Efficiency of SIDH-based signatures (yes, SIDH)

Wissam Ghantous¹, Federico Pintore² and Mattia Veroni³

¹ Mathematical Institute, University of Oxford, UK, wissam.ghantous@maths.ox.ac.uk
² Department of Mathematics, University of Bari, IT, federico.pintore@uniba.it
³ NTNU - Norwegian University of Science and Technology, Trondheim, NO, mattia.veroni@ntnu.no

Abstract. In this note we assess the efficiency of a SIDH-based digital signature built on a diminished variant of a recent identification protocol proposed by Basso et al. Despite the devastating attacks against (the mathematical problem underlying) SIDH, this identification protocol remains secure, as its security is backed by a different (and more standard) isogeny-finding problem. We conduct our analysis by applying some known cryptographic techniques to decrease the signature size by about 70% for all parameter sets (obtaining signatures of approximately 21 KB for SIKEp434). Moreover, we propose a minor optimisation to compute many isogenies in parallel from the same starting curve. Our assessment confirms that the problem of designing a practical isogeny-based signature scheme remains largely open. However, concretely determine the current state of the art which future optimisations can compare to appears to be of relevance for a problem which has witnessed only small steps towards a solution.

Keywords: Post-quantum Cryptography · Isogeny-based Cryptography · Digital Signature

1 Introduction

Isogenies between supersingular elliptic curves have been used to construct cryptosystems supposed to be secure even in the presence of quantum attackers. The family of such cryptosystems is named isogeny-based cryptography, and its most appealing members enjoy short keys and ciphertexts. At the time of writing, the most prominent example of this attractive feature is the digital signature SQISign [DFKL⁺19], which is the most compact post-quantum signature scheme. On the other hand, isogeny-based cryptosystems incur high execution times, with SQISign making no exception (despite the recent improvements in [DFLW22]). The most promising results in terms of computational efficiency have been obtained for the key-exchange SIDH [DFJP14] and the corresponding key-encapsulation mechanism SIKE [JAC⁺17]. However, not all schemes built on SIDH share the same quality. An example are the SIDH-based digital signatures proposed in [GPS17,YAJ⁺17], for which no substantial amelioration has
appeared since their publication. Nevertheless, they represented an alternative starting point for a practical isogeny-based digital signature building on existing schemes. However, three major classical attacks \cite{CD22a,MM22,Rob22} were devised in 2022, which make SIDH, SIKE and most of the SIDH-based cryptosystems — including the signature schemes based on SIDH mentioned above — completely insecure. As a consequence, SQISign and Sea-Sign/CSI-FiSh\textsuperscript{4} were, until recently, the only isogeny-based digital signature schemes still secure.

Fortunately, two isogeny-based $\Sigma$-protocols that were recently proposed have restored the family of SIDH-based digital signatures (and non-interactive zero-knowledge proofs). The first one \cite[Sec. 5.3]{DDGZ21} — denoted by $\Sigma_{\text{wSIDH}}$ in the following — was originally designed for the SIDH setting, while for the second one \cite[Sec. 4]{BCC+22} — which we denote by $\Sigma_{\text{base SECUER}}$ — the SIDH parameters are (probably the most) favorable in terms of practical efficiency, despite it being designed for a general scenario. Consequently, their implementations can take advantage of the optimised implementations for determining and evaluating isogenies in the SIDH configuration. Even so, both $\Sigma_{\text{wSIDH}}$ and $\Sigma_{\text{base SECUER}}$ are not affected by the attacks in \cite{CD22a,MM22,Rob22} and hence can still be the base for constructing digital signature schemes as well as non-interactive zero-knowledge proofs (NIZKPs in short).

\textbf{Our contribution.} In this note we assess the compactness and efficiency that can be currently reached by digital signatures (and NIZKPs) based on the SIDH setting. In doing so, we restrict our attention to a digital signature built on a slightly weaker variant of $\Sigma_{\text{base SECUER}}$. We talk about weaker variant because $\Sigma_{\text{base SECUER}}$ was designed to satisfy statistical honest-verifier zero-knowledge, which is not necessary for our case study. The variant we consider — denoted by $\Sigma_{\text{SEC}}$ — only achieves computational honest-verifier zero-knowledge, but it allows for shorter isogenies, therefore a better efficiency. The main reasons to work with $\Sigma_{\text{SEC}}$ instead of other SIDH-based $\Sigma$-protocols are three. First of all, a similar assessment focused on $\Sigma_{\text{wSIDH}}$ was recently conducted in \cite{CD22b}. Even more importantly, despite the similarities between $\Sigma_{\text{wSIDH}}$ and $\Sigma_{\text{SEC}}$, the latter has a more lightweight design, leading to smaller transcripts and faster execution times. Last but not least, the optimisations we apply to ($t$ parallel executions of) $\Sigma_{\text{SEC}}$ are applicable also to $\Sigma_{\text{base SECUER}}$ and are relevant to any application of these Zero-Knowledge Proof systems. In fact, using the Fiat-Shamir transform to remove interactivity from a $\Sigma$-protocol has applications beyond digital signatures. For example, $\Sigma_{\text{base SECUER}}$ has been used to prove random generation of supersingular curves of unknown endomorphism rings in a distributed and trusted manner \cite{BCC+22}.

We conduct our analysis by applying some known signature-shortening techniques. By doing so, we can shorten the signatures produced by means of $\Sigma_{\text{SEC}}$ by

\textsuperscript{4} Sea-Sign \cite{DG19} is an isogeny-based digital signature scheme which works with isogenies and elliptic curves over prime fields. CSI-FiSh \cite{BKV19} is an optimisation of Sea-Sign for a specific set of parameters, named CSIDH-512. It is worth noticing that the security provided by CSIDH-512 is still an active area of research \cite{CSCJR22}, and instantiating SeaSign with larger parameters leads to long execution times.
Efficiency of SIDH-based signatures (yes, SIDH) 3

approximately 69%. For example, we obtain signatures of approximately 21 KB for the parameter set SIKEp434. In addition, we propose minor optimisations to compute many isogenies in parallel from the same starting curve.

One of the techniques we consider to shorten the signatures is the unbalanced challenge space technique, firstly proposed in [BKP20] for a Σ-protocol obtained by running parallel executions of a base Σ-protocol with soundness error 1/2. In this work, however, we apply it to a base Σ-protocol with soundness error 2/3, which requires a non-trivial generalisation of the original proposal. In fact, in order to determine the number of parallel executions which are required in such case for an unbalanced challenge space, we deduce some combinatorial results. Our findings can be readily applied to every possible soundness error of the base Σ-protocol, and therefore are of independent interest.

Our assessment confirms that the problem of designing a practical isogeny-based signature scheme remains largely open. Nonetheless, the proposed optimisations can be applied to the distributed trusted-setup protocol [BCC+22, Sec. 5] built on top of ΣbaseSECUER to collaboratively produce a random supersingular elliptic curve whose endomorphism ring is hard to compute even for the involved parties. Moreover, concretely determine the current state of the art of isogeny-based signatures which future optimisations can compare to appears to be of relevance for a problem which has witnessed only small steps towards a solution.

Related Work. The SIDH-based digital signature scheme proposed in [GPS17] produces signatures of approximately 12 KB when targeting 128 bits of classical security. For the same security target, the signature scheme in [CDMP22] (deduced from a different SIDH-based Σ-protocol proposed in [DDGZ21, Sec. 6]) outputs signatures of approximately 61 KB. Both these signature schemes are no longer secure after the cryptanalytic attacks against SIDH. The analysis conducted in [CD22b] on a still-secure digital signature built on top of ΣwSIDH achieves signatures of size approximately 74 KB for the parameter set SIKEp434. Note that the protocol in [GPS17] and ΣwSIDH in [DDGZ21, Sec. 5] are 2-special sound. The protocols in [DDGZ21, Sec. 6] is instead 3-special sound.

Roadmap. The paper is organised as follows. In Section 2 we recall some cryptographic preliminaries and we provide a description of the Σ-protocol ΣSEC, which is a diminished version of ΣbaseSECUER from [BCC+22, Sec. 4]. By applying the Fiat-Shamir transform [FS86] on ΣSEC, an SIDH-based signature scheme DSSEC is obtained. In Section 3 we apply some optimisation techniques to reduce the size of the signatures produced by DSSEC: commitment recoverability (Section 3.1), response compression (Section 3.2) and seed trees (Section 3.3). In (Section 3.4), unbalanced challenge spaces are also taken into account. We conclude the section highlighting the overall gain in applying these optimisations with respect to the original scheme. In Section 4 we suggest two optimisations for the computation of several isogenies of the same degree from the same starting curve. They consists in a pre-computation of repeated initial steps (Section 4.1) and in the parallelisation of kernel generators computation (Section 4.2). Section 5 contains some closing remarks.
2 Preliminaries

In this section we list some definitions and results regarding isogenies between supersingular elliptic curves, \(\Sigma\)-protocols and digital signatures. We then detail the \(\Sigma\)-protocol \(\Sigma_{\text{SEC}}\), a diminished variant of \(\Sigma_{\text{SECURE}}\) from [BCC+22, Sec. 4].

Remark 1. In the following, commitment schemes (\(C\)) and pseudorandom number generators (\(\text{Expand}\)) are instantiated with a hash function modeled as a random oracle \(O\). We always assume the input domain of the random oracle is appropriately separated when instantiating different cryptographic primitives. With an abuse of notation, we will write \(O(\text{Expand}||\cdot)\) instead of \(\text{Expand}(\cdot)\) and \(O(\text{Com}||\cdot)\) instead of \(\text{Com}(\cdot)\) to make the usage of the random oracle explicit. Here, we identify \(\text{Expand}\) and \(\text{Com}\) with unique strings.

2.1 Supersingular Elliptic Curves, Isogenies, and Hardness Assumptions

We refer the reader to [Sil09,Gal12] for a more detailed introduction to the topic.

Let \(q\) be a power of a prime \(p \geq 5\), and let \(F_q\) be a finite field with \(q\) elements. An isogeny \(\phi: E \rightarrow E'\) between two elliptic curves \(E\) and \(E'\) over \(F_q\), with points at infinity denoted by \(0_E\) and \(0_{E'}\) respectively, is a non-constant regular rational map mapping \(\phi(0_E)\) into \(0_{E'}\). Every isogeny \(\phi\) can be written in its polynomial form \((F_1(x)/F_2(x), yG_1(x)/G_2(x))\), where \(F_1, F_2, G_1, G_2\) are polynomials over the algebraic closure of \(F\), \(F_1\) is coprime with \(F_2\), and \(G_1\) is coprime with \(G_2\). The isogeny \(\phi\) is said to be defined over \(F_{q^k}\) if the coefficients of the above polynomials are contained in \(F_{q^k}\); in this case, we say that \(E, E'\) are isogenous over \(F_{q^k}\). Tate’s theorem states that \(E, E'\) are isogenous over \(F_{q^k}\) if and only if \(\#E(F_{q^k}) = \#E'(F_{q^k})\).

An invertible isogeny is an isomorphism; in addition, if its domain and image coincide, it is an endomorphism. The set of all endomorphisms of an elliptic curve \(E\) together with the zero map form a ring under pointwise addition and composition, called the endomorphism ring of \(E\) and denoted by \(\text{End}(E)\). If \(\text{End}(E)\) is not commutative, then \(E\) is said to be supersingular. Every supersingular elliptic curve defined over \(F_{p^k}\) for some \(k \in \mathbb{N}\) is isomorphic to an elliptic curve defined over \(F_p\). The degree \(d(\phi)\) of an isogeny \(\phi\) is the maximum among \(\{\deg(F_1), \deg(F_2)\}\); we say that \(\phi\) is a \(d\)-isogeny. Two elliptic curves \(E\) and \(E'\) are \(d\)-isogenous if there exists an isogeny \(\phi: E \rightarrow E'\) of degree \(d\). Given a power \(q\) of a prime \(p > 5\) and a prime number \(\ell \neq p\), we denote by \(\mathcal{G}_q(\ell)\) the graph whose vertices are \(F_q\)-isomorphism classes of supersingular elliptic curves over \(\mathbb{F}_q\) and whose edges are equivalence classes of \(\ell\)-isogenies (two isogenies are in the same class if they have the same kernel). The composition of two isogenies of degrees \(d_1\) and \(d_2\) is an isogeny of degree \(d_1d_2\).

The kernel of an isogeny is finite, and its size is equal to the degree of the isogeny itself if the isogeny is separable. Vice versa, if \(H\) is a finite subgroup of an elliptic curve \(E\), then the elliptic curve \(E/H\) and a separable isogeny \(\psi: E \rightarrow E'\) of kernel \(\ker(\psi) = H\) are unique (modulo isomorphism). Both
$E/H$ and $\psi$ can be computed with complexity $O(\#H)$ using Velu’s formulas. We say that $\phi$ is a cyclic isogeny when $\ker(\phi)$ is a cyclic group. Given $\ell \in \mathbb{N}$, we denote by $E[\ell]$ the $\ell$-torsion subgroup $\{P \in E \mid [\ell]P = 0_E\}$ of $E$. When $\ell$ and $p$ are relatively prime, $E[\ell] \simeq (\mathbb{Z}/\ell\mathbb{Z}) \times (\mathbb{Z}/\ell\mathbb{Z})$.

2.2 $\Sigma$-protocols

Let $X$ and $Y$ be two sets whose sizes depend on a security parameter $\lambda$. Then $\mathcal{R} \subset X \times Y$ is a polynomially-computable binary relation over $X$ and $Y$ if, for any $(x, w) \in X \times Y$, whether $(x, w) \in \mathcal{R}$ can be decided in time $\text{poly}(|x|)$. If $(x, w) \in \mathcal{R}$, we call $w$ a witness for the statement $x$. The language corresponding to $\mathcal{R}$ is $L_\mathcal{R} = \{x \in X \mid \exists w \in Y : (x, w) \in \mathcal{R}\}$.

A $\Sigma$-protocol for a polynomially-computable binary relation $\mathcal{R}$ is a public-coin three-move interactive protocol between a prover and a verifier. Informally, a prover can demonstrate knowledge of a valid witness for a certain statement without revealing any information about the witness itself. Below, we define a relaxed version of sigma protocols where the special-soundness extractor only extracts a witness for a slightly larger relation $\tilde{\mathcal{R}}$, with $\mathcal{R} \subseteq \tilde{\mathcal{R}}$. Furthermore, the definition is given in the random oracle model, i.e., prover and verifier have access to a random oracle $O$. We may occasionally omit the superscript $O$ when the meaning is clear from the context.

**Definition 2 ($\Sigma$-protocols).** A $\Sigma$-protocol $S$ for polynomially-computable binary relations $\mathcal{R} \subseteq \tilde{\mathcal{R}}$ consists of five polynomial-time algorithms $(\text{Gen}, P = (P_1, P_2), V = (V_1, V_2))$ with oracle-access, where $V_2$ is deterministic, $\text{Gen}$, $P_1$, $P_2$ and $V_1$ are probabilistic, and $P_1, P_2$ share states. We denote by $\text{ComSet}$, $\text{ChSet}$, and $\text{ResSet}$ the commitment space, challenge space, and response space respectively. In the random-oracle model, the protocol goes as follows:

- The key-generation algorithm $\text{Gen}^{O}(1^\lambda)$ takes the security parameter $1^\lambda$ as input, and outputs a statement-witness pair $(x, w) \in \mathcal{R}$.
- On input $(x, w) \in \mathcal{R}$, the prover computes $\text{com} \leftarrow P_1^O(x, w)$ and sends the commitment $\text{com}$ to the verifier.
- The verifier runs $\text{ch} \leftarrow V_1^O(\text{com})$ to obtain a random challenge, and sends $\text{ch}$ to the prover.
- Given $\text{ch}$, the prover computes $\text{resp} \leftarrow P_2^O(x, w, \text{com}, \text{ch})$ and returns the response $\text{resp}$ to the verifier.
- The verifier runs $V_2^O(x, \text{com}, \text{ch}, \text{resp})$ and outputs 1 if they accept, 0 otherwise.

Here $O$ is modelled as a random oracle. Moreover, a transcript $(x, \text{com}, \text{ch}, \text{resp}) \in X \times \text{ComSet} \times \text{ChSet} \times \text{ResSet}$ of the protocol is said to be valid (relative to $x$) in case $V_2(x, \text{com}, \text{ch}, \text{resp})$ outputs 1.
We require the following properties of a $\Sigma$-protocol:

1. **Correctness**: all honestly generated transcripts must be valid. Formally, it is required that

   \[
   \Pr \left[ V_2^O(x, \text{com}, \text{ch}, \text{resp}) = 1 \right] = 1.
   \]

2. **Relaxed $\kappa$-Special Soundness**: there exists a polynomial-time extraction algorithm $Ex$ such that, given any $\kappa$ valid transcripts $(x, \text{com}, \text{ch}_1, \text{resp}_1), \ldots, (x, \text{com}, \text{ch}_\kappa, \text{resp}_\kappa)$ relative to the same statement $x \in L_R$, with the same commitment $\text{com}$ and $\kappa$ distinct challenges $\text{ch}_1, \ldots, \text{ch}_\kappa$, outputs $w$ such that $(x, w) \in \tilde{R}$ (note that $Ex$ is only required to recover a witness in $\tilde{R} \supseteq R$).

3. **Statistical and Computational Honest-Verifier Zero-Knowledge (HVZK)**: within this definition, we allow the adversary, the prover and the simulator to make queries to a common random oracle $O$. We say the $\Sigma$-protocol is statistically HVZK if there exists a PPT simulator algorithm $\text{Sim}_O$ such that, for any $(x, w) \in R$, any honestly chosen $\text{ch} \in \text{ChSet}$ and any computationally unbounded adversary $A$ that makes at most a polynomial number of queries to $O$, we have

   \[
   \Pr \left[ A^O(\text{com}, \text{ch}, \text{resp}) = 1 \right] \leq \Pr \left[ \text{com} \leftarrow P_1^O(x, w); \text{resp} \leftarrow P_2^O(x, \text{com}, w, \text{ch}) \right] = \text{negl}(\lambda).
   \]

If the above relation holds only for computationally bounded adversaries, the protocol is said to be computationally HVZK.

### 2.3 Digital signatures

Below we recall the definition of digital signature schemes, correctness and unforgeability.

**Definition 3 (Digital signature schemes).** A digital signature scheme $DS$ consists of three algorithms $(\text{KeyGen}, \text{Sign}, \text{Vrfy})$ defined as follows:

- $(vk, sk) \leftarrow \text{KeyGen}(1^\lambda)$: on input a security parameter $\lambda$, the key-generation algorithm outputs a pair of verification and signing keys $(vk, sk)$.
- $\sigma \leftarrow \text{Sign}(sk, M)$: on input a signing key $sk$ and a message $M$, the signing algorithm outputs a signature $\sigma$.
- $b \in \{0, 1\} \leftarrow \text{Vrfy}(vk, M, \sigma)$: on input a verification key $vk$, a message $M$ and a signature $\sigma$, the verification algorithm outputs 1 (accept) or 0 (reject).
Correctness. For every security parameter $\lambda \in \mathbb{N}$ and every message $M$, a signature scheme is correct if the following holds:

$$\Pr \left[ \text{Vrfy}(vk, M, \sigma) = 1 \left| (vk, sk) \leftarrow \text{KeyGen}(1^\lambda), \sigma \leftarrow \text{Sign}(sk, M) \right. \right] = 1.$$ 

Security. We define existential unforgeability under chosen message attack (EUF-CMA) with the following game between an adversary $A$ and a challenger.

**Setup:** The challenger runs $(vk, sk) \leftarrow \text{KeyGen}(1^\lambda)$ and provides the adversary $A$ with the verification key $vk$. It also prepares an empty set $S = \emptyset$.

**Signing Queries:** The adversary $A$ may adaptively submit messages $M$ to the challenger. The challenger responds with $\sigma \leftarrow \text{Sign}(sk, M)$ to $A$’s query on a message $M$ and updates the set $S \leftarrow S \cup \{(M, \sigma)\}$.

**Output:** Finally, $A$ outputs a forgery $(M^*, \sigma^*)$. We say that the adversary $A$ wins if $(M^*, \sigma^*) \notin S$ and $\text{Vrfy}(vk, M^*, \sigma^*) = 1$.

We then say that the signature scheme $DS$ is EUF-CMA-secure if, for all PPT adversaries $A$, the advantage of $A$ in winning the above game is negligible in the security parameter $\lambda$:

$$\text{Adv}_{A}^{\text{EUF-CMA}}(\lambda) := \Pr[A \text{ wins}] = \text{negl}(\lambda).$$

Via the Fiat-Shamir transform [FS86], a $\Sigma$-protocol $S$ for a binary relation $R$ can be turned into a digital signature scheme. The resulting scheme $FS(S)$ differs from $S$ in the challenge computation, as the challenge is set equal to the digest $H(\text{com}, M)$ - where $M$ is the message to sign and $H$ a hash function - instead of being randomly produced by the verifier. If the binary relation $R$ is based on a hard problem, then $FS(S)$ can be proved EUF-CMA secure.

### 2.4 The $\Sigma_{\text{SEC}}$ protocol

In this section we describe the $\Sigma$-protocol $\Sigma_{\text{SEC}}$, a weaker variant of $\Sigma_{\text{SECUE}}$ from [BCC+22, Sec. 4] (see Remark 6 for the differences between the two protocols).

For every possible value of the security parameter $\lambda$, $p$ will denote a prime of the form $p = \ell_1 \ell_2^2 f \pm 1$ (where $\ell_1, \ell_2$ are small primes such that $\ell_1 \approx \ell_2^2$ and $f \in \mathbb{N}$ is a small cofactor), $E_0$ a fixed supersingular elliptic curve over $\mathbb{F}_p^2$ such that $\#E_0(\mathbb{F}_p^2) = (\ell_1 \ell_2^2 f)^2$ (when considering SIKE parameters, we will take $E_0 : y^2 = x^3 + 6x^2 + x)$, $\{P_1, Q_1\}$ and $\{P_2, Q_2\}$ basis for $E_0[\ell_1]$ and $E_0[\ell_2^2]$, respectively. Then, the tuple $pp = (p, \ell_1, \ell_2, e_1, e_2, f, E_0, P_1, Q_1, P_2, Q_2)$ forms the public parameter for the protocol.

The $\Sigma$-protocol $\Sigma_{\text{SEC}}$ consists of five oracle-calling algorithms $(Gen, P = (P_1, P_2), V = (V_1, V_2))$, where:

- $(E_1, \varphi) \leftarrow Gen(1^\lambda)$: on input a security parameter, the key-generation algorithm uniformly samples $s$ from $\mathbb{Z}/\ell_1^2 \mathbb{Z}$ and computes the cyclic isogeny $\varphi : E_0 \to E_1 := E_0/\langle P_1 + [s]Q_1 \rangle$ having $\langle P_1 + [s]Q_1 \rangle$ as kernel. It returns the statement-witness pair $(E_1, \varphi)$.
domain and the relaxed relation \( E \) number of rational points of 
\[ R \] has relaxed 3-special soundness for the relations \( DSSP \) one of the following distributions, the decisional supersingular product problem 
\[ \Sigma \] Proposition 5 that 
mor, under the assumption that the following problem is hard, we prove in 
\[ P ( \mathcal{E}_k ) \] sampled with probability 1 and 
\[ \ell_2^2 \] computing \( \mathsf{com}_1 \leftarrow \mathcal{O}(\mathsf{Com}|E|)[b_2] \) and \( \mathsf{com}_2 \leftarrow \mathcal{O}(\mathsf{Com}|E|)[b_3] \) via the random oracle \( \mathcal{O} \). The output is \( \mathsf{com} = (\mathsf{com}_1, \mathsf{com}_2) \). 
\[ \{ -1, 0, 1 \} \leftarrow \mathcal{V}_1 ( \mathsf{com} ) : \text{on input a commitment } \mathsf{com}, \mathcal{V}_1 \text{ outputs a random challenge } \mathsf{ch} \in \{ -1, 0, 1 \} . \]
\[ \mathsf{resp} \leftarrow \mathcal{P}_2 ( \mathcal{E}_1, \varphi, \mathsf{com}, \mathsf{ch} ) : \text{on input a statement } \mathcal{E}_1, \text{ a corresponding witness } \varphi, \text{ a commitment } \mathsf{com} = (\mathsf{com}_1, \mathsf{com}_2) \text{ and a challenge } \mathsf{ch} \in \{ -1, 0, 1 \} , \text{ it outputs a response } \mathsf{resp} \text{ defined as follows.} \]
\[ 1/0 \leftarrow \mathcal{V}_2 ( \mathcal{E}_1, \mathsf{com}, \mathsf{ch}, \mathsf{resp} ) : \text{it takes as input a statement } \mathcal{E}_1, \text{ a commitment } \mathsf{com} = (\mathsf{com}_1, \mathsf{com}_2), \text{ a challenge } \mathsf{ch} \in \{ -1, 0, 1 \} \text{ and a response } \mathsf{resp}. \]

Let \( X \) be the set of supersingular elliptic curves \( E_1 \) over \( \mathbb{F}_{p^2} \) having the same number of rational points of \( E_0 \), and \( Y \) be the set of all separable isogenies with domain \( E_0 \). Define the relation 
\[ \mathcal{R}_{\mathsf{SEC}} = \{ (E_1, \varphi) \mid E_1 \in X, \varphi \in Y, \varphi : E_0 \rightarrow E_1, \deg(\varphi) = \ell_1^{e_1} \} \]
and the relaxed relation 
\[ \hat{\mathcal{R}}_{\mathsf{SEC}} = \left\{ (E_1, w) \mid \begin{array}{l}
E_1 \in X \\
\varphi = \varphi : E_0 \rightarrow E_1, \varphi \in Y, \deg(\varphi) = \ell_2^2 \ell_1^{e_1} \text{ with } 0 \leq i \leq e_2 \\
or w = (x, x') \text{ s.t. } x \neq x', \mathcal{O}(\mathsf{Com}|x) = \mathcal{O}(\mathsf{Com}|x')
\end{array} \right\}. \]

It is not difficult to see that the \( \Sigma \)-protocol \( \Sigma_{\mathsf{SEC}} \) described above is correct and has relaxed 3-special soundness for the relations \( \mathcal{R}_{\mathsf{SEC}} \) and \( \hat{\mathcal{R}}_{\mathsf{SEC}} \). Furthermore, under the assumption that the following problem is hard, we prove in 
Proposition 5 that \( \Sigma_{\mathsf{SEC}} \) is computationally HVZK.

**Problem 4 (Decisional Supersingular Product Problem).** Let \( \varphi : E_0 \rightarrow E_1 \) be an isogeny of degree \( \ell_1^{e_1} \). Given \( (E_2, E_3, \varphi') \) sampled with probability 1/2 from one of the following distributions, the decisional supersingular product problem \( \mathcal{DSSP}_{pp} \) requires to determine which distribution it is from:
compute the isogeny

Proof.

- choose a random point $R \in E_0[\ell^2_2]$ of order $\ell^2_2$. Let $\psi : E_0 \rightarrow E_2$ and $\psi' : E_1 \rightarrow E_3$ be the isogenies with kernels $\langle R \rangle$ and $\langle \varphi(R) \rangle$, respectively. Then let $\varphi' : E_1 \rightarrow E_2$ be the isogeny having $\langle \psi(\text{Ker}(\varphi)) \rangle$ as kernel, where $\text{deg}(\varphi') = \ell^2_1$.

- choose $E_2$ randomly among all the supersingular elliptic curves defined over $\mathbb{F}_p$, having the same number of rational points as $E_0$. Then, choose a random point $U \in E_2$ of order $\ell^2_1$ and compute the isogeny $\varphi' : E_2 \rightarrow E_3$ having $\langle U \rangle$ as kernel.

Proposition 5. Let $\lambda$ be a security parameter and let $\text{pp} = (p, \ell_1, \ell_2, e_1, e_2, f, E_0, P_1, Q_1, P_2, Q_2)$ be the public parameters. The $\Sigma$-protocol $\Sigma_{\text{SEC}}$ is computationally HVZK for the relation $R_{\text{SEC}}$ under DSSP$_{\text{pp}}$, assuming that the commitment oracle is computationally hiding.

Proof. For $\text{ch} = -1$, the simulator $\text{Sim}^{O}$ uniformly samples $r$ from $\mathbb{Z}/\ell^2_2\mathbb{Z}$, computes the isogeny $\psi$ of kernel generator $R = P_2 + rQ_2$ and the elliptic curve $E_2 = E_0\langle R \rangle$. It then uniformly samples $b_2, b_3 \leftarrow \{0, 1\}^\lambda$ and sets $\text{com}_1 \leftarrow \mathcal{O}(\text{Com}||E_2||b_2)$ and $\text{com}_2 \leftarrow \mathcal{O}(\text{Com}||E_3||b_3)$. The isogeny $\psi$ is computed as in the original protocol, so the transcript is valid. Under the assumption that the commitment oracle is computationally hiding, an adversary cannot distinguish between the simulated $\text{com}_2$ and a commitment computed following the protocol.

For $\text{ch} = 1$, the simulator uniformly samples $r$ from $\mathbb{Z}/\ell^2_2\mathbb{Z}$, computes a basis $\{P'_2, Q'_2\}$ of $E_1[\ell^2_2]$ and the isogeny $\psi'$ of kernel generator $R' = P'_2 + rQ'_2$, with $\text{com}_1 = \text{Sim}^{O}(E_1)$ and $\text{com}_2 \leftarrow \mathcal{O}(\text{Com}||E_3||b_3)$. The transcript is valid, since by construction we have that $\text{Sim}^{O}(E_1)$, that its order is $\ell^2_2$ and that the resulting $\varphi'$ has image $E_3$. Without knowing the witness, an adversary cannot tell $R'$, of order $\ell^2_2$, from a point of order $\ell^2_2$ that is an image through $\varphi$. The mixing properties of supersingular isogeny graphs ensure that $E_3$ is randomly distributed, as it would be if it was computed following the protocol. Finally, if the commitment oracle is computationally hiding, the adversary cannot distinguish the simulated $\text{com}_1$ from a properly formed one.

For $\text{ch} = 0$, the simulator samples $r$ from $\mathbb{Z}/\ell^2_2\mathbb{Z}$, computes the isogeny $\psi$ of kernel generator $R = P_2 + rQ_2$ and the elliptic curve $E_2 = E_0\langle R \rangle$. Then it computes a basis $\{P'_1, Q'_1\}$ of $E_2[\ell^2_1]$, it samples $s$ from $\mathbb{Z}/\ell^2_1\mathbb{Z}$ and the isogeny $\varphi'$ of kernel generator $S' = P'_1 + sQ'_1$, with codomain $E_3 = E_2\langle S' \rangle$. It then uniformly samples $b_2, b_3 \leftarrow \{0, 1\}^\lambda$ and sets $\text{com}_1 \leftarrow \mathcal{O}(\text{Com}||E_2||b_2)$ and $\text{com}_2 \leftarrow \mathcal{O}(\text{Com}||E_3||b_3)$. The transcript is valid, since $S' \in E_2$ has order $\ell^2_1$ and the kernel it generates is that of an isogeny from $E_2$ to $E_3$. The adversary cannot tell the simulated point $S'$ from a properly formed $\psi(\text{Ker}(\varphi'))$, otherwise it would have solved the DSSP$_{\text{pp}}$ instance.

Remark 6. Within the protocol $\Sigma_{\text{SEC}}$, the degree of the isogenies from $E_0$ to $E_1$ and from $E_2$ to $E_3$ is equal to $\ell^2_1$ (with $d_1$ a suitable natural number bigger than $e_1$), while the degree of the isogenies from $E_0$ to $E_2$ and from $E_1$ to $E_3$ is equal to $\ell^2_2$ (with $d_2$ a suitable natural number bigger than $e_2$). In this way,
the protocol can be proved to be statistically HVZK [BCC+22, Prop. 17]. The conditions to satisfy this stronger property heavily affect both the transcript size and the execution times. Since in our work we are only interested in standard digital signatures, it is enough to rely on the computational HVZK property of the $\Sigma$-protocol. This allows us to preserve in full the SIDH parameters, including the degrees of the isogenies.

As the protocol $\Sigma_{\text{SEC}}$ has a soundness error $\epsilon = 2/3$, it is necessary to repeat its execution in parallel $t$ times in order to obtain a negligible soundness error. It is customary to set $t$ as the minimum positive integer such that $\epsilon^t < 2^{-\lambda}$. Therefore, we obtain

$$t > \frac{1}{\log_2(3) - 1} \lambda \approx 1.7 \cdot \lambda. \quad (1)$$

The $\Sigma$-protocol that results from repeating $\Sigma_{\text{SEC}}$ in parallel $t$-times, which will be denoted by $\Sigma_{\text{SEC}}^t$ in the following, is depicted in Figure 1.

Assuming a supersingular elliptic curve over $\mathbb{F}_{p^2}$ is identified by a single element of $\mathbb{F}_{p^2}$, the average size (in bits) of a transcript of $\Sigma_{\text{SEC}}^t$ (excluding the statement) is approximated by

$$|\text{transcript}| = 4\lambda t + \left\lceil \log(3) \right\rceil t + \frac{t}{3} \left( (2\lceil \log p \rceil + \lceil \log \ell_2 \rceil + \lambda) + (4\lceil \log p \rceil + 1 + \lambda) + (6\lceil \log p \rceil + 1 + 2\lambda) \right)$$

where the terms indicate commitment, challenge and response sizes (the terms within brackets correspond to the sizes of the responses to challenge -1,1 and 0).

Within an execution of $\Sigma_{\text{SEC}}^t$, the prover computes $2t$ elliptic-curve scalar multiplications, $2t$ isogenies and $2t$ commitments to produce the commitment. In addition, to produce the response, the prover evaluates an $\ell_1^t$-isogeny (on a point of order $\ell_1^t$) $t/3$ times, and an $\ell_2^t$-isogeny (on a point of order $\ell_2^t$) $t/3$ times, on average.

When $\Sigma_{\text{SEC}}^t$ is turned into the digital signature scheme $\text{DS}_{\text{SEC}}$ via the Fiat-Shamir transform, the security of $\text{DS}_{\text{SEC}}$ is guaranteed by the hardness of the relation $\mathcal{R}_{\text{SEC}}$, as the problem of finding an isogeny between two given isogenous elliptic curves is still believed to be hard (and it has not been affected by the recent cryptanalytic attacks on SIDH). A signature produced by $\text{DS}_{\text{SEC}}$ is just a transcript of $\Sigma_{\text{SEC}}^t$ without the statement and the challenge (as the latter can be easily recovered, being it the digest of a hash function on the message $m$ to sign and the commitment $\text{com}$). Consequently, the signature size is slightly smaller than the size of a transcript (without the statement) of $\Sigma_{\text{SEC}}^t$, and the computational cost to sign is that undergone by the prover in an execution of $\Sigma_{\text{SEC}}^t$, plus one hash-function evaluation. Therefore, the efficiency analysis presented above applies almost directly to $\text{DS}_{\text{SEC}}$ (see the last column of Table 1 for the signature sizes of $\text{DS}_{\text{SEC}}$ for different SIDH/SIKE parameters).
$P_1(E, \varphi)$:  
1: $(r_1, r_2, \ldots, r_t) \leftarrow (\mathbb{Z}/\ell^2 \mathbb{Z})^t$  
2: $(b_{2,1}, \ldots, b_{2,t}) \leftarrow \{0, 1\}^{ht}$  
3: $(b_{3,1}, \ldots, b_{3,t}) \leftarrow \{0, 1\}^{ht}$  
4: for $i = 1, 2, \ldots, t$ do  
5: $E_{2,i} \leftarrow E_0/(P_2 + [r_i]Q_2)$  
6: $E_{3,i} \leftarrow E_1/(\varphi(P_2 + [r_i]Q_2))$  
7: $\text{com}_{i,1} \leftarrow O(\text{Com}||b_{2,i})$  
8: $\text{com}_{i,2} \leftarrow O(\text{Com}||b_{3,i})$  
9: return $\text{com} \leftarrow (\text{com}_{i,1}, \text{com}_{i,2})_{i=1}^t$  

$P_2(E, \varphi, \text{com}, \text{ch})$:  
1: for $i = 1, 2, \ldots, t$ do  
2: if $ch_i = -1$ then  
3: $\text{resp}_i \leftarrow (E_{2,i}, r_i, b_{2,i})$  
4: else if $ch_i = 1$ then  
5: $\text{resp}_i \leftarrow (E_{3,i}, \varphi(P_2 + [r_i]Q_2), b_{2,i})$  
6: else  
7: $\psi_i \leftarrow \text{IsogenyFromKernel}(E_0, (P_2 + [r_i]Q_2))$  
8: $\text{resp}_i \leftarrow (E_{2,i}, \psi_i(\text{Ker}(\varphi)), E_{3,i}, b_{2,i}, b_{3,i})$  
9: return $\text{resp} \leftarrow (\text{resp}_i)_{i=1}^t$  

$\mathcal{V}_2(E, \text{com}, \text{ch}, \text{resp})$:  
1: for $i = 1, 2, \ldots, t$ do  
2: if $ch_i = -1$ then  
3: $\text{resp}_i \leftarrow (E, r, b)$  
4: if $E_0/(P_2 + [r]Q_2) \neq E$ or $O(\text{Com}||b) \neq \text{com}_{i,1}$ then  
5: return 0  
6: else if $ch_i = 1$ then  
7: $\text{resp}_i \leftarrow (E, T, b)$  
8: if $T \notin E[\ell^2 \mathbb{Z}]_{\text{max}}$ or $E_1/(T) \neq E$ or $O(\text{Com}||b) \neq \text{com}_{i,2}$ then  
9: return 0  
10: else  
11: $\text{resp}_i \leftarrow (E, T, \tilde{E}, b, \tilde{b})$  
12: if $T \notin E[\ell^2 \mathbb{Z}]_{\text{max}}$ or $E/(T) \neq \tilde{E}$ or $O(\text{Com}||b) \neq \text{com}_{i,1}$ or $O(\text{Com}||\tilde{E}) \neq \text{com}_{i,2}$ then  
13: return 0  
14: return 1

Fig. 1. Algorithms in $\Sigma^e_{\text{SEC}}$. Given a supersingular elliptic curve $E$, $\text{IsogenyFromKernel}(E, \cdot)$ denotes an algorithm which, on input a subgroup $S \subset E$, computes an isogeny from $E$ with kernel $S$. Moreover, $E[\ell^2 \mathbb{Z}]_{\text{max}}$ denotes the points of order $\ell^2$ in $E[\ell^e \mathbb{Z}]$, when $\ell$ is prime and $e \in \mathbb{N}$.

| SIKE parameters | $|\log p|$ | $\lambda$ | $t$ | Transcript length | Signature length |
|-----------------|---------|------|----|------------------|-----------------|
| SIKEp344        | 434     | 128  | 218| 543692           | 543256          |
| SIKEp503        | 503     | 128  | 218| 606404           | 605968          |
| SIKEp610        | 610     | 192  | 326| 1163277          | 1162625         |
| SIKEp751        | 751     | 256  | 435| 1956775          | 1955905         |

Table 1. Average sizes (in bits) of the transcripts (excluding the statement) produced by $\Sigma^e_{\text{SEC}}$, and the average length of the signatures produced by DS_{SEC}, working with different SIDH/SIKE parameters. The signature sizes are obtained from Equation (2) minus the challenge length $|\log(3)|t$. 
3 Signature-size Optimisations

In this section, we apply some known cryptographic techniques to DS\textsubscript{SEC} in order to decrease the size of the signatures it produces. We start with some optimisations that determine a reduction of the signature size without causing any increase of the signing computations, and then we discuss those that have an impact on the signing time. We stress that none of the considered optimisations affects the security of DS\textsubscript{SEC}.

3.1 Challenge and commitment recoverability

A $\Sigma$-protocol $(\text{Gen}, P = (P_1, P_2), V = (V_1, V_2))$ is said to be commitment-recoverable if, with overwhelming probability over the random choice of a pair $(x, w) \leftarrow \text{Gen}(1^\lambda)$, for any $ch \in \text{ChSet}$ and $\text{resp} \in \text{ResSet}$, there exists a unique commitment $\text{com} \in \text{ComSet}$ that makes $(x, \text{com}, ch, \text{resp})$ a valid transcript, and such a commitment can be publicly computed by means of an algorithm taking $(x, ch, \text{resp})$ as input. This property allows for shorter signatures by omitting $\text{com}$ from them, and letting the verifier re-compute it. Its correctness is then checked by means of the challenge $ch$.

The original version of the $\Sigma$-protocol $\Sigma\text{t\textsubscript{SEC}}$, described in Figure 1, does not satisfy commitment recoverability (for example, the response $\text{resp}_i$ when $ch_i = -1$ does not allow to recover $\text{com}_{i,2}$). However, we can modify ($\Sigma\text{SEC}$ and) $\Sigma\text{t\textsubscript{SEC}}$ in such a way that the new protocol(s) are commitment-recoverable.

The modification of $\Sigma\text{t\textsubscript{SEC}}$ which we suggest\footnote{We stress that our modification is analogous to the one proposed in [CD22b, Sec. 3.2] for the protocol $\Sigma\text{wSIDH}$.} is detailed in Figure 2. In particular, $P_1$ remains unchanged, while the algorithm $P_2$ outputs the response $\text{resp}_i = (r_i, b_{2,i}, \text{com}_{i,2})$ when $ch_i = -1$; the response $\text{resp}_i = (\varphi(P_2 + [r_i]Q_2), b_{3,i}, \text{com}_{i,1})$ when $ch_i = 1$; the response $\text{resp}_i = (E_{2,i}, \psi_i(\text{Ker}(\varphi)), b_{2,i}, b_{3,i})$ when $ch_i = 0$, where $\psi_i$ is the isogeny with kernel $\langle P_2 + [r_i]Q_2 \rangle$ from $E_0$. The verifier then re-computes part of the commitment, and checks whether it corresponds to that received by $P_1$.

The expected sizes (in bits) of the elements in a transcript of the modified $\Sigma\text{SEC}$ protocol (excluding the statement) are approximated by

$$|	ext{com}| = 4\lambda t, \quad |\text{ch}| = \lceil \log(3) t \rceil,$$

$$|	ext{resp}| = \frac{4}{3} \left( (\lceil \log(\ell_2^p) \rceil + 3\lambda) + (2\lceil \log p \rceil + 1 + 3\lambda) + (4\lceil \log p \rceil + 1 + 2\lambda) \right)$$

(the terms within the brackets corresponds to the size of the responses to challenge -1,1 and 0, respectively).
Efficiency of SIDH-based signatures (yes, SIDH)

\[ P_1(E_1, \varphi); \]

1: \((r_1, r_2, \ldots, r_t) \leftarrow (\mathbb{Z}/\ell t Z)^t \)
2: \((b_{2,1}, \ldots, b_{2,t}) \leftarrow \{0, 1\}^{\lambda t} \)
3: \((b_{3,1}, \ldots, b_{3,t}) \leftarrow \{0, 1\}^{\lambda t} \)
4: for \(i = 1, 2, \ldots, t \) do
5: \( E_{2,i} \leftarrow E_0/(P_2 + [r_i]Q_2) \)
6: \( E_{3,i} \leftarrow E_1/(\varphi(P_2 + [r_i]Q_2)) \)
7: \( \text{com}_{i,1} \leftarrow O(\text{Com}||E_{2,i}||b_{2,i}) \)
8: \( \text{com}_{i,2} \leftarrow O(\text{Com}||E_{3,i}||b_{3,i}) \)
9: return \( \text{com} \leftarrow (\text{com}_{i,1}, \text{com}_{i,2})_{i=1}^t \)

\[ P_2(E_1, \varphi, \text{com}, \text{ch}); \]

1: for \(i = 1, 2, \ldots, t \) do
2: if \( \text{ch}_i = -1 \) then
3: \( \text{resp}_i \leftarrow (r_i, b_{2,i}, \text{com}_{i,2}) \)
4: else if \( \text{ch}_i = 1 \) then
5: \( \text{resp}_i \leftarrow (\varphi(P_2 + [r_i]Q_2), b_{3,i}, \text{com}_{i,1}) \)
6: else
7: \( \psi_i \leftarrow \text{IsogenyFromKernel}(E_0, \text{ch}_i, \text{com}_{i,2}) \)
8: \( \text{resp}_i \leftarrow (E_{2,i}, \psi_i(\text{Ker}(\varphi)), b_{2,i}, b_{3,i}) \)
9: return \( \text{resp} \leftarrow (\text{resp}_i)_{i=1}^t \)

\[ V(E_1, \text{com}, \text{ch}, \text{resp}); \]

1: for \(i = 1, 2, \ldots, t \) do
2: if \( \text{ch}_i = -1 \) then
3: \( \text{resp}_i \leftarrow (r, b, c) \)
4: \( E \leftarrow E_0/(\langle P_2 + [r]Q_2 \rangle \)
5: if \( O(\text{Com}||E||b) \neq \text{com}_{i,1} \) or \( c \neq \text{com}_{i,2} \) then
6: return 0
7: else if \( \text{ch}_i = 1 \) then
8: \( \text{resp}_i \leftarrow (T, b, c) \)
9: if \( T \notin E_1(\ell t Z)_{\text{max}} \) or \( O(\text{Com}||E_1/\langle T \rangle||b) \neq \text{com}_{i,2} \) or \( c \neq \text{com}_{i,1} \) then
10: return 0
11: else
12: \( \text{resp}_i \leftarrow (E, T, b, \bar{b}) \)
13: if \( T \notin E_1(\ell t Z)_{\text{max}} \) or \( O(\text{Com}||E||b) \neq \text{com}_{i,1} \) or \( O(\text{Com}||E/\langle T \rangle||\bar{b}) \neq \text{com}_{i,2} \) then
14: return 0
15: return 1

Fig. 2. Modified \( \Sigma_{\text{SEC}} \) protocol which enjoys commitment recoverability. The blue text marks the differences with the original scheme depicted in Figure 1.

In Table 2 we list the approximated sizes (in bits) of commitments, challenges and responses for the four SIKE parameter sets.

<table>
<thead>
<tr>
<th>SIKE param.</th>
<th>([\log p])</th>
<th>(\lambda)</th>
<th>(t)</th>
<th>(\text{com})</th>
<th>(\text{ch})</th>
<th>(\text{resp}) =</th>
<th>\text{signature}</th>
<th>gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIKEp434</td>
<td>434</td>
<td>128</td>
<td>218</td>
<td>1111616</td>
<td>10</td>
<td>279622</td>
<td>48.53%</td>
<td></td>
</tr>
<tr>
<td>SIKEp503</td>
<td>503</td>
<td>128</td>
<td>218</td>
<td>1111616</td>
<td>10</td>
<td>312249</td>
<td>48.47%</td>
<td></td>
</tr>
<tr>
<td>SIKEp610</td>
<td>610</td>
<td>192</td>
<td>326</td>
<td>250368</td>
<td>10</td>
<td>507993</td>
<td>48.57%</td>
<td></td>
</tr>
<tr>
<td>SIKEp751</td>
<td>751</td>
<td>256</td>
<td>435</td>
<td>445440</td>
<td>11</td>
<td>1005575</td>
<td>48.59%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Sizes (in bits) of the signatures of DSSEC (i.e. the transcripts of \( \Sigma_{\text{SEC}} \) excluding statements and challenges) produced by the modified \( \Sigma_{\text{SEC}} \) after applying challenge and commitment recoverability, for different SIDH/SIKE parameters. The "gain" column indicates by how much the signature lengths have reduced compared to those in Table 1.
Remark 7. The gain column in Table 2 indicates how many bits (in percentage) we save when storing the response computed by the modified $\Sigma^t_{\text{SEC}}$ after applying commitment and challenge recoverability. Each value is computed as $(s_0 - s_1)/s_0$, where $s_0$ is the average length of the response output by $\Sigma^t_{\text{SEC}}$, and $s_1$ is the average length of the response output by the modified version. Every time we will introduce a new optimisation, we will compute the gain it provides as $(s_i - s_{i-1})/s_0$, where $s_i$ is the average response length produced by the version of $\Sigma^t_{\text{SEC}}$ with the current and all previous modifications, and $s_{i-1}$ the response length with only the previous modifications. Note that in this way the overall gain provided by our optimisations can be obtained by simply adding together all the intermediate gains.

Thanks to commitment recoverability, when the modified $\Sigma$-protocol is turned into a digital signature, the commitment $\text{com}$ does not need to be part of the corresponding signature. In principle, the challenge $\text{ch}$ should now be part of the signature, as it necessary to recover the commitment $\text{com}$. However, the modified $\Sigma^t_{\text{SEC}}$ protocol is also challenge-recoverable, and then $\text{ch}$ can be excluded from the signature. We recall that a $\Sigma$-protocol is challenge-recoverable if the challenge $\text{ch}$ in a transcript $(x, \text{com}, \text{ch}, \text{resp})$ can be reconstructed from $\text{resp}$. This is the case for both $\Sigma^t_{\text{SEC}}$ and its modification, since the type of each $\text{resp}$ in $\text{resp}$ uniquely determines the corresponding challenge bit $\text{ch}_i$ in $\text{ch}$. Thus, one can simply omit the challenge from a transcript and let it be deduced from the response. Consequently, both challenge and commitment can be reconstructed by the verifier. Therefore, the signature sizes for different SIDH/SIKE parameters are equal to the response sizes in Table 2. Moreover, we note that the computational effort made by the prover in the modified $\Sigma^t_{\text{SEC}}$ protocol is exactly the same as in the original $\Sigma^t_{\text{SEC}}$ protocol.

### 3.2 Compressed responses

In $\Sigma^t_{\text{SEC}}$, when $\text{ch}_i = 1$ the prover responds with the point $\varphi(P_2 + [r_i]Q_2)$ generating the kernel of the commitment isogeny $\psi'_1 : E_1 \rightarrow E_{3,i}$. Being $P_2$ and $Q_2$ over $F_{p^2}$ by construction, this requires the transmission of $2 \cdot \log p + 1$ bits.

Following the algorithmic improvements proposed in [CLN16,AJK+16], we can deterministically compute a torsion basis $\{P', Q'\}$ of $E_1[\ell_{2}^{e_2}]$ for any statement/public key $E_1$. Then, the response can be set as the result $(\alpha_i, \beta_i)$ of a double discrete logarithm, with $\alpha_i, \beta_i$ such that $[\alpha_i]P' + [\beta_i]Q' = \varphi(P_2 + [r_i]Q_2)$. Since $\varphi(P_2 + [r_i]Q_2)$ is of order $\ell_{2}^{e_2}$, one of the two coefficients $\alpha_i, \beta_i$ must be invertible modulo $\ell_{2}^{e_2}$; if it is $\alpha_i$, we let $\iota_i := 1$ and $\gamma_i := \alpha_i^{-1} \beta_i$, otherwise we let $\gamma_i := 1$ and $\iota_i := \beta_i^{-1} \alpha_i$. The response can then be set as $(\iota_i, \gamma_i)$, and the kernel generator computed as $[\iota_i]P' + [\gamma_i]Q'$.

With this method, the size of the response is therefore reduced to $[\log(\ell_{2}^{e_2})] + 1$ bits, at the cost of computing a deterministic torsion basis both by signer and verifier, and determining a double discrete logarithm only on the signer’s side. Note that the basis $P', Q'$ can be computed once for all by adding it to the statement/public key $E_1$. The new public key would look exactly like an old SIDH
Efficiency of SIDH-based signatures (yes, SIDH)

public key, with the crucial difference that the basis is computed independently of the secret isogeny $\varphi$, preventing the applicability of the attacks on SIDH to this context. Moreover, the following pre-computation would allow us to use the method of compressed responses at no additional computation cost. The prover needs to determine $\varphi(P_2)$ and $\varphi(Q_2)$, and then compute their respective components $\alpha, \beta$ and $\gamma, \omega$ in the basis $\{P', Q'\}$. Then, to compute a response, the prover would only need to calculate one multiplication and two sums in $\mathbb{Z}/eZ$, since $\varphi(P_2 + [r_1]Q_2) = [\alpha + \gamma]P' + [r_1 \cdot (\beta + \omega)]Q'$. We stress that, since the number of bits of the response on challenge $c_h = -1$ is one bit shorter than the compressed response for the challenge $c_h = 1$, the challenge recoverability of the protocol is preserved. The new response length is then computed as

$$|\text{resp}| = \frac{t}{3} \left( \lceil \log(e) \rceil + 3\lambda + \left( \lceil \log(e) \rceil + 1 + 3\lambda \right) + (4\lceil \log p \rceil + 1 + 2\lambda) \right)$$

Remark 8. The above compression method could also be applied for the case where $c_h = 0$, at the cost of a slow down, since a new canonical basis would need to be computed. Indeed, unlike for $c_h \in \{\pm 1\}$, the curve $E_2$ varies for each challenge. We will therefore not consider this compression method for $c_h = 0$ in Table 3.

| SIKE param. | $\lceil \log p \rceil$ | $\lambda$ | $t$ | $|\text{resp}| = |\text{signature}|$ | gain |
|-------------|-----------------|-----------|-----|-------------------------------|------|
| SIKEp434    | 434             | 128       | 218 | 232388                         | 8.69%|
| SIKEp503    | 503             | 128       | 218 | 257531                         | 9.03%|
| SIKEp610    | 610             | 192       | 326 | 498563                         | 8.55%|
| SIKEp751    | 751             | 256       | 435 | 842740                         | 8.33%|

Table 3. Sizes (in bits) of the responses produced by the modified $\Sigma_\text{SIKE}$ with compressed responses for different SIDH/SIKE parameters. Commitment and challenge lengths remain unchanged w.r.t. Table 2. The “gain” column is computed as the difference between the new signature lengths and the ones in Table 2 divided by the signature lengths from Table 1.

3.3 Seed trees

A primitive called seed tree [BKP20] can be used to first generate a number of pseudorandom values and later efficiently disclose an arbitrary subset of them, without revealing any information on the values which are not disclosed. More precisely, a seed tree is a complete binary tree (i.e. a binary tree in which every level, except possibly the last, is completely filled, and all nodes are as far left as possible) of $\lambda$-bit seed values such that the left (resp. right) child of a $\lambda$-bit seed $\text{seed}$ is the left (resp. right) half of the bit string $\text{Expand}(\text{seed}||h)$, where $h$ is a unique identifier for the position of $\text{seed}$ in the binary tree. The seed values of a subset of the set of leaves can be efficiently revealed by sharing the appropriate set of internal seeds in the tree. As a simple example, if the sender (who created
the complete binary tree) only provides the seed value associated to the left child of the root of the tree, then the recipient will only be able to recover the seed values associated to the leaves in the left half of the tree. Notably, the recipient will not learn any information about the leaves in the right half of the tree. A seed tree consists of four oracle-calling algorithms: SeedTree, ReleaseSeeds, RecoverLeaves, SimulateSeeds. Below, we recall the formal definitions of the first three algorithms, where $\text{Expand} : \{0,1\}^\lambda + \{\text{log}_2((t-1)^2)\} \rightarrow \{0,1\}^{2\lambda}$ is a Pseudo-random Generator (PRG, is short) for any $\lambda, t \in \mathbb{N}$, instantiated by a random oracle $O$.

- $\text{SeedTree}^O(\text{seed}_\text{root}, t) \rightarrow \{\text{leaf}_i\}_{i \in \{1, \ldots, t\}}$: on input a root seed $\text{seed}_\text{root} \in \{0,1\}^\lambda$ and an integer $t \in \mathbb{N}$, the algorithm constructs a complete binary tree with $t$ leaves by recursively expanding each seed to obtain its children seeds. Calls to the random oracle are of the form $O(\text{Expand}[\text{seed}][b])$, where $b \in \{1, \ldots, t-1\}$ identifies the position of seed in the binary tree. The algorithm finally outputs the list of seeds associated with the $t$ leaves.

- $\text{ReleaseSeeds}^O(\text{seed}_\text{root}, c, j) \rightarrow \text{seeds}_\text{internal}$: on input a root seed $\text{seed}_\text{root} \in \{0,1\}^\lambda$, a bit string $c \in \{-1,0,1\}^\lambda$, and $j \in \{-1,0,1\}$, it outputs the list of seeds $\text{seeds}_\text{internal}$ that covers all the leaves with index $i$ such that $c_i = j$. Here, we say that a set of nodes $F$ covers a set of leaves $S$ if the union of the leaves of the subtrees rooted at each node $v \in F$ is exactly the set $S$. Here we note that each seed in $\text{seeds}_\text{internal}$ is coupled with an index which identifies its position in the binary tree.

- $\text{RecoverLeaves}^O(\text{seeds}_\text{internal}, c, j) \rightarrow \{\text{leaf}_i\}_{i \text{ s.t. } c_i = j}$: on input a set $\text{seeds}_\text{internal}$, a bit string $c \in \{0,1\}^t$ and a chosen $j \in \{-1,0,1\}$, it computes and outputs all the leaves of the subtrees rooted at the seeds in $\text{seeds}_\text{internal}$.

By construction, the leaves $\{\text{leaf}_i\}_{i \text{ s.t. } c_i = j}$ output by $\text{SeedTree}(\text{seed}_\text{root}, t)$ are the same as those output by $\text{RecoverLeaves}(\text{ReleaseSeeds}(\text{seed}_\text{root}, c, j), c, j)$ for any $c \in \{0,1\}^t$ and $j \in \{-1,0,1\}$. We observe that the last algorithm SimulateSeeds can be used to argue that the seeds associated with all the leaves with index $i$ such that $c_i \neq j$ are indistinguishable from uniformly random values for a recipient that is only given $\text{seeds}_\text{internal}$, $c$ and $j$.

We now describe how seed trees can be used to optimise the modified $\sum_{\text{EC}}$ on two fronts. The optimisation we propose in the following is motivated by the observation that in the first three lines of $P_1$ (Figure 2), all the elements necessary to compute the inputs to the commitment oracle are sampled: the $t$ random coefficients $\{r_1, \ldots, r_t\} \leftarrow \mathcal{S} (\mathbb{Z}/t_r \mathbb{Z})^t$ for the commitment curves $(E_{2,i}, E_{3,i})_{i=1}^t$ and two sets of random strings $(b_{2,1}, \ldots, b_{2,t}, (b_{3,1}, \ldots, b_{3,t}) \leftarrow \mathcal{S} \{0,1\}^{4t}$. Of these inputs, only the coefficients need to be selectively opened, while the entirety of the random strings can be revealed (as their disclosure does not impact the computational HVZK property of the protocol).

Therefore, instead of independently choosing $t$ coefficients and $2t$ random bit-strings, $3t$ seeds could be generated using two distinct seed trees, one for the coefficients which originates from the root $\text{seed}_\text{coeff}_\text{root}$ and one for the random bit-strings which originates from the root $\text{seed}_\text{str}_\text{root}$. Then, instead of selectively...
revealing a subset of the $2t$ bit strings according to the response algorithm $P_2$, the prover could directly send the initial seed $\text{seed}_{\text{root}}^{\text{str}}$, used to generate them, letting the verifier compute them all. On the other hand, instead of responding with the random coefficients $r_i$ for the challenge bits $c_h = -1$, the prover could output $\text{seeds}_{\text{internal}} \leftarrow \text{ReleaseSeeds}(\text{seed}_{\text{root}}^{\text{str}}, c_h, -1)$. The verifier would then use $\text{seeds}_{\text{internal}}$ along with $c_h$ and $j = -1$ to recover the required seeds by running RecoverLeaves.

Let us analyse how generating all random strings from a single root seed $\text{seed}_{\text{root}}^{\text{str}}$ and revealing it to the verifier affects the response length. Each random string is represented by $\lambda$ bits, and without the use of seed trees, one of them is communicated if $c_h = -1$ or $c_h = 1$, and two of them if $c_h = 0$. For $t$ responses $\text{resp}_i$ on challenges evenly distributed over $\{-1, 0, 1\}$, this amounts to $\frac{1}{3}t(\lambda + 2\lambda) = \frac{4}{3}t\lambda$ bits. If all random strings are generated with a seed tree from a root seed $\text{seed}_{\text{root}}^{\text{str}}$, releasing just the root seed requires only $\lambda$ bits.

Such neat analysis cannot be performed on the application of the seed tree primitive to the generation of the coefficients $r_1, \ldots, r_t$, since the amount of internal seeds that need to be revealed depends on how $-1, 0$ and $1$ are distributed over the challenge string. In the worst-case scenario, i.e. when all the leaf seeds need to be revealed, instead of $\frac{\log p}{2}$ bits for the coefficient $r_i$, only $\lambda$ bits for the generating seed need to be communicated.

The following Equation (5) determines how seed trees affect the lengths of the responses produced by the modified $\Sigma^t_{\text{SEC}}$ of Figure 2 when it also incorporates compressed responses (for challenges $c_h = 1$). In parenthesis we add response lengths for $c_h = -1, 1, 0$ respectively, where the $2\lambda$ addends represent the necessary information for commitment recoverability; the lengths of random strings $(b_{2,1}, \ldots, b_{2,t}), (b_{3,1}, \ldots, b_{3,t})$ is removed from each response and replaced by a unique $\lambda$ addend representing $\text{seed}_{\text{root}}^{\text{str}}$.

$$|\text{resp}| = \lambda + \frac{t}{3}(\lambda + 2\lambda) + (\lceil \log (\ell_2^c) \rceil + 1 + 2\lambda) + (4 \lceil \log p \rceil + 1)$$ (5)

In Table 4 we report the numbers produced by Equation (5) for different SIKE parameters. Table 4 differs from Table 3 only in the last two columns.

| SIKE param. | $\lceil \log p \rceil$ | $\lambda$ | $t$ | $|\text{com}|$ | $|\text{ch}|$ | $|\text{resp}| = |\text{signature}|$ | gain |
|-------------|----------------------|---------|-----|----------------|---------|-----------------|------|
| SIKEp434    | 434                  | 128     | 218 | 111616        | 10      | 188771           | 8.03%|
| SIKEp503    | 503                  | 128     | 218 | 111616        | 10      | 211370           | 7.62%|
| SIKEp610    | 610                  | 192     | 326 | 250368        | 10      | 403020           | 8.22%|
| SIKEp751    | 751                  | 256     | 435 | 445440        | 11      | 676681           | 8.49%|

Table 4. Sizes (in bits) of the different components of the transcripts (excluding the statement) produced by the modified $\Sigma^t_{\text{SEC}}$ of Figure 2 when it also incorporates compressed responses (for challenges $c_h = 1$) and seed trees. The “gain” indicates by how much the signature lengths have reduced compared to those in Table 3.
3.4 Unbalanced challenge space

Equation (5) clearly shows that the response $\text{resp}_i$ when $\text{ch}_i = 0$ is significantly bigger than when $\text{ch}_i \in \{-1, 1\}$. As a consequence, one might consider to unbalance the challenge string $\text{ch}$ in order to decrease the overall size of $\text{resp} = (\text{resp}_i)_{i=1}^t$. Such modification was proposed in [BKP20, Sec. 3.4.1], and has the extra positive effect of making the transcript/signature size constant.

To be more concrete, the modification consists in choosing a positive integer $K$ and performing $M$ parallel executions of $\Sigma_{\text{SEC}}$, exactly $K$ of which use the unfavourable challenge bit 0. The number of challenges in $\{-1, 0, 1\}$ $M$ having exactly $K$ components equal to 0 is $\binom{M}{K} \cdot 2^{M-K}$. Equation (1) dictates that, for a given $K$, $M$ should be selected in such a way that the success probability of a dishonest prover is bounded above by $2^{-\lambda}$, i.e.

$$\binom{M}{K} \cdot \frac{2^{M-K}}{n} \geq 2^\lambda. \quad (6)$$

Therefore, for generic $M$ and $K$, it is necessary to find the maximal number $n = n_{M,K}$ of challenges to which a dishonest prover would be able to correctly reply. Afterwards, we will find the optimal $M \in \mathbb{N}$ and $K \in \{0, \ldots, \lceil t/3 \rceil\}$ such that

$$|\text{resp}| = \lambda + \left\lceil \frac{M-K}{2} \right\rceil \left( (\lambda + 2\lambda) + (\lceil \log(\ell_e^2) \rceil + 1 + 2\lambda) \right) + K (4\lceil \log p \rceil + 1)$$

is minimal. We first start by finding $n$.

**Lemma 9.** We can express $n$ as follows

$$n = \max \left\{ \binom{h}{K} : 2^{M-h} : h \in \{K, \ldots, M\} \right\}. \quad (8)$$

**Proof.** Let $S$ be the set of all subsets $U$ of $\{-1, 0, 1\}^M$ consisting of elements of Hamming weight $M-K$ (i.e. elements with $M-K$ non-zero components) and such that, for any index $i \in \{1, \ldots, M\}$, $U$ does not contain three elements whose $i$-th components are all distinct. Then $n$ is the maximum cardinality among the sets in $S$. Given a set $U \in S$, let $h_U$ denote the number of indices $i$ such that there exists a sequence in $U$ that has a zero at index $i$, i.e.

$$h_U := \# \{i \in \{1, \ldots, M\} : \exists x \in U, x_i = 0 \}.$$

Hence, for a set $U \in S$, we can have $\binom{h_U}{K}$ choices for the entries that are zero. The remaining $M - h_U$ entries can be either 1 or −1, giving us $2^{M-h_U}$ choices. Therefore, the maximal size of a set $U \in S$ is $n$ as in Equation (8). \qed
Proposition 10. Let \( h_{\text{max}} \) denote a value of \( h \) realising the maximum of the set in Equation (8). Then \( h_{\text{max}} = 2K \) and

\[
n = \binom{2K}{K} \cdot 2^{M-2K}
\]

Proof. Let us study the behaviour of the discrete function 

\[
f(h) := \binom{h}{K} \cdot 2^{M-h}
\]

taking values in \{\( K, \ldots, M \)\}, with parametrised integers \( K < M \). We start by noticing that the left factor \( \binom{h}{K} \) is monotonically increasing, while the right factor \( 2^{M-h} \) is monotonically decreasing, with ratio \( 2^{M-h} / 2^{M-(h+1)} = 2 \) for any value of \( h \).

The function \( f(h) \) is initially increasing, since the left factor grows faster than how the right factor decreases. In fact, for any \( h \geq K \),

\[
\frac{\binom{h+1}{K}}{\binom{h}{K}} = \frac{h+1}{h+1-K} > 2 \iff h < 2K - 1.
\]

For \( h = 2K - 1 \) the ratio between the binomial coefficients for \( h \) and \( h+1 \) is exactly 2, so \( f(2K-1) = f(2K) \).

For any \( h > 2K - 1 \) the function \( f \) is decreasing, since \( \frac{\binom{h+1}{K}}{\binom{h}{K}} < 2 \) for \( 2K \leq h < M \).

We conclude by arbitrarily choosing \( h_{\text{max}} = 2K \) as a value of \( h \) maximising \( f \) (one can equivalently set \( h_{\text{max}} = 2K - 1 \), obtaining a less neat formula), and thus \( n = \binom{2K}{K} \cdot 2^{M-2K} \).

As an example, when \( \lambda = 128 \) and \( K \) is set to 75, it is sufficient to have \( M = 247 \) parallel runs of \( \Sigma_{\text{SEC}} \), since here the value \( h_{\text{max}} \) giving us the maximal size of \( U \) is 150. The values of \( M \) and \( K \) that optimally minimise the length of the response for different SIDH/SIKE parameters are collected in Table 5.

<table>
<thead>
<tr>
<th>SIKE parameters</th>
<th>( \lceil \log p \rceil )</th>
<th>( \lambda )</th>
<th>( M )</th>
<th>( K )</th>
<th>( h_{\text{max}} )</th>
<th>( \text{resp} )</th>
<th>gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIKEp434</td>
<td>434</td>
<td>128</td>
<td>250</td>
<td>48</td>
<td>96</td>
<td>154783</td>
<td>6.26%</td>
</tr>
<tr>
<td>SIKEp503</td>
<td>503</td>
<td>128</td>
<td>250</td>
<td>48</td>
<td>96</td>
<td>169041</td>
<td>6.99%</td>
</tr>
<tr>
<td>SIKEp610</td>
<td>610</td>
<td>192</td>
<td>362</td>
<td>76</td>
<td>152</td>
<td>330095</td>
<td>5.76%</td>
</tr>
<tr>
<td>SIKEp751</td>
<td>751</td>
<td>256</td>
<td>478</td>
<td>103</td>
<td>206</td>
<td>571453</td>
<td>5.38%</td>
</tr>
</tbody>
</table>

Table 5. Values of \( M \) and \( K \) for the unbalanced challenge space that minimise the response length of the modified \( \Sigma_{\text{SEC}}^{M} \) for different SIDH/SIKE parameters still guaranteeing a negligible soundness error. The size of \( \text{resp} \) is in bits, while the “gain” column reports by how much the signature lengths have reduced compared to those in Table 4.

In order to obtain the values in Table 5, we simply run through all values of \( M \), up to a very large upper bound (say twice the value of the corresponding \( t \)), and all values of \( K \in \{0, \ldots, \lfloor t/3 \rfloor \} \) and pick out the values \((M,K)\) minimizing \( |\text{resp}| \). As expected, the values of \( M \) obtained end up being very close (just a little bit bigger) than the corresponding values of \( t \) (which can be found, for example, in Table 4).
3.5 Summary

We conclude this section highlighting the overall gain in applying challenge and commitment recoverability, compressed responses, seed trees and unbalanced challenge space optimisations to the $\Sigma_{SEC}$ protocol. We phrase the results in terms of signature sizes:

- for $\text{SIKEp434}$ we shorten the signature from 66.31KB to at most 18.89KB, corresponding to a reduction of at least 71.51%
- for $\text{SIKEp503}$ we shorten the signature from 73.97KB to at most 20.63KB, corresponding to a reduction of at least 72.10%
- for $\text{SIKEp610}$ we shorten the signature from 141.92KB to at most 41.03KB, corresponding to a reduction of at least 71.09%
- for $\text{SIKEp751}$ we shorten the signature from 238.76KB to at most 69.76KB, corresponding to a reduction of at least 70.78%

We also note that all the optimisations discussed in this section could be extended to the distributed trusted-setup protocol $[\text{BCC}^+22, \text{Sec. 5}]$ built on top of $\Sigma_{base}^{SEC}$ to collaboratively produce a random supersingular elliptic curve whose endomorphism ring is hard to compute even for the parties who did the sampling.

<table>
<thead>
<tr>
<th>SIKE parameters</th>
<th>$\log p$</th>
<th>$\lambda$</th>
<th>$\text{resp}$</th>
<th>total gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{SIKEp434}$</td>
<td>434</td>
<td>128</td>
<td>18.89KB</td>
<td>71.51%</td>
</tr>
<tr>
<td>$\text{SIKEp503}$</td>
<td>503</td>
<td>128</td>
<td>20.63KB</td>
<td>72.10%</td>
</tr>
<tr>
<td>$\text{SIKEp610}$</td>
<td>610</td>
<td>192</td>
<td>41.03KB</td>
<td>71.09%</td>
</tr>
<tr>
<td>$\text{SIKEp751}$</td>
<td>751</td>
<td>256</td>
<td>69.76KB</td>
<td>70.78%</td>
</tr>
</tbody>
</table>

Table 6. Overall gains in applying all the proposed optimisations to the $\Sigma_{SEC}$ protocol.

4 Running-time optimisations

In an execution of the $\Sigma$-protocol $\Sigma_{SEC}^t$, $2t$ commitment isogenies need to be computed. The same holds for all the modified protocols introduced in Section 3, including the one that considers fixed-weight challenges, which we denote by $\Sigma_{SEC}^M$ (see Section 3.4). All such isogenies have degree $\ell_2^e$; half of them originate from $E_0$, half from $E_1$.

We now present two optimisations that take advantage of the computation of several isogenies of the same degree from the same supersingular elliptic curve. Despite focusing on $\Sigma_{SEC}^t$ (and, implicitly, on $\Sigma_{SEC}^M$), both optimisations could be extended to the distributed trusted-setup protocol $[\text{BCC}^+22, \text{Sec. 5}]$ built on top of $\Sigma_{base}^{SEC}$.

In order to better explain such optimisations, we recall the fastest generic method to compute a cyclic isogeny of degree $\ell_2^e$ from its kernel. For simplicity, we specialise our presentation to the SIKE parameters, and therefore in the following we will replace $\ell_1$ with 2, $\ell_2$ with 3, $e_1$ with $a$ and $e_2$ with $b$. 
Let $\psi$ be an isogeny of degree $3^b$ from a supersingular elliptic curve $E$ over $\mathbb{F}_{p^2}$, with kernel generated by $R := P + [r]Q$ for some basis $\{P, Q\}$ of $E[3^b]$ and some $r \in \mathbb{Z}/3^b\mathbb{Z}$. The isogeny $\psi$ can be expressed as the composition $\psi = \psi^{(b)} \circ \psi^{(b-1)} \circ \cdots \circ \psi^{(1)} = \prod_{j=0}^{b-1} \psi^{(b-j)}$ where each $\psi^{(j)}$ has degree 3. The first isogeny $\psi^{(1)}$ of such decomposition is the isogeny whose kernel is generated by $[3^b-1]R$. Then $\psi^{(2)}$ is the isogeny with kernel generated by $[3^b-2]\psi^{(1)}(R)$, and so on until $\psi^{(b)}$, the last 3-isogeny with kernel generated by $\psi^{(b-1)}(\ldots(\psi^{(1)}(R))\ldots)$.

The strategies described in [ACC+20, Appendix D] speed up the computation of $\psi = \psi^{(b)} \circ \psi^{(b-1)} \circ \cdots \circ \psi^{(1)}$ by minimising the number of operations to execute. We give a high-level description of these strategies in the following lines. In order to recursively determine the kernels of and computing the 3-isogenies in the decomposition $\psi = \prod_{j=0}^{b-1} \psi^{(b-j)}$, the strategies combine two operations: scalar multiplication and isogeny evaluation. Each of these two operations runs in a certain time, with the latter slightly faster than the former. The goal of the strategies is that of minimising the overall computational cost. Referring to Figure 3, we can graphically describe their goal and how they operate. In particular, they aim to obtain the points on the hypotenuse of the right-angled triangle using the least amount of arrows — with blue arrows representing scalar multiplications by 3 and red arrows representing 3-isogeny evaluations — under the condition that the $(i+1)$-th line from the top cannot be accessed before reaching the rightmost point on the $i$-th line from the top, i.e. before the computation of the 3-isogeny $\psi^{(i)}$. In fact, the elements on the hypotenuse, from the top-right corner to the bottom-left corner, represent the kernels of $\psi^{(1)}, \psi^{(2)}, \ldots , \psi^{(b)}$, respectively. The naive (standard) approach to compute $\psi = \prod_{j=0}^{b-1} \psi^{(b-j)}$ would be to start at the first line, go all the way to the right, then move down to the next line, and go all the way to the right, and so on. However, there exist alternative strategies which accelerate the computations. In particular, Figure 3 depicts the optimal strategy proposed in [ACC+20, Appendix D] for the case $b = 6$ (the general strategy for a generic $b$ is just a generalisation of it).

### 4.1 Computing several isogenies in parallel

We now go back to the computation of $2t$ commitment isogenies within an execution of $\Sigma^2$SEC (or within one of its variations). To outline the first optimisation we propose, we restrict our attention to the isogenies $\psi_1, \ldots , \psi_t$ which originate from $E_0$. Analogous considerations hold for the isogenies originating from $E_1$.

As we saw above, for $i \in \{1, \ldots , t\}$, the fastest method to compute $\psi_i$ from its kernel $\langle P_2 + [r_i]Q_2 \rangle$ — where $\{P_2, Q_2\}$ is a basis of $E_0[3^b]$ — is that of determining the composition $\psi^{(b)}_i \circ \psi^{(b-1)}_i \circ \cdots \circ \psi^{(1)}_i$ of 3-isogenies. We then observe that there are 3 possible values for $\psi^{(1)}_i$, $3^2$ for $\psi^{(2)}_i$, $3^3$ for $\psi^{(3)}_i$, and so on. For example, the three possible values for $\psi^{(1)}_i$ have $\langle [3^b-1]P_2 + [0]3^bQ_2 \rangle$, $\langle [3^b-1]P_2 + [1]3^bQ_2 \rangle$ and $\langle [3^b-1]P_2 + [2]3^bQ_2 \rangle$ as kernels. Since $t = 218$ for SIKEp434 and SIKEp503, $t = 326$ for SIKEp610 and $t = 435$ for SIKEp751,
some of the possible 3-isogenies will surely occur multiple times. In particular, for $j \in \{1, \ldots, b\}$ and two different $i, i' \in \{1, \ldots, t\}$, it holds that
\[ \psi_i^{(j)} \circ \cdots \circ \psi_i^{(1)} = \psi_{i'}^{(j)} \circ \cdots \circ \psi_{i'}^{(1)} \iff r_i \equiv r_{i'} \pmod{3^j}. \] (9)
After a few steps, however, repetitions are expected to stop occurring (for the rapid-mixing property of supersingular isogeny graphs). For example, considering the SIKEp434 parameters, since there are $3^5 = 243$ possible values for $r_i \pmod{3^5}$ and $t = 218$, the fourth is the last factor where we can still expect a good amount of repetitions.

Given the above observations, a speed-up in computing the $t$ isogenies of degree $3^b$ from $E_0$ can be obtained by avoiding repeatedly computing 3-isogenies which occur multiple times. To be more precise, this can be achieved by pre-
computing all possible values for $\psi^{(1)}_i, \ldots, \psi^{(a)}_i$ — with $a$ being the biggest positive integer such that $3^a < t$ — and then, for every $i \in \{1, \ldots, t\}$, calculating the congruence classes of $r_i$ modulo $3, 3^2, \ldots, 3^a$ to determine $\psi^{(1)}_i, \ldots, \psi^{(a)}_i$, respectively. Alternatively, for each $i \in \{1, \ldots, t\}$ and $j \in \{1, \ldots, a\}$, the congruence class of $r_i$ modulo $3^j$ can be determined before computing the kernel of $\psi^{(j)}_i$ and the isogeny itself. If such congruence class matches the one of a coefficient $r_k$ with $k \in \{1, \ldots, j-1\}$, then the kernel of $\psi^{(j)}_i$ and the isogeny itself do not need to be re-computed, since $\psi^{(j)}_i = \psi^{(j)}_k$.

Figure 4 shows how the modular-arithmetic checks can be exploited within the optimal strategies proposed in [ACC+20, Appendix D]. In particular, they grant the possibility of moving from line 1 to line $a + 1$ (where $a$ is equal to 3 in the toy example depicted in the figure) without the need to reach the rightmost of the first $a$ lines. In other words, the vertical orange arrows can be simply determined form the modular-arithmetic checks.

---

Fig. 4. By doing some pre-computation or avoiding the multiple computation of 3-isogenies, the optimal strategy from [ACC+20, Appendix D] to compute and evaluate a 3$^a$-isogeny from a kernel generator $R$ is granted the possibility to move from line 1 to line $a + 1$ (with $a = 3$ in the figure) without the need to reach the rightmost of the first $a$ lines. In particular, the points in boxes can obtained instantly from modular-arithmetic checks, and they determine, in turn, the vertical orange arrows.
In order to evaluate the advantage in applying the described tweak, we first deduce formulas for the number of horizontal and vertical arrows, respectively, required in the optimal strategies from \[\text{[ACC}^+\ 20, \text{Appendix D]}\] to compute isogenies of degree 3^b:

\[
H_{\text{os}}(b) = \left\lfloor \frac{3b - 4}{2} \right\rfloor, \quad V_{\text{os}}(b) = \left\lfloor \frac{b + 1}{2} \right\rfloor^2 - (b \mod 2).
\] (10)

We borrow from \[\text{[ACC}^+\ 20, \text{Appendix D]}\] the costs (in cycles) of a point-tripling operation, \(p_3\), and of computation and evaluation of a 3-isogeny, \(q_3\). With them and Equation (10), we calculate the cost of the optimal strategies from \[\text{[ACC}^+\ 20, \text{Appendix D]}\] to compute \(t\) isogenies of degree 3^b, and we compare it with the cost when pre-computation is performed. The results are presented in Table 7. For example, for SIKEp434, \(t\) is equal to 218, \(b = 137\), \(p_3 = 5322\) and \(q_3 = 5282\), with the total cost for computing \(t\) isogenies of degree 3^{137} via the optimal strategy being

\[
218(V_{\text{os}}(137) \cdot q_3 + H_{\text{os}}(137) \cdot p_3) = 5\,716\,545\,548.
\]

If isogenies are pre-computed and selected using modular arithmetics according to Equation (9), we can start each isogeny computation from the fifth line (as noted above, we expect a good amount of repetitions in the first four factors when \(t = 218\)), thus reducing the amount of arrows to compute. In this case, \(H'_{\text{os}}(137) = H_{\text{os}}(137 - 4)\) horizontal arrows and \(V'_{\text{os}}(137) = V_{\text{os}}(137 - 4) + 4\) vertical arrows are required to compute a single 3^{137}-isogeny. The total cost of computing 218 isogenies of degree 3^{137} in parallel is

\[
218(V_{\text{os}}(133) \cdot q_3 + 4 \cdot q_3 + H_{\text{os}}(133) \cdot p_3) = 5\,396\,382\,900
\]

which corresponds to an efficiency increase of 5.60% for SIKEp434 parameters.

We can analogously compute the savings determined by our tweak for the other SIKE parameter sets, pre-computing the first 5 steps for SIKEp610 and SIKEp751 (in which cases 3^5 < \(t\)). The results are summarised in Table 7, where we also indicate the cost of storing the pre-computed kernel generators of all possible initial steps \(\psi_1, \ldots, \psi_4\).

For comparison with the parallel strategy with pre-computation, we now describe a sequential approach to speed up \(\Sigma_{\text{SEC}}\) by avoiding recomputing the same steps several times, but without introducing any pre-computations. When we compute the first isogeny \(\psi_1\) from its coefficient \(r_1\), we store the kernels of its first four initial steps. When computing the second isogeny, we first check whether \(r_2 \equiv r_1 (\mod 3^j)\) for any \(j = 1, 2, 3, 4\): if so, we already have the kernel for that step, and we can save all horizontal lines that would be required to compute it; if not, we compute the kernel corresponding to that step and store it. Repeating this simple procedure would allow us to compute the initial steps of all \(t\) isogenies \(\psi_1, \ldots, \psi_t\) without repeating the same unnecessary scalar multiplications, at the cost of evaluating modular equivalences. To analyse the cost of this strategy, let us start from the one when we considered pre-computations, and let us increase it by the extra cost of performing scalar multiplications in
the first 4 steps. In particular, we need $3 \cdot (b - 1)$ scalar multiplications to obtain the 3 possibilities for $\psi_i^{(1)}$; as per the optimal strategy, we need $3^2$ isogeny evaluations on the kernels $3^{b-2} R_i$ to obtain the 9 possibilities for $\psi_i^{(2)}$. Then again, we perform $3^3 \cdot (b - 3)$ scalar multiplications to obtain the 27 possibilities for $\psi_i^{(3)}$, and finally $3^4$ isogeny evaluations to get the 81 possibilities for $\psi_i^{(4)}$. The total cost of this strategy is therefore

$$t \cdot (V_{\text{os}}(b - 4) \cdot q_3 + 4 \cdot q_3 + H_{\text{os}}(b - 4) \cdot p_3) + (30 \cdot b - 84) \cdot p_3 + 90 \cdot q_3,$$

for SIKEp434 and SIKEp503 and

$$t \cdot (V_{\text{os}}(b - 5) \cdot q_3 + 5 \cdot q_3 + H_{\text{os}}(b - 5) \cdot p_3) + (264 \cdot b - 1254) \cdot p_3 + 90 \cdot q_3,$$

for SIKEp610 and SIKEp751 (considering the fact that repetitions occur in the fifth step as well) and it requires the same amount of extra storage as the pre-computation strategy. The results for all SIKE parameter sets are presented in Table 8.

### 4.2 Computing multiple scalar multiplications in parallel

Following the optimisations presented in the previous section, we note that many points belonging to $E_0$ must be computed. These points are used in the commitment generation of $\Sigma_{\text{SEC}}$ (or in one of its variants), and are the kernel generators $R_1, R_2, \ldots R_t$, where each $R_i = P_2 + [r_i]Q_2$ is obtained by randomly sampling $r_i$ from $\mathbb{Z}/\ell^2 \mathbb{Z}$ (or by using a seed tree, as shown in Section 3.3).

We now discuss how to calculate all $R_i$’s in parallel and obtain some computational savings. An analogous strategy can be applied to parallelise the computation of the kernel generators of the commitment isogenies which originate from $E_1$. 

---

### Table 7.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>SIKE parameters</th>
<th>$t$</th>
<th>$b$</th>
<th>old cost (cc)</th>
<th>new cost (cc)</th>
<th>gain</th>
<th>storage (KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Sigma_{\text{SEC}}$ SIKEp434</td>
<td>218</td>
<td>137</td>
<td>5716545548</td>
<td>5400988804</td>
<td>5.52%</td>
<td>13.035</td>
<td></td>
</tr>
<tr>
<td>SIKEp503</td>
<td>218</td>
<td>159</td>
<td>7642101180</td>
<td>7275879492</td>
<td>4.79%</td>
<td>15.105</td>
<td></td>
</tr>
<tr>
<td>SIKEp610</td>
<td>326</td>
<td>192</td>
<td>16365527304</td>
<td>15704201096</td>
<td>4.04%</td>
<td>55.41</td>
<td></td>
</tr>
<tr>
<td>SIKEp751</td>
<td>435</td>
<td>239</td>
<td>33908315250</td>
<td>32272252410</td>
<td>4.82%</td>
<td>68.2</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{\text{SEC}}$ SIKEp434</td>
<td>250</td>
<td>137</td>
<td>6555671500</td>
<td>6019478500</td>
<td>8.18%</td>
<td>39.43</td>
<td></td>
</tr>
<tr>
<td>SIKEp503</td>
<td>250</td>
<td>159</td>
<td>8763877500</td>
<td>8140531500</td>
<td>7.11%</td>
<td>45.65</td>
<td></td>
</tr>
<tr>
<td>SIKEp610</td>
<td>362</td>
<td>192</td>
<td>18172763448</td>
<td>17438407352</td>
<td>4.04%</td>
<td>55.41</td>
<td></td>
</tr>
<tr>
<td>SIKEp751</td>
<td>478</td>
<td>239</td>
<td>37260171700</td>
<td>35462383108</td>
<td>4.82%</td>
<td>68.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Costs (in clock cycles and kilobytes of storage) and percentage gains in using the pre-computation tweak to compute $t$ commitment isogenies in $\Sigma_{\text{SEC}}$ and in its unbalanced challenge variant $\Sigma_{\text{SEC}}$ compared to those of using the optimal strategies from [ACC+20, Appendix D], for different SIKE parameter sets.
<table>
<thead>
<tr>
<th>SIKE parameters</th>
<th>t</th>
<th>b</th>
<th>old cost (cc)</th>
<th>new cost (cc)</th>
<th>gain</th>
<th>storage (KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIKEp434</td>
<td>218</td>
<td>137</td>
<td>5716545548</td>
<td>5422890556</td>
<td>5.14%</td>
<td>13.035</td>
</tr>
<tr>
<td>SIKEp503</td>
<td>218</td>
<td>159</td>
<td>7642101180</td>
<td>7301293764</td>
<td>4.46%</td>
<td>15.105</td>
</tr>
<tr>
<td>SIKEp610</td>
<td>326</td>
<td>192</td>
<td>16365527304</td>
<td>15967764224</td>
<td>4.35%</td>
<td>55.41</td>
</tr>
<tr>
<td>SIKEp751</td>
<td>435</td>
<td>239</td>
<td>33908315250</td>
<td>32601850914</td>
<td>3.85%</td>
<td>68.2</td>
</tr>
</tbody>
</table>

Table 8. Costs (in clock cycles and kilobytes of storage) and percentage gains in computing \( t \) commitment isogenies in \( \Sigma \) following the strategy that avoids pre-computation but performs modular-arithmetic checks on-the-fly, compared to those of the optimal strategy from [ACC+20, Appendix D], for different SIKE parameter sets.

For each \( R_i \), the scalar multiplication \([r_i]Q_2\) can be performed using the classical double-and-add strategy, in which the points that get doubled and (possibly) added at each step are multiples of \( Q_2 \). Hence, we can perform the multiple doublings of \( Q_2 \) only once for all \( R_i \)’s, as detailed in Figure 5, where \( m \) is the minimum number of bits necessary to represent any coefficient in \( \mathbb{Z}/\ell^2 \mathbb{Z} \), and \( r_i = (r_i^0, r_i^1, \ldots, r_i^{m-1})_2 \) is the little-endian binary representation of \( r_i \in \mathbb{Z}/\ell^2 \mathbb{Z} \).

**Parallel scalar multiplication** \((P_2, Q_2, (r_1, r_2, \ldots, r_t))\):

1: **for** \( i = 1, 2, \ldots, t \) **do**
2: \( R_i \leftarrow 0_{E_0} \)
3: **for** \( j = 0, 1, \ldots, m-2 \) **do**
4: **for** \( i = 1, 2, \ldots, t \) **do**
5: **if** \( r_i^j \) **then**
6: \( R_i \leftarrow R_i + Q_2 \)
7: \( Q_2 \leftarrow 2Q_2 \)
8: **for** \( i = 1, 2, \ldots, t \) **do**
9: **if** \( r_i^{m-1} \) **then**
10: \( R_i \leftarrow R_i + P_2 \)
11: \( R_i \leftarrow R_i + P_2 \)
12: **return** \((R_1, R_2, \ldots, R_t)\)

**Fig. 5.** Algorithm to compute \( t \) coefficients \( R_i = P_2 + [r_i]Q_2 \) in parallel.

In analysing how much this strategy saves us, let \( \text{Hw}(r_i) \) denote the Hamming weight (i.e. the number of non-zero components) of the binary representation of the coefficient \( r_i \), and let \( \text{cADD} \) and \( \text{cDBL} \) denote the cost (in cycles) of adding and doubling points over an elliptic curve, respectively. With a naive approach, we would perform \( m-1 \) doublings and \( \text{Hw}(r_i)+1 \) additions (with “+1” counting
for the last addition by \( P_2 \) for each \( R_i \), at a total cost of

\[
t(m - 1) \cdot \text{cDBL} + \sum_{i=1}^{t} (\text{Hw}(r_i) + 1) \cdot \text{cADD}.
\]

With our parallelised approach presented in Figure 5, we still perform the same amount \( \text{Hw}(r_i) + 1 \) of additions for each \( R_i \), but the \( m - 1 \) doublings are performed once for all the \( R_i \)'s, which saves us \( (t - 1)(m - 1) \cdot \text{cDBL} \) (at least 99% of the doublings for any \( t \geq 100 \), as in our case-study).

5 Conclusions

In this note we have assessed the efficiency of a SIDH-based digital signature built on a weaker but more efficient variant of the recent identification protocol \( \Sigma_{\text{SECUER}}^{\text{base}} \) from [BCC+22, Sec. 4]. The \( \Sigma \)-protocol we consider only achieves computational honest-verifier zero-knowledge instead of the stronger notion of statistical honest-verifier zero-knowledge, but it allows for shorter isogenies. We have conducted our analysis by applying some known cryptographic techniques to decrease the signature size and proposing a minor optimisation to compute many isogenies in parallel from the same starting curve. In addition, we provide novel results on unbalanced challenge space with ternary challenges. Our assessment confirms that the problem of designing a practical isogeny-based signature scheme remains largely open. Nonetheless, the proposed optimisations can be applied to the distributed trusted-setup protocol [BCC+22, Sec. 5] built on top of \( \Sigma_{\text{SECUER}}^{\text{base}} \) to collaboratively produce a random supersingular elliptic curve whose endomorphism ring is hard to compute even for the parties who did the sampling.

References


Efficiency of SIDH-based signatures (yes, SIDH)


