An Algebraic Approach to Non-Malleability

Vipul Goyal* Silas Richelson† Alon Rosen‡ Margarita Vald§ ¶

Abstract

In their seminal work on non-malleable cryptography, Dolev, Dwork and Naor, showed how to construct a non-malleable commitment with logarithmically-many "rounds"/"slots", the idea being that any adversary may successfully maul in some slots but would fail in at least one. Since then new ideas have been introduced, ultimately resulting in constant-round protocols based on any one-way function. Yet, in spite of this remarkable progress, each of the known constructions of non-malleable commitments leaves something to be desired.

In this paper we propose a new technique that allows us to construct a non-malleable protocol with only a single "slot", and to improve in at least one aspect over each of the previously proposed protocols. Two direct byproducts of our new ideas are a four round non-malleable commitment and a four round non-malleable zero-knowledge argument, the latter matching the round complexity of the best known zero-knowledge argument (without the non-malleability requirement). The protocols are based on the existence of one-way functions and admit very efficient instantiations via standard homomorphic commitments and sigma protocols.

Our analysis relies on algebraic reasoning, and makes use of error correcting codes in order to ensure that committers' tags differ in many coordinates. One way of viewing our construction is as a method for combining many atomic sub-protocols in a way that simultaneously amplifies soundness and non-malleability, thus requiring much weaker guarantees to begin with, and resulting in a protocol which is much trimmer in complexity compared to the existing ones.

Keywords: Non-malleability, commitments, zero-knowledge

*Microsoft Research, Bangalore. Email: vipul@microsoft.com. Part of this work done while visiting IDC Herzliya.

†UCLA. Email: SiRichel@math.ucla.edu. Work done while visiting IDC Herzliya. Supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 307952.

‡Efi Arazi School of Computer Science, IDC Herzliya, Israel. Email: alon.rosen@idc.ac.il. Work supported by ISF grant no. 1255/12 and by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 307952.

§The Blavatnik School of Computer Science, Tel Aviv University, Israel. Email: margarita.vald@cs.tau.ac.il. Work supported by ISF grant no. 1255/12 and by the Check Point Institute for Information Security.

¶Preliminary version of this paper was published in proceedings of FOCS, 2014.

1 Introduction

The notion of non-malleability is central in cryptographic protocol design. Its objective is to protect against a man-in-the-middle (MIM) attacker that has the power to intercept messages and transform them in order to harm the security in other instantiations of the protocol. Commitment is often used as the paragon example for non-malleable primitives because of its ability to almost "universally" secure higher-level protocols against MIM attacks.

Commitments allow one party, called the committer, to probabilistically map a message m into a string, $\mathrm{Com}(m;r)$, which can be then sent to another party, called the receiver.[1] In the statistically binding variant, the string $\mathrm{Com}(m;r)$ should be *binding*, in that it cannot be later "opened" into a message $m' \neq m$. It should also be *hiding*, meaning that for any pair of messages, m, m', the distributions $\mathrm{Com}(m;r)$ and $\mathrm{Com}(m';r')$ are computationally indistinguishable.

A commitment scheme is said to be *non-malleable* if for every message m, no MIM adversary, intercepting a commitment $\mathrm{Com}(m;r)$ and modifying it at will, is able to efficiently generate a commitment $\mathrm{Com}(\tilde{m};\tilde{r})$ to a related message $\tilde{m}$. Interest in non-malleable commitments is motivated both by the central role that they play in securing protocols under composition (see for example [CLOS02, LPV09]) and by the unfortunate reality that many widely used commitment schemes are actually highly malleable. Indeed, man-in-the-middle (MIM) attacks occur quite naturally when multiple concurrent executions of protocols are allowed, and can be quite devastating.

Beyond protocol composition, non-malleable commitments are known to be applicable in secure multi-party computation [KOS03, Wee10, Goy11], authentication [NSS06], as well as a host of other non-malleable primitives (e.g., coin flipping, zero-knowledge, etc.), and even into applications as diverse as position based cryptography [CGMO09].

1.1 Prior Work

Since their conceptualization by Dolev, Dwork and Naor [DDN91], non-malleable commitments have been studied extensively, and with increasing success in terms of characterizing their round-efficiency and the underlying assumptions required. By now, we know how to construct constant-round non-malleable commitments based on any one-way function, and moreover the constructions are fully black-box. While this might give the impression that non-malleable commitments are well understood, each of the currently known constructions leaves something to be desired.

The first construction, due to DDN is perhaps the simplest and most efficient, mainly because it can in principle be instantiated with highly efficient cryptographic "sub-protocols". This, however, comes at the cost of round-complexity that is logarithmic in the maximum overall number of possible committers. Subsequent works, due to Barak [Bar02], Pass [Pas04], and, Pass and Rosen [PR05] are constant-round, but rely on highly inefficient non-black-box techniques. Wee [Wee10] (relying on [PW10]) gives a constant-round black-box construction under the assumption that sub-exponentially hard one-way functions exist. This construction employs a generic (and costly) transformation that is designed to handle general "non-synchronizing" MIM adversaries.

Finally, recent works by Goyal [Goy11] and Lin and Pass [LP11] attain non-malleable commitment with constant round-complexity via the minimal assumption that polynomial-time hard to invert one-way functions exist. The Lin-Pass protocol makes highly non-black-box use of the under-

[1] We consider non-interactive commitments in this discussion, in general commitment schemes may be interactive.

lying one-way function (though not of the adversary), along with a concept called signature chains; resulting in significant overhead. Most relevant to the current work is the work of Goyal [Goy11]. Goyal's protocol, using a later result of Goyal, Lee, Ostrovsky and Visconti [GLOV12], can be made fully black-box, with its only shortcomings being high-communication complexity and the use of the Wee transformation (or alternatively a similarly costly transformation due to Goyal [Goy11]) for handling non-synchronizing adversaries. To construct non-malleable commitments, our work follows the blueprint proposed by Goyal, and introduces new proof techniques to significantly trim down its complexity, making various parts of the original protocol unnecessary.

The current state of affairs is such that in spite of all the remarkable advances, the DDN construction and its analysis remain the simplest and arguably most appealing candidate for non-malleable commitments. This is both due to its black-boxness and because it does not require transformations for handling a non-synchronizing MIM (in fact, the protocol is purposefully designed to introduce asynchronicity in message scheduling, which can be then exploited in the analysis).

1.2 Our Results

In this work we introduce a new algebraic technique for obtaining non-malleability, resulting in a simple and elegant non-malleable commitment scheme. The scheme's analysis contains many fundamentally new ideas allowing us to overcome substantial obstacles without sacrificing efficiency. The protocol is constructed using any statistically binding commitment scheme as a building block, and hence requires the minimal assumption that one way functions exist.

Theorem. *Assume the existence of one-way functions. Then there is a $4-$round non-malleable commitment scheme.*

Our protocol enjoys the following appealing features, each of which makes it preferable in at least one way over any of the previously proposed protocols for non malleable commitment:

Simplicity. Compared to all previous protocols, ours is significantly simpler to describe and to instantiate (though not to analyze). The simplicity of the protocol also means that there is no need to introduce costly transformations for handling non-synchronizing adversaries.

Efficiency. In particular, ours is significantly more efficient than all prior protocols both in terms of round complexity, and in the sense that we use a surprisingly small number of sub-protocols, each of which can be instantiated in a very efficient way (e.g. using standard sigma protocols).

Assumption. The assumption underlying our main protocol is the existence of one-way functions, which is necessary for non-malleable commitments.

A direct consequence of our protocol is a 4-round non-malleable zero-knowledge argument based on any OWF. This demonstrates that for zero-knowledge, non-malleability does not necessarily come at the cost of extra rounds of interaction or complexity assumptions.

Theorem. *Assume the existence of one-way functions. Then there is a 4-round black-box non-malleable zero-knowledge argument for every language in NP.*

Beyond the above virtues, we believe that our new techniques are actually the most significant contributions of this work. In addition to our use of algebra, we make novel combinatorial use of error correcting codes in order to ensure that different committers' tags differ in many coordinates (more

on that later on). Whereas prior work relied on "worst-case" analysis of differences in committers' tags, ours follows from an "average-case" claim.

One way of viewing our construction is as a method for combining n atomic sub-protocols in a way that simultaneously amplifies their soundness and non-malleability properties, thus requiring much weaker soundness and non-malleability to begin with. We hope that this paradigm will become the norm for future work on in the area as, despite requiring more careful and strenuous analysis, it leads to pleasantly lightweight protocols.

Another payoff of the algebraic techniques we employ is that our protocol only has one "slot". Nearly all of the non-malleable commitment schemes in the literature use multiple slots of interaction as a way to set up imbalances between the two different protocol instantiations that the MIM is involved in. The well known "two slot trick" of [Pas04, PR05, Goy11], for example, is a way to turn an arbitrary asymmetry between the instantiations into two: one which is heavy on the right and one on the left. The inability of the MIM to align the imbalances is crucial to the proof of non-malleability. Running the two slots in parallel introduces several technical problems, most notably "if the two imbalances are side by side, won't they just cancel each other out?" Our analysis uses a computational version of the "linear independence of polynomial evaluation" mantra in order to argue that the MIM cannot combine the two imbalances and must deal with each one separately.

We stress that the use of algebra and error correcting codes does not yield such reward for free: the analysis required becomes substantially more difficult. In the next section we describe and briefly discuss our new protocol and extractor. We then outline our techniques, keeping it informal but pointing out several of the challenges faced and new ideas required to overcome them.

Subsequent Work. Shortly after this work, Brenner *et al.* [BGR+15] give an efficient implementation of a version of our non-malleable commitment scheme assuming the hardness of DDH over elliptic curve groups. More recently, Goyal, Pandey and Richelson [GPR16] give a three round non-malleable commitment scheme, matching the lower bound of [Pas13], assuming quasi-polynomially hard one-to-one OWFs exist. Their scheme uses the same method of extraction as our scheme, along with many new ideas. A recent work of Ciampi, Ostrovsky, Siniscalchi and Visconti [COSV17] improves the results in this work by enhancing our protocol to attain stronger security. The original version of this paper claimed incorrectly to achieve concurrent non-malleability (*i.e.*, non-malleability against a MIM adversary who can execute many protocol executions, instead of just two) in four rounds. The work [COSV17] pointed out and fixed this problem. Their ideas also allowed for constructing the first four-round delayed-input non-malleable zero-knowledge arguments from (standard) OWFs. Another work by the same authors [COSV16] enhances the protocol of [GPR16] to achieve concurrent non-malleable commitments, and a concurrent non-malleable identification scheme in three rounds assuming sub-exponentially secure one-way functions. Another line of works bypass the three-round lower bound by relying on sub-exponential hardness assumptions [GKS16, KS17, LPS17]. Khurana [Khu17] constructs three round non-malleable commitments by relying on the polynomial hardness of the decisional Diffie-Hellman (DDH) assumption. Finally, Goyal and Richelson [GR19] construct three-round non-malleable commitments based on polynomially hard one-to-one OWFs.

Outline of Paper. The remainder of the introduction is a technical overview of our main scheme and proof of non-malleability. Section 2 contains preliminaries and cryptographic definitions. In Section 3 we present our most basic new protocol $\langle \mathrm{C}, \mathrm{R} \rangle$: an eight-round commitment scheme

which is non-malleable against a synchronizing adversary (*i.e.*, a MIM who plays the corresponding rounds of the left and right sessions one after another). This protocol already exhibbits our "single-slot" technique. In Sections 4 and 5 we prove security of $\langle \mathrm{C}, \mathrm{R} \rangle$. Finally, in Section 6 we show how to reduce the round complexity of $\langle \mathrm{C}, \mathrm{R} \rangle$ and give constructions of four round non-malleable commitments and non-malleable zero-knowledge.

1.3 The New Protocol

Suppose that committer C wishes to commit to message m, and let $t_1, \ldots, t_n \in \mathbb{Z}$ be a sequence of tags that uniquely correspond to C's identity (more on the tags later). Let Com be a statistically binding commitment scheme, and suppose that $m \in \mathbb{F}_q$ where $q > \max_i 2^{t_i}$. The protocol proceeds as follows:

1. C chooses random $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{F}_q^n$ and sends $\mathrm{Com}(m)$ and $\{\mathrm{Com}(r_i)\}_{i=1}^n$ to R;
2. R sends C a query vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ where each α_i is drawn randomly from $\left[2^{t_i}\right] \subset \mathbb{F}_q$;
3. C sends R the response $\mathbf{a} = (a_1, \ldots, a_n)$ where $a_i = r_i \alpha_i + m$ for all $i \in [n]$;
4. C proves in ZK that $\mathbf{a}$ (from step 3) is consistent with m and $\mathbf{r}$ (from step 1).

The statistical binding property of the protocol follows directly from the binding of Com. The hiding property follows from the hiding of Com, the zero-knowledge property of the protocol used in step 4, and from the fact that for every i the receiver R observes only a single pair of the form (α_i, a_i), where $a_i = r_i \alpha_i + m$.

Note the role of C's tags in the protocol: t_i determines the size of the $i-$th coordinate's challenge space. Historically, non-malleable commitment schemes have used the tags as a way for the committer to encode its identity into the protocol as a mechanism to prevent M (whose tag is different from C's tag) from mauling. In our protocol the tags play the same role, albeit rather passively. For example, though the size of the $i-$th challenge space depends on t_i, the size of the total challenge space depends only on the sum $\sum_{i=1}^n t_i$ of the tags. In particular, our scheme leaves open the possibility that the left and right challenge spaces might have the same size (in fact this will be ensured by our choice of tags). This raises a red flag, as previous works go to great lengths to set up imbalances between the left and right challenge spaces in order to force M to "give more information than it gets". Nevertheless, we are able to prove that any mauling attack will fail.

Readers familiar with [Goy11] will recognize elements of this prior work in the above protocol. Indeed, it is possible to view our protocol, from a very high level, as an algebraic version of the protocol from [Goy11]. However, the fundamental difference we should emphasize, is that Goyal's work uses two slots and crucially relies on the challenge space in the left interaction being much smaller than the challenge space in the right. For us, the challenge spaces in the two interactions are exactly the same size and so new techniques are needed.

1.4 Proving Non-Malleability

Consider a MIM adversary M that is playing the role of the receiver in a protocol using tags $t_1, \ldots, t_n$ while playing the role of the committer in a protocol using tags $\tilde{t}_1, \ldots, \tilde{t}_n$ (we describe explicitly how to construct the tags from C's identity in Section 2). We refer to the former as the "left" interaction

and to the latter as the "right" interaction. We let m and $\tilde{m}$ denote the messages committed to in the left and right interactions respectively. Figure 1 below shows the two copies of our protocol played by a synchronizing MIM.

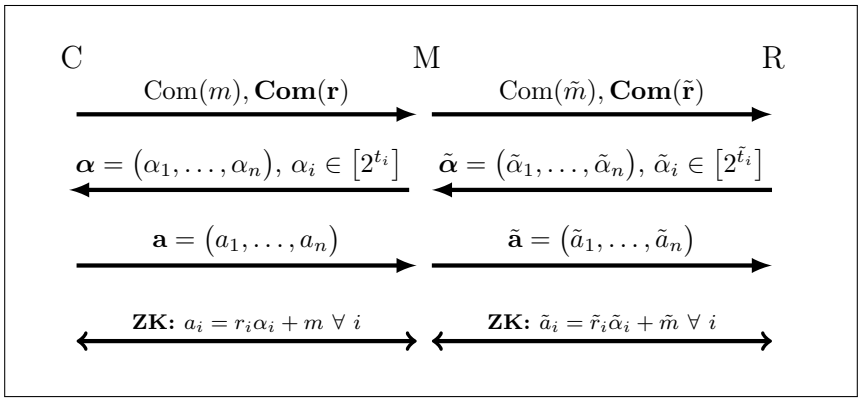

Figure 1: Protocol with Man-in-the-Middle

Our proof of non-malleability involves demonstrating the existence of an extractor, E, who is able to rewind M and extract $\tilde{m}$ without needing to rewind C in the left instantiation. Our extractor: (1) rewinds M to where $\tilde{\boldsymbol{\alpha}}$ was sent and asks a new query $\tilde{\boldsymbol{\beta}}$ instead, and (2) responds to M's left query randomly (it cannot do better without rewinding C as it does not know m), hoping that M's answer $\tilde{\mathbf{b}}$ on the right is correct. In general, there is no way for E to know whether M answered correctly or not, so our extractor: (3) rewinds again to the beginning of step 2 and asks another query $\tilde{\boldsymbol{\gamma}}$; (4) again, answers randomly on the left and receives $\tilde{\mathbf{c}}$. At this point E holds $\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\}$, where $(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}})$ is from the main thread. (5) For each $i \in [n]$, E interpolates $\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i)\}$ and $\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\gamma}_i, \tilde{c}_i)\}$, recovering candidate messages $\tilde{m}_i$ and $\tilde{m}'_i$. If $\tilde{m}_i = \tilde{m}'_i$, E outputs $\tilde{m} = \tilde{m}_i$ and halts. We refer to this method of checking consistency as the "collinearity test" as $\tilde{m}_i = \tilde{m}'_i$ iff the three points $\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\}$ are collinear. We prove that with very high probability, this test is sound; *i.e.*, M cannot answer "incorrectly but collinearly", except with probability $\approx 1/q$. The main challenge is showing that M answers correctly on the right with non-negligible probability *even though* E provides random answers in the third round of step 3.

Tags in Error Corrected Form. This discussion is meant for readers who are familiar with the roles of tags in previous non-malleable commitment schemes, for a more thorough introduction see Section 2. Just as in many of the existing NMC schemes, our protocol consists of n "atomic subprotocols", one for each tag. Previous schemes use the so called "DDN trick" [DDN91] in order to turn C's $k-$bit identity into a list of n $(= k)$ tags $t_1, \ldots, t_n$, satisfying the properties: (1) each t_i is of length $\log n + 1$; and (2) if $\{t_i\}_i$ and $\{\tilde{t}_j\}_j$ are the tags resulting from two distinct identities then there exists some i such that t_i is completely distinct from $\{\tilde{t}_j\}_j$, meaning that $t_i \neq \tilde{t}_j$ for all j.

Previous schemes' security proofs require the extractor to be able to use any completely distinct left subprotocol (i.e., one whose tag is completely distinct from $\{\tilde{t}_j\}_j$) to extract M's commitment $\tilde{m}$ with high probability. This ensures that extraction is possible even in the worst case when there is a single such subprotocol. It also introduces a good deal of redundancy into the protocol.

While one would generally expect most pairs of distinct identities to result in pairs of tags such that property (2) holds for many i, all the DDN trick can guarantee in the worst case is that it holds for a single i (since M is allowed to choose his identity adversarily, this worst case situation might very well be realized). If however, one first applies an error correcting code to C's identity obtaining, say, a codeword in $\mathbb{F}^n$ for suitably chosen finite field $\mathbb{F}$ with $|\mathbb{F}| = \text{poly}(n)$, then applying the DDN trick to this codeword would yield tags such that (1) t_i is of length $\mathcal{O}(\log n)$; and (2) t_i is completely distinct from $\{\tilde{t}_j\}_j$ for a constant fraction of the $i \in \{1, \ldots, n\}$.

Our "completely distinct on average" property requires only that extraction is possible from a completely distinct left subprotocol with constant probability, since there now are guaranteed to be many extraction opportunities. This allows us to remove much of the artificial redundancy resulting in a more efficient protocol compared to prior work.

Non-malleability against a copying M**.** To get a sense of why we might expect our scheme to be non-malleable, let us examine the situation against an M who attempts to maul C's commitment by simply copying its messages from the left interaction to the right. Let m be the message committed to on the left and let $\{t_i\}_{i=1}^n$ and $\{\tilde{t}_i\}_{i=1}^n$ be the corresponding tags.

After the first message, M will have copied C's commitments over to the right interaction, successfully committing to the coefficients of the linear polynomials $\tilde{f}_i(x) = r_i x + m$, $i = 1, \ldots, n$. The hiding of Com ensures it does not know the polynomials themselves, and so when it receives the right query vector $\tilde{\boldsymbol{\alpha}}$, its only hope of coming up with the correct valuations $\tilde{f}_i(\tilde{\alpha}_i)$ is to copy R's challenge to the left interaction and copy C's response back. However, it is unlikely that this will be possible. Indeed, M can only copy $\tilde{\alpha}_i$ over to the left when $\tilde{\alpha}_i \in \left[2^{t_i}\right]$. If $\tilde{t}_i > t_i$ then the $i-$th challenge space on the right is at least twice as big as the $i-$th challenge space on the left, which means that the probability $\tilde{\alpha}_i$ can be copied is at most $1/2$. We will use a code which ensures that $\tilde{t}_i > t_i$ for a constant fraction of the i, making the probability that M can copy every coordinate of R's query $2^{-\Omega(n)}$, ruling out the "copying" attack.

Non-malleability against general M**.** Establishing security against a general man-in-the-middle adversary is significantly more challenging, and this is where the bulk of the new ideas are required. Our proof of non-malleability will require us to delve into the full range of possibilities for M's behavior. In each case, we will show that one of three things happen:

1. M does not correctly answer its queries with good enough probability;
2. E succeeds in extracting $\tilde{m}$ with sufficient probability;
3. an M with such behavior can be used to break the hiding of Com.

The core of our result can be seen as a reduction from a PPT M who correctly answers its queries with non-negligible probability and yet causes E to fail, to a machine $\mathcal{A}$ who breaks the hiding of Com. The following is a very high level outline of our proof.

We define USEFUL to be the set of transcripts which do not lead to situation 1 above; that is, transcripts for which M has a good chance of completing the protocol given the prefix. This is important in order for E to have any chance of successfully extracting $\tilde{m}$. Indeed, if M just aborts in every rewind, E will have no chance. From this standpoint, USEFUL is the set of transcripts which give E "something to work with." We prove that most transcripts are in USEFUL in Claim 3.

We then define EXT, the set of "extractable" transcripts, on which E will succeed with high probability. These are the transcripts which lead to situation 2. Intuitively, EXT is the set of transcripts such that M has good probability of correctly answering a query in a rewind despite the fact that E provides random answers to M's queries. We prove that indeed, if a transcript is in EXT then E succeeds in extracting $\tilde{m}$.

Finally, we define TRB, the set of "troublesome" transcripts which are both useful and not extractable. Transcripts in TRB are problematic as on the one hand, usefulness ensures that the prefix is such that if M receives correct responses to its queries on the left, it gives correct responses to the queries on the right. At the same time however, transcripts in TRB are not extractable and so the prefix is also such that if M receives random responses to its queries on the left it answers the right queries incorrectly. Certainly, the hiding of Com ensures that M cannot *know* whether it receives correct or random responses to its queries on the left. So this difference in behavior suggests that we may be able to use M to violate the hiding of Com, leading to situation 3 above.

Our main claim in this part of our proof is Claim 8, which says that if the left challenge $\boldsymbol{\alpha}$ has a superpolynomial number of preimage right challenges $\tilde{\boldsymbol{\alpha}}$ then either E succeeds in extracting $\tilde{m}$, or M can be used to break hiding. We are implicitly viewing M as a mapping from $\tilde{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}$. Such a claim has been at core of the analysis of some previous NMC schemes. In particular, this idea plays a role whenever the "two-slot trick" is used, since typically this *ensures* that some slot has a right challenge space that is much bigger than the left. In our case, we have some work still left as there is only a single slot and the right and left challenge spaces have the same size. Nevertheless, we are able to prove, using a series of combinatorial arguments, that any mauling attack will wind up with M's left query having exponentially many preimage right queries.

To see these techniques in action, define the set $S = \{i \in [n] : \tilde{t}_i \leq t_i\}$, and consider an M who simply copies the right challenges $\tilde{\alpha}_i$ for $i \in S$ over to the left but who makes sure to produce a legal query in the coordinates not in S on the left. As $\left[2^{\tilde{t}_i}\right] \subset \left[2^{t_i}\right]$ for all $i \in S$, copying $\tilde{\alpha}_i$ when $i \in S$ is fine. If we think of M as a map sending right challenge $\tilde{\boldsymbol{\alpha}}$ to left challenge $\boldsymbol{\alpha}$, then for any $\tilde{\boldsymbol{\alpha}}_S = (\tilde{\alpha}_i)_{i \in S}$, M sends $\tilde{\boldsymbol{\alpha}}'$ such that $\tilde{\boldsymbol{\alpha}}'_S = \tilde{\boldsymbol{\alpha}}_S$ to $\boldsymbol{\alpha}'$ such that $\boldsymbol{\alpha}'_S = \tilde{\boldsymbol{\alpha}}_S$. In other words, M maps the set of right query vectors whose $S-$coordinates are fixed to $\tilde{\boldsymbol{\alpha}}_S$ to the set of left query vectors whose $S-$coordinates are also fixed to $\tilde{\boldsymbol{\alpha}}_S$. However, the sizes of these subsets of right and left challenges are

$$\prod_{i \notin S} 2^{\tilde{t}_i} \text{ and } \prod_{i \notin S} 2^{t_i},$$

respectively, and $\prod_{i \notin S} 2^{\tilde{t}_i} = 2^{\Omega(n)} \prod_{i \notin S} 2^{t_i}$ (we are using that our tags are in error-corrected form, which ensures $\left|[n] \setminus S\right| = \Omega(n)$). So we see that M, when restricted to the right challenges with $S-$coordinates fixed to $\tilde{\boldsymbol{\alpha}}_S$, is exponentially many to one on average, and so $\boldsymbol{\alpha}$ has exponentially many preimages with high probability.

4−Round Non-Malleability. The protocol in Figure 1 is explained sequentially, and as written, consists of 8 rounds: two for Naor's commitment, two for the query/response phase, and four for the ZK argument. However, it can be parallelized down to four rounds using a special four-round ZK argument system and some careful analysis. This requires running the entire ZK argument in parallel with the commit, query and response messages. We use a delayed-input zero-knowlege argument of knowledge which has an additional technical security feature, having to do with an adversary who is able to rewind the challenger in the security game one time. Very recently, Goyal

and Richelson [GR19] gave a three-round, delayed input witness-indistinguishable argument with security against a rewinding adversary. We compile their protocol into a four-round ZK using the Feige-Shamir technique [FS90], and use this in all of our four-round schemes. This subroutine was not used in the original submission, instead an ad hoc solution was presented, which required a difficult statistical analysis. We have chosen to modify the presentation for simplicity.

Using the OWF in a Blackbox Fashion. The protocol described in Figure 1 makes non-blackbox use of the OWF during the ZK part of the protocol. It is often desirable for protocols to make only blackbox use of their building blocks, as the alternative tends to be vastly less efficient. To this end, the work of [GLOV12] replaces the ZK proof in the [Goy11] NMC scheme with an "MPC in the head" computation [IKOS07], resulting in a constant round NMC scheme which makes blackbox use of a OWF. The same transformation works for our protocol as well, and results in a six round scheme. We point out, however, that all the ZK argument in our protocol has to do is prove "knowledge of committed values" and that these values satisfy a linear equation, both of which can be proved very efficiently (*i.e.*, without resorting to costly $\mathcal{NP}-$reductions), assuming DDH (or other widely used hardness assumptions). Therefore, if a statistically binding commitment scheme is available that has an efficient proof of knowledge of committed value, our protocol will be much more efficient than the generic transformation of [GLOV12], which requires C to imagine an entire MPC in his head.

2 Preliminaries

For positive $n \in \mathbb{N}$, let $[n] = \{1, \ldots, n\}$. A function $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ is *negligible* if it tends to 0 faster than any inverse polynomial *i.e.*, for all constants c there exists $n_c \in \mathbb{N}$ such that for every $n > n_c$ it holds that $\varepsilon(n) < n^{-c}$. We use $\mathbf{negl}(\cdot)$ to specify a generic negligible function. We abbreviate "probabilistic polynomial time" with PPT. We assume familiarity with computational indistinguishability and zero-knowledge proofs (and related protocols).

2.1 Commitment schemes

Commitment schemes are protocols which enable a party, known as the committer C, to commit himself to a value while keeping it secret from the (potentially cheating) receiver, R. This property is known as hiding. Additionally, upon receiving the commitment from C, R is ensured that even if C cheated, there is at most one value that C can decommit to during a later, decommitment phase (binding). In this work, we consider commitment schemes that are statistically-binding which means that the hiding property only holds against computationally bounded adversaries.

Definition 1 (**Statistically Binding Commitment Scheme**). *Let* $\langle \mathrm{C}, \mathrm{R} \rangle$ *be an interactive protocol between* C *and* R. *We say that* $\langle \mathrm{C}, \mathrm{R} \rangle$ *is a* statistically binding commitment scheme *if the following properties hold:*

Syntax: $\langle \mathrm{C}, \mathrm{R} \rangle$ has two phases, the *commit phase* and the *decommit phase*. The commit phase is interactive, where C uses its input message $m \in \{0,1\}^\lambda$ and uses randomness ω, R uses no input; the resulting transcript is denoted $c = \mathrm{Com}(m; \omega)$. We assume that decommitment consists of a single round where C sends (m, ω) to R; we write $\mathrm{Decom}(c) = (m, \omega)$. After the decommitment phase is complete, R outputs $\mathrm{R}(c, m, \omega) \in \{0, 1\}$ indicating either acceptance or rejection.

Correctness: For all $m \in \{0,1\}$, if C and R do not deviate from the protocol, then R accepts with probability 1.

Binding: For every C^*, there exists a negligible function $\mathbf{negl}(\cdot)$ such that C^* succeeds in the following game with probability at most $\mathbf{negl}(\lambda)$: On security parameter 1^λ: C^* first interacts with R in the commit phase to produce commitment c. Then C^* outputs two decommitments (c, m_0, ω_0) and (c, m_1, ω_1), and succeeds if $m_0 \neq m_1$ and R accepts both decommitments.

Hiding: For every PPT receiver R^* and every two messages m_0, m_1, the view of R^* after participating in the commitment phase, where C committed to m_0 is indistinguishable from its view after participating in a commitment to m_1.

[Nao91] gives a 2-round, statistically binding bit commitment scheme that can be built from any OWF [HILL99].

2.2 Non-malleable commitments

We wish for our commitment scheme to be impervious to a MIM adversary, M, who takes part in two protocol executions (in the left interaction M acts as the receiver while in the right, M plays the role of the committer), and tries to use the left interaction to affect the right. The security property we desire can be summarized:

> *For any MIM adversary* M*, there exists a standalone machine who plays only one execution as the committer, yet whose commitment is indistinguishable from* M*'s commitment on the right.*

At first glance, non-malleability seems impossible as surely nothing can be done to protect against a MIM who simply copies messages from one protocol execution to another. For this reason, non-malleable security offers protection only against any MIM who tries to change messages in a meaningful way. In this work, just as in [DDN91, PR05], we assume that the committer has an identity $id \in \{0,1\}^\lambda$. In order to perform a successful mauling attack, a MIM has to maul a commitment corresponding to C's identity into a commitment of his own, distinct identity. We allow M to choose its identity, subject to the condition that it is not equal to C's

In this work, we consider the notion of non-malleability with respect to commitment and we will frequently refer to the "message committed to by a MIM adversary M during the commitment phase". It is uniquely defined with high probability as all commitment schemes in this work are statistically binding. The definition below is essentially the same as the one in [LPV08].

The man-in-the-middle execution. In the man-in-the-middle execution, the MIM adversary M is simultaneously participating in two interactions called the left and the right interaction. In the left interaction M is the receiver and interacts with a honest committer whereas in the right interaction M is the committer and interacts with a honest receiver. We define a random variable $\mathbf{MIM}_{\langle \mathrm{C,R} \rangle}(id, m, z)$, indexed by $id \in \{0,1\}^\lambda$, $m \in \{0,1\}^\lambda$ and auxiliary information z; the variable outputs $(\tilde{id}, \tilde{m}, v)$: the identity and value M commits to in the right interaction, and M's view in the full experiment. M attempts to commit to a value $\tilde{m}$ that is related to m using an identity $\tilde{id}$ of its choice. If the right commitment (as determined by the transcript) is invalid or undefined, or $id = \tilde{id}$, $\tilde{m}$ is set to $\perp$.

The simulated execution. In the simulated execution a simulator $\mathcal{S}$ receives auxiliary information z and interacts with the honest receiver R. Let $\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C},\mathrm{R}\rangle}(id, z)$ denote the random variable describing $(\tilde{id}, \tilde{m}, v)$: the value $\mathcal{S}$ commits to in the right interaction, and $\mathcal{S}$'s view during the entire experiment. If the commitment produced by $\mathcal{S}$ is invalid or undefined, or if $id = \tilde{id}$, then $\tilde{m}$ is set to $\perp$.

Definition 2 (**Non-Malleable Commitments**). *A commitment scheme* $\langle \mathrm{C}, \mathrm{R}\rangle$ *is* non-malleable with respect to commitment *if for every PPT MIM adversary* M*, there exists a PPT simulator* $\mathcal{S}$ *such that the following ensembles are indistinguishable for all* $id, m \in \{0,1\}^\lambda$*:*

$$\{\mathbf{MIM}_{\langle \mathrm{C},\mathrm{R}\rangle}(id, m, z)\}_{z\in\{0,1\}^\star}, \text{ and } \{\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C},R\rangle}(id, z)\}_{z\in\{0,1\}^\star}$$

2.3 Zero Knowledge

Definition 3 (**Interactive Arguments of Knowledge**). *We say that a protocol* $\langle \mathrm{P}, \mathrm{V}\rangle$ *between a prover* P *and verifier* V *is an* interactive argument of knowledge *for a language* L *if it satisfies the following properties:*

- **Syntax.** *Both parties get a statement* $x \in \{0,1\}^\lambda$ *as input,* P *additionally gets a witness* w *such that* $(x, w) \in R_{\mathsf{L}}$*. The players then engage in an interactive protocol producing the transcript* τ *using respective randomness* r_{P} *and* r_{V}*.*[2] *After the protocol is complete,* V *verifies the transcript and outputs the bit* $\mathrm{V}(x, \tau; r_{\mathrm{V}})$ *indicating whether or not it accepts* P*'s proof* τ*.*
- **(Perfect) Completeness.** *For every any* $(x, w) \in R_{\mathsf{L}}$*,* $\Pr[\mathrm{V}(x, \tau) = 1] = 1$*, where the probability is over* r_{P} *and* r_{V}*, and where it is assumed that* τ *is produced by* P *and* V *using inputs* (x, w) *and* x*, and following the protocol specifications.*
- **Argument of Knowledge.** *There exists a knowledge extractor* Ext *which satisfies the following syntax, running time, and extraction guarantees.*
  - **Syntax:** Ext *takes as input a statement/transcript/randomness tuple* $(x, \tau, r_{\mathrm{V}})$ *and, using oracle access to a cheating* P^**, outputs a witness* w*.*
  - **Running Time:** Ext *runs in expected time* $\mathsf{poly}(\lambda, \mathsf{T}_{\mathrm{P}^*})$*, where* $\mathsf{T}_{\mathrm{P}^*}$ *denotes the expected running time of* P^**.*
  - **Extraction:** *Consider the experiment* E *which is parametrized by* x *and uses oracle access to a cheating PPT* P^**, and works as follows: 1) plays as honest* V *with input* x *against* P^**, obtaining* (τ, r_{V})*; 2) extracts* $w \leftarrow \mathsf{Ext}^{\mathrm{P}^*}(x, \tau, r_{\mathrm{V}})$*. Then for all* P^* *and* x*,*

$$\Pr_{\mathsf{E}^{\mathrm{P}^*}(x)}\big[\mathrm{V}(x, \tau; r_{\mathrm{V}}) = 1 \ \&\ (x, w) \notin R_{\mathsf{L}}\big] = 2^{-\Omega(\lambda)}.$$

Definition 4 (**Zero-Knowledge Arguments of Knowledge**). *We say that an interactive argument of knowledge for a language* L*,* $\langle \mathrm{P}, \mathrm{V}\rangle$*, is* zero-knowledge *if there exists a PPT simulator* SIM *satisfying the following syntax and security guarantees:*

- **Syntax:** SIM *takes* $x \in \{0,1\}^\lambda$ *as input, gets oracle access to a possibly cheating* V^**, and outputs a transcript* τ*.*

[2] Formally, P and V interact over k rounds with one player (say V) sending $\tau_1 = \mathrm{V}(x; r_{\mathrm{V}})$, then P sending $\tau_2 = \mathrm{P}(x, w, \tau_1; r_{\mathrm{P}})$, and so on. The final transcript is $\tau = (\tau_1, \ldots, \tau_k)$.

· **Security:** *For any PPT* V^* *and* $x \in \{0,1\}^\lambda$ *and witness* w *such that* $(x,w) \in R_{\mathsf{L}}$,

$$\mathsf{REAL}^{\mathrm{V}^*}(x,w) \approx_{\mathsf{c}} \mathsf{SIM}^{\mathrm{V}^*}(x),$$

where the distributions are as follows:

· $\mathsf{REAL}^{\mathrm{V}^*}(x,w)$**:** V^* *and* P *execute* $\langle \mathrm{P}, \mathrm{V}\rangle$ *honestly on instance* x *with* P *using witness* w*. The resulting transcript* τ *is output.*

· $\mathsf{SIM}^{\mathrm{V}^*}(x)$**:** *Simulated transcript* $\tau \leftarrow \mathsf{SIM}^{\mathrm{V}^*}(x)$ *is drawn and output.*

It is known [FS90] how to construct a four-round zero-knowledge argument of knowledge based on any one-way function.

2.4 Non-Malleable Zero Knowledge

The man-in-the-middle execution. Non-malleable zero-knowledge considers a MIM adversary M who plays two instatiations of $\langle \mathrm{P}, \mathrm{V}\rangle$: one on the left where M plays as a receiver with an honest prover P who proves $x \in \mathsf{L}$; one on the right where he plays as the prover proving the statement $\tilde{x} \in \mathsf{L}$ to an honest V. As in the non-malleable commitment definition, we assume that P and M have distinct identities $id, \tilde{id} \in \{0,1\}^\lambda$. We denote by $\mathbf{MIM}^{\mathrm{M}}_{\langle \mathrm{P},\mathrm{V}\rangle}(id,x,w,z)$, the random variable indexed by the identity/statement/witness/auxiliary-information tuple which outputs the triple $(\tilde{id}, \tilde{x}, \tau, \tilde{\tau})$ consisting of M's identity, the statement M tries to prove to V and the protocol transcripts $(\tau, \tilde{\tau})$ in the left and right executions.

Definition 5 (**Non-Malleable Zero-Knowledge Arguments of Knowledge).** *We say that a zero-knowledge argument of knowledge,* $\langle \mathrm{P}, \mathrm{V}\rangle$ *is* non-malleable *if there exists a simulation-extractor* $\mathsf{SIM.Ext}$ *which satisfies the following syntax, running time, security and extraction guarantees:*

· **Syntax:** $\mathsf{SIM.Ext}$ *takes* $id, x \in \{0,1\}^\lambda$ *and auxiliary information* $z \in \{0,1\}^*$ *as input, gets oracle access to a MIM* M^**, and outputs a tuple* $(\tilde{id}, \tilde{x}, \tilde{w}, \tau, \tilde{\tau}, \tilde{r}_{\mathrm{V}})$*, where* $\tilde{id} \neq id$*.*

· **Running Time:** $\mathsf{SIM.Ext}$ *runs in expected time* $\mathsf{poly}(\lambda, \mathsf{T}_{\mathrm{M}^*})$*, where* $\mathsf{T}_{\mathrm{M}^*}$ *denotes the expected running time of* M^**.*

· **Security:** *For any PPT* M *and* $id, x \in \{0,1\}^\lambda$*, witness* w *such that* $(x,w) \in R_{\mathsf{L}}$ *and auxiliary information* $z \in \{0,1\}^*$*,*

$$\left\{(\tilde{id}, \tilde{x}, \tau, \tilde{\tau})\right\}_{\mathbf{MIM}^{\mathrm{M}}_{\langle \mathrm{P},\mathrm{V}\rangle}(id,x,w,z)} \approx_{\mathsf{c}} \left\{(\tilde{id}, \tilde{x}, \tau, \tilde{\tau})\right\}_{\mathsf{SIM.Ext}^{\mathrm{M}}(id,x,z)}.$$

· **Extraction:** *For any PPT* M*,* $id, x \in \{0,1\}^\lambda$ *and auxiliary information* $z \in \{0,1\}^*$*,*

$$\Pr_{\mathsf{SIM.Ext}^{\mathrm{M}}(id,x,z)}\Big[\mathrm{V}(\tilde{x}, \tilde{\tau}; \tilde{r}_{\mathrm{V}}) = 1 \ \&\ (\tilde{x}, \tilde{w}) \notin R_{\mathsf{L}}\Big] = \mathsf{negl}(\lambda).$$

2.5 Tags in Error Corrected Form

In this section, we describe how to derive the tags from C's identity, highlighting the properties we will use moving forward. Let $id \in \{0,1\}^k$ be C's identity and let $\mathbf{y} \in \mathbb{F}^{n/2}$ be the image of id under an error correcting code with constant distance, for a suitable finite field $\mathbb{F}$. Constant distance implies that if $id, \tilde{id} \in \{0,1\}^k$ are distinct identities then $\mathbf{y}$ and $\tilde{\mathbf{y}}$ differ on a constant fraction of their coordinates. Now, set

$$t_i = \begin{cases} 2i|\mathbb{F}| + y_i, & i \leq n/2 \\ (2n+3)|\mathbb{F}| - t_{n-i+1}, & i > n/2 \end{cases}$$

Note that $2i|\mathbb{F}| \leq t_i < (2i+1)|\mathbb{F}|$ for all i. The following is a list of useful properties that the tags satisfy. Let $\{t_i\}_i$ and $\{\tilde{t}_i\}_i$ be the tags resulting from distinct identities $id \neq \tilde{id}$.

1. **Ordered:** $t_1 < t_2 < \cdots < t_n$;
2. **Well Spaced:** $t_1 = \omega(\log \lambda)$ and $t_{i+1} - t_i = \omega(\log \lambda)$ for all $i \in [n]$; moreover $t_{i+1} - \tilde{t}_i = \omega(\log \lambda)$.
3. **Good Distance and Balance:** if $i \neq j$ then $t_i \neq \tilde{t}_j$; moreover $t_i < \tilde{t}_i$ holds for a constant fraction of $i \in [n]$ (as does $t_i > \tilde{t}_i$).

Properties 1 and 2 follow immediately as long as $|\mathbb{F}| = \omega(\log \lambda)$. Property 3 follows from 1) the distance of the error correcting code as $t_i = \tilde{t}_i$ iff $y_i = \tilde{y}_i$ which must not be the case for a constant fraction of the $i \in [n]$; along with 2) if $t_i \neq \tilde{t}_i$ then either $t_i < \tilde{t}_i$ or else $t_{n-i} < \tilde{t}_{n-i}$. This is reminiscent of the two slot trick of [Pas04, PR05].

It remains to select parameters. Note that we have already touched on the role that the tags play in our protocol: the size of the challenge space in coordinate i is 2^{t_i}. This means that we would like to make the tags as small as possible, while still allowing our security proof to go through. We make the conservative selection $n = \mathcal{O}(\lambda)$ and $|\mathbb{F}| = \log^2(\lambda)$ to ensure both that the above properties hold and that all that is required of the error correcting code is that it has constant distance and constant rate. Codes with such properties are known to exist. We could use, for example polynomial based codes such as Reed-Muller codes, the multivariate generalization of Reed-Solomon codes. This results in the overall communication complexity of our non-malleable commitment scheme being $\tilde{\mathcal{O}}(\lambda^2)$. Slightly better communication complexity might be available through more agressive choices of parameters or better codes. We do not press the issue further.

3 The Protocol

In this section, we describe our 8-round protocol which is non-malleable against a synchronizing MIM. Our optimized protocols with full security appear in Section 6. The protocol of this section is given tags $t_1, \ldots, t_n$ in error corrected form as described in Section 2.5. We use Naor's two round, statistically binding bit commitment scheme [Nao91] as a building block.[3] We use boldface to denote

[3]Briefly recall Naor's scheme: 1) R sends random initialization message σ, and 2) C responds with $\mathrm{Com}_\sigma(m; s)$, a commitment to $m \in \{0,1\}$ using randomness s (we will feel free to just write $\mathrm{Com}(m)$, surpressing σ and s for simplicity). We comment that the same initialization message σ can be used for polynomially many parallel instantiations of the scheme, allowing C to commit to $m \in \mathbb{Z}_q$ one bit at a time (actually [Nao91] shows how to commit to longer messages more efficiently).

vectors; in particular a challenge vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and a response vector $\mathbf{a} = (a_1, \ldots, a_n)$. We write **Com** for the entire first commitment message, so $\mathbf{Com} = \big(\mathrm{Com}(m), \mathrm{Com}(r_1), \ldots, \mathrm{Com}(r_n)\big)$. Our non-malleable commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle$ between a committer C trying to commit to m and a receiver R appears in Figure 2. The decommitment phase is done by having the committer C send m and the randomness it used during the protocol.

Public Parameters: Let λ be the security parameter, $n = \mathsf{poly}(\lambda) \in \mathbb{N}$, an integer parameter, and tags $t_1, \ldots, t_n$ and a large prime q such that $q > 2^{t_i}$ for all i.

Commiter's Private Input: Message $m \in \mathbb{F}_q$ to be committed to.

Commit Phase:

0. $\mathrm{R} \to \mathrm{C}$ **Initialization message:** Send the first message σ of the Naor commitment scheme.
1. $\mathrm{C} \to \mathrm{R}$ **Commit message:** Sample random $r_1, \ldots, r_n \in \mathbb{F}_q$ and $s, s_1, \ldots, s_n \leftarrow \$$.
   - Define linear functions $f_1, \ldots, f_n$ by $f_i(x) = r_i x + m$ for each $i \in [n]$.
   - Send commitments $\mathbf{Com} = \big(\mathrm{Com}_\sigma(m; s), \mathrm{Com}_\sigma(r_1; s_1), \ldots, \mathrm{Com}_\sigma(r_n; s_n)\big)$.
2. $\mathrm{R} \to \mathrm{C}$ **Query:**
   - Send random challenge vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \in [2^{t_i}] \subset \mathbb{F}_q$.
3. $\mathrm{C} \to \mathrm{R}$ **Response:**
   - Send evaluation vector $\mathbf{a} = (a_1, \ldots, a_n)$, $a_i = f_i(\alpha_i)$.
4. $\mathrm{C} \longleftrightarrow \mathrm{R}$ **Consistency proof:** Parties engage in a zero-knowledge argument protocol where C proves to R that $\exists \big((m, s), (r_1, s_1), \ldots, (r_n, s_n)\big)$ such that:
   - $\mathbf{Com} = \big(\mathrm{Com}_\sigma(m; s), \mathrm{Com}_\sigma(r_1; s_1), \ldots, \mathrm{Com}_\sigma(r_n; s_n)\big)$; and
   - $a_i = r_i \alpha_i + m \;\forall\; i = 1, \ldots, n$.

Decommit Phase:

$\mathrm{C} \to \mathrm{R}$ **Decommit Message:** Send $\big((m, s), (r_1, s_1), \ldots, (r_n, s_n)\big)$.

Verification: R checks the correctness of the commit and response messages and verifies the consistency proof. If all checks pass, R outputs 1, otherwise 0.

Figure 2: The non-malleable commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle$.

Proposition 1. *The commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle$ is computationally hiding and statistically binding.*

Proof Sketch. Statistical binding follows from the statistical binding property of the underlying

commitment scheme Com. To prove computational hiding, we consider the following hybrid experiments.

1. Simulate the ZK consistency proof step. Indistinguishability follows from the ZK property.
2. For each $i \in [n]$, replace the commitment $\mathrm{Com}(r_i)$ to be a commitment to random value. Indistinguishability follows from the hiding of Com. Note that after this change, the responses in Step 3 are now random values which are unrelated to the committed values in Step 1.
3. Change the commitment $\mathrm{Com}(m)$ to be a commitment to a random string (as opposed to a commitment to m). Indistinguishability follows from the hiding of Com.

□

Theorem 1 (Main theorem)**.** *The commitment scheme* $\langle \mathrm{C}, \mathrm{R} \rangle$ *is non-malleable against a synchronizing adversary.*

We comment that the protocol in Figure 2 is actually already non-malleability against a general *non-synchronizing* adversary, (provided we choose a ZK with suitable properties, such as [FS90]). However, we only prove non-malleability against a synchronizing MIM (*i.e.*, one who plays corresponding messages of the two instantiations one after the other) because the large number of messages of the protocol above make it cumbersome to examine all possibilities for M's scheduling. We defer the proof of non-malleability against non-synchronizing M until after we parallelize our protocol down to four rounds (see Section 6).

4 Proof of Non-Malleability

In this section we prove Theorem 1. Recall from Definition 2 that we must show that for any PPT MIM M there exists a PPT simulator $\mathcal{S}$ such that for all $id \in \{0,1\}^\lambda$ and $m \in \mathbb{F}_q$,

$$\left\{\mathbf{MIM}_{\langle \mathrm{C},\mathrm{R}\rangle}(id, m, z)\right\}_z \approx_c \left\{\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C},\mathrm{R}\rangle}(id, z)\right\}_z,$$

where the distributions output $(\tilde{m}, v)$: the commitment in the right interaction and view after the commit phases of both executions are complete in the real and ideal worlds, respectively. Our simulator is a very simple machine who runs M internally, committing honestly to $0 \in \mathbb{Z}_q$ on the left and forwarding M's messages on the right to an honest receiver R.

We prove indistinguishability of the above distributions for any M by constructing an extractor E which takes M's view after the commit phases of the left and right executions are complete and outputs its commitment $\tilde{m}$ in the right execution whp. It follows that an algorithm which distinguishes $\mathcal{D}_0 = \left\{\mathbf{MIM}_{\langle \mathrm{C},\mathrm{R}\rangle}(id, m, z)\right\}_z$ from $\mathcal{D}_1 = \left\{\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C},\mathrm{R}\rangle}(id, z)\right\}_z$ can be used to break the hiding of $\langle \mathrm{C}, \mathrm{R}\rangle$ in the following way: 1) let v be M's view after completing the commit phases of the left and right executions in either the real or ideal world; 2) use E to obtain the pair $(\tilde{m}, v)$; 3) use the distinguisher to determine whether M's interaction took place in the real or ideal world. This breaks the hiding of the left commitment as the only difference between the worlds is that in the real, C commits to m while in the ideal, $\mathcal{S}$ commits to 0.

Formally, we assume that there exists a PPT distinguisher D such that

$$\left|\Pr_{(\tilde{m},v)\leftarrow\mathcal{D}_0}\big(\mathrm{D}(\tilde{m}, v) = 1\big) - \Pr_{(\tilde{m},v)\leftarrow\mathcal{D}_1}\big(\mathrm{D}(\tilde{m}, v) = 1\big)\right| \geq 2p$$

for some non-negligible $p = p(\lambda)$. We prove that E succeeds with probability at least $1 - p$. Note this suffices for proving non-malleability since it means that E extracts $\tilde{m}$ AND the D will use $(\tilde{m}, v)$ to determine whether M is interacting with C committing to m, or $\mathcal{S}$ committing to 0. We also assume without loss of generality that M is deterministic and that M's probability of successfully completing the protocol (over C's and R's random coins) is at least p.

4.1 The Extractor E

The high level description of our extractor (described formally in Figure 3) is quite simple. Intuitively, our protocol begins by C committing to n, threshold 2, Shamir secret sharings [Sha79] of m; R then asks for one random share from each sharing, which C gives. All E does is rewind M to the beginning of the right session's query phase ask for a new random share. Since E gets one share as part of its input, this will allow E to reconstruct $\tilde{m}$.

The problem with this approach is that E does not know the value C has committed to on the left and so it does not know how to answer M's query on the left correctly. The best E can do is give a random response on the left and hope that M will give a correct response on the right anyway. On the one hand, the hiding of Com dictates that M cannot distinguish a correct response from a random one. On the other hand, M doesn't actually need to know whether the response on the left is correct or not in order to perform a successful mauling attack. Imagine, for example, the MIM who mauls R's challenge to the left execution and mauls C's response back. Such an M will prevent E from extracting $\tilde{m}$ because M only correctly answers E's query if given a correct response to its own left query, which E cannot give. Of course we will prove that no M with such behavior can exist, but this proof is highly non-trivial.

Another question which our extractor raises is "how can E tell a correct response from an incorrect one?" As we have described it, the hiding of Com ensures that it cannot. However, a small modification to the E described above fixes this. Instead of asking for one new share, E rewinds twice to the beginning of the right query phase and asks for two different new shares. The key observation is that if M answers both queries correctly then the three shares it holds (the two it received plus the one it got as input) are collinear, whereas if M answers at least one incorrectly they are overwhelmingly likely to NOT be collinear.

E is given as input a transcript of a complete commit phase in both the left and right interactions. We denote the transcript with the letter $\mathbb{T}$. Specifically,

$$\mathbb{T} = \big(\mathbf{Com}, \tilde{\mathbf{Com}}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}, \tilde{\mathbf{a}}, \pi, \tilde{\pi}\big).$$

We assume WLOG that M is deterministic (and so $\tilde{\mathbf{Com}}$, $\boldsymbol{\alpha}$, $\tilde{\mathbf{a}}$ and $\tilde{\pi}$ are uniquely determined by $\mathbf{Com}, \tilde{\boldsymbol{\alpha}}$, $\mathbf{a}$, and π) we will often just write $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}, \pi\big)$.

Definition 6 (**Accepting Transcript**). *We say that* $\mathbb{T} \in \mathsf{ACC}$ *if both* π *and* $\tilde{\pi}$ *are accepting proofs.*

E is not interested in the proofs $(\pi, \tilde{\pi})$ (other than whether $\tilde{\pi}$ is accepting; if not extraction is trivial, E outputs $\perp$), so we simplify notations further, writing $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big)$.

The soundness of the ZK ensures that if $\mathbb{T} \in \mathsf{ACC}$ then query vectors $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}$ are answered correctly. We say that M aborts if M behaves in such a way as to make $\mathbb{T} \notin \mathsf{ACC}$. Note this includes the case when M acts in an obviously corrupt fashion, causing C or R to abort the protocol early.

The extractor E gets $\mathbb{T} \in \mathsf{ACC}$ as input so the probabilities which arise in our analysis often are conditioned on the event $\mathbb{T} \in \mathsf{ACC}$. We denote this with the convenient shorthand $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\cdots)$ instead of $\Pr_{\mathbb{T}}(\cdots \mid \mathbb{T} \in \mathsf{ACC})$. For fixed **Com**, M can be thought of as a deterministic map, mapping right query vectors to left ones. We write $\boldsymbol{\alpha} = \mathrm{M}(\tilde{\boldsymbol{\alpha}})$ to be consistent with this point of view. We assume that the transcript E gets as input is consistent with exactly one right commitment $\tilde{m}$. As $\langle \mathrm{C}, \mathrm{R} \rangle$ is statistically binding, this happens with overwhelming probability.

See Figure 3 below for a formal description of the extractor. Note that there are two ways for E to fail to output $\tilde{m}$. The first is if E fails to extract any value and outputs **FAIL**. The other is if E accidentally extracts an incorrect value $\tilde{m}' \neq \tilde{m}$.

Tags: Let $\{t_i\}_i$ and $\{\tilde{t}_i\}_i$ be the left and right tags, respectively, in error corrected form.

Input: $\mathbb{T} = (\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}) \in \mathsf{ACC}$, and a large value $N = \mathrm{poly}(\lambda)$. E is given oracle access to M.

Extraction procedure: For $j \in [N]$:

1. Rewind M to the beginning of step 2 of the protocol:
   - generate a random right challenge vector $\tilde{\boldsymbol{\beta}}_j = (\tilde{\beta}_{1,j}, \ldots, \tilde{\beta}_{n,j})$, where $\tilde{\beta}_{i,j} \in [2^{\tilde{t}_i}]$.
   - Feed M with $\tilde{\boldsymbol{\beta}}_j$ and receive challenge $\boldsymbol{\beta}_j = (\beta_{1,j}, \ldots, \beta_{n,j})$; if M aborts, continue.
2. Feed $\mathbf{b}_j = (b_{1,j}, \ldots, b_{n,j})$ to M where $b_{i,j} = \begin{cases} a_i, & \beta_{i,j} = \alpha_i \\ r \xleftarrow{\mathrm{R}} \mathbb{Z}_q, & \beta_{i,j} \neq \alpha_i \end{cases}$. Get $\tilde{\mathbf{b}}_j = (\tilde{b}_{1,j}, \ldots, \tilde{b}_{n,j})$; if M aborts, continue
3. For each $i \in [n]$ use $\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j})\}$ to interpolate a line and recover candidate $\tilde{m}_{i,j}$.
4. Repeat steps 1-3. Let $\tilde{\boldsymbol{\gamma}}_j = (\tilde{\gamma}_{1,j}, \ldots, \tilde{\gamma}_{n,j})$ be new right challenge vector and $\tilde{\mathbf{c}}_j = (\tilde{c}_{1,j}, \ldots, \tilde{c}_{n,j})$ be corresponding response. Let $(\tilde{m}'_{1,j}, \ldots, \tilde{m}'_{n,j})$ be recovered candidates.
5. If for some $i \in [n]$, $\tilde{m}_{i,j} = \tilde{m}'_{i,j}$, output $\tilde{m}_{i,j}$ and halt.

Output: Output **FAIL**.

Figure 3: The Extractor E.

Theorem 2 (Sufficient for Theorem 1). *Let* E *be the extractor described in Figure 3, and let* $\mathbb{T}$ *be the transcript it is given as input. Let* $\tilde{m}$ *be* M*'s commitment in the right interaction of* $\mathbb{T}$*. Then*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathrm{E}(\mathbb{T}) \neq \tilde{m}) \leq p,$$

where the probability is over $\mathbb{T} \in \mathsf{ACC}$ *and the randomness of* E *(when we wish to be explicit about this randomness, we denote it* r_{E}*).*

4.2 Extractable, Useful and Troublesome Transcripts

We now begin to chip away at Theorem 2 by examining special classes of transcripts on which a mauling attack will fail. This allows us to gather properties which the remaining pertinent transcripts must satisfy which will aid our future analysis. In this section we focus on the commitment message of the protocol.

Recall the two ways E can fail: by outputting **FAIL** or by outputting incorrect $\tilde{m}' \neq \tilde{m}$. Note that the second way requires M to answer a pair of queries incorrectly but in such a way so that they yield the same candidate message and they pass the collinearity test. In this case we say that M answers *incorrectly but collinearly.*

Definition 7 (**Incorrect but Collinear**). *Fix a main thread transcript* $\mathbb{T} = (\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{a})$ *and an* $i \in \{1, \ldots, n\}$. *Let* $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{b}})$ *and* $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{c}})$ *denote two query/response pairs arising during the execution of* E *while rewinding* M, *using randomness* r_{E}. *Suppose that interpolating* $(\tilde{\beta}_i, \tilde{b}_i)$ *and* $(\tilde{\gamma}_i, \tilde{c}_i)$ *against the main thread's point* $(\tilde{\alpha}_i, \tilde{a}_i)$ *produces the same candidate message* $\tilde{m}'$. *We say that* M *answers* $(\tilde{\beta}_i, \tilde{\gamma}_i)$ incorrectly but collinearly *if:*

1. $\tilde{m}' \neq \tilde{m}$; *and*
2. $\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\}$ *are collinear.*

We define the set

$$\mathsf{IBC}^i(\mathbf{Com}, \tilde{\alpha}_i; r_{\mathrm{E}}) = \big\{(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) : \mathrm{M} \textit{ answers } (\tilde{\beta}_i, \tilde{\gamma}_i) \textit{ incorrectly but collinearly during } \mathsf{EXT}^{\mathrm{M}}(\mathbb{T}; r_{\mathrm{E}})\big\}.$$

Finally, define

$$\mathsf{IBC}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}; r_{\mathrm{E}}) = \big\{\mathbb{T} \in \mathsf{ACC} : (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \in \mathsf{IBC}^i(\mathbf{Com}, \tilde{\alpha}_i; r_{\mathrm{E}}) \textit{ for some } i\big\}.$$

The set $\mathsf{IBC}^i(\mathbf{Com}, \tilde{\alpha}_i; r_{\mathrm{E}})$ is defined in terms of $\mathbb{T}$ and the randomness, r_{E}, that E uses to compute the third round responses. Intuitively, it consists of the second round queries which will lead to E outputting the wrong $\tilde{m}$. The following claim shows that the "incorrect but collinear" event which causes E to output the wrong $\tilde{m}$ occurs very rarely.

Claim 1. *For any* $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$, $\Pr_{\mathbb{T} \in \mathsf{ACC}, r_{\mathrm{E}}}\big(\mathbb{T} \in \mathsf{IBC}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}; r_{\mathrm{E}})\big) = \mathbf{negl}(\lambda)$.

Proof. Fix $i \in \{1, \ldots, n\}$ and let $\mathbb{T}, \mathbb{T}' \in \mathsf{ACC}$ be main threads with the same prefix **Com** but different $i-$th right queries $\tilde{\alpha}_i$ and $\tilde{\alpha}'_i$. Moreover, fix E's randomness arbitrarily making it deterministic, so that the sets $\mathsf{IBC}^i(\tilde{\alpha}_i)$ and $\mathsf{IBC}^i(\tilde{\alpha}'_i)$ are defined. Note that $\mathsf{IBC}^i(\tilde{\alpha}_i)$ and $\mathsf{IBC}^i(\tilde{\alpha}'_i)$ are disjoint. Indeed, suppose $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \in \mathsf{IBC}^i(\tilde{\alpha}_i) \cap \mathsf{IBC}^i(\tilde{\alpha}'_i)$. Then the four points

$$\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\alpha}'_i, \tilde{a}'_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\big\}$$

are collinear. This means that the line they all lie on is correct because $(\tilde{\alpha}_i, \tilde{a}_i)$ and $(\tilde{\alpha}'_i, \tilde{a}'_i)$ are correct ($\mathbb{T}, \mathbb{T}' \in \mathsf{ACC}$) and so $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \notin \mathsf{IBC}^i(\tilde{\alpha}_i) \cup \mathsf{IBC}^i(\tilde{\alpha}'_i)$ as M answered $\tilde{\beta}_i$ and $\tilde{\gamma}_i$ correctly. Therefore, for a fixed prefix **Com** and extractor queries $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$, there is at most one value of $\tilde{\alpha}_i$ such that $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \in \mathsf{IBC}^i(\tilde{\alpha}_i)$. As the set of possible $\tilde{\alpha}_i$ is superpolynomial, the chances that R's query $\tilde{\boldsymbol{\alpha}}$ in $\mathbb{T}$ is such that $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \in \bigcup_i \mathsf{IBC}^i(\tilde{\alpha}_i)$ for any extractor query $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$ is negligible. The result follows. □

As our extractor only asks polynomially many pairs of new queries $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$, we see that E outputs the wrong message $\tilde{m}' \neq \tilde{m}$ with negligible probability. This means that if E fails, it does so because it does not receive correct answers to its queries. We define EXT, the set of "extractable" transcripts, on which M has a non-negligible chance of answering a query correctly even given that its queries are answered by E.

Definition 8 (**Extractable Transcripts**). *Fix* $\varepsilon^* = \big(\lambda/N\big)^{1/2}$. *We define*

$$\mathsf{EXT}_i = \Big\{(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}) : \Pr_{\tilde{\boldsymbol{\beta}}}\big(\mathrm{M} \textit{ correctly answers } \tilde{\beta}_i \big| \mathbf{Com} \ \&\ \mathrm{M}\textit{'s queries answered by } \mathrm{E}\big) \geq \varepsilon^*\Big\}.$$

Set $\mathsf{EXT} = \big\{\mathbb{T} \in \mathsf{ACC} : (\mathbf{Com}, \tilde{\boldsymbol{\alpha}}) \in \mathsf{EXT}_i \textit{ for some } i\big\}$.

Intuitively, EXT is the set of transcripts such that M has good probability of providing at least one pair of correct answers to a pair of queries asked in a rewind despite the fact that E provides random answers to M's queries. We now prove that if a transcript is in EXT then E succeeds in extracting $\tilde{m}$ whp.

Claim 2. $\Pr_{\mathbb{T}}\big(\mathrm{E}(\mathbb{T}) = \mathbf{FAIL} \big| \mathbb{T} \in \mathsf{EXT}\big) = \mathbf{negl}(\lambda)$, *where the probability is over* $\mathbb{T}$ *and the randomness of* E.

Proof. Let $\mathbf{E}_j$ be the event that there exists an i such that M answers both $i-$th queries correctly in rewind j. Since $\mathbb{T} \in \mathsf{EXT}$ we have that $\Pr(\mathbf{E}_j) \geq (\varepsilon^*)^2 = \lambda/N$ for all j. As the $\mathbf{E}_j$ are independent,

$$\Pr_{\mathbb{T}}\big(\mathrm{E}(\mathbb{T}) = \mathbf{FAIL} \big| \mathbb{T} \in \mathsf{EXT}\big) = \Pr\big(\text{not } \mathbf{E}_j \ \forall\ j \big| \mathbb{T} \in \mathsf{EXT}\big) \leq \left(1 - \frac{\lambda}{N}\right)^N = \mathbf{negl}(\lambda).$$

□

Having looked at transcripts on which E succeeds whp, we next examine a set of transcripts on which E trivially fails. These are transcripts which M was lucky to complete given the commitment phase. Indeed, if every time E rewinds M simply aborts, E will have no chance of extracting $\tilde{m}$.

Definition 9 (**Useful Transcripts**). *Fix non-negligible* $\delta < \frac{1}{3}$ *and (temporarily) define*

$$W = \big\{\mathbf{Com} : \Pr_{\mathbb{T}}\big(\mathbb{T} \in \mathsf{ACC} \big| \mathbf{Com}\big) \leq \delta p^2\big\}.$$

Set $\mathsf{USEFUL} := \big\{\mathbb{T} \in \mathsf{ACC} : \mathbf{Com} \notin W\big\}$.

Informally, W is the set of partial transcripts for which M is unlikely to complete the protocol, so USEFUL is the set of transcripts such that if M is rewound and executed again on a different query, the protocol will complete successfully with good probability. We note that most transcripts are indeed useful.

Claim 3. $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \notin \mathsf{USEFUL}\big) \leq \delta p$.

Proof. We have

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbf{Com} \in W\big) = \Pr_{\mathbb{T}}\big(\mathbf{Com} \in W \big| \mathbb{T} \in \mathsf{ACC}\big) \leq \frac{\Pr_{\mathbb{T}}\big(\mathbb{T} \in \mathsf{ACC} \big| \mathbf{Com} \in W\big)}{\Pr_{\mathbb{T}}(\mathbb{T} \in \mathsf{ACC})} \leq \delta p,$$

using the definition of W and the fact that $\Pr_{\mathbb{T}}(\mathbb{T} \in \mathsf{ACC}) \geq p$. □

Transcripts in EXT are those for which M is likely to correctly answer a right query even given incorrect responses to its own left queries. On the other hand, USEFUL can be thought of as the transcripts for which M answers the right queries correctly if given correct answers to its left queries. This leads us to the following definition.

Definition 10 (**Troublesome Transcripts**). *We define* $\mathsf{TRB} = \mathsf{USEFUL} \setminus \mathsf{EXT}$.

Transcripts in TRB are troublesome as essentially, they are transcripts for which M answers the right queries correctly if given correct answers to its left queries, but incorrectly if given incorrect answers to its left queries. Certainly, the hiding of Com ensures that M cannot *know* whether it receives correct or random responses to its queries on the left. So this difference in behavior suggests that we may be able to use M to break the hiding of Com. However, it is not so easy. Keep in mind, M does not have to know whether it is giving a correct or incorrect answer on the left. Indeed, almost all mauling attacks one could imagine have the property that M answers correctly on the right if and only if it gets correct answers on the left. The following lemma comprises the heart of our analysis.

Lemma 1. *If* Com *is computationally hiding then there exists a constant* $\delta' < \frac{1}{3}$ *such that*

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB}\big) \leq \delta' p.$$

Lemma 1 combined with Claims 1 through 3 give us

$$\begin{aligned}
\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathrm{E}(\mathbb{T}) \neq \tilde{m}\big) &\leq \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \notin \mathsf{USEFUL}\big) + \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB}\big) \\
&+ \Pr_{\mathbb{T}}\big(\mathrm{E}(\mathbb{T}) = \mathbf{FAIL} \big| \mathbb{T} \in \mathsf{EXT}\big) \\
&+ \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{IBC}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \text{ for some } (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) \text{ asked by E}\big) \\
&\leq \delta p + \delta' p + \mathbf{negl}(\lambda) < p,
\end{aligned}$$

proving Theorem 2.

5 Proof of Lemma 1

5.1 Proof Overview

We prove Lemma 1 by defining the notion of "query dependence", and then considering the possible different ways in which M's left queries $\boldsymbol{\alpha}$ can depend on right queries $\tilde{\boldsymbol{\alpha}}$. Intuitively, $\alpha_{i'}$ being dependent on $\tilde{\alpha}_i$ is the result of M performing a mauling attack. Suppose that M mauls $\mathrm{Com}(f_{i'})$ in order to obtain $\mathrm{Com}(\tilde{f}_i)$. Then M does not know $\tilde{f}_i$ and so cannot hope to answer $\tilde{\alpha}_i$ except by mauling C's answer to $\alpha_{i'}$. Therefore, if M is rewound to the beginning of step 2 and asked a different query vector $\tilde{\boldsymbol{\beta}}$ such that $\tilde{\beta}_i = \tilde{\alpha}_i$, M will have to ask $\boldsymbol{\beta}$ such that $\beta_{i'} = \alpha_{i'}$ if it wants to answer successfully. This is the idea of query dependence: if $\tilde{\alpha}_i$ is asked on the right, then $\alpha_{i'}$ must be asked on the left.

Recall that in the introduction we considered a copying MIM who attempts to maul C's commitment by simply copying and pasting messages between the left and right sessions. Such an attack is a very simple example of a mauling attack in which each α_i is dependent on $\tilde{\alpha}_i$. We saw this attack is foiled by the large number of left tags which differ from all right tags, preventing the right query

$\tilde{\boldsymbol{\alpha}}$ from being a legal left query except with negligible probability. In fact, we prove in Claim 7 that all mauling attacks in which each α_i depends on $\tilde{\alpha}_i$ will fail whp.

This encourages us to investigate what else can happen. We arrive at three possibilities.

- UNBAL: There exist $i' > i$ such that $\alpha_{i'}$ depends on $\tilde{\alpha}_i$.
- $1-2$: There exist (i_1, i_2, i') such that $\alpha_{i'}$ depends on both $\tilde{\alpha}_{i_1}$ and $\tilde{\alpha}_{i_2}$.
- IND: There exists i such that each $\alpha_{i'}$ does *not* depend on $\tilde{\alpha}_i$.

In the actual proof we formalize the above possibilities using precise conditional probability statements. We keep it informal here, however, in order to convey as much intuition as possible.

Note that if none of the above three events occur then α_i depends on $\tilde{\alpha}_i$ for all i which is what we hope happens. We complete the proof by showing that each of the three events cannot happen except with very small probability. However, this is easier said than done. Consider, for example, the mauling attack which results in $1-2$. Intuitively, if $\alpha_{i'}$ is dependent on both $\tilde{\alpha}_{i_1}$ and $\tilde{\alpha}_{i_2}$ then M is using C's response $f_{i'}(\alpha_{i'})$ on the left to produce both $\tilde{f}_{i_1}(\tilde{\alpha}_{i_1})$ and $\tilde{f}_{i_2}(\tilde{\alpha}_{i_2})$ on the right. On the one hand it is extremely unlikely that a single polynomial evaluation on the left contains enough information to allow M to correctly give two random evaluations on the right. On the other hand, this intuition alone isn't enough to say that $1-2$ can't occur as the argument is information theoretic in nature. Indeed, any statment one wishes to make about M's behavior in the query phase must have a computational proof as an unbounded M can query however it wants to and then simply break the hiding of the commitments in the first message to learn the $\tilde{f}_i$ and answer correctly.

The key claim which allows us to capitalize on our information theoretic intuition is Claim 8 which states that if the left query $\boldsymbol{\alpha}$ has a superpolynomial number of preimage right queries $\tilde{\boldsymbol{\alpha}}$ then either E succeeds in extracting $\tilde{m}$ or M can be used to break the hiding of Com, the building block commitment scheme. The proof is technical; at this point we give only some intuition which speaks to the truth of Claim 8. Full details can be found in Section 5.3. If there are superpolynomially many $\tilde{\boldsymbol{\alpha}}$ such that $\mathrm{M}(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{\alpha}$, the chances that M can use C's response by itself to answer $\tilde{\boldsymbol{\alpha}}$ are negligible. It follows that either M must be content to not answer most of the $\tilde{\boldsymbol{\alpha}}$ such that $\mathrm{M}(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{\alpha}$ (the probability of which can be bounded using a straightforward conditional probability argument) or M must know some "extra information" about the $\tilde{f}_i$ which allows him to provide a correct response to $\tilde{\boldsymbol{\alpha}}$. But this means that either M will use this extra information to correctly answer $\tilde{\boldsymbol{\alpha}}$ even when given a random answer to $\boldsymbol{\alpha}$ on the left (in which case E succeeds in extracting $\tilde{m}$), or M is choosing to utilize this extra information only when C answers correctly on the left. However, the hiding of the commitment in the first message ensures that M cannot *know* whether he receives correct responses on the left or not, and this difference in behavior will allow us to use M to break hiding.

Armed with Claim 8, we can now make definitive statements about UNBAL and $1-2$. For example, if UNBAL occurs then $\alpha_{i'}$ is dependent on $\tilde{\alpha}_i$ for some $i' > i$, and so if R asks a new right challenge with the same $i-$th query, M will fix $\alpha_{i'}$ on the left. However, as $i' > i$, $\alpha_{i'}$ is drawn from a much larger challenge space than $\tilde{\alpha}_i$, and so M is "wasting challenge space". Specifically, the residual right challenge space with the $i-$th query fixed to $\tilde{\alpha}_i$ is superpolynomially larger than the residual left challenge space with $\alpha_{i'}$ fixed, and so with high probability, we will find ourselves in a situation where the left query has superpolynomially many right query preimages. By Claim 8, this must not happen except with negligible probability. This simple combinatorial argument is essentially the

content of Claim 5. In Section 5.2 we prove Claims 5 through 7 which show that if either UNBAL or $1-2$ or "not (UNBAL or $1-2$ or IND)" occur, then the left query will have superpolynomially many right query preimages. The proofs of Claims 6 and 7 are more involved than that of Claim 5, but they are still purely combinatorial.

Finally, we prove in Claim 9 that IND cannot happen using another reduction to hiding. It uses the same framework as Claim 8 and has similar underlying intuition (again, we differ the technical discussion and formal proof to Section 5.3). Here the main point is that if IND occurs then there exists a right query $\tilde{\alpha}_i$ on which no $\alpha_{i'}$ on the left is dependent. Intuitively this means that M does not need any of the left challenges in order to correctly return $\tilde{f}_i(\tilde{\alpha}_i)$, implying that he knows some information about the polynomial $\tilde{f}_i$. As in the intuition for Claim 8 this means either that extraction is successful, or that M is breaking hiding.

5.2 Analyzing Dependencies

In Section 4.2, we looked at the commitment message of $\langle \mathrm{C}, \mathrm{R} \rangle$, and established that it suffices to consider only transcripts $\mathbb{T} \in \mathsf{TRB}$ in order to prove Theorem 1. We now consider the query message of $\langle \mathrm{C}, \mathrm{R} \rangle$. Let R and L be the sets of right and left query vectors respectively. In this section we will often fix a commitment message $\mathbf{Com}$ (implicitly fixing $\tilde{\mathbf{Com}} = \mathrm{M}(\mathbf{Com})$) in which case M can be thought of as a deterministic function $\mathrm{M} : R \to L$ mapping $\tilde{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}$. In the rest of this section we will frequently consider subsets of R and L. Whenever we do so, we assume that $\mathbf{Com}$ is fixed (even if do not mention it explicitly). This is because we are really interested in how M behaves on these subsets, and M is not defined as a function until $\mathbf{Com}$ is fixed.

Definition 11 (Honest Queries). *For fixed* $\mathbf{Com}$*, we say that a right query vector* $\tilde{\boldsymbol{\alpha}} \in R$ *is* honest *if* M *answers* $\tilde{\boldsymbol{\alpha}}$ *honestly in the right interaction given correct responses to its queries* $\boldsymbol{\alpha} = \mathrm{M}(\tilde{\boldsymbol{\alpha}})$ *in the left interaction. We denote the set of honest right query vectors by* $\mathsf{HON}_{\mathbf{Com}}$*, or just* HON *when* $\mathbf{Com}$ *is clear from context.*

Let $R^i(\tilde{\alpha}_i)$ and $L^{i'}(\alpha_{i'})$ denote the sets of right and left query vectors whose $i-$th and $i'-$th coordinates are fixed on $\tilde{\alpha}_i$ and $\alpha_{i'}$, respectively. We write $\mathrm{M} : R^i_\tau(\tilde{\alpha}_i) \longrightarrow L^{i'}(\alpha_{i'})$ if M maps a $\tau-$fraction of $R^i(\tilde{\alpha}_i)$ to $L^{i'}(\alpha_{i'})$. Similarly, define $\mathsf{HON}^i(\tilde{\alpha}_i) := R^i(\tilde{\alpha}_i) \cap \mathsf{HON}$. Finally, we write $\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}(\cdots)$ as shorthand for $\Pr_{\tilde{\boldsymbol{\alpha}}}(\cdots | \tilde{\boldsymbol{\alpha}} \in \mathsf{HON})$.

Claim 4. *Let* $\mathbf{Com}$ *be the prefix of a transcript* $\mathbb{T} \in \mathsf{USEFUL}$*. Then*

1. $\big|\mathsf{HON}\big| \geq \delta p^2 \big|R\big|$*;*
2. *for any* $i \in [n]$*, if we (temporarily) define* $Z^i_\tau = \big\{\tilde{\alpha}_i \in \big[2^{\tilde{t}_i}\big] : \big|\mathsf{HON}^i(\tilde{\alpha}_i)\big| \leq \tau \big|R^i(\tilde{\alpha}_i)\big|\big\}$*, then*

$$\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\alpha}_i \in Z^i_\tau\big) \leq \frac{\tau}{\delta p^2}.$$

Intuitively, 2 says that with good probability, for all values $\tilde{\alpha}_i$ which appear in an honest $\tilde{\boldsymbol{\alpha}}$, $\mathsf{HON}^i(\tilde{\alpha}_i)$ comprises at least a $\tau-$fraction of $R^i(\tilde{\alpha}_i)$.

Proof. 1 follows immediately from the definition of USEFUL. For 2, we have

$$\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\alpha}_i \in Z^i_\tau\big) \leq \frac{\Pr_{\tilde{\boldsymbol{\alpha}}}\big(\tilde{\boldsymbol{\alpha}} \in \mathsf{HON} \big| \tilde{\alpha}_i \in Z^i_\tau\big)}{\Pr_{\tilde{\boldsymbol{\alpha}}}\big(\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}\big)} \leq \frac{\tau}{\delta p^2}$$

□

Parameters. We have already introduced parameters $n = \mathcal{O}(\lambda)$, non-negligible $p = p(\lambda)$, constants $\delta, \delta' < 1/3$, and $\varepsilon^* = \big(\lambda/N\big)^{1/2}$ for $N = \text{poly}(\lambda)$, a yet unspecified polynomial. Shortly we will introduce the values $\varepsilon = 1/n - \varepsilon'$ where $\varepsilon' = 1/2n^2$. We will require that $\varepsilon^* \leq \sigma\delta^2 p^5/16$ and also that $\varepsilon^* \leq n\varepsilon'(\varepsilon'\delta\delta' p^3)^2/2048$, where $\sigma = \varepsilon'(\delta')^2 p^4/257n^3$ is defined for convenience. All in all, setting $N = \omega\big(\lambda n^{10} p^{-18}\big)$ will suffice. We stress that there is no reason to believe that N must be such a large polynomial; it arises due to our analysis, which is not concerned with minimizing N. We now formally define $\varepsilon-$dependence.

Definition 12 ($\varepsilon-$**dependence**). *For fixed* $\mathbb{T} \in \mathsf{ACC}$ *and* $i, i' \in \{1, \ldots, n\}$*, we say* $\alpha_{i'}$ *is* $\varepsilon-$dependent *on* $\tilde{\alpha}_i$ *if* $\Pr_{\tilde{\boldsymbol{\beta}} \in \mathsf{HON}}\big(\beta_{i'} = \alpha_{i'} \big| \tilde{\beta}_i = \tilde{\alpha}_i\big) \geq \varepsilon$.

We stress that it is important to condition on the event $\tilde{\boldsymbol{\beta}} \in \mathsf{HON}$ because any statement about M's behavior during the query/response phase is useless unless M actually plans to successfully complete the right protocol.

Note that if $\varepsilon > \varepsilon'$ and $\alpha_{i'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_i$, then $\alpha_{i'}$ is automatically also $\varepsilon'-$dependent on $\tilde{\alpha}_i$. Additionally, notice that though our definition does leave open the possibility that there could be more than one value which is $\varepsilon-$dependent on $\tilde{\alpha}_i$, there can only be polynomially many (at most ε^{-1} to be exact). We call these values the $\varepsilon-$*dependencies* of $\tilde{\alpha}_i$. This notion is different from $\varepsilon-$dependence defined above only because the $\varepsilon-$dependencies exist regardless of what queries are asked in $\mathbb{T}$, whereas we only say that $\alpha_{i'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_i$ if both $\tilde{\alpha}_i$ and $\alpha_{i'}$ appear in $\mathbb{T}$. For the remainder of the proof we fix non-negligible values ε and ε' such that $\varepsilon = 1/n - \varepsilon'$ and $\varepsilon' = 1/2n^2$.

Definition 13 (**Special Sets of Transcripts**). *Fix (as a function of* λ*),* $\omega = \omega(1)$*. Define the following sets of transcripts:*

1. $\mathsf{UNBAL} := \big\{\mathbb{T} \in \mathsf{ACC} : \exists\ i' > i \text{ st } \alpha_{i'} \text{ is } \varepsilon' - \text{dependent on } \tilde{\alpha}_i\big\}$;
2. $\mathsf{1{-}2} := \big\{\mathbb{T} \in \mathsf{ACC} : \exists\ (i_1, i_2, i') \text{ st } \alpha_{i'} \text{ is } \varepsilon' - \text{dependent on both } \tilde{\alpha}_{i_1} \text{ and } \tilde{\alpha}_{i_2}\big\}$;
3. $\mathsf{IND} := \big\{\mathbb{T} \in \mathsf{ACC} : \exists\ i \text{ st } \Pr_{\tilde{\boldsymbol{\beta}} \in \mathsf{HON}}\big(\beta_{i'} \neq \alpha_{i'}\ \forall\ i' \big| \tilde{\beta}_i = \tilde{\alpha}_i\big) \geq \varepsilon' n\big\}$;
4. $\mathsf{SUPER{-}POLY} := \big\{\mathbb{T} \in \mathsf{ACC} : {}^{\#}\{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON} : \mathrm{M}(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{\alpha}\} \geq \lambda^{\omega}\big\}$.

Note that if $\mathbb{T} \notin \mathsf{IND}$ then for all i, there exists an i' such that $\alpha_{i'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_i$.

What follows is a sequence of claims which sheds light on the relationships between the special sets of transcripts defined above. The statements all resemble one another and their proofs are similar, and are in order of increasing complexity. We recommend those readers who are interested in understanding the proofs to read them in order as it will make the later ones much easier to understand. Readers who are interested in understanding the general flow of our overall proof will most likely find reading the proof of Claim 5 and the statements of Claims 6 and 7 more than sufficient.

Claim 5. *Fix* $\sigma = \frac{\varepsilon'(\delta')^2 p^4}{257n^3}$*. If* $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{UNBAL}\big) \geq \frac{\delta' p}{4}$*, then*

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER{-}POLY}\big) \geq \sigma.$$

Proof. We begin with the inequality $\Pr_{\mathbb{T}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{UNBAL}\big) \geq \delta' p^2/4$ (using the fact that $\Pr_{\mathbb{T}}(\mathbb{T} \in \mathsf{ACC}) \geq p$). Fix a random commit message $\mathbf{Com}$. With probability at least $\delta' p^2/8$ over $\mathbf{Com}$, we have that $\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{UNBAL} \big| \mathbf{Com}\big) \geq \delta' p^2/8$. Now let $i' > i$ be such that $\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\alpha_{i'} \text{ is } \varepsilon' - \text{dependent on } \tilde{\alpha}_i \;\&\; \mathbb{T} \in \mathsf{TRB} \big| \mathbf{Com}\big) \geq \delta' p^2/8n^2$. Such (i, i') must exist by definition of UNBAL. Temporarily define the sets X and Z as follows:

- $X = \big\{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON} : \alpha_{i'} \text{ is } \varepsilon' - \text{dependent on } \tilde{\alpha}_i \;\&\; \mathbb{T} \in \mathsf{TRB}\big\}$;
- $Z = \big\{\tilde{\alpha}_i \in \big[2^{\tilde{t}_i}\big] : \big|\mathsf{HON}^i(\tilde{\alpha}_i)\big| \leq \tau \big|R^i(\tilde{\alpha}_i)\big|\big\}$, where $\tau = \frac{\delta \delta' p^4}{16n^2}$.

Remark. Defining temporary sets X, Y and Z will be a recurring theme throughout the proofs in this section (though in this first proof we only need X and Z). X will be a set of queries which display evidence of a particular type of query dependence; and Y and Z will be sets of queries in a particular coordinate in the right session which display some certain bad behavior. We will lower bound the probability that $\tilde{\boldsymbol{\alpha}} \in X$ using the claim's hypotheses, and we will upper bound the probability that $\tilde{\alpha}_i \in Z$ using Claim 4 (in fact, Z is the same set as Z^i_τ in the statement of Claim 4, just with the indices omitted for simplicity). Though it doesn't appear here, we will also upper bound the probability that $\tilde{\alpha}_i \in Y$ using simple conditional probability. We now proceed.

We have

$$\begin{aligned}\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_i \notin Z\big) &\geq \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\big) - \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\alpha}_i \in Z\big) \\ &\geq \frac{\delta' p^2}{8n^2} - \frac{\delta' p^2}{16n^2} = \frac{\delta' p^2}{16n^2},\end{aligned}$$

using Claim 4. However, if $\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}$ is such that $\tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_i \notin Z$, then $\mathbb{T} \in \mathsf{TRB}$ and M maps an $\varepsilon'-$fraction of $\mathsf{HON}^i(\tilde{\alpha}_i)$ into $L^{i'}(\alpha_{i'})$. Furthermore, as

$$\big|\mathsf{HON}^i(\tilde{\alpha}_i)\big| \geq \tau \big|R^i(\tilde{\alpha}_i)\big| \geq \tau 2^{\omega(\log \lambda)} \big|L^{i'}(\alpha_{i'})\big|$$

(using $i' > i$ and that the tags are well spaced), we see that M, when restricted appropriately, is superpolynomially many to one on average. This means that $\mathbb{T} \in \mathsf{TRB}$ and that $\boldsymbol{\alpha}$ has superpolynomially many preimages in HON whp, and so

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER-POLY}\big) \geq \frac{\delta' p^2}{8} \cdot \frac{\delta' p^2}{16n^2} \cdot \big(1 - \mathbf{negl}(\lambda)\big) = \frac{(\delta')^2 p^4}{128n^2} - \mathbf{negl}(\lambda) > \sigma.$$

□

Claim 6. *Fix* $\sigma = \frac{\varepsilon'(\delta')^2 p^4}{257 n^3}$. *If* $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap 1\!-\!2\big) \geq \frac{\delta' p}{4}$, *then*

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER-POLY}\big) \geq \sigma.$$

Proof. Fix a commitment message $\mathbf{Com}$. With probability at least $\delta' p^2/8$ over the choice of $\mathbf{Com}$, we have $\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\mathbb{T} \in \mathsf{TRB} \cap 1\!-\!2 \big| \mathbf{Com}\big) \geq \delta' p^2/8$. Let (i_1, i_2, i') be such that

$$\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\alpha_{i'} \text{ is } \varepsilon' - \text{dependent on } \tilde{\alpha}_{i_1} \text{ and } \tilde{\alpha}_{i_2} \;\&\; \mathbb{T} \in \mathsf{TRB} \big| \mathbf{Com}\big) \geq \frac{\delta' p^2}{8n^3}.$$

Such (i_1, i_2, i') must exist by definition of $1\!-\!2$. Temporarily define sets X, Y and Z:

- $X = \big\{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON} : \alpha_{i'} \text{ is } \varepsilon' - \text{dependent on both } \tilde{\alpha}_{i_1} \text{ and } \tilde{\alpha}_{i_2} \;\&\; \mathbb{T} \in \mathsf{TRB}\big\}$;
- $Y = \big\{\tilde{\alpha}_{i_1} \in \big[2^{\tilde{t}_{i_1}}\big] : \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\boldsymbol{\alpha} \in X \big| (\mathbf{Com}, \tilde{\alpha}_{i_1})\big) \leq \frac{\varepsilon'\delta' p^2}{16n^3}\big\}$;
- $Z = \big\{\tilde{\alpha}_{i_1} \in \big[2^{\tilde{t}_{i_1}}\big] : \big|\mathsf{HON}^{i_1}(\tilde{\alpha}_{i_1})\big| \leq \tau \big|R^{i_1}(\tilde{\alpha}_{i_1})\big|\big\}$, where $\tau = \frac{\varepsilon'\delta\delta' p^4}{32n^3}$.

Note that with $\mathbf{Com}$ fixed as above we have

$$\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\big) \geq \frac{\delta' p^2}{8n^3};\; \Pr_{\tilde{\boldsymbol{\alpha}}}\big(\tilde{\alpha}_{i_1} \in Y \big| \tilde{\boldsymbol{\alpha}} \in X\big) \leq \frac{\varepsilon'}{2};\; \text{and } \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\alpha}_{i_1} \in Z\big) \leq \frac{\varepsilon'\delta' p^2}{32n^3}.$$

Now, for $v \in \big[2^{t_{i'}}\big]$, let $\mathbf{E}_v$ be the event "$\alpha_{i'} = v$." Note that if $\Pr_{\tilde{\boldsymbol{\alpha}}}\big(\mathbf{E}_v \big| \tilde{\boldsymbol{\alpha}} \in X\big) > 0$ then v is an $\varepsilon'-$dependency of $\tilde{\alpha}_{i_1}$. Let $D^{i'}(\tilde{\alpha}_{i_1}) \subset \big[2^{t_{i'}}\big]$ be the set of all $\varepsilon'-$dependencies of $\tilde{\alpha}_{i_1}$. Then for fixed $\tilde{\alpha}_{i_1}$, we define a probability mass function on $D^{i'}(\tilde{\alpha}_{i_1})$ by $P(v) = \Pr_{\tilde{\boldsymbol{\alpha}}}\big(\mathbf{E}_v \big| \tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_{i_1}\big)$. We say that $v^* \in D^{i'}(\tilde{\alpha}_{i_1})$ is *maximal* if $P(v^*) \geq P(v)$ for all $v \in D^{i'}(\tilde{\alpha}_{i_1})$. Clearly for a random $\tilde{\boldsymbol{\alpha}} \in X$, the resulting $\alpha_{i'}$ is maximal with probability at least ε' as $\big|D^{i'}(\tilde{\alpha}_{i_1})\big| \leq (\varepsilon')^{-1}$. We now lower bound the quantity $V = \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_{i_1} \notin Y \cup Z \;\&\; \alpha_{i'} \text{ maximal}\big)$. We have

$$\begin{aligned}
V &\geq \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_{i_1} \notin Y \;\&\; \alpha_{i'} \text{ maximal}\big) - \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\alpha}_{i_1} \in Z\big) \\
&\geq \Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\big) \cdot \Big[\Pr_{\tilde{\boldsymbol{\alpha}}}\big(\tilde{\alpha}_{i_1} \notin Y \;\&\; \alpha_{i'} \text{ maximal} \big| \tilde{\boldsymbol{\alpha}} \in X\big)\Big] - \frac{\varepsilon'\delta' p^2}{32n^3} \\
&\geq \frac{\delta' p^2}{8n^3} \cdot \Big[\Pr_{\tilde{\boldsymbol{\alpha}}}\big(\alpha_{i'} \text{ maximal} \big| \tilde{\boldsymbol{\alpha}} \in X\big) - \Pr_{\tilde{\boldsymbol{\alpha}}}\big(\tilde{\alpha}_{i_1} \in Y \big| \tilde{\boldsymbol{\alpha}} \in X\big)\Big] - \frac{\varepsilon'\delta' p^2}{32n^3} \\
&\geq \frac{\delta' p^2}{8n^3} \cdot \frac{\varepsilon'}{2} - \frac{\varepsilon'\delta' p^2}{32n^3} = \frac{\varepsilon'\delta' p^2}{32n^3}.
\end{aligned}$$

Finally we show that if $\tilde{\boldsymbol{\alpha}}$ is such that "$\tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_{i_1} \notin Y \cup Z \;\&\; \alpha_{i'}$ is maximal", then $\mathbb{T} \in \mathsf{TRB}$ (where $\mathbb{T}$ is the transcript resulting from $\tilde{\boldsymbol{\alpha}}$) and with probability at least $\tau' = \delta(\delta')^2(\varepsilon')^4 p^6/512n^6$ over $\tilde{\boldsymbol{\beta}} \in \mathsf{HON}$, we will have $\beta_{i'} = \alpha_{i'}$. This completes the proof of Claim 6 as it means that M maps a $\tau'-$fraction of HON into $L^{i'}(\alpha_{i'})$ and since

$$\big|\mathsf{HON}\big| \geq \delta p^2 \big|R\big| \geq \delta p^2 2^{\omega(\log\lambda)} \big|L^{i'}(\alpha_{i'})\big|$$

(using the "well spaced" property of the tags), M is superpolynomially many to one on average when restricted appropriately. Just like in the proof of Claim 5, this gives

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER-POLY}\big) \geq \frac{\delta' p^2}{8} \cdot \frac{\varepsilon'\delta' p^2}{32n^3} - \mathbf{negl}(\lambda) > \sigma.$$

So all that remains is to prove that if $\tilde{\boldsymbol{\alpha}}$ is such that "$\tilde{\boldsymbol{\alpha}} \in X \;\&\; \tilde{\alpha}_{i_1} \notin Y \cup Z \;\&\; \alpha_{i'}$ is maximal" then $\Pr_{\tilde{\boldsymbol{\beta}} \in \mathsf{HON}}\big(\beta_{i'} = \alpha_{i'}\big) \geq \tau'$. The maximality of $\alpha_{i'}$ combined with the fact that $\tilde{\alpha}_{i_1} \notin Y$ ensure that if $\tilde{\boldsymbol{\gamma}} \in \mathsf{HON}^{i_1}(\tilde{\alpha}_{i_1})$ is chosen at random, then with probability at least $(\varepsilon')^2\delta' p^2/16n^3$ over $\tilde{\boldsymbol{\gamma}}$ we will have "$\gamma_{i'}$ is $\varepsilon' -$ dependent on $\tilde{\gamma}_{i_1}$ and $\tilde{\gamma}_{i_2} \;\&\; \gamma_{i'} = \alpha_{i'}$". Moreover, as $\tilde{\alpha}_{i_1} \notin Z$, a random $\tilde{\boldsymbol{\gamma}} \in R^{i_1}(\tilde{\alpha}_{i_1})$ will be such that "$\gamma_{i'}$ is $\varepsilon' -$ dependent on $\tilde{\gamma}_{i_1}$ and $\tilde{\gamma}_{i_2} \;\&\; \gamma_{i'} = \alpha_{i'}$" with probability at least $\delta(\delta')^2(\varepsilon')^3 p^6/512n^6 = \tau'/\varepsilon'$.

So choose a random $\tilde{\boldsymbol{\gamma}} \in R^{i_1}(\tilde{\alpha}_{i_1})$ and then choose a random $\tilde{\boldsymbol{\beta}} \in \mathsf{HON}^{i_2}(\tilde{\gamma}_{i_2})$. Clearly such a $\tilde{\boldsymbol{\beta}}$ is a random element of HON. As "$\gamma_{i'}$ is $\varepsilon' -$ dependent on $\tilde{\gamma}_{i_2} \;\&\; \gamma_{i'} = \alpha_{i'}$" with probability at least τ'/ε', the definition of $\varepsilon'-$dependence ensures that $\beta_{i'} = \gamma_{i'} = \alpha_{i'}$ with probability at least τ', as desired. $\square$

Claim 7. *Fix* $\sigma = \frac{\varepsilon'(\delta')^2 p^4}{257n^3}$. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \setminus (\mathsf{UNBAL} \cup 1{-}2 \cup \mathsf{IND})\big) \geq \frac{\delta' p}{4}$, *then*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER{-}POLY}\big) \geq \sigma.$$

Proof. Fix a commitment message **Com**. With probability at least $\delta' p^2/8$ over the choice of **Com**, we have $\Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\mathbb{T} \in \mathsf{TRB}\setminus(\mathsf{UNBAL}\cup 1{-}2\cup\mathsf{IND})\big|\mathbf{Com}\big) \geq \delta' p^2/8$. Now consider the consequences of $\mathbb{T} \notin (\mathsf{UNBAL} \cup 1{-}2 \cup \mathsf{IND})$:

- if $\mathbb{T} \notin \mathsf{UNBAL}$, then for all $i' > i$, $\alpha_{i'}$ cannot be $\varepsilon-$dependent on $\tilde{\alpha}_i$ (since $\varepsilon \geq \varepsilon'$);
- if $\mathbb{T} \notin 1{-}2$, then there do not exist (i_1, i_2, i') such that $\alpha_{i'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_{i_1}$ and $\tilde{\alpha}_{i_2}$;
- if $\mathbb{T} \notin \mathsf{IND}$ then for every i, there exists at least one i' such that $\alpha_{i'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_i$.

It follows that if $\mathbb{T} \notin (\mathsf{UNBAL}\cup 1{-}2\cup\mathsf{IND})$ then for each i, α_i must be $\varepsilon-$dependent on $\tilde{\alpha}_i$. Indeed, α_1 must be $\varepsilon-$dependent on $\tilde{\alpha}_1$ as something must depend on $\tilde{\alpha}_1$ and it cannot be $\alpha_{i'}$ for $i' > 1$. Next, either α_1 or α_2 must be $\varepsilon-$dependent on $\tilde{\alpha}_2$ and it cannot be α_1 as that is already dependent on $\tilde{\alpha}_1$. Continuing in this fashion, we deduce that each α_i is $\varepsilon-$dependent on $\tilde{\alpha}_i$.

Now, going one step further in examining the consequences of $\mathbb{T} \notin (\mathsf{UNBAL} \cup 1{-}2 \cup \mathsf{IND})$, since each α_i is $\varepsilon-$dependent on $\tilde{\alpha}_i$ and $\mathbb{T} \notin 1{-}2$, it must be that $\alpha_{i'}$ is not $\varepsilon'-$dependent on $\tilde{\alpha}_i$ for all $i' \neq i$. It follows that for all i, $\Pr_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}\big(\exists\ i' \neq i \text{ st } \beta_{i'} = \alpha_{i'}\big|\tilde{\beta}_i = \tilde{\alpha}_i\big) \leq \varepsilon' n$. As $\mathbb{T} \notin \mathsf{IND}$, we have that for all i,

$$\begin{aligned}
\Pr_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}\big(\beta_i = \alpha_i\big|\tilde{\beta}_i = \tilde{\alpha}_i\big) &\geq \Pr_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}\big(\exists\ i' \text{ st } \beta_{i'} = \alpha_{i'}\big|\tilde{\beta}_i = \tilde{\alpha}_i\big) \\
&- \Pr_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}\big(\exists\ i' \neq i \text{ st } \beta_{i'} = \alpha_{i'}\big|\tilde{\beta}_i = \tilde{\alpha}_i\big). \\
&\geq 1 - \varepsilon' n - \varepsilon' n = 1 - 2\varepsilon' n,
\end{aligned}$$

so we see that, in fact, each α_i is $(1-2\varepsilon' n)-$dependent on $\tilde{\alpha}_i$. As $2\varepsilon' n < \frac{1}{2}$, each $\tilde{\alpha}_i$ has a unique $(1-2\varepsilon' n)-$depencence.

Now, choose a random $\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}$ and let $S = \{i \in [n] : \tilde{t}_i \leq t_i\}$, $\tilde{\boldsymbol{\alpha}}_S = (\tilde{\alpha}_i)_{i\in S}$ and define $\mathsf{HON}^S(\tilde{\boldsymbol{\alpha}}_S) = \bigcap_{i\in S}\mathsf{HON}^i(\tilde{\alpha}_i)$. Define $R^S(\tilde{\boldsymbol{\alpha}}_S)$ and $L^S(\boldsymbol{\alpha}_S)$ similarly. Now, temporarily define sets X, Y, Z as follows:

- $X = \big\{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON} : \alpha_i \text{ is } (1-2\varepsilon' n) - \text{dependent on } \tilde{\alpha}_i\ \forall\ i\ \&\ \mathbb{T} \in \mathsf{TRB}\big\}$;
- $Y = \big\{\tilde{\boldsymbol{\alpha}}_S : \Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\big|\tilde{\boldsymbol{\alpha}}_S\big) \leq \frac{\delta' p^2}{16}\big\}$;
- $Z = \big\{\tilde{\boldsymbol{\alpha}}_S : \big|\mathsf{HON}^S(\tilde{\boldsymbol{\alpha}}_S)\big| \leq \tau\big|R^S(\tilde{\boldsymbol{\alpha}}_S)\big|\big\}$, where $\tau = \frac{\delta\delta' p^4}{32}$.

Note that

$$\Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\big) \geq \frac{\delta' p^2}{8};\ \Pr_{\tilde{\boldsymbol{\alpha}}}\big(\tilde{\boldsymbol{\alpha}}_S \in Y\big|\tilde{\boldsymbol{\alpha}} \in X\big) \leq \frac{1}{2};\ \text{and } \Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}}_S \in Z\big) \leq \frac{\delta' p^2}{32},$$

and so $\Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\ \&\ \tilde{\boldsymbol{\alpha}}_S \notin Y \cup Z\big) \geq \delta' p^2/32$. Now suppose that some $\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}$ is such that "$\tilde{\boldsymbol{\alpha}} \in X\ \&\ \tilde{\boldsymbol{\alpha}}_S \notin Y \cup Z$". Then $\mathbb{T} \in \mathsf{TRB}$ and for a randomly selected $\tilde{\boldsymbol{\beta}} \in \mathsf{HON}^S(\tilde{\boldsymbol{\alpha}}_S)$, $\tilde{\boldsymbol{\beta}} \in X$ with probability at least $\delta' p^2/16$. But if $\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}} \in X$ and $\tilde{\boldsymbol{\alpha}}_S = \tilde{\boldsymbol{\beta}}_S$, then $\boldsymbol{\alpha}_S = \boldsymbol{\beta}_S$. Indeed, all $\tilde{\alpha}_i$ have a unique $(1-2\varepsilon' n)-$dependency, meaning that if α_i and β_i are dependent on $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ and $\tilde{\alpha}_i = \tilde{\beta}_i$ for all $i \in S$, then it must be that $\alpha_i = \beta_i$ for all $i \in S$.

It follows that if "$\tilde{\boldsymbol{\alpha}} \in X$ & $\tilde{\boldsymbol{\alpha}}_S \notin Y \cup Z$" then $\mathbb{T} \in \mathsf{TRB}$ and M maps a $\tau'-$fraction of $\mathsf{HON}^S(\tilde{\boldsymbol{\alpha}}_S)$ into $L^S(\boldsymbol{\alpha}_S)$ where $\tau' = \delta' p^2/16$. Moreover,

$$\big|\mathsf{HON}^S(\tilde{\boldsymbol{\alpha}}_S)\big| \geq \tau \big|R^S(\tilde{\boldsymbol{\alpha}}_S)\big| \geq \tau 2^{\omega(\log \lambda)} \big|L^S(\boldsymbol{\alpha}_S)\big|$$

(using the "good distance and balance" property of the tags). As in the proofs of Claims 5 and 6, we have

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER-POLY}\big) \geq \frac{\delta' p^2}{8} \cdot \frac{\delta' p^2}{32} - \mathbf{negl}(\lambda) > \sigma.$$

□

5.3 Reductions to the Hiding of Com

Overview. We begin with a brief and oversimplified description of where we currently stand in the proof of Lemma 1, and the outline of how we complete it. Recall we must show that the event $\mathbb{T} \in \mathsf{TRB}$ is unlikely. In the previous section we identified four sub-events:

$$1)\mathsf{TRB} \cap \mathsf{UNBAL};\ 2)\mathsf{TRB} \cap 1\text{-}2;\ 3)\mathsf{TRB} \cap \mathsf{IND};\ 4)\mathsf{TRB} \cap \mathsf{SUPER-POLY},$$

and we showed that sub-event 4 essentially contains 1 and 2 and $\mathsf{TRB} \cap \big(\neg(1 \text{ or } 2 \text{ or } 3)\big)$. Therefore, it remains to show that sub-events 3 and 4 are unlikely. These correspond to the two claims proved in this section. Both are similar in that they are reductions to hiding, we give a very brief overview of the intuition. First, recall the meaning of $\mathbb{T} \in \mathsf{TRB}$: it means that M answers $\tilde{\boldsymbol{\alpha}}$ correctly/incorrectly on the right if given correct/incorrect answers to $\boldsymbol{\alpha}$ on the left. Roughly speaking, if IND occurs then there exists some $i \in [n]$ such that if M is rewound and a new $\tilde{\boldsymbol{\beta}}$ is asked such that $\tilde{\beta}_i = \tilde{\alpha}_i$, the $\boldsymbol{\beta}$ that M asks on the left shares no coordinates with $\boldsymbol{\alpha}$. Thus if sub-event 3 is likely, then we might hope to break the hiding of $\langle \mathrm{C}, \mathrm{R} \rangle$ as follows:

1. choose $m_0, m_1 \in \mathbb{Z}_q$ randomly and receive a commitment to one of them, and use M to generate $\mathbb{T}$ containing $(\boldsymbol{\alpha}, \mathbf{a})$;
2. let $\mathbf{f}_0$ and $\mathbf{f}_1$ be the polynomial vectors such that $\mathbf{f}_b(\boldsymbol{\alpha}) = \mathbf{a}$ and $\mathbf{f}_b(0) = m_b$;
3. rewind M and send $\tilde{\boldsymbol{\beta}}$ such that $\tilde{\beta}_i = \tilde{\alpha}_i$, obtain $\boldsymbol{\beta}$ which shares no coordinates with $\boldsymbol{\alpha}$;
4. prepare $\mathbf{b}$ using $\mathbf{f}_b$, obtain $\tilde{\mathbf{b}}$; guess m_b if $\tilde{b}_i = \tilde{a}_i$.

The point is that if the correct/incorrect $\mathbf{f}$ is used, then $\mathbf{b}$ is correct/random. In the first case, $\mathbb{T} \in \mathsf{TRB}$ says M answers $\tilde{\mathbf{b}}$ correctly so $\tilde{b}_i = \tilde{a}_i$; in the second case, M answers $\tilde{b}_i$ incorrectly, so $\tilde{b}_i \neq \tilde{a}_i$. This captures the intuition for how we show that sub-event 3 is unlikely. Sub-event 4 is similar but more complicated as it uses an algebraic argument analogous to the proof of Claim 1 where we bounded the likelihood of the "incorrect but collinear" event. We now proceed formally by describing a general adversary $\mathcal{A}$ who participates in the hiding game for $\langle \mathrm{C}, \mathrm{R} \rangle$ as follows. Our two claims of this section are proven by further specifying this "template" $\mathcal{A}$.

- $\mathcal{A}$ chooses $m_0, m_1 \in \mathbb{Z}_q$ randomly, and sends (m_0, m_1) to a challenger $\mathcal{C}$, signaling the beginning of the hiding game of $\langle \mathrm{C}, \mathrm{R} \rangle$.

- $\mathcal{A}$ instantiates M and runs two sessions of $\langle \mathrm{C}, \mathrm{R} \rangle$ until the end of the commit phase of both executions, forwarding the messages it receives as C to $\mathcal{C}$. In the left execution, $\mathcal{C}$ commits to m_u for secret $u \in \{0, 1\}$. More specifically:
  - $\mathcal{A}$, acting as R, sends $\tilde{\sigma}$ to M, and receives σ which it forwards to $\mathcal{C}$.
  - $\mathcal{A}$ then receives **Com** from $\mathcal{C}$ which it forwards to M, and receives $\tilde{\mathbf{Com}}$.
  - $\mathcal{A}$ sends random $\tilde{\boldsymbol{\alpha}}$ such that $\tilde{\alpha}_i \in \left[2^{\tilde{t}_i}\right]$ to M, receiveing $\boldsymbol{\alpha}$ which it forwards to $\mathcal{C}$.
  - $\mathcal{A}$ receives $\mathbf{a}$ from $\mathcal{C}$ which it forwards to M, obtaining $\tilde{\mathbf{a}}$.
  - $\mathcal{A}$ continues forwarding messages between M and $\mathcal{C}$ during the zero-knowledge proof phase of $\langle \mathrm{C}, \mathrm{R} \rangle$, playing honestly as R in the right interaction.
  - When the proofs are finished, $\mathcal{A}$ verifies both π and $\tilde{\pi}$. If either is not accepted, $\mathcal{A}$ aborts. Let $\mathbb{T} = \left(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\right)$ be the resulting transcript.
- $\mathcal{A}$ chooses random $u' \in \{0, 1\}$ and defines polynomial vector $\mathbf{f}$ such that $\mathbf{f}(\boldsymbol{\alpha}) = \mathbf{a}$ and every coordinate of $\mathbf{f}$ has constant term $m_{u'}$.
- $\mathcal{A}$ rewinds M to the beginning of the query phase of the right execution and sends a new query $\tilde{\boldsymbol{\beta}}$, receiving left query $\boldsymbol{\beta}$. It can do this many times, resulting in a set of new right queries $\{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \dots\}$.
- $\mathcal{A}$ answers the left queries it obtained in the previous step with $\mathbf{f}$, and receives a right response. It collects the points it receives on the right into the set $\left\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}}), \dots\right\}$.
- $\mathcal{A}$ tests whether the points $\left\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}}), \dots\right\}$ satisfy some condition. If so, then $\mathcal{A}$ outputs u', if not it outputs $1 - u'$.

Exactly what condition $\mathcal{A}$ tests for will change between the two proofs. In the proof of Claim 8, $\mathcal{A}$ checks that the points $\left\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\right\}$ are collinear, while in the proof of Claim 9, $\mathcal{A}$ checks that $\tilde{b}_i = \tilde{a}_i$ for some preselected i. The important thing however, is that the condition be satisfied when M answers correctly, but not when M answers incorrectly. Note that if $u' = u$ then responses generated with $\mathbf{f}$ are correct and so if $\mathbb{T} \in \mathsf{USEFUL}$, then we can lower bound the probability that M answers correctly on the right using Claim 4. On the other hand, if $u' \neq u$ then the responses on the left are random. If $\mathbb{T} \notin \mathsf{EXT}$ then we have an upper bound on the probability that M answers any right query correctly. These observations together tell us that there is a non-negligible gap between the probability that the condition is satisfied when $u' = u$ and when $u' \neq u$. This gap translates to $\mathcal{A}$ having a noticeable advantage in winning the hiding game.

There are two main issues with the above outline which need to be addressed. We discuss them informally here in order to exhibit the difficulties faced when trying to push the above intuition through. The first is that we have assumed that $\mathbb{T} \in \mathsf{TRB}$ when in reality we are only allowed to assume that $\mathbb{T} \in \mathsf{TRB}$ with probability at least $\frac{\delta' p}{4}$. Fact 1 below says essentially that if the gap between the condition being satisfied when $u' = u$ and not when $u' \neq u$ is large enough, this does not matter.

A second, more subtle, issue is that we can only use $\mathbb{T} \notin \mathsf{EXT}$ to upper bound the probability that M answers correctly on the right when $u' \neq u$ if the answers on the left are distributed as if they were answered by the extractor, E. Recall that E is instructed to answer randomly on the left

unless the left query is the same as in the main thread, in which case E reuses the main thread's answer. Note that this process is exactly the same as answering one query according to $\mathbf{f}$ when $u' \neq u$. However, if M is rewound more than once and asks left challenges $\{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$, the responses it receives will no longer be random. Indeed, $\big\{(\boldsymbol{\alpha}, \mathbf{a}), (\boldsymbol{\beta}, \mathbf{b}), (\boldsymbol{\gamma}, \mathbf{c})\big\}$ will be collinear so certainly not random (and hence, not distributed as E's responses). This will mean, for example, that we will not be able to use Claim 1 to argue that M's responses on the right cannot be incorrect but collinear. In fact, if M receives random but collinear responses on the left, it might well be the case that M's right responses are incorrect but collinear (consider for example the copying MIM). Instead, we will have to use the additional hypothesis that $\mathbb{T} \in \mathsf{SUPER}-\mathsf{POLY}$ along with the observation that $\boldsymbol{\beta}$ is answered identically to how E would answer it to bound the probability that $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\big\}$ are collinear when $u' \neq u$. For details see Claim 8 below.

In the proof of Claim 9, $\mathcal{A}$ rewinds M and asks a new challenge $\tilde{\boldsymbol{\beta}}$ such that $\tilde{\beta}_i = \tilde{\alpha}_i$ for some i. Note that if $\boldsymbol{\beta}$ is such that $\beta_{i'} = \alpha_{i'}$ for some i', then M will receive at least one correct answer on the left regardless of whether $u' = u$ or not. If $u' \neq u$, this will mean that the answers M receives on the left are not distributed identically to the answers M would receive from E. Indeed, suppose that some $\alpha_{i'}$ is dependent on $\tilde{\alpha}_i$. Then if $\tilde{\boldsymbol{\beta}}$ such that $\tilde{\beta}_i = \tilde{\alpha}_i$ is asked on the right by $\mathcal{A}$, M will ask $\boldsymbol{\beta}$ on the left with $\beta_{i'} = \alpha_{i'}$, and get at least one correct response. If, on the other hand, $\tilde{\boldsymbol{\beta}}$ is asked on the right by E, then with overwhelming probability, $\tilde{\boldsymbol{\beta}}$ does not share any query with the query vector asked in the main thread as E draws its queries randomly, independent of $\mathbb{T}$. This means that $\beta_{i'}$ will likely not equal $\alpha_{i'}$, and so M will get a random response instead of a correct one. This inherent difference between $\mathcal{A}$ and E means that we cannot use Claim 2 to upper bound the probability that M answers correctly on the right. Instead we have to use the additional assumption that $\mathbb{T} \notin \mathsf{IND}$ to ensure that $\boldsymbol{\beta}$ is completely distinct from $\boldsymbol{\alpha}$ *even though* $\tilde{\beta}_i = \tilde{\alpha}_i$ on the right. Even with this assumption, the proof requires some delicacy to ensure that in fact the answers $\mathcal{A}$ gives to M are the same as the ones E would give. For details see the proof of Claim 9.

Fact 1. *Consider an efficiently testable condition that the set* $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{c}}), \dots\big\}$ *either satisfies or not, as described in the above paragraphs. Let* $\mathbf{E}$ *be an event such that:*

- $\Pr_{\mathbb{T} \in \mathsf{ACC}}(\mathbf{E}) \geq \xi$*;*
- $\Pr\big(\text{Condition satisfied} \big| u' = u \ \&\ \mathbf{E}\big) \geq \xi'$*;*
- $\Pr\big(\text{Condition satisfied} \big| u' \neq u \ \&\ \mathbf{E}\big) \leq \xi''$*,*

for non-negligible values ξ, ξ', ξ'' *satisfying* $\xi'' \leq (p\xi\xi')/8$*. Then there exists a PPT algorithm* $\mathcal{A}$ *that breaks the hiding of* $\langle \mathrm{C}, \mathrm{R} \rangle$*.*

Proof. Fix $\ell = 1/2\xi''$ and let $\mathcal{A}$ play in an $\ell-$way version of the usual hiding game of $\langle \mathrm{C}, \mathrm{R} \rangle$ as follows:

- $\mathcal{A}$ chooses random $m_1, \dots, m_\ell \in \mathbb{Z}_q$ and sends $(m_1, \dots, m_\ell)$ to $\mathcal{C}$.
- $\mathcal{A}$ instantiates M and runs two sessions of $\langle \mathrm{C}, \mathrm{R} \rangle$ until the end of the commit phase of both executions, forwarding the messages it receives as C to $\mathcal{C}$. In the left execution, $\mathcal{C}$ commits to $m_{j'}$ for secret $j' \in [\ell]$.
- For each $j \in [\ell]$, $\mathcal{A}$ defines polynomial vectors $\mathbf{g}_j$ such that $\mathbf{g}_j(\boldsymbol{\alpha}) = \mathbf{a}$ and every coordinate of $\mathbf{g}_j$ has constant term m_j.

- $\mathcal{A}$ rewinds M to the beginning of the query phase of the right execution and sends new queries $\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \ldots$, receiving left queries $\boldsymbol{\beta}, \boldsymbol{\gamma}, \ldots$.
- For each $j \in [\ell]$, $\mathcal{A}$ answers the left queries it obtained in the previous step with $\mathbf{g}_j$, and receives a right response. It collects the set $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}_j), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}}_j), \ldots\big\}_{j \in [\ell]}$.
- For each $j \in [\ell]$, $\mathcal{A}$ tests whether the points $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}_j), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}}_j), \ldots\big\}$ satisfy the condition. If so, then $\mathcal{A}$ outputs $j^* = j$ and halts.

Note that

$$\begin{aligned}\Pr(j^* = j') &\geq \Pr_{\mathbb{T}}\big(\mathbb{T} \in \mathsf{ACC}\big) \cdot \Pr_{\mathbb{T} \in \mathsf{ACC}}(\mathbf{E}) \\ &\cdot \Pr\big(\text{Condition satisfied when } j = j' \big| \mathbf{E}\big) \\ &\cdot \Pr\big(\text{Condition not satisfied whenever } j \neq j' \big| \mathbf{E}\big) \\ &\geq (p\xi\xi') \cdot \Pr\big(\text{Not } \mathbf{E}'_j \text{ for all } j \neq j' \big| \mathbf{E}\big).\end{aligned}$$

where $\mathbf{E}'_j$ is the event

$$\mathbf{E}'_j : \text{ “Conditions are satisfied when } \mathbf{g}_j \text{ is used to answer left queries.”}$$

We are given that $\Pr\big(\mathbf{E}'_j \big| \mathbf{E}\big) \leq \xi''$ for all $j \neq j'$, and as the $\mathbf{E}'_j$ are independent this means that the expected number of $\mathbf{E}'_j$ which occur is at most $\xi''\ell = 1/2$. It follows that

$$\Pr(j^* = j') \geq (p\xi\xi') \cdot \Pr\big(\text{No } \mathbf{E}'_j \text{ occur when } j \neq j' \big| \mathbf{E}\big) \geq \frac{p\xi\xi'}{2} \geq \frac{2}{\ell},$$

which means that $\mathcal{A}$'s chances of winning the hiding game are noticeably greater than $1/\ell$, violating the hiding of $\langle \mathrm{C}, \mathrm{R} \rangle$. □

Claim 8. *Fix* $\sigma = \frac{\varepsilon'(\delta')^2 p^4}{257 n^3}$. *If* $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER{-}POLY}\big) \geq \sigma$ *then there exists a PPT algorithm* $\mathcal{A}$ *who breaks the hiding of* $\langle \mathrm{C}, \mathrm{R} \rangle$.

Proof. Our $\mathcal{A}$ proceeds as follows.

- $\mathcal{A}$ chooses random $m_0, m_1 \in \mathbb{Z}_q$ and begins the hiding game, sending (m_0, m_1) to $\mathcal{C}$. Then $\mathcal{A}$ instantiates M and runs two sessions of $\langle \mathrm{C}, \mathrm{R} \rangle$ forwarding the messages it receives as C to $\mathcal{C}$. In the left interaction, $\mathcal{C}$ commits to m_u for unknown $u \in \{0, 1\}$. Let $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big)$ be the resulting transcript. Additionally, $\mathcal{A}$ chooses random $u' \in \{0, 1\}$ and defines the polynomial vector $\mathbf{f}$, to be the unique such vector so that $\mathbf{f}(\boldsymbol{\alpha}) = \mathbf{a}$ and so that every coordinate of $\mathbf{f}$ has constant term $m_{u'}$.
- $\mathcal{A}$ chooses two new random challenge vectors $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ such that each $\tilde{\beta}_i, \tilde{\gamma}_i \in \big[2^{\tilde{t}_i}\big]$. It rewinds M back to the beginning of the right execution's query message and sends $\tilde{\boldsymbol{\beta}}$, receiving left query $\boldsymbol{\beta}$. It responds with $\mathbf{b} = \mathbf{f}(\boldsymbol{\beta})$ and receives right response $\tilde{\mathbf{b}}$. It repeats this process, sending challenge $\tilde{\boldsymbol{\gamma}}$, answering $\boldsymbol{\gamma}$ with $\mathbf{c} = \mathbf{f}(\boldsymbol{\gamma})$ and receiving $\tilde{\mathbf{c}}$.
- $\mathcal{A}$ checks whether the points $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\big\}$ are collinear (by checking for collinearity in each coordinate). If so, $\mathcal{A}$ outputs u', if not $\mathcal{A}$ outputs $1 - u'$.

In light of Fact 1, it suffices to construct an event $\mathbf{E}$ such that:

1. $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbf{E}\big) \geq \sigma$;
2. $\Pr\big(\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\} \text{ collinear}\big|u'=u \;\&\; \mathbf{E}\big) \geq \delta^2 p^4$;
3. $\Pr\big(\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\} \text{ collinear}\big|u'\neq u \;\&\; \mathbf{E}\big) \leq 2\varepsilon^*$,

since $\varepsilon^* \leq \sigma\delta^2 p^5/16$. Let $\mathbf{E}$ (temporarily) be the event "$\mathbb{T} \in \mathsf{TRB}\cap\mathsf{SUPER}-\mathsf{POLY}$." By hypothesis of Claim 8, $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbf{E}) \geq \sigma$. Also, if $\mathbb{T}\in\mathsf{USEFUL}$ and $u'=u$ then Claim 4 ensures that M answers $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ correctly on the right with probability at least $(\delta p^2)^2$, which means that the probability that $\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\}$ are collinear given $u'=u$ & $\mathbf{E}$ is at least as high. On the other hand,

$$\begin{aligned}\Pr\big(\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\} \text{ collinear}\big|u'\neq u \;\&\; \mathbf{E}\big) \;&\leq\; \Pr\big(\text{collinear}\big|\tilde{\mathbf{b}} \text{ incorrect}\big)\\ &+\; \Pr\big(\tilde{\mathbf{b}} \text{ correct}\big|u'\neq u \;\&\; \mathbf{E}\big)\\ &\leq\; \Pr\big(\text{collinear}\big|\tilde{\mathbf{b}} \text{ incorrect}\big)+\varepsilon^*,\end{aligned}$$

as if $u'\neq u$ then the answer M receives to $\boldsymbol{\beta}$ is distributed identically to the answer it would have received from E, and $\mathbb{T}\notin\mathsf{EXT}$. Therefore, it suffices to show that

$$\Pr\big(\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\} \text{ collinear}\big|\tilde{\mathbf{b}} \text{ incorrect}\big) = \mathbf{negl}(\lambda).$$

Suppose that $\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\alpha}}'\in\mathsf{HON}$ are such that $\mathrm{M}(\tilde{\boldsymbol{\alpha}})=\boldsymbol{\alpha}=\mathrm{M}(\tilde{\boldsymbol{\alpha}}')$. Note that it cannot be the case that $\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\}$ and $\big\{(\tilde{\boldsymbol{\alpha}}',\tilde{\mathbf{a}}'),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\gamma},\tilde{\mathbf{c}})\big\}$ are collinear as this would mean that the four points

$$\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\alpha}}',\tilde{\mathbf{a}}'),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\}$$

lie on the same line, and moreover, that this is the correct line as it contains the correct points $(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}})$ and $(\tilde{\boldsymbol{\alpha}}',\tilde{\mathbf{a}}')$. This contradicts the hypothesis that $\tilde{\mathbf{b}}$ is an incorrect answer. So we see that there exists at most one $\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}$ such that

1. $\mathrm{M}(\tilde{\boldsymbol{\alpha}})=\boldsymbol{\alpha}$;
2. $\big\{(\tilde{\boldsymbol{\alpha}},\tilde{\mathbf{a}}),(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{b}}),(\tilde{\boldsymbol{\gamma}},\tilde{\mathbf{c}})\big\}$ are collinear.

As $\mathbb{T}\in\mathsf{SUPER}-\mathsf{POLY}$, there are at least λ^ω values of $\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}$ such that number 1 holds, so the probability that $\mathcal{A}$ chose the unique $\tilde{\boldsymbol{\alpha}}$ such that both 1 and 2 hold is negligible. □

Claim 9. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\big) \geq \frac{\delta' p}{4}$ *then there exists a PPT algorithm* $\mathcal{A}$ *who breaks the hiding of* $\langle \mathrm{C},\mathrm{R}\rangle$.

Proof. For each $i'\in[n]$, define the set

$$\mathsf{FIXED}^{i'} = \big\{\mathbf{Com} : \exists\; v\in\big[2^{t_{i'}}\big] \text{ st } \Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\alpha_{i'}=v\big|\mathbf{Com}\big)\geq\varepsilon\big\},$$

and let $\mathsf{FIXED}=\big\{\mathbb{T}\in\mathsf{ACC} : \mathbf{Com}\in\mathsf{FIXED}^{i'} \text{ for some } i'\in[n]\big\}$.

Fact 2. *Fix* $\sigma=\frac{\varepsilon'(\delta')^2p^4}{257n^3}$. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{FIXED}\big)\geq\frac{\delta' p}{8}$, *then*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{SUPER}-\mathsf{POLY}\big)\geq\sigma.$$

Proof of Fact 2. This proof is similar to (and easier than) the proofs of Claims 5 through 7. Fix commitment message $\mathbf{Com}$. Just as in the previous proofs, with probability at least $\delta' p^2/16$ over $\mathbf{Com}$, $\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{FIXED} \big| \mathbf{Com}\big) \geq \delta' p^2/16$. Let $i' \in [n]$ and $v \in \big[2^{t_{i'}}\big]$ be such that

$$\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\alpha_{i'} = v \;\&\; \mathbb{T} \in \mathsf{TRB} \big| \mathbf{Com}\big) \geq \frac{\varepsilon \delta' p^2}{16n}.$$

Such (i', v) must exist by definition of FIXED. But this means that $\mathbb{T} \in \mathsf{TRB}$ and M maps at least a $\tau-$fraction of HON into $L^{i'}(v)$, where $\tau = \varepsilon \delta' p^2/16n$. As

$$\big|\mathsf{HON}\big| \geq \delta p^2 \big|R\big| \geq \delta p^2 2^{\omega(\log \lambda)} \big|L^{i'}(v)\big|,$$

(using the "well spaced" property of the tags), we see that M, when restricted appropriately, is superpolynomially many to one on average. It follows that

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER-POLY}\big) \geq \frac{\delta' p^2}{16} \cdot \frac{\varepsilon \delta' p^2}{16n} - \mathbf{negl}(\lambda) > \sigma.$$

□

In light of Fact 2 and Claim 8, it suffices to show that if $\Pr_{\tilde{\boldsymbol{\alpha}} \in \mathsf{HON}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{IND} \setminus \mathsf{FIXED}\big) \geq \delta' p/8$ then there exists a PPT $\mathcal{A}$ who breaks the hiding of $\langle \mathrm{C}, \mathrm{R} \rangle$. Therefore, assume that this probability is at least $\delta' p/8$ and define $\mathcal{A}$ as follows.

- $\mathcal{A}$ chooses random $m_0, m_1 \in \mathbb{Z}_q$ and begins the hiding game, sending (m_0, m_1) to $\mathcal{C}$. Then $\mathcal{A}$ instantiates M and runs two sessions of $\langle \mathrm{C}, \mathrm{R} \rangle$ forwarding the messages it receives as C to $\mathcal{C}$. In the left interaction, $\mathcal{C}$ commits to m_u for unknown $u \in \{0, 1\}$. Let $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big)$ be the resulting transcript. Additionally, $\mathcal{A}$ chooses random $u' \in \{0, 1\}$ and defines the polynomial vector $\mathbf{f}$, to be the unique such vector so that $\mathbf{f}(\boldsymbol{\alpha}) = \mathbf{a}$ and so that every coordinate of $\mathbf{f}$ has constant term $m_{u'}$.
- $\mathcal{A}$ chooses random $i \in [n]$ and random legal challenge vector $\tilde{\boldsymbol{\beta}}$ such that $\tilde{\beta}_i = \tilde{\alpha}_i$. It rewinds M back to the beginning of the right execution's query message and sends $\tilde{\boldsymbol{\beta}}$, receiving left query $\boldsymbol{\beta}$. If $\beta_{i'} = \alpha_{i'}$ for any $i' \in [n]$ then $\mathcal{A}$ aborts. If not, $\mathcal{A}$ responds with $\mathbf{b} = \mathbf{f}(\boldsymbol{\beta})$ receiving right response $\tilde{\mathbf{b}}$.
- $\mathcal{A}$ checks whether $\tilde{b}_i = \tilde{a}_i$. If so, $\mathcal{A}$ outputs u', if not $\mathcal{A}$ outputs $1 - u'$.

Just as in the proof of Claim 8, it suffices (by Fact 1) to construct an event $\mathbf{E}$ such that:

1. $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbf{E}\big) \geq \frac{\varepsilon' \delta' p}{16}$;
2. $\Pr\big(\tilde{b}_i = \tilde{a}_i \big| u' = u \;\&\; \mathbf{E}\big) \geq \frac{\varepsilon' \delta \delta' p^3}{16}$;
3. $\Pr\big(\tilde{b}_i = \tilde{a}_i \big| u' \neq u \;\&\; \mathbf{E}\big) \leq \frac{\varepsilon^*}{n \varepsilon' \delta p^2}$,

since $\varepsilon^* \leq n\varepsilon'(\varepsilon' \delta \delta' p^3)^2/2048$. Temporarily let $Z = \big\{\tilde{\alpha}_i \in \big[2^{\tilde{t}_i}\big] : \big|\mathsf{HON}^i(\tilde{\alpha}_i)\big| \leq \tau \big|R^i(\tilde{\alpha}_i)\big|\big\}$, where i is the index chosen by $\mathcal{A}$ and $\tau = \varepsilon' \delta \delta' p^3/16$. Define the event

$$\mathbf{E} : \text{"}\mathbb{T} \in \mathsf{TRB} \cap \mathsf{IND} \setminus \mathsf{FIXED} \;\&\; \mathcal{A} \text{ does not abort } \&\; \tilde{\alpha}_i \notin Z\text{."}$$

Note that

$$\begin{aligned}\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbf{E}) &\geq \Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\setminus\mathsf{FIXED}\ \&\ \mathcal{A}\text{ not abort}) - \Pr_{\mathbb{T}\in\mathsf{ACC}}(\tilde{\alpha}\in Z)\\ &\geq -\frac{\varepsilon'\delta' p}{16} + \Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\setminus\mathsf{FIXED})\cdot\\ &\cdot\ \Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}(\mathcal{A}\text{ not abort}\big|\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\setminus\mathsf{FIXED})\\ &\geq \frac{\delta' p}{8}\cdot\frac{1}{n}\cdot\varepsilon' n - \frac{\varepsilon'\delta' p}{16} = \frac{\varepsilon'\delta' p}{16},\end{aligned}$$

(the first line used Claim 4) by definition of IND (the $1/n$ appears because $\mathcal{A}$ must guess the right value of $i\in[n]$). Moreover, as $\tilde{\alpha}_i\notin Z$, if $u'=u$ then M answers $\tilde{\boldsymbol{\beta}}$ and correctly on the right with probability at least $\varepsilon'\delta\delta' p^3/16$, which means that the probability that $\tilde{b}_i=\tilde{a}_i$ given $u'=u$ & $\mathbf{E}$ is at least as high.

Finally, we bound $\Pr(\tilde{b}_i=\tilde{a}_i\big|u'\neq u\ \&\ \mathbf{E})$. It does not quite work to try to use Claim 2 directly to argue that M does not answer $\tilde{\beta}_i$ correctly if $u'\neq u$. This is because the answers M receives if $u'\neq u$ are randomly distributed (this is ensured by $\mathcal{A}$ aborting in case $\beta_{i'}=\alpha_{i'}$ for any i'), whereas the answers M receives to $\boldsymbol{\beta}$ from E are random only in the case that $\boldsymbol{\beta}$ differs in every coordinate from the $\boldsymbol{\alpha}$ asked in the main thread. For this reason, we must also use the fact that $\mathbb{T}\notin\mathsf{FIXED}$.

Consider now the interaction between M and E where the main thread E receives as input has **Com** as the commitment message but has unspecified query and response messages. By definition, if $\mathbf{Com}\notin\mathsf{FIXED}^{i'}$ for all i' (ensuring that the main thread E receives is not in FIXED), then for any $\gamma_{i'}\in\left[2^{t_{i'}}\right]$, $\Pr_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}(\beta_{i'}=\gamma_{i'})\leq\varepsilon$. It follows by the union bound that no matter what main thread left query $\boldsymbol{\gamma}$ occurs (we use $\boldsymbol{\gamma}$ so as not to be confused with the $\boldsymbol{\alpha}$ that was asked by M during its interaction with $\mathcal{A}$ above),

$$\Pr_{\tilde{\boldsymbol{\beta}}}(\beta_{i'}\neq\gamma_{i'}\ \forall\ i')\geq(1-n\varepsilon)\cdot\Pr_{\tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\beta}}\in\mathsf{HON})\geq n\varepsilon'\delta p^2$$

(assuming also that **Com** is such that the transcript is in USEFUL). So we see that if the transcript E receives as input is in $\mathsf{TRB}\setminus\mathsf{FIXED}$, then a good portion of the left queries which M asks during its interaction with E will not share any coordinate with the main thread query, and so M will be given truly random responses. If in addition, the transcript given to E is not in EXT then

$$\begin{aligned}\varepsilon^* &\geq \Pr_{\tilde{\boldsymbol{\beta}}}(\text{M answers }\tilde{\beta}_i\text{ correctly}\big|\text{E answers }\boldsymbol{\beta})\\ &\geq \Pr_{\tilde{\boldsymbol{\beta}}}(\text{M answers }\tilde{\beta}_i\text{ correctly}\big|\text{E answers }\boldsymbol{\beta}\ \&\ \beta_{i'}\neq\gamma_{i'}\ \forall\ i')\cdot\Pr_{\tilde{\boldsymbol{\beta}}}(\beta_{i'}\neq\gamma_{i'}\ \forall\ i')\\ &\geq (n\varepsilon'\delta p^2)\cdot\Pr_{\tilde{\boldsymbol{\beta}}}(\text{M answers }\tilde{\beta}_i\text{ correctly}\big|\boldsymbol{\beta}\text{ answered randomly}).\end{aligned}$$

And so we have

$$\begin{aligned}\Pr(\tilde{b}_i=\tilde{a}_i\big|u'\neq u\ \&\ \mathbf{E}) &= \Pr(\text{M answers }\tilde{\beta}_i\text{ corr.}\big|\boldsymbol{\beta}\text{ answered rand. }\&\ \mathbb{T}\in\mathsf{TRB}\setminus\mathsf{FIXED})\\ &\leq \frac{\varepsilon^*}{n\varepsilon'\delta p^2},\end{aligned}$$

completing the proof of Claim 9. □

Claims 5 through 9 combine to give that if Com is computationally hiding, then

$$\begin{aligned}\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}) &\leq \Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{UNBAL}) + \Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}\cap 1{-}2)\\ &+ \Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}) + \Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T}\in\mathsf{TRB}\setminus(\mathsf{UNBAL}\cup 1{-}2\cup\mathsf{IND}))\\ &\leq \frac{\delta' p}{4}+\frac{\delta' p}{4}+\frac{\delta' p}{4}+\frac{\delta' p}{4}=\delta' p,\end{aligned}$$

completing the proof of Lemma 1, Theorem 2 and Theorem 1.

6 Non-Malleability in Four-Rounds

6.1 Four-Round, Delayed-Input, Rewind-Secure Arguments of Knowledge

Our four-round non-malleable protocols in the next two sections make use of a four-round, delayed-input, rewind-secure zero-knowledge argument scheme for NP. The formal definition is as follows.

Definition 14. *We say that a 4-round, interactive argument of knowledge $\langle \mathrm{P}, \mathrm{V} \rangle$ for a language L is* delayed-input, rewind-secure zero-knowledge *if it satisfies the following extra properties:*

- **Delayed-Input.** *The only time the instance/witness pair (x, w) is used by P is during the fourth message; the only time x is used by V is during verification. In particular, the first three rounds can be played before (x, w) is known, or even defined.*
- **Rewind-Secure Zero-Knowledge.** *There exists a PPT simulator SIM satisfying the following syntax and security guarantees:*
  - **Syntax:** SIM *begins by taking 1^λ as input and using oracle access to V^*, outputs (π_1, π_2). Then at this point, SIM can take inputs of the form (x, π_3) where $x \in \{0,1\}^\lambda$, and output π_4, again using oracle access to V^*.*
  - **Security:** *Let $\mathcal{D}$ be a distribution which, given (π_1, π_2, π_3), outputs $(x, w) \in \mathcal{R}_\mathsf{L}$. For any PPT V^*, $\mathsf{SIM}_\mathcal{D}^{\mathrm{V}^*} \approx_\mathsf{c} \mathsf{REAL}_\mathcal{D}^{\mathrm{V}^*}$, where the distributions are as follows:*
    - $\mathsf{REAL}_\mathcal{D}^{\mathrm{V}^*}$**:** *$\mathrm{V}^*$ and P play to obtain (π_1, π_2); then V^* sends π_3 and $(x, w) \leftarrow \mathcal{D}(\pi_1, \pi_2, \pi_3)$ is drawn and P computes the proof that $x \in \mathsf{L}$ using witness w by sending π_4; then V^* rewinds P sending $\hat{\pi}_3$ and $(\hat{x}, \hat{w}) \leftarrow \mathcal{D}(\pi_1, \pi_2, \hat{\pi}_3)$ is drawn and P sends $\hat{\pi}_4$ proving $\hat{x} \in \mathsf{L}$ via witness $\hat{w}$; the tuple $(x, \hat{x}; \pi_1, \pi_2, \pi_3, \pi_4, \hat{\pi}_3, \hat{\pi}_4)$ is output.*
    - $\mathsf{SIM}^{\mathrm{V}^*}$**:** *Generates (π_1, π_2), then V^* computes two third messages π_3 and $\hat{\pi}_3$; then $(x, w) \leftarrow \mathcal{D}(\pi_1, \pi_2, \pi_3)$ and $(\hat{x}, \hat{w}) \leftarrow \mathcal{D}(\pi_1, \pi_2, \hat{\pi}_3)$ are drawn and SIM uses (π_3, x) and $(\hat{\pi}_3, \hat{x})$ to generate π_4 and $\hat{\pi}_4$; the tuple $(x, \hat{x}; \pi_1, \pi_2, \pi_3, \pi_4, \hat{\pi}_3, \hat{\pi}_4)$ is output.*

Very recently, Goyal and Richelson [GR19] constructed a three-round delayed-input rewind-secure witness-indistinguishable argument system for NP based on the assumption that one-to-one one-way functions exist. A four-round version of their protocol is also possible assuming only OWFs. Using the Feige-Shamir compiler, we can convert this scheme into a four-round delayed-input, rewind-secure zero-knowledge argument system based on one-way functions. We describe this in more detail in Appendix A.

6.2 Four-Round Non-Malleable Commitments

In this section we show how to squeeze our non-malleable protocol $\langle \mathrm{C}, \mathrm{R} \rangle$ into 4 rounds. The new protocol, denoted $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$, is the same as the old except that the zero-knowledge messages are lifted up and sent together with the commit, challenge and response messages. Moreover, we use the four-round, delayed-input, rewind-secure zero-knowledge argument of knowledge from the previous section.

Notation. Let Π be the four-round zero-knowledge argument from Section 6.1, and let Σ denote the first four rounds of $\langle \mathrm{C}, \mathrm{R} \rangle$ from Section 3. We will denote the messages of Π by $(\pi_1, \ldots, \pi_4)$ and of Σ by $(\sigma_1, \ldots, \sigma_4)$. So in the notation of Section 3, $(\sigma_2, \sigma_3, \sigma_4) = (\mathbf{Com}, \boldsymbol{\alpha}, \mathbf{a})$. When we want to consolidate, we write τ_i instead of (σ_i, π_i). Consider now a MIM M participating in two executions of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$, producing a transcript $\mathbb{T} = \big(\tau_1, \tau_2, \tau_3, \tau_4; \tilde{\tau}_1, \tilde{\tau}_2, \tilde{\tau}_3, \tilde{\tau}_4\big)$. The protocol is shown in Figure 4.

Input and Subroutines: Let Π denote the four-round, delayed-input, rewind-secure zero-knowledge argument of knowledge from Section 6.1. Let Σ denote the first four messages of $\langle \mathrm{C}, \mathrm{R} \rangle$. We denote by $(\Pi_1, \Pi_2, \Pi_3, \Pi_4)$ the subroutines used to generate the rounds of Π, and similarly for Σ. C uses input $m \in \mathbb{Z}_q$ for a large prime q.

Commit Phase:

1. $\mathrm{R} \to \mathrm{C}$: R sends $(\sigma_1, \pi_1) \leftarrow \Sigma_1 \times \Pi_1$ to C.
2. $\mathrm{C} \to \mathrm{R}$: C sends $(\sigma_2, \pi_2) \leftarrow \Sigma_2(m) \times \Pi_2$ to R.
3. $\mathrm{R} \to \mathrm{C}$: R sends $(\sigma_3, \pi_3) \leftarrow \Sigma_3 \times \Pi_3$ to C.
4. $\mathrm{C} \to \mathrm{R}$: C sends $(\sigma_4, \pi_4) \leftarrow \Sigma_4 \times \Pi_4$ to R.
   - $(\pi_1, \pi_2, \pi_3, \pi_4)$ proves the same statement as in $\langle \mathrm{C}, \mathrm{R} \rangle$.

Decommitment and Output: To decommit, C sends the decommitment information for all commitments in σ_2. If all decommitments are valid and if the verification of $(\pi_1, \pi_2, \pi_3, \pi_4)$ accepts, R outputs $m \in \mathbb{Z}_q$, otherwise $\perp$.

Figure 4: : 4-round non-malleable commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$.

Theorem 3. *If OWFs exist then $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ is a $4-$round statistically binding, non-malleable commitment scheme.*

Proof. Binding and hiding are proved the same as for $\langle \mathrm{C}, \mathrm{R} \rangle$ in Section 4. We prove non-malleability. Consider a MIM M participating in two executions of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$, producing a transcript $\mathbb{T} = \big(\tau_1, \tau_2, \tau_3, \tau_4; \tilde{\tau}_1, \tilde{\tau}_2, \tilde{\tau}_3, \tilde{\tau}_4\big)$. We consider the possible ways M can schedule the messages in $\mathbb{T}$. First note that if $\tilde{\tau}_3$ and $\tilde{\tau}_4$ appear one after another then $\tilde{m}$ can be trivially extracted using the knowledge extractor of Π. The second observation is that if the message $\tilde{\tau}_2$ appears before τ_2 then M is not mauling as m has not been specified at the time $\tilde{m}$ is fixed. One subtle point is that the right commitment is not exactly fixed at the time $\tilde{\tau}_2$ is sent, since M can still choose to abort (*i.e.*, commit to $\perp$ instead of $\tilde{m}$). However, commitments to $\perp$ can be efficiently recognized because the proof $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4)$ will not verify. This intuition is easily formalized into a reduction from an M who mauls and sends $\tilde{\tau}_2$ before receiving τ_2 to an adversary who breaks the hiding of the left execution of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$. We leave the details to the reader. Finally, the two observations: $\tilde{\tau}_3$ and $\tilde{\tau}_4$ cannot be consecutive and $\tilde{\tau}_2$ appears before τ_2 means that the scheduling must be synchronizing. Therefore, it suffices to prove non-malleability against the synchronizing MIM.

Our proof that $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ is non-malleable against a synchronizing adversary follows the same extraction paradigm as the proof of Theorem 1. Our extractor, much like before, rewinds M twice sending two new values of $(\tilde{\sigma}_3, \tilde{\pi}_3)$, receiving two new values of (σ_3, π_3), answering with two values of (σ_4, π_4) where the σ_4 are random, and receiving two new values of $(\tilde{\sigma}_4, \tilde{\pi}_4)$. In this way E obtains the data $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\big\}$, and succeeds in recovering $\tilde{m}$ whenever a collinearity check passes.

One important point is that because σ_4 is sent along with π_4, the final message of Π which proves correctness of σ_4, if E wants to send a random value instead of the correct σ_4, then E must send the final message of a simulated argument. This means that E will only succeed in extracting $\tilde{m}$ if M gives correct answers on the right with non-negligible probability when given random answers on the left and a simulated argument. The following definition characterizes the extractable transcripts. Recall we said that $\mathbb{T} \in \mathsf{ACC}$ if the proof $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4)$ verifies. We work with two non-negligible parameters $\varepsilon, \delta > 0$ which can both be defined as in Section 4 in terms of M's mauling advantage; we do not explicitly define them here, the important relationship is that $\varepsilon \ll \delta$.

Definition 15 (Extractable Transcripts). *Fix non-negligible $\varepsilon > 0$ and $\mathbb{T} \in \mathsf{ACC}$. We say that $\mathbb{T}$ is* extractable *and write $\mathbb{T} \in \mathsf{EXT}$ if there exists $i \in [n]$ such that*

$$\Pr\Big[\tilde{b}_i \text{ correct}\Big] \geq \varepsilon,$$

where the probability is over the experiment: random $\tilde{\boldsymbol{\beta}} \leftarrow \mathbb{Z}_q^n$ and $\tilde{\pi}_3$ are drawn and $(\tilde{\boldsymbol{\beta}}, \tilde{\pi}_3)$ is sent to M*, who returns $(\boldsymbol{\beta}, \pi_3)$; then random $\boldsymbol{b} \leftarrow \mathbb{Z}_q^n$ is drawn and returned to* M *along with a simulated proof π_4; then* M *outputs $(\tilde{\boldsymbol{b}}, \tilde{\pi}_4)$. By "correct" we mean that $\tilde{b}_i = \tilde{f}_i(\tilde{\beta}_i)$ where $\tilde{f}_i$ is the committed linear polynomial in $\tilde{\sigma}_2$.*

Recall we wrote $\mathbb{T} \in \mathsf{USEFUL}$ if $\Pr\big[\tilde{\mathbf{b}} \text{ correct}\big] \geq 2\delta$, where the probability is over the experiment: random $\tilde{\boldsymbol{\beta}} \leftarrow \mathbb{Z}_q^n$ and $\tilde{\pi}_3$ are drawn and $(\tilde{\boldsymbol{\beta}}, \tilde{\pi}_3)$ is sent to M, who returns $(\boldsymbol{\beta}, \pi_3)$; then $(\mathbf{b}, \pi_4)$ is returned to M where $\mathbf{b} = \mathbf{f}(\boldsymbol{\beta})$ and π_4 is an honest proof; then M outputs $(\tilde{\mathbf{b}}, \tilde{\pi}_4)$. Just as in Section 4, non-malleability of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ follows from the following bound:

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\Big[\mathbb{T} \in \mathsf{USEFUL} \setminus \mathsf{EXT}\Big] \leq p, \tag{1}$$

where p is some non-negligible quantity which is related to M's mauling probability. We establish (1) in two steps; our analysis is centered around the following definition.

Definition 16 (Half Useful Transcripts). *We say that $\mathbb{T} \in \mathsf{HALF.USEFUL}$ if*

$$\Pr\big[\tilde{\boldsymbol{b}} \text{ correct}\big] \geq \delta,$$

where the probability is over the experiment: random $\tilde{\boldsymbol{\beta}} \leftarrow \mathbb{Z}_q^n$ and $\tilde{\pi}_3$ are drawn and $(\tilde{\boldsymbol{\beta}}, \tilde{\pi}_3)$ is sent to M*, who returns $(\boldsymbol{\beta}, \pi_3)$; then $(\boldsymbol{b}, \pi_4)$ is returned to* M *where $\boldsymbol{b} = \boldsymbol{f}(\boldsymbol{\beta})$ and π_4 is a simulated proof; then* M *outputs $(\tilde{\boldsymbol{b}}, \tilde{\pi}_4)$.*

In words, $\mathbb{T} \in \mathsf{HALF.USEFUL}$ if M's chance of answering correctly on the right is non-negligible given that his queries are answered correctly but he is given a simulated proof. We describe such transcripts as "half useful" because they are half way between useful and extractable transcripts. The bound

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\Big[\mathbb{T} \in \mathsf{HALF.USEFUL} \setminus \mathsf{EXT}\Big] \leq p/2$$

follows from the argument used to prove Lemma 1. Recall that, intuitively, Lemma 1 says that if M is able to answer correctly on the right with high/low probability given that its queries on the left are answered correctly/randomly, then we can use M to break the hiding of $\langle \mathrm{C}, \mathrm{R} \rangle$. The same argument goes through here because M is given a simulated proof in both the experiments defining $\mathsf{HALF.USEFUL}$ and EXT. The following Claim thus completes the proof of non-malleability. □

Claim 10. *If* $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big[\mathbb{T} \in \mathsf{USEFUL} \setminus \mathsf{HALF.USEFUL}\big] > p/2$ *then there exists a PPT* $\mathcal{A}$ *who breaks the rewind-secure zero-knowledge of* Π.

Proof. We use M to construct an adversary $\mathcal{A}$ who wins the following game against challenger $\mathcal{C}$ with non-negligible advantage. This violates the rewind-secure zero-knowledge of Π.

1. $\mathcal{A}$ sends π_1 to $\mathcal{C}$, receives π_2; $\mathcal{A}$ sends (π_3, x, w) and $(\hat{\pi}_3, \hat{x}, \hat{w})$ to $\mathcal{C}$, where $(x, w), (\hat{x}, \hat{w}) \in R_{\mathsf{L}}$; $\mathcal{C}$ produces π_4 honestly using w and chooses a bit $b \leftarrow \{0, 1\}$ and returns $(\pi_4, \hat{\pi}_4)$ to $\mathcal{A}$, where $\hat{\pi}_4$ is prepared according to b as follows:
   - if $b = 0$, $\hat{\pi}_4$ is computed honestly using $\hat{w}$;
   - if $b = 1$, $\hat{\pi}_4$ is simulated: $\hat{\pi}_4 \leftarrow \mathsf{SIM}(\hat{x}, \pi_1, \pi_2, \hat{\pi}_3)$.
2. $\mathcal{A}$ outputs b' and wins if $b' = b$.

Before specifying how $\mathcal{A}$ works, we set some convenient shorthand. Let X and Y denote the random variables which instantiate M and play the first two rounds, obtaining $(\tilde{\sigma}_1, \tilde{\pi}_1; \sigma_1, \pi_1; \sigma_2, \pi_2; \tilde{\sigma}_2, \tilde{\pi}_2)$, then they choose random $\tilde{\sigma}_3 = \tilde{\boldsymbol{\alpha}} \leftarrow \mathbb{Z}_q^n$ and output the quantity $\Pr\big[\tilde{\mathbf{a}} \text{ correct}\big]$, where the probabilities for X (resp. Y) are over the experiment: draw $\tilde{\pi}_3$, send $(\tilde{\sigma}_3, \tilde{\pi}_3)$ to M, receive (σ_3, π_3); compute correct σ_4 and return (σ_4, π_4) where π_4 is an honest (resp. simulated) proof; receive $(\tilde{\sigma}_4, \tilde{\pi}_4)$ where $\tilde{\sigma}_4 = \tilde{\mathbf{a}}$. Notice that $\mathbb{T} \in \mathsf{USEFUL}$ iff $\mathsf{X} \geq 2\delta$, while $\mathbb{T} \notin \mathsf{HALF.USEFUL}$ iff $\mathsf{Y} < \delta$. Assuming the Claim hypothesis, we get

$$\begin{aligned} p/2 &< \Pr_{\mathbb{T} \in \mathsf{ACC}}\Big[\mathbb{T} \in \mathsf{USEFUL} \setminus \mathsf{HALF.USEFUL}\Big] \leq p^{-1} \cdot \Pr_{\mathbb{T}}\Big[\big|\mathsf{X} - \mathsf{Y}\big| > \delta\Big] \\ &\leq \frac{1}{p\delta^2} \cdot \mathbb{E}_{\mathbb{T}}\Big[\mathsf{X}^2 + \mathsf{Y}^2 - 2\mathsf{XY}\Big]. \end{aligned}$$

We have used that $\Pr_{\mathbb{T}}\big[\mathbb{T} \in \mathsf{ACC}\big] \geq p$ and Markov's inequality. It follows that either $\mathbb{E}\big[\mathsf{X}^2 - \mathsf{XY}\big]$ or $\mathbb{E}\big[\mathsf{Y}^2 - \mathsf{XY}\big]$ is at least $(p\delta)^2/4$; we assume the former is the case, the proof in case the latter holds is similar. Now we can define $\mathcal{A}$; it works as follows.

1. $\mathcal{A}$ sends $(\tilde{\sigma}_1, \tilde{\pi}_1)$ to M, receives (σ_1, π_1) and sends π_1 to $\mathcal{C}$, receives π_2; $\mathcal{A}$ draws $\sigma_2 \leftarrow \Sigma_2(m)$ and sends (σ_2, π_2) to M, receives $(\tilde{\sigma}_2, \tilde{\pi}_2)$; $\mathcal{A}$ draws random $\tilde{\boldsymbol{\alpha}} \leftarrow \mathbb{Z}_q^n$ and returns $(\tilde{\sigma}_3, \tilde{\pi}_3)$ where $\tilde{\sigma}_3 = \tilde{\boldsymbol{\alpha}}$, receives (σ_3, π_3); $\mathcal{A}$ computes σ_4 and sends (π_3, x, w) to $\mathcal{C}$ where (x, w) is the statement/witness pair proving correctness of σ_4, $\mathcal{C}$ returns π_4 and $\mathcal{A}$ sends (σ_4, π_4) to M receiving $(\tilde{\sigma}_4, \tilde{\pi}_4)$ where $\tilde{\sigma}_4 = \tilde{\mathbf{a}}$. If the proof $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4)$ does not verify, $\mathcal{A}$ halts outputting a random $j' \in [N]$.
2. Now $\mathcal{A}$ rewinds M and draws a new $\tilde{\pi}_3$ and sends $(\tilde{\boldsymbol{\alpha}}, \tilde{\pi}_3)$ to M ($\tilde{\boldsymbol{\alpha}}$ reused from Step 1), receives $(\hat{\sigma}_3, \hat{\pi}_3)$; $\mathcal{A}$ computes $\hat{\sigma}_4$ and sends $(\hat{\pi}_3, \hat{x}, \hat{w})$ to $\mathcal{C}$ where $(\hat{x}, \hat{w})$ is the statement/witness pair proving correctness of $\hat{\sigma}_4$; $\mathcal{A}$ receives $\hat{\pi}_4$ from $\mathcal{C}$ and sends $(\hat{\sigma}_4, \hat{\pi}_4)$ to M, and receives $\tilde{\mathbf{b}}$ and another $\tilde{\pi}_4$.
3. If $\tilde{\mathbf{b}} = \tilde{\mathbf{a}}$, $\mathcal{A}$ outputs 0, otherwise $\mathcal{A}$ outputs a random bit.

Analysis of $\mathcal{A}$. If the transcript $\mathbb{T}$ generated in Step 1 is such that the proof $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4)$ verifies, then $\tilde{\mathbf{a}}$ is correct with overwhelming probability by the soundness of Π. Thus, if $\mathcal{A}$ makes it to Step 2, then $\tilde{\mathbf{b}}$ is correct if and only if $\tilde{\mathbf{b}} = \tilde{\mathbf{a}}$. A simple calculation shows

$$\Pr\big[\mathcal{A} \text{ wins}\big] = \frac{1}{2} + \frac{1}{4} \cdot \mathbb{E}\Big[\mathsf{X}^2 - \mathsf{XY}\Big].$$

The result follows. □

6.3 Four-Round Non-Malleable Zero-Knowledge

Using the commitment scheme from the previous section, we can construct a simple 4-round non-malleable zero knowledge argument $\langle \mathrm{P}, \mathrm{V} \rangle$ for any language $L \in \mathcal{NP}$. Our scheme appears in Figure 5.

Input and Subroutines: Let Π denote the four-round, delayed-input, rewind-secure zero-knowledge argument of knowledge from Section 6.1. Let $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ be the four-round non-malleable commitment scheme of Section 6.2. We denote by $(\Pi_1, \Pi_2, \Pi_3, \Pi_4)$ the subroutines used to generate the rounds of Π, and similarly $(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4)$ the subroutines used to generate the messages of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$. P and V both get a statement $s \in \{0,1\}^\lambda$ as input; P also takes a witness $w \in \{0,1\}^{\mathsf{poly}(\lambda)}$ such that $(x, w) \in R_{\mathsf{L}}$ as input.

Protocol:

1. $\mathrm{V} \to \mathrm{P}$: V sends $(\sigma_1, \pi_1) \leftarrow \Sigma_1 \times \Pi_1$ to P.
2. $\mathrm{P} \to \mathrm{V}$: P sends $(\sigma_2, \pi_2) \leftarrow \Sigma_2(w) \times \Pi_2$ to V.
3. $\mathrm{V} \to \mathrm{P}$: V sends $(\sigma_3, \pi_3) \leftarrow \Sigma_3 \times \Pi_3$ to P.
4. $\mathrm{P} \to \mathrm{V}$: C sends $(\sigma_4, \pi_4) \leftarrow \Sigma_4 \times \Pi_4$ to R, where $(\pi_1, \pi_2, \pi_3, \pi_4)$ proves the statement:

   $\exists w \in \{0,1\}^{\mathsf{poly}(\lambda)}$ st $(x, w) \in R_{\mathsf{L}}$ and $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ is a commitment to w using $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$.

Verification: V simply verifies the proof $(\pi_1, \pi_2, \pi_3, \pi_4)$.

Figure 5: 4-round non malleable zero-knowledge argument of knowledge $\langle \mathrm{P}, \mathrm{V} \rangle$.

Theorem 4. *If OWFs exist then $\langle \mathrm{P}, \mathrm{V} \rangle$ is a $4-$round non-malleable zero knowledge argument of knowledge for any $L \in \mathcal{NP}$.*

Proof. Zero-knowledge and soundness follow by the ZK and soundness of π and the hiding and binding of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$. Suppose M is a MIM adversary who plays two executions of $\langle \mathrm{P}, \mathrm{V} \rangle$. We denote by $\mathbb{T}$ the transcript of M's interaction:

$$\mathbb{T} = \big(\tau_1, \tau_2, \tau_3, \tau_4, \tilde{\tau}_1, \tilde{\tau}_2, \tilde{\tau}_3, \tilde{\tau}_4\big),$$

where $\tau_i = (\sigma_i, \pi_i)$ and similarly for $\tilde{\tau}_i$. The simulation extractor $\mathsf{SIM.Ext}$ works as follows:

Input and Subroutines: $\mathsf{SIM.Ext}$ takes a statement $x \in \{0,1\}^\lambda$ as input and gets oracle access to M and outputs a tuple $(\tilde{x}, \tilde{w}, \mathbb{T})$. $\mathsf{SIM.Ext}$ will use the standard zero-knowledge simulator SIM.

1. Generate $\mathbb{T}$: $\mathsf{SIM.Ext}$ generates $\mathbb{T}$ by running two executions of $\langle \mathrm{P}, \mathrm{V} \rangle$ with M, playing honestly as V on the right and generating messages on the left as follows:

- $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ is an honest execution of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ where the committed message is the all-zero string;
- $(\pi_1, \pi_2, \pi_3, \pi_4) \leftarrow \mathsf{SIM}^{\mathrm{M}}(x)$ is a simulated transcript.

Let $\tilde{x}$ be the statement proved in the right execution. If the right proof $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4)$ does not verify, $\mathsf{SIM.Ext}$ halts and outputs $(\tilde{x}, \perp, \mathbb{T})$.

2. Extract $\hat{w}$: Now $\mathsf{SIM.Ext}$ extracts the committed value inside $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3, \tilde{\sigma}_4)$ by running the following extraction procedure sufficiently (polynomially) many times.

- rewind M, draw $\tilde{\boldsymbol{\beta}} \leftarrow \mathbb{Z}_q^n$ and a new $\tilde{\pi}_3$, send $(\tilde{\boldsymbol{\beta}}, \tilde{\pi}_3)$ to M, receive $(\hat{\sigma}_3, \hat{\pi}_3)$; send $(\hat{\sigma}_4, \hat{\pi}_4)$ where $\hat{\sigma}_4 \leftarrow \mathbb{Z}_q^n$ is random and $\hat{\pi}_4 \leftarrow \mathsf{SIM}^{\mathrm{M}}(\hat{x}, \pi_1, \pi_2, \hat{\pi}_3)$ where $\hat{x}$ is the statement proving correctness of $\hat{\sigma}_4$, receive $\tilde{\mathbf{b}}$.
- repeat the above with a new random $\tilde{\boldsymbol{\gamma}} \leftarrow \mathbb{Z}_q^n$, obtaining $\tilde{\mathbf{c}}$; collect the quantities $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\big\}$ where $(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}})$ is from $\mathbb{T}$.
- if there exists $i \in [n]$ such that $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\big\}$ are collinear, let $\tilde{w}$ be the $y-$intercept of the line they span and break out of the extraction loop.

Output: Output $(\tilde{x}, \tilde{w}, \mathbb{T})$.

It is clear that $\mathsf{SIM.Ext}$ satisfies the syntax requirements of Definition 5. The expected time spent each time through the loop is $\mathsf{poly}(\lambda, \mathsf{T}_{\mathrm{M}})$, where T_{M} denotes the expected runtime of M. The loop is run a polynomial (in λ) number of times which depends on M's mauling probability. Thus, the running time requirement of $\mathsf{SIM.Ext}$ is also met. The security property follows immediately from the zero-knowledge of Π, so it remains to establish the extraction guarantee. The argument is basically the same as the proof of non-malleability in Section 6.2, we describe it again briefly and informally. For this discussion, we fix two non-negligible parameters $\varepsilon, \delta > 0$; the reader should keep the relationship $\varepsilon \ll \delta$ in mind.

To begin with, consider a modified version of $\mathsf{SIM.Ext}$ where during the generation step, the messages $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ are a commitment to a valid witness w (instead of to 0^λ), and where the messages $(\hat{\sigma}_4, \hat{\pi}_4)$ send during extraction are such that $\hat{\sigma}_4$ is correct (instead of random) and $\hat{\pi}_4$ is an honest (instead of simulated) proof. In this case, extraction will succeed in recovering $\hat{w}$ (the committed value in the right execution) whenever there exists $i \in [n]$ such that $\Pr\big[\tilde{b}_i \text{ correct}\big] \geq \varepsilon$, where the probability is over the random choices made by this modified $\mathsf{SIM.Ext}$. Moreover, conditioned on $\mathbb{T} \in \mathsf{ACC}$ (*i.e.*, the proof $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4)$ being valid), the bound $\Pr\big[\tilde{\mathbf{b}} \text{ correct}\big] \geq 2\delta \gg \varepsilon$ holds with high probability (*i.e.*, $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big[\mathbb{T} \notin \mathsf{USEFUL}\big]$ is small). It follows that when $\mathbb{T} \in \mathsf{ACC}$, $\mathsf{SIM.Ext}$ extracts the committed value $\hat{w}$, and by the soundness of Π, $(\hat{x}, \hat{w}) \in R_{\mathsf{L}}$. We now proceed to change the description of $\mathsf{SIM.Ext}$ until it matches the program described above; as we make changes, we maintain the invariant that there exists $i \in [n]$ such that $\Pr\big[\tilde{b}_i \text{ correct}\big] \geq \varepsilon$ (which ensures that $\mathsf{SIM.Ext}$ extracts the committed value $\hat{w}$ in $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3, \tilde{\sigma}_4)$) and $(\hat{x}, \hat{w}) \in R_{\mathsf{L}}$.

Now consider a modification to $\mathsf{SIM.Ext}$ where the messages $(\hat{\sigma}_4, \hat{\pi}_4)$ send during extraction are such that $\hat{\sigma}_4$ is correct, but $\hat{\pi}_4$ is simulated. The same argument used to prove Claim 10 (that $\Pr_{\mathbb{T} \in \mathsf{ACC}}\left[\mathbb{T} \in \mathsf{USEFUL} \setminus \mathsf{HALF.USEFUL}\right]$ is small) applies here to show $\Pr\left[\tilde{\mathbf{b}} \text{ correct}\right] \geq \delta \gg \varepsilon$ and $(\hat{x}, \hat{w}) \in R_{\mathsf{L}}$ hold.

Now change $\mathsf{SIM.Ext}$ so that $\hat{\sigma}_4$ is random and $\hat{\pi}_4$ is simulated. The argument used to prove Lemma 1 shows that there exists $i \in [n]$ such that $\Pr\left[\tilde{b}_i \text{ correct}\right] > \varepsilon$ and $(\hat{x}, \hat{w}) \in R_{\mathsf{L}}$.

Finally, change $\mathsf{SIM.Ext}$ so that $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ is a commitment to 0^λ instead of to w. Now $\mathsf{SIM.Ext}$ is exactly the program described above. The non-malleability of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ implies that the success probability of extracting $\hat{w}$ such that $(\hat{x}, \hat{w}) \in R_{\mathsf{L}}$ cannot change, except by a negligible amount. This completes the proof. □

Acknowledgements

We would like to thank the anonymous FOCS referees for very helpful comments and suggestions to earlier versions of this paper. Also thanks to Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi and Ivan Visconti for pointing out an error in Section 6 in a prior version of this paper. In particular, we thank them for pointing out that our 4-round construction can only satisfy a weak form of concurrent non-malleability: namely concurrent non-malleability against non-aborting adversaries. Their subsequent work [COSV17] fills this gap by constructing a four-round concurrent non-malleable commitment scheme.

References

[Bar02] Boaz Barak. Constant-Round Coin-Tossing with a Man in the Middle or Realizing the Shared Random String Model. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, FOCS '02, pages 345–355, 2002.

[BGJ+18] Saikrishna Badrinarayanan, Vipul Goyal, Abhishek Jain, Yael Tauman Kalai, Dakshita Khurana, and Amit Sahai. Promise zero knowledge and its applications to round optimal MPC. In *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part II*, pages 459–487, 2018.

[BGR+15] Hai Brenner, Vipul Goyal, Silas Richelson, Alon Rosen, and Margarita Vald. Fast non-malleable commitments. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12-6, 2015*, pages 1048–1057, 2015.

[CGMO09] Nishanth Chandran, Vipul Goyal, Ryan Moriarty, and Rafail Ostrovsky. Position based cryptography. In Shai Halevi, editor, *CRYPTO*, volume 5677 of *Lecture Notes in Computer Science*, pages 391–407. Springer, 2009.

[CLOS02] Ran Canetti, Yehuda Lindell, Rafail Ostrovsky, and Amit Sahai. Universally composable two-party and multi-party secure computation. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, STOC '02, pages 494–503, 2002.

[COSV16] Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi, and Ivan Visconti. Concurrent non-malleable commitments (and more) in 3 rounds. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*, pages 270–299, 2016.

[COSV17] Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi, and Ivan Visconti. Four-round concurrent non-malleable commitments from one-way functions. In *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part II*, pages 127–157, 2017.

[DDN91] Danny Dolev, Cynthia Dwork, and Moni Naor. Non-Malleable Cryptography (Extended Abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, STOC '91, pages 542–552, 1991.

[FS90] Uriel Feige and Adi Shamir. Witness indistinguishable and witness hiding protocols. In *STOC*, pages 416–426. ACM, 1990.

[GKS16] Vipul Goyal, Dakshita Khurana, and Amit Sahai. Breaking the three round barrier for non-malleable commitments. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 21–30, 2016.

[GLOV12] Vipul Goyal, Chen-Kuei Lee, Rafail Ostrovsky, and Ivan Visconti. Constructing non-malleable commitments: A black-box approach. In *FOCS*, pages 51–60. IEEE Computer Society, 2012.

[Goy11] Vipul Goyal. Constant Round Non-malleable Protocols Using One-way Functions. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, STOC '11, pages 695–704. ACM, 2011.

[GPR16] Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In *Proceedings of ACM Symposium on Theory of Computing*, STOC '16, 2016.

[GR19] Vipul Goyal and Silas Richelson. Non-malleable commitments using goldreich-levin list decoding. *IACR Cryptology ePrint Archive*, 2019:1195, 2019.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A Pseudorandom Generator from any One-way Function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[IKOS07] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Zero-knowledge from Secure Multiparty Computation. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, STOC '07, pages 21–30, 2007.

[Khu17] Dakshita Khurana. Round optimal concurrent non-malleability from polynomial hardness. In *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part II*, pages 139–171, 2017.

[KOS03] Jonathan Katz, Rafail Ostrovsky, and Adam Smith. Round Efficiency of Multi-party Computation with a Dishonest Majority. In *Advances in Cryptology — EUROCRYPT '03*, volume 2656 of *Lecture Notes in Computer Science*, pages 578–595. Springer, 2003.

[KS17] Dakshita Khurana and Amit Sahai. How to achieve non-malleability in one or two rounds. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 564–575, 2017.

[LP11] Huijia Lin and Rafael Pass. Constant-round Non-malleable Commitments from Any One-way Function. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, STOC '11, pages 705–714, 2011.

[LPS17] Huijia Lin, Rafael Pass, and Pratik Soni. Two-round and non-interactive concurrent non-malleable commitments from time-lock puzzles. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 576–587, 2017.

[LPV08] Huijia Lin, Rafael Pass, and Muthuramakrishnan Venkitasubramaniam. Concurrent Non-malleable Commitments from Any One-Way Function. In *Theory of Cryptography, 5th Theory of Cryptography Conference, TCC 2008*, pages 571–588, 2008.

[LPV09] Huijia Lin, Rafael Pass, and Muthuramakrishnan Venkitasubramaniam. A Unified Framework for Concurrent Security: Universal Composability from Stand-alone Non-malleability. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, STOC '09, pages 179–188, 2009.

[Nao91] Moni Naor. Bit Commitment Using Pseudorandomness. *J. Cryptology*, 4(2):151–158, 1991.

[NSS06] Moni Naor, Gil Segev, and Adam Smith. Tight bounds for unconditional authentication protocols in the manual channel and shared key models. In Cynthia Dwork, editor, *CRYPTO*, volume 4117 of *Lecture Notes in Computer Science*, pages 214–231. Springer, 2006.

[Pas04] Rafael Pass. Bounded-Concurrent Secure Multi-Party Computation with a Dishonest Majority. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, STOC '04, pages 232–241, 2004.

[Pas13] Rafael Pass. Unprovable security of perfect NIZK and non-interactive non-malleable commitments. In *Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings*, pages 334–354, 2013.

[PR05] Rafael Pass and Alon Rosen. New and improved constructions of non-malleable cryptographic protocols. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, STOC '05, pages 533–542, 2005.

[PW10] Rafael Pass and Hoeteck Wee. Constant-Round Non-malleable Commitments from Sub-exponential One-Way Functions. In *Advances in Cryptology — EUROCRYPT '10*, pages 638–655, 2010.

[Sha79] Adi Shamir. How to Share a Secret. *Commun. ACM*, 22(11):612–613, 1979.

[Wee10] Hoeteck Wee. Black-Box, Round-Efficient Secure Computation via Non-malleability Amplification. In *Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, 2010.

A Four-Round Delayed-Input, Rewind-Secure, Zero-Knowledge

"Delayed-input" and "rewind-secure" are two extra properties for a zero-knowledge AoK, that are useful for designing larger protocols. It is known how to build ZK arguments with either of these properties, however the four-round non-malleable protocols in Section 6 required a ZK AoK with both properties together. Very recently, a three-round WI argument with both properties was constructed in [GR19] based on any one-to-one one-way function. In this section, we describe how to modify their construction to get a four-round, delayed-input, rewind-secure ZK AoK from any one-way function. We begin with a general "WI to ZK" compiler for four-round, delayed-input, rewind-secure AoKs, then we modify the construction of [GR19] to get the needed WI protocol.

A.1 Rewind-Secure ZK from Rewind-Secure WI

Definition 17 (**Four-Round, Delayed-Input, Rewind-Secure WI AoK** [BGJ+18])**.** *We say that a* 4*-round, interactive argument system* $\langle \mathrm{P}, \mathrm{V} \rangle$*, for a language* L *is* delayed-input, rewind-secure witness-indistinguishable *if it satisfies the delayed-input completeness and soundness properties of Definition 18 as well as the following:*

- **Rewind-Secure Witness-Indistinguishability.** *for every PPT* V^**, it holds that*

$$\mathsf{REAL}_0^{V^*}(1^\lambda) \approx_{\mathsf{c}} \mathsf{REAL}_1^{V^*}(1^\lambda)$$

  where for $b \in \{0,1\}$ *the random variable* $\mathsf{REAL}_b^{V^*}(1^\lambda)$ *is defined via the following experiment:*

  - V^* *and* P *begin playing honestly,* V^* *sends* π_1*,* P *returns* π_2*, then* V^* *sends* (π_3, x, w) *and* $(\hat{\pi}_3, \hat{x}, \hat{w}_0, \hat{w}_1)$ *such that* $(x,w), (\hat{x}, \hat{w}_0), (\hat{x}, \hat{w}_1) \in R_{\mathsf{L}}$*, and* P *returns* π_4 *and* $\hat{\pi}_4$ *computed honestly using* (x,w) *and* $(\hat{x}, \hat{w}_b)$*, respectively. The tuple* $(x, \hat{x}; \pi_1, \pi_2, \pi_3, \pi_4, \hat{\pi}_3, \hat{\pi}_4)$ *is output.*

We give a construction of four-round delayed-input, rewind-secure zero-knowledge AoK using a four-round rewind-secure WI AoK. This construction is similar to the "WI to ZK" compiler of Feige and Shamir [FS90]. The idea is to have V prove to P in the first three rounds that it knows a OWF preimage; then have P use the delayed-input rewind-secure WI to prove the statement: "either $x \in L$ or I know the OWF preimage". The zero-knowledge simulator first extracts the OWF preimage and then completes the delayed-input WI protocol using the "I know the OWF preimage" part of the witness; rewind security of the ZK protocol follows from rewind security of the WI protocol.

Claim 11. *Assume* f *is one-way and that* Σ *is a four-round, delayed-input, rewind-secure WI AoK. Then the protocol* Π *in Figure 6 is four-round, delayed-input, rewind secure, ZK AoK.*

Parameters and Subroutines: Let λ be the security parameter, Σ be a four-round delayed-input rewind-secure WI AoK. We denote the subroutines which generate the messages in the rounds as $(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4)$. Let $f : \{0,1\}^\lambda \to \{0,1\}^{\mathsf{poly}(\lambda)}$ be a one-way function.

Protocol:

1. $\mathrm{V} \to \mathrm{P}$**:** Choose 2λ random strings $r_{0,i}, r_{1,i} \leftarrow \{0,1\}^\lambda$ for $i \in [\lambda]$ and set $s_{b,i} = f(r_{b,i})$ for all b, i. Also, draw $\sigma_1 \leftarrow \Sigma_1$ and send $\left(\{s_{0,i}, s_{1,i}\}_{i\in[n]}; \sigma_1\right)$ to P.
2. $\mathrm{P} \to \mathrm{V}$**:** Draw $c \leftarrow \{0,1\}^\lambda$ and $\sigma_2 \leftarrow \Sigma_2$ and send (c, σ_2) to V.
3. $\mathrm{V} \to \mathrm{P}$**:** Draw $\sigma_3 \leftarrow \Sigma_3$ and send $\left(\{r_{c_i,i}\}_{i\in[\lambda]}, \sigma_3\right)$ to P.

Receive Inputs: Now both P and V receive an instance $x \in \{0,1\}^\lambda$, and additionally P receives witness w such that $(x, w) \in R_{\mathsf{L}}$.

4. $\mathrm{P} \to \mathrm{V}$**:** For each $i \in [\lambda]$ check that $s_{c_i,i} = f(r_{c_i,i})$; if not abort. Otherwise, draw $\sigma_4 \leftarrow \Sigma_4(x, w)$ where x is the statement:

$$\text{EITHER } \exists\ (i, r_0, r_1) \text{ st } (s_{0,i}, s_{1,i}) = \big(f(r_0), f(r_1)\big) \text{ OR } \exists\ w \text{ st } (x, w) \in R_{\mathsf{L}}.$$

   Send σ_4 to V.

Verification: V verifies $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ and outputs the result.

Figure 6: Four-Round Delayed-Input Rewind-Secure Zero-Knowledge.

Proof. Note Π is delayed-input since Σ is. Similarly, Π is an AoK since Σ is: the extractor for Π simply runs the extractor for Σ. Note that if P completes Π with V's verification accepting, then since Σ is AoK, the extractor recovers (with overwhelming probability), a witness for the statement

$$\text{EITHER } \exists\ (i, r_0, r_1) \text{ st } (s_{0,i}, s_{1,i}) = \big(f(r_0), f(r_1)\big) \text{ OR } \exists\ w \text{ st } (x, w) \in R_{\mathsf{L}}.$$

If the witness has the form (i, r_0, r_1) with noticeable probability, then P can be used to invert f. Thus, the extractor outputs w such that $(x, w) \in R_{\mathsf{L}}$ with probability $1 - \mathsf{negl}(\lambda)$. We complete the proof by constructing a simulator SIM. SIM works as follows.

1. SIM plays honestly against V^* through the first three rounds. Specifically:
   - SIM receives $\left(\{s_{0,i}, s_{1,i}\}_{i\in[\lambda]}, \sigma_1\right)$ from V^*;
   - SIM chooses a random $c \leftarrow \{0,1\}^\lambda$ and $\sigma_2 \leftarrow \Sigma_2$ and sends (c, σ_2) to V^*;
   - SIM receives $\left(\{r_{c_i,i}\}_{i\in[\lambda]}, \sigma_3\right)$ from V^*; if $s_{c_i,i} \neq f(r_{c_i,i})$ holds for some i, SIM halts and outputs the transcript so far.
2. SIM repeats the following loop indefinitely:

- $\cdot$ SIM rewinds V^* and draws another $\hat{c} \leftarrow \{0,1\}^\lambda$ and $\hat{\sigma}_2 \leftarrow \Sigma_2$ and sends $(\hat{c}, \hat{\sigma}_2)$ to V^*;
- $\cdot$ upon receiving $\{r_{\hat{c}_i,i}\}_{i\in[\lambda]}$, if there exists $i \in [\lambda]$ such that $\hat{c}_i \neq c_i$ and $s_{\hat{c}_i,i} = f(r_{\hat{c}_i,i})$, then SIM saves the trapdoor witness $(i, r_{0,i}, r_{1,i})$ and breaks out of the loop, proceeding to Step 3. If not, SIM continues back to the beginning of the loop and tries again.

3. Now SIM receives the statement x and draws $\sigma_4 \leftarrow \Sigma_4$, a proof of the statement

$$\text{EITHER } \exists\ (i, r_0, r_1) \text{ st } (s_{0,i}, s_{1,i}) = \big(f(r_0), f(r_1)\big) \text{ OR } \exists\ w \text{ st } (x, w) \in R_{\mathsf{L}},$$

computed with witness $(i, r_{0,i}, r_{1,i})$.

It is clear that SIM has the required syntax and running time (since the expected number of times SIM repeats the loop in Step 2 is $\mathsf{poly}(\lambda)/\mathsf{p}$ where p is the probability that SIM completes Step 1 without aborting). For the security requirement, let $\mathcal{D}$ be a distribution which, given (π_1, π_2, π_3), outputs $(x, w) \in \mathsf{L}$. The difference between $\mathsf{REAL}_{\mathcal{D}}^{\mathrm{V}^*}$ and $\mathsf{SIM}_{\mathcal{D}}^{\mathrm{V}^*}$ – recall both distributions output $(x, \hat{x}; \pi_1, \pi_2, \pi_3, \hat{\pi}_3, \pi_4, \hat{\pi}_4)$ – is that in the real distribution, the "honest" witnesses to $x, \hat{x} \in \mathsf{L}$ are used by P to generate π_4 and $\hat{\pi}_4$, while in the simulated distribution, the "trapdoor" witness (i, r_0, r_1) is used in both π_4 and $\hat{\pi}_4$. Thus, a distinguisher for $\mathsf{REAL}_{\mathcal{D}}^{\mathrm{V}^*}$ and $\mathsf{SIM}_{\mathcal{D}}^{\mathrm{V}^*}$ can be used to break the rewind-secure WI of Σ. □

Definition 18. *We say that a 4-round, interactive argument of knowledge $\langle \mathrm{P}, \mathrm{V}\rangle$ for a language L is* delayed-input, rewind-secure zero-knowledge *if it satisfies the following extra properties:*

- **Delayed-Input.** *The only time the instance/witness pair (x, w) is used by* P *is during the fourth message; the only time x is used by* V *is during verification. In particular, the first three rounds can be played before (x, w) is known, or even defined.*
- **Rewind-Secure Zero-Knowledge.** *There exists a PPT simulator* SIM *satisfying the following syntax and security guarantees:*
  - $\cdot$ **Syntax:** SIM *begins by taking 1^λ as input and using oracle access to V^*, outputs (π_1, π_2). Then at this point,* SIM *can take inputs of the form (x, π_3) where $x \in \{0,1\}^\lambda$, and output π_4, again using oracle access to V^*.*
  - $\cdot$ **Security:** *Let $\mathcal{D}$ be a distribution which, given (π_1, π_2, π_3), outputs $(x, w) \in \mathcal{R}_{\mathsf{L}}$. For any PPT V^*, $\mathsf{SIM}_{\mathcal{D}}^{\mathrm{V}^*} \approx_{\mathsf{c}} \mathsf{REAL}_{\mathcal{D}}^{\mathrm{V}^*}$, where the distributions are as follows:*
    - $\cdot$ $\mathsf{REAL}_{\mathcal{D}}^{\mathrm{V}^*}$**:** *$\mathrm{V}^*$ and* P *play to obtain (π_1, π_2); then V^* sends π_3 and $(x, w) \leftarrow \mathcal{D}(\pi_1, \pi_2, \pi_3)$ is drawn and* P *computes the proof that $x \in \mathsf{L}$ using witness w by sending π_4; then V^* rewinds* P *sending $\hat{\pi}_3$ and $(\hat{x}, \hat{w}) \leftarrow \mathcal{D}(\pi_1, \pi_2, \hat{\pi}_3)$ is drawn and* P *sends $\hat{\pi}_4$ proving $\hat{x} \in \mathsf{L}$ via witness $\hat{w}$; the tuple $(x, \hat{x}; \pi_1, \pi_2, \pi_3, \pi_4, \hat{\pi}_3, \hat{\pi}_4)$ is output.*
    - $\cdot$ $\mathsf{SIM}^{\mathrm{V}^*}$**:** *Generates (π_1, π_2), then V^* computes two third messages π_3 and $\hat{\pi}_3$; then $(x, w) \leftarrow \mathcal{D}(\pi_1, \pi_2, \pi_3)$ and $(\hat{x}, \hat{w}) \leftarrow \mathcal{D}(\pi_1, \pi_2, \hat{\pi}_3)$ are drawn and* SIM *uses (π_3, x) and $(\hat{\pi}_3, \hat{x})$ to generate π_4 and $\hat{\pi}_4$; the tuple $(x, \hat{x}; \pi_1, \pi_2, \pi_3, \pi_4, \hat{\pi}_3, \hat{\pi}_4)$ is output.*

A.2 Four Round, Delayed-Input, Rewind-Secure WI AoK from OWF

Very recently, [GR19] gave a three-round, delayed-input, rewind-secure WI argument based on any one-to-one one-way function. In this section, we slightly modify their protocol to obtain a four-round, delayed-input, rewind-secure WI AoK from any OWF. We refer the reader to [GR19] and

the references therein for more details. The original version of this paper made use of a ZK protocol with weaker properties which still sufficed for the proofs in Section 6; we have rewritten this section to build on the work of [GR19] in order to make for a clearer and simpler presentation.

High Level Intuition. We begin with an oversimplified discussion where we present the intuition; the full protocol is in the next section. Roughly speaking, our four-round, delayed-input, rewind-secure WI AoK consists of λ (security parameter) parallel repetitions of the main subprotocol. For this discussion, think of the main subprotocol as consisting of two four-round argument systems (the reality is a bit more complex): 1) a four-round, delayed-input WI AoK; 2) a four-round, rewind-secure ZK argument with constant soundness. We denote these protocols by Σ and Π, respectively, and their messages by $(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ and $(\pi_1, \pi_2, \pi_3, \pi_4)$. The main subprotocol begins by playing both protocols Σ and Π for two rounds: $(\sigma_1, \sigma_2), (\pi_1, \pi_2)$. Then in the third round, V chooses one of them at random to continue (the other is abandoned). The protocol Π is used to prove that various other aspects (undisccussed so far) of the main subprotocol are set up correctly (so the statement of Π is determined at round 1); Σ is used to prove $x \in \mathsf{L}$. The key point for extractability is that whenever the statement of Π is true (*i.e.*, whenever the protocol has been set up correctly), a witness certifying $x \in \mathsf{L}$ can be extracted by rewinding Σ as long as P doesn't abort when asked to use Σ. The λ repetitions ensure that extraction succeeds with high probability $1 - 2^{-\Omega(\lambda)}$.

Suppose now that a PPT V^* plays the first two rounds $(\sigma_1, \sigma_2, \pi_1, \pi_2)$, then sends (x, w, τ_3) and $(\hat{x}, \hat{w}_0, \hat{w}_1, \hat{\tau}_3)$ where τ_3 and $\hat{\tau}_3$ are third messages (so τ_3 is a bit and then a third message of either Σ or Π). Then V^* receives τ_4 and $\hat{\tau}_4$, where τ_4 proves $x \in \mathsf{L}$ using witness w and $\hat{\tau}_4$ proves $\hat{x} \in \mathsf{L}$ using witness $\hat{w}_u$ for a random $u \in \{0, 1\}$. Consider the protocol messages: $(\sigma_1, \sigma_2, \pi_1, \pi_2, \tau_3, \tau_4, \hat{\tau}_3, \hat{\tau}_4)$. Rewind-secure WI is proved via a hybrid argument. The first step is to generate the messages of Π using the ZK simulator instead of honestly; rewind-secure ZK of Π ensures the resulting transcript is indistinguishable. Once Π is being simulated, the next change is to set up the protocol incorrectly so that there is a trapdoor witness for P to use in Σ (though P still proves $x, \hat{x} \in \mathsf{L}$ using w and w_u). The final step is to switch to using the trapdoor witness in both executions of Σ. This step presents complications as Σ does not have rewind security. For this reason, the actual protocol is more complicated than described so far. In actuality, P begins many executions of Σ in the first two rounds and uses a random subset of them in rounds three and four to complete the proofs $x, \hat{x} \in \mathsf{L}$. This ensures that for most of the executions of Σ begun in rounds 1 and 2, they will be run to completion at most once, and so for these executions of Σ, standard witness indistinguishability (*i.e.*, WI with no rewinding) holds. It is possible, however, that a small number of the Σ's will be run to completion twice (*i.e.*, if there is non-empty intersection between the random subsets chosen by P in τ_4 and $\hat{\tau}_4$). The witness indistinguishability guarantees are compromised in the copies of Σ which are run twice. So rather than using each copy to prove x or $\hat{x} \in \mathsf{L}$ directly, the proofs are set up so that x or $\hat{x} \in \mathsf{L}$ are proved jointly by all copies of Σ, but if a small number of the witnesses used in the individual copies of Σ are revealed, nothing is revealed about w or $\hat{w}_u$. The "MPC in the head" idea of [IKOS07] is used for this purpose. More details are given below. For even more detail, the reader should refer to [GR19].

The Protocol. Let Com be a two round commitment scheme based on a OWF [Nao91, HILL99]. Let Σ be a four-round, delayed-input WI AoK [FS90] based on OWF. Let Π be a four-round, rewind-secure ZK argument with constant soundness based on OWF [IKOS07]. Let λ be the security parameter, and let $N, n = \mathrm{poly}(\lambda)$ be such that ??. The delayed-input, rewind-secure WI AoK

repeats the following main subprotocol λ times.

1. Over rounds 1 and 2, P sends N commitments $c_1, \ldots, c_N$ to 0 using Com.

2. Also over rounds 1 and 2, P and V play the first two messages of Π, and the first two messages of N copies of Σ. We denote these messages by (π_1, π_2) and $\{(\sigma_1^i, \sigma_2^i)\}_{i\in[N]}$. The statement proven by Π is that all c_i are commitments to 0. The statements proven by the delayed-input proofs will come after the third round.

3. In the third round, V sends P a random bit $b \in \{0,1\}$; if $b = 1$, V also sends π_3; if $b = 0$, V sends $\{\sigma_3^i\}_{i\in[N]}$.

4. At this point both parties receive the statement $x \in \{0,1\}^\lambda$ and P receives a witness w such that $(x, w) \in R_{\mathsf{L}}$.

5. If $b = 1$, P receives π_3 and sends π_4 completing the proof that the c_i are commitments to 0.

6. If $b = 0$, P receives $\{\sigma_3^i\}_{i\in[N]}$ and does the following:

   - P imagines an $n-$party MPC computation proving $(x, w) \in R_{\mathsf{L}}$. We use an MPC protocol with perfect security against an adversary corrupting at most $n/3$ parties. For $u \in [n]$, let X_u denote the $u-$th party's view during this computation. P computes commitments $\{z_u\}_{u\in[n]}$ to V, where $z_u = \mathrm{Com}(X_u)$.
   - P chooses a random subset $S \subset [N]$ of size $|S| = \binom{n}{2}$, implicitly identifying S with the set of unordered pairs of elements in $[n]$.
   - For each $1 \le u < u' \le n$, let $i \in [N]$ be the corresponding element of S. For each such (u, u'), P computes σ_4^i where the transcript $(\sigma_1^i, \sigma_2^i, \sigma_3^i, \sigma_4^i)$ proves the statement: EITHER c_i is a commitment to 1 OR the committed views X_u and $X_{u'}$ inside z_u and $z_{u'}$ are consistent with one another and both outputs are 1.
   - P sends $\left(\{z_u\}_{u\in[n]}, S, \{\sigma_4^i\}_{i\in S}\right)$ to V.

The knowledge extractor rewinds P repeatedly, sending new fresh third messages, and receiving new fourth messages. The constant soundness of Π ensures that with probability $1 - 2^{-\Omega(\lambda)}$, in many of the main subprotocols, the c_i will all be commitments to 0. For such subprotocols, the EITHER part of the statement proved in Σ is not true, thus the OR part must be used. Since Σ is an AoK, the witnesses used (*i.e.*, the committed views X_u) can be extracted, and the witness w for $x \in \mathsf{L}$ can be recovered.

Rewind-secure WI is proven via a hybrid argument where one-by-one, the main subprotocols are changed so that they do not use the witness for $x \in \mathsf{L}$. The changes for one copy of the main subprotocol are as described in the previous paragraph. Specifically, suppose V* plays the first two rounds (τ_1, τ_2), then sends (x, w, τ_3) and $(\hat{x}, \hat{w}_0, \hat{w}_1, \hat{\tau}_3)$ where τ_3 and $\hat{\tau}_3$ are third messages (so τ_3 is a bit and then a third message of either Σ or Π). Then V* receives τ_4 and $\hat{\tau}_4$, where τ_4 proves $x \in \mathsf{L}$ using witness w and $\hat{\tau}_4$ proves $\hat{x} \in \mathsf{L}$ using witness $\hat{w}_u$ for a random $u \in \{0,1\}$. In the first hybrid, the messages $(\tau_1, \tau_2, \tau_3, \tau_4, \hat{\tau}_4)$ are all specified according to the protocol. In the second hybrid, all data pertaining to Π is simulated rather than produced honestly; the rewind-secure ZK of Π ensures this hybrid is indistinguishable to the first. In the third hybrid, the c_i are changed to commitments to 1 instead of 0; indistinguishability follows from the hiding of Com. Over the course

of the remaining hybrids, the witnesses used in the executions of Σ are changed. At the beginning of the fourth hybrid, $S, \hat{S} \subset [N]$ of size $|S| = |\hat{S}| = \binom{n}{2}$ are chosen at random and if $|S \cap \hat{S}| > n/6$ (happens with $2^{-\Omega(\lambda)}$ probability), the experiment is aborted. Additionally in the fourth hybrid, for all $i \in S \cup \hat{S} \setminus (S \cap \hat{S})$ (*i.e.*, for all i such that the $i-$th copy of Σ is played to completion at most once), the EITHER part of the witness (that c_i is a commitment to 1) is used instead of the OR part. Indistinguishability follows from the standard WI of Σ. In the final hybrid, the views used as witnesses in all $i \in S \cap \hat{S}$ (this includes the views of at most $n/3$ parties) are prepared using the MPC simulator; indistinguishability follows from perfect security of the MPC protocol against $n/3$ corruptions.