Quantum Simulated Annealing

Howard Barnum[1] (presenting)
&
Rolando Somma[2] Sergio Boixo[1,3] Manny Knill [4]

[1]Los Alamos National Laboratory
[2]Perimeter Institute
[3]University of New Mexico
[4]National Institute of Standards and Technology (Boulder)

March 28, 2008 / Classical and Quantum Information Theory, Santa Fe, NM

Introduction and outline

Motivation

Markov chain Monte-Carlo (MCMC) algorithms, such as simulated annealing (SA) are important in statistical physics and optimization. Can we get a quantum speed-up?

Combinatorial optimization

Problem instance

Minimize a nonnegative objective function E, over a search space $\mathcal{S}$.

- $\mathcal{S}$ is a set of d elements, called *configurations*.
- $\mathcal{S}$ may have additional structure, e.g. providing a notion of locality.
- d typically is exponential in the size of the problem instance.

Notation:

- $\mathcal{S}_0$: the set of minimal-energy configurations.
- E_M: $E_{max} - E_{min}$; typically polynomial in instance size
- γ: the classical gap between E_{min} and the next-lowest energy.

E_M/γ is a natural parameter: dynamic range / resolution.

Los Alamos NATIONAL LABORATORY EST. 1943

Combinatorial Optimization II: Examples

- $\mathcal{S}$ the set of paths on a graph with n vertices whose edges are assigned lengths, $E(\sigma)$ the length of path σ.
- $\mathcal{S}$ the set of spin configurations (assignments $\sigma = (\sigma_1, \sigma_2, ..., \sigma_N)$ of $\sigma_i = \pm 1$ to each vertex) on an N-site lattice L whose edges are colored with nearest-neighbor interaction energies h_{ij}, $d = 2^N$, E the Ising Hamiltonian

$$E(\sigma) = \sum_{<i,j>\in L} h_{ij}\sigma_i\sigma_j + \sum_i B_i\sigma_i \ . \tag{1}$$

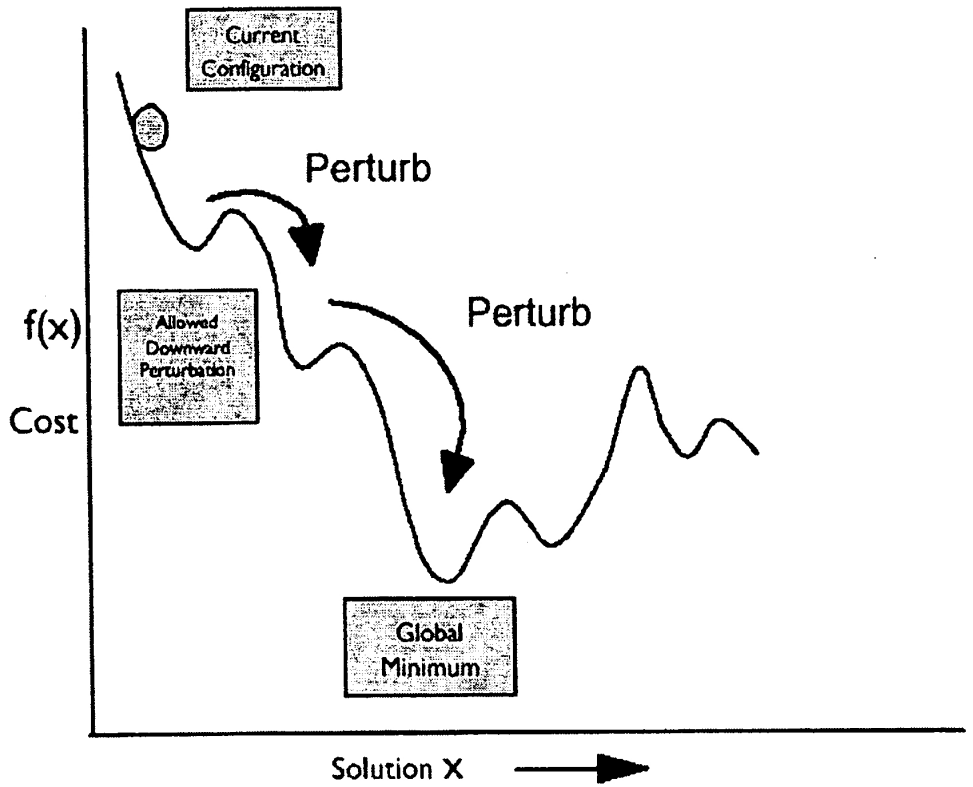
Typically NP-hard.

Combinatorial optimization (III)

With d exponential in problem size, exhaustive search is bad news.
When objective function reflects structure—e.g. notion of locality or distance—in $\mathcal{S}$, other methods may outperform exhaustive search.
E.g., with something like global convexity, gradient-like methods can help.
Can we do better when some "smoothness", but no "global convexity"?

Simulated annealing

Simulated annealing starts system at high "temperature", and cools to "ground space" $\mathcal{S}_0$. "Thermal excitations" at intermediate temperatures intended to pop system out of local minima.

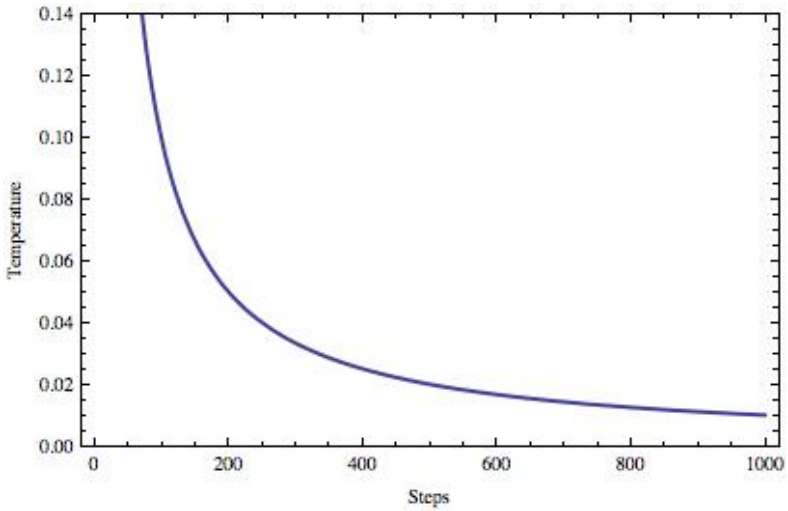


Simulated Annealing (II)

Start hot: cool and sample from the resulting Boltzmann-Gibbs distribution concentrated on $\mathcal{S}_0$.

- Choose an annealing schedule $\beta_1 < \beta_2 < \ldots < \beta_P$.
- Choose $\{M(\beta_k)\}$. Each $M(\beta_k)$ is a stochastic matrix with Boltzmann-Gibbs equilibrium distribution $\vec{\pi}$,

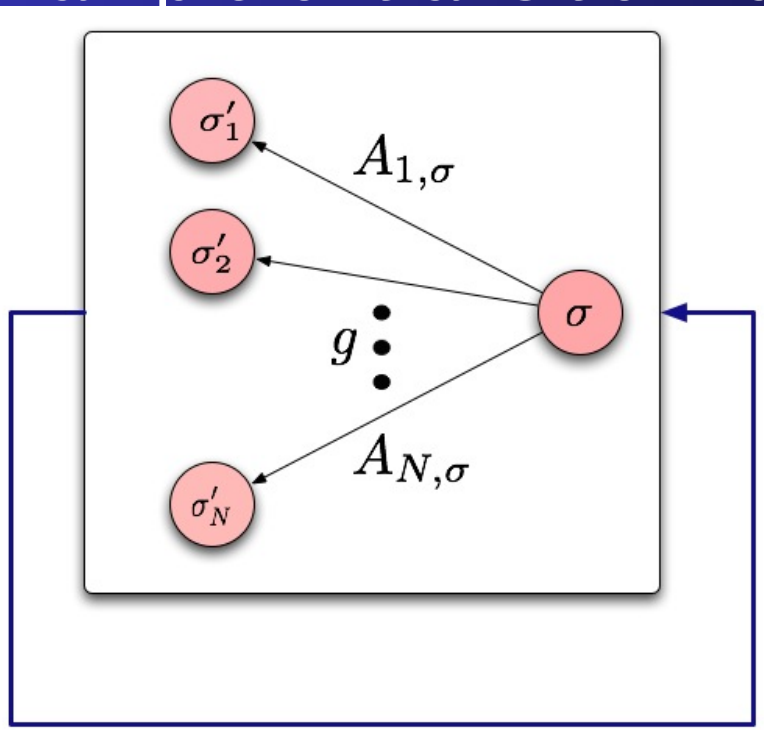
$$\vec{\pi} = M(\beta_k)\vec{\pi} \ , \quad \pi^\sigma = e^{-\beta E[\sigma]}/\mathcal{Z} \ .$$



- At each step

$$\sigma^{(k-1)} \xrightarrow{M(\beta_k)} \sigma^{(k)} \ .$$

Example of transition rule: Metropolis



Start with σ_0 and repeat:

1. Select test state σ' with rule g
2. Jump to σ' with rule $A_{\sigma'\sigma}$

$$A_{\sigma'\sigma} = \begin{cases} e^{-\beta\Delta E} & \Delta E > 0 \\ 1 & \text{otherwise} \end{cases}$$

Figure: Markov chain Monte-Carlo with Metropolis rule

Average after mixing

$$\langle O \rangle = \frac{\sum O(\sigma_f)}{\#\text{runs}}$$

Mixing time τ_{mix} (Aldous)

For δ the gap of $M(\beta) \equiv Ag$.

$$\delta^{-1} \leq \tau_{\text{mix}} \leq \delta^{-1}(\ln 1/\pi*)$$

SA convergence and cost

$$\text{Final state} \quad \vec{\mu}(\beta_f) = \left(\prod_{k=1}^{P} M(\beta_k) \right) \vec{\mu}(0) \; .$$

- Convergence: Cool slowly enough (rate proportional to gap).

$$\Delta\beta = \mathcal{O}(\delta / E_M) \quad \Rightarrow \quad \sum_{\sigma=1}^{d} \frac{(\mu^\sigma(\beta_f))^2}{\pi^\sigma(\beta_f)} \leq 2 \; .$$

- Error from the temperature: Cool enough.

$$\sum_{\sigma \notin \mathcal{S}_0} \pi^\sigma(\beta_f) \leq d\, e^{-\beta_f \gamma} \leq \epsilon^2 \; .$$

SA cost

$$\mathcal{O}(\beta_f / \Delta\beta) = \mathcal{O}(\beta_f E_M / \delta) = \mathcal{O}\left(\frac{E_M}{\gamma} \frac{\log(d/\epsilon^2)}{\delta} \right)$$

Adiabatic Computation Farhi et al., quant-ph/0001106

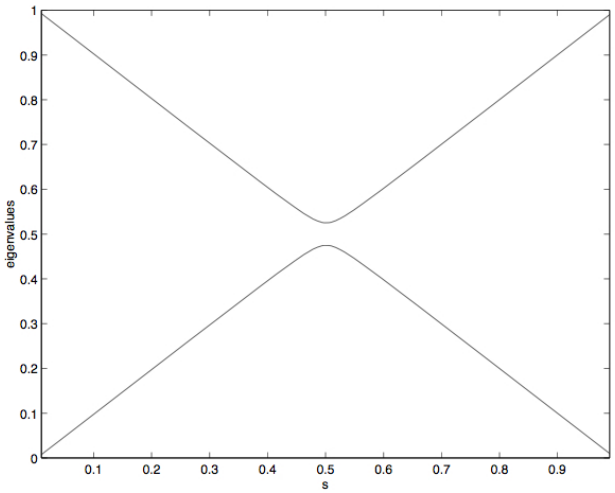
Use the adiabatic theorem to go from an "easy" ground state to the solution ground state.

$$H(s) = \left(1 - \frac{s}{T}\right) H_{\text{Easy}} + \frac{s}{T} H_{\text{Problem}} \ .$$

Adiabatic condition

Move slower than the slowest frequencies (gap squared).

$$\|\dot{H}\|/\delta^2 \leq \epsilon \quad \Rightarrow \quad T = \mathcal{O}(1/(\epsilon\delta^2)) \ .$$



AQC ≡ BQP

Adiabatic quantum computation is general quantum computation.

Doing SA with AQC

Somma, Batista, & Ortiz, PRL 99, 030603 (2007)

Define the operator $H_c = \sum_\sigma E[\sigma] \, |\sigma\rangle\langle\sigma|$.
Let $M(\beta)$ be an ergodic transition matrix satisfying detailed balance whose stationary state π_σ is the thermal distribution for E at β.
Define a Hamiltonian

$$H_q(\beta) \equiv \mathbb{1} - e^{\beta H_c/2} M(\beta) e^{-\beta H_c/2} \ .$$

Its ground state is the quantum Gibbs state,

$$|\phi_0\rangle \equiv \sum_\sigma \sqrt{\pi^\sigma} \, |\sigma\rangle \equiv \frac{e^{-\beta H_c/2}}{\sqrt{\mathcal{Z}}} \sum_\sigma |\sigma\rangle \ .$$

Measuring in the $|\sigma\rangle$ basis gives $|\sigma\rangle$ with the same probability as in the classical thermal state.

Cost of SA with AQC

Apply the adiabatic condition and the definition of the Gibbs state: $T = \mathcal{O}(1/(\epsilon\delta))$. Same cost as SA.

Quantum Walks

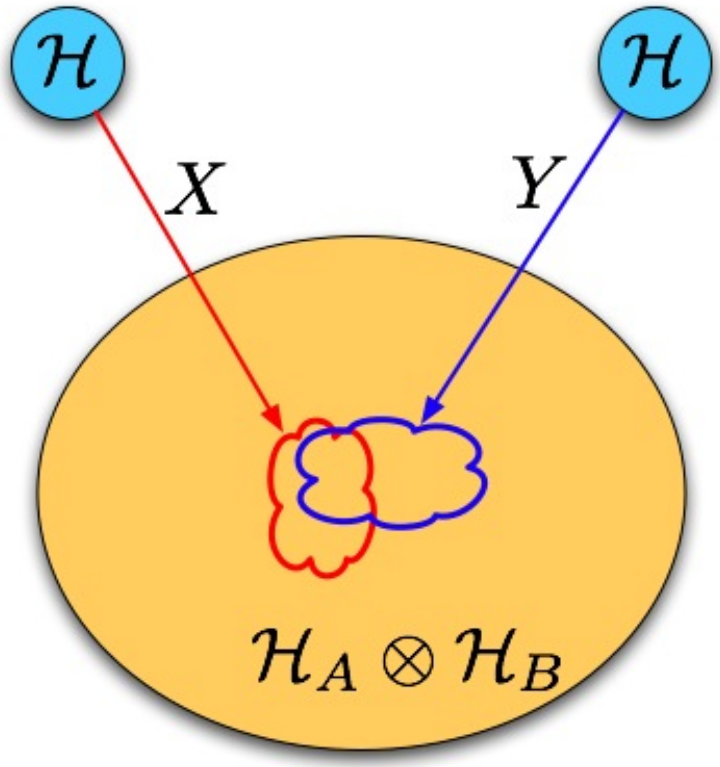
Ambainis, SIAM Journal on Computing 37, 210-239 (2007), and FoCS 2004 22-31. Szegedy, FoCS, (2004).

Szegedy's walks: Duplicate the Hilbert space $\mathcal{H}$ into $\mathcal{H}_A \otimes \mathcal{H}_B$.

1. Define the isometries K and Y via:

$$X|\sigma\rangle = |\sigma\rangle \sum_{\sigma'} \sqrt{m_{\sigma\sigma'}} |\sigma'\rangle ,$$

$$Y|\sigma'\rangle = \sum_{\sigma} \sqrt{m_{\sigma'\sigma}} |\sigma\rangle |\sigma'\rangle \ .$$



2. Define the reflections

$$\mathrm{ref}_1 = 2XX^\dagger - \mathbb{1}$$

$$\mathrm{ref}_2 = 2YY^\dagger - \mathbb{1} \ .$$

3. Rotate with

$$W_{\mathrm{sz}} = \mathrm{ref}_2 \mathrm{ref}_1 \ .$$

Spectrum of Szegedy's Quantum Walk

- Note that $H \equiv e^{\beta H_c/2} M e^{-\beta H_c/2} = X^\dagger Y$.
- Define eigenphases φ_j for the eigenvalues $|\phi_j\rangle$

$$H|\phi_j\rangle = \cos\varphi_j |\phi_j\rangle = X^\dagger Y |\phi_j\rangle \ .$$

  Note that $\varphi_0 = 0$ and $|\phi_0\rangle$ is the *quantum Gibbs state*.
- Because

$$XX^\dagger \left(Y|\phi_j\rangle\right) = \cos\varphi_j X|\phi_j\rangle$$
$$YY^\dagger \left(X|\phi_j\rangle\right) = \cos\varphi_j Y|\phi_j\rangle \ ,$$

  the rotation

$$W = \left(2YY^\dagger - \mathbb{1}\right)\left(2XX^\dagger - \mathbb{1}\right)$$

  preserves each of the (at most) 2-dimensional subspaces $\{Y|\phi_j\rangle, X|\phi_j\rangle\}$, where it acts as a rotation by $2\varphi_j$. So the spectrum of W is $e^{\pm i2\varphi_j}$.

Speedup of Quantum Simulated Annealing

Somma, Boixo, & Barnum, arxiv:0712.1008.

- The Hamiltonian H of the rotation W has eigenvalues φ_j. The gap is $|\varphi_0 - \varphi_1| = \varphi_1$.

$$\delta = 1 - \lambda_1 = 1 - \cos\varphi_1 \approx \varphi_1^2/2 \ .$$

  We get a better gap $\varphi_1 \approx \sqrt{\delta}$.
- The quantum Gibbs state

$$|\phi_0\rangle \equiv \frac{e^{-\beta H_c/2}}{\sqrt{\mathcal{Z}}} \sum_{\sigma} |\sigma\rangle$$

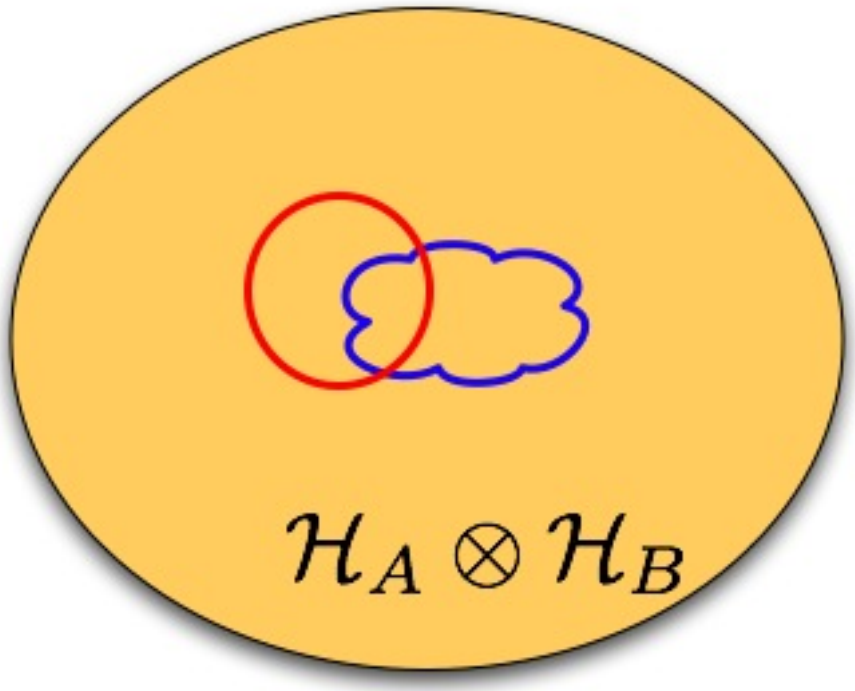
  has always eigenvalue 0.
- Rotating to the Gibbs state from some easy to prepare state does not work: the cost depends on the initial overlap, which can be exponentially small Richter, PRA 76, 042306 (2007). Use the adiabatic algorithm to get rid of the overlap.

Dealing with the degenerate 0-eigenspace (I)

We want to use the adiabatic theorem for $W(M(\beta))$ to follow $|\phi_0(\beta)\rangle$.

Problem

The Gibbs state $|\phi_0\rangle$ has eigenphase 0. So do most of the states, which do not rotate.



Solution

Mark $|\phi_0(\beta)\rangle$ to make it orthogonal to all other 0-eigenvectors.

Dealing with the degenerate 0-eigenspace (II)

- Define

$$U_X |\sigma\, o\rangle \equiv X |\sigma\rangle ,$$
$$U_Y |o\, \sigma\rangle \equiv Y |\sigma\rangle ,$$

  for some with $|o\rangle$ in $\mathcal{H}$.
- Rotate with

$$W(M(\beta)) = U_X^\dagger U_Y P_{oA} U_Y^\dagger U_X P_{oB} ,$$

  P_{oA} and P_{oB} are the selective sign change operations on the states $|o\rangle$ of $\mathcal{H}_A$ and $\mathcal{H}_B$

$$P_{oB} = 2\mathbb{1} \otimes |o\rangle \langle o| - \mathbb{1} \otimes \mathbb{1} .$$

- The two dimensional subspaces of $W(M(\beta))$ are $\{|\phi_j\, o\rangle , U_X^\dagger U_Y |o\, \phi_j\rangle\}$. The quantum Gibbs state is always $|\phi_0(\beta)\, o\rangle$, and any other 0-eigenvector orthogonal to it does not have $|o\rangle$ in $\mathcal{H}_B$.

QSA with adiabatic condition

- To first order in $\Delta\beta$,

$$|\phi_0(\beta_{k-1})\rangle \approx \left(1 - \frac{\Delta\beta}{2}\left(\langle E\rangle_{\beta_k} - H_c\right)\right)|\phi_0(\beta_k)\rangle \ .$$

- The adiabatic condition gives

$$\partial_t\beta(t)\left|\frac{\langle\psi_{\pm j}(\beta)|\partial_\beta\psi_0(\beta)\rangle}{2\varphi_j}\right| \leq \partial_t\beta(t)\frac{E_M}{2\varphi_1} \leq \epsilon \ ,$$

which gives a cost $T = \mathcal{O}(1/\epsilon\sqrt{\delta})$.
Nevertheless, this is not a rigorous version of the adiabatic theorem. We can give a rigorous proof of convergence, by using the Zeno effect in a way inspired by *Childs et al., PRA, 66, 032314, (2002).* .

QSA with Zeno effect (I)

Instead of evolving adiabatically with $H(\beta)$, project successively onto the quantum Gibbs states of $H(0), H(\Delta\beta), \ldots, H(\beta_f)$.

Quantum Zeno

$$\langle \phi_0(\beta_k) | \, \phi_0(\beta_{k-1}) \rangle \approx 1 - (\Delta\beta E_M)^2 \, .$$

For $\mathcal{O}(1/\Delta\beta)$ steps, each with fidelity $1 - \mathcal{O}\left(\Delta\beta^2\right)$, the final fidelity is $1 - \mathcal{O}(\Delta\beta)$.

- It is enough to choose $\Delta\beta = \mathcal{O}(\epsilon/(\beta_f E_M^2))$. $\Delta\beta$ can be much bigger than the gap.
- The cost of each "projection", using phase estimation, goes with the difference of the phases. That is, goes with $\varphi_1 \approx \sqrt{\delta}$. This dominates the overall cost.

QSA with Zeno effect (II)

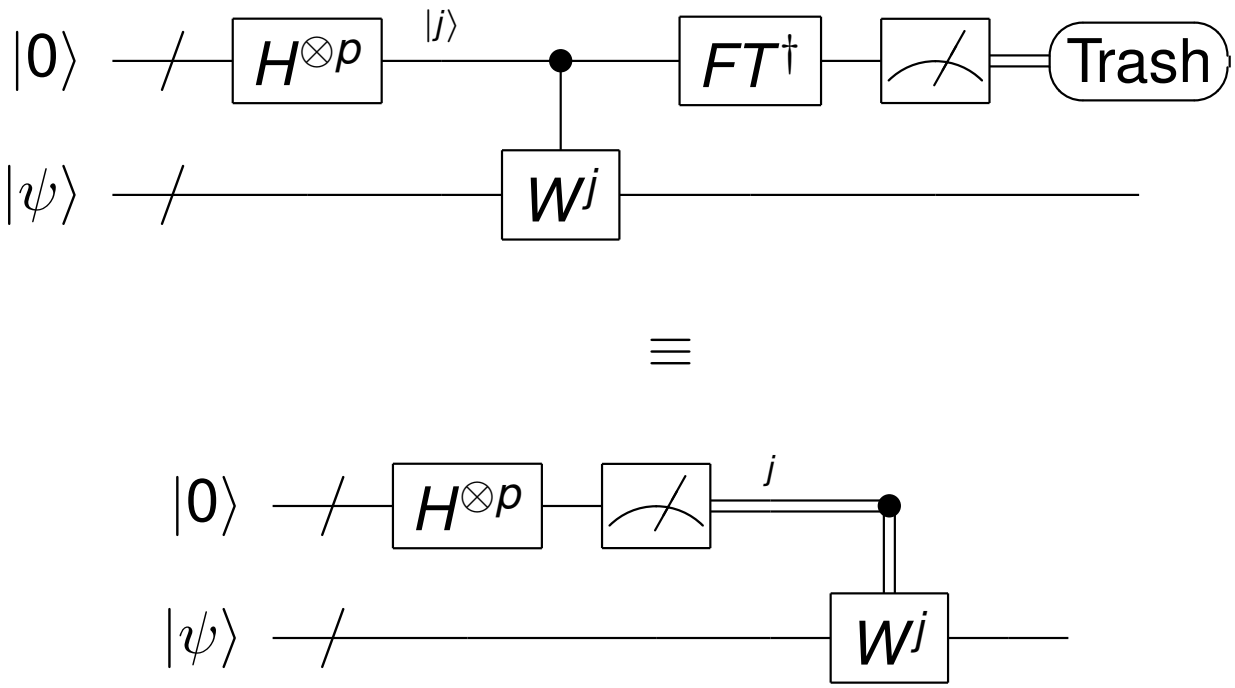


Figure: Phase estimation algorithm and randomization. The first (p-qubit) register encodes a p-bit approximation to an eigenphase of $W(M(\beta_k))$ on readout. The second register is $\mathcal{H}_A \otimes \mathcal{H}_B$. The first register is initialized by $H^{\otimes p}$ to a superposition of all states in the computational basis; a sequence of $W(M(\beta_{k+1}))$ operations, controlled by the first register, is performed on the second register; the inverse quantum Fourier transform is applied to the first register; the first register is measured. When measurement projects the first register onto $|0\rangle$, the second register is approximately projected into the 0-phase eigenspace. Tracing out the first register after the controlled operations, the overall effect on the second register is identical to the randomization procedure depicted in the lower part of the figure, where double horizontal lines indicate classical bits, double vertical lines classical control.

Cost of QSA

- The final temperature needs to be low. $\beta_f = \log(d/2\epsilon)/\gamma$.
- For Zeno effect, we need to go slow enough. $\Delta\beta = \mathcal{O}(\epsilon/(\beta_f E_M^2))$.
- This gives a total number of steps $\mathcal{O}((\beta_f E_M)^2/\epsilon)$.
- The cost of each step, from phase estimation, is $\mathcal{O}((\Delta\beta E_M\sqrt{\delta})^{-1})$.
- The total cost is

$$\mathcal{O}\left(\left(\frac{E_M}{\gamma}\right)^2 \frac{\log^2(d/\epsilon)}{\epsilon\sqrt{\delta}}\right) .$$

- The polynomial dependence on the error ϵ can be avoided by $\mathcal{O}(\log 1/\epsilon$ repetitions of the algorithm with initial error target $\epsilon_0 = 1/2$. Similar improvement can be obtained for preparation of the final $|\psi_0\rangle$.

Conclusion

1. Simulated Annealing goes like $\mathcal{O}(1/\delta)$.
2. Our algorithm mimics SA and goes like $\mathcal{O}(1/\sqrt{\delta})$.