Linearly Homomorphic Signatures over Binary Fields and New Tools for Lattice-Based Signatures

DAN BONEH*
Stanford University, USA
dabo@cs.stanford.edu

DAVID MANDELL FREEMAN†
Stanford University, USA
dfreeman@cs.stanford.edu

October 15, 2010

Abstract

We propose a linearly homomorphic signature scheme that authenticates vector subspaces of a given ambient space. Our system has several novel properties not found in previous proposals:

- It is the first such scheme that authenticates vectors defined over *binary fields*; previous proposals could only authenticate vectors with large or growing coefficients.
- It is the first such scheme based on the problem of *finding short vectors in integer lattices*, and thus enjoys the worst-case security guarantees common to lattice-based cryptosystems.

Our scheme can be used to authenticate linear transformations of signed data, such as those arising when computing mean and Fourier transform or in networks that use network coding. Our construction gives an example of a cryptographic primitive — homomorphic signatures over $\mathbb{F}_2$ — that can be built using lattice methods, but cannot currently be built using bilinear maps or other traditional algebraic methods based on factoring or discrete-log type problems.

Security of our scheme (in the random oracle model) is based on a new hard problem on lattices, called k-SIS, that reduces to standard average-case and worst-case lattice problems. Our formulation of the k-SIS problem adds to the "toolbox" of lattice-based cryptography and may be useful in constructing other lattice-based cryptosystems.

As a second application of the new k-SIS tool, we construct an ordinary signature scheme and prove it k-time unforgeable in the standard model assuming the hardness of the k-SIS problem. Our construction, which can be viewed as "removing the random oracle" from the signatures of Gentry, Peikert, and Vaikuntanathan, has shorter public keys than other lattice-based signatures with the same properties.

Keywords. Lattice-based cryptography, homomorphic signatures, k-time signatures.

*Supported by NSF.
†Supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship.

1 Introduction

A *linearly homomorphic signature scheme* signs n-dimensional vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ defined over some finite field $\mathbb{F}_p$ and outputs one signature per vector. The linear homomorphic property is that given these k signatures, anyone can produce a signature on any vector $\mathbf{v}$ in the $\mathbb{F}_p$-linear span of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. The signature is secure if it is difficult to produce a signature on any vector $\mathbf{v} \in \mathbb{F}_p^n$ outside the linear span of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. We give precise definitions in Section 2.

The original motivation for linearly homomorphic signatures comes from the *network coding* routing mechanism [14, 11, 28, 6, 15]. In a computer network that uses network coding, a message sender signs a number of "augmented" message vectors and transmits the resulting vector-signature pairs to a recipient through the network. Each router along the way receives a number of signed vectors and creates a random linear combination $\mathbf{v}$ of the vectors it receives. The router uses the homomorphic property to derive a signature on $\mathbf{v}$ and forwards $\mathbf{v}$ and its signature to the next router, which then does the same with the signed vectors it receives. The ultimate recipient obtains several random linear combinations of the original message vectors, discards all vectors that are not properly signed, and recovers the original message by solving a full-rank linear system over $\mathbb{F}_p$. Security of the signature scheme ensures that the recipient obtains the originally transmitted message vectors. In implementations there is a desire to use network coding with addition over $\mathbb{F}_2$, so that computations on messages are simple XORs and decoding amounts to solving a linear system over $\mathbb{F}_2$.

Beyond network coding, linearly homomorphic signatures enable linear computations on authenticated data. For example, consider a server that stores signed data samples $s_1, \ldots, s_n$ in $\mathbb{F}_p$. The signature on sample s_i is actually a signature on the vector $(s_i|\mathbf{e}_i) \in \mathbb{F}_p^{n+1}$, where $\mathbf{e}_i$ the ith unit vector in $\mathbb{F}_p^n$. The server stores (i, s_i) and a signature on $(s_i|\mathbf{e}_i)$. (The vector $\mathbf{e}_i$ need not be stored with the data and can be reconstructed from i when needed.) Using the homomorphic property, the server can compute a signature σ on the sum $(\sum_{i=1}^n s_i, 1, \ldots, 1)$. If σ reveals no other information about the original samples, then the server can publish the sum $\sum_{i=1}^n s_i$ and the signature σ on the sum while maintaining privacy of the original data. The "augmentation" $(1, \ldots, 1)$ proves that the published message really is the claimed sum of the original samples.[1] More generally, the server can publish an authenticated inner product of the samples $\mathbf{s} := (s_1, \ldots, s_n)$ with any known vector $\mathbf{c} \in \mathbb{F}_p^n$ without leaking additional information about the samples. This is needed, for example, to publish an authenticated Fourier coefficient from the Fourier transform of $\mathbf{s}$. It is also needed for computing an authenticated least squares fit for a given set of signed data points.

Previous results on linearly homomorphic signatures make use of groups in which the discrete logarithm problem is hard [20, 11, 28, 6] or the RSA assumption holds [15]. In the former case, signatures are linearly homomorphic over $\mathbb{F}_p$ for some large p, while in the latter case, signatures are homomorphic over the integers (with some bound on the size of the coefficients allowed in linear combinations).

In particular, no previous scheme can support linear operations over a small field such as $\mathbb{F}_2$. This appears to be an inherent limitation of discrete-log type systems, since the discrete log problem is not hard in $\mathbb{F}_2$. A similar limitation prevents an RSA-based system over $\mathbb{F}_2$.

More distantly related to our work is the notion of "redactable" signatures [27, 18, 17, 5, 23, 22, 10, 8, 7]. These schemes have the property that given a signature on a message, anyone can derive a signature on subsets of the message. Our focus here is quite different — we look at linear operations on tuples of authenticated vectors rather than a subset operation on a single message.

[1] Strictly speaking, in order to prevent mix-and-match attacks between different data sets one needs to link the n samples with a random tag that uniquely identifies the data set. See Section 2 for details.

Our contributions.

- **Homomorphic signatures over $\mathbb{F}_2$:** We construct the first unforgeable, private, linearly homomorphic signature scheme that authenticates vectors with coordinates in $\mathbb{F}_2$. Our construction gives an example of a cryptographic primitive that can be built using lattice methods, but cannot currently be built using bilinear maps or other traditional algebraic methods based on factoring or discrete-log type problems. Our scheme can be easily modified to authenticate vectors with coefficients in other small fields, including prime fields and extension fields such as $\mathbb{F}_{2^d}$. Privacy here means that a signature on a vector $\mathbf{v}$ in span$(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ that is derived from signatures on $\mathbf{v}_1, \ldots, \mathbf{v}_k$ should leak no information about $\mathbf{v}_1, \ldots, \mathbf{v}_k$ beyond what is revealed by $\mathbf{v}$.

- **Hash-and-sign signatures without random oracles:** We describe a stateless signature scheme and prove it secure in the standard model when used to sign at most k messages, for small values of k. The public key of our scheme is significantly smaller than that of any other stateless lattice-based signature scheme that can sign multiple large messages and is secure in the standard model. Our construction can be viewed as "removing the random oracle" from the signature scheme of Gentry, Peikert, and Vaikuntanathan [16], but only for signing k messages.

- **New tools for lattice-based signatures:** Unforgeability of both of our schemes is based on a new hard problem on lattices, which we call the *k-Small Integer Solutions* (k-SIS) problem. We show that k-SIS reduces to the standard *Small Integer Solution* (SIS) problem, which is known to be as hard as standard worst-case lattice problems [21].

  Privacy of our linearly homomorphic scheme depends on a new result on discrete Gaussian distributions, namely, that the distribution of a sum of samples from a discrete Gaussian is statistically close to a discrete Gaussian distribution that depends *only on the sum* and not on the individual samples. While the analogous result for *continuous* Gaussians is well-known, this is (to our knowledge) the first such result for discrete Gaussians.

  Unforgeability of our hash-and-sign signature scheme also depends on bounds for the length of vectors sampled from discrete Gaussian distributions. We prove both upper and lower bounds that are essentially as tight as possible. Our upper bound improves on a result of Micciancio and Regev [21, Lemma 4.4], and our lower bound is (to our knowledge) the first such bound in the literature.

Overview of the homomorphic signature scheme. Our construction builds on the signature scheme of Gentry, Peikert, and Vaikuntanathan [16], in which signatures are short vectors σ in lattices defined modulo some large integer q. The key idea in our construction is to use short vectors σ in (cosets of) lattices defined modulo $2q$, which allows us to encode different information modulo 2 and modulo q: $\sigma \bmod 2$ encodes information about the vector being signed, while $\sigma \bmod q$ encodes a solution to a hard problem, ensuring that an adversary cannot forge the signature.

The fact that σ is a short *integer* vector ensures that the two parts cannot be attacked independently. Specifically, applying the Chinese remainder theorem to two vectors σ_2 and σ_q that are correct mod 2 and mod q, respectively, does not produce a short integer vector. This property appears to be unique to lattice-based cryptography: if we attempted a similar construction in discrete log groups of order $2q$, we would easily be able to attack the order 2 and order q parts independently.

Concretely, our construction works as follows. Let q be an odd prime. To sign a vector subspace $V = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ of $\mathbb{F}_2^n$, we define a matrix $\mathbf{A}_V \in \mathbb{Z}_{2q}^{m \times n}$ and then sign each basis vector $\mathbf{v}_i$. A

signature on $\mathbf{v}_i \in \mathbb{F}_2^n$ is a low-norm vector $\sigma_i \in \mathbb{Z}^m$ such that

$$\mathbf{A}_V \cdot \sigma_i = q \cdot \mathbf{v}_i \mod 2q$$

A signature $\sigma \in \mathbb{Z}^m$ on a vector $\mathbf{y} \in \mathbb{F}_2^n$ is valid if σ has small norm and $\mathbf{A}_V \cdot \sigma = q \cdot \mathbf{y} \bmod 2q$.

Producing such a signature requires knowing a short basis for the lattice defined by the matrix $\mathbf{A}_V$; to obtain such a basis we combine the trapdoor generation algorithm of Alwen and Peikert [4] with the basis delegation mechanism of Cash, Hofheinz, Kiltz, and Peikert [9].

The homomorphic property of our scheme is now immediate: if we are given arbitrary vector-signature pairs $(\mathbf{v}_j, \sigma_j) \in \mathbb{F}_2^n \times \mathbb{Z}^m$ for $j = 1, \ldots, \ell$, we can create a signature on $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_\ell \in \mathbb{F}_2^n$ by computing $\sigma = \sigma_1 + \cdots + \sigma_\ell \in \mathbb{Z}^m$. Since the σ_j are all valid signatures on the $\mathbf{v}_j$, we see that $\mathbf{A}_V \cdot \sigma = q \cdot \mathbf{v} \bmod 2q$ and σ has low norm (if ℓ is sufficiently small), so σ is a valid signature on $\mathbf{v}$.

Security and the k-SIS problem. To prove unforgeability, we need to show that given signatures on basis vectors of V, it is impossible to generate a signature on a vector outside of V. To do so we define the k-SIS problem, which, roughly speaking, is as follows:

> Given a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and k short vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k \in \mathbb{Z}^m$ satisfying $\mathbf{A} \cdot \mathbf{e}_i = 0 \bmod q$, find a short vector $\mathbf{e} \in \mathbb{Z}^m$ satisfying $\mathbf{A} \cdot \mathbf{e} = 0 \bmod q$, such that $\mathbf{e}$ is not in $\mathbb{Q}$-span$(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$.

When $k = 0$ this is the standard SIS problem [21].

In Section 5 we show that an adversary that breaks the homomorphic signature scheme (defined mod $2q$) in the random oracle model can be used to solve the k-SIS problem (defined mod q). In Section 4 we show that the k-SIS problem is as hard as the SIS problem. Our reduction degrades exponentially in k, which forces us to use a constant-size k if we want our linearly homomorphic scheme to be provably secure based on worst-case lattice problems. It is an interesting open problem to give either a tighter reduction to SIS or a direct reduction from k-SIS to worst-case lattice problems.

For some applications of linearly homomorphic signatures it is desirable that the derived signatures be private; that is, a derived signature on a vector $\mathbf{v}$ in span$(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ should not leak information about the original vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ beyond what is revealed by $\mathbf{v}$. For our construction, to prove privacy it suffices to show that the distribution obtained by summing independent discrete Gaussians depends only on the coset of the sum and the linear combination being computed. We prove this statement in Section 4.

Overview of the k-time signature scheme. Our goal is to use the same mechanism as in the homomorphic signature scheme to construct an ordinary signature scheme. Since homomorphic signatures are not existentially unforgeable, we must find a way to remove the homomorphic property. To do this, we require that the bound on the length of a valid signature $\sigma \in \mathbb{Z}^m$ be very close to the expected length of the vector produced by the signing algorithm. We then show that any linear combination of valid signatures will be too long to satisfy this tight bound, so the homomorphic property is of no use to an adversary.

As with the homomorphic scheme, the security of our k-time signature scheme follows from hardness of the k-SIS problem. We prove security (in the standard model) against a static attacker, who submits all of his message queries before receiving the public key. By a standard transformation using *chameleon hashes* [19], this implies the existence of a scheme secure against an adaptive attacker, also in the standard model. Our security proof also requires tight bounds on the length of a vector sampled from a discrete Gaussian distribution. We use new upper and lower bounds that are essentially as tight as possible.

Outline of the paper. Section 2 gives a formal definition and security model (adapted from [6]) for linearly homomorphic signatures. In Section 3 we review facts about lattices that we will use in our construction and security proof; additional facts can be found in Appendix A. Section 4 describes the k-SIS problem and gives our reduction of k-SIS to SIS. We also prove our new bounds on the length of vectors sampled from discrete Gaussians and our result about the distribution of sums of discrete Gaussian samples. In Section 5 we present our homomorphic scheme and prove its security, and in Section 6 we present our k-time signature scheme and prove its security. Finally, in Section 7 we describe extensions of our scheme to vector spaces over more general fields and pose some open problems.

2 Linearly Homomorphic Signatures

We define linearly homomorphic signatures over any principal ideal domain R. These signatures authenticate tuples (a.k.a. vectors) of elements of R. This definition encompasses the homomorphic signatures over finite fields defined by Boneh et al. [6] as well as the signatures over $\mathbb{Z}$ and $\mathbb{Z}_N$ defined by Gennaro et al. [15]. While we describe the system in terms of a fixed ring R, it may be that R is determined by the Setup algorithm, as in the case where the size of R depends on the system's security parameter.

To prevent "mix-and-match" attacks, each set of vectors signed is given a unique identifier id, which serves to tie together all vectors that belong to the same file or data set. Our security model requires that this identifier be unpredictable; in our scheme it is chosen at random by the signer.

Definition 2.1 (adapted from [6])**.** Let R be a ring. A *linearly homomorphic signature scheme over* R is a tuple of probabilistic, polynomial-time algorithms $(\mathsf{Setup}, \mathsf{Sign}, \mathsf{Combine}, \mathsf{Verify})$ with the following functionality:

- $\mathsf{Setup}(n, \mathsf{params})$. On input a security parameter n (in unary) and additional public parameters params that include the dimension N of the ambient space and the dimension k of subspaces to be signed, this algorithm outputs a public key pk and a secret key sk.
- $\mathsf{Sign}(\mathsf{sk}, \mathsf{id}, \mathbf{v})$. On input a secret key sk, an identifier $\mathsf{id} \in \{0,1\}^n$, and a vector $\mathbf{v} \in R^N$, this algorithm outputs a signature σ.
- $\mathsf{Combine}(\mathsf{pk}, \mathsf{id}, \{(\alpha_i, \sigma_i)\}_{i=1}^{\ell})$. On input a public key pk, an identifier id, and a set of tuples $\{(\alpha_i, \sigma_i)\}_{i=1}^{\ell}$ with $\alpha_i \in R$, this algorithm outputs a signature σ. (this σ is intended to be a signature on $\sum_{i=1}^{\ell} \alpha_i \mathbf{v}_i$.)
- $\mathsf{Verify}(\mathsf{pk}, \mathsf{id}, \mathbf{y}, \sigma)$. On input a public key pk, an identifier $\mathsf{id} \in \{0,1\}^n$, a vector $\mathbf{y} \in R^N$, and a signature σ, this algorithm outputs either 0 (reject) or 1 (accept).

We require that for each $(\mathsf{pk}, \mathsf{sk})$ output by $\mathsf{Setup}(n, \mathsf{params})$, the following hold:

1. For all id and all $\mathbf{y} \in R^N$, if $\sigma \leftarrow \mathsf{Sign}(\mathsf{sk}, \mathsf{id}, \mathbf{y})$ then $\mathsf{Verify}(\mathsf{pk}, \mathsf{id}, \mathbf{y}, \sigma) = 1$.
2. For all $\mathsf{id} \in \{0,1\}^n$ and all sets of triples $\{(\alpha_i, \sigma_i, \mathbf{v}_i)\}_{i=1}^{\ell}$, if it holds that $\mathsf{Verify}(\mathsf{pk}, \mathsf{id}, \mathbf{v}_i, \sigma_i) = 1$ for all i, then

$$\mathsf{Verify}\left(\mathsf{pk},\ \mathsf{id},\ \textstyle\sum_i \alpha_i \mathbf{v}_i,\ \mathsf{Combine}\left(\mathsf{pk},\ \mathsf{id},\ \{(\alpha_i, \sigma_i)\}_{i=1}^{\ell}\right)\right) = 1.$$

In our lattice-based linearly homomorphic signature scheme, we cannot combine arbitrarily many valid signatures and still guarantee successful verification. We capture this property by saying that the scheme is *L-limited* if correctness property (2) holds for all $\ell \leq L$ whenever the σ_i are output by the Sign algorithm.

2.1 Unforgeability

The security model for linearly homomorphic signatures allows an adversary to make adaptive signature queries on files of his choosing, with the signer randomly choosing the identifier id for each file queried. The winning condition captures the fact that there are two distinct types of forgeries: a vector-signature pair $(\mathbf{y}^*, \sigma^*)$ that verifies for some file *not* queried to the signer (a *type 1 forgery*), or a pair $(\mathbf{y}^*, \sigma^*)$ that verifies for some file that *was* queried to the signer, but for which $\mathbf{y}^*$ is not a linear combination of the vectors queried (a *type 2 forgery*).

Definition 2.2 (adapted from [6])**.** A homomorphic signature scheme $\mathcal{S} = (\mathsf{Setup}, \mathsf{Sign}, \mathsf{Combine}, \mathsf{Verify})$ over R is *unforgeable* if the advantage of any probabilistic, polynomial-time adversary $\mathcal{A}$ in the following security game is negligible in the security parameter n:

Setup: The challenger runs $\mathsf{Setup}(n, \mathsf{params})$ to obtain $(\mathsf{pk}, \mathsf{sk})$, and gives pk to $\mathcal{A}$.

Queries: Proceeding adaptively, $\mathcal{A}$ specifies a sequence of k-dimensional subspaces $V_i \subset R^N$, represented as a k-tuples of basis vectors $\mathbf{v}_{i1}, \ldots, \mathbf{v}_{ik}$. For each i, the challenger chooses id_i uniformly from $\{0,1\}^n$ and gives to $\mathcal{A}$ the identifier id_i and the j signatures $\sigma_{ij} \leftarrow \mathsf{Sign}(\mathsf{sk}, \mathsf{id}_i, \mathbf{v}_{ij})$ for $j = 1, \ldots, k$.

Output: $\mathcal{A}$ outputs $\mathsf{id}^* \in \{0,1\}^n$, a *non-zero* vector $\mathbf{y}^* \in R^N$, and a signature σ^*.

The adversary *wins* if $\mathsf{Verify}(\mathsf{pk}, \mathsf{id}^*, \mathbf{y}^*, \sigma^*) = 1$, and either (1) $\mathsf{id}^* \neq \mathsf{id}_i$ for all i (a *type 1 forgery*), or (2) $\mathsf{id}^* = \mathsf{id}_i$ for some i but $\mathbf{y}^* \notin V_i$ (a *type 2 forgery*). The *advantage* $\mathrm{HomSig\text{-}Adv}[\mathcal{A}, \mathcal{S}]$ of $\mathcal{A}$ is defined to be the probability that $\mathcal{A}$ wins the game.

2.2 Privacy

Given signatures on vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in R^N, it is desirable that derived signatures on a vector $\mathbf{v}$ in $\mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ not leak any information about $\mathbf{v}_1, \ldots, \mathbf{v}_k$ beyond what is revealed by $\mathbf{v}$. We are not trying to hide the fact the derivation took place or the function that was used to compute $\mathbf{v}$, merely the inputs to the function.

More precisely, we define privacy for linearly homomorphic signatures using a variation of a definition from [8]. The definition captures the idea that given signatures on a number of derived vectors in one of two different vector spaces, the attacker cannot tell which space the derived signatures came from. We call signatures with this property *weakly context hiding*. The reason for "weak" is that we are not hiding the fact that derivation took place or the computed function and we assume the original signatures are not public.

Definition 2.3. A homomorphic signature scheme $\mathcal{S} = (\mathsf{Setup}, \mathsf{Sign}, \mathsf{Combine}, \mathsf{Verify})$ over R is *weakly context hiding* if the advantage of any probabilistic, polynomial-time adversary $\mathcal{A}$ in the following security game is negligible in the security parameter n:

Setup: The challenger runs $\mathsf{Setup}(n, \mathsf{params})$ to obtain $(\mathsf{pk}, \mathsf{sk})$ and gives pk to $\mathcal{A}$.

Queries: Proceeding adaptively, $\mathcal{A}$ specifies a sequence of k-dimensional subspaces $V_i \subset R^N$, represented as a k-tuples of basis vectors $\mathbf{v}_{i1}, \ldots, \mathbf{v}_{ik}$. For each i, the challenger chooses id_i uniformly from $\{0,1\}^n$ and gives to $\mathcal{A}$ the identifier id_i and the j signatures $\sigma_{ij} \leftarrow \mathsf{Sign}(\mathsf{sk}, \mathsf{id}_i, \mathbf{v}_{ij})$ for $j = 1, \ldots, k$.

Challenge: $\mathcal{A}$ outputs $(V_0, V_1, f_1, \ldots, f_s)$ where V_0 and V_1 are linear spaces over R^N represented as k-tuples of vectors $V_b = \mathrm{span}(\mathbf{v}_1^{(b)}, \ldots, \mathbf{v}_k^{(b)})$ for $b = 0, 1$. The functions $f_1, \ldots, f_s$ are linear functions[2] on $(R^N)^k$

[2]If the scheme is L-limited, we require the f_i to have at most L nonzero coefficients.

satisfying

$$f_i\big(\mathbf{v}_1^{(0)},\ldots,\mathbf{v}_k^{(0)}\big) = f_i\big(\mathbf{v}_1^{(1)},\ldots,\mathbf{v}_k^{(1)}\big) \qquad \text{for all } i = 1,\ldots,s.$$

In response, the challenger generates a random bit $b \in \{0,1\}$ and a random tag $\tau \in \{0,1\}^n$, signs the vector space V_B using the tag τ, and derives signatures σ_i on $f_i(\mathbf{v}_1^{(b)},\ldots,\mathbf{v}_k^{(b)})$ for $i = 1,\ldots,s$. It sends $\sigma_1,\ldots,\sigma_s$ to $\mathcal{A}$. Note that the functions $f_1,\ldots,f_s$ can be output adaptively after V_0, V_1 are output.

Output: $\mathcal{A}$ outputs a bit b'.

The adversary $\mathcal{A}$ wins the game if $b = b'$. The *advantage* of $\mathcal{A}$ is the probability that $\mathcal{A}$ wins the game.

Winning the context hiding game game means that the attacker was able to determine whether the challenge signatures were derived from signatures on V_0 or from signatures on V_1. We note that for discrete-log based linearly homomorphic signatures such as those of [6], weak context hiding follows from the uniqueness of the signature.

3 Background on Lattices

In this section we describe the lattices we will be using and their properties. Results from probability that we will need appear in Appendix A.

Notation. For any integer $q \geq 2$, we let $\mathbb{Z}_q$ denote the ring of integers modulo q. When q is prime, $\mathbb{Z}_q$ is a field and is sometimes denoted $\mathbb{F}_q$. We let $\mathbb{Z}_q^{n\times m}$ denote the set of $n \times m$ matrices with entries in $\mathbb{Z}_q$. We denote matrices by capital boldface letters and vectors by lowercase boldface letters. We say a function $f : \mathbb{Z} \to \mathbb{R}^+$ is *negligible* if it is $O(n^{-c})$ for all $c > 0$, and we use $\mathrm{negl}(n)$ to denote a negligible function of n. The function $\lg x$ is the base 2 logarithm of x.

Lattices. An m-dimensional lattice Λ is a full-rank discrete subgroup of $\mathbb{R}^m$. We will be interested in *integer lattices* Λ, i.e., those whose points have coordinates in $\mathbb{Z}^m$. The lattices we consider consist of vectors either generated by or orthogonal to a certain "arity check" matrix modulo some integer q. More precisely, for any integer $q \geq 2$ and any $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$, we define

$$\begin{aligned}
\Lambda_q^{\perp}(\mathbf{A}) &:= \big\{\mathbf{e} \in \mathbb{Z}^m : \mathbf{A}\cdot\mathbf{e} = \mathbf{0} \bmod q\big\}\\
\Lambda_q^{\mathbf{u}}(\mathbf{A}) &:= \big\{\mathbf{e} \in \mathbb{Z}^m : \mathbf{A}\cdot\mathbf{e} = \mathbf{u} \bmod q\big\}\\
\Lambda_q(\mathbf{A}) &:= \big\{\mathbf{e} \in \mathbb{Z}^m : \exists\, \mathbf{s} \in \mathbb{Z}_q^n \text{ with } \mathbf{A}^t\cdot\mathbf{s} = \mathbf{e} \bmod q\big\}.
\end{aligned}$$

The lattice $\Lambda_q^{\mathbf{u}}(\mathbf{A})$ is a coset of $\Lambda_q^{\perp}(\mathbf{A})$; namely, $\Lambda_q^{\mathbf{u}}(\mathbf{A}) = \Lambda_q^{\perp}(\mathbf{A}) + \mathbf{t}$ for any $\mathbf{t}$ such that $\mathbf{A}\cdot\mathbf{t} = \mathbf{u} \bmod q$.

Length of a basis. Let $\mathbf{S}$ be an ordered set of linearly independent (column) vectors $\mathbf{S} = \{\mathbf{s}_1,\ldots,\mathbf{s}_k\}$ in $\mathbb{R}^m$. We use the following standard notation:

- $\|\mathbf{S}\|$ denotes the length (using the ℓ_2 norm) of the longest vector in $\mathbf{S}$, i.e. $\|\mathbf{S}\| := \max_i\|\mathbf{s}_i\|$ for $1 \leq i \leq k$.
- $\widetilde{\mathbf{S}} := \{\tilde{\mathbf{s}}_1,\ldots,\tilde{\mathbf{s}}_k\} \subset \mathbb{R}^m$ denotes the Gram-Schmidt orthogonalization of the vectors $\mathbf{s}_1,\ldots,\mathbf{s}_k$ taken in that order. We refer to $\|\widetilde{\mathbf{S}}\|$ as the *Gram-Schmidt norm* of $\mathbf{S}$.

Generating a short basis. The public key for our signature scheme will be a random matrix $\mathbf{A}$, and the secret key will be a basis $\mathbf{S}$ for $\Lambda^{\perp}(\mathbf{A})$ with low Gram-Schmidt norm. We can generate $\mathbf{A}$ and $\mathbf{S}$ using an algorithm of Alwen and Peikert [4], which improves on an algorithm of Ajtai [3].

Theorem 3.1 ([4, Theorem 3.2 with $\delta = 1/3$])**.** *Let q be an integer*[3] *and $m := \lceil 6n \lg q \rceil$. There is a probabilistic polynomial-time algorithm* $\mathsf{TrapGen}(q, n)$ *that outputs a pair $(\mathbf{A} \in \mathbb{Z}_q^{n \times m},\ \mathbf{S} \in \mathbb{Z}^{m \times m})$ such that $\mathbf{A}$ is statistically close to a uniform matrix in $\mathbb{Z}_q^{n \times m}$ and $\mathbf{S}$ is a basis for $\Lambda_q^{\perp}\mathbf{A}$ such that $\|\widetilde{\mathbf{S}}\| \leq 30\sqrt{n \lg q}$ with all but negligible probability in n.*

By Lemma A.2, we may assume without loss of generality that the matrix $\mathbf{A}$ generated by $\mathsf{TrapGen}$ has rank n.

Delegating a basis. In our signature scheme, the lattice used to sign a file will need to be derived from two sources: the public key, which is a matrix $\mathbf{A}$ generated using $\mathsf{TrapGen}$, and the file identifier id, which is random. To combine the two, we hash the file identifier to a second matrix and derive a short basis for $\mathbf{A}\|H(\mathsf{id})$. To derive this new basis we use the delegation mechanism of Cash et al.'s identity-based encryption scheme [9].

Theorem 3.2 ([9, Lemma 3.2])**.** *Let $\mathbf{S} \in \mathbb{Z}^{m \times m}$ be an arbitrary basis of $\Lambda^{\perp}(\mathbf{A})$ for a rank n matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, and let $\mathbf{A}' \in Z_q^{n \times m'}$ be arbitrary. There is a deterministic polynomial-time algorithm* $\mathsf{ExtBasis}(\mathbf{S}, \mathbf{B} := \mathbf{A}\|\mathbf{A}')$ *that outputs a basis $\mathbf{T}$ of $\Lambda^{\perp}(\mathbf{B}) \subset \mathbb{Z}^{(m+m') \times (m+m')}$ such that $\|\widetilde{\mathbf{T}}\| = \|\widetilde{\mathbf{S}}\|$.*

Gaussian distributions. Let L be a subset of $\mathbb{Z}^m$. For any vector $\mathbf{c} \in \mathbb{R}^m$ and any positive parameter $\sigma \in \mathbb{R}_{>0}$, let $\rho_{\sigma,\mathbf{c}}(\mathbf{x}) := \exp\left(-\pi\|\mathbf{x} - \mathbf{c}\|^2/\sigma^2\right)$ be a Gaussian function on $\mathbb{R}^m$ with center $\mathbf{c}$ and parameter σ. Let $D_{\sigma,\mathbf{c}}$ be the continuous Gaussian distribution over $\mathbb{R}^m$ with center $\mathbf{c}$ and parameter σ, with $D_{\sigma,\mathbf{c}}(\mathbf{x}) = \rho_{\sigma,\mathbf{c}}(\mathbf{x})/\sigma^n$. Let $\rho_{\sigma,\mathbf{c}}(L) := \sum_{\mathbf{x} \in L} \rho_{\sigma,\mathbf{c}}(\mathbf{x})$ be the discrete integral of $\rho_{\sigma,\mathbf{c}}$ over L. Finally, let $\mathcal{D}_{L,\sigma,c}$ be the discrete Gaussian distribution over L with center $\mathbf{c}$ and parameter σ. In particular, for all $\mathbf{y} \in L$, we have $\mathcal{D}_{L,\sigma,\mathbf{c}}(\mathbf{y}) = \frac{\rho_{\sigma,\mathbf{c}}(\mathbf{y})}{\rho_{\sigma,\mathbf{c}}(L)}$. For notational convenience, $\rho_{\sigma,\mathbf{0}}$ and $\mathcal{D}_{L,\sigma,\mathbf{0}}$ are abbreviated as ρ_σ and $\mathcal{D}_{L,\sigma}$, respectively.

Sampling from a discrete Gaussian. Gentry et al. [16] construct algorithms for sampling from discrete Gaussians.

Theorem 3.3.

(a) [16, Theorem 4.1] There is a probabilistic polynomial-time algorithm $\mathsf{SampleGaussian}$ *that, given a basis $\mathbf{T}$ of an n-dimensional lattice Λ, a parameter $\sigma \geq \|\widetilde{\mathbf{T}}\| \cdot \omega(\sqrt{\log n})$, and a center $\mathbf{c} \in \mathbb{R}^n$, outputs a sample from a distribution that is statistically close to $\mathcal{D}_{\Lambda,\sigma,\mathbf{c}}$.*

(b) [16, Theorem 5.9] There is a probabilistic polynomial-time algorithm $\mathsf{SamplePre}$ *that, given a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, a basis $\mathbf{T}$ of $\Lambda_q^{\perp}(\mathbf{A})$, a parameter $\sigma \geq \|\widetilde{\mathbf{T}}\| \cdot \omega(\sqrt{\log n})$, and a vector $\mathbf{u} \in \mathbb{Z}^n$, outputs a sample from a distribution that is statistically close to $\mathcal{D}_{\Lambda_q^{\mathbf{u}}(\mathbf{A}),\sigma}$.*

[3] The result in the published version of [4] is stated and proved for odd q, with a note that this restriction can be lifted. The result in the full version has no restriction on q.

Recall that if $\Lambda_q^{\mathbf{u}}(\mathbf{A})$ is not empty, then $\Lambda_q^{\mathbf{u}}(\mathbf{A}) = \mathbf{t} + \Lambda_q^{\perp}(\mathbf{A})$ for any $\mathbf{t} \in \Lambda_q^{\mathbf{u}}(\mathbf{A})$. Algorithm $\mathsf{SamplePre}(\mathbf{A}, \mathbf{T}, \mathbf{u}, \sigma)$ simply calls $\mathsf{SampleGaussian}(\mathbf{T}, \sigma, -\mathbf{t})$ and adds $\mathbf{t}$ to the result. For Gaussians centered at the origin, we use $\mathsf{SampleGaussian}(\mathbf{T}, \sigma)$ to denote $\mathsf{SampleGaussian}(\mathbf{T}, \sigma, \mathbf{0})$. We use the notation $\mathsf{SampleGaussian}(\mathbb{Z}^m, \sigma)$ to denote sampling from the lattice $\mathbb{Z}^m$ with a basis consisting of the m unit vectors.

The smoothing parameter. For an n-dimensional lattice Λ and positive real $\epsilon > 0$, the *smoothing parameter* $\eta_\epsilon(\Lambda)$ of Λ is defined to be the smallest positive s such that $\rho_{1/s}(\Lambda^* \setminus \{0\}) \leq \epsilon$ [21]. The key property of the smoothing parameter is that if $\sigma > \eta_\epsilon(\Lambda)$, then every coset of Λ has roughly equal mass. More precisely, for any such σ, if $\epsilon \in (0, 1)$ and $\mathbf{c} \in \mathbb{R}^n$, then we have [16, Lemma 2.7]

$$\tfrac{1-\epsilon}{1+\epsilon} \cdot \rho_\sigma(\Lambda) \leq \rho_{\sigma,\mathbf{c}}(\Lambda) \leq \rho_\sigma(\Lambda). \tag{3.1}$$

For almost all matrices $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, there is a negligible ϵ such that the smoothing parameter $\eta_\epsilon(\Lambda_q^{\perp}(\mathbf{A}))$ is less than $\omega(\sqrt{\log m})$:

Lemma 3.4 ([16, Lemma 5.3])**.** *Let q be a prime and n, m integers such that $m > 2n \lg q$. Let f be some $\omega(\sqrt{\log m})$ function. Then there is a negligible function $\epsilon(m)$ such that for all but at most a q^{-n} fraction of $\mathbf{A}$ in $\mathbb{Z}_q^{n \times m}$ we have $\eta_{\epsilon(m)}(\Lambda_q^{\perp}(\mathbf{A})) < f(m)$.*

Hardness assumption. The security of our signature scheme is based on the problem of finding short vectors in $\Lambda_q^{\perp}(\mathbf{A})$ for random $\mathbf{A}$. This is known as the *Small Integer Solution* (SIS) problem, and is defined as follows.

Definition 3.5. An instance of the $\mathsf{SIS}_{q,m,\beta}$ problem is a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$. A solution to the problem is a nonzero vector $\mathbf{v} \in \mathbb{Z}^m$ such that $\|\mathbf{v}\| \leq \beta$ and $\mathbf{A} \cdot \mathbf{v} = \mathbf{0} \bmod q$ (i.e., $\mathbf{v} \in \Lambda_q^{\perp}(\mathbf{A})$). If $\mathcal{A}$ is an algorithm that takes as input a uniformly random matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, we define $\mathsf{SIS}\text{-}\mathrm{Adv}[\mathcal{B}, (q, m, \beta)]$ to be the probability that $\mathcal{A}$ outputs a solution to a uniformly random $\mathsf{SIS}_{q,m,\beta}$ problem instance $\mathbf{A}$.

Micciancio and Regev [21] and Gentry et al. [16] show that the (average case) SIS problem for $\beta = \mathrm{poly}(n)$ is hard assuming worst-case hardness of certain standard approximation problems on lattices, such as the *shortest independent vector problem* SIVP and the *shortest vector problem* GapSVP.

4 New Tools

4.1 A "One-More" SIS Problem

The security of most lattice-based signature schemes depends on the adversary's inability to find a short vector in $\Lambda_q^{\perp}(\mathbf{A})$ for some public matrix $\mathbf{A}$. However, for our linearly homomorphic signatures this criterion is insufficient. Roughly speaking, an adversary in our scheme will be given several short vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k \in \Lambda_q^{\perp}(\mathbf{A})$ and must produce a short vector in $\Lambda_q^{\perp}(\mathbf{A})$ that is not in the span of the $\mathbf{e}_i$. This is a "one-more" variant of the standard SIS problem, analogous to the "one-more discrete logarithm" problem in group-based cryptography (see e.g., [24]). We will see in Section 4.3 below that for certain choices of parameters the problem is equivalent to finding *any* short vector in $\Lambda_q^{\perp}(\mathbf{A})$ distinct from $\{\pm\mathbf{e}_i\}$, making the "one-more" analogy even more appropriate.

We now formally define the problem.

Definition 4.1. For any integer $k \geq 0$, an instance of the k-$\mathsf{SIS}_{q,m,\beta,\sigma}$ problem is a uniformly random matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and a set of k vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k \in \Lambda_q^\perp(\mathbf{A})$ drawn from the distribution $\mathcal{D}_{\Lambda_q^\perp(\mathbf{A}),\sigma}$. A solution to the problem is a nonzero vector $\mathbf{v} \in \mathbb{Z}^m$ such that

1. $\|\mathbf{v}\| \leq \beta$,
2. $\mathbf{A} \cdot \mathbf{v} = \mathbf{0} \bmod q$ (i.e., $\mathbf{v} \in \Lambda_q^\perp(\mathbf{A})$), and
3. $\mathbf{v} \notin \mathbb{Q}\text{-span}(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$.

If $\mathcal{A}$ is an algorithm that takes as input a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and vectors $\mathbf{e}_i \in \mathbb{Z}^m$ for $i = 1, \ldots, k$, we define k-SIS-$\mathrm{Adv}[\mathcal{B}, (q, m, \beta, \sigma)]$ to be the probability that $\mathcal{A}$ outputs a solution to a k-$\mathsf{SIS}_{q,m,\beta,\sigma}$ problem instance $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$ drawn at random from the appropriate distribution.

When $k = 0$ the k-SIS problem is identical to the SIS problem. The main result of this section is to show that an adversary $\mathcal{A}$ that solves the k-SIS problem in dimension m can be used to solve the SIS problem in dimension $m - k$. More precisely, we have the following:

Theorem 4.2. *Let q be a prime, and let m, β, σ, and k, be polynomial functions of a security parameter n. Suppose that $m \geq 2n \lg q$, $m/k > n$, $\sigma > \omega(\sqrt{\log m})$, $t > \omega(\sqrt{\log n})$, and $q > \sigma \cdot \omega(\sqrt{\log m})$.*

Let $\beta' = \beta \cdot (k^{3/2} + 1)k!(t\sigma)^k$. Let $\mathcal{A}$ be a polynomial-time adversary for the k-$\mathsf{SIS}_{q,m,\beta,\sigma}$ problem. Then there exists a polynomial-time algorithm $\mathcal{B}$ that solves $\mathsf{SIS}_{q,m-k,\beta'}$, such that

$$\mathsf{SIS}\text{-}\mathrm{Adv}[\mathcal{B}, (q, m-k, \beta')] \geq k\text{-}\mathsf{SIS}\text{-}\mathrm{Adv}[\mathcal{A}, (q, m, \beta, \sigma)] - \epsilon,$$

where ϵ is a negligible function of n.

Since the SIS problem is only assumed to be hard for parameters $\beta \in \mathrm{poly}(n)$, the fact that the above reduction degrades exponentially in k means that k must be chosen to be small enough so that β' is still polynomial in n. In our application the parameter σ is $\omega(\sqrt{n})$, which means that k must be chosen to be $O(1)$. In this case, if we take $t = O(\log \sigma)$ and $\beta' = \beta \cdot O(\sigma^k \log^k \sigma)$, then Theorem 4.2 shows that if the $\mathsf{SIS}_{q,m-k,\beta'}$ problem is hard, then the k-$\mathsf{SIS}_{q,m,\beta,\sigma}$ problem is also hard.

The idea of the proof of Theorem 4.2 is as follows: given an SIS challenge $\mathbf{A}' \in \mathbb{Z}_q^{n \times (m-k)}$, we can choose k random vectors $\mathbf{e}_i$ from a Gaussian distribution over $\mathbb{Z}^m$ and append k columns to $\mathbf{A}'$ to create a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ such that the $\mathbf{e}_i$ are in $\Lambda_q^\perp(\mathbf{A})$. If the k-SIS adversary $\mathcal{A}$ outputs a short vector $\mathbf{e}^* \in \Lambda_q^\perp(\mathbf{A})$ that is $\mathbb{Q}$-linearly independent of the $\{\mathbf{e}_i\}$, then we can compute a short vector $\mathbf{v} \in \Lambda_q^\perp(\mathbf{A})$ with zeroes in the last k entries. Reading off the first $m - k$ entries of $\mathbf{v}$ gives us a short vector in $\Lambda_q^\perp(\mathbf{A}')$.

To turn this idea into a formal proof, we need to show that the tuple $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$ generated in this manner is indistinguishable from a tuple $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$ selected from the distribution of k-SIS challenge instances. To show this, we define two distributions on $\mathbb{Z}_q^{n \times m} \times \mathbb{Z}^{m \times k}$.

For positive integers $m > n > k$, a prime q, and a real $\sigma > 2$, define $\mathbf{DIST}_0(n, m, k, q, \sigma)$ as:

1. For $i = 1, \ldots, k$, sample independent $\mathbf{e}_i \xleftarrow{\mathrm{R}} \mathcal{D}_{\mathbb{Z}^m,\sigma}$.
2. Choose a random $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ subject to the condition $\mathbf{A} \cdot \mathbf{e}_i = 0 \bmod q$ for all i.
3. Output $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$.

Define $\mathbf{DIST}_1(n, m, k, q, \sigma)$ as:

1. Sample a random matrix $\mathbf{A} \xleftarrow{\mathrm{R}} \mathbb{Z}_q^{n \times m}$.
2. For $i = 1, \ldots, k$ sample independent $\mathbf{e}_i \xleftarrow{\mathrm{R}} \mathcal{D}_{\Lambda_q^\perp(\mathbf{A}),\sigma}$.

3. Output $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$.

To prove Theorem 4.2, we will use the fact that for appropriate choices of parameters, the distributions $\mathbf{DIST}_0$ and $\mathbf{DIST}_1$ are statistically close.

Theorem 4.3. *Suppose $m \geq 2n \lg q$, $m > 2k$, and $\sigma > \omega(\sqrt{\log m})$. Then the distributions $\mathbf{DIST}_0(n, m, k, q, \sigma)$ and $\mathbf{DIST}_1(n, m, k, q, \sigma)$ are statistically close.*

Before proving this theorem, we need several preparatory lemmas.

Lemma 4.4. *Suppose $m \geq 2n \lg q$ and $\sigma > \omega(\sqrt{\log m})$. Then for all but a negligible fraction of $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, we have*

$$\rho_\sigma(\mathbb{Z}^m) = q^n \cdot \rho_\sigma(\Lambda_q^\perp(\mathbf{A})) \cdot (1 - \mathrm{negl}(n)).$$

Proof. By Lemma 3.4, there is a negligible $\epsilon(n)$ such that $\sigma > \eta_\epsilon(\Lambda_q^\perp(\mathbf{A}))$ for all but a q^{-n} fraction of $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$. By (3.1), for all such $\mathbf{A}$ and all $\mathbf{c} \in \mathbb{R}^n$, we have $\rho_{\sigma,\mathbf{c}}(\Lambda_q^\perp(\mathbf{A})) = \rho_\sigma(\Lambda_q^\perp(\mathbf{A}))(1 - \mathrm{negl}(n))$. If we choose a set of coset representatives $\mathbf{c}$ for $\mathbb{Z}^m / \Lambda_q^\perp(\mathbf{A})$, then we have

$$\rho_\sigma(\mathbb{Z}^m) = \sum_{\mathbf{c} \in \mathbb{Z}^m / \Lambda_q^\perp(\mathbf{A})} \rho_{\sigma,\mathbf{c}}(\Lambda_q^\perp(\mathbf{A})) = [\mathbb{Z}^m : \Lambda_q^\perp(\mathbf{A})] \cdot \rho_\sigma(\Lambda_q^\perp(\mathbf{A}))(1 - \mathrm{negl}(n)).$$

If $\mathrm{rank}(\mathbf{A}) = n$, then $[\mathbb{Z}^m : \Lambda_q^\perp(\mathbf{A})] = q^n$ and the result follows. Since $m > 2n$, the fraction of matrices $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ with $\mathrm{rank}(\mathbf{A}) < n$ is negligible (cf. Lemma A.2). □

Lemma 4.5. *Suppose $m \geq 2n \lg q$, $m > 2k$, $\sigma > \omega(\sqrt{\log m})$, and $q > \sigma \cdot \omega(\sqrt{\log m})$ with q prime. Let $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ be a matrix satisfying the conclusion of Lemma 4.4. Let $\mathbf{e}_1, \ldots, \mathbf{e}_k$ be vectors sampled from $\mathcal{D}_{\Lambda_q^\perp(\mathbf{A}),\sigma}$. Then with overwhelming probability, the set $\{\mathbf{e}_i\}$ has $\mathbb{Z}_q$-rank k.*

Proof. For $i = 1, \ldots, m$, let $\mathbf{E}_i \in \mathbb{Z}_q^{i \times m}$ be the matrix whose rows are the vectors $\mathbf{e}_j \bmod q$ for $j = 1, \ldots, i$. Then the probability that $\mathbf{e}_1, \ldots, \mathbf{e}_k$ are not $\mathbb{Z}_q$-linearly independent is at most

$$\sum_{i=0}^{k-1} \Pr[\mathbf{e}_{i+1} \in \Lambda_q(\mathbf{E}_i)] \leq \frac{1}{\rho_\sigma(\Lambda_q^\perp(\mathbf{A}))} \sum_{i=0}^{k-1} \rho_\sigma(\Lambda_q(\mathbf{E}_i)). \tag{4.1}$$

By [1, Lemma 31], we have $\rho_\sigma(\Lambda_q(\mathbf{E}_i)) \leq \rho_\sigma(\mathbb{Z}^i)/(1 - \epsilon)$, where $\epsilon = 2m \cdot \exp(-(\pi/4)(q/\sigma)^2)$. Using this result and Lemma 4.4, we see that the quantity (4.1) is bounded above by

$$\frac{q^n}{\rho_\sigma(\mathbb{Z}^m)(1 - \epsilon)} \sum_{i=0}^{k-1} \rho_\sigma(\mathbb{Z}^i). \tag{4.2}$$

By [1, Lemma 21] and the assumption $\sigma > \omega(\sqrt{\log m})$, there is a constant $\delta > 0$ such that for all i, we have $\sigma^i \leq \rho_\sigma(\mathbb{Z}^i) \leq \sigma^i(1 + \delta)$. Thus the quantity (4.2) is bounded above by

$$\frac{q^n(1 + \delta)}{\sigma^m(1 - \epsilon)} \left(\frac{\sigma^k - 1}{\sigma - 1} \right) \leq \delta' \cdot \frac{q^n}{\sigma^{m-k}},$$

for some constant δ'. Since $m \geq 2n \lg q$ and $k < m/2$, this last quantity is less than $\delta' \cdot q^{-n}$ whenever $\sigma > 4$. □

Proof of Theorem 4.3. We compute the statistical distance directly. First we note that if the vectors $\{\mathbf{e}_i\}$ chosen in $\mathbf{DIST}_0$ have $\mathbb{Z}_q$-rank ℓ, then there are $q^{n(m-\ell)}$ possible choices for $\mathbf{A}$, with each choice equally likely. We thus see that

$$\Pr\left[X = (\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k) : X \xleftarrow{\mathrm{R}} \mathbf{DIST}_0\right] = \left(\frac{1}{q}\right)^{n(m-\ell)} \prod_{i=1}^{k} \frac{\rho_\sigma(\mathbf{e}_i)}{\rho_\sigma(\mathbb{Z}^m)}.$$

On the other hand, a sample from $\mathbf{DIST}_1$ is nm independent uniform samples from $\mathbb{Z}_q$ and k independent samples from $\mathcal{D}_{\Lambda_q^\perp(\mathbf{A}),\sigma}$, and thus

$$\Pr\left[X = (\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k) : X \xleftarrow{\mathrm{R}} \mathbf{DIST}_1\right] = \left(\frac{1}{q}\right)^{nm} \prod_{i=1}^{k} \frac{\rho_\sigma(\mathbf{e}_i)}{\rho_\sigma(\Lambda_q^\perp(\mathbf{A}))}.$$

Let $S \subset \mathbb{Z}_q^{n\times m}$ be the set of matrices for which the conclusion of Lemma 4.4 holds, and for any $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ let $T_\mathbf{A} \subset \Lambda_q^\perp(\mathbf{A})^k$ be those sets of vectors $\{\mathbf{e}_i\}$ with $\mathbb{Z}_q$-rank k. We first calculate the distance between $\mathbf{DIST}_0$ and $\mathbf{DIST}_1$ when restricted to those tuples for which $\mathbf{A} \in S$ and $\{\mathbf{e}_i\} \in T_\mathbf{A}$:

$$\Delta_0 := \frac{1}{2} \sum_{\mathbf{A}\in S} \sum_{\{\mathbf{e}_i\}\in T_\mathbf{A}} \left| \left(\frac{1}{q}\right)^{n(m-k)} \prod_{i=1}^{k} \frac{\rho_\sigma(\mathbf{e}_i)}{\rho_\sigma(\mathbb{Z}^m)} - \left(\frac{1}{q}\right)^{nm} \prod_{i=1}^{k} \frac{\rho_\sigma(\mathbf{e}_i)}{\rho_\sigma(\Lambda_q^\perp(\mathbf{A}))} \right|.$$

Then by Lemma 4.4, we have

$$\Delta_0 = \frac{1}{2} \sum_{\mathbf{A}\in S} \frac{q^{nk}}{q^{nm}} \cdot \mathrm{negl}(n) \cdot \sum_{\{\mathbf{e}_i\}\in T_\mathbf{A}} \prod_{i=1}^{k} \frac{\rho_\sigma(\mathbf{e}_i)}{\rho_\sigma(\mathbb{Z}^m)}. \tag{4.3}$$

If we relax the restriction on the rank of the $\{\mathbf{e}_i\}$, then the total sum does not decrease, and the numerator of the last sum now includes all terms of the form $\prod_{i=1}^k \rho_\sigma(\mathbf{e}_i)$ with $\{\mathbf{e}_i\} \in \Lambda_q^\perp(\mathbf{A})^k$. Since the expression $(\sum_{\mathbf{e}\in\Lambda_q^\perp(\mathbf{A})} \rho_\sigma(\mathbf{e}))^k$ contains all of these terms and more, we have

$$\Delta_0 \le \sum_{\mathbf{A}\in S} \frac{q^{nk}}{q^{nm}} \cdot \mathrm{negl}(n) \cdot \left(\sum_{\mathbf{e}\in\Lambda_q^\perp(\mathbf{A})} \frac{\rho_\sigma(\mathbf{e})}{\rho_\sigma(\mathbb{Z}^m)} \right)^k = \sum_{\mathbf{A}\in S} \frac{q^{nk}}{q^{nm}} \cdot \mathrm{negl}(n) \cdot \left(\frac{\rho_\sigma(\Lambda_q^\perp(\mathbf{A}))}{\rho_\sigma(\mathbb{Z}^m)} \right)^k$$

By Lemma 4.4 and since $|S| \le q^{nm}$, we conclude that $\Delta_0 \le \mathrm{negl}(n)$.

Next, we claim that

$$\Pr[\mathbf{A} \in S \text{ and } \{\mathbf{e}_i\} \in T_\mathbf{A} : (\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k) \leftarrow \mathbf{DIST}_1] \ge 1 - \mathrm{negl}(n). \tag{4.4}$$

Given this claim, the theorem now follows from statement (2) of Lemma A.1, choosing the set A of the Lemma to be tuples with $\mathbf{A} \in S$ and $\{\mathbf{e}_i\} \in T_\mathbf{A}$ and using the fact that $\Delta_0 \le \mathrm{negl}(n)$.

To show (4.4), it suffices to show that both $\Pr[\mathbf{A} \notin S]$ and $\Pr[\mathbf{A} \in S \text{ and } \{\mathbf{e}_i\} \notin T_\mathbf{A}]$ are both negligible for tuples chosen from $\mathbf{DIST}_1$. The first quantity is negligible because all matrices $\mathbf{A}$ are equally likely to be chosen, and by Lemma 4.4 a negligible fraction of all matrices are not in S. The second quantity is negligible by Lemma 4.5. □

We can now prove our main theorem.

Proof of Theorem 4.2. Let $\mathcal{A}$ be an algorithm that takes as input a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and k vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k \in \mathbb{Z}^m$ and outputs a vector $\mathbf{e} \in \mathbb{Z}^m$. We construct an algorithm $\mathcal{B}$ that takes as input a matrix $\mathbf{B} \in \mathbb{Z}_q^{n \times (m-k)}$ and outputs a vector $\mathbf{w} \in \mathbb{Z}_q^{m-k}$.

At a high level, algorithm $\mathcal{B}$ begins by sampling k vectors $\mathbf{e}_i$ at random from a Gaussian over $\mathbb{Z}^m$. It then uses the SIS challenge $\mathbf{B}$ to create a random matrix $\mathbf{A}$ such that $\mathbf{A} \cdot \mathbf{e}_i = \mathbf{0} \bmod q$ for all i. By Theorem 4.3, the tuple $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$ created in this manner is statistically indistinguishable from a k-SIS challenge. Algorithm $\mathcal{B}$ can thus use the k-SIS solver $\mathcal{A}$ to find a short vector $\mathbf{e} \in \Lambda_q^{\perp}(\mathbf{A})$ and do Gaussian elimination over $\mathbb{Z}$ to find a vector in $\Lambda_q^{\perp}(\mathbf{A})$ whose last k entries are zero.

On a technical level, algorithm $\mathcal{B}$ works as follows:

1. Set $\mathbf{e}_i \leftarrow \mathsf{SampleGaussian}(\mathbb{Z}^m, \sigma)$ for $i = 1, \ldots, n$.
2. Let $\mathbf{E}$ be the $m \times k$ matrix whose columns are the vectors $\mathbf{e}_i$. If $\mathbf{E}$ has $\mathbb{Z}_q$-rank less than k, then abort (the simulation has failed). If the simulation does not fail, then without loss of generality[4] assume that the last k rows of $\mathbf{E}$ are linearly independent mod q.
3. Write $\mathbf{E} = \frac{\mathbf{F}}{\mathbf{G}}$, where $\mathbf{F} \in \mathbb{Z}^{(m-k) \times k}$, and $\mathbf{G} \in \mathbb{Z}^{k \times k}$ has determinant prime to q.
4. Set $\mathbf{U} \leftarrow (-\mathbf{B}) \cdot \mathbf{F} \cdot \mathbf{G}^{-1} \in \mathbb{Z}_q^{n \times k}$.
5. Set $\mathbf{A} \leftarrow \mathbf{B} \| \mathbf{U}$.
6. Run $\mathcal{A}$ on inputs $\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k$, and let $\mathbf{e} \in \mathbb{Z}^m$ be the output.
7. Write $\mathbf{e} = \mathbf{f} \| \mathbf{g}$ with $\mathbf{f} \in \mathbb{Z}^{m-k}$ and $\mathbf{g} \in \mathbb{Z}^k$.
8. Set $\mathbf{x} \leftarrow \det(\mathbf{G}) \cdot \mathbf{G}^{-1} \cdot \mathbf{g} \in \mathbb{Z}^k$.
9. Compute $\mathbf{w} \leftarrow \mathbf{F} \cdot \mathbf{x} - \det(\mathbf{G}) \cdot \mathbf{f} \in \mathbb{Z}^{m-k}$, and output $\mathbf{w}$.

We begin by observing that the selection of the $\mathbf{e}_i$ in Step (1) can be viewed as choosing m vectors from $\mathsf{SampleGaussian}(\mathbb{Z}^k, \sigma)$. If we partition these m vectors into $\lfloor m/k \rfloor$ sets of k vectors (plus some extras), then by [1, Theorem 30] the probability that any one of these sets has $\mathbb{Z}_q$-rank less than k is bounded above by some constant $\delta < 1$. Thus the probability that the matrix $\mathbf{E}$ has rank less than k is bounded above by $\delta^{\lfloor m/k \rfloor}$, which is negligible in n since $m/k \geq n$. Thus the probability that $\mathcal{B}$ aborts in Step (2) is negligible.

Since $\mathbf{E}$ has rank k with overwhelming probability, the distribution $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$ produced by the simulator is statistically close to $\mathbf{DIST}_0(n, m, k, q, \sigma)$. By Theorem 4.3, this distribution is statistically close to that of $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k)$ in a k-SIS challenge. Thus even a computationally unbounded adversary cannot tell if it is interacting with a real k-SIS challenge or with our simulation, except with negligible probability.

To conclude the proof, it suffices to show that $\mathbf{w}$ is a solution to the SIS problem for $\mathbf{B}$; namely, that (a) $\mathbf{w}$ is nonzero, (b) $\mathbf{B} \cdot \mathbf{w} = \mathbf{0} \bmod q$, and (c) $\|\mathbf{w}\| \leq \beta'$.

To show (a), first observe that $\mathbf{w} = \mathbf{0}$ if and only if $\mathbf{F} \cdot \mathbf{G}^{-1} \cdot \mathbf{g} = \mathbf{f}$ in $\mathbb{Q}^{m-k}$. Let $\mathbf{y} = \mathbf{G}^{-1} \cdot \mathbf{g} \in \mathbb{Q}^k$. If $\mathbf{w} = \mathbf{0}$, then $\mathbf{f} = \mathbf{F} \cdot \mathbf{y}$ and $\mathbf{g} = \mathbf{G} \cdot \mathbf{y}$, and therefore $\mathbf{E} \cdot \mathbf{y} = \mathbf{e}$. Thus $\mathbf{e}$ is a $\mathbb{Q}$-linear combination of the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$, contradicting the fact that $\mathbf{e}$ is a solution the k-SIS challenge.

To show (b), observe that since $\mathbf{e} = \mathbf{f} \| \mathbf{g}$ is a solution to the k-SIS challenge, we have $\mathbf{A} \cdot \mathbf{e} = \mathbf{B} \cdot \mathbf{f} + \mathbf{U} \cdot \mathbf{g} = \mathbf{0} \bmod q$. The construction of $\mathbf{U}$ then implies that $\mathbf{B} \cdot \mathbf{f} = \mathbf{B} \cdot \mathbf{F} \cdot \mathbf{G}^{-1} \cdot \mathbf{g} \bmod q$. It follows that

$$\mathbf{B} \cdot \mathbf{w} = \det(\mathbf{G})(\mathbf{B} \cdot \mathbf{F} \cdot \mathbf{G}^{-1} \cdot \mathbf{g} - \mathbf{B} \cdot \mathbf{f}) = 0 \bmod q.$$

[4] More precisely, we apply a permutation π to the rows of $\mathbf{E}$ to obtain a matrix $\mathbf{E}'$ whose last k rows have $\mathbb{Z}_q$-rank k, and we apply π^{-1} to the columns of the matrix $\mathbf{A}$ produced in Step (5).

Finally, we bound the length of $\mathbf{w}$. By a standard tail inequality [16, Lemma 4.2], the absolute value of each entry of $\mathbf{E}$ is less than $t\sigma$ with overwhelming probability. Furthermore, since $\|\mathbf{e}\| \leq \beta$ we know that each entry of $\mathbf{e}$ has absolute value bounded by β. Since $\mathbf{G} \cdot \mathbf{x} = \det(\mathbf{G}) \cdot \mathbf{g}$, Cramer's rule [12, Ch. 11, Theorem 26] implies that the ith entry of $\mathbf{x}$ is the determinant of the matrix constructed by replacing the ith column of $\mathbf{G}$ with the vector $\mathbf{g}$. There are $k!$ terms in this determinant, each of which consists of a product of $k-1$ entries from $\mathbf{G}$ and one entry from $\mathbf{g}$. Thus with overwhelming probability, each entry of $\mathbf{x}$ is bounded in absolute value by $\beta \cdot k!(t\sigma)^{k-1}$. It follows that each entry of $\mathbf{F} \cdot \mathbf{x}$ is bounded by $\beta \cdot k \cdot k!(t\sigma)^k$, and thus $\|\mathbf{F} \cdot \mathbf{x}\| \leq \beta \cdot k^{3/2} \cdot k!(t\sigma)^k$. Since $|\det(\mathbf{G})| \leq k!(t\sigma)^k$ with overwhelming probability, we have $\|\det(\mathbf{G}) \cdot \mathbf{f}\| \leq \beta \cdot k!(t\sigma)^k$. We conclude that $\|\mathbf{w}\| \leq \beta \cdot (k^{3/2}+1)k!(t\sigma)^k$ with overwhelming probability. □

Our linearly homomorphic signature scheme will rely on properties of the signature vectors mod 2. We will need the following result, which shows that for small ℓ and appropriate choices of parameters, a sample from $\mathcal{D}_{\Lambda_q^\perp(\mathbf{A}),\sigma}$ looks uniformly random mod ℓ.

Proposition 4.6. *Let q and ℓ be relatively prime integers, let $m > 2n \lg q$, and let $\sigma > \ell \cdot \omega(\sqrt{\log m})$. Suppose $(\mathbf{A}, \mathbf{e}_1, \ldots, \mathbf{e}_k) \in \mathbb{Z}_q^{n \times m} \times \mathbb{Z}^{m \times k}$ is a tuple selected from the distribution of k-$\mathsf{SIS}_{q,m,\beta,\sigma}$ challenge instances (for any β). Then the distribution of $(\mathbf{A}, \mathbf{e}_1 \bmod \ell, \ldots, \mathbf{e}_k \bmod \ell) \in \mathbb{Z}_q^{n \times m} \times \mathbb{Z}_\ell^{m \times k}$ is statistically close to the uniform distribution on $\mathbb{Z}_q^{n \times m} \times \mathbb{Z}_\ell^{m \times k}$.*

Proof. For any $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, let $\Lambda'(A) := \Lambda_q^\perp(\mathbf{A}) \cap (\ell\mathbb{Z})^m$. We first claim that $\Lambda'(\mathbf{A}) = \ell\Lambda_q^\perp(\mathbf{A})$. To see this, observe that $\Lambda'(\mathbf{A})$ is contained in both $\Lambda_q^\perp(\mathbf{A})$ and $(\ell\mathbb{Z})^m$, and $\ell\Lambda_q^\perp(\mathbf{A})$ is contained in $\Lambda'(\mathbf{A})$. Let $r = \operatorname{rank}(\mathbf{A})$; then $\det(\Lambda_q^\perp(\mathbf{A})) = q^r$. It follows that $q^r \mid \det(\Lambda'(\mathbf{A}))$ and $\ell^m \mid \det(\Lambda'(\mathbf{A}))$, and $\det(\Lambda'(\mathbf{A})) \mid \ell^m q^r$. Since q and ℓ are relatively prime, we conclude that $\det(\Lambda'(\mathbf{A})) = \ell^m q^\ell$, and the claim follows.

Since $\Lambda'(\mathbf{A}) = \ell\Lambda_q^\perp(\mathbf{A})$, we have $\eta_\epsilon(\Lambda'(\mathbf{A})) = \ell \cdot \eta_\epsilon(\Lambda_q^\perp(\mathbf{A}))$. By Lemma 3.4, there is a negligible $\epsilon(n)$ such that for all but at most a q^{-n} fraction of $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ we have $\sigma > \eta_\epsilon(\Lambda_q^\perp(\mathbf{A}))$. By [16, Corollary 2.8], for all such $\mathbf{A}$ the distribution of the $\mathbf{e}_i$ is statistically close to uniform over $\Lambda_q^\perp(\mathbf{A})/\Lambda'(\mathbf{A}) = \Lambda_q^\perp(\mathbf{A})/\ell\Lambda_q^\perp(\mathbf{A}) \cong \mathbb{Z}_\ell^m$, and the proposition follows for all $\mathbf{A}$ satisfying the bound $\eta_\epsilon(\Lambda_q^\perp(\mathbf{A})) < \sigma$. Since all matrices $\mathbf{A}$ are equally likely and all but a q^{-n} fraction satisfy this bound, the proposition follows. □

4.2 Tight Bounds on the Length of Gaussian Samples

Signatures in our schemes will be "short" vectors sampled from Gaussian distributions over cosets of a particular lattice. Signatures will be accepted as valid if and only if they are sufficiently short and satisfy some congruence condition. To quantify what "short" means, we must demonstrate an upper bound on the length of a vector sampled from a Gaussian.

Micciancio and Regev [21, Lemma 4.4] show that if σ is larger than the smoothing parameter of the n-dimensional lattice Λ, then with overwhelming probability the length of a vector sampled from $\mathcal{D}_{\Lambda,\sigma,\mathbf{c}}$ is at most $\sigma\sqrt{n}$. However, they also show that the *expected* length of such a sample is at most $\sigma\sqrt{n/2\pi} + \operatorname{negl}(n)$. Our result below "bridges the gap" between the factor of 1 in the upper bound and the factor of $1/\sqrt{2\pi}$ in the expected length. Furthermore, we show an equally strong *lower* bound on the length of the Gaussian sample.

Proposition 4.7. *Let $\Lambda \subset \mathbb{R}^n$ be a lattice. Let $\mathbf{T}$ be a basis for Λ and suppose $\sigma = \operatorname{poly}(n)$ with $\sigma \geq \|\widetilde{\mathbf{T}}\| \cdot \omega(\sqrt{\log n})$. Let $\mathbf{c} \in \mathbb{R}^m$ be any vector. Then for any constant $\epsilon > 0$ we have*

$$\Pr\left[(1-\epsilon)\sigma\sqrt{n/2\pi} \;\leq\; \|\mathbf{x} - \mathbf{c}\| \;\leq (1+\epsilon)\sigma\sqrt{n/2\pi} : x \xleftarrow{\mathrm{R}} \mathcal{D}_{\Lambda,\sigma,\mathbf{c}}\right] \;\geq\; 1 - \operatorname{negl}(n)$$

Proof. Fix a constant $\delta > 0$, and suppose $\mathbf{y} \leftarrow D_{\delta\sigma}$ is sampled from a continuous Gaussian centered at the origin with parameter $\delta\sigma$. Then by [26, Claim 3.9] (using [16, Lemma 3.1] to bound the smoothing parameter), the distribution on $\mathbf{x} + \mathbf{y}$ is statistically close to the continuous Gaussian $D_{\tau,\mathbf{c}}$ centered at $\mathbf{c}$ with parameter $\tau := \sigma\sqrt{1+\delta^2}$. By Lemma A.3, for $\hat{\mathbf{x}}$ sampled from $D_{\tau,\mathbf{c}}$ and any fixed $\gamma_1 < 1$ and $\gamma_2 > 1$, we have

$$\Pr\left[\gamma_1 \cdot \tau\sqrt{n/2\pi} \;\le\; \|\hat{\mathbf{x}} - \mathbf{c}\| \;\le\; \gamma_2 \cdot \tau\sqrt{n/2\pi}\right] \;\ge\; 1 - \mathrm{negl}(n). \tag{4.5}$$

If we apply (4.5) to $\hat{\mathbf{x}} \approx \mathbf{x} + \mathbf{y}$ sampled from $D_{\tau,\mathbf{c}}$ and to $\hat{\mathbf{x}} = \mathbf{y}$ sampled from $D_{\delta\sigma,\mathbf{0}}$, then the triangle inequality implies that with overwhelming probability we have

$$\left(\gamma_1\sqrt{1+\delta^2} - \gamma_2\delta\right)\sigma\sqrt{n/2\pi} \;\le\; \|\mathbf{x} - \mathbf{c}\| \;\le \left(\gamma_2\sqrt{1+\delta^2} + \gamma_1\delta\right)\sigma\sqrt{n/2\pi}.$$

The lemma now follows by choosing δ sufficiently close to 0 and γ_1, γ_2 sufficiently close to 1 so that

$$\gamma_1\sqrt{1+\delta^2} - \gamma_2\delta > 1 - \epsilon \quad \text{and} \quad \gamma_2\sqrt{1+\delta^2} + \gamma_1\delta < 1 + \epsilon.$$

□

4.3 Removing Linear Independence from the k-SIS Problem

In this section we show that for small values of k and tight length bounds, we can relax the linear independence condition in the statement of the k-SIS problem. Specifically, if we can find *any* nonzero vector $\mathbf{e}^*$ of the required length not equal to $\pm\mathbf{e}_i$ for any of the k vectors $\mathbf{e}_i$ in the problem statement, then with overwhelming probability $\mathbf{e}^*$ is not in the linear span of the $\mathbf{e}_i$.

Our main result is that if we sample about $k < n^{1/4}$ short vectors $\mathbf{e}_i$ from a Gaussian distribution on a lattice, then the probability that there is an additional short vector in the k-dimensional sublattice spanned by the $\mathbf{e}_i$ is negligible.

Proposition 4.8. *Let $\Lambda \subset \mathbb{Z}^n$ be a lattice. Let $\mathbf{T}$ be a basis for Λ and suppose $\sigma = \mathrm{poly}(n)$ with $\sigma \ge \|\widetilde{\mathbf{T}}\| \cdot \omega(\sqrt{\log n})$. Let $k = k(n)$ be such that $2k \cdot \omega(\log^{1/4} n) < n^{1/4}$ and choose $\mathbf{e}_1, \ldots, \mathbf{e}_k \leftarrow \mathcal{D}_{\Lambda,\sigma}$. Then with overwhelming probability, the only nonzero vectors of length at most $1.1 \cdot \sigma\sqrt{n/2\pi}$ in $\mathbb{Z}$-span$(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$ are the vectors $\pm\mathbf{e}_i$ for $i = 1, \ldots, k$.*

Before proving Proposition 4.8 we first state a simple corollary of Lemma 5.1 from [25].

Lemma 4.9. *Let $\Lambda \subset \mathbb{Z}^n$ be a lattice and $\mathbf{v} \in \mathbb{R}^n$ a vector of norm 1. Then for $\sigma > 0$ and $\mathbf{e}$ sampled from $\mathcal{D}_{\Lambda,\sigma}$, the inner product $\langle \mathbf{v}, \mathbf{e}\rangle$ satisfies*

$$\Pr\left[\ \left|\langle \mathbf{v}, \mathbf{e}\rangle\right| > \sigma r\ \right] < 2e^{-\pi r^2}.$$

In particular,

$$\Pr\left[\ \left|\langle \mathbf{v}, \mathbf{e}\rangle\right| > \sigma \cdot \omega(\sqrt{\log n})\ \right] < \mathrm{negl}(n)$$

Proof. Define the rescaled lattice $\Lambda' = \Lambda/\sigma$ and observe that $\Pr[\mathcal{D}_{\Lambda',1} = \mathbf{e}]$ is the same as $\Pr[\mathcal{D}_{\Lambda,\sigma} = \sigma\mathbf{e}]$ for all $\mathbf{e} \in \mathbb{R}^n$. Therefore it suffices to prove the lemma for $\sigma = 1$ which follows directly from [25, Lemma 5.1] (taking $d = 1$ and $c = 0$). The general case follows by scaling the lattice by a factor of σ. □

Proof of Proposition 4.8. By Proposition 4.7 and Lemma 4.9, we know that with overwhelming probability

$$|\langle \mathbf{e}_i, \mathbf{e}_j \rangle| = \|\mathbf{e}_i\| \cdot \left|\langle \tfrac{\mathbf{e}_i}{\|\mathbf{e}_i\|}, \mathbf{e}_j \rangle\right| \leq 1.2\, \sigma^2 \sqrt{n/2\pi} \cdot \omega(\sqrt{\log n}) \qquad \text{for all } i \neq j. \tag{4.6}$$

Now, let $\mathbf{e}$ be a nonzero vector in $\mathbb{Z}$-span$(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$, namely $\mathbf{e} = \sum_{i=1}^k c_i \mathbf{e}_i$ for some $c_i \in \mathbb{Z}$. Then

$$\|\mathbf{e}\|^2 = \langle \mathbf{e}, \mathbf{e} \rangle = \sum_{i=1}^k c_i^2 \|\mathbf{e}_i\|^2 + \sum_{1 \leq i < j \leq k} 2 c_i c_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle \tag{4.7}$$

Let $\mathbf{c} := (c_1, \ldots, c_k)$ and define $\|\mathbf{c}\|_\infty = \max(\{|c_i|\}_{i=1}^k)$. Then by (4.6) and since $k^2\sqrt{n} \cdot \omega(\log^{1/2} n) < n/4$ we have

$$\left| \sum_{1 \leq i < j \leq k} 2 c_i c_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle \right| \leq k^2 \|c\|_\infty^2 (1.2\, \sigma^2 \sqrt{n/2\pi}) \cdot \omega(\log^{1/2} n) \leq 0.3\, \|\mathbf{c}\|_\infty^2 \frac{\sigma^2 n}{2\pi}.$$

Similarly, by Proposition 4.7 we know that with overwhelming probability

$$\sum_{i=1}^k c_i^2 \|\mathbf{e}_i\|^2 \geq \sum_{i=1}^k c_i^2 \left(0.9\, \frac{\sigma^2 n}{2\pi}\right) = 0.9\, \frac{\sigma^2 n}{2\pi} \|\mathbf{c}\|^2.$$

Plugging the last two bounds into (4.7) we obtain

$$\|\mathbf{e}\|^2 \geq \frac{\sigma^2 n}{2\pi} \left(0.9\, \|\mathbf{c}\|^2 - 0.3\, \|c\|_\infty^2\right)$$

Now, if $\mathbf{c} \in \mathbb{Z}^k$ has more than one non-zero coordinate, then the quantity $0.9\|\mathbf{c}\|^2 - 0.3\|c\|_\infty^2$ is greater than 1.21 (the minimum case is the vector $(1, 1, 0, \ldots, 0)$, for which the difference is 1.5). Therefore, if $\mathbf{c}$ has more than one non-zero coordinate the vector $\mathbf{e}$ is longer than $1.1 \cdot \sigma\sqrt{n/2\pi}$. Finally, if $\mathbf{e}$ has one non-zero coordinate, then by Proposition 4.7 that coordinate must be 1 or -1 if $\mathbf{e}$ is to have length at most $\sigma\sqrt{n/2\pi}$. □

While Proposition 4.8 shows that there is no additional short vector in the *integer* span of k short vectors sampled from a Gaussian, a short integer vector that is in the *rational* span of the given vectors is also a valid solution to the k-SIS problem. our signature scheme. The following lemma shows that with overwhelming probability, there is no such vector that is not already in the $\mathbb{Z}$-span.

Lemma 4.10. *Suppose $\sigma > f(n)$ and $k < \sigma/g(n)$ for some $\omega(\sqrt{\log n})$ functions f, g. Let $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ be chosen from $(\mathcal{D}_{\mathbb{Z}^n,\sigma})^k$. Then with overwhelming probability we have*

$$\mathbb{Z}^n \cap \mathbb{Q}\text{-span}(S) = \mathbb{Z}\text{-span}(S).$$

Proof. Suppose there is an integer vector $\mathbf{v}$ in the $\mathbb{Q}$-span of S but not in the $\mathbb{Z}$-span of S. Then there is a prime p such that the set $\{\mathbf{e}_i \bmod p\}$ is linearly dependent over $\mathbb{F}_p$, so it suffices to show that for all primes p, the set S is linearly independent mod p with overwhelming probability.

Next note that the set S is determined by nk independent samples from the one-dimensional Gaussian $\mathcal{D}_{\mathbb{Z},\sigma}$. For a given $a \in \mathbb{Z}$, the probability that a sample from $\mathcal{D}_{\mathbb{Z},\sigma}$ is congruent to a mod p is

$$\frac{\rho_{\sigma,-a}(p\mathbb{Z})}{\rho_{\sigma,0}(\mathbb{Z})} = \frac{\rho_{\sigma/p,-a/p}(\mathbb{Z})}{\rho_{\sigma,0}(\mathbb{Z})} < \frac{\int_{-\infty}^{\infty} e^{-\pi p^2 x^2/\sigma^2} dx + 2}{\int_{-\infty}^{\infty} e^{-\pi x^2/\sigma^2} dx - 1} = \frac{\frac{1}{p} + \frac{2}{\sigma}}{1 - \frac{1}{\sigma}},$$

which is close to $1/p$. Since $\sigma > \omega(\sqrt{\log n})$, if n is sufficiently large, then this probability is less than $1/2k$ for all $p > 2k$.

Now let $\mathbf{E}$ be the matrix whose rows are the vectors $\mathbf{e}_i$. Suppose $p > 2k$ and consider a $k \times k$ submatrix of $\mathbf{E}$. Pick an entry of this matrix and assume that the determinant of the associated $(k-1) \times (k-1)$ minor is nonzero mod p. Then there is a unique value $a \bmod p$ of the remaining entry that makes the determinant of the $k \times k$ submatrix zero mod p. By our analysis above, the probability that this value is a mod p is at most $1/2k$. Applying the same argument inductively to the $(k-1) \times (k-1)$ minor along with a union bound, we see that the probability that the determinant is zero mod p is at most $1/2$. Since there are $\lfloor n/k \rfloor = O(n)$ independent $k \times k$ submatrices of $\mathbf{E}$ and all must have determinant zero mod p if S is linearly dependent mod p, the probability that S is linearly dependent mod p is negligible in n.

Finally, suppose $p \leq 2k$. Since $\sigma > p \cdot \omega(\sqrt{\log n})$, it follows from [16, Corollary 2.8] that the distribution of $\mathbf{e}_i \bmod p\mathbb{Z}^n$ is statistically close to uniform. Thus by Lemma A.2, the probability that $\mathbf{E}$ has rank less than k is at most $1/p^{n-k} + \mathrm{negl}(n)$, which is negligible in n. □

4.4 Linear Combinations of Discrete Gaussians

The privacy property of our linearly homomorphic scheme will follow from the fact that the distribution obtained by summing independent discrete Gaussians is itself a discrete Gaussian distribution that depends only on the coset of the sum and the linear combination being computed. We first prove this statement for a sum of two vectors; the more general statement follows by induction.

Lemma 4.11. *Let $\Lambda_1, \Lambda_2 \subset \mathbb{R}^m$ be two lattices and $\sigma_1, \sigma_2 \in \mathbb{R}$. Let X and Y be independent random variables distributed as $\mathcal{D}_{\Lambda_1+\mathbf{t}_1,\sigma_1}$ and $\mathcal{D}_{\Lambda_2+\mathbf{t}_2,\sigma_2}$ respectively. Define*

$$\tau := \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1/2}$$

and suppose that $\tau > \eta_\epsilon(\Lambda_1 \cap \Lambda_2)$ for some negligible ϵ. Then the random variable $Z := X + Y$ is sampled from a distribution statistically close to $\mathcal{D}_{(\Lambda_1+\mathbf{t}_1)+(\Lambda_2+\mathbf{t}_2),\sqrt{\sigma_1^2+\sigma_2^2}}$.

Proof. Let $\mathbf{z}$ be some vector in $(\Lambda_1 + \mathbf{t}_1) + (\Lambda_2 + \mathbf{t}_2)$. This means that there are $\mathbf{t}_1' \in \Lambda_1 + \mathbf{t}_1$ and $\mathbf{t}_2' \in \Lambda_2 + \mathbf{t}_2$ such that $\mathbf{t}_1' + \mathbf{t}_2' = \mathbf{z}$. Now, the set

$$\left\{ (\mathbf{t}_1' + \mathbf{s},\ \mathbf{t}_2' - \mathbf{s}) \ :\ \mathbf{s} \in \Lambda_1 \cap \Lambda_2 \right\} \subseteq (\Lambda_1 + \mathbf{t}_1) \times (\Lambda_2 + \mathbf{t}_2)$$

consists of all pairs of vectors in $\mathbb{Z}^m$ that sum up to $\mathbf{z}$ where the first element is in $\Lambda_1 + \mathbf{t}_1$ and the second element is in $\Lambda_2 + \mathbf{t}_2$. Therefore, setting $\Lambda := \Lambda_1 \cap \Lambda_2$ and $w := \rho_{\sigma_1}(\Lambda_1 + \mathbf{t}_1)^{-1}\rho_{\sigma_2}(\Lambda_2 + \mathbf{t}_2)^{-1}$ we obtain:

$$\begin{aligned}
\Pr[Z = \mathbf{z}] &= \sum_{\mathbf{y} \in \Lambda + \mathbf{t}_2'} \Pr[Y = \mathbf{y}] \cdot \Pr[X = \mathbf{z} - \mathbf{y}] \\
&= \sum_{\mathbf{y} \in \Lambda + \mathbf{t}_2'} \exp\left(-\pi \frac{\sigma_1^2\|\mathbf{y}\|^2 + \|\mathbf{z}-\mathbf{y}\|^2\sigma_2^2}{\sigma_1^2\sigma_2^2}\right) \cdot w \\
&= \sum_{\mathbf{y} \in \Lambda + \mathbf{t}_2'} \exp\left(-\pi \frac{\|\mathbf{y} - \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\,\mathbf{z}\|^2}{\tau^2} - \pi \frac{\|\mathbf{z}\|^2}{\sigma_1^2+\sigma_2^2}\right) \cdot w \\
&= e^{-\pi\|\mathbf{z}\|^2/(\sigma_1^2+\sigma_2^2)} \cdot \rho_\tau\left(\Lambda + \mathbf{t}_2' - \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\,\mathbf{z}\right) \cdot w
\end{aligned}$$

where τ is defined in the statement of the lemma. Now, using equation (3.1) we obtain that for all $\mathbf{z}$ in $(\Lambda_1 + \mathbf{t}_1) + (\Lambda_2 + \mathbf{t}_2)$:

$$\Pr[Z = \mathbf{z}] = e^{-\pi|\mathbf{z}|^2/(\sigma_1^2+\sigma_2^2)} \cdot \rho_\tau(\Lambda) \cdot (1-\delta) \cdot w = \rho_{\sqrt{\sigma_1^2+\sigma_2^2}}(\mathbf{z}) \cdot (1-\delta) \cdot w'$$

for some $\delta = \delta(\mathbf{z}) \in [0, \frac{2\epsilon}{1+\epsilon}]$ and a scalar $w' := w \cdot \rho_\tau(\Lambda)$ that is independent of $\mathbf{z}$. Since ϵ is negligible it follows that Z is statistically close to $\mathcal{D}_{(\Lambda_1+\mathbf{t}_1)+(\Lambda_2+\mathbf{t}_2),\sqrt{\sigma_1^2+\sigma_2^2}}$. □

Theorem 4.12. *Let $\Lambda \subseteq \mathbb{Z}^m$ be a lattice and $\sigma \in \mathbb{R}$. For $i = 1, \ldots, k$ let $\mathbf{t}_i \in \mathbb{Z}^n$ and let X_i be mutually independent random variables sampled from $\mathcal{D}_{\Lambda+\mathbf{t}_i,\sigma}$. Let $\mathbf{c} \in \mathbb{Z}^k$ and define*

$$\tau := \|\mathbf{c}\|\sigma, \quad \ell := \mathrm{lcm}(c_1, \ldots, c_k), \quad g := \gcd(c_1, \ldots, c_k), \quad b := \left(\frac{1}{c_1^2} + \cdots + \frac{1}{c_k^2}\right)^{-1/2}.$$

Suppose that $\sigma \cdot \ell^{-1} b > \eta_\epsilon(\Lambda)$ for some negligible ϵ. Then $Z = \sum_{i=1}^k c_i X_i$ is statistically close to $\mathcal{D}_{g\Lambda+\mathbf{t},\tau}$ where $\mathbf{t} = \sum_{i=1}^k c_i \mathbf{t}_i$.

Proof. We induct on k. The base case is trivial: multiplying a sample from $\mathcal{D}_{\Lambda+\mathbf{t}_1,\sigma}$ by c_1 is the same as sampling from $\mathcal{D}_{c_1\Lambda+c_1\mathbf{t}_1,c_1\sigma}$. The inductive step follows from Lemma 4.11, using the fact that $\bigcap_{i=1}^k c_i\Lambda = \ell\Lambda$ and $\sum_{i=1}^k c_i\Lambda = g\Lambda$. □

5 A Linearly Homomorphic Signature Scheme over $\mathbb{F}_2$

We now describe our linearly homomorphic signature scheme over $\mathbb{F}_2$. Our construction is inspired by the signature scheme of Gentry, Peikert, and Vaikuntanathan [16]. In the GPV scheme, signatures are short vectors in $\Lambda_q^{\mathbf{u}}(\mathbf{A})$, where $\mathbf{u}$ is the hash of the message to be signed. The key idea in our construction of homomorphic signatures is to work simultaneously modulo 2 and modulo an odd prime q. Specifically, a signature on a vector $\mathbf{v} \in \mathbb{F}_2^n$ is a short vector $\mathbf{e} \in \mathbb{Z}^m$ such that $\mathbf{e}$ is in both $\Lambda_q^\perp(\mathbf{A})$ and $\Lambda_2^{\mathbf{v}}(\mathbf{A})$. The mod 2 part ties the signature to the message, while the mod q part ensures that the signature cannot be forged. By the Chinese remainder theorem, such a vector $\mathbf{e}$ is in the lattice $\Lambda_{2q}^{q\cdot\mathbf{v}}(\mathbf{A})$.

In order to be able to sign multiple files, the matrix $\mathbf{A}$ must be different for every file, yet still have a trapdoor that allows us to generate signatures using the SamplePre algorithm. To achieve this, we divide $\mathbf{A}$ into two parts. The left half is a public matrix generated by the TrapGen algorithm, while the right half depends on the identifier of the file being signed. Given the secret basis output by TrapGen, we can use the ExtBasis algorithm to compute a short basis for $\Lambda_{2q}^\perp(\mathbf{A})$.

Our scheme is as follows:

$\mathsf{Setup}(n, \mathsf{params})$. Given a security parameter n and parameters $\mathsf{params} = (N, k, L, m, q, \sigma)$, where $N = n$ is the dimension of vectors to be signed, $k < n$ is the dimension of subspaces to be signed, $L \geq 1$ is the maximum number of linear combinations that can be authenticated, $m(n, L) > n$ is an integer, $q(n, L)$ is an odd prime, and $\sigma(n, L)$ is a real number, do the following:

1. Run $\mathsf{TrapGen}(n, m, 2q)$ to generate a matrix $\mathbf{A} \in \mathbb{Z}_{2q}^{n\times m}$ and a basis $\mathbf{T}$ of $\Lambda_{2q}^\perp(\mathbf{A})$ such that $\|\widetilde{\mathbf{T}}\| \leq 30\sqrt{n \lg 2q}$.
2. Let $H : \{0,1\}^* \to \mathbb{Z}_{2q}^{n\times m}$ be a hash function, viewed as a random oracle.
3. Output the public key $\mathsf{pk} \leftarrow (\mathbf{A}, H)$, and the private key $\mathsf{sk} \leftarrow (\mathbf{A}, H, \mathbf{T})$.

$\mathsf{Sign}(\mathsf{sk}, \mathsf{id}, \mathbf{v})$. Given secret key $\mathsf{sk} = (\mathbf{A}, H, \mathbf{T})$, identifier $\mathsf{id} \in \{0,1\}^n$, and a vector $\mathbf{v} \in \mathbb{F}_2^n$, do the following:

1. Set $\mathbf{B} \leftarrow \mathbf{A} \| H(\mathsf{id}) \in \mathbb{Z}_{2q}^{n \times 2m}$.
2. Let $\mathbf{S} \leftarrow \mathsf{ExtBasis}(\mathbf{T}, \mathbf{B})$ be a basis for $\Lambda_{2q}^{\perp}(\mathbf{B})$ with $\|\widetilde{\mathbf{S}}\| = \|\widetilde{\mathbf{T}}\|$.
3. Output $\mathbf{e} \leftarrow \mathsf{SamplePre}(\mathbf{B}, \mathbf{S}, \sigma, q \cdot \mathbf{v})$.

$\mathsf{Combine}(\mathsf{pk}, \mathsf{id}, \{(\alpha_i, \mathbf{e}_i)\}_{i=1}^{\ell})$. Given a public key pk, an identifier id, and pairs $\{(\alpha_i, \mathbf{e}_i)\}_{i=1}^{\ell}$ with $\alpha_i \in \mathbb{F}_2 = \{0,1\}$, output $\mathbf{e} \leftarrow \sum_{i=1}^{\ell} \alpha_i \mathbf{e}_i \in \mathbb{Z}^{2m}$.

$\mathsf{Verify}(\mathsf{pk}, \mathsf{id}, \mathbf{y}, \mathbf{e})$. Given a public key $\mathsf{pk} = (\mathbf{A}, H)$, an identifier id, a signature $\mathbf{e} \in \mathbb{Z}^{2m}$, and a vector $\mathbf{y} \in \mathbb{F}_2^n$, do the following:

1. Set $\mathbf{B} \leftarrow \mathbf{A} \| H(\mathsf{id}) \in \mathbb{Z}_{2q}^{n \times 2m}$.
2. If (a) $\|\mathbf{e}\| \le L \cdot \sigma\sqrt{m}$ and (b) $\mathbf{B} \cdot \mathbf{e} = q \cdot \mathbf{y} \bmod 2q$, output 1. Otherwise output 0.

Proposition 5.1. *Suppose $\sigma \ge 30\sqrt{n \lg 2q} \cdot \omega(\sqrt{\log n})$. Then the scheme described above is an L-limited linearly homomorphic signature scheme over $\mathbb{F}_2$.*

Proof. We must show that the correctness conditions of Definition 2.1 hold, with (2) holding for all $\ell \le L$ and σ_i produced by the Sign algorithm. By Theorem 3.3 (b), the vector $\mathbf{e}$ output by the Sign algorithm satisfies $\mathbf{B} \cdot \mathbf{e} = q \cdot \mathbf{v} \bmod 2q$ and is drawn from a distribution statistically close to $\mathcal{D}_{\Lambda_{2q}^{q \cdot \mathbf{v}}(\mathbf{B}), \sigma}$. By Proposition 4.7 (with $\epsilon = 1$), we have $\|\mathbf{e}\| \le \sigma\sqrt{m}$ with overwhelming probability. It follows that if $\mathbf{e}$ is output by $\mathsf{Sign}(\mathsf{sk}, \mathsf{id}, \mathbf{v})$, then $\mathsf{Verify}(\mathsf{pk}, \mathsf{id}, \mathbf{v}, \mathbf{e}) = 1$.

Since the coefficients α_i of linear combinations of messages are in $\{0,1\}$, the length of the vector $\mathbf{e}$ output by $\mathsf{Combine}(\mathsf{pk}, \mathsf{id}, \{(\alpha_i, \mathbf{e}_i)\}_{i=1}^{\ell})$ when given signatures $\mathbf{e}_i$ output by Sign is at most $\ell\sigma\sqrt{m}$, so this vector passes verification test (a) whenever $\ell \le L$. As for verification test (b), suppose we have vectors $\mathbf{v}_i$ such that $\mathsf{Verify}(\mathsf{pk}, \mathsf{id}, \mathbf{v}_i, \mathbf{e}_i) = 1$ for all i. Since q is odd, this implies that $\mathbf{B} \cdot \mathbf{e}_i = 0 \bmod q$ and $\mathbf{B} \cdot \mathbf{e}_i = \mathbf{v}_i \bmod 2$ for all i. It follows that $\mathbf{B} \cdot \mathbf{e} = 0 \bmod q$ and $\mathbf{B} \cdot \mathbf{e} = \sum \alpha_i \mathbf{v}_i \bmod 2$, and therefore $\mathbf{B} \cdot \mathbf{e} = q \cdot \sum \alpha_i \mathbf{v}_i \bmod 2q$. □

5.1 Unforgeability

We prove unforgeability of our linearly homomorphic signature scheme over $\mathbb{F}_2$ in the random oracle model. Given an adversary that breaks the signature scheme over $\mathbb{Z}_{2q}$, we construct an adversary that simulates the signature scheme and the hash function H and solves the k-SIS problem over $\mathbb{Z}_q$. By Theorem 4.2, this adversary can in turn be used to solve the SIS problem over $\mathbb{Z}_q$.

Our simulation begins by guessing which of the adversary's signature and hash queries will correspond to the file identifier id^* associated with the adversary's forgery and outputting a public key $\mathbf{A}$ derived from the k-SIS challenge matrix. For queries *not* associated with id^*, the simulator "swaps the roles" of the public key and hash function as follows: we use TrapGen to program the random oracle with a matrix $H(\mathsf{id})$ for which we know a short basis, and we use ExtBasis to compute a short basis for $\mathbf{A} \| H(\mathsf{id})$. We can then compute the signatures as in the real system.

For the query id^*, we construct $H(\mathsf{id}^*)$ so that the k-SIS challenge vectors are valid signatures for the vectors queried by the adversary. We construct the mod q part of $H(\mathsf{id}^*)$ using the fact that valid signatures are elements of $\Lambda_q^{\perp}(\mathbf{A} \| H(\mathsf{id}^*))$, and we construct the mod 2 part of $H(\mathsf{id}^*)$ using the fact that the k-SIS challenge vectors are statistically close to random mod 2.

With this setup, a forged signature is exactly a solution to the k-SIS problem mod q. We now give the theorem.

Theorem 5.2. *Let $\mathcal{N}$ be the linearly homomorphic signature scheme over $\mathbb{F}_2$ described above. Suppose that $m = \lceil 6n \lg 2q \rceil$ and $\sigma = 30\sqrt{n \lg 2q} \log n$. Let $\beta = L \cdot \sigma\sqrt{m}$. Then $\mathcal{N}$ is unforgeable in the random oracle model assuming that k-$\mathsf{SIS}_{q,2m,\beta}$ is infeasible.*

In particular, let $\mathcal{A}$ be a polynomial-time adversary as in Definition 2.2. Then there exists a polynomial-time algorithm $\mathcal{B}$ that solves k-$\mathsf{SIS}_{q,2m,\beta,\sigma}$, such that

$$k\text{-SIS-Adv}[\mathcal{B}, (q, 2m, \beta, \sigma)] \geq \frac{1}{Q_s + Q_h + 1} \cdot \text{HomSig-Adv}[\mathcal{A}, \mathcal{N}] - \frac{Q_s(Q_s + Q_h)}{2^n} - 2^{k-2m} - \epsilon,$$

where Q_s and Q_h are the number of signature and hash queries made by $\mathcal{A}$, respectively, and ϵ is a negligible function of the security parameter n.

Proof. Let $\mathcal{A}$ be an adversary as in Definition 2.2 that makes at most Q_s signature queries and at most Q_h hash queries. We construct an algorithm $\mathcal{B}$ that takes as input a random matrix $\mathbf{B} \in \mathbb{Z}_q^{n \times 2m}$ and k vectors $\mathbf{e}_1^*, \ldots, \mathbf{e}_k^* \leftarrow \mathcal{D}_{\Lambda_q^\perp(\mathbf{B}),\sigma}$, and outputs a short vector $\mathbf{e}^* \in \mathbb{Z}^{2m}$ that is $\mathbb{Q}$-linearly independent of the $\mathbf{e}_i^*$.

Algorithm $\mathcal{B}$ simulates the hash function H and the Setup and Sign algorithms of $\mathcal{N}$ as follows:

Setup. $\mathcal{B}$ does the following:

1. Choose random $\mathbf{A}_2 \xleftarrow{\text{R}} \mathbb{F}_2^{n \times m}$.
2. Let $\mathbf{A}_q$ be the left m columns of $\mathbf{B}$.
3. Use the Chinese remainder theorem to compute $\mathbf{A} \in \mathbb{Z}_{2q}^{n \times m}$ such that $\mathbf{A} = \mathbf{A}_2 \bmod 2$ and $\mathbf{A} = \mathbf{A}_q \bmod q$.
4. Choose random $J \xleftarrow{\text{R}} \{1, \ldots, Q_s + Q_h + 1\}$.
5. Output the public key $\mathbf{A}$.

Hash query. When $\mathcal{A}$ requests the value of $H(\mathsf{id})$, algorithm $\mathcal{B}$ does the following:

1. If id has already been queried, return $H(\mathsf{id})$.
2. If id is the Jth hash query, do the following:
   (a) Let $\mathbf{E}$ be the $2m \times k$ matrix whose columns are the vectors $\mathbf{e}_i^*$. If the last m rows of $\mathbf{E}$ have $\mathbb{F}_2$-rank less than k, then abort. (The simulation has failed.)
   (b) If id was chosen by the simulator to answer a signature query, do the following:
      i. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be the vectors queried by the adversary, and let $\mathbf{V}$ be the $n \times k$ matrix whose columns are the vectors $\mathbf{v}_i$.
      ii. Choose a random matrix $\mathbf{U}_2 \leftarrow \mathbb{F}_2^{n \times m}$ such that $(\mathbf{A}_2 \| \mathbf{U}_2) \cdot \mathbf{E} = \mathbf{V} \bmod 2$.[5]

      Otherwise, choose a uniformly random matrix $\mathbf{U}_2 \xleftarrow{\text{R}} \mathbb{F}_2^{n \times m}$.
   (c) Let $\mathbf{U}_q$ be the right m columns of $\mathbf{B}$.

[5]E.g., by choosing $m - k$ columns of $\mathbf{U}_2$ at random and solving for the remaining ones. The fact that the last m rows of $\mathbf{E}$ have $\mathbb{F}_2$-rank k guarantees that for an appropriate choice of columns, a solution exists and is unique.

(d) Use the Chinese remainder theorem to compute $\mathbf{U}_{\mathsf{id}} \in \mathbb{Z}_{2q}^{n \times m}$ such that $\mathbf{U}_{\mathsf{id}} = \mathbf{U}_q \bmod q$ and $\mathbf{U}_{\mathsf{id}} = \mathbf{U}_2 \bmod 2$.

(e) Return $H(\mathsf{id}) \leftarrow \mathbf{U}_{\mathsf{id}}$.

3. Otherwise, do the following:

(a) Run $\mathsf{TrapGen}(n, m, 2q)$ to generate a matrix $\mathbf{U}_{\mathsf{id}} \in \mathbb{Z}_{2q}^{n \times m}$ and a short basis $\mathbf{T}_{\mathsf{id}}$ of $\Lambda_{2q}^{\perp}(\mathbf{U}_{\mathsf{id}})$.

(b) Return $H(\mathsf{id}) \leftarrow \mathbf{U}_{\mathsf{id}}$ and store $\mathbf{T}_{\mathsf{id}}$.

Sign. When $\mathcal{A}$ requests a signature on a file V represented by k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}_2^n$, algorithm $\mathcal{B}$ does the following:

1. Choose a random $\mathsf{id} \xleftarrow{\mathrm{R}} \{0,1\}^n$. If id has already been queried to the hash function H, then abort. (The simulation has failed.)
2. Simulate a query to $H(\mathsf{id})$. If this query is the Jth distinct hash query, let $\{\mathbf{e}_i\}$ be the k-SIS challenge vectors $\{\mathbf{e}_i^*\}$. Otherwise,

(a) Set $\mathbf{B} \leftarrow \mathbf{A} \| H(\mathsf{id})$.

(b) Let $\mathbf{S} \leftarrow \mathsf{ExtBasis}(\mathbf{T}_{\mathsf{id}}, \mathbf{B})$ be a basis for $\Lambda_{2q}^{\perp}(\mathbf{B})$ with $\|\widetilde{\mathbf{S}}\| = \|\widetilde{\mathbf{T}_{\mathsf{id}}}\|$.

(c) For $i = 1, \ldots, n$, set $\mathbf{e}_i \leftarrow \mathsf{SamplePre}(\mathbf{B}, \mathbf{S}, \sigma, q \cdot \mathbf{v}_i) \in \mathbb{Z}^{2m}$.

3. Output $\mathbf{e}_1, \ldots, \mathbf{e}_k$.

Output. Eventually $\mathcal{A}$ outputs a signature $\mathbf{e}^*$, an identifier id^*, and a nonzero vector $\mathbf{y}^* \in \mathbb{F}_2^n$. Algorithm $\mathcal{B}$ outputs $\mathbf{e}^*$.

We first analyze the situations where the simulator can abort without outputting an answer and show that the probability that this happens is negligible, assuming Q_s and Q_h are polynomial in n.

- **Step** (2a) **of hashing.** By Proposition 4.6, the entries of $\mathbf{E}$ are statistically close to uniformly random mod 2. It follows from Lemma A.2 that the probability that the last m rows of $\mathbf{E}$ have $\mathbb{F}_2$-rank less than k is bounded above by $1/2^{m-k} + \mathrm{negl}(n)$.

- **Step** (1) **of signing.** In this case, the simulator aborts if the identifier chosen for the signature query is the same as an identifier previously queried to the hash oracle. This happens with probability at most $Q_s(Q_h + Q_s)/2^n$.

Next, we observe that if the simulator does not abort, the distribution of the simulator's outputs are (statistically) indistinguishable from the distribution of the outputs in the real signature scheme, under the assumption that H is a random oracle.

- **The public key.** The matrix $\mathbf{A}$ in the simulation is uniformly random in $\mathbb{Z}_{2q}^{n \times m}$. By Theorem 3.1, the matrix $\mathbf{A}$ in the real system is statistically close to uniform.

- **The output of** H **(on most queries).** If id is not the Jth hash query, then by Theorem 3.1, the matrices $\mathbf{U}_{\mathsf{id}}$ output by all other hash queries are statistically close to uniform. We defer the remaining case.

- **The output of** Sign **(on most queries).** If id is not the Jth hash query, then Theorem 3.3 (b) shows that the $\mathbf{e}_i$ in both the real construction and the simulation come from a distribution that is statistically close to $\mathcal{D}_{\Lambda_{2q}^{q \cdot \mathbf{v}_i}(\mathbf{B}), \sigma}$. We have shown above that the matrices $\mathbf{B} = \mathbf{A} \| H(\mathsf{id})$ in the real scheme and

simulation come from statistically close distributions. It follows that the output distributions of the signatures are statistically close.

- **The output of H and** Sign **when** id **is the Jth hash query.** By Proposition 4.6, the entries of the vectors $\mathbf{e}_i^*$ are statistically close to uniform mod 2. It follows that $\mathbf{U}_2$ is statistically close to uniform mod 2 given the adversary's view, and therefore $H(\mathsf{id})$ is uniformly random in $\mathbb{Z}_{2q}^{n\times m}$.

  As for the signature, observe that the vectors $\mathbf{e}_i^*$ come from the distribution $\mathcal{D}_{\Lambda_q^\perp(\mathbf{A}\|\mathbf{U}_{\mathsf{id}}),\sigma}$ and $\mathbf{A}\|\mathbf{U}_{\mathsf{id}}$ is constructed so that $\mathbf{e}_i^* \in \Lambda_2^{\mathbf{v}_i}(\mathbf{A}\|\mathbf{U}_{\mathsf{id}})$. Since $\Lambda_{2q}^{q\mathbf{v}_i}(\mathbf{A}\|\mathbf{U}_{\mathsf{id}}) = \Lambda_q^\perp(\mathbf{A}\|\mathbf{U}_{\mathsf{id}}) \cap \Lambda_2^{\mathbf{v}_i}(\mathbf{A}\|\mathbf{U}_{\mathsf{id}})$, the vectors $\mathbf{e}_i^*$ come from the distribution $\mathcal{D}_{\Lambda_{2q}^{q\mathbf{v}_i}(\mathbf{A}\|\mathbf{U}_{\mathsf{id}}),\sigma}$, which by Theorem 3.3 (b) is statistically close to the distribution of signatures output by the real system.

Now the probability that the simulator "guesses right"; i.e., that that id^* is the Jth query to the hash function, is $1/(Q_s + Q_h + 1)$. If the simulator guesses right, then $\mathbf{e}^*$ is a valid signature for the nonzero vector $\mathbf{y}^*$ belonging to the file with identifier id^*. In particular, this implies that $\mathbf{e}^*$ is nonzero and $\|\mathbf{e}^*\| < \beta$. Furthermore, our construction implies that $(\mathbf{A}\|\mathbf{U}_{\mathsf{id}^*}) \cdot \mathbf{e}^* = q \cdot \mathbf{y}^* \bmod 2q$, and therefore $\mathbf{e}^*$ is a nonzero vector in $\Lambda_q^\perp(\mathbf{A}\|\mathbf{U}_{\mathsf{id}^*}) = \Lambda_q^\perp(\mathbf{B})$.

It remains to show that $\mathbf{e}^*$ is not a $\mathbb{Q}$-linear combination of the vectors $\mathbf{e}_i^*$. Suppose the contrary; then $\mathbf{e}^* = \sum_{i=1}^k c_i\mathbf{e}_i^*$ for some $c_i \in \mathbb{Q}$. We claim that with overwhelming probability, the c_i all have odd denominator when written in lowest terms. To see this, let d be the least common denominator of the c_i. If some c_i has even denominator, then when we clear denominators we obtain an equation $d\mathbf{e}^* = \sum c_i'\mathbf{e}_i^*$ where d is even and at least one of the $c_i' \in \mathbb{Z}$ is odd. Reducing mod 2, we see that this implies that the $\mathbf{e}_i^*$ are $\mathbb{Z}_2$-linearly dependent. However, by Proposition 4.6 the $\mathbf{e}_i^*$ are statistically close to uniform in $\mathbb{Z}_2^m$, and therefore the probability that the $\mathbf{e}_i^*$ are $\mathbb{Z}_2$-linearly dependent is at most $1/2^{m-k}$, which is negligible (cf. Lemma A.2). This contradicts the hypothesis that some c_i has even denominator with non-negligible probability. It follows that the values $\{c_i \bmod 2\}$ are well-defined with overwhelming probability.

Now suppose that id^* was not produced by the simulator during a signature query (type 1 forgery). Since the vectors $\mathbf{e}_i$ are statistically close to uniform mod 2, their $\mathbb{F}_2$-span is a random subspace of $\mathbb{F}_2^m$ of dimension at most k. Because the adversary has no information about the vectors $\mathbf{e}_i^*$, the probability that $\mathbf{e}^*$ is in the $\mathbb{F}_2$-span of the $\mathbf{e}_i^*$ is at most $1/2^{m-k}$. By our claim above, if $\mathbf{e}^* = \sum c_i\mathbf{e}_i^*$ with $c_i \in \mathbb{Q}$, then this equation holds mod 2 with overwhelming probability, in which case $\mathbf{e}^*$ is in the $\mathbb{F}_2$-span of the $\mathbf{e}_i^*$, a contradiction.

On the other hand, suppose that id^* was produced by the simulator during a signature query (type 2 forgery). Then since $(\mathbf{A}\|\mathbf{U}_{\mathsf{id}^*}) \cdot \mathbf{e}_i^* = \mathbf{v}_i \bmod 2$, we have $\mathbf{y}^* = \sum c_i\mathbf{v}_i \bmod 2$. Thus $\mathbf{y}^*$ is not a valid forgery, and we have again obtained a contradiction. □

Corollary 5.3. *Let $\mathcal{N}$ be the linearly homomorphic signature scheme over $\mathbb{F}_2$ described above. Suppose that $m = \lceil 6n \lg 2q \rceil$, $\sigma = 30\sqrt{n \lg 2q} \log n$, and $m/k > n$. Let $\beta = L \cdot \sigma \cdot \sqrt{m}$, let $t > \omega(\sqrt{\log n})$, and let $\beta' = \beta \cdot (k^{3/2}+1)k!(t\sigma)^k$. Then $\mathcal{N}$ is unforgeable in the random oracle model assuming that $\mathsf{SIS}(q, 2m-k, \beta')$ is infeasible. In particular, if $k = O(1)$ and $t = \log n$, then we may take $\beta' = O(L \cdot (n \lg 2q)^{k/2+1}(\log n)^{2k+1})$.*

Since the SIS problem is only assumed to be hard for $\beta' \in \mathrm{poly}(n)$, our choice of σ forces k to be $O(1)$ to ensure security based on SIS.

5.2 Privacy

In our linearly homomorphic signature scheme, one derives a signature on a linear combination $\mathbf{v}$ of messages by taking a linear combination of the signatures on the original messages $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Thus the derived

signature on $\mathbf{v}$ is a linear combination of short vectors in cosets of some lattice Λ. To show that this derived signature does not leak information about the original signatures, we show that a linear combination of signatures generated by our signing algorithm is itself a short vector sampled from a distribution that depends only on the message $\mathbf{v}$ and the computed function. In particular, the derived signature does not depend on the original vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ (up to negligible statistical distance). Consequently, the derived signature provably does not leak any information about $\mathbf{v}_1, \ldots, \mathbf{v}_k$ beyond what is revealed by $\mathbf{v}$. We note that the length of σ reveals information about the computed linear function, but not about the original messages.

We now show that our linearly homomorphic signature scheme is weakly context hiding. Specifically, a derived signature on a linear combination $\mathbf{v} = \sum c_i \mathbf{v}_i$ depends (up to negligible statistical distance) only on $\mathbf{v}$ and the c_i, and not on the initial messages $\mathbf{v}_i$.

Theorem 5.4. *Let $\mathcal{N}$ be the linearly homomorphic signature scheme over $\mathbb{F}_2$ described above. Suppose that k is constant, $m = \lceil 6n \lg 2q \rceil$ and $\sigma = 30\sqrt{n \lg 2q} \log n$. Then $\mathcal{N}$ is weakly context-hiding.*

Proof. It suffices to prove that even a computationally unbounded adversary cannot win the context-hiding game with non-negligible probability. Such an adversary can recover a short basis for $\Lambda_{2q}^{\perp}(\mathbf{A})$ from the public key $\mathbf{A}$ and thus produce properly distributed signatures without asking the challenger. Therefore it suffices to argue that the adversary cannot win the challenge phase.

Let $(V_0, V_1, f_1, \ldots, f_s)$ be the adversary's output in the challenge phase, where $V_b = \operatorname{span}(\mathbf{v}_1^{(b)}, \ldots, \mathbf{v}_k^{(b)})$ for $b = 0, 1$. We know that

$$\mathbf{u}_i := f_i\big(\mathbf{v}_1^{(0)}, \ldots, \mathbf{v}_k^{(0)}\big) = f_i\big(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_k^{(1)}\big) \in \mathbb{F}_2^n \qquad \text{for all } i = 1, \ldots, s.$$

For $j = 1, \ldots, k$, let $\mathbf{e}_j^{(0)}$ and $\mathbf{e}_j^{(1)}$ be the challenger's signatures (in $\mathbb{Z}^m$) on the basis vectors of V_0 or V_1, respectively. The challenger chooses a bit b and gives the adversary signatures on $\mathbf{u}_1, \ldots, \mathbf{u}_s$ computed using the $\mathsf{Combine}$ algorithm on the signatures $\mathbf{e}_j^{(b)}$.

Suppose $b = 0$. Consider a signature $\mathbf{e}$ given to the adversary that authenticates the vector $\mathbf{u} = \mathbf{u}_i$ for some i Then $\mathbf{u} = \sum_{j=1}^{k} c_j \mathbf{v}_j^{(0)}$ for some $c_j \in \{0, 1\}$. (Recall that we identify $\mathbb{F}_2$ with $\{0, 1\}$.) Then our $\mathsf{Combine}$ algorithm computes $\mathbf{e} = \sum_{j=1}^{k} c_j \mathbf{e}_j^{(0)}$ for the same c_j. By Theorem 4.12, the combined signature $\mathbf{e}$ is generated from a distribution statistically close to $\mathcal{D}_{\Lambda_{2q}^{\mathbf{u}}, \|\mathbf{c}\|\sigma}$. (Note that the smoothing parameter bound holds since we have $g = \ell = 1$ and $\sigma > \sqrt{k} \cdot \eta_\epsilon(\Lambda_{2q}^{\perp}(\mathbf{A}))$ for constant k.) Since the same argument applies to the case $b = 1$, the adversary cannot distinguish an $\mathbf{e}_i$ generated from V_0 from an $\mathbf{e}_i$ generated from V_1. By applying this argument to all s signatures and using the triangle inequality for statistical distance, we obtain that the adversary cannot distinguish s signatures derived from V_0 from s signatures derived from V_1, as required. □

6 k-Time GPV Signatures Without Random Oracles

In this section we give a second application of the k-SIS mechanism described in Section 4, namely, a stateless variant of the the signature scheme of Gentry, Peikert, and Vaikuntanathan [16] that is k-time unforgeable in the standard model. The notion of k-time security means that a signing key can only be used to sign k messages. In particular, a forger is allowed at most k signing queries.

The main idea is to construct signatures as in our homomorphic scheme of Section 5, but remove the homomorphic property by setting the bound on the length of a valid signature to be very close to the expected

length of the signature. Since we expect a small number of such vectors to form a set that is nearly orthogonal, any linear combination of signatures will produce a vector that is too long to be accepted as a valid signature.

We prove our signature scheme *weakly unforgeable*; i.e., unforgeable under a static chosen-message attack, in which the adversary must submit all signature queries before seeing the public key. A standard transformation using *chameleon hashes* [19] produces a scheme that is unforgeable under the usual notion of adaptive chosen-message attack.

We now describe our weakly unforgeable signature scheme, which is essentially a GPV signature in which the hash function is replaced with the Chinese remaindering of the message (interpreted as a vector in $\mathbb{F}_2^n$) with the zero vector in $\mathbb{Z}_q^n$.

$\mathsf{Setup}(n, \mathsf{params})$. Given a security parameter n that is also the bit length of messages to be signed, do the following:

1. Choose an odd prime q. Set $m \leftarrow \lceil 6n \lg 2q \rceil$. Set $\sigma \leftarrow 30\sqrt{n \lg 2q} \log n$.
2. Run $\mathsf{TrapGen}(n, m, 2q)$ to generate a matrix $\mathbf{A} \in \mathbb{Z}_{2q}^{n \times m}$ and a basis $\mathbf{T}$ of $\Lambda_{2q}^{\perp}(\mathbf{A})$ such that $\|\widetilde{\mathbf{T}}\| \leq \sigma / \log n$.
3. Output the public key $\mathsf{pk} \leftarrow (\mathbf{A}, \sigma)$, and the private key $\mathsf{sk} \leftarrow (\mathbf{A}, \sigma, \mathbf{T})$.

$\mathsf{Sign}(\mathsf{sk}, \mathbf{v})$. Given secret key $\mathsf{sk} = (\mathbf{A}, \mathbf{T})$, and a message $\mathbf{v}$ (interpreted as a vector in $\mathbb{F}_2^n$), output $\mathbf{e} \leftarrow \mathsf{SamplePre}(\mathbf{A}, \mathbf{T}, \sigma, q \cdot \mathbf{v})$.

$\mathsf{Verify}(\mathsf{pk}, \mathbf{e}, \mathbf{v})$. Given a public key $\mathsf{pk} = \mathbf{A}$, a signature $\mathbf{e} \in \mathbb{Z}^{2m}$, and a message $\mathbf{v} \in \mathbb{F}_2^n$, output 1 if (a) $0 < \|\mathbf{e}\| \leq 1.1 \cdot \sigma\sqrt{m/2\pi}$ and (b) $\mathbf{A} \cdot \mathbf{e} = q \cdot \mathbf{v} \bmod 2q$. Otherwise output 0.

Correctness of our scheme follows from Proposition 4.7 (with $\epsilon = 0.1$). In our fully unforgeable scheme, the vector $\mathbf{v}$ used in signing is not the message but rather $H(m, r)$, where m is the message, r is random, and H is a chameleon hash function. The signature includes the randomness r in addition to the vector $\mathbf{e}$. For a discussion of lattice-based chameleon hash functions, see [9, §2.2].

6.1 Security

An adversary attacking our k-time signature scheme requests k signatures $\mathbf{e}_i$ on messages of his choice, receives a public key and the signatures, and then outputs a message $\mathbf{v}^*$ and a signature $\mathbf{e}^*$. The adversary wins the game if $\mathsf{Verify}(\mathsf{pk}, \mathbf{e}^*, \mathbf{v}^*) = 1$ and $\mathbf{v}^*$ is not equal to any of the messages queried. We denote the probability of an adversary $\mathcal{A}$ winning the game by $\mathrm{WeakSig\text{-}Adv}[\mathcal{A}, \mathcal{S}]$.

As was the case for our homomorphic scheme of Section 5, a valid forgery is a short vector $\mathbf{e}^*$ in $\Lambda_q^{\perp}(\mathbf{A})$. The main idea of our security proof is to show that the length bound on a valid $\mathbf{e}^*$ is so tight that the only nonzero integer vectors of comparable length in the $\mathbb{Q}$-span of the requested signatures $\mathbf{e}_i$ are the vectors $\pm\mathbf{e}_i$. (Since $-\mathbf{e}_i$ authenticates the same message as $\mathbf{e}_i$, the signature $-\mathbf{e}_i$ is not a valid forgery.) Thus $\mathbf{e}^*$ is outside the linear span of the $\mathbf{e}_i$, and we can use it to solve the k-SIS instance in which the $\mathbf{e}_i$ are the challenge vectors.

Our security theorem as follows:

Theorem 6.1. *Let $\mathcal{S}$ be the signature scheme described above. Suppose k is constant and $\beta = 1.1 \cdot \sigma\sqrt{m/2\pi}$. Then $\mathcal{S}$ is a weakly unforgeable k-time signature scheme assuming that k-$\mathsf{SIS}_{q,m,\beta,\sigma}$ is infeasible.*

In particular, let $\mathcal{A}$ be a polynomial-time adversary that plays the security game described above. Then there exists a polynomial-time algorithm $\mathcal{B}$ that solves k-$\mathsf{SIS}_{q,m,\beta,\sigma}$, such that

$$k\text{-}\mathsf{SIS}\text{-}\mathrm{Adv}[\mathcal{B}, (q, m, \beta, \sigma)] \geq \mathrm{WeakSig\text{-}Adv}[\mathcal{A}, \mathcal{S}] - \mathrm{negl}(n).$$

Proof. We use the same strategy as in the proof of Theorem 5.2. Suppose we are given a k-SIS challenge consisting of a matrix $\mathbf{B} \in \mathbb{Z}_q^{n \times m}$ and k vectors $\mathbf{e}_1^*, \ldots, \mathbf{e}_k^* \leftarrow \mathcal{D}_{\Lambda_q^\perp(\mathbf{B}),\sigma}$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}_2^n$ be the messages queried by the adversary $\mathcal{A}$. (If fewer than k messages are queried, we choose random $\mathbf{v}_i$ to bring the total up to k.) Our algorithm $\mathcal{B}$ does the following:

1. Let $\mathbf{E}$ be the $m \times k$ matrix whose columns are the vectors $\mathbf{e}_i^*$. If $\mathbf{E}$ has $\mathbb{F}_2$-rank less than k, then abort. (The simulation has failed.)
2. Let $\mathbf{V}$ be the $n \times k$ matrix whose columns are the vectors $\mathbf{v}_i$.
3. Choose a random matrix $\mathbf{A}_2 \leftarrow \mathbb{F}_2^{n \times m}$ such that $\mathbf{A}_2 \cdot \mathbf{E} = \mathbf{V} \bmod 2$.
4. Use the Chinese remainder theorem to compute $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ such that $\mathbf{A} = \mathbf{A}_2 \bmod 2$ and $\mathbf{A} = \mathbf{B} \bmod q$.
5. Output the public key $\mathbf{A}$ and the signatures $\mathbf{e}_i^*$.

Now suppose that after receiving this information, the adversary $\mathcal{A}$ outputs a signature $\mathbf{e}^*$ on a message $\mathbf{v}^* \in \mathbb{F}_2^n$. Algorithm $\mathcal{B}$ outputs $\mathbf{e}^*$.

By Proposition 4.6, the entries of $\mathbf{E}$ are statistically close to uniformly random mod 2. By Lemma A.2, the probability that $\mathbf{E}$ has $\mathbb{F}_2$-rank less than k is bounded above by $1/2^{m-k} + \mathrm{negl}(n)$. Thus the probability that the simulator aborts is negligible.

We next observe that if the simulator does not abort, the distribution of the simulator's outputs are (statistically) indistinguishable from the distribution of the outputs in the real signature scheme.

- **The public key.** By Proposition 4.6, the entries of the vectors $\mathbf{e}_i^*$ are statistically close to uniform mod 2. It follows that $\mathbf{A}_2$ is statistically close to uniform mod 2 given the adversary's view, and therefore $\mathbf{A}$ is uniformly random in $\mathbb{Z}_{2q}^{n \times m}$.
- **The signatures.** Observe that the vectors $\mathbf{e}_i^*$ come from the distribution $\mathcal{D}_{\Lambda_q^\perp(\mathbf{A}),\sigma}$ and $\mathbf{A}$ is constructed so that $\mathbf{e}_i^* \in \Lambda_2^{\mathbf{v}_i}(\mathbf{A})$. Since $\Lambda_{2q}^{q\mathbf{v}_i}(\mathbf{A}) = \Lambda_q^\perp(\mathbf{A}) \cap \Lambda_2^{\mathbf{v}_i}(\mathbf{A})$, the vectors $\mathbf{e}_i^*$ come from the distribution $\mathcal{D}_{\Lambda_{2q}^\perp(\mathbf{A}),\sigma}$, which by Theorem 3.3 (b) is statistically close to the distribution of signatures output by the real system.

Finally, since the adversary's forgery $\mathbf{e}^*$ is in $\Lambda_{2q}^{q\mathbf{v}^*}(\mathbf{A})$ and $\mathbf{A} = \mathbf{B} \bmod q$, we have $\mathbf{e}^* \in \Lambda_q^\perp(\mathbf{B})$. Furthermore, the fact that $\mathbf{e}^*$ is a valid signature implies $0 < \|\mathbf{e}^*\| \leq 1.1 \cdot \sigma\sqrt{m/2\pi}$. Since $\mathbf{v}^* \notin \{\mathbf{v}_i\}$, we see that $\mathbf{e}^* \neq \pm\mathbf{e}_i$ for all i. (Here we use the fact that since the messages are defined mod 2, the signature $-\mathbf{e}_i$ authenticates the same message as $\mathbf{e}_i$.) Proposition 4.8 now shows that with overwhelming probability, $\mathbf{e}^* \notin \mathbb{Z}\text{-span}(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$.

By Theorem 4.3, the distribution of a k-SIS challenge is statistically close to the distribution obtained by sampling $\mathbf{e}_i$ from a Gaussian over $\mathbb{Z}^m$ and choosing random $\mathbf{A}$ such that $\mathbf{A} \cdot \mathbf{e}_i = 0 \bmod q$. Lemma 4.10 now shows that for the latter distribution, any integer vector in $\mathbb{Q}\text{-span}(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$ is in fact in $\mathbb{Z}\text{-span}(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$ with overwhelming probability. Thus we can apply the same conclusion to our k-SIS challenge and see that $\mathbf{e}^* \notin \mathbb{Q}\text{-span}(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\})$. It follows that $\mathbf{e}^*$ is a solution to the k-SIS challenge. □

7 Further Directions

Extending the linearly homomorphic system. While our linearly homomorphic scheme in Section 5 authenticates vectors with coordinates in $\mathbb{F}_2$, the same construction works for any field $\mathbb{F}_p$ where p is a small

prime. We simply set $m = \lceil 6n \lg pq \rceil$ and $\sigma = 30\sqrt{n \lg pq} \log n$, and sign a vector $\mathbf{v} \in \mathbb{F}_p^n$ using the lattice $\Lambda_{pq}^{q \cdot \mathbf{v}}(\mathbf{A})$. If p is odd and we identify $\mathbb{F}_p$ with $\{-(p-1)/2, \ldots, (p-1)/2\}$, then the output of Combine on ℓ vectors can be up to $\ell(p-1)$ times as long as the largest input vector. An argument as in Proposition 5.1 shows that the resulting system is $L/(p-1)$-limited. We must therefore increase L accordingly to allow for as many linear combinations as in the system over $\mathbb{F}_2$.

More interestingly, our system can also be used to authenticate vector spaces defined over non-prime fields. Suppose for concreteness that our vectors live in $(\mathbb{F}_{2^d})^n$. If we fix a basis for $\mathbb{F}_{2^d}$ over $\mathbb{F}_2$, then when computing signatures we may view the vectors as elements of $(\mathbb{F}_2)^{nd}$ and compute signatures in exactly the same manner as above. The difference comes when computing linear combinations over $\mathbb{F}_{2^d}$: in our representation multiplying an element $x \in \mathbb{F}_{2^d}$ by an element $\alpha \in \mathbb{F}_{2^d}$ is multiplying the corresponding vector $\mathbf{x} \in \mathbb{F}_2^d$ by a matrix $M_\alpha \in \mathbb{F}_2^{d \times d}$. To compute this action on the signature vector $\mathbf{e} \in \mathbb{Z}^m$, we lift M_α to an integer matrix with entries in $\{0, 1\}$ and group the elements of $\mathbf{e}$ into d-tuples corresponding to the underlying elements of $\mathbb{F}_{2^d}$. Multiplying each d-tuple by M_α now has the effect of multiplying the underlying elements of $\mathbb{F}_{2^d}$ by α. We see that this action increases the length of $\mathbf{e}$ by a factor of at most d, so combining ℓ vectors gives an output that is up to ℓd times as long as the largest input vector. By the same argument as above, the system over $\mathbb{F}_{2^d}$ is L/d-limited.

Parameter selection. We now consider how large the parameter q needs to be in order to guarantee security of our signature schemes. We use the result of Gentry, Peikert, and Vaikuntanathan [16, Section 9] which gives a reduction of $\mathsf{SIS}_{q,m,\beta}$ to standard worst-case lattice problems. The reduction requires that m and β be polynomial in n and that

$$q \geq \beta\sqrt{n} \cdot \omega(\sqrt{\log n}). \tag{7.1}$$

By Corollary 5.3, if k is constant and $t = \log n$, then there is some constant c (depending on k) such that our linearly homomorphic scheme over $\mathbb{F}_2$ is unforgeable whenever

$$\frac{q}{(\lg 2q)^{k/2+1}} > c(k) \cdot L \cdot n^{(k+3)/2} (\log n)^{2k+2}.$$

Even if the constant c and the complexity parameter L were equal to 1, for reasonable values of n and k the modulus q must be very large: at least 2^{170} for $n = 1000$ and $k = 10$. (For our k-time signature scheme the same analysis holds, with the parameter L replaced by the constant $1.1/\sqrt{2\pi}$.)

The requirement that q be large is due entirely to the fact that the reduction from k-SIS to SIS is exponential in k. If a better reduction could be found and/or k-SIS could be reduced directly to worst-case lattice problems in a way that gave a bound on q similar to (7.1), then $q/\lg q$ would grow like $L \cdot n^{3/2} \log^2 n$, and we could use values of q in the range of 2^{40}.

Open problems. An important open problem inspired by our construction is to find a tight reduction of k-SIS to worst-case lattice problems, either by improving on the reduction to SIS given by Theorem 4.2 or by a direct argument. An improved reduction would support the use of the k-SIS problem in developing cryptosystems for other applications, and would also allow us to implement our systems with smaller parameters.

Acknowledgments

We thank Chris Peikert for helpful discussions, and in particular for providing the idea of using continuous Gaussians to prove Proposition 4.7. We also thank the anonymous reviewers for helpful suggestions.

References

[1] S. Agrawal, D. Boneh, and X. Boyen. "Lattice basis delegation in fixed dimension and shorter-ciphertext hierarchical IBE." In *Advances in Cryptology — CRYPTO 2010*, Springer LNCS **6223** (2010), 98–115.

[2] L. V. Ahlfors. *Complex analysis*. Third edition. McGraw-Hill Book Co., New York (1978).

[3] M. Ajtai. "Generating hard instances of the short basis problem." In *ICALP*, ed. J. Wiedermann, P. van Emde Boas, and M. Nielsen, Springer LNCS **1644** (1999), 1–9.

[4] J. Alwen and C. Peikert. "Generating shorter bases for hard random lattices." In *STACS* (2009), 75–86. Full version available at `http://www.cc.gatech.edu/~cpeikert/pubs/shorter.pdf`.

[5] G. Ateniese, D. H. Chou, B. de Medeiros, and G. Tsudik. "Sanitizable signatures." In *Computer Security — ESORICS '05*, Springer LNCS **3679** (2005), 159–177.

[6] D. Boneh, D. Freeman, J. Katz, and B. Waters. "Signing a linear subspace: Signature schemes for network coding." In *Public-Key Cryptography — PKC '09*, Springer LNCS **5443** (2009), 68–87.

[7] C. Brzuska, H. Busch, Ö. Dagdelen, M. Fischlin, M. Franz, S. Katzenbeisser, M. Manulis, C. Onete, A. Peter, B. Poettering, and D. Schröder. "Redactable signatures for tree-structured data: Definitions and constructions." In *Applied Cryptography and Network Security — ACNS '10*, Springer LNCS **6123** (2010), 87–104.

[8] C. Brzuska, M. Fischlin, T. Freudenreich, A. Lehmann, M. Page, J. Schelbert, D. Schröder, and F. Volk. "Security of sanitizable signatures revisited." In *Public Key Cryptography — PKC '09*, Springer LNCS **5443** (2009), 317–336.

[9] D. Cash, D. Hofheinz, E. Kiltz, and C. Peikert. "Bonsai trees, or, how to delegate a lattice basis." In *Advances in Cryptology — EUROCRYPT 2010*, Springer LNCS **6110** (2010), 523–552.

[10] E.-C. Chang, C. L. Lim, and J. Xu. "Short redactable signatures using random trees." In *Topics in Cryptology — CT-RSA '09*, Springer LNCS **5473** (2009), 133–147.

[11] D. Charles, K. Jain, and K. Lauter. "Signatures for network coding." *International Journal of Information and Coding Theory* **1** (2009), 3–14.

[12] D. Dummit and R. Foote. *Abstract Algebra*. 2nd edition. Prentice-Hall, Upper Saddle River, NJ (1999).

[13] M. Evans, N. Hastings, and B. Peacock. *Statistical distributions*. Third edition. Wiley Series in Probability and Statistics: Texts and References Section, Wiley-Interscience, New York (2000).

[14] C. Fragouli and E. Soljanin. "Network coding fundamentals." *Found. Trends Netw.* **2** (2007), 1–133.

[15] R. Gennaro, J. Katz, H. Krawczyk, and T. Rabin. "Secure network coding over the integers." In *Public Key Cryptography — PKC '10*, Springer LNCS **6056** (2010), 142–160.

[16] C. Gentry, C. Peikert, and V. Vaikuntanathan. "Trapdoors for hard lattices and new cryptographic constructions." In *STOC*, ed. R. E. Ladner and C. Dwork. ACM (2008), 197–206.

[17] S. Haber, Y. Hatano, Y. Honda, W. Horne, K. Miyazaki, T. Sander, S. Tezoku, and D. Yao. "Efficient signature schemes supporting redaction, pseudonymization, and data deidentification." In *ACM Symposium on Information, Computer and Communications Security — ASIACCS '08* (2008), 353–362.

[18] R. Johnson, D. Molnar, D. Song, and D. Wagner. "Homomorphic signature schemes." In *Topics in Cryptology — CT-RSA 2002*, Springer LNCS **2271** (2002), 244–262.

[19] H. Krawczyk and T. Rabin. "Chameleon signatures." In *Network and Distributed System Security Symposium (NDSS)* (2000).

[20] M. Krohn, M. Freedman, and D. Mazières. "On-the-fly verification of rateless erasure codes for efficient content distribution." In *Proc. of IEEE Symposium on Security and Privacy* (2004), 226–240.

[21] D. Micciancio and O. Regev. "Worst-case to average-case reductions based on Gaussian measures." In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society, Washington, DC, USA (2004), 372–381.

[22] K. Miyazaki, G. Hanaoka, and H. Imai. "Digitally signed document sanitizing scheme based on bilinear maps." In *ACM Symposium on Information, Computer and Communications Security — ASIACCS '06* (2006), 343–354.

[23] K. Miyazaki, M. Iwamura, T. Matsumoto, R. Sasaki, H. Yoshiura, S. Tezuka, and H. Imai. "Digitally signed document sanitizing scheme with disclosure condition control." *IEICE Transactions on Fundamentals* **E88-A** (2005), 239–246.

[24] P. Paillier and D. Vergnaud. "Discrete-log-based signatures may not be equivalent to discrete log." In *Advances in Cryptology — ASIACRYPT '05*, Springer LNCS **3788** (2005), 1–20.

[25] C. Peikert. "Limits on the hardness of lattice problems in ℓ_p norms." *Comput. Complex.* **17** (2008), 300–351.

[26] O. Regev. "On lattices, learning with errors, random linear codes, and cryptography." *Journal of the ACM* **56** (2009), Art. 34. Preliminary version in *STOC '05*.

[27] R. Steinfeld, L. Bull, and Y. Zheng. "Context extraction signatures." In *Information Security and Cryptology (ICISC)*, Springer LNCS **2288** (2001), 285–304.

[28] F. Zhao, T. Kalker, M. Médard, and K. Han. "Signatures for content distribution with network coding." In *Proc. Intl. Symp. Info. Theory (ISIT)* (2007).

A Probability

Statistical distance. Let X and Y be two random variables taking values in some countable set Ω. Define the *statistical distance*, denoted $\Delta(X;Y)$, to be

$$\Delta(X;Y) := \frac{1}{2}\sum_{s\in\Omega}\big|\Pr[X=s]-\Pr[Y=s]\big|$$

If $X(n)$ and $Y(n)$ are ensembles of random variables, we say X and Y are *statistically close* if $\Delta(X;Y)$ is a negligible function of n.

Lemma A.1. *Let X, Y be random variables taking values in a finite set Ω. Let $A \subseteq \Omega$, let X_A be a random variable taking values in A defined by*

$$\Pr[X_A = s] := \Pr[X = s] / \Pr[X \in A] \quad \textit{for all } s \in A,$$

and let Y_A be defined similarly. Then:

1. *If $\Pr[X \in A] \geq 1 - \epsilon$, then $\Delta(X; X_A) \leq \epsilon$.*
2. *If $\Pr[X \in A] \geq 1 - \epsilon$ and $\Delta(X_A; Y_A) \leq \epsilon$, then $\Delta(X, Y) \leq 4\epsilon$.*

Proof. The first statement is Property (5) of [1, Lemma 12]. For the second statement, we have

$$\begin{aligned}
\Pr[Y \in A] &= \sum_{s \in A} \Pr[Y = s] \\
&= \sum_{s \in A} \Pr[X = s] + \Pr[Y = s] - \Pr[X = s] \\
&\geq \sum_{s \in A} \Pr[X = s] - \Big| \Pr[Y = s] - \Pr[X = s] \Big| \\
&= \Pr[X \in A] - \Delta(X_A; Y_A) \\
&\geq 1 - 2\epsilon.
\end{aligned}$$

It now follows from the first statement that $\Delta(Y; Y_A) \leq 2\epsilon$, and therefore

$$\Delta(X; Y) \leq \Delta(X; X_A) + \Delta(X_A; Y_A) + \Delta(Y_A; Y) \leq 4\epsilon. \qquad \square$$

Random matrices over finite fields.

Lemma A.2. *Let m, n, q be integers with $m > n$ and q prime. Then the probability that a uniformly random matrix $\mathbf{A} \xleftarrow{\mathrm{R}} \mathbb{F}_q^{n \times m}$ has $\mathbb{F}_q$-rank less than n is at most $1/q^{m-n}$.*

Proof. We view $\mathbf{A}$ as a set of n independent vectors $\mathbf{v}_i \xleftarrow{\mathrm{R}} \mathbb{F}_q^m$. The probability that $\mathbf{A}$ has less than full rank is bounded above by

$$\sum_{i=0}^{n-1} \Pr[\mathbf{v}_{i+1} \in \mathrm{span}(\mathbf{v}_1, \ldots, \mathbf{v}_i)] = \sum_{i=0}^{n-1} \frac{1}{q^{m-i}} < \frac{1}{q^{m-n}}.$$

$\square$

Length of a vector sampled from a continuous Gaussian.

Lemma A.3. *Let $\tau > 0$ and $\mathbf{c} \in \mathbb{R}^n$. For $\hat{\mathbf{x}} \in \mathbb{R}^n$ sampled from $D_{\tau, \mathbf{c}}$ and any fixed $\gamma_1 < 1$ and $\gamma_2 > 1$, we have*

$$\Pr\left[\gamma_1 \cdot \tau\sqrt{n/2\pi} \;\leq\; \|\hat{\mathbf{x}} - \mathbf{c}\| \;\leq\; \gamma_2 \cdot \tau\sqrt{n/2\pi}\right] \;\geq\; 1 - \mathrm{negl}(n). \tag{A.1}$$

Proof. Let $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{y}$. The length of $\hat{\mathbf{x}}$ is distributed according to the chi distribution with parameter $\tau/\sqrt{2\pi}$, which is given by

$$\Pr[\|\hat{\mathbf{x}} - \mathbf{c}\| = x] = \frac{2^{1-n/2}\left(\frac{x\sqrt{2\pi}}{\tau}\right)^{n-1} e^{-\pi x^2 \tau^2}}{\Gamma(\frac{n}{2})} \tag{A.2}$$

(see [13, §8.3]). By Stirling's formula [2, p. 204] we have $\Gamma(z) > z^z e^{-z}\sqrt{2\pi/z}$ for positive real z. If we let $f_n(x)$ denote the right hand side of (A.2), then for any $\gamma > 0$ we have

$$f_n\left(\gamma \cdot \tau\sqrt{n/2\pi}\right) \;<\; \frac{2^{1-n/2} n^{(n-1)/2} \gamma^{n-1} e^{-\gamma^2 n/2}}{\sqrt{4\pi/n} \cdot n^{n/2} 2^{-n/2} e^{-n/2}} \;=\; \frac{1}{\gamma\sqrt{\pi}} e^{-\frac{n}{2}(\gamma^2 - 1 - 2\log\gamma)}. \tag{A.3}$$

Next, observe that $f_n(x)$ is increasing on $(0, \tau\sqrt{(n-1)/2\pi})$ and decreasing thereafter. Since for any fixed $\gamma < 1$ we have $\gamma \le \sqrt{(n-1)/n}$ for sufficiently large n, it follows that for any $\gamma > 1$ and sufficiently large n we have

$$\Pr\left[\|\hat{\mathbf{x}} - \mathbf{c}\| \le \gamma \cdot \frac{\tau\sqrt{n}}{\sqrt{2\pi}}\right] \;=\; \int_0^{\gamma\tau\sqrt{n/2\pi}} f_n(x)dx \;<\; \frac{\gamma\tau\sqrt{n}}{\sqrt{2\pi}} \cdot \frac{1}{\gamma\sqrt{\pi}} e^{-\frac{n}{2}(\gamma^2 - 1 - 2\log\gamma)}. \tag{A.4}$$

Since $\gamma^2 - 1 - 2\log\gamma > 0$ for any $\gamma < 1$, the right hand side of (A.4) is a negligible function of n, proving the lower bound in the Lemma.

To show the upper bound, we use the substitution $y = x/(\tau\sqrt{n/2\pi})$ and the fact that $\gamma > 1$ to compute

$$\begin{aligned}
\Pr\left[\|\hat{\mathbf{x}} - \mathbf{c}\| \ge \gamma \cdot \tau\sqrt{n/2\pi}\right] &= \int_{\gamma\tau\sqrt{n/2\pi}}^{\infty} f_n(x)\, dx \\
&= \tau\sqrt{n/2\pi} \int_{\gamma}^{\infty} f_n(y \cdot \tau\sqrt{n/2\pi})\, dy \\
&< \tau\sqrt{n/2\pi} \int_{\gamma}^{\infty} \frac{1}{y\sqrt{\pi}} e^{-\frac{n}{2}(y^2 - 1 - 2\log y)} dy \qquad \text{(A.5)} \\
&< \frac{\tau}{\pi}\sqrt{\frac{n}{2}} \left(\int_{\gamma}^{3\gamma} e^{-\frac{n}{2}(\gamma^2 - 1 - 2\log\gamma)} dy + \int_{3\gamma}^{\infty} e^{-\frac{n}{2}y} dy \right) \qquad \text{(A.6)} \\
&= \frac{\tau}{\pi}\sqrt{\frac{n}{2}} \left(2\gamma \cdot e^{-\frac{n}{2}(\gamma^2 - 1 - 2\log\gamma)} + \frac{2}{n} e^{-3\gamma\frac{n}{2}} \right). \qquad \text{(A.7)}
\end{aligned}$$

The inequality (A.5) follows from the bound in (A.3). The first term in (A.6) uses the fact that $x^2 - 1 - \log x$ is increasing for $x > 1$, while the second term uses the fact that $x^2 - 1 - \log x > x$ for $x > 3$ (and we eliminate the $1/y$ in both terms using the fact that $\gamma > 1$). Finally, since $x^2 - 1 - 2\log x > 0$ for any $x > 1$, both terms in (A.7) are negligible functions of n, proving the upper bound in the Lemma. □