A Generic Time Hierarchy for Semantic Models with One Bit of Advice

Dieter van Melkebeek *
University of Wisconsin - Madison
Department of Computer Sciences
1210 West Dayton Street, Madison, WI 53706-1685, USA
dieter@cs.wisc.edu

Konstantin Pervyshev
St. Petersburg State University
Department of Mathematics and Mechanics
Universitetskiy pr. 28, 198504 St. Petersburg, Russia
pervyshev@logic.pdmi.ras.ru

Abstract

We show that for any reasonable semantic model of computation and for any positive integer a and rationals $1 \leq c < d$, there exists a language computable in time n^d with a bits of advice but not in time n^c with a bits of advice. A semantic model is one for which there exists a computable enumeration that contains all machines in the model but may also contain others. We call such a model reasonable if it has an efficient universal machine that can be complemented within the model in exponential time and if it is efficiently closed under deterministic transducers.

Our result implies the first such hierarchy theorem for randomized machines with zero-sided error, quantum machines with one- or zero-sided error, unambiguous machines, symmetric alternation, Arthur-Merlin games of any signature, etc. Our argument yields considerably simpler proofs of known hierarchy theorems with one bit of advice for randomized and quantum machines with two-sided error.

Our paradigm also allows us to derive stronger separation results in a unified way. For models that have an efficient universal machine that can be simulated deterministically in exponential time and that are efficiently closed under randomized reductions with two-sided error, we establish the following: For any constants a and c, there exists a language computable in polynomial time with one bit of advice but not in time n^c with $a \log n$ bits of advice. The result applies to randomized and quantum machines with two-sided error. For randomized machines with one-sided error, our approach yields that for any constants a and c there exists a language computable in polynomial time with one bit of advice but not in time n^c with $a(\log n)^{1/c}$ bits of advice.

1. Introduction

Hierarchy theorems address one of the most fundamental questions in computational complexity: Can we decide more languages on a certain model of computation when given a bit more of a certain resource? In fact, a time hierarchy for deterministic Turing machines constitutes the main technical contribution in the paper by Hartmanis and Stearns [9] that founded the field. Later on, Cook [2], Seiferas, Fischer and Meyer [15], and Žàk[16] established time hierarchies for nondeterministic Turing machines. Their techniques apply to virtually any *syntactic* model of interest, i.e., one for which there exists a computable enumeration of exactly the machines in the model.

Several models we care about are not syntactic, though. Examples include randomized or quantum machines with two-, one-, or zero-sided error, unambiguous machines, symmetric alternation, Arthur-Merlin games of any signature, etc. Each of these models has a computable enumeration that contains all machines of the model but may also contain other machines. For example, we can computably enumerate all randomized machines; the enumeration contains all randomized machines with two-sided error but also contains machines that violate the promise of bounded er-

*Partially supported by NSF Career award CCR-0133693, EIA-0205236, and EMT-0523680.

ror. We dub models with such an enumeration as *semantic*. See Section 4.1 for more on nomenclature.

To date, except for a few cases in which a non-syntactic model is known to be equivalent in power to a syntactic one, no hierarchy is known for any non-syntactic model[1]. In particular, it remains open whether for every constant c there exists a language that can be solved on randomized machines with two-sided error in polynomial time but not in time n^c.

In 2002, Barak [1] used instance checkers for exponential-time complete languages to prove the latter statement in a slightly nonuniform version of the model, namely a model in which the machines get $a(n)$ bits of advice for some function $a(n) = O(\log\log n)$. In other words, he established the result for randomized machines with two-sided error whose descriptions can depend on the input length n in such a way that the size of the variable part is bounded by $a(n)$. Subsequently, several authors tried to get as close as possible to the desired uniform result and managed to reduce the amount of advice to a single bit [3, 7]. Barak's argument also applies to quantum machines with two-sided error but not to any of the other semantic models on our list. Roughly speaking, due to the use of instance checkers, the model has to be closed in an efficient way under randomized reductions with two-sided error for the proof to carry through.

More recently, Fortnow, Santhanam, and Trevisan [4] gave a specific argument for randomized machines with one-sided error and one bit of advice. They also developed an approach that works for all of the above models but needs considerably more advice: They obtain a hierarchy theorem for any reasonable semantic model of computation with $a(n)$ bits of advice where $a(n)$ is some function in $O(\log n \cdot \log\log n)$.

As our main result, we manage to get the best of both worlds and thereby improve both lines of research.

Theorem 1 *For any reasonable semantic model of computation and any constants a and c, there exists a language computable in polynomial time with one bit of advice but not in time n^c with a bits of advice.*

As a corollary to Theorem 1, we obtain the following hierarchy with a bits of advice for any constant $a \geq 1$.

Theorem 2 *For any reasonable semantic model of computation and any positive integer a and rationals $1 \leq c < d$, there exists a language computable in time n^d with a bits of advice but not in time n^c with a bits of advice.*

We refer to Section 4.1 for a precise definition of "reasonable" but all of the specific models listed above fall under the notion.

We use the technique of delayed diagonalization adapted to the setting of computations with advice. Our approach differs from Barak's as well as the one by Fortnow et al. Like the latter but unlike the former, our proof relativizes. Since instance checkers are the sole culprit of nonrelativization in Barak's argument, our proof shows that that component is not critical for obtaining a time hierarchy for randomized machines with two-sided error and one bit of advice. Apart from yielding stronger results and being more widely applicable, our approach also provides considerably simpler proofs for all the hierarchy theorems with one bit of advice that were known before [3, 7, 4]. We refer to Section 2 for a more detailed comparison of techniques.

As is clear from the statement of Theorem 1, the proof of our main result actually yields more than a hierarchy theorem because we can accommodate up to a bits of advice for any constant a at the smaller time bound while still only needing a single bit of advice at the larger time bound. Barak's argument goes further along that road and handles up to $a \log n$ instead of a bits of advice but only for a more restrictive subclass of semantic models. We show how to match Barak's bound of $a \log n$ using our approach.

Theorem 3 *For any reasonable randomized semantic model of computation that is efficiently closed under randomized reductions with two-sided error, and any constants a and c, there exists a language computable in polynomial time with one bit of advice but not in time n^c with $a \log n$ bits of advice.*

We refer to Section 4.3 for a full specification of the models to which Theorem 3 applies; the list includes randomized and quantum machines with two-sided error. Our proof of Theorem 3 uses instance checkers again but in a different way than Barak and for a more limited purpose. Thus, we further relegate the use of instance checkers in this context.

Theorem 3 does not seem to apply to randomized machines with one-sided error. For that specific model, Fortnow et al.'s argument[2] yields a somewhat weaker separation theorem, namely for $a(\log n)^{1/c}$ bits of advice instead of $a \log n$ bits at the smaller time bound of n^c. We show how to obtain that result using our approach, too.

Theorem 4 *For any constants a and c there exists a language computable by randomized machines with one-sided error in polynomial time with one bit of advice but not in time n^c with $a(\log n)^{1/c}$ bits of advice.*

[1]Here, we are assuming that one interprets "a bit more time" as implying "at most a polynomial amount more time." The exceptions we are aware of follow from the characterizations IP = PSPACE, MIP = NEXP, and PCP($\log n$, 1) = NP, BP$\cdot\oplus$P = $\Sigma_2^{\oplus \mathrm{P}}$.

[2]Fortnow et al. [4] actually only prove the result for $(\log n)^{1/2c}$ bits of advice but a small modification of their argument works up to $a(\log n)^{1/c}$ bits of advice at the smaller time bound of n^c.

A modification of the proof of Theorem 4 also allows us to establish Theorem 3 for the specific model of randomized machines with two-sided error in a *relativizable* way, i.e., without the use of instance checkers. Thus, the paradigm we present offers a unified way for deriving new as well as known separation results within non-syntactic models of computation.

The rest of this paper is organized as follows. In Section 2, we provide an overview of the arguments that have been used for deriving hierarchy theorems in the past. Section 3 describes the intuition behind our constructions and develops them in an informal way. Section 4 contains the formal presentation of our generic hierarchy theorem as well as the precise statements of our separation theorems. Due to space limitations, except for our main result (Theorem 1), we defer the detailed proofs to the journal version of our paper. Finally, in Section 5, we present some possible directions for further research.

2. Previous work

In this section, we survey the arguments that have been used in hierarchy theorems and that exhibit a close relationship to ours. We focus on techniques and qualitative improvements rather than quantitative ones. Readers who would like to skip to Section 3 for a description of our constructions can do so without loss of continuity.

For their seminal hierarchy theorem, Hartmanis and Stearns [9] used a diagonalization technique rooted in Cantor's proof that the reals are not countable. They assume the model of computation has a computable enumeration of machines and a universal machine U. They pick an infinite sequence of inputs $x_1, x_2, \ldots$, and use x_i to diagonalize against the ith machine M_i of the enumeration by running the universal machine on $\langle M_i, x_i, 0^t\rangle$, where t denotes the allotted amount of time, and doing the opposite. This approach results in a time hierarchy for essentially any syntactic model with an efficient universal machine for which "doing the opposite" is easy.

We don't know whether "doing the opposite" is easy in models like nondeterministic machines. We can run a deterministic simulation and complement the result but that involves an exponential slowdown. Cook [2] was the first to get around the need for easy complementation. His proof works by contradiction and goes as follows.

Assume the hierarchy theorem for nondeterministic machines fails. Then for every polynomial-time nondeterministic machine there exists an equivalent nondeterministic machine that runs in time n^c for some fixed c. Applying this speedup $O(\log n)$ times in a uniform way (exploiting the existence of a universal machine) shows that even every exponential-time nondeterministic machine has an equivalent nondeterministic machine that runs in time n^c. We can simulate the latter nondeterministic machine on a deterministic one in time 2^{n^c}. On the other hand, deterministic machines are also nondeterministic machines. Thus, we obtain a simulation of every exponential-time deterministic machine by another deterministic machine that runs in time 2^{n^c} – a contradiction with the time hierarchy for deterministic machines.

Seiferas et al. [15] use a more direct argument and explicitly construct a language L that witnesses the nondeterministic time hierarchy for a given constant c. They start from any computable language L' that cannot be decided by nondeterministic machines in time n^{c+1}, e.g., a complete language for double exponential time. They define L as the language accepted by the nondeterministic machine M that acts as follows on strings of the form $\langle x, i, 0^k\rangle$. Let M' denote a fixed deterministic machine that decides L'. If k is larger than the running time of M' on x, then M outputs the result of that computation. Otherwise, M uses the universal machine to simulate M_i on input $\langle x, i, 0^{k+1}\rangle$ for n^c steps. M runs in polynomial time but the language L it defines cannot be accepted by nondeterministic machines that run in time n^c. Indeed, suppose that M_i were such a machine. For small k, we have that $M_i(\langle x, i, 0^k\rangle) = M(\langle x, i, 0^k\rangle) = M_i(\langle x, i, 0^{k+1}\rangle)$, and for large k that $M_i(\langle x, i, 0^k\rangle) = M(\langle x, i, 0^k\rangle) = M'(x)$. It follows that $M'(x) = M_i(\langle x, i, \epsilon\rangle)$ for each x. Since M_i runs in time n^c, this contradicts the fact that the language L' decided by M' cannot be accepted by a nondeterministic machine in time n^{c+1}.

Žàk's argument [16] is similar but replaces the use of a difficult language L' by delayed diagonalization. Essentially, on inputs of the form $\langle x, i, 0^k\rangle$, the role of M' is taken over by the complement of the deterministic simulation of M_i for n^c steps. The rest of the argument is analogous: Suppose that M_i runs in n^c steps and is equivalent to M, and let k be the first large value (for the given i). We have on the one hand that $M_i(\langle x, i, 0^k\rangle) = M_i(\langle x, i, 0^{k-1}\rangle) = \ldots = M_i(\langle x, i, \epsilon\rangle)$ and on the other hand that $M_i(\langle x, i, 0^k\rangle) = \neg M_i(\langle x, i, \epsilon\rangle)$. Thus, M_i is not equivalent to M or takes more than n^c steps.

As a side note, we point out that it suffices for the machine M in Žàk's construction to act as described on *some* input x, say $x = \epsilon$, whereas Seiferas et al. in principle need the behavior on *every* x. Thus, Žàk's argument naturally leads to a *unary* language L that can be accepted by nondeterministic machines in polynomial time but not in time n^c.

The constructions by Cook, Seiferas et al., and Žàk work for any syntactic model that has an efficient universal machine and is efficiently closed under deterministic transducers. For Cook's argument, we also need the existence of deterministic simulations that incur a non-exorbitant slowdown; exponential overhead as in the case of nondeterministic machines is fine. This essentially corresponds to what we mean by a "reasonable" syntactic model of computation. See Section 4.1 for the formal definitions.

Unfortunately, none of these techniques seem to extend to *semantic* models because they all involve simulations of arbitrary machines of the enumeration. For example, in the case of randomized machines with two-sided error, simulating a randomized machine M_i on an input on which M_i accepts with probability 50% would take M outside of the model because its error probability is not bounded away from 50%.

Instance checkers are tools that enable us to refrain from making errors. Recall that an instance checker for a language L' is a polynomial-time randomized oracle machine C that can output 0, 1, or "I don't know" on any input x such that the following properties hold: (i) $C^{L'}(x)$ outputs $L'(x)$ with probability 1, and (ii) for any oracle P, $C^{P}(x)$ outputs $\neg L'(x)$ with exponentially small probability. Barak [1] had the insight that an instance checker for a language L' in exponential time yields a randomized machine M' with two-sided error that decides L' and has a running time that is optimal up to a polynomial factor. The machine M' acts as follows: For $k = 1, 2, \ldots$ and for $i = 1, \ldots, k$, run C^{M_i} for k steps and halt as soon as one of the runs of the instance checker comes to a 0/1 conclusion; then output that conclusion. Let $t(n)$ denote the worst-case high-confidence running time of M' on inputs of length n. The properties of the instance checker imply that (a) $t(n)$ is exponentially bounded, (b) M' decides L' with exponentially small two-sided error, and (c) for some positive constant α, no machine M_i can do the same in $(t(n))^{\alpha}$ steps. The details of the argument are not relevant for us but the intuition for the optimality property (c) is that M' would start running the instance checker with oracle M_i as soon as $k \geq i$; if M_i were to decide L' with high confidence within $(t(n))^{\alpha}$ steps for some sufficiently small positive constant α, then M' would halt with high confidence within fewer than $t(n)$ steps.

If L' is complete for exponential time and $t(n)$ is polynomially bounded then we can efficiently transform every exponential-time deterministic machine into an equivalent polynomial-time randomized machine with two-sided error. We can trivially transform a polynomial-time randomized machine into an equivalent exponential-time deterministic machine. The desired hierarchy theorem for randomized machines with two-sided error (at the polynomial-time level) then follows from the hierarchy theorem for deterministic machines (at the exponential-time level).

If $t(n)$ is not polynomially bounded then for any constant c there are infinitely many input lengths n such that $(t(n))^{\alpha/4c} \geq n + 1$. Suppose we could efficiently compute a value $t^*(n)$ such that $(t(n))^{\alpha/4c} < t^*(n) \leq (t(n))^{\alpha/2c}$. Then padding strings of length n in L' to length $t^*(n)$ would yield a language $L = \{x10^{t^*(|x|)-|x|-1} \mid x \in L'$ and $t^*(|x|) \geq |x| + 1\}$ computable by randomized machines with two-sided error in polynomial time but not in time n^c. We chose the range for $t^*(n)$ such that there exists a (unique) value of the form $t^*(n) = 2^{2^{\tau^*(n)}}$ in that range with $\tau^*(n)$ integer. Computing $\tau^*(n)$ may be difficult but its value can be specified using $\log\log t^*(n)$ bits. Therefore, L can be decided by a randomized machine M with two-sided error in polynomial time with $a(n) = \log\log n$ bits of advice but not by such machines in time n^c without advice.

This isn't a fair time hierarchy theorem yet – for that, the time n^c machines should be allowed the same amount of advice as M. We can satisfy that requirement by tweaking the construction of the machine M' such that it runs each of the machines M_i with every possible advice string of length $\log\log k$. In fact, we can accommodate up to $a \log k$ bits of advice for the M_i's for any constant a. Both the case where $t(n)$ is polynomially bounded (now needing a hierarchy theorem for deterministic machines with advice) and the other case carry through.

Moreover, once the advice for the witnessing machine M is under $\log n$ bits, we can apply a translation technique and obtain a hierarchy theorem with a single bit of advice. This involves another level of padding to encode the $a(n) < \log n$ bits of the original advice for M in the padding length and using the one bit of new advice to indicate whether the padding length is valid. See [3, 7] for the details. This way, we obtain a language which randomized machines with two-sided error can decide in polynomial time and one bit of advice but not in time n^c and $a \log n$ bits of advice. The same strong separation holds for any reasonable semantic model of computation with the additional property of being efficiently closed under randomized reductions with two-sided error. We refer to Section 4.3 for the formal definitions.

Semantic classes with one-sided error typically do not exhibit the latter additional closure property. For the specific model of randomized machines with one-sided error, Fortnow et al. [4] use a modification of the above two-case approach to derive a somewhat weaker separation result, namely with $a(\log n)^{1/c}$ instead of $a \log n$ bits of advice at the smaller time bound of n^c. See Theorem 4 and footnote 2 for the precise statement. Instead of an exponential-time complete language L' and Barak's optimal algorithm based on instance checkers, Fortnow et al. consider an NP-complete language L and Levin's optimal algorithm based on searching for NP-witnesses [11]. The more restrictive advice bound of $a(\log n)^{1/c}$ is dictated by the separation result for nondeterministic machines with advice, which is needed for the case where $t(n)$ is polynomially bounded.

For their actual hierarchy theorem (where the length of the advice is the same for both time bounds considered), Fortnow et al. manage to eliminate the need for additional model requirements but they can only do so for some advice function in $O(\log n \cdot \log\log n)$. Their approach can be viewed as running Cook's argument with advice. The $\log n$ term in the advice bound comes from the $O(\log n)$ levels in Cook's argument. The $\log\log n$ term per level comes from a padding argument similar to Barak's.

Using a different strategy, we manage to get the advice down to a single bit. In fact, we obtain a hierarchy theorem with a bits of advice for any reasonable semantic model and any constant $a \geq 1$. We view our approach as extending Žàk's delayed diagonalization argument to machines with a bits of advice. A similar extension of Seiferas et al.'s argument leads to the same result [13]. Our approach à la Žàk seems more appropriate for deriving the unconditional hierarchy theorems we obtain; the approach à la Seiferas et al. seems more suitable for obtaining hierarchy theorems that are conditional on a complexity class separation such as P $\neq$ NP [8].

3. Intuition and informal derivation

In this section, we first sketch the construction of our generic hierarchy theorem with a constant number of bits of advice, and then the argument for our separation theorems. The formal treatment will be given in Section 4.

3.1. Hierarchy theorem

Consider a semantic model of computation with enumeration $M_1, M_2, \ldots$. We assume that there exists some underlying notion of "promise" which allows us to tell whether M_i with advice sequence $\alpha = \alpha_0, \alpha_1, \alpha_2, \ldots$, satisfies the promise on a given input x. Whether the latter is the case only depends on the behavior on input x; in particular, it is determined by M_i and the component $\alpha_{|x|}$ of the advice sequence α. We use the notation M_i/α to denote M_i with advice sequence α, and $M_i//\alpha_n$ to denote M_i with advice α_n at a fixed length n. M_i/α falls within the model iff $M_i//\alpha_n$ satisfies the promise at every length n.

Let us try to use straightforward diagonalization to establish a hierarchy theorem with $a \geq 0$ bits of advice. For a given constant $c \geq 1$, we would like to construct a machine M and an advice sequence α of modulus a (i.e., $|\alpha_n| = a$ for each length n), such that M/α falls within the model, takes not much more than n^c time, and disagrees with each M_i/β for each advice sequence β of modulus a for which M_i/β falls within the model and runs in time n^c.

With each M_i we associate a length n_i and distinct strings $x_{i,b}$ of length n_i for each value of $b \in \{0,1\}^a$. If $M_i//b$ satisfies the promise on $x_{i,b}$ and runs in time n_i^c, we would like to have M/α do the opposite of $M_i//b$ on that input. Assuming the existence of an efficient universal machine U, we would set

$$M/\alpha\,(x_{i,b}) = \neg U(\langle M_i//b, x_{i,b}, 0^{n_i^c}\rangle). \tag{1}$$

There are two problems with this approach. First, complementation may not be easy within the model. Second, even if complementation is easy, the simulation (1) may violate the promise. Recall that M/α has to satisfy the promise everywhere, whereas $M_i//b$ (run for n_i^c steps) may violate the promise on input $x_{i,b}$ for some values of b. Of course, there is no need to diagonalize in the case where b does not work for M_i on input $x_{i,b}$, i.e., if $M_i//b$ does not satisfy the promise on $x_{i,b}$ or takes more than n_i^c time. In that case, M/α can do something trivial, e.g., reject irrespective of the input. However, figuring out whether b works for M_i on input $x_{i,b}$ may not be easy. We could tell M for each value of b whether b works for M_i on $x_{i,b}$ but that would require $2^a > a$ bits of advice for M_i at length n_i. In fact, with 2^a bits of advice we could tell M explicitly how to behave like (1) on the 2^a strings $x_{i,b}$.

By adapting the technique of delayed diagonalization, we can cut the advice M needs to a single bit, implying a hierarchy theorem for any constant $a \geq 1$. Delayed diagonalization consists of a slow complementation executed at a larger input length n_i^* and a process to copy down the complementary behavior to length n_i. We will use a slow but safe simulation of $\neg U$ and exploit the freedom the copying process offers to link the behavior on various input lengths and — in some sense — spread the 2^a bits of advice needed at some length n' over different smaller lengths n. By a safe simulation of $\neg U$ we mean a machine S which always satisfies the promise and agrees with $\neg U$ on input x whenever U satisfies the promise on x. M may not have enough time to run S on $\langle M_i//b, x_{i,b}, 0^{n_i^c}\rangle$ at length n_i but it certainly does at a sufficiently larger length n_i^*, typically $n_i^* = 2^{n_i^c}$. We then set up M and α on lengths between n_i and n_i^* in such a way that if M_i/β satisfies the promise, runs in time n^c, and agrees with M/α for some advice sequence β of modulus a, then M/α "copies" its behavior at length n_i^* down to certain smaller and smaller lengths. If we can reach length $n = n_i$, we have the following contradiction for $b = \beta_n$:

$$\begin{aligned} M_i/\beta\,(x_{i,b}) = M/\alpha\,(x_{i,b}) = S(\langle M_i//b, x_{i,b}, 0^{n^c}\rangle) = \\ = \neg U(\langle M_i//b, x_{i,b}, 0^{n^c}\rangle) = \neg M_i/\beta\,(x_{i,b}). \end{aligned} \tag{2}$$

Thus, we succeeded in diagonalizing against M_i/β for any advice sequence β of modulus a. Due to the spreading, for a given M_i and b, we actually need strings $x_{i,b} = x_{i,b,n}$ of many comparable but different lengths n in order to guarantee that we can reach at least one of those lengths again while copying down.

The copying process capitalizes on M's ability to spend polynomially more time than the n^c steps M_i is allotted.

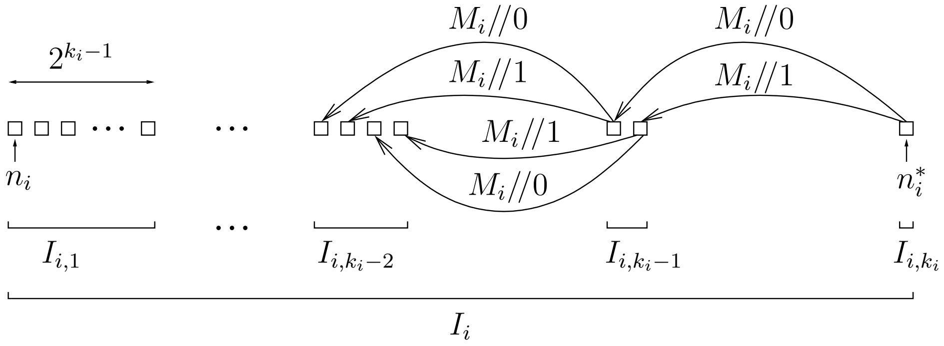

Figure 1. Construction of M on I_i for $a = 1$ in Theorem 1. An arrow from length n' to length n labeled $M_i//b$ denotes that $M//1$ at length n copies $M_i//b$ at length n'.

This allows M to simulate M_i on polynomially larger inputs. Consider length $n' = n_i^*$ and each possible value of $b \in \{0,1\}^a$. We say that b works for M_i at length n' if b works for M_i on all inputs of length n', i.e., $M_i//b$ satisfies the promise on all inputs of length n' and runs in time $(n')^c$. In that case, we pick some smaller but polynomially related length n and allow M/α on inputs x of length n to run $M_i//b$ on the input $0^{n'-n}x$ of length n'. As a result, we have that

$$(\forall\, x \in \{0,1\}^n)\ M/\alpha\,(x) = M_i//b\,(0^{n'-n}x).$$

We say that M/α at length n copies $M_i//b$ at length n'. If b does not work for M_i at length n', we let M/α act trivially at length n'. We use different lengths n for different values of b in such a way that b and n' are efficiently recoverable from n. Thus, M only needs a single bit of advice α_n at each length n, namely whether or not b works for M_i at length n'.

We then recursively apply the process to all 2^a lengths n we introduced, each time fixing the behavior of M/α at new lengths n. Provided we do not run out of lengths, we reach a point where the lengths n become so small that $S(\langle M_i//b, x, 0^{n^c}\rangle)$ runs in time polynomial in n_i^* for strings x of length n. At that point, the copying process bottoms out and we try to diagonalize as indicated above: For each $b \in \{0,1\}^a$, we pick a different string $x_{i,b,n}$ of length n, e.g., $x_{i,b,n} = 10^{n-a-1}b$, and define

$$M/\alpha\,(0^{n_i^*-n}x_{i,b,n}) = S(\langle M_i//b, x_{i,b,n}, 0^{n^c}\rangle). \quad (3)$$

The pattern $1(0+1)^*$ for the strings $x_{i,b,n}$ ensures the compatibility of (3) for different lengths n. On strings of length n_i^* that are not of the form $0^*x_{i,b,n}$, M/α acts trivially. If we make sure that n_i^* and the bottom-out lengths n are efficiently recognizable, M does not need any advice at length n_i^*.

One can think of the copying process as constructing a tree from the root n_i^* to the leaves. Each copying step creates 2^a siblings that are connected to their parent n' through an edge labeled with a corresponding value of $b \in \{0,1\}^a$. The process associates a unique length to each non-root node and determines the behavior of M/α at that length by specifying the corresponding advice bit. It leaves the behavior at the root length n_i^* free to be used for the diagonalization.

Now, suppose that for some advice sequence β of modulus a, M_i/β falls within the model, runs in time n^c, and agrees with M/α. Consider the path from the root n_i^* to a leaf n obtained by selecting at every non-leaf node n' the edge labeled $\beta_{n'}$. For each edge on that path, its label works for M_i at the parent node n' so M/α at the child node n copies M_i/β at the parent node n'. Since M/α and M_i/β agree, this means that the behavior of M/α is copied down along that path and that M/α at length n copies M/α at length n_i^*. However, (3) then leads to the contradiction (2) for $b = \beta_n$ and $x_{i,b} = x_{i,\beta_n,n}$.

To finish the argument, we need to argue that we have enough lengths n available to execute the above process. We can assign subsequent lengths from left to right to any given level of the copying tree, with gaps between the intervals used for adjacent levels. Let n_i denote the start of the first interval and k_i denote the number of intervals, i.e., the number of levels of the tree. The jump from the start of any interval to the start of the next one can be an arbitrary but fixed polynomial, say from n to n^d. Assuming the safe simulation S runs in exponential time, we need $\Theta(\log n_i/\log d)$ such jumps to go from n_i to n_i^* so we set $k_i = \Theta(\log n_i/\log d)$. The first interval forms the bottleneck for the embedding because it is the largest one and the gap that is available for it is smallest. The first interval contains $a^{k_i-1} = n_i^{\Theta(\log a/\log d)}$ elements, which fit within the gap between n_i and n_i^d provided d is a sufficiently large. Thus, we can accommodate all intervals without overlap.

We refer to the formal proof in Section 4.2 for a more detailed calculation.

Figure 1 illustrates the process for $a = 1$. In that case, the tree is binary; interval $I_{i,j}$ in the figure contains the $2^{k_i - j}$ nodes at depth $k_i - j$ of the copying tree, $1 \leq j \leq k_i$.

We managed to let M/α diagonalize against M_i/β for any β of modulus a. We did so by specifying the behavior of M/α on some lengths n in the interval $I_i = [n_i, n_i^*]$, while always making sure that M/α satisfies the promise and runs in some fixed polynomial amount of time. To handle all machines M_i in one construction, we use disjoint intervals I_i for different machines M_i and let M/α act trivially on those lengths n we do not use during the process.

The above technique applies to any semantic model that has an efficient universal machine which can be complemented within the model in exponential time, and that is efficiently closed under deterministic transducers. Taking these properties as the definition for a reasonable semantic model, we obtain Theorem 1. Theorem 2 follows from Theorem 1 by a standard padding argument. We refer to Section 4.2 and the journal version for more details.

Before moving on to our stronger separation results, let us point out the intuitive role the one bit of advice for M plays: It allows us to prevent M/α from simulating machines M_i/β on inputs where they do not satisfy the promise – a critical issue in semantic non-syntactic models.

3.2. Separation theorems

The above approach only works for bounded modulus $a(n)$. For unbounded modulus $a(n)$, the number of leaves of the copying tree becomes super-polynomial in the largest length ℓ associated to a leaf, which is incompatible with the requirement that each leaf maps to a unique length. Even if we are willing to give M $a(n)$ bits of advice at length n, the issue remains.

We get around the problem by restricting the behavior of M/α in such a way that it can be safely recovered at length n' from any list of machines at least one of which works appropriately at length n'. By the latter we mean: satisfying the promise at length n', running in time $(n')^c$, and agreeing with M/α at length n'. We can then modify the process for copying from length n' to length n as follows. At length n, M/α gets as advice whether there exists a string $b \in \{0,1\}^{a(n')}$ such that $M_i//b$ works appropriately at length n'. In case the advice bit is set, on an input x of length n, M/α runs the above recovery procedure for M/α on input $0^{n'-n}x$ using the list of machines $M_i//b$ for each $b \in \{0,1\}^{a(n')}$; as a result, M/α at length n copies M/α at length n'. Otherwise, M/α acts trivially at length n.

Notice that there no longer is a need for multiple lengths n to map to the same length n'. The copying tree becomes a line with root at length n_i^* and a unique leaf at length n_i. There also no longer is a need to make large (polynomially bounded) jumps from n to n'. We needed those in Section 3.1 just to ensure enough space for embedding the intervals. Since the intervals are now of length 1, we could set $n' = n + 1$. Since there is only one leaf, the structure of the copying tree on its own does not impose any limitations on the size of the modulus. As the recovery procedure needs to consider $M_i//b$ for each possible $b \in \{0,1\}^{a(n')}$, $a(n')$ has to be logarithmically bounded for M/α to run in polynomial time. Therefore, logarithmic moduli are the best one can hope for using this approach.

Safe recovery is only possible in some settings. We know of two basic mechanisms, namely instance checking and membership proof recovery. Both severely restrict the behavior of M/α and take away the freedom to define M/α at length n_i^* so as to complement $M_i//b$ at length n_i. Thus, for each mechanism we need new strategies to diagonalize. The models of computation also need to have the necessary closure properties to accommodate the recovery process based on instance checkers or membership proofs, respectively.

We use an instance checker to copy down EXP-complete behavior and then exploit that to diagonalize assuming the model allows complementation in EXP. We develop this approach in Section 3.2.1. It works up to the limit of logarithmic modulus.

We use membership proofs to copy down NP-complete behavior. Assuming the model allows an efficient simulation in NP, we obtain an efficient safe simulation which we then use to simplify the construction from Section 3.1. We develop this approach in Section 3.2.2. It works up to modulus $\Theta((\log n)^{1/c})$ for randomized machines with one-sided error and up to modulus $\Theta(\log n)$ for randomized machines with two-sided error.

3.2.1 Copying using instance checking

Recall that an instance checker for a language L is a polynomial-time randomized oracle machine C that can output 0, 1, or "I don't know" on any input x such that the following properties hold: (i) $C^L(x)$ outputs $L(x)$ with probability 1, and (ii) for any oracle P, $C^P(x)$ outputs $\neg L(x)$ with exponentially small probability. There exist instance checkers for certain paddable exponential-time complete languages L that only make queries of length $f \cdot n$ on inputs of length n for some constant $f \geq 1$. For ease of exposition, we assume in this section that $f = 1$. The formal proof in the journal version will show how to eliminate that assumption.

The key for safe recovery of L is roughly the following computation: For each possible advice string b of length $a(n)$, run the instance checker C with the oracle defined by n^c computation steps of $M_i//b$ at length n; halt as soon as one of these runs produces a 0/1 conclusion and then output that conclusion. Provided the model of computation is closed under randomized reductions with two-sided error, the properties of the instance checker guarantee that this computation works appropriately as long as there exists at least one advice string b for which $M_i//b$ works appropriately.

Let us be a bit more precise. Let $n_{i,j}$, $1 \leq j \leq k_i$, denote the lengths associated to the nodes of the copying line, where $n_{i,1} = n_i$ and $n_{i,k_i} = n_i^*$. On input $0^{n_i^* - n_i}x$, where x is a string of length n_i, M runs a fixed deterministic exponential-time algorithm for L on input x. For any $1 \leq j < k_i$, $M//1$ acts as follows on inputs of the form $0^{n_{i,j} - n_i}x$ where x is a string of length n_i: For each advice string b of length $a(n_{i,j+1})$, run the instance checker C on input x answering each query y by taking the majority vote of a linear number of independent runs of $U(\langle M_i//b, 0^{n_{i,j+1}-n_i}y, 0^{(n_{i,j+1})^c}\rangle)$; halt as soon as one of these computations yields a 0/1 conclusion and then output that conclusion. $M//1$ acts trivially on other inputs of length $n_{i,j}$, as does $M//0$ on all inputs.

We say that $M_i//b$ works appropriately at length n if b works for M_i at that length and L at length n_i is a copy of $M_i//b$ at length n, i.e., for each string x of length n_i, $M_i//b\,(0^{n_{i,j}-n_i}x) = L(x)$. We set $\alpha_{n_{i,j}}$ for $1 \leq j < k_i$ to indicate whether there exists a string b of length $a(n_{i,j+1})$ such that $M_i//b$ works appropriately at length $n_{i,j+1}$. If so, we know that L at length n_i is a copy of M/α at length $n_{i,j}$.

If the copying process succeeds, we have that $\alpha_{n_i} = 1$ and therefore M/α agrees with the exponential-time complete language L at length n_i. We exploit this fact to accomplish the desired diagonalization as follows. We introduce a new length $\tilde{n}_i$ smaller than n_i. For any string b of length $a(\tilde{n}_i)$, consider the complement of the deterministic simulation of $U(\langle M_i//b, 0^{\tilde{n}_i - a(\tilde{n}_i)}b, 0^{\tilde{n}_i^c}\rangle)$. Assuming that computation runs in deterministic exponential time, we can compute in polynomial time a string $z_{i,b}$ such that $L(z_{i,b}) = \neg U(\langle M_i//b, 0^{\tilde{n}_i - a(\tilde{n}_i)}b, 0^{\tilde{n}_i^c}\rangle)$. Using the paddability of L, we can set up things such that the length of $z_{i,b}$ equals n_i.

$M//1$ on input $0^{\tilde{n}_i - a(\tilde{n}_i)}b$ then runs $M//1$ on input $z_{i,b}$. Like before, $M//1$ acts trivially on other strings of length $\tilde{n}_i$, as does $M//0$ on all strings. We set $\alpha_{\tilde{n}_i}$ to indicate whether $M//1$ agrees with L on inputs of length n_i.

Now, suppose there exists an advice sequence β of modulus $a(n)$ such that M_i/β falls within the model, runs in time n^c, and agrees with M/α. Then the copying process is guaranteed to succeed and we obtain a contradiction similar to (2): For $b = \beta_{\tilde{n}_i}$ and $x_{i,b} = 0^{\tilde{n}_i - a(\tilde{n}_i)}b$,

$$\begin{aligned} M_i/\beta(x_{i,b}) = M/\alpha\,(x_{i,b}) = M/\alpha\,(z_{i,b}) = L(z_{i,b}) = \\ = \neg U(\langle M_i//b, x_{i,b}, 0^{\tilde{n}_i^c}\rangle) = \neg M_i/\beta\,(x_{i,b}). \end{aligned}$$

Note that M/α at length n runs the instance checker C at most $2^{a(n')}$ times, where $n' = n^{O(1)}$. It follows that M/α runs in polynomial time as long as $a(n') \leq a \log n'$ for some constant a and the model is efficiently closed under randomized reductions with two-sided error. This approach works for any reasonable randomized semantic model with the latter closure property, thus establishing Theorem 3.

Let us end the informal treatment by reiterating the role of the instance checkers in our construction: They provide us an advice efficient way to realize the desired copying by M while always satisfying the promise. We want the copying to happen as soon as there exists at least one advice string b for which $M_i//b$ behaves appropriately at length n'. Before, M needed a separate bit of advice for each possible advice string b, namely to indicate whether $M_i//b$ behaves appropriately at length n'. Now, we can handle all possibilities for b at once using a single bit of advice for M, namely whether there exists at least one choice of b for which $M_i//b$ behaves appropriately at length n'.

3.2.2 Copying using membership proof recovery

Consider a language L that has membership proofs and for which the search for a membership proof at length n reduces to L at length n. Satisfiability is an example of such a language L. The crux for the safe recovery of L is the following computation: For each possible string b of length $a(n)$, run the reduction using the oracle defined by n^c computation steps of $M_i//b$ at length n; verify those candidate membership proofs and accept iff at least one of them is valid. Models like randomized machines with one- or two-sided error allow the efficient simulation of the above process. Provided the model has the latter closure property, we can develop a copying process with one bit of advice in a similar way as in Section 3.2.1. It uses a sequence of lengths from m_i to $m_i^* = 2^{m_i^{O(1)}}$ with jumps bounded by some fixed polynomial, and allows us to assume that $M//1$ decides L at length m_i.

Now, assume that our model of computation has a universal machine U that can be mimicked by a nondeterministic polynomial-time machine N. This is the case, for example, for the model of randomized machines with one-sided error: For a randomized machine M_i, string x, and integer $t \geq 0$, we can let $N(\langle M_i, x, 0^t\rangle)$ check whether there exists a random string that makes M_i accept input x in t steps; whenever M_i satisfies the promise on input x and runs in t steps, $N(\langle M_i, x, 0^t\rangle) = M_i(x)$. Suppose also that L is paddable and NP-complete, as satisfiability is. Then, for some length $\tilde{m}_i$ polynomially related to m_i, there exists an efficient translation of queries to U of length $\tilde{m}_i$ into queries to L of length m_i. Since we can assume that $M//1$ satisfies the promise at length m_i, runs in polynomial time, and agrees with L at length m_i, we obtain an efficient safe simulation T of U at length $\tilde{m}_i$.

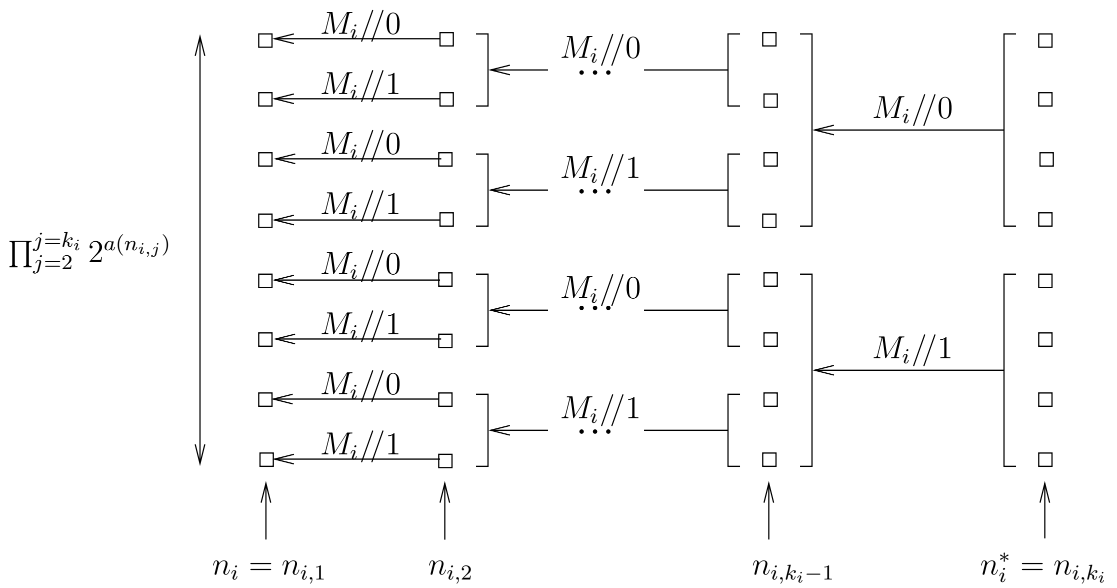

Figure 2. Partial construction of M on I_i for $a(n) = 1$ in Theorem 3. Each box contains two distinct strings, one corresponding to each value in $\{0,1\}^{a(n_i)}$. An arrow from boxes at length n' to boxes at length n labeled $M_i//b$ denotes that $M//1$ copies $M_i//b$ on the corresponding inputs.

An efficient safe simulation of U can be used as a substitute for U in the construction from Section 3.1. In that case, there no longer is a need for advice as each advice bit in that construction indicates whether U satisfies the promise on a certain set of inputs — T satisfies the promise everywhere! As a consequence, we no longer have to use different lengths for all the nodes of the copying tree. We still need to assign $2^{a(\ell)}$ strings of length ℓ to each leaf of length ℓ such that these strings are distinct for all leaves.

Suppose the length we assign to a node only depends on its depth in the tree. As before, let us use the notation $n_{i,j}$ to denote the length corresponding to depth $k_i - j$, $1 \le j \le k_i$, with $n_i \doteq n_{i,1}$ and $n_i^* \doteq n_{i,k_i}$. The resulting copying process is illustrated for the case $a(n) = 1$ in Figure 2.

The logarithm of the number of distinct strings of length n_i we need can then be expressed as

$$a(n_i) + \sum_{j=2}^{k_i} a(n_{i,j}) = \sum_{j=1}^{k_i} a(n_{i,j}). \tag{4}$$

The question is how large we can make $a(n)$ such that (4) does not exceed n_i.

If a computation of n^c steps can be complemented within the model in time 2^{n^c} (as in the case of randomized machines with one-sided error), we have the condition $2^{n_i^c} \le (n_i^*)^{O(1)}$ in order to guarantee that M/α runs in polynomial time at length n_i^*. Since $a(n_i^*) \le n_i$ follows from our upper bound on (4), we conclude that $a(n_i^*) = O((\log n_i^*)^{1/c})$.

We can actually achieve modulus $a(\log n)^{1/c}$ for any constant a. By setting $n_{i,j+1} = n_{i,j}^d$, $1 \le j < k_i$, where d is any constant, (4) becomes a linear function in n_i with a coefficient that is a geometric sum $\sum_{j=1}^{k_i} r^j$ and such that the ratio r converges to 0 when d grows. By picking d large enough, we can bound (4) by n_i.

We developed our simplification of the copying process from Section 3.1 assuming free access to an efficient safe simulation T of U at all the levels we need it. In reality, we have to build T at all those levels using the recovery procedure. We can apply the recovery as described above to obtain T at length $\tilde{m}_i = n_i^*$ by building L at length m_i. Once we have L at length m_i, we can exploit the paddability properties of L and apply the recovery procedure to obtain L at any smaller length in I_i except the few lengths that are reserved for the simplified copying process. This effectively makes T available at all lengths up to n_i^*, which is (more than) what we need for the simplified copying process. See Figure 3 for an illustration.

Finally, we point out that copying using membership proof recovery also leads to a *relativizing* proof of Theorem 3 for the specific case of randomized machines with two-sided error.

Our strong separation result for randomized machines with one-sided error hinges on the efficient simulation of a universal machine by a nondeterministic machine. For randomized machines with two-sided error we do not know an efficient simulation of a universal machine U by a nondeter-

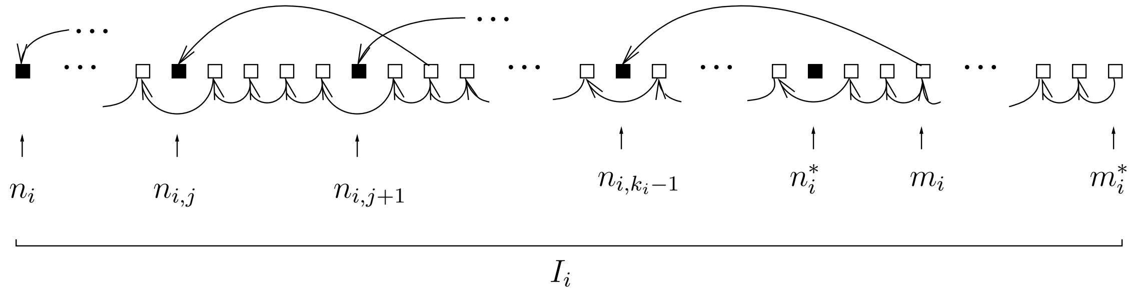

Figure 3. Full construction of M on I_i in Theorem 3, combining simplified copying nodes (black) with efficient safe simulation nodes (white). An arrow from length n' to length n denotes that $M//1$ (above line) or $M_i//b$ (below line) at length n' is used to construct $M//1$ at length n.

ministic machine but do by a nondeterministic machine N^L with oracle access to satisfiability, L. Moreover, the two-sided error model has the following closure property: Given randomized machines R_i and R'_i that efficiently solve satisfiability with two-sided error on inputs of certain lengths m_i and m'_i, respectively, we can easily construct a randomized machine M that efficiently simulates N^L with two-sided error on inputs of a polynomially related length $\tilde{m}_i$. The property essentially follows from the standard argument that if NP $\subseteq$ BPP then $\Sigma_2^p \subseteq$ BPP. More precisely, suppose wlog. that N^L only makes oracle queries of length exactly m_i on inputs of length $\tilde{m}_i$. Let R''_i be a sufficiently amplified version of R_i such that for most random strings ρ of the appropriate length, $R''_i(q, \rho)$ computes satisfiability correctly for all queries q of length m_i. For any fixed ρ and input x of length $\tilde{m}_i$, $N^{R''_i(\cdot,\rho)}(x)$ defines a nondeterministic computation, which we can efficiently translate into a satisfiability question y, say of length m'_i. Our simulation M of $N^L(x)$ first picks ρ uniformly at random, then computes y, and finally runs R'_i on y.

Combined with copying down satisfiability to lengths m_i and m'_i, we obtain our efficient safe simulation T of U at length $\tilde{m}_i$. By the analysis for the one-sided error case, this leads to a strong separation result with up to $a(\log n)^{1/c}$ bits of advice on the time n^c side. In fact, we can do better by exploiting the efficient closure of the two-sided error model under complementation. We can actually get an efficient safe *complementation* S of U instead of an efficient safe *simulation* T. In that case, we only need to apply S once, namely at the beginning of the interval I_i, and furthermore copy down satisfiability as needed. In terms of Figure 3, there is only one black node, namely at length n_i. The tree in Figure 2 reduces to a single node and (4) to the single term $a(n_i)$. The restriction on the advice length due to the simulation/complementation is relaxed from $a(n) = O((\log n)^{1/c})$ to $a(n) \leq n$. The restriction due to copying down satisfiability remains $a(n) = O(\log n)$.

Note that the above argument does not use instance checkers or any other nonrelativizing ingredient. Thus, we obtain a relativizing proof of Theorem 3 for the case of randomized machines with two-sided error. The argument fails for quantum machines with two-sided error because the simulation of BQP in the polynomial-time hierarchy remains open.

4. Formal treatment

In this section we provide the formal definitions of the notions involved in the statements of our results. We introduce the notion of a (randomized) semantic model of computation with advice and list the modest properties we need for our hierarchy and separations theorems to apply. We also formally prove our main result (Theorem 1). We refer to the journal version of the paper for the proofs of our other results.

4.1. Semantic models

Fix an alphabet Σ containing the symbols 0 and 1. We abstractly view a *model of computation* as consisting of a set $\mathcal{M} \subseteq \Sigma^*$ of "machines" (or "programs"), and a partial computable function $\gamma : \Sigma^* \times \Sigma^* \to \Sigma^*$. For any $M \in \mathcal{M}$ and $x \in \Sigma^*$, $\gamma(M, x)$ determines the output of M on input x (possibly undefined). We also use the shorthand $M(x)$ for $\gamma(M, x)$. A language $L \subseteq \Sigma^*$ is said to be "accepted" or "decided" by M if $M(x) = L(x)$ for each $x \in \Sigma^*$, where $L(x)$ denotes the indicator for the property "$x \in L$", i.e., $L(x) = 1$ if $x \in L$ and $L(x) = 0$ otherwise.

We assume there is an underlying notion of *time*. Whenever $\gamma(M, x)$ is defined, M halts and produces its output after a finite number of steps, denoted $t_M(x)$. We say that M runs in time t at length n if $t_M(x) \leq t$ for each $x \in \Sigma^n$, and that M runs in time $t(n)$ if M runs in time $t(n)$ at each length n.

We call a model of computation *syntactic* if $\mathcal{M}$ is computably enumerable. We call the model *semantic* if there exists a computably enumerable set $\mathcal{M}' \subseteq \Sigma^*$ and a predicate $\pi \subseteq \Sigma^* \times \Sigma^*$ such that

$$\mathcal{M} = \{M \in \mathcal{M}' \mid (\forall\, x \in \Sigma^*)\, \pi(M, x)\}.$$

The predicate π can be thought of as a condition on or *promise* about the behavior of M on input x. A machine $M \in \mathcal{M}'$ has to satisfy the promise on each input x in order to fall within the computation model $\mathcal{M}$. Note that we could abstract away the predicate π at this point and just consider the model as defined by $\mathcal{M}$ and γ. However, the predicate π will play a critical role once we introduce advice. We assume the notion of running time extends to every machine M in $\mathcal{M}'$.

Examples of syntactic models include deterministic, nondeterministic, and randomized machines, as well as alternating machines of any fixed signature. Every syntactic model is also semantic but not vice versa. For example, randomized machines with two-sided error form a semantic non-syntactic model $\mathcal{M}$. There does not exist a computable enumeration of $\mathcal{M}$ but the model $\mathcal{M}'$ of all randomized machines is syntactic and we can obtain $\mathcal{M}$ as those machines of $\mathcal{M}'$ that satisfy the promise of two-sided error. Other examples of semantic non-syntactic models include randomized machines with one-sided or zero-sided error, quantum machines with two-, one-, or zero-sided error, unambiguous machines, symmetric alternation, Arthur-Merlin games of any signature, etc.

We point out that similar formalizations of the intuitive difference between syntactic and semantic computation have been proposed before in the literature [12, 4]. However, the earlier attempts all seem to associate these notions with complexity classes rather than models of computation. For example, BPP (the class of languages decidable by polynomial-time randomized machines with two-sided error) is considered a semantic non-syntactic class, whereas P is considered syntactic. This leads to inconsistencies since BPP may coincide with P. Our approach based on machines rather than languages does not suffer from that pitfall.

An *advice* sequence α of modulus $a(n)$ is an infinite sequence of strings $\alpha_0, \alpha_1, \alpha_2, \ldots$, one for each length n, such that $|\alpha_n| = a(n)$ for each n. We define the behavior of a machine $M \in \mathcal{M}'$ with advice α, denoted M/α, on a given input x as equal to the behavior of M on input $\langle x, \alpha_{|x|}\rangle$, where $\langle \cdot, \cdot \rangle$ denotes a standard pairing function. In particular, M/α satisfies the promise on input x iff $\pi(M, \langle x, \alpha_{|x|}\rangle)$ holds, and $M/\alpha(x) = M(\langle x, \alpha_{|x|}\rangle)$. Whenever we talk about a property of M/α at length n (like satisfying the promise, running time, etc.), we refer to that property on all inputs of the form $\langle x, \alpha_n\rangle$ where x is a string of length n. Note that the behavior of M/α at length n depends on the component α_n but not on the other components of α. We use the shorthand $M//\alpha_n$ to denote that behavior.

We consider M/α to fall within the model iff $M \in \mathcal{M}'$ and M/α satisfies the promise at each length. We point out that, apart from the predicate π, the choice of the encapsulating syntactic model $\mathcal{M}'$ and the actual advice string α play a role. This differs from the Karp-Lipton notion of computation with advice [10], who essentially only consider those machines $M \in \mathcal{M}'$ that robustly satisfy the promise, i.e., the machines in $\mathcal{M}$. More precisely, M/α falls within their model iff M/β falls within our model for each advice sequence β of the same modulus as α.

We now introduce the additional requirements a semantic model of computation has to satisfy for our hierarchy theorem to apply. The first one deals with the existence of an efficient universal machine.

Definition 1 *A* universal machine *is a machine $U \in \mathcal{M}'$ such that for each $M \in \mathcal{M}'$, $x \in \Sigma^*$, and $t \geq t_M(x)$, U satisfies the promise on input $\langle M, x, 0^t\rangle$ whenever M satisfies the promise on input x, and if so, $U(\langle M, x, 0^t\rangle) = M(x)$. We call U* efficient *if it runs in polynomial time.*

The second condition states that the model can be complemented within the model in exponential time. We phrase the condition in terms of the universal machine U.

Definition 2 *We say that U can be complemented within the model in exponential time if there exists a machine S that runs in time $2^{n^{O(1)}}$, satisfies the promise on every input, and such that $S(x) = \neg U(x)$ for every input $x \in \Sigma^*$ on which U satisfies the promise.*

The final property states that the model is closed under deterministic transducers. By the latter, we mean deterministic machines D that, on input x, output either an answer $a(x)$, or else a query $q(x)$. Note that a transducer that always outputs an answer is equivalent to a standard Turing machine, and that a transducer that always outputs a query is equivalent to a many-one reduction. For any $M' \in \mathcal{M}'$, we use the following notation:

$$D^{M'}(x) = \begin{cases} a(x) & \text{if } D \text{ outputs an answer on input } x \\ M'(q(x)) & \text{otherwise} \end{cases}$$

and

$$t_{D^{M'}}(x) = \begin{cases} t_D(x) & \text{if } D \text{ outputs an answer on input } x \\ t_D(x) + t_{M'}(q(x)) & \text{otherwise.} \end{cases}$$

We are now ready to formally state the closure property we need.

Definition 3 *A semantic model is* closed under deterministic transducers *if for each deterministic transducer D and* *each machine $M' \in \mathcal{M}'$, there exists a machine $M \in \mathcal{M}'$ such that the following holds for all inputs x: If $D(x)$ outputs an answer or if M' satisfies the promise on input $q(x)$, then M satisfies the promise on input x, and $M(x) = D^{M'}(x)$. We say that the closure is efficient if M runs in time $t_{D^{M'}}(x)$ on input x.*

Our hierarchy theorem applies to any semantic model with the above three properties.

Definition 4 *A semantic model of computation is called* reasonable *if it has an efficient universal machine that can be complemented deterministically in exponential time and if it is efficiently closed under deterministic transducers.*

All the concrete models mentioned in this paper are reasonable semantic models.

We point out that for the proof of Theorem 1, we can relax the efficiency requirement in Definition 3 to time $(t_{D^{M'}}(x))^{O(1)}$ instead of time $t_{D^{M'}}(x)$. However, for the strong hierarchy of Theorem 2, we seem to need the efficiently requirement as stated in Definition 3.

4.2. Proof of theorem 1

We now formally prove our main result. We refer to Section 3.1 for the intuition behind the proof.

Assume a reasonable semantic model of computation. Let $\mathcal{M}$ be the set of the machines belonging to the model. This set is contained in some other set $\mathcal{M}'$ that has a computable enumeration $(M_i)_{i=1}^{\infty}$. This reasonable semantic model has an efficient universal machine U which runs in time n^u for some constant u (Definition 1) and has a safe complementation S within the model running in time 2^{n^s} for some constant s (Definition 2). Without loss of generality, we assume that c is a positive integer.

A disjoint interval $I_i = [n_i, n_i^*]$ of input lengths is reserved for every machine M_i. Interval I_i contains the subintervals $I_{i,j} = [l_{i,j}, r_{i,j}]$, $1 \le j \le k_i$, where $l_{i,1} = n_i$, $r_{i,j} < l_{i,j+1}$ and $r_{i,k_i} = n_i^*$. We set

$$r_{i,j} = l_{i,j} + (2^a)^{(k_i-j)} - 1 \qquad k_i = \lceil \log n_i \rceil .$$

Thus, for every $n \in I_{i,j}$ we have $n = l_{i,j} + \Delta_n$, $0 \le \Delta_n < (2^a)^{(k_i-j)}$. We can think of $I_{i,j}$ as the nodes at level $k_i - j$ of a full 2^a-ary tree with root at n_i^*. Let us fix $l_{i,j}$, $1 \le j \le k_i$, such that

$$l_{i,j} = {n_i}^{d^{(j-1)}} \qquad d = \max(\lceil 4^{a\cdot c\cdot s} \rceil, 2a).$$

It remains to fix the starting input lengths n_i of the intervals I_i taking the following into account. For any number n we want to efficiently compute a number i such that $n \in I_i$ and the description of the machine M_i that corresponds to interval I_i. Since the enumeration $(M_i)_{i=1}^{\infty}$ of machines in $\mathcal{M}'$ can be very ineffective, we allow the intervals I_i to be sparsely distributed over input lengths, and we let $n_i = \max(n_{i-1}^* + 1, m)$ where m is such that the description of machine M_i is produced after m steps of the enumerating procedure. As for the starting length n_1, some of the inequalities in the proof below require that every input length n of interest (that is, belonging to some interval) is greater than some constant. We choose n_1 larger than all these constants. Notice that now, given a number n, we can compute in linear time the numbers i and j, if any, such that $n \in I_{i,j}$ and produce the description of machine M_i.

To guarantee the disjointness of the subintervals $I_{i,j}$ we need to check that $r_{i,j} < l_{i,j+1}$ for any i and any $1 \le j < k_i$. If n_1 is big enough, we have

$$\begin{aligned} (2^a)^{(k_i-j)} &\le (2^a)^{(k_i-1)} \le (2^a)^{\log n_i} \le {n_i}^a \\ r_{i,j} &= {n_i}^{d^{(j-1)}} + (2^a)^{(k_i-j)} - 1 < \\ &< {n_i}^{d^{(j-1)}} + {n_i}^a < {n_i}^{d^j} = l_{i,j+1}. \end{aligned}$$

Let $x_{i,b,n} = 10^{n-a-1}b$ where b is some string of length a. This works provided $n_1 > a$ as then any input length $n \in I_{i,j}$ is greater than a.

Given an input x of length n, machine M/α does the following.

1. Compute numbers i and j such that $n \in I_{i,j}$. If no such numbers exist, output 0 and halt.
2. If $j < k_i$ and $\alpha_n = 1$ then
   (a) Compute Δ_n such that $n = l_{i,j} + \Delta_n$.
   (b) Let $n' = l_{i,j+1} + \lfloor \Delta_n / 2^a \rfloor$ and let $b = \Delta_n \bmod 2^a$.
   (c) Call U on $\langle M_i//b, 0^{n'-n}x, 0^{(n')^c} \rangle$.
3. If $j = k_i$ and $x = 0^{n-m}x_{i,b,m}$ for some $m \in I_{i,1}$ then
   (a) Call S on $\langle M_i//b, x_{i,b,m}, 0^{m^c} \rangle$.
4. Output 0.

M uses its advice α_n at length n only if n belongs to some subinterval $I_{i,j}$, $1 \le j < k_i$. For such an input length n, let $\alpha_n = 1$ if $M_i//b$ satisfies the promise at length n' and runs in time $(n')^c$ (see the above algorithm for definitions of n' and b). Otherwise, let $\alpha_n = 0$.

Let us verify that the resulting machine M and advice α are such that M/α:

(A) falls within the model,

(B) runs in polynomial time, and

(C) disagrees with any M_i/β for any advice sequence β of modulus a for which M_i/β falls within the model and runs in time n^c.

Note that we can translate a query y of length m to S into the query $\langle S, y, 0^t \rangle$ to U with $t = 2^{m^s}$. Using that translation, M becomes a deterministic transducer to machine $U \in \mathcal{M}'$. The possible queries to U occur in steps 2(c) and 3(a) of the algorithm. Step 2(c) makes the query $\langle M_i//b, 0^{n'-n}x, 0^{(n')^c} \rangle$ to U. By the choice of the advice α, that step is only executed if $M_i//b$ satisfies the promise at length n' and runs in time $(n')^c$. As for step 3(a), by Definition 2, S satisfies the promise on every input. It follows from Definition 1 that machine U satisfies the promise on every query the transducer M makes. Thus, by Definition 3, M/α falls within the model.

The length of the query $\langle M_i//b, 0^{n'-n}x, 0^{(n')^c} \rangle$ to U in step 2(c) is polynomial in n since M_i is produced in time linear in n and $n' \leq n^d$. Step 3(a) runs S on input $\langle M_i//b, x_{i,b,m}, 0^{m^c} \rangle$ for some $m \in I_{i,1}$ and is only executed if the input to M is of length n_i^*. If n_1 is big enough, we have

$$|\langle M_i//b, x_{i,b,m}, 0^{m^c} \rangle| \leq 2m^c$$

$$m \leq n_i + (2^a)^{k_i - 1} - 1 < n_i + (n_i)^a \leq 2{n_i}^a$$

$$n_i^{2 \cdot a \cdot c \cdot s} \leq n_i^{\log d} = d^{\log n_i} \leq d^{k_i} \leq \log(n_i^{d^{k_i}}) = \log(n_i^*)^d,$$

and step 3(a) using the simulation by U takes time at most

$$(2 \cdot 2^{(2m^c)^s})^u \leq 2^{2^{s+1}(2 \cdot n_i^a)^{c \cdot s} u} < 2^{n_i^{2a \cdot c \cdot s}} \leq (n_i^*)^d.$$

The efficiency requirement in Definition 3 then implies that M runs in polynomial time.

For property (C), consider an arbitrary machine M_i with an advice sequence β of modulus a such that M_i/β falls within the model and runs in time n^c. Let us assume that M_i/β agrees with M/α at each length. Then we can prove by induction on j from k_i down to 1 that there exists an input length $n \in I_{i,j}$ such that M/α at length n copies M/α at length n_i^*, i.e.,

$$(\forall\, x \in \{0,1\}^n)\ M/\alpha\,(x) = M/\alpha\,(0^{n_i^*-n}x).$$

The case when $j = k_i$ holds trivially. For any $j < k_i$, by the induction hypothesis, there is an input length $n' \in I_{i,j+1}$ such that M/α at length n' copies M/α at length n_i^*. Then consider $n = l_{i,j} + \Delta_{n'} \cdot 2^a + \beta_{n'}$. We have that $n \in I_{i,j}$ and $n' = l_{i,j+1} + \lfloor \Delta_n / 2^a \rfloor$. By the specification of M and by the choice of the advice sequence α, M/α at length n copies M/α at length n' and, consequently, copies M/α at length n_i^*.

Hence, for some $n \in I_{i,1}$, M/α at length n copies M/α at length n_i^*. At the same time, M_i/β at length n fails to copy M/α at length n_i^* since $M/\alpha\,(0^{n_i^*-n}\, x_{i,\beta_n,n}) = \neg M_i/\beta\,(x_{i,\beta_n,n})$ whenever M_i/β satisfies the promise at length n and runs in time n^c. Therefore, M_i/β does not agree with M/α at length n, which contradicts our assumption.

This finishes the proof of Theorem 1.

Let us point out that we do not really need the strong form of efficiency stated in Definition 3 for the above proof. The place where it plays a role is in our argument for property (B); requiring the running time of M to be $(t_{D^{M'}}(x))^{O(1)}$ suffices for that argument.

4.3. Randomized semantic models

In order to apply an instance checker C in a semantic model of computation $(\mathcal{M}', \gamma, \pi)$, we need to augment the notion we introduced in Section 4.1. Intuitively, we would like to run C with an "oracle" P that is the result of running a machine $M' \in \mathcal{M}'$ on the queries y of the instance checker. For that to make sense and interact well with the properties of the instance checker, we need to associate a random variable $\underline{M}'(y)$ with the behavior of M' on input y. We call a model equipped with such random variables a *randomized model*. Natural examples for $\underline{M}'(y)$ include the acceptance indicator for randomized machines or for Arthur-Merlin games under an optimal strategy for Merlin.

Once we have such an underlying random process, for any randomized oracle machine D, we can define the random variable $D^{M'}(x)$ as the outcome of a run of D where each query y is answered according to a sample of $\underline{M}'(y)$. We require that we can efficiently simulate such a process in our model of computation and that the simulation be sound whenever $D^{M'}$ has two-sided error on input x. More precisely, we stipulate the following.

Definition 5 *A randomized semantic model of computation is* closed under randomized reductions with two-sided error *if for every randomized oracle machine D and every machine $M' \in \mathcal{M}'$, there exists a machine $M \in \mathcal{M}'$ such that the following holds for any string x: If $D^{M'}$ has two-sided error on input x, then M satisfies the promise on input x and $M(x)$ equals the majority outcome of $D^{M'}$ on input x. We say that the closure is efficient if M runs in time $(t_D(n) \cdot \max_{0 \leq m \leq t_D(n)} t_{M'}(m))^{O(1)}$.*

Another condition we need is that the model has an efficient universal machine U (see Definition 1) which can be simulated deterministically in exponential time.

Definition 6 *We say that U can be deterministically simulated in exponential time if there exists a deterministic machine T which runs in time $2^{n^{O(1)}}$ and such that $T(x) = U(x)$ for each $x \in \Sigma^*$ on which U satisfies the promise.*

Our generic separation theorem applies to any reasonable randomized semantic model defined as follows.

Definition 7 *A randomized semantic model of computation is called* reasonable *if it has an efficient universal machine* *that can be simulated deterministically in exponential time and if it is efficiently closed under randomized reductions with two-sided error.*

Reasonable randomized semantic models include randomized and quantum machines with two-sided error.

5. Further research

In this paper, we established a hierarchy theorem that applies to any "reasonable" semantic model of computation with one bit of advice (Theorems 1 and 2). The most pertinent open problem is to eliminate the need for the one bit of advice. Ideally, we would like to do that without further restricting the meaning of "reasonable" but the question remains open for any semantic model which is not known to be equivalent in power to a syntactic one.

For randomized machines with two-sided error, the question whether a hierarchy theorem would require nonrelativizing techniques is still up for debate [5, 6, 14]. Prior to our work, a hierarchy theorem with one bit of advice was established using nonrelativizing techniques. Our proof shows that the result itself as well as the strong separation stated in Theorem 3 do relativize for this specific model. Whether our generic separation theorem (Theorem 3) relativizes for every model remains open.

Improving the advice bound in our separation results (Theorems 3 and 4) forms another possible direction for further research. As for Theorem 4, one can abstract the properties the model needs for our proof to carry through, just as we did for our other arguments. We refrained from stating Theorem 4 in such generality because randomized machines with one-sided error are the only interesting application we could think of. Are there others?

Finally, one can ask about strong hierarchy theorems, in which the more restricted machines fail to decide the hard language for almost all input lengths (instead of just one or infinitely many). Even with advice, no such hierarchy theorems are known for a non-syntactic model.

Acknowledgments

We thank Edward A. Hirsch, Dimitri Grigoriev, Rahul Santhanam, and the anonymous Complexity reviewers for helpful comments on an earlier version of the paper.

References

[1] B. Barak. A probabilistic time hierarchy theorem for slightly nonuniform algorithms. In *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques in Computer Science*, volume 2483 of *Lecture Notes in Computer Science*, pages 194–208. Springer-Verlag, 2002.

[2] S. Cook. A hierarchy theorem for nondeterministic time complexity. *Journal of Computer and System Sciences*, 7:343–353, 1973.

[3] L. Fortnow and R. Santhanam. Hierarchy theorems for probabilistic polynomial time. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 316–324. IEEE, 2004.

[4] L. Fortnow, R. Santhanam, and L. Trevisan. Hierarchies for semantic classes. In *Proceedings of the 37th ACM Symposium on the Theory of Computing*, pages 348–355. ACM, 2005.

[5] L. Fortnow and M. Sipser. Probabilistic computation and linear time. In *Proceedings of the 21st ACM Symposium on the Theory of Computing*, pages 148–156. ACM, 1989.

[6] L. Fortnow and M. Sipser. Retraction of "Probabilistic computation and linear time". In *Proceedings of the 29th ACM Symposium on the Theory of Computing*, page 750. ACM, 1997.

[7] O. Goldreich, M. Sudan, and L. Trevisan. From logarithmic advice to single-bit advice. Technical Report TR-04-093, Electronic Colloquium on Computational Complexity, 2004.

[8] D. Grigoriev, E. Hirsch, and K. Pervyshev. Time hierarchies for cryptographic function inversion with advice. Technical Report TR-05-076, Electronic Colloquium on Computational Complexity, 2005.

[9] J. Hartmanis and R. Stearns. On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117:285–306, 1965.

[10] R. Karp and R. Lipton. Turing machines that take advice. *L'Enseignement Mathématique*, 28(2):191–209, 1982. A preliminary version appeared in STOC 1980.

[11] L. Levin. Universal search problems. *Problems of Information Transmission*, 9(3):265–266, 1973. In Russian.

[12] C. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and System Sciences*, 48:498–532, 1994.

[13] K. Pervyshev. Time hierarchies for computations with a bit of advice. Technical Report TR-05-054, Electronic Colloquium on Computational Complexity, 2005.

[14] R. Rettinger and R. Verbeek. Monte-Carlo polynomial versus linear time - the truth-table case. In *Proceedings of the 13th International Sympsoium on Fundamentals of Computation Theory*, pages 311–322. Springer-Verlag, 2001.

[15] J. Seiferas, M. Fischer, and A. Meyer. Separating nondeterministic time complexity classes. *Journal of the ACM*, 25:146–167, 1978.

[16] S. Žàk. A Turing machine time hierarchy. *Theoretical Computer Science*, 26:327–333, 1983.