

Lecture 8: QMA, Circuit-to-Hamiltonian Constructions, and Adiabatic Quantum Computation

Computational Aspects of Quantum Simulation
ENS de Lyon

2026

Abstract

In the last lecture we introduced quantum ground-state problems and several methods for attacking them. Today we formalize the complexity class that captures the difficulty of estimating ground-state energies: QMA. We then sketch the QMA-completeness of the k -local Hamiltonian problem, which introduces two key ideas: history states and circuit-to-Hamiltonian constructions.

Learning objectives. After this lecture, students should be able to:

- state the definition of QMA and its basic properties;
- explain the proof strategy showing that the k -local Hamiltonian problem is QMA-complete;
- describe the circuit-to-Hamiltonian (Feynman–Kitaev) construction and the role of history states;

Main message. QMA is the quantum analogue of NP: the proof is a quantum state, and the k -local Hamiltonian problem is its canonical complete problem. The reason is the circuit-to-Hamiltonian construction, which encodes an entire quantum computation, as a *history state*, into the ground state of a local Hamiltonian.

All exercises are optional unless stated otherwise.

Contents

1	The classes NP and QMA	1
1.1	The class NP	1
1.2	The class QMA	2
2	The k-local Hamiltonian problem	3
2.1	Problem statement	3
2.2	Containment in QMA	3
2.3	QMA-hardness: the quantum Cook–Levin theorem (proof sketch)	4
3	Assigned Exercise	6
	References	6

1 The classes NP and QMA

1.1 The class NP

Definition 1.1 (NP). A language $L \subseteq \{0, 1\}^*$ is in NP if and only if there exists a polynomial-time uniform circuit family $\{C_n\}_{n \in \mathbb{N}}$ such that, for all x with $n := |x|$,

$$x \in L \iff \exists y \in \{0, 1\}^{\text{poly}(n)} \text{ such that } C_n(x, y) = 1.$$

In this formulation, NP describes problems for which every YES instance admits a short classical proof that can be verified efficiently. The verifier need not know how to *find* the witness; it only needs to check it once it is supplied.

Recall that MAX-Cut is NP-Complete, and that from a previous class we showed that we can write the MAX-Cut optimization problem on a graph $G(V, E)$ in the form:

$$\lambda_{\min}(H), \quad H := \sum_{(i,j) \in E} Z_i Z_j.$$

In physics this is called the anti-ferromagnetic Ising model. Thus this shows that the class NP captures the difficulty of the *classical* local Hamiltonian problem. In this class we will explore which class captures the difficulty of the quantum analog of this problem.

1.2 The class QMA

QMA is obtained by allowing the proof to be a polynomial-size quantum state and the verifier to perform a polynomial-size quantum computation.

Definition 1.2 (QMA). A promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ is in QMA if there is a polynomial-time uniformly generated family of quantum circuits $\{V_x\}$ (one for each input x , with $n := |x|$) such that:

- (completeness) if $x \in A_{\text{yes}}$, there exists $|\psi\rangle \in (\mathbb{C}^2)^{\otimes \text{poly}(n)}$ such that $\Pr[V_x(|\psi\rangle) = 1] \geq 2/3$;
- (soundness) if $x \in A_{\text{no}}$, then for all $|\psi\rangle \in (\mathbb{C}^2)^{\otimes \text{poly}(n)}$ we have $\Pr[V_x(|\psi\rangle) = 1] \leq 1/3$.

Definition 1.3 (Acceptance operator). Let V_x be a quantum verifier acting on a witness register $\mathcal{H}_w$ and an ancilla register $\mathcal{H}_a$, initialized in $|0^m\rangle$. Let Π_{acc} denote the projector onto states for which the output qubit is 1. The *acceptance operator* associated with V_x is the operator on the witness register

$$A_{\text{acc}}^{(x)} := (I_w \otimes \langle 0^m |_a) V_x^\dagger \Pi_{\text{acc}} V_x (I_w \otimes |0^m\rangle_a).$$

For every witness $|\psi\rangle$,

$$\Pr[V_x(|\psi\rangle) = 1] = \langle \psi | A_{\text{acc}}^{(x)} | \psi \rangle.$$

Since

$$0 \preceq A_{\text{acc}}^{(x)} \preceq I,$$

the maximum acceptance probability is

$$\max_{|\psi\rangle} \Pr[V_x(|\psi\rangle) = 1] = \lambda_{\max}(A_{\text{acc}}^{(x)}).$$

Remark 1.4 (Amplification). The constants 2/3 and 1/3 are conventional. The completeness and soundness errors can be reduced exponentially: for any $\epsilon = 2^{-\text{poly}(n)}$ one can obtain completeness at least $1 - \epsilon$ and soundness at most ϵ . Unlike the classical case, one must be careful with repetition, because an unknown quantum witness cannot simply be copied. Nevertheless, QMA admits efficient error amplification.

Exercise 1.5 (In class). Recall that MA (Merlin–Arthur) is the randomized analogue of NP: Merlin sends a classical witness, which Arthur verifies with a probabilistic polynomial-time algorithm. In MA, error can be reduced by independently repeating the verification and taking a majority vote. Why is the corresponding argument less immediate for QMA? What obstruction prevents Arthur from simply making many copies of the witness?

Proposition 1.6. $\text{NP} \subseteq \text{QMA}$.

Exercise 1.7 (In class). Why?

Remark 1.8 (Why promise problems?). Unlike NP, QMA is most naturally formulated in terms of promise problems. The verifier is required to behave correctly only on inputs in $A_{\text{yes}} \cup A_{\text{no}}$; no condition is imposed outside this set. This will matter for the local Hamiltonian problem, where we are promised that the ground energy lies either below a or above b , with a gap between the two cases.

2 The k -local Hamiltonian problem

2.1 Problem statement

Definition 2.1 (k -local Hamiltonian problem). Fix a constant k . An instance consists of an n -qubit Hamiltonian

$$H = \sum_{j=1}^m h_j, \quad \|h_j\|_\infty \leq 1,$$

where $m = \text{poly}(n)$ and each h_j is Hermitian and acts nontrivially on at most k qubits, together with two real numbers $a < b$ satisfying

$$b - a \geq \frac{1}{\text{poly}(n)}.$$

We are promised that exactly one of the following holds:

- **YES:** the ground-state energy satisfies $\lambda_{\min}(H) \leq a$;
- **NO:** the ground-state energy satisfies $\lambda_{\min}(H) \geq b$.

The problem is to decide whether the instance is a YES or a NO instance.

Here k -local means that each term h_j involves at most k of the n qubits, although the full Hamiltonian H acts on the entire system. We assume that the local terms and the numbers a, b are specified using polynomially many bits, so that the instance has polynomial size.

Theorem 2.2 (Kitaev; Kempe–Kitaev–Regev). *The k -local Hamiltonian problem is QMA-complete for every constant $k \geq 2$.*

Kitaev proved this for $k \geq 5$ [3]; the improvement to $k \geq 2$ is due to Kempe, Kitaev, and Regev [4]. In this lecture we show membership in QMA and sketch the hardness proof; see [1] for a broader survey.

2.2 Containment in QMA

Recall that the energy of a state $|\psi\rangle$ with respect to H is

$$E(\psi) = \langle \psi | H | \psi \rangle.$$

To prove containment, it suffices to give a verifier that accepts with probability at least $2/3$ when given a state of energy at most a , and accepts with probability at most $1/3$ on every state of energy at least b .

We make a simplification: expand each h_j in the Pauli basis. Since a k -local term has at most 4^k Pauli components, this costs only a factor 4^k in the number of terms when k is a constant (or even $O(\log n)$), and each coefficient $c_P = \text{tr}(Ph_j)/2^k$ has $|c_P| \leq \|h_j\| \leq 1$. We may therefore assume

$$H = \sum_{j=1}^m c_j P_j, \quad P_j \in \{\mathbb{I}, X, Y, Z\}^{\otimes k}, \quad c_j \in [-1, 1], \quad m = \text{poly}(n).$$

Exercise 2.3. Let $P \in \{\mathbb{I}, X, Y, Z\}^{\otimes n}$ and let $\text{supp}(P)$ denote the set of qubits on which P acts as X , Y , or Z (e.g. $\text{supp}(X_1 Y_2 \mathbb{I}_3) = \{1, 2\}$). Show that there is a depth-one circuit U such that $U^\dagger Z^{\otimes \text{supp}(P)} U = P$.

Exercise 2.4. For a bit string x , let $h(x)$ be its Hamming weight, and for $\Omega \subseteq [n]$ let x_Ω be the restriction of x to the coordinates in Ω . Argue that

$$Z^{\otimes \Omega} = \sum_x (-1)^{h(x_\Omega)} |x\rangle\langle x|.$$

From these exercises we obtain the following procedure for estimating $\langle \psi | H | \psi \rangle$ from N copies of $|\psi\rangle$:

1. Set $q = 0$.

2. Pick $j \in [m]$ uniformly at random.
3. Apply U_j , where $U_j^\dagger Z^{\otimes \text{supp}(P_j)} U_j = P_j$, to a fresh copy of $|\psi\rangle$ and measure in the computational basis, obtaining a bit string x .
4. If the Hamming weight of $x_{\text{supp}(P_j)}$ is even, add mc_j to q ; if it is odd, add $-mc_j$ to q .
5. Repeat steps 2–4 a total of N times (with fresh randomness and a fresh copy each time), then return q/N .

Correctness. Let j be the sampled index and x the measurement outcome. Each round contributes the random variable

$$Y_{j,x} := (-1)^{h(x_{\text{supp}(P_j)})} mc_j,$$

whose expectation is

$$\mathbb{E}[Y] = \sum_{j=1}^m \sum_x \Pr(j) \Pr(x | j) (-1)^{h(x_{\text{supp}(P_j)})} mc_j,$$

with

$$\Pr(j) = \frac{1}{m}, \quad \Pr(x | j) = \langle \psi | U_j^\dagger |x\rangle \langle x | U_j | \psi \rangle.$$

By the two exercises above,

$$\sum_x (-1)^{h(x_{\text{supp}(P_j)})} \Pr(x | j) = \langle \psi | U_j^\dagger Z^{\otimes \text{supp}(P_j)} U_j | \psi \rangle = \langle \psi | P_j | \psi \rangle,$$

and therefore

$$\mathbb{E}[Y] = \sum_{j=1}^m c_j \langle \psi | P_j | \psi \rangle = \langle \psi | H | \psi \rangle.$$

The above shows that the procedure will converge to the expectation value of H . In order to show completeness and soundness we must estimate the expectation value to enough precision to reliably distinguish between YES and NO instances. Analyzing this will be the content of the assigned exercise at the end of the document.

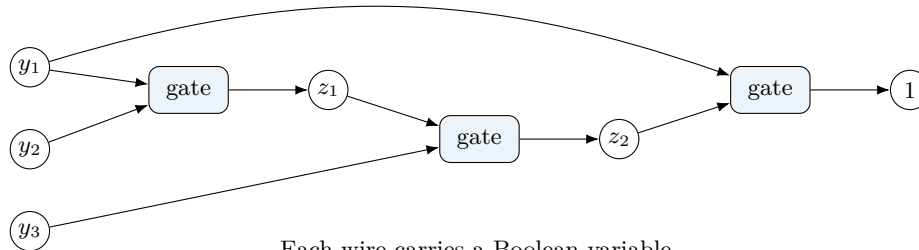
2.3 QMA-hardness: the quantum Cook–Levin theorem (proof sketch)

Proof sketch of the classical Cook–Levin theorem. The Cook–Levin theorem states that SAT is NP-complete. The key idea is to encode the execution of a polynomial-time verifier as a polynomial-size Boolean formula. We give the circuit version of the argument, which motivates the quantum construction below.

Suppose $L \in \text{NP}$. By definition, there is a polynomial-size Boolean circuit C_x such that

$$x \in L \iff \exists y \in \{0, 1\}^{\text{poly}(|x|)} \text{ such that } C_x(y) = 1.$$

Here x is fixed and y is the witness. We convert the computation of C_x into a Boolean formula.



Each wire carries a Boolean variable,
while each gate gives a local consistency constraint.

Introduce one Boolean variable for the value carried by every wire of the circuit. For each gate g , introduce a constant-size formula ϕ_g that is satisfied exactly when the input and output variables of g are consistent with its action. For example, an AND gate with inputs a, b and output c imposes $c = a \wedge b$, which can be written in CNF as

$$(\neg a \vee \neg b \vee c) \wedge (a \vee \neg c) \wedge (b \vee \neg c).$$

The complete formula imposes all gate constraints together with the condition that the output wire equals one:

$$\Phi_x = \left(\bigwedge_{g \in C_x} \phi_g \right) \wedge (z_{\text{out}} = 1).$$

This construction has two key properties.

1. If $x \in L$, there is a witness y for which the circuit accepts. Assigning every internal wire its value during this computation satisfies all constraints, so Φ_x is satisfiable.
2. Conversely, if Φ_x is satisfiable, the gate constraints force the wire values to describe a valid execution of C_x . Since $z_{\text{out}} = 1$, the corresponding witness y is accepted, and therefore $x \in L$.

Thus $x \in L \iff \Phi_x$ is satisfiable. The circuit has polynomially many gates and each gate contributes a constant number of clauses, so Φ_x has polynomial size. This is a polynomial-time reduction from every problem in NP to SAT.

If we think of each constraint as a function that outputs either 0 or 1 (with 1 for true), through the same argument we can define a local Hamiltonian whose ground state is zero if and only if the circuit has an accepting witness via

$$H = z_{\text{out}} + \sum_{g \in C_x} H_g$$

where $H_g = 1 - \phi_g$.

History states and the Feynman–Kitaev Hamiltonian. We now give the quantum analogue. Let the verifier be a circuit $V_x = V_L \cdots V_1$ with L gates, acting on a witness register $\mathcal{H}_w$ of n_w qubits and an ancilla register $\mathcal{H}_a$ of n_a qubits. We add a *clock register* $\mathcal{H}_c$ with orthonormal basis $\{|0\rangle, \dots, |L\rangle\}$, so that the full Hilbert space is $\mathcal{H}_w \otimes \mathcal{H}_a \otimes \mathcal{H}_c$. The role played by the computation history in the classical proof is taken by the *history state*: for a witness $|\xi\rangle$,

$$|\eta_\xi\rangle := \frac{1}{\sqrt{L+1}} \sum_{t=0}^L \left(V_t \cdots V_1 (|\xi\rangle_w |0^{n_a}\rangle_a) \right) \otimes |t\rangle_c.$$

The *Feynman–Kitaev Hamiltonian* is

$$H = H_{\text{prop}} + H_{\text{init}} + H_{\text{out}}.$$

- H_{init} penalizes improperly initialized states in which the ancillas are not all 0:

$$H_{\text{init}} = \mathbb{I}_w \otimes (\mathbb{I}_a - |0^{n_a}\rangle\langle 0^{n_a}|) \otimes |0\rangle\langle 0|_c.$$

- H_{out} penalizes computations that do not end in the accept state at time L . We assume the first witness qubit is the read-out qubit, with $|1\rangle$ meaning “accept”:

$$H_{\text{out}} = |0\rangle\langle 0|_1 \otimes \mathbb{I} \otimes |L\rangle\langle L|_c.$$

- H_{prop} , the most interesting term, penalizes incorrect propagation of the circuit:

$$H_{\text{prop}} = \frac{1}{2} \sum_{t=0}^{L-1} \left(\mathbb{I} \otimes |t\rangle\langle t|_c + \mathbb{I} \otimes |t+1\rangle\langle t+1|_c - V_{t+1} \otimes |t+1\rangle\langle t|_c - V_{t+1}^\dagger \otimes |t\rangle\langle t+1|_c \right).$$

Exercise 2.5. Show that $(H_{\text{prop}} + H_{\text{init}}) |\eta_\xi\rangle = 0$ for every $|\xi\rangle$.

Exercise 2.6. The Hamiltonian H_{init} is not *local*, since its ancilla projector acts on all ancillas at once. Find a replacement that is positive semidefinite, annihilates $|\eta_\xi\rangle$, assigns an energy penalty of at least one to every initial ancilla string other than 0^n , and consists of terms each involving only a single ancilla qubit (together with the clock register).

Exercise 2.7. Argue that if all gates of V_x are 2-local and the clock is encoded in *unary*, then H_{prop} is 5-local. (Proving soundness with a unary clock requires an extra term, so below we proceed as if we had a binary clock.)

Exercise 2.8. Define $W = \sum_{t=0}^L V_t \cdots V_1 \otimes |t\rangle\langle t|_c$, and

$$H_{\text{prop}}^0 = \frac{1}{2} \sum_{t=0}^{L-1} \left(\mathbb{I} \otimes |t\rangle\langle t|_c + \mathbb{I} \otimes |t+1\rangle\langle t+1|_c - \mathbb{I} \otimes |t+1\rangle\langle t|_c - \mathbb{I} \otimes |t\rangle\langle t+1|_c \right).$$

Show that $W^\dagger H_{\text{prop}} W = H_{\text{prop}}^0$.

Completeness. Suppose some $|\xi\rangle$ makes the circuit accept with probability at least $1 - \epsilon$. Only the $t = L$ component of the history state is penalized by H_{out} , and it carries weight $1/(L+1)$, so

$$\langle \eta_\xi | H_{\text{out}} | \eta_\xi \rangle \leq \frac{\epsilon}{L+1} \quad \text{and thus} \quad \langle \eta_\xi | H | \eta_\xi \rangle \leq \frac{\epsilon}{L+1}.$$

Exercise 2.9. Prove both inequalities.

We therefore set $a = \epsilon/(L+1)$.

Soundness. Now suppose every state $|\xi\rangle$ makes V_x accept with probability at most ϵ . This direction is much more involved: the terms of H do not commute, so we cannot simply analyze the spectrum of each term separately. The proof rests on the following lemma, whose proof is beyond the scope of this class (see [3, 4]).

Lemma 2.10 (Kitaev). *There is a constant $c > 0$ such that, if V_x accepts every state with probability at most ϵ , then*

$$\lambda_{\min}(H) \geq \frac{c(1 - \sqrt{\epsilon})}{(L+1)^3}.$$

Setting $b = c(1 - \sqrt{\epsilon})/L^3$ and using amplification so that ϵ is exponentially small, we get $b - a \geq c/(2L^3) \geq 1/\text{poly}(n)$. This completes the sketch of the reduction from any QMA problem to the local Hamiltonian problem. (Combined with the previous subsection, this gives Theorem 2.2 for $k = 5$.)

3 Assigned Exercise

Exercise 3.1. We will complete the argument in Section 2.2.

1. Let $Y = \frac{1}{N} \sum_{i=1}^N X_i$ where X_i are i.i.d. random variables (all identical to X). Show that

$$\Pr(|Y - \mathbb{E}[X]| > \epsilon) \leq \frac{\text{Var}[X]}{N\epsilon^2}.$$

2. Use the above to show that in order for the procedure defined in Section 2.2 to meet the QMA completeness and soundness requirements, it suffices to set N to take

$$N \geq \frac{48m^2}{(b-a)^2}.$$

References

- [1] S. Gharibian, *Approximation, Proof Systems, and Correlations in a Quantum World*, arXiv:1301.2632 (2013).
- [2] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, *Adiabatic quantum computation is equivalent to standard quantum computation*, SIAM J. Comput. **37**(1), 166–194 (2007); arXiv:quant-ph/0405098.
- [3] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*, Graduate Studies in Mathematics **47**, American Mathematical Society (2002).
- [4] J. Kempe, A. Kitaev, and O. Regev, *The complexity of the local Hamiltonian problem*, SIAM J. Comput. **35**(5), 1070–1097 (2006).