

Public-coin Differing-input Obfuscation from KOALA and iO

Liyan Chen

Vinod Vaikuntanathan

Abstract

We show that the recent construction of witness encryption (Hair and Sahai, arXiv:2609.18275v1) is a public-coin extractable witness encryption scheme under the Knowledge of Orthogonality Assumption, or KOALA for short (Buellens-Wee, PKC 2019). As an immediate application, we obtain public-coin differing-input obfuscation (diO) from KOALA and indistinguishability obfuscation (iO).

Contents

1	Introduction	3
1.1	Our Results	3
2	Preliminaries	4
2.1	The Bounded-Factor MinRank Problem and Its NP-Completeness	4
2.2	The KOALA Assumption	5
2.3	Extractable Witness Encryption	5
2.4	Indistinguishability Obfuscation and Differing-Input Obfuscation	6
3	Extractable Witness Encryption from KOALA	7
3.1	Construction	8
4	Boosting Extractable WE to diO	10
4.1	Construction and Analysis	10

1 Introduction

The notion of program obfuscation [BGI⁺01, GGH⁺13] has long become a central primitive in cryptography. So far, the cryptography community has largely focused on indistinguishability obfuscation (iO) which guarantees that the obfuscations of any pair of functionally equivalent programs are computationally indistinguishable. This is due to its wide application, starting from [SW14], and because it can be built from well-founded assumptions [JLS21, JLS22, RVV24].

On the other hand, there is a stronger notion of program obfuscation called differing-input obfuscation (diO), or extractable obfuscation [ABG⁺13, BCP14]. It says that if there exists an adversary that distinguishes obfuscation of circuits C_0, C_1 , then there is an algorithm that extracts an input x on which the two circuits differ. In particular, consider any (efficiently samplable) distribution

$$(C_0, C_1, \text{aux}) \leftarrow D_\lambda,$$

diO says that if there exists an adversary that distinguishes $(\text{diO}(C_0), \text{aux})$ from $(\text{diO}(C_1), \text{aux})$, then there exists an adversary that on input (C_0, C_1, aux) outputs x such that $C_0(x) \neq C_1(x)$.

Unfortunately, it is known [GGHW14, BSW16] that diO does not exist for general distributions D_λ . Roughly speaking, this is because the auxiliary information aux could be a structured trapdoor information (usually an obfuscated program) that helps to distinguish $\text{diO}(C_0), \text{diO}(C_1)$, but not reveal a differing input x for which $C_0(x) \neq C_1(x)$. Later work [IPS15] observes that to avoid these impossibility results, we may restrict the samplers to be *public-coin*: aux is uniformly random, the sampler takes aux and deterministically generates C_0, C_1 , and the adversary knows aux (and hence C_0 and C_1). This is the notion of *public-coin* differing-input obfuscation: diO whose security only holds for *public-coin* samplers.

So far very little is known about the existence of public-coin diO. The only work (to our knowledge) addressing it is [BP15]: they show any of the following primitives rules out public-coin diO:

- Extractable one-way functions w.r.t. some structured auxiliary input distribution $\mathcal{Z}$; or
- SNARKs for NP w.r.t. some structured auxiliary input distribution $\mathcal{Z}'$.¹

Similar to diO, both of these primitives have “extraction-type” security guarantee, and [BP15] needs their security w.r.t. structured auxiliary input distribution. In fact, we believe it is more likely that these two primitives does not exist.

1.1 Our Results

In this work we give positive evidence that public coin differing-input obfuscation exists, by constructing it from two assumptions: the Knowledge of OrthogonalLity Assumption (KOALA) [BW19, AWY20] and indistinguishability obfuscation (iO). KOALA is a knowledge-type assumption that says in prime-order group, if an adversary can distinguish $g^{\mathbf{u}}$ and $g^{\mathbf{v}}$, where $\mathbf{u}$ is a random vector and $\mathbf{v}$ is sampled from a hidden low-dimensional subspace, then the adversary must know a vector orthogonal to the subspace. KOALA is believed to be true since it holds in the generic group model, but it is a concrete, albeit not falsifiable, assumption.

Theorem 1.1 (Informal; see Theorem 4.2). *Assume KOALA (Definition 2.3) and iO, there exists public-coin differing-inputs obfuscation.*

¹We refer curious readers to [BP15] for the exact definitions.

Our proof starts from the observation that the recent Hair-Sahai witness encryption (WE) scheme [HS26], originally proven secure in the generic group model, is an extractable witness encryption scheme for NP statements sampled from public coins, under the (public-coin) KOALA assumption.

Theorem 1.2 (Informal; see Theorem 3.2). *Assuming KOALA, there exists public coin extractable witness encryption for NP.*

Additionally, inspired by the (succinct) WE to (succinct) iO transform in [JJMP25], we show a generic transform from extractable witness encryption to differing-input obfuscation, assuming the existence of iO.

Theorem 1.3 (Informal; see Theorem 4.1). *Assuming public coin extractable witness encryption for NP and indistinguishability obfuscation, there exists differing-input obfuscation for public coin samplers.*

Our main result follows as a direct consequence of the above two theorems.

2 Preliminaries

2.1 The Bounded-Factor MinRank Problem and Its NP-Completeness

Let p be a prime. Given a basis of a linear space of matrices $\mathcal{S} \subseteq \mathbb{F}_p^{m \times m}$, homogeneous MinRank asks for the minimum rank of a nonzero matrix of $\mathcal{S}$. We use the promise version of [HS26], which says either there is no small-rank matrix in $\mathcal{S}$, or there is a rank 1 matrix $\mathbf{u}\mathbf{v}^T$ in $\mathcal{S}$ and furthermore $\mathbf{v} \in \{0, 1\}^m$.

Problem 2.1 (Bounded-Factor MinRank). Let m and r be positive integers, let p be a prime. An instance is a basis $\mathbf{M}_1, \dots, \mathbf{M}_k$ of a linear space $\mathcal{S} \subseteq \mathbb{F}_p^{m \times m}$, promised to satisfy one of

YES: $\mathcal{S}$ contains a nonzero matrix $\mathbf{u}\mathbf{v}^T$ in which $\mathbf{u} \in \mathbb{F}_p^m$ and of $\mathbf{v} \in \{0, 1\}^m$.

NO: $\text{rank}(\mathbf{M}) > r$ for every nonzero $\mathbf{M} \in \mathcal{S}$;

and the task is to decide which of the two holds.

Recently, Hair and Sahai [HS26] proved that the promise Bounded-Factor MinRank is NP-hard, for $r = \Theta(\log m)$.

Theorem 2.2 (Hardness of Promise Bounded-Factor MinRank). *The Promise Bounded-Factor MinRank Problem is NP-hard under one-to-one randomized polynomial-time reductions, with the following parameters:*

- $p = \Theta(2^m)$;
- $r = \Theta(\log m)$;

where m is polynomial in the circuit size of the SAT statement. The reduction runs in polynomial time and has negligible error probability.

Furthermore, the reduction translates witness in the following sense: given any witness of the original SAT statement, one can efficiently compute nonzero coefficients $a_1, \dots, a_k$ and vectors $\mathbf{u}, \mathbf{v}$ with $\sum_i a_i \mathbf{M}_i = \mathbf{u}\mathbf{v}^T$ and $\mathbf{v} \in \{0, 1\}^m$; given any nonzero coefficients $a_1, \dots, a_k$ that $\sum_i a_i \mathbf{M}_i$ is of rank $\leq r$, the witness of the original SAT statement can be efficiently computed.

The only randomness behind Theorem 2.2 is to pick a large prime $p = \Theta(2^m)$; the reduction is deterministic once a sufficiently large prime is supplied.

2.2 The KOALA Assumption

Consider a cyclic group described as $\mathcal{G} = (p, \mathbb{G}, g)$, where the order p is of $\Theta(\lambda)$ bits, the group is generated by g and group operations can be efficiently computed. For a matrix $\mathbf{A}$ over $\mathbb{Z}_p$, write $[\mathbf{A}] := g^{\mathbf{A}}$, component-wise the exponentiation.

The Knowledge of OrthogonALity Assumption (KOALA) of Beullens and Wee [BW19] says that an adversary able to distinguish $[\mathbf{s}^T \mathbf{V}]$ from a random vector of group elements, where $\mathbf{s}^T \mathbf{V}$ is a vector sampled from a low-dimensional subspace $\mathbf{V}$, must already know a nonzero vector orthogonal to the subspace. In our statement, $\mathbf{V}$ is produced by a sampler from public random string aux .

Definition 2.3 (KOALA). Let $\text{Samp} = \{\text{Samp}_\lambda\}_\lambda$ be an efficient non-uniform sampling algorithm that on input a prime p and a string aux (and potentially additional randomness) outputs a matrix $\mathbf{V} \in \mathbb{Z}_p^{\ell_1 \times \ell_2}$ with $\ell_1 < \ell_2$. For an efficient adversary $\mathcal{A} = \{\mathcal{A}_\lambda\}_\lambda$ set

$$\text{Adv}_{\mathcal{A}, \text{GroupGen}, \text{Samp}}^{\text{dist}}(\lambda) := |\Pr[\mathcal{A}_\lambda(\mathcal{G}, \text{aux}, [\mathbf{sV}]) \rightarrow 1] - \Pr[\mathcal{A}_\lambda(\mathcal{G}, \text{aux}, [\mathbf{r}]) \rightarrow 1]|,$$

where the probabilities are taken over the choice of uniformly random aux , $\mathcal{G} = (p, \mathbb{G}, g) \leftarrow \text{GroupGen}(1^\lambda)$, $\mathbf{V} \leftarrow \text{Samp}_\lambda(p, \text{aux})$, $\mathbf{s} \leftarrow \mathbb{Z}_p^{\ell_1}$, $\mathbf{r} \leftarrow \mathbb{Z}_p^{\ell_2}$, and the randomness of Adv_λ .

For an efficient adversary $\mathcal{B} = \{\mathcal{B}_\lambda\}_\lambda$ set

$$\text{Adv}_{\mathcal{B}, \text{GroupGen}, \text{Samp}}^{\text{find}}(\lambda) := \Pr[\mathcal{B}_\lambda(\mathcal{G}, \text{aux}) \rightarrow \mathbf{x} \wedge \mathbf{xV}^\top = \mathbf{0} \wedge \mathbf{x} \neq \mathbf{0}],$$

where the probabilities are taken over a uniformly random aux , $\mathcal{G} \leftarrow \text{GroupGen}(1^\lambda)$, $\mathbf{V} \leftarrow \text{Samp}_\lambda(p, \text{aux})$, $\mathbf{s} \leftarrow \mathbb{Z}_p^{\ell_1}$, $\mathbf{r} \leftarrow \mathbb{Z}_p^{\ell_2}$, and the randomness of the $\mathcal{B}_\lambda$.

The *KOALA assumption* holds for GroupGen if for every efficient adversary $\mathcal{A}$ and every efficient sampler Samp , there is an efficient adversary $\mathcal{B}$ and a polynomial Q such that

$$\text{Adv}_{\mathcal{B}, \text{GroupGen}, \text{Samp}}^{\text{find}}(\lambda) \geq \text{Adv}_{\mathcal{A}, \text{GroupGen}, \text{Samp}}^{\text{dist}}(\lambda)/Q(\lambda) - \text{negl}(\lambda).$$

Remark 2.4. Here aux is public uniform randomness. As pointed out by [AWY20], a stronger variant of KOALA where Samp chooses aux along with $\mathbf{V}$ is unlikely to be true.

2.3 Extractable Witness Encryption

Let $\mathcal{R}$ be an NP relation, $L_{\mathcal{R}} = \{x : \exists w, (x, w) \in \mathcal{R}\}$ be the NP language it defines, and $\mathcal{R}_x = \{w : (x, w) \in \mathcal{R}\}$ be the witness for x . A witness encryption scheme [GGSW13] encrypts a message to a statement x : anyone holding a witness for x can decrypt; on the other hand, if x is a NO instance, the message remains hidden. Extractable witness encryption gives the stronger guarantee that if an adversary can decrypt, then it must know a witness for x . We state the formal definition of *public-coin* extractable witness encryption.

Syntax. An (extractable) witness encryption scheme for $\mathcal{R}$ is a pair of polynomial-time algorithms (Enc, Dec) satisfying the following properties:

- $\text{Enc}(1^\lambda, x, b) \rightarrow \text{ct}$. This is a probabilistic algorithm. It takes the security parameter 1^λ , a statement x and a bit $b \in \{0, 1\}$, and outputs a ciphertext ct .
- $\text{Dec}(x, w, \text{ct}) \rightarrow b'$. This is a deterministic algorithm. It takes a statement x , a witness w and a ciphertext ct , and outputs a bit b' or the failure symbol $\perp$.

Definition 2.5 (Public-coin hard-statement sampler). An efficient non-uniform sampling algorithm $\Phi = \{\Phi_\lambda\}_\lambda$ is a *public-coin hard-statement sampler* for $\mathcal{R}$ if $\Phi_\lambda(\text{aux})$ is a statement for every aux , and for every efficient non-uniform algorithm $\mathcal{A} = \{\mathcal{A}_\lambda\}_\lambda$ there is a negligible function ε such that for all λ ,

$$\Pr_{\text{aux}}[(x, w) \in \mathcal{R} : x \leftarrow \Phi_\lambda(\text{aux}), w \leftarrow \mathcal{A}_\lambda(\text{aux})] \leq \varepsilon(\lambda).$$

The sampler and the attacker receive the same random coins aux , so Φ can keep no secret from $\mathcal{A}$.

Definition 2.6 (Public-coin extractable witness encryption). A pair (Enc, Dec) with the syntax above is a *public-coin extractable witness encryption scheme* for $\mathcal{R}$ if the following hold.

- *Correctness*: for every NP statement x and every $b \in \{0, 1\}$,

$$\Pr[\forall w \in \mathcal{R}_x, \text{Dec}(x, w, \text{ct}) = b : \text{ct} \leftarrow \text{Enc}(1^\lambda, x, b)] \geq 1 - \text{negl}(\lambda),$$

the probability being over the coins of Enc .

- *Security*: for every public-coin hard-statement sampler Φ for $\mathcal{R}$ and every efficient non-uniform distinguisher $\mathcal{D} = \{\mathcal{D}_\lambda\}_\lambda$ there is a negligible function ε such that for all λ ,

$$\left| \Pr[\mathcal{D}_\lambda(\text{aux}, \text{ct}) = 1 : x \leftarrow \Phi_\lambda(\text{aux}), \text{ct} \leftarrow \text{Enc}(1^\lambda, x, 0)] \right. \\ \left. - \Pr[\mathcal{D}_\lambda(\text{aux}, \text{ct}) = 1 : x \leftarrow \Phi_\lambda(\text{aux}), \text{ct} \leftarrow \text{Enc}(1^\lambda, x, 1)] \right| \leq \varepsilon(\lambda),$$

where the probability is over aux and the coins of Enc .

In other words, the security definition says that any distinguisher can be transformed into a violation of the “hard-statement” property of Φ_λ , i.e. an algorithm that takes as input x sampled from Φ_λ and outputs a witness w for it.

2.4 Indistinguishability Obfuscation and Differing-Input Obfuscation

Syntax. An obfuscator for $\mathcal{C}$ is a polynomial-time algorithm Obf .

- $\text{Obf}(1^\lambda, C) \rightarrow C'$. This is a probabilistic algorithm. It takes the security parameter 1^λ and a circuit C , and outputs a circuit C' .

Definition 2.7 (Correctness). An obfuscator Obf for $\mathcal{C}$ is *correct* if for every λ , every C of size at most 2^λ ,

$$\Pr[\forall x, C'(x) = C(x) : C' \leftarrow \text{Obf}(1^\lambda, C)] \geq 1 - \text{negl}(\lambda),$$

the probability being over the coins of Obf .

Definition 2.8 (Indistinguishability obfuscation). An obfuscator Obf is an *indistinguishability obfuscator* if for every non-uniform circuit family $\{C_{\lambda,0}, C_{\lambda,1}\}_\lambda$ with $C_{\lambda,0} \equiv C_{\lambda,1}$ and $|C_{\lambda,0}| = |C_{\lambda,1}|$ for all $\lambda \in \mathbb{N}$, and for every efficient non-uniform distinguisher $\mathcal{D} = \{\mathcal{D}_\lambda\}_\lambda$, there is a negligible function ε such that for all λ ,

$$\left| \Pr[\mathcal{D}_\lambda(C') = 1 : C' \leftarrow \text{Obf}(1^\lambda, C_0)] \right. \\ \left. - \Pr[\mathcal{D}_\lambda(C') = 1 : C' \leftarrow \text{Obf}(1^\lambda, C_1)] \right| \leq \varepsilon(\lambda),$$

the probability being over the coins of Obf .

Definition 2.9 (Public-coin differing-inputs sampler). An efficient non-uniform sampling algorithm $\Phi = \{\Phi_\lambda\}_\lambda$ is a *public-coin differing-inputs sampler* if it outputs two circuits of same size, and for every efficient non-uniform algorithm $\mathcal{A} = \{\mathcal{A}_\lambda\}_\lambda$ there is a negligible function ε such that for all λ ,

$$\Pr_{\text{aux}}[C_0(x) \neq C_1(x) : (C_0, C_1) \leftarrow \Phi_\lambda(\text{aux}), x \leftarrow \mathcal{A}_\lambda(\text{aux})] \leq \varepsilon(\lambda).$$

Definition 2.10 (Public-coin differing-inputs obfuscation). An obfuscator Obf for $\mathcal{C}$ is a *public-coin differing-inputs obfuscator* if for every public-coin differing-inputs sampler Φ and every efficient non-uniform distinguisher $\mathcal{D} = \{\mathcal{D}_\lambda\}_\lambda$ there is a negligible function ε such that for all λ ,

$$\left| \Pr[\mathcal{D}_\lambda(\text{aux}, C') = 1 : (C_0, C_1) \leftarrow \Phi_\lambda(\text{aux}), C' \leftarrow \text{Obf}(1^\lambda, C_0)] \right. \\ \left. - \Pr[\mathcal{D}_\lambda(\text{aux}, C') = 1 : (C_0, C_1) \leftarrow \Phi_\lambda(\text{aux}), C' \leftarrow \text{Obf}(1^\lambda, C_1)] \right| \leq \varepsilon(\lambda),$$

the probability being over aux and the coins of Obf .

3 Extractable Witness Encryption from KOALA

We show that the witness encryption construction of Hair and Sahai [HS26] is in fact a *public-coin extractable* witness encryption scheme, under the KOALA assumption (Definition 2.3).

Theorem 3.1 (Extractable Witness encryption for promise Bounded-Factor MinRank). *Assume the KOALA assumption holds for some GroupGen (Definition 2.3).*

Let $r = \Theta(\log m)$, and $p = \Theta(2^m)$. There exists a witness encryption scheme for promise Bounded-Factor MinRank with correctness probability at least $1 - \text{poly}(m)/p$ for each witness. Moreover, for every public coin statement sampler $(\mathbf{M}_1, \dots, \mathbf{M}_k) \leftarrow \Phi_\lambda(\text{aux})$, and every PPT adversary $\text{Adv} = \{\text{Adv}_\lambda\}$ that distinguishes encryption of 0 or 1 under the statement sampled by Φ_λ with non-negligible advantage $\delta = \delta(\lambda)$, there is an efficient adversary $\mathcal{B} = \{\mathcal{B}_\lambda\}$ that takes aux and the statement as input, and outputs, with non-negligible probability, nonzero coefficients $a_1, \dots, a_k \in \mathbb{F}_p$, satisfying

$$\text{rank} \left(\sum_i a_i \mathbf{M}_i \right) \leq r.$$

In other words, $\mathcal{B}$ outputs a witness violating the NO case.

Using the NP-hardness of Bounded-Factor MinRank, one can easily construct an extractable witness encryption scheme for general NP statements.

Theorem 3.2 (Witness encryption for NP). *Assume the KOALA assumption holds for some GroupGen. Then SAT has a public-coin extractable witness encryption scheme.*

Proof. The witness encryption runs the randomized reduction (Theorem 2.2) from SAT to promise Bounded-Factor MinRank, and publishes the ciphertext under MinRank as well as the randomness used in the reduction.

For correctness, the reduction has negligible error probability; furthermore, we set $m > n + O(\lambda)$ and $p = \Theta(2^m)$ such that correctness holds for all witnesses simultaneously with probability $1 - 2^n \text{poly}(m)/p \geq 1 - 2^{-\lambda}$.

For extractability, an extracted witness violating the NO case (as in Theorem 3.1) efficiently yields one for the original statement. $\square$

3.1 Construction

Here we state the construction for [Theorem 3.1](#). We note that the construction is identical to [\[HS26\]](#).

An instance x is a basis $\mathbf{M}_1, \dots, \mathbf{M}_k$ of a linear space $\mathcal{S} \subseteq \mathbb{F}_p^{m \times m}$. write $\mathbf{M}_{\mathbf{a}} = \sum_{j=1}^k a_j \mathbf{M}_j$. The *decryption* witness is a rank-one element of $\mathcal{S}$ together with its decomposition:

$$\mathcal{R}_{\text{dec}} = \left\{ (x, (\mathbf{a}, \mathbf{u}, \mathbf{v})) : \mathbf{M}_{\mathbf{a}} = \mathbf{u}\mathbf{v}^\top \neq \mathbf{0}, \mathbf{v} \in \{0, 1\}^m \right\},$$

while the *extraction* gives coefficients witnessing low rank, with r the threshold of [Problem 2.1](#):

$$\mathcal{R}_{\text{ext}} = \left\{ (x, \mathbf{a}) : \mathbf{a} \neq \mathbf{0}, \text{rank}(\mathbf{M}_{\mathbf{a}}) \leq r \right\}.$$

Construction 3.3 (Witness encryption for MinRank).

- Enc($1^\lambda, x, \beta$): Given an instance $x = (\mathbf{M}_1, \dots, \mathbf{M}_k)$ of [Problem 2.1](#), and a message $\beta \in \{0, 1\}$, sample a cyclic group $\mathbb{G} = \langle g \rangle$ of prime order p . Then sample

$$\mathbf{s} \leftarrow \mathbb{F}_p^m, \quad \mathbf{r} \leftarrow \{0, \dots, m-1\}^m, \quad \eta \leftarrow \mathbb{F}_p^k,$$

set, for $j = 1, \dots, k$,

$$y_j = \begin{cases} \mathbf{s}^\top \mathbf{M}_j \mathbf{r}, & \beta = 0, \\ \eta_j, & \beta = 1, \end{cases}$$

and output

$$\text{ct} = (\mathbb{G}, p, g, [\mathbf{s}], [\mathbf{y}]),$$

where $[\mathbf{s}] := (g^{s_1}, \dots, g^{s_m})$ and $[\mathbf{y}] := (g^{y_1}, \dots, g^{y_k})$.

- Dec(x, w, ct): Given a valid witness $w = (\mathbf{a}, \mathbf{u}, \mathbf{v})$ for $\mathcal{R}_{\text{dec}}$, parse $\text{ct} = (\mathbb{G}, p, g, [\mathbf{s}], [\mathbf{y}])$ and compute

$$H = \prod_{i=1}^m (g^{s_i})^{u_i}, \quad Z = \prod_{j=1}^k (g^{y_j})^{a_j}, \quad B = (m-1) \text{wt}(\mathbf{v}).$$

Output 0 if $Z \in \{1, H, H^2, \dots, H^B\}$, and output 1 otherwise.

Lemma 3.4 (Correctness). *Let x be an Bounded-Factor MinRank instance with witness $w = (\mathbf{a}, \mathbf{u}, \mathbf{v})$ and $\beta \in \{0, 1\}$.*

$$\Pr[\text{Dec}(x, w, \text{ct}) = \beta : \text{ct} \leftarrow \text{Enc}(1^\lambda, x, \beta)] \geq 1 - \frac{m^2}{p}.$$

the probability being over the coins of Enc.

Proof. Write

$$\mathbf{M}_{\mathbf{a}} = \sum_j a_j \mathbf{M}_j = \mathbf{u}\mathbf{v}^\top$$

and $H = g^{\mathbf{u}^\top \mathbf{s}}$.

If $\beta = 0$ then

$$Z = g^{\sum_j a_j \mathbf{s}^\top \mathbf{M}_j \mathbf{r}} = g^{\mathbf{s}^\top \mathbf{M}_{\mathbf{a}} \mathbf{r}} = g^{(\mathbf{u}^\top \mathbf{s})(\mathbf{v}^\top \mathbf{r})} = H^{\mathbf{v}^\top \mathbf{r}}$$

and since $\mathbf{v} \in \{0, 1\}^m$ and $\mathbf{r} \in \{0, \dots, m-1\}^m$ we have $0 \leq \mathbf{v}^\top \mathbf{r} \leq (m-1) \text{wt}(\mathbf{v}) = B$. So Z lies in the list and the output is 0. Therefore, in this case the decryption is correct with probability 1.

If $\beta = 1$ then $Z = g^{\mathbf{a}^\top \eta}$. As $\mathbf{a} \neq \mathbf{0}$, $\mathbf{a}^\top \eta$ is uniform on $\mathbb{F}_p$ and independent of H . The decryption output is 0 only if Z is one of the at most $B + 1 \leq m^2$ elements of the list, which happens with probability at most m^2/p . $\square$

Lemma 3.5 (Extractability). *Assume the KOALA assumption holds for some GroupGen (Definition 2.3). Then Construction 3.3 is a public-coin extractable witness encryption scheme (Definition 2.6) for the NP relation $\mathcal{R}_{\text{ext}}$.*

Proof. The proof proceeds via the following outline: first, assume that there is a public coin statement sampler $\Phi = \{\Phi_\lambda\}_{\lambda \in \mathbb{N}}$ which generates an NP statement $x := (\mathbf{M}_1, \dots, \mathbf{M}_k) = \Phi_\lambda(\text{aux})$, and a PPT distinguisher $\mathcal{A} = \{\mathcal{A}_\lambda\}_{\lambda \in \mathbb{N}}$ that takes (aux, ct) and distinguishes with non-negligible advantage $\delta = \delta(\lambda)$ encryptions of 0 or 1 under the statement x sampled by Φ_λ . We will first show how to turn this into a sampler-distinguisher pair for the KOALA assumption (Step 1 below). Then, we apply the KOALA assumption to get out of them a KOALA extractor. Finally, we show how to turn the KOALA extractor into an extractor for the witness encryption scheme (Step 2 below). This extractor $\mathcal{B} = \{\mathcal{B}_\lambda\}_{\lambda \in \mathbb{N}}$ takes aux and the statement x as input, and outputs, with non-negligible probability, a witness w that x is *not* a NO instance. That is, nonzero coefficients $a_1, \dots, a_k \in \mathbb{F}_p$ satisfying

$$\text{rank} \left(\sum_i a_i \mathbf{M}_i \right) \leq r.$$

Step 1. Starting from $(\Phi, \mathcal{A})$, we construct the KOALA sampler $\Psi = \{\Psi_\lambda\}_{\lambda \in \mathbb{N}}$ as follows. Given aux that determines an instance $x = (\mathbf{M}_1, \dots, \mathbf{M}_k)$, Ψ_λ samples $\mathbf{r} \in \{0, \dots, m-1\}^m$ and outputs

$$\mathbf{V}_x(\mathbf{r}) = (\mathbf{I}_m \mid \mathbf{M}_1 \mathbf{r} \mid \dots \mid \mathbf{M}_k \mathbf{r}) \in \mathbb{F}_p^{m \times (m+k)}.$$

Note that for every $\mathbf{s} \in \mathbb{F}_p^m$,

$$\mathbf{s}^\top \mathbf{V}_x(\mathbf{r}) = (s_1, \dots, s_m, \mathbf{s}^\top \mathbf{M}_1 \mathbf{r}, \dots, \mathbf{s}^\top \mathbf{M}_k \mathbf{r}).$$

Hence, the vector $[\mathbf{s}^\top \mathbf{V}_x(\mathbf{r})]$ is exactly the pair $([\mathbf{s}], [\mathbf{y}])$ that forms an encryption of 0 under x .

We also construct a KOALA distinguisher $\mathcal{A}'$ using the witness encryption distinguisher $\mathcal{A}$: $\mathcal{A}'$ gets as input $(\text{aux}, [\mathbf{t}])$ and splits $[\mathbf{t}]$ into its first m and last k entries, $[\mathbf{t}_1]$ and $[\mathbf{t}_2]$, sets $\text{ct} = (\mathbb{G}, p, g, [\mathbf{t}_1], [\mathbf{t}_2])$, and outputs $\mathcal{A}_\lambda(\text{aux}, \text{ct})$. This concludes Step 1.

We are now in a position to apply the KOALA assumption with $\ell_1 = m < m+k = \ell_2$ to the KOALA sampler Ψ and KOALA adversary $\mathcal{A}'$. Note that $\mathcal{A}'$ runs $\mathcal{A}$ on an encryption of 0 under x when $\mathbf{t} = \mathbf{s}^\top \mathbf{V}$, and on an encryption of 1 under x when $\mathbf{t}$ is uniform. So, by assumption, $\text{Adv}_{\mathcal{A}', \text{GroupGen, Samp}}^{\text{dist}}(\lambda) = \delta(\lambda)$, and the KOALA assumption gives an efficient extractor $\mathcal{B}'$ and a polynomial Q such that

$$\text{Adv}_{\mathcal{B}', \text{GroupGen, Samp}}^{\text{find}}(\lambda) \geq \delta(\lambda)/Q(\lambda) - \text{negl}(\lambda).$$

Step 2. Recall that $\mathcal{B}'$, on input $\mathcal{G} \leftarrow \text{GroupGen}(1^\lambda)$ and aux , outputs $(\mathbf{c}, \mathbf{a}) \in \mathbb{F}_p^m \times \mathbb{F}_p^k$ such that $\mathbf{V}_x(\mathbf{r}) \cdot (\mathbf{c} \mid \mathbf{a})^t = \mathbf{0}$. By our construction of $\mathbf{V}_x$, this means that

$$\mathbf{c} + \sum a_i \mathbf{M}_i \mathbf{r} = \mathbf{c} + \mathbf{M}_a \mathbf{r} = \mathbf{0}. \quad (3.1)$$

We now construct the witness encryption extractor $\mathcal{B}$. $\mathcal{B}$, on input aux which determines the statement x , samples $\mathcal{G} \leftarrow \text{GroupGen}(1^\lambda)$, runs $\mathcal{B}'(\mathcal{G}, \text{aux})$, parses its output as $(\mathbf{c}, \mathbf{a}) \in \mathbb{F}_p^m \times \mathbb{F}_p^k$, and outputs $\mathbf{a}$.

First, note that if Equation (3.1) holds and $(\mathbf{c}, \mathbf{a}) \neq \mathbf{0}$, then $\mathbf{a} \neq \mathbf{0}$, as $\mathbf{a} = \mathbf{0}$ would force $\mathbf{c} = \mathbf{0}$. It remains to show that if Equation (3.1) holds, then with high probability $\mathbf{M}_a$ has rank at most r .

Crucially, $\mathcal{B}'$ only knows $\mathbf{aux}$ (hence the statement x) but does not know $\mathbf{r}$ at all. If it chooses $(\mathbf{c}, \mathbf{a}) \neq \mathbf{0}$, then for this $\mathbf{c}$ and $\mathbf{M}_{\mathbf{a}}$, with noticeable probability Equation (3.1) holds, over a random $\mathbf{r} \leftarrow \{0, \dots, m-1\}^m$.

On the other hand, for every fixed $\mathbf{M} \in \mathbb{F}_p^{m \times m}$ of rank at least ρ and every fixed $\mathbf{c} \in \mathbb{F}_p^m$,

$$\Pr_{\mathbf{r} \leftarrow \{0, \dots, m-1\}^m} [\mathbf{c} + \mathbf{M}\mathbf{r} = \mathbf{0}] \leq m^{-\rho}. \quad (3.2)$$

Indeed, choose ρ linearly independent columns of $\mathbf{M}$ and fix the coordinates of $\mathbf{r}$ outside them arbitrarily. As the chosen columns are independent, at most one of the m^ρ equally likely values of the remaining ρ coordinates satisfies the equation.

Applying Equation (3.2) to $\mathbf{M} = \mathbf{M}_{\mathbf{a}}$ gives

$$\Pr[\text{Equation (3.1) holds} \wedge \text{rank}(\mathbf{M}_{\mathbf{a}}) > r] \leq m^{-(r+1)}.$$

Every other successful output of $\mathcal{B}'$ has $\text{rank}(\mathbf{M}_{\mathbf{a}}) \leq r$, and therefore

$$\Pr[\mathbf{a} \neq \mathbf{0} \wedge \text{rank}(\mathbf{M}_{\mathbf{a}}) \leq r] \geq \frac{\delta(\lambda)}{Q(\lambda)} - \text{negl}(\lambda) - m^{-(r+1)},$$

For the parameters $r = \Theta(\log m)$ and $p = \Theta(2^m)$ of Theorem 3.1, we have $m = \text{poly}(\lambda)$, so $m^{-(r+1)} = \lambda^{-\Omega(\log \lambda)}$ is negligible. Hence $\mathcal{B}$ succeeds with non-negligible probability whenever δ is non-negligible. $\square$

4 Boosting Extractable WE to diO

We show that public-coin extractable witness encryption, together with indistinguishability obfuscation, implies public-coin differing-inputs obfuscation. Combined with Theorem 3.2, it follows that indistinguishability obfuscation and the KOALA assumption together imply public-coin differing-inputs obfuscation.

Theorem 4.1 (Differing-inputs obfuscation from extractable witness encryption). *Assume there is a public-coin extractable witness encryption scheme for NP (Definition 2.6), and there is a indistinguishability obfuscation (Definitions 2.7 and 2.8). Then there is a public-coin differing-inputs obfuscation (Definitions 2.7 and 2.10).*

With Theorem 3.2, there exists public coin diO from KOALA and iO.

Theorem 4.2. *Assume KOALA (Definition 2.3) and iO, there exists public-coin differing-inputs obfuscation.*

4.1 Construction and Analysis

Our construction adapts the transformation from (succinct) witness encryption to (succinct) indistinguishability obfuscation of Jain, Jin, Mathialagan and Paneth [JJMP25].

Let iO be an indistinguishability obfuscator (Definitions 2.7 and 2.8), and let (Enc, Dec) be a public-coin extractable witness encryption scheme (Definition 2.6) for the NP relation

$$\mathcal{R}_{\neq} = \{((C_0, C_1), x) : C_0(x) \neq C_1(x)\},$$

whose statements are pairs of circuits of the same size, and whose witnesses are the inputs on which the two circuits differ.

Padding. Let s be the size of the circuit to be obfuscated. Looking ahead, we remark that the security proof replaces it by a circuit $P_{C_0, C_1, \text{ct}}$, defined in $\text{Hyb}_{b,2}$ below, that has two circuits of size s and a ciphertext hardwired. The size of $P_{C_0, C_1, \text{ct}}$ is bounded by a fixed polynomial $L(\lambda, s)$ whenever $|C_0| = |C_1| = s$. We choose L large enough to also accommodate a circuit of size s , and pad both circuits to this size: for a circuit G of size at most $L(\lambda, s)$, we write G^{pad} for a circuit of size exactly $L(\lambda, s)$ that computes the same function.

Construction 4.3 (Differing-inputs obfuscator).

- $\text{diO}(1^\lambda, C)$: Let $s = |C|$, and let C^{pad} be C padded to size $L(\lambda, s)$. Output $C' \leftarrow \text{iO}(1^\lambda, C^{\text{pad}})$.

Lemma 4.4 (Correctness). *For every λ and every circuit C with $L(\lambda, |C|) \leq 2^\lambda$,*

$$\Pr[\forall x, C'(x) = C(x) : C' \leftarrow \text{diO}(1^\lambda, C)] \geq 1 - \text{negl}(\lambda),$$

the probability being over the coins of diO .

Proof. Immediate by the correctness of iO (Definition 2.7). □

Security analysis. Fix a public-coin differing-inputs sampler Φ (Definition 2.9) and an efficient non-uniform distinguisher $\mathcal{D} = \{\mathcal{D}_\lambda\}_\lambda$. For $b \in \{0, 1\}$ we consider the following hybrids. In the following aux is uniformly random, $(C_0, C_1) \leftarrow \Phi_\lambda(\text{aux})$, and $s = |C_0| = |C_1|$.

$\text{Hyb}_{b,1}$: Output (aux, C') for $C' \leftarrow \text{iO}(1^\lambda, C_b^{\text{pad}})$, that is, for $C' \leftarrow \text{diO}(1^\lambda, C_b)$.

$\text{Hyb}_{b,2}$: Sample the obfuscation as follows.

- Sample $\text{ct} \leftarrow \text{Enc}(1^\lambda, (C_0, C_1), b)$.
- Construct the following circuit $P_{C_0, C_1, \text{ct}}$. On input x :
 - If $C_0(x) = C_1(x)$, output $C_0(x)$.
 - Else, compute $\beta \leftarrow \text{Dec}((C_0, C_1), x, \text{ct})$. Output $C_1(x)$ if $\beta = 1$, and $C_0(x)$ otherwise.
- Output (aux, C') for $C' \leftarrow \text{iO}(1^\lambda, P_{C_0, C_1, \text{ct}}^{\text{pad}})$.

For a hybrid Hyb , write $p_\lambda(\text{Hyb}) = \Pr[\mathcal{D}_\lambda(\text{aux}, C') = 1 : (\text{aux}, C') \leftarrow \text{Hyb}]$.

Lemma 4.5 (Switching to the hybrid circuit). *For each $b \in \{0, 1\}$ there is a negligible function ε such that for all λ ,*

$$|p_\lambda(\text{Hyb}_{b,1}) - p_\lambda(\text{Hyb}_{b,2})| \leq \varepsilon(\lambda).$$

Proof. The two circuits in $\text{Hyb}_{b,1}$ and $\text{Hyb}_{b,2}$ are equivalent as long as the witness encryption ciphertext ct has correctness: for every valid witness w , it decrypts to b . By correctness of witness encryption (Definition 2.6), the later happens with probability $1 - \text{negl}(\lambda)$. Then the security of iO guarantees indistinguishability of $\text{Hyb}_{b,1}$ and $\text{Hyb}_{b,2}$. □

Lemma 4.6 (Switching the encrypted bit). *There is a negligible function ε such that for all λ ,*

$$|p_\lambda(\text{Hyb}_{0,2}) - p_\lambda(\text{Hyb}_{1,2})| \leq \varepsilon(\lambda).$$

Proof. We treat the differing-inputs sampler Φ as a public-coin hard-statement sampler for $\mathcal{R}_\neq$ (Definition 2.5). Apply the security of the extractable witness encryption scheme (Definition 2.6) to Φ and the following distinguisher.

- $\mathcal{D}'_\lambda(\text{aux}, \text{ct})$: Compute $(C_0, C_1) \leftarrow \Phi_\lambda(\text{aux})$, construct the circuit $P_{C_0, C_1, \text{ct}}$ as in the hybrids, sample $C' \leftarrow \text{iO}(1^\lambda, P_{C_0, C_1, \text{ct}}^{\text{pad}})$, and output $\mathcal{D}_\lambda(\text{aux}, C')$.

When $\text{ct} \leftarrow \text{Enc}(1^\lambda, (C_0, C_1), \beta)$ for $\beta \in \{0, 1\}$, the pair (aux, C') on which $\mathcal{D}'_\lambda$ runs $\mathcal{D}_\lambda$ is distributed as $\text{Hyb}_{\beta, 2}$. Hence $|p_\lambda(\text{Hyb}_{0, 2}) - p_\lambda(\text{Hyb}_{1, 2})|$ is the advantage of $\mathcal{D}'$ in Definition 2.6, which is negligible. $\square$

Proof of Theorem 4.1. The obfuscator diO is correct by Lemma 4.4, and is secure by Lemma 4.5 and Lemma 4.6. $\square$

AI use statement

The human authors came up with the idea that the Hair-Sahai witness encryption might be an extractable witness encryption under the KOALA assumption, and asked Codex to work out the details (and it turns out to be immediate). The rest of the ideas were generated by the human authors, who also thoroughly rewrote the paper. The human authors take full responsibility for this paper and all its contents.

References

- [ABG⁺13] Prabhakaran Ananth, Dan Boneh, Sanjam Garg, Amit Sahai, and Mark Zhandry. Differing-inputs obfuscation and applications. Cryptology ePrint Archive, Paper 2013/689, 2013. <https://eprint.iacr.org/2013/689>.
- [AWY20] Shweta Agrawal, Daniel Wichs, and Shota Yamada. Optimal broadcast encryption from LWE and pairings in the standard model. In *Theory of Cryptography – TCC 2020, Part I*, volume 12550 of *LNCS*, pages 149–178, 2020. <https://eprint.iacr.org/2020/1179>.
- [BCP14] Elette Boyle, Kai-Min Chung, and Rafael Pass. On extractability obfuscation. In *Theory of Cryptography – TCC 2014*, volume 8349 of *LNCS*, pages 52–73, 2014. <https://eprint.iacr.org/2013/650>.
- [BGI⁺01] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In *Advances in Cryptology – CRYPTO 2001*, volume 2139 of *LNCS*, pages 1–18, 2001. <https://eprint.iacr.org/2001/069>.
- [BP15] Elette Boyle and Rafael Pass. Limits of extractability assumptions with distributional auxiliary input. In *Advances in Cryptology – ASIACRYPT 2015, Part II*, volume 9453 of *LNCS*, pages 236–261, 2015. <https://eprint.iacr.org/2013/703>.
- [BSW16] Mihir Bellare, Igors Stepanovs, and Brent Waters. New negative results on differing-inputs obfuscation. In *Advances in Cryptology – EUROCRYPT 2016, Part II*, volume 9666 of *LNCS*, pages 792–821, 2016. <https://eprint.iacr.org/2016/162>.
- [BW19] Ward Beullens and Hoeteck Wee. Obfuscating simple functionalities from knowledge assumptions. In *Public-Key Cryptography – PKC 2019, Part II*, volume 11443 of *LNCS*, pages 254–283, 2019. <https://eprint.iacr.org/2019/056>.

- [GGH⁺13] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 40–49, 2013. <https://eprint.iacr.org/2013/451>.
- [GGHW14] Sanjam Garg, Craig Gentry, Shai Halevi, and Daniel Wichs. On the implausibility of differing-inputs obfuscation and extractable witness encryption with auxiliary input. In *Advances in Cryptology – CRYPTO 2014, Part I*, volume 8616 of *LNCS*, pages 518–535, 2014. <https://eprint.iacr.org/2013/860>.
- [GGSW13] Sanjam Garg, Craig Gentry, Amit Sahai, and Brent Waters. Witness encryption and its applications. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC)*, pages 467–476, 2013. <https://eprint.iacr.org/2013/258>.
- [HS26] Isaac M Hair and Amit Sahai. Witness encryption via prime-order generic groups. arXiv:2609.18275, 2026. <https://arxiv.org/abs/2609.18275>.
- [IPS15] Yuval Ishai, Omkant Pandey, and Amit Sahai. Public-coin differing-inputs obfuscation and its applications. In *Theory of Cryptography – TCC 2015, Part II*, volume 9015 of *LNCS*, pages 668–697, 2015. <https://eprint.iacr.org/2014/942>.
- [JJMP25] Abhishek Jain, Zhengzhong Jin, Surya Mathialagan, and Omer Paneth. On succinct obfuscation via propositional proofs. In *Proceedings of the 66th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1703–1740, 2025. <https://eprint.iacr.org/2026/2009>.
- [JLS21] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from well-founded assumptions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 60–73, 2021. <https://eprint.iacr.org/2020/1003>.
- [JLS22] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from LPN over $\mathbb{F}_p$, DLIN, and PRGs in NC^0 . In *Advances in Cryptology – EUROCRYPT 2022, Part I*, volume 13275 of *LNCS*, pages 670–699, 2022. <https://eprint.iacr.org/2021/1334>.
- [RVV24] Seyoon Ragavan, Neekon Vafa, and Vinod Vaikuntanathan. Indistinguishability obfuscation from bilinear maps and LPN variants. In *Theory of Cryptography – TCC 2024, Part IV*, volume 15367 of *LNCS*, pages 3–36, 2024. <https://eprint.iacr.org/2024/856>.
- [SW14] Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: Deniable encryption, and more. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 475–484, 2014. <https://eprint.iacr.org/2013/454>.