NEW LIMITS TO CLASSICAL AND QUANTUM INSTANCE COMPRESSION∗

ANDREW DRUCKER†

Abstract. Given an instance of a hard decision problem, a limited goal is to compress that instance into a smaller, equivalent instance of a second problem. As one example, consider the problem where, given Boolean formulas ψ₁, . . . , ψ₄, we must determine if at least one ψₙ is satisfiable. An OR-compression scheme for SAT is a polynomial-time reduction R that maps (ψ₁, . . . , ψ₄) to a string z, such that z lies in some “target” language L′ if and only if ∨[ψₙ ∈ SAT] holds. (Here, L′ can be arbitrarily complex.) AND-compression schemes are defined similarly. A compression scheme is strong if |z| is polynomially bounded in n = maxₖ |ψₖ|, independent of t. Strong compression for SAT seems unlikely. Work of Harnik and Naor [SIAM J. Comput., 39 (2010), pp. 1667–1713] and Bodlaender, Downey, Fellows, and Hermelin [J. Comput. System Sci., 75 (2009), pp. 423–434] showed that the infeasibility of strong OR-compression for SAT would show limits to instance compression for a large number of natural problems. Bodlaender et al. also showed that the infeasibility of strong AND-compression for SAT would have consequences for a different list of problems. Motivated by this, Fortnow and Santhanam [J. Comput. System Sci., 77 (2011), pp. 91–106] showed that if SAT is strongly OR-compressible, then NP ⊆ coNP/poly. Finding similar evidence against AND-compression was left as an open question. We provide such evidence: we show that strong AND- or OR-compression for SAT would imply nonuniform, statistical zero-knowledge proofs for SAT—an even stronger and more unlikely consequence than NP ⊆ coNP/poly. Our method applies against probabilistic compression schemes of sufficient “quality” with respect to the reliability and compression amount (allowing for tradeoff). This greatly strengthens the evidence given by Fortnow and Santhanam against probabilistic OR-compression for SAT. We also give variants of these results for the analogous task of quantum instance compression, in which a polynomial-time quantum reduction must output a quantum state that, in an appropriate sense, “preserves the answer” to the input instance. The central idea in our proofs is to exploit the information bottleneck in an AND-compression scheme for a language L in order to fool a cheating prover in a proof system for L. Our key technical tool is a new method to “disguise” information being fed into a compressive mapping; we believe this method may find other applications.

1. Introduction.

1.1. Instance compression and parametrized problems. Given an instance of a hard decision problem, we may hope to compress that instance into a smaller, equivalent instance, either of the same or of a different decision problem. Here, we do not ask to be able to recover the original instance from the smaller instance; we only require that the new instance have the same (yes/no) answer as the original. Such instance compression may be the first step toward obtaining a solution; this has been a central technique in the theory of fixed-parameter-tractable algorithms [GN07, DF13]. Sufficiently strong compression schemes for certain problems would also have important implications for cryptography [HN10]. Finally, compressing an instance of

*Received by the editors July 1, 2013; accepted for publication (in revised form) April 14, 2015; published electronically October 27, 2015. This work was done in the EECS Department of MIT, appeared in Proceedings of FOCS, 2012, and formed part of the author’s Ph.D. dissertation. Supported by an NSF Career grant of Scott Aaronson.

http://www.siam.org/journals/sicomp/44-5/92711.html

†Department of Computer Science, University of Chicago, Chicago, IL 60637 (andy.drucker@gmail.com).
a difficult problem may also be a worthwhile goal in its own right, since it can make
the instance easier to store and communicate [HN10].

It is unknown whether one can efficiently, significantly compress an arbitrary in-
stance of a natural \NP-complete language like SAT – the set of satisfiable Boolean
formulas.\footnote{If we could efficiently reduce instances of some \NP-complete problem to shorter instances of the \emph{same} problem, then we could iterate the reduction to solve our problem in polynomial time, implying \P = \NP. However, even if \P \neq \NP, SAT might still have an efficient compressive reduction to a different target problem—say, to the Halting problem.} A more limited goal is to design an efficient reduction that achieves compre-
sion on instances that are particularly “simple” in some respect. To explore this
idea, one needs a formal model defining simple instances; the versatile framework of
\emph{parametrized problems} [DF13] is one such model and has been extensively used to
study instance compression. A parametrized problem is a decision problem in which
ev\text{ery instance has an associated \emph{parameter value} }k, giving some measure of the com-
plexity of a problem instance. For example, one can parametrize a Boolean formula
\psi by the number of distinct variables appearing in \psi.

An ambitious goal for a parametrized problem \(P\) is to compress an arbitrary
instance \(x\) of the decision problem for \(P\) into an equivalent instance \(x'\) of a second,
“target” decision problem, where the output length \(|x'|\) is bounded by a \emph{polynomial}
in \(k = k(x)\). If \(P\) has such a reduction running in time \text{poly}(|x| + k), we will say \(P\)
is \emph{strongly compressible}; we say \(P\) is \emph{strongly self-compressible} if the target problem
of the reduction is \(P\) itself. (In the literature of parametrized problems, a strong
self-compression reduction is usually referred to as a \emph{polynomial kernelization}. More
generally, a \emph{kernelization} is a polynomial-time self-compression reduction whose out-
put size is bounded by \emph{some} function of the parameter \(k\) alone.) All of these definitions
make sense both for deterministic and randomized compression algorithms, where in
the randomized case we specify some allowed probability of an incorrect output. For
randomized compression, we will require that the output-size bound be obeyed in
every outcome, although this is not essential.

1.2. Previous work: Results and motivation. Let \text{VAR-SAT denote the}
Satisfiability problem for Boolean formulas, parametrized by the number of distinct
variables in the formula. In their study of instance compression for \NP-hard problems,
Harnik and Naor [HN10] asked whether \text{VAR-SAT} is strongly compressible.\footnote{Technically, they asked a slightly different question whose equivalence to this one was pointed out in [FS11].} They
showed that a positive answer would have significant consequences for cryptography.
Notably, they showed that given an \emph{errorless} strong compression reduction for \text{VAR-
SAT} (with any target problem), we could construct collision-resistant hash functions
based on any one-way function—a long-sought goal.

In fact, Harnik and Naor showed that for their applications, it would suffice
to achieve strong compression for a different parametrized problem, the \emph{“OR(SAT)
problem,”} in which one is given a collection of Boolean formulas \(\psi_1, \ldots, \psi_t\) and asked
to determine whether at least one is satisfiable; the parameter used is the maximum
hit-length of any subformula \(\psi_j\). Strong compression for \text{VAR-SAT} easily implies
strong compression for \text{OR(SAT)}. Harnik and Naor defined a hierarchy of decision
problems called the “\text{VC hierarchy},” which can be modeled as a class of parametrized
problems (see [FS11]). They showed that a strong compression reduction for any of
the problems \emph{“above”} \text{OR(SAT)} in this hierarchy would also imply strong compression
for \text{OR(SAT)}; this includes parametrized versions of natural problems like the Clique
and Dominating Set problems. While Harnik and Naor’s main motivation was to find a strong compression scheme for OR(SAT) to use in their cryptographic applications, their work also gives a basis for showing negative results: by the reductions in [HN10], any evidence against strong compression for OR(SAT) is also evidence against strong compression for a variety of other parametrized problems.

In subsequent, independent work, Bodlaender et al. [BDFH09] also studied the compressibility of OR(SAT) and of related problems; these authors’ motivations came from the theory of fixed-parameter tractable (FPT) algorithms [DF13]. An FPT algorithm for a parametrized problem $P$ is an algorithm that solves an arbitrary instance $x$ with parameter $k = k(x)$ in time $g(k) \cdot \text{poly}(|x| + k)$ for some function $g(\cdot)$. The idea is that even if $P$ is hard in general, an FPT algorithm for $P$ may be practical on instances where the parameter $k$ is small. Now as long as $P$ is decidable, a kernelization reduction for $P$ provides the basis for an FPT algorithm for $P$: on input $x$, first compress $x$, then solve the equivalent, compressed instance. The kernelization approach is one of the most widely used schemas for developing FPT algorithms. For algorithmic efficiency, strong self-compression reductions are especially desirable.

Strong self-compression reductions are known for parametrized versions of many natural NP-complete problems, such as the Vertex Cover problem; see, e.g., the survey [GN07]. But for many other such parametrized problems, including numerous problems known to admit FPT algorithms (such as OR(SAT)), no strong compression reduction is known to any target problem. Bodlaender et al. [BDFH09] conjectured that no strong self-compression reduction exists for OR(SAT). They made the same negative conjecture for the analogous “AND(SAT) problem,” in which one is given Boolean formulas $\psi_1, \ldots, \psi_t$ and asked to decide whether $\bigwedge_{j=1}^t [\psi_j \in \text{SAT}]$ holds—that is, whether every $\psi_j$ is individually satisfiable. These conjectures are sometimes referred to as the “OR-” and “AND-conjectures.”

Bodlaender et al. showed that these conjectures would have considerable explanatory power. First, they showed [BDFH09, Theorem 1] that the nonexistence of strong self-compression reductions for OR(SAT) would rule out strong self-compression for a large number of other natural parametrized problems; these belong to a class we call “OR-expressive problems.” Under the assumption that AND(SAT) does not have strong self-compression, Bodlaender et al. ruled out strong self-compression reductions for a second substantial list of problems [BDFH09, Theorem 2], belonging to a class we will call “AND-expressive.” Despite the apparent similarity of OR(SAT) and AND(SAT), no equivalence between the compression tasks for these two problems is known.

In light of their results, Bodlaender et al. asked for complexity-theoretic evidence against strong self-compression for OR(SAT) and AND(SAT). Fortnow and Santhanam [FS11] provided the first such evidence: they showed that if OR(SAT) has a strong compression reduction (to any target problem), then $\text{NP} \subseteq \text{coNP}/\text{poly}$ and the Polynomial Hierarchy collapses to its third level.

The techniques of [BDFH09, FS11] were refined and extended by many researchers to give further evidence against efficient compression for various parametrized problems, e.g., in [DLS09, DvM10, BTY11, BJK11a, BJK11b, BJK11c, CFM11, HW12, 3In fact, every problem with an FPT algorithm is kernelizable [CCDF97]. However, the most efficient FPT algorithms need not arise from the kernelization approach.

4See section 3.2. The class of OR-expressive problems is not identical to the class described in [BDFH09], but it is closely related and contains their class, as well as other classes of problems identified in [HN10, BJK11a].
DM12, Kra12]. (See [DM12] for further discussion and references.) As one notable
development that is relevant to our work, Dell and Van Melkebeek [DvM10] combined
the techniques of [BDFH09, FS11] with new ideas to provide tight compression-size
lower bounds for certain problems that do admit polynomial kernelizations. Re-
searchers also used ideas from [BDFH09, FS11] in other areas of complexity, giving
new evidence of lower bounds for the length of PCPs [FS11, DvM10] and for the
density of NP-hard sets [BH08].

Finding evidence against strong compression for AND(SAT) was left as an open
question by these works, however. The limits of probabilistic compression schemes for
OR(SAT) and for OR-expressive problems (including VAR-SAT) also remained un-
clear. The results and techniques of [FS11] give evidence only against some restrictive
subclasses of probabilistic compression schemes for OR(SAT): schemes with one-sided
error, avoiding false negatives; schemes whose error probability is exponentially small
in the length of the entire input; and schemes using \(O(\log n)\) random bits, where
\(n = \max_j |\psi_j|\).

1.3. Our results.

1.3.1. Results on classical compression. We complement the results of [FS11]
by providing evidence against strong compression for AND(SAT): we prove that such
a compression scheme, to any target problem, would also imply \(NP \subseteq \text{coNP}/\text{poly}\). Our
proof technique applies naturally to the probabilistic setting, with two-sided error. We
show (in Theorem 5.4, item 1) that any sufficiently “high-quality” compression scheme
for AND(SAT) would imply \(NP \subseteq \text{coNP}/\text{poly}\). Here, quality is defined by a certain
relationship between the reliability and the compression amount of the reduction, and
it allows for tradeoff.

We also show (in Theorem 5.4, item 2, and Theorem 5.5) that beyond a second,
somewhat more demanding quality threshold, probabilistic compression reductions
either for AND(SAT) or for OR(SAT) would imply the existence of nonuniform, sta-
tistical zero-knowledge proofs for NP languages—a stronger (and even more unlikely)
consequence than \(NP \subseteq \text{coNP}/\text{poly}\). The more-demanding quality threshold in this
second set of results is still rather modest and allows us to derive (in section 5.3) the
following result as a special case.

**Theorem 1.1 (Informal).** Suppose that either AND(SAT) or OR(SAT) is
strongly compressible with success probability \(\geq 0.5 + 1/\text{poly}(n)\) for an AND or OR of
length-\(n\) formulas. Then there are nonuniform, statistical zero-knowledge proofs for
all languages in NP.

At the other extreme, where we consider compression schemes with more modest
compression amounts, but with greater reliability, our techniques yield the following
result.

**Theorem 1.2 (Informal).** Let \(t(n) : \mathbb{N}^+ \to \mathbb{N}^+\) be any polynomially bounded
function. Suppose there is a compression scheme compressing an AND of \(t(n)\) length-
\(n\) SAT instances into an instance \(z\) of a second decision problem \(L'\), where \(|z| \leq C \cdot t(n)\log t(n)\) for some \(C > 0\). If the scheme’s error probability on such inputs is
bounded by a sufficiently small inverse-polynomial in \(n\) (depending on \(t(n)\) and \(C\)),
then there are nonuniform, statistical zero-knowledge proofs for all languages in NP.
The corresponding result also holds for OR-compression.

Our results give the first strong evidence of hardness for compression of
AND(SAT). They also greatly strengthen the evidence given by Fortnow and San-
thanam against probabilistic compression for OR(SAT) and provide the first strong
evidence against probabilistic compression for VAR-SAT. For deterministic (or error-
free) compression of OR(SAT), the limits established by our techniques also follow
NEW LIMITS TO INSTANCE COMPRESSION

from the techniques of [FS11], which apply given an OR-compression scheme with compression bound of form $|z| \leq O(t(n) \log t(n))$.\footnote{This is not explicitly shown in [FS11] but follows from the technique of [FS11, Theorem 3.1]; see also [DvM10, Lemma 3] for a more general result that makes the achievable bounds clear.} On the other hand, we provide somewhat stronger complexity-theoretic evidence for these limits to compression.

Using our results on the infeasibility of compression for AND(SAT) and OR(SAT) and building on [HN10, BDFH09, FS11], we give new complexity-theoretic evidence against strong compressibility for a list of interesting parametrized problems with FPT algorithms (see Theorem 5.6). This is the first strong evidence against strong compressibility for any of the ten “AND-expressive” problems identified in [BDFH09] (and listed in section 3.2). For the numerous “OR-expressive” problems identified in [HN10, BDFH09] and other works, this strengthens the negative evidence given by [FS11].

Our methods also extend the known results on limits to compression for parametrized problems that do possess polynomial kernelizations: we can partially extend the results of Dell and Van Melkebeek [DvM10] to the case of probabilistic algorithms with two-sided error. For example, for $d > 1$ and any $\varepsilon > 0$, Dell and Van Melkebeek proved that if the SATisfiability problem for $N$-variable $d$-CNFs has a polynomial-time compression reduction with output-size bound $O(N^{d-\varepsilon})$, then $\text{NP} \subseteq \text{coNP}/\text{poly}$. Their result applies to co-nondeterministic reductions and to probabilistic reductions without false negatives; we prove (in Theorem 5.10) that the result also holds for probabilistic reductions with two-sided error, as long as the success probability of the reduction is at least $.5 + N^{-\beta}$ for some $\beta = \beta(d, \varepsilon) > 0$. Using reductions described in [DvM10], we also obtain quantitatively sharp limits to probabilistic compression for several other natural $\text{NP}$-complete problems, including the Vertex Cover and Clique problems on graphs and hypergraphs. (However, the limits we establish do not give lower bounds on the cost of oracle communication protocols; these protocols are a generalization of compression reductions, studied in [DvM10], to which that work’s results do apply. Extending our results to this model seems like an interesting challenge for further study.)

Our results about AND(SAT) and OR(SAT) follow from more general results about arbitrary languages. For any language $L$, we follow previous authors and consider the “OR($L$) problem,” in which one is given a collection $x^1, \ldots, x^t$ of strings and is asked to determine whether at least one of them is a member of $L$. We show (in Theorem 5.1, item 1) that if a sufficiently high-quality probabilistic polynomial-time compression reduction exists for the OR($L$) problem, then $L \in \text{NP}/\text{poly}$. As before, high-quality is defined by a relation between the reliability of the reduction and the compression amount.

We also prove (in Theorem 5.1, item 2) that a polynomial-time compression scheme for OR($L$) meeting a more demanding standard of quality implies that $L$ possesses nonuniform statistical zero-knowledge proof systems and therefore lies in $\text{NP}/\text{poly} \cap \text{coNP}/\text{poly}$. (For deterministic compression, the conclusion $L \in \text{coNP}/\text{poly}$ was established earlier in [FS11].) Applying these results to $L := \text{SAT}$ gives our hardness-of-compression results for AND(SAT); applying the second set of results to $L := \text{SAT}$ gives our improved negative results for OR(SAT).

We note that, as part of his subsequent work, Dell [Del14] has recently shown that the same quality of compression for the OR($L$) problem which we show to imply $L \in \text{NP}/\text{poly}$ also implies that $L \in \text{coNP}/\text{poly}$. Using this, he establishes limits of achievable probabilistic compression for OR(SAT) that are just as strong as those we
show for AND(SAT), under the assumption \( \text{NP} \not\subseteq \text{coNP/poly} \). Dell’s improvement can be obtained directly from our work by plugging in a known result on the complexity of the Statistical Difference problem that was unknown to us during our research but used by Dell; see Theorem 2.16 and section 5.1.

In an unpublished work, Buhrman [Buh] constructed an oracle \( A \) such that, for every \( \text{NP}^A \)-complete language \( L \), the decision problem \( \text{AND}(L) \) does not have a \( \text{P}^A \)-computable strong compression reduction. This gave earlier, indirect evidence against efficient strong compression for the \( \text{AND} \)(SAT) problem—or at least it indicated that exhibiting such a compression reduction would require novel techniques. Now, inspection of the proofs reveals that our new results on compression for \( \text{OR}(L) \) are all perfectly relativizing. This allows us to identify many more oracles obeying the property of Buhrman’s oracle: namely, we may take any \( A \) for which \( \text{NP}^A \not\subseteq \text{coNP}^A/\text{poly} \).

For example, this holds with probability 1 for a random oracle \[\text{BG81}\].

Such an oracle can also be obtained through a simple diagonalization argument.

1.3.2. Results on quantum compression. Up to this point, we have discussed compression reductions in which the input and output are both “classical” bit-strings. However, from the perspective of quantum computing and quantum information [NC00], it is natural to ask about the power of compression reductions that output a quantum state. A “\( n \)-qubit state” is a quantum superposition over classical \( n \)-bit strings; a vast body of research has explored the extent to which information can be succinctly encoded within and retrieved from such quantum states. If quantum computers become a practical reality, quantum instance compression schemes could help to store and transmit hard computational problems; compressing an instance might also be a first step toward its solution by a quantum algorithm.

We propose the following quantum generalization of classical instance compression: a quantum compression reduction for a language \( L \) is a quantum algorithm that, on input \( x \), outputs a quantum state \( \rho \) on some number \( q \) of qubits—hopefully with \( q \ll |x| \), to achieve significant compression. Our correctness requirement is that there should exist some fixed quantum measurement \( \mathcal{M}_q \) on \( q \)-qubit states for each \( q > 0 \), such that \( \mathcal{M}_q(\rho) = L(x) \) holds with high probability over the inherent randomness in the measurement \( \mathcal{M}_q(\rho) \). We do not require that \( \mathcal{M}_q \) be an efficiently performable measurement; this is by analogy to the general version of the classical compression task, in which the target language of the reduction may be arbitrarily complex.

Our results for quantum compression are closely analogous to our results in the classical case. First, we show that for any language \( L \), if a sufficiently high-quality quantum polynomial-time compression reduction exists for the \( \text{OR}(L) \) problem, then \( L \) possesses a nonuniform, 2-message quantum interactive proof system (with a single prover). Second, we show that a sufficiently higher-quality quantum polynomial-time compression reduction for \( \text{OR}(L) \) implies that \( L \) possesses a nonuniform quantum statistical zero-knowledge proof system. In fact, the two “quality thresholds” in our quantum results are essentially the same as in the corresponding results for the classical case.\(^7\)

It follows that, unless there exist surprisingly powerful quantum

\(^6\)In [BG81], it is shown that \( \text{NP}^A \not\subseteq \text{coNP}^A \) for random \( A \); the technique readily extends to give the stronger claim above.

\(^7\)We do place a minor additional restriction on quantum compression reductions for \( \text{OR}(L) \): we require that the reduction, on input \( (x^1, \ldots, x^t) \), outputs a quantum state of size determined by \( (\max_j |x^j|) \) and \( t \).
NEW LIMITS TO INSTANCE COMPRESSION

proofs of unsatisfiability for Boolean formulas, the limits we establish for probabilistic compression of AND(SAT) and OR(SAT) hold just as strongly for quantum compression.\(^8\)

1.4. Our techniques.

1.4.1. The overall approach. We first describe our techniques for the classical case; these form the basis for the quantum case as well. Our first two general results, giving complexity upper bounds on any language \(L\) for which OR(\(L\)) has a sufficiently high-quality compression reduction (Theorem 5.1, items 1 and 2), are both based on a single reduction that we describe next. This reduction applies to compression reductions mapping some number \(t(n) \leq \text{poly}(n)\) of inputs of length \(n\) to an output string \(z\) of length \(|z| = O(t(n) \log t(n))\).

Fix any language \(L\) such that OR(\(L\)) has a possibly-probabilistic compression reduction

\[ R(x^1, \ldots, x^t) : \{0, 1\}^{t \times n} \rightarrow \{0, 1\}^{t'} \]

with some target language \(L'\), along with parameters \(t', t\) satisfying \(t' \leq O(t \log t) \leq \text{poly}(n)\).\(^9\) We will use \(R\) to derive upper bounds on the complexity of \(L\). (The reader may keep in mind the main intended setting \(L = \text{SAT}\), which we will use to derive our hardness results for the compression of AND(SAT). No special properties of this language will be used in the argument, however.)

A simple, motivating observation is that if we take a string \(y \in L\) and “insert” it into a tuple \(\overline{x} = (x^1, \ldots, x^t)\) of elements of \(\overline{L}\), replacing some \(x^j\) to yield a modified tuple \(\overline{x}'\), then the values \(R(\overline{x}), R(\overline{x}')\) are different with high probability—for, by the “OR-respecting” property of \(R\), we will with high probability have \(R(\overline{x}) \in \overline{L}'\), \(R(\overline{x}') \in L'\). More generally, for any distribution \(D\) over \(t\)-tuples of inputs from \(\overline{L}\), let \(D[y, j]\) denote the distribution obtained by sampling \(\overline{x} \sim D\) and replacing \(x^j\) with \(y\); then the two output distributions

\[ R(D), R(D[y, j]) \]

are far apart in statistical distance. (Of course, the strength of the statistical-distance lower bound we get will depend on the reliability of our compression scheme.)

We want this property to serve as the basis for an interactive proof system by which a computationally powerful prover can convince a skeptical polynomial-time (but nonuniform) verifier that a string \(y\) lies in \(L\). The idea for our initial, randomized protocol (which can then be derandomized by standard techniques) is that the prover will make his case by demonstrating his ability to distinguish between the two \(R\)-output distributions described above, when the verifier privately chooses one of the two distributions, samples from it, and sends the sample to the prover.\(^10\) But then to make our proof system meaningful, the verifier also needs to fool a cheating prover in the case \(y \notin L\). To do this, we want to choose \(D, j\) in such a way that the distributions \(R(D), R(D[y, j])\) are as close as possible whenever \(y \notin L\).

---

\(^8\)We remark that 3-message quantum interactive proofs are known to be fully as powerful as quantum interactive proofs in which polynomially many messages are exchanged [Wat03] and that these proof systems are equal in power to PSPACE in the uniform setting [JJuW11]. However, 2-message quantum proof systems seem much weaker and are not known to contain coNP.

\(^9\)Here, we pay exclusive attention to \(R\)’s behavior on tuples of strings of some equal length \(n\).

\(^10\)Interactive proofs based on distinguishing tasks have seen many uses in theoretical computer science, and indeed we will rely upon known protocols of this kind in our work; see section 2.4.
We may not be able to achieve this for an index \( j \) that is poorly chosen. For instance, \( R(\pi) \) may always copy the first component \( x^1 \) as part of the output string \( z \), so taking \( j = 1 \) would fail badly. To get around this, we choose our replacement index \( j \) uniformly at random, aiming in this way to make \( R \) “insensitive” to the insertion of \( y \).\(^{11}\) As \( R \) is a compression scheme, it doesn’t have room in its output string to replicate its entire input, so there is reason for hope.

This invites us to search for a distribution \( D^* \) over \((\mathcal{L}_n)^t\) with the following properties:

(i) For every \( y \in \mathcal{L}_n \), if we select \( j \in [t] \) uniformly, then the expected statistical distance \( \mathbb{E}_j [||R(D^*) - R(D^*[y,j])||_{\text{stat}}] \) is “not too large.”\(^{12}\)

(ii) \( D^* \) is efficiently sampleable, given nonuniform advice of length \( \text{poly}(n) \).

Condition (i) is quite demanding: we need a single distribution \( D^* \) rendering \( R \) insensitive to the insertion of \textit{any} string \( y \in \mathcal{L}_n \)—a set which may be of exponential size. Condition (ii) is also a strong restriction: \( \mathcal{L}_n \) may be a complicated set, and in general we can only hope to sample from distributions over \((\mathcal{L}_n)^t\) in which \( t \)-tuples are formed out of a fixed “stockpile” of \( \text{poly}(n) \) elements of \( \mathcal{L}_n \), hard-coded into the nonuniform advice.

Remarkably, it turns out that such a distribution \( D^* \) can always be found. In fact, in item (i), we can force the two distributions to be nonnegligible close (with expected statistical distance \( \leq 1 - \frac{1}{\text{poly}(n)} \)) whenever the output-size bound \( t' \) obeyed by \( R \) is \( O(t \log t) \); the distributions will be much closer when \( t' \ll t \). We call our key technical result (Lemma 4.6), guaranteeing the existence of such a \( D^* \), the “Disguising-Distribution Lemma.”

Assuming this lemma for the moment, we can use \( D^* \) as above to reduce any membership claim for \( L \) to a distinguishing task for a prover-verifier protocol. Given any input \( y \), we’ve constructed two distributions \( \mathcal{R} = R(D^*) \) and \( \mathcal{R}' = R(D^*[y,j]) \) (with \( j \) uniform), where each distribution is sampleable in nonuniform polynomial time. Our analysis guarantees some lower bound \( D = D(n) \) on \( ||\mathcal{R} - \mathcal{R}'||_{\text{stat}} \) in the case \( y \in L \) and some upper bound \( d = d(N) \) on this distance when \( y \notin L \). (These parameters depend on the reliability and compression guarantees of \( R \).)

If \( D(n) - d(n) \geq \frac{1}{\text{poly}(n)} \), then the prover can demonstrate that the input \( y \) is in \( L \) by a distinguishing protocol as described above. (Formally, our proof establishes a reduction from the membership question for \( L \) to the so-called Statistical Difference promise problem, for which distinguishing protocols are already known. Thus, we will not need to directly define any prover-verifier protocol. In his subsequent work, Dell [Del14] gives a different, but related, reduction from OR-\( (L) \)-compression to the Statistical Difference problem.) By known results, this distinguishing protocol can be converted to public-coin protocols and then nonuniformly derandomized to show that \( L \in \text{NP}/\text{poly} \). Also, if \( D(n)^2 - d(n) \geq \frac{1}{\text{poly}(n)} \), then, using a powerful result due to Sahai and Vadhan [SV03], we can derive a nonuniform, statistical zero-knowledge proof system for \( L \). This also implies \( L \in \text{NP}/\text{poly} \cap \text{coNP}/\text{poly} \).

\(^{11}\)We emphasize that the “insensitivity” we are looking for is \textit{statistical}; we are not asking that \( y \) have a small effect on the output of \( R \) for most \textit{particular} outcomes to \( \pi \sim D \). This latter goal may not be achievable, e.g., if \( R \) outputs the sum of all its input strings \( x^i \) taken as vectors over \( \mathbb{F}_2^t \).

\(^{12}\)For our purposes, it actually suffices to bound \( ||R(D^*) - R(D^*[y,j])||_{\text{stat}} \), where \( j \) is a uniform value sampled “internally” as part of the distribution. However, our techniques will yield the stronger property in condition (i) above, and this is the course we will follow in proving our general results.
1.4.2. The Disguising-Distribution lemma. The Disguising-Distribution lemma, informally described in section 1.4.1, is a statement about the behavior of $R(x^1, \ldots, x^t)$ on a specified product subset $S^t$ of inputs ($S = \mathcal{T}_n$ in our application). This lemma is a “generic” result about the behavior of compressive mappings; it uses no properties of $R$ other than $R$’s output-size bound. In view of its generality and interest, we are hopeful that the lemma will find other applications.

Our proof of this lemma uses two central ideas. First, we interpret the search for the disguising distribution $D^*$ as a two-player game between a disguising player (choosing $D^*$) and an opponent who chooses $y$; we can then apply simple yet powerful principles of game theory. Second, to build a winning strategy for the disguising player, we will exploit an information bottleneck in $R$ stemming from its compressive property.\[13\]

To describe the proof, it is helpful to first understand how one may obtain the distribution $D^*$ if we drop the efficient-sampleability requirement on $D^*$ and focus on the disguising requirement (condition (i)). To build $D^*$ in this relaxed setting, we will appeal to the minimax theorem for two-player, zero-sum games; applied here, it tells us that to guarantee the existence of a $D^*$ that succeeds in disguising all strings $y \in S$, it is enough to show how to build a $D^*_Y$ that succeeds in expectation, when $y$ is sampled from some fixed (but arbitrary) distribution $Y$ over $S$.

Here, a natural idea springs to mind: let $D^*_Y$ just be a product distribution over $t$ copies of $Y$! In this case, inserting $y \sim Y$ into $D^*_Y$ at a random location is equivalent to conditioning on the outcome of a randomly chosen coordinate of a sample from $D^*_Y$. The intuition here is that, due to the output-size bound on $R$, the distribution $R(D^*_Y)$ shouldn’t have enough “degrees of freedom” to be affected much by this conditioning. Indeed, we show (in Lemma 4.2) that if $\mathcal{P} = (x^1, \ldots, x^t) \sim D^*_Y$, then conditioning on the value of $x^j$ for a uniformly chosen index $j \in [t]$ has bounded expected effect on the output distribution $R(\mathcal{P})$. More precisely, the expected statistical distance between the pre- and post-conditioned distributions is bounded nonnegligibly away from 1 (provided that $t' \leq O(t \log t)$). We refer to this important property of $R$ as “distributional stability.”

Our original proof of the distributional property used a coding-theoretic argument and Fano’s inequality. Several researchers suggested an alternative proof using Kullback–Leibler (KL) divergence and an inequality due to Pinsker. This gives slightly better bounds than our original proof when $t' \leq t$. Also, by using an inequality of Vajda’s in place of Pinsker, this approach allows us to handle values of $t'$ as large as $O(t \log t)$ in a simpler way. Thus, this is the approach we will follow here.

These researchers also pointed out that the distributional stability property can also be established using other similar, known results that also follow from divergence-based techniques: a lemma of Raz [Raz98] and the “average encoding theorem” of Klauck et al. [KNTSZ07]. The latter was used in [KNTSZ07] to identify a stability property for trace and Hellinger distance metrics, for the inputs to a problem in quantum communication complexity; this was used for a round-elimination argument.

\[13\]This is hardly the first paper in which such a bottleneck plays a crucial and somewhat unexpected role. For example, an interesting and slightly similar application of information-theoretic tools to the study of metric embeddings was found recently by Regev [Reg13].
Their proof is for inputs drawn from the uniform distribution but extends readily to general distributions and can be used to derive the kind of lemma we need.\textsuperscript{14, 15}

Using the distributional-stability property of compressive mappings under product input distributions, we then establish a certain “sparsified variant” of this property (Lemma 4.3), which allows us to replace the set $Y$ above with a small set sampled from $Y$; this is an important tool in addressing the efficient-sampleability requirement on our desired $D^*$. From this variant, we use the minimax theorem to derive (in Lemma 4.4) a distribution $\mathcal{D}$ over product input-distributions to $R$—with each product distribution defined over small subsets of $S$—such that, in expectation, $\mathcal{D}$ disguises the random insertion of any string $y \in S$ at a uniformly chosen position $j$. Finally, in Lemma 4.6 we obtain our desired disguising distribution $D^*$ as a sparsified version of $\mathcal{D}$, using a result due to Lipton and Young [LY94] and, independently, to Althöfer [Alt94], that guarantees the existence of sparsely supported, nearly optimal strategies in two-player, zero-sum games.

1.4.3. Extension to the quantum case. Our techniques for studying quantum compression are closely analogous to the classical case. The main technical difference is that the output $R(D)$ of our compression reduction, on any input distribution $D$, is now a (mixed) quantum state. In this setting, to carry out an analogue of the argument sketched in sections 1.4.1 and 1.4.2 and fool a cheating prover, we need a disguising distribution for $R$ that meets a modified version of condition (i) from section 1.4.1:

(i') For every $y \in \mathcal{T}_n$, if we select $j \in [t]$ uniformly, then for any quantum measurement $\mathcal{M}$, the expected statistical distance $\mathbb{E}_j[||\mathcal{M}(R(D^*)) - \mathcal{M}(R(D'[y,j]))||_{\text{stat}}]$ is not too large.

A basic measure of distance between quantum states, the trace distance, is relevant here: if two states $\rho, \rho'$ are at trace distance $||\rho - \rho'||_\text{tr} \leq \delta$, then for any measurement $\mathcal{M}$, the statistical distance $||\mathcal{M}(\rho) - \mathcal{M}(\rho')||_{\text{stat}}$ is at most $\delta$. (In fact, this property characterizes the trace distance.) Thus, to satisfy condition (i'), it will be enough to construct $D^*$ so as to upper-bound $\mathbb{E}_j[||R(D^*) - R(D'[y,j])||_{\text{tr}}]$ for uniformly chosen $j$. We do this by essentially the same techniques as in the classical case.\textsuperscript{16} The one significant difference is that here we need to establish a stability property for trace distance, analogous to the stability property for statistical distance described in

\textsuperscript{14} This is true both for classical distributional stability and for the quantum variant that we will discuss. Maneva [Man01] applies the techniques of [KNTSZ07] to round reduction in classical communication, showing distributional stability for uniform input-tuples. Sen and Venkatesh [SV08] approach round elimination using a variant of the distributional stability property, in which multiple input coordinates are first set to fixed strings; they show a useful property for all input distributions, yielding round elimination under an arbitrary distribution on inputs. Applying the minimax theorem, they obtain a round-elimination property for randomized protocols with worst-case error bounds. By making small changes to their argument (in particular, by retaining the randomization in their distributional algorithms and applying the minimax theorem to the statistical distance bound in their stability property, rather than to the protocol error), one can obtain an alternative kind of disguising distribution that does not strictly obey property (i) described earlier but something similar that is as useful for our purposes. Our important efficient-sampleability property (ii) requires further ideas, to be discussed shortly, that have no analogue in [SV08].

\textsuperscript{15} Russell Impagliazo suggested the use of Raz’s lemma; Salil Vadhan also helped me to understand the connection. Ashwin Nayak and S. Vadhan suggested direct proofs of distributional stability based on divergence and Pinsker’s inequality. Dieter van Melkebeek also suggested the relevance of Pinsker’s inequality, and James Lee and Avi Wigderson suggested finding a more direct information-theoretic proof.

\textsuperscript{16} Again, we follow similar steps as Klauck et al. [KNTSZ07], who proved an instance of this property in the setting of quantum communication protocols.
section 1.4.2. This can be obtained using the same basic divergence-based techniques as in the classical case, with the help of suitable tools from quantum information theory.

1.5. Organization of the paper. In section 2, we present definitions and facts that will be used throughout our work. In section 3, we formally introduce parametrized problems and AND- and OR-expressive problems. In section 4, we prove the main technical lemmas we use to obtain our results on limits of efficient instance compression. Our results for the classical setting are proved in section 5, and the quantum results are proved in section 6 (along with some needed quantum background). Finally, in section 7 we present questions for future study.

2. Preliminaries.

2.1. Statistical distance and distinguishability. All distributions in this paper will take finitely many values; let supp(\(D\)) be the set of values assumed by \(D\) with nonzero probability, and let \(D(u) := \Pr[D = u]\). For a probability distribution \(D\) and \(t \geq 1\), we let \(D^\otimes t\) denote a \(t\)-tuple of outputs sampled independently from \(D\). We let \(\mathcal{U}_K\) denote the uniform distribution over a multiset \(K\).

The statistical distance of two distributions \(D, D'\) over a shared universe of outcomes is defined as \(||D - D'||_{\text{stat}} := \frac{1}{2} \sum_{u \in \text{supp}(D) \cup \text{supp}(D')} ||D(u) - D'(u)||\). The statistical distance between random variables is defined as the statistical distance between their governing distributions. We will use the following familiar distinguishability interpretation of the statistical distance. Suppose a value \(b \in \{0, 1\}\) is selected uniformly, unknown to us, and a sample \(u \in U\) is drawn from \(D\) if \(b = 0\) or from \(D'\) if \(b = 1\). We observe \(u\), and our goal is to correctly guess \(b\). It is a basic fact that, for any \(D, D'\), our maximum achievable success probability in this distinguishing experiment is precisely \(\frac{1}{2}(1 + ||D - D'||_{\text{stat}})\).

We will also use the following easy facts.

**Fact 2.1.** If \(X, Y\) are random variables taking values in some set \(S\), and \(R(X)\) is any (possibly randomized) function whose domain contains \(S\), then \(||R(X) - R(Y)||_{\text{stat}} \leq ||X - Y||_{\text{stat}}\).

**Fact 2.2.** (see [SV03, Fact 2.3]). Let \((X_1, X_2, Y_1, Y_2)\) be random variables such that \(X_1\) is independent of \(X_2\) and that \(Y_1\) is independent of \(Y_2\). Then, \(||(X_1, X_2) - (Y_1, Y_2)||_{\text{stat}} \leq ||X_1 - Y_1||_{\text{stat}} + ||X_2 - Y_2||_{\text{stat}}\).

2.2. Information theory background. For a finitely supported random variable \(Z\), we let \(H(Z) := \sum_z \in \text{supp}(Z) \Pr[Z = z] \log_2 \Pr[Z = z]\) denote the Shannon entropy of \(Z\). Then, for two possibly dependent random variables \(Y, Z\), \(H(Z|Y) := \mathbb{E}_{y \sim Y}[H(Z|Y=y)] = H((Y,Z)) - H(Y)\) is called the entropy of \(Z\) conditional on \(Y\). (For \(y \in \text{supp}(Y)\), \(Z|Y=y\) denotes \(Z\) conditioned on the event \(Y = y\).)

**Fact 2.3.** For all \(X, Y\), \(H((X,Y)) \leq H(X) + H(Y)\) and \(H(X|Y) \leq H(X)\) with equality holding in each case if \(X, Y\) are independent. Similarly, \(H(X|(Y,Z)) \leq H(X|Y)\).

**Definition 2.4** (Mutual information). The mutual information between random variables \(X, Y\) is defined as \(I(X;Y) := H(X) + H(Y) - H((X,Y))\).

The next fact follows easily from the definitions.

**Fact 2.5.** Mutual information obeys the following properties for all random variables \(X, Y, Z\):

1. \(I(X; Y) = I(Y; X)\).
2. \(I(X; (Y, Z)) = I(X; Y) + I((X, Y); Z) - I(Y; Z)\).
3. \(I(X; (Y, Z)) \geq I(X; Y)\).
4. \(I(X; Z) = 0\) if \(X, Z\) are independent.

**Lemma 2.6.** If \(X^1, \ldots, X^t\) are independent, then \(I(Y; (X^1, \ldots, X^t)) \geq \sum_{j \in [t]} I(Y; X^j)\).

Our proof of this standard claim follows steps in [Nay99, p. 33].

**Proof.** We have

\[
I(Y; (X^1, \ldots, X^t)) = I(Y; X^t) + I((Y, X^t); (X^1, \ldots, X^{t-1})) - I(X^t; (X^1, \ldots, X^{t-1})) = 0, \text{ by Fact 2.5, item 4}
\]

where we used item 2 of Fact 2.5 in the first step and items 1 and 3 in the second step. Iterating in this way gives the lemma. \(\Box\)

The next definition is a useful, nonsymmetric measure of difference between random variables.

**Definition 2.7 (KL divergence).** The (binary) KL divergence, or KL divergence between random variables \(X, Y\), is denoted \(D_{KL}(X||Y)\) and defined as

\[
D_{KL}(X||Y) := \sum_{x \in \text{supp}(X)} \Pr[X = x] \cdot \log_2 \left( \frac{\Pr[X = x]}{\Pr[Y = x]} \right).
\]

The convention is that for \(p \neq 0\), we have \(p \log_2(p/0) = +\infty\). So \(D_{KL}\) may be infinite. We have the following basic equivalence (see [CT06, Chapter 2]).

**Fact 2.8.** Let \(X, Y\) be two random variables in a common probability space; let \(X'\) be identically distributed to \(X\) and independent of \(Y\). We have \(I(X; Y) = D_{KL}((X, Y)||(X', Y))\).

A proof of the following important result can be found in [CT06, Lemma 11.6.1, p. 370].

**Theorem 2.9** (Pinsker’s inequality, stated for binary KL divergence). For any random variables \(Z, Z'\), \(D_{KL}(Z||Z') \geq (2/\ln 2) \cdot ||Z - Z'||_{\text{stat}}^2\).

In particular, \(D_{KL}(Z||Z')\) is always nonnegative. When \(||Z - Z'||_{\text{stat}} \approx 1\), the following bound, known as Vajda’s inequality (see [FHT03, RW09]), gives a better bound on the divergence.

**Theorem 2.10** (Vajda’s inequality, stated for binary KL divergence). For any random variables \(Z, Z'\), let \(\Delta := ||Z - Z'||_{\text{stat}}\). Then,

\[
D_{KL}(Z||Z') \geq \frac{1}{\ln 2} \left( \ln \left( \frac{1 + \Delta}{1 - \Delta} \right) - \frac{2\Delta}{1 + \Delta} \right) \geq \frac{1}{\ln 2} \left( \ln \left( \frac{1}{1 - \Delta} \right) - 1 \right).
\]

**2.3. Basic complexity classes and promise problems.** We will assume familiarity with basic concepts in complexity, such as languages and promise problems. For the needed background in complexity theory, see [AB09]. For a language \(L \subseteq \{0, 1\}^*\), we will use \(L(x)\) to denote the characteristic function of \(L\) evaluated on \(x\), i.e., \(L(x) := 1\) if \(x \in L\), otherwise \(L(x) := 0\). We assume that the reader is familiar with the basic complexity classes \(\text{NP}\) and \(\text{coNP}\) and the higher levels \(\Sigma_k^p, \Pi_k^p\) of the polynomial hierarchy \(\text{PH}\). In this paper, we define \(\text{NP, coNP, etc. as classes of languages (not promise problems)}\).

We also assume familiarity with the general model of polynomial-size, nonuniform advice and with the nonuniform classes \(\text{NP/poly and coNP/poly}\). It is considered
unlikely that $\text{NP} \subseteq \text{coNP}/\text{poly}$. In particular, this would imply a collapse of the polynomial hierarchy:

**THEOREM 2.11** (see [Yap83]). If $\text{NP} \subseteq \text{coNP}/\text{poly}$, then $\text{PH} = \Sigma^p_3 = \Pi^p_3$.

We use $\text{pr-NP}$, $\text{pr-coNP}$, etc. to denote the analogous complexity classes for promise problems. Recall that, for a class $C$ of promise problems, $\text{co}C = \{(\Pi^r_Y, \Pi^r_N) : (\Pi^r_N, \Pi^r_Y) \in C\}$. A *many-to-one* reduction $B$ from the promise problem $\Pi = (\Pi^r_Y, \Pi^r_N)$ to $\Pi' = (\Pi'^r_Y, \Pi'^r_N)$ is a mapping $B$ satisfying $B(\Pi^r_Y) \subseteq \Pi'^r_Y$, $B(\Pi^r_N) \subseteq \Pi'^r_N$. This definition applies as well to the special case where one or both of the promise problems are languages. When we refer to $\text{NP}$-complete languages in this paper, we mean languages complete for $\text{NP}$ under deterministic, polynomial-time, many-to-one reducibility.

All of the results we prove in this paper about limits of compression for languages $L$ and language complexity classes readily extend to the setting of compression for promise problems (under the analogous definitions). However, for notational simplicity, we will state our main results for languages and will only use promise problems and promise classes where doing so helps to streamline our proofs and our result statements.

We also assume familiarity with the model of *Arthur–Merlin protocols* [BM88, AB09]. Recall that these are two-message protocols in which Arthur’s random bits are public (visible to Merlin) and Arthur sends the first message. We let $\text{pr-AM}_{c(n), s(n)}$ denote the class of promise problems definable by a polynomial-time Arthur–Merlin protocol with completeness $c(n)$ and soundness $s(n)$; let $\text{pr-AM} := \text{pr-AM}_{1, 1/3}$. Then, $\text{pr-coAM} = \{(\Pi_Y, \Pi_N) : (\Pi_N, \Pi_Y) \in \text{pr-AM}\}$. The following standard result follows from the nonuniform derandomization technique of Adleman [Adl78].

**THEOREM 2.12.** $\text{pr-AM} \subseteq \text{pr-NP}/\text{poly}$. Similarly, $\text{pr-coAM} \subseteq \text{pr-coNP}/\text{poly}$.

### 2.4. Statistical zero-knowledge and the SD problem

Next, we will define the *statistical zero-knowledge* class $\text{SZK}$. Actually, we will only work with its promise-problem analogue $\text{pr-SZK}$.

Informally, these are the promise problems $(\Pi^r_Y, \Pi^r_N)$ for which a (private-coin) interactive proof of membership in $\Pi^r_Y$ can be given, in which the verifier *learns (almost) nothing—except* to become convinced that the input $y$ indeed lies in $\Pi^r_Y$! The “learns nothing” requirement is cashed out by requiring that the verifier be able to *simulate* interactions with the intended prover strategy on any input $y$, such that if $y \in \Pi^r_Y$, the resulting distribution is negligibly close in statistical distance to the true distribution generated by their interaction.

Making this definition formal is somewhat delicate. For details, and for more information on these and related classes, see [SV03]. Fortunately, there is a simple (but nontrivial) alternative characterization of $\text{pr-SZK}$. First, given a Boolean circuit $C = C(r)$ with $k$ output gates, and an ordering on these gates, let $D_C$ denote the output distribution of $C$ on a uniformly random input $r$. (This is a random variable over $\{0, 1\}^k_\text{.)}$ We use the following problem.

**DEFINITION 2.13** (The Statistical Difference problem). For parameters $0 \leq d < D \leq 1$, define the promise problem $\text{SD}^D_d = (\Pi^r_Y, \Pi^r_N)$ as follows:

$$\Pi^r_Y := \{(C, C') : ||D_C - D_{C'}||_{\text{stat}} \geq D\}, \quad \Pi^r_N := \{(C, C') : ||D_C - D_{C'}||_{\text{stat}} \leq d\}.$$

Define $\text{SD}^D_d$ analogously, switching the “yes” and “no” cases. In this definition, both $d = \bar{d}(n)$ and $D = D(n)$ may be parameters depending on the input length $n = |(C, C')|$.  

---

17 Often the promise class is denoted $\text{SZK}$. 

It is shown in [SV03] that the standard, complicated definition of \( \text{pr-SZK} \) is equivalent to the following simpler one, which we take as our definition.

**Definition 2.14.** Let \( \text{pr-SZK} \) be defined as the class of promise problems for which there is a many-to-one,\(^{18}\) deterministic polynomial-time reduction from \( \Pi \) to \( \text{SD}_{\leq}^{2/3} \).

The constants \( 2/3, 1/3 \) in the above definition are not arbitrary; it is unknown whether we get the same class if we replace them by \( .51, .49 \). However, we have the following result.

**Theorem 2.15 (Follows from [SV03]; described as Theorem 1 in [GV11]).** Suppose that \( 0 \leq d = d(n) < D = D(n) \leq 1 \) are polynomial-time computable and satisfy \( D^2 > d + \frac{1}{\text{poly}(n)} \). Then, \( \text{SD}_{\leq}^{D} \in \text{pr-SZK} \).

If we keep the gap requirement but do not require that \( d(n), D(n) \) are polynomial-time computable, there is a many-to-one nonuniform polynomial time reduction from \( \text{SD}_{\leq}^{D} \) to a problem in \( \text{pr-SZK} \).

The nonuniform part of the result is proved by the same techniques as the uniform claim. When we merely have \( D - d \geq \frac{1}{\text{poly}(n)} \), the following weaker, standard result holds.

**Theorem 2.16.** Suppose that \( 0 \leq d = d(n) < D = D(n) \leq 1 \) are polynomial-time computable and satisfy \( D > d + \frac{1}{\text{poly}(n)} \). Then, \( \text{SD}_{\leq}^{D} \in \text{AM} \cap \text{co-AM} \).

If the gap holds but we do not require that \( d(n), D(n) \) are polynomial-time computable, there is a many-to-one reduction, computable in nonuniform polynomial time, from \( \text{SD}_{\leq}^{D} \) to a problem in \( \text{AM} \cap \text{co-AM} \).

The uniform part of Theorem 2.16 is proved in [Xia09, Lemma A.4.1, p. 144], which notes that it can be derived from results of [GVW02]. The nonuniform part of the claim follows by the same techniques. The membership in \( \text{co-AM} \) in Theorem 2.16 was unknown to this author during our research and will not be used in our stated theorems (although we will point out its applicability). This fact was used by Dell to extend our results [Del14].

From Theorems 2.15 and 2.16, one can observe that \( \text{pr-SZK} \subseteq \text{AM} \cap \text{co-AM} \), but this was known before Theorem 2.16 was proved. The containment \( \text{pr-SZK} \subseteq \text{co-AM} \) is due to Fortnow [For87]; the fact \( \text{pr-SZK} \subseteq \text{AM} \) was first shown by Aiello and H˚astad [AH91].\(^{19}\) It was proved by Okamoto that \( \text{pr-SZK} \) is closed under complement [Oka00].

**2.5. \( f \)-compression reductions.** Here, we define a class of compression reductions for the problem in which one is given \( (x^1, \ldots, x^m) \) and must compute \( f(L(x^1), \ldots, L(x^m)) \), for some Boolean “combining function” \( f \). Our focus will be the case where \( f \) is the OR or AND function of its input bits. This is closely related to the parametrized problem \( f \circ L \) that will be formally defined in section 3.1, but the definition below doesn’t need to refer to that definition.

Our definition is modeled on ones in [BDFH09, FS11], with some differences. Notably, we will consider reductions where a quantitative compression guarantee is only made when all the input strings \( x^j \) are of some equal length \( n \), and the number of input strings \( x^j \) is equal to some value \( t_1(n) \) determined by \( n \). The error bound will

\(^{18}\)Recall the definition in section 2.3.

\(^{19}\)These works treat language classes, but the proofs extend without change to the promise-problem setting. Also, these works analyze a so-called honest-verifier model of statistical zero-knowledge proofs; these were shown to have the same expressive power as “cheating-verifier” statistical zero-knowledge proofs in [GSV98].
also be a function of $n$. This specialization is mostly to reduce clutter in our work and will not lead to loss of generality: we will be ruling out the existence of compression reductions (under complexity-theoretic assumptions and for all polynomially bounded $t_1(n)$ that are sufficiently large compared to other parameters), so ruling out even compression algorithms that work only in narrow input-regimes will lead to stronger results.

**Definition 2.17 (Probabilistic $f$-compression reductions).** Let $L, L'$ be two languages, and let $f : \{0, 1\}^* \rightarrow \{0, 1\}$ be a Boolean function. Let $t_1(n), t_2(n) : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and $\xi(n) : \mathbb{N}^+ \rightarrow [0, 1]$ be given, with $t_1, t_2$ polynomially bounded in $n$.

A probabilistic $f$-compression reduction for $L$ with parameters $(t_1(n), t_2(n), \xi(n))$ and target language $L'$ is a randomized mapping $R(x^1, \ldots, x^m)$ outputting a string $z$, such that for all $(x^1, \ldots, x^{t_1(n)}) \in \{0, 1\}^{t_1(n) \times n}$

1. $\Pr[R[L(x^1), \ldots, L(x^{t_1(n)})]] \geq 1 - \xi(n)$;
2. $|z| \leq t_2(n)$.

If some reduction $R$ as above is computable in probabilistic polynomial time, we say that $L$ is PPT-$f$-compressible with parameters $(t_1(n), t_2(n), \xi(n))$. (This does not require that $(t_1(n), t_2(n), \xi(n))$ themselves be computable.)

While our definition does not rule out that the output size bound $t_2(n)$ may in some cases be larger than the input size $t_1(n) \cdot n$, our interest will of course be in cases where the output size bound is (at least for large $n$) smaller than the input size. Note that we have restricted our definition to the case where $t_1(n), t_2(n)$ are both polynomially bounded in $n$. This does exclude some interesting cases from consideration. However, our study of this special case will still enable us to prove limits on strong compression reductions as described in the introduction, where there is no corresponding restriction on the size of the input-tuple.

### 3. Parametrized problems and parametrized compression.

A central aim of our work is to better understand the limitations of efficient compressive reductions for a variety of parametrized problems. However, some readers may be satisfied to understand our work on the limits of efficient AND- and OR-compression (as defined in section 2.5) for SAT and other NP-complete languages. To prove these results, including Theorem 1.2 in the introduction, we will not need the definitions of this section, and readers may choose to skip ahead to section 4. (We will prove Theorem 1.1 using the definitions below, but this result can also be derived directly from our Theorem 5.1, item 2, with little trouble.)

#### 3.1. Parametrized problems

We will use the following definition.

**Definition 3.1.** A *parametrized problem* is a subset of binary strings of the form $\langle x, 1^k \rangle$ for $x \in \{0, 1\}^*$ and $k > 0$ (under some natural binary encoding of such tuples).

Thus, our convention is that a parametrized problem is just a particular type of decision problem, i.e., a language.\footnote{In this definition, we are following [FS11]. In [BDFH09, DF13] and many other works, parametrized problems are defined as a subset of $\{0, 1\}^* \times \mathbb{N}^+$ (the parameter is still presented as part of the input); they refer to the corresponding subset of strings of form $\langle x, 1^k \rangle$ as the “un-parametrized version” or “classical version” of the problem.} However, we will use $P$ to denote a generic parametrized problem, as opposed to an “ordinary” language, denoted $L$. Sometimes, as in the introduction, we speak of “parametrized versions” of an ordinary decision problem $L$, but there is no single, canonical way to go from a decision problem to a parametrized problem; different parameters may be selected.
We now formally define the VAR-SAT, OR(SAT), and AND(SAT) problems discussed in the introduction, and some more general problems.

**Definition 3.2.** Fix some natural encoding of tuples of bit-strings and of Boolean formulas $\psi$ as bit-strings. Define $\text{VAR-SAT} := \{ \langle \psi, 1^k \rangle \mid \psi \text{ is satisfiable and contains } \leq k \text{ distinct variables} \}$. For any $L \subseteq \{0, 1\}^*$ and $f : \{0, 1\}^* \to \{0, 1\}$, define
\[
f \circ L := \{ \langle x^1, \ldots, x^t \rangle, 1^k \mid f(L(x^1), \ldots, L(x^t)) = 1 \text{ and } |x^j| \leq k \text{ for each } j \}\.
\]

We let $\text{OR}(L)$ denote $\text{OR} \circ L$, and similarly for $\text{AND}(L)$.

### 3.2. OR-expressive and AND-expressive parametrized problems.

Our compression lower bounds will apply to two classes of parametrized problems, which are closely related to classes identified earlier in [HN10, BDFH09, BJK11a, BTY11]. The classes we introduce will help to apply our techniques uniformly to these earlier classes.

**Definition 3.3 (OR- and AND-expressive problems).** A parametrized problem $P$ is OR-expressive, with parameter $S(n) \leq \text{poly}(n)$, if there exists an NP-complete language $L$ and a deterministic polynomial-time reduction $B$ acting as follows. Whenever $B$ receives an input of form $\langle x^1, \ldots, x^t, 1^n \rangle$, for any $t, n \in \mathbb{N}^+$, $B$ outputs a tuple $\langle y^1, 1^{k_1}, \ldots, y^t, 1^{k_t} \rangle$ with the following properties:

1. $\langle x^1, \ldots, x^t, 1^n \rangle \in \text{OR}(L) \iff \exists i \in [s] : \langle y^i, 1^{k_i} \rangle \in P$.
2. $s \leq S(n)$ (in particular, the bound is independent of $t$).
3. For each $i \in [s]$, $|y^i| \leq (t + n)^{O(1)}$ and $k_i \leq n^{O(1)}$.

Define AND-expressive problems identically, except that we replace condition 1 above by

1'. $\langle x^1, \ldots, x^t, 1^n \rangle \in \text{AND}(L) \iff \forall i \in [s] : \langle y^i, 1^{k_i} \rangle \in P$.

The results of [BDFH09] imply that a variety of natural parametrized problems are expressive.

**Theorem 3.4 (follows from [BDFH09]).**

1. OR(SAT) is OR-expressive with $S(n) = 1$. Also, each of the following parametrized problems (whose specific parameters are defined in [BDFH09]) are OR-expressive with $S(n) \leq \text{poly}(n)$:

- $k$-Path, $k$-Cycle, $k$-Exact Cycle and $k$-Short Cheap Tour,
- $k$-Graph Minor Order Test and $k$-Bounded Treewidth Subgraph Test,
- $k$-Planar Graph Subgraph Test and $k$-Planar Graph Induced Subgraph Test,
- $(k, \sigma)$-Short Nondeterministic Turing Machine Computation,
- $w$-Independent Set, $w$-Clique and $w$-Dominating Set.

2. AND(SAT) is AND-expressive with $S(n) = 1$. Also, each of the following parametrized problems (defined in [BDFH09]) are AND-expressive with $S(n) \leq \text{poly}(n)$:

- $k$-Cutwidth, $k$-Modified Cutwidth, and $k$-Search Number,
- $k$-Pathwidth, $k$-Treewidth, and $k$-Branchwidth,
- $k$-Gate Matrix Layout, $k$-Front Size, $w$-3-Coloring, and $w$-3-Domatic Number.

In work subsequent to ours, Fellows and Jansen showed that $k$-Pathwidth is also OR-expressive [FJ14].

---

21 Sometimes one can take $S(n) = 1$; loosely speaking, this can be done for problems with “paddable parameters.”
In [BDFH09], the authors define a notion of *compositionality* for parametrized problems. If a parametrized problem \( P \) is compositional and \( \mathsf{NP} \)-complete, then it is \( \mathsf{OR} \)-expressive with respect to the \( \mathsf{NP} \)-complete language \( L = P \). Also, if \( P \) is \( \mathsf{NP} \)-complete and \( \overline{P} \) is compositional, then \( P \) is \( \mathsf{AND} \)-expressive. These facts follow almost immediately from the definitions. Theorem 3.4 then follows from the compositionality results proved in [BDFH09].

Bodlaender, Jansen, and Kratsch [BJK11a] introduced a more general notion of *cross-compositionality* of parametrized problems in order to show new hardness results. If an \( \mathsf{NP} \)-complete problem “cross-composes” to a target parametrized problem \( P \), then \( P \) is \( \mathsf{OR} \)-expressive, as follows from the definitions [BJK11a, section 3].\(^{22}\) Problems \( P \) of this type include certain parametrized versions of the Clique, Chromatic Number, and Feedback Vertex Set problems [BJK11a].

We also have the following result, derived from the earlier work of [HN10].

**Theorem 3.5** (follows from [HN10, FS11]). Each of the problems Clique, Dominating Set,\(^{23}\) and Integer Programming, described in [HN10] and modeled as parametrized problems in [FS11] (with slightly distinctive, but natural, parametrizations), are \( \mathsf{OR} \)-expressive.

A class of reductions between parametrized problems, called \( \mathsf{W} \)-reductions, is used in these works (see [FS11, Definition 2.10]); \( \mathsf{OR} \)(\( \mathsf{SAT} \)) is shown to \( \mathsf{W} \)-reduce to each of the problems listed in Theorem 3.5. This immediately implies that these problems are \( \mathsf{OR} \)-expressive with \( S(n) = 1 \). Also, if an \( \mathsf{OR} \)-expressive parametrized problem \( P \) \( \mathsf{W} \)-reduces to a second problem \( Q \), then \( Q \) is also \( \mathsf{OR} \)-expressive. This technique was used in [BTY11] to derive additional hardness-of-compression results for problems not easily captured by the compositionality framework; our new results apply to these problems as well.

\( \mathsf{AND} \)-expressiveness results are fewer in number, although this may be due in part to the fact that, after the results of [FS11] appeared, \( \mathsf{OR} \)-expressiveness results were preferentially sought. Another example of an \( \mathsf{AND} \)-expressive problem (not known to be \( \mathsf{OR} \)-expressive) is given in [BJK11c] (see also [KW11]).

### 3.3. Parametrized compression

We define compression reductions for parametrized problems as follows, following [FS11] (but with some added flexibility in our definitions).

**Definition 3.6** (Probabilistic parametrized compression reductions). Let \( P \) be a parametrized problem and \( L' \) be a language, and say we are given two functions

\[
c(m, k, w) : (\mathbb{N}^+)^3 \rightarrow \mathbb{N}^+, \quad \xi(m, k, w) : (\mathbb{N}^+)^3 \rightarrow [0, 1] .
\]

A randomized mapping \( R : \{0, 1\}^* \rightarrow \{0, 1\}^* \) is called a \((c, \xi)\)-parametrized compression reduction for \( P \) with target language \( L' \) if for all inputs of form \((y, 1^k, 1^w)\), \( R((y, 1^k, 1^w)) \) outputs a \( z \) such that

1. \( \Pr[R(L'(z) = P((y, 1^k)))] \geq 1 - \xi(|y|, k, w) \);
2. \( |z| \leq c(|y|, k, w) \).

We call \( c \) the compression bound and \( \xi \) the error bound of the reduction; we call \( w \) the confidence parameter. For a parametrized problem \( P \), if some reduction \( R \) as above is computable in probabilistic polynomial time, we say that \( P \) is PPT-compressible with parameters \((c, \xi)\).

\(^{22}\)A small observation needed to confirm this claim is that one can always preprocess an instance of \( \mathsf{OR}(L) \) to ensure \( t \leq 2^{n+1} \) by dropping duplicate strings.

\(^{23}\)These are different parametrized problems than \( w \)-Clique and \( w \)-Dominating Set in Theorem 3.4 above.
We will not be exploring the full range of possible parameter values in the above definition, but we feel that it provides a reasonable framework for future work. (Only a few interesting examples of randomized parametrized compression reductions seem to be known; see [HN10, KW12, Wah13].) The idea of a confidence parameter \( w \), which one can use to increase the reliability of the compression at the expense of a potentially larger output size, is natural for probabilistic compression and will be useful in our work. (The same basic notion was used earlier in [FS11].)

Next, we define a notion of “strong” compressibility as in the introduction, preserving flexibility in the error bound.

**Definition 3.7.** Say that \( L \) is strongly PPT-compressible with error bound \( \varepsilon(m,k,w) \) if \( L \) is PPT-compressible (to some target language \( L' \)) with error bound \( \varepsilon \) and some compression bound \( c \) satisfying \( c(m,k,1) \leq k^{O(1)} \) with the polynomial bound independent of \( m \).

Using the majority-vote technique of [FS11, Proposition 5.1], we have the following easy result.

**Lemma 3.8.** Let \( a > 0 \). Suppose that \( L \) is strongly PPT-compressible with error bound satisfying \( \varepsilon(m,k,1) \leq \frac{5}{2} - k^{-O(1)} \). Then, \( L \) is also PPT-compressible with compression bound \( c'(m,k,w) \leq k^{O(1)} \cdot w \) and error bound \( \varepsilon'(m,k,w) \leq 2^{-w} \).

### 3.4. Connecting parametrized compression and \( f \)-compression.

Next, we show that evidence against efficient AND- and OR-compression for NP-complete languages implies evidence against efficient compression for “expressive” parametrized problems. The following lemma is modeled on [BDFH09, Lemma 2] but adapted slightly to the probabilistic setting.

**Lemma 3.9.** Let \( L \) be an NP-complete language.

1. Suppose that the parametrized problem \( L \) is OR-expressive with respect to \( L \) with parameter \( S(n) \leq \text{poly}(n) \). If \( P \) is strongly PPT-compressible with error bound \( \varepsilon(m,k,1) \leq \frac{5}{2} - k^{-O(1)} \), then there is a \( C > 0 \) such that, for any polynomially bounded function \( T(n) : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \), \( L \) is PPT-OR-compressible with parameters
   \[
   t_1(n) = T(n), \quad t_2(n) \leq S(n) \cdot n^C, \quad \varepsilon'(n) \leq 2^{-n}.
   \]

2. Suppose that \( P \) is AND-expressive with respect to \( L \). If \( P \) is strongly PPT-compressible with error bound \( \varepsilon(m,k,1) \leq \frac{5}{2} - k^{-O(1)} \), then \( L \) is PPT-AND-compressible with parameters \( t_1(n), t_2(n), \varepsilon'(n) \) as in item 1.

**Proof of Lemma 3.9.** We will prove item 1 above; item 2 is similar. Let \( R \) be the PPT compression reduction for \( P \) given by Lemma 3.8. Let \( L' \) be the target language of \( R \), and let \( B \) be the reduction for \( (P,L) \) as in Definition 3.3. We define an OR-compression reduction \( R' \) for \( L \) with target language \( L' := \text{OR}(L_0) \) as follows. First, we take \( t_1(n) := T(n) \). On inputs \( x^1, \ldots, x^{T(n)} \in \{0,1\}^{T(n) \times n} \), the reduction first applies \( B \) to \( \langle x^1, \ldots, x^{T(n)}, 1^n \rangle \), yielding a tuple \( \langle y^1, k^1, \ldots, y^s, k^s \rangle \). Next, for each \( i \in [s] \), \( R' \) applies \( R \) to the string \( \langle y^i, 1^{k^i}, 1^{2n} \rangle \) (here, we are selecting the confidence parameter \( w := 2n \) for \( R \)), yielding an output \( z' \). Then \( R' \) outputs \( \langle z^1, \ldots, z^s, 1^M \rangle \), where \( M := \max_i |z^i| \).

\( R' \) is clearly polynomial-time computable. Let us analyze its compression property. First, each \( y^i \) is of bit-length \( |y^i| \leq \langle T(n) + n \rangle^c \), and \( k_i \leq n^a \) for some \( a > 0 \) independent of the choice of \( T(\cdot) \); this holds by item 3 of Definition 3.3. Then, by the compression guarantee for \( R \), we have \( |z^i| \leq n^b \cdot w \leq n^{b+2} \) for some \( b > 0 \) also independent of \( T \). Thus, we can take \( t_2(n) \) as needed.

© 2015 Andrew Drucker
Now, we bound the error of \( R' \). Using Definition 3.3, item 1, applied to \( B \), the equivalence \( \langle (z^1, \ldots, z^s), 1^M \rangle \in OR(L_0) \iff \bigwedge_{j=1}^{O(n)} \exists j \in L \) holds provided each application of \( R \), namely, \( R((y^1, 1^k)) \) for \( i \in [s] \), is successful. By a union bound, this occurs with probability \( \geq 1 - S(n) \cdot 2^{-2n} \), which is larger than \( 1 - 2^{-n} \) for sufficiently large \( n \). (For smaller \( n \), \( R' \) may solve its input problem directly by brute force.) This proves the lemma. \( \square \)

4. Technical lemmas. In this section, we present our main technical lemmas. Our final goal in this section will be the Disguising-Distribution lemma, our key technical tool for our main results.

4.1. Distributional stability. Here, we define the notion of “distributional stability” described in section 1.4.2.

**Definition 4.1.** Let \( U \) be some finite universe, and let \( T, n \geq 1 \) be integers. Given a possibly randomized mapping \( F(x^1, \ldots, x^T) : \{0, 1\}^{T \times n} \rightarrow U \) and a collection \( D_1, \ldots, D_T \) of mutually independent distributions over \( \{0, 1\}^n \), for \( j \in [T] \) let

\[
\gamma_j := \mathbb{E}_{y \sim D_j} \left[ ||F(D_1, \ldots, D_{j-1}, y, D_{j+1}, \ldots, D_T) - F(D_1, \ldots, D_T)||_{\text{stat}} \right].
\]

For \( \delta \in [0, 1] \), say that \( F \) is \( \delta \)-distributionally stable (or \( \delta \)-DS) for \( D_1, \ldots, D_T \) if

\[
\frac{1}{T} \sum_{j=1}^{T} \gamma_j \leq \delta.
\]

**Lemma 4.2.** Let \( R(x^1, \ldots, x^t) : \{0, 1\}^{t \times n} \rightarrow \{0, 1\} \leq t' \) be any possibly randomized mapping for any \( n, t, t' \in \mathbb{N}^+ \). \( R \) is \( \delta \)-distributionally stable with respect to any independent input distributions \( D_1, \ldots, D_t \), where we may take either of the following two bounds:

1. \( \delta := \sqrt{\frac{\ln 2 \cdot t' + 1}{t}} \).
2. \( \delta := 1 - 2^{-\frac{t' + 3}{2}} \).

Our proof of Lemma 4.2, item 1, essentially follows suggestions by Ashwin Nayak and Salil Vadhan; \(^{24} \) item 2 is a small modification using Vajda’s inequality.

**Proof of Lemma 4.2.** Define independent random variables \( X^j \sim D_j \) over \( \{0, 1\}^n \) for \( j \in [t] \). Let \( R := R(X^1, \ldots, X^t) \). The entropy of \( R \) is at most \( \log_2 \left( \left| \{0, 1\} \leq t' \right| \right) < t' + 1 \). Thus, the mutual information \( I((X^1, \ldots, X^t); R) \) is less than \( t' + 1 \). As the \( X^j \)’s are independent, Lemma 2.6 gives

\[
\sum_{j \in [t]} I(X^j; R) < t' + 1.
\]

By Fact 2.8,

\[
I(X^j; R) = D_{\text{KL}} \left( (X^j; R) \mid \mid (Y^j; R) \right).
\]

\(^{24}\) As discussed earlier, the proof is related to work in [Raz98, KNTSZ07, Man01, SV08].
where \( Y_j \sim D_j \) is independent of \( R \). By Theorem 2.9,

\[
D_{\text{KL}} ((X^j, R) \| (Y^j, R)) \geq \frac{2}{\ln 2} \cdot \left\| (X^j, R) - (Y^j, R) \right\|_{\text{stat}}^2
\]

\[
= \frac{2}{\ln 2} \cdot E_{x^j \sim D_j} \left[ \left\| R(D_1, \ldots, D_{j-1}, x^j, D_{j+1}, \ldots, D_t) - R(D_1, \ldots, D_t) \right\|_{\text{stat}}^2 \right],
\]

where the equality follows from the distinguishability interpretation of statistical distance. Using this, we find

\[
\left( \frac{1}{t} \sum_{j \in [t]} E_{x^j \sim D_j} [ \left\| R(D_1, \ldots, D_{j-1}, x^j, D_{j+1}, \ldots, D_t) - R(D_1, \ldots, D_t) \right\|_{\text{stat}}^2 ] \right)^{\frac{1}{2}}
\]

\[
\leq \frac{1}{t} \sum_{j \in [t]} E_{x^j \sim D_j} [ \left\| R(D_1, \ldots, D_{j-1}, x^j, D_{j+1}, \ldots, D_t) - R(D_1, \ldots, D_t) \right\|_{\text{stat}}^2 ]
\]

\[
< \frac{\ln 2}{2} \cdot \frac{t' + 1}{t},
\]

where in line 2 we used Jensen’s inequality. Thus, \( R \) is \( \sqrt{\frac{\ln 2}{2}} \cdot \frac{t' + 1}{t} \)-distributionally stable with respect to \( D^1, \ldots, D^t \). This proves item 1 of the Lemma.

For item 2, we apply Vajda’s inequality (Theorem 2.10) to each \( j \in [t] \) to find

\[
D_{\text{KL}} ((X^j, R) \| (Y^j, R)) \geq \frac{1}{\ln 2} \left( \ln \left( \frac{1}{1 - \left\| (X^j, R) - (Y^j, R) \right\|_{\text{stat}}} - 1 \right) \right)
\]

\[
= \frac{1}{\ln 2} \left( \ln \left( \frac{1}{\varepsilon_j} - 1 \right) \right),
\]

where we define \( \varepsilon_j := 1 - E_{x^j \sim D_j} [ R(D_1, \ldots, D_{j-1}, x^j, D_{j+1}, \ldots, D_t) - R(D_1, \ldots, D_t) ] \), and note that \( \varepsilon_j > 0 \). Averaging over \( j \in [t] \) and applying (1) and (2),

\[
\frac{t' + 1}{t} \geq \frac{1}{t} \sum_{j \in [t]} \frac{1}{\ln 2} \left( \ln \left( \frac{1}{\varepsilon_j} - 1 \right) \right), \quad \text{i.e.,} \quad \frac{1}{t} \sum_{j \in [t]} \ln \left( \frac{1}{\varepsilon_j} \right) \leq \frac{(\ln 2)(t' + 1)}{t} - 1.
\]

The function \( f(x) = \ln(1/x) \) has second derivative \( x^{-2} > 0 \) for \( x > 0 \), so by Jensen’s inequality,

\[
\ln \left( \frac{1}{\frac{1}{t} \sum_{j \in [t]} \varepsilon_j} \right) \leq \frac{(\ln 2)(t' + 1)}{t} + 1, \quad \text{so that} \quad \frac{1}{t} \sum_{j \in [t]} \varepsilon_j \geq \left( e^{\frac{(\ln 2)(t' + 1)}{t}} + 1 \right)^{-1}
\]

\[
\geq 2^{-\frac{t'}{t} - 3},
\]

which proves item 2. \( \square \)

### 4.2. Sparsified distributional stability

Here, we prove a technical lemma showing that if a mapping \( F \) is distributionally stable with respect to i.i.d. inputs, then \( F \) also obeys a slightly different stability property, in which we replace an input distribution \( D \) with a “sparsified” version of \( D \).

**Lemma 4.3.** Let \( U \) be a finite set, and let \( F(x^1, \ldots, x^T) : \{0, 1\}^{T \times n} \rightarrow U \) be given. Suppose that \( F \) is \( \delta \)-DS with respect to input distribution \( D^{\otimes T} \) for every distribution \( D \).
over \(\{0,1\}^n\). Fix some distribution \(D\) over \(\{0,1\}^n\), and let \(x^1, \ldots, x^d\) be independently sampled from \(D\).

Let \(k^* \sim U_{[d]}\). Let \(\hat{D}\) be the distribution that samples uniformly from the multiset \(\{x^k\}_{k \in [d]}\). (This distribution is itself a random variable, determined by \(x^1, \ldots, x^d\) and by \(k^*\).) Define

\[
\beta_j := \mathbb{E}_{k^*, x^1, \ldots, x^d} \left[ \left\| F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) - F \left( \hat{D} \otimes T \right) \right\|_{stat} \right],
\]

where the \(\hat{D}\)'s are mutually independent (for fixed \(x^1, \ldots, x^d\) and \(k^*\)). Then, \(\frac{1}{T} \sum_{j=1}^d \beta_j \leq \delta + 2T/d\).

**Proof.** Let \(\hat{D}\) denote the distribution, determined by \(x^1, \ldots, x^d\), that samples uniformly from the multiset \(\{x^k\}_{k \in [d]}\). By an easy calculation, for any values of \(x^1, \ldots, x^d\) and \(k^*\) we can bound \(\left\| \hat{D} - \hat{D} \right\|_{stat} \leq 1/d\). It follows that

\[
\left\| F \left( \hat{D} \otimes T \right) - F \left( \hat{D} \otimes T \right) \right\|_{stat} \leq \left\| \hat{D} \otimes T - \hat{D} \otimes T \right\|_{stat} \leq T/d,
\]

where in the last step we used Fact 2.2 and the fact that for any assignment to \(x^1, \ldots, x^d\) and to \(k^*\), the \(T\) copies of \(D\) used are mutually independent, as are the copies of \(\hat{D}\). By identical reasoning, for any assignment to \(x^1, \ldots, x^d\) and to \(k^*\), and for any index \(j \in [T]\), we have

\[
\left\| F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) - F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) \right\|_{stat} \leq (T-1)/d.
\]

Using the triangle inequality for \(\| \cdot \|_{stat}\), for any values \(x^1, \ldots, x^d, k^*\) and any index \(j \in [T]\) we always have

\[
(3) \quad \left\| F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) - F \left( \hat{D} \otimes T \right) \right\|_{stat} \leq \left\| F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) - F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) \right\|_{stat} + \left\| F \left( \hat{D} \otimes T \right) - F \left( \hat{D} \otimes T \right) \right\|_{stat} + 2T/d.
\]

Now suppose that we fix any values \(x^1, \ldots, x^d\), leaving \(k^*\) undetermined. The value \(k^*\) is uniform on \([d]\), so that \(x^{k^*}\) is distributed exactly according to \(\hat{D}\). Under our conditioning, let

\[
\gamma_j = \gamma_j \left( \{x^k\}_{k \in [d]} \right) := \mathbb{E}_{k^*} \left[ \left\| F \left( \hat{D} \otimes (j-1), x^{k^*}, \hat{D} \otimes (T-j) \right) - F \left( \hat{D} \otimes T \right) \right\|_{stat} \right].
\]

By our original assumption, \(F\) is \(\delta\)-DS with respect to input distribution \(\hat{D} \otimes T\). Thus, for any \(x^1, \ldots, x^d\), we have \(\frac{1}{T} \sum_{j=1}^d \gamma_j \leq \delta\). Now, \(\gamma_j\) is itself a random variable, determined by \(x^1, \ldots, x^d\), and from (3) we have \(\beta_j \leq \mathbb{E}[\gamma_j] + 2T/d\). Using linearity of expectation, it follows that \(\frac{1}{T} \sum_{j=1}^d \beta_j \leq \delta + 2T/d\).

**4.3. Building disguising distributions.** In the next lemmas, we show how the distributional stability of a mapping \(F\) can be used to obtain a disguising distribution for \(F\). In Lemma 4.6, we will apply this to give disguising distributions for any sufficiently compressive mapping \(R\).
Recall that $U_K$ denotes the uniform distribution over a multiset $K$.

**Lemma 4.4**. Suppose that $F(x_1, \ldots, x_T) : \{0,1\}^{T \times n} \to U$ is $\delta$-distributionally stable with respect to input distribution $D^\otimes T$ for every distribution $D$ over $\{0,1\}^n$.

Let $S \subseteq \{0,1\}^n$, and fix some value $d > 0$. There exists a distribution $K$ over size-$d$ multisets $K \subseteq S$, such that for every $y \in S$, the following holds:

$$\mathbb{E}_{K \sim K^*} \left[ \left\| F \left( U_K^{(j^*-1)}, y, U_K^{(T-j^*)} \right) - F \left( U_K^{(T)} \right) \right\|_{\text{stat}} \right] \leq \delta + 2T/(d+1).$$

(Here, the copies of $U_K$ are to be mutually independent for fixed $K$, although the set $K \sim K$ used is the same for each copy.)

**Proof.** Consider the following two-player, simultaneous-move, zero-sum game:

- Player 1 chooses a size-$d$ multiset $K \subseteq S$.
- Player 2 chooses a string $y \in S$.
- Payoff: Player 2 receives a payoff equal to

$$\mathbb{E}_{j^* \sim U_{\{1, \ldots, T\}}} \left[ \left\| F \left( U_K^{(j^*-1)}, y, U_K^{(T-j^*)} \right) - F \left( U_K^{(T)} \right) \right\|_{\text{stat}} \right].$$

(Note that this payoff is a determinate value, given $(K,y)$.)

Consider any randomized strategy by player 2, specified by a distribution $y \sim Y$ over $S$. In response, let $K_Y$ be the randomized player-1 strategy that chooses a size-$d$ multiset $K$ of elements sampled independently from $Y$.

To bound the expected payoff under the strategy-pair $(K_Y, Y)$, note that we can equivalently generate $(K,y) \sim (K_Y,Y)$ as follows. First, sample $x_1, \ldots, x_{d+1}$ independently from $Y$. Sample $k^* \sim U_{\{d+1\}}$, set $y := x_{k^*}$, and let $K := \{x_1, \ldots, x_{k^*-1}, x_{k^*+1}, \ldots, x_{d+1}\}$. It is easily verified that $(K,y) \sim (K_Y,Y)$ as desired. Then Lemma 4.3, applied to our initial distributional-stability assumption on $F$, informs us that

$$\mathbb{E}_{j^* \sim U_{\{1, \ldots, T\}}, K, y} \left[ \left\| F \left( U_K^{(j^*-1)}, y, U_K^{(T-j^*)} \right) - F \left( U_K^{(T)} \right) \right\|_{\text{stat}} \right] \leq \delta + 2T/(d+1).$$

Thus, Player 2’s expected payoff against $K_Y$ is at most $\delta + 2T/(d+1)$.

As $Y$ was arbitrary, the minimax theorem tells us that there is a distribution $K$ over player 1’s moves such that player 2’s expected payoff under every strategy is at most $\delta + 2T/(d+1)$. The result follows.

**Lemma 4.5.** Let $U$ be a finite set, and let $F(x_1, \ldots, x_T) : \{0,1\}^{T \times n} \to U$ be given. Suppose that $F$ is $\delta$-distributionally stable with respect to input distribution $D^\otimes T$ for each distribution $D$ over $\{0,1\}^n$.

Let $S \subseteq \{0,1\}^n$, and fix $d > 0$. Given any $\varepsilon > 0$, let $s := \lceil (5 \ln 2)n/\varepsilon^2 \rceil$. Then there is a collection $K_1, \ldots, K_s$ of size-$d$ multisets contained in $S$, such that for every $y \in S$, we have

$$\mathbb{E}_{a \sim U_{\{1, \ldots, s\}}, j^* \sim U_{\{1, \ldots, T\}}} \left[ \left\| F \left( U_{K_a}^{(j^*-1)}, y, U_{K_a}^{(T-j^*)} \right) - F \left( U_{K_a}^{(T)} \right) \right\|_{\text{stat}} \right] \leq \delta + 2T/(d+1) + \varepsilon.$$  

**Proof.** This is an immediate application (to the game in Lemma 4.4) of a general result due to Lipton and Young [LY94, Theorem 2], showing that all two-player, zero-sum games have sparsely supported, nearly optimal player strategies. (Essentially the same result was proved independently by Althöfer [Alt94].) The support size required in the Lipton–Young–Althöfer result depends logarithmically on the number.
of pure strategies available to the player we are opposing; in our case, player 2 has a choice of $|S| \leq 2^n$ strings $y$, so we get $s = O(n/\varepsilon^2)$. In their proof technique applied to our setting, the $K_1, \ldots, K_s$ are obtained by sampling independently from the distribution $\mathcal{K}$ given by Lemma 4.4, giving a suitable choice of $K_1, \ldots, K_s$ with nonzero probability.

Lemma 4.6 (Disguising-Distribution lemma). Let $R(x^1, \ldots, x^t) : \{0, 1\}^{t \times n} \to \{0, 1\}^{\leq t'}$ be any possibly randomized mapping for $t, t' \in \mathbb{N}^+$. Let $S \subseteq \{0, 1\}^n$, and fix $d > 0$. Given any $\varepsilon > 0$, let $s := \lfloor (\ln 2)n/\varepsilon^2 \rfloor$. Let $\hat{\delta} := \min \left\{ \sqrt{\frac{\ln 2}{n}} \cdot \frac{t^2 + 1}{t}, 1 - 2^{-\frac{d^2}{4} - 3} \right\}$.

Then there exists a collection $K_1, \ldots, K_s$ of size-$d$ multisets contained in $S$, such that for every $y \in S$, we have

$$E_{a \sim \mathcal{U}(s^t), j^t \sim \mathcal{U}(t)} \left[ \left\| R \left( U_{K_a}^{\otimes (j^t - 1)}, y, U_{K_a}^{\otimes (t - j^t)} \right) - R \left( U_{K_a}^{\otimes t} \right) \right\|_{\text{stat}} \right] \leq \hat{\delta} + 2t/(d + 1) + \varepsilon .$$

Proof. This follows immediately from the combination of Lemmas 4.2 and 4.5, applied to $F := R$ (and with $T := t$). □

5. Limits to efficient (classical) compression. In this section, we show that a sufficiently high-quality PPT-OR-compression reduction for any language $L$ implies $L \in \text{NP/poly}$. (The same reduction also establishes $L \in \text{coNP/poly}$, as observed by Dell [Del14] and as we will explain.) We also show that above a higher threshold of quality, such a compression reduction implies that $L$ has nonuniform, statistical zero-knowledge proofs.

We will then apply these results to give evidence against efficient probabilistic compression for AND(SAT) and OR(SAT), as described in the introduction, and for other parametrized problems with either of the two “expressiveness” properties described in section 3.2. We will also present our result extending the work of Dell and Van Melkebeek [DvM10] on problems with polynomial kernelizations.

5.1. Complexity upper bounds from OR-compression schemes.

Theorem 5.1. Let $L$ be any language. Suppose that $t_1(n), t_2(n) : \mathbb{N}^+ \to \mathbb{N}^+$ are (not necessarily computable, but polynomially bounded) functions. Suppose that there exists a PPT-OR-compression reduction $R(x^1, \ldots, x^t) : \{0, 1\}^{t_1(n) \times n} \to \{0, 1\}^{\leq t_2(n)}$ for $L$ with parameters $t_1(n), t_2(n)$, error bound $\xi(n) < .5$, and some target language $L'$. Let

$$\hat{\delta} := \min \left\{ \sqrt{\frac{\ln 2}{2} \cdot \frac{t_2(n) + 1}{t_1(n)}}, 1 - 2^{-\frac{t_2(n)}{\ln 2} - 3} \right\} .$$

1. If for some constant $c > 0$ we have

$$1 - 2\xi(n) - \hat{\delta} \geq \frac{1}{n^c} ,$$

then $L \in \text{NP/poly}$.

2. If for some $c > 0$ we have the (stronger) bound

$$\left( 1 - 2\xi(n) \right)^2 - \hat{\delta} \geq \frac{1}{n^c} ,$$

© 2015 Andrew Drucker
then there is a many-to-one reduction from \( L \) to a promise problem in \( \text{pr-SZK} \).

The reduction is computable in nonuniform polynomial time (so in particular \( L \in \text{NP/poly} \cap \text{coNP/poly} \)).

We remark that, using the majority-vote technique of [FS11, Proposition 5.1], one can reduce the error bound \( \xi(n) \) of an OR-compression scheme, at the cost of increasing the output-length bound \( t_2(n) \). With this technique we can in some cases apply Theorem 5.1, where its assumptions do not hold for the original scheme.

Proof of Theorem 5.1. We will use the same basic reduction to prove items 1 and 2.

First, with nonuniformity, it is easy to handle length-\( n \) inputs whenever \( L_n = \{0,1\}^n \), so let us assume from this point on that \( T_n \) is nonempty. Using \( R \), we define a deterministic, nonuniform polynomial-time reduction \( R \) that, on input \( y \in \{0,1\}^n \), builds a description of two circuits \( C,C' \). The aim is that, for the distributions \( D_C,D_{C'} \) as described in section 2.4, the distance \( ||D_C - D_{C'}||_{\text{stat}} \) should be large if \( y \in L \) and small if \( y \notin L \). \( R \) works as follows:

- **Nonuniform advice for length \( n \).** A description of the value \( t_1(n) \) and the multisets \( K_1,\ldots,K_s \subseteq T_n \) given by Lemma 4.5 with
  \[
  (t,t') := (t_1(n),t_2(n)), \quad S := T_n, \quad d := [8t_1(n) \cdot n^c], \quad \varepsilon := \frac{1}{4n^c}.
  \]

  (Here, \( c > 0 \) is as in (4) or (5), according to which item of the theorem we are proving.) Note that \( d \) and the value \( s \) given by Lemma 4.5 are both \( \leq \text{poly}(n) \) under these settings, so our advice is of polynomial length.

- **On input \( y \in \{0,1\}^n \).** Let \( R \) output descriptions \( \langle C,C' \rangle \) of the following two randomized circuits:
  - Circuit \( C \): samples \( a \sim U[a] \), then samples \( \pi = (x^1,\ldots,x^{t_1(n)}) \sim U_{K_a}^{\otimes t_1(n)} \) and outputs \( z := R(\pi) \).
  - Circuit \( C' \): samples values \( a \sim U[a], j^* \sim U[t_1(n)] \), then, samples \( \pi \sim (U_{K_a}^{\otimes (j^*-1)}, y, U_{K_{t_1(n)-j^*}}) \), and outputs \( z := R(\pi) \).

**Claim 5.2.** The following holds:

1. If \( y \in L \), then \( ||D_C - D_{C'}||_{\text{stat}} \geq D(n) := 1 - 2\xi(n) \).
2. If \( y \notin L \), then

  \[
  (6) \quad ||D_C - D_{C'}||_{\text{stat}} \leq d(n) := \hat{\delta} + \frac{1}{2n^c}.
  \]

We defer the proof of Claim 5.2 and use it to prove the two items of Theorem 5.1.

For item 1 of Theorem 5.1, if (4) holds (for sufficiently large \( n \)), then \( D(n) - d(n) \geq \frac{1}{2n^c} \).

Now, \( D(n),d(n) \) were parametrized in terms of \( n = |y| \), but the gap \( D(n) - d(n) \) is also at least inverse-polynomial in the length \( N \leq \text{poly}(n) \) of the output description \( \langle C,C' \rangle \). Thus, our reduction \( R \) reduces any instance \( y \) of the decision problem for \( L \) to an equivalent instance \( R(y) = \langle C,C' \rangle \) of the promise problem \( \text{SD}_{d'(N)}^{D'(N)} \) with different parameters \( D'(N),d'(N) \) still satisfying the gap condition \( D' - d' \geq \frac{1}{\text{poly}(N)} \).

By Theorems 2.12 and 2.16, \( \text{SD}_{d'}^{D'} \in \text{pr-NP/poly} \). Let \( (A,\{a_N\}_{N>0}) \) be a nondeterministic polynomial-time algorithm and nonuniform advice family solving \( \text{SD}_{d'}^{D'} \). Then by applying \( (A,\{a_N\}) \) to \( R(y) \), we obtain a nondeterministic, nonuniform polynomial-time algorithm for solving \( L \). This shows \( L \in \text{NP/poly} \), proving item 1 of the theorem.
Next, for item 2 of Theorem 5.1, if (5) holds for sufficiently large \( n \), then \( D(n)^2 - d(n) \geq 1/2n \). Arguing as in the previous case, we exhibit a nonuniform polynomial-time reduction from \( L \) to \( SD_{\leq d'} \), where this time \( D'(N)^2 - d'(N) \geq 1/\text{poly}(N) \). The problem \( SD_{\leq d'} \) is in \( \text{pr-SZK} \), by Theorem 2.15.

In item 1, one can also obtain \( L \in \text{coNP/poly} \) simply by substituting the fact that, by Theorem 5.16, \( SD_{\leq d'} \in \text{pr-coNP/poly} \) in this case [Del14].

Proof of Claim 5.2. (1) First, suppose \( y \in L \). We will use the distinguishing interpretation of statistical distance (see section 2.1) to argue that \( ||D_C - D_{C'}||_{\text{stat}} \) is large. Suppose that an unbiased coin \( b \sim U_{\{0,1\}} \) is flipped, unseen by us, and we receive a sample \( z \sim D_C \) if \( b = 0 \), or \( z \sim D_{C'} \) if \( b = 1 \). Consider the distinguisher that outputs the guess \( \tilde{b} := 0 \) if \( z \in \overline{L} \), or \( \tilde{b} := 1 \) if \( z \in L' \).

We lower-bound the success probability \( \Pr[\tilde{b} = b] \) as follows. Say we condition on \( b = 0 \), so that \( z \sim D_C \). The distributions \( U_{K_n} \) are supported on \( \overline{L_n} \), so in the execution of \( C \) we get \( \pi \in (\overline{L_n}) \). Then it follows from the OR-compression property of \( R \) for \( L \) that \( \Pr[z \in L' \mid \pi] \geq 1 - \xi(n) \). On the other hand, suppose that we condition on \( b = 1 \), so that \( z \sim D_{C'} \). In an execution of \( C' \), the input tuple \( \pi \) contains \( y \in L_n \); thus, by the OR-compression property of \( R \), we have \( \Pr[z \in L' \mid \pi] \geq 1 - \xi(n) \). So regardless of the value of \( b \), our distinguisher succeeds with probability \( \geq 1 - \xi(n) \). Thus \( 1 - \xi(n) \leq \frac{1}{2}(1 + ||D_C - D_{C'}||_{\text{stat}}) \). This proves item 1.

(2) Now suppose \( y \notin L \); we must upper-bound \( ||D_C - D_{C'}||_{\text{stat}} \). Consider the distinguishing experiment between \( C \) and \( C' \) as in item 1. If we regard the random variables \( a \) and \( j^* \) (the latter used only by \( C' \)) to be part of the joint probability space of both algorithms (noting that \( a \) is identically distributed in the two circuits), then revealing the values \( a, j^* \) along with \( z \) to the distinguisher cannot decrease the distinguisher’s maximum achievable success probability. Now conditioned on revealed values \( a, j^* \), the maximum achievable success probability in the modified distinguishing experiment is

\[
\frac{1}{2} \left( 1 + \left| \left| R \left( U_{K_n}^{(t_1(n))} \right) - R \left( U_{K_n}^{(t_1(n)) \oplus (j^*)} \right) \right|_{\text{stat}} \right| \right),
\]

from which we conclude that

\[
||D_C - D_{C'}||_{\text{stat}} \leq \sum_{a \sim U_{\{0,1\}}, j^* \sim U_{t_1(n)}} \left[ \left| \left| R \left( U_{K_n}^{(t_1(n))} \right) - R \left( U_{K_n}^{(t_1(n)) \oplus (j^*)} \right) \right|_{\text{stat}} \right| \right].
\]

By our choice of \( K_1, \ldots, K_n \) and Lemma 4.6, the right-hand side of (7) is upper-bounded by \( \hat{\delta} + 2t_1(n)/(d + 1) + \varepsilon < \hat{\delta} + 2 \cdot \frac{1}{4n} \), by our settings to \( d, \varepsilon \). This proves (6).

Next, we state a useful consequence of Theorem 5.1 for when \( t_2(n) \leq O(t_1(n) \cdot \log_2(t_1(n))) \).

Theorem 5.3. Let \( L \) be any language. Suppose that \( t_1(n), t_2(n) : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \) satisfy \( t_2(n) \leq C \cdot t_1(n) \log t_1(n) \) and \( t_1(n) \leq n^{C'} \) for some \( C, C' > 0 \). Suppose that \( R \) is a PPT-OR-compression reduction \( R(x_1, \ldots, x_t(n)) : \{0,1\}^{t_1(n) \times n} \rightarrow \{0,1\}^{\leq t_2(n)} \) for \( L \) with parameters \( t_1(n), t_2(n), t_2(n), \) error bound \( \xi(n) < .5 \), and some target language \( L' \).

If \( \xi(n) < n^{-C-C'}/32 \), then there is a nonuniform polynomial-time many-to-one reduction from \( L \) to a promise problem in \( \text{pr-SZK} \).

Proof. We bound the quantity \( \hat{\delta} \) from Theorem 5.1:

\[
\hat{\delta} \leq 1 - 2^{-t_1(n)/t_1(n) - 3} \leq 1 - 2^{-C \log_2(t_1(n))/8} \leq 1 - t_1(n)^{-C}/8 \leq 1 - n^{-C-C'}/8.
\]

© 2015 Andrew Drucker
Throughout this section, for parameters $\xi(n) \in L$ and error bound $\xi(n) < \frac{1}{\text{poly}(n)}$, and the desired conclusion then follows from Theorem 5.1, item 2.

5.2. Application to AND- and OR-compression of NP-complete languages. Throughout this section, for parameters $t_1(n), t_2(n)$, we will use the shorthand

$$\hat{\delta} := \min \left\{ \sqrt{\frac{\ln 2}{2}}, \frac{2t_2(n) + 1}{t_1(n)}, \left(1 - 2^{-\frac{t_2(n)}{t_1(n)}) - 3}\right\}.$$ 

Here is our first main result giving evidence against efficient AND-compression for NP-complete languages.

**Theorem 5.4.** Suppose that for some NP-complete language $L$, any target language $L'$, and an error bound $\xi(n) < 0.5$, $L$ has a PPT-AND-compression reduction $R$ with target language $L'$ with polynomially bounded parameters $t_1(n), t_2(n) : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and error bound $\xi(n) < 0.5$.

1. If

$$1 - 2\xi(n) - \hat{\delta} \geq \frac{1}{\text{poly}(n)},$$

then $\text{NP} \subseteq \text{coNP}/\text{poly}$ and $\text{PH} = \Sigma_3^p = \Pi_3^p$.

2. If we have the bound

$$\left(1 - 2\xi(n)\right)^2 - \hat{\delta} \geq \frac{1}{\text{poly}(n)},$$

then $L$ (and every other language in NP) is many-to-one reducible in nonuniform polynomial time to a problem in PPT-$\text{SZK}$, and $\text{NP} \subseteq \text{coNP}/\text{poly}$.

3. The conclusion of item 2 also holds if $t_2(n) \leq C \cdot t_1(n) \log(t_1(n))$ and if $\xi(n)$ is a sufficiently small inverse-polynomial function of $n$ (determined by $t_1$ and the constant $C$).

**Proof of Theorem 5.4.** (1.) The reduction $R$ is also a PPT-OR-compression for $\overline{L}$ with target language $\overline{L'}$ and with the same parameters.

If (8) holds in case 1, we apply item 1 of Theorem 5.1 to $\overline{L}$, concluding that $\overline{L} \in \text{NP}/\text{poly}$, i.e., $L \in \text{coNP}/\text{poly}$. The consequence for $\text{PH}$ is from Theorem 2.11.

(2.) Similarly, if (9) holds in case 2, we apply item 2 of Theorem 5.1 to $\overline{L}$, giving a nonuniform many-to-one reduction from $\overline{L}$ to a problem $\Pi = (\Pi_Y, \Pi_Y) \in \text{PPT-}\text{SZK}$. This is also a reduction from $L$ to $(\Pi_Y, \Pi_Y)$, which by Theorem 2.15 also lies in $\text{PPT-}\text{SZK}$. The extension to other languages in NP follows from the NP-completeness of $L$.

(3.) In this case, we apply Theorem 5.3, item 1, to $\overline{L}$.

The next theorem gives evidence for the infeasibility of efficient OR-compression for NP-complete languages.

**Theorem 5.5.** Suppose that for some NP-complete language, $L$ has a PPT-OR-compression reduction $R$ with arbitrary target language $L'$ with polynomially bounded parameters $t_1(n), t_2(n) : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and error bound $\xi(n) < 0.5$.

1. If

$$(1 - 2\xi(n))^2 - \hat{\delta} \geq \frac{1}{\text{poly}(n)},$$

then $L$ (and every other language in NP) is reducible in nonuniform polynomial time to a problem in PPT-$\text{SZK}$, and $\text{NP} \subseteq \text{coNP}/\text{poly}$.

© 2015 Andrew Drucker
2. The conclusion of item 1 also holds if \( t_2(n) \leq C \cdot t_1(n) \log(t_1(n)) \) and if \( \xi(n) \) is a sufficiently small inverse-polynomial function of \( n \) (determined by \( t_1 \) and \( C \)).

Item 2 above, together with item 3 of Theorem 5.4, yields Theorem 1.2 from the introduction.

Proof of Theorem 5.5. For item 1, if (10) holds, we apply item 2 of Theorem 5.1 to \( L \) itself. For item 2, we apply item 1 of Theorem 5.3 to \( L \). \( \square \)

5.3. Limits to strong compression for parametrized problems. Next, we use Theorem 5.1 to give evidence against strong compressibility for “expressive” parametrized problems. The result we give below is a simple-to-state, representative example; the quantitative settings studied here are not the only interesting ones our techniques can handle.

Theorem 5.6. Say that \( P \) is OR-expressive or AND-expressive, e.g., one of the problems listed in Theorems 3.4 and 3.5. Suppose additionally that \( P \) is strongly PPT-compressible\(^{25} \) with error bound \( \xi(m, k, w) \) satisfying \( \xi(m, k, 1) \leq 0.5 - k^{-O(1)} \) (independent of \( m \)), i.e., with success probability \( \geq 0.5 + k^{-O(1)} \). Then, every language in \( \text{NP} \) is many-to-one reducible in nonuniform polynomial time to a problem in \( \text{pr-SZK} \) (and \( \text{NP} \subseteq \text{coNP/poly} \)).

Theorem 1.1 from the introduction follows by considering the special cases \( P = \text{OR(SAT)} \) and \( P = \text{AND(SAT)} \).

Proof of Theorem 5.6. Suppose first that \( P \) is OR-expressive with respect to the \( \text{NP} \)-complete language \( L \) and with some parameter \( S(n) \leq \text{poly}(n) \). We apply item 1 of Lemma 3.9 to \( L \) and the assumed strong compression reduction for \( P \). Using some function \( T(n) \leq \text{poly}(n) \) to be determined, and with \( w(n) := 1 \), we obtain a PPT-OR-compression for \( L \) with parameters

\[
t_1(n) = T(n), \quad t_2(n) \leq S(n) \cdot n^C, \quad \xi'(n) \leq 2^{-n}.
\]

(\( S(n) \leq \text{poly}(n) \) and is independent of the choice of \( T(n) \), as is the constant \( C > 0 \).)

We evaluate

\[
(1 - 2\xi'(n))^2 - \sqrt{\frac{\ln 2}{2}} \cdot \frac{t_2(n) + 1}{t_1(n)} \geq (1 - 4 \cdot 2^{-n}) - \sqrt{\frac{\ln 2}{2}} \cdot \frac{S(n) \cdot n^C + 1}{T(n)}.
\]

The expression above can be made greater than 0.5 for large \( n \) by choosing a sufficiently fast-growing \( T(n) \leq \text{poly}(n) \). Under such a setting, (10) holds for \( (t_1(n), t_2(n), \xi'(n)) \). We can then apply the first assertion of Theorem 5.5, item 1, to our PPT-OR-compression for \( L \), which yields the desired conclusion.

The case where \( P \) is AND-expressive is handled analogously; in this case, we apply Lemma 3.9, item 2, and the first assertion of Theorem 5.4, item 2. \( \square \)

5.4. Application to problems with polynomial kernelizations. In this section, we prove new limits to efficient compression for the Satisfiability problem on \( d \)-CNFs and for some problems on graphs and hypergraphs, partially extending results of Dell and Van Melkebeek [DvM10] to handle two-sided error. First, we need some background.

Definition 5.7 (Hypergraphs, vertex covers, and cliques). For any integer \( d \geq 2 \), a \( d \)-uniform hypergraph, or \( d \)-hypergraph, is a set \( H \) of size-\( d \) subsets of a vertex

\(^{25}\)This is as in Definition 3.7.
set \( V = [N] \). A vertex cover in a \( d \)-uniform hypergraph \( H \) is a subset of vertices that intersects all hyperedges in \( H \). A subset \( V' \subseteq V \) is a clique in \( H \) if every size-\( d \) subset of \( V' \) is a member of \( H \).

Clearly, \( H \) has a vertex cover of size \( s \) exactly if the “complement” hypergraph \( \mathcal{H} := \{ e : |e| = d \land e \notin H \} \) contains a clique of size \( N - s \).

**Definition 5.8.** Define the parametrized problems

\[ d\text{-Vertex Cover} := \{ \langle (H, s), 1^N \rangle : H \text{ is a } d\text{-hypergraph on } [N] \text{ and contains a vertex cover of size } s \} \]

\[ d\text{-Clique} := \{ \langle (H, s), 1^N \rangle : H \text{ is a } d\text{-hypergraph on } [N] \text{ and contains a clique of size } s \} \]  

Also define the parametrized \( d\)-CNF Satisfiability problem

\[ d\text{-SAT}_{\text{par}} := \{ \langle \psi, 1^N \rangle : \psi \text{ is a satisfiable } d\text{-CNF on } N \text{ variables} \} \]

We will prove new limits on efficient compression for these problems with the help of the following powerful, ingenious reduction of Dell and Van Melkebeek.

**Theorem 5.9** (see [DvM10, Lemma 2]). Fix \( d \geq 2 \), and let \( T(n) : \mathbb{N}^+ \to \mathbb{N}^+ \) be polynomially bounded. There is a deterministic polynomial-time OR-compression reduction\(^{27}\) \( R^* \) for \( L = 3\text{-SAT} \)\(^{28}\) with target language \( L' = d\text{-Clique} \). For the first parameter, we have \( t_1(n) = T(n) \). The \( d\)-Clique instance \( \langle (H, s), 1^N \rangle \) output by \( R^* \) satisfies \( N = O \left( n \cdot \max \left( n, T(n)^{1/d+o(1)} \right) \right) \).

By combining Theorem 5.9 with our Theorem 5.4, we will prove the following result.

**Theorem 5.10.** Let \( d \geq 2, \varepsilon > 0 \) be given. There is a \( \beta = \beta(d, \varepsilon) > 0 \) for which the following holds. Suppose that \( d\text{-Clique} \) has a polynomial-time compression reduction with output-size bound \( O(N^{d-\varepsilon}) \) and success probability \( .5 + N^{-\beta} \); that is (in the terms of Definition 3.6), suppose that \( d\text{-Clique} \) is PPT-compressible with parameters \( c, \xi \) satisfying

\[ c(M, N, 1) \leq O(N^{d-\varepsilon}), \quad \xi(M, N, 1) \leq .5 - N^{-\beta} \]

with any target language \( L' \). Then, every language in \( \text{NP} \) is many-to-one reducible in nonuniform polynomial time to a problem in \( \text{pr-SZK} \) (and \( \text{NP} \subseteq \text{coNP/poly} \)).

The same result holds if we replace \( d\text{-Clique} \) with \( d\text{-Vertex Cover} \) or \( d\text{-SAT}_{\text{par}} \).

Theorem 5.10 gives a version of [DvM10, Theorems 1 and 2] that applies to probabilistic reductions with two-sided error. However, our result does not apply to the more general setting of oracle communication protocols, to which those earlier results do apply (for co-nondeterministic protocols and protocols avoiding false negatives).

Dell and Van Melkebeek use their techniques to show compression lower bounds for several other interesting graph problems (including the Feedback Vertex Set, Bounded-Degree Deletion, and Nonplanar Deletion problems) via reductions from 2-vertex cover [DvM10, section 5.2]. Using our results and the reductions in [DvM10], one can also obtain similarly strong compression lower bounds for these problems for the two-sided error setting.

\(^{26}\)This is a different parametrized problem than the two clique-based problems mentioned in section 3.2.

\(^{27}\)This is as in Definition 2.17.

\(^{28}\)Here, 3-SAT is just the usual language \( \{ \langle \psi \rangle : \psi \text{ is a satisfiable 3-CNF} \} \).
Proof of Theorem 5.10. We already described a simple reduction (in both directions) between the $d$-Vertex Cover and $d$-Clique problems that preserves the parameter $N$. Also, an instance of $d$-Vertex Cover on $N$ vertices is efficiently reducible to a $d$-SAT instance over $O(N)$ variables [DvM10, Lemma 5]. Thus, it suffices to prove the result for $d$-Clique.

Let $R$ be the compression reduction assumed to exist for $d$-Clique, with the value $\beta > 0$ to be determined later. Let $C > d$ be a large integer value, also to be determined.

We will define an OR-compression reduction $R'$ for $L = 3$-SAT and target language $L'$ from our assumption; this will allow us to apply Theorem 5.5. $R'$ works as follows. We let $t_1(n) := n^C$. On input formulas $\psi_1, \ldots, \psi_n$, each of bit-length $n$, the reduction first computes $\langle (H, s), 1^N \rangle := R^*(\psi_1, \ldots, \psi_n)$, where $R^*$ is as in Theorem 5.9. Next, $R'$ outputs the value $z := R((H, s), 1^N)$.

$R'$ is clearly polynomial-time computable. To analyze $R'$, fix length-$n$ formulas $\psi_1, \ldots, \psi_n$, and let $b := \bigvee_{j=1}^n [\psi_j \in 3$-$SAT]$. By the OR-compression property of the deterministic mapping $R^*$, we have $[b = 1] \iff \langle (H, s), 1^N \rangle \in d$-Clique. Then by the assumed reliability guarantee of $R$,

$$\Pr[L'(z) = b] \geq .5 + N^{-\beta} \geq .5 + \left(O\left(n \cdot \max\left(n, n^{C/d + o(1)}\right)\right)\right)^{-\beta} \geq .5 + n^{-\beta(1 + C/d + o(1))}.$$  

Thus the error bound $\xi(n)$ of our reduction $R'$ is at most $5 - n^{-\beta(1 + C/d) + o(1)}$. Also, by the compression guarantee of $R$, the output $z$ satisfies

$$|z| \leq O(N^{d - \epsilon}) \leq O\left(\left(n^{1 + C/d + o(1)}\right)^{d - \epsilon}\right) \leq O\left(n^{C - 1 + o(1)}\right)$$

with the last step valid provided we take $C > d(d + 1)/\epsilon$. Thus, as an output-size bound for $R'$, we may take $t_2(n) = O\left(n^{C - 1 + o(1)}\right)$. We evaluate

$$(1 - 2\xi(n))^2 - \sqrt{\frac{\ln 2}{2} \cdot \frac{t_2(n) + 1}{t_1(n)}} \geq 4n^{-2(\beta(1 + C/d) - o(1))} - O(n^{-5 + o(1)}) \geq n^{-\Omega(1)},$$

provided we take $\beta < .25(1 + C/d)^{-1}$. Thus, under these settings, (10) holds. Then Theorem 5.5, item 1, gives the desired conclusion, since $L = 3$-SAT is NP-complete.

6. Extension to quantum compression. In this section, we will show that our results on OR- and AND-compression have analogues for the model in which the compression scheme is allowed to be a quantum algorithm, outputting a quantum state.

We assume familiarity with the basics of quantum computing and quantum information. (For the needed background, consult [NC00].) However, readers without this background should be able to follow the overall structure of the argument if they regard “qubits,” “quantum operations,” “quantum algorithms,” and “quantum measurements” as certain types of black-box objects and accept some known facts about them. In particular, a “mixed state on $m$ qubits” is a kind of “superposition” over classical $m$-bit strings. Let $M_2^n$ denote the collection of $m$-qubit mixed states. $M_2^n$ can be identified with the set of $2^m$-by-$2^m$, trace-1, positive-semidefinite complex matrices.
A quantum operation is a certain type of mapping \( \text{OP} : \text{MS}_m \to \text{MS}_{m'} \) for some \( m, m' > 0 \). (The operations allowed by quantum physics are the completely positive, trace-preserving maps; these are a subset of the linear transformations mapping \( \text{MS}_m \subset \mathbb{C}^{2^m \times 2^m} \) into \( \text{MS}_{m'} \subset \mathbb{C}^{2^{m'} \times 2^{m'}} \).) We let \( \text{OP}_{m, m'} \) denote the valid quantum operations from \( m \)-qubit into \( m' \)-qubit states.

Quantum measurements are measurements performed on quantum states to yield classical information about these states; in the quantum setting, measurements are inherently probabilistic and alter the states being measured. See [NC00, Chapter 2] for formal details. Quantum states turn out to inherit some of the information-theoretic limitations of their classical counterparts; this fact will be the basis for our results on quantum compression.

### 6.1. Trace distance and distinguishability of quantum states.

The **trace distance** is a metric on mixed quantum states from a shared state space [NC00]; we denote the trace distance between \( \rho, \rho' \in \text{MS}_m \) by \( \|\rho - \rho'\|_{\text{tr}} \in [0, 1] \). Formally, treating \( \rho, \rho' \) as matrices, and using the square root operation for the positive semidefinite matrix \( (\rho - \rho')^2 \), \( \|\rho - \rho'\|_{\text{tr}} := \frac{1}{2} \text{Tr} \sqrt{\rho - \rho'}^2 \). This distance is intimately related to the distinguishability of quantum states. Suppose that \( \rho, \rho' \) are two known states, and we are sent one or the other, depending on the outcome of an unbiased coin flip \( b \in \{0, 1\} \). We want to guess \( b \), by applying some series of quantum operations and measurements. For any \( \rho, \rho' \), it is known [NC00, Theorem 9.1] that our success probability at this task is maximized by using a single binary measurement, chosen based on \( \rho, \rho' \), and that our maximum achievable success probability equals \( \frac{1}{2} (1 + \|\rho - \rho'\|_{\text{tr}}) \).

A probability distribution over mixed states is again a mixed state. Thus, for a distribution \( D \) over a finite universe \( U \) and a mapping \( R : U \to \text{MS}_m \), \( R(D) \) defines a quantum state. We use the following standard claim concerning such states, which follows from the distinguishability characterization of \( \|\cdot\|_{\text{tr}} \).

**Claim 6.1.** For any distributions \( D, D' \) over a shared finite universe \( U \), and any mapping \( R : U \to \text{MS}_m \), we have \( \|R(D) - R(D')\|_{\text{tr}} \leq \|D - D'\|_{\text{stat}} \). Similarly, for any valid quantum operation \( \text{OP} \in \text{OP}_{m, m'} \) and states \( \rho, \rho' \in \text{MS}_m \), we have \( \|\text{OP}(\rho) - \text{OP}(\rho')\|_{\text{tr}} \leq \|\rho - \rho'\|_{\text{tr}} \).

### 6.2. Quantum f-compression.

The following notion of quantum compression is modeled on Definition 2.17. We no longer will have a target language for our reduction; instead, we will require that the answer to our original instance of the decision problem \( f \circ L \) be recoverable by some quantum measurement performed on the quantum state output by the reduction. (This measurement need not be efficiently performable.)

**Definition 6.2.** (Quantum f-compression reductions). Let \( L \) be a language, and let \( f : \{0, 1\}^* \to \{0, 1\} \) be a Boolean function. Let \( t_1(n), t_2(n) : \mathbb{N}^+ \to \mathbb{N}^+ \) and \( \xi(n) : \mathbb{N}^+ \to [0, 1] \) be given, with \( t_1, t_2 \) polynomially bounded. A quantum f-compression reduction for \( L \) with parameters \( t_1(n), t_2(n), \xi(n) \) is a mapping \( R(x^1, \ldots, x^m) \) outputting a mixed state \( \rho \), along with a family of (not necessarily efficiently performable) binary quantum measurements \( \{\text{MS}_n\}_{n \geq 0} \) on \( t_2(n) \)-qubit states. We require the following properties: for all \( (x^1, \ldots, x^{t_1(n)}) \in \{0, 1\}^{t_1(n)} \times n \),

1. the state \( \rho = R(x^1, \ldots, x^{t_1(n)}) \) is on \( t_2(n) \) qubits;
2. we have \( \Pr\left[ \text{MS}_n(\rho) = f(L(x^1), \ldots, L(x^{t_1(n)})) \right] \geq 1 - \xi(n) \).

If \( R \) as above is computable in quantum polynomial time, we say \( L \) is QPT-f-compressible with parameters \( (t_1(n), t_2(n), \xi(n)) \).

---

29A binary measurement is a measurement with two possible outcomes.
A consequence of item 2 in the definition above and the quantum background we mentioned is that, for inputs \(\pi,\pi\) with \((f \circ L)(\pi) \neq (f \circ L)(\pi)\), we have \(\|R(\pi) - R(\pi)\|_{1r} \geq 1 - 2\xi(n)\). For our negative results (giving evidence against quantum compression), this is the only property of quantum measurements we need to work with.

6.3. Quantum complexity classes. We will be using the class \(\text{QIP}[k]\) of languages definable by \(k\)-message, quantum interactive proof systems [Wat03]. These are proof systems in which a computationally unbounded prover exchanges quantum messages with a quantum polynomial-time verifier; a total of \(k = k(n)\) messages are exchanged. The verifier sends the first message if \(k\) is even, or the prover if \(k\) is odd, and the parties alternate thereafter.

Our treatment of these classes will be informal, since all the technical properties we need are summarized in results from prior work (for details see [Wat03, KW00, Wat02, JUW09]). In the definition, we will require completeness \(2/3\) and soundness \(1/3\) (as is typically done). We take \(\text{QIP} := \bigcup_{k \geq 0} \text{QIP}[k]\). It was shown in [Wat03, KW00] that for any \(3 \leq k(n) \leq \text{poly}(n)\), \(\text{PSPACE} \subseteq \text{QIP}[k(n)] = \text{QIP}[3]\); the latter class was recently shown to equal \(\text{PSPACE} [JUW11]\). Importantly for us, however, the class \(\text{QIP}[2]\) is not known to contain even \(\text{coNP}\).

In what follows, we will actually want to work with the promise-problem classes \(\text{pr-\text{QIP}}[k]\). In this section, we will freely state more general results about promise classes when they are provable by the same techniques as used in the cited papers to study languages.

For our work, it is important to know that a quantum interactive proof system with even a small completeness-soundness gap can be amplified without increasing the number of rounds of interaction [KW00, JUW09]. In particular, we will use this fact for \(k = 2\).

**Theorem 6.3** (follows from [JUW09, section 3.2], building on [KW00]). Suppose a promise problem \(\Pi\) has a two-message quantum interactive proof system with completeness and soundness guarantees \(c(n), s(n)\) obeying \(c(n) - s(n) \geq 1/\text{poly}(n)\).

Then there is a nonuniform polynomial-time many-to-one reduction from \(\Pi\) to a promise problem in \(\text{pr-\text{QIP}}[2]\). If \(c, s\) are computable in time \(\text{poly}(n)\), then \(\Pi \in \text{pr-\text{QIP}}[2]\).

A model of quantum statistical zero-knowledge proofs was proposed by Watrous [Wat02] and used to define the class \(\text{QSZK}\) of promise problems having polynomial-time proof systems of this type.\(^3\) We will use \(\text{pr-\text{QSZK}}\) to denote this class. Watrous showed in [Wat02] that Sahai and Vadhan’s “statistical distance characterization” of \(\text{pr-\text{SZK}}\), embodied in Definition 2.14, has a quantum analogue. For a quantum circuit \(C\) with an \(m\)-qubit output register, let \(\rho_C\) denote the output state of \(C\) when given the all-zeros state as input. We consider circuits built from a fixed, finite “universal” gate-set (see [NC00, Chapter 4]).

**Definition 6.4.** For parameters \(0 \leq d < D \leq 1\), define the promise problem \(\text{TD}^D_d \subseteq \text{QSZK}\) by \(\Pi_Y := \{(C, C') : \|\rho_C - \rho_{C'}\|_{1r} \geq D\}\), \(\Pi_N := \{(C, C') : \|\rho_C - \rho_{C'}\|_{1r} \leq d\}\). Here, both \(d = d(n)\) and \(D = D(n)\) may be parameters depending on the input length \(n = \|\langle C, C'\rangle\|\).

Then, appealing to the result of [Wat02], we can use the following definition.

---

\(^3\)Watrous’s original model was of honest-verifier quantum statistical zero-knowledge proof systems; he later showed that these proof systems are equivalent in power to “cheating-verifier” ones [Wat09].

© 2015 Andrew Drucker
DEFINITION 6.5. Let $\text{pr-QSZK}$ be defined as the class of promise problems for which there is a many-to-one (classical, deterministic) polynomial-time reduction from $\Pi$ to $\text{TD}_{\leq 1/3}^{2/3}$.

It was shown in [Wat02] that $\text{pr-QSZK}$ is contained in $\text{pr-QIP}[2] \cap \text{pr-coQIP}[2]$ and is closed under complement. We do not know if, e.g., $\text{TD}_{\leq 0.49}^{2/3}$ is in $\text{pr-QSZK}$, but we have the following.

**Theorem 6.6.** Suppose that $0 \leq d = d(n) < D = D(n) \leq 1$ satisfy $D > d + \frac{1}{\text{poly}(n)}$. Then, $\text{TD}_{\leq d}^{D}$ is many-to-one reducible in nonuniform (classical, deterministic) polynomial time to a problem in $\text{pr-QIP}[2]$. If $d(n), D(n)$ are computable in time $\text{poly}(n)$, then $\text{TD}_{\leq d}^{D} \in \text{pr-QIP}[2]$.

**Proof (sketch).** Let $(C, C')$ be the input quantum circuit descriptions. Consider the two-message “distinguishing” protocol, in which the verifier privately chooses one of the two circuits at random and generates its output quantum state $\rho$ (on the all-zeros input state) by direct simulation. The verifier then sends the state to the prover, who is asked to guess which of the two circuits was used.

By the distinguishability interpretation of trace distance, this gives a two-message proof for $\text{TD}_{\leq d}^{D}$ with completeness-soundness gap $\geq 1/\text{poly}(n)$. The desired result follows from Theorem 6.3.

Unlike the classical case, there is no known “nonuniform derandomization” result known for $\text{QIP}[2]$ (or for other quantum classes). However, we do have a satisfying analogue of Theorem 2.15 as follows.

**Theorem 6.7** (follows from [Wat02]). Suppose that $0 \leq d = d(n) < D = D(n) \leq 1$ satisfy $D^{2} > d + \frac{1}{\text{poly}(n)}$. Then $\text{TD}_{\leq d}^{D}$ is many-to-one reducible in nonuniform polynomial time to a problem in $\text{pr-QSZK}$. If $d(n), D(n)$ are computable in time $\text{poly}(n)$, then $\text{TD}_{\leq d}^{D} \in \text{pr-QSZK}$.

### 6.4. Quantum distributional stability

We will use a quantum analogue of the distributional stability property.

**Definition 6.8.** Let $t, t', n \in \mathbb{N}^{+}$. Given a mapping $F : \{0, 1\}^{t \times n} \to \text{MS}_{t'}$, and a collection $D_{1}, \ldots, D_{t}$ of mutually independent distributions over $\{0, 1\}^{n}$ for $j \in [t]$, let

$$
\gamma_{j} := \mathbb{E}_{y \sim D_{j}} \left[ ||F(D_{1}, \ldots, D_{j-1}, y, D_{j+1}, \ldots, D_{t}) - F(D_{1}, \ldots, D_{t})||_{1} \right].
$$

For $\delta \in [0, 1]$, we say that $F$ is $\delta$-quantum-distributionally stable (or $\delta$-QDS) for $D_{1}, \ldots, D_{t}$ if $\frac{1}{t} \sum_{j=1}^{t} \gamma_{j} \leq \delta$.

**Lemma 6.9.** Let $t, t', n \in \mathbb{N}^{+}$. Let $R : \{0, 1\}^{t \times n} \to \text{MS}_{t'}$ be given. Then, $R$ is $\delta$-QDS with respect to any input distributions $D_{1}, \ldots, D_{t}$, where we may take $\delta := \min \left\{ \sqrt{\frac{\ln 2}{t}} \frac{|t'|}{t}, 1 - 2^{\frac{|t'|}{t} - 2} \right\}$.

The slight improvement in the bounds comes from the fact that $R$ outputs exactly $t'$ qubits. The proof of Lemma 6.9 is very similar to that of Lemma 4.2 but requires further background.\footnote{A result of this kind for the uniform distribution, with a bound similar to the first bound given above, was shown in [KNTSZ07], and we follow similar steps.} In the rest of section 6.4, we will describe how Lemma 6.9 is proved.

We use various concepts and results of quantum information theory. See [NC00, and Chapter 11 in particular] for more background. In particular, we assume familiarity with the notion of bipartite and reduced states. For a bipartite state...
\(\rho_{A,B}\) on subsystems \(A,B\), we let \(\rho_A\) (resp., \(\rho_B\)) denote the reduced state over \(A\) (resp., \(B\)) We let \(S(\rho) := -\operatorname{Tr}(\rho \log_2 \rho)\) denote the Von Neumann entropy of a quantum state (here, identifying \(\rho\) with its density matrix). In analogy to the classical case, the entropy of a \(d\)-qubit state can be shown to be at least 0 and at most \(d\). We define the quantum mutual information between subsystems \(A,B\) of \(\rho_{A,B}\) as \(I_q(A;B) := S(\rho_A) + S(\rho_B) - S(\rho_{A,B})\).

**Lemma 6.10** (see [NC00, Theorem 11.8.5, p. 513]). Consider a quantum system \(\rho_{X,Y}\) with a subsystem that is a classical random variable \(X\), that is, a state of form \(\rho_{X,Y} = \sum_{x \in \text{supp}(X)} \Pr[X = x] |x\rangle \langle x| \otimes \sigma_x\). Then, \(S(\rho_{X,Y}) = H(X) + \sum_x \Pr[X = x] S(\sigma_x)\).

We have the following basic bound on the quantum mutual information between a classical message and its quantum encoding.

**Lemma 6.11.** For a classical random variable \(X\), and a state \(\rho_{X,Y}\) as in Lemma 6.10, with the states \(\{\sigma_x\}\) on \(d\) qubits, we have \(I_q(X;Y) \leq d\).

**Proof.** Using Lemma 6.10, we calculate

\[
I_q(X;Y) = S(\rho_X) + S(\rho_Y) - S(\rho_{X,Y}) = H(X) + S(\rho_Y) - H(X) - \sum_x \Pr[X = x] S(\sigma_x)
\]

\[
= S(\rho_Y) - \sum_x \Pr[X = x] S(\sigma_x) \leq d,
\]

since \(\rho_Y\) consists of \(d\) qubits and \(S(\sigma_x) \geq 0\) for each \(x\).

Not all properties of classical entropy and mutual information are inherited by their quantum counterparts.\(^3\) However, we have [Nay99, p. 33 and Appendix A].

**Fact 6.12.** Quantum mutual information obeys the following properties for all \(X,Y,Z\):

1. \(I_q(X;Y) = I_q(Y;X) \geq 0\).
2. \(I_q(X;Y,Z) = I_q(X;Y) + I_q((X,Y);Z) - I_q(Y;Z)\).
3. *(Strong subadditivity)* \(I_q(X;Y,Z) \geq I_q(X;Y)\).
4. \(I_q(X;Z) = 0\) if the subsystems \(X,Z\) are independent classical random variables.

Item 3 is a quite nontrivial fact in the quantum setting, with multiple formulations [NC00, Chapter 11]. With these facts in hand, the proof of Lemma 6.13 below mimics that of Lemma 2.6.

**Lemma 6.13.** If \(X^1,\ldots,X^t\) are independent classical random variables and \(Y\) a quantum subsystem, then \(I_q(Y;(X^1,\ldots,X^t)) \leq \sum_{j \in [t]} I_q(Y;X^j)\).

Next, we need quantum analogues of Pinsker’s and Vajda’s inequalities. For mixed states \(\rho,\sigma\) over the same number of qubits, define the relative entropy (a quantum analogue of KL divergence) as \(S(\rho|\sigma) := \operatorname{Tr}(\rho \log_2 \rho) - \operatorname{Tr}(\rho \log_2 \sigma)\).

We also have the following analogue of Fact 2.8 [KNTSZ07, p. 10].

**Fact 6.14.** \(I_q(A;B) = S(\rho_{AB}) - S(\rho_A \otimes \rho_B)\).

A quantum Pinsker inequality was explicitly proved in [KNTSZ07, Theorem III.1].\(^3\) However, that proof actually demonstrates a more general principle as follows.

---

\(^3\)For example, it is not generally true that \(S(\rho_{AB}) \leq S(\rho_A)\) by analogy with Fact 2.3, which tells us that \(H(X,Y) \geq H(X)\). Note, though, that we don’t use this classical fact in proving Lemma 4.2.

\(^3\)An earlier version appears in [OP93]. We note that [KNTSZ07] states its results in terms of the trace norm, which is twice the trace distance.
**Theorem 6.15** (see [KNTSZ07]). Suppose that for some \(\alpha, \beta \geq 0\), the statistical distance and KL divergence obey the relation \(\|X - Y\|_{\text{stat}} \geq \alpha \implies D_{\text{KL}}(X \| Y) \geq \beta\) for every pair of classical distributions \(X, Y\). Then, for any pair \(\rho, \sigma\) of quantum states, \(\|\rho - \sigma\|_{\text{tr}} \geq \alpha \implies S(\rho \| \sigma) \geq \beta\).

Combining this principle with the classical Pinsker and Vajda inequalities, we obtain the following.

**Corollary 6.16** (Quantum Pinsker and Vajda inequalities). For any states \(\rho, \sigma\),

\[
S(\rho \| \sigma) \geq \max \left\{ \frac{2}{\ln 2} \cdot \|\rho - \rho'\|^2_{\text{tr}}, \frac{1}{\ln 2} \left( \ln \left( \frac{1}{1 - \|\rho - \sigma\|_{\text{tr}}} \right) - 1 \right) \right\}.
\]

In the quantum setting, we let \(R\) denote the mixed quantum state \(R(X^1, \ldots, X^t)\), where \(X^j \sim D_j\). The inequality \(I_q((X^1, \ldots, X^t); R) \leq t'\) follows from Lemma 6.11, since \(R \in \mathcal{MS}_p\). With the assembled tools in hand, the proof of Lemma 6.9 is essentially identical to that of Lemma 4.2. The one difference is that the classical equality

\[
\|(X^j, R) - (Y^j, R)\|_{\text{stat}} = E_{x^j \sim D_j} \left[ || R(D_1, \ldots, D_{j-1}, x^j, D_{j+1}, \ldots, D_t) - R(D_1, \ldots, D_t) ||_{\text{stat}} \right]
\]

that we used there is replaced by the inequality

\[
\|(X^j, R) - (Y^j \otimes R)\|_{\text{tr}} \geq E_{x^j \sim D_j} \left[ || R(D_1, \ldots, D_{j-1}, x^j, D_{j+1}, \ldots, D_t) - R(D_1, \ldots, D_t) ||_{\text{tr}} \right].
\]

This inequality follows by considering the experiment that first measures the \(X^j\) register and then performs an optimal distinguishing measurement on \(R\) conditioned on the outcome of the first measurement. Note that this inequality goes in the needed direction.

### 6.5. Building quantum disguising distributions

Next, we give quantum analogues of our Disguising-Distribution lemmas. First, we have the following analogue of Lemma 4.3.

**Lemma 6.17.** Let \(t, t' \in \mathbb{N}^+\), and let \(F(x^1, \ldots, x^T) : \{0,1\}^{T \times n} \rightarrow \mathcal{MS}_{t'}\) be given. Suppose that \(F\) is \(\delta\)-QDS with respect to input distribution \(D^\otimes T\) for every distribution \(D\) over \(\{0,1\}^n\).

Fix some distribution \(D\) over \(\{0,1\}^n\), and let \(x^1, \ldots, x^d\) be independently sampled from \(D\). Let \(k^* \sim \mathcal{U}_d\). Let \(\hat{D}\) be the distribution that samples uniformly from the multiset \(\{k^*\}_{k \neq k^*}\). Let \(\beta_j := E_{k^*, x^1, \ldots, x^d} \left[ || F(\hat{D}^\otimes (j-1), x^{k^*}, \hat{D}^\otimes (T-j)) - F(\hat{D}^\otimes T) ||_{\text{tr}} \right]\), where all the \(\hat{D}\)s are to be mutually independent (for fixed values of \(x^1, \ldots, x^k\) and \(k^*\)). Then, \(\frac{1}{T} \sum_{j=1}^T \beta_j \leq \delta + 2T/d\).

**Proof.** The proof is identical to that of Lemma 4.3, except that we replace statistical distance with trace distance (where appropriate—the input distributions we manipulate still are to be compared in statistical distance) and appeal to Claim 6.1 to argue that applying \(F\) does not increase trace distance between states.

After giving quantum analogues of Lemmas 4.4 and 4.5 (by closely analogous proofs), we have the following.
LEMMA 6.18 (Quantum Disguising-Distribution lemma). Let \( R(x^1, \ldots, x^t) : \{0,1\}^{t \times n} \rightarrow \text{MS}_d \) be any possibly randomized mapping, where \( n, t, t' \in \mathbb{N}^+ \). Let \( S \subseteq \{0,1\}^n \), and fix \( d > 0 \). Given any \( \varepsilon > 0 \), let \( s := \lceil (5 \ln 2)n/\varepsilon^2 \rceil \). Let \( \hat{\delta} := \min \left\{ \sqrt{\ln 2 \cdot \frac{n}{t'}}, \ 1 - 2^{t'/2} - 2^{-s} \right\} \). Then there exists a collection \( K_1, \ldots, K_s \) of size-\( d \) multisets contained in \( S \), such that for every \( y \in S \), we have

\[
\mathbb{E}_{a \sim \mathcal{U}_{\{0,1\}^t}, y^* \sim \mathcal{U}_{\{0,1\}^t}} \left[ \left\| R \left( \mathcal{U}_{K_a}^{(s-1)}, y, \mathcal{U}_{K_a}^{(t-j)} \right) - R \left( \mathcal{U}_{K_a}^{(s)} \right) \right\|_{\text{tr}} \right] \leq \hat{\delta} + 2t/(d+1) + \varepsilon.
\]

6.6. Complexity upper bounds from quantum compression schemes.

Our lemmas imply a quantum analogue of Theorem 5.1 as follows.

**Theorem 6.19.** Let \( L \) be any language. Suppose there is a QPT-\( \text{OR} \)-compression reduction \( R(x^1, \ldots, x^t) : \{0,1\}^{t_1(n) \times n} \rightarrow \text{MS}_{t_2(n)} \) for \( L \) with polynomially bounded parameters \( t_1(n), t_2(n) : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \), and with error bound \( \xi(n) < 0.5 \). Let \( \hat{\delta} := \min \left\{ \sqrt{\ln 2 \cdot \frac{t_2(n)}{t_1(n)}}, \ 1 - 2^{t_1(n) - t_2(n)} - 2^{-s} \right\} \).

1. If for some \( c > 0 \) we have \( (1 - 2\xi(n)) - \hat{\delta} \geq 1/n^c \), then there is a nonuniform (classical, deterministic) polynomial-time many-to-one reduction from \( L \) to a problem in \( \text{pr-QIP}[2] \).
2. If we have the stronger bound \( (1 - 2\xi(n))^2 - \hat{\delta} \geq 1/n^c \), then \( L \) has a nonuniform (classical, deterministic) polynomial-time many-to-one reduction to a problem in \( \text{pr-QSZK} \).

**Proof of Theorem 6.19.** The proof is closely analogous to that of Theorem 5.1, except that our nonuniform reduction, on input \( y \), outputs a description \( \langle C, C' \rangle \) of a pair of quantum circuits. If \( y \in L \), then \( \|\rho_C - \rho_{C'}\|_{\text{tr}} \geq D(n) := 1 - 2\xi(n) \), while if \( y \notin L \), we have \( \|\rho_C - \rho_{C'}\|_{\text{tr}} \leq d(n) := \hat{\delta} + 1/n^c \). Applying Theorems 6.6 and 6.7 gives us the complexity upper bounds in items 1 and 2. \( \square \)

Using Theorem 6.19, we can prove quantum versions of Theorems 5.4 and 5.5, giving evidence against efficient quantum OR- and AND-compression for NP-complete languages, under the assumption that no coNP-complete language is nonuniformly reducible to a problem in \( \text{pr-QIP}[2] \), or alternatively, to a problem in \( \text{pr-QSZK} \).

7. Questions for further study.

1. Can we extend our results on the infeasibility of compression for AND(SAT) to give corresponding lower bounds on the cost of solving this problem by oracle communication protocols as studied in [DvM10]? Can we do the same for two-sided error protocols for OR(SAT)?
2. Using our results on the infeasibility of compression for AND(SAT), can we extend the work of [DvM10] to prove new kernel-size lower bounds for interesting problems with polynomial kernels, under the assumption \( \text{NP} \not\subseteq \text{coNP/poly} \)?
3. Can we find other applications for the Disguising-Distribution lemma?

**Acknowledgments.** I thank Scott Aaronson, Hans Bodlaender, Holger Dell, Lance Fortnow, Russell Impagliazzo, Bart Jansen, James Lee, Dieter van Melkebeek, Ashwin Nayak, Karolina Soltys, Salil Vadhan, Thomas Vidick, Avi Wigderson, Ryan Williams, and several reviewers for helpful comments. Thanks especially to Russell, Ashwin, and Salil for allowing me to include their alternative proof suggestions.

© 2015 Andrew Drucker
REFERENCES


[Buh] H. Buhrman, Personal communication.


NEW LIMITS TO INSTANCE COMPRESSION


