Eva: Efficient IVC-Based Authentication of Lossy-Encoded Videos

Chengru Zhang[1,3], Xiao Yang[2], David Oswald[2], Mark Ryan[2], and Philipp Jovanovic[3]

[1]The University of Hong Kong, Hong Kong
[2]University of Birmingham, UK
[3]University College London, UK

Abstract

With the increasing spread of fake videos for misinformation, proving the provenance of an edited video (without revealing the original one) becomes critical. To this end, we introduce Eva, the first cryptographic protocol for authenticating lossy-encoded videos. Compared to previous cryptographic methods for image authentication, Eva supports significantly larger amounts of data that undergo complex transformations during encoding. We achieve this by decomposing repetitive and manageable components from video codecs, which can then be handled using Incrementally Verifiable Computation (IVC). By providing a formal definition and security model for proofs of video authenticity, we demonstrate the security of Eva under well-established cryptographic assumptions.

To make Eva efficient, we construct an IVC based on folding schemes that incorporate lookup arguments, resulting in a linear-time prover whose proofs can be compressed to a constant size. We further improve the performance of Eva through various optimizations, including tailored circuit design and GPU acceleration. The evaluation of our implementation shows that Eva is practical: for a 1-minute HD (1280×720) video encoded in H.264 at 30 frames per second, Eva generates a proof in about 2.5 hours on consumer-grade hardware at a speed of 5.5 µs per pixel, surpassing previous cryptographic image authentication schemes that support arbitrary editing operations by more than an order of magnitude.

1 Introduction

Disinformation campaigns frequently target visual multimedia content, like images and videos, due to their popularity and ease of distribution on social media platforms [1, 2]. This trend has been further exacerbated recently by the rapid evolution of (generative) AI tools [3–5] that enable the manipulation, generation, and dissemination of (fake) multimedia content with a few clicks, presenting a significant challenge to content moderation and fact-checking systems.

To combat maliciously generated multimedia content, the two primary defenses include 1) the *detection* of fake content, by humans [6–8] or automated methods [9–11], and 2) the *authentication* of genuine content in which a *prover* tries to convince a *verifier* of the content's provenance by providing some authentication information [12–15]. Among authentication-based approaches, the Coalition for Content Provenance and Authenticity (C2PA) standard [13] is an industry-wide effort to authenticate multimedia content based on digital signatures.

The issue with these existing approaches is that they either lack flexibility or raise security and privacy concerns. In practice, raw multimedia content often needs to be edited and encoded before publishing, but authentication-based methods [14–17] typically allow only a limited set of predefined transformations. While C2PA permits arbitrary edits, it requires trusted editing software to sign the editing operations, introducing trust assumptions that are difficult to meet, as an adversary may be able to extract the signing key from the software and generate legitimate signatures for arbitrary content. In addition, C2PA's metadata may expose sensitive information that is not intended for disclosure, such as the thumbnail of the original footage. Furthermore, many methods based on detection [6–11] and authentication [14–17] are prone to false positives or false negatives with a non-negligible probability, an issue which may be even worse in the presence of active attackers, who can bypass these mechanisms by exploiting their vulnerabilities [18–21]. Researchers have proposed cryptographic solutions for *image* authentication [22–26] tackling some of these challenges. Still, the problem of *video* authentication remains largely unaddressed, as it is a significantly harder task mainly due to the following two challenges:

- First, while lossless image encoding is common in practice, video encoding is usually lossy. Hence, a video authentication prover has to support lossy encoding which involves significant complexities. In particular, given a lossy-encoded video, the verifier cannot recover the edited video that exactly matches the prover's claim because of the information loss caused by encoding. In contrast, with lossless encoding, the prover's claim can be reconstructed accurately by the verifier.
- Second, videos extend images by adding a time dimension, increasing data sizes significantly. However, authenticating large amounts of (edited) data, which is usually achieved through advanced zero-knowledge proofs, imposes heavy computational and memory costs on the prover.

In this work, we introduce Eva, the first efficient cryptographic protocol for authentication of lossy-encoded videos that supports arbitrary editing operations. The core protocol works as follows: 1) After recording a video footage $\boldsymbol{V}$, the recorder signs its hash $\mathsf{H}(\boldsymbol{V})$ and produces signature σ. 2) After the prover edits $\boldsymbol{V}$ and encodes the edited video $\boldsymbol{V}'$ to obtain ζ, a proof π is generated, demonstrating that σ is a valid signature on $\mathsf{H}(\boldsymbol{V})$ and that $\boldsymbol{V}$ has been honestly transformed into ζ. 3) When the verifier receives ζ and π, it can verify the authenticity of ζ even without access to $\boldsymbol{V}$.

To address the first challenge, a naive solution is to prove that the edited video $\boldsymbol{V}'$ and the video $\widetilde{\boldsymbol{V}}$ reconstructed from the encoded bitstream ζ are "similar", but it is difficult to define a metric to quantify such similarity without introducing false positives or false negatives. Proving the correctness of video encoding is thus inevitable to achieve non-negligible error rates. However, converting the highly complex encoding process into a circuit is intricate and computationally expensive. Our key insight is that, despite the video being encoded in a lossy format, the verifier can still accurately recover some of the encoder's intermediate data from the encoded video. By treating the data as public inputs, we can bypass the proof of the most complicated parts of the encoding process and instead focus on manageable components, thereby significantly reducing the difficulty of circuit construction.

To address the second challenge, we exploit the highly repetitive structure of videos and video processing algorithms: they are usually based on *macroblocks*, *i.e.*, small and fixed-size blocks of pixels that can be processed independently. This allows us to leverage *Incrementally Verifiable Computation* (IVC) [27] constructed from *folding schemes* [28–32] to reduce the circuit size and memory requirements. In each IVC step, the prover only needs to 1) generate an *incoming proof* of the honest editing and encoding for a few macroblocks, and then 2) accumulate it into the *running proof*. Finally, we also leverage *lookup arguments* [33–36] to avoid expensive bit operations in the arithmetic circuits for video processing.

1.1 Contribution

As the core contributions of our work, we formalize the notion of *proofs of video authenticity* along with its security model, and we propose Eva, the first cryptographic protocol to ensure multimedia authenticity for videos with lossy encoding.

By introducing Eva, we not only show that it is *feasible* to construct cryptographic protocols for video authentication, but further demonstrate that they can be *efficient* and *secure*.

Feasibility. Due to the complexity of video encoders and the large size of videos, it is unrealistic to naively prove video authenticity in arithmetic circuits. Eva makes it feasible by leveraging the shared data between encoders and decoders and the repetitive structure of videos, which allows us to incrementally process videos with manageable costs per step using IVC. Eva is not only capable of handling arbitrarily large video files with a constant memory footprint, but also allows for arbitrary editing operations. We showcase Eva's compatibility with H.264 [37], and it can be extended to support other macroblock-based video and (lossy) image codecs.

Efficiency. Eva is efficient and practical for real-world applications.For a 1-minute HD (1280×720) video encoded by H.264 with 30 frames per second, a proof can be generated in ~ 2.5 hours on a consumer-grade desktop. During IVC proof generation, the memory cost is kept at a constant ~ 10 GB, and compressing the IVC proof requires $50 \sim 60$ GB of RAM. The final proofs are succinct: they have a constant size of 448 bytes and can be verified by resource-constrained devices like mobile phones or blockchain validators. Our efforts to enhance the efficiency and succinctness of Eva are twofold:

- As a theoretical contribution that might be of independent interest, we propose an IVC scheme that incorporates lookup arguments [33–36], which lies at the core of Eva. Our scheme is built upon Nova [28] and its implementation in the `sonobe` library [38]. Given that our circuits for video encoding and editing require extensive bitwise operations, we integrate LogUp [39], an efficient lookup argument, into our variant of Nova, thereby significantly reducing the number of constraints. Additionally, we improve `sonobe` by lowering the folding prover's complexity from $O(n \log(n))$ to

Table 1: Comparison between Eva and cryptographic protocols for image authentication.[a]

	Format	Compression	Editing Operations	Prover time	Prover RAM	Proof size	Max dimensions[b]
PhotoProof [22]	Image	Lossless	Arbitrary	$O(P^3 \log P)$ (< 18676 µs/px[c])	$O(P)$	$O(1)$ (2.67 KB)	128×128 $< \infty$
VIR [23]	Image	Lossless	Masking	$O(P \log P)$ (~ 16 µs/px)	$O(P)$	$O(1)$ (223 B)	3840×2160 $< \infty$
ZK-IMG [24]	Image	Lossless	Arbitrary	$O(P \log P)$ (> 355 µs/px[c])	$O(P)$	$O(\log P)$ ($\sim$10 KB)	1280×720 $< \infty$
VIMz [25]	Image	Lossless	Arbitrary	$O(P)$ (~ 167 µs/px)	$O(N)$	$O(\log^2 N)$ ($\sim$10 KB)	3840×2160 $< \infty$
VerITAS [26]	Image	Lossless	Arbitrary	$O(P \log P)$ (~ 95 µs/px)	$O(P)$	$O(\log^2 P)$ ($\sim$100 KB)	6632×4976 $< \infty$
Eva (this work)	Video	Lossy (H.264)	Arbitrary	$O(P)$ (~ 5.5 µs/px)	$O(1)$	$O(1)$ (448 B)	$1280 \times 720 \times 1800$ ∞

[a] Asymptotic complexity is measured w.r.t. the number of pixels $P = MNL$, where M is the height, N is the width, and L is the time. Concrete results are reported inside the parentheses.

[b] Each cell displays the empirical (upper value) and theoretical (lower value) maximum dimensions. ∞ refers to unlimited dimensions, while $< \infty$ means that unlimited dimensions are unsupported (due to bounded RAM).

[c] Reported by original authors due to source code unavailability. $>$ and $<$ indicate the estimated performance on our machine.

$O(n)$ and by optimizing the proof compression (known as decider) circuit via *commit-and-prove SNARKs* (CP-SNARKs) [40], while maintaining the ability of `sonobe`'s decider to generate zero-knowledge and constant sized proofs for recursive computations.

- In terms of practical improvements, we provide a concrete implementation of Eva equipped with hand-crafted circuits and various optimizations. To minimize circuit size, we utilize both general techniques, such as lookup arguments and non-deterministic advice, and tailored approaches, including dedicated non-native operations and efficient handling of branches based on dynamic conditions. In our implementation, we also exploit GPU acceleration, amortization via batching, among other optimizations, to further boost the prover's performance.

Security. Eva is proven secure in our model under well-established cryptographic assumptions, providing soundness against attackers and zero-knowledge of the original video except with negligible probability. By incorporating Eva into the C2PA standard, we can not only improve the security of C2PA by eliminating the trust assumptions on editing software but also provide better privacy guarantees by hiding the original footage from the verifier.

1.2 Related Work

By regarding videos as a generalization of images, we summarize the comparison of Eva to cryptographic protocols for image authentication in Table 1. PhotoProof [22] is a pioneering work in this direction that uses *Proof-Carrying Data* (PCD) [41] to offer authenticity of edited images. Due to the high computational cost of proof generation, it only supports tiny images. In [23], Ko et al. propose VIR, which utilizes CP-SNARKs [40] to generate constant-sized proofs of redaction on images (masking secret parts with black tiles). VIR significantly reduces the prover time while supporting much larger images. Built upon `halo2` [42], a more efficient proof system, ZK-IMG [24] also has faster prover than PhotoProof while maintaining support for arbitrary editing operations.

Concurrent with our work, Dziembowski et al. introduce VIMz [25], and Datta et al. propose VerITAS [26], which share several common ideas with Eva. For instance, VIMz also employs folding schemes to reduce prover RAM costs, and VerITAS, like Eva, utilizes lookup argument to improve prover time. However, alongside these general techniques, Eva incorporates customized IVC, tailored circuit design, and dedicated optimizations to minimize prover time, achieving optimal time complexity ($O(P)$) and the fastest prover time (~ 5 µs/px) among all the protocols.

For comparison of the concrete performance, we refer the reader to Section 7, and an extended review of related work can be found in Appendix A.

2 Preliminaries

2.1 Notations

In this paper, $y \coloneqq F(x)$ denotes the output of a deterministic algorithm F on input x. For a randomized algorithm F, we write $y \leftarrow F(x)$, or $y \coloneqq F(x; r)$ when it is supplied with external randomness r. With security parameter λ (or in the unary representation 1^λ), a negligible function in λ is denoted by $\varepsilon(\lambda)$.

Vectors and matrices are denoted by boldface italic lowercase (e.g., $\boldsymbol{x} = (x_0, x_1, \dots)$) and uppercase letters (e.g., $\boldsymbol{X} = \begin{bmatrix} x_{0,0} & \cdots \\ \vdots & \ddots \end{bmatrix}$), respectively. $\boldsymbol{x}[i, j]$ is the subvector of $\boldsymbol{x}$ from index i to j, and $\boldsymbol{X}[i, j; k, l]$ is the submatrix of $\boldsymbol{X}$ from row i to j and column k to l, inclusive. When it is clear from the context, we write $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{z})$ to indicate that $\boldsymbol{x}$ is the concatenation of $\boldsymbol{y}$ and $\boldsymbol{z}$.

We consider a *half-pairing cycle* of elliptic curve groups $(\mathbb{G}, \widehat{\mathbb{G}}, \mathbb{G}_T), \mathbb{H}$, where $\mathbb{G}$ and $\mathbb{H}$ form a 2-cycle. In this cycle, $\mathbb{F}_q$, the base field (*i.e.*, the field over which the curve is defined) of $\mathbb{G}$, is also the scalar field (*i.e.*, the prime field modulo the order of the curve) of $\mathbb{H}$; and $\mathbb{F}_p$, the scalar field of $\mathbb{G}$, is also the base field of $\mathbb{H}$. Further, $(\mathbb{G}, \widehat{\mathbb{G}}, \mathbb{G}_T)$ is a pairing-friendly group, *i.e.*, there is a bilinear map $e : \mathbb{G} \times \widehat{\mathbb{G}} \rightarrow \mathbb{G}_T$.

Algorithms are written in pseudocode, and we distinguish between the operations inside and outside an arithmetic circuit by using "Circuit" and "Algorithm" prefixes, respectively. Also, "Gadget" refers to a small circuit that performs a specific operation, which is often used as a building block in larger circuits. "$cond\,?\,x : y$" is a conditional expression that evaluates to x if the condition $cond$ is true, and y otherwise. The notation "**assert** $cond$" represents an operation that returns 0 if $cond$ is not satisfied and does nothing otherwise. Its in-circuit equivalent, "**enforce** $x = y$", adds a constraint to the constraint system to enforce equality between x and y. Additionally, *hints* refer to the non-deterministic advice [43] provided by the prover to the circuit.

2.2 Cryptographic Primitives

We rely on two collision-resistant hash functions H and ρ, an existentially unforgeable signature scheme $\mathsf{Sig} = (\mathsf{Sig}.\mathcal{K}, \mathsf{Sig}.\mathcal{S}, \mathsf{Sig}.\mathcal{V})$ under chosen-message attack, and a commitment scheme $\mathsf{CM} = (\mathsf{CM}.\mathcal{K}, \mathsf{CM}.\mathcal{C}, \mathsf{CM}.\mathcal{V})$ that is binding and hiding, which we assume the reader is familiar with.

Here, we consider ρ as a random oracle in the random oracle model. The signing key and verification key in Sig are denoted by sk and vk, respectively. The commitment key in CM is denoted by ck. For simplicity, we treat the randomness in $\mathsf{CM}.\mathcal{C}$ and $\mathsf{CM}.\mathcal{V}$ as implicit and omit it from the notation.

2.3 SNARKs, CP-SNARKs, and Lookup Arguments

Consider a relation R with an associated NP-language L_R. For a *statement* $\boldsymbol{x} \in L_R$ and a *witness* $\boldsymbol{w}$, we have $R(\boldsymbol{x}, \boldsymbol{w}) = 1$ if $\boldsymbol{x}$ and $\boldsymbol{w}$ satisfy R, and $R(\boldsymbol{x}, \boldsymbol{w}) = 0$ otherwise.

An *argument system* Π for R is a protocol between a prover $\mathcal{P}$ and a verifier $\mathcal{V}$, where $\mathcal{P}$ convinces $\mathcal{V}$ that $R(\boldsymbol{x}, \boldsymbol{w}) = 1$. We say Π is *interactive* if it involves interaction between $\mathcal{P}$ and $\mathcal{V}$, while Π is *non-interactive* if $\mathcal{P}$ sends a single message to $\mathcal{V}$.

A *SNARK* [44] is a non-interactive argument system that produces short proofs, as defined below.

Definition 1 (SNARK)**.** *A* succinct non-interactive argument of knowledge (SNARK) *for relation R consists of a tuple of algorithms $\Pi = (\mathcal{K}, \mathcal{P}, \mathcal{V})$:*

- $\mathcal{K}(1^\lambda, R) \to \mathsf{srs}$

  On input security parameter 1^λ and relation R, the key generation algorithm outputs the structured reference string $\mathsf{srs} = (\mathsf{pk}, \mathsf{vk})$, which includes a proving key pk and a verification key vk. We also require the key generation algorithm to output the secret trapdoor td, which is usually omitted from the notation for simplicity.

- $\mathcal{P}(\mathsf{pk}, \boldsymbol{x}, \boldsymbol{w}) \to \pi$

  On input proving key pk, statement $\boldsymbol{x}$, and witness $\boldsymbol{w}$, the proof generation algorithm outputs a proof π.

- $\mathcal{V}(\mathsf{vk}, \boldsymbol{x}, \pi) =: b$

  On input verification key vk, statement $\boldsymbol{x}$, and proof π, the verification algorithm outputs a bit b, indicating whether the proof is valid.

A SNARK Π should be *succinct*, *complete*, *knowledge-sound*, and optionally, *zero-knowledge*.

Succinctness. Succinctness holds if the size of any proof π satisfies

$$|\pi| = \mathrm{poly}(\lambda)\,\mathrm{polylog}(|\boldsymbol{x}| + |\boldsymbol{w}|)$$

Completeness. Completeness holds if for every pair of $(\boldsymbol{x}, \boldsymbol{w})$ such that $R(\boldsymbol{x}, \boldsymbol{w}) = 1$,

$$\Pr\left[\mathsf{srs} \leftarrow \mathcal{K}(1^\lambda, R), \pi \leftarrow \mathcal{P}(\mathsf{pk}, \boldsymbol{x}, \boldsymbol{w}) : \mathcal{V}(\mathsf{vk}, \boldsymbol{x}, \pi) = 1\right] = 1$$

Knowledge Soundness. Knowledge soundness holds if for every polynomial-time adversary $\mathcal{A}$, there exists a polynomial-time extractor Ext such that for all input randomness r,

$$\Pr\left[\begin{array}{l}(\mathsf{srs},\mathsf{td})\leftarrow\mathcal{K}(1^\lambda,R)\\(\boldsymbol{x},\pi):=\mathcal{A}(\mathsf{srs};r)\\\boldsymbol{w}:=\mathsf{Ext}(\mathsf{srs},\mathsf{td};r)\\\mathcal{V}(\mathsf{vk},\boldsymbol{x},\pi)=1\end{array}:R(\boldsymbol{x},\boldsymbol{w})=0\right]\le\varepsilon(\lambda)$$

Zero-knowledge. Intuitively, Π is zero-knowledge (*i.e.*, Π is a zkSNARK) if, even without knowing the witness $\boldsymbol{w}$, it is still possible to simulate a proof that is indistinguishable from honestly generated ones.

Formally, statistical (or computational) zero-knowledge holds if there exists a simulator Sim such that for every unbounded (or polynomial-time) distinguisher $\mathcal{A}$,

$$\Pr\left[\begin{array}{l}(\mathsf{srs},\mathsf{td})\leftarrow\mathcal{K}(1^\lambda,R)\\(\boldsymbol{x},\boldsymbol{w})\leftarrow\mathcal{A}(\mathsf{srs})\\\pi\leftarrow\mathcal{P}(\mathsf{pk},\boldsymbol{x},\boldsymbol{w})\end{array}:\mathcal{A}(\pi)=1\right]\approx\Pr\left[\begin{array}{l}(\mathsf{srs},\mathsf{td})\leftarrow\mathcal{K}(1^\lambda,R)\\(\boldsymbol{x},\boldsymbol{w})\leftarrow\mathcal{A}(\mathsf{srs})\\\pi\leftarrow\mathsf{Sim}(\mathsf{pk},\mathsf{td},\boldsymbol{x})\end{array}:\mathcal{A}(\pi)=1\right]$$

Below we introduce two types of SNARKs that are going to be used in our construction, namely, *CP-SNARKs* and *lookup arguments*.

It is desirable if a SNARK for R can be augmented in the following way: when proving $R(\boldsymbol{x},\boldsymbol{w})=1$, we can additionally demonstrate that the commitment to some portion of $\boldsymbol{w}$ is c. *CP-SNARKs* [40,45] are proposed to achieve this goal.

Generally, CP-SNARKs consider $\boldsymbol{w}=(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})=(\boldsymbol{v}_0,\ldots,\boldsymbol{v}_{\ell-1},\boldsymbol{\omega})$, and $\boldsymbol{c}=(c_0,\ldots,c_{\ell-1})$, where the i-th c_i is claimed to be the commitment to the i-th vector $\boldsymbol{v}_i$. Now, the original relation we are interested in becomes $R(\boldsymbol{x},(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega}))$, and the augmented relation that we aim to prove is $R^{\mathsf{cp}}\left((\boldsymbol{x},\boldsymbol{c}),(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})\right)$, which returns 1 if $\left(\boldsymbol{x},(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})\right)$ satisfies R and c_i is indeed the commitment to $\boldsymbol{v}_i$. Formally, $R^{\mathsf{cp}}\left((\boldsymbol{x},\boldsymbol{c}),(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})\right)=1$ if and only if

$$R\left(\boldsymbol{x},(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})\right)=1\wedge\bigwedge_{i\in[0,\ell-1]}\mathsf{CM}.\mathcal{V}(\mathsf{ck},c_i,\boldsymbol{v}_i)=1$$

Definition 2 (CP-SNARK)**.** *For a commitment scheme* CM*, a commitment key* $\mathsf{ck}\leftarrow\mathsf{CM}.\mathcal{K}(1^\lambda)$*, and a relation* R *for statement* $\boldsymbol{x}$ *and witness* $\boldsymbol{w}=(\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})$*,* a commit-and-prove SNARK (CP-SNARK) *for* R *is a SNARK for relation* R^{cp} *(as defined above). A CP-SNARK consists of a tuple of algorithms* $\mathsf{CP}=(\mathcal{K},\mathcal{P},\mathcal{V})$*:*

- $\mathcal{K}(1^\lambda,\mathsf{ck},R)\rightarrow(\mathsf{pk},\mathsf{vk})$

  On input the security parameter λ*, a commitment key* ck*, and a relation* R*, the key generation algorithm outputs a pair of proving and verification key* $\mathsf{srs}=(\mathsf{pk},\mathsf{vk})$*.*

- $\mathcal{P}(\mathsf{pk},\boldsymbol{x},\boldsymbol{c},\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega})\rightarrow\pi$

  On input the proving key pk*, the statement* $\boldsymbol{x}$*, the commitments* $\boldsymbol{c}=(c_i)_{i=0}^{\ell-1}$*, and the witness* $\{\boldsymbol{v}_i\}_{i=0}^{\ell-1},\boldsymbol{\omega}$*, the proof generation algorithm outputs a proof* π*.*

- $\mathcal{V}(\mathsf{vk}, \boldsymbol{x}, \boldsymbol{c}, \pi) =: b$

  On input the verification key vk*, the statement* $\boldsymbol{x}$*, the commitments* $\boldsymbol{c}$*, and a proof* π*, the verification algorithm outputs a bit* b*, indicating whether the proof is valid.*

If a CP-SNARK for R further satisfies zero-knowledge (*i.e.*, if it is a zkSNARK for R^{cp}), then we denote it by $\mathsf{ZKCP} = (\mathcal{K}, \mathcal{P}, \mathcal{V})$.

We are also interested in *lookup arguments* [33–36], which prove that all elements in $\boldsymbol{\alpha} = \{\alpha_i\}_{i=0}^{\mu-1}$, a set of *queries*, are in a *lookup table* $\boldsymbol{\tau} = \{\tau_j\}_{j=0}^{\nu-1}$. Formally, we consider a lookup relation $R^{\mathsf{lookup}}(\boldsymbol{\tau}, \boldsymbol{\alpha})$, which returns 1 if and only if

$$\boldsymbol{\alpha} \subseteq \boldsymbol{\tau}$$

Definition 3 (Lookup Arguments)**.** *A* lookup argument *is a SNARK for relation* R^{lookup} *(as defined above).*

2.4 Folding Schemes

Intuitively, a non-interactive folding scheme [28] for relation R folds two instances into a single instance such that the correctness of the folded instance implies that of the original ones.

Definition 4 (NIFS)**.** *A* non-interactive folding scheme (NIFS) *consists of a tuple of algorithms* $\mathsf{NIFS} = (\mathcal{G}, \mathcal{K}, \mathcal{P}, \mathcal{V})$*:*

- $\mathcal{G}(1^\lambda) \to \mathsf{pp}$

  On input security parameter 1^λ*, the setup algorithm outputs public parameters* pp*.*

- $\mathcal{K}(\mathsf{pp}, R) =: (\mathsf{pk}, \mathsf{vk})$

  On input public parameters pp *and a relation* R*, the key generation algorithm outputs a pair of proving key* pk *and verification key* vk*.*

- $\mathcal{P}(\mathsf{pk}, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)) \to (\mathbb{U}, \mathbb{W}, \overline{T})$

  On input the proving key pk *and two instance-witness pairs* $(\mathbb{U}_1, \mathbb{W}_1)$ *and* $(\mathbb{U}_2, \mathbb{W}_2)$*, the proof generation algorithm outputs a folded instance-witness pair* $(\mathbb{U}, \mathbb{W})$ *and a folding proof* $\overline{T}$*.*

- $\mathcal{V}(\mathsf{vk}, \mathbb{U}_1, \mathbb{U}_2, \overline{T}) =: \mathbb{U}$

  On input of the verification key vk*, two instances* $\mathbb{U}_1$ *and* $\mathbb{U}_2$*, and the folding proof* $\overline{T}$*, the verification algorithm outputs a folded instance* $\mathbb{U}$*.*

A folding scheme NIFS satisfies the following properties.

Perfect Completeness. Completeness holds if for every PPT adversary $\mathcal{A}$,

$$\Pr\left[\begin{array}{l} \mathsf{pp} \leftarrow \mathcal{G}(1^\lambda) \\ (R, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)) \leftarrow \mathcal{A}(\mathsf{pp}) \\ R(\mathbb{U}_1, \mathbb{W}_1) = 1, R(\mathbb{U}_2, \mathbb{W}_2) = 1 \\ (\mathsf{pk}, \mathsf{vk}) := \mathcal{K}(\mathsf{pp}, R) \\ (\mathbb{U}_{\mathcal{P}}, \mathbb{W}, \overline{T}) \leftarrow \mathcal{P}(\mathsf{pk}, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)) \\ \mathbb{U}_{\mathcal{V}} := \mathcal{V}(\mathsf{vk}, \mathbb{U}_1, \mathbb{U}_2, \overline{T}) : \\ \qquad\qquad\qquad \mathbb{U}_{\mathcal{P}} = \mathbb{U}_{\mathcal{V}} \wedge R(\mathbb{U}_{\mathcal{P}}, \mathbb{W}) = 1 \end{array}\right] = 1$$

Knowledge Soundness. Knowledge soundness holds if, for every polynomial-time adversary $\mathcal{A}$, there exists a polynomial-time extractor Ext such that for all input randomness r,

$$\Pr\left[\begin{array}{l} \mathsf{pp} \leftarrow \mathcal{G}(1^\lambda) \\ (R, \mathbb{U}_1, \mathbb{U}_2, \mathbb{W}, \overline{T}) := \mathcal{A}(\mathsf{pp}; r) \\ (\mathsf{pk}, \mathsf{vk}) := \mathcal{K}(\mathsf{pp}, R) \\ \mathbb{U} := \mathcal{V}(\mathsf{vk}, \mathbb{U}_1, \mathbb{U}_2, \overline{T}) \\ R(\mathbb{U}, \mathbb{W}) = 1 \\ (\mathbb{W}_1, \mathbb{W}_2) := \mathsf{Ext}(\mathsf{pp}; r) : \\ \qquad R(\mathbb{U}_1, \mathbb{W}_1) = 0 \vee R(\mathbb{U}_2, \mathbb{W}_2) = 0 \end{array}\right] \leq \varepsilon(\lambda)$$

Zero-knowledge. Statistical (or computational) zero-knowledge holds if there exists a simulator Sim such that for every unbounded (or polynomial-time) distinguisher $\mathcal{A}$,

$$\Pr\left[\begin{array}{l} \mathsf{pp} \leftarrow \mathcal{G}(1^\lambda) \\ (R, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)) \leftarrow \mathcal{A}(\mathsf{pp}) \\ R(\mathbb{U}_1, \mathbb{W}_1) = 1 \wedge R(\mathbb{U}_2, \mathbb{W}_2) = 1 \\ (\mathsf{pk}, \mathsf{vk}) := \mathcal{K}(\mathsf{pp}, R) \\ (\cdot, \cdot, \overline{T}) \leftarrow \mathcal{P}(\mathsf{pk}, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)) : \\ \hfill \mathcal{A}(\overline{T}) = 1 \end{array}\right]$$

$$\approx \Pr\left[\begin{array}{l} \mathsf{pp} \leftarrow \mathcal{G}(1^\lambda) \\ (R, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)) \leftarrow \mathcal{A}(\mathsf{pp}) \\ R(\mathbb{U}_1, \mathbb{W}_1) = 1 \wedge R(\mathbb{U}_2, \mathbb{W}_2) = 1 \\ (\mathsf{pk}, \mathsf{vk}) := \mathcal{K}(\mathsf{pp}, R) \\ \overline{T} \leftarrow \mathsf{Sim}(\mathsf{pk}, \mathsf{vk}, \mathbb{U}_1, \mathbb{U}_2) : \\ \hfill \mathcal{A}(\overline{T}) = 1 \end{array}\right]$$

2.5 Incrementally Verifiable Computation

Incrementally verifiable computation [27] allows one to verify the repeated execution of a function $\mathcal{F}$, dubbed *step function*. Specifically, the prover can generate a proof π_i demonstrating that the current state $\boldsymbol{z}_i$ is the result of i invocations of $\mathcal{F}$ starting from an initial state $\boldsymbol{z}_0$, given the proof π_{i-1} attesting to $\boldsymbol{z}_{i-1}$. This notion is formalized as below.

Definition 5 (IVC)**.** *An* incrementally verifiable computation (IVC) *scheme is composed of four algorithms* $\mathsf{IVC} = (\mathcal{G}, \mathcal{K}, \mathcal{P}, \mathcal{V})$*:*

- $\mathcal{G}(1^\lambda) \rightarrow \mathsf{pp}$

  On input security parameter 1^λ*, the setup algorithm* $\mathcal{G}$ *outputs the public parameters* pp.

- $\mathcal{K}(\mathsf{pp}, \mathcal{F}) =: (\mathsf{pk}, \mathsf{vk})$

  On input public parameters pp *and a polynomial-time computable function* $\mathcal{F}$*, the key generation algorithm* $\mathcal{K}$ *outputs a pair of proving key* pk *and verification key* vk.

- $\mathcal{P}(\mathsf{pk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{aux}_i, \pi_i) \rightarrow \pi_{i+1}$

On input the proving key pk*, an index* i*, an initial input* $\boldsymbol{z}_0$*, the claimed output* $\boldsymbol{z}_i$ *in the* i*-th iteration, the non-deterministic advice* aux_i*, and a proof* π_i *attesting to* $\boldsymbol{z}_i$*, the proof generation algorithm* $\mathcal{P}$ *outputs a new proof* π_{i+1} *that attests to* $\boldsymbol{z}_{i+1} = \mathcal{F}(\boldsymbol{z}_i; \mathsf{aux}_i)$*.*

- $\mathcal{V}(\mathsf{vk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i) =: b$

  On input the verification key vk*, an index* i*, an initial input* $\boldsymbol{z}_0$*, the claimed output* $\boldsymbol{z}_i$ *in the* i*-th iteration, and a proof* π_i *attesting to* $\boldsymbol{z}_i$*, the verification algorithm* $\mathcal{V}$ *outputs a bit* b*, indicating whether the proof is valid.*

An IVC scheme IVC satisfies the following properties.

Perfect Completeness. Completeness holds if for any PPT adversary $\mathcal{A}$,

$$\Pr\left[\begin{array}{l} \mathsf{pp} \leftarrow \mathcal{G}(\lambda), \\ (\mathcal{F}, i, \boldsymbol{z}_0, \boldsymbol{z}_i, \mathsf{aux}_i, \pi_i) \leftarrow \mathcal{A}(\mathsf{pp}) \\ (\mathsf{pk}, \mathsf{vk}) := \mathcal{K}(\mathsf{pp}, \mathcal{F}) \\ \boldsymbol{z}_{i+1} := \mathcal{F}(\boldsymbol{z}_i; \mathsf{aux}_i) \\ \mathcal{V}(\mathsf{vk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i) = 1 \\ \pi_{i+1} \leftarrow \mathcal{P}(\mathsf{pk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{aux}_i, \pi_i) : \\ \qquad \mathcal{V}(\mathsf{vk}, (i+1, \boldsymbol{z}_0, \boldsymbol{z}_{i+1}), \pi_{i+1}) = 1 \end{array}\right] = 1$$

Knowledge Soundness[1]. Knowledge soundness holds if, for every polynomial-time adversary $\mathcal{A}$, there exists a polynomial-time extractor Ext such that for all input randomness r,

$$\Pr\left[\begin{array}{l} \mathsf{pp} \leftarrow \mathcal{G}(\lambda), \\ (\mathcal{F}, (i \geq 1, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i) := \mathcal{A}(\mathsf{pp}; r) \\ (\mathsf{pk}, \mathsf{vk}) := \mathcal{K}(\mathsf{pp}, \mathcal{F}) \\ \mathcal{V}(\mathsf{vk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i) = 1 \\ (\boldsymbol{z}_{i-1}, \mathsf{aux}_{i-1}, \pi_{i-1}) := \mathsf{Ext}(\mathsf{pp}; r) : \\ \qquad\qquad\qquad \boldsymbol{z}_i = \mathcal{F}(\boldsymbol{z}_{i-1}; \mathsf{aux}_{i-1}) \\ \qquad \wedge \mathcal{V}(\mathsf{vk}, (i-1, \boldsymbol{z}_0, \boldsymbol{z}_{i-1}), \pi_{i-1}) = 1 \end{array}\right] \geq 1 - \varepsilon(\lambda)$$

Succinctness. Succinctness holds if the size of π_i and the run time of $\mathcal{P}$ and $\mathcal{V}$ are independent of the number of iterations.

Note that, unlike the definition of succinctness in SNARKs, a succinct IVC may have proof size and verifier time that are linear in the size of $\mathcal{F}$. In addition, an IVC is not necessarily zero-knowledge.

To achieve a fully succinct and zero-knowledge IVC, one can include an additional zkSNARK that compresses the proof while hiding the witnesses [28,31]. This concept is formalized as *decider*.

In the decider, we are interested in a relation R^{Decider} that encodes the IVC's verification algorithm. Formally, given statement $\boldsymbol{x} = (k, \boldsymbol{z}_0, \boldsymbol{z}_k)$ and witness $\boldsymbol{w} = (\pi_k)$, $R^{\mathsf{Decider}}(\boldsymbol{x}, \boldsymbol{w}) = 1$ if and only if $\mathsf{IVC}.\mathcal{V}(\mathsf{vk}_\Phi, (k, \boldsymbol{z}_0, \boldsymbol{z}_k), \pi_k) = 1$, where vk_Φ is the IVC verification key. With this relation, we can define a decider as follows, who has the same syntax and security properties as a zkSNARK.

Definition 6 (Decider)**.** *For an IVC scheme* IVC *whose verification algorithm* $\mathsf{IVC}.\mathcal{V}$ *is expressed as a relation* R^{Decider} *(as defined above), a step function* $\mathcal{F}$*,*

[1]We adopt [46]'s definition of knowledge soundness, which implies the notions in [28,30,31] that require Ext to extract all previous auxiliary values.

and a pair of proving and verification key $(\mathsf{pk}_\Phi, \mathsf{vk}_\Phi) \leftarrow \mathsf{IVC}.\mathcal{K}(\mathsf{pp}, \mathcal{F})$*, a decider* $\mathsf{Decider} = (\mathcal{K}, \mathcal{P}, \mathcal{V})$ *is a zkSNARK for* R^{Decider}.

Note that in our definition, the decider requires circuit-specific setup, which suffices for our application. That is, every structured reference string srs generated by $\mathsf{Decider}.\mathcal{K}$ is only for a specific $\mathcal{F}$ and vk_Φ.

3 Proofs of Video Authenticity

In this section, we formalize *proofs of video authenticity*, a category of video authentication protocols that are provably secure. We begin by describing the data types and operations involved, followed by the algorithm definition as well as the security properties.

3.1 Data Types and Operations

We first consider two forms of *video* data: the raw video $\boldsymbol{V}$ and the video stream ζ.

Raw Video. A *raw video* is usually for being displayed on a screen or edited by video processing software. It is composed of a series of *frames* ordered by time. Each frame is a still image described as a two dimensional matrix of *pixels*.

A pixel consists of several *components* that carry the properties of the pixel's color or luminance. For instance, three color components R, G, and B that respectively indicate the relative proportions of red, green, and blue make up the RGB color space. It is more common to use the YCbCr color space in video processing, where a pixel is represented by a luma component Y, a blue chroma component Cb, and a red chroma component Cr, each of which is an 8-bit[2] integer.

The chroma components are usually *subsampled* in practice to reduce the data size, and we assume a $4:2:0$ subsampling ratio, where both the horizontal and vertical resolutions of Cb and Cr are halved. Hence, in a frame with M rows and N columns, there are $M \times N$ Y components, $M/2 \times N/2$ Cb components, and $M/2 \times N/2$ Cr components.

The *resolution* of a frame with M rows and N columns of pixels is defined as $N \times M$[3], where N and M are also called the *width* and the *height* of the frame. The frequency at which the frames in a video are displayed is dubbed the frame rate, which is typically measured in frames per second (FPS). High image resolution and frame rate typically appear as higher quality video.

Moreover, in video processing, a frame is usually partitioned into *macroblocks* of size 16×16, which contains 16×16 bytes for Y and 8×8 bytes for both Cb and Cr, due to subsampling. Formally, we define a macroblock as $\boldsymbol{X} := (\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}}) \in \mathcal{B}$, where $\mathcal{B}$ is the set of all possible macroblocks, *i.e.*, $\mathcal{B} := \mathbb{Z}_{2^8}^{16\times16} \times \mathbb{Z}_{2^8}^{8\times8} \times \mathbb{Z}_{2^8}^{8\times8}$. In consequence, for a video with L frames, each of which has M rows and N columns, we write $\boldsymbol{V} := \{\boldsymbol{X}_i\}_{i=0}^{M/16\times N/16\times L-1} \in \mathcal{B}^{M/16\times N/16\times L}$.

Video Stream. Due to the large size, a raw video is compressed into a *video stream* when being transmitted over the network or stored in a file to reduce

[2]It is possible to have a higher bit depth (e.g., 10-bit) in video codecs, but we only discuss 8-bit color components for clarity.

[3]Note that while the frame resolution $N \times M$ is column-first, the frame matrix is still written in row-major order.

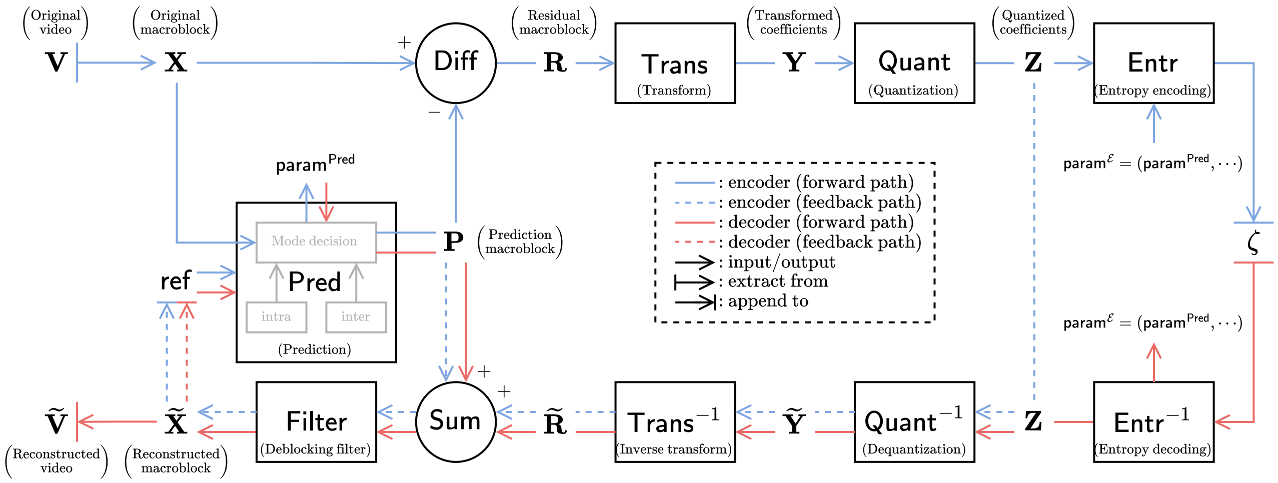


Figure 1: Block diagram of a macroblock-based video codec. Blue lines: data flow during encoding. Red lines: data flow during decoding. Solid lines: forward paths. Dashed lines: feedback paths for updating the reference information. Arrows with a vertical bar at the start: the data is extracted from the source. Arrows with a vertical bar at the end: the data is appended to the destination.

communication and storage costs. As a more compact form, a video stream is interpreted as a sequence of bits encapsulated in a *multimedia container* such as MP4, which may additionally include audio and subtitles. For simplicity, we focus solely on the visual part in this work.

Encoding and Decoding. In video codecs, a raw video $\boldsymbol{V}$ is converted to a video stream ζ by the *encoder*, whereas the *decoder* reconstructs a video $\widetilde{\boldsymbol{V}}$ from a video stream ζ. The codec is *lossless* if the decoder can reconstruct the original video exactly, and *lossy* if some information is discarded, resulting in a lower quality video in exchange for a smaller video stream.

Now we briefly review the general workflow of macroblock-based video codecs, e.g., H.264/AVC [37], H.265/HEVC [47], and AOMedia Video 1 (AV1) [48], as illustrated in Figure 1.

The encoder aims to reduce the file size by removing *redundant* and *non-essential* information from $\boldsymbol{V}$ while maintaining as much visual quality as possible. To this end, the encoder employs four stages for every macroblock $\boldsymbol{X}$ in $\boldsymbol{V}$: *prediction*, *transform*, *quantization*, and *entropy coding*, where only quantization may introduce loss of information, while the other stages are lossless.

1. During prediction Pred, the encoder generates a *prediction macroblock* $\boldsymbol{P}$ for $\boldsymbol{X}$, so that the difference between the original macroblock and the prediction macroblock is minimized. There are two types of prediction: *intra-frame prediction* that removes spatial redundancy within a frame (e.g., background areas with uniform colors or patterns), and *inter-frame prediction* that avoids temporal redundancy among multiple frames (e.g., stationary areas with no motion or moving objects with simple patterns) by leveraging motion estimation and motion compensation. Both prediction modes rely on some reference information ref, which we will discuss later.

   With all possible prediction results, we decide the final prediction mode by selecting the best result $\boldsymbol{P}$ whose difference between $\boldsymbol{X}$ is minimal. The final difference $\boldsymbol{X} - \boldsymbol{P}$ is output as the *residue macroblock* $\boldsymbol{R}$, and the prediction parameters $\mathsf{param}^{\mathsf{Pred}} = (\mathsf{mode}, \cdots)$ are also returned, where mode is the selected prediction mode ("inter" or "intra").

2. During transform Trans, the encoder further reduces the spatial redundancy by transforming the pixel data in the residue macroblock $\boldsymbol{R}$ to the frequency domain. In this way, we can obtain the low-frequency components representing essential features and the high-frequency components containing non-essential details.

   This process usually involves Discrete Cosine Transform (DCT) and Hadamard Transform, after which the *transformed coefficients* $\boldsymbol{Y}$ are forwarded to the next stage.

3. During quantization Quant, the precision of transformed coefficients $\boldsymbol{Y}$ is reduced, in order to discard non-essential information (e.g., perceptually hard-to-notice details) in the coefficients.

   Quantization is done by scaling and rounding the coefficients, obtaining the *quantized coefficients* $\boldsymbol{Z}$, where rounding is the main reason of information loss.

4. Finally, entropy coding Entr minimizes statistical redundancy in quantized coefficients $\boldsymbol{Z}$ by assigning shorter codes to more frequent elements, whereas less frequent data is mapped to longer codes. The encoding parameters $\mathsf{param}^{\mathcal{E}}$ are also compressed by Entr, which contains parameters used in the encoding process such as the prediction parameters $\mathsf{param}^{\mathsf{Pred}}$.

   Examples of entropy coding include Huffman coding and arithmetic coding. The output of entropy coding is appended to the bitstream ζ.

Note that the reference information ref used for predicting subsequent marcoblocks needs to be computed by *reconstructing* from already encoded data. This is generally done by reversing the encoding algorithm. Given the quantized coefficients $\boldsymbol{Z}$ encoded from $\boldsymbol{X}$ and its prediction $\boldsymbol{P}$, $\boldsymbol{Z}$ is first *dequantized* via Quant^{-1}, which returns the dequantized coefficients $\widetilde{\boldsymbol{Y}}$. Due to the loss of information during quantization, they are close but may not be equal to the original transformed coefficients, *i.e.*, $\widetilde{\boldsymbol{Y}} \approx \boldsymbol{Y}$. $\widetilde{\boldsymbol{Y}}$ is then *inverse transformed* via Trans^{-1} to obtain the residual macroblock $\widetilde{\boldsymbol{R}} \approx \boldsymbol{R}$. Next, we compute the sum of $\widetilde{\boldsymbol{R}}$ and the prediction macroblock $\boldsymbol{P}$, which is fed to an optional *deblocking filter* to get the reconstructed macroblock $\widetilde{\boldsymbol{X}} \approx \boldsymbol{X}$.

Reconstruction is also the core subroutine of the decoding process. Before reconstruction, the decoder extracts a subsequence from the bitstream ζ and applies entropy decoding on the subsequence to get the quantized coefficients $\boldsymbol{Z}$. The decoder then reconstructs the macroblock $\widetilde{X}$ in the same way as encoder, where the prediction macroblock $\boldsymbol{P}$ used for reconstruction is generated by performing the prediction operation Pred on input the previously reconstructed reference information ref. Since the prediction mode mode is encoded in the bitstream ζ, it is unnecesary to have $\boldsymbol{X}$ when choosing the prediction mode in Pred.

Formally, we define a block-wise encoding operation $\mathcal{E} : \mathcal{B} \times \{0,1\}^* \times \{0,1\}^* \to \mathcal{B} \times \{0,1\}^*$, which takes a macroblock $\boldsymbol{X}$, some reference information $\mathsf{ref} \in \{0,1\}^*$, and some encoding parameters $\mathsf{param}^{\mathcal{E}}$ as input, encodes $\boldsymbol{X}$ under $\mathsf{param}^{\mathcal{E}}$ with the help of ref, and outputs the reconstructed macroblock $\boldsymbol{X}'$ and the encoded bitstream $y \in \{0,1\}^*$. The encoding parameters $\mathsf{param}^{\mathcal{E}}$ control the quality and performance of the encoding process. In addition to prediction parameters $\mathsf{param}^{\mathsf{Pred}}$, we also include other configurations in $\mathsf{param}^{\mathcal{E}}$. For instance, in H.264, $\mathsf{param}^{\mathcal{E}}$ contains the quantization parameter qp, which determines the precision of quantized coefficients.

Conversely, the block-wise decoding operation $\mathcal{D} : \{0,1\}^* \times \{0,1\}^* \times \{0,1\}^* \to \mathcal{B}$ takes the encoded bitstream y, the reference information ref, and the encoding parameters $\mathsf{param}^{\mathcal{E}}$ as input, decodes y under $\mathsf{param}^{\mathcal{E}}$ with the help of ref, and outputs the reconstructed macroblock $\boldsymbol{X}'$.

Abusing the notation slightly, we allow applying $\mathcal{E}$ to the entire video $\boldsymbol{V}$ to obtain the encoded video stream $\zeta \coloneqq \mathcal{E}(\boldsymbol{V}, \mathsf{param}^{\mathcal{E}})$, and $\mathcal{D}$ to the encoded video stream ζ to obtain the decoded video $\boldsymbol{V}' \coloneqq \mathcal{D}(\zeta, \mathsf{param}^{\mathcal{E}})$. Further, we assume that one can extract intermediate data from $\mathcal{E}$ and $\mathcal{D}$, such as the prediction macroblock $\boldsymbol{P}$ and quantized coefficients $\boldsymbol{Z}$.

Metadata. Metadata meta is a set of information associated with the video, such as the author name, the recording device ID, the location and time of recording. We assume that meta is immutable.

Editing. A block-wise editing operation Δ is defined as $\Delta : \mathcal{B} \times \{0,1\}^* \to \mathcal{B}$, which takes a macroblock $\boldsymbol{X}$ and some editing parameters $\mathsf{param}^{\Delta} \in \{0,1\}^*$ as input, edits $\boldsymbol{X}$, and outputs the edited macroblock $\boldsymbol{X}'$. param^{Δ} contains configurations specific to the editing operation, such as the brightness level, the position of an overlay mask, etc.

While one may also edit the metadata meta of a video in practice, we assume that meta is immutable in our definition. This assumption does not invalidate editing operations such as cropping and cutting, since the resolution and frame rate are regarded as part of the encoded video stream ζ.

3.2 Algorithm and Security Definitions

A proof of video authenticity involves four parties: the *trusted party*, the *recorder*, the *prover*, and the *verifier*.

- The trusted party (e.g., a manufacturer) runs the key generation algorithms $\mathcal{K}_\Sigma$ and $\mathcal{K}_\Pi$, where $\mathcal{K}_\Sigma$ generates signing keys for the recorders, and $\mathcal{K}_\Pi$ produces necessary parameters for proof generation and verification. The signing keys are then securely provisioned to the recorders and are safely protected using mechanisms such as secure enclaves.

- The recorder (e.g., a camcorder) records the original video, generates the metadata, and runs the recording algorithm $\mathcal{R}$, which signs the video and the metadata under the signing key.

- The prover (e.g., a content creator) edits and encodes the original video, publishes the processed video, and generates a proof of authenticity using the proof generation algorithm $\mathcal{P}$.

- The verifier (e.g., a website visitor) checks if the proof is valid w.r.t. the video by executing the verification algorithm $\mathcal{V}$.

Now we formally define the algorithms discussed above in a proof of video authenticity.

Definition 7 (Proof of Video Authenticity)**.** *A* proof of video authenticity *is defined as* $\mathsf{VA} = (\mathcal{K}_\Sigma, \mathcal{K}_\Pi, \mathcal{R}, \mathcal{P}, \mathcal{V})$*:*

- $\mathcal{K}_\Sigma(1^\lambda) \to (\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma)$

  $\mathcal{K}_\Pi(1^\lambda) \to (\mathsf{pk}_\Pi, \mathsf{vk}_\Pi)$

  Both key generation algorithms $\mathcal{K}_\Sigma$ and $\mathcal{K}_\Pi$ take as input security parameter 1^λ. $\mathcal{K}_\Sigma$ outputs a pair of secret signing key sk_Σ and public signature verification key vk_Σ, and $\mathcal{K}_\Pi$ outputs a pair of public proving key pk_Π and public proof verification key vk_Π. $\mathcal{K}_\Pi$ also returns the secret trapdoor td, which is omitted from the notation for simplicity but is used in security definitions.

- $\mathcal{R}(\mathsf{sk}_\Sigma, \boldsymbol{V}, \mathsf{meta}) \to \sigma$

  The recording algorithm $\mathcal{R}$ takes as input signing key sk_Σ, video $\boldsymbol{V}$ and its metadata meta, and outputs a signature σ on $\boldsymbol{V}$ and meta.

- $\mathcal{P}(\mathsf{pk}_\Pi, \mathsf{vk}_\Sigma, \boldsymbol{V}, \mathsf{meta}, \mathsf{param}, \sigma) \to (\zeta, \pi)$

  The proof generation algorithm $\mathcal{P}$ takes as input proving key pk_Π, signature verification key vk_Σ, video $\boldsymbol{V}$, metadata meta, editing and encoding parameters $\mathsf{param} = (\mathsf{param}^\Delta, \mathsf{param}^\mathcal{E})$, and signature σ. It outputs a video stream ζ and a proof π that attests to 1) the honesty of the editing and encoding process from $\boldsymbol{V}$ to ζ under param, and 2) the validity of σ on $(\boldsymbol{V}, \mathsf{meta})$ under vk_Σ.

- $\mathcal{V}(\mathsf{vk}_\Pi, \mathsf{vk}_\Sigma, \zeta, \mathsf{meta}, \mathsf{param}, \pi) =: b$

  The verification algorithm $\mathcal{V}$ takes as input proof verification key vk_Π, signature verification key vk_Σ, processed video stream ζ and its metadata meta', editing and encoding parameters param, and proof π, and outputs a bit b indicating if the proof is valid for ζ, meta and vk_Σ.

Now we formalize the security of VA. Consider the relation $R^{\mathsf{VA}}(\boldsymbol{x}, \boldsymbol{w})$ for the authenticity of a video, where $\boldsymbol{x} = (\zeta, \mathsf{meta}, \mathsf{param}, \mathsf{vk}_\Sigma), \boldsymbol{w} = (\sigma, \boldsymbol{V})$. For a signature scheme Sig, an editing operation Δ, and an encoder $\mathcal{E}$, $R^{\mathsf{VA}}(\boldsymbol{x}, \boldsymbol{w}) = 1$ if and only if

$$\mathsf{Sig}.\mathcal{V}(\mathsf{vk}_\Sigma, \sigma, (\boldsymbol{V}, \mathsf{meta})) = 1 \wedge \zeta = \mathcal{E}(\Delta(\boldsymbol{V}, \mathsf{param}^\Delta), \mathsf{param}^\mathcal{E})$$

The security of VA is defined below, which can be regarded as the security of zkSNARKs for R^{VA}.

COMPLETENESS. Completeness holds if for every video $\boldsymbol{V}$, metadata meta, and editing and encoding parameters param,

$$\Pr\left[\begin{array}{l} (\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma) \leftarrow \mathcal{K}_\Sigma(1^\lambda) \\ (\mathsf{pk}_\Pi, \mathsf{vk}_\Pi) \leftarrow \mathcal{K}_\Pi(1^\lambda) \\ \sigma \leftarrow \mathcal{R}(\mathsf{sk}_\Sigma, \boldsymbol{V}, \mathsf{meta}) \\ (\zeta, \pi) \leftarrow \mathcal{P}(\mathsf{pk}_\Pi, \mathsf{vk}_\Sigma, \boldsymbol{V}, \mathsf{meta}, \mathsf{param}, \sigma) \\ R^{\mathsf{VA}}((\zeta, \mathsf{meta}, \mathsf{param}, \mathsf{vk}_\Sigma), (\sigma, \boldsymbol{V})) = 1 : \\ \qquad\qquad \mathcal{V}(\mathsf{vk}_\Pi, \mathsf{vk}_\Sigma, \zeta, \mathsf{meta}, \mathsf{param}, \pi) = 1 \end{array}\right] = 1$$

KNOWLEDGE SOUNDNESS. Knowledge soundness holds if for every polynomial-time adversary $\mathcal{A}$, there is a polynomial-time extractor Ext such that for all input randomness r,

$$\Pr\left[\begin{array}{l}(\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma) \leftarrow \mathcal{K}_\Sigma(1^\lambda) \\ (\mathsf{pk}_\Pi, \mathsf{vk}_\Pi, \mathsf{td}) \leftarrow \mathcal{K}_\Pi(1^\lambda) \\ (\zeta, \mathsf{meta}, \mathsf{param}, \pi) := \mathcal{A}(\mathsf{pk}_\Pi, \mathsf{vk}_\Pi, \mathsf{vk}_\Sigma; r) \\ (\sigma, \boldsymbol{V}) := \mathsf{Ext}(\mathsf{pk}_\Pi, \mathsf{vk}_\Pi, \mathsf{vk}_\Sigma, \mathsf{td}; r) \\ \mathcal{V}(\mathsf{vk}_\Pi, \mathsf{vk}_\Sigma, \zeta, \mathsf{meta}, \mathsf{param}, \pi) = 1 : \\ \qquad R^{\mathsf{VA}}((\zeta, \mathsf{meta}, \mathsf{param}, \mathsf{vk}_\Sigma), (\sigma, \boldsymbol{V})) = 0\end{array}\right] \leq \varepsilon(\lambda)$$

Zero-Knowledge. Optionally, VA may satisfy the zero-knowledge property, which holds if there exists a simulator Sim such that for every polynomial-time distinguisher $\mathcal{A}$,

$$\begin{aligned}&\Pr\left[\begin{array}{l}(\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma) \leftarrow \mathcal{K}_\Sigma(1^\lambda) \\ (\mathsf{pk}_\Pi, \mathsf{vk}_\Pi, \mathsf{td}) \leftarrow \mathcal{K}_\Pi(1^\lambda) \\ ((\zeta, \mathsf{meta}, \mathsf{param}, \mathsf{vk}_\Sigma), (\sigma, \boldsymbol{V})) \leftarrow \mathcal{A}(\mathsf{vk}_\Sigma, \mathsf{pk}_\Pi, \mathsf{vk}_\Pi) \\ (\cdot, \pi) \leftarrow \mathcal{P}(\mathsf{pk}_\Pi, \mathsf{vk}_\Sigma, \boldsymbol{V}, \mathsf{meta}, \mathsf{param}, \sigma) : \\ \hfill \mathcal{A}(\pi) = 1\end{array}\right] \\ \approx &\Pr\left[\begin{array}{l}(\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma) \leftarrow \mathcal{K}_\Sigma(1^\lambda) \\ (\mathsf{pk}_\Pi, \mathsf{vk}_\Pi, \mathsf{td}) \leftarrow \mathcal{K}_\Pi(1^\lambda) \\ ((\zeta, \mathsf{meta}, \mathsf{param}, \mathsf{vk}_\Sigma), (\sigma, \boldsymbol{V})) \leftarrow \mathcal{A}(\mathsf{vk}_\Sigma, \mathsf{pk}_\Pi, \mathsf{vk}_\Pi) \\ \pi \leftarrow \mathsf{Sim}(\mathsf{td}, \mathsf{pk}_\Pi, \mathsf{vk}_\Sigma, \mathsf{meta}, \mathsf{param}, \zeta) : \\ \hfill \mathcal{A}(\pi) = 1\end{array}\right]\end{aligned}$$

Succinctness. Optionally, VA may produce succinct proofs, if for every video $\boldsymbol{V}$ of dimension $M \times N \times L$, the proof π for $\boldsymbol{V}$ is of size $|\pi| = \mathrm{poly}(\lambda)\,\mathrm{polylog}(MNL)$.

4 Nova-Based IVC with Lookup Argument

What lies at the heart of Eva is an IVC scheme based on folding that supports efficient lookup. Our IVC scheme is heavily inspired by the implementation of Nova [28] based IVC in the `sonobe` library [38]. However, we extend it with LogUp [39], an efficient lookup argument, and equip it with CP-SNARKs to achieve linear time prover while maintaining the constant proof size and verifier time in `sonobe`'s IVC.

In this section, we briefly review the approaches in Nova and `sonobe` while highlighting our techniques for integrating LogUp and CP-SNARKs into our IVC scheme. The security proofs are deferred to Appendix B.

The high-level idea behind our IVC scheme is to 1) split off $\boldsymbol{q}$, the queries to the lookup table, from the witnesses $\boldsymbol{w}$, 2) fold their commitments $\overline{Q}$ and $\overline{W}$ separately, and 3) check the folded instances as well as the lookup relation against the same $\overline{Q}$ in IVC's augmented step circuit, thereby linking the folding scheme with the lookup argument.

4.1 NIFS

To elaborate on the intuition above, we first describe our folding scheme modified from Nova.

Recall that in Nova, we consider *committed relaxed R1CS*, which is a variant of the Rank-1 Constraint System (R1CS) [49]. Similar to R1CS, a committed relaxed R1CS over $\mathbb{F}$ with n constraints and m variables (among which 1 variable is constant and l variables are public inputs) is defined by three matrices $\mathsf{CS} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \in (\mathbb{F}^{n\times m}, \mathbb{F}^{n\times m}, \mathbb{F}^{n\times m})$. A witness $\mathbb{W}$ to CS not only consists of the witness $\boldsymbol{w} \in \mathbb{F}^{m-l-1}$ to the original R1CS relation, but also includes an *error term* $\boldsymbol{e} \in \mathbb{F}^n$. The instance $\mathbb{U}$ corresponding to $\mathbb{W}$ is a tuple $(u, \boldsymbol{x}, \overline{W}, \overline{E})$, where $u \in \mathbb{F}$ is a scalar for absorbing constant terms, $\boldsymbol{x} \in \mathbb{F}^l$ is the public input, and $\overline{W}, \overline{E}$ are the commitments to $\boldsymbol{w}, \boldsymbol{e}$ respectively. We say $(\mathbb{U}, \mathbb{W})$ satisfies CS if $\mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{w}, \overline{W}) = 1$, $\mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{e}, \overline{E}) = 1$, and $\boldsymbol{A}\boldsymbol{v} \circ \boldsymbol{B}\boldsymbol{v} = u \cdot \boldsymbol{C}\boldsymbol{v} + \boldsymbol{e}$, where $\boldsymbol{v} = (u, \boldsymbol{x}, \boldsymbol{w})$.

In our modification, we consider $\boldsymbol{\tau} = \{\tau_j\}_{j=0}^{\nu-1}$, a read-only lookup table with ν entries. We assume among the witnesses in CS, there are μ queries $\boldsymbol{\alpha} = \{\alpha_i\}_{i=0}^{\mu-1}$ to the lookup table. Further, for the j-th table entry τ_j, we count o_j, the number of τ_j's occurrences in the query vector $\boldsymbol{\alpha}$. The witness $\mathbb{W}$ is now defined as $\mathbb{W} = (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e})$, where $\boldsymbol{q} = \{\alpha_i\}_{i=0}^{\mu-1} \cup \{o_j\}_{j=0}^{\nu-1} \in \mathbb{F}^{\mu+\nu}$, $\boldsymbol{w} \in \mathbb{F}^{m-\mu-\nu-l-1}$, and $\boldsymbol{e} \in \mathbb{F}^n$. Consequently, the instance $\mathbb{U} = (u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E})$ also contains an additional term $\overline{Q}$, *i.e.*, the commitment to $\boldsymbol{q}$, and the list of variables in the constraint system becomes $\boldsymbol{v} = (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$. For notational convenience, we write the part of field elements in $\mathbb{U}$ as $\mathbb{U}^{\mathbb{F}} = (u, \boldsymbol{x})$, and the part of group elements (commitments) as $\mathbb{U}^{\mathbb{G}} = (\overline{Q}, \overline{W}, \overline{E})$.

We adapt Nova's construction accordingly. In the interactive setting, to fold $(\mathbb{U}_1, \mathbb{W}_1)$ and $(\mathbb{U}_2, \mathbb{W}_2)$, $\mathcal{P}$ first computes the cross term $\boldsymbol{t}$ and sends the commitment $\overline{T}$ to $\mathcal{V}$, who samples and sends back a challenge r. Then, both parties output the folded instance $\mathbb{U}$ by computing the random linear combination of $\mathbb{U}_1$ and $\mathbb{U}_2$ component-wise. $\mathcal{P}$ further output the folded witness $\mathbb{W}$, which is also a random linear combination of $\mathbb{W}_1$ and $\mathbb{W}_2$. One can apply the Fiat-Shamir transform [50] to obtain the non-interactive construction, as given in Algorithm 1.

4.2 IVC

The next step is to design an IVC scheme utilizing our variant of Nova and integrate it with lookup arguments.

To this end, we first recap how a folding scheme is converted to an IVC in [28,51]. Given a step circuit $\mathcal{F}$, we construct an *augmented* step circuit $\mathcal{F}^{\mathsf{aug}}$, which is associated with two types of instance-witness pairs $(\mathbb{U}, \mathbb{W})$ and $(\mathbbm{u}, \mathbbm{w})$, where $(\mathbb{U}, \mathbb{W})$ is a *running* instance-witness pair, and $(\mathbbm{u}, \mathbbm{w})$ is an *incoming* instance-witness pair.

In the i-th step of IVC, $(\mathbbm{u}_i, \mathbbm{w}_i)$ that represents the execution of $\mathcal{F}^{\mathsf{aug}}$ in the $(i-i)$-th step is folded into $(\mathbb{U}_i, \mathbb{W}_i)$, producing an updated running instance-witness pair $(\mathbb{U}_{i+1}, \mathbb{W}_{i+1})$ that represents all previous invocations of $\mathcal{F}^{\mathsf{aug}}$. Then, $\mathcal{F}^{\mathsf{aug}}$ is invoked again, which not only includes the original step circuit $\mathcal{F}$, but also enforces the correct folding of $\mathbb{U}_i, \mathbbm{u}_i$ by running $\mathsf{NIFS}.\mathcal{V}$.

Recall that $\mathsf{NIFS}.\mathcal{V}$ computes the random linear combination of two commitments, *i.e.*, $\overline{X} := \overline{X}_1 + r\overline{X}_2$. However, in our setting, $\overline{X}$'s coordinates are over the base field of $\mathbb{G}$, while the circuit is defined over the scalar field of $\mathbb{G}$, meaning that expensive non-native operations are necessary to compute $\overline{X}$ in-circuit. A common solution [46] is to deploy two augmented step circuits $\mathcal{F}_1^{\mathsf{aug}}, \mathcal{F}_2^{\mathsf{aug}}$ on a cycle of curves $(\mathbb{G}, \mathbb{H})$, where the circuit on one curve is responsible for folding instances from the other curve. Nevertheless, this approach is suboptimal, as it requires additional costs for encoding $\mathsf{NIFS}.\mathcal{V}$, a necessary component in the augmented step circuit, on the secondary curve $\mathbb{H}$.

Algorithm 1: NIFS

Fn $\mathsf{NIFS}.\mathcal{G}(1^\lambda)$**:**
  $\mathsf{ck} \leftarrow \mathsf{CM}.\mathcal{K}(1^\lambda)$
  return $\mathsf{pp} := \mathsf{ck}$

Fn $\mathsf{NIFS}.\mathcal{K}(\mathsf{pp}, \mathsf{CS})$**:**
  Parse $\mathsf{ck} := \mathsf{pp}$
  return $\mathsf{pk} := (\mathsf{ck}, \mathsf{CS}), \mathsf{vk} := \perp$

Fn $\mathsf{NIFS}.\mathcal{P}(\mathsf{pk}, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2))$**:**
  Parse $(\mathsf{ck}, (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})) := \mathsf{pk}$
  for $i \in \{1, 2\}$ **do**
    Parse $(u_i, \boldsymbol{x}_i, \overline{Q}_i, \overline{W}_i, \overline{E}_i) := \mathbb{U}_i, (\boldsymbol{q}_i, \boldsymbol{w}_i, \boldsymbol{e}_i) := \mathbb{W}_i$
    $\boldsymbol{v}_i := (u_i, \boldsymbol{x}_i, \boldsymbol{q}_i, \boldsymbol{w}_i)$
  $\boldsymbol{t} := \boldsymbol{A}\boldsymbol{v}_1 \circ \boldsymbol{B}\boldsymbol{v}_2 + \boldsymbol{A}\boldsymbol{v}_2 \circ \boldsymbol{B}\boldsymbol{v}_1 - u_1 \cdot \boldsymbol{C}\boldsymbol{v}_2 - u_2 \cdot \boldsymbol{C}\boldsymbol{v}_1$
  $\overline{T} \leftarrow \mathsf{CM}.\mathcal{C}(\mathsf{ck}, \boldsymbol{t})$
  $r := \rho(\mathbb{U}_1, \mathbb{U}_2, \overline{T})$ ▷ Compute challenge
  Fold instances:
    $u := u_1 + ru_2, \boldsymbol{x} := \boldsymbol{x}_1 + r\boldsymbol{x}_2$
    $\overline{Q} := \overline{Q}_1 + r\overline{Q}_2, \overline{W} := \overline{W}_1 + r\overline{W}_2, \overline{E} := \overline{E}_1 + r\overline{T} + r^2\overline{E}_2$
  Fold witnesses:
    $\boldsymbol{q} := \boldsymbol{q}_1 + r\boldsymbol{q}_2, \boldsymbol{w} := \boldsymbol{w}_1 + r\boldsymbol{w}_2, \boldsymbol{e} := \boldsymbol{e}_1 + r\boldsymbol{t} + r^2\boldsymbol{e}_2$
  return $\mathbb{U} := (u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E}), \mathbb{W} := (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}), \overline{T}$

Fn $\mathsf{NIFS}.\mathcal{V}(\mathsf{vk}, \mathbb{U}_1, \mathbb{U}_2, \overline{T})$**:**
  for $i \in \{1, 2\}$ **do**
    Parse $(u_i, \boldsymbol{x}_i, \overline{Q}_i, \overline{W}_i, \overline{E}_i) := \mathbb{U}_i$
  $r := \rho(\mathbb{U}_1, \mathbb{U}_2, \overline{T})$ ▷ Compute challenge
  Fold instances:
    $u := u_1 + ru_2, \boldsymbol{x} := \boldsymbol{x}_1 + r\boldsymbol{x}_2$
    $\overline{Q} := \overline{Q}_1 + r\overline{Q}_2, \overline{W} := \overline{W}_1 + r\overline{W}_2, \overline{E} := \overline{E}_1 + r\overline{T} + r^2\overline{E}_2$
  return $\mathbb{U} := (u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E})$

CycleFold [51] aims to minimize the costs on $\mathbb{H}$, which is adopted by `sonobe` [38]. In this paradigm, we offload the heavy lifting of non-native operations on the primary curve $\mathbb{G}$ to a lightweight circuit on $\mathbb{H}$, which can handle them natively, thereby avoiding the need for duplicating the entire $\mathsf{NIFS}.\mathcal{V}$.

It starts by splitting $\mathsf{NIFS}.\mathcal{V}$ into two parts:

- $\mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}$ folds field elements in $\mathbb{U}_1^{\mathbb{F}}$ and $\mathbb{U}_2^{\mathbb{F}}$.

  On input vk, $\mathbb{U}_1^{\mathbb{F}} = (u_1, \boldsymbol{x}_1)$, $\mathbb{U}_2^{\mathbb{F}} = (u_2, \boldsymbol{x}_2)$ and r, it computes $u := u_1 + ru_2$, $\boldsymbol{x} := \boldsymbol{x}_1 + r\boldsymbol{x}_2$ and returns $\mathbb{U}^{\mathbb{F}} := (u, \boldsymbol{x})$.

- $\mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}$ folds group elements in $\mathbb{U}_1^{\mathbb{G}}$ and $\mathbb{U}_2^{\mathbb{G}}$.

  On input vk, $\mathbb{U}_1^{\mathbb{G}} = (\overline{Q}_1, \overline{W}_1, \overline{E}_1)$, $\mathbb{U}_2^{\mathbb{G}} = (\overline{Q}_2, \overline{W}_2, \overline{E}_2)$, r and $\overline{T}$, it computes $\overline{Q} := \overline{Q}_1 + r\overline{Q}_2, \overline{W} := \overline{W}_1 + r\overline{W}_2, \overline{E} := \overline{E}_1 + r\overline{T} + r^2\overline{E}_2$ and returns $\mathbb{U}^{\mathbb{G}} := (\overline{Q}, \overline{W}, \overline{E})$.

Then, we construct a CycleFold circuit $\mathcal{F}^{\mathsf{cf}}$ on $\mathbb{H}$ (see Circuit 2) that performs the check $\mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}$ natively, which requires only ~ 4500 constraints[4]. Denote $(\mathbb{U}^{\mathsf{cf}}, \mathbb{W}^{\mathsf{cf}})$ and $(\mathbb{u}^{\mathsf{cf}}, \mathbb{w}^{\mathsf{cf}})$ respectively as the running and incoming instance-witness pairs for $\mathcal{F}^{\mathsf{cf}}$.

Circuit 2: $\mathcal{F}^{\mathsf{cf}}$

Statement: $r, \mathbb{U}_i^{\mathbb{G}}, \mathbb{u}_i^{\mathbb{G}}, \mathbb{U}_{i+1}^{\mathbb{G}}, \overline{T}$
1 **enforce** $\mathbb{U}_{i+1}^{\mathbb{G}} = \mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}(\mathsf{vk}, \mathbb{U}_i^{\mathbb{G}}, \mathbb{u}_i^{\mathbb{G}}, r, \overline{T})$

With the help of $\mathcal{F}^{\mathsf{cf}}$, the augmented circuit $\mathcal{F}^{\mathsf{aug}}$ only needs to fold the field parts of primary instances. As a trade-off, $\mathcal{F}^{\mathsf{aug}}$ becomes responsible for enforcing the correct folding of CycleFold instances $\mathbb{U}_i^{\mathsf{cf}}, \mathbb{u}_i^{\mathsf{cf}}$ using $\mathsf{NIFS}.\mathcal{V}$. Since $\mathbb{U}_i^{\mathsf{cf}}, \mathbb{u}_i^{\mathsf{cf}}$ are over $\mathbb{H}$, the group operations in $\mathsf{NIFS}.\mathcal{V}$ can be handled natively by $\mathcal{F}^{\mathsf{aug}}$ over $\mathbb{G}$. Although the field elements in $\mathbb{U}_i^{\mathsf{cf}}, \mathbb{u}_i^{\mathsf{cf}}$ become non-native, emulating non-native field operations is much cheaper than non-native group operations, thanks to the techniques in [52].

To summarize, with $\boldsymbol{z}_i$ as the state of IVC in the i-th step, the $\mathcal{F}^{\mathsf{aug}}$ circuit computes $\boldsymbol{z}_{i+1} \coloneqq \mathcal{F}(\boldsymbol{z}_i)$, folds $\mathbb{u}_i^{\mathbb{F}}$ into $\mathbb{U}_i^{\mathbb{F}}$ by running $\mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}(\mathsf{vk}, \mathbb{U}_i^{\mathbb{F}}, \mathbb{u}_i^{\mathbb{F}}, r)$, and folds $\mathbb{u}_i^{\mathsf{cf}}$ into $\mathbb{U}_i^{\mathsf{cf}}$ by running $\mathsf{NIFS}.\mathcal{V}(\mathsf{vk}^{\mathsf{cf}}, \mathbb{U}_i^{\mathsf{cf}}, \mathbb{u}_i^{\mathsf{cf}}, \overline{T}^{\mathsf{cf}})$.

Furthermore, to ensure the consistency of instances between two steps, $\mathcal{F}^{\mathsf{aug}}$ computes $\mathsf{H}(\mathbb{U}_{i+1}, i+1, \boldsymbol{z}_0, \boldsymbol{z}_{i+1})$ and $\mathsf{H}(\mathbb{U}_{i+1}^{\mathsf{cf}}, i+1)$ in the i-th step, treats the digests as outputs, and checks them in the $i+1$-th step by running H in a similar way.

Then we add support for LogUp [39] to $\mathcal{F}^{\mathsf{aug}}$. Inspired by `gnark` [53], which incorporates LogUp into Groth16 [54] and Plonk [55], our $\mathcal{F}^{\mathsf{aug}}$ additionally checks the set inclusion identity [39, Lemma 5] in-circuit.

Suppose $\mathcal{F}$, during its execution, makes queries $\boldsymbol{\alpha} = \{\alpha_i\}_{i=0}^{\mu-1}$ to a lookup table with entries $\boldsymbol{\tau} = \{\tau_j\}_{j=0}^{\nu-1}$. As per LogUp, $\{\alpha_i\}_{i=0}^{\mu-1} \subseteq \{\tau_j\}_{j=0}^{\nu-1}$ holds if and only if there is a set of multiplicities $\boldsymbol{o} = \{o_j\}_{j=0}^{\nu-1}$ such that the below identity for set inclusion holds:

$$\sum_{i=0}^{\mu-1} \frac{1}{X - \alpha_i} = \sum_{j=0}^{\nu-1} \frac{o_j}{X - \tau_j}.$$

By Schwartz-Zippel Lemma, we can check this polynomial identity by evaluating it at a random point $X = c$. Here, c can be the random message from $\mathcal{V}$ after receiving the commitment $\overline{Q}$ to $\boldsymbol{q} = \{\alpha_i\}_{i=0}^{\mu-1} \cup \{o_j\}_{j=0}^{\nu-1}$ from $\mathcal{P}$. Thanks to Fiat-Shamir transform, we can eliminate the interaction and compute $c \coloneqq \rho(\overline{Q})$ instead. To do the final check in-circuit, $\mathcal{P}$ needs to feed $\{o_j\}_{j=0}^{\nu-1}$ and c as hints to $\mathcal{F}^{\mathsf{aug}}$, and $\mathcal{F}^{\mathsf{aug}}$ can then enforce $\sum_{i=0}^{\mu-1} \frac{1}{c-\alpha_i} = \sum_{j=0}^{\nu-1} \frac{o_j}{c-\tau_j}$.

Note that to ensure c is honestly chosen, $\mathcal{F}^{\mathsf{aug}}$ needs to check if $c = \rho(\overline{Q})$. However, $\overline{Q}$ is a part of $\mathbb{u}_{i+1}$, which is unknown to $\mathcal{F}^{\mathsf{aug}}$ in the i-th step. Thus, while c is a hint dynamically computed by $\mathcal{P}$ in the i-th step, we still mark it as a public input, which is going to be included in $\mathbb{u}_{i+1}.\boldsymbol{x}$. With $\mathbb{u}_{i+1}.\overline{Q}$ and $\mathbb{u}_{i+1}.\boldsymbol{x}$, $\mathcal{F}^{\mathsf{aug}}$ in the $(i+1)$-th step can now check the honesty of c in the i-th step.

[4]In `sonobe`, $\mathcal{F}^{\mathsf{cf}}$ is split into three parts, each with ~ 1500 constraints, and this is also the case for our variant. We omit this detail for clarify.

We provide the construction of the augmented step circuit $\mathcal{F}^{\mathsf{aug}}$ in Circuit 3. Note that because $\mathcal{F}^{\mathsf{cf}}$ makes no queries to the lookup table $\boldsymbol{\tau}$, we have $\mathbb{w}^{\mathsf{cf}}.\boldsymbol{q} = \varnothing$, $\mathbb{u}^{\mathsf{cf}}.\overline{Q} = \overline{0}$.

Circuit 3: $\mathcal{F}^{\mathsf{aug}}$

Witness: $i, \boldsymbol{z}_i, \mathbb{U}_i, \mathbb{u}_i, \mathbb{U}^{\mathbb{G}}_{i+1}, \overline{T}, \mathbb{U}^{\mathsf{cf}}_i, \mathbb{u}^{\mathsf{cf}}_i, \overline{T}^{\mathsf{cf}}, \mathsf{aux}_i$
Statement: h_1, h_2, c
Constant: $\{\tau_j\}_{j=0}^{\nu-1}$

$\boldsymbol{z}_{i+1} \coloneqq \mathcal{F}(\boldsymbol{z}_i; \mathsf{aux}_i)$ ▷ Let $\{\alpha_i\}_{i=0}^{\mu-1}$ be queries made by $\mathcal{F}$
Check $\mathbb{u}_i$:
- **enforce** $\mathbb{u}_i.u = 1$
- **enforce** $\mathbb{u}_i.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_i, i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{H}(\mathbb{U}^{\mathsf{cf}}_i, i), \rho(\mathbb{u}_i.\overline{Q}))$
- **enforce** $\mathbb{u}_i.\overline{E} = \overline{0}$

$r \coloneqq \rho(\mathbb{U}_i, \mathbb{u}_i, \overline{T})$
$\mathbb{U}^{\mathbb{F}}_{i+1} \coloneqq \mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}(\mathsf{vk}, \mathbb{U}^{\mathbb{F}}_i, \mathbb{u}^{\mathbb{F}}_i, r)$
Check $\mathbb{u}^{\mathsf{cf}}_i$:
- **enforce** $\mathbb{u}^{\mathsf{cf}}_i.u = 1$
- **enforce** $\mathbb{u}^{\mathsf{cf}}_i.\boldsymbol{x} = (r, \mathbb{U}^{\mathbb{G}}_i, \mathbb{u}^{\mathbb{G}}_i, \mathbb{U}^{\mathbb{G}}_{i+1}, \overline{T})$
- **enforce** $\mathbb{u}^{\mathsf{cf}}_i.\overline{E} = \overline{0}$

$\mathbb{U}^{\mathsf{cf}}_{i+1} \coloneqq \mathsf{NIFS}.\mathcal{V}(\mathsf{vk}^{\mathsf{cf}}, \mathbb{U}^{\mathsf{cf}}_i, \mathbb{u}^{\mathsf{cf}}_i, \overline{T}^{\mathsf{cf}})$
Check lookup queries:
- $\{o_j\}_{j=0}^{\nu-1} \leftarrow \mathsf{Hint}(\{\alpha_i\}_{i=0}^{\mu-1})$
- $c \leftarrow \mathsf{Hint}(\{\alpha_i\}_{i=0}^{\mu-1} \cup \{o_j\}_{j=0}^{\nu-1})$
- **enforce** $\sum_{i=0}^{\mu-1} \frac{1}{c-\alpha_i} = \sum_{j=0}^{\nu-1} \frac{o_j}{c-\tau_j}$

Check public inputs:
- **enforce** $h_1 = \mathsf{H}((i = 0)\ ?\ \mathbb{U}_\perp : \mathbb{U}_{i+1}, i+1, \boldsymbol{z}_0, \boldsymbol{z}_{i+1})$
- **enforce** $h_2 = \mathsf{H}((i = 0)\ ?\ \mathbb{U}^{\mathsf{cf}}_\perp : \mathbb{U}^{\mathsf{cf}}_{i+1}, i+1)$

Now we are ready to present the full construction of IVC, which is illustrated in Algorithm 4. In the setup algorithm $\mathsf{IVC}.\mathcal{G}$, two commitment keys ck and $\mathsf{ck}^{\mathsf{cf}}$ are generated, one for primary instances, and the other for CycleFold instances. The key generation algorithm $\mathsf{IVC}.\mathcal{K}$ takes a step function $\mathcal{F}$ as input, creates the augmented function $\mathcal{F}^{\mathsf{aug}}$ for $\mathcal{F}$, and converts $\mathcal{F}^{\mathsf{aug}}$ and $\mathcal{F}^{\mathsf{cf}}$ to committed relaxed R1CS instances $\mathsf{CS}^{\mathsf{aug}}$ and $\mathsf{CS}^{\mathsf{cf}}$, respectively. Then, $\mathsf{IVC}.\mathcal{K}$ invokes $\mathsf{NIFS}.\mathcal{K}$ for each R1CS instance to obtain the proving and verification keys for them.

Before the proof generation $\mathsf{IVC}.\mathcal{P}$ starts, $\mathcal{P}$ first prepares two empty running instance-witness pairs $(\mathbb{U}_0 \coloneqq \mathbb{U}_\perp, \mathbb{W}_0 \coloneqq \mathbb{W}_\perp)$ and $(\mathbb{U}^{\mathsf{cf}}_0 \coloneqq \mathbb{U}^{\mathsf{cf}}_\perp, \mathbb{W}^{\mathsf{cf}}_0 \coloneqq \mathbb{W}^{\mathsf{cf}}_\perp)$. Moreover, the incoming instance-witness pair in the 0-th step is also $(\mathbb{u}_0 \coloneqq \mathbb{U}_\perp, \mathbb{w}_0 \coloneqq \mathbb{W}_\perp)$. $\mathcal{P}$ then proceeds to the incremental proof generation.

In the i-th iteration, $\mathcal{P}$ first folds the incoming $(\mathbb{u}_i, \mathbb{w}_i)$ into $(\mathbb{U}_i, \mathbb{W}_i)$ and obtains $(\mathbb{U}_{i+1}, \mathbb{W}_{i+1})$, during which the challenge r is computed. As an edge case, for $i = 0$, $\mathbb{U}_{i+1}$ and $\mathbb{W}_{i+1}$ are respectively set to $\mathbb{U}_\perp, \mathbb{W}_\perp$. Then the CycleFold circuit $\mathcal{F}^{\mathsf{cf}}$ is executed for statement $r, \mathbb{U}^{\mathbb{G}}_i, \mathbb{u}^{\mathbb{G}}_i, \mathbb{U}^{\mathbb{G}}_{i+1}, \overline{T}$, whose execution trace $\mathcal{F}^{\mathsf{cf}}$ is also collected to construct the incoming CycleFold instance-witness pair $(\mathbb{u}^{\mathsf{cf}}_i, \mathbb{w}^{\mathsf{cf}}_i)$. Later, $(\mathbb{u}^{\mathsf{cf}}_i, \mathbb{w}^{\mathsf{cf}}_i)$ is folded into $(\mathbb{U}^{\mathsf{cf}}_i, \mathbb{W}^{\mathsf{cf}}_i)$ and results in $(\mathbb{U}^{\mathsf{cf}}_{i+1}, \mathbb{W}^{\mathsf{cf}}_{i+1})$. Note that, for $i = 0$, we instead set $\mathbb{U}^{\mathsf{cf}}_{i+1} = \mathbb{U}^{\mathsf{cf}}_\perp, \mathbb{W}^{\mathsf{cf}}_{i+1} = \mathbb{W}^{\mathsf{cf}}_\perp$. Now, $\mathcal{P}$ invokes the augmented step circuit $\mathcal{F}^{\mathsf{aug}}$. When asked for hints $\{o_j\}_{j=0}^{\nu-1}$ w.r.t. $\{\alpha_i\}_{i=0}^{\mu-1}$, $\mathcal{P}$ sets o_j as the number of occurrences of τ_j in $\{\alpha_i\}_{i=0}^{\mu-1}$ for all $j \in [0, \nu-1]$. When asked for hint c w.r.t. $\boldsymbol{q} = \{\alpha_i\}_{i=0}^{\mu-1} \cup \{o_j\}_{j=0}^{\nu-1}$, $\mathcal{P}$ computes $\mathbb{u}_{i+1}.\overline{Q} \leftarrow \mathsf{CM}.\mathcal{C}(\mathsf{ck}, \boldsymbol{q})$ and $c := \rho(\mathbb{u}_{i+1}.\overline{Q})$, and dynamically marks c as a statement. Finally, $\mathcal{P}$ constructs the incoming primary instance-witness pair $(\mathbb{u}_{i+1}, \mathbb{w}_{i+1})$ from the in-circuit variables in $\mathcal{F}^{\mathsf{aug}}$. The counter $i := i + 1$ and the state $\boldsymbol{z}_{i+1} := \mathcal{F}(\boldsymbol{z}_i; \mathsf{aux}_i)$ are updated as well for the next iteration.

The verification of an IVC proof $\pi_i = ((\mathbb{U}_i, \mathbb{W}_i), (\mathbb{u}_i, \mathbb{w}_i), (\mathbb{U}_i^{\mathsf{cf}}, \mathbb{W}_i^{\mathsf{cf}}))$ is straightforward. $\mathcal{V}$ simply verifies the digests and challenges in $\mathbb{u}_i$ and checks if $\mathbb{w}_i, \mathbb{W}_i, \mathbb{W}_i^{\mathsf{cf}}$ satisfy $\mathbb{u}_i, \mathbb{U}_i, \mathbb{U}_i^{\mathsf{cf}}$, respectively.

4.3 Decider

Finally, we introduce a decider Decider that compresses the final IVC proof π_k into a succinct zero-knowledge proof ϖ via a zkSNARK for the relation R^{IVC}. Given statement $\boldsymbol{x} = (k, \boldsymbol{z}_0, \boldsymbol{z}_k)$ and witness $\boldsymbol{w} = \pi_k$, $R^{\mathsf{IVC}}(\boldsymbol{x}, \boldsymbol{w}) = 1$ if and only if $\mathsf{IVC}.\mathcal{V}(\mathsf{vk}, (k, \boldsymbol{z}_0, \boldsymbol{z}_k), \pi_k) = 1$.

Before diving into the details, we first review two different methods for constructing Decider for Nova-based IVC.

In Nova [28], the authors construct a dedicated Polynomial IOP [56] for relaxed R1CS and compile it into a zkSNARK for R^{IVC} using a polynomial commitment scheme (PCS). Two choices of the PCS are presented: a Pedersen-based PCS with Bulletproofs [57] as the IPA, and a two-tiered PCS (e.g., Dory-PC [58]) with Dory-IPA [58]. For an augmented step circuit $\mathcal{F}^{\mathsf{aug}}$ with n constraints, the former achieves $O(\log n)$ proof size and $O(n)$ verification time, while the latter makes both proof size and verification time logarithmic in n.

`sonobe` [38] instead expresses R^{IVC} as an arithmetic circuit, whose satisfiability is proven with Groth16 [54], yielding constant proof size and verifier time. Nevertheless, `sonobe`'s decider only supports compressing proofs that use KZG commitment [59], where $\mathsf{IVC}.\mathcal{P}$ needs to interpolate the polynomial from the input vector. This results in an $O(n \log n)$ prover due to number-theoretic transforms (NTT).

Our goal is to design a decider that improves both approaches. Specifically, it should produce constant-sized proofs that can be verified in constant time w.r.t. n, while keeping the prover time linear in n in each step of $\mathsf{IVC}.\mathcal{P}$. To this end, we follow `sonobe`'s approach and prove the satisfiability of the decider circuit $\mathcal{F}^{\mathsf{Decider}}$, which encodes R^{IVC} over the primary curve $\mathbb{G}$. However, instead of using the plain Groth16, we leverage LegoGro16 [40], a ZKCP that establishes a bridge between Groth16 and Pedersen commitment [60], allowing us to choose Pedersen commitment as CM. As a result, the prover time in each iteration of incremental proof generation is linear, and both the final compressed proof size and the verifier time are constant. It is worth noting that, due to Groth16, the prover time in Decider is $O(n' \log n')$, where n' is the number of constraints in $\mathcal{F}^{\mathsf{Decider}}$ and is linear in n. But we stress that $\mathsf{Decider}.\mathcal{P}$ is a one-time cost at the end of multiple steps of $\mathsf{IVC}.\mathcal{P}$, and it thus can be relatively cheap in practice.

Instead of constructing a decider circuit $\mathcal{F}^{\mathsf{Decider}}$ for the entire $\mathsf{IVC}.\mathcal{V}$ algorithm, we utilize various techniques and design a more efficient $\mathcal{F}^{\mathsf{Decider}}$.

First, as in [28], we require $\mathsf{Decider}.\mathcal{P}$ to run $\mathsf{NIFS}.\mathcal{P}$ once more to absorb $(\mathbb{u}_k, \mathbb{w}_k)$ into $(\mathbb{U}_k, \mathbb{W}_k)$, and $\mathcal{F}^{\mathsf{Decider}}$ only needs to check the output $(\mathbb{U}_{k+1}, \mathbb{W}_{k+1})$ instead of both inputs.

Furthermore, we get rid of verifying the $\mathbb{U}_{k+1}^{\mathbb{G}}$ part (*i.e.*, the commitments

Algorithm 4: IVC

Fn $\mathsf{IVC}.\mathcal{G}(1^\lambda)$**:**

  return $(\mathsf{ck} \leftarrow \mathsf{NIFS}.\mathcal{G}(1^\lambda), \mathsf{ck}^{\mathsf{cf}} \leftarrow \mathsf{NIFS}.\mathcal{G}(1^\lambda))$

Fn $\mathsf{IVC}.\mathcal{K}((\mathsf{ck}, \mathsf{ck}^{\mathsf{cf}}), \mathcal{F})$**:**

  $(\mathsf{pk}, \mathsf{vk}) := \mathsf{NIFS}.\mathcal{K}(\mathsf{ck}, \mathsf{CS}^{\mathsf{aug}})$

  $(\mathsf{pk}^{\mathsf{cf}}, \mathsf{vk}^{\mathsf{cf}}) := \mathsf{NIFS}.\mathcal{K}(\mathsf{ck}^{\mathsf{cf}}, \mathsf{CS}^{\mathsf{cf}})$

  return $(\mathsf{pk}, \mathsf{pk}^{\mathsf{cf}}), (\mathsf{vk}, \mathsf{vk}^{\mathsf{cf}})$

Fn $\mathsf{IVC}.\mathcal{P}((\mathsf{pk}, \mathsf{pk}^{\mathsf{cf}}), (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{aux}_i, \pi_i)$**:**

  Parse $((\mathbb{U}_i, \mathbb{W}_i), (\mathbbm{u}_i, \mathbbm{w}_i), (\mathbb{U}_i^{\mathsf{cf}}, \mathbb{W}_i^{\mathsf{cf}})) := \pi_i$

  $(\mathbb{U}_{i+1}, \mathbb{W}_{i+1}, \overline{T}) := (i = 0)\ ?\ (\mathbb{U}_\perp, \mathbb{W}_\perp, \overline{0}) : \mathsf{NIFS}.\mathcal{P}(\mathsf{pk}, (\mathbb{U}_i, \mathbb{W}_i), (\mathbbm{u}_i, \mathbbm{w}_i))$

  Run $\mathcal{F}^{\mathsf{cf}}$ and extract the in-circuit variables $\boldsymbol{v}^{\mathsf{cf}}$:

    $r := \rho(\mathbb{U}_i, \mathbbm{u}_i, \overline{T})$

    $\mathcal{F}^{\mathsf{cf}}(r, \mathbb{U}_i^{\mathbb{G}}, \mathbbm{u}_i^{\mathbb{G}}, \mathbb{U}_{i+1}^{\mathbb{G}}, \overline{T})$

  Construct $\mathbbm{u}_i^{\mathsf{cf}}, \mathbbm{w}_i^{\mathsf{cf}}$:

    Parse $(u, \boldsymbol{x}, \varnothing, \boldsymbol{w}) := \boldsymbol{v}^{\mathsf{cf}}$ ▷ $u = 1, \boldsymbol{x} = (r, \mathbb{U}_i^{\mathbb{G}}, \mathbbm{u}_i^{\mathbb{G}}, \mathbb{U}_{i+1}^{\mathbb{G}})$

    $\mathbbm{w}_i^{\mathsf{cf}} := (\varnothing, \boldsymbol{w}, \varnothing), \quad \mathbbm{u}_i^{\mathsf{cf}} \leftarrow (u, \boldsymbol{x}, \overline{0}, \mathsf{CM}.\mathcal{C}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{w}), \overline{0})$

  $(\mathbb{U}_{i+1}^{\mathsf{cf}}, \mathbb{W}_{i+1}^{\mathsf{cf}}, \overline{T}^{\mathsf{cf}}) := (i = 0)\ ?\ (\mathbb{U}_\perp^{\mathsf{cf}}, \mathbb{W}_\perp^{\mathsf{cf}}, \overline{0}) : \mathsf{NIFS}.\mathcal{P}(\mathsf{pk}^{\mathsf{cf}}, (\mathbb{U}_i^{\mathsf{cf}}, \mathbb{W}_i^{\mathsf{cf}}), (\mathbbm{u}_i^{\mathsf{cf}}, \mathbbm{w}_i^{\mathsf{cf}}))$

  Run $\mathcal{F}^{\mathsf{aug}}$ and extract the in-circuit variables $\boldsymbol{v}$:

    $h_1 := \mathsf{H}(\mathbb{U}_{i+1}, i+1, \boldsymbol{z}_0, \boldsymbol{z}_{i+1}), \quad h_2 := \mathsf{H}(\mathbb{U}_{i+1}^{\mathsf{cf}}, i+1)$

    $\mathcal{F}^{\mathsf{aug}}((i, \boldsymbol{z}_i, \mathbb{U}_i, \mathbbm{u}_i, \mathbb{U}_{i+1}^{\mathbb{G}}, \overline{T}, \mathbb{U}_i^{\mathsf{cf}}, \mathbbm{u}_i^{\mathsf{cf}}, \overline{T}^{\mathsf{cf}}), (h_1, h_2, \perp); \mathsf{aux}_i)$

  Construct $\mathbbm{u}_{i+1}, \mathbbm{w}_{i+1}$:

    Parse $(u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w}) := \boldsymbol{v}$ ▷ $u = 1, \boldsymbol{x} = (h_1, h_2, c)$

    $\mathbbm{w}_{i+1} := (\boldsymbol{q}, \boldsymbol{w}, \varnothing), \quad \mathbbm{u}_{i+1} \leftarrow (u, \boldsymbol{x}, \mathsf{CM}.\mathcal{C}(\mathsf{ck}, \boldsymbol{q}), \mathsf{CM}.\mathcal{C}(\mathsf{ck}, \boldsymbol{w}), \overline{0})$

  return $\pi_{i+1} := ((\mathbb{U}_{i+1}, \mathbb{W}_{i+1}), (\mathbbm{u}_{i+1}, \mathbbm{w}_{i+1}), (\mathbb{U}_{i+1}^{\mathsf{cf}}, \mathbb{W}_{i+1}^{\mathsf{cf}}))$

Fn $\mathsf{IVC}.\mathcal{V}((\mathsf{vk}, \mathsf{vk}^{\mathsf{cf}}), (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i)$**:**

  Parse $((\mathbb{U}_i, \mathbb{W}_i), (\mathbbm{u}_i, \mathbbm{w}_i), (\mathbb{U}_i^{\mathsf{cf}}, \mathbb{W}_i^{\mathsf{cf}})) := \pi_i$

  Check $\mathbbm{u}_i$:

    assert $\mathbbm{u}_i.u = 1 \wedge \mathbbm{u}_i.\overline{E} = \overline{0}$

    assert $\mathbbm{u}_i.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_i, i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{H}(\mathbb{U}_i^{\mathsf{cf}}, i), \rho(\mathbbm{u}_i.\overline{Q}))$

  Check $\mathbbm{w}_i$ against $\mathbbm{u}_i$:

    Parse $(u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E}) := \mathbbm{u}_i, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbbm{w}_i$

    $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$

    assert $\boldsymbol{Av} \circ \boldsymbol{Bv} = \boldsymbol{Cv}$

    assert $\mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{q}, \overline{Q}) \wedge \mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{w}, \overline{W}) \wedge \mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{e}, \overline{E})$

  Check $\mathbb{W}_i$ against $\mathbb{U}_i$:

    Parse $(u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E}) := \mathbb{U}_i, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbb{W}_i$

    $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$

    assert $\boldsymbol{Av} \circ \boldsymbol{Bv} = u \cdot \boldsymbol{Cv} + \boldsymbol{e}$

    assert $\mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{q}, \overline{Q}) \wedge \mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{w}, \overline{W}) \wedge \mathsf{CM}.\mathcal{V}(\mathsf{ck}, \boldsymbol{e}, \overline{E})$

  Check $\mathbb{W}_i^{\mathsf{cf}}$ against $\mathbb{U}_i^{\mathsf{cf}}$:

    Parse $(u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E}) := \mathbb{U}_i^{\mathsf{cf}}, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbb{W}_i^{\mathsf{cf}}$

    $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$

    assert $\boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v} = u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e}$

    assert $\boldsymbol{q} = \varnothing \wedge \overline{Q} = \overline{0} \wedge \mathsf{CM}.\mathcal{V}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{w}, \overline{W}) \wedge \mathsf{CM}.\mathcal{V}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{e}, \overline{E})$

  return 1

$\mathbb{U}_{k+1}.\overline{Q}, \mathbb{U}_{k+1}.\overline{W}, \mathbb{U}_{k+1}.\overline{E})$ in-circuit, since the commitment verification involves non-native group operations that are prohibitively expensive (a single scalar multiplication costs $\sim 10^5$ constraints as per [61]). In Sigmabus [61] and its follow-up work [62], the authors observe that it is possible to trade non-native group operations for computing circuit-friendly hash functions by leveraging a sigma protocol. However, this approach is still suboptimal, as the hash preimage, containing the values to be committed, can be long for a complex step circuit $\mathcal{F}$. Thanks to CP-SNARKs [40], we can completely eliminate the constraints for commitment verification, because they are able to prove the commitment to a subset of witnesses without running $\mathsf{CM}.\mathcal{V}$ or adding any other trade-off in-circuit.

Finally, $\mathcal{F}^{\mathsf{Decider}}$ checks the satisfiability of $(\mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}})$ against $\mathsf{CS}^{\mathsf{cf}}$, where the commitment verification becomes native operations, while the check $\boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v} \equiv u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e} \pmod{q}$ requires non-native field operations. Although one can perform the check natively by employing a SNARK (e.g., Spartan [63]) solely for this equation on the non-pairing curve $\mathbb{H}$, the final proof size in this case is no longer constant. We observe that it is feasible to compute this equation in-circuit, because 1) non-native field operations are relatively cheap in comparison to non-native group operations, and 2) the size of $\mathcal{F}^{\mathsf{cf}}$ (and thus $\mathsf{CS}^{\mathsf{cf}}$) is constant and small. However, we further apply several optimizations to improve the efficiency of this check, whose details are presented in Section 6.

We summarize the construction of the decider circuit $\mathcal{F}^{\mathsf{Decider}}$ in Circuit 5.

Circuit 5: $\mathcal{F}^{\mathsf{Decider}}$

Witness: $\mathbb{W}_{k+1}, \mathbb{W}_k^{\mathsf{cf}}$

Statement: $\mathbb{U}_{k+1}^{\mathbb{F}}, \mathbb{U}_k^{\mathsf{cf}}$

Constant: $\mathsf{CS}^{\mathsf{aug}} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}), \mathsf{CS}^{\mathsf{cf}} = (\boldsymbol{A}^{\mathsf{cf}}, \boldsymbol{B}^{\mathsf{cf}}, \boldsymbol{C}^{\mathsf{cf}}), \mathsf{ck}^{\mathsf{cf}}$

1 Check $\mathbb{W}_{k+1}$ against $\mathbb{U}_{k+1}^{\mathbb{F}}$:
- Parse $(u, \boldsymbol{x}) := \mathbb{U}_{k+1}^{\mathbb{F}}, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbb{W}_{k+1}$
- $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$
- **enforce** $\boldsymbol{A}\boldsymbol{v} \circ \boldsymbol{B}\boldsymbol{v} = u \cdot \boldsymbol{C}\boldsymbol{v} + \boldsymbol{e}$

2 Check $\mathbb{W}_k^{\mathsf{cf}}$ against $\mathbb{U}_k^{\mathsf{cf}}$:
- Parse $(u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E}) := \mathbb{U}_k^{\mathsf{cf}}, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbb{W}_k^{\mathsf{cf}}$
- $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$
- **enforce** $\boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v} \equiv u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e} \pmod{q}$
- **enforce** $\boldsymbol{q} = \varnothing \wedge \overline{Q} = \overline{0}$
- **enforce** $\mathsf{CM}.\mathcal{V}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{w}, \overline{W})$
- **enforce** $\mathsf{CM}.\mathcal{V}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{e}, \overline{E})$

We present the decider algorithm $\mathsf{Decider}$ in Algorithm 6, which shares the design of decider in Nova and `sonobe` but replaces the underlying SNARK with $\mathsf{LegoGro16}$. As discussed above, $\mathsf{Decider}.\mathcal{P}$ first runs $\mathsf{NIFS}.\mathcal{P}$ to fold $(\mathbb{u}_k, \mathbb{w}_k)$ into $(\mathbb{U}_k, \mathbb{W}_k)$, and then generates a proof ϖ with $\mathsf{ZKCP}.\mathcal{P}$, attesting that $\mathcal{F}^{\mathsf{Decider}}$ is satisfiable, and that the commitments $\overline{Q}, \overline{W}, \overline{E}$ in $\mathbb{U}_{k+1}^{\mathbb{G}}$ open to $\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}$ in $\mathbb{W}_{k+1}$. Correspondingly, the verifier $\mathsf{Decider}.\mathcal{V}$ folds $\mathbb{u}_k$ into $\mathbb{U}_k$ as well, ensures $\mathbb{u}_k$ is a (strict) R1CS instance, and then verifies the proof ϖ and the commitments in $\mathbb{U}_{k+1}$ using $\mathsf{ZKCP}.\mathcal{V}$.

Algorithm 6: Decider

Fn $\mathsf{Decider}.\mathcal{K}(1^\lambda, (\mathsf{ck}, \mathsf{CS}^{\mathsf{Decider}}))$**:**
- $(\mathsf{pk}, \mathsf{vk}) \leftarrow \mathsf{ZKCP}.\mathcal{K}(1^\lambda, \mathsf{ck}, \mathsf{CS}^{\mathsf{Decider}})$
- **return** $(\mathsf{pk}, \mathsf{vk})$

Fn $\mathsf{Decider}.\mathcal{P}((\mathsf{pk}, \mathsf{pk}_\Phi), (k, \boldsymbol{z}_0, \boldsymbol{z}_k), \pi_k)$**:**
- Parse $((\mathbb{U}_k, \mathbb{W}_k), (\mathbb{u}_k, \mathbb{w}_k), (\mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}})) := \pi_k$
- $(\mathbb{U}_{k+1}, \mathbb{W}_{k+1}, \overline{T}) := \mathsf{NIFS}.\mathcal{P}(\mathsf{pk}_\Phi, (\mathbb{U}_k, \mathbb{W}_k), (\mathbb{u}_k, \mathbb{w}_k))$
- $\boldsymbol{x} := (\mathbb{U}_{k+1}^{\mathbb{F}}, \mathbb{U}_k^{\mathsf{cf}}), \boldsymbol{c} := (\mathbb{U}_{k+1}^{\mathbb{G}})$
- $\boldsymbol{\upsilon} := (\mathbb{W}_{k+1}), \boldsymbol{\omega} := (\mathbb{W}_k^{\mathsf{cf}})$
- $\varpi \leftarrow \mathsf{ZKCP}.\mathcal{P}(\mathsf{pk}, \boldsymbol{x}, \boldsymbol{c}, \boldsymbol{\upsilon}, \boldsymbol{\omega})$
- **return** $(\varpi, \mathbb{U}_k, \mathbb{u}_k, \mathbb{U}_k^{\mathsf{cf}}, \overline{T})$

Fn $\mathsf{Decider}.\mathcal{V}((\mathsf{vk}, \mathsf{vk}_\Phi), (k, \boldsymbol{z}_0, \boldsymbol{z}_k), (\varpi, \mathbb{U}_k, \mathbb{u}_k, \mathbb{U}_k^{\mathsf{cf}}, \overline{T}))$**:**
- $\mathbb{U}_{k+1} := \mathsf{NIFS}.\mathcal{V}(\mathsf{vk}_\Phi, \mathbb{U}_k, \mathbb{u}_k, \overline{T})$
- Check $\mathbb{u}_k$:
  - **assert** $\mathbb{u}_k.u = 1, \mathbb{u}_k.\overline{E} = \overline{0}$
  - **assert** $\mathbb{u}_k.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_k, k, \boldsymbol{z}_0, \boldsymbol{z}_k), \mathsf{H}(\mathbb{U}_k^{\mathsf{cf}}, k), \rho(\mathbb{u}_k.\overline{Q}))$
- $\boldsymbol{x} := (\mathbb{U}_{k+1}^{\mathbb{F}}, \mathbb{U}_k^{\mathsf{cf}}), \boldsymbol{c} := (\mathbb{U}_{k+1}^{\mathbb{G}})$
- **assert** $\mathsf{ZKCP}.\mathcal{V}(\mathsf{vk}, \boldsymbol{x}, \boldsymbol{c}, \varpi)$
- **return** 1

5 The Eva Protocol

In this section, we introduce the construction of Eva, our proof of video authenticity based on IVC.

Recall that in a proof of video authenticity, $\mathcal{P}$ aims to convince $\mathcal{V}$ that the processed video stream ζ is honestly edited and encoded from some original video $\boldsymbol{V}$ whose signature σ is valid with respect to the public key vk_Σ. Due to the nature of video processing algorithms, we can view the editing and encoding operation as a sequence of sub-procedures on each macroblock of the video. Thus, we only need to construct the gadgets for encoding and editing a single macroblock, which can be naturally extended to handle the entire video $\boldsymbol{V}$ by utilizing our folding-based IVC.

In Section 5.1, we elaborate on $\mathcal{F}^{\mathcal{E}}$, the gadget for video encoding, as well as its building blocks. Next, we present several instantiations of $\mathcal{F}^{\Delta}$ for several video editing operations in Section 5.2. With $\mathcal{F}^{\mathcal{E}}$ and $\mathcal{F}^{\Delta}$ as two key components, Section 5.3 provides the construction of our IVC step circuit $\mathcal{F}^{\mathsf{Eva}}$, which incorporates the checks for the validity of signature σ. Finally, we build upon $\mathcal{F}^{\mathsf{Eva}}$ the full construction of Eva and discuss its security in Section 5.4.

5.1 Gadgets for Video Encoding

First, we construct $\mathcal{F}^{\mathcal{E}}$, a gadget for encoding a single macroblock in a video. Specifically, we focus on supporting H.264/AVC [37], but our methodology can be extended to other block-based video codecs such as H.265/HEVC, AV1, etc.

Naively, one may translate the entire encoder to the $\mathcal{F}^{\mathcal{E}}$ gadget. However, due to the reasons below, this would require a prohibitive number of constraints, which is infeasible in practice.

- The encoding process involves complex operations, such as motion estimation, entropy coding, etc.
- The encoding of a macroblock may depend on other macroblocks in the current frame (when the prediction mode is mode = "intra") or even in the neighboring frames (when mode = "inter").

To address these challenges, we make extensive use of verifier's knowledge. Note that although video codecs are generally lossy, the decoder can still accurately extract from ζ some information that appears in the encoding process as well. In fact, the prediction macroblock $\boldsymbol{P}$ decoded by $\mathcal{D}$ is identical to the original prediction macroblock computed by $\mathcal{E}$, which is also the case for the quantized coefficients $\boldsymbol{Z}$.

Thus, we save the prover's cost by treating $\boldsymbol{P}$ and $\boldsymbol{Z}$ as public inputs, which can be recovered by $\mathcal{V}$. Now, to prove the honest encoding of a macroblock $\boldsymbol{X}$ with encoding parameters $\mathsf{param}^{\mathcal{E}}$, $\mathcal{F}^{\mathcal{E}}$ no longer includes the entire $\mathcal{E}$. Instead, $\mathcal{F}^{\mathcal{E}}$ only has to enforce the honest execution of differing, transform, and quantization, as depicted in Figure 2, while it becomes unnecessary to prove prediction and entropy coding.

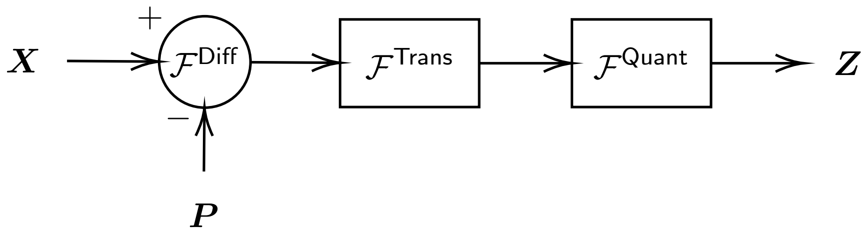


Figure 2: Overview of in-circuit operations for video encoding

As summarized in Gadget 7, the encoding gadget $\mathcal{F}^{\mathcal{E}}$ takes as input the current macroblock $\boldsymbol{X}$, the current prediction macroblock $\boldsymbol{P}$, and additionally the encoding parameters $\mathsf{param}^{\mathcal{E}}$, and returns the quantized coefficients $\boldsymbol{Z}$ by running the gadgets $\mathcal{F}^{\mathsf{Diff}}$, $\mathcal{F}^{\mathsf{Trans}}$, and $\mathcal{F}^{\mathsf{Quant}}$. Here, $\mathcal{F}^{\mathsf{Diff}}$ for residual macroblock computation simply returns $\boldsymbol{R} \coloneqq \boldsymbol{X} - \boldsymbol{P}$, while the constructions of $\mathcal{F}^{\mathsf{Trans}}$ and $\mathcal{F}^{\mathsf{Quant}}$ are elaborated in the following sections.

Gadget 7: $\mathcal{F}^{\mathcal{E}}(\boldsymbol{X}, \boldsymbol{P}, \mathsf{param}^{\mathcal{E}})$

```
R := F^Diff(X, P)                ▷ Compute residual macroblock R
Y := F^Trans(R)                  ▷ Compute transformed coefficients Y
Z ← F^Quant(Y, param^E)          ▷ Compute quantized coefficients Z
return Z
```

5.1.1 Transform

The transform operation in H.264 is based on 4×4 DCT (Discrete Cosine Transform). For efficiency, this process is slightly different from the original DCT: it only involves integer operations, while the fractional part of the DCT coefficients is deferred to the quantization step. More specifically, with the core transform matrix $\boldsymbol{C}_4 := \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$, the transform on a 4×4 block $\boldsymbol{\Lambda}$ is computed as $\boldsymbol{\Gamma} := \boldsymbol{C}_4 \boldsymbol{\Lambda} \boldsymbol{C}_4^\intercal$. Hence, the construction of the corresponding gadget $\mathcal{F}^{\mathsf{Trans}}$ is straightforward: for a residual macroblock $\boldsymbol{R}$ with color components $\boldsymbol{R}^{\mathsf{Y}}, \boldsymbol{R}^{\mathsf{Cb}}, \boldsymbol{R}^{\mathsf{Cr}}$, we divide them into blocks of 4×4 and apply the transform operation on each block. Since every entry $r_{i,j}$ in $\boldsymbol{R}$ is in $[-255, 255]$, the matrix multiplication can be natively performed in $\mathbb{F}_p$ without overflow.

After the core transform, the DC (i.e., the first) coefficients of all blocks from every color component are collected into a 4×4 matrix $\boldsymbol{B}^{\mathsf{Y}}$ and two 2×2 matrices $\boldsymbol{B}^{\mathsf{Cb}}, \boldsymbol{B}^{\mathsf{Cr}}$, while the AC (i.e., the remaining) coefficients are unchanged. These matrices are transformed again using the Hadamard matrices $\boldsymbol{H}_4 := \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$ and $\boldsymbol{H}_2 := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, respectively. Finally, the transformed DC coefficients $\boldsymbol{D}^{\mathsf{Y}}, \boldsymbol{D}^{\mathsf{Cb}}, \boldsymbol{D}^{\mathsf{Cr}}$ as well as the AC coefficients $\{\boldsymbol{A}_i^{\mathsf{Y}}\}_{i=0}^{15}, \{\boldsymbol{A}_i^{\mathsf{Cb}}\}_{i=0}^{4}, \{\boldsymbol{A}_i^{\mathsf{Cr}}\}_{i=0}^{4}$ are returned. The entire in-circuit transform process is depicted in Gadget 8.

Gadget 8: $\mathcal{F}^{\mathsf{Trans}}(\boldsymbol{R})$

$\boldsymbol{C}_4 := \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}, \boldsymbol{H}_4 := \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \boldsymbol{H}_2 := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

for $k \in [0, 16)$ **do**
- $i := \lfloor k/4 \rfloor, j := k \bmod 4$
- $\boldsymbol{A}_k^{\mathsf{Y}} := \boldsymbol{C}_4 \boldsymbol{R}^{\mathsf{Y}}[4i, 4i+4; 4j, 4j+4] \boldsymbol{C}_4^\intercal$
- $b_{i,j}^{\mathsf{Y}} := a_{k,0,0}^{\mathsf{Y}}$

for $k \in [0, 4)$ **do**
- $i := \lfloor k/2 \rfloor, j := k \bmod 2$
- $\boldsymbol{A}_k^{\mathsf{Cb}} := \boldsymbol{C}_4 \boldsymbol{R}^{\mathsf{Cb}}[4i, 4i+4; 4j, 4j+4] \boldsymbol{C}_4^\intercal$
- $\boldsymbol{A}_k^{\mathsf{Cr}} := \boldsymbol{C}_4 \boldsymbol{R}^{\mathsf{Cr}}[4i, 4i+4; 4j, 4j+4] \boldsymbol{C}_4^\intercal$
- $b_{i,j}^{\mathsf{Cb}} := a_{k,0,0}^{\mathsf{Cb}}$
- $b_{i,j}^{\mathsf{Cr}} := a_{k,0,0}^{\mathsf{Cr}}$

$\boldsymbol{D}^{\mathsf{Y}} := \boldsymbol{H}_4 \boldsymbol{B}^{\mathsf{Y}} \boldsymbol{H}_4^\intercal$

$\boldsymbol{D}^{\mathsf{Cb}} := \boldsymbol{H}_2 \boldsymbol{B}^{\mathsf{Cb}} \boldsymbol{H}_2^\intercal$

$\boldsymbol{D}^{\mathsf{Cr}} := \boldsymbol{H}_2 \boldsymbol{B}^{\mathsf{Cr}} \boldsymbol{H}_2^\intercal$

$\boldsymbol{Y} := (\{\boldsymbol{A}_i^{\mathsf{Y}}\}_{i=0}^{15}, \{\boldsymbol{A}_i^{\mathsf{Cb}}\}_{i=0}^{4}, \{\boldsymbol{A}_i^{\mathsf{Cr}}\}_{i=0}^{4}, \boldsymbol{D}^{\mathsf{Y}}, \boldsymbol{D}^{\mathsf{Cb}}, \boldsymbol{D}^{\mathsf{Cr}})$

return $\boldsymbol{Y}$

5.1.2 Quantization

The quantization step maps a coefficient v from the transform step to a quantized value u. This process is lossy, and how much information is preserved is controlled by the quantization parameter qp in H.264. A large qp leads to a higher compression ratio but also more distortion, while a small qp results in larger file sizes but better quality.

In general, the quantized coefficient is computed by $u := \lfloor v \times \psi / 2^\delta \rceil$, i.e., we first scale the transformed coefficient v by $\psi / 2^\delta$ and then round the result to the nearest integer. Here, ψ is the multiplication factor that takes the fractional part of the DCT coefficients into account. The shift δ is $15 + \lfloor \mathsf{qp}/6 \rfloor$ for AC coefficients and $16 + \lfloor \mathsf{qp}/6 \rfloor$ for DC coefficients.

In H.264, floating-point operations such as rounding are further replaced with approximate integer operations, as the latter are more efficient in hardware. This is also beneficial for reducing the circuit size: even with the state-of-the-art techniques [64], the in-circuit floating-point operations are still expensive, e.g., a single FP32 division would cost ~ 76 constraints. Now, the absolute value and the sign of the quantized coefficient u is computed as $\begin{cases} \mathrm{abs}(u) := (\mathrm{abs}(v) \times \psi + \phi) \gg \delta \\ \mathrm{sign}(u) := \mathrm{sign}(v) \end{cases}$, where v's absolute value is scaled by the multiplication factor ψ, added with an offset ϕ, and then right shifted by δ bits. The offset ϕ equals f for AC coefficients, and is $2f$ for DC coefficients, where $f := 2^{4+\lfloor \mathsf{qp}/6 \rfloor} \times \begin{cases} 682, & \mathsf{mode} = \text{“intra”} \\ 342, & \mathsf{mode} = \text{“inter”} \end{cases}$. The values 682 and 342 are taken from the H.264 `JM` reference software [65], which are the approximate values of $2^{11}/3$ and $2^{11}/6$ respectively.

When dealing with scaling and rounding in-circuit, we leverage the efficient gadgets $\mathcal{F}^{\mathsf{SignAbs}}$ and $\mathcal{F}^{\gg}$ in [64] for computing absolute values and shifting operations. Both gadgets make queries to a lookup table, which can be efficiently checked by our IVC with lookup arguments integrated. The constructions of these gadgets are given in Appendix C. With these gadgets, we give the construction of the $\mathcal{F}^{\mathsf{ScaleRound}}$ gadget for scaling and rounding an input coefficient v in-place in Gadget 9.

Gadget 9: $\mathcal{F}^{\mathsf{ScaleRound}}(v, \psi, \phi, \delta)$

1. $(s, u) \leftarrow \mathcal{F}^{\mathsf{SignAbs}}(v)$
2. $t := \mathcal{F}^{\gg}(u \times \psi + \phi, \delta)$
3. $v := s\ ?\ t : -t$

Now we are finally ready to present the quantization gadget $\mathcal{F}^{\mathsf{Quant}}$. As per `JM`, the matrix of multiplication factors is defined as $\boldsymbol{\Psi} := \begin{bmatrix} 13107 & 5243 & 8066 \\ 11916 & 4660 & 7490 \\ 10082 & 4194 & 6554 \\ 9362 & 3647 & 5825 \\ 8192 & 3355 & 5243 \\ 7282 & 2893 & 4559 \end{bmatrix}$. For an AC coefficient $a_{i,j}$, the multiplication factor is in the ($\mathsf{qp} \bmod 6$)-th row and the $p_{i,j}$-th column of $\boldsymbol{\Psi}$, where $p_{i,j} = \begin{cases} 0, & (i,j) \in \{(0,0),(0,2),(2,0),(2,2)\} \\ 1, & (i,j) \in \{(1,1),(1,3),(3,1),(3,3)\} \\ 2, & \text{otherwise} \end{cases}$. We use a matrix $\boldsymbol{P}$ to represent the mapping from (i,j) to $p_{i,j}$. On the other hand, the multiplication factor for DC coefficients are always $\psi_{0,0}$. Then, for all AC and DC blocks, we apply the $\mathcal{F}^{\mathsf{ScaleRound}}$ gadget to quantize their coefficients with the corresponding parameters, except that the DC coefficients for luma are right shifted by 1 bit before quantization.

The entire quantization process is summarized in Gadget 10, with qp and mode included in $\mathsf{param}^{\mathcal{E}}$.

5.2 Gadgets for Video Editing

We showcase various gadgets for video editing, including color manipulations (e.g., conversion to grayscale, brightness adjustment, color inversion), spatial

Gadget 10: $\mathcal{F}^{\mathsf{Quant}}(\boldsymbol{Y}, \mathsf{param}^{\mathcal{E}})$

```
Parse $(\{\boldsymbol{A}_i^{\mathsf{Y}}\}_{i=0}^{15}, \{\boldsymbol{A}_i^{\mathsf{Cb}}\}_{i=0}^{4}, \{\boldsymbol{A}_i^{\mathsf{Cr}}\}_{i=0}^{4}, \boldsymbol{D}^{\mathsf{Y}}, \boldsymbol{D}^{\mathsf{Cb}}, \boldsymbol{D}^{\mathsf{Cr}}) \coloneqq \boldsymbol{Y}$
Parse $(\mathsf{qp}, \mathsf{mode}, \cdot) \coloneqq \mathsf{param}^{\mathcal{E}}$
$q \coloneqq \lfloor \mathsf{qp}/6 \rfloor, r \coloneqq \mathsf{qp} \bmod 6$
$f \coloneqq ((\mathsf{mode} = \text{“intra”})\ ?\ 682 : 342) \times 2^{4+q}$
$\boldsymbol{\Psi} \coloneqq \begin{bmatrix} 13107 & 5243 & 8066 \\ 11916 & 4660 & 7490 \\ 10082 & 4194 & 6554 \\ 9362 & 3647 & 5825 \\ 8192 & 3355 & 5243 \\ 7282 & 2893 & 4559 \end{bmatrix}, \boldsymbol{P} \coloneqq \begin{bmatrix} 0 & 2 & 0 & 2 \\ 2 & 1 & 2 & 1 \\ 0 & 2 & 0 & 2 \\ 2 & 1 & 2 & 1 \end{bmatrix}$
for $\boldsymbol{A} \in \{\boldsymbol{A}_i^{\mathsf{Y}}\}_{i=0}^{15} \cup \{\boldsymbol{A}_i^{\mathsf{Cb}}\}_{i=0}^{4} \cup \{\boldsymbol{A}_i^{\mathsf{Cr}}\}_{i=0}^{4}$ do
    for $i \in [0,4), j \in [0,4)$ do
        $\mathcal{F}^{\mathsf{ScaleRound}}(a_{i,j}, \psi_{r,p_{i,j}}, f, 15+q)$
for $i \in [0,4), j \in [0,4)$ do
    $\mathcal{F}^{\mathsf{ScaleRound}}(\mathcal{F}^{\gg}(d_{i,j}^{\mathsf{Y}}, 1), \psi_{0,0}, 2f, 16+q)$
for $\boldsymbol{D} \in \{\boldsymbol{D}^{\mathsf{Cb}}, \boldsymbol{D}^{\mathsf{Cr}}\}$ do
    for $i \in [0,2), j \in [0,2)$ do
        $\mathcal{F}^{\mathsf{ScaleRound}}(d_{i,j}, \psi_{0,0}, 2f, 16+q)$
$\boldsymbol{Z} \coloneqq (\{\boldsymbol{A}_i^{\mathsf{Y}}\}_{i=0}^{15}, \{\boldsymbol{A}_i^{\mathsf{Cb}}\}_{i=0}^{4}, \{\boldsymbol{A}_i^{\mathsf{Cr}}\}_{i=0}^{4}, \boldsymbol{D}^{\mathsf{Y}}, \boldsymbol{D}^{\mathsf{Cb}}, \boldsymbol{D}^{\mathsf{Cr}})$
return $\boldsymbol{Z}$
```

operations (e.g., masking, cropping), and temporal operations (e.g., cutting). Additionally, we explain how to perform complex editing operations that involve multiple macroblocks in-circuit.

5.2.1 Color Manipulations

Thanks to the use of the YCbCr color space, it is straightforward to perform common color manipulations for videos encoded in H.264 or in many other video codecs. In contrast, color operations on RGB often involve the conversion between color spaces, demanding for in-circuit fixed-point or floating-point computation.

For instance, converting pixels in RGB to grayscale requires computing the luminance from the RGB components, which is given by $0.299 \times \mathsf{R} + 0.587 \times \mathsf{G} + 0.114 \times \mathsf{B}$. In YCbCr, the luma component already represents the luminance, and thus we can simply keep the luma component unchanged while setting the chroma components to 128.

We depict the grayscale conversion gadget for a macroblock $\boldsymbol{X}$ in Gadget 11.

Gadget 11: $\mathcal{F}^{\Delta_{\mathsf{gray}}}(\boldsymbol{X})$

```
Parse $(\boldsymbol{X}^{\mathsf{Y}}, \cdot) \coloneqq \boldsymbol{X}$
return $\boldsymbol{X}' \coloneqq (\boldsymbol{X}^{\mathsf{Y}}, \mathbf{128}, \mathbf{128})$
```

When adjusting the brightness, we only need to focus on the luma component, which is scaled by a factor $\mathsf{param}^{\mathsf{bright}}$ and clamped to $[0, 255]$, as shown in Gadget 12. We support 65536 levels of brightness adjustment, with $\mathsf{param}^{\mathsf{bright}} \in \{\frac{0}{256}, \frac{1}{256}, \ldots, \frac{65535}{256}\}$. Given a luma component x^{Y} and a parameter $\mathsf{param}^{\mathsf{bright}} = \frac{\beta}{256}$, we handle the in-circuit scaling operation by first computing $\beta \times x^{\mathsf{Y}}$ and then right shifting the product by 8 bits. Next, in order to clamp the product to

$[0, 255]$, we again shift the result to the right by 8 bits. If the remaining bits are all 0, then the original result is returned as it is smaller than 256. Otherwise, we return 255.

Gadget 12: $\mathcal{F}^{\Delta_{\mathsf{bright}}}(\boldsymbol{X}, \mathsf{param}^{\mathsf{bright}} = \frac{\beta}{256})$

Parse $(\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}}) \coloneqq \boldsymbol{X}$
for $i \in [0, 16), j \in [0, 16)$ **do**
  $u \coloneqq \mathcal{F}^{\gg}(x^{\mathsf{Y}}_{i,j} \times \beta, 8)$
  $v \coloneqq \mathcal{F}^{\gg}(u, 8)$
  $x^{\mathsf{Y}}_{i,j} \coloneqq (v = 0) \mathbin{?} u : 255$
return $\boldsymbol{X}' \coloneqq (\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}})$

Gadget 13 illustrates the gadget for color inversion, where we subtract all components in each pixel value from 255.

Gadget 13: $\mathcal{F}^{\Delta_{\mathsf{inv}}}(\boldsymbol{X})$

Parse $(\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}}) \coloneqq \boldsymbol{X}$
for $i \in [0, 16), j \in [0, 16)$ **do**
  $x^{\mathsf{Y}}_{i,j} \coloneqq 255 - x^{\mathsf{Y}}_{i,j}$
for $i \in [0, 8), j \in [0, 8)$ **do**
  $x^{\mathsf{Cb}}_{i,j} \coloneqq 255 - x^{\mathsf{Cb}}_{i,j}$
  $x^{\mathsf{Cr}}_{i,j} \coloneqq 255 - x^{\mathsf{Cr}}_{i,j}$
return $\boldsymbol{X}' \coloneqq (\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}})$

5.2.2 Spatial and Temporal Operations

Now we present the gadgets for spatial and temporal operations.

To mask a macroblock $\boldsymbol{X}$ with a layer $\boldsymbol{L}$, we additionally require a binary matrix $\boldsymbol{B}$, where each bit $b_{i,j}$ indicates whether we should replace the pixel in $\boldsymbol{X}$ with the corresponding pixel in $\boldsymbol{L}$. More specifically, if $b_{i,j}$ is true, then $x_{i,j}$ is updated to $l_{i,j}$, while $x_{i,j}$ remains unchanged otherwise. With $(\boldsymbol{B}, \boldsymbol{L})$ as the masking parameter $\mathsf{param}^{\mathsf{mask}}$, the masking gadget $\mathcal{F}^{\Delta_{\mathsf{mask}}}$ is given in Gadget 14. Note that different macroblocks may have different $\mathsf{param}^{\mathsf{mask}}$, which allows for arbitrary overlays with dynamic content and position (e.g., subtitles) without incurring additional costs.

Cropping and cutting both work similarly to each other, where the former removes data in the horizontal and vertical directions, while the latter removes data in the temporal direction. We unify both cases via the removal parameter $\mathsf{param}^{\mathsf{remove}}$, which consists of a boolean value b that indicates if the macroblocks needs to be removed. By specifying b according to the operation type, we can support both operations with the same gadget $\mathcal{F}^{\Delta_{\mathsf{remove}}}$. For instance, cropping requires $b = 1$ for macroblocks outside the cropped region, while for cutting, all macroblocks in removed frames have $b = 1$. The construction of $\mathcal{F}^{\Delta_{\mathsf{remove}}}$ is shown in Gadget 15, where $\perp$ is a dummy macroblock. Although the process seems straightforward, we omit an important detail in the description: how to handle $\perp$ is in fact non-trivial, and we defer the discussion to Section 5.3.

Gadget 14: $\mathcal{F}^{\Delta_{\mathsf{mask}}}(\boldsymbol{X}, \mathsf{param}^{\mathsf{mask}} = (\boldsymbol{B}, \boldsymbol{L}))$

Parse $(\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}}) \coloneqq \boldsymbol{X}$
Parse $(\boldsymbol{B}^{\mathsf{Y}}, \boldsymbol{B}^{\mathsf{Cb}}, \boldsymbol{B}^{\mathsf{Cr}}) \coloneqq \boldsymbol{B}$
Parse $(\boldsymbol{L}^{\mathsf{Y}}, \boldsymbol{L}^{\mathsf{Cb}}, \boldsymbol{L}^{\mathsf{Cr}}) \coloneqq \boldsymbol{L}$
for $i \in [0, 16), j \in [0, 16)$ **do**
  $x^{\mathsf{Y}}_{i,j} \coloneqq b^{\mathsf{Y}}_{i,j} \mathrel{?} l^{\mathsf{Y}}_{i,j} : x^{\mathsf{Y}}_{i,j}$
for $i \in [0, 8), j \in [0, 8)$ **do**
  $x^{\mathsf{Cb}}_{i,j} \coloneqq b^{\mathsf{Cb}}_{i,j} \mathrel{?} l^{\mathsf{Cb}}_{i,j} : x^{\mathsf{Cb}}_{i,j}$
  $x^{\mathsf{Cr}}_{i,j} \coloneqq b^{\mathsf{Cr}}_{i,j} \mathrel{?} l^{\mathsf{Cr}}_{i,j} : x^{\mathsf{Cr}}_{i,j}$
return $\boldsymbol{X}' \coloneqq (\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}})$

Gadget 15: $\mathcal{F}^{\Delta_{\mathsf{remove}}}(\boldsymbol{X}, \mathsf{param}^{\mathsf{remove}} = b)$

Parse $(\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}}) \coloneqq \boldsymbol{X}$
return $\boldsymbol{X}' \coloneqq b \mathrel{?} (\perp, \perp, \perp) : (\boldsymbol{X}^{\mathsf{Y}}, \boldsymbol{X}^{\mathsf{Cb}}, \boldsymbol{X}^{\mathsf{Cr}})$

We also want to point out that while we require each macroblock to have its own $\mathsf{param}^{\mathsf{remove}}$, we can avoid linear communication complexity when transmitting the parameters from the prover $\mathcal{P}$ to the verifier $\mathcal{V}$. In fact, $\mathcal{P}$ can simply send the dimensions of the original video and the offset of the cropped or cut video with respect to the original one, and $\mathcal{V}$ can recover the parameters from these values.

5.2.3 More Complicated Operations

We discuss how to build gadgets for more complex editing operations that involve multiple macroblocks, such as rotation. While $\mathcal{F}^{\Delta}$ handles macroblocks one-by-one in our design, it still allows such advanced functionalities. To this end, we can leverage vector commitment schemes [66] such as Merkle trees, where one can commit to the entire vector of messages and later open the commitment to the message at a specific position.

Now, $\mathcal{F}^{\Delta}$ additionally takes as input the vector commitment to the original video $\boldsymbol{V}$. For an editing operation that *reads* both the current macroblock $\boldsymbol{X}_i$ and another macroblock $\boldsymbol{X}_j$, the prover can feed $\boldsymbol{X}_j$ to $\mathcal{F}^{\Delta}$ as a hint, and $\mathcal{F}^{\Delta}$ enforces that $\boldsymbol{X}_j$ is indeed the j-th macroblock in the video by checking the vector commitment against $\boldsymbol{X}_j$ and j. Similarly, we can also support operations that *update* macroblocks in different positions by including the vector commitment to the edited video $\boldsymbol{V}'$ as input. In this way, $\mathcal{F}^{\Delta}$ is able to access other macroblocks in the video. without affecting its macroblock-wise design.

5.3 Building the Step Circuit

With the gadgets for video encoding and editing in place, we are now ready to construct the step circuit $\mathcal{F}^{\mathsf{Eva}}$. We discuss how $\mathcal{F}^{\mathsf{Eva}}$ achieves the proof of correct editing and encoding and the proof of valid signature separately.

Proof of Editing and Encoding. For the correctness of editing and encoding, $\mathcal{F}^{\mathsf{Eva}}$ can utilize the $\mathcal{F}^{\Delta}$ and $\mathcal{F}^{\mathcal{E}}$ gadgets. Since both gadgets extensively use bitwise operations, we fill the lookup table $\boldsymbol{\tau}$ with 2^8 entries in $\mathbb{Z}_{2^s}$ to maximize the efficiency. Then, for a macroblock $\boldsymbol{X}$, $\mathcal{F}^{\mathsf{Eva}}$ runs $\mathcal{F}^{\Delta}$ on $\boldsymbol{X}$ to obtain the edited macroblock $\boldsymbol{X}'$, and then invokes $\mathcal{F}^{\mathcal{E}}$ on $\boldsymbol{X}'$ to get the quantized coefficients $\boldsymbol{Z}$.

We further extend $\mathcal{F}^{\mathsf{Eva}}$ to handle b macroblocks $\{\boldsymbol{X}_j\}_{j=0}^{b-1}$ in a batch, where each $\boldsymbol{X}_j$ is associated with public inputs $\boldsymbol{P}_j$ and $\boldsymbol{Z}_j$. As we will see in Section 6 and Section 7, with a reasonably large b, we can amortize the constraints for $\mathsf{NIFS}.\mathcal{V}$ in the augmented circuit $\mathcal{F}^{\mathsf{aug}}$, thereby enabling a more efficient IVC prover.

Nevertheless, the naive combination of $\mathcal{F}^{\Delta}$ and $\mathcal{F}^{\mathcal{E}}$ is suboptimal. Recall that in IVC, the circuit $\mathcal{F}^{\mathsf{aug}}$ computes $\mathsf{H}(\mathbb{U}_i, \cdot)$ and $\mathsf{H}(\mathbb{U}_{i+1}, \cdot)$, where $\mathbb{U}_i.\boldsymbol{x}$ and $\mathbb{U}_{i+1}.\boldsymbol{x}$ contain all the public inputs to the step circuit, which are $\{\boldsymbol{P}_j\}_{j=0}^{b-1}$ and $\{\boldsymbol{Z}_j\}_{j=0}^{b-1}$ in our case. Thus, the circuit needs to hash these data twice, which becomes expensive when b is large.

We tackle this problem by finding the balance point between the advantage of utilizing verifier's knowledge and the drawback of handling public inputs in IVC. In fact, it's possible to avoid treating $\{\boldsymbol{P}_j\}_{j=0}^{b-1}$ and $\{\boldsymbol{Z}_j\}_{j=0}^{b-1}$ as public inputs while enjoying the shared information between the encoder and the decoder. Instead, we only treat them as witnesses, and the public input is now their digest $\hbar$. More specifically, in the i-th step of IVC, we absorb $\{\boldsymbol{P}_{bi+j}\}_{j=0}^{b-1}$ and $\{\boldsymbol{Z}_{bi+j}\}_{j=0}^{b-1}$ into $\hbar_i$ via H, thereby obtaining the next state $\hbar_{i+1}$. In this way, $\boldsymbol{P}$ and $\boldsymbol{Z}$ are no longer involved the digest computation $\mathsf{H}(\mathbb{U}_i, \cdot)$ and $\mathsf{H}(\mathbb{U}_{i+1}, \cdot)$ in $\mathcal{F}^{\mathsf{aug}}$. Instead, the prover only needs to compute their digest once per step in IVC, which occurs in $\mathcal{F}^{\mathsf{Eva}}$.

The soundness is unaffected: the verifier can derive $\hbar_{i+1}$ from $\hbar_i, \boldsymbol{P}, \boldsymbol{Z}$ as well and check the proof against $\hbar_{i+1}$, but the collision resistance of H prevents a malicious prover from providing incorrect $\boldsymbol{P}'$ and $\boldsymbol{Z}'$ that lead to the same $\hbar_{i+1}$.

Proof of Valid Signature. Now, we discuss how to prove the validity of σ. Due to the hash-and-sign paradigm, the signature σ is actually for the digest of $\boldsymbol{V}$ and meta. By regarding the digest computation of $\boldsymbol{V}$ as an iterative invocation of H on each macroblock in the video, we can extend $\mathcal{F}^{\mathsf{Eva}}$ by hashing the original macroblock $\boldsymbol{X}$ as well. On the other hand, the hash of meta and the execution of $\mathsf{Sig}.\mathcal{V}$ are deferred to the end of IVC, as we will see in Section 5.4.

Now, the i-th state of IVC $\boldsymbol{z}_i$ not only contains the digest $\hbar_i$ of prediction macroblocks and quantized coefficients, but it also records h_i, the hash of macroblocks in the original video. In each step, $\mathcal{F}^{\mathsf{Eva}}$ additionally updates the digest h_{i+1} by absorbing the incoming macroblocks $\{\boldsymbol{X}_{bi+j}\}_{j=0}^{b-1}$ into h_i.

Furthermore, to increase parallelism, we compute the digests h_{i+1} and $\hbar_{i+1}$ in two steps: 1) calculate the partial digests $h'_{bi+j} := \mathsf{H}(\boldsymbol{X}_{bi+j})$ and $\hbar'_{bi+j} := \mathsf{H}(\boldsymbol{P}_{bi+j}, \boldsymbol{Z}_{bi+j}, \mathsf{param}_{bi+j})$ for all $j \in [0, b-1]$, and 2) derive the final digests h_{i+1} and $\hbar_{i+1}$ by hashing the partial digests $\{h'_{bi+j}\}_{j=0}^{b-1}$ and $\{\hbar'_{bi+j}\}_{j=0}^{b-1}$.

The final construction of $\mathcal{F}^{\mathsf{Eva}}$ is given in Circuit 16.

In addition to the points discussed above, we take extra care to handle the possible removal of macroblocks due to the cropping or cutting operations ($\mathcal{F}^{\Delta_{\mathsf{remove}}}$ in Section 5.2). For a removed macroblock $\boldsymbol{X}'$, $\mathcal{F}^{\mathcal{E}}$ and subsequent operations should not be performed, since $\boldsymbol{X}'$ is no longer encoded by $\mathcal{E}$.

Such a design introduces different control flows depending on a dynamic parameter $\mathsf{param}^{\mathsf{remove}}$, resulting in a non-uniform circuit that is not directly supported by our IVC. A common technique to avoid this non-uniformity is to run all possible control flows, and then select among the results based on the

Circuit 16: $\mathcal{F}^{\mathsf{Eva}}$

Witness: $\boldsymbol{z}_i, \{\boldsymbol{X}_{bi+j}\}_{j=0}^{b-1}, \{\mathsf{param}_{bi+j}\}_{j=0}^{b-1}$

$(h_i, \hbar_i) \coloneqq \boldsymbol{z}_i$

for $j \in [0, b-1]$ **do**

  $\boldsymbol{X}'_{bi+j} \leftarrow \mathcal{F}^{\Delta}(\boldsymbol{X}_{bi+j}, \mathsf{param}^{\Delta}_{bi+j})$

  $\boldsymbol{P}_{bi+j} \leftarrow \mathsf{Hint}(\boldsymbol{X}'_{bi+j})$

  $\boldsymbol{Z}_{bi+j} \leftarrow \mathcal{F}^{\mathcal{E}}(\boldsymbol{X}'_{bi+j}, \boldsymbol{P}_{bi+j}, \mathsf{param}^{\mathcal{E}}_{bi+j})$

  $h'_{bi+j} \coloneqq \mathsf{H}(\boldsymbol{X}_{bi+j})$

  $\hbar'_{bi+j} \coloneqq \mathsf{H}(\boldsymbol{P}_{bi+j}, \boldsymbol{Z}_{bi+j}, \mathsf{param}_{bi+j})$

  if $\Delta = \Delta_{\mathsf{remove}}$ **then**

    $\hbar'_{bi+j} \coloneqq \mathsf{param}^{\mathsf{remove}}_{bi+j}\ ?\ \mathsf{param}^{\mathsf{remove}}_{bi+j} : \hbar'_{bi+j}$

$h_{i+1} \coloneqq \mathsf{H}(h_i, \{h'_{bi+j}\}_{j=0}^{b-1})$

$\hbar_{i+1} \coloneqq \mathsf{H}(\hbar_i, \{\hbar'_{bi+j}\}_{j=0}^{b-1})$

if $\Delta = \Delta_{\mathsf{remove}}$ **then**

  $\hbar_{i+1} \coloneqq (\bigwedge_{j\in[0,b-1]} \mathsf{param}^{\mathsf{remove}}_{bi+j})\ ?\ \hbar_i : \hbar_{i+1}$

return $\boldsymbol{z}_{i+1} \coloneqq (h_{i+1}, \hbar_{i+1})$

dynamic branching condition: 1) Perform $\mathcal{F}^{\mathcal{E}}$ and H to derive $\hbar'$, as if $\boldsymbol{X}'$ is not removed. We can use dummy values for $\boldsymbol{X}'$, $\boldsymbol{P}$, $\boldsymbol{Z}$, and $\mathsf{param}^{\mathcal{E}}$ if they do not exist. 2) Compute $\hbar'$ without $\boldsymbol{P}$, $\boldsymbol{Z}$, and $\mathsf{param}^{\mathcal{E}}$, *i.e.*,, $\hbar' \coloneqq \mathsf{H}(\mathsf{param}^{\mathsf{remove}})$. We further get rid of the hash and set $\hbar' \coloneqq \mathsf{param}^{\mathsf{remove}}$, as the parameter only has a single bit. Later, we select between 1) and 2) based on the branching condition $\mathsf{param}^{\mathsf{remove}}$.

Nevertheless, this approach is still deficient. Recall that $\hbar$ is a public input computed by both $\mathcal{P}$ and $\mathcal{V}$. Thus, for a very large original footage $\boldsymbol{V}$, even the cropped (or cut) video ζ is very small, $\mathcal{V}$ still needs to compute the hash of dummy values for the non-existent $\boldsymbol{P}$ and $\boldsymbol{Z}$. In fact, $\mathcal{V}$'s costs are the same as if nothing is removed.

To save verification cost, one may consider running all control flows in-circuit when computing $\hbar_{i+1}$, *i.e.*, computing $\mathsf{H}(\hbar_i, S)$ for all $S \in 2^{\{\hbar'_{bi+j}\}_{j=0}^{b-1}}$, where $2^{\{\hbar'_{bi+j}\}_{j=0}^{b-1}}$ is the power set of $\{\hbar'_{bi+j}\}_{j=0}^{b-1}$, and then the correct result can be selected. It is straightforward to see the downside of this approach: it significantly increases the prover's complexity.

We take a hybrid approach by reducing the number of branches to 2, depending on whether *all* macroblocks in a batch of size b are discarded. If this is the case, the circuit selects the previous digest $\hbar_i$ as the next digest $\hbar_{i+1}$. Otherwise, the circuit selects $\mathsf{H}(\hbar_i, \{\hbar'_{bi+j}\}_{j=0}^{b-1})$ as $\hbar_{i+1}$. As a result, $\mathcal{P}$ only needs to additionally handle 3 constraints while making it unnecessary for $\mathcal{V}$ to hash all dummy values. In fact, what $\mathcal{V}$ computes now is the hash of $\boldsymbol{P}$ and $\boldsymbol{Z}$ for a cropped (or cut) video whose size is padded to a multiple of the batch size b, which is pretty close to the actual size of ζ.

5.4 Final Protocol

Having built an IVC scheme based on our variant of Nova with support for lookup arguments in Section 4 and the IVC step circuit in Section 5.3, we now present Eva, a succinct, efficient, and secure proof of video authenticity.

In $\mathcal{K}_\Pi$, the trusted party instantiates $(\mathsf{pk}_\Pi, \mathsf{vk}_\Pi)$ with the proving and verification keys for the IVC scheme and the corresponding decider, and in $\mathcal{K}_\Sigma$, $(\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma)$ is obtained by invoking $\mathsf{Sig}.\mathcal{K}$, where Sig is instantiated with Schnorr signature [67]. Then, given the signing key sk_Σ, the recorder $\mathcal{R}$ computes the signature σ on the video $\boldsymbol{V}$ and its metadata meta as $\sigma \leftarrow \mathsf{Sig}.\mathcal{S}(\mathsf{sk}_\Sigma, \mathsf{H}(\boldsymbol{V}, \mathsf{meta}))$.

Next, we dive into the details of our prover $\mathcal{P}$ and verifier $\mathcal{V}$. $\mathcal{P}$ aims to convince $\mathcal{V}$ the satisfiability of $\mathcal{F}^{\mathsf{Eva}}$. To this end, $\mathcal{P}$ first instantiates the lookup table $\boldsymbol{\tau}$ with 2^8 entries $\{0, \ldots, 255\}$. Then $\mathcal{P}$ prepares the inputs to $\mathcal{F}^{\mathsf{Eva}}$ by transforming $\boldsymbol{V}$ into $\boldsymbol{V}'$ via Δ, and using $\mathcal{E}$ to encode $\boldsymbol{V}'$, during which the quantized coefficients $\{\boldsymbol{Z}_i\}$ and prediction macroblocks $\{\boldsymbol{P}_i\}$ are extracted. Next, $\mathcal{P}$ incrementally proves the satisfiability of $\mathcal{F}^{\mathsf{Eva}}$ using our Nova-based IVC.After $k = (M/16 \times N/16 \times L)/b$ steps, the final IVC state becomes $\boldsymbol{z}_k = (h_k, \hbar_k)$, where h_k is the digest of $\{\boldsymbol{X}_i\}_{i=0}^{bk-1}$, and $\hbar_k$ is the digest of $\{\boldsymbol{P}_i\}_{i=0}^{bk-1}$, $\{\boldsymbol{Z}_i\}_{i=0}^{bk-1}$, and $\{\mathsf{param}_i\}_{i=0}^{bk-1}$. Finally, $\mathcal{P}$ compresses the IVC proof with a decider based on ZKCP and returns the compressed zero-knowledge proof as well as the video stream ζ. These data are sent to $\mathcal{V}$, together with the metadata meta and editing and encoding parameters param.

As mentioned in Section 5.3, it is still left to compute $\mathsf{H}(h_k, \mathsf{meta})$ and run $\mathsf{Sig}.\mathcal{V}$ on the digest. Since $\mathsf{Decider}.\mathcal{V}$ takes the final state as public inputs, we can give $\mathcal{V}$ the hash h_k and ask $\mathcal{V}$ to handle the rest of verification. However, this approach is suboptimal because of the weak security guarantee: h_k leaks information about the original video $\boldsymbol{V}$, leading to compromise of the zero-knowledge property.

To achieve full zero-knowledge, we exploit the flexibility of the decider circuit $\mathcal{F}^{\mathsf{Decider}}$ and hide h_k from $\mathcal{V}$. More specifically, we 1) verify σ on $\mathsf{H}(h_k, \mathsf{meta})$ under vk_Σ in $\mathcal{F}^{\mathsf{Decider}}$, and 2) move the computations related to $\mathbb{u}_k.\boldsymbol{x}$ in $\mathsf{Decider}.\mathcal{V}$ to $\mathcal{F}^{\mathsf{Decider}}$, as the first component of $\mathbb{u}_k.\boldsymbol{x}$, *i.e.*, $\mathsf{H}(\mathbb{U}_k, k, \boldsymbol{z}_0, \boldsymbol{z}_k)$, also leaks h_k.

In our adapted decider circuit $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$, the statement $\mathbb{U}'_k$ and $\mathbb{u}'_k$ now no longer include $\boldsymbol{x}$. Instead, the prover provides h_k and $\boldsymbol{x}_k$ as witnesses, and the circuit reconstructs $\mathbb{U}_k$ by merging $\mathbb{U}'_k$ with $\boldsymbol{x}_k$, and $\mathbb{u}_k$ by merging $\mathbb{u}'_k$ with $(\mathsf{H}(\mathbb{U}_k, k, \boldsymbol{z}_0, (h_k, \hbar_k)), \mathsf{H}(\mathbb{U}_k^{\mathsf{cf}}, k), \rho(\mathbb{u}_k.\overline{Q}))$. Then, the circuit computes $\mathbb{U}_{k+1}^{\mathbb{F}}$ using the field-only operation $\mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}$, and finally, checks $\mathbb{W}_{k+1}$ against $\mathbb{U}_{k+1}^{\mathbb{F}}$. The final construction of $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$ is given in Circuit 17.

To verify the proof, $\mathcal{V}$ checks if the metadata meta and parameters param are acceptable. Similar to $\mathcal{P}$, $\mathcal{V}$ runs the decoding algorithm $\mathcal{D}$ on ζ to obtain $\{\boldsymbol{P}_i\}_{i=0}^{bk-1}$ and $\{\boldsymbol{Z}_i\}_{i=0}^{bk-1}$. After that, $\mathcal{V}$ computes $\hbar_k$ by hashing $\{\boldsymbol{P}_i\}_{i=0}^{bk-1}$, $\{\boldsymbol{Z}_i\}_{i=0}^{bk-1}$, and $\{\mathsf{param}_i\}_{i=0}^{bk-1}$. It is also $\mathcal{V}$'s task to check the commitments in $\mathbb{U}_k^{\mathbb{G}}, \mathbb{u}_k^{\mathbb{G}}$ and $\mathbb{U}_{k+1}^{\mathbb{G}}$, which are not included in the decider circuit $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$ due to the complexity of non-native group operations. With the randomness r and the cross term commitment $\overline{T}$, $\mathcal{V}$ derives $\mathbb{U}_{k+1}^{\mathbb{G}}$ by calling $\mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}$ on $\mathbb{U}_k^{\mathbb{G}}, \mathbb{u}_k^{\mathbb{G}}$. The commitments $\overline{Q}, \overline{W}, \overline{E}$ in $\mathbb{U}_{k+1}^{\mathbb{G}}$ are linked to the in-circuit witnesses $\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}$ in $\mathbb{W}_{k+1}$ via ZKCP. Note that $\mathcal{V}$ cannot learn h_k from $r := \rho(\mathbb{U}_k, \mathbb{u}_k, \overline{T})$, since $\mathbb{U}_k.\boldsymbol{x}$, the random linear combination of all previous public inputs, is also kept secret. Finally, by running $\mathsf{ZKCP}.\mathcal{V}$, the verifier can check the authenticity of the video.

We summarize the complete Eva protocol in Algorithm 18.

Circuit 17: $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$

Witness: $h_k, \sigma, \boldsymbol{x}_k, \mathbb{W}_{k+1}, \mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}}$
Statement: $\mathsf{vk}_\Sigma, \mathsf{meta}, k, \boldsymbol{z}_0, \hbar_k, r, \mathbb{u}'_k, \mathbb{U}'_k, \overline{T}$
Constant: $\mathsf{CS}^{\mathsf{aug}} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}), \mathsf{CS}^{\mathsf{cf}} = (\boldsymbol{A}^{\mathsf{cf}}, \boldsymbol{B}^{\mathsf{cf}}, \boldsymbol{C}^{\mathsf{cf}}), \mathsf{ck}^{\mathsf{cf}}$

enforce $\mathsf{Sig}.\mathcal{V}(\mathsf{vk}_\Sigma, \sigma, \mathsf{H}(h_k, \mathsf{meta}))$ ▷ Verify σ

Reconstruct $\mathbb{U}_k$ and $\mathbb{u}_k$:
- $\mathbb{U}_k := \mathbb{U}'_k$
- $\mathbb{U}_k.\boldsymbol{x} := \boldsymbol{x}_k$
- $\mathbb{u}_k := \mathbb{u}'_k$
- $\mathbb{u}_k.\boldsymbol{x} := (\mathsf{H}(\mathbb{U}_k, k, \boldsymbol{z}_0, (h_k, \hbar_k)), \mathsf{H}(\mathbb{U}_k^{\mathsf{cf}}, k), \rho(\mathbb{u}_k.\overline{Q}))$

enforce $r = \rho(\mathbb{U}_k, \mathbb{u}_k, \overline{T})$ ▷ Check r

$\mathbb{U}_{k+1}^{\mathbb{F}} := \mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}(\mathsf{vk}, \mathbb{U}_k^{\mathbb{F}}, \mathbb{u}_k^{\mathbb{F}}, r)$ ▷ Compute $\mathbb{U}_{k+1}^{\mathbb{F}}$

Check $\mathbb{W}_{k+1}$ against $\mathbb{U}_{k+1}^{\mathbb{F}}$:
- Parse $(u, \boldsymbol{x}) := \mathbb{U}_{k+1}^{\mathbb{F}}, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbb{W}_{k+1}$
- $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$
- **enforce** $\boldsymbol{Av} \circ \boldsymbol{Bv} = u \cdot \boldsymbol{Cv} + \boldsymbol{e}$

Check $\mathbb{W}_k^{\mathsf{cf}}$ against $\mathbb{U}_k^{\mathsf{cf}}$:
- Parse $(u, \boldsymbol{x}, \overline{Q}, \overline{W}, \overline{E}) := \mathbb{U}_k^{\mathsf{cf}}, (\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{e}) := \mathbb{W}_k^{\mathsf{cf}}$
- $\boldsymbol{v} := (u, \boldsymbol{x}, \boldsymbol{q}, \boldsymbol{w})$
- **enforce** $\boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v} \equiv u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e} \pmod q$
- **enforce** $\boldsymbol{q} = \varnothing \wedge \overline{Q} = \overline{0}$
- **enforce** $\mathsf{CM}.\mathcal{V}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{w}, \overline{W})$
- **enforce** $\mathsf{CM}.\mathcal{V}(\mathsf{ck}^{\mathsf{cf}}, \boldsymbol{e}, \overline{E})$

5.5 Security

We formally capture the security properties of Eva in Theorem 1, whose proof is deferred to Appendix D.

Theorem 1. *Eva is a succinct and zero-knowledge proof of video authenticity.*

6 Implementation and Optimization

We rely on the H.264 reference implementation `JM` [65] to encode and decode videos with the H.264 Main profile. We modify its source code and hook the encoding and decoding processes to extract the prediction macroblocks $\{\boldsymbol{P}_i\}$ and the quantized coefficients $\{\boldsymbol{Z}_i\}$, which are necessary for proof generation and verification.

Then we develop Eva in Rust over the BN254/Grumpkin half-pairing cycle of curves. The architecture of our implementation is illustrated in Figure 3, where we highlight the efforts of our own and the improvements to existing work with solid and dashed shapes. In the implementation, we make heavy use of the `arkworks` library [68] for algebraic operations and circuit constructions. Our variant of Nova is built upon the folding schemes implemented in `sonobe` [38], but we add support for LogUp and introduce various improvements that we will discuss soon. We also provide an alternative implementation of LegoGro16. Unlike the original implementation [69], ours is more flexible and performant: it allows for shared witnesses $\{\boldsymbol{v}\}_{i=0}^{\ell-1}$ with arbitrary length and supports increased parallelism.

Algorithm 18: Eva

Fn $\mathit{Eva}.\mathcal{K}_\Sigma(1^\lambda)$**:**
  return $(\mathsf{sk}_\Sigma, \mathsf{vk}_\Sigma) \leftarrow \mathsf{Sig}.\mathcal{K}(1^\lambda)$

Fn $\mathit{Eva}.\mathcal{K}_\Pi(1^\lambda)$**:**
  $\mathsf{pp} \leftarrow \mathsf{IVC}.\mathcal{G}(1^\lambda)$ ▷ pp contains ck
  $(\mathsf{pk}_\Phi, \mathsf{vk}_\Phi) := \mathsf{IVC}.\mathcal{K}(\mathsf{pp}, \mathcal{F}^{\mathsf{Eva}})$
  $(\mathsf{pk}, \mathsf{vk}) \leftarrow \mathsf{ZKCP}.\mathcal{K}(1^\lambda, \mathsf{ck}, \mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}})$
  return $(\mathsf{pk}_\Pi := (\mathsf{pk}, \mathsf{pk}_\Phi), \mathsf{vk}_\Pi := (\mathsf{vk}, \mathsf{vk}_\Phi))$

Fn $\mathit{Eva}.\mathcal{R}(\mathsf{sk}_\Sigma, \boldsymbol{V}, \mathsf{meta})$**:**
  return $\sigma \leftarrow \mathsf{Sig}.\mathcal{S}(\mathsf{sk}_\Sigma, \mathsf{H}(\boldsymbol{V}, \mathsf{meta}))$

Fn $\mathit{Eva}.\mathcal{P}(\mathsf{pk}_\Pi, \mathsf{vk}_\Sigma, \boldsymbol{V}, \mathsf{meta}, \mathsf{param}, \sigma)$**:**
  $\boldsymbol{V}' := \Delta(\boldsymbol{V}, \{\mathsf{param}_i^\Delta\}_{i=0}^{bk-1})$
  Encode $\boldsymbol{V}'$ and extract $\{\boldsymbol{P}_i\}_{i=0}^{bk-1}, \{\boldsymbol{Z}_i\}_{i=0}^{bk-1}$:
    $\zeta := \mathcal{E}(\boldsymbol{V}', \{\mathsf{param}_i^{\mathcal{E}}\}_{i=0}^{bk-1})$
  $\boldsymbol{z}_0 := (0,0)$, $\pi_0 := ((\mathbb{U}_\perp, \mathbb{W}_\perp), (\mathbb{U}_\perp, \mathbb{W}_\perp), (\mathbb{U}_\perp^{\mathsf{cf}}, \mathbb{W}_\perp^{\mathsf{cf}}))$
  for $j \in [0, k)$ **do**
    $\mathsf{aux}_j := (\{(\boldsymbol{X}_i, \boldsymbol{P}_i, \boldsymbol{Z}_i, \mathsf{param}_i)\}_{i=bj}^{bj+b-1})$
    $\pi_{j+1} \leftarrow \mathsf{IVC}.\mathcal{P}(\mathsf{pk}_\Phi, (j, \boldsymbol{z}_0, \boldsymbol{z}_j), \mathsf{aux}_j, \pi_j)$
    $\boldsymbol{z}_{j+1} := \mathcal{F}(\boldsymbol{z}_j; \mathsf{aux}_j)$
  Parse $(h_k, \hbar_k) := \boldsymbol{z}_k$, $((\mathbb{U}_k, \mathbb{W}_k), (\mathbb{u}_k, \mathbb{w}_k), (\mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}})) := \pi_k$
  $(\mathbb{U}_{k+1}, \mathbb{W}_{k+1}, \overline{T}) := \mathsf{NIFS}.\mathcal{P}(\mathsf{pk}_\Phi, (\mathbb{U}_k, \mathbb{W}_k), (\mathbb{u}_k, \mathbb{w}_k))$
  $r := \rho(\mathbb{U}_k, \mathbb{u}_k, \overline{T})$
  $\boldsymbol{x} := (\mathsf{vk}_\Sigma, \mathsf{meta}, k, \boldsymbol{z}_0, \hbar_k, r, \mathbb{u}_k', \mathbb{U}_k', \overline{T})$, $\boldsymbol{c} := (\mathbb{U}_{k+1}^{\mathbb{G}})$
  $\boldsymbol{\upsilon} := (\mathbb{W}_{k+1})$, $\boldsymbol{\omega} := (h_k, \sigma, \mathbb{U}_k.\boldsymbol{x}, \mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}})$
  $\varpi \leftarrow \mathsf{ZKCP}.\mathcal{P}(\mathsf{pk}, \boldsymbol{x}, \boldsymbol{c}, \boldsymbol{\upsilon}, \boldsymbol{\omega})$
  return $\zeta, \pi := (\varpi, \mathbb{U}_k', \mathbb{u}_k', \overline{T}, r)$

Fn $\mathit{Eva}.\mathcal{V}(\mathsf{vk}_\Pi, \mathsf{vk}_\Sigma, \zeta, \mathsf{meta}, \mathsf{param}, \pi)$**:**
  Parse $(\varpi, \mathbb{U}_k', \mathbb{u}_k', \overline{T}, r) := \pi$
  Decode ζ and extract $\{\boldsymbol{P}_i\}_{i=0}^{bk-1}, \{\boldsymbol{Z}_i\}_{i=0}^{bk-1}$:
    $\widetilde{\boldsymbol{V}} := \mathcal{D}(\zeta, \{\mathsf{param}_i^{\mathcal{E}}\}_{i=0}^{bk-1})$
  $\hbar_0 := 0$
  for $j \in [0, k)$ **do**
    $\hbar_{j+1} := \mathsf{H}(\hbar_j, \{\mathsf{H}(\boldsymbol{P}_i, \boldsymbol{Z}_i, \mathsf{param}_i)\}_{i=bj}^{bj+b-1})$
  $\mathbb{U}_{k+1}^{\mathbb{G}} := \mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}(\mathsf{vk}_\Phi, \mathbb{U}_k^{\mathbb{G}}, \mathbb{u}_k^{\mathbb{G}}, r, \overline{T})$
  assert $\mathbb{u}_k'.u = 1, \mathbb{u}_k'.\overline{E} = \overline{0}$ ▷ Check $\mathbb{u}_k'$
  $\boldsymbol{x} := (\mathsf{vk}_\Sigma, \mathsf{meta}, k, (0,0), \hbar_k, r, \mathbb{u}_k', \mathbb{U}_k', \overline{T})$, $\boldsymbol{c} := (\mathbb{U}_{k+1}^{\mathbb{G}})$
  return $\mathsf{ZKCP}.\mathcal{V}(\mathsf{vk}, \boldsymbol{x}, \boldsymbol{c}, \varpi)$

In addition, as elaborated below, a bunch of optimizations are applied to maximize the efficiency of the prover.

GPU Acceleration. The prover's cost in our construction is dominated by the computation of commitments $\overline{Q}$, $\overline{W}$, and $\overline{E}$, which involves a multi-scalar multiplication (MSM) operation on $\mathbb{G}$ in Pedersen commitment. Due to the parallelizable nature of MSM, many existing works have investigated the acceleration of MSM on hardware that supports a high degree of parallelism, such as GPUs [70, 71], FPGAs [72], and ASICs [73].

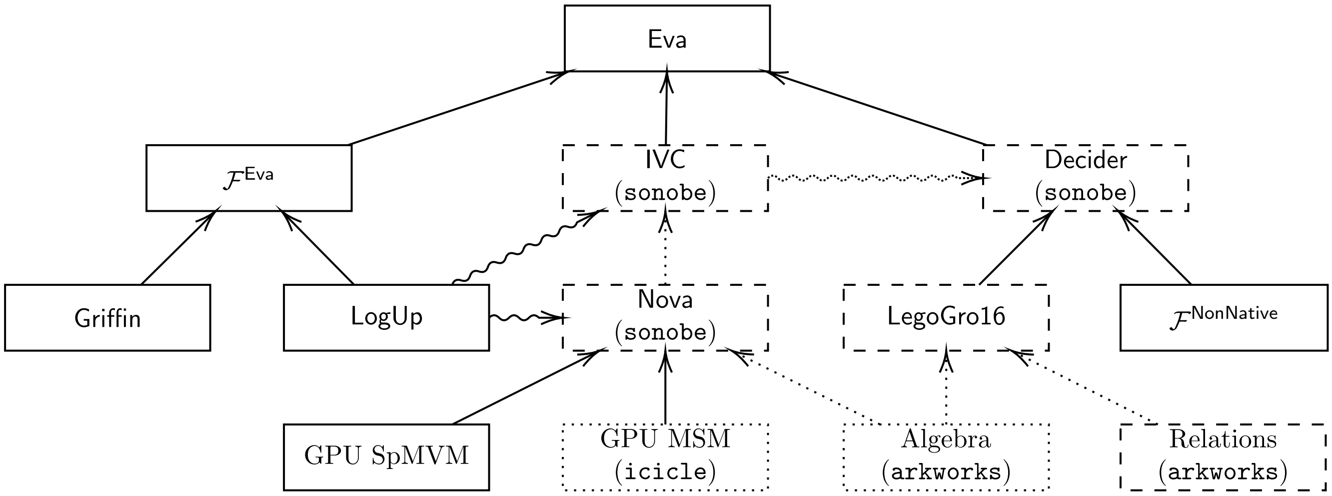


Figure 3: Architecture of Eva's implementation. A box represents a building block, a straight line from X to Y stands for "Y is built upon X", and a waved line from X to Y denotes "Y supports X". Solid shapes are implemented by ourselves from scratch, dashed shapes are our forks of third-party implementations but with significant modifications, and dotted shapes are provided by existing libraries.

We integrate `icicle` [74]'s GPU implementation of MSM with precomputation into our prover, which provides a 6-7x speedup over the original CPU implementation. While this optimization necessitates extra hardware, GPUs are more accessible and cost-effective than FPGAs and ASICs, especially in our setting where the prover already relies on powerful GPUs for video editing tasks.

We also note that the computation of cross term $\boldsymbol{t} := \boldsymbol{A}\boldsymbol{v}_1 \circ \boldsymbol{B}\boldsymbol{v}_2 + \boldsymbol{A}\boldsymbol{v}_2 \circ \boldsymbol{B}\boldsymbol{v}_1 - u_1 \cdot \boldsymbol{C}\boldsymbol{v}_2 - u_2 \cdot \boldsymbol{C}\boldsymbol{v}_1$ is another important factor in prover time. Observing that the right hand side essentially requires matrix and vector operations, we can further optimize the prover by leveraging GPU-accelerated linear algebra. Since the R1CS matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are sparse, we implement sparse matrix-vector multiplication (SpMVM) over prime fields in CUDA. With $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ represented in the compressed sparse row (CSR) format, we improve the prover's time for computing $\boldsymbol{t}$ by 2x compared to the CPU implementation.

Choice of Hash Function. It is common to use *circuit-friendly* hash functions [75–78] in SNARKs, among which Poseidon [75] is a popular choice. However, as we need to hash a large amount of data in our circuits for verifying the signature and avoiding complex prediction operations, selecting a more efficient one in our context would greatly reduce the circuit size. We choose Griffin [77] as H and ρ, which is the most efficient hash function to our knowledge in terms of the R1CS circuit size, thereby saving up to 50% of constraints compared to Poseidon. Concretely, we instantiate Griffin with degree $d = 5$, state size $t = 24$, and the number of rounds $R = 9$. Note that a large state size is necessary for improving the plain (i.e., bare-metal) performance of Griffin. Otherwise, computing Griffin hashes outside the circuit would be slower than Poseidon due to the high degree exponentiation $x^{1/d}$.

Amortizing Constraints for Folding Verification. Compared with recursive SNARKs [79], folding-based IVC reduces the prover's overhead by avoiding the in-circuit verification of a SNARK proof. However, in addition to the evaluation of $\mathcal{F}$, the prover still needs to prove the folding verification algorithm $\mathsf{NIFS}.\mathcal{V}$ in each step circuit $\mathcal{F}^{\mathsf{aug}}$. Existing techniques, such as cycle of curves [46] and CycleFold [51], makes the folding verification circuit practically small (less than 10^5 constraints), but this additional cost would become prohibitive for a small $\mathcal{F}$. In fact, when handling one macroblock per step, our $\mathcal{F}^{\mathsf{Eva}}$ circuit has ~ 3500 constraints, while proving $\mathsf{NIFS}.\mathcal{V}$ requires ~ 67000 constraints, which is a 19x increase in the prover's cost.

To minimize such overhead, we amortize the constraints for $\mathsf{NIFS}.\mathcal{V}$ by processing b macroblocks in batch in each step, so that the prover only needs to prove $\mathsf{NIFS}.\mathcal{V}$ once for every b macroblocks. The larger b is, the more the prover can save on the cost of $\mathsf{NIFS}.\mathcal{V}$. On the downside, a large b would increase the circuit size of $\mathcal{F}^{\mathsf{Eva}}$, imposing a higher memory requirement on the prover. As a trade-off between time and space, we set the batch size to $b = 256$ in our implementation, thereby reducing the cost of $\mathsf{NIFS}.\mathcal{V}$ to ~ 260 constraints per macroblock, while $\mathcal{F}^{\mathsf{Eva}}$ has a reasonable size (~ 760000 constraints).

Parallel Circuit Synthesis. Another efficiency bottleneck in our implementation is the synthesis (*i.e.*, creation) of the step circuit. As discussed above, the prover now needs to process $b = 256$ macroblocks per step. Hence, the $\mathcal{F}^{\mathsf{Eva}}$ circuit essentially consists of b copies of the logic for processing a single macroblock, which are sequentially converted into constraints when synthesizing the circuit in `arkworks`. It is natural to ask whether we can parallelize the processing of these b macroblocks, which would significantly reduce the time for circuit generation. Unfortunately, `arkworks` does not support such parallelism, since a constraint in general may depend on previously computed variables, although in our case, the constraints for each macroblock are independent of each other.

As a workaround, we 1) synthesize a *dummy circuit* for a dummy macroblock, 2) create b *partial circuits*, each of which handles one macroblock in the video, and then 3) merge the the partial circuits into the *final circuit* by concatenating the variables and constraints. The indices of variables and constraints in each partial circuit are offset by the number of variables and constraints in the dummy circuit, in order to avoid overlapping variables and constraints in the final circuit. In this way, we can parallelize 2), the most time-consuming step, without breaking the internal sequential dependencies of variables and constraints in partial circuits.

Efficient Non-Native Field Operations. Recall that in $\mathcal{F}^{\mathsf{Decider}}$, we need to check $\boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v} \equiv u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e} \pmod{q}$ in a non-native field $\mathbb{F}_q$. To this end, we can apply the in-circuit big integer arithmetic proposed in [52], which allows for efficient operations with arbitrary precision. The high-level idea behind [52] is to represent a big integer as a vector of *limbs* in the native field, and then perform limb-wise arithmetic operations. Note that since the native field cannot contain arbitrarily large limbs, we need to *align* the bitwidth of each limb after a certain number of operations, which is done by performing the expensive bit decomposition operation. Thus, when checking the equality of two big integers that are not necessarily aligned, the circuit size would become very large if we naively decompose and align the limbs before the actual comparison. While [52] decreases the number of bit decompositions in equality checks, they are still the most costly operation in the circuit.

Utilizing [52], a straightforward approach to emulation of field operations in $\mathbb{F}_q$ is to perform every operation modulo q. For instance, to multiply two non-native field elements a, b, we need to compute $c \bmod q$ after performing the big integer multiplication $c \coloneqq a \cdot b$, so that the resulting big integer c is in $\mathbb{F}_q$.

The in-circuit modulo operation can be implemented by asking the prover to provide the quotient s and the remainder r as hints, whose validity is checked by enforcing $c = sq + r$, $0 \le s < c$, and $0 \le r < q$. Nevertheless, due to the complexity of the equality check, the final circuit relying on modulo operations would have $\sim 4 \times 10^7$ constraints[5] and require more than 200 GB of memory.

To further improve the efficiency, we defer the modulo operation to the end of the circuit. That is, we avoid performing modulo operations during the computation of $LHS \coloneqq \boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v}$ and $RHS \coloneqq u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e}$. All intermediate results are treated as big integers, and LHS and RHS are converted to elements in $\mathbb{F}_q$ only before the final equality check.

In addition, we observe that it is unnecessary to compute both $LHS \bmod q$ and $RHS \bmod q$ when checking $LHS \equiv RHS \pmod{q}$. Instead, we can compute $LHS - RHS$ as a big integer, and then check whether $LHS - RHS$ is a multiple of q.

Now, we successfully get rid of all modulo operations in the circuit. Although the intermediate values during the computation of LHS and RHS contain more limbs, resulting in more expensive multiplication operations, the overall cost is still much lower than the naive approach thanks to the elimination of modulo operations. In fact, the number of constraints for checking $\boldsymbol{A}^{\mathsf{cf}}\boldsymbol{v} \circ \boldsymbol{B}^{\mathsf{cf}}\boldsymbol{v} \equiv u \cdot \boldsymbol{C}^{\mathsf{cf}}\boldsymbol{v} + \boldsymbol{e} \pmod{q}$ is reduced to $\sim 2.5 \times 10^6$, which is a 16x improvement over the naive approach.

7 Evaluation

To evaluate our implementation of Eva, we compile it with multi-threading and AVX2 enabled, and run it on a consumer-grade PC equipped with an Intel Core i9-12900K CPU (16 cores, 24 threads) with 64 GB of RAM and an NVIDIA GeForce RTX 3080 GPU with 12 GB of VRAM.

For testing purposes, we utilize two raw video files that are widely used for video codec benchmarking, as shown in Figure 4: the first, "foreman.yuv," contains 256 frames with a resolution of 352×288, while the second, "bunny.yuv," consists of 1800 frames (equivalent to 1 minute at 30 FPS) with a resolution of 1280×720.

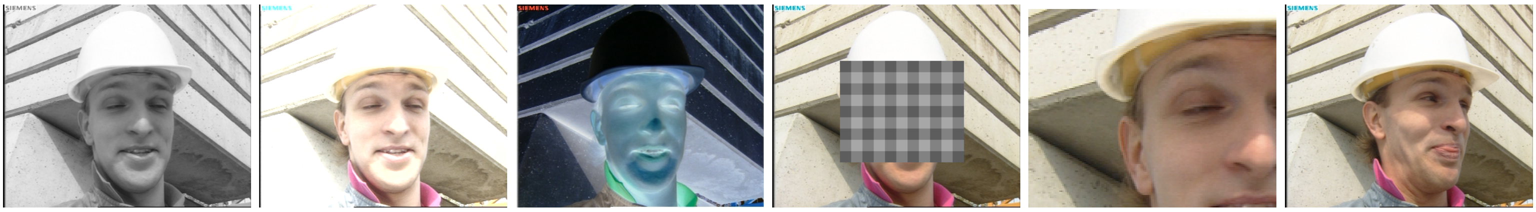
(a) foreman.yuv
352×288, 256 frames

(b) bunny.yuv
1280×720, 1800 frames

Figure 4: Preview of original videos in the test dataset

The former video is used to demonstrate Eva's capability to handle a variety of editing operations. On the other hand, the latter is for showcasing that Eva is

[5] See https://hackmd.io/x82lTH5oTcKE3uPHniuefw.



(a) Δ_{gray} (b) Δ_{bright} $\mathsf{param}^{\mathsf{bright}} = 416/256$ (c) Δ_{inv} (d) Δ_{mask} Mask of size 176×144 (e) Δ_{remove} (cropping) $352 \times 288 \to 160 \times 128$ (f) Δ_{remove} (cutting) 256 frames $\to$ 128 frames

Figure 5: Preview of edited videos in the test dataset

able to process very large videos, and we do not perform any edit operations on it. We apply several editing operations to "foreman.yuv", including grayscale conversion, brightness adjustment, color inversion, masking, cropping, and cutting, and we give the preview of the edited videos in Figure 5.

Circuit efficiency. An important metric for evaluating the performance of a protocol based on general-purpose SNARKs is the efficiency of the arithmetic circuits. We measure the circuit efficiency of Eva by the number of R1CS constraints in our augmented step circuit $\mathcal{F}^{\mathsf{aug}}$ and decider circuit $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$.

We first report the number of constraints in $\mathcal{F}^{\mathsf{aug}}$ in Table 2, where Δ_{id} is an identity function (i.e., no edits are performed). We can observe that the sizes of $\mathcal{F}^{\mathsf{Eva}}$ with all editing operations is between 600K and 1M constraints. When augmenting the circuit for the use of IVC, the additional cost is dominated by the check of lookup identity (500K to 860K constraints), as our $\mathcal{F}^{\mathsf{Eva}}$ makes heavy use of lookup tables for efficient in-circuit editing and encoding. In fact, if we replace the lookup arguments with bit decompositions, the constraints for the lookup identity check would increase by approximately eightfold (since our $\boldsymbol{\tau}$ contains 8-bit entries) in $\mathcal{F}^{\mathsf{Eva}}$, resulting in nearly an order of magnitude increase in the overall circuit size.

Next, we also list the number of constraints in our decider circuit $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$, as shown in Table 3. The size of $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$ also depends on the editing operation, because we need to check the satisfiability of $\mathbb{W}_k$ and $\mathbb{U}_k^{\mathbb{F}}$ against $\mathsf{CS}^{\mathsf{aug}}$, which has different dimensions for different Δ. The dominant parts of the circuit are the checks of relaxed R1CS satisfiability and commitment verification, each of which introduces $\sim$ 3M constraints, resulting in a total of $\sim$ 10M constraints for $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$.

Microbenchmarks. We conduct microbenchmarks to evaluate the performance of the prover in Eva.

First, we study the impact of GPU acceleration on the prover's performance by measuring the time for computing the cross term $\boldsymbol{t} := \boldsymbol{A}\boldsymbol{v}_1 \circ \boldsymbol{B}\boldsymbol{v}_2 + \boldsymbol{A}\boldsymbol{v}_2 \circ \boldsymbol{B}\boldsymbol{v}_1 - u_1 \cdot \boldsymbol{C}\boldsymbol{v}_2 - u_2 \cdot \boldsymbol{C}\boldsymbol{v}_1$ on CPU and GPU. Here, R1CS matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are generated from the augmented step circuit $\mathcal{F}^{\mathsf{aug}}$ for $\mathcal{F}^{\mathsf{Eva}}$, with batch size b ranging from 2^0 to 2^9. The results are shown in Figure 6, from which we can observe that the GPU outperforms the CPU by a factor of $1.8 \sim 2.4$. Because our SpMVM implementation is for general sparse matrices, we expect further improvements by exploiting the specific sparsity pattern of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$.

Second, we evaluate the running time of $\mathsf{IVC}.\mathcal{P}$ with respect to the batch size b. As we can see in Figure 7, when the batch size b is small, doubling b nearly

Table 2: Breakdown of the number of R1CS constraints in $\mathcal{F}^{\mathsf{aug}}$ with $b = 256$.

		Subroutine \ Editing Op.	Δ_{id}	Δ_{gray}	Δ_{bright}	Δ_{inv}	Δ_{mask}	Δ_{remove}	
$\mathcal{F}^{\mathsf{Eva}}$	Process $b = 256$ blocks	Create variables	42	42	42	42	426	43	
		$\mathcal{F}^{\Delta}$	0	0	1024	0	384	384	
	$\mathcal{F}^{\mathcal{E}}$	$\mathcal{F}^{\mathsf{Diff}}$	0						
	$\mathcal{F}^{\mathcal{E}}$	$\mathcal{F}^{\mathsf{Trans}}$	0						
	$\mathcal{F}^{\mathcal{E}}$	$\mathcal{F}^{\mathsf{Quant}}$ Y	1328						×256
	$\mathcal{F}^{\mathcal{E}}$	$\mathcal{F}^{\mathsf{Quant}}$ U	320	0	320	320	320	320	
	$\mathcal{F}^{\mathcal{E}}$	$\mathcal{F}^{\mathsf{Quant}}$ V	320	0	320	320	320	320	
		$\mathsf{H}(\boldsymbol{X})$	306						
		$\mathsf{H}(\boldsymbol{P}, \boldsymbol{Z}, \mathsf{param})$	612	612	612	612	918	613	
		$\mathsf{H}(h_i, \cdots)$	5508						
		$\mathsf{H}(\hbar_i, \cdots)$	5508	5508	5508	5508	5508	5511	
		Subtotal	760584	596744	1022728	760584	1035528	859403	
Augmentation		Create variables	6865						
		Fold $\mathbb{u}_i^{\mathbb{F}}$ into $\mathbb{U}_i^{\mathbb{F}}$	13361						
		Fold $\mathbb{u}_i^{\mathsf{cf}}$ into $\mathbb{U}_i^{\mathsf{cf}}$	45108						
		Check lookup identity	594177	430337	790785	594177	692481	594177	
		Check public inputs	9117						
		Subtotal	668628	504788	865236	668628	766932	668628	
		Total	1429212	1101532	1887964	1429212	1802460	1528031	

Table 3: Breakdown of the number of R1CS constraints in $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$ with $b = 256$.

Subroutine \ Editing Op.	Δ_{id}	Δ_{gray}	Δ_{bright}	Δ_{inv}	Δ_{mask}	Δ_{remove}
Create variables	695968					
Verify signature	4590					
Reconstruct instances	8173					
Check r	5186					
Fold $\mathbb{u}_k^{\mathbb{F}}$ into $\mathbb{U}_k^{\mathbb{F}}$	3					
Check $\mathsf{CS}^{\mathsf{aug}}$ satisfiability	2705351	2082759	3557319	2705351	3353543	2902732
Check $\mathsf{CS}^{\mathsf{cf}}$ satisfiability	2575044					
Verify commitments in $\mathbb{U}_k^{\mathsf{cf}}$	3544918					
Total	9539233	8916641	10391201	9539233	10187425	9736614

halves the average running time of $\mathsf{IVC}.\mathcal{P}$ for each macroblock. This is because for a small b, the dominant part of the augmented step circuit is still the in-circuit folding verification, thereby demonstrating the effectiveness of amortizing the constraints for $\mathsf{NIFS}.\mathcal{V}$ through batching.

Finally, we fix $b = 256$ and study how the editing operation Δ affects the prover. We report the running time and RAM usage of $\mathsf{IVC}.\mathcal{P}$ and $\mathsf{Decider}.\mathcal{P}$ in Table 4, which also includes the sizes of $\mathcal{F}^{\mathsf{aug}}$ and $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$ summarized in Table 2 and Table 3, in order to illustrate how the number of constraints affects prover time and RAM usage. Across different operations, the running time of $\mathsf{IVC}.\mathcal{P}$ ranges from 290 to 420 ms, and the peak RAM usage varies from 7 to 11 GB. $\mathsf{Decider}.\mathcal{P}$ takes much more time ($80 \sim 100$ s) and RAM ($40 \sim 50$

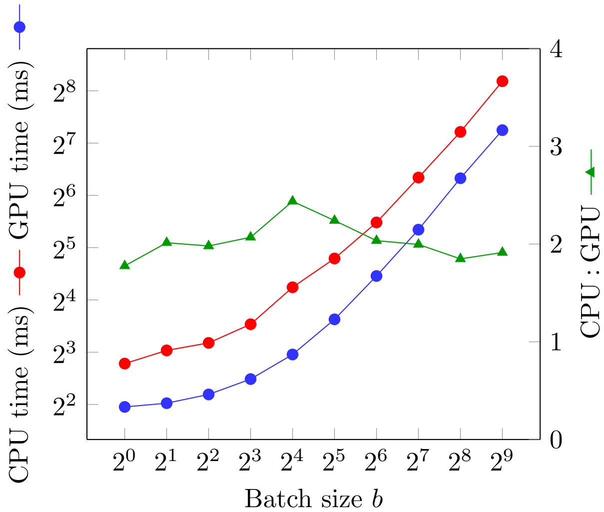


Figure 6: Running time of cross term computation on CPU and GPU w.r.t. the batch size b.

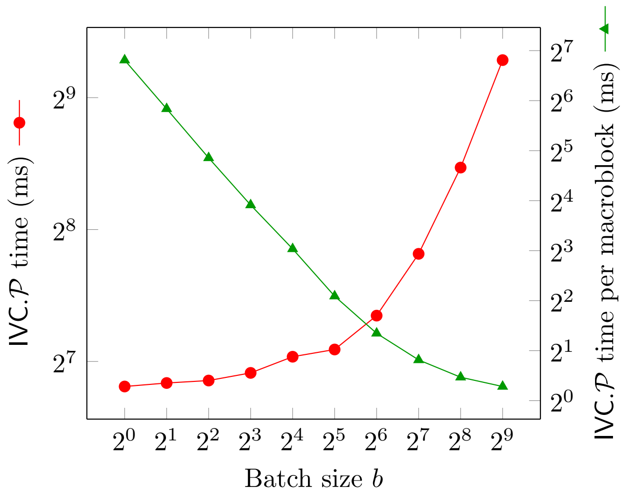


Figure 7: Running time of $\mathsf{IVC}.\mathcal{P}$ w.r.t. the batch size b, where the editing operation is fixed to $\Delta = \Delta_{\mathsf{id}}$.

GB) to generate a proof, but it only happens once at the end of $\mathsf{Eva}.\mathcal{P}$.

End-to-end performance. The end-to-end performance of Eva is evaluated based on the running time of each algorithm. We test Eva with $b = 256$ on both videos in the dataset, with different editing operations applied to “foreman.yuv”. The results are presented in Table 5.

In $\mathsf{Eva}.\mathcal{P}$, generating the final IVC proof for “foreman.yuv” needs 2 to 3 minutes. For the 1-minute HD video “bunny.yuv”, it takes ~ 2.5 hours to prove all IVC steps. An additional 80 to 100 seconds is needed to make the proof fully succinct and zero-knowledge by running $\mathsf{ZKCP}.\mathcal{P}$ in our decider, which

Table 4: Benchmarking results of $\mathsf{IVC}.\mathcal{P}$ and $\mathsf{Decider}.\mathcal{P}$ for different editing operation Δ with $b = 256$.

		Δ_{id}	Δ_{gray}	Δ_{bright}	Δ_{inv}	Δ_{mask}	Δ_{remove}
$\mathsf{IVC}.\mathcal{P}$	$\lvert\mathcal{F}^{\mathsf{aug}}\rvert$	1429212	1101532	1887964	1429212	1802460	1528031
	Time (ms)	350.408	289.478	426.646	352.887	413.162	359.715
	Peak RAM (GB)	8.493	6.845	9.644	8.566	11.084	8.694
$\mathsf{Decider}.\mathcal{P}$	$\lvert\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}\rvert$	9539233	8916641	10391201	9539233	10187425	9736614
	Time (s)	80.747	74.283	87.669	80.286	98.973	94.637
	Peak RAM (GB)	44.226	39.560	48.601	45.705	52.766	43.797

Table 5: End-to-end performance of Eva with $b = 256$.

		$\mathcal{K}_\Sigma$	$\mathcal{K}_\Pi$	$\mathcal{R}$		$\mathcal{P}$		$\mathcal{V}$	
				H	$\mathsf{Sig}.\mathcal{S}$	$\mathsf{IVC}.\mathcal{P}$ (all steps)	$\mathsf{ZKCP}.\mathcal{P}$	H	$\mathsf{ZKCP}.\mathcal{V}$
		(μs)	(s)	(s)	(μs)	(s)	(s)	(s)	(ms)
foreman	Δ_{id}	63.709	116.996	1.271	92.204	138.098	80.747	1.803	4.870
	Δ_{gray}	63.325	101.187	1.276	92.137	113.767	74.283	1.784	4.652
	Δ_{bright}	63.402	127.364	1.283	92.921	164.931	87.669	1.794	4.561
	Δ_{inv}	63.223	115.292	1.274	92.932	138.243	80.286	1.802	4.483
	Δ_{mask}	63.481	136.183	1.308	92.538	160.298	98.973	2.347	4.870
	Δ_{remove}	63.089	119.093	1.280	92.726	138.919	94.637	0.990 (crop) 0.883 (cut)	4.496
bunny	Δ_{id}	63.250	128.515	81.713	92.488	9113.536	104.997	114.301	7.288

produces a constant-sized proof of 448 bytes. The entire proof generation process is completed within 60 GB of RAM, which is primarily due to the additional $\mathsf{ZKCP}.\mathcal{P}$ step.

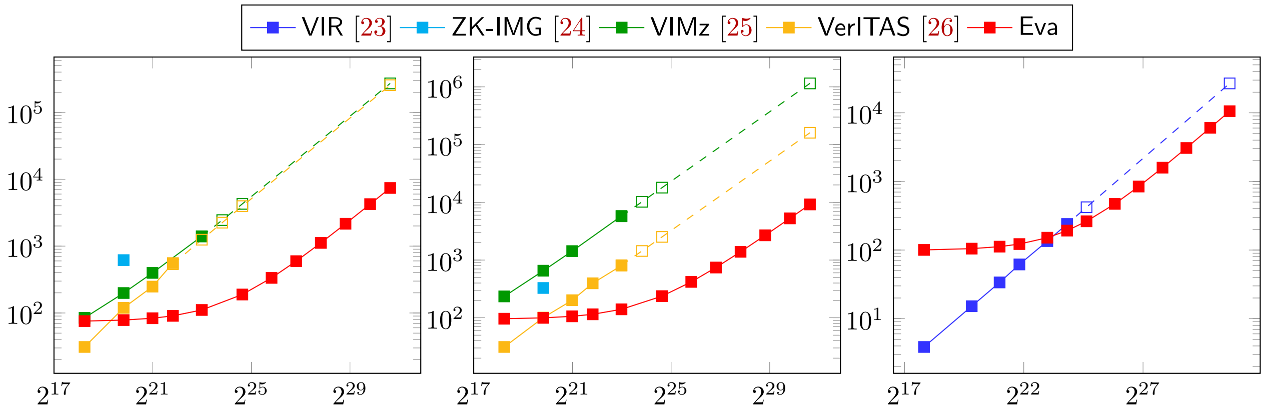
The prover time for other editing operations on "bunny.yuv", as well as for other videos with different resolutions and frame rates, can be estimated by scaling the time of $\mathsf{IVC}.\mathcal{P}$ in Table 4 according to the number of steps required. This is confirmed by the results for both "foreman.yuv" with Δ_{id} and "bunny.yuv": for the former, the estimated total time is $\frac{350.408}{1000} \times \frac{352\times288\times256}{256\times256} = 138.761$ s, and the actual time is 138.098 s; similarly, for the latter, the estimated total time is $\frac{350.408}{1000} \times \frac{1280\times720\times1800}{256\times256} = 8869.702$ s, with a relative error of 2.675% compared to the actual time of 9113.536 s. We observe that the error is majorly due to page faults and context switches when proving "bunny.yuv".

Another conclusion from the results is that, while the time required for $\mathsf{ZKCP}.\mathcal{P}$ constitutes a significant portion of the prover time for smaller videos, it is constant with respect to the number of IVC steps and becomes relatively insignificant for larger videos. In spite of the theoretical performance of $\mathsf{ZKCP}.\mathcal{P}$, the actual time for "bunny.yuv" is slower, again due to the costs of memory management and process scheduling.

The recorder $\mathcal{R}$'s running time is dominated by the computation of $\mathsf{Griffin}$ hash, whose complexity is linear in the size of the original video $\boldsymbol{V}$. Similarly, H is also the bottleneck of $\mathcal{V}$, but it depends on the size of prediction macroblocks $\boldsymbol{P}$ and quantized coefficients $\boldsymbol{Z}$ of the edited video $\boldsymbol{V}'$. Thus, with Δ_{remove}, the verifier takes less time for computing H than other editing operations. In addition, $\mathcal{V}$ needs to validate the ZKCP proof by running $\mathsf{ZKCP}.\mathcal{V}$, which takes $4 \sim 7$ ms.

Comparison with related work. Finally, we compare the performance of

Eva with related work on image authentication based on zkSNARKs [23–26], focusing specifically on the prover time. PhotoProof [22] is not included in the comparison, as it only supports tiny images of size up to 128 × 128. Due to differences in image and video encoding formats, a common dataset cannot be used across all protocols. Consequently, the prover time is evaluated based on the number of pixels in the image or video.



(a) Grayscale conversion (b) Cropping (c) Masking

Figure 8: Comparison of prover time (y-axis, in seconds) across different protocols w.r.t. the number of pixels (x-axis).

For protocols that support arbitrary editing operations [24–26], we select two representative operations for comparison: grayscale conversion, representing color manipulations, and cropping, representing spatial modifications. The target resolution for the cropping operation is set to 640 × 480. Since the source code of ZK-IMG is not available, we rely on existing results for comparison. Specifically, we adopt the prover time for operations "RGB2YCbCr" and "Crop (HD → SD)" on HD (1280 × 720) images reported in [24, Table 4]. Note that ZK-IMG was evaluated on a powerful server with 64 CPU cores and 512 GB of RAM, suggesting that the prover time would likely be slower on our machine. All remaining protocols are evaluated on the same machine as Eva.

The results of prover time are given in Figure 8a and Figure 8b, respectively. The total prover time of VIMz, VerITAS, and Eva for both Δ_{gray} and Δ_{crop} increases nearly linearly with the number of pixels. Because our figures have a logarithmic scale on both axes, the prover time for Eva, which includes the ZKCP.$\mathcal{P}$ time, does not appear as a straight line when the number of pixels is small.

We observe that Eva is generally faster than ZK-IMG, VIMz, and VerITAS, except at SD resolution (640 × 480), where VerITAS outperforms Eva due to our relatively long ZKCP.$\mathcal{P}$ time. However, as the number of pixels increases, the ZKCP-based decider is no longer the dominant factor in our prover, allowing the advantages of our efficient IVC.$\mathcal{P}$ to become apparent. In particular, for 4K resolution (3840 × 2160), Eva is 5.7 ∼ 41 times faster than VIMz and VerITAS. We estimate that if we apply VIMz and VerITAS to "bunny.yuv" (1280 × 720 × 1800 pixels) even without considering memory constraints and lossy encoding, the proof generation would be respectively 17 ∼ 33 and 37 ∼ 126 times longer than Eva.

For VIR [23], their redaction operation is equivalent to our masking operation with black tiles as the mask. Thus, we compare Eva with VIR in terms of Δ_{mask} and present the results in Figure 8c. To ensure a fair comparison, we set the granularity of redaction (*i.e.*, the minimum size of black tiles) in VIR to 1×1, matching the granularity of our masking operation. For relatively small number of pixels, our prover takes longer than VIR due to the one-time cost of the decider. However, when the data size increases, the prover time of VIR increases more rapidly than ours, and Eva begin to outperform VIR for 5K (5120×2880) and larger resolution. We also estimate that, even with unlimited RAM, VIR is 2.570x slower than Eva when proving the masking operation for "bunny.yuv".

8 Discussion

We further explore practical considerations for deploying Eva in real-world scenarios.

On-chain verification. It is possible to deploy Eva on blockchains to provide on-chain verification of video authenticity. More specifically, given $\hbar_k$ computed by the user, the smart contract can check if the proof π is valid. This is practical because π is on the BN254 curve, which is natively supported by Ethereum and its Layer 2 solutions. Also, thanks to our design of decider based on $\mathsf{LegoGroth16}$ [40], π is very small and only require 2 pairings for verification. We estimate that verifying π on EVM would require ~ 362000 gas, or equivalently ~ 16 USD as of August 2024.

In comparison, although Dziembowski et al. claim that VIMz [25] supports on-chain verification, the concrete costs of their smart contracts are not provided in their paper, which turn out to be prohibitively high. In fact, they choose $\mathsf{Spartan}$ as the zkSNARK for decider and rely on the `solidity-verifier` library [80] for verifying $\mathsf{Spartan}$ proofs on EVM, requiring $\sim$ 200M gas[6] or ~ 9000 USD.

Implementation of the recorder. Note that in Eva, both $\mathcal{R}$ and $\mathcal{P}$ take the raw footage $\boldsymbol{V}$ as input. This imply that $\mathcal{R}$ should send to $\mathcal{P}$ the recorded video $\boldsymbol{V}$ *as is*, in an *uncompressed* or *losslessly encoded* manner.

However, $\mathcal{R}$ is usually resource-constrained and only allows lossily encoded videos in practice. In this case, $\mathcal{P}$ needs to decode $\widetilde{\boldsymbol{V}}$ from the encoded video stream ζ before editing and proving. But due to information loss, $\widetilde{\boldsymbol{V}}$ is not exactly the same as the original video $\boldsymbol{V}$ that $\mathcal{R}$ signs, leading to a mismatch between the signed video and the video to be proven.

To address this mismatch, an intuitive solution is to require $\mathcal{R}$ to sign ζ. Then, $\mathcal{P}$ needs to prove 1) $\mathsf{Sig}.\mathcal{V}(\mathsf{vk}_\Sigma, \zeta, \sigma)$, 2) honest editing and encoding on $\widetilde{\boldsymbol{V}}$, and additionally 3) $\widetilde{\boldsymbol{V}} = \mathcal{D}(\zeta)$ to connect 1) with 2). Nevertheless, this approach is impractical due to the complex decoding algorithm $\mathcal{D}$.

We adopt a more strategic approach, where $\mathcal{R}$ takes an additional step of decoding ζ and signs the decoded $\widetilde{\boldsymbol{V}}$ instead of ζ. This ensures that the video that $\mathcal{R}$ signs is exactly the one proven by $\mathcal{P}$, thereby eliminating the need for proving correct decoding. Here, $\mathcal{R}$ does not need to store the decoded $\widetilde{\boldsymbol{V}}$ for signature generation, since $\mathcal{R}$ can hash $\widetilde{\boldsymbol{V}}$ on-the-fly: $\mathcal{R}$ maintains a short digest as the accumulated hash, and once a new macroblock is populated by the decoder, $\mathcal{R}$ absorbs it into the accumulated digest, which can then be discarded.

[6] For details, see https://github.com/lurk-lab/solidity-verifier/issues/29.

Acknowledgments

This research was partially funded by the Engineering and Physical Sciences Research Council (EPSRC) under grant EP/R012598/1. The authors would like to thank Yihua Cheng for the discussion on related work based on detection, and the developers of `sonobe` for reviewing the pull request for optimizing non-native field operations in the decider circuit.

References

[1] Dan Milmo. Youtube is major conduit of fake news, factcheckers say. https://www.theguardian.com/technology/2022/jan/12/youtube-is-major-conduit-of-fake-news-factcheckers-say.

[2] Emma Tucker. Tiktok's search engine repeatedly delivers misinformation to its majority-young user base, report says. https://edition.cnn.com/2022/09/18/business/tiktok-search-engine-misinformation/index.html.

[3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[4] Zach Evans, CJ Carr, Josiah Taylor, Scott H Hawley, and Jordi Pons. Fast timing-conditioned latent audio diffusion. In *Forty-first International Conference on Machine Learning*, 2024.

[5] Dan Kondratyuk, Lijun Yu, Xiuye Gu, Jose Lezama, Jonathan Huang, Grant Schindler, Rachel Hornung, Vighnesh Birodkar, Jimmy Yan, Ming-Chang Chiu, et al. Videopoet: A large language model for zero-shot video generation. In *Forty-first International Conference on Machine Learning*, 2024.

[6] Rashid Tahir, Brishna Batool, Hira Jamshed, Mahnoor Jameel, Mubashir Anwar, Faizan Ahmed, Muhammad Adeel Zaffar, and Muhammad Fareed Zaffar. Seeing is believing: Exploring perceptual differences in deepfake videos. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–16, 2021.

[7] Pavel Korshunov and Sébastien Marcel. Subjective and objective evaluation of deepfake videos. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2510–2514. IEEE, 2021.

[8] Matthew Groh, Ziv Epstein, Chaz Firestone, and Rosalind Picard. Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, 119(1):e2110013119, 2022.

[9] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3207–3216, 2020.

[10] Iacopo Masi, Aditya Killekar, Royston Marian Mascarenhas, Shenoy Pratik Gurudatt, and Wael AbdAlmageed. Two-branch recurrent network for isolating deepfakes in videos. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, pages 667–684. Springer, 2020.

[11] Alexandros Haliassos, Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic. Lips don't lie: A generalisable and robust approach to face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5039–5049, 2021.

[12] Rosario Gennaro and Pankaj Rohatgi. How to sign digital streams. In *Annual International Cryptology Conference*, pages 180–197. Springer, 1997.

[13] Coalition for Content Provenance and Authenticity. *Content Credentials: C2PA Technical Specification*, 2.0 edition, 2024.

[14] Baris Coskun, Bulent Sankur, and Nasir Memon. Spatio–temporal transform based video hashing. *ieee Transactions on Multimedia*, 8(6):1190–1208, 2006.

[15] Fouad Khelifi and Ahmed Bouridane. Perceptual video hashing for content identification and authentication. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1):50–67, 2017.

[16] Mahmoud E Farfoura, Shi-Jinn Horng, Jing-Ming Guo, and Ali Al-Haj. Low complexity semi-fragile watermarking scheme for h. 264/avc authentication. *Multimedia Tools and Applications*, 75:7465–7493, 2016.

[17] Paarth Neekhara, Shehzeen Hussain, Xinqiao Zhang, Ke Huang, Julian McAuley, and Farinaz Koushanfar. Facesigns: Semi-fragile watermarks for media authentication. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.

[18] Paarth Neekhara, Brian Dolhansky, Joanna Bitton, and Cristian Canton Ferrer. Adversarial threats to deepfake detection: A practical perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 923–932, 2021.

[19] Shehzeen Hussain, Paarth Neekhara, Malhar Jere, Farinaz Koushanfar, and Julian McAuley. Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3348–3357, 2021.

[20] Qingying Hao, Licheng Luo, Steve TK Jan, and Gang Wang. It's not what it looks like: Manipulating perceptual hashing based applications. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 69–85, 2021.

[21] Jonathan Prokos, Neil Fendley, Matthew Green, Roei Schuster, Eran Tromer, Tushar Jois, and Yinzhi Cao. Squint hard enough: Attacking perceptual hashing with adversarial machine learning. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 211–228, 2023.

[22] Assa Naveh and Eran Tromer. Photoproof: Cryptographic image authentication for any set of permissible transformations. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 255–271. IEEE, 2016.

[23] Hankyung Ko, Ingeun Lee, Seunghwa Lee, Jihye Kim, and Hyunok Oh. Efficient verifiable image redacting based on zk-SNARKs. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*, pages 213–226, 2021.

[24] Daniel Kang, Tatsunori Hashimoto, Ion Stoica, and Yi Sun. Zk-img: Attested images via zero-knowledge proofs to fight disinformation. *arXiv preprint arXiv:2211.04775*, 2022.

[25] Stefan Dziembowski, Shahriar Ebrahimi, and Parisa Hassanizadeh. VIMz: Verifiable image manipulation using folding-based zkSNARKs. *Cryptology ePrint Archive*, 2024.

[26] Trisha Datta, Binyi Chen, and Dan Boneh. Veritas: Verifying image transformations at scale. *Cryptology ePrint Archive*, 2024.

[27] Paul Valiant. Incrementally verifiable computation or proofs of knowledge imply time/space efficiency. In *Theory of Cryptography: Fifth Theory of Cryptography Conference, TCC 2008, New York, USA, March 19-21, 2008. Proceedings 5*, pages 1–18. Springer, 2008.

[28] Abhiram Kothapalli, Srinath Setty, and Ioanna Tzialla. Nova: Recursive zero-knowledge arguments from folding schemes. In *Annual International Cryptology Conference*, pages 359–388. Springer, 2022.

[29] Abhiram Kothapalli and Srinath Setty. Supernova: Proving universal machine executions without universal circuits. *Cryptology ePrint Archive*, 2022.

[30] Abhiram Kothapalli and Srinath Setty. Hypernova: Recursive arguments for customizable constraint systems. *Cryptology ePrint Archive*, 2023.

[31] Benedikt Bünz and Binyi Chen. Protostar: generic efficient accumulation/folding for special-sound protocols. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 77–110. Springer, 2023.

[32] Liam Eagen and Ariel Gabizon. Protogalaxy: Efficient protostar-style folding of multiple instances. *Cryptology ePrint Archive*, 2023.

[33] Jonathan Bootle, Andrea Cerulli, Jens Groth, Sune Jakobsen, and Mary Maller. Arya: Nearly linear-time zero-knowledge proofs for correct program execution. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 595–626. Springer, 2018.

[34] Ariel Gabizon and Zachary J Williamson. plookup: A simplified polynomial protocol for lookup tables. *Cryptology ePrint Archive*, 2020.

[35] Arantxa Zapico, Vitalik Buterin, Dmitry Khovratovich, Mary Maller, Anca Nitulescu, and Mark Simkin. Caulk: Lookup arguments in sublinear time. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 3121–3134, 2022.

[36] Srinath Setty, Justin Thaler, and Riad Wahby. Unlocking the lookup singularity with Lasso. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 180–209. Springer, 2024.

[37] ITU Telecommunication Standardization Sector. *H.264: Advanced video coding for generic audiovisual services*, 14 edition, 2021.

[38] Privacy & Scaling Explorations. sonobe: Experimental folding schemes library. https://github.com/privacy-scaling-explorations/sonobe, 2023.

[39] Ulrich Haböck. Multivariate lookups based on logarithmic derivatives. *Cryptology ePrint Archive*, 2022.

[40] Matteo Campanelli, Dario Fiore, and Anaïs Querol. Legosnark: Modular design and composition of succinct zero-knowledge proofs. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 2075–2092, 2019.

[41] Alessandro Chiesa and Eran Tromer. Proof-carrying data and hearsay arguments from signature cards. In *Innovations in Computer Science - ICS 2010*, volume 10, pages 310–331, 2010.

[42] Zcash. The Halo2 zero-knowledge proving system. https://github.com/zcash/halo2, 2022.

[43] Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, Eran Tromer, and Madars Virza. Snarks for c: Verifying program executions succinctly and in zero knowledge. In *Annual cryptology conference*, pages 90–108. Springer, 2013.

[44] Nir Bitansky, Ran Canetti, Alessandro Chiesa, and Eran Tromer. From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 326–349, 2012.

[45] Matteo Campanelli, Antonio Faonio, Dario Fiore, Anaïs Querol, and Hadrián Rodríguez. Lunar: a toolbox for more efficient universal and updatable zkSNARKs and commit-and-prove extensions. In *Advances in Cryptology–ASIACRYPT 2021: 27th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 6–10, 2021, Proceedings, Part III 27*, pages 3–33. Springer, 2021.

[46] Wilson Nguyen, Dan Boneh, and Srinath Setty. Revisiting the Nova proof system on a cycle of curves. *Cryptology ePrint Archive*, 2023.

[47] ITU Telecommunication Standardization Sector. *H.265: High efficiency video coding*, 9 edition, 2023.

[48] Alliance for Open Media. *AV1 Bitstream & Decoding Process Specification*, 1 edition, 2019.

[49] Rosario Gennaro, Craig Gentry, Bryan Parno, and Mariana Raykova. Quadratic span programs and succinct nizks without pcps. In *Advances in Cryptology–EUROCRYPT 2013: 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings 32*, pages 626–645. Springer, 2013.

[50] Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *Conference on the theory and application of cryptographic techniques*, pages 186–194. Springer, 1986.

[51] Abhiram Kothapalli and Srinath Setty. Cyclefold: Folding-scheme-based recursive arguments over a cycle of elliptic curves. *Cryptology ePrint Archive*, 2023.

[52] Ahmed Kosba, Charalampos Papamanthou, and Elaine Shi. xjsnark: A framework for efficient verifiable computation. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 944–961. IEEE, 2018.

[53] Gautam Botrel, Thomas Piellard, Youssef El Housni, Ivo Kubjas, and Arya Tabaie. gnark is a fast zk-snark library that offers a high-level api to design circuits. https://github.com/Consensys/gnark, 2023.

[54] Jens Groth. On the size of pairing-based non-interactive arguments. In *Advances in Cryptology–EUROCRYPT 2016: 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II 35*, pages 305–326. Springer, 2016.

[55] Ariel Gabizon, Zachary J Williamson, and Oana Ciobotaru. Plonk: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge. *Cryptology ePrint Archive*, 2019.

[56] Benedikt Bünz, Ben Fisch, and Alan Szepieniec. Transparent snarks from dark compilers. In *Advances in Cryptology–EUROCRYPT 2020: 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10–14, 2020, Proceedings, Part I 39*, pages 677–706. Springer, 2020.

[57] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Greg Maxwell. Bulletproofs: Short proofs for confidential transactions and more. In *2018 IEEE symposium on security and privacy (SP)*, pages 315–334. IEEE, 2018.

[58] Jonathan Lee. Dory: Efficient, transparent arguments for generalised inner products and polynomial commitments. In *Theory of Cryptography Conference*, pages 1–34. Springer, 2021.

[59] Aniket Kate, Gregory M Zaverucha, and Ian Goldberg. Constant-size commitments to polynomials and their applications. In *Advances in Cryptology-ASIACRYPT 2010: 16th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 5-9, 2010. Proceedings 16*, pages 177–194. Springer, 2010.

[60] Torben Pryds Pedersen. Non-interactive and information-theoretic secure verifiable secret sharing. In *Annual international cryptology conference*, pages 129–140. Springer, 1991.

[61] George Kadianakis, Mary Maller, and Andrija Novakovic. Sigmabus: Binding sigmas in circuits for fast curve operations. *Cryptology ePrint Archive*, 2023.

[62] Michele Orrù, George Kadianakis, Mary Maller, and Greg Zaverucha. Beyond the circuit: How to minimize foreign arithmetic in zkp circuits. *Cryptology ePrint Archive*, 2024.

[63] Srinath Setty. Spartan: Efficient and general-purpose zksnarks without trusted setup. In *Annual International Cryptology Conference*, pages 704–737. Springer, 2020.

[64] Jens Ernstberger, Chengru Zhang, Luca Ciprian, Philipp Jovanovic, and Sebastian Steinhorst. Zero-knowledge location privacy via accurate floating point SNARKs. *arXiv preprint arXiv:2404.14983*, 2024.

[65] MPEG & VCEG Joint Video Team. H.264/AVC JM reference software. https://vcgit.hhi.fraunhofer.de/jvet/JM, 2019.

[66] Dario Catalano and Dario Fiore. Vector commitments and their applications. In *Public-Key Cryptography–PKC 2013: 16th International Conference on Practice and Theory in Public-Key Cryptography, Nara, Japan, February 26–March 1, 2013. Proceedings 16*, pages 55–72. Springer, 2013.

[67] Claus-Peter Schnorr. Efficient identification and signatures for smart cards. In *Advances in Cryptology—CRYPTO'89 Proceedings 9*, pages 239–252. Springer, 1990.

[68] arkworks contributors. `arkworks` zksnark ecosystem. https://arkworks.rs, 2022.

[69] Lovesh Harchandani. legogro16: LegoGroth16 implementation on top of Zexe. https://github.com/lovesh/legogro16, 2023.

[70] Tao Lu, Chengkun Wei, Ruijing Yu, Chaochao Chen, Wenjing Fang, Lei Wang, Zeke Wang, and Wenzhi Chen. cuZK: Accelerating zero-knowledge proof with a faster parallel multi-scalar multiplication algorithm on GPUs. *Cryptology ePrint Archive*, 2022.

[71] Xudong Zhu, Haoqi He, Zhengbang Yang, Yi Deng, Lutan Zhao, and Rui Hou. Elastic MSM: A fast, elastic and modular preprocessing technique for multi-scalar multiplication algorithm on GPUs. *Cryptology ePrint Archive*, 2024.

[72] Kaveh Aasaraai, Don Beaver, Emanuele Cesena, Rahul Maganti, Nicolas Stalder, and Javier Varela. FPGA acceleration of multi-scalar multiplication: Cyclonemsm. *Cryptology ePrint Archive*, 2022.

[73] Ye Zhang, Shuo Wang, Xian Zhang, Jiangbin Dong, Xingzhong Mao, Fan Long, Cong Wang, Dong Zhou, Mingyu Gao, and Guangyu Sun. Pipezk: Accelerating zero-knowledge proof with a pipelined architecture. In *2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA)*, pages 416–428. IEEE, 2021.

[74] Ingonyama. icicle: a GPU library for zero-knowledge acceleration. `https://github.com/ingonyama-zk/icicle`, 2023.

[75] Lorenzo Grassi, Dmitry Khovratovich, Christian Rechberger, Arnab Roy, and Markus Schofnegger. Poseidon: A new hash function for Zero-Knowledge proof systems. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 519–535, 2021.

[76] Clémence Bouvier, Pierre Briaud, Pyrros Chaidos, Léo Perrin, Robin Salen, Vesselin Velichkov, and Danny Willems. New design techniques for efficient arithmetization-oriented hash functions: anemoi permutations and jive compression mode. In *Annual International Cryptology Conference*, pages 507–539. Springer, 2023.

[77] Lorenzo Grassi, Yonglin Hao, Christian Rechberger, Markus Schofnegger, Roman Walch, and Qingju Wang. Horst meets fluid-spn: Griffin for zero-knowledge applications. In *Annual International Cryptology Conference*, pages 573–606. Springer, 2023.

[78] Lorenzo Grassi, Dmitry Khovratovich, Reinhard Lüftenegger, Christian Rechberger, Markus Schofnegger, and Roman Walch. Reinforced concrete: a fast hash function for verifiable computation. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 1323–1335, 2022.

[79] Eli Ben-Sasson, Alessandro Chiesa, Eran Tromer, and Madars Virza. Scalable zero knowledge via cycles of elliptic curves. In *Advances in Cryptology–CRYPTO 2014: 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II 34*, pages 276–294. Springer, 2014.

[80] Lurk Lab. solidity-verifier: Solidity implementation of nova proving system verifier. `https://github.com/lurk-lab/solidity-verifier`, 2023.

[81] Benedikt Bünz, Alessandro Chiesa, Pratyush Mishra, and Nicholas Spooner. Proof-carrying data from accumulation schemes. 2020.

[82] Polygon Zero. Plonky2: a SNARK implementation based on techniques from plonk and FRI. `https://github.com/0xPolygonZero/plonky2`, 2021.

[83] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Fast reed-solomon interactive oracle proofs of proximity. In *45th international colloquium on automata, languages, and programming (icalp 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[84] Microsoft. PhotoDNA. `https://www.microsoft.com/en-us/photodna`.

Appendix A Additional Related Work

In the following, we provide an overview of related work. First, we examine cryptographic protocols for image authentication since they have a close relationship to video authentication and rely heavily on cryptographic proofs as well. Second, we discuss non-cryptographic methods for authenticating genuine videos and detecting fake videos.

Image authentication based on cryptographic proofs. The advance of succinct proof systems like zkSNARKs has made it possible to prove statements previously deemed infeasible, enabling the development of cryptographic protocols for image authentication. The pioneering work in this direction, PhotoProof [22], uses Proof-Carrying Data (PCD) [41] to prove the authenticity of edited images. Specifically, the proof demonstrates that the edited image m' is derived from the original image m, and the signature on the hash of m is valid. Due to the high computational cost of the proof generation, the authors only evaluate PhotoProof on images with a maximum resolution of 128×128 pixel. In [23], Ko et al. propose VIR, a verifiable image redacting protocol based on CP-SNARKs [40]. By focusing solely on the operation of redacting images (masking secret parts with black tiles), VIR significantly reduces prover time ($\sim$ 300x smaller) and supports much larger images, up to 3840×2160 resolution. Built upon a more efficient proof system `halo2` [42], ZK-IMG [24] also has $\sim$ 100x faster prover than PhotoProof, while maintaining support for arbitrary editing operations.

Concurrent with our work, Dziembowski et al. introduce VIMz [25], and Datta et al. propose VerITAS [26] which share several common ideas with Eva. For instance, VIMz also employs folding schemes to reduce prover RAM costs, and VerITAS, like Eva, utilizes lookup argument to improve prover time. However, alongside these general techniques, Eva incorporates a range of tailored optimizations to minimize prover time, resulting in better performance than both protocols. In terms of image size, both schemes support high-resolution images. VerITAS even showcases proof generation for an image of resolution 6632×4976, which is made practical due to its custom proof system for proving pre-image of lattice-based hash functions.

Considering that videos can be seen as a generalization of images, we provide a comparison between Eva and cryptographic image authentication protocols in Table 1, in terms of supported format and compression modes, allowed editing operations, prover time and RAM usage, proof size, and maximum dimensions of the input data. We compare the prover time complexity for the number of pixels of the image or video P and the per-pixel prover time for an editing operation with average complexity. If source code is available, we ran experiments ourselves and provide the concrete prover time on our machine. Otherwise, we refer to the authors' evaluation. For more details, we refer the reader to Section 7. Eva is not only the first cryptographic protocol for providing authenticity of lossily encoded video but also has additional advantages over related work.

As discussed above, the prover performance in PhotoProof is suboptimal. With a time complexity of $O(P^3 \log P)$, PhotoProof takes $\sim$ 18676 µs/px based on the authors' evaluation on a lower-spec machine compared to ours. In contrast, VIR demonstrates significant improvement in prover time due to its use of dedicated proof systems for specific editing operations, resulting in a per-pixel prover time of $\sim$ 16 µs and $O(P \log P)$ time complexity. While ZK-IMG also achieves performance gain over PhotoProof, the proof generation still takes a considerable amount of time, especially when proof of hash is involved ($>$ 355 µs, where $>$ is used because this was evaluated on a very high-performance server). Compared to ZK-IMG, VIMz further reduces prover time by $2 \sim 3$x, averaging around 167 µs/px with a linear complexity. Similarly, VerITAS offers fast prover performance at about 95 µs/px by optimizing the time for proof of hash, although its complexity is still $O(P \log P)$ due to polynomial interpolation. Eva, due to the combination of customized folding scheme, tailored circuit design, and various optimizations in our implementation, achieves optimal time complexity ($O(P)$) and the fastest prover time (~ 5 µs/px) among all the protocols.

Thanks to the macroblock-based structure of video encoding, Eva only needs to handle a fixed number of macroblocks in each incremental step of IVC, controlled by a constant batch size b. In comparison, all prior works [22–24, 26] require RAM proportional to the image size. This is because they either load the entire image into the arithmetic circuit or, in the case of VIR, use a structured reference string srs containing commitment keys for the entire image. While VIMz achieves lower memory costs through folding schemes, its memory usage remains linear in the image width N, because of its row-by-row proof generation process.

Regarding proof size, both [22, 23] generate proofs of constant size (2.67 KB and 223 B, respectively). This is also the case for Eva, which produces proofs of size 448 B. In contrast, the proofs in ZK-IMG [24] have a size of $O(\log P)$. This is due to ZK-IMG utilizing `Halo2` [42], which is based on the inner-product argument from [81] that generates proofs of size $O(\log n)$ in the number of constraints n. Here, n is linear in the number of pixels P, according to the circuit design described in [24, Section 7]. Meanwhile, both VIMz and VerITAS produce proofs of size $O(\log^2 n)$, because the former leverages Spartan [63] as the decider, and the latter is powered by `Plonky2` [82] with FRI [83] as the commitment scheme. For VIMz, n is linear in the image width N, while for VerITAS, n is proportional to P. Concretely, the proofs from ZK-IMG, VIMz, and VerITAS are at least 20 times larger than those of Eva.

Furthermore, due to its constant RAM consumption, Eva supports videos with unlimited resolution and frame count, while the other protocols cannot achieve infinite resolution, because their RAM usage scales with image dimensions, but the prover only has bounded RAM in practice.

Prior video authentication protocols. In video authentication, the prover generates authentication information for a claimed video, which can later be verified either publicly or privately. Existing work regarding video authentication follows two technique routes. The first is based on robust hash [14,15] (sometimes referred to as perceptual hash [15,84]), a digest extraction algorithm whose output is robust against benign transformations (e.g., resizing, (re-)encoding, cropping) but fragile to malicious manipulations (e.g., object replacement). After the robust hash is extracted, the prover generates authentication information by feeding the resulting hash value to, e.g., signing and watermarking.

However, it is challenging to define transformations that achieve the balance between robustness and fragility. Consider a toy example: if fragility is determined by the number of altered pixels, then minor but malicious edits (e.g., changing a number on a banknote) might pass as acceptable, while significant but benign edits (e.g., cropping to remove a person) might be rejected. In fact, these protocols experience non-negligible false positive or false negative rates [14–17] and active attackers can bypass some of these mechanisms [20, 21]. Additionally, having predefined legal transformations may not be practical, because whether a transformation is benign or malicious can be subjective and context-dependent.

The second is adopted by *Coalition for Content Provenance and Authenticity (C2PA)* [13], an industry standard for multimedia authentication based on digital signatures. C2PA requires that recording devices, such as mobile phones or cameras, have built-in signing keys certified by the device manufacturer. When multimedia content is recorded, the device generates a signature for both the content and its metadata, which may include thumbnails, capture date, location, etc. The multimedia content can later be edited by trusted editing software, which also has signing keys embedded. Analogously, the editing software signs the processed content, the metadata, and the editing operations performed. Upon publishing the processed content along with these signatures, the verifier (e.g., a news consumer) can check the provenance and authenticity of the content by verifying the associated signatures.

The trust model of C2PA assumes that both recording devices and editing software are trustworthy. However, the trust assumption regarding editing software is problematic in practice: while recording devices may utilize trusted execution environments (TEEs) or hardware security module (HSMs) to protect the signing keys, these mechanisms are not available for editing software. Consequently, attackers could potentially extract signing keys via reverse engineering, enabling them to generate valid signatures for malicious content.

Furthermore, C2PA may inadvertently leak sensitive information. For instance, during the editing process, the thumbnail of the original content might be signed by the editing software and published along with the processed content for verification. This may expose data that was not intended for disclosure, such as the faces of individuals that were blurred in the processed content, thereby raising privacy concerns.

Detection of fake videos. Another direction to fight misinformation is detecting fake videos. Human eyes are not always reliable in distinguishing real videos from fake ones, especially with the rise of deepfake technology [6–8]. Focusing on the detection of AI-generated videos, machine learning models have been developed [9–11], achieving promising results. However, the inherent characteristics of human eyes and neural networks inevitably produce false positives and false negatives with non-negligible probability. This is especially evident when active attackers manipulate videos to exploit some vulnerabilities in a specific detection method. For instance, it is demonstrated in [18,19] that several existing deepfake detectors can be bypassed by adversarial examples.

Appendix B Security of Our Nova-Based IVC

B.1 Security of NIFS

We provide the intuition to prove the security of our NIFS in terms of completeness, knowledge soundness, and zero-knowledge.

Essentially, our NIFS modifies Nova by splitting the vector of witnesses into $\boldsymbol{q}$ and $\boldsymbol{w}$. Thus, the core step in our proof is the conversion from the instance-witness pair $\mathbb{U}, \mathbb{W}$ in our NIFS to the one $\mathbb{U}', \mathbb{W}'$ in Nova, where $\mathbb{U}' := (\mathbb{U}.u, \mathbb{U}.\boldsymbol{x}, \mathbb{U}.\overline{Q} + \mathbb{U}.\overline{W}, \mathbb{U}.\overline{E})$, $\mathbb{W}' := (\mathbb{W}.\boldsymbol{q} \cup \mathbb{W}.\boldsymbol{w}, \mathbb{W}.\boldsymbol{e})$.

Proof of completeness. Given an adversary $\mathcal{A}$ who breaks the completeness of NIFS, we can construct an adversary $\mathcal{A}'$ who breaks the completeness of Nova. Whenever $\mathcal{A}$ outputs $R, (\mathbb{U}_1, \mathbb{W}_1), (\mathbb{U}_2, \mathbb{W}_2)$, $\mathcal{A}'$ constructs the corresponding instances and witnesses $(\mathbb{U}'_1, \mathbb{W}'_1), (\mathbb{U}'_2, \mathbb{W}'_2)$ for Nova. Then, $\mathcal{A}'$ outputs $R, (\mathbb{U}'_1, \mathbb{W}'_1), (\mathbb{U}'_2, \mathbb{W}'_2)$, thereby breaking the completeness of Nova. $\square$

Proof of knowledge soundness. With Nova's extractor Ext, we build an extractor

Ext' for NIFS. Provided $R, \mathbb{U}_1, \mathbb{U}_2, \mathbb{W}, \overline{T}$, which are the output of $\mathcal{A}$, Ext' converts $\mathbb{U}_1, \mathbb{U}_2, \mathbb{W}$ to Nova's instances and witness $\mathbb{U}'_1, \mathbb{U}'_2, \mathbb{W}'$, feeds $R, \mathbb{U}'_1, \mathbb{U}'_2, \mathbb{W}', \overline{T}$ to Ext, converts the returned $\mathbb{W}'_1, \mathbb{W}'_2$ back to $\mathbb{W}_1, \mathbb{W}_2$, and outputs $\mathbb{W}_1, \mathbb{W}_2$. □

Proof of zero-knowledge. Intuitively, NIFS is zero-knowledge because the commitment $\overline{T}$ in the transcript is hiding. Formally, the simulator Sim uniformly samples a random value r_t and computes $\overline{T} \leftarrow \mathsf{CM}.\mathcal{C}(\mathsf{ck}, r_t)$. $\overline{T}$ is indistinguishable from honestly generated commitments, because CM is a hiding commitment scheme. □

B.2 Security of IVC

Below we provide proofs of the security of IVC.

Proof of succinctness. Our IVC is succinct, because π_i only consists of two running instance-witness pairs and one incoming instance-witness pair, whose sizes are independent of the number of steps. □

Proof of completeness. To prove the completeness of IVC, we rely on the completeness of NIFS. We focus on the non-base case where $i > 0$, with the output of $\mathcal{A}$ consisting of $\mathcal{F}, i, \boldsymbol{z}_0, \boldsymbol{z}_i, \mathsf{aux}_i, \pi_i$.

Given that $\mathcal{V}(\mathsf{vk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i) = 1$, the following conditions hold: 1) $\mathbb{u}_i$ is a valid incoming instance with $\mathbb{u}_i.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_i, i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{H}(\mathbb{U}_i^{\mathsf{cf}}, i), \rho(\mathbb{u}_i.\overline{Q}))$, 2) $\mathbb{w}_i$ and $\mathbb{u}_i$ satisfy $\mathsf{CS}^{\mathsf{aug}}$, 3) $\mathbb{W}_i$ and $\mathbb{U}_i$ satisfy $\mathsf{CS}^{\mathsf{aug}}$, and 4) $\mathbb{W}_i^{\mathsf{cf}}$ and $\mathbb{U}_i^{\mathsf{cf}}$ satisfy $\mathsf{CS}^{\mathsf{cf}}$.

Our goal is to show that $\mathcal{V}(\mathsf{vk}, (i+1, \boldsymbol{z}_0, \boldsymbol{z}_{i+1}), \pi_{i+1}) = 1$. According to the construction of $\mathsf{IVC}.\mathcal{P}$, $\mathbb{U}_{i+1}$ and $\mathbb{W}_{i+1}$ are obtained by folding $\mathbb{u}_i, \mathbb{w}_i$ into $\mathbb{U}_i, \mathbb{W}_i$. Since NIFS is complete, $\mathbb{U}_{i+1}$ and $\mathbb{W}_{i+1}$ also satisfy $\mathsf{CS}^{\mathsf{aug}}$.

Moreover, when running $\mathcal{F}^{\mathsf{cf}}$ in $\mathsf{IVC}.\mathcal{P}$, the check in Circuit 2 passes due to the completeness of NIFS. Therefore, $\mathbb{u}_i^{\mathsf{cf}}$ and $\mathbb{w}_i^{\mathsf{cf}}$, constructed from the variables in $\mathcal{F}^{\mathsf{cf}}$, satisfy $\mathsf{CS}^{\mathsf{cf}}$. As a result, both $(\mathbb{u}_i^{\mathsf{cf}}, \mathbb{w}_i^{\mathsf{cf}})$ and $(\mathbb{U}_i^{\mathsf{cf}}, \mathbb{W}_i^{\mathsf{cf}})$ satisfy $\mathsf{CS}^{\mathsf{cf}}$. Again, by the completeness of NIFS, $\mathbb{U}_{i+1}^{\mathsf{cf}}$ and $\mathbb{W}_{i+1}^{\mathsf{cf}}$ also satisfy $\mathsf{CS}^{\mathsf{cf}}$.

To complete the proof, we demonstrate that $\mathbb{u}_{i+1}$ and $\mathbb{w}_{i+1}$ satisfy $\mathsf{CS}^{\mathsf{aug}}$ as well, where $\mathbb{u}_{i+1}$ is a valid incoming instance. Consider the checks in Circuit 3. First, Line 2 passes because $\mathbb{u}_i$ is known to be valid. In addition, according to the construction of $\mathbb{u}_i^{\mathsf{cf}}$, Line 6 is also satisfied. Line 7 checks the equation for set inclusion, which holds since the queries $\boldsymbol{\alpha}$ are supposed to be a subset of the lookup table $\boldsymbol{\tau}$. Moreover, Line 8 ensures that the statements h_1 and h_2 match the in-circuit variables calculated via $\mathsf{NIFS}.\mathcal{V}$ and H, which is guaranteed by the completeness of NIFS. Consequently, the vector of variables $\boldsymbol{z}$ is valid for $\mathcal{F}^{\mathsf{aug}}$, implying that $\mathbb{u}_{i+1}$ and $\mathbb{w}_{i+1}$ satisfy $\mathsf{CS}^{\mathsf{aug}}$. Since c is computed as $c := \rho(\mathbb{u}_{i+1}.\overline{Q})$, it follows that $\mathbb{u}_{i+1}.\boldsymbol{x} = (h_1, h_2, c)$ is well-formed. □

Proof of knowledge soundness. Now we prove the knowledge soundness of IVC by only considering the non-base case $i > 1$.

Similarly, for $\mathcal{A}$'s output $\mathcal{F}, i, \boldsymbol{z}_0, \boldsymbol{z}_i, \mathsf{aux}_i, \pi_i$, we have $\mathcal{V}(\mathsf{vk}, (i, \boldsymbol{z}_0, \boldsymbol{z}_i), \pi_i) = 1$, which again indicates 1) $\mathbb{u}_i$ is a valid incoming instance with $\mathbb{u}_i.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_i, i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{H}(\mathbb{U}_i^{\mathsf{cf}}, i), \rho(\mathbb{u}_i.\overline{Q}))$, 2) $\mathbb{w}_i$ and $\mathbb{u}_i$ satisfy $\mathsf{CS}^{\mathsf{aug}}$, 3) $\mathbb{W}_i$ and $\mathbb{U}_i$ satisfy $\mathsf{CS}^{\mathsf{aug}}$, and 4) $\mathbb{W}_i^{\mathsf{cf}}$ and $\mathbb{U}_i^{\mathsf{cf}}$ satisfy $\mathsf{CS}^{\mathsf{cf}}$.

With these conditions, Ext works as below:

1. Reconstruct $\boldsymbol{v}$ from $\mathbb{w}_i$ and $\mathbb{u}_i$ by computing $\boldsymbol{v} \coloneqq (\mathbb{u}_i.u, \mathbb{u}_i.\boldsymbol{x}, \mathbb{w}_i.\boldsymbol{q}, \mathbb{w}_i.\boldsymbol{w})$. Because $\mathbb{w}_i$ and $\mathbb{u}_i$ satisfy $\mathsf{CS}^{\mathsf{aug}}$, $\boldsymbol{v}$ is also a satisfying vector of variables for $\mathcal{F}^{\mathsf{aug}}$.

2. Obtain the witnesses $j, \boldsymbol{z}_j, \mathbb{U}_j, \mathbb{u}_j, \mathbb{U}^{\mathbb{G}}_{j+1}, \overline{T}_j, \mathbb{U}^{\mathsf{cf}}_j, \mathbb{u}^{\mathsf{cf}}_j, \overline{T}^{\mathsf{cf}}_j$ from $\boldsymbol{v}$.

3. By the checks in Line 8 of Circuit 3, we have $h_1 = \mathsf{H}(\mathbb{U}_{j+1}, j+1, \boldsymbol{z}_0, \boldsymbol{z}_{j+1})$, and $h_2 = \mathsf{H}(\mathbb{U}^{\mathsf{cf}}_{j+1}, j+1)$. Also, since h_1, h_2 are parts of $\mathbb{u}_i.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_i, i, \boldsymbol{z}_0, \boldsymbol{z}_i), \mathsf{H}(\mathbb{U}^{\mathsf{cf}}_i, i), \rho(\mathbb{u}_i.\overline{Q}))$ and H is collision-resistant, we can deduce that the two preimages are equal, *i.e.*,

   - $j + 1 = i$.
   - $\mathbb{U}^{\mathsf{cf}}_{j+1} = \mathbb{U}^{\mathsf{cf}}_i$, where $\mathbb{U}^{\mathsf{cf}}_{j+1} \coloneqq \mathsf{NIFS}.\mathcal{V}(\mathsf{vk}^{\mathsf{cf}}, \mathbb{U}^{\mathsf{cf}}_j, \mathbb{u}^{\mathsf{cf}}_j, \overline{T}^{\mathsf{cf}}_j)$. Consequently, except with negligible probability, Ext can invoke the extractor of NIFS on input $\mathsf{CS}^{\mathsf{cf}}, \mathbb{U}^{\mathsf{cf}}_j, \mathbb{u}^{\mathsf{cf}}_j, \mathbb{W}^{\mathsf{cf}}_i, \overline{T}^{\mathsf{cf}}_j$ and obtain $\mathbb{W}^{\mathsf{cf}}_j$ and $\mathbb{w}^{\mathsf{cf}}_j$ such that $(\mathbb{U}^{\mathsf{cf}}_j, \mathbb{W}^{\mathsf{cf}}_j)$ and $(\mathbb{u}^{\mathsf{cf}}_j, \mathbb{w}^{\mathsf{cf}}_j)$ satisfy $\mathsf{CS}^{\mathsf{cf}}$.

     Now, we reconstruct $\boldsymbol{v}^{\mathsf{cf}}$ from $\mathbb{w}^{\mathsf{cf}}_j$ and $\mathbb{u}^{\mathsf{cf}}_j$ analogously, such that $\boldsymbol{v}^{\mathsf{cf}}$ satisfies $\mathcal{F}^{\mathsf{cf}}$. Due to the checks in Line 5 of Circuit 3, the statements in $\boldsymbol{v}^{\mathsf{cf}}$ are $\mathbb{u}^{\mathsf{cf}}_j.\boldsymbol{x} = (r_j, \mathbb{U}^{\mathbb{G}}_j, \mathbb{u}^{\mathbb{G}}_j, \mathbb{U}^{\mathbb{G}}_{j+1}, \overline{T}_j)$. Combining this with the check in Circuit 2, we know that $\mathbb{U}^{\mathbb{G}}_{j+1} = \mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}(\mathsf{vk}, \mathbb{U}^{\mathbb{G}}_j, \mathbb{u}^{\mathbb{G}}_j, r_j, \overline{T}_j)$.
   - $\mathbb{U}_{j+1} = \mathbb{U}_i$. Note that $\mathbb{U}^{\mathbb{F}}_{j+1} = \mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}(\mathsf{vk}, \mathbb{U}^{\mathbb{F}}_j, \mathbb{u}^{\mathbb{F}}_j, r_j)$, and $\mathbb{U}^{\mathbb{G}}_{j+1} = \mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}(\mathsf{vk}, \mathbb{U}^{\mathbb{G}}_j, \mathbb{u}^{\mathbb{G}}_j, r_j, \overline{T}_j)$. Hence, $\mathbb{U}_i = \mathsf{NIFS}.\mathcal{V}(\mathsf{vk}, \mathbb{U}_j, \mathbb{u}_j, \overline{T}_j)$. Consequently, except with negligible probability, Ext can invoke the extractor of NIFS on input $\mathsf{CS}^{\mathsf{aug}}, \mathbb{U}_j, \mathbb{u}_j, \mathbb{W}_i, \overline{T}_j$ and obtain $\mathbb{W}_j$ and $\mathbb{w}_j$ such that $(\mathbb{U}_j, \mathbb{W}_j)$ and $(\mathbb{u}_j, \mathbb{w}_j)$ satisfy $\mathsf{CS}^{\mathsf{aug}}$.
   - $\boldsymbol{z}_{j+1} = \boldsymbol{z}_i$. This implies that $\mathcal{F}(\boldsymbol{z}_j, \mathsf{aux}_j) = \boldsymbol{z}_i$.

4. By the checks in Line 2 of Circuit 3, we know that $\mathbb{u}_j$ is an incoming instance, and that $\mathbb{u}_j.\boldsymbol{x} = (\mathsf{H}(\mathbb{U}_j, j, \boldsymbol{z}_0, \boldsymbol{z}_j), \mathsf{H}(\mathbb{U}^{\mathsf{cf}}_j, j), \rho(\mathbb{u}_j.\overline{Q}))$.

5. By the checks in Line 7 of Circuit 3, LogUp's identity for set inclusion holds, given a uniform challenge $c \coloneqq \rho(\mathbb{u}_i.\overline{Q})$. Therefore, the queries are in the lookup table, *i.e.*, $\boldsymbol{\alpha} \subseteq \boldsymbol{\tau}$.

6. Finally, Ext computes $\pi_j \coloneqq ((\mathbb{U}_j, \mathbb{W}_j), (\mathbb{u}_j, \mathbb{w}_j), (\mathbb{U}^{\mathsf{cf}}_j, \mathbb{W}^{\mathsf{cf}}_j))$ and outputs $\boldsymbol{z}_j, \mathsf{aux}_j, \pi_j$ as well as $\boldsymbol{\alpha}$.

We can observe from the analysis above that the outputs of Ext satisfy the checks in IVC's knowledge soundness definition and the lookup relation R^{lookup}, thereby concluding the proof.

□

B.3 Security of Decider

Due to the similar design behind our and Nova's deciders, we refer the reader to [28, Appendix D] for the proof that Decider is a zkSNARK for R^{IVC}, which satisfies completeness, knowledge soundness, zero-knowledge and succinctness. In a nutshell, the security of Decider is powered by the the corresponding properties of ZKCP and NIFS. Plus, Decider is succinct because both the LegoGro16 proof ϖ and the Pedersen commitments in committed instances are of constant size.

Appendix C Gadgets for Integer Operations

We review the gadgets in [64] for the computation of sign and absolute value as well as the right shifting operation.

Sign and absolute value. We cannot directly compute the sign and the absolute value of a variable x in an arithmetic circuit over $\mathbb{F}_p$. Intuitively, a number is positive if it is greater than 0, and is negative otherwise. However, as the field $\mathbb{F}_p$ is not *ordered*, we cannot *compare* between its elements. Therefore, we manually define elements in the set $\{1, 2, \ldots, (p-1)/2\}$ as positive, and those in the set $\{(p+1)/2, \ldots, p-2, p-1\}$ as negative.

Before explaining the computation of sign and absolute value under this definition, we introduce $\mathcal{F}^{\mathsf{EnforceBitLen}}$, a gadget for ensuring the bit length of a variable x is at most W, i.e., $x \in [0, 2^W - 1]$. Powered by lookup arguments, $\mathcal{F}^{\mathsf{EnforceBitLen}}$ is the key to efficiency of the in-circuit operations in quantization. As depicted in Gadget 19, on input a variable x and a constant W, the gadget first asks the prover to provide $\{x_i\}_{i=0}^{W/\log\nu-1}$, the limbs of x in base-ν (recall that ν is the size of the lookup table). Here, we assume ν is a power of 2. Then, the gadget enforces x indeed decomposes into these limbs by comparing their concatenation, expressed as $\sum_{i=0}^{W/\log\nu-1} 2^{i\log\nu}x_i$, with x. Finally, the limbs of x are appended to $\boldsymbol{\alpha}$, the list of queries, to make sure every limb is in base-ν. We reemphasize that lookup argument is critical to the performance of $\mathcal{F}^{\mathsf{EnforceBitLen}}$: without the lookup table, circuits in R1CS can only handle the range check bit-by-bit (instead of limb-by-limb), which is done by enforcing $x_i(1-x_i) = 0$ for each claimed bit x_i. Consequently, $\mathcal{F}^{\mathsf{EnforceBitLen}}$ would cost $W+1$ constraints, which is much more expensive than the $W/\log\nu + 1$ constraints with the lookup table.

Gadget 19: $\mathcal{F}^{\mathsf{EnforceBitLen}}(x, W)$
1 $\{x_i\}_{i=0}^{W/\log\nu-1} \leftarrow \mathsf{Hint}(x)$
2 **enforce** $\sum_{i=0}^{W/\log\nu-1} 2^{i\log\nu}x_i = x$
3 $\boldsymbol{\alpha} \coloneqq \boldsymbol{\alpha} \cup \{x_i\}_{i=0}^{W/\log\nu-1}$

Now, as long as the upper bound of x's absolute value satisfies $x < 2^W < (p-1)/2$, we can extract the sign and the absolute value of x using $\mathcal{F}^{\mathsf{SignAbs}}$, as depicted in Gadget 20. The gadget first asks the prover to determine if x is positive. The prover checks which set x belongs to, and provides s as a hint. The gadget enforces that s is boolean, and computes x's absolute value $y \coloneqq s \,?\, x : -x$. Finally, the gadget enforces that y has at most W bits by invoking $\mathcal{F}^{\mathsf{EnforceBitLen}}(y, W)$ and returns s and y. Soundness holds because if an adversary feeds the incorrect s to the gadget, then y's value belongs to the negative set and is hence greater than $(p-1)/2$, but $\mathcal{F}^{\mathsf{EnforceBitLen}}$ guarantees that $0 \le y < 2^W < (p-1)/2$.

Right shift. It is also non-trivial to implement the gadget $\mathcal{F}^{\gg}(x, \delta)$ for shifting x to the right by δ bits. Here, we assume that $x \in [0, 2^W - 1]$, $\delta \in [U, V]$, and $2^{W+V-U} < p$. Intuitively, we could treat the right shift operation as integer division, i.e., $x \gg \delta = x/2^\delta$. The prover computes the quotient q and the remainder r such that $x = q \cdot 2^\delta + r$, and feeds q, r as hints to the gadget. Then the gadget enforces $x = q \cdot 2^\delta + r$. In addition, it is also required to check that $q \in [0, 2^{W-U} - 1]$, $r \in [0, 2^\delta - 1]$ to ensure q and r are well-formed. Here, since δ is not a

Gadget 20: $\mathcal{F}^{\mathsf{SignAbs}}(x \in [-2^W+1, 2^W-1])$

$s \leftarrow \mathsf{Hint}(x)$
$y := s \,?\, x : -x$
enforce $s(1-s) = 0$
$\mathcal{F}^{\mathsf{EnforceBitLen}}(y, W)$
return s, y

constant, it requires two $\mathcal{F}^{\mathsf{EnforceBitLen}}$ calls to enforce r's range, one for checking $r \in [0, 2^V-1]$ and another for checking $2^\delta - 1 - r \in [0, 2^V-1]$, introducing $2V/\log\nu$ queries to the lookup table. We are convinced that $r \in [0, 2^\delta - 1]$ only when both conditions are satisfied.

However, it is possible to eliminate one $\mathcal{F}^{\mathsf{EnforceBitLen}}$ call. As presented in Gadget 21, $\mathcal{F}^{\gg}$ first computes $x' := x \ll (V-\delta) = x \cdot 2^{V-\delta}$. Since $\delta \in [U, V]$, we have $V - \delta \in [0, V-U]$, and thus $x \cdot 2^{V-\delta} < 2^{W+V-U} < (p-1)/2$ does not overflow. Then we handle $x' \gg V$ analogously: the prover provides the quotient q and the remainder r for $x'/2^V$ as hints, and the gadget checks if $q \in [0, 2^{W-U}-1]$, $r \in [0, 2^V-1]$, and $x' = q \cdot 2^V + r$. This optimized approach only adds $V/\log\nu$ queries to the lookup table for checking r, thereby saving $V/\log\nu$ constraints compared to the naive construction.

Gadget 21: $\mathcal{F}^{\gg}(x \in [0, 2^W-1], \delta \in [U, V])$

$x' := x \cdot 2^{V-\delta}$
$q, r \leftarrow \mathsf{Hint}(x')$
enforce $x' = q \cdot 2^V + r$
$\mathcal{F}^{\mathsf{EnforceBitLen}}(q, W-U)$
$\mathcal{F}^{\mathsf{EnforceBitLen}}(r, V)$

Appendix D Security of Eva

Below we prove Theorem 1 by showing that Eva satisfies *succinctness*, *completeness*, *knowledge soundness*, and *zero-knowledge*.

Proof of succinctness. Eva satisfy succinctness because its proofs are of constant size. Specifically, the LegoGro16 proof ϖ has 4 $\mathbb{G}$ elements and 1 $\hat{\mathbb{G}}$ element, the partial running instance $\mathbb{U}'_k$ has 3 $\mathbb{G}$ elements and 1 $\mathbb{F}_p$ element, the partial incoming instance $\mathbb{u}'_k$ has 2 $\mathbb{G}$ elements, $\overline{T}$ is in $\mathbb{G}$, and r is in $\mathbb{F}_p$. In total, the proof π consists of 10 $\mathbb{G}$ elements, 1 $\hat{\mathbb{G}}$ element, and 2 $\mathbb{F}_p$ elements. □

Proof of completeness. We omit the proof of completeness for Eva, as it is straightforward to see from the design of our circuits and the completeness of IVC, NIFS, and ZKCP. □

Proof of knowledge soundness. We prove the knowledge soundness of Eva by constructing an efficient extractor Ext. Given public parameters $\mathsf{pk}_\Pi, \mathsf{vk}_\Pi, \mathsf{vk}_\Sigma$, the trapdoor td, and $\mathcal{A}$'s output $(\zeta, \mathsf{meta}, \mathsf{param}, \pi)$, we have $\mathcal{V}(\mathsf{vk}_\Pi, \mathsf{vk}_\Sigma, \zeta, \mathsf{meta}, \mathsf{param}, \pi) = 1$ by condition. Hence, $\mathsf{ZKCP}.\mathcal{V}(\mathsf{vk}, \boldsymbol{x}, \boldsymbol{c}, \varpi) = 1$, for $\boldsymbol{x} := (\mathsf{vk}_\Sigma, \mathsf{meta}, k, \boldsymbol{z}_0, \hbar_k, r, \mathbb{u}'_k, \mathbb{U}'_k, \overline{T}), \boldsymbol{c} := (\mathbb{U}^{\mathbb{G}}_{k+1})$. With this condition, Ext works as below:

1. Invoke the extractor of ZKCP on input $\boldsymbol{x}, \boldsymbol{c}, \varpi$. Except with negligible probability, Ext can obtain $\boldsymbol{v} := (\mathbb{W}_{k+1}), \boldsymbol{\omega} := (h_k, \sigma, \mathbb{U}_k.\boldsymbol{x}, \mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}})$, such that $(\boldsymbol{x}, \boldsymbol{c})$ and $(\boldsymbol{v}, \boldsymbol{\omega})$ satisfy $\mathcal{F}^{\mathsf{Decider}_{\mathsf{Eva}}}$, and $\boldsymbol{v} = \mathbb{W}_{k+1}$ opens $\boldsymbol{c} = \mathbb{U}_{k+1}^{\mathbb{G}}$.

2. Reconstruct $\mathbb{U}_k$ from $\mathbb{U}'_k$ and $\mathbb{U}_k.\boldsymbol{x}$.

3. Reconstruct $\mathbb{u}_k$ from $\mathbb{u}'_k$ and $\mathbb{u}_k.\boldsymbol{x} := (\mathsf{H}(\mathbb{U}_k, k, \boldsymbol{z}_0, (h_k, \hbar_k)), \mathsf{H}(\mathbb{U}_k^{\mathsf{cf}}, k), \rho(\mathbb{u}_k.\overline{Q}))$.

4. Line 4 of Circuit 17 enforces that $\mathbb{U}_{k+1}^{\mathbb{F}} := \mathsf{NIFS}.\mathcal{V}^{\mathbb{F}}(\mathsf{vk}, \mathbb{U}_k^{\mathbb{F}}, \mathbb{u}_k^{\mathbb{F}}, r)$. Also, we have $\mathbb{U}_{k+1}^{\mathbb{G}} := \mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}(\mathsf{vk}_{\Phi}, \mathbb{U}_k^{\mathbb{G}}, \mathbb{u}_k^{\mathbb{G}}, r, \overline{T})$. Thus, $\mathbb{U}_{k+1} := \mathsf{NIFS}.\mathcal{V}(\mathsf{vk}_{\Phi}, \mathbb{U}_k, \mathbb{u}_k, \overline{T})$.

   Moreover, Line 5 of Circuit 17 and the commit-and-prove relation w.r.t. $\boldsymbol{v} = \mathbb{W}_{k+1}$ and $\boldsymbol{c} = \mathbb{U}_{k+1}^{\mathbb{G}}$ imply that $\mathbb{W}_{k+1}$ and $\mathbb{U}_{k+1}$ satisfy $\mathsf{CS}^{\mathsf{aug}}$.

   Consequently, except with negligible probability, Ext can invoke the extractor of NIFS on input $\mathbb{U}_k, \mathbb{u}_k, \mathbb{W}_{k+1}, \overline{T}$ and obtain $\mathbb{W}_k, \mathbb{w}_k$ such that $(\mathbb{U}_k, \mathbb{W}_k)$ and $(\mathbb{u}_k, \mathbb{w}_k)$ satisfy $\mathsf{CS}^{\mathsf{aug}}$.

5. By the checks in Line 6 of Circuit 17, we can deduce that $\mathbb{U}_k^{\mathsf{cf}}$ and $\mathbb{W}_k^{\mathsf{cf}}$ satisfy $\mathsf{CS}^{\mathsf{cf}}$. At this point, Ext can recover $\pi_k := ((\mathbb{U}_k, \mathbb{W}_k), (\mathbb{u}_k, \mathbb{w}_k), (\mathbb{U}_k^{\mathsf{cf}}, \mathbb{W}_k^{\mathsf{cf}}))$ such that all checks in $\mathsf{IVC}.\mathcal{V}(\mathsf{vk}, (k, \boldsymbol{z}_0, \boldsymbol{z}_k), \pi_k) = 1$ pass.

   Consequently, except with negligible probability, Ext can invoke the extractor of IVC on input $\mathcal{F}^{\mathsf{aug}}, k, \boldsymbol{z}_0, \boldsymbol{z}_k, \pi_k$ and obtain the state and proof at $k-1$-th step.

6. Repeatedly invoke the extractor of IVC on the last state and proof, and obtain the previous state and proof, until reaching the initial step.

7. Parse the original video $\boldsymbol{V}$ from all the auxiliary states $\{\mathsf{aux}_i\}$ and return $\sigma, \boldsymbol{V}$.

By the satisfiability of $\mathcal{F}^{\mathsf{aug}}$, we can conclude that, with param, $\{\boldsymbol{Z}_i\}$ is the correct encoding of an video $\boldsymbol{V}'$ edited from the original video $\boldsymbol{V}$ whose digest is h_k. Also, by construction of $\mathsf{ZKCP}.\mathcal{V}$, ζ is the entropy coded bitstream of $\{\boldsymbol{Z}_i\}$. Furthermore, by Line 1 of Circuit 17, σ is a valid signature on $\mathsf{H}(h_k, \mathsf{meta})$. Thus, Ext successfully extracts $\boldsymbol{V}$ and σ such that $R^{\mathsf{VA}}((\zeta, \mathsf{meta}, \mathsf{param}, \mathsf{vk}_{\Sigma}), (\sigma, \boldsymbol{V})) = 1$, except with negligible probability, thereby completing the proof. □

Proof of zero-knowledge. For zero-knowledge, we leverage the technique in [28, Appendix D] for constructing a simulator Sim who can produce $\mathbb{U}_k, \mathbb{u}_k$ that are indistinguishable from the outputs of the honest prover, if H and CM are hiding.

First, Sim uniformly samples several random values r_1, r_2, r_q, r_w, and initiates $\mathbb{U}_1, \mathbb{u}_1$, where $\mathbb{U}_1 = \mathbb{U}_{\perp}$, $\mathbb{u}_1.u = 1$, $\mathbb{u}_1.\overline{Q} = \mathsf{CM}.\mathcal{C}(\mathsf{ck}, r_q)$, $\mathbb{u}_1.\overline{W} = \mathsf{CM}.\mathcal{C}(\mathsf{ck}, r_w)$, $\mathbb{u}_1.\overline{E} = \overline{0}$, $\mathbb{u}_1.\boldsymbol{x} = (\mathsf{H}(r_1), \mathsf{H}(r_2), \rho(\mathbb{u}_1.\overline{Q}))$. Here, $\mathbb{U}_1$, $\mathbb{u}_1.u$, and $\mathbb{u}_1.\overline{E}$ are equal to real ones. Also, since we assume H and CM are hiding, $\mathbb{u}_1.\overline{Q}$, $\mathbb{u}_i.\overline{W}$, and $\mathbb{u}_1.\boldsymbol{x}$ are indistinguishable from real ones.

Then, we show that for every i, Sim can generate $\mathbb{U}_{i+1}$ and $\mathbb{u}_{i+1}$ that are indistinguishable from real ones, given that $\mathbb{U}_i$ and $\mathbb{u}_i$ are indistinguishable. To this end, Sim uniformly samples randomness r_1, r_2, r_q, r_w, r_t and computes $\overline{T} := \mathsf{CM}.\mathcal{C}(\mathsf{ck}, r_t)$, which is indistinguishable from real commitments due to the hiding property of CM. With $\overline{T}$, $\mathbb{U}_{i+1}$ is computed by $\mathbb{U}_{i+1} := \mathsf{NIFS}.\mathcal{V}(\mathsf{vk}_\Phi, \mathbb{U}_i, \mathbb{u}_i, \overline{T})$. Further, $\mathbb{u}_{i+1}$ is computed in the same way as the base case. In this way, both $\mathbb{U}_{i+1}$ and $\mathbb{u}_{i+1}$ are indistinguishable from real instances.

After k steps, $\mathbb{U}_k, \mathbb{u}_k$ are indistinguishable from the honestly generated ones. Again, Sim computes $\overline{T} := \mathsf{CM}.\mathcal{C}(\mathsf{ck}, r_t)$ for a random r_t and $r := \rho(\mathbb{U}_k, \mathbb{u}_k, \overline{T})$, which are indistinguishable from real ones.

Next, Sim computes $\hbar_k$ by hashing the prediction macroblocks and quantized coefficients decoded from ζ, and derives $\mathbb{U}^{\mathbb{G}}_{k+1}$ by running $\mathbb{U}^{\mathbb{G}}_{k+1} := \mathsf{NIFS}.\mathcal{V}^{\mathbb{G}}(\mathsf{vk}_\Phi, \mathbb{U}^{\mathbb{G}}_k, \mathbb{u}^{\mathbb{G}}_k, r, \overline{T})$.

Finally, Sim invokes the ZKCP simulator on input $\boldsymbol{x}$ and $\boldsymbol{c}$, where $\boldsymbol{x} := (\mathsf{vk}_\Sigma, \mathsf{meta}, k, \boldsymbol{z}_0, \hbar_k, r, \mathbb{u}'_k, \mathbb{U}'_k, \overline{T}), \boldsymbol{c} := (\mathbb{U}^{\mathbb{G}}_{k+1})$. The ZKCP simulator returns a simulated proof ϖ that is indistinguishable from the honestly generated ones, and the proof $\pi := (\varpi, \mathbb{U}'_k, \mathbb{u}'_k, \overline{T}, r)$ that Sim returns is therefore also indistinguishable from the honest proofs. □

Discussion on security. For a raw video $\boldsymbol{V}$ signed by the recorder and an encoded video stream ζ, our current security model guarantees that $\zeta = \mathcal{E}(\Delta(\boldsymbol{V}), \mathsf{param}^{\mathcal{E}})$, where Δ is the editing operation, and $\mathsf{param}^{\mathcal{E}}$ is the encoding parameters. Below we discuss two potential issues with our current security model and possible solutions.

First, our model does not ensure $\mathsf{param}^{\mathsf{Pred}}$ to be the best prediction parameters for encoding $\boldsymbol{V}' := \Delta(\boldsymbol{V})$. Recall that, in order to reduce the circuit size, the prediction process is removed from our $\mathcal{F}^{\mathcal{E}}$, and thus the choice of prediction parameters $\mathsf{param}^{\mathsf{Pred}}$ is not enforced.

Consequently, for an original macroblock $\boldsymbol{X} \in \boldsymbol{V}$ and two editing operations Δ and $\widehat{\Delta}$, a malicious prover $\mathcal{A}$ may produce $\boldsymbol{P}$ and $\boldsymbol{Z}$ by encoding $\boldsymbol{X}' := \Delta(\boldsymbol{X})$, but later prove that $\boldsymbol{Z}$ is encoded from $\widehat{\boldsymbol{X}}' := \widehat{\Delta}(\boldsymbol{X})$. This is possible if $\mathcal{A}$ can find some prediction parameters $\widehat{\mathsf{param}}^{\mathsf{Pred}}$ such that the prediction macroblock for $\widehat{\boldsymbol{X}}'$ becomes $\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{X}}' - \boldsymbol{X}' + \boldsymbol{P}$. By feeding $\widehat{\boldsymbol{P}}$ to the circuit, the residual macroblock is now computed as $\widehat{\boldsymbol{X}}' - \widehat{\boldsymbol{P}} = \boldsymbol{X}' - \boldsymbol{P} = \boldsymbol{R}$, and the final output becomes $\boldsymbol{Z}$. In this way, $\mathcal{A}$'s claimed editing operation is $\widetilde{\Delta} = \widehat{\Delta}$, and the claimed encoding parameters are $\widetilde{\mathsf{param}}^{\mathsf{Pred}} = \widehat{\mathsf{param}}^{\mathsf{Pred}}$, but $\boldsymbol{Z}$ is actually the encoding of $\Delta(\boldsymbol{X})$ under $\mathsf{param}^{\mathsf{Pred}}$.

Our model allows for such an "attack", because both $\Delta, \mathsf{param}^{\mathsf{Pred}}$ and $\widehat{\Delta}, \widehat{\mathsf{param}}^{\mathsf{Pred}}$ are valid configurations for encoding $\boldsymbol{X}$ as $\boldsymbol{Z}$. A stronger security model might require the claimed editing operation $\widetilde{\Delta}$ to closely resemble the actual one Δ (as we will discuss soon, it is impossible to guarantee exactly equality between $\widetilde{\Delta}$ and Δ).

To achieve security under this enhanced model, we propose the following approaches:

- If the video codec generates similar predictions for different prediction parameters, or if we can restrict the prover to use prediction parameters with similar effects, then $\mathcal{A}$ can no longer find a prediction macroblock that balances the large difference between the claimed $\widetilde{\Delta}$ and the actual Δ. This would ensure that $\widetilde{\Delta} \approx \Delta$.

- If the above conditions are not met, it is still possible to mitigate this issue. Observe that while lossy encoding reduces the video quality, the encoded video ζ is still similar to the original $\boldsymbol{V}'$. Thus, with the correct prediction parameters $\widetilde{\mathsf{param}}^{\mathsf{Pred}} = \mathsf{param}^{\mathsf{Pred}}$ for $\boldsymbol{V}'$, encoding ζ under $\mathsf{param}^{\mathsf{Pred}}$ again will produce a video that resembles ζ. On the other hand, with cheating prediction parameters $\widetilde{\mathsf{param}}^{\mathsf{Pred}} \not\approx \mathsf{param}^{\mathsf{Pred}}$, the re-encoded video will significantly differ from ζ.

Therefore, to detect if $\mathcal{A}$ is cheating, the verifier can re-encode ζ with the claimed prediction parameters and inspect if the re-encoded video is close to the encoded video. By rejecting significantly different videos, $\mathcal{V}$ can ensure that the claimed prediction parameters satisfy $\widetilde{\mathsf{param}}^{\mathsf{Pred}} \approx \mathsf{param}^{\mathsf{Pred}}$, thereby guaranteeing $\widetilde{\Delta} \approx \Delta$.

- The most robust solution is to extend Eva and generate proofs of best encoding. Intuitively, one can achieve this by incorporating the prediction process into the circuit. However, designing a more efficient approach remains an open problem.

The second issue is that the quantization parameter qp may enable a malicious prover $\mathcal{A}$ to find two editing operations Δ and $\widetilde{\Delta}$ that produce the same video stream ζ after quantization.

We regard the design of a stronger security model that guarantees the strict equality between Δ and $\widetilde{\Delta}$ as out of scope. The reason is that, due to the inherent loss of information in lossy encoding, it is always possible for $\mathcal{A}$ to find two modifications on the original video that have the same encoded stream, even qp is fixed to some small values. In some scenarios such as the detection of fake news, the verifier can simply reject low-quality videos with large qp (e.g., $\mathsf{qp} \geq 50$), as we expect videos published by news agencies to have more reasonable qp values (e.g., $\mathsf{qp} \approx 30$).