

NP-Completeness of 3-Color and SAT

Lecture 24
April 23, 2015

Recap

NP: languages that have non-deterministic polynomial time algorithms

Recap

NP: languages that have non-deterministic polynomial time algorithms

A language **L** is **NP-Complete** iff

- **L** is in **NP**
- for every **L'** in **NP**, $L' \leq_P L$

Recap

NP: languages that have non-deterministic polynomial time algorithms

A language **L** is **NP-Complete** iff

- **L** is in **NP**
- for every **L'** in **NP**, $L' \leq_P L$

L is **NP-Hard** if for every **L'** in **NP**, $L' \leq_P L$.

Recap

NP: languages that have non-deterministic polynomial time algorithms

A language **L** is **NP-Complete** iff

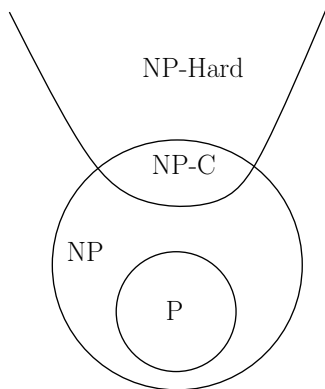
- **L** is in **NP**
- for every **L'** in **NP**, $L' \leq_P L$

L is **NP-Hard** if for every **L'** in **NP**, $L' \leq_P L$.

Theorem (Cook-Levin)

SAT is **NP-Complete**.

Pictorial View



P and NP

Possible scenarios:

① $P = NP$.

② $P \neq NP$

P and NP

Possible scenarios:

- 1 $P = NP$.
- 2 $P \neq NP$

Question: Suppose $P \neq NP$. Is every problem in $NP \setminus P$ also **NP-Complete**?

P and NP

Possible scenarios:

- 1 $P = NP$.
- 2 $P \neq NP$

Question: Suppose $P \neq NP$. Is every problem in $NP \setminus P$ also **NP-Complete**?

Theorem (Ladner)

*If $P \neq NP$ then there is a problem/language $X \in NP \setminus P$ such that X is not **NP-Complete**.*

Part I

NP-Completeness of Graph Coloring

Graph Coloring

Problem: Graph Coloring

Instance: $G = (V, E)$: Undirected graph, integer k .

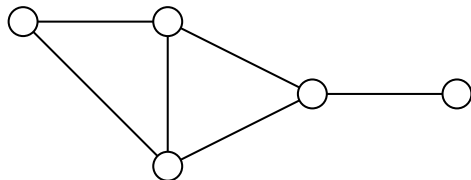
Question: Can the vertices of the graph be colored using k colors so that vertices connected by an edge do not get the same color?

Graph 3-Coloring

Problem: 3 Coloring

Instance: $G = (V, E)$: Undirected graph.

Question: Can the vertices of the graph be colored using 3 colors so that vertices connected by an edge do not get the same color?

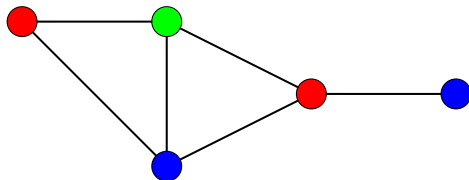


Graph 3-Coloring

Problem: 3 Coloring

Instance: $G = (V, E)$: Undirected graph.

Question: Can the vertices of the graph be colored using 3 colors so that vertices connected by an edge do not get the same color?



Graph Coloring

Observation: If G is colored with k colors then each color class (nodes of same color) form an independent set in G . Thus, G can be partitioned into k independent sets iff G is k -colorable.

Graph Coloring

Observation: If G is colored with k colors then each color class (nodes of same color) form an independent set in G . Thus, G can be partitioned into k independent sets iff G is k -colorable.

Graph 2-Coloring can be decided in polynomial time.

Graph Coloring

Observation: If G is colored with k colors then each color class (nodes of same color) form an independent set in G . Thus, G can be partitioned into k independent sets iff G is k -colorable.

Graph **2**-Coloring can be decided in polynomial time.

G is **2**-colorable iff G is bipartite!

Graph Coloring

Observation: If G is colored with k colors then each color class (nodes of same color) form an independent set in G . Thus, G can be partitioned into k independent sets iff G is k -colorable.

Graph 2-Coloring can be decided in polynomial time.

G is 2-colorable iff G is bipartite! There is a linear time algorithm to check if G is bipartite using **BFS**

Graph Coloring and Register Allocation

Register Allocation

Assign variables to (at most) k registers such that variables needed at the same time are not assigned to the same register

Interference Graph

Vertices are variables, and there is an edge between two vertices, if the two variables are “live” at the same time.

Observations

- [Chaitin] Register allocation problem is equivalent to coloring the interference graph with k colors
- Moreover, $3\text{-COLOR} \leq_P k\text{-Register Allocation}$, for any $k \geq 3$

Class Room Scheduling

Given n classes and their meeting times, are k rooms sufficient?

Class Room Scheduling

Given n classes and their meeting times, are k rooms sufficient?

Reduce to Graph k -Coloring problem

Create graph G

- a node v_i for each class i
- an edge between v_i and v_j if classes i and j *conflict*

Class Room Scheduling

Given n classes and their meeting times, are k rooms sufficient?

Reduce to Graph k -Coloring problem

Create graph G

- a node v_i for each class i
- an edge between v_i and v_j if classes i and j *conflict*

Exercise: G is k -colorable iff k rooms are sufficient

Frequency Assignments in Cellular Networks

Cellular telephone systems that use Frequency Division Multiple Access (FDMA) (example: GSM in Europe and Asia and AT&T in USA)

- Breakup a frequency range $[a, b]$ into disjoint *bands* of frequencies $[a_0, b_0], [a_1, b_1], \dots, [a_k, b_k]$
- Each cell phone tower (simplifying) gets one band
- Constraint: nearby towers cannot be assigned same band, otherwise signals will interference

Frequency Assignments in Cellular Networks

Cellular telephone systems that use Frequency Division Multiple Access (FDMA) (example: GSM in Europe and Asia and AT&T in USA)

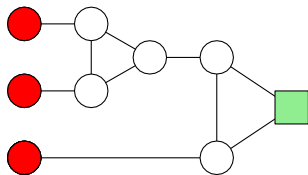
- Breakup a frequency range $[a, b]$ into disjoint *bands* of frequencies $[a_0, b_0], [a_1, b_1], \dots, [a_k, b_k]$
- Each cell phone tower (simplifying) gets one band
- Constraint: nearby towers cannot be assigned same band, otherwise signals will interference

Problem: given k bands and some region with n towers, is there a way to assign the bands to avoid interference?

Can reduce to k -coloring by creating interference/conflict graph on towers.

3 color this gadget.

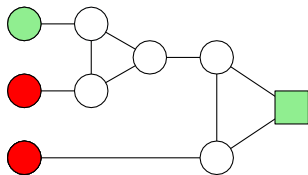
You are given three colors: red, green and blue. Can the following graph be three colored in a valid way (assuming that some of the nodes are already colored as indicated).



- (A) Yes.
- (B) No.

3 color this gadget II

You are given three colors: red, green and blue. Can the following graph be three colored in a valid way (assuming that some of the nodes are already colored as indicated).



- (A) Yes.
- (B) No.

3-Coloring is NP-Complete

- **3-Coloring** is in **NP**.
 - Non-deterministically guess a 3-coloring for each node
 - Check if for each edge (u, v) , the color of u is different from that of v .
- **Hardness:** We will show $3\text{-SAT} \leq_P 3\text{-Coloring}$.

Reduction Idea

Start with **3SAT** formula (i.e., **3CNF** formula) φ with n variables x_1, \dots, x_n and m clauses C_1, \dots, C_m . Create graph G_φ such that G_φ is 3-colorable iff φ is satisfiable

- need to establish truth assignment for x_1, \dots, x_n via colors for some nodes in G_φ .

Reduction Idea

Start with **3SAT** formula (i.e., **3CNF** formula) φ with n variables x_1, \dots, x_n and m clauses C_1, \dots, C_m . Create graph G_φ such that G_φ is 3-colorable iff φ is satisfiable

- need to establish truth assignment for x_1, \dots, x_n via colors for some nodes in G_φ .
- create triangle with node True, False, Base

Reduction Idea

Start with **3SAT** formula (i.e., **3CNF** formula) φ with n variables x_1, \dots, x_n and m clauses C_1, \dots, C_m . Create graph G_φ such that G_φ is 3-colorable iff φ is satisfiable

- need to establish truth assignment for x_1, \dots, x_n via colors for some nodes in G_φ .
- create triangle with node True, False, Base
- for each variable x_i two nodes v_i and \bar{v}_i connected in a triangle with common Base

Reduction Idea

Start with **3SAT** formula (i.e., **3CNF** formula) φ with n variables x_1, \dots, x_n and m clauses C_1, \dots, C_m . Create graph G_φ such that G_φ is 3-colorable iff φ is satisfiable

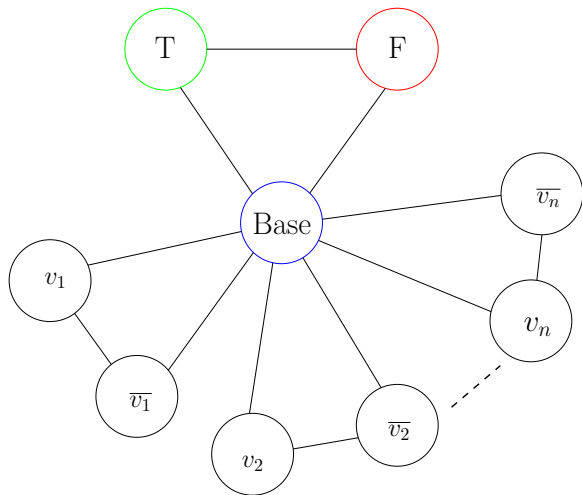
- need to establish truth assignment for x_1, \dots, x_n via colors for some nodes in G_φ .
- create triangle with node True, False, Base
- for each variable x_i two nodes v_i and \bar{v}_i connected in a triangle with common Base
- If graph is 3-colored, either v_i or \bar{v}_i gets the same color as True. Interpret this as a truth assignment to v_i

Reduction Idea

Start with **3SAT** formula (i.e., **3CNF** formula) φ with n variables x_1, \dots, x_n and m clauses C_1, \dots, C_m . Create graph G_φ such that G_φ is 3-colorable iff φ is satisfiable

- need to establish truth assignment for x_1, \dots, x_n via colors for some nodes in G_φ .
- create triangle with node True, False, Base
- for each variable x_i two nodes v_i and \bar{v}_i connected in a triangle with common Base
- If graph is 3-colored, either v_i or \bar{v}_i gets the same color as True. Interpret this as a truth assignment to v_i
- Need to add constraints to ensure clauses are satisfied (next phase)

Figure

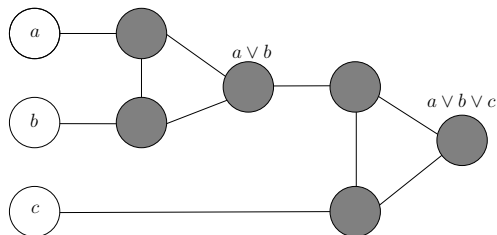


Clause Satisfiability Gadget

For each clause $C_j = (a \vee b \vee c)$, create a small gadget graph

- gadget graph connects to nodes corresponding to a, b, c
- needs to implement OR

OR-gadget-graph:



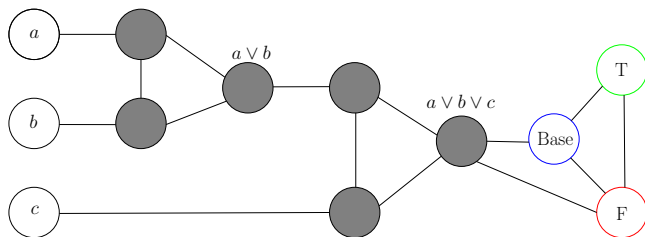
OR-Gadget Graph

Property: if **a**, **b**, **c** are colored False in a 3-coloring then output node of OR-gadget has to be colored False.

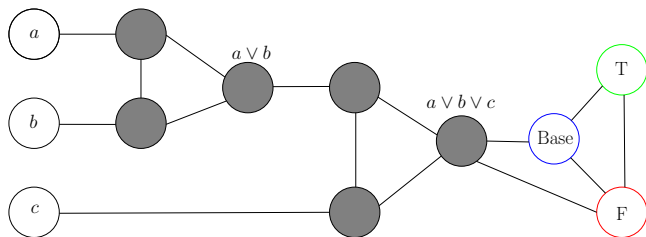
Property: if one of **a**, **b**, **c** is colored True then OR-gadget can be 3-colored such that output node of OR-gadget is colored True.

Reduction

- create triangle with nodes True, False, Base
- for each variable x_i two nodes v_i and \bar{v}_i connected in a triangle with common Base
- for each clause $C_j = (a \vee b \vee c)$, add OR-gadget graph with input nodes a, b, c and connect output node of gadget to both False and Base



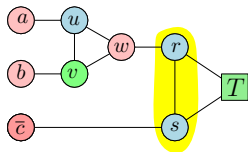
Reduction



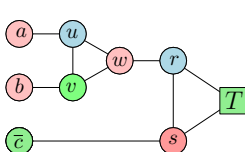
Claim

No legal **3**-coloring of above graph (with coloring of nodes T, F, B fixed) in which a, b, c are colored False. If any of a, b, c are colored True then there is a legal **3**-coloring of above graph.

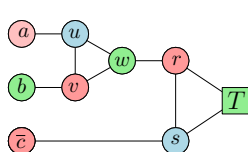
3 coloring of the clause gadget



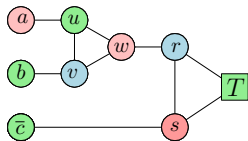
FFF - **BAD**



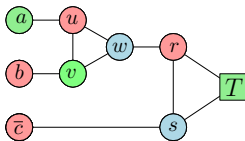
FFT



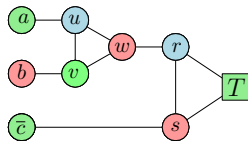
FTF



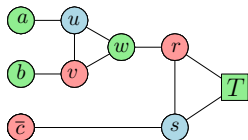
FTT



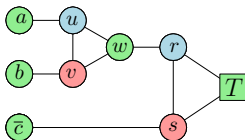
TFF



TFT



TTF

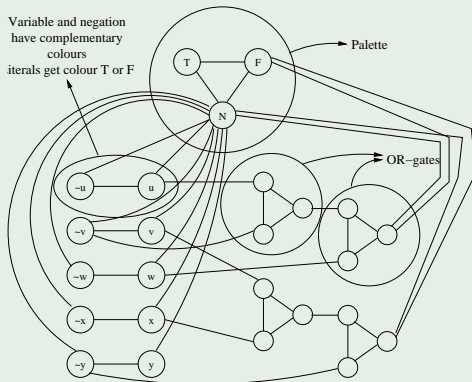


TTT

Reduction Outline

Example

$$\varphi = (u \vee \neg v \vee w) \wedge (v \vee x \vee \neg y)$$



Correctness of Reduction

φ is satisfiable implies G_φ is 3-colorable

- if x_i is assigned True, color v_i True and \bar{v}_i False

Correctness of Reduction

φ is satisfiable implies \mathbf{G}_φ is 3-colorable

- if x_i is assigned True, color v_i True and \bar{v}_i False
- for each clause $\mathbf{C}_j = (\mathbf{a} \vee \mathbf{b} \vee \mathbf{c})$ at least one of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is colored True. OR-gadget for \mathbf{C}_j can be 3-colored such that output is True.

Correctness of Reduction

φ is satisfiable implies \mathbf{G}_φ is 3-colorable

- if x_i is assigned True, color v_i True and \bar{v}_i False
- for each clause $\mathbf{C}_j = (\mathbf{a} \vee \mathbf{b} \vee \mathbf{c})$ at least one of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is colored True. OR-gadget for \mathbf{C}_j can be 3-colored such that output is True.

Correctness of Reduction

φ is satisfiable implies \mathbf{G}_φ is 3-colorable

- if \mathbf{x}_i is assigned True, color \mathbf{v}_i True and $\bar{\mathbf{v}}_i$ False
- for each clause $\mathbf{C}_j = (\mathbf{a} \vee \mathbf{b} \vee \mathbf{c})$ at least one of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is colored True. OR-gadget for \mathbf{C}_j can be 3-colored such that output is True.

\mathbf{G}_φ is 3-colorable implies φ is satisfiable

- if \mathbf{v}_i is colored True then set \mathbf{x}_i to be True, this is a legal truth assignment

Correctness of Reduction

φ is satisfiable implies \mathbf{G}_φ is 3-colorable

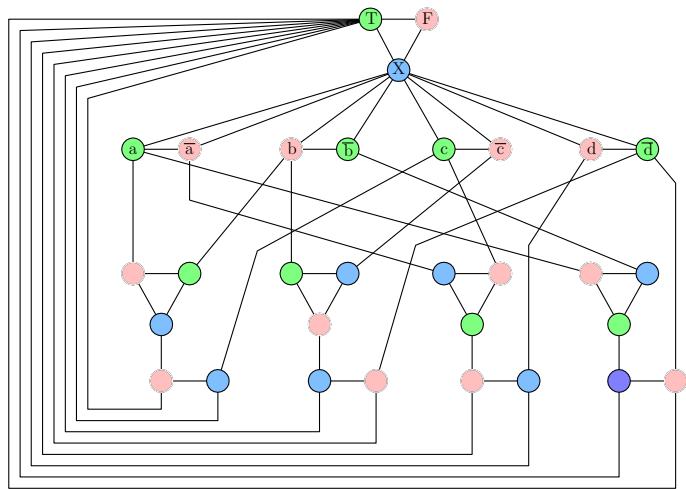
- if \mathbf{x}_i is assigned True, color \mathbf{v}_i True and $\bar{\mathbf{v}}_i$ False
- for each clause $\mathbf{C}_j = (\mathbf{a} \vee \mathbf{b} \vee \mathbf{c})$ at least one of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is colored True. OR-gadget for \mathbf{C}_j can be 3-colored such that output is True.

\mathbf{G}_φ is 3-colorable implies φ is satisfiable

- if \mathbf{v}_i is colored True then set \mathbf{x}_i to be True, this is a legal truth assignment
- consider any clause $\mathbf{C}_j = (\mathbf{a} \vee \mathbf{b} \vee \mathbf{c})$. it cannot be that all $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are False. If so, output of OR-gadget for \mathbf{C}_j has to be colored False but output is connected to Base and False!

Graph generated in reduction...

... from 3SAT to 3COLOR



Part II

Proof of Cook-Levin Theorem

Cook-Levin Theorem

Theorem (Cook-Levin)

SAT is **NP-Complete**.

We have already seen that **SAT** is in **NP**.

Need to prove that every language $L \in \mathbf{NP}$, $L \leq_P \mathbf{SAT}$

Cook-Levin Theorem

Theorem (Cook-Levin)

SAT is **NP-Complete**.

We have already seen that **SAT** is in **NP**.

Need to prove that every language $L \in \mathbf{NP}$, $L \leq_P \mathbf{SAT}$

Difficulty: Infinite number of languages in **NP**. Must *simultaneously* show a *generic* reduction strategy.

High-level Plan

What does it mean that $L \in NP$?

$L \in NP$ implies that there is a non-deterministic TM M and polynomial $p()$ such that

$$L = \{x \in \Sigma^* \mid M \text{ accepts } x \text{ in at most } p(|x|) \text{ steps}\}$$

High-level Plan

What does it mean that $L \in NP$?

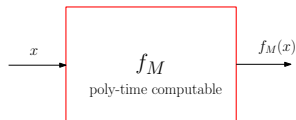
$L \in NP$ implies that there is a non-deterministic TM M and polynomial $p()$ such that

$$L = \{x \in \Sigma^* \mid M \text{ accepts } x \text{ in at most } p(|x|) \text{ steps}\}$$

We will describe a reduction f_M that depends on M, p such that:

- f_M takes as input a string x and outputs a SAT formula $f_M(x)$
- f_M runs in time polynomial in $|x|$
- $x \in L$ if and only if $f_M(x)$ is satisfiable

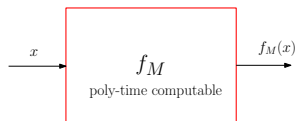
Plan continued



$\mathbf{f}_M(\mathbf{x})$ is satisfiable if and only if $\mathbf{x} \in \mathbf{L}$

$\mathbf{f}_M(\mathbf{x})$ is satisfiable if and only if non-det \mathbf{M} accepts \mathbf{x} in $\mathbf{p}(|\mathbf{x}|)$ steps

Plan continued



$f_M(x)$ is satisfiable if and only if $x \in L$

$f_M(x)$ is satisfiable if and only if non-det M accepts x in $p(|x|)$ steps

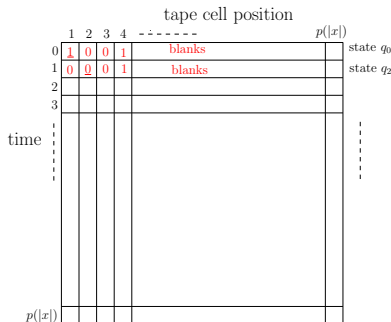
BIG IDEA

- $f_M(x)$ will express “ M on input x accepts in $p(|x|)$ steps”
- $f_M(x)$ will encode a computation history of M on x

$f_M(x)$ will be a carefully constructed CNF formula s.t if we have a satisfying assignment to it, then we will be able to see a complete accepting computation of M on x *down to the last detail* of where the head is, what transition is chosen, what the tape contents are, at each step.

Tableau of Computation

M runs in time $p(|x|)$ on x . Entire computation of **M** on x can be represented by a “tableau”



Row **i** gives contents of all cells at time **i**

At time **0** tape has input x followed by blanks

Each row long enough to hold all cells **M** might ever have scanned.

Variable of $f_M(x)$

Four types of variable to describe computation of M on x

- $T(b, h, i)$: tape cell at position h holds symbol b at time i .
 $1 \leq h \leq p(|x|)$, $b \in \Gamma$, $0 \leq i \leq p(|x|)$
- $H(h, i)$: read/write head is at position h at time i .
 $1 \leq h \leq p(|x|)$, $0 \leq i \leq p(|x|)$
- $S(q, i)$ state of M is q at time i $q \in Q$, $0 \leq i \leq p(|x|)$
- $I(j, i)$ instruction number j is executed at time i
 M is non-deterministic, need to specify transitions in some way.
Number transitions as $1, 2, \dots, \ell$ where j 'th transition is
 $\langle q_j, b_j, q'_j, b'_j, d_j \rangle$ indication $(q'_j, b'_j, d_j) \in \delta(q_j, b_j)$,
direction $d_j \in \{-1, 0, 1\}$.

Number of variables is $O(p(|x|)^2)$ where constant in $O()$ hides dependence on fixed machine M .

Notation

Some abbreviations for ease of notation

$\bigwedge_{k=1}^m x_k$ means $x_1 \wedge x_2 \wedge \dots \wedge x_m$

$\bigvee_{k=1}^m x_k$ means $x_1 \vee x_2 \vee \dots \vee x_m$

$\bigoplus(x_1, x_2, \dots, x_k)$ is a formula that means exactly one of x_1, x_2, \dots, x_m is true. Can be converted to CNF form

Clauses of $f_M(x)$

$f_M(x)$ is the conjunction of **8** clause groups:

$$f_M(x) = \varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4 \wedge \varphi_5 \wedge \varphi_6 \wedge \varphi_7 \wedge \varphi_8$$

where each φ_i is a CNF formula. Described in subsequent slides.

Property: $f_M(x)$ is satisfied iff there is a truth assignment to the variables that simultaneously satisfy $\varphi_1, \dots, \varphi_8$.

φ_1 asserts (is true iff) the variables are set T/F indicating that **M** starts in state \mathbf{q}_0 at time **0** with tape contents containing \mathbf{x} followed by blanks.

Let $\mathbf{x} = \mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_n$

$\varphi_1 = \mathbf{S}(\mathbf{q}, \mathbf{0})$ state at time **0** is \mathbf{q}_0

\bigwedge and

$\bigwedge_{h=1}^n \mathbf{T}(\mathbf{a}_h, \mathbf{h}, \mathbf{0})$ at time **0** cells **1** to **n** have \mathbf{a}_1 to \mathbf{a}_n

$\bigwedge_{h=n+1}^{p(|x|)} \mathbf{T}(\mathbf{B}, \mathbf{h}, \mathbf{0})$ at time **0** cells **n + 1** to $p(|x|)$ have blanks

\bigwedge and

$\mathbf{H}(\mathbf{1}, \mathbf{0})$ head at time **0** is in position **1**

φ_2

φ_2 asserts **M** in exactly one state at any time **i**

$$\varphi_2 = \bigwedge_{i=0}^{p(|x|)} (\oplus(\mathbf{S}(q_0, i), \mathbf{S}(q_1, i), \dots, \mathbf{S}(q_{|Q|}, i)))$$

φ_3 asserts that each tape cell holds a unique symbol at any given time.

$$\varphi_3 = \bigwedge_{i=0}^{p(|x|)} \bigwedge_{h=1}^{p(|x|)} \bigoplus (\mathbf{T}(\mathbf{b}_1, \mathbf{h}, \mathbf{i}), \mathbf{T}(\mathbf{b}_2, \mathbf{h}, \mathbf{i}), \dots, \mathbf{T}(\mathbf{b}_{|\Gamma|}, \mathbf{h}, \mathbf{i}))$$

For each time \mathbf{i} and for each cell position \mathbf{h} exactly one symbol $\mathbf{b} \in \Gamma$ at cell position \mathbf{h} at time \mathbf{i}

φ_4 asserts that the read/write head of **M** is in exactly one position at any time **i**

$$\varphi_4 = \bigwedge_{i=0}^{p(|x|)} (\oplus (\mathbf{H}(1, i), \mathbf{H}(2, i), \dots, \mathbf{H}(p(|x|), i)))$$

φ_5 asserts that **M** accepts

- Let q_a be unique accept state of **M**
- without loss of generality assume **M** runs all $p(|x|)$ steps

$$\varphi_5 = S(q_a, p(|x|))$$

State at time $p(|x|)$ is q_a the accept state.

If we don't want to make assumption of running for all steps

$$\varphi_5 = \bigvee_{i=1}^{p(|x|)} S(q_a, i)$$

which means **M** enters accepts state at some time.

φ_6 asserts that **M** executes a unique instruction at each time

$$\varphi_6 = \bigwedge_{i=0}^{p(|x|)} \oplus (l(1, i), l(2, i), \dots, l(m, i))$$

where **m** is max instruction number.

φ_7 ensures that variables don't allow tape to change from one moment to next if the read/write head was not there.

"If head is **not** at position **h** at time **i** then at time **i + 1** the symbol at cell **h** must be unchanged"

$$\varphi_7 = \bigwedge_i \bigwedge_h \bigwedge_{b \neq c} \left(\overline{\mathbf{H}(h, i)} \Rightarrow \overline{\mathbf{T}(b, h, i) \wedge \mathbf{T}(c, h, i + 1)} \right)$$

since $\mathbf{A} \Rightarrow \mathbf{B}$ is same as $\neg \mathbf{A} \vee \mathbf{B}$, rewrite above in CNF form

$$\varphi_7 = \bigwedge_i \bigwedge_h \bigwedge_{b \neq c} (\mathbf{H}(h, i) \vee \neg \mathbf{T}(b, h, i) \vee \neg \mathbf{T}(c, h, i + 1))$$

φ_8 asserts that changes in tableau/tape correspond to transitions of **M** (as Lenny says, this is the big cookie).

Let **j**'th instruction be $\langle \mathbf{q}_j, \mathbf{b}_j, \mathbf{q}'_j, \mathbf{b}'_j, \mathbf{d}_j \rangle$

$\varphi_8 = \bigwedge_i \bigwedge_j (\mathbf{I}(j, i) \Rightarrow \mathbf{S}(\mathbf{q}_j, i))$ If instr **j** executed at time **i** then state must be correct to do **j**

$\bigwedge_i \bigwedge_j (\mathbf{I}(j, i) \Rightarrow \mathbf{S}(\mathbf{q}'_j, i + 1))$ and at next time unit, state must be the proper next state for instr **j**

$\bigwedge_i \bigwedge_h \bigwedge_j [(\mathbf{I}(j, i) \wedge \mathbf{H}(h, i)) \Rightarrow \mathbf{T}(\mathbf{b}_j, h, i)]$ if **j** was executed and head was at position **h**,

then cell **h** has correct symbol for **j**

$\bigwedge_i \bigwedge_j \bigwedge_h [(\mathbf{I}(j, i) \wedge \mathbf{H}(h, i)) \Rightarrow \mathbf{T}(\mathbf{b}'_j, h, i + 1)]$ if **j** was done then at time **i** with head at **h** then at next time step symbol \mathbf{b}'_j was indeed written in position **h**

$\bigwedge_i \bigwedge_j \bigwedge_h [(\mathbf{I}(j, i) \wedge \mathbf{H}(h, i)) \Rightarrow \mathbf{H}(h + \mathbf{d}_j, i + 1)]$ and head is moved properly according to instr **j**.

Proof of Correctness

(Sketch)

- Given \mathbf{M} , \mathbf{x} , poly-time algorithm to construct $\mathbf{f}_{\mathbf{M}}(\mathbf{x})$
- if $\mathbf{f}_{\mathbf{M}}(\mathbf{x})$ is satisfiable then the truth assignment completely specifies an accepting computation of \mathbf{M} on \mathbf{x}
- if \mathbf{M} accepts \mathbf{x} then the accepting computation leads to an "obvious" truth assignment to $\mathbf{f}_{\mathbf{M}}(\mathbf{x})$. Simply assign the variables according to the state of \mathbf{M} and cells at each time i .

Thus \mathbf{M} accepts \mathbf{x} if and only if $\mathbf{f}_{\mathbf{M}}(\mathbf{x})$ is satisfiable