Title: Some recent stuff about the Quantum Approximate Optimization Algorithm

Speakers: Edward Farhi

Series: Colloquium

Date: January 18, 2023 - 2:00 PM

URL: https://pirsa.org/23010095

Abstract: I will introduce the QAOA and discuss some recent developments. These might include the application of the QAOA to the Sherrington-Kirkpatrick model, landscape independence, and the odd behavior when starting in a good place.

Zoom link: https://pitp.zoom.us/j/94804346887?pwd=c1hkRHBZZmRHS1RpM1BqU0R6TCtEdz09

Combinatorial Optimization

n bits $\quad$ m clauses

$$C(z) = \sum_{\alpha=1}^{m} C_\alpha(z) \qquad z_1 \ldots z_n = z$$

$$C_\alpha(z) = \begin{cases} 1 & \text{satisfies} \\ 0 & \text{does not} \end{cases}$$

$$C_{max} = \max_z C(z)$$

Want $\boxed{\dfrac{C(z)}{C_{max}}}$ big.

Quantum Algorithm Ingredients:

$$U(C,\gamma) = e^{-i\gamma C} = \prod_{\alpha=1}^{m} e^{-i\gamma C_\alpha}$$

$$B = \sum_{j=1}^{n} X_j \qquad X_j = \sigma_j^x$$

$$U(B,\beta) = e^{-i\beta B} = \prod_{j=1}^{n} e^{-i\beta X_j}$$

$$|s\rangle = \frac{1}{\sqrt{2^n}} \sum_z |z\rangle$$

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$$

All easy to construct!

For any integer $p \geq 1$ $\quad \underline{\gamma_1 \ldots \gamma_p = \vec{\gamma}}$ $\quad \underline{\beta_1 \ldots \beta_p = \vec{\beta}}$

$$\underline{|\vec{\gamma}, \vec{\beta}\rangle} = U(B,\beta_p)\,U(C,\gamma_p) \cdots U(B,\beta_1)\,U(C,\gamma_1)\,|s\rangle$$

Required Circuit Depth at most $mp + p$.

$$F_p(\vec{\gamma},\vec{\beta}) = \langle \vec{\gamma},\vec{\beta}\,|\,C\,|\,\vec{\gamma},\vec{\beta}\rangle$$

$$M_p = \max_{\vec{\gamma},\vec{\beta}} F_p(\vec{\gamma},\vec{\beta})$$

$$\boxed{M_p \geq M_{p-1}}$$

Can Show $\boxed{\lim_{p\to\infty} M_p = C_{max}}$

Quantum Algorithm with Angle Search

Fix p. Start with angles $(\vec{\gamma}, \vec{\beta})$

Use the Quantum Computer to make

$$|\vec{\gamma}, \vec{\beta}\rangle$$

Measure to get a string z and $C(z)$.

Repeat with same angles to get a good estimate of $F_p(\vec{\gamma}, \vec{\beta})$

Repeat with new angles to get near

$$M_p = \max_{\vec{\gamma}, \vec{\beta}} F_p(\vec{\gamma}, \vec{\beta})$$

For p fixed we can classically preprocess and determine the best angles in advance.

Example: Max Cut on 3-regular graphs

$$C = \sum_{\langle jk \rangle} C_{\langle jk \rangle}$$

$$C_{\langle jk \rangle} = \frac{1}{2}\left(-Z_j Z_k + 1\right) \qquad Z_j = \pm 1$$

p = 1 Look at contribution from edge $\langle jk \rangle$

$$\langle s | e^{i\gamma C} e^{i\beta B} Z_j Z_k e^{-i\beta B} e^{-i\gamma C} | s \rangle$$

$$|s\rangle = |+\rangle_1 |+\rangle_2 \cdots |+\rangle_n$$

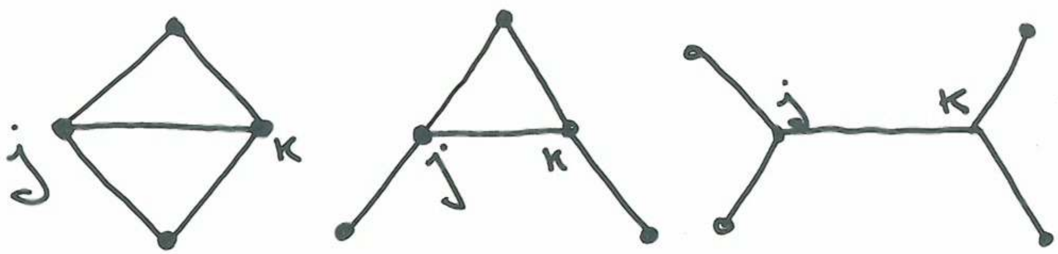
$$e^{i\beta B} Z_j Z_k e^{-i\beta B} = e^{i\beta(X_j+X_k)} Z_j Z_k e^{-i\beta(X_j+X_k)}$$

$$= [\cos 2\beta Z_j + \sin 2\beta Y_j][\cos 2\beta Z_k + \sin 2\beta Y_k]$$

only bits j k involved

Conjugate with $e^{i\gamma C}$

only bits connected to j,k involved



3 possible subgraphs

For p fixed we can classically preprocess and determine the best angles in advance.

Example: Max Cut on 3-regular graphs

$$C = \sum_{\langle jk \rangle} C_{\langle jk \rangle}$$

$$C_{\langle jk \rangle} = \frac{1}{2}\left(-z_j z_k + 1\right) \qquad z_j = \pm 1$$

$p=1$ Look at contribution from edge $\langle jk \rangle$

$$\langle s| e^{i\gamma C} e^{i\beta B} z_j z_k e^{-i\beta B} e^{-i\gamma C} |s\rangle$$

$$|s\rangle = |+\rangle_1 |+\rangle_2 \cdots |+\rangle_n$$

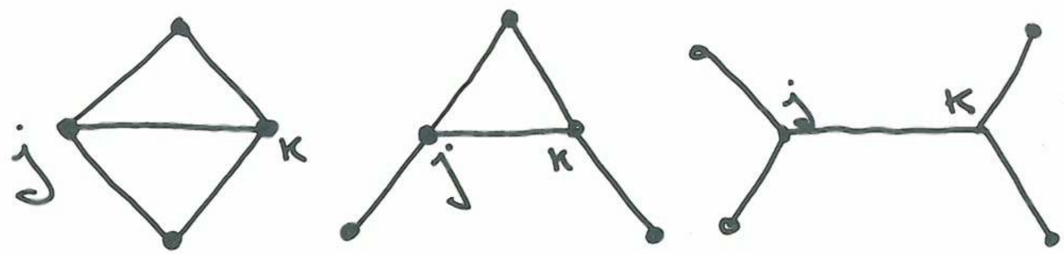
$$e^{i\beta B} Z_j Z_k e^{-i\beta B} = e^{i\beta(X_j+X_k)} Z_j Z_k e^{-i\beta(X_j+X_k)}$$

$$= [\cos 2\beta Z_j + \sin 2\beta Y_j][\cos 2\beta Z_k + \sin 2\beta Y_k]$$

only bits $j\ k$ involved

Conjugate with $e^{i\gamma C}$

only bits connected to j, k involved



3 possible subgraphs

Each subgraph type gives a function of γ, β which does not depend on n or m.

These can be evaluated on a classical computer looking at a 4, 5 or 6 qubit system. Then

$$F_1(\gamma, \beta)$$

can be evaluated on a classical computer and optimal angles chosen.

At $p=1$, the QAOA will produce a cut that is at least .6924 times the optimal cut. For ALL instances of 3-regular Max Cut!

Max E3LIN2

n variables, m equations each with 3 variables $(z_i + z_j + z_k) \bmod 2 = \begin{cases} 1 \\ 0 \end{cases}$ $z_i = 0, 1$

Instance is specified by a collection of triples and a 0 or 1 for each triple

Task Find a string that maximizes the number of satisfied equations. NP hard to find the optimal solution.

Try to find a "good" solution.

Approximate Optimization

Algorithm: Guess a Random String
Achieves $\frac{1}{2}$

Limit: $\left(\frac{1}{2}+\varepsilon\right)$ would imply P=NP 2001

Bounded Occurrence: Every variable is in no more than D equations

Algorithm: $\left(\frac{1}{2}+\frac{\text{const}}{D}\right)$ 2000

Quantum Algorithm: $\left(\frac{1}{2}+\frac{\text{const}}{D^{3/4}}\right)$ 2014

Classical Algorithm: $\left(\frac{1}{2}+\frac{\text{const}}{D^{1/2}}\right)$ 2015

Boaz Barak, Ankur Moitra,
Ryan O'Donnell, Prasad Raghavendra,
Oded Regev, David Steurer,
Luca Trevisan,
Aravindan Vijayaraghavan,
David Witmer, John Wright,
Johan Håstad

Quantum Approximate Optimization Algorithm E.F., Jeffrey Goldstone, Sam Gutmann

Better Analysis: $\left(\frac{1}{2} + \frac{1}{101\, D^{1/2} \ln D}\right)$

2015

Typical: $\left(\frac{1}{2} + \frac{1}{2\sqrt{3e}\, D^{1/2}}\right)$

Can we get rid of the $\ln D$???

If $\exists$ algorithm $\left(\frac{1}{2} + \frac{\text{const}}{D^{1/2}}\right)$
for a sufficiently large constant
then P=NP.

QAOA Typical Performance

Random Graphs with some distribution

e.g. * 3 regular graphs

or * each edge included with probability $3/(n-1)$

or * each edge included with probability $d/(n-1)$

or * whatever

Not Worst Case but Typical

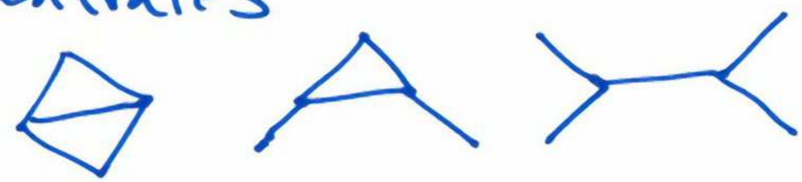
We demonstrate that if we fix parameters $(\vec{\gamma}, \vec{\beta})$ then for typical instances

$$\left[F(\vec{\gamma}, \vec{\beta}) = \langle \vec{\gamma}, \vec{\beta} | C | \vec{\gamma}, \vec{\beta} \rangle \right]$$

is the same for all instances (up to $\sqrt{m}$ fluctuations)

F.G.S.L. Brandao, M. Broughton
S. Gutmann, H. Neven

Consider $p=1$, Max-Cut on 3 regular graphs, uniformly generated. For large n fraction of subgraph types concentrates



$\Longrightarrow$ $\boxed{\langle \gamma, \beta | C | \gamma, \beta \rangle \text{ is instance independent}}$ $n \to \infty$

20 vertex 3-regular graphs
25 samples

	Low		Random		High	
p	Mean	Std.	Mean	Std.	Mean	Std.
2	6.636	0.319	14.691	0.036	22.409	0.228
3	5.218	0.294	15.125	0.042	23.109	0.175
4	3.933	0.259	14.627	0.157	23.822	0.272
5	3.132	0.159	15.725	0.113	24.349	0.179
6	2.550	0.100	16.404	0.140	24.918	0.266
7	1.954	0.088	15.975	0.096	25.110	0.221

angles are not optimal

Why? $F(\vec{\gamma},\vec{\beta}) = \sum_{\alpha=1}^{m} \langle \vec{\gamma},\vec{\beta} | C_\alpha | \vec{\gamma},\vec{\beta} \rangle$

Fix $\vec{\gamma},\vec{\beta}$ $\quad F = \sum_{\alpha=1}^{m} C_\alpha \qquad 0 \le C_\alpha \le 1$

There is a distribution over instances.

Think of C_α as a random variable.

F is of order m.

If the terms are independent, then regardless of the distribution, by the Law of Large Numbers, the standard deviation is of order $\sqrt{m}$ $\Longrightarrow$ concentration !!!!

To test this we measured the correlation coefficients between different $\mathcal{C}_\alpha$ on random instances.

Always very small. Like 0.04

This was at 20 bits.

We bet that the correlation coefficients at higher bit number will aslo be small.

Sherrington - Kirkpatrick Model

n bits $z_a = \pm 1$

$$\left[C_J(z) = \frac{1}{\sqrt{n}} \sum_{a<b} J_{ab} z_a z_b \right]$$

mean $J_{ab} = 0$ var $J_{ab} = 1$

$$\left[\lim_{n \to \infty} E_J \max_z \frac{C_J(z)}{n} = .763\ldots \right]$$

Parisi!

$$\langle s| e^{i\gamma C} e^{i\beta B} \frac{C}{n} e^{-i\beta B} e^{-i\gamma C} |s\rangle$$

p=1
QAOA

$$= \frac{1}{2^n} \sum_{z^1} \sum_{z^2} \sum_{z^3} e^{i\gamma C(z^1)} \langle z^1| e^{+i\beta B} |z^2\rangle \cdot$$

$$\frac{1}{n} C(z^2) \langle z^2| e^{-i\beta B} |z^3\rangle e^{-i\gamma C(z^3)}$$

$$C(z) = \frac{1}{\sqrt{n}} \sum_{a<b} J_{ab} z_a z_b$$

Now for ϕ small

$$\left[E_J e^{i\phi J} = (1 - \phi^2/2) + \cdots \right]$$

$$\left[E_J e^{i\phi J} J = i\phi + \cdots \right]$$

Upstairs $\gamma \sum_{a<b} J_{ab} (z_a^1 z_b^1 - z_a^3 z_b^3)$

Downstairs $\gamma \sum_{c<d} J_{cd} \, z_c^2 z_d^2$

Turn Crank

$$\left[E_J \langle \gamma, \beta | \frac{C_J}{n} | \gamma, \beta \rangle_J \text{ as } n \to \infty \right.$$
$$\left. = \underline{2 \sin 2\beta \cos 2\beta \; \gamma e^{-2\gamma^2}} \right]$$

Optimum at $\beta = \pi/8$ $\gamma = 1/2$

get $\boxed{\frac{1}{2\sqrt{e}} \simeq .303}$

For any fixed p we have an iterative procedure to evaluate

$$\left[\lim_{n\to\infty} E_J \; {}_J\langle \vec{\gamma}, \vec{\beta} \,|\, \frac{C_J}{n} \,|\, \vec{\gamma}, \vec{\beta} \rangle_J \right]$$

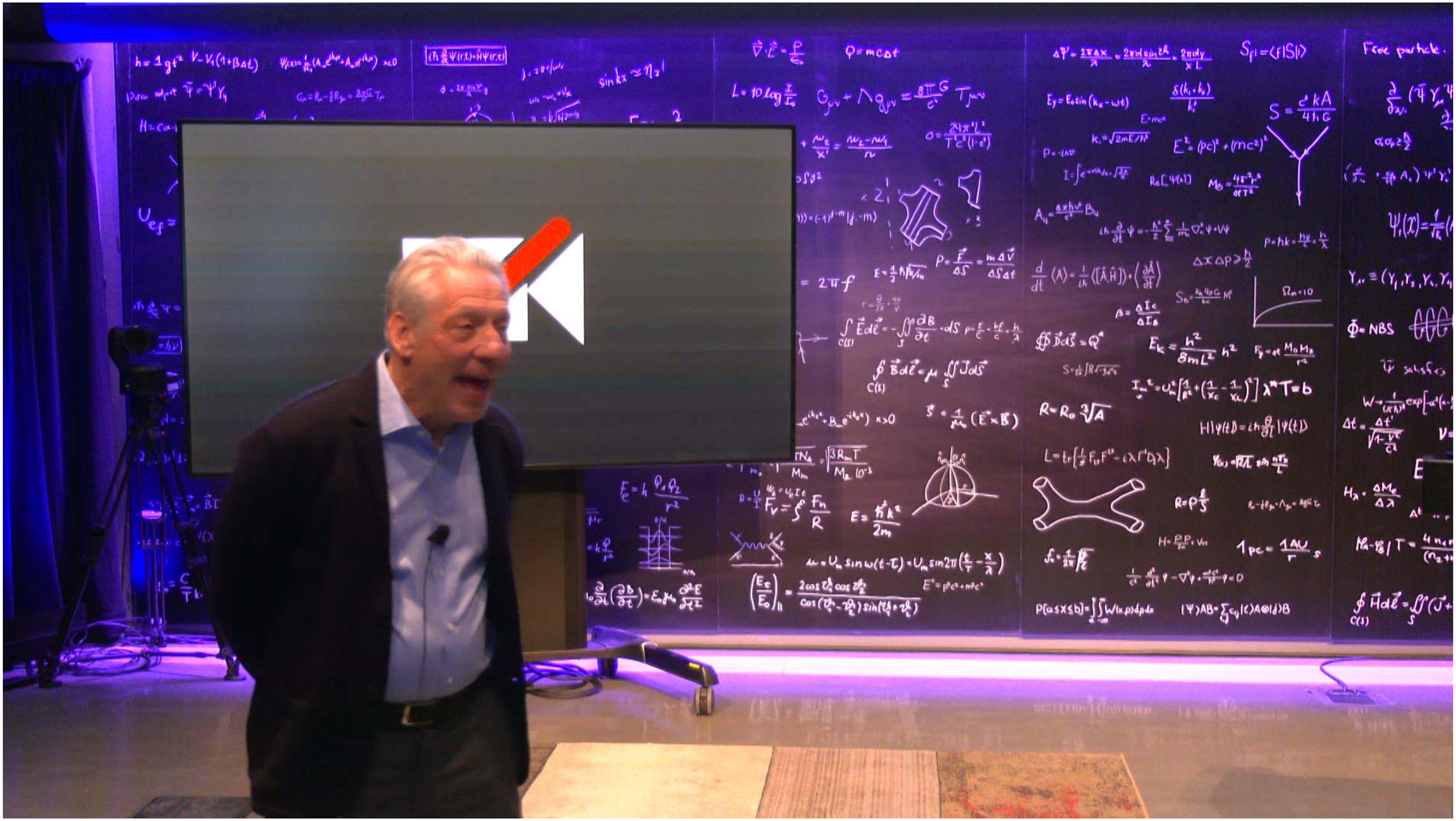
which takes $\sim 4^p$ steps

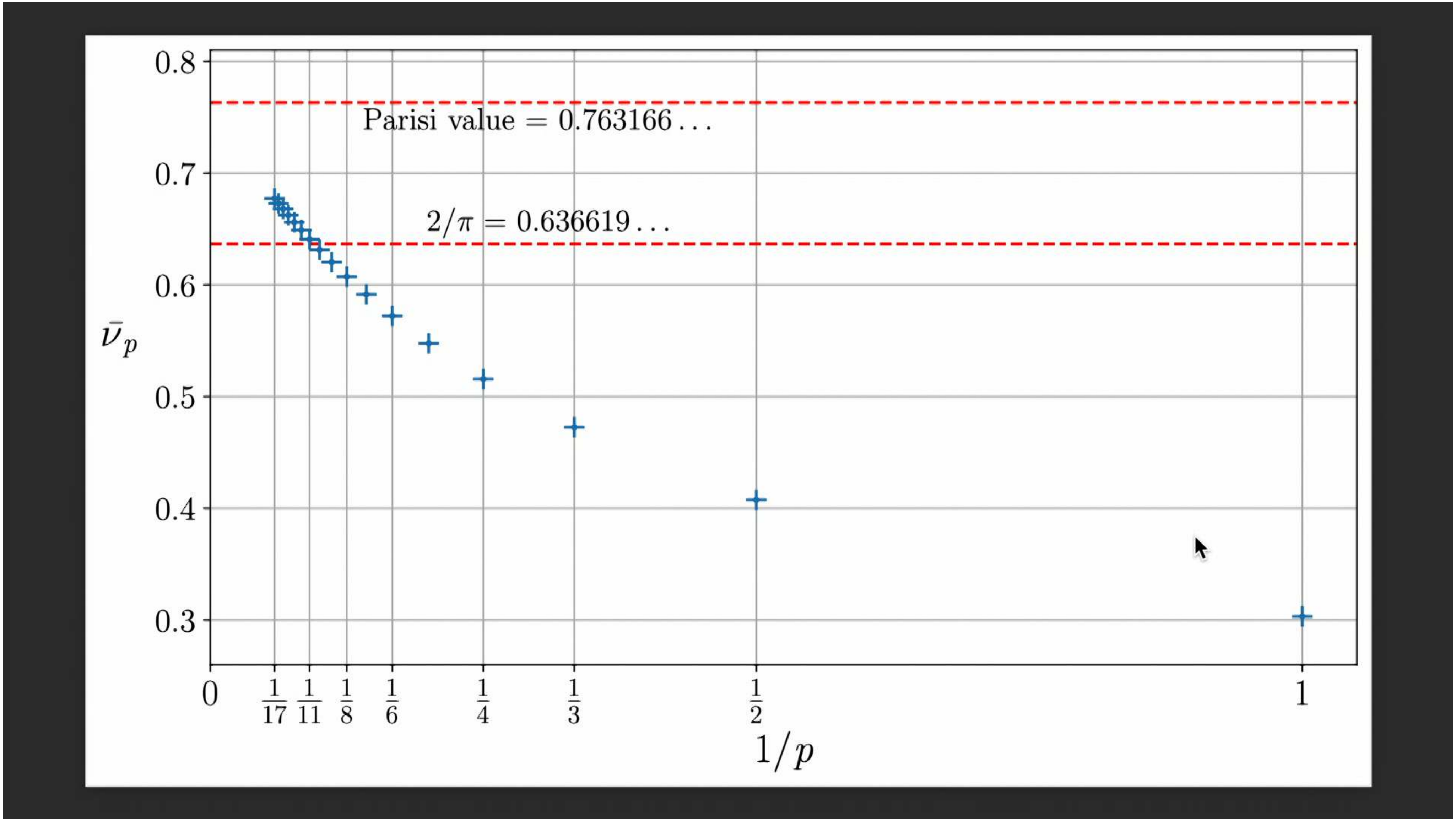
E.F.
S. Gutmann
J. Goldstone
L. Zhou

Again p fixed $n \to \infty$.

Call this quantity

$$V_p^{SK}(\vec{\gamma}, \vec{\beta})$$

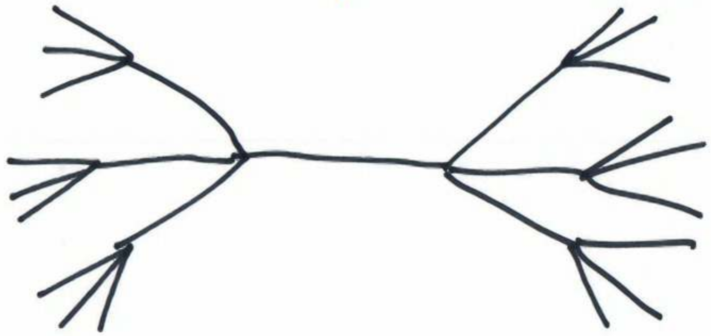




Parisi value = 0.763166...
2/π = 0.636619...
$\bar{\nu}_p$
0.8
0.7
0.6
0.5
0.4
0.3
0
$\frac{1}{17}$ $\frac{1}{11}$ $\frac{1}{8}$ $\frac{1}{6}$ $\frac{1}{4}$ $\frac{1}{3}$ $\frac{1}{2}$ 1
$1/p$

Consider Max-Cut on large-girth D-regular graphs. Apply QAOA

For fixed p see only a tree neighborhood



One subgraph

$$\frac{1}{|E|} \langle \vec{\gamma}, \vec{\beta} | C_{MC} | \vec{\gamma}, \vec{\beta} \rangle$$

$$= \frac{1}{2} + \frac{\nu_p(D, \vec{\gamma}, \vec{\beta})}{\sqrt{D}}$$

Now

$$\boxed{V_p^{SK}(\vec{\gamma},\vec{\beta}) = \lim_{D\to\infty} \mathcal{V}_p(\vec{\gamma},\vec{\beta})}$$

average over instances = one subgraph !

J. Basso, E.F., K. Marwaha
B. Villalonga and L. Zhou

Also Show

$$\lim_{n\to\infty} E_J\left[\langle\vec{\gamma},\vec{\beta}|\left(\frac{C}{n}\right)^2|\vec{\gamma},\vec{\beta}\rangle\right]$$

$$= \lim_{n\to\infty} E_J^2\left[\langle\vec{\gamma},\vec{\beta}|\frac{C}{n}|\vec{\gamma},\vec{\beta}\rangle\right]$$

which proves Landscape Independence for the SK Model.

Can also show that for $n\to\infty$, for typical instances all measured strings z have

$$\left\{\frac{C(z)}{n} \text{ close to } \langle\vec{\gamma},\vec{\beta}|\frac{C}{n}|\vec{\gamma},\vec{\beta}\rangle\right\}$$

Want to show that for $n \to \infty$,
and then $p \to \infty$ we achieve the
Parisi value of .763...

Want to apply our techniques
to other problems where
instances are drawn from a
distribution.... . . . . ?

The QAOA gets stuck starting from a good classical string!

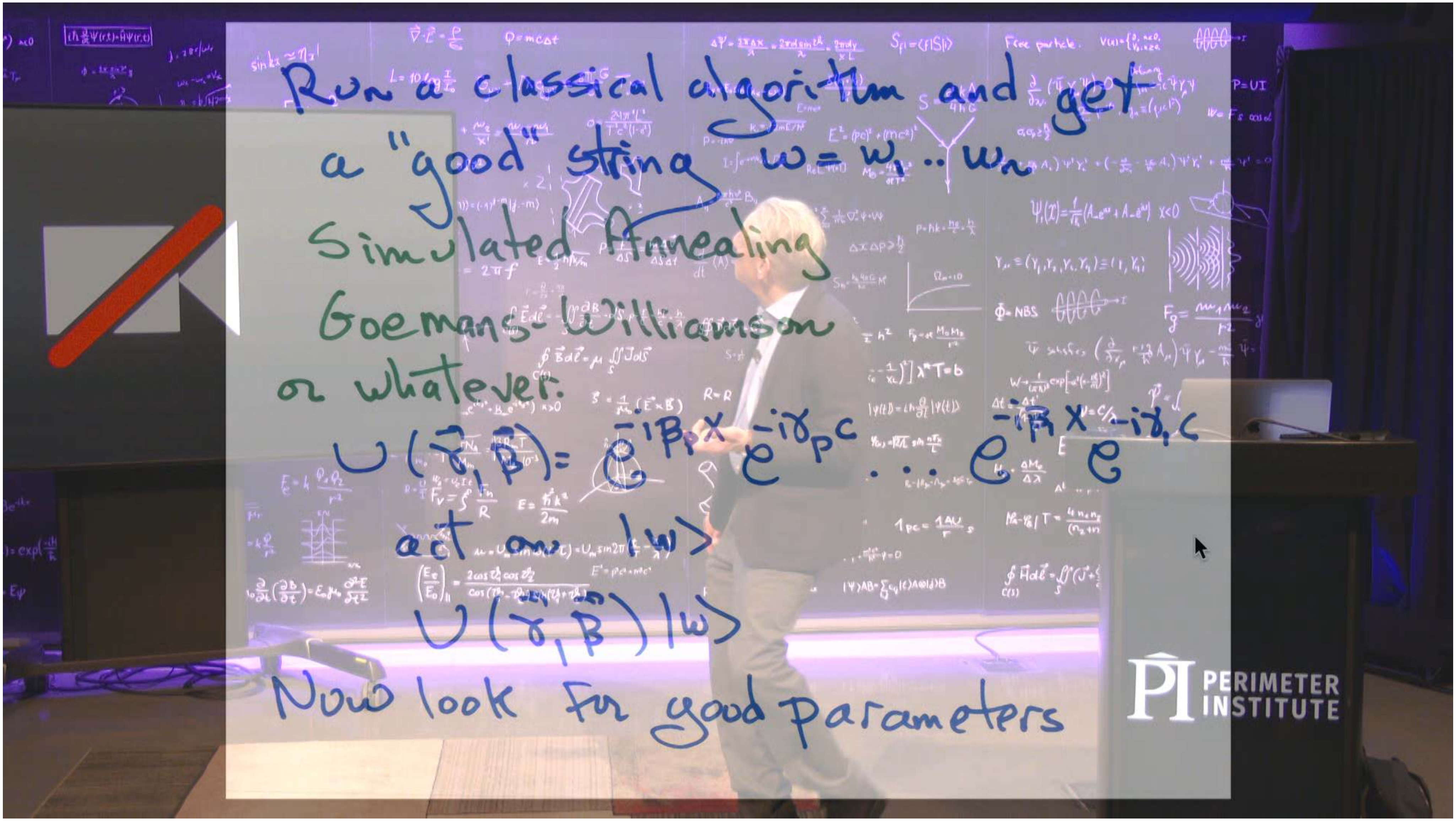
M. Cain, E.F., S. Gutmann, D. Ranard, E. Tang

Usual QAOA starts in

$$|s\rangle = \frac{1}{\sqrt{2^n}} \sum_z |z\rangle \quad \text{superposition of all inputs}$$

$$= |+\rangle_1 |+\rangle_2 \ldots |+\rangle_n \quad \text{product state}$$

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$$



Run a classical algorithm and get a "good" string $w = w_1 \dots w_n$

Simulated Annealing

Goemans-Williamson

or whatever.

$$U(\vec{\gamma}, \vec{\beta}) = e^{-i\beta_p X} e^{-i\gamma_p C} \cdots e^{-i\beta_1 X} e^{-i\gamma_1 C}$$

act on $|w\rangle$

$$U(\vec{\gamma}, \vec{\beta})|w\rangle$$

Now look for good parameters

Run a classical algorithm and get a "good" string $\omega = \omega_1 \ldots \omega_n$

Simulated Annealing

Goemans-Williamson

or whatever.

$$U(\vec{\gamma}, \vec{\beta}) = e^{-i\beta_p X} e^{-i\gamma_p C} \cdots e^{-i\beta_1 X} e^{-i\gamma_1 C}$$

act on $|\omega\rangle$

$$U(\vec{\gamma}, \vec{\beta})|\omega\rangle$$

Now look for good parameters

Dramatic Failure

! At low bit number simulation we can not !
! find any improvement beyond $C(\omega)$. !

For example. 16 bit 3-regular graph

24 edges

average cut value 12

Start on $|\omega\rangle$'s where $C(\omega) = 15$

Look at $p = 1, 2, 3 \ldots$

Can not find parameters that improve cost beyond 15 (maybe by .2)

Slight Intuition

$$C_{MC} = \sum_{\langle ij \rangle} \frac{1}{2}(1 - z_i z_j) = \sum_{\langle ij \rangle} C_{ij}$$

$\langle w | U^\dagger C_{ij} U | w \rangle$

$U = 1$ 1 if $w_i \neq w_j$

$\quad\quad\quad\ \ 0$ if $w_i = w_j$

$U \neq 1$ $1 - \varepsilon$ if $w_i \neq w_j$

$\quad\quad\quad\ \ \delta$ if $w_i = w_j$

Suppose $\varepsilon = \delta$. In a good string there are more satisfied edges than unsatisfied. More edges get worse than those that improve. Net loss

Decoupled graph

For any QAOA unitary U we can prove that $\varepsilon = \delta$ for each

At any depth starting in a good string fails. However $p=1$ QAOA starting in $|s\rangle$ works perfectly!

(Even with $\pm Z_i Z_j$ for each edge!)

We also have beautiful thermal arguments for more general cases!

Conclusion

* QAOA is an easy to implement all purpose optimizer
* Worst case performance guarantees
* Performance improves with depth
* Can analyze random instances in the infinite size limit
* Not a hint of any obstacle to performance starting in |s>
* Numerics are uniformly encouraging
* Need to learn more!!