf{x}$ are either 0 or 1,

$\mathbf{Q}$ is an $n \times n$ symmetric matrix, $\mathbf{h}$ is an $n \times 1$ vector.

Ising problem: $x_i' = \pm 1 \Rightarrow$ use $x_i = (1 + x_i')/2$

Quantum computing

- Superposition
  - Conventional computing: a *bit* exists in one state at a time, either 0 or 1
  - A *qubit* exists in two states at one time, these states are probabilistic → quantum computer can manipulate vast data sets simultaneously
  - Number of quantum states: $2^{2000} \sim 10^{600}$ for a computer with 2000 qubits (Number of atoms in the universe - ?)
- Entanglement:
  - Conventional bits interact only in a linear sequence, changing each other's state one at a time in a chain of binary operations
  - Qubits can interact directly with each other, even at great distances, altering each other's states without intermediate causal connections → potential for using computational "shortcuts" (*quantum tunneling*)
  - Conceptually counterintuitive

    Albert Einstein's quote: "God does not play dice with the universe"

Superposition: Schrödinger's cat

Life-size cat figure in the garden of Erwin Schrödinger's house in Zurich

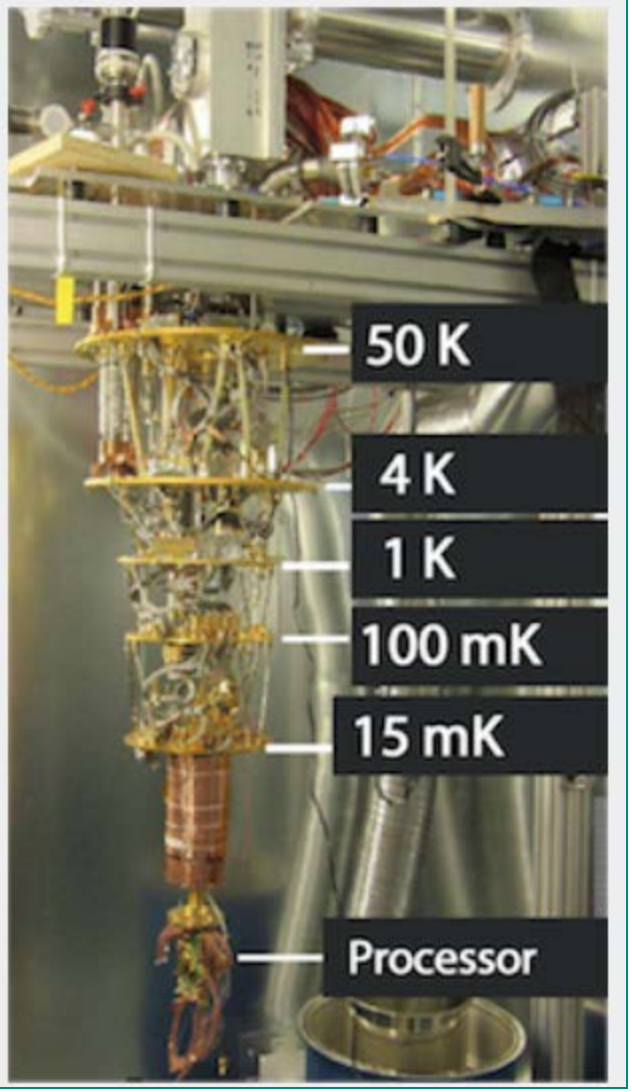
(depending on the light conditions, the cat appears either alive or dead)

Huttenstrasse 9
8006 Zürich

Quantum computing (cont.)

- Hardware challenges: providing a *super-cool* environment
  - Technological challenge: "complete" isolation from the environment (vibrations, heat, light, electromagnetism,….)
  - Any energy input changes quantum states: an ideal quantum mechanical system can only exist at a temperature of absolute zero (0°K = -273.15°C = 459.67°F).
  - A processor must operate in the "near-absolute-zero" space
  - The latest generation D-Wave system: 15 millikelvin

https://www.dwavesys.com/tutorials/background-reading-series/introduction-d-wave-quantum-hardware

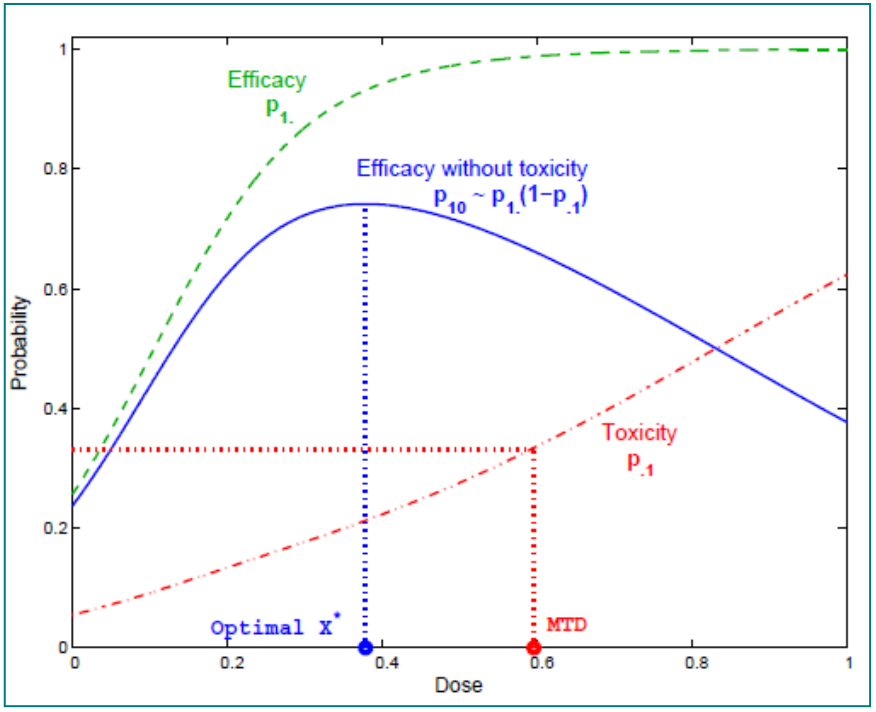


Motivation 2: oncology studies

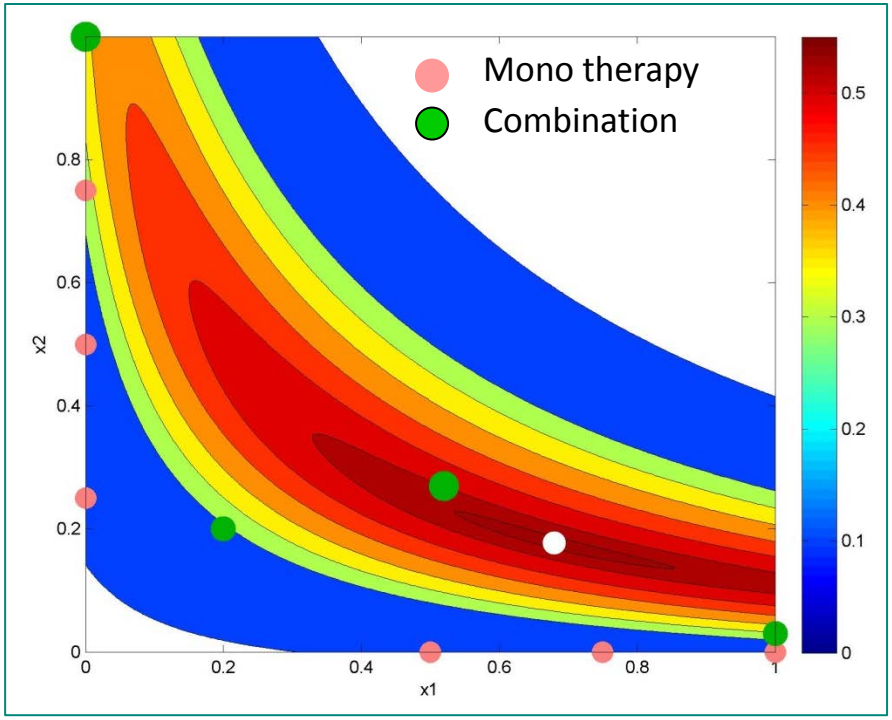
- Dozens of compounds (mono/combination therapies) and cancer types
- Need to screen multiple drug combinations for each cancer type
- Sponsors competing for resources (patients, research sites)
- Hundreds of clinical trials needed within the traditional setting (one treatment, one disease, one study at a time)
- Modeling is needed
  - Statistical, mechanistic
  - Operational processes (enrollment, drug supply)
  - Borrowing information between studies
- Example: dose finding (simultaneous modeling of efficacy and toxicity, including drug combinations)

Dose finding: efficacy – toxicity balance

Single drug: maximize probability of efficacy w/out toxicity



Drug combination: borrowing information between studies (composite designs)



Fedorov, Leonov (2013), Chapters 6, 8

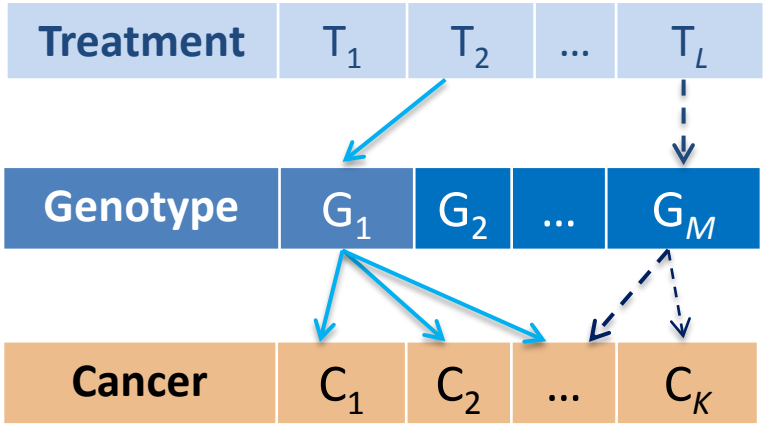
Trial classifiers in oncology

Genotype	G_1	G_2	…	G_m	…	G_M
Cancer	C_1	C_2	…	C_k	…	C_K
Treatment	T_1	T_2	…	T_l	…	T_L
???						

- Number of all possible combinations = M x K x L x (?)
- Interested in drug combinations: L ~N(N-1)/2, N - size of the drug portfolio
- Different designs: different order of layers

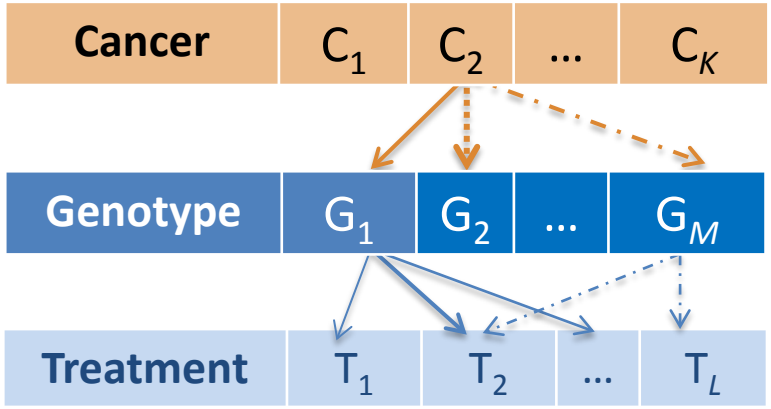
Oncology: types of trials

Basket



Testing treatment effect on a specific biomarker across cancer types

Umbrella



Testing treatment effect on different biomarkers in a single cancer type

Modeling

Fusion of combinatorial and model-based design techniques

- Different type of designs within a single hierarchical framework
- Discussions with Rosemary Bailey: workshop "*Design and Analysis of Experiments in Healthcare*" at the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, July 2015

Starting point: ANOVA-type model, factorial designs, how to tackle the problem of a large number of drug combinations

- No dose finding today: *doses have been selected*

$$y_{ijk\ell\ell'} = \mu_{jk\ell\ell'} + \varepsilon_{ijk\ell\ell'} \ , \quad i = 1, ..., n_{jk\ell\ell'} \ , \tag{1}$$

Indices: i - subject, j - subtrial, k - cancer type, ℓ, ℓ' - drugs

Modeling: combinatorial designs

J - no. of sub-trials, K – no. of cancer types, L – no. of drugs

Constraints to make a design balanced (Rosemary):

1. All sub-trials involve the same number of cancer types
2. All sub-trials use the same number of drugs
3. Each pair of distinct cancer types are involved together at the same non-zero number of sub-trials
4. Each pair of distinct drugs are used together at the same non-zero number of sub-trials
5. Each drug is used on each type of cancer at the same number of sub-trials

BIBD

Example: J = 10, K = 6, L = 5 (Bailey, Cameron, 2018)

(a) Block	Cancer (subpopulation)	Drugs
1	C123	D15
2	C156	D12
3	C134	D23
4	C126	D34
5	C145	D45
6	C245	D13
7	C235	D24
8	C356	D35
9	C346	D14
10	C246	D25

(b)	Cancers					
Block	C1	C2	C3	C4	C5	C6
1	D15	D15	D15			
2	D12				D12	D12
3	D23		D23	D23		
4	D34	D34				D34
5	D45			D45	D45	
6		D13		D13	D13	
7		D24	D24		D24	
8			D35		D35	D35
9			D14	D14		D14
10		D25		D25		D25

From combinatorial to model-based designs

$$\mu_{jk\ell\ell'} = a + b_j + c_k, \tag{2}$$

$$\mu_{jk\ell\ell'} = a + b_j + c_k + \{d_\ell \bigvee d_{\ell'}\},$$

$$\mu_{jk\ell\ell'} = a + b_j + c_k + d_\ell + d_{\ell'} + d_{\ell\ell'},\ \ell < \ell'.$$

Goal: regression-type model,

$$y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + \varepsilon,\ \mathbf{x} \in \mathcal{X}, \tag{3}$$

$\mathbf{x}$ - vector of experimental conditions; $f(\mathbf{x})$ - basis functions;

$\boldsymbol{\beta}$ - unknown parameters; $\mathcal{X}$ - set of candidate (support) points

Candidate points, model (2), BIBD design

No	Block effects b_j				Cancer effects c_k						Drug effects d_ℓ			
	b_1	b_2	...	b_{10}	c_1	c_2	c_3	c_4	c_5	c_6	d_1	d_2	...	d_5
1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2	1	0	0	0	1	0	0	0	0	0	0	0	0	1
3	1	0	0	0	1	0	0	0	0	0	1	0	0	0
4	1	0	0	0	1	0	0	0	0	0	1	0	0	1
5	1	0	0	0	0	1	0	0	0	0	0	0	0	0
6	1	0	0	0	0	1	0	0	0	0	0	0	0	1
7	1	0	0	0	0	1	0	0	0	0	0	1	0	0
8	1	0	0	0	0	1	0	0	0	0	0	1	0	1
9	1	0	0	0	0	0	1	0	0	0	0	0	0	0
10	1	0	0	0	0	0	1	0	0	0	0	0	0	1
11	1	0	0	0	0	0	1	0	0	0	0	1	0	0
12	1	0	0	0	0	0	1	0	0	0	0	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
109	0	0	0	1	0	1	0	0	0	0	0	0	0	0
110	0	0	0	1	0	1	0	0	0	0	0	0	0	1
111	0	0	0	1	0	1	0	0	0	0	0	1	0	0
112	0	0	0	1	0	1	0	0	0	0	0	1	0	1
113	0	0	0	1	0	0	0	1	0	0	0	0	0	0
114	0	0	0	1	0	0	0	1	0	0	0	0	0	1
115	0	0	0	1	0	0	0	1	0	0	0	1	0	0
116	0	0	0	1	0	0	0	1	0	0	0	1	0	1
117	0	0	0	1	0	0	0	0	0	1	0	0	0	0
118	0	0	0	1	0	0	0	0	0	1	0	0	0	1
119	0	0	0	1	0	0	0	0	0	1	0	1	0	0
120	0	0	0	1	0	0	0	0	0	1	0	1	0	1
	x_{b1}	x_{b2}	...	x_{b10}	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	...	x_{11}

Set $\mathcal{X}$: 120 distinct support points

(b)	Cancers					
Block	C1	C2	C3	C4	C5	C6
1	D15	D15	D15			
2	D12				D12	D12
3	D23		D23	D23		
4	D34	D34				D34
5	D45			D45	D45	
6		D13		D13	D13	
7		D24	D24		D24	
8			D35		D35	D35
9			D14	D14		D14
10		D25		D25		D25

Regression model, BIBD

The simplest linear regression model:

$$\boldsymbol{\beta} = (\beta_0; \beta_{b1}, \ldots, \beta_{b10}; \beta_1 \ldots, \beta_6; \ \beta_7, \ldots, \beta_{11})^T, \tag{4}$$

with $\beta_0 = a$, $\beta_{b1 \to b10} = b_{1 \to 10}$, $\beta_{1 \to 6} = c_{1 \to 6}$, $\beta_{7 \to 11} = d_{1 \to 5}$,

$$\mathbf{f}(\mathbf{x}) = (1; x_{b1}, \ldots, x_{b10}; x_1, \ldots x_6; x_7, \ldots, x_{11})^T.$$

All parameters $\boldsymbol{\beta}$ cannot be estimated: constraints needed, e.g.

$$\sum_{j=1}^{10} \beta_{bj} = 0$$ (Wu, Hamada (2000, Chapter 1.6)).

Simplification: exclude block terms β_{bj}, instead add interaction terms:

$$\boldsymbol{\beta} = (\beta_1 \ldots, \beta_6; \ \beta_7, \ldots, \beta_{11}, \beta_{12}, \ldots, \beta_{21})^T, \tag{5}$$

$$\mathbf{f}(\mathbf{x}) = (x_1, \ldots x_6; x_7, \ldots, x_{11}, x_7x_8, \ x_7x_9, \ \ldots \ x_{10}x_{11})^T.$$

Candidate points, model (5)

Points	Cancers						Drugs					Comments
No	c1	c2	c3	c4	c5	c6	d1	d2	d3	d4	d5	
1	1	0	0	0	0	0	0	0	0	0	0	Placebo
2	1	0	0	0	0	0	1	0	0	0	0	Single drugs,
3	1	0	0	0	0	0	0	1	0	0	0	**Cancer 1**
4	1	0	0	0	0	0	0	0	1	0	0	
5	1	0	0	0	0	0	0	0	0	1	0	
6	1	0	0	0	0	0	0	0	0	0	1	
7	1	0	0	0	0	0	1	1	0	0	0	Pairs with D1
8	1	0	0	0	0	0	1	0	1	0	0	
9	1	0	0	0	0	0	1	0	0	1	0	
10	1	0	0	0	0	0	1	0	0	0	1	
11	1	0	0	0	0	0	0	1	1	0	0	Pairs with D2
12	1	0	0	0	0	0	0	1	0	1	0	
13	1	0	0	0	0	0	0	1	0	0	1	
14	1	0	0	0	0	0	0	0	1	1	0	Pairs with D3
15	1	0	0	0	0	0	0	0	1	0	1	
16	1	0	0	0	0	0	0	0	0	1	1	Pairs with D4
17	0	1	0	0	0	0	0	0	0	0	0	Placebo
18	0	1	0	0	0	0	1	0	0	0	0	Single drugs,
19	0	1	0	0	0	0	0	1	0	0	0	**Cancer 2**
...						...	...					
22	0	1	0	0	0	0	0	0	0	0	1	
23	0	1	0	0	0	0	1	1	0	0	0	Pairs with D1
...						...	...					
26	0	1	0	0	0	0	1	0	0	1	0	
...						...	...					
32	0	1	0	0	0	0	0	0	0	1	1	Pairs with D4
...						...	...					
81	0	0	0	0	0	1	0	0	0	0	0	**Cancer 6**
...						...	...					
96	0	0	0	0	0	1	0	0	0	1	1	Pairs with D4
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	

Set $\mathcal{X}_2$: 96 distinct support points

Comparison of designs

Designs ξ_n: n support points with equal weight $1/n$.

- ξ^*_{96}: all 96 support points from $\mathcal{X}_2$
- ξ^*_{48}: 48 points removed from ξ_{96} via backward elimination, i.e., on step 1 find

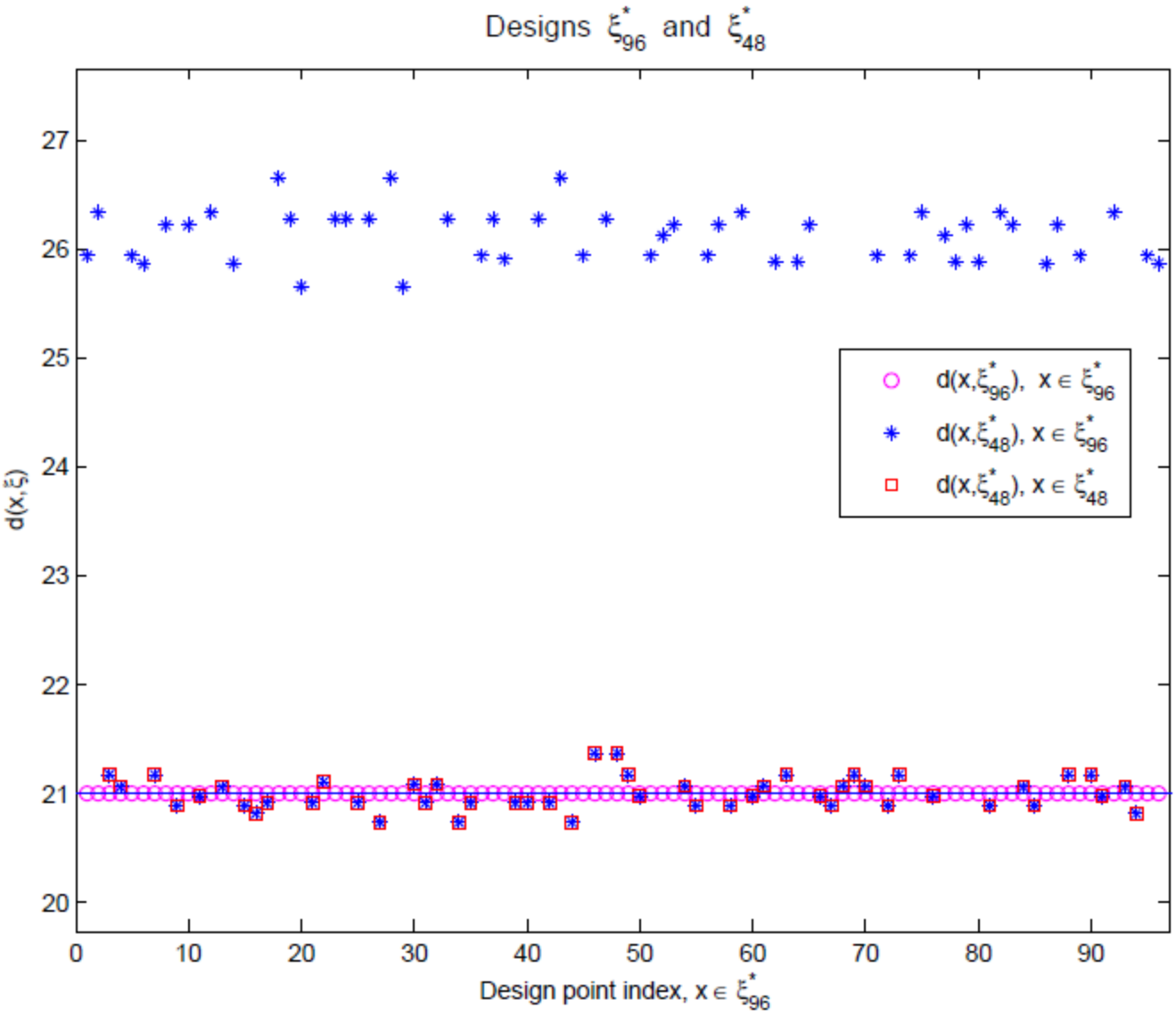
$$\mathbf{x}_1^- = \arg\min_{\mathbf{x}\in\mathcal{X}_2} \mathbf{f}^T(\mathbf{x})\mathbf{M}^{-1}(\xi_{96})\mathbf{f}(\mathbf{x}),$$

  and remove $\mathbf{x}_1^-$ from then "current" design $\xi_{96} = \xi^*_{96}$, etc.

- ξ_{66}: 66 points obtained from BIBD design by removing replications

Design	$\lvert\mathbf{D}\rvert^{1/21}$	Efficiency (%)	$\mathrm{tr}\mathbf{D}/21$	$d_{av}(\mathbf{x}_i, \xi)$
ξ^*_{96}	13.80	100	44.1	21
ξ^*_{48}	14.56	94.8	48.3	23.6
ξ_{66}	14.76	93.5	36.7	24.06

Comparison of designs



ξ_{96}*: D-optimal (*Equivalence Theorem, sensitivity function*)

ξ_{48}: small loss of efficiency (~5%), but possibly more practical

Cost analysis: which option is better

- Minor increase in sample size
- Substantial reduction in number of treatment arms

Back to quantum computing

- Similarities between QUBO and iterative algorithms of construction of model-based design
- Example: construction of saturated designs on a hypercube

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i,\ \mathbf{x}_i \in \mathcal{X} = \left\{ \mathbf{x} \mid \mathbf{x} \in \{0,1\}^{m-1} \right\}.$$

Second-order exchange algorithm (Fedorov, 1972, Ch. 3.2):

Given $\xi_{N,(s)}$ find $\mathbf{x}_i^{(s)} \in \xi_{N,(s)}$ and $\mathbf{x}^{(s+1)} \in \mathcal{X} \setminus \xi_{N,(s)}$ such that their swapping maximizes the increment of

$$\frac{|\mathbf{M}(\xi_{N,(s+1)})|}{|\mathbf{M}(\xi_{N,(s)})|} = 1 + \Delta_s \ .$$

One can verify that

$$\Delta_s = \mathbf{x}^T \mathbf{M}^{-1}(\xi_{N,(s)}) \mathbf{x} - \mathbf{x}_i^T \mathbf{M}^{-1}(\xi_{N,(s)}) \mathbf{x}_i$$

$$+ \mathbf{x}^T \mathbf{M}^{-1}(\xi_{N,(s)}) \mathbf{x}_i \mathbf{x}_i^T \mathbf{M}^{-1}(\xi_{N,(s)}) \mathbf{x} - \mathbf{x}^T \mathbf{M}^{-1}(\xi_{N,(s)}) \mathbf{x} \mathbf{x}_i^T \mathbf{M}^{-1}(\xi_{N,(s)}) \mathbf{x}_i$$

Second-order exchange algorithm

$$\Delta_s = \mathbf{x}^T\mathbf{M}^{-1}(\xi_{N,(s)})\mathbf{x} - \mathbf{x}_i^T\mathbf{M}^{-1}(\xi_{N,(s)})\mathbf{x}_i$$

$$+\mathbf{x}^T\mathbf{M}^{-1}(\xi_{N,(s)})\mathbf{x}_i\mathbf{x}_i^T\mathbf{M}^{-1}(\xi_{N,(s)})\mathbf{x} - \mathbf{x}^T\mathbf{M}^{-1}(\xi_{N,(s)})\mathbf{x}\mathbf{x}_i^T\mathbf{M}^{-1}(\xi_{N,(s)})\mathbf{x}_i$$

$\mathbf{x}_i$ - current design point, $\mathbf{x}$ - candidate point:

$\tilde{\Delta}_{s,i}(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x}$, where $\mathbf{Q} = (1 - \mathbf{x}_i^T\mathbf{M}^{-1}\mathbf{x}_i)\mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{x}_i\mathbf{x}_i^T\mathbf{M}^{-1}$.

On each step s:

- Maximize $\tilde{\Delta}_{s,i}(\mathbf{x})$ over $\mathbf{x} \in \mathcal{X}$ for each design point $\mathbf{x}_i$.
- Find maximum $\Delta_{s,i}$ over design points.
- Iterative procedure (multiple steps)

Saturated design, $m = 7$: number of possible combinations $\sim 0.6 \times 10^9$

Iterative quantum annealing on D-Wave

- Mapping the original optimization procedure to the QUBO problem wired at D-Wave machine
- Embedding algorithms to match “logical” and “physical” qubits: linking hardware and software
- Probabilistic nature: multiple solutions are typically returned
- Solutions: values that correspond to the optimal configuration of qubits found (*lowest points in the energy landscape*)
- Web API: libraries available for C, Python, and MATLAB

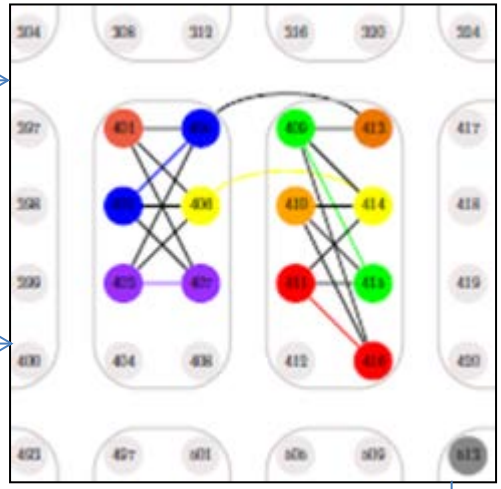
QMI - Quantum Machine Instruction

Interaction between classical components and quantum annealing core

- Embedding: depends on matrix Q

```
%    ------  Use D-Wave heuristic embedder    ------
     load('dw2x_adjacency.mat');
     embedding = sapiFindEmbedding(Q, A);
%    ---------------------------------------------
```

```
%    -------------------
%    Select quantum chip
%    -------------------
 solver = sapiSolver(conn,'DW2X');
```



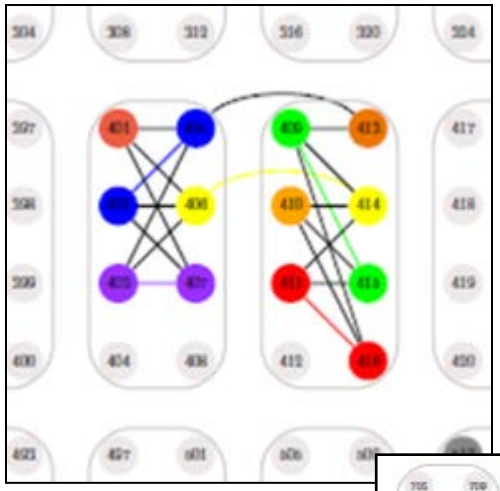
- Select a solution

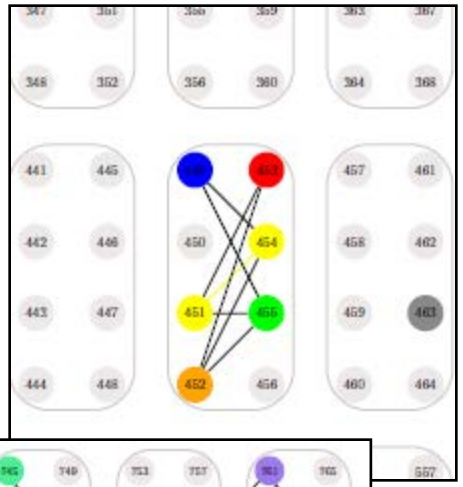
Annealing time

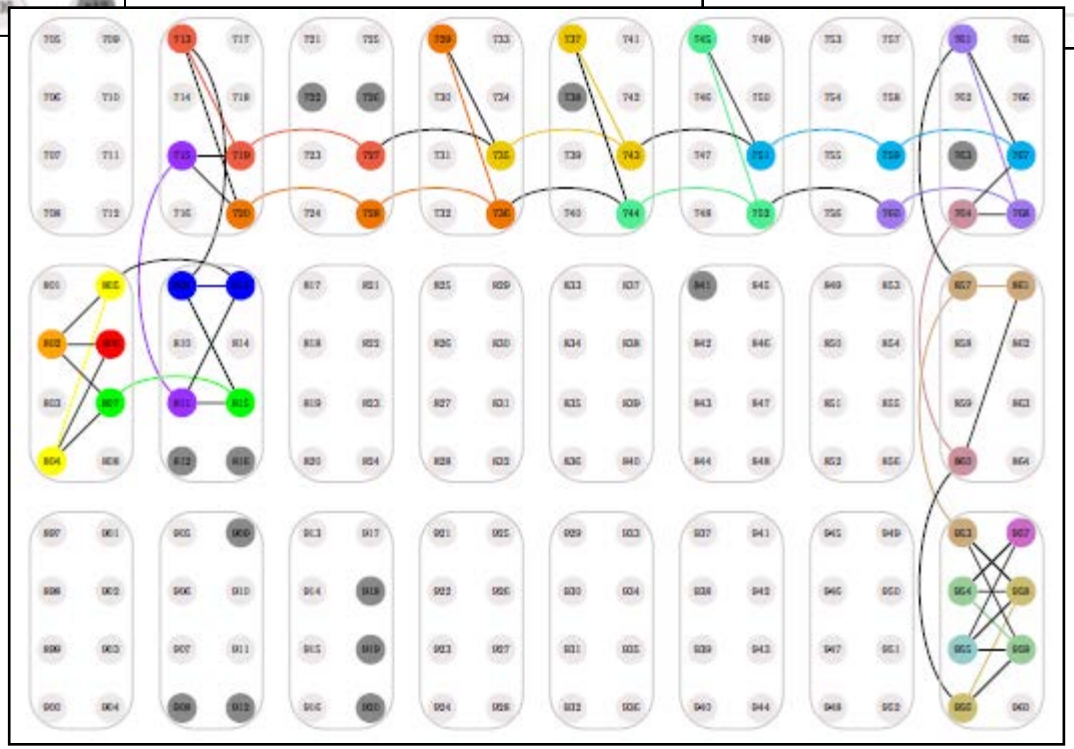
Number of runs

- Return to the classical component and update design

Embedding: empirical solutions







Problem (5): how to deal with constraints

$$\mathbf{l}_1^T\mathbf{x} = 1, \;\; l_{1i} = 1, \; i = 1, \ldots, 6; \; l_{1i} = 0, \; i = 7, \ldots, 11; \tag{8}$$

$$0 \leq \mathbf{l}_2^T\mathbf{x} \leq 2, \; \mathbf{l}_2 = \mathbf{1} - \mathbf{l}_1, \tag{9}$$

To reduce the optimization problem to QUBO:

(a) Introduce auxiliary variables $u_1, u_2 \in \{0, 1\}$ to reduce (9) to the equality

$$\mathbf{l}_2^T\mathbf{x} + u_1 + u_2 = 2. \tag{10}$$

(b) Introduce penalty functions

$$\Psi_1(\mathbf{x}) = w_1 \left[\mathbf{l}_1^T\mathbf{x} - 1\right]^2, \Psi_2(\mathbf{x}, u_1, u_2) = w_2 \left[\mathbf{l}_2^T\mathbf{x} + u_1 + u_2 - 2\right]^2.$$

(c) Identify "reasonable" values of weights w_1, w_2 to solve the problem

$$\{\mathbf{x}_s^+, \mathbf{x}_s^-\} = \arg \min_{\mathbf{x}\in\mathcal{X}, \mathbf{x}'\in\mathcal{X}_{N,s}; u_1, u_2} \left[\Psi_1(\mathbf{x}) + \Psi_2(\mathbf{x}, u_1, u_2) - \Delta_s(\mathbf{x}, \mathbf{x}')\right] \; ;$$

Recent activities

- Workshop “Quantum computing and its application in drug development”, George Washington University, March 2017, co-organized by GWU Department of Statistics, Lockheed Martin and ICON Innovation Center
- A session at JSM 2017, July 2017 (Baltimore, MD)
- A session at CEN-ISBS 2017 conference in Vienna, August 2017
- Scientific Interest Group formed in March 2018 within Statistical Section of ASA
- A session planned at JSM 2018, August 2018 (Vancouver, BC)

References

- Bohachevsky, I., Johnson, M.E., Stein, M.L. (1986). Generalized simulated annealing for function optimization. *Technometrics,* **28**(3), 209–217.
- Bailey, R.A., Cameron, P.J. (2018). Designs which allow each medical centre to treat only a limited number of cancer types with only a limited number of drugs. Available at: https://arxiv.org/pdf/1803.00006.pdf
- Fedorov, V.V, Leonov, S.L. (2013). *Optimal Design for Nonlinear Response Models*. Chapman & Hall/CRC Biostatistics Series, Boca Raton, FL.
- Haines, L.M. (1987). The application of the annealing algorithm to the construction of exact optimal designs for linear regression models. *Technometrics,* **29**(4), 439–447.
- Wu, C.F.J., and Hamada, M. (2000). *Experiments. Planning, Analysis, and Parameter Design Optimization*. Wiley, New York.

Questions:

Sergei.Leonov@iconplc.com