THE HALTING PROBLEM - PROOF

Solving Problems

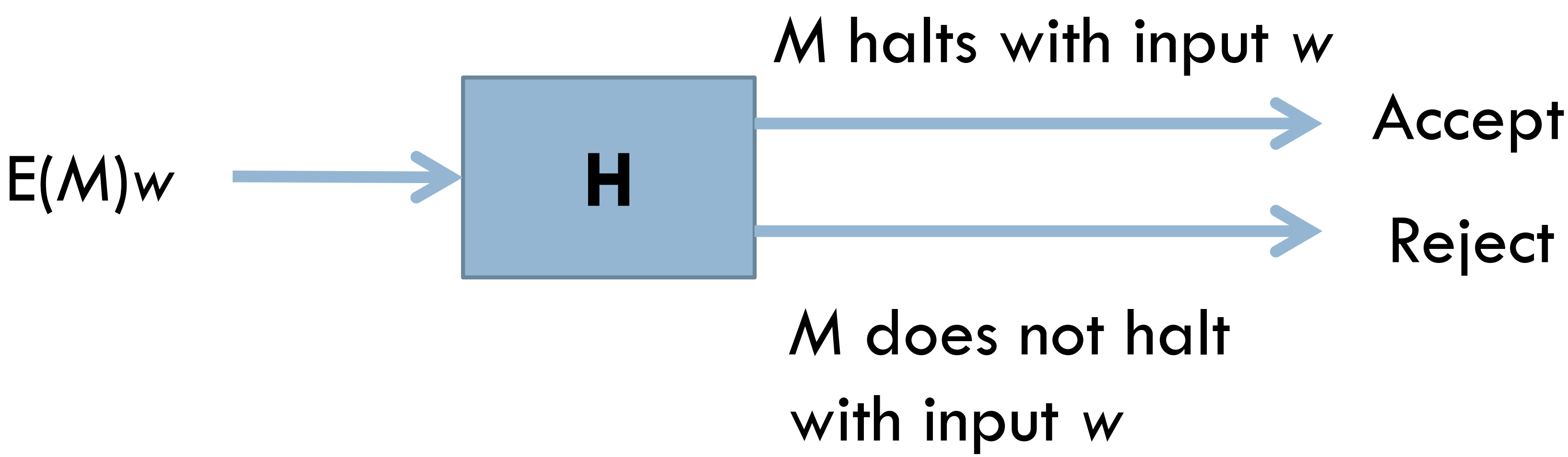
- What makes a problem *decidable*?
  - Is a problem such that there exist a TM that *halts* (accept/reject) on every input
- What makes an unsolvable problem?
  - Is a problem such that there does not exist any TM that can solve the problem
- What is the Halting Problem for TM?
  - Given an arbitrary TM M with input alphabet Σ and a string w over Σ^*, will the computation of M with w halt?

Halting Problem

- Is the *Halting Problem* decidable/undecidable?
  - There is no algorithm that solves the halting problem, thus the halting problem is *undecidable*
  - A solution to the halting problem requires a general algorithm that answers the halting question (i.e., *equivalent to the "acceptance" question of a string for a TM*) for each combination of TM and input string
    - The proof of the fact that the halting problem is undecidable is done by **contradiction** and using the **encoding** of an arbitrary TM and a string as the starting point

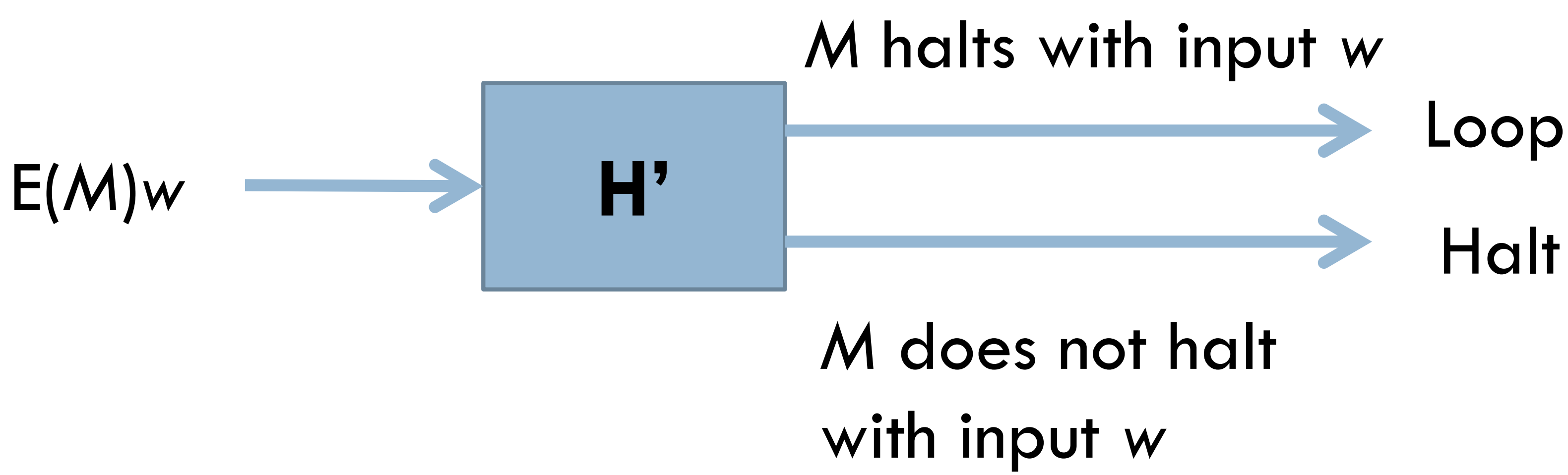
Halting problem is undecidable - Proof

- Proof by contradiction
  - Assume that there is a TM *H* that solves the halting problem
  - The computation of *H* can be depicted as follows:



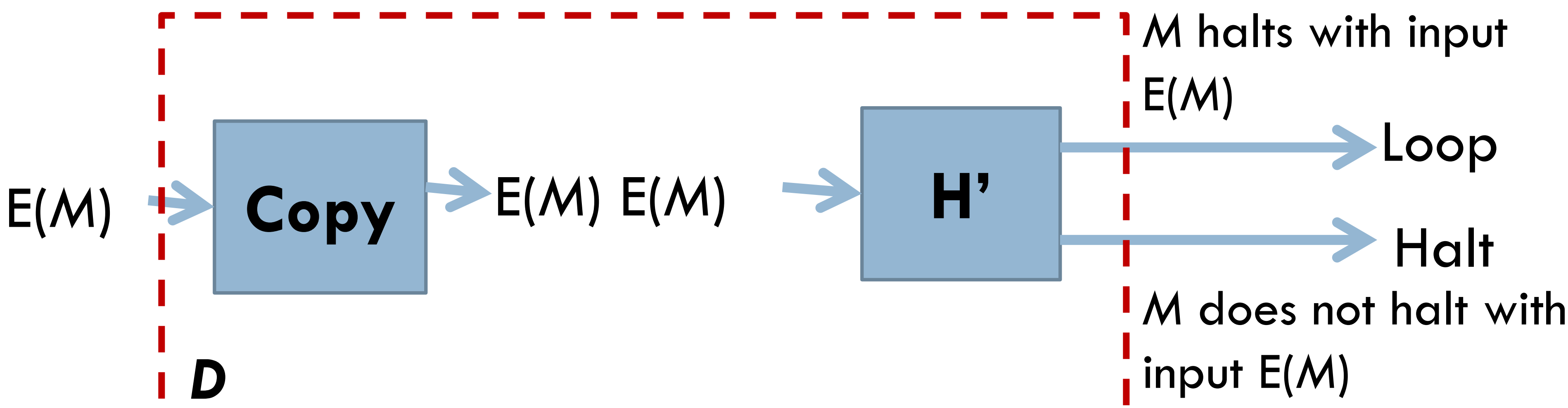
Halting problem is undecidable - Proof

- We modify *H* to construct a new TM *H'*, which behaves very much as *H*:
  - The computations of *H'* are the same as *H*, except that *H'* loops indefinitely whenever *H* terminates in an accepting state, i.e., whenever *H* halts on input *w*, and halts otherwise



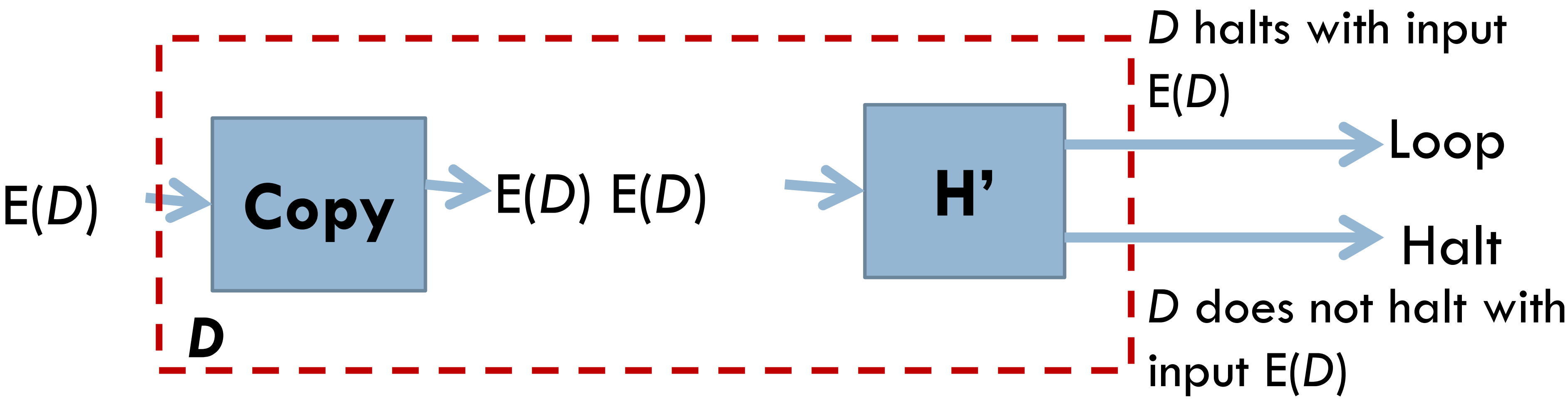
Halting problem is undecidable - Proof

- *H*' is combined with a *copy machine* to construct a new TM *D*:
  - The input to *D* is a TM representation E(*M*)
  - A computation of *D* begins with copying the string E(*M*) to yield E(*M*)E(*M*)
  - The computation continues by running *H*' on E(*M*)E(*M*)



Halting problem is undecidable Proof

- Consider a computation of *D* with input E(*D*):
  - The input to *D* is the representation to any arbitrary TM



- Examining the preceding computation, we see that:
  - *D halt* on input E(*D*) iff *D does not halt* on input E(*D*), which is a **contradiction**
  - Therefore, the original assumption is wrong and the halting problem is **undecidable**