Manipulating Statistical Difference

Amit Sahai and Salil Vadhan

ABSTRACT. We give several efficient transformations for manipulating the statistical difference (variation distance) between a pair of probability distributions. The effects achieved include increasing the statistical difference, decreasing the statistical difference, "polarizing" the statistical relationship, and "reversing" the statistical relationship. We also show that a boolean formula whose atoms are statements about statistical difference can be transformed into a single statement about statistical difference. All of these transformations can be performed in polynomial time, in the sense that, given circuits which sample from the input distributions, it only takes polynomial time to compute circuits which sample from the output distributions.

By our prior work (see FOCS 97), such transformations for manipulating statistical difference are closely connected to results about SZK, the class of languages possessing statistical zero-knowledge proofs. In particular, some of the transformations given in this paper are equivalent to the closure of SZK under complement and under certain types of Turing reductions. This connection is also discussed briefly in this paper.

1. Introduction

Statistical difference, also known as variation distance, is a fundamental measure of similarity between probability distributions. This measure is the most appropriate choice for many applications in algorithms and cryptography, so it is natural to seek efficient means of manipulating it. In this paper, we exhibit polynomial-time transformations mapping pairs of distributions to pairs of distributions which have the following effects:

A **(Increasing Statistical Difference)** "Noticeable" statistical difference is driven exponentially close to 1 (whereas negligible statistical difference remains negligible). This is a well-known technique — we simply repeat each distribution many times independently.

B **(Decreasing Statistical Difference)** Statistical difference that is bounded away from 1 is driven exponentially close to 0 (whereas statistical difference that is negligibly close to 1 remains as such). This is achieved by an XOR construction inspired by a technique of [**DDPY94**].

1991 *Mathematics Subject Classification.* Primary 68Q25, 60E05; Secondary 68Q15.

Both authors are supported by DOD/NDSEG Graduate Fellowships and partially by DARPA grant DABT63-96-C-0018.

A longer paper, containing the results of this paper along with those of [**SV97**], plus some additional results, will be made available at `http://www-math.mit.edu/~salil` and `http://theory.lcs.mit.edu/~amits`.

C **(Polarizing the Statistical Relationship)** If the original distributions have (moderately) large statistical difference, then the resulting distributions will have statistical difference exponentially close to 1, and if the original distributions have (moderately) small statistical difference, then the resulting ones will have statistical difference exponentially close to 0.

D **(Reversing the Statistical Relationship)** If the original distributions have small statistical difference, the resulting ones will have large statistical difference, and if the original ones have large statistical difference, the resulting ones will have small statistical difference.

Of these, Transformations C and D are the main new contributions of this work, though we also describe Transformations A and B in detail, as we make use of them for the former. The final result of this paper is an application of all of these techniques to show that a boolean formula whose atoms are statements about statistical difference can be efficiently transformed into a single statement about statistical difference. All of these results are discussed in more detail below.

Our initial motivation for addressing these questions arose from our recent work [**SV97**] showing a close relationship between statistical difference and statistical zero-knowledge (SZK) proofs. Specifically, the problem of distinguishing between pairs of (succinctly described) distributions with large statistical difference and pairs with small statistical difference was shown to be *complete* for SZK, the class of languages possessing statistical zero-knowledge proofs [**SV97**]. By that result, efficient transformations for manipulating statistical difference yield closure properties of SZK, and conversely. For example, one of the transformations given here is equivalent to the closure of SZK under complementation, and another is equivalent to the closure of SZK under certain types of Turing (or Cook) reductions. In fact, some of these transformations were developed in [**SV97**] in order to prove things about SZK, whereas others were obtained by extracted ideas which appeared in [**DDPY94, Oka96, SV97**] stated in terms of SZK.

Another reason for looking at statistical difference is that many computational problems of interest can be cast as statistical difference problems; examples include Quadratic Residuosity [**GMR89**], Graph Isomorphism [**GMW91**], and approximate versions of the Closest and Shortest Vector Problems [**GG98**]. Indeed, statistical zero-knowledge proofs are often constructed for such problems based on this observation.

1.1. Formal Setting.

If X and Y are probability distributions (or random variables) on a discrete space D, the *statistical difference* between X and Y is defined to be

$$\|X - Y\| = \max_{S \subset D} |\Pr[X \in S] - \Pr[Y \in S]|.$$

In this paper, we focus on distributions Z which have a "succinct description" which enables them to be sampled efficiently. By "succinct description" we mean a circuit C which, when fed the uniform distribution, has output distribution Z. For example, if C has m input gates and n output gates, Z would be the distribution induced on $\{0,1\}^n$ when C is fed the uniform distribution on $\{0,1\}^m$. Thus, when we speak of "efficient transformations" of pairs of distributions, we mean that there is a polynomial-time computable function on pairs of circuits that achieves the desired transformation on the corresponding pair of distributions. For notational convenience, we write C for both the circuit and the distribution Z it defines.

1.2. Contrast with the "Standard" Setting. In the setting of randomized algorithms, there is typically a single probability space at hand, and one is interested in the probability of some event (such as the algorithm giving the correct answer.) To achieve various effects on the probability of this event (such as decreasing the error probability), a Chernoff bound argument is often all that is needed. In the setting of statistical difference, there are *two* probability spaces at hand, and we are interested in the maximum *difference* in probabilities, over all possible events. Because of this, we are limited in what we can achieve with Chernoff bounds alone. Chernoff bounds do, however, enable us to analyze the following *direct product* construction: Suppose we have a pair of distributions X_0 and X_1, and we let Y_0 and Y_1 consist of k independent copies of X_0 and X_1, respectively. Then a Chernoff bound argument tells us that

$$\|X_0 - X_1\| > \epsilon \Rightarrow \|Y_0 - Y_1\| > 1 - e^{-\Omega(k\epsilon^2)}. \tag{1.1}$$

Moreover, one can also show that

$$\|X_0 - X_1\| < \epsilon \Rightarrow \|Y_0 - Y_1\| < k\epsilon.$$

This gives us Transformation A — noticeable statistical differences are driven exponentially close to 1, whereas negligible statistical differences remain negligible.

1.3. Polarizing the Statistical Relationship. Our first main result is a transformation which "polarizes" the statistical relationship between two distributions. That is, pairs of distributions which are statistically close become much closer and pairs of distributions which are statistically far apart become much further apart. That is, we exhibit a polynomial-time computable transformation which takes a triple $(C_0, C_1, 1^k)$, where C_0 and C_1 are circuits and produces a new pair of circuits (D_0, D_1) such that

$$\begin{aligned} \|C_0 - C_1\| > 2/3 \quad &\Rightarrow \quad \|D_0 - D_1\| > 1 - 2^{-k} \\ \|C_0 - C_1\| < 1/3 \quad &\Rightarrow \quad \|D_0 - D_1\| < 2^{-k}. \end{aligned}$$

Note that this is *not* achieved by the direct product construction described above. Looking back at Equation 1.1, we see that the statistical difference will go to 1 in both cases, whereas we want the statistical difference to go to 1 in the first case and 0 in the second case. Thus, the Polarization Lemma[1] is achieved by carefully combining the direct product construction with another construction which decreases statistical difference.

To decrease statistical difference, we show that, given two pairs of distributions, we can efficiently produce a third pair of distributions whose statistical difference is exactly the *product* of the original two statistical differences. Iterating this drives statistical differences which are bounded away from 1 to 0 as desired, whereas statistical differences that are negligibly close to 1 remain as such. The construction is based on the intuition that the hardness of guessing the XOR of two bits should be the "product" of the hardnesses of the bits individually, and is inspired by [**DDPY94**].

[1] The Polarization Lemma stated here is called the Amplification Lemma in [**SV97**]. We change the name here to stress that the Polarization Lemma does not merely increase statistical difference.

1.4. Reversing the Statistical Relationship. Our second result is perhaps even more unexpected than the Polarization Lemma — we exhibit an efficient transformation that "reverses" the statistical relationship between a pair of distributions. That is, we show that there is polynomial-time computable function that maps pairs of circuits (C_0, C_1) to pairs of circuits (D_0, D_1) such that

$$\begin{aligned} \|C_0 - C_1\| < 1/3 &\quad\Rightarrow\quad \|D_0 - D_1\| > 2/3 \\ \|C_0 - C_1\| > 2/3 &\quad\Rightarrow\quad \|D_0 - D_1\| < 1/3 \end{aligned}$$

The techniques used in this transformation come from work on statistical zero-knowledge [**Oka96, SV97**], as discussed below.

1.5. Statistical Zero-Knowledge. Informally, zero-knowledge proofs [**GMR89**] are interactive proofs in which the verifier "learns nothing" other than the assertion being proven. A *statistical* zero-knowledge proof is a type of zero-knowledge proof in which the "learns nothing" condition is interpreted in a strong information-theoretic sense. Statistical zero-knowledge proofs are of interest both in cryptography and complexity, and the set of languages possessing such proofs, SZK, has been studied in a number of recent works.

In prior work [**SV97**], we related SZK to statistical difference by proving that the problem of deciding whether the statistical difference between two distributions is large or small is *complete* for SZK. This can be formally described as a "promise" problem[2] Statistical Difference (abbreviated SD) whose YES instances are pairs of circuits whose statistical difference is greater than 2/3 and whose NO instances are pairs whose statistical difference is less than 1/3:

$$\begin{aligned} \mathrm{SD}_Y &= \left\{(C_0, C_1) : \|C_0 - C_1\| > \frac{2}{3}\right\} \\ \mathrm{SD}_N &= \left\{(C_0, C_1) : \|C_0 - C_1\| < \frac{1}{3}\right\} \end{aligned}$$

The main theorem in relating SD to SZK is the following:

Theorem 1.1 ([**SV97**]). *SD is complete for* SZK*. That is,* $\mathrm{SD} \in \mathsf{SZK}$*, and every problem in* SZK *reduces to* SD *(via a many-one polynomial-time reduction).*

By Theorem 1.1, efficient transformations for manipulating statistical difference can yield closure properties of SZK, and conversely. For example, by Theorem 1.1, the Reversal Mapping described above is equivalent to the closure of SZK under complement. In fact, the existence of such a transformation was originally deduced from the fact that SZK is closed under complement [**Oka96, SV97**]. This result motivated our search for a more explicit description of such a mapping. By extracting ideas used in the transformations of statistical zero-knowledge proofs given in [**Oka96**] and [**SV97**], we obtained the description of this transformation given in this paper. The Polarization Lemma, on the other hand, was originally developed for the purpose of proving things about SZK [**SV97**], but now also serves as an essential tool in the construction of our other transformations on statistical difference.

[2] A *promise problem* is simply a decision problem in which some inputs are excluded [**ESY84**].

1.6. Boolean Closure. The final result of this paper is an application of all of the above transformations to show that a boolean formula whose atoms are statements about membership in SD can be efficiently transformed into a single statement about SD. This is a strengthening of a result of [**DDPY94**] who show how to do this for *monotone* formulae whose atoms are statements about *random self-reducible languages* (which can be reduced to an extreme case of SD in which the thresholds are 1 and 0).

This result is based on the observation that some of transformations described above can be interpreted as boolean operations on statistical difference. For example, suppose we let distribution Z_0 consist of a copy of of distribution X_0 followed by an independent copy of distribution Y_0, and similarly let Z_1 consist of X_1 followed by Y_1. Then we see that if either X_0 and X_1 or Y_0 and Y_1 are statistically far apart, then Z_0 and Z_1 will be statistically far apart. Similarly, if both pairs (X_0, X_1) and (Y_0, Y_1) are statistically very close, then Z_0 and Z_1 will be statistically close. Thus, this operation in some sense represents OR. Similarly, the XOR construction mentioned earlier represents AND, and the Reversal Mapping represents negation. Combining these operations with the Polarization Lemma, we see that, given a k-ary formula ϕ and k pairs of input distributions, we can produce a pair of distributions whose statistical difference indicates whether or not the formula is true when its variables are set according to whether the corresponding pairs of input distributions are statistically far or statistically close. How efficient is this procedure? We show that a careful implementation of this procedure, using these particular AND or OR operations, can be performed in time polynomial in the size of ϕ and the circuits describing the input distributions.

By Theorem 1.1, this implies a very strong boolean closure property of SZK, one that does not necessarily follow from the fact that SZK is closed under complement, union, and intersection. As explained in Section 6, this can also be viewed as closure under a weak form of polynomial-time Turing reductions, and a step towards determining whether SZK is closed under general polynomial-time Turing reductions.

2. Notation and Basic Facts

First, we introduce some notation that will be used throughout the paper. If X is a probability distribution (or random variable), we write $x \leftarrow X$ to indicate that x is a sample taken from X. If S is a set, we write $x \in_R S$ to indicate that x is uniformly selected from S. In this paper, we focus on probability distributions that have a "succinct description" which enables them to sampled efficiently. By "succinct distribution" we mean a circuit C which, when fed the uniform distribution, has output distribution Z. For example, if C has m input gates and n output gates, Z would be the distribution induced on $\{0,1\}^n$ by feeding C the uniform distribution on $\{0,1\}^m$. For notational convenience, we write C for both the circuit and the distribution Z it defines.

Recall the definition of statistical difference given in Section 1.1. For probability distributions (or random variables) X and Y on a discrete set D,

$$\|X - Y\| = \max_{S \subset D} |\Pr[X \in S] - \Pr[Y \in S]|.$$

This is often also called the *variation distance* between X and Y. There is an equivalent formulation of statistical difference in terms of the ℓ_1 norm $|\cdot|_1$ that will sometimes be more convenient for us. To every probability distribution X on a discrete set D, the *mass function* of X is a vector in $\mathbb{R}^D$ whose x'th coordinate is $\Pr[X = x]$. For the sake of elegance, we also denote this vector by X. With this notation, we can state the following well-known fact.

Fact 2.1. $\|X - Y\| = \frac{1}{2}|X - Y|_1$.

Proof. For any set $S \subset D$,

$$\begin{aligned} & 2\,|\Pr[X \in S] - \Pr[Y \in S]| \\ &\quad = |\Pr[X \in S] - \Pr[Y \in S]| + |\Pr[X \notin S] - \Pr[Y \notin S]| \\ &\quad = \left|\sum_{x \in S}(\Pr[X = x] - \Pr[Y = x])\right| + \left|\sum_{x \notin S}(\Pr[X = x] - \Pr[Y = x])\right| \\ &\quad \leq \sum_{x \in S}|\Pr[X = x] - \Pr[Y = x]| + \sum_{x \notin S}|\Pr[X = x] - \Pr[Y = x]| \\ &\quad = |X - Y|_1 . \end{aligned}$$

Equality is achieved by taking $S = \{x\colon\ \Pr[X = x] > \Pr[Y = x]\}$. □

It is immediate from this characterization of statistical difference that it is a metric (as long as we identify random variables that are identically distributed). In particular, it satisfies the Triangle Inequality.

Fact 2.2 (Triangle Inequality). *For any probability distributions X, Y, and Z, $\|X - Y\| \leq \|X - Z\| + \|Z - Y\|$.*

Recall that for any two vectors $v \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$, their *tensor product* $v \otimes w$ is the vector in $\mathbb{R}^{nm}$, whose (i, j)'th component is $v_i w_j$. Now, if we have a pair of random variables (X, Y) (on the same probability space) taking values in $D \times E$, then X is independent from Y iff the corresponding mass functions satisfy $(X, Y) = X \otimes Y$, where we view the mass functions of X and Y as elements of $\mathbb{R}^D$ and $\mathbb{R}^E$, respectively. For this reason, if we have random variables X and Y taking values in sets D and E, respectively, we write $X \otimes Y$ for the random variable taking values in $D \times E$ which consists of independent samples of X and Y.

Now, for any two vectors v and w, $|v \otimes w|_1 = |v|_1 \cdot |w|_1$. In addition, for any mass function X, $|X|_1 = 1$. These facts enable one to show that the statistical difference behaves well with respect to independent random variables:

Fact 2.3. *Suppose X_1 and X_2 are independent random variables on one probability space and Y_1 and Y_2 are independent random variables on another probability space. Then,*

$$\|(X_1, X_2) - (Y_1, Y_2)\| \leq \|X_1 - Y_1\| + \|X_2 - Y_2\| .$$

Proof.

$$\begin{aligned} \|(X_1, X_2) - (Y_1, Y_2)\| &\leq \|(X_1, X_2) - (Y_1, X_2)\| + \|(Y_1, X_2) - (Y_1, Y_2)\| \\ &= \frac{1}{2}|X_1 \otimes X_2 - Y_1 \otimes X_2|_1 + \frac{1}{2}|Y_1 \otimes X_2 - Y_1 \otimes Y_2|_1 \\ &= \frac{1}{2}|(X_1 - Y_1) \otimes X_2|_1 + \frac{1}{2}|Y_1 \otimes (X_2 - Y_2)|_1 \\ &= \frac{1}{2}|X_1 - Y_1|_1 \cdot |X_2|_1 + \frac{1}{2}|Y_1|_1 \cdot |X_2 - Y_2|_1 \\ &= \|X_1 - Y_1\| + \|X_2 - Y_2\| \end{aligned}$$

□

One basic fact about statistical difference is that it cannot be created out of nothing. That is, for any procedure A, even if it be randomized, the statistical difference between $A(X)$ and $A(Y)$ is no greater than the statistical difference betewen X and Y. Formally, if D is any set, a *randomized procedure* on D is a a pair $A = (f, R)$, where R is a probability distribution on some set E and f is a function from $D \times E$ to any set F. Think of the distribution R as providing a "random seed" to the procedure A. If X is a probability distribution on D, then $A(X)$ denotes the probability distribution on F obtained by sampling $X \otimes R$ and applying f to the result. Note that applying a *function* is a special case of applying a randomized procedure.

FACT 2.4. *If X and Y are random variables and A is any randomized procedure, then $\|A(X) - A(Y)\| \leq \|X - Y\|$.*

PROOF. Let $A = (f, R)$. Then, for any set $S \subset F$,

$$\begin{aligned}
&|\Pr[A(X) \in S] - \Pr[A(Y) \in S]| \\
&\quad = \quad |\Pr[f(X \otimes R) \in S] - \Pr[f(Y \otimes R) \in S]| \\
&\quad = \quad \left|\Pr\left[X \otimes R \in f^{-1}(S)\right] - \Pr\left[Y \otimes R \in f^{-1}(S)\right]\right| \\
&\quad \leq \quad \|X \otimes R - Y \otimes R\| \\
&\quad \leq \quad \|X - Y\| + \|R - R\| \\
&\quad = \quad \|X - Y\|.
\end{aligned}$$

Taking the maximum over all sets S completes the proof. □

The next fact is useful when arguing that the statistical difference between two distributions is small.

FACT 2.5. *Suppose $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ are probability distributions on a set $D \times E$ such that*

1. *X_1 and Y_1 are identically distributed, and*
2. *With probability $> 1 - \epsilon$ over $x \leftarrow X_1$ (equivalently, $x \leftarrow Y_1$),*

$$\|X_2|_{X_1=x} - Y_2|_{Y_1=x}\| < \delta \tag{2.1}$$

(where $B|_{A=a}$ denotes the conditional distribution of B given that $A = a$ for jointly distributed random variables A and B).

Then $\|X - Y\| < \epsilon + \delta$.

PROOF. Let $T \subset D$ be the set of x's for which Equation 2.1 holds. Now, let S be an arbitrary subset of $D \times E$ and, for every $x \in D$, define $S_x = \{y \in E : (x, y) \in S\}$. Then,

$$\begin{aligned}
\Pr[X \in S] &\leq \Pr[X_1 \notin T] + \sum_{x \in T} \Pr[X_2 \in S_x | X_1 = x] \Pr[X_1 = x] \\
&< \epsilon + \sum_{x \in T} (\Pr[Y_2 \in S_x | Y_1 = x] + \delta) \Pr[Y_1 = x] \\
&\leq \epsilon + \delta + \Pr[Y \in S].
\end{aligned}$$

By symmetry, we also have $\Pr[Y \in S] < \epsilon + \delta + \Pr[X \in S]$. Since S was arbitrary, $\|X - Y\| < \epsilon + \delta$. □

The next fact formalizes the intuition that if two distributions have small statistical difference, then their mass functions must be close at most points.

FACT 2.6. *If X and Y are any two distributions such that $\|X-Y\|<\epsilon$, then with probability $>1-2\sqrt{\epsilon}$ over $x \leftarrow X$,*

$$(1-\sqrt{\epsilon})\Pr[X=x]<\Pr[Y=x]<(1+\sqrt{\epsilon})\Pr[X=x]. \tag{2.2}$$

PROOF. Let $S=\{x\colon (1-\sqrt{\epsilon})\Pr[X=x]\geq \Pr[Y=x]\}$, *i.e.* the set of x's for which the left-hand inequality in Equation 2.2 is violated. Then,

$$\begin{aligned}\Pr[Y\in S] &\leq (1-\sqrt{\epsilon})\Pr[X\in S]\\ &= \Pr[X\in S]-\sqrt{\epsilon}\Pr[X\in S].\end{aligned}$$

Thus, $\|X-Y\|\geq\sqrt{\epsilon}\Pr[X\in S]$,so we must have $\Pr[X\in S]<\sqrt{\epsilon}$. A similar argument show that the right-hand inequality in Equation 2.2 is violated with probability less than $\sqrt{\epsilon}$. □

3. Direct Product and XOR Lemmas

In this section, we describe two simple constructions, and their effect on the statistical difference between a pair of distributions. These are essential building blocks for the more complex transformations in later sections. The first construction is known as the *direct product* construction. In this construction, one samples a distribution independently k times. When applied to a pair of distributions, this construction has the effect of increasing any noticeable statistical difference to one exponentially close to 1. This is formalized in the following lemma:

LEMMA 3.1 (Direct Product Lemma). *Let X and Y be distributions such that $\|X-Y\|=\epsilon$. Then for all k,*

$$k\epsilon\geq\|\otimes^k X-\otimes^k Y\|\geq 1-2e^{-k\epsilon^2/2}$$

PROOF. The upper bound of $k\epsilon$ follows immediately from Fact 2.3, so we proceed to the proof of the lower bound. Recall, from the definition of statistical difference, that there must exist a set S such that

$$\Pr[X\in S]-\Pr[Y\in S]=\epsilon.$$

Let $p=\Pr[Y\in S]$. Then, $\Pr[X\in S]=p+\epsilon$. Hence, in k independent samples of X, the expected number of samples that lie in S is $(p+\epsilon)k$, whereas in k independent samples of Y, the expected number of samples that lie in S is pk. The Chernoff bound[3] tells us that the probability that *at least* $(p+\frac{\epsilon}{2})k$ components of $\otimes^k Y$ lie in S is at most $\exp(-k\epsilon^2/2)$, whereas the probability that *at most* $(p+\frac{\epsilon}{2})k$ components of $\otimes^k X$ lie in S is at most $\exp(-k\epsilon^2/2)$. Let S' be the set of all k-tuples that contain more than $(p+\frac{\epsilon}{2})k$ components that lie in S. Then we have,

$$\|\otimes^k X-\otimes^k Y\|\geq\Pr\left[\otimes^k X\in S'\right]-\Pr\left[\otimes^k Y\in S'\right]\geq 1-2e^{-k\epsilon^2/2}.$$

□

[3] For the formulation of the Chernoff bound we use, see, for example, the formulation of Hoeffding's inequality in [**Hof95**, Sec. 7.2.1].

At first glance, it may seem that the proof above is unnecessarily loose, and that one might be able to prove that the statistical difference increases even for small values of k. However, Madhu Sudan [**Sud97**] has pointed out that for any $p \in [0,1]$, there exist distributions X and Y such that $\|X \otimes X - Y \otimes Y\| = \|X - Y\| = p$. Consider the following two distributions:

$$\begin{aligned} X &= \begin{cases} 1 & \text{with probability } (1+p)/2 \\ 0 & \text{with probability } (1-p)/2 \end{cases} \\ Y &= \begin{cases} 0 & \text{with probability } (1+p)/2 \\ 1 & \text{with probability } (1-p)/2 \end{cases} \end{aligned}$$

Here, $\|X - Y\| = ((1+p) - (1-p))/2 = p$, and also

$$\|X \otimes X - Y \otimes Y\| = ((1+p)^2 - (1-p)^2)/4 = p.$$

Nevertheless, the direct product construction gives us an efficient and effective technique for increasing the statistical difference between two distributions. The two bounds in Lemma 3.1 show that large values go to 1 faster than small values. The following lemma provides a complementary technique which decreases the statistical difference to 0, with small values going to 0 faster than large values.

Lemma 3.2 (XOR Lemma). *There is a polynomial-time computable function that maps a triple $(C_0, C_1, 1^k)$, where C_0 and C_1 are circuits, to a pair of circuits (D_0, D_1) such that $\|D_0 - D_1\| = \|C_0 - C_1\|^k$. Specifically, D_0 and D_1 are defined as follows:*

D_0**:** *Uniformly select $(b_1, \ldots, b_k) \in \{0,1\}^k$ such that $b_1 \oplus \cdots \oplus b_k = 0$, and output a sample of $C_{b_1} \otimes \cdots \otimes C_{b_k}$.*
D_1**:** *Uniformly select $(b_1, \ldots, b_k) \in \{0,1\}^k$ such that $b_1 \oplus \cdots \oplus b_k = 1$, and output a sample of $C_{b_1} \otimes \cdots \otimes C_{b_k}$.*

In order to prove this lemma, we employ a generalization of the technique used in [**DDPY94**] to represent the logical AND of statements about Graph Nonisomorphism. This tool is described in the following Proposition.

Proposition 3.3. *Let X_0, X_1, Y_0, Y_1 be any random variables, and define the following pair of random variables:*

Z_0: *Choose $a, b \in_R \{0,1\}$ such that $a \oplus b = 0$. Output a sample of $X_a \otimes Y_b$.*
Z_1: *Choose $a, b \in_R \{0,1\}$ such that $a \oplus b = 1$. Output a sample of $X_a \otimes Y_b$.*

Then $\|Z_0 - Z_1\| = \|X_0 - X_1\| \cdot \|Y_0 - Y_1\|$.

The statistical difference between X_0 and X_1 (or Y_0 and Y_1) measures the advantage a computationally unbounded party has, over random guessing, of guessing b given a sample from X_b, where b is selected uniformly from $\{0,1\}$. Intuitively, the above Proposition says that the advantage one has in guessing the XOR of two independent bits is the product of the advantages one has for guessing each individual bit.

PROOF.

$$\begin{aligned}
\|Z_0 - Z_1\| &= \frac{1}{2}\left|Z_0 - Z_1\right|_1 \\
&= \frac{1}{2}\left|\left(\frac{1}{2}X_0 \otimes Y_0 + \frac{1}{2}X_1 \otimes Y_1\right) - \left(\frac{1}{2}X_1 \otimes Y_0 + \frac{1}{2}X_0 \otimes Y_1\right)\right|_1 \\
&= \frac{1}{4}\left|(X_0 - X_1) \otimes (Y_0 - Y_1)\right|_1 \\
&= \left(\frac{1}{2}\left|X_0 - X_1\right|_1\right) \cdot \left(\frac{1}{2}\left|Y_0 - Y_1\right|_1\right) \\
&= \|X_0 - X_1\| \cdot \|Y_0 - Y_1\|
\end{aligned}$$

□

Proposition 3.3 and an induction argument establish Lemma 3.2. Yao's XOR Lemma [**Yao82**] (see also [**GNW95**]) can be seen as an analogue of Lemma 3.2 in the computational setting, where the analysis is much more difficult.

4. Polarizing the Statistical Relationship

In this section, we combine the techniques from the previous section to yield the following lemma:

LEMMA 4.1 (Polarization Lemma).[4] *There is a polynomial-time computable function that takes a triple $(C_0, C_1, 1^k)$, where C_0 and C_1 are circuits, and outputs a pair of circuits (D_0, D_1) such that*

$$\begin{aligned}
\|C_0 - C_1\| < 1/3 &\quad\Rightarrow\quad \|D_0 - D_1\| < 2^{-k} \\
\|C_0 - C_1\| > 2/3 &\quad\Rightarrow\quad \|D_0 - D_1\| > 1 - 2^{-k}
\end{aligned}$$

The usefulness of the Polarization Lemma comes from the fact that the two distributions it produces can be treated almost as if they were identically distributed or disjoint (*i.e.* statistical difference 0 and 1, respectively). Indeed, it will be used in the constructions of the next two sections, and it was essential in proving that SD (with thresholds of 2/3 and 1/3, as we've defined it) is in SZK [**SV97**].

Recall that the Direct Product construction of Lemma 3.1 gives a way to increase statistical difference with large values going to 1 faster than small values. Similarly, the XOR Lemma (Lemma 3.2) shows how to decrease statistical difference with small values going to 0 faster than large values. Intuitively, alternating these procedures should "polarize" large and small values of statistical difference, pushing them closer to 1 and 0, respectively. A similar alternation between procedures with complementary effects was used by Ajtai and Ben-Or [**AB84**] to amplify the success probability of randomized constant-depth circuits.

PROOF. Let $\ell = \lceil \log_{4/3} 6k \rceil$. Apply the XOR Lemma (Lemma 3.2) to the triple $(C_0, C_1, 1^\ell)$ to produce (C_0', C_1') such that if

$$\begin{aligned}
\|C_0 - C_1\| < 1/3 &\quad\Rightarrow\quad \|C_0' - C_1'\| < (1/3)^\ell \\
\|C_0 - C_1\| > 2/3 &\quad\Rightarrow\quad \|C_0' - C_1'\| > (2/3)^\ell.
\end{aligned}$$

[4] The Polarization Lemma stated here is called the Amplification Lemma in [**SV97**]. We change the name here to stress that the Polarization Lemma does not merely increase statistical difference.

Let $m = 3^{\ell-1}$. Then apply the direct product construction, letting $C_0'' = \otimes^m C_0'$ and let $C_1'' = \otimes^m C_1'$. Then, by Lemma 3.1,

$$\begin{aligned} \|C_0 - C_1\| < 1/3 &\quad\Rightarrow\quad \|C_0'' - C_1''\| < 1/3 \\ \|C_0 - C_1\| > 2/3 &\quad\Rightarrow\quad \|C_0'' - C_1''\| > 1 - 2e^{-3^{\ell-1}(2/3)^{2\ell}/2} > 1 - 2e^{-k}. \end{aligned}$$

Finally, apply the XOR Lemma (Lemma 3.2) one more time to $(C_0'', C_1'', 1^k)$ to produce (D_0, D_1) such that

$$\begin{aligned} \|C_0 - C_1\| < 1/3 &\quad\Rightarrow\quad \|D_0 - D_1\| < 3^{-k} < 2^{-k} \\ \|C_0 - C_1\| > 2/3 &\quad\Rightarrow\quad \|D_0 - D_1\| > (1 - 2e^{-k})^k > 1 - 2ke^{-k} > 1 - 2^{-k}. \end{aligned}$$

□

Notice that the above analysis relies on the fact that $(2/3)^2 > (1/3)$, so it will not work if 1/3 and 2/3 are replaced by, say, .49 and .51. We do not know how to prove such a polarization lemma for arbitrary constant thresholds. We can however extend it to thresholds α and β, where β^2/α is greater than 1. More precisely, we have the following lemma:

Lemma 4.2 (General Polarization Lemma). *There is a function that takes as input a 5-tuple $(C_0, C_1, \alpha, \beta, 1^k)$, where $\beta^2 = \lambda \cdot \alpha$, with $\lambda > 1$, and C_0 and C_1 are circuits. The function is computable in time polynomial in the input size and $\alpha^{-1/\log(\lambda)}$, and outputs a pair of circuits (D_0, D_1) such that*

$$\begin{aligned} \|C_0 - C_1\| < \alpha &\quad\Rightarrow\quad \|D_0 - D_1\| < 2^{-k} \\ \|C_0 - C_1\| > \beta &\quad\Rightarrow\quad \|D_0 - D_1\| > 1 - 2^{-k} \end{aligned}$$

Proof. Let $\ell = \lceil \log_\lambda(6 \ln 6) \rceil$. Apply the XOR Lemma (Lemma 3.2) to the triple $(C_0, C_1, 1^\ell)$ to produce (C_0', C_1') such that if

$$\begin{aligned} \|C_0 - C_1\| < \alpha &\quad\Rightarrow\quad \|C_0' - C_1'\| < \alpha^\ell \\ \|C_0 - C_1\| > \beta &\quad\Rightarrow\quad \|C_0' - C_1'\| > \beta^\ell. \end{aligned}$$

Let $m = 1/3\alpha^\ell = \lambda^\ell/(3\beta^{2\ell})$. Then apply the direct product construction, letting $C_0'' = \otimes^m C_0'$ and let $C_1'' = \otimes^m C_1'$. Then, by Lemma 3.1,

$$\begin{aligned} \|C_0 - C_1\| < \alpha &\quad\Rightarrow\quad \|C_0'' - C_1''\| < 1/3 \\ \|C_0 - C_1\| > \beta &\quad\Rightarrow\quad \|C_0'' - C_1''\| > 1 - 2e^{-\frac{\lambda^\ell}{3\beta^{2\ell}} \cdot \frac{\beta^{2\ell}}{2}} \geq 2/3 \end{aligned}$$

An application of Lemma 4.1 finishes the proof. □

5. Reversing the Statistical Relationship

Proposition 5.1 (Reversal Mapping). *There is a polynomial-time computable function that maps pairs of circuits (C_0, C_1) to pairs of circuits (D_0, D_1) such that*

$$\begin{aligned} \|C_0 - C_1\| < 1/3 &\quad\Rightarrow\quad \|D_0 - D_1\| > 2/3 \\ \|C_0 - C_1\| > 2/3 &\quad\Rightarrow\quad \|D_0 - D_1\| < 1/3. \end{aligned}$$

That is, SD *reduces to* $\overline{\text{SD}}$.

By Theorem 1.1 (and the closure of SZK under reductions [**SV97**]), Proposition 5.1 is equivalent to the closure of SZK under complement. In fact, the existence of such a transformation was originally deduced from the fact that SZK is closed under complement [**Oka96, SV97**]. This result motivated our search for a more explicit description of such a mapping. By extracting ideas used in the transformations of statistical zero-knowledge proofs given in [**Oka96**] and [**SV97**], we obtained the description of this transformation given below.

The Construction. Let (C_0, C_1) be any pair of circuits and let $n = |(C_0, C_1)|$. By the Polarization Lemma (Lemma 4.1), we can produce in polynomial time a pair of circuits (C_0', C_1') such that

$$\begin{aligned} \|C_0 - C_1\| < 1/3 \quad &\Rightarrow \quad \|C_0' - C_1'\| > 1 - 2^{-n} \\ \|C_0 - C_1\| > 2/3 \quad &\Rightarrow \quad \|C_0' - C_1'\| < 2^{-n} \end{aligned}$$

Let $q = \mathrm{poly}(n)$ be the number of input gates of C_0' and C_1' (w.l.o.g. we may assume they have the same number) and let $\ell = \mathrm{poly}(n)$ be the number of output gates. For notational convenience, let $R = \{0,1\}^q$ and $L = \{0,1\}^\ell$. Let $m = n^3 q^2$ and define a new distribution $\overline{C}\colon \{0,1\}^m \times R^m \to L^m$ as follows:

$$\overline{C}(\overline{b}, \overline{r}) = (C_{b_1}'(r_1), \ldots, C_{b_m}'(r_m)).$$

We use the notation $\overline{z} \leftarrow \overline{C}$ to denote $\overline{z}$ chosen according to $\overline{C}$, *i.e.* select $\overline{b}$ and $\overline{r}$ uniformly and let $\overline{z} = \overline{C}(\overline{b}, \overline{r})$.

Let $\mathcal{H}$ be a 2-universal family of hash functions from $\{0,1\}^m \times R^m \times L^m$ to $T = \{0,1\}^{(q+1)m - 2\Delta - n}$, where $\Delta = \sqrt{nmq^2} = m/n$. We can now describe the new distributions:

D_0: Let $(\overline{b}, \overline{r}) \in_R \{0,1\}^m \times R^m$, $\overline{y} \leftarrow \overline{C}$, and $h \in_R \mathcal{H}$. Output $(\overline{C}(\overline{b}, \overline{r}), \overline{b}, h, h(\overline{b}, \overline{r}, \overline{y}))$.

D_1: Let $(\overline{b}, \overline{r}) \in_R \{0,1\}^m \times R^m$, $h \in_R \mathcal{H}$, and $t \in_R T$. Output $(\overline{C}(\overline{b}, \overline{r}), \overline{b}, h, t)$.

The important things to note about these distributions are that $\overline{b}$ is part of the output, and that D_0 and D_1 only differ in the last component, where D_0 has the value of the hash function and D_1 has a truly random element of T. Also note that the size of T is chosen to be $|\{0,1\}^m \times R^m| / 2^{2\Delta + n}$, which is essentially $|\{0,1\}^m \times R^m|$, scaled down by a factor of $2^{2\Delta+n}$, which can be thought of as a "fudge factor" needed to make the proof work. The introduction of the sample $\overline{y}$ in D_0 may at first seem superfluous; we explain below.

Intuition. For intuition, consider the case that $\overline{C}$ is a uniform distribution; that is, for every $\overline{z} \in \mathrm{range}(\overline{C})$, the size of the preimage set $|\{(\overline{b}, \overline{r})\colon \overline{C}(\overline{b}, \overline{r}) = \overline{z}\}|$ is the same value N. (It turns out that $\overline{C}$ is actually "close enough" to uniform for these arguments to work.) Then the range of $\overline{C}$ has size $2^{(q+1)m}/N$. So, in D_0, conditioned on a value for $\overline{C}(\overline{b}, \overline{r})$, the triple $(\overline{b}, \overline{r}, \overline{y})$ is selected uniformly from a set of size $2^{(q+1)m}$. Since this is much greater than $|T|$, the Leftover Hash Lemma of [**ILL89**] implies that conditioned on any value for the first component of D_0, the last two components $(h, h(\overline{b}, \overline{r}, \overline{y}))$ are distributed close to the uniform distribution on $\mathcal{H} \times T$, which is the distribution that D_1 has in its last two components.[5] Thus, if their second components were missing, D_0 and D_1 would be statistically close.

[5] Here we see the importance of $\overline{y}$: Without $\overline{y}$, conditioned on some value of $\overline{C}(\overline{b}, \overline{r})$, the pair $(\overline{b}, \overline{r})$ would be selected uniformly from a space of size N. If we were only hashing this pair, for the distribution $h(\overline{b}, \overline{r})$ to be uniform by the Leftover Hash Lemma, T would have had to be chosen so that $|T| \ll N$. The value of N, however, depends on the inner workings of the circuit C, and is in general unknown. By including $\overline{y}$, which comes uniformly from a space of size $2^{(q+1)m}/N$, we balance the arguments to h so that they come from a space of size $2^{(q+1)m}$, a known quantity.

Now, consider the case that $\|C_0' - C_1'\| \approx 1$. Then $\overline{b}$ is essentially "determined" by $\overline{C}(\overline{b},\overline{r})$. So the presence of $\overline{b}$ can be ignored, and the above argument says that D_0 and D_1 are statistically very close. Now, consider the case that $\|C_0' - C_1'\| \approx 0$. Then $\overline{b}$ is essentially "unrestricted" by $\overline{C}(\overline{b},\overline{r})$. Since there are 2^m choices for $\overline{b}$, conditioning on $\overline{b}$ in addition to $\overline{C}(\overline{b},\overline{r})$, cuts the number of triples $(\overline{b},\overline{r},\overline{y})$ down from $2^{m(q+1)}$ to roughly $2^{m(q+1)}/2^m$. Since $2^{m(q+1)}/2^m$ is much smaller than $|T|$, $h(\overline{b},\overline{r},\overline{y})$ will cover only a small fraction of $|T|$ and thus will be far from uniform (conditioned on values for $\overline{C}(\overline{b},\overline{r})$, $\overline{b}$, and h).

Proof of Proposition 5.1. First we will argue that $\overline{C}$ is close to uniform, so that we can apply arguments like those given above. This is the case because $\overline{C}$ is composed of many independent, identically distributed random variables. For $\overline{z} \in L^m$, we say the weight of $\overline{z}$ is the logarithm of the size of the preimage set of $\overline{z}$. Formally, let $\mathrm{wt}(\overline{z}) = \log_2 |\{(\overline{b},\overline{r})\colon \overline{C}(\overline{b},\overline{r}) = \overline{z}\}|$. Let w be the expected weight of an image, $w = \mathrm{E}_{\overline{z}\leftarrow\overline{C}}[\mathrm{wt}(\overline{z})]$. Then we can show the following:

Lemma 5.2. $\Pr_{\overline{z}\leftarrow\overline{C}}[|\mathrm{wt}(\overline{z}) - w| > \Delta] < 2^{-\Omega(n)}$.

Proof. For $z \in L$, let $\mathrm{wt}_0(z) = \log_2 |\{(b,r)\colon C_b(r) = z\}|$. Then, for $\overline{z} \in L^m$, $\mathrm{wt}(\overline{z}) = \mathrm{wt}_0(z_1) + \cdots + \mathrm{wt}_0(z_m)$. Observe that when $\overline{z}$ is selected according to $\overline{C}$, $z_1, \ldots, z_m$ are independent and identically distributed. Moreover, for any $z \in L$, $0 \le \mathrm{wt}_0(z) \le q$. So, by the Hoeffding inequality [**Hof95**, Sec. 7.2.1], we have

$$\Pr_{\overline{z}\leftarrow\overline{C}}[|\mathrm{wt}(\overline{z}) - w| > \Delta] < 2e^{-2\Delta^2/mq^2} = 2e^{-2n}.$$

□

It will be convenient to eliminate those $\overline{z} \in L^m$ that have weight far above or below the mean. Let $G = \{(\overline{b},\overline{r})\colon |\mathrm{wt}(\overline{C}(\overline{b},\overline{r})) - w| \le \Delta\}$ be the set of *good* pairs $(\overline{b},\overline{r})$. The above Lemma says that $|G| \ge (1 - 2^{-\Omega(n)})|\{0,1\}^m \times R^m|$. Thus $\|G - \{0,1\}^m \times R^m\| \le 2^{-\Omega(n)}$, where for simplicity of notation, we let the name of a set also refer to the uniform distribution on the same set. Define $\overline{C}'$ to be the distribution obtained by selecting $(\overline{b},\overline{r}) \leftarrow G$ and outputting $\overline{C}(\overline{b},\overline{r})$. Then, since $\overline{C}$ is a function, Fact 2.4 tells us that $\|\overline{C} - \overline{C}'\| = 2^{-\Omega(n)}$. Similarly, we define variants of D_0 and D_1 that sample from G instead of $\{0,1\}^m \times R^m$:

D_0': Let $(\overline{b},\overline{r})\in_R G$, $\overline{y} \leftarrow \overline{C}'$, and $h \in_R \mathcal{H}$. Output $(\overline{C}'(\overline{b},\overline{r}), \overline{b}, h, h(\overline{b},\overline{r},\overline{y}))$.

D_1': Let $(\overline{b},\overline{r})\in_R G$, $h \in_R \mathcal{H}$, and $t \in_R T$. Output $(\overline{C}'(\overline{b},\overline{r}), \overline{b}, h, t)$.

Since D_0' (or D_1') is a randomized procedure applied to two (or one) independent samplings from G, Fact 2.4 tells us that $\|D_0 - D_0'\| = 2^{-\Omega(n)}$ (and $\|D_1 - D_1'\| = 2^{-\Omega(n)}$). Hence, it suffices to prove that these modified distributions have the properties we want in each case. For the case when C_0 and C_1 are statistically far, we prove the following claim:

Claim 5.3. *If* $\|C_0' - C_1'\| > 1 - 2^{-n}$, *then* $\|D_0' - D_1'\| < 2^{-\Omega(n)}$.

This use of "dummy" samples to form a space whose size is known is the "complementary usage of messages" technique of Okamoto [**Oka96**].

Proof. First we formalize the idea that $\overline{b}$ is "determined" by $\overline{C}$. Define $f\colon L \to \{0,1\}$ by

$$f(z) = \begin{cases} 0 & \text{if } \Pr\left[C_0' = z\right] > \Pr\left[C_1' = z\right] \\ 1 & \text{otherwise} \end{cases}$$

Then

$$\Pr_{b,r}[f(C_b'(r)) = b] = \frac{1}{2}\Pr_r[f(C_0'(r)) = 0] + \frac{1}{2}\Pr_r[f(C_1'(r)) = 1].$$

Now, by the definition of statistical difference, $\Pr_r[C_0'(r) \in f^{-1}(0)] \geq 1 - 2^{-n}$ and $\Pr_r[C_1'(r) \in f^{-1}(1)] \geq 1 - 2^{-n}$. Thus, $\Pr_{b,r}[f(C_b'(r)) = b] > 1 - 2^{-n}$. Now define $\overline{f}\colon L^m \to \{0,1\}^m$ by $\overline{f}(\overline{z}) = (f(z_1), \ldots, f(z_m))$. Then

$$\Pr_{\overline{b},\overline{r}}[\overline{f}(\overline{C}(\overline{b},\overline{r})) = \overline{b}] > (1 - 2^{-n})^m = 1 - 2^{-\Omega(n)}.$$

Since G is a $1 - 2^{-\Omega(n)}$ fraction of $\{0,1\}^m \times R^m$, the same is true when $(\overline{b},\overline{r})$ is selected uniformly from G. Thus, if we define:

D_0'': Let $(\overline{b},\overline{r}) \in_R G$, $\overline{y} \leftarrow \overline{C}'$, and $h \in_R \mathcal{H}$. Output $(\overline{C}'(\overline{b},\overline{r}), \overline{f}(\overline{C}'(\overline{b},\overline{r})), h, h(\overline{b},\overline{r},\overline{y}))$.

D_1'': Let $(\overline{b},\overline{r}) \in_R G$, $h \in_R \mathcal{H}$, and $t \in_R T$. Output $(\overline{C}'(\overline{b},\overline{r}), \overline{f}(\overline{C}'(\overline{b},\overline{r})), h, t)$.

Then, by Fact 2.5, $\|D_0' - D_0''\| = 2^{-\Omega(n)}$ and $\|D_1' - D_1''\| = 2^{-\Omega(n)}$. So it suffices to show that $\|D_0'' - D_1''\| = 2^{-\Omega(n)}$. Since the first components of D_0'' and D_1'' are identically distributed and the second components are determined by the first ones, it suffices to show (by Fact 2.5) that, conditioned on any value for the first coordinate, the third and fourth components have statistical difference $2^{-\Omega(n)}$. This will follow from the Leftover Hash Lemma [**ILL89**]:

Lemma 5.4 (Leftover Hash Lemma [**ILL89**]). *Let $\mathcal{H}$ be a family of 2-universal hash functions from D to T. Let X by a probability distribution on D such that for all $x \in D$, $\Pr\left[X = x\right] \leq \epsilon/|T|$. Then the following two distributions have statistical difference at most $\epsilon^{1/3}$.*

1. *Choose $x \leftarrow X$, $h \in_R \mathcal{H}$. Output $(h, h(x))$.*
2. *Choose $h \in_R \mathcal{H}$, $t \in_R T$. Output (h, t).*

By the above argument and the Leftover Hash Lemma, it suffices to show that conditioned on any value $\overline{z}$ for $\overline{C}'(\overline{b},\overline{r})$, no triple $(\overline{b},\overline{r},\overline{y})$ has probability more than $2^{-\Omega(n)}/|T|$. The pair $(\overline{b},\overline{r})$ comes uniformly from a set of size $2^{\mathrm{wt}(\overline{z})} \geq 2^{w-\Delta}$, and $\overline{y}$ is selected independently according to $\overline{C}'$, so the probability of any triple $(\overline{b},\overline{r},\overline{y})$ is at most

$$\left(\frac{1}{2^{w-\Delta}}\right)\left(\frac{2^{w+\Delta}}{|G|}\right) \leq \frac{2^{2\Delta}}{(1 - 2^{-\Omega(n)})2^{(q+1)m}} = \frac{2^{-\Omega(n)}}{|T|}.$$

Thus, $\|D_0'' - D_1''\| \leq 2^{-\Omega(n)}$, and the claim is established. □

Now we treat the other case, when C_0 and C_1 are stastically close.

Claim 5.5. *If $\|C_0' - C_1'\| < 2^{-n}$, then $\|D_0' - D_1'\| > 1 - 2^{-\Omega(n)}$.*

Proof. First, we formalize the idea that $\overline{b}$ is almost completely "undetermined" by $\overline{C}(\overline{b},\overline{r})$. Since $\|C_0' - C_1'\| < 2^{-n}$, it follows from Fact 2.6 that with probability $1 - 2^{-\Omega(n)}$ over $z \leftarrow C_0'$,

$$(1 - 2^{-\Omega(n)}) \Pr[C_1' = z] \leq \Pr[C_0' = z] \leq (1 + 2^{-\Omega(n)}) \Pr[C_1' = z].$$

In other words,

$$1 - 2^{-\Omega(n)} \leq \frac{|\{r : C_0'(r) = z\}|}{|\{r : C_1'(r) = z\}|} \leq 1 + 2^{-\Omega(n)}.$$

The same is true with probability $1 - 2^{-\Omega(n)}$ when the roles of C_0' and C_1' are reversed. Thus, with probability $1 - m2^{-\Omega(n)} = 1 - 2^{-\Omega(n)}$ over $\overline{z} \leftarrow \overline{C}$, we have for *every* pair $\overline{b}, \overline{c} \in \{0,1\}^m$,

$$1 - 2^{-\Omega(n)} = (1 - 2^{\Omega(n)})^m \leq \frac{\left|\{\overline{r} : \overline{C}(\overline{b},\overline{r}) = \overline{z}\}\right|}{\left|\{\overline{r} : \overline{C}(\overline{c},\overline{r}) = \overline{z}\}\right|} \leq (1 + 2^{-\Omega(n)})^m = 1 + 2^{-\Omega(n)}.$$

Since there are 2^m choices for $\overline{c}$, this, combined with Lemma 5.2, implies that, with probability $1 - 2^{-\Omega(n)}$ over $\overline{z} \leftarrow \overline{C}$, the following holds for *every* $\overline{b} \in \{0,1\}^m$:

$$\left|\{\overline{r} : \overline{C}(\overline{b},\overline{r}) = \overline{z}\}\right| \leq (1 + 2^{-\Omega(n)}) \cdot \frac{2^{\mathrm{wt}(\overline{z})}}{2^m} \leq (1 + 2^{-\Omega(n)}) \cdot 2^{w+\Delta-m}.$$

Since this is true with probability $1 - 2^{-\Omega(n)}$ for $\overline{z}$ selected according to $\overline{C}$, it is also true with probability $1 - 2^{-\Omega(n)}$ for $\overline{z}$ selected according to $\overline{C}'$. Fix any such $\overline{z}$ and fix any $\overline{b} \in \{0,1\}^m$ and $h \in \mathcal{H}$. Then, in D_0', conditioned on $\overline{C}'(\overline{b},\overline{r}) = \overline{z}$, $\overline{b}$, and h, there are at most

$$\begin{aligned}(1 + 2^{-\Omega(n)}) \cdot 2^{w+\Delta-m}\left(\frac{|G|}{2^{w-\Delta}}\right) &\leq (1 + 2^{-\Omega(n)}) \cdot 2^{2\Delta-m}(2^{m+mq}) \\ &= (1 + 2^{-\Omega(n)}) \cdot 2^{4\Delta+n-m}|T| \\ &= 2^{-\Omega(m)}|T|\end{aligned}$$

possible values for $(\overline{r},\overline{y})$. Thus, with probability $1 - 2^{-\Omega(n)}$, conditioned on values for the first three components of D_0', the fourth component $h(\overline{b},\overline{r},\overline{y})$ can cover at most a $2^{-\Omega(m)} \leq 2^{-\Omega(n)}$ fraction of T. In contrast, conditioned on values for the first three components of D_1', the fourth component is uniformly distributed on T. Therefore, $\|D_0' - D_1'\| \geq 1 - 2^{-\Omega(n)}$. □

6. An Application – Boolean Closure

In this section, we use the transformations developed in the prior sections to prove that any boolean formulae whose atoms are statements about membership in SD can be efficiently transformed into a single statement about SD. By Theorem 1.1 [**SV97**], this implies a very strong boolean closure property of SZK: given an arbitrary boolean formula whose atoms are statements about membership in *any* language in SZK, one can efficiently construct a statistical zero-knowledge interactive proof for its validity. Note that such a property does not follow immediately from the fact that a class is closed under intersection, union, and complementation, because applying these more than a constant number of times could incur a superpolynomial cost in efficiency, while we ask that the construction can be done efficiently with respect to the size of the formula. The procedure for doing this is based on work by De Santis, Di Crescenzo, Persiano, and Yung [**DDPY94**].

They show how to construct statistical zero-knowledge proofs for all monotone boolean formulae whose atoms are statements about a random self-reducible language. Their zero-knowledge proofs are constructed by producing two distributions which are either disjoint or identical, depending on whether or not the formula is true. Hence, their construction can be viewed as a reduction to an extreme case of SD, in which the thresholds are 1 and 0.

Using the direct product construction, the XOR Lemma, and the Polarization Lemma, we generalize their result to monotone formulae whose atoms are statements about membership in SD. Then, using our Reversal Mapping, we further generalize to non-monotone formulae.

Definition 6.1. Let Π be any promise problem. Then we define a new promise problem $\Phi(\Pi)$ whose instances are $(\phi, x_1, \ldots, x_k)$ where $k \geq 0$, $x_1, \ldots, x_k \in \Pi_Y \cup \Pi_N$ (*i.e.* $x_1, \ldots, x_k$ satisfy the promise for Π), and $\phi(v_1, \ldots, v_k)$ is a k-ary propositional formula. The YES instances of $\Phi(\Pi)$ are those instances for which $\phi((x_1 \in \Pi_Y), \ldots, (x_k \in \Pi_Y))$ is true and the NO instances are those for which it is false.

Mon(Π) is defined analogously, except that only monotone ϕ are considered.[6]

The main result of this section follows:

Theorem 6.2. $\Phi(\mathrm{SD})$ *reduces to* SD. *That is, there is a polynomial-time computable function that maps an instance* $x = (\phi, (C_0^1, C_1^1), \ldots, (C_0^k, C_1^k))$ *of* $\Phi(\mathrm{SD})$ *to an instance* $y = (D_0, D_1)$ *of* SD *such that*

$$\begin{aligned} x \in \Phi(\mathrm{SD})_Y &\quad\Rightarrow\quad y \in \mathrm{SD}_Y \\ x \in \Phi(\mathrm{SD})_N &\quad\Rightarrow\quad y \in \mathrm{SD}_N. \end{aligned}$$

The main step in proving Theorem 6.2 is the following Lemma, which mimics the construction of [**DDPY94**] for Mon(Graph Nonisomorphism):

Lemma 6.3. Mon(SD) *reduces to* SD.

Proof. For intuition, consider two instances of statistical difference (C_0, C_1) and (D_0, D_1), both of which have statistical difference very close to 1 or very close to 0 (which can be achieved by the Polarization Lemma). Then $(C_0 \otimes D_0, C_1 \otimes D_1)$ will have statistical difference very close to 1 if either of the original statistical differences is very close to 1 and will have statistical difference very close to 0 otherwise. Thus, this operation represents OR. Similarly, the XOR operation in Proposition 3.3 represents AND. To obtain Lemma 6.3, we will recursively apply these constructions, taking care to keep the running time polynomial.

Let $w = (\phi, (C_0^1, C_1^1), \ldots, (C_0^k, C_1^k))$ be an instance of Mon(SD) and let $n = |w|$. By applying the Polarization Lemma (Lemma 4.1), we can constuct in polynomial time pairs of circuits $(D_0^1, D_1^1), \ldots, (D_0^k, D_1^k)$ such that the statistical difference between D_0^i and D_1^i is greater than $1 - 2^{-n}$ if $(C_0^i, C_1^i) \in \mathrm{SD}_Y$ and is less than 2^{-n} if $(C_0^i, C_1^i) \in \mathrm{SD}_N$.

Consider the randomized recursive procedure Sample(ψ, b) in Figure 6 which takes a subformula $\psi(v_{i_1}, \ldots, v_{i_k})$ of ϕ and a bit $b \in \{0, 1\}$ as input.

Executing Sample(ϕ, b) for $b \in \{0, 1\}$ takes time polynomial in n, because the number of recursive calls is equal to the number of subformulas of ϕ. For a

[6] In [**DDPY94**], only monotone formulae are treated. What they call $\Phi(L)$ is what we call Mon(L).

```
Sample($\psi, b$)
    If $\psi = v_i$, sample $z \leftarrow D_b^i$.
    If $\psi = \tau \vee \mu$,
        Sample $z_1 \leftarrow$ Sample($\tau, b$);
        Sample $z_2 \leftarrow$ Sample($\mu, b$);
        Let $z = (z_1, z_2)$.
    If $\psi = \tau \wedge \mu$,
        Choose $c, d \in_R \{0,1\}$ subject to $c \oplus d = b$;
        Sample $z_1 \leftarrow$ Sample($\tau, c$);
        Sample $z_2 \leftarrow$ Sample($\tau, d$);
        Let $z = (z_1, z_2)$.
    Output $z$.
```

FIGURE 1

subformula τ of ϕ, let $\mathrm{Dif}(\tau) = \|\mathrm{Sample}(\tau, 0) - \mathrm{Sample}(\tau, 1)\|$. Then we can prove the following about Dif:

CLAIM 6.4. *For every subformula $\tau = \tau(v_{i_1}, \ldots, v_{i_j})$ of ϕ,* $\mathrm{Dif}(\tau) > 1 - m2^{-n}$ *if*

$$\tau((C_0^{i_1}, C_1^{i_1}) \in \mathrm{SD}_Y, \ldots, (C_0^{i_j}, C_1^{i_j}) \in \mathrm{SD}_Y)$$

is true and $\mathrm{Dif}(\tau) < m2^{-n}$ *if it is false, where $m = |\tau|$.*

PROOF OF CLAIM. By induction on subformulae of ψ. It holds for atomic subformulae (*i.e.* the variables v_i) by the properties of the D_b^i's.

Consider the case when $\psi = \tau \vee \mu$. If ψ is true (with the appropriate arguments), either τ or μ must be true. Without loss of generality, say τ is true. Then, by Fact 2.4 and induction,

$$\mathrm{Dif}(\psi) \geq \mathrm{Dif}(\tau) > 1 - |\tau|2^{-n} > 1 - |\psi|2^{-n}.$$

If ψ is false, then both τ and μ are false. By Fact 2.3 and induction,

$$\mathrm{Dif}(\psi) \leq \mathrm{Dif}(\tau) + \mathrm{Dif}(\mu) < |\tau|2^{-n} + |\mu|2^{-n} \leq |\psi|2^{-n}.$$

Now consider the case when $\psi = \tau \wedge \mu$. By Proposition 3.3, $\mathrm{Dif}(\psi) = \mathrm{Dif}(\tau) \cdot \mathrm{Dif}(\mu)$. If ψ is true, then, by induction,

$$\mathrm{Dif}(\psi) \geq (1 - |\tau|2^{-n})(1 - |\mu|2^{-n}) > 1 - (|\tau| + |\mu|)2^{-n} \geq 1 - |\psi|2^{-n}.$$

If ψ is false, then, without loss of generality, say τ is false. By induction,

$$\mathrm{Dif}(\psi) \leq \mathrm{Dif}(\tau) < |\tau|2^{-n} \leq |\psi|2^{-n}.$$

□

Now, let A and B be circuits describing the computations of $\mathrm{Sample}(\phi, 0)$ and $\mathrm{Sample}(\phi, 1)$, respectively, (which take the random bits each procedure uses as input). By the above claim, $\|A - B\| > 1 - n2^{-n} > 2/3$ if ϕ is true with the appropriate arguments, and $\|A - B\| < n2^{-n} < 1/3$ if ϕ is false. In other words, the construction of A and B from w describes a many-one reduction from Mon(SD) to SD. This reduction can be computed in polynomial time because Sample runs in polynomial time. □

Now it is straightforward to deduce Theorem 6.2

PROOF. Let $(\phi, x_1, \ldots, x_k)$ be any instance of $\Phi(\mathrm{SD})$, where $\phi = \phi(v_1, \ldots, v_k)$. Use DeMorgan's laws to propagate all negations of ϕ to its variables. Now replace all occurrences of the literal $\neg v_i$ with a new variable w_i. Let $\psi(v_1, \ldots, v_k, w_1, \ldots, w_k)$ be the resulting (monotone) formula. Then, letting f be the Reversal Mapping of Proposition 5.1 which reduces SD to $\overline{\mathrm{SD}}$,

$$(\phi, x_1, \ldots, x_k) \mapsto (\psi, x_1, \ldots, x_k, f(x_1), \ldots, f(x_k))$$

is a reduction from $\Phi(\mathrm{SD})$ to $\mathrm{Mon}(\mathrm{SD})$. Composing this with the reduction in Lemma 6.3, we obtain Theorem 6.2. □

By Theorems 1.1 and 6.2 (along with the fact that SZK is closed under many-one reductions [**SV97**]), we obtain the following:

COROLLARY 6.5. *For every language L, $L \in \mathsf{SZK} \Rightarrow \Phi(L) \in \mathsf{SZK}$.*

Corollary 6.5 is a strengthening of several previous results. In [**DDPY94**], it was shown that $\mathrm{Mon}(L) \in \mathsf{SZK}$ for any language L which is random self-reducible, whose complement is self-reducible, or whose complement has a noninteractive statistical zero-knowledge proof. They also gave statistical zero-knowledge proofs for some simple statements involving a random-self-reducible language and its complement. Damgård and Cramer [**DC96**] extended these results by showing that $\mathrm{Mon}(L) \in \mathsf{SZK}$ as long as L or its complement has a 3-round public coin statistical zero-knowledge proof, and also treat a larger class of monotone functions.

Corollary 6.5 can be generalized to work for all boolean formulae whose atoms are statements about membership in any finite set of languages in SZK, but we omit the notationally cumbersome formal statement. Corollary 6.5 can be viewed as demonstrating that SZK is closed under a weak form of (polynomial-time) Turing reducibility. In particular, if $\Phi(L)$ were defined in terms of circuits rather than formulae, then the analogue of Corollary 6.5, together with the closure of SZK under many-one reductions, would imply that SZK is closed under *nonadaptive* Turing reductions. We do not know whether this is true, but we can prove that SZK is closed under a weaker form of Turing reductions:

PROPOSITION 6.6. *SZK is closed under (adaptive) polynomial-time Turing reductions which make a $O(\log n)$ of oracle queries on inputs of length n. That is, if A is a language*[7] *in SZK and B reduces to A via such a reduction, then $B \in \mathsf{SZK}$.*

PROOF. Let M be the polynomial-time oracle machine such that M decides B when given oracle access to A, and on every input x, M^A asks at most $m = c \log |x|$ oracle queries. We define a *transcript* to be a sequence $t = (y_1, \sigma_1, \ldots, y_m, \sigma_m)$ such that $M(x)$ would ask oracle queries $y_1, \ldots, y_m$ (in that order) when given oracle responses $\sigma_1, \ldots, \sigma_m \in \{0, 1\}$. (Note that we are not requiring $\sigma_1, \ldots, \sigma_m$ to be the *correct* responses which would indicate whether or not $x_i \in A$.) We call such a transcript *accepting* if it would make $M(x)$ accept. Notice that, given x, we can enumerate in polynomial time all accepting transcripts for x by simulating M on all 2^m possible sequences of oracle responses. Let $t_i = (y_1^i, \sigma_1^i, \ldots, y_m^i, \sigma_m^i)$ be the ith accepting transcript in this enumeration, and let s be the number of accepting

[7] This proposition extends to the case of promise problems if we assume the oracle queries in the reduction never violate the promise.

transcripts. Now consider the following formula ϕ on a total of $s \cdot m$ variables:

$$\phi(v_1^1, \ldots, v_m^1, \ldots, v_1^s, \ldots, v_m^s) = \bigvee_{i=1}^{s} \left(\ell_1^i \wedge \cdots \wedge \ell_m^i\right),$$

where the literal ℓ_j^i is v_j^i if $\sigma_j^i = 1$ and is $\neg v_j^i$ if $\sigma_j^i = 0$. Now we claim that

$$M^A(x) \text{ accepts} \Leftrightarrow \phi((y_1^1 \in A), \ldots, (y_m^1 \in A), \ldots, (y_1^s \in A), \ldots, (y_m^s \in A)) = 1.$$

If $M^A(x)$ accepts then the term corresponding to the transcript with the correct oracle responses will evaluate to true. However, if $M^A(x)$ does not accept, then any accepting transcript must have at least one σ_j which disagrees the response oracle A would give on input y_j, so all the terms evaluate to false. Thus,

$$x \mapsto (\phi, y_1^1, \ldots, y_m^1, \ldots, y_1^s, \ldots, y_m^s)$$

is a (many-one) reduction from B to $\Phi(A)$. By Corollary 6.5 and the fact that SZK is closed under many-one reductions [**SV97**], we conclude that $B \in$ SZK. □

It would be interesting to prove that SZK is closed under Turing reductions, adaptive or nonadaptive, that make polynomially many oracle calls, or give evidence that this is not the case.

Acknowledgments

We are grateful to Oded Goldreich for his guidance throughout this work, and for countless insightful comments and conversations. We also thank our advisor, Shafi Goldwasser, for encouragement throughout; Madhu Sudan, for helping us better understand the effect of parallel repetition on statistical difference; Luca Trevisan for a conversation which helped us see that Proposition 6.6 follows from Corollary 6.5; and Avi Wigderson for encouraging us to seek a more explicit description of the Reversal Mapping.

References

[AB84] Miklos Ajtai and Michael Ben-Or, *A theorem on probabilistic constant depth computations*, Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing (Washington, D.C.), 1984, pp. 471–474.

[DC96] Ivan Damgård and Ronald Cramer, *On monotone function closure of perfect and statistical zero-knowledge*, Theory of Cryptography Library: Record 96-03, 1996, `http://theory.lcs.mit.edu/~tcryptol`.

[DDPY94] Alfredo De Santis, Giovanni Di Crescenzo, Giuseppe Persiano, and Moti Yung, *On monotone formula closure of SZK*, Proceedings of the Thirty Fifth Annual Symposium on Foundations of Computer Science, 1994, pp. 454–465.

[ESY84] Shimon Even, Alan L. Selman, and Yacov Yacobi, *The complexity of promise problems with applications to public-key cryptography*, Information and Control **61** (1984), no. 2, 159–173.

[GG98] Oded Goldreich and Shafi Goldwasser, *On the limits of non-approximability of lattice problems*, Proceedings of the 30th Annual ACM Symposium on Theory of Computing (Dallas, TX), ACM, May 1998, To appear.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff, *The knowledge complexity of interactive proof systems*, SIAM Journal on Computing **18** (1989), no. 1, 186–208.

[GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson, *Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems*, Journal of the Association for Computing Machinery **38** (1991), no. 1, 691–729.

[GNW95] Oded Goldreich, Noam Nisan, and Avi Wigderson, *On Yao's XOR-lemma*, ECCC Report TR95-050, March 1995, Available from `http://www.eccc.uni-trier.de/eccc/`.

[Hof95] Micha Hofri, *Analysis of algorithms: Computational methods and mathematical tools*, Oxford University Press, 1995.

[ILL89] Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *Pseudo-random generation from one-way functions (extended abstract)*, Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing (Seattle, Washington), 15–17 May 1989, pp. 12–24.

[Oka96] Tatsuaki Okamoto, *On relationships between statistical zero-knowledge proofs*, Proceedings of the Twenty Eighth Annual ACM Symposium on the Theory of Computing, 1996, See also preprint of full version, August 1997.

[Sud97] Madhu Sudan, September 1997, Personal Communication.

[SV97] Amit Sahai and Salil Vadhan, *A complete promise problem for statistical zero-knowledge*, Proceedings of the 38th Annual Symposium on the Foundations of Computer Science, IEEE, October 1997, pp. 448–457.

[Yao82] Andrew C. Yao, *Theory and application of trapdoor functions*, Proceedings of the Twenty Third Annual Symposium on Foundations of Computer Science, 1982, pp. 80–91.

MIT Laboratory for Computer Science, 545 Technology Sq., Cambridge, MA 02139
E-mail address: amits@theory.lcs.mit.edu, salil@math.mit.edu