Quantum Neural Architecture Search with Quantum Circuits Metric and Bayesian Optimization

Trong Duong
Department of Mathematical Sciences
KAIST
trongduong@kaist.ac.kr

Sang T. Truong
Department of Computer Science
Stanford University
sttruong@cs.stanford.edu

Minh Tam
Department of Applied Physics
Aalto University
minh.phamnguyen@aalto.fi

Bao Bach
Department of Computer Science and Engineering
Ho Chi Minh City University of Technology
bao.bachbbace12@hcmut.edu.vn

June-Koo Kevin Rhee
School of Electrical Engineering
KAIST
rhee.jk@kaist.edu

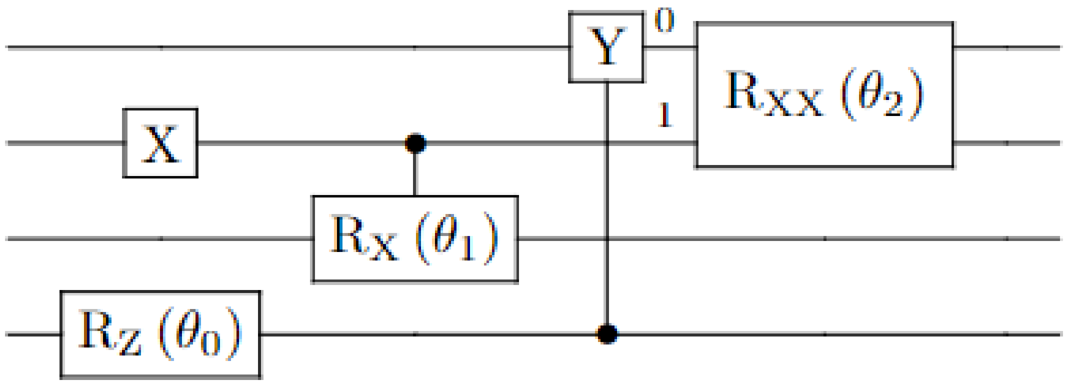
I. Preliminaries QNN

A QNN can be realized by a quantum circuit that contains a sequence of fixed and parametrized unitary operators, so-called quantum gates acting on its qubits $q_0, \ldots, q_n$, where n is the number of processing qubit. Usually, it is restricted to sequence of fixed gates W_i and one-parameter gates $U_i(\theta_i)$

$$U(\theta) = U_L(\theta_L)W_L \ldots U_2(\theta_i)W_2U_1(\theta_1)W_1.$$

where

$$W_i \in \{H, X, Y, Z; CX, CY, CZ\}$$
$$U_i \in \{R_X(\cdot), R_Y(\cdot), R_Z(\cdot), CR_X(\cdot), CR_Y(\cdot), CR_Z(\cdot), R_{XX}(\cdot), R_{YY}(\cdot), R_{ZZ}(\cdot)\}$$



II. Core idea

Denote the performance as $y \equiv f(x)$ with x is a QNN. We model $f(x)$ as drawing from a Gaussian process

$$f(x) \sim GP(m(x), k(x, X))$$

where $m: \chi \to R$ is the mean of the GP, the positive definite kernel $k: \chi \times \chi \to R$ (covariance matrix). $X = [x_1, \ldots, x_D]$ and $Y = [y_1 \ldots y_D]$ are our data set of QNN and their respective performance. The main goal is that we find hyperparameters of the mean function and the kernel k such that the likelihood that Y happens given X is the largest

$$\max_{\alpha} P(Y|X) \sim N(m_X, K_X)$$

For simplicity, we assume that $m_X = 0$. We generate new QNN x during optimization using Expected Improvement (EI) acquisition function, and Evolutionary Algorithm (EA)

$$m(x, X) = K(x, X) K_X^{-1} (Y - m(X))$$
$$k(x, X) = K(x, x) - K(x, X) K_X^{-1} K(X, x)$$
$$y \sim \mathcal{N}(m(x, X), k(x, X))$$

$$\mathrm{EI}_t(x) = \mathbb{E}[(f(x) - f_{t-1}^*)^+ | (x_i, y_i)_{i=1}^{t-1}],$$

$$k(x,x') = \alpha e^{-\sum_i \beta_i d_i(x,x')} + \bar{\alpha} e^{-\sum_i \bar{\beta}_i \bar{d}_i(x,x')},$$

If x is a number, then the Gaussian process is easy to understand. However, since x are QNN here, we need to assign a distance meaning between different QNNs. For instance, how to quantify two similar QNNs.

The distance $d(G_1, G_2)$ of two QNNs consists of several components:

1. Gate mass is a number representing the amount of computation a gate can perform.
2. Gate-type Mismatch: is the cost incurred when one attempts to match a gate $u \in QNN_1$ another gate $v \in QNN_2$ of different types.
3. Structural Dissimilarity Cost This term determines how different the relative position of two matched gates in their respective QNNs are

1. Gate mass

Gate mass is a number representing the amount of computation a gate can perform. Motivated from the fact that a unitary matrix of size $p \times p$ has $p^2 - 1$ real degrees of freedom, we define the mass parametrized gates as

$$lm(u) = para\,\dim(u) \times \ (uni\dim(u)^2 \ - \ 1)$$

For the fixed gate, we assign to them partial fraction of the parameterized gates

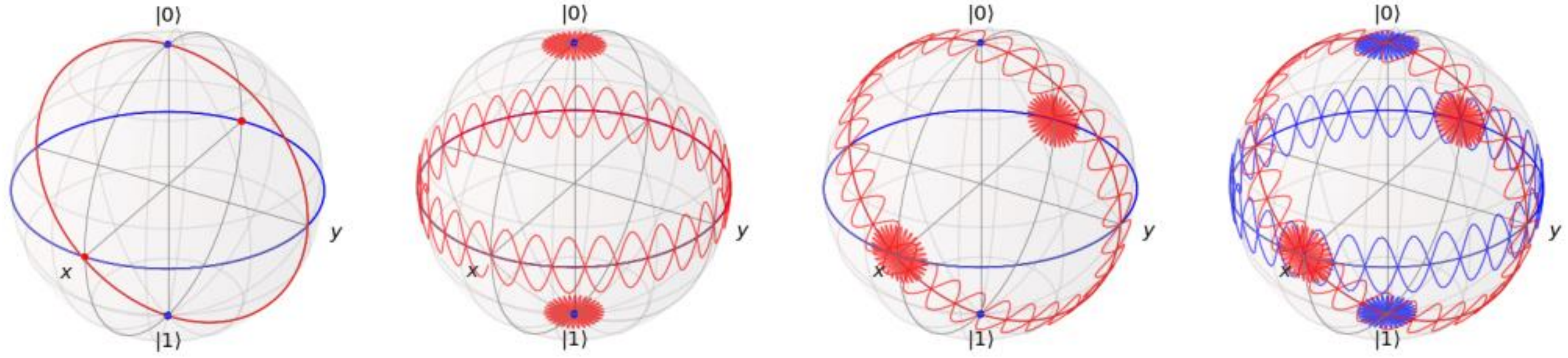
$$lm(u) = \frac{\eta}{|\mathcal{DL}|} \sum_{s \in \mathcal{VL}} lm(s)$$

where DL and VL are the sets of fixed layers and parametrized layers, respectively, and η is a fixed ratio. The total mass of the QNN is sum of all mass.

2. Quantum gates metric

We propose a similarity measure between two (possibly parametrized) quantum gates. Although the two unitary operations induced by the gates can be different, their effect might follow similar patterns throughout the Hilbert space. The difference between the exact effects will be captured by core distance d_{core}, while the difference between their patterns by shape distance d_{shape}. We define the distance between two quantum gates to be the average between those two types of distance:

$$d_{\text{gate}}(U^1, U^2) = \frac{d_{\text{core}}(U^1, U^2) + d_{\text{shape}}(U^1, U^2)}{2}$$



(a) R_Z and R_X (b) R_Z and R_Z^Y (c) R_Z and R_X^Y (d) R_Z^X and R_X^Y

For a quantum gate U, fixed or parametrized, there exists a unique scalar $t > 0$ and a unique Hermitian operator H such that $|H|_* = 1$ and $U = e^{iHt}$. The core distance between the two unitary operators is given by

$$d_{\text{core}}(U^1, U^2) = \frac{\|H^1 - H^2\|_*}{2}$$

The two gates are said to have the same shape when their codomains obtained as θ varies have the same shape up to a unitary transformation.

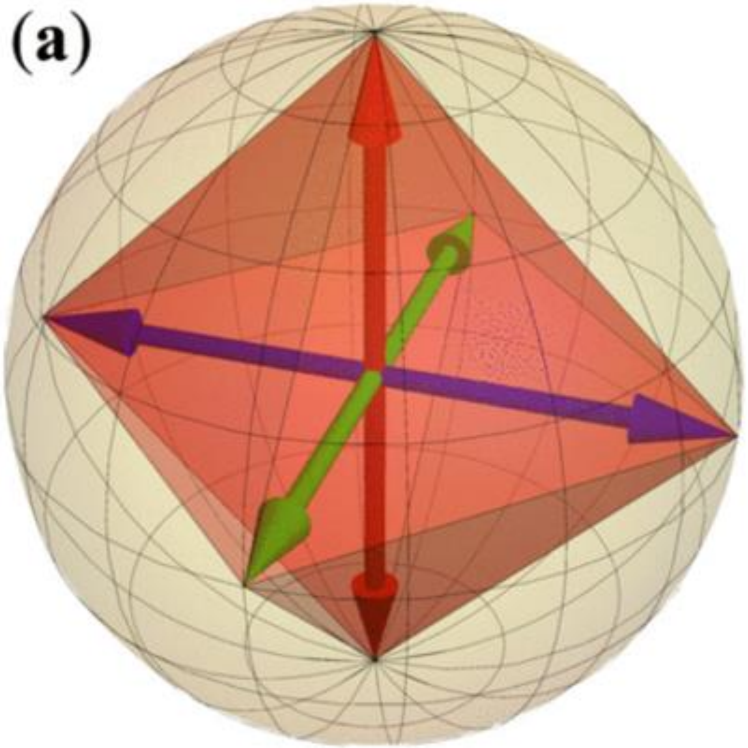
$$\left|\langle\phi| U^{2\dagger}(\theta) V U^1(\theta) |\psi\rangle\right|^2 = 1 \quad \forall\theta$$

This allows us to define the shape distance to be the minimal disparity induced by any unitary mapping The shape distance should consider

$$\int_{\mathcal{H}} \int_{\Theta} \left|\langle\phi| U^{2\dagger}(\theta) V U^1(\theta) |\psi\rangle\right|^2 d\theta d\mu(\psi)$$

We use a collection of quantum states $\{|\psi_k\rangle$ with $k = 1, \ldots, K$ to approximate the integral. Members of the collection must represent the Hilbert space in some sense. We decide to use a collection called Mutually Unbiased Bases (MUB) that contains $K = d(d + 1)$ members for a qubit system of dimension $d = 2^n$. A key property of MUB is that its members spread out evenly in the Hilbert space.

$$|\langle\psi_k|\psi_{k'}\rangle|^2 = 1/d$$



Therefore we attempt to define the shape distance motivated by the following optimization problem

$$\min_{V,\phi_k} \frac{1}{KT} \sum_{k,t} \left(1 - \left|\langle\phi_k| U^{2\dagger}(\theta_t) V U^1(\theta_t) |\psi_k\rangle\right|^2\right)$$

$$= \min_{V,M} \frac{1}{KT} \sum_{k,t} \left(1 - \left|\langle e_k| M^\dagger U^{2\dagger}(\theta_t) V U^1(\theta_t) |\psi_k\rangle\right|^2\right),$$

It turns out that the problem of optimizing the fidelity is similar to the following problem involving a subsystem and the environment,

$$\max_M \mathrm{Tr}\left(\rho_2'(M^\dagger \otimes I)\rho_1'(M \otimes I)\right)$$

Which is extremely difficult, and we did not find much work on this problem. Hence, we consider the root of fidelity in the definition of d_{shape}

$$d_{\text{shape}}(U^1, U^2) = \min_{V,M} \frac{1}{KT} \sum_{k,t} \left(1 - \left|\langle\phi_k| U^{2\dagger}(\theta_t) V U^1(\theta_t) |\psi_k\rangle\right|\right)$$

$$= \min_{V,M} \frac{1}{KT} \sum_{k,t} \left(1 - \left|\langle e_k| M^\dagger U^{2\dagger}(\theta_t) V U^1(\theta_t) |\psi_k\rangle\right|\right)$$

Theorem 1. *The shape distance given by the integral vanishes if and only if the shape distance given by a finite sum over anchor states vanishes. That is, for $T \geq 2$ distinct parameters $\theta_1, \ldots, \theta_T$, if there exist a unitary operator V and quantum states $|\phi\rangle$ such that*

$$\sum_{t=1}^{T} \int_{\mathcal{H}} \left(1 - \left|\langle\phi| U^{2\dagger}(\theta_t) V U^1(\theta_t) |\psi\rangle\right|\right) d\mu(\psi) = 0 \tag{15}$$

if and only if for the same V,

$$\frac{1}{KT} \sum_{k,t} \left(1 - \left|\langle\phi_k| U^{2\dagger}(\theta_t) V U^1(\theta_t) |\psi_k\rangle\right|\right) = 0, \tag{16}$$

where $|\phi_k\rangle$ is counterpart to $|\psi_k\rangle$ in the same way as $|\phi\rangle$ is counterpart to $|\psi\rangle$.

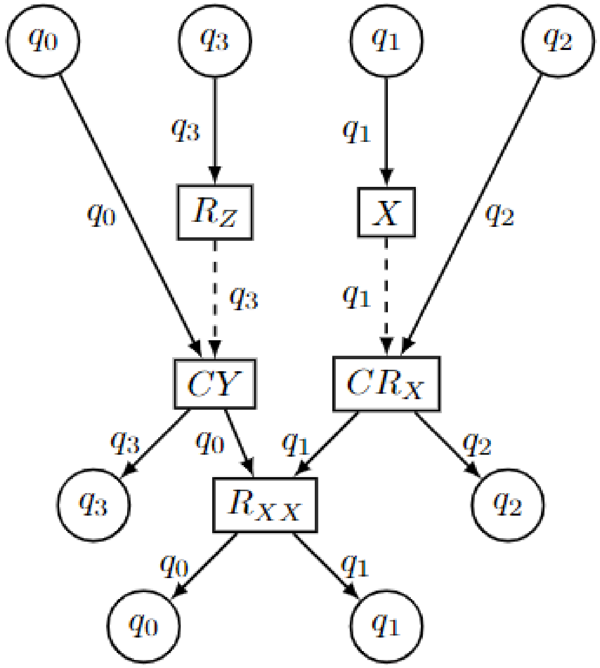
Hence the gate-type mismatch cost matrix $C_{gtm} \in R^{n_1 \times n_2}$ is given by $\left(C_{gtm}\right)_{ij} = d_{gate}\left(u_i, v_j\right)$.

3. Structural Dissimilarity Cost

Structural Dissimilarity Cost This term determines how different the relative position of two matched gates in their respective QNNs are. A small $(C_{\text{str}})_{i,j}$ value means the gate $u_i \in \mathcal{L}_1$ and the gate $v_j \in \mathcal{L}_2$ are at structurally similar position. The cost matrix is computed by

$$(C_{\text{str}})_{i,j} = \frac{1}{6n} \sum_{s \in \{\text{sp},\text{lp},\text{avg}\}} \sum_{t \in \{\text{ip},\text{op}\}} \sum_{q=1}^{n} |\delta_t^{s,q}(i) - \delta_t^{s,q}(j)|, \tag{17}$$

where "sp", "lp", "avg", "ip", "op" abbreviate for shortest path, longest path, random walk, input, and output. The value $\delta_t^{s,q}(\cdot)$ measures the shortest/longest/average path length from/to the input/output node of the q-th qubit in the respective DAG circuit. When there is no path between two nodes, we assign to it the longest path length in the entire DAG.

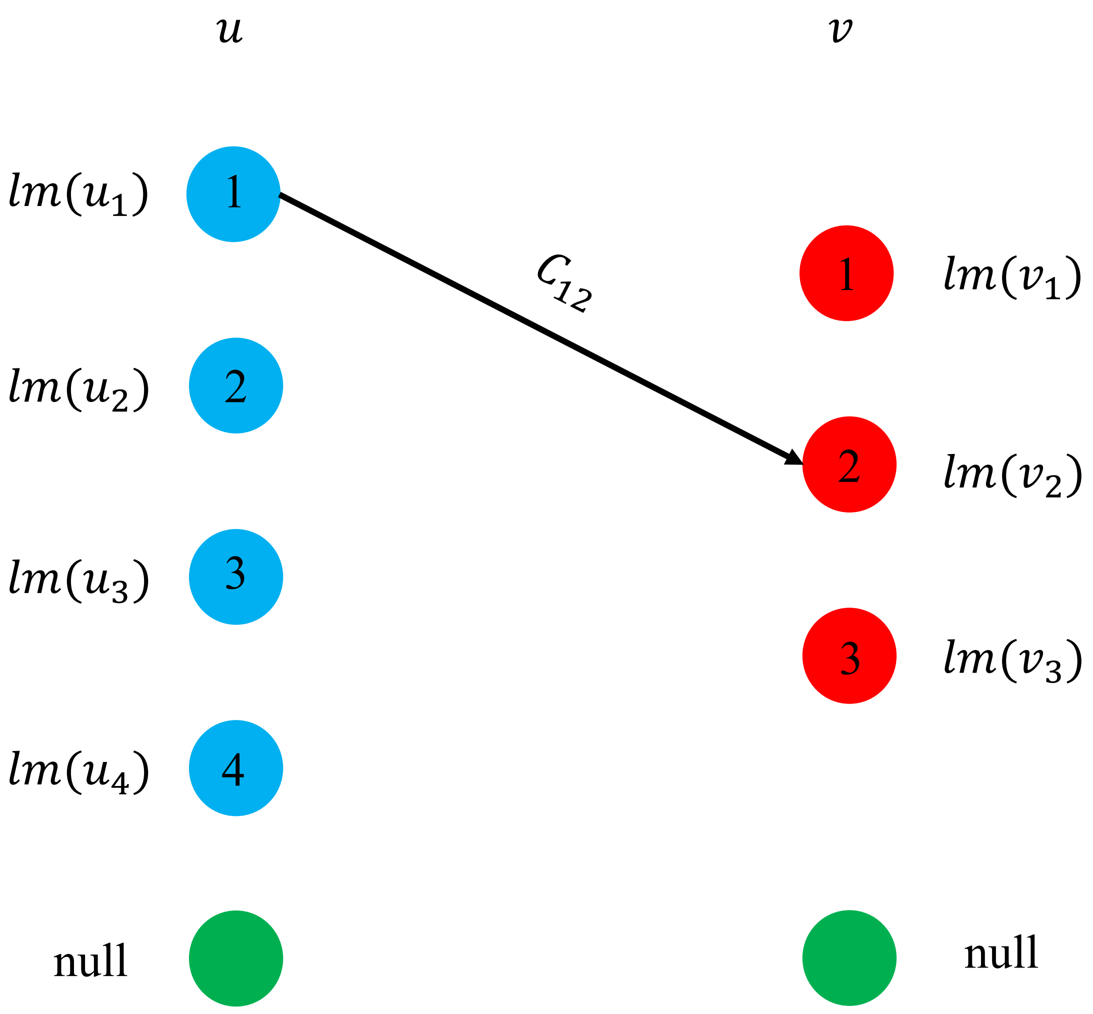


Optimal Transport Distance Let $\langle \cdot | \cdot \rangle$ denote the component-wise dot product and $\mathbf{1}_n$ an all-one vector. The optimal solution of the Kantorovich's formulation of Optimal Transport defines the distance between two QNNs. Let $\bar{n}_i = n_i + 1; Z, C \in \mathbb{R}^{\bar{n}_1 \times \bar{n}_2}; y_1 \in \mathbb{R}^{\bar{n}_1}, y_2 \in \mathbb{R}^{\bar{n}_2}$ such that $\mathbf{1}_{\bar{n}_1}^T y_1 = \mathbf{1}_{\bar{n}_2}^T y_2$

$$d(\mathcal{G}_1, \mathcal{G}_2) = \underset{Z}{\text{minimize}} \, \langle Z, C \rangle \text{ s.t. } Z \geq 0, Z\mathbf{1}_{\bar{n}_2} = y_1, Z^T \mathbf{1}_{\bar{n}_1} = y_2 \tag{18}$$

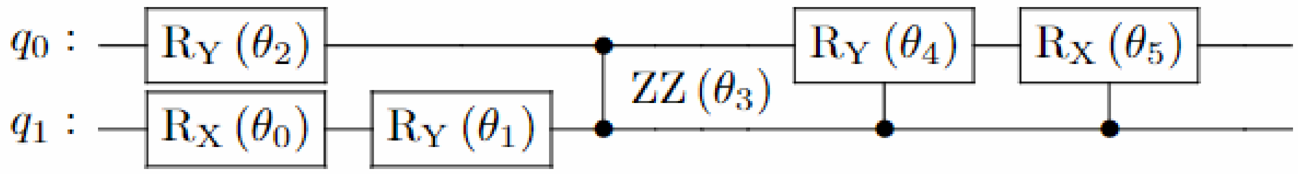
In the formulation, C is the matrix containing pairwise dissimilarity between component gates of the two circuits, or the "ground distance" in the optimal transport literature, and $Z = [z_{ij}]$ is the amount of mass of $u_i \in \mathcal{L}_1$ matched to $v_j \in \mathcal{L}_2$. We set $C = \begin{bmatrix} C_{\text{gtm}} + \nu C_{\text{str}} & \mathbf{1}_{n_1} \\ \mathbf{1}_{n_2}^T & 0 \end{bmatrix}$, $y_1 = [\{lm(u)\}_{u \in \mathcal{L}_1}, tm(\mathcal{G}_2)]^T$, and $y_2 = [\{lm(u)\}_{u \in \mathcal{L}_2}, tm(\mathcal{G}_1)]^T$. The last row and column in the cost matrix C are dedicated to a made-up *null gate* in each QNN. The weight of structural cost ν is a hyperparameter for the distance, usually set to 0.1. Matching two gates is subject to gate-type mismatch and structural dissimilarity cost while matching a gate to the null gate induces to a cost of 1, an upper bound for a finite d_{gate} value. The null gate of a circuit is where a leftover mass of any gates in the other is collected when there is no possible match for it.

$$d(\mathcal{G}_1, \mathcal{G}_2) = \underset{Z}{\text{minimize}} \langle Z, C \rangle \text{ s.t. } Z \geq 0, Z\mathbf{1}_{\bar{n}_2} = y_1, Z^T\mathbf{1}_{\bar{n}_1} = y_2$$

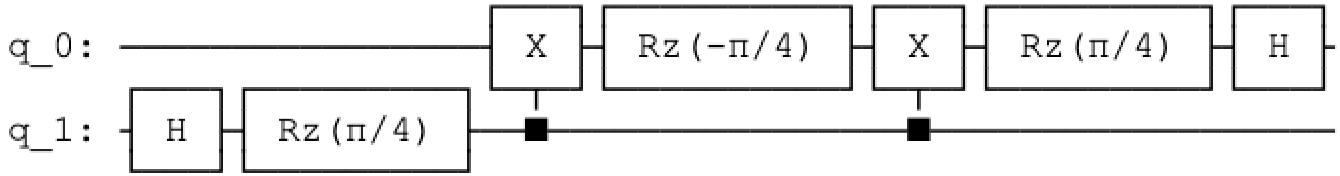


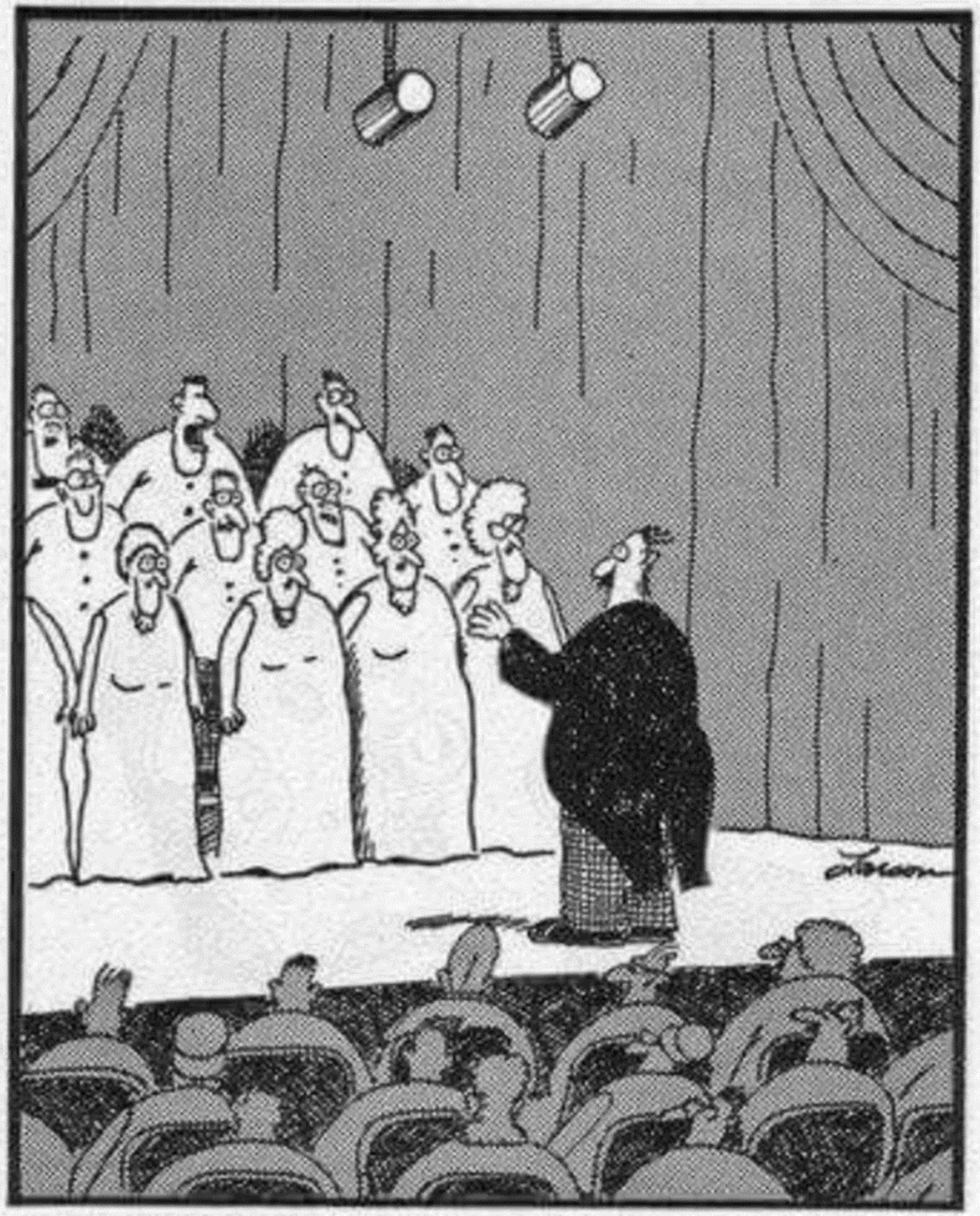
	QFT	MaxCut	QGAN		QFT	MaxCut	QGAN
Bayes. Opt.				**19 az.** [30]			
EI	**1**	**0.94**	**0.002**	depth 1	0.98		0.007
random	0.88	0.90	0.065	depth 2	1		
MaxCut az. [7]				**QGAN az.** [35]			
depth 1		0.746		depth 1	0.7		0.476
depth 2		0.751		depth 2	0.7		0.277
depth 3		0.762		depth 3	0.7		0.092

In the QFT experiment, the algorithm found some circuits with 6 gates that simulate the QFT operator perfectly. They outperform common ansatzs with many more gates and parameters, and requires less gates (6 gates) than the most optimized QFT circuit (7 gates) in Qiskit using the same pool of component gates



(a) QFT





In that one split second, when the choir's last note had ended, but before the audience could respond, Vinnie Conswego belches the phrase, "That's all, folks."