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. 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 underlying 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 protocol of Goyal [Goy11] 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. For example, this technique alone allows for an immediate linear reduction in communication complexity compared with its nearest relative, Goyal's protocol.

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 instantiation of a sequential 7-round version of our protocol 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]. Their scheme uses the same method of extraction as our scheme, along with many new ideas. Even more recently, Ciampi, Ostrovsky, Siniscalchi and Visconti [COSV16a] construct three round concurrent non-malleable commitments assuming subexponentially secure OWF. In another recent work, the same authors give the first four-round concurrent non-malleable commitment scheme based on (standard) OWF, using the scheme in this work as a building block [COSV16b]. They observe that our scheme already satisfies a weak form of concurrent non-malleability, where no MIM who commits to a valid message in each session on the right can be mauling. They then show how to compile it into a scheme with full concurrent non-malleability without incurring a cost to the round complexity or the underlying assumption.

1.3 The New Protocol

Suppose that committer C wishes to commit to message m, and let $t_1, \dots, t_n \in \mathbb{Z}$ 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 $\mathbf{r} = (r_1, \dots, r_n) \in \mathbb{F}_q^n$ and sends $\text{Com}(m)$ and $\{\text{Com}(r_i)\}_{i=1}^n$ to R;
2. R sends C a query vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ where each α_i is drawn randomly from $\left[2^{t_i}\right] \subset \mathbb{F}_q$;
3. C sends R the response $\mathbf{a} = (a_1, \dots, a_n)$ where $a_i = r_i\alpha_i + m$;

4. C proves in ZK that the values $\mathbf{a}$ (from step 3) are 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.

At a very high level, our protocol can be seen as an algebraic abstraction of Goyal's protocol. However, the fundamental difference we should emphasize from [Goy11] is that he 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 the techniques of [Goy11] do not apply to our setting−at least at first. Our protocol does have small imbalances between the challenge spaces of individual coordinates, which is what we will eventually use to prove non-malleability. However, proving that the coordinates are sufficiently independent so that these imbalances accrue to something usable is completely new to this work.

1.4 Proving Non-Malleability

Consider a MIM adversary M that is playing the role of the receiver in a protocol using tags $t_1, \dots, t_n$ while playing the role of the committer in a protocol using tags $\tilde{t}_1, \dots, \tilde{t}_n$ (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. One nice feature of our protocol is that it is automatically secure against a non-synchronizing adversary, simply because there are so few rounds, there is no way for the MIM to benefit by changing the message order: any scheduling but the synchronous one can be dealt with trivially. So the only scheduling our proof actually needs to handle is a synchronizing one, as depicted in Figure 1 below.

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 is modeled after Goyal's extractor which: (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 answers correctly on the right.

In Goyal's protocol there is no way for E to know whether M answered correctly or not, and so it must have a verification message after the query response phase so E can compare M's answer with the main thread to verify correctness. We sidestep this necessity in the following way. We rewind to the beginning of step 2 twice and ask two new query vectors $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$, we answer randomly on the

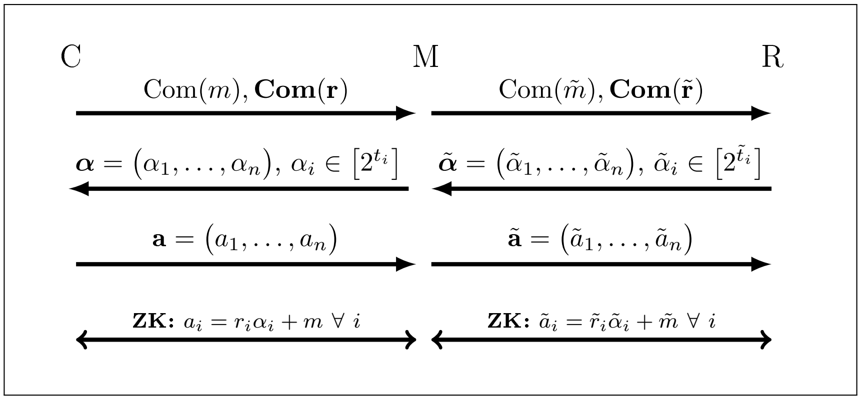

Figure 1: Protocol with Man-in-the-Middle

left obtaining $\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\}$, where $(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}})$ is from the main thread. Comparing both $(\tilde{\beta}_i, b_i)$ and $(\tilde{\gamma}_i, c_i)$ with $(\tilde{\alpha}_i, a_i)$ will result in candidate values $\tilde{m}_i$ and $\tilde{m}_i'$, but with no verification message it is not clear how E should verify which one (if either) is correct. We accomplish this with the following "collinearity test". If $\tilde{m}_i = \tilde{m}_i'$ then E checks whether the points $\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\}$ are collinear. If so, E deems that $\tilde{m}_i$ was the correct value. This requires proving that M cannot answer "incorrectly but collinearly".

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 $|\mathbb{F}| = \mathrm{poly}(n)$, 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 an incredibly trim protocol.

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 $\tilde{\alpha}_i \in [2^{t_i}]$. 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 vector $\tilde{\boldsymbol{\alpha}}$ negligible. So M will not be able to successfully answer R's query and complete the proof when performing 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. Such a claim has been at core of the analysis of some previous NMC schemes. In fact, as many previous schemes (such as [Goy11], for example) use multiple slots in order to *ensure* that some slot has a right challenge space that is much bigger than the left, such a claim often encompasses nearly the entire analysis. 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 the Feige-Shamir four round ZK argument system [FS90]. This requires running the entire ZK argument in parallel with the commit, query and response messages. We make use of some standard properties of the [FS90] scheme; namely, that it is delayed-input zero-knowledge and that it is an argument of knowledge. Additionally we make a further requirement that the ZK argument is instantiated on top of 3 round WI proofs which remain WI even if the adversary gets to rewind the challenger one time. This technical property is non-standard and is crucial for our proof. Most of the difficulty in Section 6 revolves around constructing a delayed-input three-round WI with this property. We also construct the first 4-round non-malleable zero-knowledge argument essentially by running a 4-round ZK argument protocol in parallel with a non-malleable commitment to the witness w.

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. 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.

It is worth noting that directly plugging in the ideas of [GLOV12] into our protocol results in a 6-round NMC scheme. We do not address the issue of trying to reduce the round complexity of this blackbox protocol to 4 because our 4-round non-blackbox protocol is so much faster in practice.

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:*

Correctness: If C and R do not deviate from the protocol, then R should accept (with probability 1) during the decommit phase.

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, d_0) and (c, m_1, d_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.

On the Existence of Identities. In this work, just as in [DDN91, PR05], we assume that the committer has an identity $id \in \{0,1\}^k$. 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. Though this sounds like a strong assumption on the network, essentially requiring that "you know who you are talking to", for our purposes, it is actually equivalent to the requirement discussed above, that the MIM do something other than simply copy messages. This is because our protocol is interactive, and the first committer message contains a statistically binding commitment to m. This means that if we set the committer's identity to be the first committer message, C's and M's identities will be distinct unless M copied C's first message.

Moving forward, we assume that the committer's id is externally given and we require that non-malleability holds only in the case when C and M's identities are different. We also assume for simplicity that player identities are known before the protocol begins, though strictly speaking this is not necessary, as the identities do not appear in the protocol until after the first committer message. We point out that M can choose his identity adversarially, as long as it is not equal to C's.

Definition of Non-Malleable Commitments. 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". We note that this is uniquely defined, as all commitment schemes in this work are statistically binding, and so for all but a negligible fraction of the possible transcripts $\mathbb{T}$ of the interaction between M and an honest receiver R, there exists at most one message m that is consistent with $\mathbb{T}$ (*i.e.*, for which there exist random coin tosses which give $\mathbb{T}$). We recall the definition of non-malleable commitments of Lin et al *et al.* [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 $\mathbf{MIM}_{\langle \mathrm{C},\mathrm{R}\rangle}(m,z)$ describing $(\tilde{m}, v)$: the value M commits to in the right interaction, and M's view in the full experiment. Specifically, M has auxiliary information z and interacts on the left with an honest committer C with input message m and identity id and on the right with honest receiver R. M attempts to commit to a value $\tilde{m}$ that is related to m using an identity $\widetilde{id}$ of its choice. If the right commitment (as determined by the transcript) is invalid or undefined, or $id = \tilde{id}$ its value is set to $\perp$.

The simulated execution. In the simulated execution a simulator $\mathcal{S}$ interacts with an honest receiver R. $\mathcal{S}$ receives security parameter 1^λ and auxiliary information z and interacts with the honest receiver R. Let $\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C,R} \rangle}(1^\lambda, z)$ denote the random variable describing $(\tilde{m}, v)$: 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, its value 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* $m \in \{0,1\}^\lambda$*:*

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

2.3 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

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

Note that $2i|\mathbb{F}| \le t_i < (2i+1)|\mathbb{F}|$ 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 protocol given tags $t_1, \ldots, t_n$ in error corrected form as described in Section 2.3. We use Naor's two round, statistically binding bit commitment scheme [Nao91] as a building block.[1] We use boldface to denote 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.

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.

3. Replace the polynomials with random polynomials $\bar{f}_1, \ldots, \bar{f}_n$ such that $\bar{f}_i(\alpha_i) = f_i(\alpha_i)$ and $\bar{f}_i(0) = m'$ for randomly sampled m' and all $i \in [n]$. The indistinguishability follows from having $n+1$ variables and only n equations.

4. 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.

In the final hybrid, the transcript of the commitment stage contains no information about the value m being committed to, and so no information about m is leaked by the protocol. □

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

We comment that non-malleability against a general *non-synchronizing* adversary actually holds in the above protocol (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). This makes it much easier (in fact trivial) to directly examine all of the non-synchronizing options for message scheduling that M has available.

[1] 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).

Public Parameters: Tags $t_1, \dots, t_n$ 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. R $\to$ C **Initialization message:** Send the first message σ of the Naor commitment scheme.
1. C $\to$ R **Commit message:** Sample random $r_1, \dots, r_n \in \mathbb{F}_q$ and $s, s_1, \dots, s_n$.
   - Define linear functions $f_1, \dots, f_n$ by $f_i(x) = r_i x + m$.
   - Send commitments $\mathbf{Com} = \big(\mathrm{Com}_\sigma(m; s), \mathrm{Com}_\sigma(r_1; s_1), \dots, \mathrm{Com}_\sigma(r_n; s_n)\big)$.
2. R $\to$ C **Query:**
   - Send random challenge vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$, $\alpha_i \in [2^{t_i}] \subset \mathbb{F}_q$.
3. C $\to$ R **Response:**
   - Send evaluation vector $\mathbf{a} = (a_1, \dots, a_n)$, $a_i = f_i(\alpha_i)$.
4. C $\longleftrightarrow$ R **Consistency proof:** Parties engage in a zero-knowledge argument protocol where C proves to R that $\exists \big((m, s), (r_1, s_1), \dots, (r_n, s_n)\big)$ such that:
   - $\mathbf{Com} = \big(\mathrm{Com}_\sigma(m; s), \mathrm{Com}_\sigma(r_1; s_1), \dots, \mathrm{Com}_\sigma(r_n; s_n)\big)$; and
   - $a_i = r_i \alpha_i + m \;\forall\; i = 1, \dots, n$.

Decommit Phase:

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

Verification: R checks the correctness of the commit and response messages.

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

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

$$\left\{\mathbf{MIM}_{\langle \mathrm{C},\mathrm{R}\rangle}(m, z)\right\}_{m,z} \approx_c \left\{\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C},\mathrm{R}\rangle}(1^\lambda, z)\right\}_{z,id},$$

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 $\mathcal{D}_0 = \left\{\mathbf{MIM}_{\langle \mathrm{C},\mathrm{R}\rangle}(m,z)\right\}_{m,z}$ from $\mathcal{D}_1 = \left\{\mathbf{SIM}^{\mathcal{S}}_{\langle \mathrm{C},\mathrm{R}\rangle}(1^\lambda,z)\right\}_{z,id}$ 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. This is the first appearance of a tangeable payoff of the algebraicity of our protocol. For example, the protocol of [Goy11] (which is similar to ours, but strictly combinatorial in nature) does not have this algebraic verification technique at its disposal and must introduce use extra rounds into the protocol to ensure its extractor can reconstruct $\tilde{m}$.

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,

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

Since E will not be interested in the proofs $(\pi, \tilde{\pi})$, and since M is deterministic (and so $\tilde{\mathbf{Com}}$, $\boldsymbol{\alpha}$, $\tilde{\mathbf{a}}$ are uniquely determined by $\mathbf{Com}, \tilde{\boldsymbol{\alpha}}$, and $\mathbf{a}$) we will often just write $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big)$.

Definition 3 (**Accepting Transcript**). *We say that* $\mathbb{T} \in \mathsf{ACC}$ *if both* π *and* $\tilde{\pi}$ *are accepting proofs.*

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 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 $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\cdots\big)$ instead of $\Pr_{\mathbb{T}}\big(\cdots \big| \mathbb{T} \in \mathsf{ACC}\big)$. For fixed $\mathbf{Com}$, 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 $\tilde{m}' \neq \tilde{m}$.

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*

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

where the probability is over $\mathbb{T} \in \mathsf{ACC}$ *and the randomness of* 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 4 (**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. *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*

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

Input: $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big) \in \mathsf{ACC}$, 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 $\boldsymbol{\beta}_j = (\beta_{1,j}, \ldots, \beta_{n,j})$ for left interaction.
2. Feed $\mathbf{b}_j = (b_{1,j}, \ldots, b_{n,j})$ to M where $b_{i,j} = \begin{cases} a_i, & \beta_{i,j} = \alpha_i \\ r \xleftarrow{\text{R}} \mathbb{Z}_q, & \beta_{i,j} \neq \alpha_i \end{cases}$. Get $\tilde{\mathbf{b}}_j = (\tilde{b}_{1,j}, \ldots, \tilde{b}_{n,j})$.
3. For each $i \in [n]$ use $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j})\big\}$ 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 $i \in [n]$, $\tilde{m}_{i,j} = \tilde{m}'_{i,j}$ and $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j}), (\tilde{\gamma}_{i,j}, \tilde{c}_{i,j})\big\}$ are collinear output $\tilde{m}_{i,j}$ and halt.

Output: Output **FAIL**.

Figure 3: The Extractor E.

2. $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\big\}$ *are collinear.*

We define the set $\mathsf{IBC}^i(\mathbf{Com}, \tilde{\alpha}_i) = \big\{(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}) : \mathrm{M}$ *answers* $(\tilde{\beta}_i, \tilde{\gamma}_i)$ *incorrectly but collinearly*$\big\}$. *Finally, define*

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

Note that $\mathsf{IBC}^i(\mathbf{Com}, \tilde{\alpha}_i)$ is well defined given $\mathbb{T}$ and E's randomness. Intuitively IBC is the set of transcripts for which E might fail because M answers incorrectly but collinearly. The following claim shows that these transcripts rarely occur.

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

Proof. Fix $i \in \{1, \ldots, n\}$ and let $\mathbb{T}, \mathbb{T}' \in \mathsf{ACC}$ be main threads with the same prefix $\mathbf{Com}$ 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 $\mathbf{Com}$ 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. $\square$

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

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

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,

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

$\square$

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 6 (**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 $\Pr_{\mathbb{T}}(\mathbb{T} \in \mathsf{ACC}) \geq p$. $\square$

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 7 (**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

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

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 $\langle \mathrm{C}, \mathrm{R} \rangle$. 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 8 (**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. $|\mathsf{HON}| \geq \delta p^2 |R|$;
2. *for any* $i \in [n]$, *if we (temporarily) define* $Z^i_\tau = \{\tilde{\alpha}_i \in [2^{\tilde{t}_i}] : |\mathsf{HON}^i(\tilde{\alpha}_i)| \leq \tau |R^i(\tilde{\alpha}_i)|\}$, *then*

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

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

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

□

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 $N = \mathrm{poly}(\lambda)$, a yet unspecified polynomial. Shortly we will introduce the values $\varepsilon = 1/n - \varepsilon'$ where $\varepsilon' = 1/2n^2$. We will require that $\varepsilon^* \le \sigma\delta^2 p^5/16$ and also that $\varepsilon^* \le n\varepsilon'(\varepsilon'\delta\delta' p^3)^2/2048$, 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 9 ($\varepsilon-$**dependence**). *For fixed* $\mathbb{T} \in \mathsf{ACC}$ *and* $i, i' \in \{1, \dots, n\}$, *we say* $\alpha_{i'}$ *is* $\varepsilon-$dependent *on* $\tilde{\alpha}_i$ *if* $\Pr_{\tilde{\boldsymbol{\beta}} \in \mathsf{HON}}\big(\beta_{i'} = \alpha_{i'} \big| \tilde{\beta}_i = \tilde{\alpha}_i\big) \ge \varepsilon$.

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

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

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

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

Proof. We begin with the inequality $\Pr_{\mathbb{T}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{UNBAL}\big) \ge \delta' p^2/4$ (using the fact that $\Pr_{\mathbb{T}}(\mathbb{T} \in \mathsf{ACC}) \ge p$). Fix a random commit message **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

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

$\square$

Claim 6. *Fix* $\sigma=\frac{\varepsilon'(\delta')^2p^4}{257n^3}$. *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:

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

- $Z = \{\tilde{\alpha}_{i_1} \in [2^{\tilde{t}_{i_1}}] : |\mathsf{HON}^{i_1}(\tilde{\alpha}_{i_1})| \le \tau |R^{i_1}(\tilde{\alpha}_{i_1})|\}$, where $\tau = \frac{\varepsilon'\delta\delta' p^4}{32n^3}$.

Note that with **Com** fixed as above we have

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

Now, for $v \in [2^{t_{i'}}]$, let $\mathbf{E}_v$ be the event "$\alpha_{i'} = v$." Note that if $\mathrm{Pr}_{\tilde{\boldsymbol{\alpha}}}\big(\mathbf{E}_v \big| \tilde{\boldsymbol{\alpha}} \in X\big) > 0$ then v is an ε'−dependency of $\tilde{\alpha}_{i_1}$. Let $D^{i'}(\tilde{\alpha}_{i_1}) \subset [2^{t_{i'}}]$ be the set of all ε'−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 $P(v) = \mathrm{Pr}_{\tilde{\boldsymbol{\alpha}}}\big(\mathbf{E}_v \big| \tilde{\boldsymbol{\alpha}} \in X\ \&\ \tilde{\alpha}_{i_1}\big)$. We say that $v^* \in D^{i'}(\tilde{\alpha}_{i_1})$ is *maximal* if $P(v^*) \ge P(v)$ 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 $|D^{i'}(\tilde{\alpha}_{i_1})| \le (\varepsilon')^{-1}$. We now lower bound the quantity $V = \mathrm{Pr}_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\tilde{\boldsymbol{\alpha}} \in X\ \&\ \tilde{\alpha}_{i_1} \notin Y \cup Z\ \&\ \alpha_{i'}\ \text{maximal}\big)$. We have

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

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 τ'−fraction of HON into $L^{i'}(\alpha_{i'})$ and since

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

(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

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

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 $\mathrm{Pr}_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}\big(\beta_{i'} = \alpha_{i'}\big) \ge \tau'$. 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 ε' − 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 ε' − 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 ε' − dependent on $\tilde{\gamma}_{i_2}$ & $\gamma_{i'} = \alpha_{i'}$" with probability at least τ'/ε', the definition of ε'−dependence ensures that $\beta_{i'} = \gamma_{i'} = \alpha_{i'}$ with probability at least τ', as desired. □

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

$$\mathrm{Pr}_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER{-}POLY}\big) \ge \sigma.$$

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

In this section we complete the proof of Lemma 1 by proving two claims which show how to use an M with unlikely behavior to break the hiding of $\langle \mathrm{C}, \mathrm{R}\rangle$. We first give an intuitive description of our method of argument. This description is slightly technical but does not get into the specifics of either Claim 8 or Claim 9.

We construct an adversary $\mathcal{A}$ who takes part in the hiding game for $\langle \mathrm{C}, \mathrm{R}\rangle$. $\mathcal{A}$ is defined as follows:

- $\mathcal{A}$ chooses random $m_0, m_1 \in \mathbb{Z}_q$ 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 $\mathbf{Com}$ from $\mathcal{C}$ which it forwards to M, and receives $\tilde{\mathbf{Com}}$.
  - $\mathcal{A}$ sends random $\tilde{\boldsymbol{\alpha}}$ such that $\tilde{\alpha}_i \in \big[2^{\tilde{t}_i}\big]$ 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 $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big)$ 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 $\{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \ldots\}$.
- $\mathcal{A}$ 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 $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}}), \ldots\big\}$.
- $\mathcal{A}$ tests whether the points $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}}), \ldots\big\}$ 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 $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\big\}$ 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 $\mathbb{T} \in \mathsf{SUPER{-}POLY}$ 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 $m_1, \ldots, m_\ell \in \mathbb{Z}_q$ and sends $(m_1, \ldots, m_\ell)$ 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* $\sigma = \frac{\varepsilon'(\delta')^2 p^4}{257n^3}$. *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 "$\mathbb{T} \in \mathsf{TRB} \cap \mathsf{SUPER\!-\!POLY}$." 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 $\mathbb{T} \in \mathsf{SUPER\!-\!POLY}$, 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*

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

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 $\Pr_{\tilde{\boldsymbol{\alpha}}\in\mathsf{HON}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\backslash\mathsf{FIXED}\big)\geq\delta' p/8$ 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

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

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 $\Pr\big(\tilde{b}_i = \tilde{a}_i \big| u' \neq u \ \&\ \mathbf{E}\big)$. 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 $\gamma_{i'} \in \big[2^{t_{i'}}\big]$, $\Pr_{\tilde{\boldsymbol{\beta}}\in\mathsf{HON}}\big(\beta_{i'} = \gamma_{i'}\big) \leq \varepsilon$. 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),

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

(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

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

And so we have

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

completing the proof of Claim 9. □

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

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

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

6 Non-Malleability in Four-Rounds

6.1 Four-Round Non-Malleable Commitments

In this section we show how to squeeze our non-malleable protocol $\langle \mathrm{C}, \mathrm{R} \rangle$ into 4 rounds. In the new protocol, the zero-knowledge messages are lifted up and sent together with the commit, challenge and response messages. It will be important for our security proof that we can extract M's commitment message from many coordinates. In order to facilitate this, we change our protocol $\langle \mathrm{C}, \mathrm{R} \rangle$ so that C commits to the coefficients of a quadratic polynomial in each coordinate and reveals two evaluations in the second and third protocol message. The constant terms of these quadratics will be shares of m in an appropriately chosen secret sharing scheme. We use a variant of the zero-knowledge argument of knowledge protocol of Feige and Shamir [FS90] in which V sets a trapdoor by proving a hard statement using a 3-round witness-hiding argument of knowledge ($\mathcal{WHAOK}$) and P uses a 3-round witness-indistinguishable proof of knowledge ($\mathcal{WIPOK}$) to prove either the original statement $x \in L$ or knowledge of V's trapdoor. We instantiate the $\mathcal{WIPOK}$ with a version of the 3-round $\mathcal{WIPOK}$ protocol of [FLS99], where the statement to be proven can be chosen in the last round, and where witness-indistinguishability holds even if the adversary is allowed to rewind the challenger once. These properties together allow our protocol to be parallelized down to four rounds. We comment that the 3-round $\mathcal{WIPOK}$ of [FLS99] requires OWP, but as it can be changed to require OWF by including an additional random string along with the first message of the $\mathcal{WHAOK}$, we ignore this issue. Finally, we note that there exist protocols for proving knowledge of commitment which are amenable to this type of parallelization, and do not require a general $\mathcal{NP}-$reduction (such as Schnorr protocols based on DDH). Such protocols make a much better choice in practice.

The above parallelization gives the first 4-round non-malleable commitment scheme. Our 4-round commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ appears in Figure 4. We discuss now the modified Feige-Shamir zero-knowledge protocol we use. First, we alter the $\mathcal{WHAOK}$ in order to require only the existence of a OWF f (the original construction required OWP). Second, we change the $\mathcal{WIPOK}$ so that it retains some security even in a version of the WI game where the adversary is allowed to rewind the challenger once. The four round ZK protocol we use goes as follows:

1. R chooses $2n$ random pairs (x_i^b, y_i^b) for $i = 1, \ldots, n$ and $b \in \{0,1\}$ such that $y_i^b = f(x_i^b)$ for a OWF f, and sends the y_i^b to C.

Public Parameters: Fix $\ell = n/4$, $k = \Omega(\lambda)$, and $n' = \Omega(\lambda^2)$. Let $\{t_{i,j}, t'_{i,j}\}_{(i,j)\in[n]\times[k]}$ be tags in error corrected form, prime q st $q > 2^{t_{i,j}}, 2^{t'_{i,j}}$ for all i,j, and OWF $f : X \to Y$. Let π be 3−round $\mathcal{WIPOK}$ whose statement may be chosen in the last round.

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

1. R $\to$ C: Sample random $x_i^0, x_i^1 \in X$ for $i = 1, \ldots, \lambda$ and send $(y_i^0, y_i^1)_i = \big(f(x_i^0), f(x_i^1)\big)_i$. Also send σ, the first message of Naor's commitment scheme.

2. C $\to$ R:

   - Let $(m_1, \ldots, m_k) \in \mathbb{Z}_q^k$ be random shares of m under an ℓ out of k Shamir secret sharing scheme. For $(i,j) \in [n] \times [k]$ choose random $r_{i,j}, s_{i,j} \leftarrow \mathbb{Z}_q$ and randomness $\omega_{i,j} \leftarrow \$$. Set $\mathbf{Com} = \big\{\mathrm{Com}_\sigma(m_j \| r_{i,j} \| s_{i,j}; \omega_{i,j})\big\}_{i,j}$.
   - Choose a random $z \in \{0,1\}^\lambda$, and kn' first messages for π whose statements will be chosen later: $\{\pi^1_{j,I_j}\}_{(j,I_j)\in[k]\times[n']}$. Send $\big(\mathbf{Com}, z, \{\pi^1_{j,I_j}\}\big)$ to R.

3. R $\to$ C: For $(i,j) \in [n] \times [k]$ choose $\alpha_{i,j}, \alpha'_{i,j} \leftarrow \big[2^{t_{i,j}}\big] \times \big[2^{t'_{i,j}}\big] \subset (\mathbb{Z}_q^*)^2$; set $(\boldsymbol{\alpha}, \boldsymbol{\alpha}') = \big(\{\alpha_{i,j}\}, \{\alpha'_{i,j}\}\big)$. Choose k second messages for π: $\{\pi_j^2\}_{j\in[k]}$. Send $\big((\boldsymbol{\alpha}, \boldsymbol{\alpha}'), \{x_i^{z_i}\}_{i=1}^\lambda, \{\pi_j^2\}\big)$.

4. C $\to$ R: If $y_i^{z_i} \neq f(x_i^{z_i})$ for some i, abort. Otherwise, for $(i,j) \in [n] \times [k]$ let $(a_{i,j}, a'_{i,j}) = \big(m_j + \alpha_{i,j} r_{i,j} + \alpha_{i,j}^2 s_{i,j}, m_j + \alpha'_{i,j} r_{i,j} + (\alpha'_{i,j})^2 s_{i,j}\big)$. Set $(\mathbf{a}, \mathbf{a}') = \big(\{a_{i,j}\}, \{a'_{i,j}\}\big)$. Also choose $I_j \leftarrow [n']$ at random and let π_j^3 be the third message of π so that $(\pi^1_{j,I_j}, \pi_j^2, \pi_j^3)$ proves the following statement. Send $\big((\mathbf{a}, \mathbf{a}'), \{(I_j, \pi_j^3)\}_j\big)$ to R.

   - EITHER: $\exists$ $\big(m_j, \{(r_{i,j}, s_{i,j}, \omega_{i,j})\}_i\big)$ such that $\mathbf{Com}_j = \big\{\mathrm{Com}_\sigma(m_j \| r_{i,j} \| s_{i,j}; \omega_{i,j})\big\}_i$ and $(a_{i,j}, a'_{i,j}) = \big(m_j + \alpha_{i,j} r_{i,j} + \alpha_{i,j}^2 s_{i,j}, m_j + \alpha'_{i,j} r_{i,j} + (\alpha'_{i,j})^2 s_{i,j}\big)$;
   - OR: $\exists$ (x^0, x^1) such that $(y^0_{\mathsf{val}}, y^1_{\mathsf{val}}) = \big(f(x^0), f(x^1)\big)$ for some $\mathsf{val} = 1, \ldots, \lambda$.

Decommitment and Output: C sends $\{(m_j, r_{i,j}, s_{i,j}, \omega_{i,j})\}_{(i,j)\in[n]\times[k]}$. R checks that these are valid decommitments to $\mathbf{Com}$ sent in round 2 and are consistent with $(\mathbf{a}, \mathbf{a}')$ sent in round 4. If so, R reconstructs and outputs $m \in \mathbb{Z}_q$ from the shares $\{m_j\}$.

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

2. C chooses a random challenge $z \in \{0,1\}^n$ and sends z, along with $\pi_1^1, \ldots, \pi_{n'}^1$ where each is the first message of a $\mathcal{WIPOK}$ for a statement to be determined later.[2]

3. R returns $x_i^{z_i}$ for all $i = 1, \ldots, n$, and additionally sends π^2, the second message of $\mathcal{WIPOK}$.

4. C checks that $y_i^{z_i} = f(x_i^{z_i})$ for all $i = 1, \ldots, n$ (aborting if not) and chooses $i' \in \{1, \ldots, n'\}$ at random and sends (i', π^3) so that $(\pi^1_{i'}, \pi^2, \pi^3)$ is a $\mathcal{WIPOK}$ transcript proving either $x \in \mathcal{L}$

[2]the security properties of the ZK hold for arbitrary polynomials $n, n' = \mathrm{poly}(\lambda)$; we choose n' larger than n for our proof of non-malleability.

or knowledge of some pair (x^0, x^1) such that $(y_i^0, y_i^1) = \big(f(x^0), f(x^1)\big)$ for some i.

Computational soundness follows from the hardness of inverting f and the soundness of $\mathcal{WIPOK}$. Just as in the original Feige-Shamir protocol, zero-knowledge follows from the witness indistinguishability of $\mathcal{WIPOK}$. Notice also that because C uses only $\pi^1_{i'}$ in the final round, if R is allowed to rewind C and get a second fourth message $\big(i'', (\pi^3)'\big)$, proving another (possibly different) statement, R is not able to distinguish which witness C is using in this second proof unless $i'' = i'$ which happens with probability $1/n'$ (over C's randomness, which will change in rewinding[3] based on changes to C's auxiliary input). This observation will be crucial in our proof of non-malleability.

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 Sketch. Statistical binding and computational hiding are immediate. Our proof that $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ is non-malleable follows the same extraction paradigm as the proof of Theorem 1, except that now we must extract from many coordinates because of the way the commitment is secret shared. Specifically, in order to extract $\tilde{m}$ we must extract $\tilde{m}_j$ for at least ℓ values of $j \in [k]$. This requires a slightly different analysis of the dependencies between the left and right queries, and is the reason why our four round protocol uses quadratic polynomials instead of linear, however one can show that essentially the same extractor used for proving Theorem 1 works here as well. The extractor rewinds M twice to the beginning of the right session's third message and asks new queries $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}')$ and $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\gamma}}')$, receiving left queries $(\boldsymbol{\beta}, \boldsymbol{\beta}')$ and $(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$, provides random answers $(\mathbf{b}, \mathbf{b}')$ and $(\mathbf{c}, \mathbf{c}')$ and receives $(\tilde{\mathbf{b}}, \tilde{\mathbf{b}}')$ and $(\tilde{\mathbf{c}}, \tilde{\mathbf{c}}')$ on the right. For each (i, j), E checks whether either

$$\big\{(\tilde{\alpha}_{i,j}, \tilde{a}_{i,j}), (\tilde{\alpha}'_{i,j}, \tilde{a}'_{i,j}), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j}), (\tilde{\gamma}_{i,j}, \tilde{c}_{i,j})\big\}; \text{ or } \big\{(\tilde{\alpha}_{i,j}, \tilde{a}_{i,j}), (\tilde{\alpha}'_{i,j}, \tilde{a}'_{i,j}), (\tilde{\beta}'_{i,j}, \tilde{b}'_{i,j}), (\tilde{\gamma}'_{i,j}, \tilde{c}'_{i,j})\big\}$$

are consistent with a quadratic polynomial. If so, E sets $\tilde{m}_j$ equal to the constant term of this quadratic. E repeats this process polynomially many times and at the end attempts to reconstruct $\tilde{m}$ from the $\tilde{m}_j$ he has extracted. If E has not extracted enough $\tilde{m}_j$ to recover $\tilde{m}$ or if recovery errs because the $\tilde{m}_j$ are not consistent with a valid sharing, E outputs $\perp$.

One important point is that because $(\mathbf{a}, \mathbf{a}')$ is sent along with the final message of the zero-knowledge argument proving correctness of $(\mathbf{a}, \mathbf{a}')$, if E wants to send random responses on the left, he must also send the final message of a simulated argument; namely, he must use M's trapdoor statement as his witness. This means that before E can start extracting, he must extract M's trapdoor and, more importantly, 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. We prove in Appendix A that if (i, j) is such that M answers either $\tilde{\alpha}_{i,j}$ or $\tilde{\alpha}'_{i,j}$ correctly on the right given correct answers and simulated argument on the left then E extracts $\tilde{m}_j$ with high probability. This portion of the proof is very similar to the proof of Theorem 1.

We know that M's chance of answering correctly on the right is non-negligible when given correct answers and correct proofs on the left (else $\mathbb{T} \notin \mathsf{USEFUL}$) so it remains to deal with an M who answers correctly on the right with non-negligible probability given correct answers and honest proofs on the left, but who answers correctly on the right only with negligible probability given correct answers and simulated proofs on the left. We will show how to use such an M to break the security of the $\mathcal{WIPOK}$. In the next section we introduce some notation and give a formal proof of this part, which completes the proof that $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ is standalone non-malleable against a synchronizing adversary.

[3] This can be achieved by C choosing the random tape in this step by applying a PRF on the view so far.

In order to prove standalone non-malleability against a non-synchronizing M, we note that extraction is trivial from an M who uses any scheduling other than the synchronizing one. The key observations are: 1) an M who mauls must play the second message on the left before the second message on the right, 2) if the third and fourth messages on the right are consecutive then E can extract $\tilde{m}$ trivially simply by rewinding and asking $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}')$, receiving $(\tilde{\mathbf{b}}, \tilde{\mathbf{b}}')$ and checking whether the coordinates of

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

are consistent with quadratics. In other words, E works as usual except it does not have to worry about sending a random response and simulated proof on the left because of the way M has scheduled the messages. □

6.2 Completing the Proof of Theorem 3

Recall that $p \leq \Pr_{\mathbb{T}}(\mathbb{T} \in \mathsf{ACC})$. Say $\mathbb{T} \in \mathsf{NICE}$ if $\Pr_{\hat{I}}\big(\text{M answers cor. on right}\big|\hat{I}\big) \geq p^2/3$ where the experiment consists of rewinding M and sending a new fourth message on the left with the same $(\mathbf{a}, \mathbf{a}')$, and random indices $\hat{I} = \{\hat{I}_{i,j}\}$ along with corresponding proofs. Just like in Claim 3, we have $\Pr_{\mathbb{T} \in \mathsf{ACC}}(\mathbb{T} \notin \mathsf{NICE}) \leq p/3$. Given $\varepsilon > 0$ and $\mathbb{T} \in \mathsf{ACC}$ we say that an index $j \in [k]$ is *required for* $\mathbb{T}$ if $\Pr_{\hat{I}}\big(\text{M answers correctly on right}\big|\hat{I}_j \neq I_j\big) \leq \varepsilon$. Given $I = \{I_j\}, J = \{J_j\} \in [n']^k$, say that $I \cap J = \emptyset$ if $I_j \neq J_j$ for all $j \in [k]$. Given $\varepsilon > 0$, $\mathbb{T} \in \mathsf{ACC}$ and $J \in [n']^k$, we say that j is *required for* $\mathbb{T}$ *away from* J if $\Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I} \cap J = \emptyset \ \&\ \hat{I}_j \neq I_j\big) \leq \varepsilon$. We will be interested in approximately computing the required indices for $\mathbb{T}$.

Note that if we are given $\mathbb{T} \in \mathsf{ACC}$, oracle access to M, and all of the decommitment information for the π^1_{j,I_j} and both witnesses for each proof, we can approximate $\Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I}_j \neq I_j\big)$ to within ε^2 with probability at least $1 - 2^{-\Omega(\lambda)}$ for all $j \in [k]$ in polynomial time by the Chernoff bound. In this way, we partition $[k]$ into three categories: 1) j such that approx is $> \varepsilon + \varepsilon^2$; 2) j st approx is $< \varepsilon - \varepsilon^2$; 3) j st approx is in $[\varepsilon - \varepsilon^2, \varepsilon + \varepsilon^2]$. Let $\mathsf{RQD}_\varepsilon(\mathbb{T})$ be the $j \in [k]$ which fall into the second and third category. Define $\mathsf{RQD}_{\varepsilon,J}(\mathbb{T})$ similarly except using approximations of $\Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I} \cap J = \emptyset \ \&\ \hat{I}_j \neq I_j\big)$. Note this does not require the decommitment information for the π^1_{j,J_j}. We usually omit the $\mathbb{T}$, writing RQD_ε and $\mathsf{RQD}_{\varepsilon,J}$. Note RQD_ε and $\mathsf{RQD}_{\varepsilon,J}$ depend slightly on the randomness of the approximations, however this will not matter for us; the properties we need will hold whp over this randomness. Given $I, \hat{I} \in [n']^k$ write $\hat{I}_{\mathsf{RQD}_\varepsilon} = I_{\mathsf{RQD}_\varepsilon}$ to mean $\hat{I}_j = I_j$ for all $j \in \mathsf{RQD}_\varepsilon$.

Definition 11 (Half Extractable Transcripts). *Fix non-negligible $\varepsilon^* = p^4/(36\lambda^2)$. We say that transcript $\mathbb{T} \in \mathsf{EXT}_{\mathsf{HALF}}$ if*

$$\Pr_{(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}'), \hat{I}}\big(\text{M answers cor. on right}\big|\pi_j \text{ sim. } \forall\, j \notin \mathsf{RQD}_\varepsilon \ \&\ \hat{I}_{\mathsf{RQD}_\varepsilon} = I_{\mathsf{RQD}_\varepsilon}\big) \geq \varepsilon^*.$$

In words, $\mathbb{T} \in \mathsf{EXT}_{\mathsf{HALF}}$ if M's chance of answering correctly on the right is non-negligible given that his queries are answered correctly but he is given simulated proofs except for the coordinates in RQD_ε.

Definition 12 (Extractable Transcripts). *Fix $\varepsilon^{**} = p(\varepsilon^*)^5/\lambda^2$ and $\mathbb{T} \in \mathsf{ACC}$. We say that the index $(i,j) \in [n] \times [k]$ is* extractable for $\mathbb{T}$ *if*

$$\Pr_{(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}'), (\boldsymbol{b}, \boldsymbol{b}'), \hat{I}}\big(\text{M answers } \ \tilde{\beta}_{i,j} \text{ or } \tilde{\beta}'_{i,j} \text{ cor.}\big|\text{fourth message on left given by E}\big) \geq \varepsilon^{**},$$

where the probability is over rewind queries $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}')$, rewind answers $(\boldsymbol{b}, \boldsymbol{b}')$ and rewind indices $\hat{I}$ such that $\hat{I}_{\mathsf{RQD}_\varepsilon} = I_{\mathsf{RQD}_\varepsilon}$. We say that $\mathbb{T} \in \mathsf{EXT}$ if $\mathbb{T} \in \mathsf{ACC}$ and

$$^\#\{j \in [k] : \exists\ i \text{ st } (i,j) \text{ extractable for } \mathbb{T}\} \geq \ell.$$

In words, the coordinate (i,j) is extractable if M answers one of $\tilde{\beta}_{i,j}$ or $\tilde{\beta}'_{i,j}$ correctly on the right with non-negligible probability when given random answers and simulated proofs on the left. Because of the way $\tilde{m}$ is secret shared, $\mathbb{T} \in \mathsf{EXT}$ implies that $\mathrm{E}(\mathbb{T}) = \tilde{m}$ with high probability (see Figure 6 in Appendix A for the formal description of E). We have already seen that $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T} \notin \mathsf{NICE}) \leq p/3$. The following two claims combine to show that $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T} \notin \mathsf{EXT}) \leq p$.

Claim 10. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T} \in \mathsf{NICE} \setminus \mathsf{EXT}_{\mathsf{HALF}}) \geq p/3$ *then there exists a PPT* $\mathcal{A}$ *who breaks the witness indistinguishability of* $\mathcal{WIPOK}$.

Claim 11. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T} \in \mathsf{EXT}_{\mathsf{HALF}} \setminus \mathsf{EXT}) \geq p/3$ *then there exists a PPT* $\mathcal{A}$ *who breaks the hiding of* Com.

Claim 11 is analogous to the Lemma 1 from Section 4, we give a proof sketch in Appendix A. Claim 10 is the main result of this section and is proven below. Let us first complete the proof of Theorem 3 assuming $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T} \notin \mathsf{EXT}) \leq p$. As usual, the existence of an extractor with high success probability gives a reduction from non-malleability to hiding. In this case, we do not get exactly such a reduction because in addition to taking $\mathbb{T}$ as input, E takes also the decommitments for the indices (i,j) such that $j \in \mathsf{RQD}_\varepsilon$. Note that as $|\mathsf{RQD}_\varepsilon| = \mathcal{O}(1)$ when $\mathbb{T} \in \mathsf{ACC}$ with probability $\gg 1-p$, the number of auxiliary shares $\tilde{m}_j$ E requires is small. E therefore transforms an M who mauls $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ to a PPT adversary who wins the following enhanced hiding game with non-negligible advantage. Claim 12 below completes the proof of Theorem 3.

Enhanced Hiding Game for $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$. Consider the following game between a challenger $\mathcal{C}$ and PPT adversary $\mathcal{A}$ specified by a constant c.

1. $\mathcal{A}$ sends $m_0, m_1 \in \mathbb{Z}_q$ to $\mathcal{C}$;
2. $\mathcal{C}$ returns a commitment to m_b for random $b \in \{0,1\}$ using $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$; let
$$\mathbf{Com} = \{\mathbf{Com}_j\}_{j\in[k]} = \big\{\mathrm{Com}(m_j \| r_{i,j} \| s_{i,j}; \omega_{i,j})\big\}_{(i,j)\in[n]\times[k]}$$
be the commitment vector in the second round of $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$.
3. $\mathcal{A}$ sends $\mathcal{C}$ a subset $S \subset [k]$ of size at most c.
4. $\mathcal{C}$ returns the decommitments $(m_j, r_{i,j}, s_{i,j}, \omega_{i,j})$ for all (i,j) such that $j \in S$.
5. $\mathcal{A}$ outputs $b' \in \{0,1\}$ and wins if $b' = b$.

Claim 12. *For all constants c and PPT $\mathcal{A}$,* $\Pr(\mathcal{A} \text{ wins}) \leq 1/2 + \mathbf{negl}$.

Claim 12 follows via a standard reduction to the hiding of Com; the key point is that since $c = \mathcal{O}(1)$, the set $S \subset [k]$ of auxiliary decommitments $\mathcal{A}$ will require can be guessed and one does not run into selective hiding issues.

Proof of Claim 10. If $\mathbb{T} \in \mathsf{NICE}$ we have
$$p^2/3 \leq \Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I}\big) \leq \Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I}_{\mathsf{RQD}_\varepsilon} = I_{\mathsf{RQD}_\varepsilon}\big) + k\varepsilon.$$

We choose ε so that $\Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I}_{\mathsf{RQD}_\varepsilon} = I_{\mathsf{RQD}_\varepsilon}\big) \geq p^2/4$ when $\mathbb{T} \in \mathsf{NICE}$. Consider the random variable $\mathsf{X}_J = \Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I} \cap J = \emptyset\big)$ over the choice of $J \leftarrow [n']^k$. We have that $\mathsf{X}_J - k\varepsilon \leq \Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I} \cap J = \emptyset \;\&\; \hat{I}_{\mathsf{RQD}_{\varepsilon,J}} = I_{\mathsf{RQD}_{\varepsilon,J}}\big)$. Also, $\mathbb{E}[\mathsf{X}_J] \geq p^2/3$ when $\mathbb{T} \in \mathsf{NICE}$. It follows that if $\mathbb{T} \in \mathsf{NICE}$ then $\mathsf{X}_J > p^2/6$ with probability at least $p^2/6$ over $J \leftarrow [n']^k$, which in turn means $\Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I} \cap J = \emptyset \;\&\; \hat{I}_{\mathsf{RQD}_{\varepsilon,J}} = I_{\mathsf{RQD}_{\varepsilon,J}}\big) \geq p^2/12$. Our $\mathcal{A}$ interacts with the challenger $\mathcal{C}$ as follows.

Setting $\mathbb{T}$: $\mathcal{A}$ instantiates two sessions of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ with M. Before sending the second message on the left, $\mathcal{A}$ chooses a random $J \in [n']^k$, and receives from $\mathcal{C}$, $\{\pi^1_{j,J_j}\}$: many first messages of $\mathcal{WIPOK}$ for a statement to be declared later. $\mathcal{A}$ completes the second message by choosing $\{\pi^1_{j,I_j}\}_{(j,I_j)\in[k]\times([n']\setminus\{J_j\})}$ on his own and sends the message to M. $\mathcal{A}$ plays honestly as R, when M sends back the third message of the left session, $\mathcal{A}$ prepares a fourth message by choosing $I \in [n']^k$ at random such that $I \cap J = \emptyset$, and sending $\{\pi^3_j\}$ so that $(\pi^1_{j,I_j}, \pi^2_j, \pi^3_j)$ is an honest proof for each j for the usual statement of $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$. $\mathcal{A}$ receives M's fourth message on the right. If any of the proofs fail, $\mathcal{A}$ outputs a random guess $b' \in \{0,1\}$. If all of the proofs pass, $\mathcal{A}$ proceeds, saving $(\tilde{\mathbf{a}}, \tilde{\mathbf{a}}')$, the evaluation portion of M's fourth message.

Computing $\mathsf{RQD}_{\varepsilon,J}$: $\mathcal{A}$ uses $\mathbb{T}$, oracle access to M and the decommitment information for the $\{\pi^1_{j,I_j}\}$ for $I_j \neq J_j$ to compute $\mathsf{RQD}_{\varepsilon,J}$ as described above.

Checking $\mathbb{T} \in \mathsf{NICE} \setminus \mathsf{EXT}_{\mathsf{HALF}}$ (to within Reasonable Doubt): $\mathcal{A}$ uses the same information as above to compute approximations of the probabilities X_J and

$$\mathsf{Y}_J = \Pr_{\hat{I}}\big(\text{M answers cor.}\big|\hat{I} \cap J = \emptyset \ \&\ \hat{I}_{\mathsf{RQD}_{\varepsilon,J}} = I_{\mathsf{RQD}_{\varepsilon,J}} \ \&\ \pi_j \text{ simulated}\big)$$

to within ε^* with probability $1 - 2^{-\Omega(\lambda)}$. If the approximation for X_J is less than $p/6 - \varepsilon^*$ or for Y_J is greater than $> \sqrt{\varepsilon^*} + \varepsilon^*$, output a random $b' \in \{0,1\}$. Otherwise proceed.

$\mathcal{A}$'s Decision: $\mathcal{A}$ now returns to his interaction with $\mathcal{C}$. Let $\{\pi^2_j\}$ be as in the third message of the left interaction of $\mathbb{T}$. $\mathcal{A}$ sends $\{\pi^2_j\}$ to $\mathcal{C}$ along with the statement from $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$ to be proven and both witnesses. $\mathcal{A}$ receives $\{\hat{\pi}^3_j\}$. Now, $\mathcal{A}$ sends $\big((\mathbf{a}, \mathbf{a}'), \{(\hat{I}_j, \pi^3_j)\}\big)$ to M where $(\hat{I}_j, \pi^3_j)$ is as in $\mathbb{T}$ when $j \in \mathsf{RQD}_{\varepsilon,J}$, and equals $(J_j, \hat{\pi}^3_j)$ otherwise. $\mathcal{A}$ checks whether the evaluation component of M's fourth message on the right is $(\tilde{\mathbf{a}}, \tilde{\mathbf{a}}')$. If so and all of M's proofs pass, output $b' = 1$; otherwise output a random $b' \in \{0,1\}$.

Computing $\mathcal{A}$'s Winning Probability. Consider the $\mathbb{T}$ which $\mathcal{A}$ computes. If $\mathbb{T} \notin \mathsf{ACC}$ $\mathcal{A}$ guesses b' randomly, so wins with prob $1/2$. With probability at least p, $\mathbb{T} \in \mathsf{ACC}$. Similarly, if $\mathsf{X}_J < p^2/6 - 2\varepsilon^*$ or $\mathsf{Y}_J > \sqrt{\varepsilon^*} + 2\varepsilon^*$ then $\mathcal{A}$ guesses b' randomly and so wins with prob $1/2$. By assumption, $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in \mathsf{NICE} \setminus \mathsf{EXT}_{\mathsf{HALF}}\big) \geq p/3$ which means we will have $\mathsf{X}_J > p^2/6$ and $\mathsf{Y}_J < \sqrt{\varepsilon^*}$ with probability at least $p^2/6 - \sqrt{\varepsilon^*}$. In this case, $\mathcal{A}$ does not guess b' randomly and moreover, there is a gap between X_J and Y_J. The claim follows since it means if M's fourth message matches $(\tilde{\mathbf{a}}, \tilde{\mathbf{a}}')$ then it is more likely that $\mathcal{C}$ chose $b = 1$ corresponding to the real witness, than that he chose $b = 0$ corresponding to the simulated witness. Therefore, $\mathcal{A}$ wins with non-negligible advantage, breaking the witness indistinguishability of $\mathcal{WIPOK}$. $\square$

6.3 Four-Round Non-Malleable Zero-Knowledge

Using our new commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle_{\mathsf{OPT}}$, we obtain a simple 4-round non-malleable zero knowledge argument $\langle \mathrm{P}, \mathrm{V} \rangle$ for any language $L \in \mathcal{NP}$. A detailed description of $\langle \mathrm{P}, \mathrm{V} \rangle$ appears in Figure 5. It builds on top of a four round ZK argument of knowledge π which remains zero-knowledge even when the adversary is allowed to rewind the prover once. Since the statement and witness are fixed before the protocol starts, such protocols can be constructed, for example, using the MPC-in-the-head technique of [IKOS07].

Proposition 2. *If OWFs exist then $\langle \mathrm{P}, \mathrm{V} \rangle$ is a $4-$round non-malleable zero knowledge argument of knowledge for any $L \in \mathcal{NP}$.*

Public Input: Tags $t_1, \dots, t_n$ in error corrected form, large prime q and OWF $f : X \to Y$.

Common input: $x \in L$.

Input to the prover: A witness w for $x \in L$.

1. $\mathrm{V} \to \mathrm{P}$: Send $\langle \mathrm{C}(w), \mathrm{R}\rangle_1$, the first message of $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ and π_1, the first message of π.
2. $\mathrm{P} \to \mathrm{V}$: Send $\langle \mathrm{C}(w), \mathrm{R}\rangle_2$ and π_2, the second message of π for the statement:
   - $\langle \mathrm{C}(w), \mathrm{R}\rangle_2$ contains valid commitment to w such that $(x, w) \in L$.
3. $\mathrm{V} \to \mathrm{P}$: Send $\langle \mathrm{C}(w), \mathrm{R}\rangle_3$ and π_3.
4. $\mathrm{P} \to \mathrm{V}$: Send $\langle \mathrm{C}(w), \mathrm{R}\rangle_4$ and π_4.

Verification and Output: If π and all proofs in $\langle \mathrm{C}(w), \mathrm{R}\rangle$ accept then accept; otherwise reject.

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

Proof Sketch. Zero-knowledge and soundness follow by the ZK and soundness of π and the hiding and binding of $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$. The extractor for $\langle \mathrm{P}, \mathrm{V}\rangle$ simply runs the extractor for $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$. To prove non-malleability, we use the extractor for $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ to extract $\tilde{w}$ from a right execution of $\langle \mathrm{P}, \mathrm{V}\rangle$ without rewinding the left. Note the extractor must simulate π on the left, which amounts to completing an inner $\mathcal{WIPOK}$ using the trapdoor statement. Recall extraction succeeds as long as M's chance of answering the linear evaluation portion of the fourth message of $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ on the right is non-negligible. Certainly if π is completed honestly on the left then M's chance of answering correctly on the right is non-negligible. Therefore, either extraction of $\tilde{w}$ succeeds or there is a gap between M's chance of answering correctly on the right given an honest proof for π and a simulated proof. Much like in the proof of Claim 10, this gap can be used to break the ZK of π, which is supposed to hold even if the adversary can rewind the challenger once. We expand on this below.

Let us fix some notation. Denote the fourth message of $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$ on the left by $\big((\mathbf{a}, \mathbf{a}'), \Gamma\big)$ where Γ is the information regarding the proofs in $\langle \mathrm{C}, \mathrm{R}\rangle_{\mathsf{OPT}}$; Γ will not be important in this discussion, except that there is some notion of a *correct* Γ which can be efficiently verified. Let $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ be a three round $\mathcal{WIPOK}$ which is a subprotocol of π used to prove either the honest statement or a trapdoor statement. We assume σ is WI even in a game where the adversary can rewind the challenger once. Specifically, no PPT adversary should be able to win the following game with non-negligible advantage: C and $\mathcal{A}$ interact, obtaining $(\sigma_1, \sigma_2, \sigma_3)$, a valid proof of a statement using witness W_0; then $\mathcal{A}$ rewinds C sending σ_2'; C chooses $b \leftarrow \{0, 1\}$ and responds with σ_3' so that $(\sigma_1, \sigma_2', \sigma_3')$ is a valid proof using witness W_b; $\mathcal{A}$ outputs $b' \in \{0, 1\}$ and wins if $b' = b$. If M's chance of giving correct $\big((\tilde{\mathbf{a}}, \tilde{\mathbf{a}}'), \tilde{\Gamma}\big)$ on the right is non-negligible when σ on the left is computed using an honest witness but negligible when σ is computed using a trapdoor witness then M can be used to win this game.

Specifically, a PPT $\mathcal{A}$ instantiates M who plays two sessions of $\langle \mathrm{P}, \mathrm{V}\rangle$ one time through honestly obtaining $(\sigma_1, \sigma_2, \sigma_3)$ on the left and $(\tilde{\mathbf{a}}, \tilde{\mathbf{a}}')$ on the right; if either session does not complete, $\mathcal{A}$ aborts and outputs a guess for b'. In particular, if $\mathcal{A}$ continues then it must be that $(\tilde{\mathbf{a}}, \tilde{\mathbf{a}}')$ is correct. Then

$\mathcal{A}$ rewinds M sending a new third message in the right execution, with the same $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\alpha}}')$ but random $\tilde{\sigma}_2'$, receiving $(\boldsymbol{\beta}, \boldsymbol{\beta}')$ and σ_2' on the right. $\mathcal{A}$ forwards σ_2' to $\mathcal{C}$ and receives σ_3' so that $(\sigma_1, \sigma_2', \sigma_3')$ is a valid proof using either the honest witness or the trapdoor witness. $\mathcal{A}$ sends σ_3' to M along with correct $\big((\mathbf{b}, \mathbf{b}'), \Gamma\big)$, and checks whether $(\tilde{\mathbf{b}}, \tilde{\mathbf{b}}')$ in M's fourth message on the right equals $(\tilde{\mathbf{a}}, \tilde{\mathbf{a}}')$. If so, $\mathcal{A}$ decides $\mathcal{C}$ used honest witness, if not $\mathcal{A}$ answers randomly. $\square$

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.

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.

[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.

[COSV16a] 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.

[COSV16b] Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi, and Ivan Visconti. On round-efficient non-malleable protocols. *IACR Cryptology ePrint Archive*, 2016:621, 2016.

[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.

[FLS99] Uriel Feige, Dror Lapidot, and Adi Shamir. Multiple non-interactive zero-knowledge proofs under general assumptions. *J. Comput. Syst. Sci.*, 29, 1999.

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

[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.

[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.

[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.

[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.

[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. 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 Extractor and Proof of Claim 11

Claim 11 (Restated). *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}(\mathbb{T} \in \mathsf{EXT}_{\mathsf{HALF}} \setminus \mathsf{EXT}) \geq p/3$ *then there exists a PPT* $\mathcal{A}$ *who breaks the hiding of* Com.

Proof Sketch. Claim 11 says that if M answers correctly on the right given correct answers on the left (and simulated proofs) then he must also answer correctly in many coordinates on the right given random answers on the left (and simulated proofs). This is the same high level statement as Lemma 1, and the proof of Claim 11 follows the same overall path. In particular, we analyze the different possibilities for the dependencies between the left and right queries and prove that each such possibility is impossible unless M is breaking the hiding of Com. The tags $\{t_{i,j}, t'_{i,j}\}$ we are given are in error-corrected form and ordered so that:

1. $t_{i_1,j_1} < t'_{i_2,j_2}$ for all $(i_1, j_1), (i_2, j_2)$;
2. $t_{i_1,j_1} < t_{i_2,j_2}$ whenever $(i_1, j_1) < (i_2, j_2)$;
3. $t'_{i_1,j_1} < t'_{i_2,j_2}$ whenever $(i_1, j_1) > (i_2, j_2)$;

where $(i_1, j_1) < (i_2, j_2)$ if $i_1 < i_2$ or $i_1 = i_2$ and $j_1 < j_2$. Intuitively, this ensures that each polynomial gets one query from a large subset and one query from a small subset. Recall the definition of $\varepsilon-$dependence.

Definition 13 ($\varepsilon-$**dependence**)**.** *For* $\mathbb{T} \in \mathsf{ACC}$ *and* $(i, j), (i', j') \in [n] \times [k]$*, we say* $\alpha_{i',j'}$ *is* $\varepsilon-$dependent *on* $\tilde{\alpha}_{i,j}$ *if* $\Pr_{(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\beta}}')\in\mathsf{HON}}\big(\beta_{i',j'} = \alpha_{i',j'} \big| \tilde{\beta}_{i,j} = \tilde{\alpha}_{i,j}\big) \geq \varepsilon$.

Now we start classifying important sets of indices; this is analogous to the important sets of transcripts in the proof of Lemma 1.

Definition 14 (**Independent Indices and Super-poly Transcripts**)**.** *Let* $\omega = \omega(1)$ *and* $\varepsilon' = 1/(kn) - \varepsilon$*. For* $\mathbb{T} \in \mathsf{ACC}$*, let*

1. $\mathsf{IND}(\mathbb{T}) := \big\{(i, j) \in [n] \times [k] : \Pr_{(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\beta}}')\in\mathsf{HON}}\big(\beta_{i',j'} \neq \alpha_{i',j'}\ \forall\ (i', j')\big|\tilde{\beta}_{i,j} = \tilde{\alpha}_{i,j}\big) \geq \varepsilon' nk\big\}$;
2. $\mathsf{SUPER{-}POLY} := \big\{\mathbb{T} \in \mathsf{ACC} : {}^{\#}\{(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\alpha}}') \in \mathsf{HON} : \mathrm{M}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\alpha}}') = (\boldsymbol{\alpha}, \boldsymbol{\alpha}')\} \geq \lambda^{\omega}\big\}$.

Note by the union bound, if $(i, j) \notin \mathsf{IND}$ then there exists (i', j') such that $\alpha_{i',j'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_{i,j}$. The following claims are analogous to Claims 8 and 9 from Section 4.

Tags: Let $\{t_{i,j}, t'_{i,j}\}$ and $\{\tilde{t}_{i,j}, \tilde{t}'_{i,j}\}$ be the left and right tags, respectively, in error corrected form.

Input: $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\alpha}}', \mathbf{a}, \mathbf{a}'\big) \in \mathsf{ACC}$, a large value $N = \text{poly}(\lambda)$, and all decommitments corresponding to left messages m_j and proofs π^1_{j,I_j} for $j \in \mathsf{RQD}_\varepsilon(\mathbb{T})$. Also E receives all decommitments corresponding to proofs π^1_{j,J_j} for $j \notin \mathsf{RQD}_\varepsilon(\mathbb{T})$ and $J_j \neq I_j$. E is given oracle access to M.

Obtain M**'s Trapdoor:** Rewind M many times to the beginning of the left execution's second message asking fresh random $z \in \{0,1\}^\lambda$. Play honestly on the right and view M's third message in the left execution until obtaining (x^0, x^1) such that $(y^0_{\mathsf{val}}, y^1_{\mathsf{val}}) = \big(f(x^0), f(x^1)\big)$ for some $\mathsf{val} \in [\lambda]$. If after N attempts no such (x^0, x^1) has been obtained, output **FAIL**.

Extraction procedure: For $\mathsf{count} \in [N]$:

1. Rewind M to the beginning of step 2 of the protocol:
   - generate a random right challenge vector $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}') = (\tilde{\beta}_{i,j}, \tilde{\beta}'_{i,j})_{(i,j)\in[n]\times[k]}$, where $(\tilde{\beta}_{i,j}, \tilde{\beta}'_{i,j}) \in [2^{\tilde{t}_{i,j}}] \times [2^{\tilde{t}'_{i,j}}]$.
   - Feed M with $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}')$ and reuse the $\tilde{\pi}^2_j$ from $\mathbb{T}$ receive $(\boldsymbol{\beta}, \boldsymbol{\beta}')$ and $\{\hat{\pi}^2_j\}$ for left interaction.
2. Set $(\mathbf{b}, \mathbf{b}') = (b_{i,j}, b'_{i,j})$, where $b_{i,j} = \begin{cases} a_{i,j}, & \beta_{i,j} = \alpha_{i,j} \\ r \xleftarrow{\text{R}} \mathbb{Z}_q, & \beta_{i,j} \neq \alpha_{i,j} \end{cases}$ and similarly for $b'_{i,j}$ when $j \notin \mathsf{RQD}_\varepsilon(\mathbb{T})$. When $j \in \mathsf{RQD}_\varepsilon(\mathbb{T})$, set $b_{i,j} = m_j + r_{i,j}\beta_{i,j} + s_{i,j}\beta^2_{i,j}$, and similarly for $b'_{i,j}$ using the decommitments given as input. For each $j \notin \mathsf{RQD}_\varepsilon(\mathbb{T})$, choose $\hat{I}_j \in [n']$ at random such that $\hat{I}_j \neq I_j$; set $\hat{I}_j = I_j$ for $j \in \mathsf{RQD}_\varepsilon$. Compute $\hat{\pi}^3_j$ so that $(\pi^1_{j,\hat{I}_j}, \pi^2_j, \hat{\pi}^3_j)$ are $\mathcal{WIPOK}$ proofs using the honest witness when $j \in \mathsf{RQD}_\varepsilon$ and trapdoor witness when $j \notin \mathsf{RQD}_\varepsilon$. Receive $(\tilde{\mathbf{b}}, \tilde{\mathbf{b}}')$.
3. Repeat steps 1-2. Let $(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\gamma}}')$ be right challenge and $(\tilde{\mathbf{c}}, \tilde{\mathbf{c}}')$ the response.
4. For each $(i,j) \in [n] \times [k]$, check whether either

   $$\big\{(\tilde{\alpha}_{i,j}, \tilde{a}_{i,j}), (\tilde{\alpha}'_{i,j}, \tilde{a}'_{i,j}), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j}), (\tilde{\gamma}_{i,j}, \tilde{c}_{i,j})\big\};\ \text{or}\ \big\{(\tilde{\alpha}_{i,j}, \tilde{a}_{i,j}), (\tilde{\alpha}'_{i,j}, \tilde{a}'_{i,j}), (\tilde{\beta}'_{i,j}, \tilde{b}'_{i,j}), (\tilde{\gamma}'_{i,j}, \tilde{c}'_{i,j})\big\}$$

   are consistent with a quadratic. If so, let $\tilde{m}_j$ be the constant term of this quadratic.

Message Reconstruction and Output: Use the $\tilde{m}_j$ to reconstruct and output the secret $\tilde{m}$. If reconstruction fails or if fewer than ℓ of the $\tilde{m}_j$ have been found, output $\perp$.

Figure 6: The Four Round Extractor E.

Claim 13. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in (\mathsf{EXT}_{\mathsf{HALF}} \setminus \mathsf{EXT}) \cap \mathsf{SUPER{-}POLY}\big) \geq \sigma$ *then there exists a PPT algorithm* $\mathcal{A}$ *who breaks the hiding of* Com.

Claim 14. *If $(i,j) \in [n] \times [k]$ is such that*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in (\mathsf{EXT_{HALF}} \setminus \mathsf{EXT}) \ \&\ (i,j) \in \mathsf{IND}(\mathbb{T}) \ \&\ (i,j) \text{ not extractable}\big) \geq \sigma$$

then there exists a PPT algorithm $\mathcal{A}$ who breaks the hiding of Com.

As in Section 4, statistical arguments allow us to rule out certain options for the dependencies. The following is exactly analogous:

Definition 15. *For $\mathbb{T} \in \mathsf{ACC}$, say $\mathbb{T} \in \mathsf{UNBAL}$ if there exists (i,j), $(i',j') \in [n] \times [k]$ such that either 1) $\alpha'_{i',j'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_{i,j}$; 2) $\alpha_{i',j'}$ is $\varepsilon-$dependent on $\tilde{\alpha}_{i,j}$ and $(i',j') > (i,j)$; or 3) $\alpha'_{i',j'}$ is $\varepsilon-$dependent on $\tilde{\alpha}'_{i,j}$ and $(i',j') < (i,j)$.*

Claim 15. *If $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in (\mathsf{EXT_{HALF}} \setminus \mathsf{EXT}) \cap \mathsf{UNBAL}\big) \geq \delta' p$, then there exists a PPT algorithm $\mathcal{A}$ who breaks the hiding of* Com.

Definition 16. *For $\mathbb{T} \in \mathsf{ACC}$, say $\mathbb{T} \in 1{-}2$ if there exists $(i_1,j_1),(i_2,j_2),(i',j') \in [n] \times [k]$ such that $\alpha'_{i',j'}$ is $\varepsilon-$dependent on both $\tilde{\alpha}_{i_1,j_1}$ and $\tilde{\alpha}_{i_2,j_2}$.*

Claim 16. *If $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in (\mathsf{EXT_{HALF}} \setminus \mathsf{EXT}) \cap 1{-}2\big) \geq \delta' p$, then there exists a PPT algorithm $\mathcal{A}$ who breaks the hiding of* Com.

Claims 15 and 16 are proven exactly like Claims 5 and 6, using Claim 13. The next definition and claim say that we cannot have a situation in which two left queries in different coordinates depend on two sibling right queries of the same coordinate. Claim 17 is proven using a reduction to the hiding of Com analogously to how Claims 8 and 9 from Section 4 are proven.

Definition 17 (Transcripts with Mixed Polynomial Dependencies). *For $\mathbb{T} \in \mathsf{ACC}$ we say that $\mathbb{T} \in \mathsf{MIXED}$ if there exist $(i'_1,j'_1),(i'_2,j'_2),(i,j) \in [n] \times [k]$ such that 1) $(i'_1,j'_1) \neq (i'_2,j'_2)$; 2) one of $\alpha_{i'_1,j'_1}, \alpha'_{i'_1,j'_1}$ is $\varepsilon-$dependent on $\tilde{\alpha}_{i,j}$ and one of $\alpha_{i'_2,j'_2}, \alpha'_{i'_2,j'_2}$ is $\varepsilon-$dependent on $\tilde{\alpha}'_{i,j}$; 3) no $\alpha_{i',j'}$ or $\alpha'_{i',j'}$ is $\varepsilon'-$dependent on either $\tilde{\alpha}_{i,j}$ or $\tilde{\alpha}'_{i,j}$ unless $(i',j') = (i'_1,j'_1)$ or (i'_2,j'_2).*

Claim 17. *If*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in (\mathsf{EXT_{HALF}} \setminus \mathsf{EXT}) \cap \mathsf{MIXED}\big) \geq \delta' p$$

then there exists a PPT $\mathcal{A}$ who breaks the hiding of Com.

The pieces required to prove Claim 11 are now in place. It suffices to bound

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T} \in (\mathsf{EXT_{HALF}} \setminus \mathsf{EXT}) \ \&\ \mathbb{T} \notin \mathsf{UNBAL} \ \&\ \mathbb{T} \notin 1{-}2 \ \&\ \mathbb{T} \notin \mathsf{MIXED} \ \&\ \mathbb{T} \notin \mathsf{SUPER{-}POLY}\big).$$

So fix $\mathbb{T} \in \mathsf{EXT_{HALF}} \setminus \mathsf{EXT}$ and let's examine the dependencies among the queries in $\mathbb{T}$. Define the dependency graph; a directed graph (V,E) with vertex set $V = [n] \times [k] \times \{0,1\}$ and $\big((i',j',b'),(i,j,b)\big) \in E$ if $\alpha^{b'}_{i',j'}$ is $\varepsilon-$dependent on $\tilde{\alpha}^b_{i,j}$, where $(\tilde{\alpha}^0_{i,j}, \tilde{\alpha}^1_{i,j}) = (\tilde{\alpha}_{i,j}, \tilde{\alpha}'_{i,j})$. We analyze the structure of this graph in light of Claims 14 $-$ 17

First, if (i,j,b) is such that $\big((i',j',b'),(i,j,b)\big) \notin E$ for all (i',j',b'), then $(i,j) \in \mathsf{IND}(\mathbb{T})$, and so E extracts $\tilde{m}_j$. As $\mathbb{T} \notin \mathsf{EXT}$, there can exist at most $\ell - 1$ such $j \in [k]$, and so most $j \in [k]$ will be such that: for all $(i,b) \in [n] \times \{0,1\}$, exists (i',j',b') st $\big((i',j',b'),(i,j,b)\big) \in E$.

Second, for each (i',j',b') there exists at most one (i,j,b) st $\big((i',j',b'),(i,j,b)\big) \in E$, otherwise $\mathbb{T} \in 1{-}2$. Let $N(i,j,b)$ (resp. $N(i',j',b')$) be the set of (i',j',b') (resp. (i,j,b)) st $\big((i',j',b'),(i,j,b)\big) \in E$. Let

$$U = \big\{(i',j',b') : \exists\ (i,j,b) \text{ st } N(i,j,b) = \{(i',j',b')\} \text{ and } N(i',j',b') = \{(i,j,b)\}\big\}.$$

It follows that $|U| \geq 2n(k-2\ell)$, and so $^{\#}\big\{(i',j') \in [n]\times[k] : (i',j',0),(i',j',1) \in U\big\} \geq 2n(k-4\ell)$.

Third, there cannot exist $(i',j') \in [n]\times[k]$ and $(i_1,j_1,b_1),(i_2,j_2,b_2) \in [n]\times[k]\times\{0,1\}$ such that $(i',j',0),(i',j',1) \in U$, and $\big((i',j',0),(i_1,j_1,b_1)\big),\big((i',j',1),(i_2,j_2,b_2)\big) \in E$ unless $(i_1,j_1)=(i_2,j_2)$ and $b_1=0$, $b_2=1$ (else $\mathbb{T} \in \mathsf{MIXED} \cup \mathsf{UNBAL}$). Let us write $(i',j') \sim (i,j)$ if $(i',j',0),(i',j',1) \in U$ and $\big((i',j',0),(i,j,0)\big),\big((i',j',1),(i,j,1)\big) \in E$. There are at least $2n(k-4\ell)$ pairs of pairs $(i',j'),(i,j)$ such that $(i',j') \sim (i,j)$.

Fourth, we cannot have $(i',j') \sim (i,j)$ unless $(i',j')=(i,j)$ or else $\mathbb{T} \in \mathsf{UNBAL}$. Since the tags are in error corrected form, a constant fraction of $(i,j) \in [n]\times[k]$ have $(i,j) \sim (i,j)$ and $t_{i,j} > \tilde{t}_{i,j}$. As in the proof of Claim 7, fixing the right queries in all such (i,j) forces M to fix an exponentially larger fraction of the queries on the right, and so $\mathbb{T} \in \mathsf{SUPER{-}POLY}$. $\square$