# Asymptotics of hybrid primal lattice attacks

Daniel J. Bernstein[1,2]

[1] Department of Computer Science, University of Illinois at Chicago, USA
[2] Horst Görtz Institute for IT Security, Ruhr University Bochum, Germany
djb@cr.yp.to

**Abstract.** The literature gives the impression that (1) existing heuristics accurately predict how effective lattice attacks are, (2) non-ternary lattice systems are not vulnerable to hybrid multi-decoding primal attacks, and (3) the asymptotic exponents of attacks against non-ternary systems have stabilized.

This paper shows that 1 contradicts 2 and that 1 contradicts 3: the existing heuristics imply that hybrid primal key-recovery attacks are exponentially faster than standard non-hybrid primal key-recovery attacks against the LPR PKE with any constant error width. This is the first report since 2015 of an exponential speedup in heuristic non-quantum primal attacks against non-ternary LPR.

Quantitatively, for dimension $n$, modulus $n^{Q_0+o(1)}$, and error width $w$, a surprisingly simple hybrid attack reduces heuristic costs from $2^{(\rho+o(1))n}$ to $2^{(\rho-\rho H_0+o(1))n}$, where $z_0 = 2Q_0/(Q_0+1/2)^2$, $\rho = z_0 \log_4(3/2)$, and $H_0 = 1/(1+(\lg w)/0.057981 z_0)$. This raises the questions of (1) what heuristic exponent is achieved by more sophisticated hybrid attacks and (2) what impact hybrid attacks have upon concrete cryptosystems whose security analyses have ignored hybrid attacks, such as Kyber-512.

**Keywords:** algorithm analysis, lattice attacks, combinatorial attacks, lattice-basis reduction, hybrid attacks, CVPP, BDDP

## 1 Introduction

The NewHope proposal [12]—the winner of the 2016 Internet Defense Prize, and the ancestor of many subsequent cryptosystems—states that some *other* proposals for lattice-based post-quantum cryptography have "combinatorial vulnerabilities [56] induced by the fact that secrets are ternary".

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Excellence Strategy of the German Federal and State Governments—EXC 2092 CASA—390781972 "Cyber Security in the Age of Large-Scale Adversaries"; by the U.S. National Science Foundation under grant 1913167; and by the Taiwan's Executive Yuan Data Safety and Talent Cultivation Project (AS-KPQ-109-DSTCP). "Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation" (or other funding agencies). Permanent ID of this document: be42ac4f84a5e0d20476d6958fb8c96259c3608b. Date: 2023.12.08.

The cited paper "[56]" is a 2007 paper "A hybrid lattice-reduction and meet-in-the-middle attack against NTRU" [80] by Howgrave-Graham. "Hybrid" refers to the fact that the attack combines ingredients from two earlier types of attacks: attacks that reduce a public lattice basis in the hope of discovering a secret lattice vector, and attacks that carry out a meet-in-the-middle combinatorial search for the secret vector. There are many other uses of the word "hybrid" in the literature; a more distinctive feature of the attack from [80] is that it formulates many guesses for portions of the secret vector and applies a decoding step to each guess, so one might call it a "multi-decoding" attack.

"Ternary" in [12], sometimes called "trinary" in the literature, refers to secret vectors with entries in $\{-1, 0, 1\}$. "Vulnerabilities" does not have a clear definition in [12] but is nevertheless clear as a guide to cryptosystem selection: surely nobody would want to use a cryptosystem with "vulnerabilities"!

The Kyber proposal—a descendant of NewHope, and NIST's first selection of a post-quantum encryption system to standardize—expresses [13, page 13] an "intuition" that the threat from "improvements to hybrid attacks [56] targeting schemes with very low noise" is much larger than the threat from "attacks attempting to exploit failures that occur with extremely low probability".

Again "[56]" is the Howgrave-Graham paper. The "threat" evaluation guides cryptosystem selection in the same way that the word "vulnerabilities" does: the reader is invited to envision the nightmare scenario of improved hybrid attacks rendering "schemes with very low noise" feasible to break. The Kyber security analysis makes no further mention of hybrid attacks. The reader is led to believe that Kyber is immune to this entire class of attacks.

A skim of other proposals for lattice-based encryption finds KINDI and LIMA similarly dismissing hybrid attacks, and finds Compact-LWE, Ding Key Exchange, EMBLEM, Frodo, HILA5, KCL, Lizard, LOTUS, SABER, and Titanium not even citing Howgrave-Graham's paper.

One *can* find hybrid-attack costs quantified, and often reported to be the best attacks known, for LAC [98, Table 3], NTRU [44, Table 6], NTRU Prime [26, Table 2], and Round5 [15, Table 4]. All of these systems use ternary secrets.

This might sound like a convincing case for the narrative that ternary secrets are the source of combinatorial vulnerabilities—perhaps worse vulnerabilities than currently known—and that, unless ternary secrets are critical for performance, one should use a non-ternary system, such as Kyber.

**1.1. Is it true that hybrid multi-decoding attacks are limited to ternary secrets?** A skeptical security reviewer will ask for *algorithm analyses* providing clear evidence that non-ternary systems are not vulnerable to hybrid attacks. Let's look at what the literature says about the analysis of these attacks.

Howgrave-Graham's paper [80] introduced hybrid attacks as combinatorially searching "some linear combinations" of some basis vectors. Nothing in this description structurally limits the combinations to coefficients $-1, 0, 1$.

The paper then specialized this to attacking *binary* secrets, taking the linear combinations specifically as $2^\sigma$ sums of subsets of $\sigma$ basis vectors, where $\sigma$ is

chosen by the attacker. There is a comment in [80, Section 8] that moving from binary secrets to ternary secrets would make hybrid attacks "substantially harder"—clearly the attacker cannot afford such a large $\sigma$ if the search space has size $3^\sigma$ instead of $2^\sigma$—but there was no claim in [80] that this change would *eliminate* hybrid speedups.

2009 Hirschhorn–Hoffstein–Howgrave-Graham–Whyte [71] and subsequently 2015 Hoffstein–Pipher–Schanck–Silverman–Whyte–Zhang [73] reported hybrid speedups for various ternary encryption systems. These papers did not claim that there was a sharp dividing line between $2^\sigma$, $3^\sigma$, $5^\sigma$, etc. The obvious reason for these papers to focus on ternary encryption systems is simply that there were concrete proposals of ternary encryption systems.

A hybrid attack against a signature system with secret entries $-2, -1, 0, 1, 2$ was outlined by 2013 Ducas–Durmus–Lepoint–Lyubashevsky [52, full version, Appendix A.5], and was claimed in [52, footnote 2] to be faster than other known attacks against the system. There was no claim that the same type of attack is inapplicable to encryption systems. One can easily argue that the analysis in [52] is too optimistic for the attacker—there are, e.g., silent assumptions that searching cost is as small as an entropy-based lower bound, and that meet-in-the-middle probabilities (see Section 4.3) are 1—but the point here is that, whether or not the paper is accurate, the paper did not claim that non-ternary KEMs are immune to hybrid attacks.

What, then, is the basis for the NewHope claim that hybrid "vulnerabilities" are "induced by the fact that secrets are ternary"? The NewHope paper [12] provides no justification for this claim, and provides no citations regarding hybrid attacks except for Howgrave-Graham's paper—a paper that does not actually make this claim.

Similarly, the Kyber documentation provides no justification for the idea that Kyber is immune to hybrid attacks. Given the standardization processes underway for Kyber, one would expect a much more clear paper trail justifying the omission of hybrid attacks from the Kyber security analysis.

The idea that hybrid attacks save time only for "sufficiently narrow" secret distributions was questioned in 2020 Bernstein–Lange [30, page 26]: "Is this true, or a calculation error in existing algorithm analyses?" Another possibility is that the requisite analyses never even took place: perhaps the evidence of hybrid attacks mattering for ternary systems was misinterpreted as evidence of hybrid attacks not mattering for non-ternary systems.

**1.2. Contributions of this paper.** Fix $w \in \{3, 5, 7, \ldots\}$. The main point of this paper is that hybrid primal key-recovery attacks against the LPR PKE with error width $w$ are, according to existing heuristics commonly used to analyze the performance of lattice attacks, exponentially faster than standard non-hybrid primal key-recovery attacks against the same system.

Quantitatively, consider the key-recovery problem reviewed in Section 2.1 with

- lattice dimension $n \to \infty$,
- modulus $q$ growing as $n^{Q_0+o(1)}$ for any constant $Q_0 > 1/2$, and
- standard deviation $s$ for any constant $s > 0$.

The existing heuristics imply, by calculations reviewed in Section 2, that the cost of the standard non-hybrid non-quantum primal key-recovery attack is $2^{(\rho+o(1))n}$ where $z_0 = 2Q_0/(Q_0+1/2)^2$ and $\rho = z_0 \log_4(3/2)$. If all entries in the secret vectors are in $\{-(w-1)/2, \ldots, (w-1)/2\}$ then the existing heuristics also imply, by new calculations in Section 3, that the cost of a *simple* width-$w$ hybrid non-quantum primal key-recovery attack is $2^{(\rho-\rho H_0+o(1))n}$ where $H_0 = 1/(1+(\lg w)/0.057981 z_0)$.

The difference $\rho H_0$ is positive for all $w$. Consequently, either the proposals [12], [13], etc. are wrong in assuming that hybrid primal speedups are limited to ternary secrets (i.e., to $w = 3$), or the heuristics used for previous attack analyses are wrong, or both.

**1.3. Which hybrid attacks?** Clarity and reviewability of attack analyses are always important, but are particularly important for a paper disputing previous work. This is why Section 3's quantification of the hybrid improvement focuses on *simple* hybrid attacks, skipping speedups that are not necessary for the main point. For example, simple hybrid attacks have 0 levels of collision search; these are much easier to analyze than hybrid attacks with more levels of collision search.

This paper is not claiming that the simple hybrid attack analyzed in Section 3 is the optimal hybrid attack. Section 4 surveys known directions of speedups in hybrid attacks; it would be unsurprising for analyses of these speedups to further reduce the heuristic asymptotic exponent. The necessary analyses will require more work and, in most cases, hybrid-specific heuristics.

**1.4. Interactions with memory-access costs.** To maximize comparability, this paper measures cost using the notion of "instructions" from [7], the source of the "gate" counts in the latest Kyber proposal [13]. This notion allows a single instruction to read or write data from an array of any size, as in introductory courses on algorithms.

In the broader cryptographic literature, attack analyses are sometimes carried out in more realistic cost metrics where the cost of communication grows with the communication distance. *Asymptotic* analyses of this type include [20], [21], [29], [17], and [32].

One can, with some work, check that all sieving algorithms in the literature have heuristic asymptotic exponents above $\log_4(3/2)+o(1)$ in these more realistic metrics, although this has to be taken with a grain of salt since there has not been much effort to optimize these exponents in those metrics. If the costs of memory access in fact push sieving to a higher exponent than $\log_4(3/2) + o(1)$ then presumably they will also increase the 0.057981 above, increasing the exponent fraction eliminated by hybrid attacks. In the opposite direction, presumably any improvements in sieving will reduce the relative influence of hybrid attacks.

The literature gives the impression that sieving-based lattice attacks fail when the optimal amount of memory for the standard primal attack is not available. For example, [116] states the following: "If you only have $2^{90}$ bits of memory, then you cannot iterate a BDGL-based sieve. So you cannot use BDGL to find

dimension 384 vectors that are as short as you like. You cannot use BDGL to break Kyber512." This is contradicted by hybrid attacks, which (1) naturally consume less memory than the standard non-hybrid attack and (2) provide smooth time-memory tradeoffs to fit into even less memory. See Section 3.6.

**1.5. Caveats regarding concrete attack costs.** Analysis of asymptotic exponents is a normal first step in understanding algorithm performance. It is important to keep in mind, however, that those exponents say nothing about performance for any specific problem size.

If $F$ has a smaller exponent than $G$, then $F(n)$ is smaller than $G(n)$ for all sufficiently large $n$, but $F(n)$ could be larger than $G(n)$ for, e.g., $n = 512$. There can also be errors in the opposite direction: the ratio $F(512)/G(512)$ can be even smaller than one might guess from the asymptotics. One needs to quantify $F$ and $G$ in more detail to draw conclusions about specific values of $n$.

The heuristic exponential speedup described in this paper implies a heuristic speedup for all sufficiently large $n$. This does *not* establish that there is a heuristic speedup for $n = 512$. Given the balancing allowed by the structure of hybrid attacks, and given various speedup ideas in the literature on hybrid attacks (see Section 4), it would not be surprising to see a speedup for $n = 512$; but such a speedup is not demonstrated, let alone quantified, in this paper.

These warnings should not be taken as endorsing the literature's omission of hybrid attacks from the analyses of a wide range of concrete lattice-based cryptosystems. On the contrary, the omission is unjustified, and could easily be disastrous. Careful analysis of the concrete impact of hybrid attacks is required.

**1.6. Evaluating attack stability.** One of the fundamental arguments for caution regarding lattice-based cryptography is that the estimated security levels of lattice systems keep dropping. Measuring the short-term instability here is difficult because of various sources of noise, but the following difference in dimensions is clear:

- FrodoKEM [**11**], which is frequently portrayed as the most conservative lattice system, says it is an "instantiation and implementation" of a 2010 cryptosystem [**97**] from Lindner and Peikert.
- The Lindner–Peikert paper proposed lattice dimension 256, saying this was as hard to break as AES-128: "For the 'medium security' ($n = 256$) parameter set, the best runtime/advantage ratio is approximately $2^{120}$ seconds, which translates on our machine to about $2^{150}$ operations. It seems reasonable to conclude that these parameters currently offer security levels at least matching those of AES-128."
- Given subsequent attacks, it is no longer reasonable to suggest that breaking this dimension-256 proposal costs "about $2^{150}$ operations". Within proposals claimed today to be as hard to break as AES-128, the most aggressive proposals use dimensions around 512, and FrodoKEM uses dimension 640.

Will users next decade be worrying about the safety of dimension 1024? 2048?

A counterargument is as follows: (1) the security loss since 2010 came from a burst of work improving the asymptotic exponents of lattice attacks, (2) those

exponents stabilized years ago, and (3) subexponential improvements have much less effect on security levels. Consider, for example, the following statement from NIST [3, page 84]: "The performance of sieving algorithms has been improving [306–314], however recent results [315] indicate that improvements in locally sensitive hash techniques, which have resulted in the largest decreases in asymptotic complexity for sieving thus far, cannot be improved further."

There are three structural flaws in this counterargument. First, while it is certainly true that using better locality-sensitive hash functions produced better exponents for sieving algorithms, an asymptotic optimality result for locality-sensitive hash functions does not imply asymptotic optimality for sieving algorithms. Formally, this logical gap is clear from the cited optimality paper [90] saying that the class of sieving algorithms it considers covers "*almost* all approaches to date" (emphasis added). Furthermore, the paper says its "lower bounds" are "$0.292d + o(d)$ for large $d$ (or $0.265d + o(d)$ quantumly)", but [43] and [69] are two papers from 2021 reporting heuristic quantum exponents around 0.257, which is below 0.265. Those exponent improvements are by only a small percentage, but they contradict the claimed "lower bounds" on the cost of sieving. The theorems also do not rule out larger improvements and non-quantum improvements.

Second, even if one ignores quantum computation and assumes that the best possible non-quantum sieving exponent is $\log_4(3/2) \approx 0.292481$ (which, again, is not proven in [90]), this does not imply that a lattice attack *using* sieving has reached optimal exponents. There are other layers in lattice attacks. This paper's results regarding hybrid attacks show that paying attention to *those* layers produces—according to the existing heuristics—exponential speedups. The heuristic asymptotic exponents of lattice attacks have not stabilized.

Third, the gap between asymptotic costs and concrete costs is a two-edged sword. It is, as noted in Section 1.5, not safe to assume that an asymptotic speedup means a speedup for, e.g., $n = 512$; but it is also not safe to assume that asymptotic stability means stability for $n = 512$. On the contrary, one would expect concrete attack costs to take even longer to stabilize than asymptotic exponents.

**1.7. Caveats regarding the heuristics.** This paper is not claiming that the existing heuristics are accurate, nor is it collecting any evidence for the accuracy of the heuristics. This paper is considering consequences of the hypothesis that the heuristics are accurate.

It is important to keep in mind that the heuristics can be wrong. See [31] for an example of a class of lattices for which earlier literature had obtained incorrect attack conclusions by applying the same heuristics.

**1.8. Priority dates.** One aspect of this paper was announced earlier: in [24], I outlined why existing heuristics imply that a simple hybrid attack produces an exponential speedup. This paper does more work to quantify the exponents, building on the asymptotic analysis from [25].

## 2 The standard primal attack

This section reviews the key-recovery problem for the LPR cryptosystem; the standard primal key-recovery attack; the standard analysis of the attack cost, assuming existing heuristics; the standard analysis of the required attack parameters, again assuming existing heuristics; and the asymptotics of those parameters.

The resulting first-order asymptotics of the heuristic attack cost are as follows. Fix a real number $Q_0 > 1/2$ and an error distribution $\chi$ supported on a finite set of integers. Assume that $q \in n^{Q_0+o(1)}$ as $n \to \infty$, where $q$ is the modulus and $n$ is the lattice dimension. Assume that $\chi$ has zero average and nonzero standard deviation. Write $z_0 = 2Q_0/(Q_0 + 1/2)^2$ and $\rho = z_0 \log_4(3/2)$. The cost is then $2^{(\rho+o(1))n}$. For example, $\rho$ is approximately 0.187188 for $Q_0 = 2$; 0.259983 for $Q_0 = 1$; and 0.292481 for $Q_0$ close enough to $1/2$.

These asymptotics are the baseline performance for this paper. Section 3 shows that hybrid primal attacks are exponentially faster than that, assuming existing heuristics.

**2.1. The target cryptosystem.** For comparability of this paper's analysis with the asymptotic analysis from [25], this paper targets exactly the PKE defined in [25, Section 2.3]:

> This PKE has three parameters: an integer $n \geq 2$; an integer $q \geq 2$; and a probability distribution $\chi$ supported on a finite set of integers. Assume for simplicity that the average of $\chi$ is 0. Write $R$ for the ring $\mathbb{Z}[x]/(x^n + 1)$.
>
> Key generation works as follows. Generate uniform random $G \in R/q$. Generate $a, e \in R$ with coefficients drawn independently at random from $\chi$. Compute $A = aG + e \in R/q$. The secret key is $(a, e)$. The public key is $(G, A) \in (R/q)^2$.

The definition in [25] also includes encryption and decryption, but this paper (except in Section 4.6) focuses on attacks that recover the secret key from the public key, i.e., attacks against "normal-form 1-sample search Ring-LWE".

As noted in [25, Section 2.3], the 2010 Lyubashevsky–Peikert–Regev PKE [99, Section 1.1] included further restrictions: it required $n$ to be a power of 2, and required $q$ to be a prime congruent to 1 modulo $2n$. This paper, like [25], does not impose these requirements: the requirements exclude various followup cryptosystems and do not interact with this paper's attack analysis. Readers are cautioned that the structure of $(n, q)$ can be relevant to more sophisticated attacks: e.g., if $n = 768$ then the equation $A = aG+e$ in $R/q$ implies the equation $A = aG + e$ in the smaller ring $(\mathbb{Z}/q)[x]/(x^{256} + 1)$.

**2.2. Details of the standard primal key-recovery attack.** This paper's baseline attack is exactly the attack reviewed in [25, Section 2.4], which is quoted here except for a change of notation in the definition of $L$:

There is an attack parameter $\kappa \leq n$. Define a function $\text{First}_\kappa : R \to \mathbb{Z}^\kappa$ that extracts the first $\kappa$ coefficients from its input. This induces a function, also written $\text{First}_\kappa$, from $R/q$ to $(\mathbb{Z}/q)^\kappa$.

Define $L$ as the set of all $(\alpha, \epsilon, \gamma) \in R \times \mathbb{Z}^\kappa \times \mathbb{Z}$ such that $\text{First}_\kappa(\gamma A - \alpha G)$ is the same as $\epsilon$ modulo $q$. This is a lattice of full rank $d = n + \kappa + 1$ and determinant $q^\kappa$. Note that $\pm(a, \text{First}_\kappa(e), 1)$ are elements of this lattice.

There is another attack parameter $\beta$. The attack writes down a basis for $L$, applies BKZ-$\beta$ to reduce this basis, and hopes that BKZ-$\beta$ outputs one of the short nonzero vectors $\pm(a, \text{First}_\kappa(e), 1)$, in particular revealing $a$.

Aside from the question of which BKZ variant is plugged in, this attack is from 1998 Hoffstein–Pipher–Silverman [74, Section 3.4.2] (see also the 1996 preprint [75]) in the case $\kappa = n$. The generalization to any $\kappa \leq n$ is from 2001 May–Silverman [103]. This paper ignores the scaling generalization from 1997 Coppersmith–Shamir [47]; that generalization is important for cryptosystems that choose $a$ and $e$ to have different sizes, but has negligible effect on the cryptosystem from Section 2.1. See [25, Section 2.4] for the history of the subsequent "Ring-LWE" naming.

**2.3. The standard heuristic analysis of the attack cost.** The standard heuristic analysis of BKZ-$\beta$ says that 2016 Becker–Ducas–Gama–Laarhoven [19] solves SVP-$\beta$ with heuristic cost $2^{(\log_4(3/2)+o(1))\beta}$, and that BKZ-$\beta$ is is within a subexponential factor of SVP-$\beta$, hence also heuristic cost $2^{(\log_4(3/2)+o(1))\beta}$.

The $\log_4(3/2)$ exponent depends on the model of computation. This paper's quantification of attack costs does not consider costs of memory access, as noted in Section 1.4, and does not consider quantum computation.

**2.4. The standard parameter choice.** Assume $\beta \geq 60$ (as explained in, e.g., [25, Section 2.6]). The standard analysis of the standard primal attack is the analysis from the NewHope paper [12, Section 6.3], which uses various heuristics to conclude that the attack works if and only if the 2-norm of $(a, \text{First}_\kappa(e), 1)$ is below $(d/\beta)^{1/2}\delta^{2\beta-d-1}q^{\kappa/d}$, where $\delta = (\beta(\pi\beta)^{1/\beta}/(2\pi \exp 1))^{1/2(\beta-1)}$.

The squared 2-norm of $(a, \text{First}_\kappa(e), 1)$ is $(n + \kappa)s^2 + 1$ on average, where $s$ is the standard deviation of $\chi$. The analyses in [12], [8], [13], etc. heuristically treat the squared 2-norm as being exactly its average.

(The squared 2-norm is almost certainly close to its average. The differences are not visible in the first-order asymptotics used in this paper and do not matter for evaluating exponential speedups. The differences *are* visible in second-order asymptotics: [25, Section 4] exploits cases where the squared 2-norm is smaller than average to obtain a subexponential speedup for one-out-of-many-ciphertext attacks.)

To summarize, the existing heuristics imply that the attack works if and only if $\text{StandardRatio}(n, q, s, \kappa, \beta) < 1$, where StandardRatio is defined as follows. This definition is copied from [25, Definition 2.5.1].

**Definition 2.4.1 (the standard ratio).** *Let $n, q, s, \kappa, \beta$ be real numbers such that $2 \leq n$; $2 \leq q$; $0 < s$; $1 \leq \kappa$; and $2 \leq \beta$. Then* StandardRatio$(n, q, s, \kappa, \beta)$ *is defined as* $((n+\kappa)s^2+1)^{1/2}/(d/\beta)^{1/2}\delta^{2\beta-d-1}q^{\kappa/d}$ *where* $d = n+\kappa+1$ *and* $\delta = (\beta(\pi\beta)^{1/\beta}/(2\pi\exp 1))^{1/2(\beta-1)}$.

For a review of the underlying heuristics, see [25, Section 2.5]. which also notes that subsequent speedups and corrections are portrayed by [6] as minor. Different heuristics might produce different asymptotics, but the NewHope heuristics have not been withdrawn, and this paper focuses on those.

The standard parameter choice minimizes $\beta$ subject to the condition StandardRatio$(n, q, s, \kappa, \beta) < 1$, and minimizes $\kappa$ subject to this choice of $\beta$. The point of prioritizing minimization of $\beta$ is that the heuristic cost of the attack grows exponentially with $\beta$; see Section 2.3.

**2.5. Asymptotics for the standard parameter choice.** [25, Theorem 2.7.1] identifies the second-order asymptotics of the minimum $\beta$ among all $(\beta, \kappa)$ satisfying StandardRatio$(n, q, s, \kappa, \beta) < 1$, assuming $\lg q$ grows as $Q_0 \lg n + Q_1 + o(1)$ and $\lg s$ grows as $S_0 \lg n + S_1 + o(1)$, with $0 \leq S_0 \leq 1/2 < Q_0 - S_0$.

Those second-order asymptotics imply simpler first-order asymptotics: the minimum $\beta$ has $\beta/n \in z_0 + o(1)$ where $z_0 = 2Q_0/(Q_0 - S_0 + 1/2)^2$. One can derive these first-order asymptotics from first-order hypotheses on the parameters, namely that $\lg q \in (Q_0 + o(1)) \lg n$ and $\lg s \in (S_0 + o(1)) \lg n$, i.e., $q \in n^{Q_0+o(1)}$ and $s \in n^{S_0+o(1)}$. (See `standardratio_below` and `standardratio_above` in the formally verified theorems accompanying [25].)

The case of interest in this paper is that $\chi$ is independent of $n$, so the standard deviation $s$ of $\chi$ is also independent of $n$, so $S_0 = 0$ and $1/2 < Q_0$. The minimum $\beta$ satisfying StandardRatio$(n, q, s, \kappa, \beta) < 1$ then has $\beta/n \in z_0 + o(1)$ with $z_0 = 2Q_0/(Q_0 + 1/2)^2$. Plugging this into the standard heuristic analyses (see Sections 2.3 and 2.4) produces the conclusion that the heuristic attack cost is $2^{(\log_4(3/2)+o(1))\beta} = 2^{(\rho+o(1))n}$ where $\rho = z_0 \log_4(3/2)$.

Readers focusing on the case that $q$ is a prime congruent to 1 modulo $2n$ should note that in this case $1 \leq Q_0$. Conversely, Heath-Brown conjectured in [68] (see also [123] for numerical evidence) that the smallest prime congruent to 1 modulo $k$ is $O(k(\log k)^2)$; this conjecture implies that taking $Q_0 = 1$ is compatible with taking $q$ specifically as the smallest prime congruent to 1 modulo $2n$.

## 3 Analysis of a simple hybrid multi-decoding attack

This section highlights a specific type of hybrid attack and computes what the existing heuristics imply about this attack. The main result is as follows.

Recall from Section 2 that the existing heuristics imply that the standard primal attack costs $2^{(\rho+o(1))n}$, where $z_0 = 2Q_0/(Q_0+1/2)^2$ and $\rho = z_0 \log_4(3/2)$, assuming $q \in n^{Q_0+o(1)}$ with $Q_0 > 1/2$.

Fix $w \in \{3, 5, 7, \ldots\}$. Assume that the error distribution $\chi$ is supported on $\{-(w-1)/2, \ldots, -1, 0, 1, \ldots, (w-1)/2\}$. The existing heuristics then imply

that the width-$w$ hybrid attack in this section costs $2^{(\rho-\rho H_0+o(1))n}$, where $H_0 = 1/(1 + (\lg w)/0.057981z_0)$.

Compared to the standard primal attack, this attack reduces heuristic costs by a factor $2^{(\rho H_0+o(1))n}$. This is always an exponential factor: $z_0$, $\rho$, and $H_0$ are positive no matter what $Q_0$ and $w$ are.

**3.1. Overview of the attack components.** Consider first the following situation: some coefficients of the secret $a$ are already known. Specifically, write $a$ uniquely as $u+v$ where $u$ is supported on the first $\sigma$ coefficients and $v$ is supported on the last $n-\sigma$ coefficients, and assume that $u$ is known. Section 3.4 reviews how the knowledge of $u$ efficiently reduces the attack problem to a "bounded-distance decoding" problem for a lattice $L'$ having smaller rank than the lattice $L$ that was used in Section 2.2. Section 3.5 quantifies the asymptotics of how much time is saved in this situation as $\sigma$ grows.

Section 3.6 moves from this situation to the actual attack situation, where $u$ is *not* known in advance, and quantifies the performance of a simple hybrid attack that runs through guesses for $u$. See Section 4 for a survey of more advanced hybrid attacks.

One might think that the number of guesses for $u$ outweighs the per-$u$ speedup unless the error width $w$ is below some small cutoff. However, suitable algorithms for the decoding problem spend most of their time on precomputations that depend only on $L'$, and spend exponentially less time on each per-$u$ computation. The definition of $L'$ does not depend on $u$; precomputations for $L'$ are reused for all of the guesses for $u$. This allows an exponential number of guesses, producing an exponential speedup for any $w$.

The necessary subroutines from the literature are explained first: Section 3.2 reviews what the literature says about the performance of an algorithm to find closest vectors after precomputation, and Section 3.3 reviews what the literature says about the performance of an algorithm for bounded-distance decoding after precomputation.

**3.2. Closest vectors after precomputation ("CVPP").** This paper uses the following result from Ducas–Laarhoven–van Woerden [**53**] (improving on results of Laarhoven [**92**] and Doulgerakis–Laarhoven–de Weger [**51**]) as a black box: a computation with heuristic cost $2^{(\log_4(3/2)-0.057981+o(1))\beta}$ finds an element of $L$ closest to $t$, given a rank-$\beta$ lattice $L$ and a vector $t$, after an $L$-dependent $t$-independent precomputation with heuristic cost $2^{(\log_4(3/2)+o(1))\beta}$.

One can almost see these numbers from [**53**] saying that one can "heuristically solve CVPP instances in $2^{0.234d+o(d)}$ amortized time, for batches of size at least $2^{0.058d+o(d)}$". The numbers "0.234" and "0.058" are rounded; one has to look at the structure of the algorithm analysis in [**53**] to see that the exact numbers have sum $\log_4(3/2)$, reflecting a precomputation with cost $2^{(\log_4(3/2)+o(1))\beta}$ as stated above. This also forces the exact second number to be at least 0.057981, since if it were below 0.057981 than the exact first number would be above 0.2345 and would not round to 0.234. (A more precise recomputation of the 0.058 would, by essentially the same argument, have to give something larger than 0.057981,

and would give slightly better hybrid exponents, but would take more effort to check.)

The structure of [53] also shows that each target vector $t$ is handled separately after the $L$-dependent precomputation, rather than actually requiring any particular batch size. However, hybrid attacks do generate batches of targets (this is clear from the algorithm structure and is highlighted in, e.g., [5]), and one can choose parameters in hybrid attacks to rely purely on the batch statement quoted above. Faster batch CVP algorithms, whether or not structured as CVP algorithms after precomputation, would accelerate hybrid attacks.

**3.3. Bounded-distance decoding after precomputation ("BDDP").** The problem of bounded-distance decoding has a real number $D > 0$ as a parameter. A problem instance provides the following information about two vectors $x, y$: the sum $x + y$; a lattice $L$ containing $y$; and the promise that $x$ has 2-norm at most $D$. Solving the problem instance means finding $(x, y)$.

One approach to decoding is "embedding" (from 1983 Kannan [87, Lemma 3]), where one converts this close-vector problem into a short-vector problem. For example, look for a short vector $(x, 1)$ in the lattice of pairs $(\alpha, \gamma)$ such that $\gamma \in \mathbb{Z}$ and $\gamma(x + y) - \alpha \in L$; more generally, pick a scale factor $\lambda$ and then look for a short vector $(x, \lambda)$ in the lattice of pairs $(\alpha, \gamma)$ such that $\gamma \in \lambda\mathbb{Z}$ and $\gamma(x + y) - \alpha\lambda \in L$.

Embedding is a reasonable starting approach—it is used inside the standard primal attack, for example—but hybrid attacks instead rely on more sophisticated decoders with the following two-stage structure:

- Perform precomputations on the lattice $L$, independently of the vector $x+y$.
- Use those precomputations to reduce $x + y$.

In the paper [80] introducing hybrid attacks, Howgrave-Graham used BKZ-$\beta$ for the first stage, and the "nearest plane" algorithm ("weak reduction") for the second stage. The nearest-plane algorithm had been analyzed implicitly by 1982 Lenstra–Lenstra–Lovász [96] and explicitly by 1986 Babai [16].

Howgrave-Graham briefly mentioned the possibility of doing better by "mixing Babai's CVP (which is essentially blocksize 1) with searching in higher blocksizes $2, 3, \ldots$". This block-size-$\beta$ generalization of the nearest-plane algorithm is now called the "nearest-colattice algorithm". This name comes from 2020 Espitau–Kirchner [57], along with an analysis of the algorithm.

The specific algorithm variant that matters here is the algorithm from [57, Section 5, third paragraph]: first apply BKZ-$\beta$ to find a "highly reduced basis" of the input lattice $L$; then compute a "CVP on the tail of the basis" to find a vector in the $\beta$-dimensional projection of $L$ closest to the projection of the input vector $x + y$, hopefully the projection of $y$; then "finish with Babai's algorithm", hopefully finding $y$.

This algorithm has the desired two-stage structure. The BKZ-$\beta$ computation, along with the precomputation for CVP-$\beta$ from Section 3.2, depends only on $L$, and has heuristic cost $2^{(\log_4(3/2)+o(1))\beta}$. Each input vector $x + y$ is then handled with heuristic cost just $2^{(\log_4(3/2)-0.057981+o(1))\beta}$.

When does the algorithm succeed? The existing heuristics are reviewed in [25, Section 3.1] and imply that the algorithm works if the distance $D$ is below $(1+o(1))(d/\beta)^{1/2}\delta^{2\beta-d-1}(\det L)^{1/d}$, where $d$ is the rank of $L$ and $\delta = (\beta(\pi\beta)^{1/\beta}/(2\pi\exp 1))^{1/2(\beta-1)}$.

Note that this heuristic distance bound is within a factor $1+o(1)$ of the heuristic shortness bound used in Section 2.4. In other words, using BKZ-$\beta$ and CVP-$\beta$ to find a close vector has, at this level of asymptotic detail, the same heuristic effectiveness as using BKZ-$\beta$ to find a short vector.

This is more detail than necessary for seeing exponential speedups: one can instead use simpler first-order asymptotics. Specifically, $\delta^{2(\beta-1)} \in \Theta(\beta)$, so the heuristics imply that the algorithm works if $D$ is below $(d/\beta)^{1/2}\Theta(\beta)^{1-d/2\beta}(\det L)^{1/d}$. The heuristics also imply that projecting from dimension $d$ to dimension $\beta$ multiplies 2-norm by $(\beta/d)^{1/2}$, and thus that the algorithm works if the projection of $x+y$ has 2-norm below $\Theta(\beta)^{1-d/2\beta}(\det L)^{1/d}$. This matches what is stated in [57, Section 5, third paragraph].

**3.4. Exploiting partial information.** Recall the public lattice $L$ from Section 2.2, the set of all $(\alpha,\epsilon,\gamma) \in R \times \mathbb{Z}^\kappa \times \mathbb{Z}$ such that $\operatorname{First}_\kappa(\gamma A - \alpha G)$ is the same as $\epsilon$ modulo $q$. The standard primal attack uses BKZ-$\beta$ to (hopefully) find the short elements $\pm(a, \operatorname{First}_\kappa(e), 1)$ of this lattice.

Define $L'$ as the set of all $(\omega,\epsilon) \in R \times \mathbb{Z}^\kappa$ such that $\operatorname{First}_\kappa(-\omega G)$ is the same as $\epsilon$ modulo $q$ and $\omega$ is supported on the last $n-\sigma$ coefficients; here $\sigma$ is another attack parameter. This is a lattice of rank $n-\sigma+\kappa$ and determinant $q^\kappa$.

(The literature often states algorithms purely for full-rank lattices. To fit that restriction, simply suppress the first $\sigma$ coefficients of $\omega$, all of which are 0 by definition, to convert $L'$ into a full-rank lattice.)

Now write $a$ uniquely as $u+v$ where $u$ is supported on the first $\sigma$ coefficients and $v$ is supported on the last $n-\sigma$ coefficients, as in Section 3.1; and consider the situation that $u$ is known.

By definition $A = aG + e = uG + vG + e$ in $R/q$, so $-vG = e + uG - A$, so $\operatorname{First}_\kappa(-vG) = \operatorname{First}_\kappa(e+uG-A)$, so $(v, \operatorname{First}_\kappa(e+uG-A)) \in L'$. The known vector $(0, \operatorname{First}_\kappa(uG-A))$ is close to this vector $(v, \operatorname{First}_\kappa(e+uG-A)) \in L'$: the difference is the short vector $(v, \operatorname{First}_\kappa(e))$.

This is the desired reduction to a decoding problem: namely, the problem of finding the vector $(v, \operatorname{First}_\kappa(e+uG-A)) \in L'$ close to the known vector $(0, \operatorname{First}_\kappa(uG-A))$.

**3.5. The speedup from partial information.** Now consider the question of how large $\beta$ needs to be asymptotically for the decoding algorithm from Section 3.3 to solve the decoding problem from Section 3.4. This analysis assumes $\sigma \in (H+o(1))n$ with $0 < H < 1$; the choice of $H$ will be optimized in Section 3.6.

The squared 2-norm of the short vector $(v, \operatorname{First}_\kappa(e))$ from Section 3.4 is $(n-\sigma+\kappa)s^2$ on average, where $s$ is the standard deviation of $\chi$. As noted in Section 2.4, the squared 2-norm is almost certainly close to its average, and the gap does not affect first-order asymptotics; in any case, the existing heuristics treat the squared 2-norm as exactly its average.

The heuristics reviewed in Section 3.3 for the decoding algorithm then imply that the algorithm works if $(n-\sigma+\kappa)^{1/2}s$ is below $(1+o(1))(d/\beta)^{1/2}\delta^{2\beta-d-1}q^{\kappa/d}$, where $d = n-\sigma+\kappa$ and again $\delta = (\beta(\pi\beta)^{1/\beta}/(2\pi\exp 1))^{1/2(\beta-1)}$.

This is equivalent to StandardRatio$(n-\sigma-1,q,s,\kappa,\beta) < 1+o(1)$, almost by definition of StandardRatio. The only change is between $(n-\sigma+\kappa)s^2$ and $(n-\sigma-1+\kappa)s^2+1$; these have constant difference, and converge to $\infty$ (since $s>0$ by assumption and $n-\sigma \in (1-H+o(1))n$ with $H<1$ by assumption), so they have ratio $1+o(1)$.

Note that $q \in (n-\sigma-1)^{Q_0+o(1)}$, since $\lg(n-\sigma-1) \in (1+o(1))\lg n$. The asymptotic analysis of StandardRatio from [25] then says that the minimum $\beta$ with StandardRatio$(n-\sigma-1,q,s,\kappa,\beta) < 1+o(1)$ has $\beta/(n-\sigma-1) \in z_0+o(1)$; equivalently, $\beta/n \in (1-H)z_0+o(1)$.

The precomputation from Section 3.3, namely BKZ-$\beta$ followed by a CVP-$\beta$ precomputation, then costs $2^{(\log_4(3/2)+o(1))\beta} = 2^{(\rho-\rho H+o(1))n}$. This is better than the standard primal attack by a factor $2^{(\rho H+o(1))n}$: in short, knowing an $H$ fraction of the $n$ entries in $a$ cuts an $H$ fraction out of the asymptotic exponent.

The main computation, decoding $(0,\text{First}_\kappa(uG-A))$, is even faster: it costs just $2^{(\log_4(3/2)-0.057981+o(1))\beta}$, better than the precomputation by a factor $2^{(0.057981+o(1))\beta} = 2^{(0.057981(1-H)z_0+o(1))n}$.

**3.6. Multiple guesses for the partial information.** Now let's drop the assumption that $u$ is known, and instead let's search through possibilities for $u$, taking advantage of the fact that only a tiny fraction of the cost in Section 3.5 is from the main $u$-dependent computation.

Fix $w \in \{3,5,7,\dots\}$. Consider the following simple width-$w$ hybrid attack:

- Apply the precomputation from Section 3.3 to the lattice $L'$ from Section 3.4.
- Search all $w^\sigma$ choices of $u \in R$ where the last $n-\sigma$ coefficients are 0 and the first $\sigma$ coefficients are in $\{-(w-1)/2,\dots,-1,0,1,\dots,(w-1)/2\}$.
- For each $u$, apply the main computation from Section 3.3 to try to find an element of $L'$ close to $(0,\text{First}_\kappa(uG-A))$.

If $\chi$ is supported on $\{-(w-1)/2,\dots,-1,0,1,\dots,(w-1)/2\}$ then this attack will at some point guess the correct $u$, matching the first $\sigma$ coefficients of $a$.

Assume $\sigma \in (H+o(1))n$ as in Section 3.5. Then the precomputation costs $2^{(\rho-\rho H+o(1))n}$ as in Section 3.5. There are also $w^\sigma = 2^{\sigma\lg w} \in 2^{(H\lg w+o(1))n}$ decoding steps, each of which costs $2^{(\rho-\rho H-0.057981(1-H)z_0+o(1))n}$, for total attack cost $2^{(\rho-\rho H+o(1))n} + 2^{(\rho-\rho H-0.057981(1-H)z_0+H\lg w+o(1))n}$.

Now define $H_0 = 1/(1+(\lg w)/0.057981z_0)$, and choose specifically $H = H_0$. Then $H\lg w = 0.057981(1-H)z_0$, so the decoding steps together cost $2^{(\rho-\rho H+o(1))n}$, just like the precomputation. The total heuristic attack cost is then $2^{(\rho-\rho H_0+o(1))n}$ as claimed.

As mentioned in Section 1, this algorithm also offers smooth time-memory tradeoffs, reducing concerns about subroutines that use memory exponential in $\beta$. Specifically, increasing $H$ beyond $H_0$ gradually increases the cost exponent $\rho-\rho H-0.057981(1-H)z_0+H\lg w$, but also gradually decreases $\beta$ since $\beta \in ((1-H)z_0+o(1))n$.

## 4 Known unknowns

Howgrave-Graham's original hybrid-attack paper [80] started by describing "an algorithm that enumerates all possible $v$" (as in Section 3, which has $u$ where [80] has "$v$"), but continued with a more sophisticated meet-in-the-middle approach. One might hope for the meet-in-the-middle approach and more sophisticated combinatorial approaches to reduce $w^\sigma$ guesses to $w^{(1/2+o(1))\sigma}$ guesses or even lower, decreasing the heuristic asymptotic exponent for the full attack cost from $\rho - \rho/(1 + (\lg w)/0.057981z_0)$ to $\rho - \rho/(1 + 0.5(\lg w)/0.057981z_0)$ or even lower. However, a full analysis of the meet-in-the-middle approach is challenging; see Section 4.3.

More broadly, the literature presents ideas in several different directions for speeding up hybrid attacks, and presents reasons to believe that many of these ideas produce speedups. This section surveys these ideas. Each direction—and each combination of directions—raises new analysis questions. It is plausible, in light of the results of Section 3, that many of these ideas produce changes in the heuristic asymptotic attack exponent. Concrete security analysis should consider all of the ideas; as noted in Section 1.5, asymptotic exponents do not say anything about any particular concrete size.

Sometimes the literature considers quantum algorithms: for example, one can trivially replace $w^\sigma$ guesses with approximately $w^{\sigma/2}$ iterations of Grover search. This requires recalculation of asymptotic exponents and concrete costs.

**4.1. Exploiting non-uniform error distributions.** As a preliminary point, note that the attack from Section 3.6 can, and almost always does, finish early. There are $w^\sigma$ guesses for $u$; as soon as the attack checks the correct guess, the attack finds $v$, recognizes that $v$ is short, and terminates.

One might think that this is a small constant-factor speedup, for two reasons: first, half the time is spent on precomputation before the guesses begin; second, the attack will try $(1 + w^\sigma)/2$ guesses on average. But neither of these reasons holds up to scrutiny:

- Balancing precomputation with main computation is just one of the options available to the attacker; see Section 3.6. Increasing $\sigma$ means that exponentially less time is spent on precomputation—and then speedups in the main computation become correspondingly more important.
- Typically $\chi$ is non-uniform on its support. Guesses then have different success probabilities, and if the attack tries guesses in decreasing order of probability then it will finish with fewer than $(1 + w^\sigma)/2$ guesses on average, perhaps far fewer.

As an example of non-uniformity, some of Kyber's proposed parameter sets define $\chi$ as a width-5 binomial distribution; i.e., a sum of 4 independent unbiased coins minus 2; i.e., $-2, -1, 0, 1, 2$ with probability $1/16, 1/4, 3/8, 1/4, 1/16$ respectively. Then $a$ matches $u$ on the first $\sigma$ coefficients with probability $(1/16)^{\sigma_{-2}+\sigma_2}(1/4)^{\sigma_{-1}+\sigma_1}(3/8)^{\sigma_0}$ if the first $\sigma$ coefficients of $u$ have exactly $\sigma_j$ entries equal to $j$. It is easy to sort the guesses for $u$ in decreasing order of this

probability. Similar comments apply to parameter sets that define $\chi$ as a sum of 6 unbiased coins minus 3.

A more general attack strategy searches a *subset* of the $w^\sigma$ possibilities for $u$. For example:

- The binomial distribution on $\{-2,-1,0,1,2\}$ produces ternary $u$ with probability $(7/8)^\sigma$, so an attack that searches only $3^\sigma$ ternary vectors, rather than all $5^\sigma$ possible vectors, succeeds with probability $(7/8)^\sigma$.
- Most of the hybrid-attack literature considers a search specifically through the set of $u$ where $j$ occurs $\sigma_j$ times, and then optimizes the attacker's choice of the $\sigma_j$ parameters. For example, Howgrave-Graham considered searching through binary secrets with an attacker-chosen weight.
- One can choose $U \in \{1,\ldots,w^\sigma\}$, list the $U$ highest-probability possibilities using some mechanism of breaking ties, and search those possibilities.

The first example is easy to state but seems inferior to the third: some ternary vectors are less likely to appear than some non-ternary vectors. For example, out of the $5^{50} \approx 2^{116}$ length-50 vectors with entries in $\{-2,-1,0,1,2\}$, there are $3^{50} \approx 2^{79.248}$ ternary vectors, and the binomial distribution produces a ternary vector with probability $(7/8)^{50} \approx 2^{-9.632}$; but a short calculation shows that the binomial distribution produces one of the $3^{50}$ most popular vectors with probability about $2^{-8.013}$. See the `guess50.sage` script attached to this PDF.

The second example, when chosen with maximum-likelihood parameters, is a special case of the third. The third example is due to 2018 Wunderer [**127**, Section 7.3], optimizing the quantum hybrid attack from 2017 Göpfert–van Vredendaal–Wunderer [**65**]. See the NTRU Prime documentation [**26**, Section 6.8] for some non-quantum examples of the same approach.

One can amplify the success probability by running the whole attack again with another selection of $\sigma$ positions in case of failure. The literature does not appear to have explored the extent to which work can be shared across the BKZ-$\beta$ computations for different selections of positions, but it seems likely that choosing suitable subsets produces a smaller heuristic asymptotic exponent for hybrid attacks even without such sharing. Another option, in the quantum context, is to use a Grover search through selections of positions.

A full analysis would have to consider correlations among selections of positions. Consider, for example, an attack that simply tries $u = 0$ for various selections of $\sigma$ positions. The chance that a selection succeeds depends on the number of zeros in the secret; the attack does a better job of targeting secrets with more zeros. For comparison, the attack of [**25**] does a better job of targeting secrets with smaller 2-norm.

**4.2. Optimizing block sizes.** As noted in Section 3.3, Howgrave-Graham used block size $\beta$ for precomputation and block size 1 for the main computation, but mentioned the possibility of using larger block sizes for the main computation.

Section 3 uses block sizes $(\beta,\beta)$ instead of $(\beta,1)$: i.e., block size $\beta$ for the precomputation and for the main computation. A small speedup in this case is

that some variants of BKZ-$\beta$ include the necessary CVP-$\beta$ precomputation, as noted in [25].

The literature typically takes $(\beta, 1)$. There is no reason to think that this is optimal; perhaps this case hides most of the power of hybrid attacks. There is also no reason to think that $(\beta, \beta)$ is optimal. The asymptotic exponents and concrete costs of general pairs $(\beta, \beta')$ are open questions. 2019 Albrecht–Curtis–Wunderer [5] estimated concrete performance of various pairs $(\beta, \beta')$, but without the preprocessing speedups reviewed in Sections 3.2 and 3.3.

**4.3. Collision searches: meet-in-the-middle and beyond.** The classic subset-sum problem asks whether there are elements $u_1, u_2, \ldots, u_\sigma$ of $\{0, 1\}$ such that $u_1C_1 + u_2C_2 + \cdots + u_\sigma C_\sigma = S$, given $C_1, C_2, \ldots, C_\sigma, S$.

There are $2^\sigma$ possible vectors $(u_1, u_2, \ldots, u_\sigma)$. A meet-in-the-middle search takes only about $2^{\sigma/2}$ operations (and 1981 Schroeppel–Shamir [117] reduced the memory usage to about $2^{\sigma/4}$). 2010 Howgrave-Graham–Joux [81] used multiple levels of collision search to reduce the heuristic asymptotic cost exponent below $1/2$, in particular claiming cost exponent 0.311. 2011 Becker–Coron–Joux [18] reached heuristic exponent 0.291, and reported heuristic exponent 0.337 for the algorithm from [81]; May and Meurer had discovered an error in the analysis from [81]. 2019 Esser–May [58] claimed 0.255 but withdrew the claim. 2020 Bonnetain–Bricout–Schrottenloher–Shen [38] then reached 0.283. For quantum exponents, see 2013 Bernstein–Jeffery–Lange–Meurer [28], 2018 Helm–May [70], and [38].

2003 Odlyzko, as reported in [82], adapted a meet-in-the-middle search to the problem of finding small $(a, e)$ given $G$ and $A = aG + e$. Odlyzko decomposed $a$ as $a_0 + a_1$, computed $a_0G$ for all possible $a_0$, and then computed $a_1G$ for all possible $a_1$. There is almost a collision between $A - a_0G$ and $a_1G$ for the correct $(a_0, a_1)$. The only issue is that there is a small difference $e$ between $A - a_0G$ and $a_1G$; Odlyzko addressed this by picking a locality-sensitive hash function and looking for collisions among hashes of $A - a_0G$ and $a_1G$. 2013 Wang–Ma–Ma [124] proposed a quantum version of Odlyzko's attack, and 2016 van Vredendaal [122] proposed time-memory tradeoffs.

Multiple levels of collision search were similarly adapted to this problem by 2021 May [102] in the ternary case and 2022 Glaser–May [64] in further cases. See also 2021 van Hoof–Kirshanova–May [79] for a quantum version in the ternary case.

Howgrave-Graham's main hybrid attack from [80] first decomposed $a$ as $u + v$ as in Section 3, and then decomposed $u$ as $u_0 + u_1$ as in Odlyzko's meet-in-the-middle search. Recall that Section 3 decoded $(0, \mathrm{First}_\kappa(uG - A))$ for each possible $u$; Howgrave-Graham instead decoded $(0, \mathrm{First}_\kappa(u_0G - A))$ for each possible $u_0$ and decoded $(0, \mathrm{First}_\kappa(-u_1G))$ for each possible $u_1$, and looked for collisions among locality-sensitive hashes of the decoding results.

This is no longer bounded-distance decoding: the decoding inputs have no reason to be close to the lattice. One might call it "bounded-distance-pair decoding": the pair of correct inputs $(0, \mathrm{First}_\kappa(u_0G - A))$ and $(0, \mathrm{First}_\kappa(-u_1G))$

has difference close to the lattice, so one can reasonably hope for the decoding outputs to also be close, and to collide after locality-sensitive hashing.

Analyzing the collision probability here is not easy. In [126, Section 4.4], [127, Section 5.3.3], and [128, Section 4.4], Wunderer pointed out various underestimates and overestimates in the meet-in-the-middle probability analyses for the nearest-plane algorithm from previous hybrid-attack papers; see also 2021 Nguyen [111]. The latest heuristics need to be generalized from the nearest-plane algorithm to higher-block-size BDD algorithms, and need to be systematically tested against experiments.

It should be possible to similarly handle multiple levels of collision search: insert a decoding step and locality-sensitive hashing before each equality test. Analyzing the performance of such an algorithm will be even more challenging than analyzing the performance of a single level of collision search.

**4.4. Dual attacks.** A dual attack is most easily phrased as a way to distinguish $A = aG + e$ from uniform: first, given $G$, find a short nonzero vector $x$ such that $Gx = 0$; then, given $A = aG + e$, observe that $Ax = aGx + ex = ex$ tends to be shorter than $Rx$ would be for a uniform random $R$.

More generally, one can use a short nonzero vector $(x, y)$ such that $Gx = y$. Then $Ax = aGx + ex = ay + ex$, which again tends to be shorter than $Rx$ would be for a uniform random $R$.

Even more generally, one can combine information from many short vectors $(x, y)$ to amplify the overall distinguishing probability. Care is required in the analysis to account for vectors not being statistically independent: see 2023.03 Ducas–Pulles [54], 2023 Wiemers–Ehlen [125], 2023.12 Ducas–Pulles [55], and 2023 Carrier–Debris-Alazard–Meyer-Hilfiger–Tillich [42].

To the extent that these distinguishers work, one can easily convert them into secret-recovery attacks by guessing some coordinates of $e$ and using the distinguishers to identify the correct guess. A closer look at the distinguishers shows that the computations for many guesses can be batched via FFTs. Once some coordinates of $e$ are known, finding the remaining coordinates of $e$ is a relatively small problem.

Like a hybrid attack, a dual attack begins with a precomputation stage that depends only on $G$. Guessing some coordinates of $e$ in a dual search attack is analogous to guessing some coordinates of $a$ in a simple hybrid attack without collision search. The guesses are used in different ways; it would be interesting to compute the heuristic asymptotic exponents of dual attacks with and without FFTs, and to compare those to the exponents of various types of hybrid primal attacks. Presumably there will be some cases where algorithms have the same first-order asymptotic exponents but different second-order asymptotic exponents, as in [25, Section 4].

Dual attacks have a long history, going back to 1962 Gallager [62], who introduced the dual-attack idea as a decoder for "low-density parity-check codes" (LDPC codes). 2001 Al Jabri [10] analyzed the performance of a dual attack against random codes. Meanwhile Bleichenbacher had used a dual lattice attack (including FFTs and various other speedup ideas) to break the NSA/NIST DSA

standard in an estimated $2^{64}$ operations, forcing an update of the standard in 2001, although details of the attack were not posted until 2004. See [109, page 72, "unpublished attack"], [121, Section 5], [37], [107], [108], and [23, "Bleichenbacher's attack" section].

Analyses of the effectiveness of simpler dual lattice attacks appeared in 2009 Micciancio–Regev [105, Section 5.4, "Security" paragraphs] for the case of a single $x$, and in the NewHope paper [12, Section 6.4] for the case of many $x$.

2017 Albrecht [4] reported that ternary secrets of constant Hamming weight allow smaller exponents for dual attacks. The algorithm in [4] guessed that some coordinates are all 0 or, more generally, have at most a few positions set, and retried if none of the guesses succeed; this is analogous to the retries in Section 4.1. 2020 Espitau–Joux–Kharchenko [56] instead guessed all possibilities (not necessarily ternary) for some coordinates, as did 2021 Bi–Lu–Luo–Wang–Zhang [36]. 2021 Guo–Johansson [66], crediting Bleichenbacher, used FFTs to merge computations across guesses.

2019 Cheon–Hhan–Hong–Son [45] and 2022 Bi–Lu–Luo–Wang [35] considered collision searches inside dual attacks. For further speedup ideas, see 2022 MATZOV [101], 2022 Albrecht–Shen [9] (using quantum computation), and 2022 Carrier–Shen–Tillich [41].

**4.5. Optimizing decoding.** The 0.057981 in this paper comes from the 2020 Ducas–Laarhoven–van Woerden [53] algorithm for CVP after precomputation, as explained in Section 3.2. Plugging that algorithm into the nearest-colattice algorithm is one way to decode after precomputation, but there are alternatives. For example, 2021 Laarhoven–Walter [93] presents and recommends dual algorithms for bounded-distance decoding, although it does not account for failures of independence and does not compare to the nearest-colattice algorithm.

**4.6. Exploiting proof gaps.** This paper has focused on the key-recovery problem for the LPR PKE, finding $(a, e)$ given $G$ and $A = aG + e$. As mentioned in Section 2.2, this is a "search ring-LWE" problem.

Finding the secret key breaks "OW-CPA", "IND-CPA", and "IND-CCA2" for the LPR PKE (typically with probability slightly below 1 because of decryption failures in the PKE, depending on how $q$ is chosen), and breaks "IND-CCA2" for a typical KEM built from that PKE. However, breaking these security properties can be easier than key recovery.

The IND-CCA2 security analysis for a typical lattice KEM—see, e.g., Kyber's IND-CCA2 security analysis [13, Theorems 1, 2, and 3]—is structured as follows:

- There are theorems proving that the ROM IND-CCA2 security level of the KEM is at least the IND-CPA security level of the underlying PKE, aside from a 2-bit probability gap. See, e.g., 2017 Hofheinz–Hövelmanns–Kiltz [76, Theorems 3.2 and 3.3].
- There are similar theorems for QROM IND-CCA2, but with a much larger probability gap between the IND-CCA2 problem and the IND-CPA problem: see, e.g., 2023 Hövelmanns–Majenz [72, Theorem 1]. (Some KEMs avoid this gap; see generally [112, Section 3.8].)

- The IND-CPA security level of the PKEs in question is trivially proven to be at least the security level of a "decision Ring-LWE" problem, namely distinguishing a pair $(aG + e, aG' + e')$ from uniform given $G$ and $G'$. (For Kyber, one has to consider a more general "decision Module-LWE" problem; the differences do not matter for this paper.)

There are three ways that the theorems allow these problems to have lower security levels than the original search-Ring-LWE problem: (1) IND-CCA2 can have lower security level than QROM IND-CCA2; (2) low-probability IND-CPA can have lower security level than high-probability IND-CPA; (3) decision Ring-LWE can have lower security level than search Ring-LWE.

Hybrid attacks naturally target lower probabilities at higher speed; see Section 4.1. The same is true for dual attacks. Furthermore, illustrating the general risk of decision problems being easier than search problems (see [**22**, Section 6]), dual attacks against decision Ring-LWE are simpler and faster than dual attacks against search Ring-LWE: one simply runs the distinguisher once, without the extra cost of handling many guesses for portions of $e$. Similarly, for dual attacks against IND-CPA for the LPR PKE, it suffices to distinguish two known guesses $e_0, e_1$ for $e$. Perhaps there is also a noticeable speedup from choosing the IND-CPA plaintexts non-randomly, for example searching through many plaintexts to find the maximum possible angle between $e_0$ and $e_1$.

It would be interesting to quantify these effects, for example to see how weak low-probability decision Ring-LWE is against known attacks compared to high-probability search Ring-LWE. Theorems saying that the QROM IND-CCA2 security level is at least the security level of low-probability decision Ring-LWE are vacuous if low-probability decision Ring-LWE is breakable. Perhaps attacks against low-probability decision Ring-LWE can also be converted into similarly efficient QROM IND-CCA2 attacks against a KEM.

## References

[1] Michel Abdalla, David Pointcheval, Pierre-Alain Fouque, Damien Vergnaud (editors), *Applied cryptography and network security, 7th international conference, ACNS 2009, Paris-Rocquencourt, France, June 2–5, 2009, proceedings*, Lecture Notes in Computer Science, 5536, Springer, 2009. See [71].

[2] Carlisle Adams, Jan Camenisch (editors), *Selected areas in cryptography—SAC 2017—24th international conference, Ottawa, ON, Canada, August 16–18, 2017, revised selected papers*, 10719, Springer, 2018. ISBN 978-3-319-72564-2. See [17].

[3] Gorjan Alagic, Daniel Apon, David Cooper, Quynh Dang, Thinh Dang, John Kelsey, Jacob Lichtinger, Yi-Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, Daniel Smith-Tone, *Status report on the third round of the NIST Post-Quantum Cryptography Standardization Process* (2022). NISTIR 8413. URL: https://web.archive.org/web/20230824124130/https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8413-upd1.pdf. Citations in this document: §1.6.

[4] Martin R. Albrecht, *On dual lattice attacks against small-secret LWE and parameter choices in HElib and SEAL*, in Eurocrypt 2017 [**48**] (2017), 103–129. URL: https://eprint.iacr.org/2017/047. Citations in this document: §4.4, §4.4.

[5] Martin R. Albrecht, Benjamin R. Curtis, Thomas Wunderer, *Exploring trade-offs in batch bounded distance decoding*, in SAC 2019 [**115**] (2019), 467–491. URL: https://eprint.iacr.org/2019/1122. Citations in this document: §3.2, §4.2.

[6] Martin R. Albrecht, Léo Ducas, *Lattice attacks on NTRU and LWE: a history of refinements* (2021). URL: https://eprint.iacr.org/2021/799. Citations in this document: §2.4.

[7] Martin R. Albrecht, Vlad Gheorghiu, Eamonn W. Postlethwaite, John M. Schanck, *Estimating quantum speedups for lattice sieves*, in Asiacrypt 2020 [**106**] (2020), 583–613. URL: https://eprint.iacr.org/2019/1161. Citations in this document: §1.4.

[8] Martin R. Albrecht, Florian Göpfert, Fernando Virdia, Thomas Wunderer, *Revisiting the expected cost of solving uSVP and applications to LWE*, in Asiacrypt 2017 [**119**] (2017). URL: https://eprint.iacr.org/2017/815. Citations in this document: §2.4.

[9] Martin R. Albrecht, Yixin Shen, *Quantum augmented dual attack* (2022). URL: https://eprint.iacr.org/2022/656. Citations in this document: §4.4.

[10] Abdulrahman Al Jabri, *A statistical decoding algorithm for general linear block codes*, in IMA 2001 [**78**] (2001), 1–8. Citations in this document: §4.4.

[11] Erdem Alkim, Joppe W. Bos, Léo Ducas, Patrick Longa, Ilya Mironov, Michael Naehrig, Valeria Nikolaenko, Chris Peikert, Ananth Raghunathan, Douglas Stebila, *FrodoKEM: Learning With Errors key encapsulation: algorithm specifications and supporting documentation* (2021). URL: https://web.archive.org/web/20220119174856/https://frodokem.org/files/FrodoKEM-specification-20210604.pdf. Citations in this document: §1.6.

[12] Erdem Alkim, Léo Ducas, Thomas Pöppelmann, Peter Schwabe, *Post-quantum key exchange—a new hope*, in USENIX 2016 [**77**] (2016), 327–343. 9 August 2016 version. URL: https://eprint.iacr.org/2015/1092. Citations in this document: §1, §1, §1, §1.1, §1.2, §2.4, §2.4, §4.4.

[13] Roberto Avanzi, Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M. Schanck, Peter Schwabe, Gregor Seiler, Damien Stehlé, *CRYSTALS-Kyber: Algorithm specifications and supporting documentation (version 3.02)* (2021). URL: https://web.archive.org/web/20211215150153/https://pq-crystals.org/kyber/data/kyber-specification-round3-20210804.pdf. Citations in this document: §1, §1.2, §1.4, §2.4, §4.6.

[14] Roberto Avanzi, Howard M. Heys (editors), *Selected areas in cryptography—SAC 2016—23rd international conference, St. John's, NL, Canada, August 10–12, 2016, revised selected papers*, Lecture Notes in Computer Science, 10532, Springer, 2017. ISBN 978-3-319-69452-8. See [92].

[15] Hayo Baan, Sauvik Bhattacharya, Jung Hee Cheon, Scott Fluhrer, Oscar Garcia-Morchon, Paul Gorissen, Thijs Laarhoven, Rachel Player, Ronald Rietman, Markku-Juhani O. Saarinen, Yongha Son, Ludo Tolhuizen, José Luis Torre-Arce, Zhenfei Zhang, *Round5: KEM and PKE based on (Ring) Learning with Rounding* (2020). URL: https://web.archive.org/web/20220120070542/https://round5.org/doc/Round5_Submission042020.pdf. Citations in this document: §1.

[16] László Babai, *On Lovász' lattice reduction and the nearest lattice point problem*, Combinatorica **6** (1986), 1–13. Citations in this document: §3.3.

[17] Gustavo Banegas, Daniel J. Bernstein, *Low-Communication parallel quantum multi-target preimage search*, in SAC 2017 [**2**] (2017), 325–335. URL: https://cr.yp.to/papers.html#groverrho. Citations in this document: §1.4.

[18] Anja Becker, Jean-Sebastien Coron, Antoine Joux, *Improved generic algorithms for hard knapsacks*, in Eurocrypt 2011 [**113**] (2011). URL: https://eprint.iacr.org/2011/474. Citations in this document: §4.3.

[19] Anja Becker, Léo Ducas, Nicolas Gama, Thijs Laarhoven, *New directions in nearest neighbor searching with applications to lattice sieving*, in SODA 2016 [**91**] (2016), 10–24. URL: https://eprint.iacr.org/2015/1128. Citations in this document: §2.3.

[20] Daniel J. Bernstein, *Circuits for integer factorization: a proposal* (2001). URL: http://cr.yp.to/papers.html#nfscircuit. Citations in this document: §1.4.

[21] Daniel J. Bernstein, *Better price-performance ratios for generalized birthday attacks*, in Workshop Record of SHARCS'07: Special-purpose Hardware for Attacking Cryptographic Systems (2007). URL: https://cr.yp.to/papers.html#genbday. Citations in this document: §1.4.

[22] Daniel J. Bernstein, *Comparing proofs of security for lattice-based encryption* (2019). Second PQC Standardization Conference. URL: https://cr.yp.to/papers.html#latticeproofs. Citations in this document: §4.6.

[23] Daniel J. Bernstein, *Why EdDSA held up better than ECDSA against Minerva* (2019). URL: https://blog.cr.yp.to/20191024-eddsa.html. Citations in this document: §4.4.

[24] Daniel J. Bernstein, *Hybrid attacks* (2020). URL: https://groups.google.com/a/list.nist.gov/g/pqc-forum/c/8DvX3yF5bZc/m/CPJjiLbtAQAJ. Citations in this document: §1.8.

[25] Daniel J. Bernstein, *Multi-ciphertext security degradation for lattices* (2023). URL: https://cr.yp.to/papers.html#lprrr. Citations in this document: §1.8, §2.1, §2.1, §2.1, §2.1, §2.1, §2.2, §2.2, §2.4, §2.4, §2.4, §2.4, §2.5, §2.5, §3.3, §3.5, §4.1, §4.2, §4.4.

[26] Daniel J. Bernstein, Billy Bob Brumley, Ming-Shing Chen, Chitchanok Chuengsatiansup, Tanja Lange, Adrian Marotzke, Bo-Yuan Peng, Nicola Tuveri, Christine van Vredendaal, Bo-Yin Yang, *NTRU Prime: round 3* (2020). URL: https://ntruprime.cr.yp.to/nist/ntruprime-20201007.pdf. Citations in this document: §1, §4.1.

[27] Daniel J. Bernstein, Johannes Buchmann, Erik Dahmen (editors), *Post-quantum cryptography*, Springer, 2009. ISBN 978-3-540-88701-0. See [105].

[28] Daniel J. Bernstein, Stacey Jeffery, Tanja Lange, Alexander Meurer, *Quantum algorithms for the subset-sum problem*, in PQCrypto 2013 [**60**] (2013), 16–33. URL: https://eprint.iacr.org/2013/199. Citations in this document: §4.3.

[29] Daniel J. Bernstein, Tanja Lange, *Batch NFS*, in SAC 2014 [**85**] (2014), 38–58. URL: https://cr.yp.to/papers.html#batchnfs. Citations in this document: §1.4.

[30] Daniel J. Bernstein, Tanja Lange, *Challenges in evaluating costs of known lattice attacks* (2020). Slides. URL: https://cr.yp.to/talks.html#2020.02.19. Citations in this document: §1.1.

[31] Daniel J. Bernstein, Tanja Lange, *Non-randomness of S-unit lattices* (2021). URL: https://cr.yp.to/papers.html#spherical. Citations in this document: §1.7.

[32] Daniel J. Bernstein, Bo-Yin Yang, *Asymptotically faster quantum algorithms to solve multivariate quadratic equations*, in PQCrypto 2018 [**94**] (2018), 487–506.

URL: https://cr.yp.to/papers.html#groverxl. Citations in this document: §1.4.

[33] Guido Bertoni, Jean-Sébastien Coron (editors), *Cryptographic hardware and embedded systems—CHES 2013—15th international workshop, Santa Barbara, CA, USA, August 20–23, 2013, proceedings*, 8086, Springer, 2013. ISBN 978-3-642-40348-4. See [107].

[34] Karthikeyan Bhargavan, Elisabeth Oswald, Manoj Prabhakaran (editors), *Progress in cryptology—INDOCRYPT 2020—21st international conference on cryptology in India, Bangalore, India, December 13–16, 2020, proceedings*, 12578, Springer, 2020. ISBN 978-3-030-65276-0. See [56].

[35] Lei Bi, Xianhui Lu, Junjie Luo, Kunpeng Wang, *Hybrid dual and Meet-LWE attack*, in ACISP 2022 [**110**] (2022), 168–188. URL: https://eprint.iacr.org/2022/1330. Citations in this document: §4.4.

[36] Lei Bi, Xianhui Lu, Junjie Luo, Kunpeng Wang, Zhenfei Zhang, *Hybrid dual attack on LWE with arbitrary secrets*, Cybersecurity **5** (2022), 15. URL: https://eprint.iacr.org/2021/152. Citations in this document: §4.4.

[37] Daniel Bleichenbaher, *On the generation of one-time keys in DL signature schemes* (2004). Slides presented privately to IEEE in 2000, posted in 2004. URL: https://blog.cr.yp.to/20191024-bleichenbacher.pdf. Citations in this document: §4.4.

[38] Xavier Bonnetain, Rémi Bricout, André Schrottenloher, Yixin Shen, *Improved classical and quantum algorithms for subset-sum*, in Asiacrypt 2020 [**106**] (2020), 633–666. URL: https://eprint.iacr.org/2020/168. Citations in this document: §4.3, §4.3.

[39] Joe P. Buhler (editor), *Algorithmic number theory, third international symposium, ANTS-III, Portland, Oregon, USA, June 21–25, 1998, proceedings*, Lecture Notes in Computer Science, 1423, Springer, 1998. ISBN 3-540-64657-4. See [74].

[40] Ran Canetti, Juan A. Garay (editors), *Advances in cryptology—CRYPTO 2013—33rd annual cryptology conference, Santa Barbara, CA, USA, August 18–22, 2013, proceedings, part I*, Lecture Notes in Computer Science, 8042, Springer, 2013. See [52].

[41] Kevin Carrier, Yixin Shen, Jean-Pierre Tillich, *Faster dual lattice attacks by using coding theory* (2022). URL: https://eprint.iacr.org/2022/1750. Citations in this document: §4.4.

[42] Kévin Carrier, Thomas Debris-Alazard, Charles Meyer-Hilfiger, Jean-Pierre Tillich, *Reduction from sparse LPN to LPN, Dual Attack 3.0* (2023). URL: https://eprint.iacr.org/2023/1852. Citations in this document: §4.4.

[43] André Chailloux, Johanna Loyer, *Lattice sieving via quantum random walks*, in Asiacrypt 2021 [**120**] (2021), 63–91. URL: https://eprint.iacr.org/2021/570. Citations in this document: §1.6.

[44] Cong Chen, Oussama Danba, Jeffrey Hoffstein, Andreas Hulsing, Joost Rijneveld, John M. Schanck, Peter Schwabe, William Whyte, Zhenfei Zhang, *NTRU: algorithm specifications and supporting documentation* (2019). URL: https://ntru.org/f/ntru-20190330.pdf. Citations in this document: §1.

[45] Jung Hee Cheon, Minki Hhan, Seungwan Hong, Yongha Son, *A hybrid of dual and meet-in-the-middle attack on sparse and ternary secret LWE* (2019). URL: https://eprint.iacr.org/2019/1114. Citations in this document: §4.4.

[46] Jung Hee Cheon, Jean-Pierre Tillich (editors), *Post-quantum cryptography—12th international workshop, PQCrypto 2021, Daejeon, South Korea, July 20–22, 2021, proceedings*, 12841, Springer, 2021. ISBN 978-3-030-81292-8. See [79].

[47] Don Coppersmith, Adi Shamir, *Lattice attacks on NTRU*, in Eurocrypt 1997 [**59**] (1997), 52–61. URL: https://link.springer.com/content/pdf/10.1007/3-540-69053-0_5.pdf. Citations in this document: §2.2.

[48] Jean-Sébastien Coron, Jesper Buus Nielsen (editors), *Advances in cryptology—EUROCRYPT 2017—36th annual international conference on the theory and applications of cryptographic techniques, Paris, France, April 30–May 4, 2017, proceedings, part II*, 10211, 2017. ISBN 978-3-319-56613-9. See [4].

[49] Jing Deng, Vladimir Kolesnikov, Alexander A. Schwarzmann (editors), *Cryptology and network security—22nd international conference, CANS 2023, Augusta, GA, USA, October 31–November 2, 2023, proceedings*, 14342, Springer, 2023. ISBN 978-981-99-7562-4. See [64].

[50] Jintai Ding, Rainer Steinwandt (editors), *Post-quantum cryptography—10th international conference, PQCrypto 2019, Chongqing, China, May 8–10, 2019, revised selected papers*, 11505, Springer, 2019. ISBN 978-3-030-25509-1. See [51].

[51] Emmanouil Doulgerakis, Thijs Laarhoven, Benne de Weger, *Finding closest lattice vectors using approximate Voronoi cells*, in PQCrypto 2019 [**50**] (2019), 3–22. URL: https://eprint.iacr.org/2016/888. Citations in this document: §3.2.

[52] Léo Ducas, Alain Durmus, Tancrède Lepoint, Vadim Lyubashevsky, *Lattice signatures and bimodal Gaussians*, in Crypto 2013 [**40**] (2013), 40–56. URL: https://eprint.iacr.org/2013/383. Citations in this document: §1.1, §1.1, §1.1.

[53] Léo Ducas, Thijs Laarhoven, Wessel P. J. van Woerden, *The randomized slicer for CVPP: sharper, faster, smaller, batchier*, in PKC 2020 [**89**] (2020), 3–36. URL: https://eprint.iacr.org/2020/120. Citations in this document: §3.2, §3.2, §3.2, §3.2, §4.5.

[54] Léo Ducas, Ludo N. Pulles, *Does the dual-sieve attack on learning with errors even work?*, in Crypto 2023 [**67**] (2023), 37–69. URL: https://eprint.iacr.org/2023/302. Citations in this document: §4.4.

[55] Léo Ducas, Ludo N. Pulles, *Accurate score prediction for dual-sieve attacks* (2023). URL: https://eprint.iacr.org/2023/1850. Citations in this document: §4.4.

[56] Thomas Espitau, Antoine Joux, Natalia Kharchenko, *On a dual/hybrid approach to small secret LWE—A dual/enumeration technique for learning with errors and application to security estimates of FHE schemes*, in Indocrypt 2020 [**34**] (2020), 440–462. URL: https://eprint.iacr.org/2020/515. Citations in this document: §4.4.

[57] Thomas Espitau, Paul Kirchner, *The nearest-colattice algorithm: time-approximation tradeoff for approx-CVP*, in ANTS 2020 [**61**] (2020), 251–266. URL: https://eprint.iacr.org/2020/694. Citations in this document: §3.3, §3.3, §3.3.

[58] Andre Esser, Alexander May, *Better sample—random subset sum in* $2^{0.255n}$ *and its impact on decoding linear codes* (2019). Withdrawn. URL: https://arxiv.org/abs/1907.04295. Citations in this document: §4.3.

[59] Walter Fumy (editor), *Advances in cryptology—EUROCRYPT '97, international conference on the theory and application of cryptographic techniques, Konstanz, Germany, May 11–15, 1997*, Lecture Notes in Computer Science, 1233, Springer, 1997. See [47].

[60] Philippe Gaborit (editor), *Post-quantum cryptography—5th international workshop, PQCrypto 2013, Limoges, France, June 4–7, 2013, proceedings*, 7932, Springer, 2013. ISBN 978-3-642-38615-2. See [28].

[61] Steven Galbraith (editor), *ANTS XIV: proceedings of the fourteenth algorithmic number theory symposium, Auckland 2020*, Open Book Series, 4, Mathematical Sciences Publishers, 2020. ISBN 978-1-935107-07-1. See [57].

[62] Robert G. Gallager, *Low-density parity-check codes*, IRE Transactions on Information Theory **8** (1962), 21–28. URL: https://doi.org/10.1109/TIT.1962.1057683. Citations in this document: §4.4.

[63] Henri Gilbert (editor), *Advances in cryptology—EUROCRYPT 2010, 29th annual international conference on the theory and applications of cryptographic techniques, French Riviera, May 30–June 3, 2010, proceedings*, Lecture Notes in Computer Science, 6110, Springer, 2010. See [81].

[64] Timo Glaser, Alexander May, *How to enumerate LWE keys as narrow as in Kyber/Dilithium*, in CANS 2023 [**49**] (2023), 75–100. URL: https://eprint.iacr.org/2022/1337. Citations in this document: §4.3.

[65] Florian Göpfert, Christine van Vredendaal, Thomas Wunderer, *A hybrid lattice basis reduction and quantum search attack on LWE*, in PQCrypto 2017 [**95**] (2017), 184–202. URL: https://eprint.iacr.org/2017/221. Citations in this document: §4.1.

[66] Qian Guo, Thomas Johansson, *Faster dual lattice attacks for solving LWE with applications to CRYSTALS*, in Asiacrypt 2021 [**120**] (2021), 33–62. Citations in this document: §4.4.

[67] Helena Handschuh, Anna Lysyanskaya (editors), *Advances in cryptology—CRYPTO 2023—43rd annual international cryptology conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20–24, 2023, proceedings, part III*, 14083, Springer, 2023. ISBN 978-3-031-38547-6. See [54].

[68] D. Roger Heath-Brown, *Almost-primes in arithmetic progressions and short intervals*, Mathematical Proceedings of the Cambridge Philosophical Society **83** (1978), 357–375. Citations in this document: §2.5.

[69] Max Heiser, *Improved quantum hypercone locality sensitive filtering in lattice sieving* (2021). URL: https://eprint.iacr.org/2021/1295. Citations in this document: §1.6.

[70] Alexander Helm, Alexander May, *Subset sum quantumly in* $1.17^n$, in TQC 2018 [**83**] (2018), 5:1–5:15. URL: https://www.dagstuhl.de/dagpub/978-3-95977-080-4. Citations in this document: §4.3.

[71] Philip S. Hirschhorn, Jeffrey Hoffstein, Nick Howgrave-Graham, William Whyte, *Choosing NTRUEncrypt parameters in light of combined lattice reduction and MITM approaches*, in ACNS 2009 [**1**] (2009), 437–455. URL: https://assets.onboardsecurity.com/static/downloads/NTRU/resources/params.pdf. Citations in this document: §1.1.

[72] Kathrin Hövelmanns, Christian Majenz, *A note on Failing gracefully: Completing the picture for explicitly rejecting Fujisaki-Okamoto transforms using worst-case correctness* (2023). URL: https://eprint.iacr.org/2023/1811. Citations in this document: §4.6.

[73] Jeff Hoffstein, Jill Pipher, John M. Schanck, Joseph H. Silverman, William Whyte, Zhenfei Zhang, *Choosing parameters for NTRUEncrypt* (2015). URL: https://eprint.iacr.org/2015/708. Citations in this document: §1.1.

[74] Jeffrey Hoffstein, Jill Pipher, Joseph H. Silverman, *NTRU: a ring-based public key cryptosystem*, in ANTS 1998 [**39**] (1998), 267–288. See also [**75**]. URL: https://ntru.org/f/hps98.pdf. Citations in this document: §2.2.

[75] Jeffrey Hoffstein, Jill Pipher, Joseph H. Silverman, *NTRU: a new high speed public key cryptosystem* (2016). Circulated at Crypto 1996, put online in 2016; see also [**74**]. URL: https://ntru.org/f/hps96.pdf. Citations in this document: §2.2.

[76] Dennis Hofheinz, Kathrin Hövelmanns, Eike Kiltz, *A modular analysis of the Fujisaki-Okamoto transformation*, in TCC 2017-1 [**86**] (2017), 341–371. URL: https://eprint.iacr.org/2017/604. Citations in this document: §4.6.

[77] Thorsten Holz, Stefan Savage (editors), *25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10–12, 2016*, USENIX Assocation, 2016. See [12].

[78] Bahram Honary (editor), *Cryptography and coding: proceedings of the 8th IMA international conference held in Cirencester, December 17–19, 2001*, Lecture Notes in Computer Science, 2260, Springer, 2001. See [10].

[79] Iggy van Hoof, Elena Kirshanova, Alexander May, *Quantum key search for ternary LWE*, in PQCrypto 2021 [**46**] (2021), 117–132. URL: https://eprint.iacr.org/2021/865. Citations in this document: §4.3.

[80] Nick Howgrave-Graham, *A hybrid lattice-reduction and meet-in-the-middle attack against NTRU*, in Crypto 2007 [**104**] (2007), 150–169. URL: https://www.iacr.org/archive/crypto2007/46220150/46220150.pdf. Citations in this document: §1, §1, §1.1, §1.1, §1.1, §3.3, §4, §4, §4.3.

[81] Nick Howgrave-Graham, Antoine Joux, *New generic algorithms for hard knapsacks*, in Eurocrypt 2010 [**63**] (2010). URL: https://eprint.iacr.org/2010/189. Citations in this document: §4.3, §4.3, §4.3.

[82] Nick Howgrave-Graham, Joseph H. Silverman, William Whyte, *A meet-in-the-middle attack on an NTRU private key* (2003). NTRU Tech Report 004v2. URL: https://ntru.org/f/tr/tr004v2.pdf. Citations in this document: §4.3.

[83] Stacey Jeffery (editor), *13th conference on the theory of quantum computation, communication and cryptography, TQC 2018, July 16–18, 2018, Sydney, Australia*, 111, Schloss Dagstuhl, 2018. ISBN 978-3-95977-080-4. URL: https://www.dagstuhl.de/dagpub/978-3-95977-080-4. See [70].

[84] David S. Johnson, Ronald Fagin, Michael L. Fredman, David Harel, Richard M. Karp, Nancy A. Lynch, Christos H. Papadimitriou, Ronald L. Rivest, Walter L. Ruzzo, Joel I. Seiferas (editors), *Proceedings of the 15th annual ACM symposium on theory of computing, 25-27 April, 1983, Boston, Massachusetts, USA*, ACM, 1983. See [87].

[85] Antoine Joux, Amr M. Youssef (editors), *Selected areas in cryptography—SAC 2014—21st international conference, Montreal, QC, Canada, August 14–15, 2014, revised selected papers*, 8781, Springer, 2014. ISBN 978-3-319-13050-7. See [29].

[86] Yael Kalai, Leonid Reyzin (editors), *Theory of cryptography—15th international conference, TCC 2017, Baltimore, MD, USA, November 12–15, 2017, proceedings, part I*, Lecture Notes in Computer Science, 10677, Springer, 2017. ISBN 978-3-319-70499-9. See [76].

[87] Ravi Kannan, *Improved algorithms for integer programming and related lattice problems*, in STOC 1983 [**84**] (1983), 193–206. Citations in this document: §3.3.

[88] Aggelos Kiayias (editor), *Topics in cryptology—CT-RSA 2011—the cryptographers' track at the RSA Conference 2011, San Francisco, CA, USA, February 14–18, 2011, proceedings*, Lecture Notes in Computer Science, 6558, Springer, 2011. ISBN 978-3-642-19073-5. See [97].

[89] Aggelos Kiayias, Markulf Kohlweiss, Petros Wallden, Vassilis Zikas (editors), *Public-key cryptography—PKC 2020—23rd IACR international conference on practice and theory of public-key cryptography, Edinburgh, UK, May 4–7, 2020, proceedings, part II*, 12111, Springer, 2020. ISBN 978-3-030-45387-9. See [53].

[90] Elena Kirshanova, Thijs Laarhoven, *Lower bounds on lattice sieving and information set decoding*, in Crypto 2021 [**100**] (2021), 791–820. URL: https://eprint.iacr.org/2021/785. Citations in this document: §1.6, §1.6.

[91] Robert Krauthgamer (editor), *Proceedings of the twenty-seventh annual ACM-SIAM symposium on discrete algorithms, SODA 2016, Arlington, VA, USA, January 10–12, 2016*, SIAM, 2016. ISBN 978-1-61197-433-1. See [19].

[92] Thijs Laarhoven, *Sieving for closest lattice vectors (with preprocessing)*, in SAC 2016 [**14**] (2016), 523–542. URL: https://arxiv.org/abs/1607.04789. Citations in this document: §3.2.

[93] Thijs Laarhoven, Michael Walter, *Dual lattice attacks for closest vector problems (with preprocessing)*, in CT-RSA 2021 [**114**] (2021), 478–502. URL: https://eprint.iacr.org/2021/557. Citations in this document: §4.5.

[94] Tanja Lange, Rainer Steinwandt (editors), *Post-quantum cryptography—9th international conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9–11, 2018, proceedings*, 10786, Springer, 2018. ISBN 978-3-319-79062-6. See [32].

[95] Tanja Lange, Tsuyoshi Takagi (editors), *Post-quantum cryptography—8th international workshop, PQCrypto 2017, Utrecht, the Netherlands, June 26–28, 2017, proceedings*, Lecture Notes in Computer Science, 10346, Springer, 2017. ISBN 978-3-319-59878-9. See [65].

[96] Arjen K. Lenstra, Hendrik W. Lenstra, Jr., László Lovász, *Factoring polynomials with rational coefficients*, Mathematische Annalen **261** (1982), 515–534. ISSN 0025-5831. MR 84a:12002. URL: https://openaccess.leidenuniv.nl/bitstream/handle/1887/3810/346_050.pdf. Citations in this document: §3.3.

[97] Richard Lindner, Chris Peikert, *Better key sizes (and attacks) for LWE-based encryption*, in CT-RSA [**88**] (2011), 319–339. URL: https://eprint.iacr.org/2010/613. Citations in this document: §1.6.

[98] Xianhui Lu, Yamin Liu, Zhenfei Zhang, Dingding Jia, Haiyang Xue, Jingnan He, Bao Li, Kunpeng Wang, *LAC: practical Ring-LWE based public-key encryption with byte-level modulus* (2018). URL: https://eprint.iacr.org/2018/1009. Citations in this document: §1.

[99] Vadim Lyubashevsky, Chris Peikert, Oded Regev, *On ideal lattices and learning with errors over rings*, Journal of the ACM **60** (2013), Article 43, 35 pages. URL: https://eprint.iacr.org/2012/230. Citations in this document: §2.1.

[100] Tal Malkin, Chris Peikert (editors), *Advances in cryptology—CRYPTO 2021—41st annual international cryptology conference, CRYPTO 2021, virtual event, August 16–20, 2021, proceedings, part II*, 12826, Springer, 2021. ISBN 978-3-030-84244-4. See [90], [102].

[101] MATZOV, *Report on the security of LWE: improved dual lattice attack* (2022). URL: https://zenodo.org/records/6493704. Citations in this document: §4.4.

[102] Alexander May, *How to meet ternary LWE keys*, in Crypto 2021 [**100**] (2021), 701–731. URL: https://eprint.iacr.org/2021/216. Citations in this document: §4.3.

[103] Alexander May, Joseph H. Silverman, *Dimension reduction methods for convolution modular lattices*, in [**118**] (2001), 110–15. URL: https://www.cits.ruhr-uni-bochum.de/personen/may/publications.html. Citations in this document: §2.2.

[104] Alfred Menezes (editor), *Advances in cryptology—CRYPTO 2007, 27th annual international cryptology conference, Santa Barbara, CA, USA, August 19–23, 2007, proceedings*, Lecture Notes in Computer Science, 4622, Springer, 2007. ISBN 978-3-540-74142-8. See [80].

[105] Daniele Micciancio, Oded Regev, *Lattice-based cryptography*, in Post-quantum cryptography [**27**] (2009), 147–191. URL: https://web.archive.org/web/20210506211055/https://cims.nyu.edu/~regev/papers/pqc.pdf. Citations in this document: §4.4.

[106] Shiho Moriai, Huaxiong Wang (editors), *Advances in cryptology—ASIACRYPT 2020—26th international conference on the theory and application of cryptology and information security, Daejeon, South Korea, December 7–11, 2020, proceedings, part II*, 12492, Springer, 2020. ISBN 978-3-030-64833-6. See [7], [38].

[107] Elke De Mulder, Michael Hutter, Mark E. Marson, Peter Pearson, *Using Bleichenbacher's solution to the hidden number problem to attack nonce leaks in 384-bit ECDSA*, in CHES 2013 [**33**] (2013), 435–452; see also newer version [**108**]. URL: https://eprint.iacr.org/2013/346. Citations in this document: §4.4.

[108] Elke De Mulder, Michael Hutter, Mark E. Marson, Peter Pearson, *Using Bleichenbacher's solution to the hidden number problem to attack nonce leaks in 384-bit ECDSA: extended version*, Journal of Cryptographic Engineering **4** (2014), 33–45; see also older version [**107**]. Citations in this document: §4.4.

[109] National Institute of Standards and Technology, *Digital Signature Standard (DSS)* (2001). FIPS Publication 186-2. URL: https://web.archive.org/web/20231007193942/https://csrc.nist.gov/files/pubs/fips/186-2/upd1/final/docs/fips186-2-change1.pdf. Citations in this document: §4.4.

[110] Khoa Nguyen, Guomin Yang, Fuchun Guo, Willy Susilo (editors), *Information security and privacy—27th Australasian conference, ACISP 2022, Wollongong, NSW, Australia, November 28–30, 2022, proceedings*, 13494, Springer, 2022. ISBN 978-3-031-22300-6. See [35].

[111] Phong Q. Nguyen, *Boosting the hybrid attack on NTRU: torus LSH, permuted HNF and boxed sphere* (2021). Third PQC Standardization Conference. URL: https://web.archive.org/web/20230903180555/https://csrc.nist.gov/CSRC/media/Events/third-pqc-standardization-conference/documents/accepted-papers/nguyen-boosting-hybridboost-pqc2021.pdf. Citations in this document: §4.3.

[112] NTRU Prime Risk-Management Team, *Risks of lattice KEMs* (2021). URL: https://ntruprime.cr.yp.to/warnings.html. Citations in this document: §4.6.

[113] Kenneth G. Paterson (editor), *Advances in cryptology—EUROCRYPT 2011, 30th annual international conference on the theory and applications of cryptographic techniques, Tallinn, Estonia, May 15–19, 2011, proceedings*, Lecture Notes in Computer Science, 6632, Springer, 2011. ISBN 978-3-642-20464-7. See [18].

[114] Kenneth G. Paterson (editor), *Topics in cryptology—CT-RSA 2021—cryptographers' track at the RSA conference 2021, virtual event, May 17–20, 2021, proceedings*, 12704, Springer, 2021. ISBN 978-3-030-75538-6. See [93].

[115] Kenneth G. Paterson, Douglas Stebila (editors), *Selected areas in cryptography—SAC 2019—26th international conference, Waterloo, ON,*

*Canada, August 12–16, 2019, revised selected papers*, 11959, Springer, 2020. ISBN 978-3-030-38470-8. See [5].

[116] John M. Schanck, *When sorting your data costs more than cracking AES-128* (2023). URL: https://web.archive.org/web/20231125213807/https://finiterealities.net/kyber512/. Citations in this document: §1.4.

[117] Richard Schroeppel, Adi Shamir, *A $T = O(2^{n/2})$, $S = O(2^{n/4})$ algorithm for certain NP-complete problems*, SIAM Journal on Computing **10** (1981), 456–464. Citations in this document: §4.3.

[118] Joseph H. Silverman (editor), *Cryptography and lattices: proceedings of the 1st International Conference (CaLC 2001) held in Providence, RI, March 29–30, 2001*, Lecture Notes in Computer Science, 2146, Springer, 2001. ISBN 3-540-42488-1. MR 2002m:11002. See [103].

[119] Tsuyoshi Takagi, Thomas Peyrin (editors), *Advances in cryptology—ASIACRYPT 2017—23rd international conference on the theory and applications of cryptology and information security, Hong Kong, China, December 3–7, 2017, proceedings, part II*, Lecture Notes in Computer Science, 10625, Springer, 2017. ISBN 978-3-319-70696-2. See [8].

[120] Mehdi Tibouchi, Huaxiong Wang (editors), *Advances in cryptology—ASIACRYPT 2021—27th international conference on the theory and application of cryptology and information security, Singapore, December 6–10, 2021, proceedings, part IV*, 13093, Springer, 2021. ISBN 978-3-030-92067-8. See [43], [66].

[121] Serge Vaudenay, *Evaluation report on DSA* (2001). URL: https://www.cryptrec.go.jp/exreport/cryptrec-ex-1002-2001.pdf. Citations in this document: §4.4.

[122] Christine van Vredendaal, *Reduced memory meet-in-the-middle attack against the NTRU private key* (2016). URL: https://eprint.iacr.org/2016/177. Citations in this document: §4.3.

[123] Samuel S. Wagstaff Jr., *Greatest of the least primes in arithmetic progressions having a given modulus*, Mathematics of Computation **33** (1979), 1073–1080. URL: https://www.ams.org/journals/mcom/1979-33-147/S0025-5718-1979-0528061-7/. Citations in this document: §2.5.

[124] Hong Wang, Zhi Ma, ChuanGui Ma, *An efficient quantum meet-in-the-middle attack against NTRU-2005*, Chinese Science Bulletin **58** (2013), 3514–3518. Citations in this document: §4.3.

[125] Andreas Wiemers, Stephan Ehlen, *A remark on the independence heuristic in the dual attack* (2023). URL: https://eprint.iacr.org/2023/1238. Citations in this document: §4.4.

[126] Thomas Wunderer, *Revisiting the hybrid attack: improved analysis and refined security estimates* (2016). URL: https://eprint.iacr.org/2016/733. Citations in this document: §4.3.

[127] Thomas Wunderer, *On the security of lattice-based cryptography against lattice reduction and hybrid attacks* (2018). URL: https://tuprints.ulb.tu-darmstadt.de/8082/. Citations in this document: §4.1, §4.3.

[128] Thomas Wunderer, *A detailed analysis of the hybrid lattice-reduction and meet-in-the-middle attack*, J. Mathematical Cryptology **13** (2019), 1–26. Citations in this document: §4.3.