On the complexity of the Arora-Ge Algorithm against LWE

Martin Albrecht, Carlos Cid, Jean-Charles Faugère, Robert Fitzpatrick, Ludovic Perret

On the complexity of the Arora-Ge Algorithm against LWE

Martin R. Albrecht, Carlos Cid, Jean-Charles Faugère, Robert Fitzpatrick, and Ludovic Perret

Abstract

Arora & Ge [5] recently showed that solving LWE can be reduced to solve a high-degree non-linear system of equations. They used a linearization to solve the systems. We investigate here the possibility of using Gröbner bases to improve Arora & Ge approach.

Introduction

The Learning With Errors (LWE) Problem was introduced by Regev in [27, 26]. It is a generalisation for large primes of the well-known LPN (Learning Parity with Noise) problem. Since its introduction, LWE has become a source of many innovative cryptosystems, such as the oblivious transfer protocol by Peikert et al. [25], a cryptosystem by Akavia et al. [1] that is secure even if almost the entire secret key is leaked, homomorphic encryption [21, 10, 4], etc... Reasons of LWE's success in cryptography include its simplicity as well as convincing theoretical arguments regarding its hardness, i.e. a reduction from (worst-case) assumed hard lattice problems to (average-case) LWE.

The purpose of this paper is to investigate whether algebraic techniques (e.g. [16, 17, 18, 19, 3, 2, 20]) can be used in the context of LWE. This is motivated by a recent result Arora & Ge [5] who showed that solving LWE can be reduced to solve a high-degree non-linear system of equations.

Learning With Errors

We reproduce below the definition of the LWE problem from [27, 26].

Definition 1 (LWE)**.** *Let $n \geq 1$ be the number of variables, q be an odd prime integer, χ be a probability distribution on $\mathbb{Z}_q$ and $\mathbf{s}$ be a secret vector in $\mathbb{Z}_q^n$. We denote by $L_{\mathbf{s},\chi}^{(n)}$ the probability distribution on $\mathbb{Z}_q^n \times \mathbb{Z}_q$ obtained by choosing $\mathbf{a} \in \mathbb{Z}_q^n$ at random, choosing $e \in \mathbb{Z}_q$ according to χ, and returning $(\mathbf{a}, c) = (\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$.* LWE *is the problem of finding $\mathbf{s} \in \mathbb{Z}_q^n$ given pairs $\mathbb{Z}_q^n \times \mathbb{Z}_q$ sampled according to $L_{\mathbf{s},\chi}^{(n)}$.*

The modulus q is typically taken to be polynomial in n, and χ is the discrete Gaussian distribution on $\mathbb{Z}_q$ with mean 0 and standard deviation $\sigma = \alpha \cdot q$, for some α. To *discretize* the Gaussian distribution $\mathbb{N}0, \sigma^2$ modulo q, we sample according to $\mathbb{N}0, \sigma^2$ and round to the nearest integer mod q. In what follows, $\chi_{\alpha,q}$ will then denote this discretized distribution.

A typical setting for the standard deviation (std) is $\sigma = n^\varepsilon$, with $\varepsilon, 0 \leq \varepsilon \leq 1$. For example, [27] suggests $q \approx n^2$ and $\alpha = 1/(\sqrt{n} \cdot \log^2 n)$. Indeed, as soon as $\varepsilon \geq 1/2$ (worst-case) GAPSVP $- \tilde{O}(n/\alpha)$ reduces to (average-case) LWE[1]. Thus, any algorithm solving LWE (when $\varepsilon \geq 1/2$) can be used for GAPSVP $- \tilde{O}(n/\alpha)$. We emphasize that it is widely believed that only exponential algorithm exists for solving GAPSVP $- \tilde{O}(n/\alpha)$.

Recently, Arora & Ge [5] introduced a variant of LWE with *structured* errors. In this setting, you have given an oracle such that given LWE samples returns polynomials which vanish on the errors.

[1] The reduction is quantum if q is polynomial but can be made [24] classical if q is super polynomial.

They showed that the (discretized) Gaussian intrinsically induced a structure on the errors. This feature can be used to reduce LWE to the problem of solving a non-linear system of multivariate of equations.

The total complexity (time and space) of their approach is $2^{\tilde{O}(n^{2\varepsilon})}$. It is then subexponential when $\varepsilon < 1/2$, but remains exponential when $\varepsilon \geq 1/2$. It is interesting that Arora&Ge reach with a completely different approach the $\varepsilon = 1/2$ hardness limit advised by Regev [27, 26].

Note that an improvement on Arora&Ge could allow to challenge the '*subexponetiality*' of GAPSVP $-\ \tilde{O}(n/\alpha)$. Remark that [5] uses linearization to solve the non-linear system. It is then natural to investigate whether more advanced tools, such as Gröbner bases [11, 12, 13], could improve the algorithm of Arora&Ge.

In this note, we will show that Gröbner bases can bring a practical improvement on the complexity of [5]. We also briefly discuss whether Gröbner bases can (or can not) allow to change the complexity class of Arora&Ge. Before that, we need to recall some basic complexity results about Gröbner bases.

Gröbner bases – Complexity Results

Gröbner basis is probably the main tool allowing to solve non-linear system of finite fields. From an algorithmic point of view, Lazard [22] showed that computing the Gröbner basis for a system of polynomials is equivalent to perform a Gaussian elimination on the *Macaulay matrices* [23] $\mathcal{M}_{d,m}^{\text{acaulay}}$ for $d, 1 \leq d \leq D$ for some integer D. Moreover, the most efficient known algorithms such as F_5 [15] reduce Gröbner basis computations to a series of Gaussian eliminations on matrices of increasing sizes.

Definition 2 (Macaulay Matrix [23])**.** *Let $f_1, \ldots, f_m \in \mathbb{Z}_q[x_1, \ldots, x_n]$. The* Macaulay matrix $\mathcal{M}_{d,m}^{\text{acaulay}}(f_1, \ldots, f_m)$ *of degree d is defined as follows: list "horizontally" all the degree d monomials from smallest to largest sorted by some fixed admissible monomial ordering. The smallest monomial comes last. Multiply each f_i by all monomials $t_{i,j}$ of degree $d - d_i$ where $d_i = \deg(f_i)$. Finally, construct the coefficient matrix for the resulting system:*

$$\mathcal{M}_{d,m}^{\text{acaulay}}(f_1, \ldots, f_m) := \begin{array}{c} \\ (t_{1,1}, f_1) \\ (t_{1,2}, f_1) \\ \vdots \\ (t_{m,1}, f_m) \\ (t_{m,2}, f_m) \\ \vdots \end{array} \begin{array}{c} \text{monomials of degree } \leq d \text{ sorted for } < \\ \begin{pmatrix} & & \\ & & \\ & & \\ & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Theorem 3 ([22])**.** *Let $\mathbf{f} = (f_1, \ldots, f_m) \in (\mathbb{Z}_q[x_1, \ldots, x_n])^m$ and $<$ be a monomial ordering. There exists a positive integer D for which Gaussian elimination on all $\mathcal{M}_{d,m}^{\text{acaulay}} = (f_1, \ldots, f_m)$ matrices for $d, 1 \leq d \leq D$ computes a Gröbner basis of $\langle f_1, \ldots, f_m \rangle$ w.r.t. to $<$. The degree D will be called* degree of regularity *of $f_1, \ldots, f_m$.*

Consequently, the complexity of computing a Gröbner basis is bounded by the complexity of performing Gaussian elimination on the Macaulay matrix in some degree D. Roughly, the complexity of computing a Gröbner basis with an algorithm based on the degree of regularity (such as – but not limited too – Buchberger's algorithm, $\mathrm{F}_4, \mathrm{F}_5$ [15, 11, 12, 14]) is:

$$O\left(\binom{n+D}{D}^{\omega}\right), \tag{1}$$

where $2 \leq \omega < 3$ is the linear algebra constant, and D is the degree of semi-regularity of the system.

In general, computing the degree of regularity of a system is a difficult problem. However, it is known for a specific family of polynomial systems [6, 8, 7, 9].

Definition 4 (Semi-regular Sequence [8])**.** *Let $m > n$, and $f_1, \ldots, f_m \in \mathbb{Z}_q[x_1, \ldots, x_n]$ be homogeneous polynomials of degrees $d_1, \ldots, d_m$ respectively and I the ideal generated by these polynomials. The system is said to be a* semi-regular sequence *if the Hilbert series [13] of I w.r.t. the grevlex order is:*

$$H_I(z) = \left[\frac{\prod_{i=1}^{m}(1-z^{d_i})}{(1-z)^n} \right]_+, \tag{2}$$

where $[S]_+$ denotes the series obtained by truncating S before the index of its first non-positive coefficient. Thus, the degree of regularity D involved in Theorem 3 *for a semi-regular sequence is:*

$$1 + \deg(H_I).$$

Improving Arora-Ge Approach

We briefly detail below the linearization approach of Arora-Ge. We then discuss whether Gröbner bases can be used in this context.

Basic Arora-Ge Algorithm – A Linerization Approach

The idea of [5] is to generate a non-linear noise-free system of equations from LWE samples. This is due to the following well-known feature of a Gaussian noise:

Lemma 5. *Let $C > 0$ be a constant. It holds that:*

$$\Pr[e \xleftarrow{\$} \chi_{\alpha,q} : |e| > C \cdot \sigma] \leq e^{O(-C^2)}.$$

As a consequence, elements sampled from a Gaussian distribution only takes values on a (small) subset $[-C \cdot \sigma, \ldots, C \cdot \sigma]$ of $\mathbb{Z}_q$ with high probability. From now on, we set $t = C \cdot \sigma$. We can re-interpret Lemma 5 algebraically by considering the polynomial:

$$P(X) = X \prod_{i=1}^{t} (X+i)(X-i).$$

Clearly P is of degree $2t+1 \in O(\sigma)$. Thus, if $e \xleftarrow{\$} \chi_{\alpha,q}$, then $P(e) = 0$ with probability at least $1 - e^{O(-C^2)}$.

For $i \geq 1$, let $(\mathbf{a}_i, \langle \mathbf{a}_i, s \rangle + e_i) = (\mathbf{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$. If $e_i \xleftarrow{\$} \chi_{\alpha,q}$, then

$$P\big(\mathbf{a}_i, \langle \mathbf{a}_i, s \rangle - b_i\big) = 0, \tag{3}$$

with probability at least $1 - e^{O(-C^2)}$. As a consequence, each sample $(\mathbf{a}_i, \langle \mathbf{a}_i, s \rangle + e_i) = (\mathbf{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$ allows to generate a non-linear equation of degree $2t+1$ in the n components of the secret $\mathbf{s}$.

The idea of Arora & Ge is then to generate sufficiently many equations as in (3) to perform a linearization. However, one has to choose the constant – denoted by C_{AG} – occurring in Lemma 5 sufficiently big so that all errors generated lies with high probability in $[-C_{\mathrm{AG}} \cdot \sigma, \ldots, C_{\mathrm{AG}} \cdot \sigma] \subseteq \mathbb{Z}_q$, i.e. the secret $\mathbf{s}$ is indeed a common solution of the M_{AG} equations constructed as in (3). To this end, we set:

$$p_f = \frac{M_{\mathrm{AG}}}{e^{O(C_{\mathrm{AG}}^2)}}.$$

This is the probably that the secret $s \in \mathbb{Z}_q^n$ is not solution of the system $\mathcal{S}_{\mathrm{AG}}$ generated from M_{AG} equations as in (3), i.e. the probability of failure of Arora-Ge approach. Let also $D_{AG} = 2C_{\mathrm{AG}} \cdot \sigma + 1$ be the degree of the equations occuring in $\mathcal{S}_{\mathrm{AG}}$. According to [5], taking $C_{\mathrm{AG}} \in \tilde{O}(\sigma)$ allows to make the probability of failure negligible.

To summarize, Arora-Ge approach reduces to linearize at system of M_{AG} equations of degree $D_{AG} = 2C_{\mathrm{AG}} \cdot \sigma + 1 \in \tilde{O}(\sigma^2)$. Moreover, correctness of this approach can be proven:

Theorem 6. *[5] Let $D_{\mathrm{AG}} < q$. The system obtained by linearizing $M_{\mathrm{AG}} = O\left(q \cdot \log(q)\binom{n+D_{\mathrm{AG}}}{D_{\mathrm{AG}}}\sigma\right) = n^{O(D_{\mathrm{AG}})} = 2^{\tilde{O}(D_{\mathrm{AG}})}$ equations as in* (3) *has at most one solution with high probability.*

The time complexity of the basic Arora-Ge approach is then

$$C_{AG}^{\mathrm{plx}} = n^{O(D_{AG})} = 2^{\tilde{O}(\sigma^2)} = 2^{\tilde{O}(n^{2\varepsilon})}.$$

Note also this algorithm also requires $2^{\tilde{O}(n^{2\varepsilon})}$ LWE samples for performing the linearization.

From Linerization to Gröbner Bases

The question we try to address here is whether the complexity C_{AG}^{plx} can be improved by using Gröbner bases instead of linearization. The rational is that you can decrease the constant C_{AG} (and so the degree of the equations) to a value smaller than $\tilde{O}(n^{2\cdot\varepsilon})$ by considering less equations (whilst keeping the probability p_f of failure similar in bother approaches). However, the cost of the solving step increase since one has to compute a Gröbner basis. The question is then to find – if any – a tradeoff allowing to improve upon linearization.

To do so, we will consider a number of equations of the form $\sqrt[\theta]{M_{\mathrm{AG}}}$, with $\theta > 1$ ($\theta = 1$ is the basic Arora-Ge). We want to keep the probability of failure similar for the linearization and Gröbner basis approaches. As a consequence, we need to take a constant C_θ such that:

$$p_f = \frac{\sqrt[\theta]{M_{\mathrm{AG}}}}{e^{O(C_\theta^2)}}.$$

An easy calculation leads to $C_\theta \in \tilde{O}\left(\frac{C_{\mathrm{AG}}}{\sqrt{\theta}}\right)$. Thus, decreasing the number of equations from M_{AG} to $\sqrt[\theta]{M_{\mathrm{AG}}}$ allows to divide the constant C_{AG} by a factor $\sqrt{\theta}$. The degree of the equations we are doing to consider is then equal to $2\sigma \cdot C_\theta + 1 \in \tilde{O}\left(\frac{\sigma^2}{\sqrt{\theta}}\right)$.

The question is now to find a good candidate for θ. Typically, if θ is too big then you will greatly decrease the number of equations, but the cost of the solving step will become prohibitive and the total complexity will be worth than for a linearization.

We have considered a θ of the form: $\theta = n^{2\cdot\beta}$, for some $\beta \geq 0$ (note that we get the basic Arora-Ge by taking $\beta = 0$). In this new setting, we get a constant $C_\beta = n^{\varepsilon-\beta}$. We have then to solve a system having $M_\beta = \sqrt[n^{2\cdot\beta}]{M_{\mathrm{AG}}} \in 2^{\tilde{O}(n^{2(\varepsilon-\beta)})}$ equations of degree $D_\beta = \tilde{O}(n^{2\cdot\varepsilon-\beta})$. We denote such system by $\mathcal{S}_{\mathrm{GB}}(\beta)$.

The question is to determine the complexity $C_{GB-AG}^{\mathrm{plx}}(\beta)$ of solving $\mathcal{S}_{\mathrm{AG}}(\beta)$. This reduces to study its degree of regularity D_{reg}^{β}. Given current algorithms, the specific structure of the system does not allow to solve it faster than random systems. As a consequence, we assume that D_{reg}^{β} is not bigger than the degree of regularity of a semi-regular system of the same size[2], namely:

$$D_{reg}^{\beta} \leq 1 + \deg(H_\beta),$$

where:

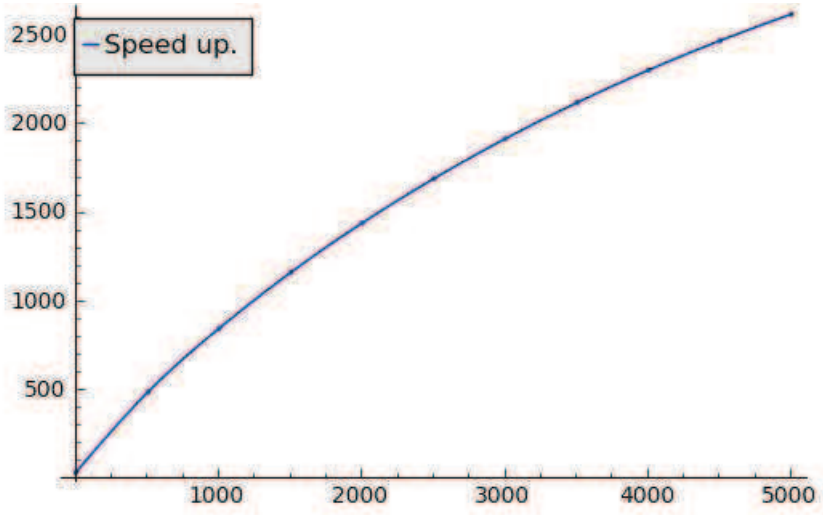
$$H_\beta(z) = \left[\frac{(1-z^{D_\beta})^{M_\beta}}{(1-z)^n}\right]_+,$$

where $[.]_+$ denotes the series obtained by truncating before the index of its first non-positive coefficient.

We present below some experiments performed for $\beta = 1/5$. We have computed explicitly the complexities for both approaches: linearization and Gröbner bases. As suggested in [27],

[2] We have performed few experiments for small parameters. The experiments seem to confirm this hypothesis.

we considered $q \approx n^2$ and $\alpha = 1/(\sqrt{n} \cdot \log^2 n)$ We plotted below the speed-up we obtained, i.e. $\log_2\left(\frac{C^{\text{plx}}_{GB-AG}(\beta)}{C^{\text{plx}}_{AG}}\right)$ (y-axis) for $n, 0 \leq n \leq 5000$. We can see that Gröbner bases allow to improve the complexity of the basic Arora-Ge when $n \leq 5000$ (x-axis). Note that further experiments are required to confirm this behavior when n tends to infinity[3]



However, the form of the speed-up also tends to suggest that we only improve from a constant C^{plx}_{AG}. change the asymptotical behavior of the Arora&Ge approach. we mention that we are currently considering several forms for the β. In particular, β which is not a constant but a function of n. As a conclusion, we also emphasize that Arora-Ge needs exponential (or subexponetial) number of LWE samples. For most cryptosystems based on LWE, you have access to much less samples, typically polynomially-many. In this situation, you have then not enough samples to perform the linearization and the only option to mount the Arora&Ge approach is to solve the system by using Gröbner bases.

References

[1] A. Akavia, S. Goldwasser, and V. Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. In O. Reingold, editor, *TCC*, volume 5444 of *Lecture Notes in Computer Science*, pages 474–495. Springer Verlag, 2009.

[2] M. R. Albrecht and C. Cid. Cold boot key recovery by solving polynomial systems with noise. In J. Lopez and G. Tsudik, editors, *ACNS*, volume 6715 of *Lecture Notes in Computer Science*, pages 57–72, 2011.

[3] M. R. Albrecht and K. G. Paterson. Breaking an identity-based encryption scheme based on dhies. In L. Chen, editor, *IMA Int. Conf.*, volume 7089 of *Lecture Notes in Computer Science*, pages 344–355. Springer Verlag, 2011.

[4] M. R. Albrecht, P. Farshim, J.-C. Faugère, and L. Perret. Polly Cracker, revisited. In *Advances in Cryptology – ASIACRYPT 2011*, volume 7073 of *Lecture Notes in Computer Science*, pages 179–196, Berlin, Heidelberg, New York, 2011. Springer Verlag. full version available as Cryptology ePrint Archive, Report 2011/289, 2011 `http://eprint.iacr.org/`.

[3] Note that the degree of the equations involved being huge, it becomes rather costly to just expand the Hilbert series for the systems considered.

[5] S. Arora and R. Ge. New algorithms for learning in presence of errors. In L. Aceto, M. Henzinger, and J. Sgall, editors, *ICALP*, volume 6755 of *Lecture Notes in Computer Science*, pages 403–415. Springer Verlag, 2011.

[6] M. Bardet. *Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie*. PhD thesis, Université de Paris VI, 2004.

[7] M. Bardet, J.-C. Faugère, and B. Salvy. Complexity study of Gröbner basis computation. Technical report, INRIA, 2002. `http://www.inria.fr/rrrt/rr-5049.html`.

[8] M. Bardet, J.-C. Faugère, and B. Salvy. On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. International Conference on Polynomial System Solving (ICPSS)*, pages 71–75, 2004.

[9] M. Bardet, J.-C. Faugère, B. Salvy, and B.-Y. Yang. Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems. In *Proc. of MEGA 2005, Eighth International Symposium on Effective Methods in Algebraic Geometry*, 2005.

[10] Z. Brakerski and V. Vaikuntanathan. Efficient fully homomorphic encryption from (standard) LWE. In R. Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 97–106. IEEE, 2011.

[11] B. Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, University of Innsbruck, 1965.

[12] B. Buchberger, G. E. Collins, R. G. K. Loos, and R. Albrecht. Computer algebra symbolic and algebraic computation. *SIGSAM Bull.*, 16(4):5–5, 1982.

[13] D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, Varieties and Algorithms*. Springer Verlag, 2005.

[14] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases (F4). *Journal of Pure and Applied Algebra*, 139:61–88, June 1999.

[15] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In T. Mora, editor, *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation ISSAC*, pages 75–83. ACM Press, July 2002. isbn: 1-58113-484-3.

[16] J.-C. Faugère and A. Joux. Algebraic cryptanalysis of hidden field equation (hfe) cryptosystems using gröbner bases. In D. Boneh, editor, *CRYPTO*, volume 2729 of *Lecture Notes in Computer Science*, pages 44–60. Springer Verlag, 2003.

[17] J.-C. Faugère and L. Perret. Polynomial equivalence problems: Algorithmic and theoretical aspects. In S. Vaudenay, editor, *Advances in Cryptology – EUROCRYPT 2006*, volume 4004 of *Lecture Notes in Computer Science*, pages 30–47. Springer Verlag, 2006.

[18] J.-C. Faugère, F. L. dit Vehel, and L. Perret. Cryptanalysis of minrank. In Wagner [28], pages 280–296.

[19] J.-C. Faugère, A. Otmani, L. Perret, and J.-P. Tillich. In *Advances in Cryptology – EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 279–298. Springer Verlag, 2010.

[20] J.-C. Faugère, L. Perret, C. Petit, and G. Renault. Improving the complexity of index calculus algorithms in elliptic curves over binary fields. In D. Pointcheval and T. Johansson, editors, *Advances in Cryptology - EUROCRYPT 2012*, volume 7237 of *Lecture Notes in Computer Science*, pages 27–44. Springer Verlag, 2012.

[21] C. Gentry. *A fully homomorphic encryption scheme*. PhD thesis, Stanford University, 2009. Available at `http://crypto.stanford.edu/craig`.

[22] D. Lazard. Gröbner-bases, Gaussian elimination and resolution of systems of algebraic equations. In *Proceedings of the European Computer Algebra Conference on Computer Algebra*, volume 162 of *Lecture Notes in Computer Science*, Berlin, Heidelberg, New York, 1983. Springer Verlag.

[23] F. S. Macaulay. On some formula in elimination. *London Mathematical Society*, 1(33):3–27, 1902.

[24] C. Peikert. Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In M. Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009*, pages 333–342. ACM, 2009.

[25] C. Peikert, V. Vaikuntanathan, and B. Waters. A framework for efficient and composable oblivious transfer. In Wagner [28], pages 554–571.

[26] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In H. N. Gabow and R. Fagin, editors, *STOC*, pages 84–93. ACM, 2005.

[27] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6), 2009.

[28] D. Wagner, editor. *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings*, volume 5157 of *Lecture Notes in Computer Science*, 2008. Springer Verlag.

M. R. Albrecht INRIA, LIP6, CNRS
malb@lip6.fr
C. Cid Royal Holloway, University of London
carlos.cid@rhul.ac.uk
J.-C. Faugère INRIA, LIP6, CNRS
jean-charles.faugere@inria.fr
R. Fitzpatrick Royal Holloway, University of London
robert.fitzpatrick.2010@live.rhul.ac.uk
L. Perret INRIA, LIP6, CNRS
ludovic.perret@lip6.fr