

Algebra & Number Theory

Volume 6

2012

No. 2



mathematical sciences publishers

Algebra & Number Theory

msp.berkeley.edu/ant

EDITORS

MANAGING EDITOR

Bjorn Poonen
Massachusetts Institute of Technology
Cambridge, USA

EDITORIAL BOARD CHAIR

David Eisenbud
University of California
Berkeley, USA

BOARD OF EDITORS

Georgia Benkart	University of Wisconsin, Madison, USA	Shigefumi Mori	RIMS, Kyoto University, Japan
Dave Benson	University of Aberdeen, Scotland	Raman Parimala	Emory University, USA
Richard E. Borcherds	University of California, Berkeley, USA	Jonathan Pila	University of Oxford, UK
John H. Coates	University of Cambridge, UK	Victor Reiner	University of Minnesota, USA
J-L. Colliot-Thélène	CNRS, Université Paris-Sud, France	Karl Rubin	University of California, Irvine, USA
Brian D. Conrad	University of Michigan, USA	Peter Sarnak	Princeton University, USA
Hélène Esnault	Universität Duisburg-Essen, Germany	Joseph H. Silverman	Brown University, USA
Hubert Flenner	Ruhr-Universität, Germany	Michael Singer	North Carolina State University, USA
Edward Frenkel	University of California, Berkeley, USA	Ronald Solomon	Ohio State University, USA
Andrew Granville	Université de Montréal, Canada	Vasudevan Srinivas	Tata Inst. of Fund. Research, India
Joseph Gubeladze	San Francisco State University, USA	J. Toby Stafford	University of Michigan, USA
Ehud Hrushovski	Hebrew University, Israel	Bernd Sturmfels	University of California, Berkeley, USA
Craig Huneke	University of Kansas, USA	Richard Taylor	Harvard University, USA
Mikhail Kapranov	Yale University, USA	Ravi Vakil	Stanford University, USA
Yujiro Kawamata	University of Tokyo, Japan	Michel van den Bergh	Hasselt University, Belgium
János Kollár	Princeton University, USA	Marie-France Vignéras	Université Paris VII, France
Yuri Manin	Northwestern University, USA	Kei-Ichi Watanabe	Nihon University, Japan
Barry Mazur	Harvard University, USA	Andrei Zelevinsky	Northeastern University, USA
Philippe Michel	École Polytechnique Fédérale de Lausanne	Efim Zelmanov	University of California, San Diego, USA
Susan Montgomery	University of Southern California, USA		

PRODUCTION

contact@msp.org

Silvio Levy, Scientific Editor


See inside back cover or www.jant.org for submission instructions.

The subscription price for 2012 is US \$175/year for the electronic version, and \$275/year (+\$40 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW[®] from Mathematical Sciences Publishers.

PUBLISHED BY

 **mathematical sciences publishers**

<http://msp.org/>

A NON-PROFIT CORPORATION

Typeset in L^AT_EX

Copyright ©2012 by Mathematical Sciences Publishers

Arithmetic of singular Enriques surfaces

Klaus Hulek and Matthias Schütt

Dedicated to the memory of Eckart Viehweg

We study the arithmetic of Enriques surfaces whose universal covers are singular K3 surfaces. If a singular K3 surface X has discriminant d , then it has a model over the ring class field $H(d)$. Our main theorem is that the same holds true for any Enriques quotient of X . It is based on a study of Néron–Severi groups of singular K3 surfaces. We also comment on Galois actions on divisors of Enriques surfaces.

1. Introduction

Enriques surfaces have formed a vibrant research area over the last 30 years. In many respects, they share the properties of K3 surfaces, yet in other aspects they behave differently. This twofold picture is illustrated in this paper which investigates arithmetic aspects of Enriques surfaces.

The arithmetic of Enriques surfaces is only partially well-understood. For instance, Bogomolov and Tschinkel [1998] proved that potential density of rational points holds on Enriques surfaces. The cited work predates all substantial progress on K3 surfaces in the same direction. In fact, until now the corresponding statement for K3 surfaces has not been proved in full generality.

In this paper, we investigate the arithmetic of those Enriques surfaces whose universal covers are singular K3 surfaces, i.e., K3 surfaces with Picard number $\rho = 20$. We will refer to them as *singular Enriques surfaces*. Singular K3 surfaces are closely related to elliptic curves with complex multiplication (CM). These structures will be crucial to our investigations; often they explain arithmetic properties of singular K3 surfaces (see Sections 3 and 6).

We point out one particular property that illustrates these relations: the field of definition. A singular K3 surface of discriminant d has a model over the ring class field $H(d)$ just like elliptic curves with CM in an order of discriminant d , by

Partial support from DFG under grant Hu 337/6-1 is gratefully acknowledged.

MSC2000: primary 14J28; secondary 11E16, 11G15, 11G35, 14J27.

Keywords: Enriques surface, singular K3 surface, elliptic fibration, Néron–Severi group, Mordell–Weil group, complex multiplication.

[Schütt 2007, Proposition 4.1]. Our main theorem states how this property carries over to Enriques surfaces:

Theorem 1.1. *Let Y be an Enriques surface whose universal cover X is a singular K3 surface. Let $d < 0$ denote the discriminant of X . Then Y admits a model over the ring class field $H(d)$.*

The proof of Theorem 1.1 consists in two steps: first we establish a general result for automorphisms of K3 surfaces over number fields (Proposition 2.1); then we extend the afore-mentioned results for fields of definition of singular K3 surfaces to include their Néron–Severi groups (Theorem 2.4). Here we combine two approaches that both rely on elliptic fibrations. In Section 3 we review the theory of singular K3 surfaces and use Inose’s pencil and the theory of Mordell–Weil lattices to deduce Theorem 2.4 for most singular K3 surfaces (see Remark 3.8). On the other hand, Section 4 provides a direct approach for those singular K3 surfaces which are Kummer (Corollary 4.2). Through Shioda–Inose structures, we then connect the two partial results and are thus able to give a full proof of Theorem 2.4 (see 4F).

In Section 5 we address explicit questions. Lattice theoretically one can determine all singular K3 surfaces that admit an Enriques involution. With 61 or 62 exceptions, we give an explicit geometric construction of an Enriques involution on these singular K3 surfaces. This construction combines Shioda–Inose structures (3B) and the base change approach from [Hulek and Schütt 2011, §3].

In Section 6 we discuss the problem of Galois action on Néron–Severi groups. In this context, a different picture arises for Enriques surfaces than for K3 surfaces. The paper concludes with a formulation of several interesting classification problems for Enriques surfaces and K3 surfaces.

2. Automorphisms of K3 surfaces

2A. Basics about K3 surfaces and Enriques surfaces. This paper is concerned with complex algebraic K3 surfaces and Enriques surfaces. Here we briefly review their basic properties. For details the reader is referred to [Barth et al. 2004, Chapter VIII]; information and examples relevant for this paper can also be found in [Hulek and Schütt 2011].

A K3 surface X is a smooth projective surface with trivial canonical bundle $\omega_X \cong \mathcal{O}_X$ that is simply connected. The classical example consists in a smooth quartic in \mathbb{P}^3 ; here we will mostly work with elliptic K3 surfaces and Kummer surfaces.

In terms of the Enriques–Kodaira classification of algebraic surfaces, a complex Enriques surface Y is a smooth projective surface with vanishing irregularity $q(Y) = h^1(Y, \mathcal{O}_Y) = 0$ and $\omega_Y^{\otimes 2} = \mathcal{O}_Y$, but $\omega_Y \neq \mathcal{O}_Y$. Equivalently Y is the quotient

of a K3 surface X by a fixed point free involution τ . Conversely the K3 surface X can be recovered as the universal covering of Y .

The *Néron–Severi group* $\text{NS}(S)$ of an algebraic surface S is the group of divisors up to algebraic equivalence. Here we identify divisors moving in families such as fibres of a fibration. The Néron–Severi group is finitely generated abelian; its rank is called the *Picard number* and denoted by $\rho(S)$. In essence, $\text{NS}(S)$ encodes the discrete structure of the Picard group of S . The intersection pairing endows $\text{NS}(S)$ with a quadratic form that also induces the notion of numerical equivalence.

On a K3 surface algebraic and numerical equivalence coincide, and $\text{NS}(S)$ is torsion-free. Equipped with the intersection form, it becomes an even lattice of signature $(1, \rho(S) - 1)$, the *Néron–Severi lattice*. On an Enriques surface, however, algebraic and numerical equivalence do not coincide, as in $\text{NS}(Y)$ there is two-torsion represented by the canonical divisor K_Y . The quotient gives the torsion-free group of divisors up to numerical equivalence:

$$\text{Num}(Y) = \text{NS}(Y) / \{0, K_Y\}.$$

The intersection pairing endows $\text{Num}(Y)$ with a lattice structure. Contrary to the K3 case, this lattice has always the same rank and abstract shape:

$$\text{Num}(Y) = U + E_8(-1), \quad \text{rank}(\text{Num}(Y)) = 10$$

where U denotes the hyperbolic plane \mathbb{Z}^2 with intersection pairing $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and E_8 is the unique even unimodular positive-definite lattice of rank 8. The -1 indicates that the sign of the intersection form is reversed so that $\text{Num}(Y)$ has signature $(1, 9)$ as predicted by the Hodge index theorem.

The Torelli theorem [Piatetski-Shapiro and Shafarevich 1971] reduces many investigations of complex K3 surfaces X to a study of $H^2(X)$ with its different structures as lattice or Hodge structure. By the cycle class map, $H^2(X)$ contains an algebraic part coming from $\text{NS}(X)$. The orthogonal complement of $\text{NS}(X)$ in $H^2(X, \mathbb{Z})$ is called the *transcendental lattice*:

$$T(X) = \text{NS}(X)^\perp \subset H^2(X, \mathbb{Z}).$$

As another characterisation, $T(X)$ is the smallest primitive sublattice of $H^2(X, \mathbb{Z})$ that contains the (up to scalar unique) 2-form η_X after complexifying.

2B. Surfaces over number fields. We will consider complex surfaces S that admit a model over some number field. This arithmetic setting brings up the natural question whether geometric objects such as $\text{NS}(S)$ or the automorphism group $\text{Aut}(S)$ are defined over the same field. The problem is as follows:

Let X be a complex K3 surface defined over a number field L . The action of its absolute Galois group $G_L = \text{Gal}(\bar{L}/L)$ on $\text{NS}(X)$ factors through a finite extension M/L . We say that $\text{NS}(X)$ is defined over L if $M = L$, i.e., if G_L acts trivially on $\text{NS}(X)$. Throughout this paper, we will verify this property by exhibiting a set of generators of $\text{NS}(X)$ each of which is defined over L . In fact, for elliptic surfaces with section (which we will mostly be concerned with), both conditions are equivalent.

The same terminology is employed for an Enriques surface Y by saying that $\text{NS}(Y)$ or $\text{Num}(Y)$ is defined over a number field L if G_L acts trivially.

Let ψ be an automorphism of a complex K3 surface X . Since we assumed X to be algebraic, the induced automorphism ψ^* acts as multiplication by a root of unity ζ on the holomorphic 2-form η_X . We assume that X is defined over some number field. The next proposition gives a criterion for the field of definition of ψ . This criterion will be crucial for the proof of Theorem 1.1.

Proposition 2.1. *Let X be a K3 surface over some number field L . Let $\psi \in \text{Aut}(X)$ and $\zeta \in \bar{\mathbb{Q}}$ such that $\psi^*\eta_X = \zeta\eta_X$. Assume that $\text{NS}(X)$ is defined over L and $\zeta \in L$. Then ψ is defined over L .*

Proof. We first need to show that ψ is defined over some number field. Essentially this holds true because the automorphism group of any algebraic K3 surface is discrete by [Sterk 1985, Theorem 0.1]. The general idea is well-known: if the field of definition of ψ were to require a transcendental extension of L , then the transcendental generators of this extension could be turned into parameters, so that ψ would come in a nondiscrete family of automorphisms.

Now suppose that ψ is defined over some finite extension M/L . We want to apply the Torelli theorem [Piatetski-Shapiro and Shafarevich 1971] to ψ and its conjugates to deduce that $M = L$. For this purpose, we assume without loss of generality that M/L is Galois. Let $\sigma \in \text{Gal}(M/L)$. Then $\psi^\sigma \in \text{Aut}(X)$, and we claim that $\psi = \psi^\sigma$. Explicitly we can write

$$\psi^\sigma = \sigma \circ \psi \circ \sigma^{-1}.$$

By the Torelli theorem, it suffices to verify the claim for the induced action on $\text{NS}(X)$ and $T(X)$. For $\text{NS}(X)$ this follows directly from the fact that σ and σ^{-1} act trivially by assumption. For $T(X)$, it suffices to check the action on the holomorphic 2-form. One has

$$(\psi^\sigma)^*(\eta_X) = (\sigma^{-1})^* \circ \psi^*(\eta_X) = (\sigma^{-1})^*(\zeta\eta_X) = \zeta^\sigma\eta_X = \psi^*(\eta_X)$$

since $\zeta \in L$. Hence $\psi^* = (\psi^\sigma)^*$ on $H^2(X, \mathbb{Z})$, and the claim $\psi = \psi^\sigma$ follows from the Torelli theorem [Piatetski-Shapiro and Shafarevich 1971]. In consequence, ψ is defined over L . \square

Remark 2.2. The conditions of Proposition 2.1 are sufficient, but not necessary. For instance, we exhibited a K3 surface with an Enriques involution over \mathbb{Q} , but with $\text{NS}(X)$ defined over $\mathbb{Q}(\sqrt{-3})$ in [Hulek and Schütt 2011, §5.3] (see also 6C).

2C. Enriques involutions. Proposition 2.1 has an immediate impact on involutions, and in particular on Enriques involutions. Namely for an involution ψ , the eigenvalue of η_X can only be $\zeta = \pm 1$, so Proposition 2.1 only requires the Néron–Severi group of the covering K3 surface to be defined over L :

Corollary 2.3. *Let X be a K3 surface over some number field L . If $\text{NS}(X)$ is defined over L , then so is every involution on X . In particular, this holds for Enriques involutions.*

Theorem 1.1 requires some concepts that we will discuss in detail in the next section. It concerns K3 surfaces with Picard number 20, the so-called *singular K3 surfaces* (see 3A). By definition, the discriminant of a singular K3 surface X is the determinant of the intersection form on $\text{NS}(X)$. For a singular K3 surface, the discriminant d gives rise to a very particular number field, the ring class field $H(d)$ as we discuss in 3D. In order to deduce Theorem 1.1, it suffices to combine Corollary 2.3 with the following result for any singular K3 surface (admitting an Enriques involution):

Theorem 2.4. *Let X be a singular K3 surface of discriminant d . Consider the ring class field $H(d)$. Then X admits a model over $H(d)$ with $\text{NS}(X)$ defined over $H(d)$.*

The statement about a model over the ring class field $H(d)$ has been known before (cf. [Schütt 2007, Proposition 4.1]), but the extension for the Néron–Severi group seems to have gone unnoted until now. A proof will be given in the next two sections after reviewing the previous relevant results on singular K3 surfaces. We conclude this section with a direct corollary:

Corollary 2.5. *Let Y be an Enriques surface whose universal cover X is a singular K3 surface. Let $d < 0$ denote the discriminant of X . Then Y admits a model over the ring class field $H(d)$ with $\text{Num}(Y)$ defined over $H(d)$.*

The corresponding statement for $\text{NS}(Y)$ does not hold true in general, as we will discuss within the framework of Galois actions on divisors in 6D. See Example 6.10 and Corollary 6.14.

3. Arithmetic of singular K3 surfaces

This section will review those parts of the theory of singular K3 surface that are relevant to our issues. The section culminates in Lemma 3.7, the main step towards the proof of Theorem 2.4. It is based on Shioda–Inose structures and Inose’s fibration. All the required techniques will be explained along the way.

3A. Singular K3 surfaces. A complex K3 surface X is called *singular* if its Picard number $\rho(X) = \text{rank NS}(X)$ equals the maximum number allowed by Lefschetz's theorem:

$$\rho(X) = h^{1,1}(X) = 20.$$

Singular K3 surfaces involve no moduli, so the terminology "singular" should be understood in the sense of exceptional (just like for singular j -invariants of elliptic curves with complex multiplications, a similarity that will become clear very soon). We will discuss fields of definition of singular K3 surfaces in 3D. Recently singular K3 surfaces over \mathbb{Q} have gained some prominence due to modularity; namely, in analogy with the Eichler–Shimura correspondence between modular forms of weight 2 and elliptic curves over \mathbb{Q} , for any suitable modular form of weight 3 there is a singular K3 surface over \mathbb{Q} associated (cf. [Elkies and Schütt 2008b]).

By the Torelli theorem [Piatetski-Shapiro and Shafarevich 1971; Shioda and Inose 1977], singular K3 surfaces are classified up to isomorphism by their transcendental lattices. For a singular K3 surface, the transcendental lattice is even and positive definite of rank two and endowed with an orientation. Up to conjugation in $\text{SL}_2(\mathbb{Z})$, we identify it with the quadratic intersection form

$$Q(X) = \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} \quad (1)$$

with integer entries $a, c \in \mathbb{N}, b \in \mathbb{Z}$ and discriminant $d = b^2 - 4ac < 0$. This number equals the determinant of the intersection form on $\text{NS}(X)$; we refer to it as the *discriminant* of X . By the Torelli theorem [Piatetski-Shapiro and Shafarevich 1971; Shioda and Inose 1977] two singular K3 surfaces are isomorphic if and only if the transcendental lattices admit an isometry preserving the orientation (or equivalently the quadratic forms are conjugate in $\text{SL}_2(\mathbb{Z})$).

The classical example for a singular K3 surface is the Fermat quartic in \mathbb{P}^3 . Here we give an alternative example in terms of an elliptic fibration that will reappear later in another context (5G). Our treatment draws on the theory of elliptic surfaces; all relevant concepts can be found in [Schütt and Shioda 2010] for instance.

Example 3.1. Consider the universal elliptic curve for $\Gamma_1(6)$:

$$\mathcal{E} : y^2 + (t - 2)xy - t(t - 1)y = x^3 - tx^2.$$

Here a point of order six is given by $(0, 0)$. \mathcal{E} gives rise to a rational elliptic surface S over \mathbb{P}^1 . By Tate's algorithm [1975], S has the following singular fibres in Kodaira's notation:

fibre	I_6	I_3	I_2	I_1
t	∞	0	1	-8

Any quadratic base change f of \mathbb{P}^1 gives rise to a K3 surface X . We generally have $\rho(X) \geq 18$ by the Shioda–Tate formula [Shioda 1990, Corollary 5.3], but one can increase the Picard number conveniently by inferring ramification points at singular fibres. For instance, setting $t = -8s^2/(s^2 - 1)$ yields an elliptic K3 surface X with three singular fibres of type I_2 and I_6 each, and thus $\rho(X) = 20$ over \mathbb{C} again by the Shioda–Tate formula and the Lefschetz inequality $\rho(X) \leq h^{1,1}(X)$. On X , there are two additional two-torsion sections with x -coordinate

$$-4s^2(3s \pm 1)(s \mp 1)/(s^2 - 1)^2.$$

General theory shows that the singular fibres do not allow any further torsion in the Mordell–Weil group. Over \mathbb{C} one obtains $\text{MW}(X) = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$. It follows that X is the universal elliptic curve for the group $\Gamma_1(6) \cap \Gamma(2)$. By [Schütt and Shioda 2010, 11.10 (22)], $\text{NS}(X)$ has discriminant -12 . With the discriminant form à la Nikulin [1980, Proposition 1.6.1 and Corollary 1.9.4], one can then compute the transcendental lattice with intersection form $Q(X) = \text{diag}(2, 6)$ (in agreement with the tables in [Shimada and Zhang 2001]).

3B. Shioda–Inose structure. In order to prove the surjectivity of the period map, mathematicians first considered Kummer surfaces. However, singular abelian surfaces (with $\rho(A) = 4$) cannot possibly yield all singular K3 surfaces as Kummer surfaces because the transcendental lattice of a Kummer surface is always two-divisible as an even lattice. In detail, the intersection form is obtained from $T(A)$ by multiplication by 2:

$$T(\text{Km}(A)) = T(A)(2).$$

This problem of nonprimitivity was overcome by Shioda and Inose [1977]. Generally they considered two elliptic curves E, E' . Their product is an abelian surface $A = E \times E'$ and yields the Kummer surface $X' = \text{Km}(E \times E')$. Over \mathbb{C} , the Picard numbers depend on whether E and E' are isogenous ($E \sim E'$) or have complex multiplication (CM):

$$\rho(A) = \begin{cases} 2 & \text{if } E \not\sim E', \\ 3 & \text{if } E \sim E' \text{ without CM,} \\ 4 & \text{if } E \sim E' \text{ with CM,} \end{cases} \tag{2}$$

$$\rho(X') = \rho(A) + 16.$$

The Kummer surface X' admits several jacobian elliptic fibrations. For instance, the projections onto the factors E and E' induce two isotrivial elliptic fibrations on the Kummer surface X' that we will analyse in Section 4. In [Shioda and Inose 1977, §2], a jacobian elliptic fibration with a fibre of type II^* was found on X' . It

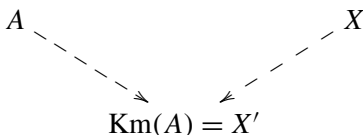
has exactly two further reducible fibres of the following types:

$$\begin{aligned}
 2I_0^* & & E \not\cong E', \\
 I_0^*, I_1^* & & E \cong E', j(E) \neq 0, 12^3, \\
 2I_1^* & & j(E) = j(E') = 12^3, \\
 I_0^*, IV^* & & j(E) = j(E') = 0.
 \end{aligned}$$

Starting from this elliptic fibration, we proceed with the quadratic base change

$$f : \mathbb{P}^1 \rightarrow \mathbb{P}^1$$

that ramifies exactly at the above two reducible singular fibres. Since both ramified fibres are nonreduced, the base change applied to X' results in another elliptic K3 surface X . By construction, the elliptic K3 surface X has two fibres of type II^* and possibly some reducible fibres of type I_2 or IV depending on the above cases. The Kummer surface X' can be recovered from X as (the desingularisation of) the quotient by the involution of the double cover $X \dashrightarrow X'$. (In [Hulek and Schütt 2011] we abused terminology by referring to this involution as deck transformation, but here we will call it base change involution.) The base change involution is a Nikulin involution that composes the involution on the base curve \mathbb{P}^1 with the hyperelliptic involution on the fibres:



The gist of this construction is that the K3 surface X recovers the transcendental lattice of the abelian surface A :

$$T(X) = T(X')(1/2) = T(A). \tag{3}$$

Morrison coined the terminology *Shioda–Inose structure* for such a setting: abelian surface and K3 surface with the same transcendental lattice such that Kummer quotient and Nikulin involution yield the same Kummer surface. He developed lattice theoretic criteria to decide which K3 surfaces of Picard number $\rho \geq 17$ admit a Shioda–Inose structure [Morrison 1984, §6].

3C. Surjectivity of the period map. The surjectivity of the period map requires to exhibit singular K3 surfaces for any quadratic form Q as in (1). By the above considerations, this can be achieved by exhibiting a singular abelian surface A with $Q(A) = Q$ because then the Shioda–Inose structure provides a suitable singular K3 surface X with $Q(X) = Q$.

Chronologically, the corresponding surjectivity statement for singular abelian surfaces was already established before Shioda–Inose’s work by Shioda and Mitani [1974]. Namely, it was shown that any singular abelian surface has product type. Given the quadratic form $Q(A)$ with coefficients as in (1), the abelian surface A admits the representation $A = E \times E'$ with the following elliptic curves given as complex tori $E_\tau = \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$:

$$E = E_\tau, \quad \tau = \frac{-b + \sqrt{d}}{2a}, \quad E' = E_{\tau'}, \quad \tau' = \frac{b + \sqrt{d}}{2}. \tag{4}$$

Note that this representation need not be unique, and in fact there can be arbitrarily many distinct representations for the same singular abelian surface (and thus also for singular K3 surfaces).

Example 3.2. The K3 surface X from Example 3.1 is not a Kummer surface, since $T(X)$ is not two-divisible as an even lattice. Through the Shioda–Inose structure, X arises from the self-product of the elliptic curve $E_{\sqrt{-3}}$ with j -invariant $2^4 3^3 5^3$.

3D. Fields of definition. We have seen that every singular abelian surface A is the product of two elliptic curves with CM in the same field. CM elliptic curves are well-understood thanks to the connection to class field theory (cf. [Shimura 1971, §5]). Indeed both curves in (4) are defined over the ring class field $H(d)$. This field is an abelian Galois extension of the imaginary quadratic field $K = \mathbb{Q}(\sqrt{d})$ with prescribed ramification and Galois group isomorphic to the class group $Cl(d)$ (see [Cox 1989, §9]). We recall one way to describe $Cl(d)$: it consists of $SL_2(\mathbb{Z})$ -conjugacy classes of primitive 2×2 matrices Q as in (1) of discriminant $d < 0$ together with Gauss composition (cf. [Cox 1989, §3] for instance). By [Shimura 1971, Theorem 5.7], $H(d)$ is generated over K by adjoining the j -invariant of E' , or in fact of any elliptic curve with CM by the given order in K of discriminant d . Here $Cl(d)$ acts naturally as a permutation on all these CM elliptic curves – abstractly on the complex tori, but also in a compatible way through the Galois action on $H(d)$ permuting j -invariants.

Shioda–Inose used these CM properties to deduce that any singular K3 surface is defined over some number field. Namely, the Kummer quotient X' respects the base field (a property that we will exploit in Section 4). Hence the only step in the Shioda–Inose structure that may require increasing the base field concerns the elliptic fibration with a fibre of type II^* .

Subsequently Inose [1978] exhibited an explicit model for X over a specific extension of $H(d)$. This model is expressed purely in terms of the j -invariants j, j' of the elliptic curves E, E' from (4):

$$X : \quad y^2 = x^3 - 3At^4x + t^5(t^2 - 2Bt + 1), \tag{5}$$

where $A^3 = jj'/12^6$ and $B^2 = (1 - j/12^3)(1 - j'/12^3)$. Thus we know that any singular K3 surface X of discriminant d admits a model over a degree six extension of $H(d)$. In [Schütt 2007, Proposition 4.1] it was then noted that the above fibration can be twisted in such a way that it is defined over $H(d)$ (cf. (14) in case $AB \neq 0$):

Theorem 3.3. *Let X be a singular K3 surface of discriminant d . Then X has a model over the ring class field $H(d)$.*

In practice, the given field of definition can be far from optimal, that is, X may admit a model over a much smaller number field. In fact, the modularity converse in [Elkies and Schütt 2008b] required to exhibit models of singular K3 surfaces over \mathbb{Q} where the ring class field had degree as large as 32 over \mathbb{Q} . We can already detect a similar behaviour on the level of the elliptic curves E, E' in (4): because of the Galois action of the class group $Cl(d)$, the elliptic curve E' can at best be defined over a quadratic subfield of $H(d)$. The factor E , however, may be defined over \mathbb{Q} even for large d by inspection of the denominators in (4).

3E. Néron–Severi group. In the remainder of this section, we derive an important intermediate result for the proof of Theorem 2.4. The remaining steps will be done in 4F. We have recalled in Theorem 3.3 that any singular K3 surface X admits a model over the ring class field $H(d)$. Here d denotes the discriminant of $T(X)$ as usual. It remains to show that there always is a model of X with $NS(X)$ defined over $H(d)$ as well.

The basic idea for the proof is to work with a model of Inose’s pencil (5) over $H(d)$ as in the proof of [Schütt 2007, Proposition 4.1]:

$$X : y^2 = x^3 + at^4x + t^5(b_2t^2 + b_1t + b_0), \quad a, b_i \in H(d). \tag{6}$$

Note that fibres of type II^* do not admit any inner Galois action (i.e. on fibre components). Hence these two singular fibres of X together with the zero section generate a sublattice $U + 2E_8(-1) \subset NS(X)$ that is fully defined over the base field $H(d)$. It remains to study the Galois action on the remaining generators of $NS(X)$ (there are two generators remaining, since $\rho(X) = 20$). Looking at the other reducible singular fibres, we distinguish four cases as in 3B:

Reducible fibres other than II^*	rank(MW)	case
–	2	$E \not\cong E'$
I_2	1	$E \cong E', j(E) \neq 0, 12^3$
$2I_2$	0	$E \cong E', j(E) = 12^3$
IV	0	$E \cong E', j(E) = 0$

Table 1. Singular fibres and MW-rank of Inose’s pencil.

Lemma 3.4. *If the singular K3 surface X admits an Inose pencil (5) of MW-rank at most one, then X has a model with $\text{NS}(X)$ defined over $H(d)$.*

Proof. For the last two surfaces in Table 1 (MW-rank zero), there are explicit models with $\text{NS}(X)$ defined over \mathbb{Q} (cf. [Schütt 2010, §10]). For the case of MW-rank one with an I_2 fibre, it is also easy to see that $\text{NS}(X)$ can be defined over $L = H(d)$. The fibre does not admit any Galois action, since the identity component is fixed by Galois. By the formula of Shioda–Tate, the Mordell–Weil group has rank one. The Mordell–Weil generator P can only be either fixed or mapped to its inverse by Galois. But if the latter is the case, then the section P is defined over some quadratic extension of L . More precisely, it is given in x, y -coordinates as $P = (U, \sqrt{\gamma}V)$ for some $\gamma \in L, U, V \in L(t)$. Consider the quadratic twist of X with respect to this quadratic extension of L :

$$\gamma y^2 = x^3 + at^4x + t^5(b_2t^2 + b_1t + b_0).$$

This is an alternative model of the fixed elliptic fibration (6) on X over L such that both models become isomorphic over $L(\sqrt{\gamma})$. This quadratic twist transforms the section to (U, V) (defined over L) without introducing any Galois action on the singular fibres (since they only have types I_1, I_2, II, II^*). Thus the Néron–Severi group of the new model of X is defined over $L = H(d)$. \square

Remark 3.5. If $T(X)$ is primitive and lies in the principal genus, then it is possible to replace the CM-curves E, E' by opposite Galois conjugates that are isomorphic: $E^\sigma \cong (E')^{\sigma^{-1}}$. By [Schütt 2007, §6] (which combines [Shimura 1971] and [Shioda and Mitani 1974]), one has $T(E^\sigma \times (E')^{\sigma^{-1}}) = T(E \times E')$. According to Table 1, the induced Inose pencil on X has MW-rank one. By Lemma 3.4 this produces a model of X with $\text{NS}(X)$ defined over $H(d)$.

3F. Mordell–Weil lattices. A similar argument goes through for almost all instances of the case where $E \not\cong E'$. Here we can argue with the Mordell–Weil lattice $\text{MWL}(X)$ of the fibration. In general, the Mordell–Weil lattice of an elliptic surface $S \rightarrow C$ with section was defined in [Shioda 1990] as follows. In $\text{NS}(S)$ consider the trivial lattice $\text{Triv}(S)$ generated by the zero section and fibre components. By [Shioda 1990, Theorem 1.3] there is an isomorphism

$$\text{MW}(S) \cong \text{NS}(S) / \text{Triv}(S).$$

The torsion in $\text{MW}(S)$ is contained in (and determined by) the primitive closure $\text{Triv}(S)'$ of $\text{Triv}(S)$ inside $\text{NS}(S)$. The quotient $\text{MW}(S) / \text{MW}(S)_{\text{tor}}$ is endowed with a lattice structure by means of the orthogonal projection φ in $\text{NS}(S)_{\mathbb{Q}}$ with respect to $\text{Triv}(S)$. Here tensoring with \mathbb{Q} is required unless $\text{Triv}(S)'$ is unimodular. By construction $\varphi(\text{MW}(S))(-1)$ is a positive definite, though not necessarily

integral lattice that one refers to as *Mordell–Weil lattice* $MWL(S)$. The Mordell–Weil lattice satisfies functorial properties for base change and Galois actions. For details the reader is referred to [Shioda 1990] or the survey paper [Schütt and Shioda 2010].

In the present situation the only reducible fibres have type II^* . The nonidentity fibre components generate the root lattice $E_8(-1)$, so $\text{Triv}(X) = U + 2E_8(-1)$. Hence $MWL(X)$ is a positive definite even integral lattice of rank two that fits into the decomposition

$$\text{NS}(X) = U + 2E_8(-1) + MWL(X)(-1).$$

Since $\text{Triv}(X)$ is unimodular, the discriminant forms of $\text{NS}(X)$ and $MWL(X)$ agree up to sign. By [Nikulin 1980, Corollary 1.9.4], this implies that $T(X)$ and $MWL(X)$ lie in the same genus (or in the same isogeny class).

3G. Binary even quadratic forms. To understand the possible Galois actions on $MWL(X)$, we shall need a simple observation about the automorphisms of such lattices. It will be phrased in terms of the corresponding quadratic form Q as in (1). Multiplication by ± 1 gives the trivial automorphisms of Q ; any other automorphism will be called nontrivial. The problem whether Q admits nontrivial automorphisms depends on its order in the class group of even positive definite binary quadratic forms with given discriminant and degree of primitivity:

Lemma 3.6. *The positive-definite quadratic form Q admits a nontrivial automorphism if and only if it is two-torsion in its class group.*

The proof is elementary, so we will omit it here although we did not find a concise reference. For later use, we shall give the possible automorphism groups. Recall that any quadratic form Q as in (1) can be transformed by conjugation in $SL_2(\mathbb{Z})$ to a reduced form where the coefficients satisfy $-a < b \leq a \leq c$ (and $b \geq 0$ if $a = c$). The inverse of a quadratic form is obtained by replacing b by $-b$. A reduced quadratic form is two-torsion if and only if

$$b = 0 \quad \text{or} \quad a = b \quad \text{or} \quad a = c.$$

We obtain the following nontrivial automorphism groups where D_{2n} denotes the dihedral group of order $2n$:

3H. Intermediate step. We conclude this section with an intermediate result towards the proof of Theorem 2.4. In the next section, we will use the Shioda–Inose structure to complete the proof.

Lemma 3.7. *In all cases of MW-rank two in Table 1, the model (5) admits a twist such that there is an $H(d)$ -rational section.*

Q	$\begin{pmatrix} 2a & 0 \\ 0 & 2c \end{pmatrix}$ $a < c$	$\begin{pmatrix} 2a & a \\ a & 2c \end{pmatrix}$ $a < c$	$\begin{pmatrix} 2a & b \\ b & 2a \end{pmatrix}$ $0 < b < a$	$\begin{pmatrix} 2a & 0 \\ 0 & 2a \end{pmatrix}$	$\begin{pmatrix} 2a & a \\ a & 2a \end{pmatrix}$
$\text{Aut}(Q)$	$(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2$	D_8	D_{12}

Table 2. Quadratic forms with nontrivial automorphisms groups.

Proof. If the automorphism group of MWL is only two-torsion, then the lemma follows after a quadratic twist for one of the MW generators. This leaves the cases of the last two quadratic forms in Table 2. Here the class number of Q is one. Hence $T(X)$ has exactly the intersection form Q . In the Shioda–Inose structure, we can choose E by (4) with j -invariant $j = 12^3$ resp. $j = 0$. The extra automorphism of E induces an extra automorphism on X that respects the elliptic fibration (5):

$$(x, y, t) \mapsto (-x, iy, -t) \text{ resp. } (x, y, t) \mapsto (\varrho x, y, t)$$

where ϱ, i denote primitive third resp. fourth roots of unity. The respective automorphism makes $\text{MWL}(X)$ into a module of rank one over $\mathbb{Z}[i]$ resp. $\mathbb{Z}[\varrho]$. This identification is compatible with the Galois action over $H(d)$, since the automorphisms are defined over $H(d)$. Hence it suffices to study the Galois action on the given modules of rank one. Their only automorphisms are the units in $\mathbb{Z}[i]$ resp. $\mathbb{Z}[\varrho]$, i.e. the group of fourth resp. sixth roots of unity. On the elliptic curves with CM by these rings, it is well-known that such a Galois action can be accounted for by biquadratic or sextic twisting (see [Silverman 1994, §II, Example 10.6 and Exercises 2.33, 2.34] or [Schütt 2008, §8]). Thanks to the special shape of the present Weierstrass form (5) with $A = 0$ or $B = 0$, this translates directly into twists of X . Thus there is a twist with $\text{MWL}(X)$ defined over $H(d)$. \square

Remark 3.8. If MWL admits no nontrivial automorphisms, then Lemma 3.7 already settles Theorem 2.4 completely. By the proof of Lemma 3.7, this also holds for MWL with nonabelian automorphism group (the last two entries in Table 2). It is the two-torsion cases of Table 2 that require an extra argument.

In the next section, we will use the Shioda–Inose structures and study Kummer surfaces of product type in detail. In this case, although we may not have any automorphisms on the Kummer surface to relate the MW-generators, we can use the endomorphisms of the abelian surface instead. This approach will enable us to complete the proof of Theorem 2.4 in 4F.

4. Singular Kummer surfaces of product type

Let E, E' be isogenous complex elliptic curves with CM. Then the abelian surface $A = E \times E'$ is singular ($\rho(A) = 4$). Let d denote its discriminant (that is the

discriminant of $T(A)$). Then E, E' have models over the ring class field $H(d)$ (obtained from the CM-field by adjoining the j -invariants).

Throughout this section, we only consider the case where $E \not\cong E'$ (MW-rank two) and no j -invariant equals 0 or 12^3 (no extra automorphisms). The same results hold in the other cases, but we would have to distinguish more subcases and also consider biquadratic/sextic twisting etc. Note that for the excluded cases we have already given a full proof of Theorem 2.4 in Lemma 3.4 (for $E \cong E'$) and in the proof of Lemma 3.7 (for j or $j' \in \{0, 12^3\}$; cf. Remark 3.8). Thus the cases considered explicitly in this section will suffice to complete the proof of Theorem 2.4.

4A. Consider the Kummer surface $X' = \text{Km}(A)$. Recall the isotrivial elliptic fibrations on X' that are induced by the projections onto E and E' from 3B. These are naturally defined over $H(d)$ as follows. Fix Weierstrass models

$$E : y^2 = f(x), \quad E' : y^2 = g(x) \quad (7)$$

with cubic polynomials $f, g \in H(d)[x]$. Then X' admits a birational model

$$X' : f(t)y^2 = g(x) \quad (8)$$

with the structure of an elliptic curve over the function field $H(d)(t)$. We denote the corresponding elliptic fibration by the pair (X', π) . This fibration has singular fibres of type I_0^* at ∞ and at the zeroes of $f(t)$. Over $\bar{\mathbb{Q}}$ we have $\text{MW}(X', \pi) = \mathbb{Z}^2 \times (\mathbb{Z}/2\mathbb{Z})^2$ with torsion sections given by the roots of $g(x)$.

Proposition 4.1. *The elliptic fibration (X', π) admits a model over $H(d)$ such that MW is generated by two-torsion and sections defined over $H(d)$. In particular, MWL is generated by sections defined over $H(d)$.*

Proof. By the Shioda–Tate formula, the Mordell–Weil lattice has rank two since $\rho(X') = 20$. Due to the singular fibre types $\text{MWL}(X', \pi)$ will not be integral, but it is positive-definite. Hence the results from 3G and 3H apply directly to prove the claim with the exception of the first three special cases from Table 2. Here we pursue an alternative uniform approach based on the fact that as in Lemma 3.7 we can find a quadratic twist with at least one MW-generator P over $H(d)$.

The crucial ingredient is the following lattice isomorphism which Shioda established in [Shioda 2007, Proposition 3.1]:

$$\text{Hom}(E, E') \cong \text{MWL}(X', \pi). \quad (9)$$

Here $\text{Hom}(E, E')$ is endowed with a norm given by the degree. The isomorphism takes a homomorphism $\phi : E \rightarrow E'$ as input. Via its graph Γ_ϕ in A and the image $\bar{\Gamma}_\phi$ in X' , one associates to ϕ the element \bar{R}_ϕ in $\text{MWL}(X', \pi)$ corresponding to $\bar{\Gamma}_\phi$ under the orthogonal projection $\text{NS}(X') \rightarrow \text{MWL}(X', \pi)$ (see 3F).

Shioda [2007] worked over an algebraically closed field, so that the isomorphism (9) is independent of the chosen model. However, for the specified models in (7), (8) the isomorphism (9) is clearly Galois-equivariant.

Following Lemma 3.7, we apply a quadratic twist on X' such that there is an $H(d)$ -rational section P (nontorsion). That is, for some $c \in H(d)$ we consider the $H(d)(\sqrt{c})$ -isomorphic model

$$X' : cf(t)y^2 = g(x).$$

In terms of the elliptic curves E, E' , this is accounted for by twisting *one* elliptic curve by \sqrt{c} , say:

$$E : y^2 = f(x), \quad E' : cy^2 = g(x). \tag{10}$$

For these models, the isomorphism (9) is by construction again Galois-equivariant. Hence the section P corresponds to a homomorphism $\phi : E \rightarrow E'$ over $H(d)$. Now pick any endomorphism ϵ of E' that is not multiplication by an integer. By CM-theory, ϵ is defined over $H(d)$, and together $\phi, \epsilon \circ \phi$ generate the lattice $\text{Hom}(E, E')$ up to finite index. In conclusion, (9) gives a section $R_{\epsilon \circ \phi}$ over $H(d)$ that is independent of P . By construction, these sections generate $\text{MWL}(X', \pi)$ up to finite index. Proposition 4.1 thus follows. \square

4B. Néron–Severi group of Kummer surfaces. We collect a few consequences of Proposition 4.1. We start with a version of Theorem 6.3 for singular Kummer surfaces. Note that since $T(X') = T(A)(2)$, the Kummer surface X' has discriminant $4d$.

Corollary 4.2. *The singular Kummer surface X' has a model over $H(d)$ with $\text{NS}(X')$ defined over $H(4d)$.*

Proof. Fix the model of the elliptic fibration (X', π) from Proposition 4.1 with MW-rank two over $H(d)$. In order to generate $\text{NS}(X')$, we have to add to these $H(d)$ -rational sections the two-torsion sections and the components of the I_0^* fibres. These rational curves are defined over the splitting field of the polynomials $f(t), g(x)$ over $H(d)$. That is, we adjoin to $H(d)$ the x -coordinates of the two-torsion points of E and E' . By the analogue of the Kronecker–Weber theorem for imaginary quadratic number fields [Silverman 1994, §II Theorem 5.6], these algebraic numbers generate exactly $H(4d)$ over $H(d)$. \square

4C. Isogenous CM-elliptic curves. Before continuing with the proof of Theorem 2.4, we note another implication of Proposition 4.1. Here we are concerned with the field of definition of the isogeny between E and E' . By the classical theory, any two elliptic curves with CM in the same field K have models over some minimal ring class field H ; moreover they are isogenous over $\bar{\mathbb{Q}}$. Here we ask whether they

admit H -isogenous models, i.e. models over H with isogeny defined over H as well. When the CM-curves are \mathbb{Q} -curves, this property comes for free, but this situation does not always persist (cf. Remark 4.4). The following result might be well-known to the experts, but we could not find a reference.

Corollary 4.3. *Let E, E' be elliptic curves with CM by orders in the same imaginary quadratic field K . Let $H = K(j(E), j(E'))$. Then E, E' have H -isogenous models.*

Proof. We can start with any two Weierstrass forms over H as in (7). The proof of Proposition 4.1 exhibits a quadratic twist of E' with a nontrivial homomorphism $\phi : E \rightarrow E'$. \square

Remark 4.4. Corollary 4.3 only seemingly conflicts with a result of Gross [1980, §11]. Namely, Gross found that there are CM-elliptic curves which are not \mathbb{Q} -curves, i.e. E is not H -isogenous to all its conjugates. Here we let $E' = E^\sigma$ be a conjugate of E . If E, E^σ are not H -isogenous (so that E is not a \mathbb{Q} -curve), then Corollary 4.3 provides us with a quadratic twist of E^σ which is H -isogenous to E . But then the quadratic twist of E^σ and E are not conjugate any more, so there is no contradiction to E 's failure of being a \mathbb{Q} -curve.

4D. Auxiliary elliptic fibration. Recall the singular K3 surface X with Inose's elliptic fibration (5). By [Shioda 2006] the quadratic base change $t = u^2$ recovers the Kummer surface X' . Since X also dominates X' by the Shioda–Inose structure, Shioda alluded to this picture as X being sandwiched by the Kummer surface X' . In the base change, the two fibres of type II^* are replaced by fibres of type IV^* . Let us explain how to find this base changed fibration on the previous model of X' :

$$X' : cf(t)y^2 = g(x).$$

Projection onto the affine coordinate $u = y$ endows X' with the structure of an elliptic fibration π' since the fibres are plane cubics in x, t . Write (X', π') for X' with this fixed elliptic fibration. Visibly (X', π') is the quadratic base change of the rational elliptic surface S' obtained by setting $u^2 = v$. S' has singular fibres of type IV at $v = 0, \infty$; in X' they are replaced by fibres of type IV^* as alluded to before. Here S' is given as a cubic pencil whose base points form sections. Recall that these sections are all defined over $H(4d)$.

By base change $\text{MWL}(S')(2)$ embeds into $\text{MWL}(X', \pi')$. Consider the orthogonal complement

$$L = [\text{MWL}(S')(2)]^\perp \subset \text{MWL}(X', \pi').$$

By construction, L is exactly the invariant sublattice of $\text{MWL}(X', \pi')$ for the involution corresponding to the base change $X' \rightarrow X$, i.e. $L = \text{MWL}(X)(2)$.

Over $\bar{\mathbb{Q}}$ (or in fact algebraically closed fields of characteristic $\neq 2, 3$), Shioda used a similar argument as for the isomorphism (9) to derive an isomorphism

$$L \cong \text{Hom}(E, E')(4), \quad \text{so that} \quad \text{MWL}(X) \cong \text{Hom}(E, E')(2). \quad (11)$$

Compared to the previous argument that gave (9), there is one subtlety here: For $\phi \in \text{Hom}(E, E')$, the orthogonal projection onto $L_{\mathbb{Q}}$ maps the divisor $\bar{\Gamma}_{\phi}$ to $\frac{1}{2}L$. This holds true since the quotient $\text{MWL}(X', \pi')/(L+L^{\perp})$ need not be trivial (hence we tensor L with \mathbb{Q} a priori), but due to the quadratic base change the quotient is always isomorphic to a finite number of copies of $\mathbb{Z}/2\mathbb{Z}$. Now instead of $\bar{\Gamma}_{\phi}$, one takes the image of the divisor $2\bar{\Gamma}_{\phi}$ in L . Computing intersection numbers using the theory of Mordell–Weil lattices, Shioda verifies the isomorphism (11). In our setting, the main problem is to find models which make the isomorphisms (11) Galois-equivariant over a suitable field.

4E. Galois equivariance. We know that E, E' admit $H(d)$ -isogenous models, so that $\text{Hom}(E, E')$ is generated by isogenies over $H(d)$. The elliptic fibration π' on X' is defined over $H(d)$ as well, but in order to endow it with a section (a base point of the cubic pencil), we may have to increase the base field to $H(4d)$. This makes the isomorphisms in (11) for the specified models Galois-equivariant over $H(4d)$. For X , however, we need a model with MWL over $H(d)$, so we have to throw in some more information. We distinguish two cases according to the degree h of the Galois extension $H(4d)/H(d)$. Note that with the Legendre symbol $(\cdot/2)$ at 2, one obtains from the class number formula

$$h = \deg(H(4d)/H(d)) = \begin{cases} 1 & \text{if } (d/2) = 1 \text{ or } d = -3, -4; \\ 2 & \text{if } 2 \mid d, d \neq -4; \\ 3 & \text{if } (d/2) = -1, d \neq -3. \end{cases}$$

4E.1. First case: $h = 1, 2$. This case is very simple. By assumption, both polynomials f, g have a root over $H(d)$. A base point of the cubic pencil gives an $H(d)$ -rational section of the elliptic fibration (X', π') . Due to the singular fibre types and the involution $u \mapsto -u$, we obtain a Weierstrass form

$$X' : y'^2 = x'^3 - 3au^4x' + u^4(b_2u^4 - 2b_1u^2 + b_0). \quad (12)$$

As quotient by the base change involution $u \mapsto -u$ of $X' \rightarrow S'$ composed with the hyperelliptic involution $y' \mapsto -y'$, we obtain a model of X over $H(d)$. Compared to (5), this Weierstrass form is not yet normalised with respect to b_0, b_2 .

By construction, the isomorphisms (11) are $H(d)$ -Galois equivariant for these specific models of E, E', X', X . That is, we have exhibited a model of X over $H(d)$ with fibration of type (5) and MW-rank two over $H(d)$. It follows that this model has $\text{NS}(X)$ defined over $H(d)$.

4E.2. *Second case:* $h = 3$. In this case, we compare two $\bar{\mathbb{Q}}$ -isomorphic models that we denote by X_1, X_2 . From (12), we obtain a model over $H(4d)$ as quotient by the Nikulin involution $(x', y', u) \mapsto (x', -y', -u)$:

$$X_1: y'^2 = x'^3 - 3au^4x' + u^5(b_2u^2 - 2b_1u + b_0) \quad (13)$$

with $\text{MWL}(X_1)$ defined over $H(4d)$ by the Galois-equivariant isomorphism (11). From (5), we derive a model over $H(d)$

$$X_2: y^2 = x^3 - 3c^2B^2A^3t^4x + c^3B^2A^3t^5(B^2t^2 - 2B^2t + 1). \quad (14)$$

Here $B^2, A^3 \in H(d)$ as given in 3B. By Lemma 3.7, we can choose $c \in H(d)$ in such a way that X_2 has an $H(d)$ -rational section P and an orthogonal section Q defined over some quadratic extension M of $H(d)$. We assume that $M \neq H(d)$ and derive a contradiction from the above two models. Essentially, this works because we compare a quadratic and a cubic extension of $H(d)$.

By assumption, we can choose Q anti-invariant under conjugation in $M/H(d)$ (so that P, Q generate $\text{MW}(X_2)$ up to finite index). Hence there are rational functions $x_Q, y_Q \in H(d)(t)$ and some constant $c_Q \in H(d)$ such that

$$Q = (x_Q, \sqrt{c_Q}y_Q) \quad \text{and} \quad M = H(d)(\sqrt{c_Q}).$$

We work out an isomorphism of the two elliptic fibrations X_1, X_2 . This can only take the shape

$$(x, y, t) \mapsto (x', y', u) = (\gamma\alpha^2x, \alpha^3\gamma^{3/2}y, \alpha t). \quad (15)$$

Thus we require

$$a = \gamma^2(c^2B^2A^3), \quad b_1 = \gamma^3(c^3B^4A^3), \quad \alpha b_2 = \gamma^3(c^3B^4A^3), \quad b_0 = \alpha\gamma^3(c^3B^2A^3).$$

The first two relations give $\gamma = b_1/(acB^2) \in H(4d)$, so that also $\alpha \in H(4d)$. The section P on X_2 with $H(d)$ -rational y -coordinate $y_P(t)$ pulls back to a section P_1 with y' -coordinate $\gamma^{3/2}\alpha^3y_P(\alpha t)$. By construction, P_1 is $H(4d)$ -rational, so $\gamma^{3/2} \in H(4d)$. But here $H(4d)$ has degree three over $H(d)$, so $\gamma^{3/2} \in H(d)$. In other words, the isomorphism (15) is defined over $H(4d)$.

In consequence, Q pulls-back to a section on X_1 with y' -coordinate $\sqrt{c_Q}$ times an $H(4d)$ -rational function. The same argument as for $\gamma^{3/2}$ then shows that $\sqrt{c_Q} \in H(d)$. This gives the required contradiction.

4F. Proof of Theorem 2.4. We collect all results necessary to prove Theorem 2.4. Let X be a singular K3 surface of discriminant d . We decided to work with Inose's pencil over $H(d)$ as in (14). Thus it suffices to check the field of definition of $\text{MW}(X)$ to verify Theorem 2.4. In many cases, this was achieved in Lemma 3.4

or in the intermediate Lemma 3.7 (as explained in Remark 3.8). For the remaining K3 surfaces, we considered the Kummer surface X' from the Shioda–Inose structure which actually sandwiches X (4D). Note that for Kummer surfaces we exhibited a proof of Theorem 2.4 that only uses the techniques from Lemma 3.7 (Proposition 4.1, Corollary 4.2). Thanks to the interplay between $H(d)$ and $H(4d)$, this suffices to deduce that $\text{MW}(X)$ is defined over $H(d)$ by 4E. This completes the proof of Theorem 2.4. \square

5. Enriques surfaces of base change type

This section provides a technique to construct explicit examples of Enriques surfaces whose covers are singular K3 surfaces. In the sequel, we refer to them as *singular Enriques surfaces*. The main idea is to invoke the base change construction from [Hulek and Schütt 2011, §3] for singular K3 surfaces. We will review the concept in 5B and then relate it to the Shioda–Inose structures from 3B.

5A. Singular K3 surfaces with Enriques involution. Our first problem concerns K3 surfaces: Which singular K3 surfaces admit an Enriques involution? Keum’s result [1990] gives a partial answer for all singular K3 surfaces that are Kummer surfaces (i.e. with transcendental lattice two-divisible). The full problem can also be solved by purely lattice-theoretic means in terms of the transcendental lattice. In fact, one finds that the discriminant almost suffices to reach a decision: it suffices for non-Kummer surfaces while for Kummer surfaces we know the answer anyway from [Keum 1990]. Sertöz [2005] gave the solution based on the techniques developed by Keum:

Theorem 5.1. *Let X be a singular K3 surface of discriminant d . Then X does not admit an Enriques involution exactly in the following cases:*

- (i) $d \equiv -3 \pmod{8}$,
- (ii) $d = -4, -8$,
- (iii) $d = -16$ and X is not Kummer, i.e. $Q(X) = \text{diag}(2, 8)$.

Note that the discriminants in case (ii) determine unique singular K3 surfaces up to isomorphism. In case (iii), we have to exempt the Kummer surface $\text{Km}(E_i \times E_i)$ with transcendental lattice of intersection form $Q = \text{diag}(4, 4)$ which admits an Enriques involution by [Keum 1990].

Sertöz’s proof is purely lattice theoretic and based on machine computations. In particular, for those singular K3 surfaces admitting some Enriques involution, it does not give any explicit geometric description of any such involution. Here we shall combine the ideas from [Hulek and Schütt 2011, §3] and Section 3 to derive explicit Enriques involutions on almost all singular K3 surfaces possible according to Theorem 5.1.

5B. Enriques involutions of base change type. We start by reviewing the set-up from [Hulek and Schütt 2011, §3]:

S	rational elliptic surface
f	quadratic base change of \mathbb{P}^1 (not ramified at nonreduced fibres of S)
X	base change of S by f : K3 surface
ι	base change involution
(-1)	hyperelliptic involution
$\boxplus P$	translation by a section $P \in \text{MW}(X)$

In this situation, the composition $j = \iota \circ (-1)$ defines a Nikulin involution on X , i.e. j has eight isolated fixed points and leaves the holomorphic two-form invariant. The quotient X/j has a resolution X' that is again K3. X' is the quadratic twist of S at the ramification points of the base change f : The induced action of ι and j gives a decomposition of $\text{MW}(X)$ up to some 2-power index:

$$\text{MW}(X)_{\mathbb{Q}} \cong \text{MW}(S)_{\mathbb{Q}} + \text{MW}(X')_{\mathbb{Q}}. \quad (16)$$

Let $P' \in \text{MW}(X')$ and P denote the induced section on X . By construction, P is anti-invariant for ι^* . In consequence,

$$\tau := \boxplus P \circ \iota$$

is an involution on X . By definition, this involution can only have fixed points on the fixed fibres of ι . If these fibres are smooth, one has

$$\text{Fix}(\tau) = \emptyset \iff P \cap O \cap \text{Fix}(\iota) = \emptyset.$$

The latter condition can be checked with P' on the ramified fibres of X' (generally of type I_0^*). Here P' has to meet nonidentity components.

Example 5.2. The prototype example for this construction is a two-torsion section P induced from X' (or equivalently from S since two-torsion is not affected by quadratic twisting). Outside characteristic two, such a section is always disjoint from O . For τ to have fixed points, one of the ramified fibres has to be singular such that it is additive or P meets the identity component.

The latter occurs for Example 3.1: There is exactly one two-torsion section induced from S . This section $(t-1, t-1)$ meets both ramified fibres (at 0 and ∞) at their identity components. The other two-torsion sections are interchanged by ι (which is why (16) only holds after tensoring with \mathbb{Q}).

5C. We ask which singular K3 surfaces admit an Enriques involution of base change type. For now we only exclude 62 or 63 singular K3 surfaces as specified in Exception 5.5 (62 assuming some special cases of ERH; see 5E).

Proposition 5.3. *Let X be a singular K3 surface admitting an Enriques involution. Assume that X is not among the 62 or 63 K3 surfaces from Exception 5.5. Then X has an Enriques involution τ of base change type where the Nikulin quotient X' is a Kummer surface.*

The proof of the proposition will be given in 5E and 5F. It is based on the Shioda–Inose structure of singular K3 surfaces to that we will return next.

One word about Exception 5.5: we do not believe this exception to be necessary, but we have not found a general argument to overcome it (cf. Remark 5.6). To illustrate this, we will show in 5G that Example 3.1 which falls under Exception 5.5 does indeed admit an Enriques involution of base change type (but we did not check whether the quotient X' is a Kummer surface).

5D. Enriques involutions and Shioda–Inose structures. Let E, E' denote elliptic curves and consider the corresponding Shioda–Inose structure as in 3B. Then $X' = \text{Km}(E \times E')$ admits an Enriques involution by [Keum 1990], but how about the K3 surface X from 3B that recovers the transcendental lattice of the abelian surface $E \times E'$?

If E and E' are not isogenous, then X has Picard number $\rho(X) = 18$ and the fibration (5) of Mordell–Weil rank zero yields

$$\text{NS}(X) = U + 2E_8(-1).$$

This lattice does not admit any primitive embedding of the Enriques lattice $U(2) + E_8(-2)$ because of the 2-length. Hence the K3 surface X cannot have an Enriques involution. We now consider the case where E and E' are isogenous, possibly with CM.

Here is our main tool to construct explicit Enriques involutions: the Shioda–Inose structure falls under the settings studied in 5B. We already chose the notation to indicate this: there is a K3 surface X with a Nikulin involution yielding the Kummer surface X' . Conversely, X is obtained from X' by a quadratic base change. In terms of the elliptic fibration (5) on X , the Nikulin involution is given as

$$J : (x, y, t) \mapsto (x/t^4, -y/t^6, 1/t).$$

Thus the quotient X/J attains singularities in the fibres at $t = \pm 1$ whose minimal resolution is X' . In general, the quotient results in fibres of type I_0^* , but there are other possibilities as sketched in 3B. Concretely, there is another involution corresponding to the base change $\mathbb{P}^1 \rightarrow \mathbb{P}^1$ induced by $X \rightarrow X'$:

$$\iota = J \circ (-1) : (x, y, t) \mapsto (x/t^4, y/t^6, 1/t).$$

The quotient X/ι gives a rational elliptic surface S . It extends the Shioda–Inose structure to the following diagram (where we could also add the induced elliptic

fibrations):

$$\begin{array}{ccc}
 E \times E' & & X \\
 \swarrow \text{---} & & \swarrow \text{---} \\
 & \text{Km}(E \times E') = X' & \\
 & \nwarrow \text{---} & \searrow \text{---} \\
 & & S
 \end{array}$$

By construction, S has a singular fibre of type II^* . From the Shioda–Tate formula [Shioda 1990, Corollary 5.3], it follows that S is extremal, i.e., it has finite Mordell–Weil group. Since a singular fibre of type II^* does not admit any torsion sections (of order relatively prime to the characteristic), we infer that $\text{MW}(S) = \{O\}$. By [Shioda 1990, Proposition 8.12] (cf. (16)), this implies that

$$\text{MWL}(X) = \text{MWL}(X')(2).$$

Hence as soon as the Mordell–Weil rank of X is positive, there is a section P (induced from X') and an involution τ as in 5B. In order to exhibit an Enriques involution on X , it remains to determine whether τ is fixed point free. In general there are three cases of positive Mordell–Weil rank to be distinguished according to the types of singular fibres. For nonsingular K3 surfaces, i.e., Mordell–Weil rank one with $E \not\cong E'$ and $\rho = 19$, this has been done in [Hulek and Schütt 2011, §4.2] (without referring to Shioda–Inose structures). The property whether τ is fixed point free or not depends on the parity of the height of the Mordell–Weil generator modulo 4. In the next sections, we will treat the singular cases and thus prove Proposition 5.3.

Remark 5.4. There is a natural continuation of this connection between Enriques involutions of base change type and Shioda–Inose structures. Recall from Section 4 that the K3 surface X is sandwiched by the Kummer surface X' in the following sense: X' can also be recovered from X by the quadratic base change $u \mapsto t = u^2$ applied to (5). As in 5B, each section of X induces an involution τ of base change type on the Kummer surface X' . Here we ask whether τ is an Enriques involution. We have seen that the base change replaces the fibres of type II^* by type IV^* (so these are fixed by τ). However, none of these fibre types admits a free involution, so there cannot be an Enriques involution on X' as in 5B for the specified base change.

5E. Mordell–Weil rank one and $E \cong E'$. In this case, E is a CM elliptic curve with $j(E) \neq 0, 12^3$. The elliptic fibration (5) on X has exactly one reducible fibre of type I_2 at $t = 1$ in addition to the two fibres of type II^* . Together with the Mordell–Weil generator P , we can write

$$\text{NS}(X) = U + 2E_8(-1) + \langle A_1(-1), P \rangle.$$

We consider two cases according to the intersection behaviour of the section P and the fibre of type I_2 .

If P meets the nonidentity component of the I_2 fibre, then P has height

$$h(P) = 4 + 2(P \cdot O) - 1/2.$$

Equivalently, the discriminant $d = -2h(P)$ of X is odd. Clearly, P and O do not intersect on the I_2 fibre which is one of the two fixed fibres of the base change involution ι . Here translation by P exchanges the fibre components including the nodes, so it acts freely on the singular fibre. It remains to check for the specialisation of P on the other fixed fibre at $t = -1$. Note that P is induced from a section P' on the Nikulin quotient X' , so

$$P \cdot O = 2(P' \cdot O') + \#P \cap O \cap \text{Fix}(\iota).$$

Since P and O can only possibly intersect on the irreducible fixed fibre of ι at $t = -1$, the parity of the intersection number $P \cdot O$ depends only the intersection behaviour at that fibre. In consequence, the discriminant d of X satisfies the congruence

$$d \equiv -7 \pmod{8} \iff P \cap O \cap \text{Fix}(\iota) = \emptyset \iff \text{Fix}(\tau) = \emptyset.$$

In comparison, Theorem 5.1 states that a singular K3 surface of odd discriminant d admits an Enriques involution if and only if $d \equiv -7 \pmod{8}$. This proves Proposition 5.3 for all odd discriminants and MW rank one cases. (As explained in3H such fibrations exist on X if and only if the transcendental lattice is primitive and lies in the principal genus.)

We now consider the case where X has even discriminant, i.e., the section P meets the identity component of the I_2 fibre. Then τ fixes both fibre components. As they are isomorphic to \mathbb{P}^1 , there are fixed points. (In fact one can see that τ fixes one component pointwise.) In conclusion, the given elliptic fibration (5) on X does not admit an Enriques involution of base change type.

This failure to produce an Enriques involution poses the problem how it can be overcome for the singular K3 surfaces in consideration for Proposition 5.3. Recall that we are in the special case where the fibration (5) corresponds to $E \cong E'$. The principal idea now is to choose an alternative elliptic fibration of the same kind on X , but for a pair (E, E') such that $E \not\cong E'$ (resembling our approach in3H). Whenever this is possible, the new fibration falls under the next case of Mordell–Weil rank two, and Proposition 5.3 can be proved along those lines. Here we can vary the pair (E, E') by conjugates $(E^\sigma, (E')^{\sigma^{-1}})$. This fails to return a fibration of MW rank two if and only $E^\sigma \cong E^{\sigma^{-1}}$ for all Galois elements σ . Equivalently, the class group is only two-torsion. Note that $E \cong E'$ implies that $T(E \times E') = T(X)$ is primitive and lies in the principal genus. Since the same applies to all conjugates,

we derive the following abstract characterisation of the singular K3 surfaces where the Shioda–Inose structure does not produce an Enriques involution of base change type:

Exception 5.5. A singular K3 surface X of even discriminant d does not admit an elliptic fibration (5) of Mordell–Weil rank two if and only if $T(X)$ is primitive and gives the full principal genus of its class group. In other words $Q(X) = \text{diag}(2, |d|/2)$ and the class group $Cl(d)$ is only two-torsion.

There are 101 known discriminants $d < 0$ such that $Cl(d)$ is only two-torsion; the discriminant of biggest absolute value is $d = -7392$. By [Weinberger 1973], there could be one more such discriminant of size $> 10^{10}$, but this is ruled out by the extended Riemann hypothesis for odd real Dirichlet characters. Out of the 101 known discriminants, 65 are even (they were already studied by Euler, cf. [Cox 1989]) and $-4, -8, -16$ are ruled out by Theorem 5.1, so the above exception concerns 62 or 63 singular K3 surfaces. We consider one of them in detail in 5G after completing the proof of Proposition 5.3.

Remark 5.6. For each of the 62 known singular K3 surfaces from Exception 5.5, one could try to exhibit an Enriques involution as in 5B for a different base change than in the Shioda–Inose structure. However, there does not seem to be a universal way to achieve this. Notably, the general K3 surface X arising from the Shioda–Inose structure for the present case $E \cong E'$ only admits four essentially different jacobian elliptic fibrations. To see this, one can argue with a gluing technique of Kneser–Witt that has been successfully applied to K3 surfaces in [Nishiyama 1996]. For these four fibrations, the fibre types reveal that only (5) and one other fibration can arise through a quadratic base change. The latter pulls back from the unique rational elliptic surface with a singular fibre of type I_9 and $MW = \mathbb{Z}/3\mathbb{Z}$ by the one-dimensional family of quadratic base changes that ramify at the reducible fibre. A case-by-case analysis (exactly as above) shows that a singular elliptic K3 surface within this family can only have an Enriques involution of base change type if it does not fall under Exception 5.5.

5F. Mordell–Weil rank two. In this case, E and E' are isogenous, but nonisomorphic elliptic curves with CM. Both fixed fibres for the base change involution ι at $t = \pm 1$ are smooth. On the Nikulin quotient X' , they correspond to fibres of type I_0^* . As explained, the fibration (5) on X has integral even Mordell–Weil lattice $MWL(X) = MWL(X')(2) = \text{Hom}(E, E')(2)$, and

$$\text{NS}(X) = U + 2E_8(-1) + MWL(X)(-1).$$

For an Enriques involution τ on X , we ask that some section $P \in MWL(X)$ meets both fixed fibres at nonidentity components. Equivalently, there is a section $P' \in$

$\text{MWL}(X')$ (inducing P) that meets both ramified fibres (type I_0^*) at nonidentity components.

Assumption: There is no such section $P' \in \text{MWL}(X')$. Equivalently, since the simple components of a fibre admit a group structure, the nonidentity components of one of the I_0^* fibres are fully avoided by $\text{MW}(X')$. Correspondingly, $\text{NS}(X')$ admits an orthogonal summand $D_4(-1)$ which we single out in the following decomposition:

$$\text{NS}(X') = U + E_8(-1) + D_4(-1) + \langle D_4(-1), \text{MWL}(X')(-1) \rangle.$$

Hence the discriminant group of $\text{NS}(X')$ contains two copies of $\mathbb{Z}/2\mathbb{Z}$ (coming from D_4^\vee/D_4). Indeed, since the length is bounded by the rank of the transcendental lattice, which is two, this gives the full 2-part of the discriminant group:

$$2\text{-part}(\text{NS}(X')^\vee / \text{NS}(X')) \cong D_4^\vee / D_4 \cong (\mathbb{Z}/2\mathbb{Z})^2. \tag{17}$$

Right away, we deduce that $\text{NS}(X')$ has discriminant d' equalling four times an odd integer. By (3), this odd integer is exactly the discriminant $d = d'/4$ of X . In particular, if d is even, $\text{MWL}(X')$ cannot fully avoid the nonidentity components of either of the I_0^* fibres. Thus there is a section of the fibration (5) inducing an Enriques involution τ on X .

To complete the proof of Proposition 5.3, we return to the case of odd discriminant d . The isomorphism (17) gives an equality of discriminant forms

$$-q_{D_4} = q_{D_4} = (q_{\text{NS}(X')})|_{2\text{-part}}.$$

By [Nikulin 1980], there is an equality $q_{\text{NS}(X')} = -q_{T(X')}$. Hence it suffices to compare the discriminant forms of $T(X')$ and D_4 . In the present situation, $T(X')$ has the quadratic form

$$\begin{pmatrix} 4a & 2b \\ 2b & 4c \end{pmatrix}$$

with odd b . Hence its discriminant form takes the following values on a set of representatives of the 2-part of $T(X')^\vee/T(X')$:

$$0, a, c, a + b + c \pmod{2\mathbb{Z}}.$$

In comparison, q_{D_4} does exclusively attain the value $1 \pmod{2\mathbb{Z}}$ on the nonzero elements of D_4^\vee/D_4 . For $T(X')$, this can only happen if all a, b, c are odd. Equivalently, the discriminant satisfies $d \equiv -3 \pmod{8}$. This is exactly the main case excluded by Theorem 5.1.

Conversely, we deduce that a singular K3 surface X admits an Enriques involution if it has an elliptic fibration (5) of Mordell–Weil rank two and if either d is even or $d \equiv -7 \pmod{8}$. The latter can be achieved unless $T(X)$ is primitive and

corresponds to the principal class in its class group which is only two-torsion (cf. Exception 5.5). This completes the proof of Proposition 5.3.

5G. Appendix: More on Example 3.1. In this subsection, we will show that the singular K3 surface X from Example 3.1 (which falls under Exception 5.5) does admit an alternative elliptic fibration with an Enriques involution of base change type. We will pursue an abstract approach following ideas of Kneser and Witt as worked out for elliptic K3 surfaces by Nishiyama [1996].

Lemma 5.7. *X has an elliptic fibration with $\mathbb{Z}/3\mathbb{Z} \subset \text{MW}$ and two fibres of type I_9 .*

Proof. By [Nishiyama 1996, §6] the elliptic fibrations on X are classified by primitive embeddings of a certain partner lattice M of $T(X)$ into Niemeier lattices. Here we can take $M = A_1(-1) + A_5(-1)$ since M and $T(X)$ have the same discriminant form. Consider the Niemeier lattice N with root lattice

$$N_{\text{root}} = A_8(-1)^3 \quad \text{and quotient} \quad N/N_{\text{root}} = (\mathbb{Z}/3\mathbb{Z})^3.$$

Embedding M primitively into one summand $A_8(-1)$, we obtain the essential lattice of an elliptic fibration of X as orthogonal complement $M^\perp \subset N$. The singular fibres of this fibration are encoded in the roots of M^\perp , i.e., in $(M^\perp)_{\text{root}} = A_8(-1)^2$. The torsion in MW for this fibration is isomorphic to the quotient of the primitive closure of $(M^\perp)_{\text{root}}$ in N by $(M^\perp)_{\text{root}}$, i.e., $\text{MW}_{\text{tor}} \cong \mathbb{Z}/3\mathbb{Z}$. \square

The given elliptic fibration is not isotrivial due to the singular fibres of type I_9 . The torsion in MW then implies that X is a base change of the universal elliptic curve with 3-torsion section and j -invariant not identical zero. This elliptic surface has singular fibres I_1, I_3, IV^* , so necessarily the base change factors through the intermediate rational elliptic surface S' with configuration I_1, I_1, I_1, I_9 and $\text{MW}(S') = \mathbb{Z}/3\mathbb{Z}$. In particular, X arises from S' by a quadratic base change. Hence we are in the set-up of 5B with base change involution ι etc.

Now we consider the quadratic twist X' . It is the desingularisation of the quotient of X by the Nikulin involution $j = \iota \circ (-1)$. We claim that this quotient exhibits another Shioda–Inose structure on X :

Lemma 5.8. *X' is a Kummer surface with $T(X') = T(X)(2)$.*

Proof. It suffices to prove that j is a Morrison–Nikulin involution; i.e., j^* exchanges two copies of $E_8(-1)$ in $\text{NS}(X)$. Here we argue with the above elliptic fibration: j exchanges the two reducible fibres of type I_9 and the three-torsion sections $Q, \square Q$. Consider these 20 rational curves on X . Omitting the component of one I_9 fibre met by Q and the component of the other I_9 fibre met by $\square Q$, we find two disjoint configurations of type $\tilde{E}_8(-1)$ that are interchanged by j . The lemma now follows from [Morrison 1984, Theorem 5.7]. \square

The induced elliptic fibration on X' has singular fibres $I_1, I_1, I_1, I_9, I_0^*, I_0^*$. Since $\rho(X) = 20$, both X and X' have MW-rank two. In particular, there are plenty of t^* -anti-invariant sections on X (induced from X'). As in 5B, each such section gives an involution τ .

Lemma 5.9. *There is a fixed-point free involution τ on X as above.*

Proof. We verify the claim on X' by assuming the contrary. This means that for one of the I_0^* fibres all nonidentity components are avoided by $\text{MW}(X')$. As in 5F, this implies that X' has discriminant four times an odd integer. But we have seen that X' has $T(X') = T(X)(2)$ with discriminant -48 . This gives a contradiction. \square

Remark 5.10. This example also shows that not every singular Enriques surface arises by the canonical Shioda–Inose structure from 5D. This fact can also be seen in terms of Enriques surfaces with finite automorphism group. Kondō classified these exceptional Enriques surfaces in [Kondō 1986]. Some are singular, but do not admit an elliptic fibration with a II^* fibre.

5H. Brauer groups. In [Hulek and Schütt 2011], we also answered a question by Beauville about Brauer groups. Namely Beauville asked for explicit examples of complex Enriques surfaces Y where the Brauer group $\text{Br}(Y) \cong \mathbb{Z}/2\mathbb{Z}$ pulls back identically zero to the covering K3 surface X via the universal cover $\pi : X \rightarrow Y$. He also raised the question whether such an example exists over \mathbb{Q} .

In [Hulek and Schütt 2011, §5], we gave affirmative solutions for both questions. Our basic objects were the singular K3 surfaces X with

$$\text{NS}(X) = U + 2E_8(-1) + \langle -4M \rangle + \langle -2N \rangle \tag{18}$$

where $M, N \in \mathbb{N}$ and $N > 1$ is odd. The above decomposition corresponds to an elliptic fibration (5) on X with MW-rank two. As in 5B, the section P of height $4M$ induces an Enriques involution τ on X . Clearly the orthogonal section of height $2N$ gives an anti-invariant divisor for τ^* . By [Beauville 2009], this implies the vanishing of $\pi^* \text{Br}(Y)$.

Previously we determined one surface (for $M = 1, N = 3$) with a model of (5) and Enriques involution τ defined over \mathbb{Q} . Here we want to point out that for any other surface X as above, this can be achieved over the class field $H(-8MN)$ by Theorem 1.1.

6. Classification problems

We conclude this paper by formulating classification problems for singular Enriques surfaces. In addition to fields of definition, we also consider Galois actions on divisors. First we review the situation for singular K3 surfaces.

6A. Obstructions for singular K3 surfaces. Although singular K3 surfaces can often be descended from the ring class field $H(d)$ to some smaller number field, there are certain obstructions to this descent. In this section we shall discuss two of them. The first comes from the transcendental lattice. Since the Néron–Severi lattice of a general K3 surface is determined by intersection numbers, it is a geometric invariant; that is, conjugate surfaces have the same NS. Since $T(X)$ and $\text{NS}(X)$ are related as orthogonal complements in the K3 lattice Λ , they share the same discriminant form up to sign by [Nikulin 1980, Proposition 1.6.1]. In particular, this fixes the genus of $T(X)$ (sometimes also called the isogeny class).

Theorem 6.1 [Shimada 2009; Schütt 2007]. *Let X be a singular K3 surface X over some number field. The transcendental lattices of X and its Galois conjugates cover the full genus of $T(X)$.*

This result has an immediate consequence on the fields of definition:

Corollary 6.2. *Let X be a singular K3 surface X of discriminant d over a number field L . Let $K = \mathbb{Q}(\sqrt{-d})$ and \bar{L} the Galois closure of L over K . Denote by $\mathcal{G}(X)$ the genus of $T(X)$. Then*

$$\#\mathcal{G}(X) \mid \deg_K L.$$

In particular, one deduces that a singular K3 surface X can only be defined over \mathbb{Q} if the genus of $T(X)$ consists of a single class.

The second obstruction stems from the Galois action on the divisors. Namely, even if a singular K3 surface X admits a model over a smaller field than $H(d)$, the ring class field is preserved through the Galois action on $\text{NS}(X)$:

Theorem 6.3 [Schütt 2010]. *Let X be a singular K3 surface of discriminant d over some number field L . Assume that $\text{NS}(X)$ is generated by divisors defined over L . Then the extension $L(\sqrt{d})$ contains the ring class field $H(d)$.*

In other words, Theorem 2.4 is not far from being optimal: at best, there is a model with $\text{NS}(X)$ defined over a quadratic subfield of $H(d)$.

Theorem 6.3 provides a direct proof of the following natural generalisation from CM elliptic curves, from [Shafarevich 1996]: Fixing $n \in \mathbb{N}$, there are only finitely many singular K3 surfaces over all number fields of degree bounded by n (up to complex isomorphism). The problem of explicit classifications, however, is still wide open. Even in the simplest case, it is not clear yet how many singular K3 surfaces there are over \mathbb{Q} —only that there are many, cf. [Elkies and Schütt 2008b]. In contrast, the restrictive setting of Theorem 6.3 is much more accessible. For instance there are exactly 13 singular K3 surfaces up to $\bar{\mathbb{Q}}$ -isomorphism with NS defined over \mathbb{Q} . By [Schütt 2010, Theorem 1], they stand in bijective correspondence with the discriminants d of class number one.

We shall now discuss how these obstructions turn out for singular Enriques surfaces. Then we formulate analogous classification problems.

6B. Fields of definition of singular Enriques surfaces. We start by pointing out that Theorem 6.1 carries over to singular Enriques surfaces directly. This fact is due to the universal property that defines the covering K3 surface X of an Enriques surface Y . Explicitly, X can be defined universally as

$$X = \text{Spec}(\mathcal{O}_Y \oplus \mathcal{H}_Y).$$

As this construction respects the base field, the obstructions from Theorem 6.1 on the field of definition of a singular K3 surface X carry over to each singular Enriques surface that is covered by X . Recall that a K3 surface may admit (arbitrarily) finitely many distinct Enriques quotients by [Ohashi 2007, Theorem 0.1], while the universal cover associates a unique K3 surface to a given Enriques surface.

Corollary 6.4. *Let $n \in \mathbb{N}$. There are only finitely many singular Enriques surfaces over all number fields of degree at most n up to complex isomorphism.*

Problem 6.5. The following two questions concern singular Enriques surfaces up to $\bar{\mathbb{Q}}$ -isomorphism:

- (1) For $n \in \mathbb{N}$, find all singular Enriques surfaces over number fields L of degree at most n over \mathbb{Q} .
- (2) Specifically classify all singular Enriques surfaces over \mathbb{Q} .

6C. Galois action on divisors. Upon translating the obstructions for singular K3 surfaces from 6A to singular Enriques surfaces, we saw in 6B that Theorem 6.1 and its corollary carry over directly to the Enriques quotients. In contrast, Theorem 6.3 has to be weakened on the Enriques side. Generally speaking, this weakening is due to the fact that (part of) the Galois action can be accommodated by a sublattice of $\text{NS}(X)$ that is killed by the Enriques involution. In support of these ideas, we shall review an example from [Hulek and Schütt 2011] (which draws heavily from [Elkies and Schütt 2008a]).

Consider the following family \mathcal{X} of elliptic K3 surfaces

$$\mathcal{X} : y^2 = x^3 + t^2x^2 + t^3(t - a)^2x, \quad a \neq 0. \tag{19}$$

This elliptic fibration has reducible singular fibres of type III^* at 0 and ∞ and I_4 at $t = a$. The general member has Picard number $\rho(\mathcal{X}) = 19$ with

$$\text{MW}(\mathcal{X}) = \{O, (0, 0)\} \cong \mathbb{Z}/2\mathbb{Z}.$$

Note that \mathcal{X} is of base change type – apply the base change $s = (t - a)^2/t$ to the rational elliptic surface S with Weierstrass form

$$S := y^2 = x^3 + x^2 + sx.$$

As in 5B, the two-torsion section induces an Enriques involution τ (unless the other singular fibres degenerate, i.e., unless $a = -1/16$). Denote the family of Enriques quotients by \mathcal{Y} . We first study the Galois action on $\text{Num}(\mathcal{Y})$:

Lemma 6.6. *Let $Y_a \in \mathcal{Y}$ ($a \neq -1/16$). Then $\text{Num}(Y_a)$ is defined over $\mathbb{Q}(a)$.*

Proof. Since $\text{Num}(Y_a)$ is torsion-free, the Galois action on $\text{Num}(Y_a)$ coincides with that on the invariant part of $\text{NS}(X_a)$. In the present situation, the I_4 fibre of \mathcal{X} is split-multiplicative, i.e., all fibre components are defined over $\mathbb{Q}(a)$. The same holds trivially for the fibres of type III^* . Together with the sections O and $(0, 0)$, these rational curves generate $\text{NS}(X_a)^*$ up to finite index. As this holds regardless of the Picard number of X_a (being 19 or 20), the lemma follows. \square

Remark 6.7. It is crucial that the lemma holds for *all* members of the family \mathcal{Y} , including the singular ones. Compare the situation for singular K3 surfaces in the family \mathcal{X} where Theorem 6.3 will often enforce a Galois action on the additional generator of NS. For the specialisations over \mathbb{Q} with $\rho = 20$, see 6E.

6D. Néron–Severi group. We point out that in this specific setting, Lemma 6.6 gives a stronger statement than Corollary 2.5. The situation gets more complicated if we consider $\text{NS}(\mathcal{Y})$ with its two-torsion because this can admit a quadratic Galois action. In particular, we can only conjecture an analogue of Corollary 2.5 for $\text{NS}(Y)$ that is more precise than saying that $\text{NS}(Y)$ is defined over some quadratic extension of $H(d)$ (Conjecture 6.11).

The main problem here lies in similar subtleties as encountered in the context of cohomologically and numerically trivial involutions (see [Hulek and Schütt 2011, §4] and the references therein). Namely, to decide about $\text{NS}(Y)$ it is necessary to work out generators of the full group (see Remark 6.9). We work this out for the family \mathcal{Y} in detail:

Proposition 6.8. *If $Y_a \in \mathcal{Y}$ ($a \neq -1/16$), then $\text{NS}(Y_a)$ is defined over $\mathbb{Q}(a, \sqrt{-a})$.*

Proof. The next remark will indicate that it is not sufficient to argue with the elliptic fibration (19) on \mathcal{X} . Instead, we consider Inose’s fibration (5) for the given family. The following Weierstrass form was derived in [Hulek and Schütt 2011, §5.3]:

$$\mathcal{X}: y'^2 = x'^3 + (9a - 1)x'/9 + \left(27\left(u - \frac{a^3}{u}\right) + 81a + 2\right)/27.$$

There is a section P of height 4 (thus disjoint from the zero section) with x' -coordinate

$$P_{x'}(u) = (3u^4 + 12u^3a + 6u^2a^3 + 4u^2a^2 - 12ua^4 + 3a^6)/(12a^2u^2).$$

The section P is anti-invariant for the base change involution ι of the Shioda–Inose structure on \mathcal{X} :

$$\iota : (x', y', u) \mapsto (x', y', -a^3/u).$$

The base change involution composed with translation by P defines an Enriques involution τ' on \mathcal{X} by 5B. Denote the family of Enriques quotients by \mathcal{Y}' . By Kondō’s classification in [Kondō 1986], \mathcal{Y}' has finite automorphism group, and in particular τ and τ' are conjugate in $\text{Aut}(\mathcal{X})$ so that $\mathcal{Y} \cong \mathcal{Y}'$.

We continue by determining an explicit basis of $\text{NS}(\mathcal{Y}')$. The induced elliptic fibration on \mathcal{Y}' has a singular fibre of type II^* , a bisection R (the push-down of O and P) and two multiple smooth fibres $F_1 = 2G_1, F_2 = 2G_2$. We claim that these twelve curves generate $\text{NS}(\mathcal{Y}')$. To see this, note that by construction R meets the simple component of the II^* fibre twice. The remaining fibre components form the root lattice of type $E_8(-1)$. Orthogonally in $\text{NS}(\mathcal{Y}')$, we find R, G_1, G_2 . Since $R^2 = -2, R \cdot G_i = 1$, we know that R, G_1 generate the hyperbolic plane U . Thus we have determined a unimodular lattice $L = U + E_8(-1)$ inside $\text{NS}(\mathcal{Y}')$ – necessarily of index two due to its rank being ten. Since $G_2 \notin L$, it follows that L and G_2 generate all of $\text{NS}(\mathcal{Y}')$.

We now consider the Galois action on these generators of $\text{NS}(Y'_a)$ for some $Y'_a \in \mathcal{Y}'$. Clearly the II^* fibre and the bisection R are defined over $\mathbb{Q}(a)$. The multiple fibres sit at the ramification points of the base change on the base curve \mathbb{P}^1 , i.e., at the roots of $u^2 + a^3$. Proposition 6.8 follows and cannot be improved since the conjugation of $\mathbb{Q}(\sqrt{-a})/\mathbb{Q}(a)$ permutes the multiple fibres if $\sqrt{-a} \notin \mathbb{Q}(a)$, and thus gives a nontrivial Galois action on $\text{NS}(Y'_a)$. □

Remark 6.9. Note that the above Galois action is not visible on the elliptic fibration (19) of \mathcal{X} yielding \mathcal{Y} . The multiple fibres of the induced elliptic fibration on \mathcal{Y} have different type $2I_0, 2I_2$. Hence they cannot be interchanged by Galois. Nonetheless there can be a nontrivial Galois action on $\text{NS}(Y_a)$. This goes undetected in the above model because the push-down of fibre components and torsion sections from \mathcal{X} to \mathcal{Y} generate $\text{NS}(\mathcal{Y})$ only up to index two.

6E. CM-points. Concretely, the family \mathcal{X} is parametrised by the Fricke modular curve $X_0(2)^+$. In [Elkies and Schütt 2008a], we list all \mathbb{Q} -rational CM-points. Two of them give singular K3 surfaces without Enriques involution (discriminant -8 at $a = -1/16$ and discriminant -4 at $a = 0$ for a suitable alternative model of \mathcal{X}).

The other 14 discriminants are:

$$-7, -12, -16, -20, -24, -28, -36, -40, -52, -72, -88, -100, -148, -232.$$

For the discriminants of class number two, the additional section can only be defined over a quadratic extension of \mathbb{Q} by Theorem 6.3. So there are indeed singular Enriques surfaces with Num defined over \mathbb{Q} where the same does not hold for the covering K3 surfaces. A detailed example where this holds even for NS is provided by the surfaces at $a = -1/144$ which corresponds to the discriminant -24 (as mentioned in 5H). Details can be found in [Hulek and Schütt 2011, §5.3]. We work out one example from the list where Num is defined over \mathbb{Q} , but NS is neither defined over \mathbb{Q} nor over $H(d)$:

Example 6.10. The specialisation X with discriminant $d = -12$ sits at $a = 1/9$. In terms of the elliptic fibration (19), there is a section of height 3 over $H(d) = \mathbb{Q}(\sqrt{-3})$ with x -coordinate $-12t^3/(9t - 1)^2$. One finds that X has transcendental lattice two-divisible, so X is the Kummer surface of $E \times E$ for E with j -invariant zero. In particular X is different from the singular K3 surface studied in Example 3.1 and 5G.

The Enriques quotient Y has multiple fibres at $\pm\sqrt{-1}/27$. Compared with Num(Y) which is defined over \mathbb{Q} , complex conjugation acts on NS(Y) as nontrivial Galois action. Note that $H(d)(\sqrt{-1}) = H(4d)$ in the present situation.

6F. In the above example (and in fact for all specialisations over \mathbb{Q} with $\rho = 20$), we have seen that NS(Y) is defined over the ring class field $H(4d)$. We conjecture that this is always the case which would give an analogue of Corollary 2.5:

Conjecture 6.11. Let Y be an Enriques surface whose universal cover X is a singular K3 surface. Let $d < 0$ denote the discriminant of X . Then Y admits a model over the ring class field $H(d)$ with NS(Y) defined over $H(4d)$.

The above one-dimensional family provides small evidence for this conjecture. Our main motivation stems from the base change construction of Enriques involutions in the framework of Shioda–Inose structures as investigated in Section 5. By Proposition 5.3, almost every possible singular K3 surface admits such an Enriques involution. In terms of the model (14), the Enriques quotient Y attains multiple fibres at the ramification points of the underlying base change, i.e., at $\pm 2B$. Recall from (5) that $B^2 = (1 - j/12^3)(1 - j'/12^3)$, so there is a quadratic Galois action on NS(Y) unless $B \in H(d)$. Note that B can be interpreted in terms of the Weber function $\sqrt{j - 12^3}$ where j now denotes the usual modular function. The values of Weber functions at CM-point have been studied extensively starting from Weber. In the present situation, Schertz [1976] proved that for singular j -values, $\sqrt{j - 12^3} \in H(4d)$. This implies:

Lemma 6.12 (Schertz). *In the above setting, one has $B \in H(4d)$.*

We sketch an alternative proof of Lemma 6.12. It is based on a geometric approach that will also carry information about the Enriques surface Y (and its elliptic fibration with fibre of type II^*). Consider the Kummer surface X' from the Shioda–Inose structure. In general, it has fibres of type I_0^* where Y has the multiple fibres (if $E \cong E'$, there could be fibres of type I_1^* or IV^* ; see 3B). By Corollary 4.2, X' has a model with $\text{NS}(X')$ defined over $H(4d)$. In particular, every elliptic fibration of X' can be defined over $H(4d)$ with all of NS defined there as well. We apply this argument to the elliptic fibration on X' induced from (14):

$$X' : y^2 = x^3 - 3c^2 B^2 A^3 (t^2 - 4B^2)^2 x + c^3 B^2 A^3 (t - 2B^2)(t^2 - 4B^2)^3. \quad (20)$$

Assume that $B \notin H(d)$ and denote $L = H(d)(B)$. By [Shioda 2006], the singular fibres of X' predict the Weierstrass form (20) (in case $AB \neq 0$) up to Möbius transformation. This property holds generally for constants A, B, c , but in the present situation, A and B are related to the j -invariants of E, E' by (5). Upon applying Möbius transformations, one can thus show that the above jacobian elliptic fibration does not admit a model over $H(d)$ without $\text{Gal}(L/H(d))$ -action interchanging the I_0^* fibres. By Corollary 4.2, one obtains that $B \in H(4d)$. This proves Lemma 6.12.

Corollary 6.13. *Conjecture 6.11 holds true for any singular Enriques surface arising from the Shioda–Inose structure as in Section 5.*

The geometric proof of Lemma 6.12 is of particular interest to us, since the statement about the Galois action on the I_0^* fibres of X' carries over to the multiple fibres of the corresponding elliptic fibration of the Enriques surface Y and vice versa. Centrally, we use once again that a model of a K3 or Enriques surface with NS defined over a fixed field has *all* elliptic fibrations (with or without section) defined over this field as well. Hence we can move freely between models and elliptic fibrations. Thus we obtain:

Corollary 6.14. *If $B \notin H(d)$, then any model over $H(d)$ of the Enriques surface Y admits a nontrivial Galois action of $\text{Gal}(H(4d)/H(d))$ on $\text{NS}(Y)$.*

We have seen an instance of this phenomenon in Example 6.10. The same reasoning implies a nontrivial action of $\text{Gal}(\mathbb{Q}(a, \sqrt{-a})/\mathbb{Q}(a))$ on $\text{NS}(Y_a)$ for all $\mathbb{Q}(a)$ -models of members Y_a of the family \mathcal{Y} .

The above results allow us to draw an analogy to the study of automorphisms of Enriques surfaces; cf. [Barth and Peters 1983; Mukai and Namikawa 1984]. Namely we have exhibited two kind of singular Enriques surfaces over $H(d)$ —one with cohomologically trivial Galois action and one with numerically, but not cohomologically trivial Galois action.

6G. We conclude this paper with the corresponding classification problem for singular Enriques surfaces. Note that by the above reasoning, at least the second problem is more complicated than for K3 surfaces (as solved in [Schütt 2010]).

Problem 6.15. The following two questions concern singular Enriques surfaces either up to $\bar{\mathbb{Q}}$ - or up to L -isomorphism:

- (1) For a given number field L (or all number fields of bounded degree), classify all singular Enriques surfaces with Num or NS defined over L .
- (2) Determine all singular Enriques surfaces over $L = \mathbb{Q}$ with trivial Galois action on Num or NS.

Acknowledgements

We thank Bas Edixhoven and Jaap Top for useful comments. We are grateful to the referees for many helpful suggestions and remarks. This project was started when the second author held a position at University of Copenhagen.

References

- [Barth and Peters 1983] W. Barth and C. Peters, “Automorphisms of Enriques surfaces”, *Invent. Math.* **73**:3 (1983), 383–411. MR 85g:14052 Zbl 0518.14023
- [Barth et al. 2004] W. P. Barth, K. Hulek, C. A. M. Peters, and A. Van de Ven, *Compact complex surfaces*, 2nd ed., *Ergebnisse der Mathematik (3)* **4**, Springer, Berlin, 2004. MR 2004m:14070 Zbl 1036.14016
- [Beauville 2009] A. Beauville, “On the Brauer group of Enriques surfaces”, *Math. Res. Lett.* **16**:6 (2009), 927–934. MR 2011b:14079 Zbl 1195.14053
- [Bogomolov and Tschinkel 1998] F. A. Bogomolov and Y. Tschinkel, “Density of rational points on Enriques surfaces”, *Math. Res. Lett.* **5**:5 (1998), 623–628. MR 99m:14040 Zbl 0957.14016
- [Cox 1989] D. A. Cox, *Primes of the form $x^2 + ny^2$: Fermat, class field theory and complex multiplication*, Wiley, New York, 1989. MR 90m:11016 Zbl 0701.11001
- [Elkies and Schütt 2008a] N. D. Elkies and M. Schütt, “K3 families of high Picard rank”, (2008).
- [Elkies and Schütt 2008b] N. D. Elkies and M. Schütt, “Modular forms and K3 surfaces”, 2008. arXiv math/0809.0830v2
- [Gross 1980] B. H. Gross, *Arithmetic on elliptic curves with complex multiplication*, *Lecture Notes in Mathematics* **776**, Springer, Berlin, 1980. MR 81f:10041 Zbl 0433.14032
- [Hulek and Schütt 2011] K. Hulek and M. Schütt, “Enriques surfaces and Jacobian elliptic K3 surfaces”, *Math. Z.* **268**:3-4 (2011), 1025–1056. MR 2818742 Zbl 1226.14052
- [Inose 1978] H. Inose, “Defining equations of singular K3 surfaces and a notion of isogeny”, pp. 495–502 in *Proceedings of the International Symposium on Algebraic Geometry (Kyoto, 1977)*, edited by M. Nagata, Kinokuniya, Tokyo, 1978. MR 81h:14021 Zbl 0411.14009
- [Keum 1990] J. H. Keum, “Every algebraic Kummer surface is the K3-cover of an Enriques surface”, *Nagoya Math. J.* **118** (1990), 99–110. MR 91f:14036 Zbl 0699.14047
- [Kondō 1986] S. Kondō, “Enriques surfaces with finite automorphism groups”, *Japan. J. Math. (N.S.)* **12**:2 (1986), 191–282. MR 89c:14058 Zbl 0616.14031

- [Morrison 1984] D. R. Morrison, “On K3 surfaces with large Picard number”, *Invent. Math.* **75**:1 (1984), 105–121. MR 85j:14071 Zbl 0509.14034
- [Mukai and Namikawa 1984] S. Mukai and Y. Namikawa, “Automorphisms of Enriques surfaces which act trivially on the cohomology groups”, *Invent. Math.* **77**:3 (1984), 383–397. MR 86i:14012 Zbl 0559.14038
- [Nikulin 1980] V. Nikulin, “Integral symmetric bilinear forms and some of their applications”, *Math. USSR, Izv.* **14** (1980), 103–167. Zbl 427.10014
- [Nishiyama 1996] K.-i. Nishiyama, “The Jacobian fibrations on some K3 surfaces and their Mordell–Weil groups”, *Japan. J. Math. (N.S.)* **22**:2 (1996), 293–347. MR 97m:14037 Zbl 0889.14015
- [Ohashi 2007] H. Ohashi, “On the number of Enriques quotients of a K3 surface”, *Publ. Res. Inst. Math. Sci.* **43**:1 (2007), 181–200. MR 2008b:14062 Zbl 1133.14038
- [Piatetski-Shapiro and Shafarevich 1971] I. I. Piatetski-Shapiro and I. R. Shafarevich, “Torelli’s theorem for algebraic surfaces of type K3”, *Izv. Akad. Nauk SSSR Ser. Mat.* **35** (1971), 530–572. In Russian; translated in *Math. USSR Izv.* **5**:3 (1971), 547–588. MR 44 #1666 Zbl 0219.14021
- [Schertz 1976] R. Schertz, “Die singulären Werte der Weberschen Funktionen $f, f_1, f_2, \gamma_2, \gamma_3$ ”, *J. Reine Angew. Math.* **286/287** (1976), 46–74. MR 54 #10205 Zbl 0335.12018
- [Schütt 2007] M. Schütt, “Fields of definition of singular K3 surfaces”, *Commun. Number Theory Phys.* **1**:2 (2007), 307–321. MR 2008g:14060 Zbl 1157.14308
- [Schütt 2008] M. Schütt, “Arithmetic of a singular K3 surface”, *Michigan Math. J.* **56**:3 (2008), 513–527. MR 2009k:11106 Zbl 1163.14022
- [Schütt 2010] M. Schütt, “K3 surfaces with Picard rank 20”, *Algebra Number Theory* **4**:3 (2010), 335–356. MR 2011c:14113 Zbl 1190.14034
- [Schütt and Shioda 2010] M. Schütt and T. Shioda, “Elliptic surfaces”, pp. 51–160 in *Algebraic geometry in East Asia* (Seoul, 2008), edited by J. Keum et al., Adv. Stud. Pure Math. **60**, Math. Soc. Japan, Tokyo, 2010. MR 2012b:14069 Zbl 1216.14036
- [Sertöz 2005] A. S. Sertöz, “Which singular K3 surfaces cover an Enriques surface”, *Proc. Amer. Math. Soc.* **133**:1 (2005), 43–50. MR 2005m:14066 Zbl 1049.14032
- [Shafarevich 1996] I. R. Shafarevich, “On the arithmetic of singular K3-surfaces”, pp. 103–108 in *Algebra and analysis* (Kazan, 1994), edited by M. M. Arslanov et al., de Gruyter, Berlin, 1996. MR 98h:14041 Zbl 0947.14020
- [Shimada 2009] I. Shimada, “Transcendental lattices and supersingular reduction lattices of a singular K3 surface”, *Trans. Amer. Math. Soc.* **361** (2009), 909–949. MR 2009m:14055 Zbl 1187.14048
- [Shimada and Zhang 2001] I. Shimada and D.-Q. Zhang, “Classification of extremal elliptic K3 surfaces and fundamental groups of open K3 surfaces”, *Nagoya Math. J.* **161** (2001), 23–54. MR 2002d:14056 Zbl 1064.14503
- [Shimura 1971] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan **11**, Iwanami Shoten, Tokyo, 1971. MR 47 #3318 Zbl 0221.10029
- [Shioda 1990] T. Shioda, “On the Mordell–Weil lattices”, *Comment. Math. Univ. St. Paul.* **39**:2 (1990), 211–240. MR 91m:14056 Zbl 0725.14017
- [Shioda 2006] T. Shioda, “Kummer sandwich theorem of certain elliptic K3 surfaces”, *Proc. Japan Acad. Ser. A Math. Sci.* **82**:8 (2006), 137–140. MR 2008b:14064 Zbl 1112.14044
- [Shioda 2007] T. Shioda, “Correspondence of elliptic curves and Mordell–Weil lattices of certain elliptic K3’s”, pp. 319–339 in *Algebraic cycles and motives*, vol. 2, edited by J. Nagel and C. Peters, London Math. Soc. Lecture Note Ser. **344**, Cambridge Univ. Press, 2007. MR 2009a:14052 Zbl 1136.14028

- [Shioda and Inose 1977] T. Shioda and H. Inose, “On singular K3 surfaces”, pp. 119–136 in *Complex analysis and algebraic geometry*, edited by W. L. Baily, Jr. and T. Shioda, Iwanami Shoten, Tokyo, 1977. MR 56 #371 Zbl 0374.14006
- [Shioda and Mitani 1974] T. Shioda and N. Mitani, “Singular abelian surfaces and binary quadratic forms”, pp. 259–287 in *Classification of algebraic varieties and compact complex manifolds*, edited by H. Popp, Lecture Notes in Mathematics **412**, Springer, Berlin, 1974. MR 52 #3174 Zbl 0302.14011
- [Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015
- [Sterk 1985] H. Sterk, “Finiteness results for algebraic K3 surfaces”, *Math. Z.* **189**:4 (1985), 507–513. MR 86j:14038 Zbl 0545.14032
- [Tate 1975] J. Tate, “Algorithm for determining the type of a singular fiber in an elliptic pencil”, pp. 33–52 in *Modular functions of one variable, IV* (Antwerp, 1972), edited by B. J. Birch and W. Kuyk, Lecture Notes in Math. **476**, Springer, Berlin, 1975. MR 52 #13850 Zbl 1214.14020
- [Weinberger 1973] P. J. Weinberger, “Exponents of the class groups of complex quadratic fields”, *Acta Arith.* **22** (1973), 117–124. MR 47 #1776 Zbl 0217.04202

Communicated by János Kollár

Received 2010-03-13 Revised 2010-11-28 Accepted 2010-12-29

hulek@math.uni-hannover.de *Institut für Algebraische Geometrie, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany*

schuett@math.uni-hannover.de *Institut für Algebraische Geometrie, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany*

An upper bound on the Abbes–Saito filtration for finite flat group schemes and applications

Yichao Tian

Let \mathbb{O}_K be a complete discrete valuation ring of residue characteristic $p > 0$, and G be a finite flat group scheme over \mathbb{O}_K of order a power of p . We prove in this paper that the Abbes–Saito filtration of G is bounded by a linear function of the degree of G . Assume \mathbb{O}_K has generic characteristic 0 and the residue field of \mathbb{O}_K is perfect. Fargues constructed the higher level canonical subgroups for a “near from being ordinary” Barsotti–Tate group \mathcal{G} over \mathbb{O}_K . As an application of our bound, we prove that the canonical subgroup of \mathcal{G} of level $n \geq 2$ constructed by Fargues appears in the Abbes–Saito filtration of the p^n -torsion subgroup of \mathcal{G} .

Let \mathbb{O}_K be a complete discrete valuation ring with residue field k of characteristic $p > 0$ and fraction field K . We denote by v_π the valuation on K normalized by $v_\pi(K^\times) = \mathbb{Z}$. Let G be a finite and flat group scheme over \mathbb{O}_K of order a power of p such that $G \otimes K$ is étale. We denote by $(G^a, a \in \mathbb{Q}_{\geq 0})$ the Abbes–Saito filtration of G . This is a decreasing and separated filtration of G by finite and flat closed subgroup schemes. We refer the readers to [Abbes and Saito 2002; 2003; Abbes and Mokrane 2004] for a full discussion, and to Section 1 for a brief review of this filtration. Let ω_G be the module of invariant differentials of G . The generic étaleness of G implies that ω_G is a torsion \mathbb{O}_K -module of finite type. Thus, there exist nonzero elements $a_1, \dots, a_d \in \mathbb{O}_K$ such that

$$\omega_G \simeq \bigoplus_{i=1}^d \mathbb{O}_K / (a_i).$$

We put $\deg(G) = \sum_{i=1}^d v_\pi(a_i)$, and call it the degree of G . The aim of this note is to prove the following:

Theorem 1. *Let G be a finite and flat group scheme over \mathbb{O}_K of order a power of p such that $G \otimes K$ is étale. Then we have $G^a = 0$ for $a > p/(p-1) \deg(G)$.*

This research was supported by a grant DMS-0635607 from the National Science Foundation.

MSC2000: primary 14L15; secondary 14G22, 11S15.

Keywords: finite flat group schemes, ramification filtration, canonical subgroups.

Our bound is optimal when G is killed by p . Let $E_\delta = \text{Spec}(\mathbb{O}_K[X]/(X^p - \delta X))$ be the group scheme of Tate–Oort over \mathbb{O}_K . We have $\deg(E_\delta) = v_\pi(\delta)$, and an easy computation by Newton polygons gives [Fargues 2009, Lemme 5]:

$$E_\delta^a = \begin{cases} E_\delta & \text{if } 0 \leq a \leq p/(p-1) \deg(E_\delta), \\ 0 & \text{if } a > p/(p-1) \deg(E_\delta). \end{cases}$$

However, our bound may be improved when G is not killed by p or G contains many identical copies of a closed subgroup. In [2006, Theorem 7], Hattori proves that if K has characteristic 0 and G is killed by p^n , then the Abbes–Saito filtration of G is bounded by that of the multiplicative group μ_{p^n} , i.e., we have $G^a = 0$ if $a > en + e/(p-1)$ where e is the absolute ramification index of K . Compared with Hattori’s result, our bound has the advantage that it works in both characteristic 0 and characteristic p , and that it is good if $\deg(G)$ is small.

The basic idea used to prove Theorem 1 is approximation of general power series over \mathbb{O}_K by linear functions. First, we choose a “good” presentation of the algebra of G such that the defining equations of G involve only terms of total degree $m(p-1) + 1$ with $m \in \mathbb{Z}_{\geq 0}$; see Proposition 1.6. The existence of such a presentation is a consequence of the classical theory on p -typical curves of formal groups. With this good presentation, we can prove in Lemma 1.9 that the neutral connected component of the a -tubular neighborhood of G is isomorphic to a closed rigid ball for $a > p/(p-1) \deg(G)$, and the only zero of the defining equations of G in the neutral component is the unit section.

The motivation of our theorem comes from the theory of canonical subgroups. We assume that K has characteristic 0, and the residue field k is perfect of characteristic $p \geq 3$. Let G be a Barsotti–Tate group of dimension $d \geq 1$ over \mathbb{O}_K . Abbes and Mokrane [2004] were the first to construct the canonical subgroup of level 1 of G in the case where G comes from an abelian scheme over \mathbb{O}_K . Then, Tian [2010] generalized their result to the Barsotti–Tate case. More specifically, it was shown that if a Barsotti–Tate group G over \mathbb{O}_K is “near from being ordinary”, a condition expressed explicitly as a bound on the Hodge height of G (see Section 2.1), then a certain piece of the Abbes–Saito filtration of $G[p]$ lifts the kernel of Frobenius of the special fiber of G [Tian 2010, Theorem 1.4]. Later on, Fargues [2009] gave another construction of the canonical subgroup of level 1 using Hodge–Tate maps, and his approach also allowed us to construct by induction the canonical subgroups of level $n \geq 2$, i.e., the canonical lifts of the kernel of the n -th iteration of the Frobenius. He proved that the canonical subgroup of higher level appears in the Harder–Narasimhan filtration of $G[p^n]$, which was introduced by him in [Fargues 2007]. It is conjectured that the canonical subgroup of higher level also appears in the Abbes–Saito filtration of $G[p^n]$. In this paper, we prove this conjecture as a corollary, Theorem 2.5, of Theorem 1. Fargues’s result on the degree of the

quotient of $G[p^n]$ by its canonical subgroup of level n (see Theorem 2.4(i)) will play an essential role in our proof.

Notation. In this paper, \mathbb{O}_K will denote a complete discrete valuation ring with residue field k of characteristic $p > 0$ and fraction field K . Let π be a uniformizer of \mathbb{O}_K , and v_π be the valuation on K normalized by $v_\pi(\pi) = 1$. Let \bar{K} be an algebraic closure of K , K^{sep} be the separable closure of K contained in \bar{K} , and \mathcal{G}_K be the Galois group $\text{Gal}(K^{\text{sep}}/K)$. We also denote by v_π the unique extension of the valuation to \bar{K} .

1. Proof of Theorem 1

First, we recall the definition of the filtration of Abbes–Saito for finite flat group schemes according to [Abbes and Mokrane 2004; Abbes and Saito 2003].

1.1. We denote the Jacobson radical of a semilocal ring R by \mathfrak{m}_R . An algebra R over \mathbb{O}_K is called *formally of finite type* if R is semilocal, complete with respect to the \mathfrak{m}_R -adic topology, Noetherian, and R/\mathfrak{m}_R is finite over k . We say an \mathbb{O}_K -algebra R formally of finite type is *formally smooth* if each of the factors of R is formally smooth over \mathbb{O}_K .

Let $\mathbf{FEA}_{\mathbb{O}_K}$ be the category of finite, flat, and generically étale \mathbb{O}_K -algebras, and $\mathbf{Set}_{\mathcal{G}_K}$ be the category of finite sets endowed with a discrete action of the Galois group \mathcal{G}_K . We have the fiber functor

$$\mathcal{F} : \mathbf{FEA}_{\mathbb{O}_K} \rightarrow \mathbf{Set}_{\mathcal{G}_K},$$

which associates to an object A of $\mathbf{FEA}_{\mathbb{O}_K}$ the set $\text{Spec}(A)(\bar{K})$ equipped with the natural action of \mathcal{G}_K . We define a filtration on the functor \mathcal{F} as follows. For each object A in $\mathbf{FEA}_{\mathbb{O}_K}$, we choose a presentation

$$0 \rightarrow I \rightarrow \mathcal{A} \rightarrow A \rightarrow 0, \tag{1}$$

where \mathcal{A} is an \mathbb{O}_K -algebra formally of finite type and formally smooth. For any $a = m/n \in \mathbb{Q}_{>0}$ with m prime to n , we define \mathcal{A}^a to be the π -adic completion of the subring $\mathcal{A}[I^n/\pi^m] \subset \mathcal{A} \otimes_{\mathbb{O}_K} K$ generated over \mathcal{A} by all the f/π^m with $f \in I^n$. The \mathbb{O}_K -algebra \mathcal{A}^a is topologically of finite type, and the tensor product $\mathcal{A}^a \otimes_{\mathbb{O}_K} K$ is an affinoid algebra over K [Abbes and Saito 2003, Lemma 1.4]. We put $X^a = \text{Sp}(\mathcal{A}^a \otimes_{\mathbb{O}_K} K)$, which is a smooth affinoid variety over K [Abbes and Saito 2003, Lemma 1.7]. We call it the *a-th tubular neighborhood of $\text{Spec}(A)$ with respect to the presentation (1)*. The \mathcal{G}_K -set of the geometric connected components of X^a , denoted by $\pi_0(X^a(A)_{\bar{K}})$, depends only on the \mathbb{O}_K -algebra A and the rational number a , but not on the choice of the presentation [Abbes and Saito

2003, Lemma 1.9.2]. For rational numbers $b > a > 0$, we have natural inclusions of affinoid varieties $\mathrm{Sp}(A \otimes_{\mathbb{O}_K} K) \hookrightarrow X^b \hookrightarrow X^a$, which induce natural morphisms $\mathrm{Spec}(A)(\bar{K}) \rightarrow \pi_0(X^b(A)_{\bar{K}}) \rightarrow \pi_0(X^a(A)_{\bar{K}})$. For a morphism $A \rightarrow B$ in $\mathbf{FEA}_{\mathbb{O}_K}$, we can choose presentations of A and B so that we have a functorial map $\pi_0(X^a(B)_{\bar{K}}) \rightarrow \pi_0(X^a(A)_{\bar{K}})$. Hence we get, for any $a \in \mathbb{Q}_{>0}$, a (contravariant) functor

$$\mathcal{F}^a : \mathbf{FEA}_{\mathbb{O}_K} \rightarrow \mathbf{Set}^{\mathcal{G}_K}$$

given by $A \mapsto \pi_0(X^a(A)_{\bar{K}})$. We have natural morphisms of functors $\phi_a : \mathcal{F} \rightarrow \mathcal{F}^a$ and $\phi_{a,b} : \mathcal{F}^b \rightarrow \mathcal{F}^a$ for rational numbers $b > a > 0$ with $\phi_a = \phi_{b,a} \circ \phi_b$. For any A in $\mathbf{FEA}_{\mathbb{O}_K}$, we have

$$\mathcal{F}(A) \xrightarrow{\sim} \varprojlim_{a \in \mathbb{Q}_{>0}} \mathcal{F}^a(A)$$

[Abbes and Saito 2002, 6.4]; if A is a complete intersection over \mathbb{O}_K , the map $\mathcal{F}(A) \rightarrow \mathcal{F}^a(A)$ is surjective for any a [Abbes and Saito 2002, 6.2].

1.2. Let $G = \mathrm{Spec}(A)$ be a finite and flat group scheme over \mathbb{O}_K such that $G \otimes K$ is étale over K , and $a \in \mathbb{Q}_{>0}$. The group structure of G induces a group structure on $\mathcal{F}^a(A)$, and the natural map $G(\bar{K}) = \mathcal{F}(A) \rightarrow \mathcal{F}^a(A)$ is a homomorphism of groups. Hence, the kernel $G^a(\bar{K})$ of $G(\bar{K}) \rightarrow \mathcal{F}^a(A)$ is a \mathcal{G}_K -invariant subgroup of $G(\bar{K})$, and it defines a closed subgroup scheme G_K^a of the generic fiber $G \otimes K$. The scheme theoretic closure of G_K^a in G , denoted by G^a , is a closed subgroup of G finite and flat over \mathbb{O}_K . Putting $G^0 = G$, we get a decreasing and separated filtration $(G^a, a \in \mathbb{Q}_{\geq 0})$ of G by finite and flat closed subgroup schemes. We call it the *Abbes–Saito filtration* of G . For any real number $a \geq 0$, we put $G^{a+} = \bigcup_{b \in \mathbb{Q}_{>a}} G^a$.

Assume G is connected, i.e., the ring A is local. Let

$$0 \rightarrow I \rightarrow \mathbb{O}_K[[X_1, \dots, X_d]] \rightarrow A \rightarrow 0 \tag{2}$$

be a presentation of A by the ring of formal power series such that the unit section of G corresponds to the point $(X_1, \dots, X_d) = (0, \dots, 0)$. Since A is a relative complete intersection over \mathbb{O}_K , I is generated by d elements f_1, \dots, f_d . For $a \in \mathbb{Q}_{>0}$, the \bar{K} -valued points of the a -th tubular neighborhood of G are given by

$$X^a(\bar{K}) = \{(x_1, \dots, x_d) \in \mathfrak{m}_{\bar{K}}^d \mid v_{\pi}(f_i(x_1, \dots, x_d)) \geq a \text{ for } 1 \leq i \leq d\}, \tag{3}$$

where $\mathfrak{m}_{\bar{K}}$ is the maximal ideal of $\mathbb{O}_{\bar{K}}$. The subset $G(\bar{K}) \subset X^a(\bar{K})$ corresponds to the zeros of the f_i 's. Let X_0^a be the connected component of X^a containing 0. Then the subgroup $G^a(\bar{K})$ is the intersection of $X_0^a(\bar{K})$ with $G(\bar{K})$.

The basic properties of Abbes–Saito filtration that we need are summarized as follows.

Proposition 1.3 [Abbes and Mokrane 2004, 2.3.2, 2.3.5]. *Let G and H be finite and flat group schemes, generically étale over \mathbb{O}_K , and $f : G \rightarrow H$ be a homomorphism of group schemes.*

- (i) *The closed subgroup G^{0+} is the connected component of G , and we have $(G^{0+})^a = G^a$ for any $a \in \mathbb{Q}_{>0}$.*
- (ii) *Given $a \in \mathbb{Q}_{>0}$, f induces a canonical homomorphism $f^a : G^a \rightarrow H^a$. If f is flat and surjective, then $f^a(\bar{K}) : G^a(\bar{K}) \rightarrow H^a(\bar{K})$ is surjective.*

Now we return to the proof of Theorem 1.

Lemma 1.4. *Let R be a \mathbb{Z}_p -algebra, \mathcal{X} be a formal group of dimension d over R such that $\text{Lie}(\mathcal{X})$ is a free R -module of rank d . Then*

- (i) *the ring \mathbb{Z}_p acts naturally on \mathcal{X} , and its image in $\text{End}_R(\mathcal{X})$ lies in the center of $\text{End}_R(\mathcal{X})$;*
- (ii) *there exist parameters (X_1, \dots, X_d) of \mathcal{X} such that*

$$[\zeta](X_1, \dots, X_d) = (\zeta X_1, \dots, \zeta X_d)$$

for any $(p-1)$ -st root of unity $\zeta \in \mathbb{Z}_p$.

Proof. This is actually a classical result on formal groups. In the terminology of [Hazewinkel 1978], the formal group \mathcal{X} comes from the base change of $\mathcal{X}^{\text{univ}}$ defined by the d -dimensional universal p -typical formal group law (denoted by $F_V(X, Y)$ in [Hazewinkel 1978, 15.2.8]) over

$$\mathbb{Z}_p[V] = \mathbb{Z}_p[V_i(j, k); i \in \mathbb{Z}_{\geq 0}, j, k = 1, \dots, d],$$

where the $V_i(j, k)$ are free variables. So we are reduced to proving the lemma for $\mathcal{X}^{\text{univ}}$. If X and Y stand for the column vectors (X_1, \dots, X_d) and (Y_1, \dots, Y_d) respectively, the formal group law on $\mathcal{X}^{\text{univ}}$ is determined by

$$F_V(X, Y) = f_V^{-1}(f_V(X) + f_V(Y)), \quad \text{with } f_V(X) = \sum_{i=0}^{\infty} a_i(V) X^{p^i},$$

where the $a_i(V)$ are certain $d \times d$ matrices with coefficients in $\mathbb{Q}_p[V]$ with $a_1(V)$ invertible, X^{p^i} stands for $(X_1^{p^i}, \dots, X_d^{p^i})$, and f_V^{-1} is the unique d -tuple of power series in (X_1, \dots, X_d) with coefficients in $\mathbb{Q}_p[V]$ such that $f_V^{-1} \circ f_V = 1$; see [Hazewinkel 1978, 10.4]. We note that $F_V(X, Y)$ is a d -tuple of power series with coefficient in $\mathbb{Z}_p[V]$, although $f_V(X)$ has coefficients in $\mathbb{Q}_p[V]$ [Hazewinkel 1978, 10.2(i)]. Via approximation by integers, we see easily that the operation of multiplication by an element $\xi \in \mathbb{Z}_p$ given by $[\xi](X) = f_V^{-1}(\xi f_V(X))$ is well defined. This proves (i). Statement (ii) is an immediate consequence of the fact that $f_V(X)$ contains only p -powers of X . □

Remark 1.5. The referee gives the following alternative proof of this lemma via the Cartier theory of formal groups. Let \mathcal{X} be the formal group over R as in the lemma. We denote by $\mathcal{X}(R[[T]])$ the group of $R[[T]]$ -valued points of \mathcal{X} whose reduction modulo T is the neutral element $0 \in \mathcal{X}(R)$. A formal group law over \mathcal{X} is a datum $(\mathcal{X}; \gamma_1, \dots, \gamma_d)$, where $\gamma_1, \dots, \gamma_d \in \mathcal{X}(R[[T]])$ are such that their image in $\mathcal{X}(R[[T]]/T^2)$ forms a basis for $\text{Lie}(\mathcal{X})$. In particular, $(\gamma_i)_{1 \leq i \leq d}$ establish an isomorphism $\mathcal{X} \simeq \text{Spf}(R[[X_1, \dots, X_d]])$ of formal schemes over R . Recall that $\mathcal{X}(R[[T]])$ is the Cartier module associated with \mathcal{X} over the big Cartier ring (denoted by $\text{Cart}(R)$ in [Chai 2004, 2.3]). Since R is a \mathbb{Z}_p -algebra, the Cartier theory [Chai 2004, 4.3, 4.4] implies that there exists a p -typical formal group law $(\mathcal{X}; \gamma_1, \dots, \gamma_d)$ over \mathcal{X} , i.e., we have $\epsilon_p \cdot \gamma_i = 0$, where

$$\epsilon_p = \prod_{\substack{\ell \text{ prime} \\ (\ell, p)=1}} (1 - \frac{1}{\ell} V_\ell F_\ell)$$

is Cartier’s idempotent in $\text{Cart}(R)$; see [Chai 2004, 4.1]. Let $\Delta : \mathbb{Z}_p = W(\mathbf{F}_p) \rightarrow W(\mathbb{Z}_p)$ be the Cartier homomorphism given by $(x_0, x_1, \dots) \mapsto ([x_0], [x_1], \dots)$, where $x_n \in \mathbf{F}_p$ and $[x_n]$ denotes its Teichmüller lift. Then we get a natural map $u : \mathbb{Z}_p \xrightarrow{\Delta} W(\mathbb{Z}_p) \rightarrow W(R)$. For a $(p-1)$ -st root of unity $\zeta \in \mathbb{Z}_p$, we have $u(\zeta) = [\zeta] \in W(R)$. Note that for any $a \in R$ and $1 \leq i \leq d$, the p -typical curve $[a] \cdot \gamma_i$ is the image of γ_i under the map $\mathcal{X}(R[[T]]) \rightarrow \mathcal{X}(R[[T]])$ induced by $T \mapsto aT$. Applying this fact to $u(\zeta) \cdot \gamma_i = [\zeta] \cdot \gamma_i$, one obtains the lemma immediately.

Proposition 1.6. *Let $G = \text{Spec}(A)$ be a connected finite and flat group scheme over \mathbb{O}_K of order a power of p . Then there exists a presentation of A of type (2) such that the defining equations f_i for $1 \leq i \leq d$ have the form*

$$f_i(X_1, \dots, X_d) = \sum_{|n| \geq 1}^{\infty} a_{i,n} X^n \quad \text{with } a_{i,n} = 0 \text{ if } (p-1) \nmid (|n| - 1),$$

where $\underline{n} = (n_1, \dots, n_d) \in (\mathbb{Z}_{\geq 0})^d$ are multiindexes, $|n| = \sum_{j=1}^d n_j$, and X^n is short for $\prod_{j=1}^d X_j^{n_j}$.

Proof. By a theorem of Raynaud [Berthelot et al. 1982, 3.1.1], there is a projective abelian variety V over \mathbb{O}_K , and an embedding of group schemes $j : G \hookrightarrow V$. Let V' be the quotient of V by G . Let \mathcal{X}, \mathcal{Y} be, respectively, the formal completions of V and V' along their unit sections. They are formal groups over \mathbb{O}_K . Since G is connected, it is identified with the kernel of the natural isogeny $\phi : \mathcal{X} \rightarrow \mathcal{Y}$. Let (X_1, \dots, X_d) (respectively (Y_1, \dots, Y_d)) be parameters of \mathcal{X} (respectively \mathcal{Y}) satisfying the preceding lemma. The isogeny ϕ is thus given by

$$(X_1, \dots, X_d) \mapsto (f_1(X_1, \dots, X_d), \dots, f_d(X_1, \dots, X_d)),$$

where $f_i = \sum_{|n| \geq 1} a_{i,n} X^n \in \mathbb{O}_K \llbracket X_1, \dots, X_d \rrbracket$. Since for any $(p - 1)$ -th root of unity $\zeta \in \mathbb{Z}_p$ we have $f_i(\zeta X_1, \dots, \zeta X_d) = \zeta f_i(X_1, \dots, X_d)$, it's easy to see that $a_{i,n} = 0$ if $(p - 1) \nmid (|n| - 1)$. \square

Remark 1.7. As pointed out by the referee, we can avoid using Raynaud’s deep theorem to realize G as the kernel of an isogeny of formal groups over \mathbb{O}_K . In fact, by the biduality formula $G \simeq (G^D)^D$, where G^D denotes the Cartier dual of G , we have a canonical closed embedding $u : G \hookrightarrow U = \text{Res}_{G^D/S}(\mathbf{G}_m)$ of group schemes over $S = \text{Spec}(\mathbb{O}_K)$. Here, “ $\text{Res}_{G^D/S}$ ” means Weil’s restriction of scalars, so U is an affine smooth group scheme over S . Since the quotient of an affine scheme by a finite flat group scheme is always representable by a scheme [Raynaud 1967], we can consider the quotient $U' = U/G$ and the formal groups \mathcal{X}, \mathcal{Y} associated with U and U' , so that G is the kernel of the natural isogeny $\phi : \mathcal{X} \rightarrow \mathcal{Y}$.

1.8. Proof of Theorem 1. Let $H = G^{0+}$ be the connected component of G . By 1.3(i), we have $G^a = H^a$ for $a \in \mathbb{Q}_{>0}$. The exact sequence of finite flat group schemes $0 \rightarrow H \rightarrow G \rightarrow G/H \rightarrow 0$ induces a long exact sequence of finite \mathbb{O}_K -modules

$$0 \rightarrow H^{-1}(\ell_{G/H}) \rightarrow H^{-1}(\ell_G) \rightarrow H^{-1}(\ell_H) \rightarrow \omega_{G/H} \rightarrow \omega_G \rightarrow \omega_H \rightarrow 0,$$

where ℓ_G means the co-Lie complex of G [Berthelot et al. 1982, 3.2.9]. Since the generic fiber of G/H is étale, it's easy to see that $0 \rightarrow \omega_{G/H} \rightarrow \omega_G \rightarrow \omega_H \rightarrow 0$ is exact. Since G/H is étale, we have $\omega_{G/H} = 0$ and hence $\deg(G) = \deg(H)$. Up to replacing G by H , we may assume that $G = \text{Spec}(A)$ is connected.

We choose a presentation of A as in Proposition 1.6 so that we have an isomorphism of \mathbb{O}_K -algebras

$$A \simeq \mathbb{O}_K \llbracket X_1, \dots, X_d \rrbracket / (f_1, \dots, f_d)$$

where

$$f_i(X_1, \dots, X_d) = \sum_{j=1}^d a_{i,j} X_j + \sum_{|n| \geq p} a_{i,n} X^n.$$

As A is finite as an \mathbb{O}_K -module, we have

$$\Omega_{A/\mathbb{O}_K}^1 = \widehat{\Omega}_{A/\mathbb{O}_K}^1 \simeq \left(\bigoplus_{i=1}^d A dX_i \right) / (df_1, \dots, df_d).$$

Since $\omega_G \simeq e^*(\Omega_{A/\mathbb{O}_K}^1)$, where e is the unit section of G , we get

$$\omega_G \simeq \left(\bigoplus_{i=1}^d \mathbb{O}_K dX_i \right) / \left(\sum_{1 \leq j \leq d} a_{i,j} dX_j \right)_{1 \leq i \leq d}.$$

In particular, if U denotes the matrix $(a_{i,j})_{1 \leq i,j \leq d}$, then $\deg(G) = v_\pi(\det(U))$.

For any rational number λ , we denote by $\mathbf{D}^d(0, |\pi|^\lambda)$ (respectively $\mathring{\mathbf{D}}^d(0, |\pi|^\lambda)$) the rigid analytic closed (respectively open) disk of dimension d over K consisting of points (x_1, \dots, x_d) with $v_\pi(x_i) \geq \lambda$ (respectively $v_\pi(x_i) > \lambda$) for $1 \leq i \leq d$; we put $\mathbf{D}^d(0, 1) = \mathbf{D}^d(0, |\pi|^0)$ and $\mathring{\mathbf{D}}^d(0, 1) = \mathring{\mathbf{D}}^d(0, |\pi|^0)$. Let $a > p/(p-1) \deg(G)$ be a rational number, X^a be the a -th tubular neighborhood of G with respect to the chosen presentation. By (3), we have a cartesian diagram of rigid analytic spaces

$$\begin{CD} X^a @>>> \mathring{\mathbf{D}}^d(0, 1) \\ @V f VV @VV f=(f_1, \dots, f_d) V \\ \mathbf{D}^d(0, |\pi|^a) @>>> \mathring{\mathbf{D}}^d(0, 1), \end{CD} \tag{4}$$

where $f(y_1, \dots, y_d) = (f_1(y_1, \dots, y_d), \dots, f_d(y_1, \dots, y_d))$ and horizontal arrows are inclusions. Let X_0^a be the connected component of X^a containing 0. By the discussion below (3), we just need to prove that 0 is the only zero of the f_i contained in X_0^a .

Let $V = (b_{i,j})_{1 \leq i,j \leq d}$ be the unique $d \times d$ matrix with coefficients in \mathbb{O}_K such that $UV = VU = \det(U)I_d$, where I_d is the $d \times d$ identity matrix. If \mathbf{A}_K^d denotes the d -dimensional rigid affine space over K , then V defines an isomorphism of rigid spaces

$$\mathbf{g} : \mathbf{A}_K^d \rightarrow \mathbf{A}_K^d, \quad (x_1, \dots, x_d) \mapsto \left(\sum_{j=1}^d b_{1,j}x_j, \dots, \sum_{j=1}^d b_{d,j}x_j \right).$$

It's clear that $\mathbf{g}(\mathring{\mathbf{D}}^d(0, 1)) \subset \mathring{\mathbf{D}}^d(0, 1)$, so that f is defined on $\mathbf{g}(\mathring{\mathbf{D}}^d(0, 1))$. The composite morphism $f \circ \mathbf{g} : \mathring{\mathbf{D}}^d(0, 1) \rightarrow \mathring{\mathbf{D}}^d(0, 1)$ is given by

$$(x_1, \dots, x_d) \mapsto (\det(U)x_1 + R_1, \dots, \det(U)x_d + R_d), \tag{5}$$

where $R_i = \sum_{|n| \geq p} a_{i,n} \prod_{j=1}^d (\sum_{k=1}^d b_{j,k}x_k)^{n_j}$ involves only terms of order $\geq p$ for $1 \leq i \leq d$. For $1 \leq i \leq d$, we have basic estimations

$$v_\pi(\det(U)x_i) = \deg(G) + v_\pi(x_i) \quad \text{and} \quad v_\pi(R_i) \geq p \min_{1 \leq j \leq d} \{v_\pi(x_j)\}. \tag{6}$$

Lemma 1.9. *For any rational number $a > p/(p-1) \deg(G)$, the map \mathbf{g} induces an isomorphism of affinoid rigid spaces*

$$\mathbf{g} : \mathbf{D}^d(0, |\pi|^{a-\deg(G)}) \xrightarrow{\sim} X_0^a.$$

Assuming this lemma for a moment, we can complete the proof of Theorem 1 as follows. Consider the composite

$$\mathbf{h} = f \circ \mathbf{g} |_{\mathbf{D}^d(0, |\pi|^{a-\deg(G)})} : \mathbf{D}^d(0, |\pi|^{a-\deg(G)}) \xrightarrow{\sim} X_0^a \hookrightarrow X^a \xrightarrow{f} \mathbf{D}^d(0, |\pi|^a).$$

To complete the proof of Theorem 1, we just need to prove that $\mathbf{h}^{-1}(0) = \{0\}$. Let (x_1, \dots, x_d) be a point of $\mathbf{D}^d(0, |\pi|^{a-\deg(G)})$, and $(z_1, \dots, z_d) = \mathbf{h}(x_1, \dots, x_d)$. We may assume $v_\pi(x_1) = \min_{1 \leq i \leq d} \{v_\pi(x_i)\}$. We have $v_\pi(x_1) \geq a - \deg(G) > 1/(p-1) \deg(G)$ by the assumption on a . It follows thus from (6) that

$$v_\pi(R_1) \geq p v_\pi(x_1) > \deg(G) + v_\pi(x_1) = v_\pi(\det(U)x_1).$$

Hence, we deduce from (5) that $v_\pi(z_1) = \deg(G) + v_\pi(x_1)$. In particular, $z_1 = 0$ if and only if $x_1 = 0$. Therefore, we have $\mathbf{h}^{-1}(0) = \{0\}$. This achieves the proof of Theorem 1.

Proof of Lemma 1.9. Let ϵ be any rational number with

$$0 < \epsilon < (p-1)/pa - \deg(G).$$

We will prove that

$$\mathbf{D}^d(0, |\pi|^{a-\deg(G)}) = \mathbf{D}^d(0, |\pi|^{a-\deg(G)-\epsilon}) \cap \mathbf{g}^{-1}(X^a).$$

This will imply that $\mathbf{D}^d(0, |\pi|^{a-\deg(G)})$ is a connected component of $\mathbf{g}^{-1}(X^a)$. Since $\mathbf{g} : \mathbf{A}_K^d \rightarrow \mathbf{A}_K^d$ is an isomorphism, the lemma will follow immediately.

We prove first the inclusion \subset . It suffices to show $\mathbf{g}(\mathbf{D}^d(0, |\pi|^{a-\deg(G)})) \subset X^a$. Let (x_1, \dots, x_d) be a point of $\mathbf{D}^d(0, |\pi|^{a-\deg(G)})$. By (4), we have to check that $(z_1, \dots, z_d) = \mathbf{f}(\mathbf{g}(x_1, \dots, x_d))$ lies in $\mathbf{D}^d(0, |\pi|^a)$. We obtain from (6) that $v_\pi(\det(U)x_i) = \deg(G) + v_\pi(x_i) \geq a$ and $v_\pi(R_i) \geq p(a - \deg(G))$. As $a > p/(p-1) \deg(G)$, we have $v_\pi(R_i) > a$. It follows from (5) that

$$v_\pi(z_i) \geq \min\{v_\pi(\det(U)x_i), v_\pi(R_i)\} \geq a.$$

This proves $(z_1, \dots, z_d) \in \mathbf{D}^d(0, |\pi|^a)$; hence $\mathbf{g}(\mathbf{D}^d(0, |\pi|^{a-\deg(G)})) \subset X^a$.

To prove the inclusion \supset , we just need to verify that every point which is in $\mathbf{D}^d(0, |\pi|^{a-\deg(G)-\epsilon})$ but outside $\mathbf{D}^d(0, |\pi|^{a-\deg(G)})$ does not lie in $\mathbf{g}^{-1}(X^a)$. Let (x_1, \dots, x_d) be such a point. We may assume that

$$a - \deg(G) - \epsilon \leq v_\pi(x_1) < a - \deg(G) \quad \text{and} \quad v_\pi(x_i) \geq a - \deg(G) - \epsilon \quad \text{for } 2 \leq i \leq d. \tag{7}$$

Let

$$(z_1, \dots, z_d) = (\det(U)x_1 + R_d, \dots, \det(U)x_d + R_d)$$

be the image of (x_1, \dots, x_d) under the composite $\mathbf{f} \circ \mathbf{g}$. According to (4), the proof will be completed if we can prove that (z_1, \dots, z_d) is not in $\mathbf{D}^d(0, |\pi|^a)$. From (6) and (7), we get $v_\pi(\det(U)x_1) = \deg(G) + v_\pi(x_1) < a$ and $v_\pi(R_1) \geq p(a - \deg(G) - \epsilon)$. Thanks to the assumption on ϵ , we have $p(a - \deg(G) - \epsilon) > a$, so $v_\pi(z_1) = v_\pi(\det(U)x_1) < a$. This shows that (z_1, \dots, z_d) is not in $\mathbf{g}^{-1}(X^a)$; hence the proof of the lemma is complete. \square

2. Applications to canonical subgroups

In this section, we suppose the fraction field K has characteristic 0 and the residue field k is perfect of characteristic $p \geq 3$. Let e be the absolute ramification index of \mathbb{O}_K . For any rational number $\epsilon > 0$, we denote by $\mathbb{O}_{K,\epsilon}$ the quotient of \mathbb{O}_K by the ideal consisting of elements with p -adic valuation greater or equal to ϵ .

2.1. First we recall some results on the from [Abbes and Mokrane 2004; Tian 2010; Fargues 2009]. Let $v_p : \mathbb{O}_K/p \rightarrow [0, 1]$ be the truncated p -adic valuation (with $v_p(0) = 1$). Let G be a truncated Barsotti–Tate group of level $n \geq 1$ nonétale over \mathbb{O}_K , and $G_1 = G \otimes_{\mathbb{O}_K} (\mathbb{O}_K/p)$. The Lie algebra of G_1 denoted by $\text{Lie}(G_1)$ is a finite free \mathbb{O}_K/p -module. The Verschiebung homomorphism $V_{G_1} : G_1^{(p)} \rightarrow G_1$ induces a semilinear endomorphism φ_{G_1} of $\text{Lie}(G_1)$. We choose a basis of $\text{Lie}(G_1)$ over \mathbb{O}_K/p , and let U be the matrix of φ under this basis. We define the Hodge height of G , denoted by $h(G)$, to be the truncated p -adic valuation of $\det(U)$. We note that the definition of $h(G)$ does not depend on the choice of U . The Hodge height of G is an analog of the Hasse invariant in mixed characteristic, and we have $h(G) = 0$ if and only if G is ordinary.

Theorem 2.2 [Fargues 2009, théorème 4]. *Let G be a truncated Barsotti–Tate group of level 1 over \mathbb{O}_K of dimension $d \geq 1$ and height h . Assume $h(G) < 1/2$ if $p \geq 5$ and $h(G) < 1/3$ if $p = 3$.*

- (i) *For any rational number $ep/(p-1)h(G) < a \leq ep/(p-1)(1-h(G))$, the finite flat subgroup G^a of G given by the Abbes–Saito filtration has rank p^d .*
- (ii) *Let C be the subgroup $G^{ep/(p-1)(1-h(G))}$ of G . We have $\deg(G/C) = eh(G)$.*
- (iii) *The subgroup $C \otimes \mathbb{O}_{K,1-h(G)}$ coincides with the kernel of the Frobenius homomorphism of $G \otimes \mathbb{O}_{K,1-h(G)}$. Moreover, for any rational number ϵ with $h(G)/(p-1) < \epsilon \leq 1-h(G)$, if H is a finite and flat closed subgroup of G such that $H \otimes \mathbb{O}_{K,\epsilon}$ coincides with the kernel of Frobenius of $G \otimes \mathbb{O}_{K,\epsilon}$, then we have $H = C$.*

The subgroup C in this theorem, when it exists, is called the *canonical subgroup* (of level 1) of G .

Remark 2.3. The conventions here are slightly different from those in [Fargues 2009]. The Hodge height is called Hasse invariant there, while we choose to follow the terminologies in [Abbes and Mokrane 2004] and [Tian 2010]. Our index of Abbes–Saito filtration and the degree of G are e times those in [Fargues 2009].

Part (iii) of Theorem 2.2 is not explicitly stated in [Fargues 2009, théorème 4], but it’s an easy consequence of Proposition 11 in that paper.

For the canonical subgroups of higher level, we have this:

Theorem 2.4 [Fargues 2009, théorème 6]. *Let G be a truncated Barsotti–Tate group of level n over \mathbb{O}_K of dimension $d \geq 1$ and height h . Assume $h(G) < 1/3^n$ if $p = 3$ and $h(G) < 1/(2p^{n-1})$ if $p \geq 5$.*

(i) *There exists a unique closed subgroup of G that is finite and flat over \mathbb{O}_K and satisfies the following:*

- $C_n(\bar{K})$ is free of rank d over $\mathbb{Z}/p^n\mathbb{Z}$.
- For each integer i with $1 \leq i \leq n$, let C_i be the scheme theoretic closure of $C_n(\bar{K})[p^i]$ in G . Then the subgroup $C_i \otimes \mathbb{O}_{K,1-p^{i-1}h(G)}$ coincides with the kernel of the i -th iterated Frobenius of $G \otimes \mathbb{O}_{K,1-p^{i-1}h(G)}$.

(ii) *We have $\deg(G/C_n) = e(p^n - 1)/(p - 1)h(G)$.*

The subgroup C_n in the theorem above is called the canonical subgroup of level n of G . Fargues actually proves that C_n is a certain piece of the Harder–Narasimhan filtration of G . The aim of this section is to show that C_n appears also in the Abbes–Saito filtration.

Theorem 2.5. *Let G be a truncated Barsotti–Tate group of level n over \mathbb{O}_K satisfying the assumptions in Theorem 2.4, and C_n be its canonical subgroup of level n . Then for any rational number a satisfying*

$$ep(p^n - 1)/(p - 1)^2h(G) < a \leq ep/(p - 1)(1 - h(G)),$$

we have $G^a = C_n$.

Proof. We proceed by induction on n . If $n = 1$, this is Theorem 2.2(i). We suppose $n \geq 2$ and the theorem is valid for truncated Barsotti–Tate groups of level $n - 1$. For each integer i with $1 \leq i \leq n$, let G_i denote the scheme theoretic closure of $G(\bar{K})[p^i]$ in G , and C_i the scheme theoretic closure of $C_n(\bar{K})[p^i]$ in C_n . By Theorem 2.4(i), it’s clear that C_i is the canonical subgroup of level i of G_i . Let a be a rational number with $(ep(p^n - 1)/(p - 1)^2)h(G) < a \leq (ep/(p - 1))(1 - h(G))$. By the induction hypothesis and the functoriality of Abbes–Saito filtration 1.3(ii), we have $C_{n-1}(\bar{K}) = G_{n-1}^a(\bar{K}) \subset G^a(\bar{K})$, and the image of $G^a(\bar{K})$ in $G_1(\bar{K})$ is exactly $C_1(\bar{K}) = G_1^a(\bar{K})$. Note that we have a commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & C_{n-1}(\bar{K}) & \longrightarrow & C_n(\bar{K}) & \longrightarrow & C_1(\bar{K}) \longrightarrow 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & G_{n-1}(\bar{K}) & \longrightarrow & G(\bar{K}) & \xrightarrow{\times p^{n-1}} & G_1(\bar{K}) \longrightarrow 0, \end{array}$$

where the rows are exact sequences of groups and the vertical arrows are natural inclusions. So we have $C_n(\bar{K}) \subset G^a(\bar{K})$. On the other hand, Theorems 1 and 2.4(ii)

imply that $(G/C_n)^a(\bar{K}) = 0$ since

$$a > \frac{ep(p^n - 1)}{(p - 1)^2} h(G) = \frac{p}{p - 1} \deg(G/C_n).$$

Therefore, we get $G^a(\bar{K}) \subset C_n(\bar{K})$ by Proposition 1.3(ii). This completes the proof. \square

Acknowledgements

I would like to thank Ahmed Abbes for his comments on an earlier version of this paper. I also express my deep gratitude to the anonymous referee for his careful reading and useful suggestions for clarifying some arguments.

References

- [Abbes and Mokrane 2004] A. Abbes and A. Mokrane, “Sous-groupes canoniques et cycles évanescents p -adiques pour les variétés abéliennes”, *Publ. Math. Inst. Hautes Études Sci.* 99 (2004), 117–162. MR 2005f:14090 Zbl 1062.14057
- [Abbes and Saito 2002] A. Abbes and T. Saito, “Ramification of local fields with imperfect residue fields”, *Amer. J. Math.* 124:5 (2002), 879–920. MR 2003m:11196 Zbl 1084.11064
- [Abbes and Saito 2003] A. Abbes and T. Saito, “Ramification of local fields with imperfect residue fields, II”, *Doc. Math. Extra Vol.* (2003), 5–72. MR 2005g:11231 Zbl 1127.11349
- [Berthelot et al. 1982] P. Berthelot, L. Breen, and W. Messing, *Théorie de Dieudonné cristalline, II*, Lecture Notes in Mathematics 930, Springer, Berlin, 1982. MR 85k:14023 Zbl 0516.14015
- [Chai 2004] C. L. Chai, “Notes on Cartier–Dieudonné theory”, 2004, available at <http://tinyurl.com/74bcra7>.
- [Fargues 2007] L. Fargues, “La filtration de Harder–Narasimhan des schémas en groupes finis et plats”, preprint, 2007. To appear in *J. Reine Angew. Math.*
- [Fargues 2009] L. Fargues, “La filtration canonique des points de torsion des groupes p -divisibles”, preprint, 2009, available at <http://www-irma.u-strasbg.fr/~fargues/canoniqueHN.pdf>.
- [Hattori 2006] S. Hattori, “Ramification of a finite flat group scheme over a local field”, *J. Number Theory* 118:2 (2006), 145–154. MR 2007b:14104 Zbl 1107.14036
- [Hazewinkel 1978] M. Hazewinkel, *Formal groups and applications*, Pure and Applied Mathematics 78, Academic Press, New York, 1978. MR 82a:14020 Zbl 0454.14020
- [Raynaud 1967] M. Raynaud, “Passage au quotient par une relation d’équivalence plate”, pp. 78–85 in *Proc. Conf. Local Fields* (Driebergen, 1966), edited by T. A. Springer, Springer, Berlin, 1967. MR 38 #1104 Zbl 0165.24003
- [Tian 2010] Y. Tian, “Canonical subgroups of Barsotti–Tate groups”, *Ann. of Math. (2)* 172:2 (2010), 955–988. MR 2012a:14105 Zbl 1203.14026

Communicated by Brian Conrad

Received 2010-05-03

Revised 2011-05-02

Accepted 2011-05-30

yichaot@math.ac.cn

Mathematics Department, Fine Hall, Washington Road,
Princeton, NJ 08544, United States

Current address:

Morningside Center of Mathematics, 55 Zhong Guan Cun
East Road, Haidian District, Beijing, 100190, China

On the smallest number of generators and the probability of generating an algebra

Rostyslav V. Kravchenko, Marcin Mazur and Bogdan V. Petrenko

In this paper we study algebraic and asymptotic properties of generating sets of algebras over orders in number fields. Let A be an associative algebra over an order R in an algebraic number field. We assume that A is a free R -module of finite rank. We develop a technique to compute the smallest number of generators of A . For example, we prove that the ring $M_3(\mathbb{Z})^k$ admits two generators if and only if $k \leq 768$. For a given positive integer m , we define the density of the set of all ordered m -tuples of elements of A which generate it as an R -algebra. We express this density as a certain infinite product over the maximal ideals of R , and we interpret the resulting formula probabilistically. For example, we show that the probability that 2 random 3×3 matrices generate the ring $M_3(\mathbb{Z})$ is equal to $(\zeta(2)^2 \zeta(3))^{-1}$, where ζ is the Riemann zeta function.

1. Introduction	243
2. Preliminary results	247
3. The density of the set of ordered k -tuples which generate an algebra	253
4. Proof of Theorem 3.3	256
5. The smallest number of generators	265
6. Generators of matrix algebras over finite fields	269
7. The numbers $ \text{Gen}_k(M_n(\mathbb{F}_q), \mathbb{F}_q) $	272
8. Finite products of matrix algebras over rings of algebraic integers	284
Acknowledgments	290
References	290

1. Introduction

Let R be a commutative ring with 1. Recall that a set S generates an associative unital R -algebra A if the set of all monomials in the elements of S (including the

MSC2000: primary 16S15, 11R45, 11R99, 15A33, 15B36, 11C20, 11C08; secondary 16P10, 16H05.

Keywords: density, smallest number of generators, probability of generating.

degree-zero monomial 1) spans A as an R -module. This paper lays a foundation for our program to investigate properties of the sets of generators of R -algebras A whose additive group is a finitely generated R -module. A substantial part of our results grew out of the following question: given a ring A whose additive group is a free abelian group of finite rank and a positive integer k , what is the probability that k random elements of A generate it as a \mathbb{Z} -algebra? We will show that this question can be stated in a rigorous way and that it has a very interesting answer. The following formulas, in which ζ denotes the Riemann zeta function, are special cases of our results (see Theorem 8.1):

- The probability that m random 2×2 matrices generate the ring $M_2(\mathbb{Z})$ is equal to $1/(\zeta(m-1)\zeta(m))$.
- The probability that 2 random 3×3 matrices generate the ring $M_3(\mathbb{Z})$ is equal to $1/(\zeta(2)^2\zeta(3))$.
- The probability that 3 random 3×3 matrices generate the ring $M_3(\mathbb{Z})$ is equal to

$$\frac{1}{\zeta(2)\zeta(3)\zeta(4)} \prod_p \left(1 + \frac{1}{p^2} + \frac{1}{p^3} - \frac{1}{p^5} \right),$$

where the product is taken over all prime numbers.

Our main results are obtained for algebras A over an order R in some number field such that A is a free R -module of finite rank. It is not hard though to extend the results to the case when R is an order in a global field of positive characteristic (we will address this in a follow-up paper). Roughly speaking, a choice of an integral basis of R and of a basis of A over R allows us to introduce integral coordinates on all Cartesian powers A^k , $k \in \mathbb{N}$. For any subset S of A^k and any N we consider the finite set $S(N)$ of all points whose coordinates are in the interval $[-N, N]$. We define the density $\text{den}(S)$ of S as the limit

$$\lim_{N \rightarrow \infty} \frac{|S(N)|}{|A^k(N)|}$$

(we do not claim that it always exists). Our goal is to calculate the density of the set of generators of A .

Definition 1.1. Let A be an algebra over a commutative ring R , and let k be a positive integer. We define the set $\text{Gen}_k(A, R)$ as follows:

$$\text{Gen}_k(A, R) = \{(a_1, \dots, a_k) \in A^k : a_1, \dots, a_k \text{ generate } A \text{ as an } R\text{-algebra}\}.$$

For the rest of the introduction, we assume that R is an order in a number field K and A is an R -algebra which is free of finite rank m as an R -module (unless stated otherwise).

In Theorem 3.2 we prove that the set $\text{Gen}_k(A, R)$ has density, which we denote by $\text{den}_k(A)$, and that it can be computed locally as follows:

$$\text{den}_k(A) = \prod_{\mathfrak{p} \in \text{m-Spec } R} \frac{|\text{Gen}_k(A/\mathfrak{p}A, R/\mathfrak{p})|}{|R/\mathfrak{p}|^{mk}}, \tag{1}$$

where $\text{m-Spec } R$ denotes the set of all maximal ideals of R . In order to prove Theorem 3.2 we had to extend this local-to-global formula for density to a substantially larger class of sets. This led us to Theorem 3.3, which is of independent interest and has potential applications to various other questions. Theorem 3.3 deals with a finite set f_1, \dots, f_s of polynomials in $R[x_1, \dots, x_n]$ and the set S of all $a \in R^n$ such that the ideal generated by $f_1(a), \dots, f_s(a)$ is R . It asserts that the set S has density $\text{den}(S)$ given by the formula

$$\text{den}(S) = \prod_{\mathfrak{p} \in \text{m-Spec } R} \left(1 - \frac{t_{\mathfrak{p}}}{|R/\mathfrak{p}|^n}\right),$$

where $t_{\mathfrak{p}}$ is the number of common zeros in $(R/\mathfrak{p})^n$ of the polynomials f_1, \dots, f_s considered as polynomials over the field R/\mathfrak{p} .

As a first application of our results we answer in Section 3A the following question posed by Ilya Kapovich: what is the probability that m random elements of a free abelian group of rank $n \leq m$ generate the group? Our results provide a rigorous proof of the following answer: the probability in question is equal to $(\prod_{k=m-n+1}^m \zeta(k))^{-1}$, where ζ is the Riemann zeta function (when $m = n$ this product should be interpreted as 0).

In Section 5 we show how (1) can be used to get information about the smallest number of generators of an R -algebra A .

Definition 1.2. Let A be a finitely generated R -algebra. By $r(A, R)$ we denote the smallest number of generators of A as an R -algebra.

In Theorem 5.2 we prove that if k is an integer such that $k - 1 \geq r_0 := r(A \otimes_R K, K)$ and $k \geq r_{\mathfrak{p}} := r(A/\mathfrak{p}A, R/\mathfrak{p}R)$ for every maximal ideal \mathfrak{p} of R then $\text{den}_k(A) > 0$. Let r_f be the largest among the numbers $r_{\mathfrak{p}}$. Clearly, if $\text{den}_k(A) > 0$ then A can be generated by k elements. Using this remark and Theorem 5.2 we show in Theorem 5.5 that the smallest number of generators of A coincides with r_f if $r_f > r_0$ and is either r_0 or $r_0 + 1$ otherwise. A special case of this result, when $R = \mathbb{Z}$, was kindly communicated to us by H. W. Lenstra (private communication, 2007). Note that when $r_f = r_0$, we only know that r is either r_0 or $r_0 + 1$. Nevertheless, it is often possible to prove that $\text{den}_{r_0}(A) > 0$ and conclude that $r = r_0$. For example, we have been unable for a long time to find the largest integer n such that the product $M_3(\mathbb{Z})^n$ of n copies of the matrix ring $M_3(\mathbb{Z})$ admits two generators as a \mathbb{Z} -algebra. We knew that $n \leq 768$, but any attempts to construct explicitly two

generators for such large values of n have been beyond our computational ability. It turns out, though, that we can prove that $\text{den}_2(\mathbf{M}_3(\mathbb{Z})^{768}) > 0$, hence we get a (nonconstructive) proof that $n = 768$ (see Theorem 8.2).

In Theorem 5.7 we extend Lenstra's original approach to obtain a similar formula for the smallest number of generators of algebras over any commutative ring R of dimension at most 1. This formula is reminiscent of the Forster–Swan theorem on the number of generators of modules over Noetherian commutative rings [Matsumura 1986, Theorem 5.8]. By analogy with this result, in Conjecture 5.8 we propose an extension of our formula to algebras over general Noetherian rings.

In order to use (1) in concrete cases one needs to be able to compute the numbers $|\text{Gen}_k(A/\mathfrak{p}A, R/\mathfrak{p})|$. This leads us to the results of Sections 6 and 7, where we study these numbers under the assumption that $A/\mathfrak{p}A$ is a product of matrix algebras. After various reductions in Section 6 we derive explicit formulas for $|\text{Gen}_k(\mathbf{M}_n(\mathbb{F}), \mathbb{F})|$, where \mathbb{F} is a finite field and $n = 2, 3$. Furthermore, we get a lower bound when $n > 3$ (Proposition 7.9). As a corollary, we prove that if $m \geq 2$ then the probability that m matrices in $\mathbf{M}_n(\mathbb{F}_q)$, chosen under the uniform distribution, generate the \mathbb{F}_q -algebra $\mathbf{M}_n(\mathbb{F}_q)$ tends to 1 as $q + m + n \rightarrow \infty$ (see Corollary 7.10). This result proves and vastly generalizes the conjectural formula [Petrenko and Sidki 2007, (17), p. 27]. The case of $n = 2$ and some of the results of Section 6 have been discussed earlier in [Kravchenko and Petrenko 2006], which was the starting point for the present work. This part of our paper has been influenced by ideas of Philip Hall [1936].

In Section 8 the results of Sections 6 and 7 are applied to finite products of matrix algebras over the ring of integers in a number field.

To state some of our remaining results, we need the following notation.

Definition 1.3. Let $m, n \geq 1$ be integers and let A be an R -algebra. We introduce the following notation:

- (i) $\text{gen}_m(A, R)$ is the largest $k \in \mathbb{Z} \cup \{\infty\}$ such that $r(A^k, R) \leq m$;
- (ii) $\text{gen}_{m,n}(q) = \text{gen}_m(\mathbf{M}_n(\mathbb{F}_q), \mathbb{F}_q)$;
- (iii) $\mathfrak{g}_{m,n}(q) = |\text{Gen}_m(\mathbf{M}_n(\mathbb{F}_q), \mathbb{F}_q)|$.

We show in Proposition 6.2 that

$$\text{gen}_{m,n}(q) = \frac{\mathfrak{g}_{m,n}(q)}{|\text{PGL}_n(\mathbb{F}_q)|}$$

and $r(\mathbf{M}_n(\mathbb{F}_q)^{1+\text{gen}_{m,n}(q)}, \mathbb{F}_q) = m + 1$ by Corollary 2.15.

Here are some special cases of our results in Section 8:

$$(1) \text{gen}_{m,2}(q) = \frac{q^{2m-1}(q^m - 1)(q^m - q)}{q^2 - 1}.$$

$$(2) \text{ gen}_m(\mathbb{M}_2(\mathbb{Z}), \mathbb{Z}) = \text{gen}_{m,2}(2) = \frac{2^{2m-1}(2^m - 2)(2^m - 1)}{3}.$$

$$(3) \text{ gen}_{m,3}(q) = \frac{q^{3m-3}(q^m - 1)(q^m - q)(q^m + q)}{(q - 1)^2(q + 1)(q^2 + q + 1)} \\ \times (q^{3m} - q^{m+2} + q^{2m} - 2q^{m+1} - q^m + q^3 + q^2).$$

$$(4) \text{ gen}_m(\mathbb{M}_3(\mathbb{Z}), \mathbb{Z}) = \text{gen}_{m,3}(2) \\ = \frac{(2^m - 2)(2^m - 1)(2^m + 2)(2^{3m} + 2^{2m} - 2^{m+3} - 2^m + 12)2^{3m-3}}{21}.$$

The techniques we have developed so far can be applied to any finitely generated \mathbb{Z} -algebra whose reduction modulo every prime is a direct sum of matrix rings over finite fields. However, among maximal orders in semisimple algebras over \mathbb{Q} the only such algebras are the maximal orders in matrix rings by the Hasse–Brauer–Noether–Albert theorem. In order to extend our results to maximal orders in other semisimple algebras we need to obtain formulas for the number of generators of algebras over finite fields which have nontrivial Jacobson radical. This will be done in a subsequent paper. Let us just mention here a special case, when A is a maximal order in the quaternion algebra $\mathbb{Q}(i, j)$ ($i^2 = -1 = j^2$). For any odd prime p we have $A/pA \cong \mathbb{M}_2(\mathbb{F}_p)$, so A and $\mathbb{M}_2(\mathbb{Z})$ differ only at the prime 2 and at infinity. Note that $A/2A$ is a commutative algebra over \mathbb{F}_2 whose quotient modulo the Jacobson radical is the field \mathbb{F}_4 . Since \mathbb{F}_4^{16} cannot be generated by two elements, we see that A^{16} requires at least three generators. It can be verified that A^{15} admits two generators. So A can be distinguished from $\mathbb{M}_2(\mathbb{Z})$ by counting the smallest number of generators of powers of these two algebras. Note that for the integral quaternions $\mathbb{Z}[i, j]$ already $\mathbb{Z}[i, j]^4$ requires at least three generators. In a subsequent paper we will extend these observations to a much larger class of orders.

In another work in progress we apply the techniques developed in the present paper to study generators of various nonassociative algebras. Our technique applies to any finitely generated R -module equipped with an R -bilinear form, but we focus mainly on Lie algebras and Jordan algebras. For example, we show that the probability that m random elements generate the Lie ring $\mathfrak{sl}_2(\mathbb{Z})$ of 2×2 integer matrices with zero trace is equal to

$$\frac{1}{\zeta(m-1)\zeta(m)}.$$

2. Preliminary results

Let R be a commutative ring with 1. Unless stated otherwise, all R -algebras are assumed to be associative, unital, and finitely generated as an R -module.

In this section we collect several fairly straightforward observations which are used through the paper. Let A be an R -algebra. Recall that elements a_1, \dots, a_k generate A as an R -algebra if all the (noncommutative) monomials in a_1, \dots, a_k , including the degree-zero monomial 1, generate A as an R -module. We say that a_1, \dots, a_k *strongly generate* A as an R -algebra if already all the (noncommutative) monomials in a_1, \dots, a_k of positive degree generate A as an R -module.

Lemma 2.1. *Suppose that there exists no R -algebra homomorphism $A \rightarrow R/I$ for any proper ideal I of R . Then any set that generates A as an R -algebra also strongly generates A .*

Proof. Suppose that a_1, \dots, a_k generate A as an R algebra and let J be the R submodule of A generated by all the (noncommutative) monomials in a_1, \dots, a_k of positive degree. Then $R \cdot 1 + J = A$. Since J is closed under multiplication, it is a two-sided ideal of A and $A/J \cong R/(R \cap J)$. By our assumption, $R \cap J$ cannot be a proper ideal of R , so $R \cdot 1 \subset J$ and $J = A$. \square

Example 2.2. Let the algebra $A = \prod_{i=1}^n M_{m_i}(R)$ be a finite product of matrix algebras over R , with each $m_i \geq 2$. Then any set which generates A as an R -algebra also strongly generates A . This is a direct consequence of Lemma 2.1 and the remark that A has no nontrivial commutative quotients.

In this paper we decided to focus on unital algebras and we do not discuss strong generators. However most of our results can be easily modified to sets of strong generators and algebras which are not necessarily unital. One can also reduce questions about strong generators to generators using the following observation. Recall that if A is an R -algebra (unital or not) we can construct a unital algebra $A^{(1)}$ which is $R \oplus A$ as an R -module with multiplication defined by $(r, a)(s, b) = (rs, ab + rb + sa)$. We have the following lemma.

Lemma 2.3. *Let $a_1, \dots, a_k \in A$. Then the following conditions are equivalent:*

- (1) a_1, \dots, a_k strongly generate A as an R -algebra.
- (2) $(r_1, a_1), \dots, (r_k, a_k)$ generate $A^{(1)}$ as an R -algebra for any elements $r_1, \dots, r_k \in R$.
- (3) $(r_1, a_1), \dots, (r_k, a_k)$ generate $A^{(1)}$ as an R -algebra for some elements $r_1, \dots, r_k \in R$.

Proof. We identify A with the ideal $\{0\} \oplus A$ in $A^{(1)}$. Assume (1) and let r_1, \dots, r_k be in R . Since $(0, a_i) = (r_i, a_i) - r_i(1, 0)$, the R -subalgebra B of $A^{(1)}$ generated by $(r_1, a_1), \dots, (r_k, a_k)$ contains all monomials in $(0, a_1), \dots, (0, a_k)$, hence it contains A . Since B also contains $R \oplus \{0\}$, we see that $A^{(1)} = B$. Thus condition (1) indeed implies (2). Condition (3) is clearly a consequence of (2). Assume (3) and let C be the subalgebra of A strongly generated by a_1, \dots, a_k . Note that

any monomial of positive degree in $(r_1, a_1), \dots, (r_k, a_k)$ is of the form (r, c) for some $r \in R$ and $c \in C$. By the assumption in (3), for any $a \in A$ there is $r \in R$ such that (r, a) is an R -linear combination of monomials of positive degree in $(r_1, a_1), \dots, (r_k, a_k)$. It follows that $a \in C$. Thus $C = A$, which shows that (1) follows from (3). \square

The following observation is straightforward.

Lemma 2.4. *Let A be an R -algebra. For any ideal I of R we have*

$$A^{(1)}/IA^{(1)} = (A/IA)^{(1)},$$

where the adjunction of unity on the right is in the category of R/I -algebras.

Definition 2.5. For an R -algebra A and positive integer k we denote by $\text{Gen}_k(A, R)$ the set of all k -tuples $(a_1, \dots, a_k) \in A^k$ which generate A as an R -algebra. When there is no danger of confusion, we write $\text{Gen}_k(A)$ for $\text{Gen}_k(A, R)$.

Lemma 2.6. *The elements a_1, \dots, a_k generate A as an R -algebra if and only if for every maximal ideal \mathfrak{m} of R the images of a_1, \dots, a_k in $A \otimes_R R/\mathfrak{m} = A/\mathfrak{m}A$ generate $A/\mathfrak{m}A$ as an R/\mathfrak{m} -algebra.*

Proof. Let J be the R submodule of A generated by all the (noncommutative) monomials in a_1, \dots, a_k . By [Matsumura 1986, Theorem 4.8], $A = J$ if and only if $A/J \otimes_R R/\mathfrak{m} = 0$ for every maximal ideal \mathfrak{m} of R . The result follows from the simple remark that $A/J \otimes_R R/\mathfrak{m} = 0$ if and only if the images of a_1, \dots, a_k in $A/\mathfrak{m}A$ generate it as an R/\mathfrak{m} -algebra. \square

Lemma 2.7. *Let R be a field and let A be an R -algebra of dimension m . The elements a_1, \dots, a_k generate A as an R -algebra if and only if the (noncommutative) monomials in a_1, \dots, a_k of degree $< m$ span A as an R -vector space.*

Proof. Let A_i be the subspace of A spanned by all the monomials in a_1, \dots, a_k of degree $\leq i$. Clearly $A_0 \subseteq A_1 \subseteq A_2 \subseteq \dots$. We also see that

$$A_{i+1} = A_i + a_1A_i + a_2A_i + \dots + a_kA_i,$$

for any i . It follows that if $A_i = A_{i+1}$ for some i , then $A_j = A_i$ for all $j \geq i$. Since $\dim_R A_m \leq m$, we must have $A_i = A_{i+1}$ for some $i < m$. Thus $A_i = A_{m-1}$ for all $i \geq m$. This proves that a_1, \dots, a_k generate A as an R -algebra if and only if $A = A_{m-1}$. \square

Lemma 2.8. *Suppose that A can be generated by m elements as an R -module. The elements a_1, \dots, a_k generate A as an R -algebra if and only if the (noncommutative) monomials in a_1, \dots, a_k of degree $< m$ generate A as an R -module.*

Proof. Suppose that a_1, \dots, a_k generate A as an R -algebra and let A_i be the R -submodule of A generated by all the monomials in a_1, \dots, a_k of degree $\leq i$. For any maximal ideal \mathfrak{m} of R the dimension of $A/\mathfrak{m}A$ over R/\mathfrak{m} does not exceed m . Thus $A/A_{m-1} \otimes_R R/\mathfrak{m} = 0$ for every maximal ideal \mathfrak{m} of R by Lemma 2.7. Hence $A = A_{m-1}$ by [Matsumura 1986, Theorem 4.8]. \square

Recall that $\text{Spec } R$ is the set of all prime ideals of R equipped with the Zariski topology and $\mathfrak{m}\text{-Spec } R$ is the subspace of $\text{Spec } R$ consisting of all maximal ideals. For $\mathfrak{p} \in \text{Spec } R$ we denote by $R_{\mathfrak{p}}$ the localization of R at the prime ideal \mathfrak{p} and we set $A_{\mathfrak{p}} = R_{\mathfrak{p}} \otimes_R A$. The residue field $R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$ is denoted by $\kappa(\mathfrak{p})$. Recall that $\kappa(\mathfrak{p})$ coincides with the field of fractions of R/\mathfrak{p} .

Definition 2.9. We say that the elements a_1, \dots, a_k generate A at a prime ideal \mathfrak{p} of R if their images in $\kappa(\mathfrak{p}) \otimes_R A$ generate $\kappa(\mathfrak{p}) \otimes_R A$ as a $\kappa(\mathfrak{p})$ -algebra. Equivalently, a_1, \dots, a_k generate A at \mathfrak{p} if their images in $A_{\mathfrak{p}}$ generate $A_{\mathfrak{p}}$ as an $R_{\mathfrak{p}}$ -algebra.

Lemma 2.10. *Let $a_1, \dots, a_k \in A$. The set of all prime ideals \mathfrak{p} such that a_1, \dots, a_k generate A at \mathfrak{p} is open.*

Proof. Let B be the R submodule of A generated by all monomials in a_1, \dots, a_k of degree $< m$, where m is such that A can be generated by m elements as an R -module. By Lemma 2.8, the images of a_1, \dots, a_k in $A_{\mathfrak{p}}$ generate $A_{\mathfrak{p}}$ as an $R_{\mathfrak{p}}$ -algebra if and only if $(A/B)_{\mathfrak{p}} = 0$. Since the support of a finitely generated R -module is closed, the result follows. \square

Corollary 2.11. *For any positive integer k the set*

$$U_k = \{\mathfrak{p} \in \text{Spec } R : A_{\mathfrak{p}} \text{ can be generated by } k \text{ elements as an } R_{\mathfrak{p}}\text{-algebra}\}$$

is open.

Proof. Suppose that $A_{\mathfrak{p}}$ is generated by k elements as an $R_{\mathfrak{p}}$ -algebra. We may choose elements a_1, \dots, a_k in A which generate A at \mathfrak{p} . By Lemma 2.10, there is an open neighborhood of \mathfrak{p} such that a_1, \dots, a_k generate A at \mathfrak{q} for each \mathfrak{q} in this neighborhood. This shows that U_k is open. \square

Proposition 2.12. *Suppose that $A = \prod_{i=1}^s A_i$ is a product of R -algebras A_1, \dots, A_s such that for any maximal ideal \mathfrak{m} of R and any $i \neq j$ the R/\mathfrak{m} -algebras $A_i \otimes_R R/\mathfrak{m}$ and $A_j \otimes_R R/\mathfrak{m}$ do not have isomorphic quotients. Then $\text{Gen}_k(A) = \prod_{i=1}^s \text{Gen}_k(A_i)$ under the natural identifications.*

Proof. The proposition says that a sequence a_1, \dots, a_k of elements in A generates A as an R -algebra if and only if for every i the projection of these sequence to A_i generates A_i as an R -algebra. The implication to the right is clear. Since $A \otimes_R R/\mathfrak{m} = \prod_{i=1}^s (A_i \otimes_R R/\mathfrak{m})$, Lemma 2.6 reduces the proof to the case when R is a field. Suppose that a sequence a_1, \dots, a_k of elements in A has the property

that for every i the projection of these sequence to A_i generates A_i as an R -algebra. Let B be the R -subalgebra of A generated by a_1, \dots, a_k . By our assumption, the projection $\pi_i : B \rightarrow A_i$ is surjective. Let $J_i = \ker \pi_i$. Since A_i and A_j have no isomorphic quotients for $i \neq j$, we conclude that $J_i + J_j = B$ for $i \neq j$ (for otherwise, $J = J_i + J_j$ would be a proper ideal of B and B/J would be isomorphic to a quotient of A_i and a quotient of A_j). The Chinese remainder theorem implies now that $B = A$. \square

Example 2.13. Let $A_i = M_{n_i}(R)^{m_i}$ be the product of m_i copies of the $n_i \times n_i$ matrix ring over R , where $n_i \neq n_j$ for $i \neq j$. Then for any maximal ideal \mathfrak{m} of R we have $A_i \otimes_R R/\mathfrak{m} = M_{n_i}(R/\mathfrak{m})^{m_i}$. Consider two distinct indices i, j . If the R/\mathfrak{m} -algebras $A_i \otimes_R R/\mathfrak{m}$ and $A_j \otimes_R R/\mathfrak{m}$ had isomorphic quotients, they would have isomorphic quotients which are simple R/\mathfrak{m} -algebras. Clearly any simple quotient of $M_{n_i}(R/\mathfrak{m})^{m_i}$ is isomorphic to $M_{n_i}(R/\mathfrak{m})$. Since $M_{n_i}(R/\mathfrak{m})$ and $M_{n_j}(R/\mathfrak{m})$ are not isomorphic (they have different dimensions over R/\mathfrak{m}), we see that the R/\mathfrak{m} -algebras $A_i \otimes_R R/\mathfrak{m}$ and $A_j \otimes_R R/\mathfrak{m}$ do not have isomorphic quotients. Therefore the assumptions of Proposition 2.12 are satisfied and

$$\text{Gen}_k \left(\prod_{i=1}^s M_{n_i}(R)^{m_i} \right) = \prod_{i=1}^s \text{Gen}_k (M_{n_i}(R)^{m_i}).$$

Recall that in Definition 1.3 we defined $\text{gen}_m(A, R)$ as the largest k such that A^k admits m generators as an R -algebra. The following proposition implies that if $\text{gen}_m(A, R)$ is finite then $\text{gen}_{m+1}(A, R) > \text{gen}_m(A, R)$.

Proposition 2.14. *Let A be an R -algebra and let n be a positive integer. If A^n can be generated by m elements as an R -algebra then A^{n+1} can be generated by $m + 1$ elements.*

Proof. Let $a_i = (a_{i,1}, \dots, a_{i,n})$, with $i = 1, \dots, m$, generate A^n . Let $b_i = (a_{i,1}, \dots, a_{i,n}, a_{i,1})$, $i = 1, \dots, m$, and set $b_{m+1} = (0, \dots, 0, 1)$. For any $w = (w_1, \dots, w_n) \in A^n$ there is a noncommutative polynomial $p(x_1, \dots, x_m)$ with coefficients in R such that $w = p(a_1, \dots, a_m)$. Then $p(b_1, \dots, b_m) = (w_1, \dots, w_m, w_1)$. It follows that $b_{m+1}p(b_1, \dots, b_m) = (0, \dots, 0, w_1)$ and $p(b_1, \dots, b_m) - b_{m+1}p(b_1, \dots, b_m) = (w_1, \dots, w_m, 0)$. Thus the algebra generated by b_1, \dots, b_{m+1} coincides with A^{n+1} . \square

Corollary 2.15. *Let A be an R -algebra. If $\text{gen}_m(A, R)$ is finite then*

$$r(A^{1+\text{gen}_m(A,R)}, R) = m + 1.$$

We end this section with a discussion of an effective method of checking if given elements generate an R -algebra A . The key observation is contained in the following simple lemma:

Lemma 2.16. *Let A be an R -algebra generated as an R -module by elements u_1, \dots, u_m and let $k \geq 1$ be an integer. For every monomial $M = M(x_1, \dots, x_k)$ in k noncommuting variables x_1, \dots, x_k there are polynomials $p_j^M(x_{1,1}, \dots, x_{k,m}) \in R[x_{1,1}, \dots, x_{k,m}]$, $j = 1, \dots, m$, such that the degree of each p_j^M does not exceed the degree of M and*

$$M(a_1, \dots, a_k) = \sum_{i=1}^m p_i^M(a_{1,1}, \dots, a_{k,m}) u_i$$

whenever $a_{i,j} \in R$ satisfy $a_i = \sum_{j=1}^m a_{i,j} u_j$.

Proof. There exist elements $c_{i,j,s} \in R$, $1 \leq i, j, s \leq m$, such that $u_i u_j = \sum_{s=1}^m c_{i,j,s} u_s$. Note that these elements are not unique, unless A is a free R -module with basis u_1, \dots, u_m (this is the case we are mainly interested in). We fix some choice of elements $c_{i,j,s}$ and call them the structure constants for A . Furthermore, choose and fix $r_i \in R$, $i = 1, \dots, m$, such that $1 = \sum r_i u_i$. We prove the lemma by induction on the degree of M . If degree of M is 0 then $M = 1$ and we can choose constant polynomials $p_i^M = r_i$. Suppose that the lemma holds for all monomials of degree less than n and let M be a monomial of degree n . Then $M = N x_t$ for some monomial N of degree $n - 1$ and some $t \in \{1, \dots, k\}$. If $a_i = \sum_{j=1}^m a_{i,j} u_j$, with $a_{i,j} \in R$, $1 \leq i \leq k$, then

$$\begin{aligned} M(a_1, \dots, a_k) &= N(a_1, \dots, a_k) \sum_{j=1}^m a_{t,j} u_j \\ &= \left(\sum_{i=1}^m p_i^N(a_{1,1}, \dots, a_{m,k}) u_i \right) \left(\sum_{j=1}^m a_{t,j} u_j \right) \\ &= \sum_{i=1}^m \sum_{j=1}^m p_i^N(a_{1,1}, \dots, a_{m,k}) a_{t,j} \sum_{s=1}^m c_{i,j,s} u_s \\ &= \sum_{s=1}^m \left(\sum_{i=1}^m \sum_{j=1}^m p_i^N(a_{1,1}, \dots, a_{m,k}) a_{t,j} c_{i,j,s} \right) u_s. \end{aligned}$$

This proves that the polynomials

$$p_s^M = \sum_{i=1}^m \sum_{j=1}^m c_{i,j,s} p_i^N x_{t,j}, \quad s = 1, \dots, m,$$

have the required properties. □

Lemma 2.17. *Let A be an R -algebra which is a free R -module with a basis u_1, \dots, u_m and let $k \geq 1$ be an integer. There is a finite set $T \subseteq R[x_{1,1}, \dots, x_{k,m}]$ of polynomials of degree not exceeding m^2 such that for any commutative R -algebra S the following two conditions are equivalent:*

- (i) The elements $a_i = \sum_{j=1}^m a_{i,j} \otimes u_j$, $1 \leq i \leq k$, of $S \otimes_R A$, where $a_{i,j} \in S$, generate $S \otimes_R A$ as an S -algebra.
- (ii) The ideal of S generated by all the values $f(a_{1,1}, \dots, a_{k,m})$, where $f \in T$, coincides with S .

Proof. Consider polynomials p_j^M described in Lemma 2.16. It is clear that the same polynomials (or rather their images in $S[x_{1,1}, \dots, x_{k,m}]$) work for the S algebra $S \otimes_R A$ and its generators $1 \otimes u_1, \dots, 1 \otimes u_m$. Let $\mathcal{M} = \mathcal{M}(x_{i,j})$ be the matrix whose rows are labeled in some way by the monomials M of degree $< m$ in noncommuting variables x_1, \dots, x_k , and whose row with label M is

$$(p_1^M(x_{1,1}, \dots, x_{k,m}), \dots, p_m^M(x_{1,1}, \dots, x_{k,m})).$$

The $m \times m$ minors of \mathcal{M} are polynomials in $R[x_{1,1}, \dots, x_{k,m}]$ of degree $\leq m^2$. Consider the set T of all these minors. Consider elements $a_i = \sum_{j=1}^m a_{i,j} \otimes u_j$ in $S \otimes_R A$, where $a_{i,j} \in S$ and $1 \leq i \leq k$. Let B be the set of all elements of the form $M(a_1, \dots, a_k)$, where M is a monomial of degree $< m$. By Lemmas 2.8 and 2.6, the elements a_1, \dots, a_k generate $S \otimes_R A$ as an S -algebra if and only if for every maximal ideal \mathfrak{m} of S , the image of the set B in $S \otimes_R A/\mathfrak{m}(S \otimes_R A)$ spans the S/\mathfrak{m} -vector space $S \otimes_R A/\mathfrak{m}(S \otimes_R A)$. This is equivalent to saying that the reduction modulo \mathfrak{m} of the matrix $\mathcal{M}(a_{i,j})$ has rank m , which in turn is equivalent to the condition that at least one of the $m \times m$ minors of $\mathcal{M}(a_{i,j})$ does not belong to \mathfrak{m} . Thus the set T of all the $m \times m$ minors of $\mathcal{M}(x_{i,j})$ has the required property. \square

3. The density of the set of ordered k -tuples which generate an algebra

The results of this section arose from our attempt to answer the following question: what is the probability that k random elements of a ring A , whose additive group is free of finite rank, generate A as a ring. Before we answer this question, we need to make it more precise. We will discuss it in a slightly more general context.

Throughout this section K will be a number field of degree d over \mathbb{Q} , with the ring of integers O_K . We work with an order R in K , that is, R is a subring of K which is free of rank d as a \mathbb{Z} -module. We fix an integral basis w_1, \dots, w_d of R over \mathbb{Z} . Any element r of R can be uniquely written as $r = \sum r_i w_i$ with $r_i \in \mathbb{Z}$. For a positive integer N we denote by $R(N)$ the set of all $r \in R$ such that $|r_i| \leq N$ for all i . Clearly $|R(N)| = (2N + 1)^d$.

Let A be an R -algebra which is free of finite rank m as an R -module. Fix a basis e_1, \dots, e_m of A over R . This choice allows us to identify A and R^m . Using this identification we define $A(N)$ as $R^m(N)$, so $|A(N)| = (2N + 1)^{dm}$. We define the density $\text{den}_k(A)$ of the set of k generators of A as an R -algebra as follows.

Definition 3.1.
$$\text{den}_k(A) = \lim_{N \rightarrow \infty} \frac{|\text{Gen}_k(A) \cap A(N)^k|}{(2N + 1)^{dmk}}.$$

At the moment it is not clear whether the limit on the right in the last formula exists. We will show, however, that it exists and is independent of the choice of an integral basis of R and the choice of a basis of A over R .

Consider a maximal ideal \mathfrak{p} of R . We denote by $\mathbb{F}_{\mathfrak{p}}$ the field R/\mathfrak{p} and by $N(\mathfrak{p})$ its cardinality. Recall that we say that elements a_1, \dots, a_k of A generate A at \mathfrak{p} if their images in $A \otimes_R \mathbb{F}_{\mathfrak{p}}$ generate $A \otimes_R \mathbb{F}_{\mathfrak{p}}$ as an $\mathbb{F}_{\mathfrak{p}}$ -algebra. Let $g_k(\mathfrak{p}, A)$ be the cardinality of the set $\text{Gen}_k(A \otimes_R \mathbb{F}_{\mathfrak{p}})$. In other words, $g_k(\mathfrak{p}, A)$ is the number of k -tuples of elements of $A \otimes_R \mathbb{F}_{\mathfrak{p}}$ which generate $A \otimes_R \mathbb{F}_{\mathfrak{p}}$ as an $\mathbb{F}_{\mathfrak{p}}$ -algebra. It is not hard to see that the density of the set $\text{Gen}_k(\mathfrak{p}, A)$ of all k -tuples in A^k which generate A at \mathfrak{p} is

$$\lim_{N \rightarrow \infty} \frac{|\text{Gen}_k(\mathfrak{p}, A) \cap A(N)^k|}{(2N + 1)^{dmk}} = \frac{g_k(\mathfrak{p}, A)}{N(\mathfrak{p})^{mk}}.$$

Note that by Lemma 2.6, a given k -tuple of elements of A generates it as an R -algebra if and only if it generates A at \mathfrak{p} for every maximal ideal \mathfrak{p} of R . Suppose now that the events “generate at \mathfrak{p} ” are independent for different maximal ideals (we use this notion in a very intuitive sense here, just to motivate our result). It would mean that the probability that random k elements of A generate it as an R -algebra is the product of the numbers $g_k(\mathfrak{p}, A)/N(\mathfrak{p})^{mk}$ over all maximal ideals \mathfrak{p} of R . One of the main results of this section is a rigorous proof that this is indeed true. In other words, we prove the following theorem.

Theorem 3.2. *Let A be an R -algebra which is free of rank m as an R -module and let $k > 0$ be an integer. For a maximal ideal \mathfrak{p} of R denote by $g_k(\mathfrak{p}, A)$ the number of k -tuples of elements of $A \otimes_R \mathbb{F}_{\mathfrak{p}}$ which generate $A \otimes_R \mathbb{F}_{\mathfrak{p}}$ as an $\mathbb{F}_{\mathfrak{p}}$ -algebra. Then*

$$\text{den}_k(A) = \prod_{\mathfrak{p} \in m\text{-Spec } R} \frac{g_k(\mathfrak{p}, A)}{N(\mathfrak{p})^{mk}}. \tag{2}$$

This result establishes in particular the existence and independence of all the choices of the limit defining the quantity $\text{den}_k(A)$.

We will derive Theorem 3.2 as a consequence of a more general result. To this end consider the set $T = \{f_1, \dots, f_s\}$ of polynomials in $R[x_{1,1}, \dots, x_{k,m}]$ established in Lemma 2.17 (under our choice of a basis of A over R). We identify A^k with the set R^{mk} so that a tuple $(a_1, \dots, a_k) \in A^k$ corresponds to $(a_{i,j}) \in R^{mk}$ if and only if $a_i = \sum_{j=1}^m a_{i,j} e_j$. Note that according to Lemma 2.17, the element $a = (a_{i,j}) \in R^{mk}$ corresponds to a k -tuple in $\text{Gen}_k(A)$ if and only if the ideal of R generated by the elements $f_1(a), \dots, f_s(a)$ coincides with R . Moreover, a corresponds to a k -tuple which generates A at \mathfrak{p} if and only if $f_i(a) \notin \mathfrak{p}$ for some i . It follows that $g_k(\mathfrak{p}, A) = N(\mathfrak{p})^{mk} - t_{\mathfrak{p}}$, where $t_{\mathfrak{p}}$ is the number of solutions to $f_1 = \dots = f_s = 0$ over the finite field $\mathbb{F}_{\mathfrak{p}}$. It is clear now that Theorem 3.2 is a consequence of the following result.

Theorem 3.3. *Let R be an order in a number field K and let $T = \{f_1, \dots, f_s\} \subset R[x_1, \dots, x_n]$ be a finite set of polynomials. Define*

$$S = S(T) = \{x = (x_1, \dots, x_n) \in R^n : \text{the ideal generated by } f_1(x), \dots, f_s(x) \text{ is } R\}.$$

For each maximal ideal \mathfrak{p} of R let $t_{\mathfrak{p}}$ be the number of solutions to $f_1 = \dots = f_s = 0$ over the finite field $\mathbb{F}_{\mathfrak{p}} = R/\mathfrak{p}$. For a positive integer N let $S_N = S_N(T) = \{x \in S : x_i \in R(N), i = 1, 2, \dots, n\}$. Then

$$\lim_{N \rightarrow \infty} \frac{|S_N|}{(2N + 1)^{dn}} = \prod_{\mathfrak{p} \in \text{m-Spec } R} \left(1 - \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n}\right). \tag{3}$$

A proof of Theorem 3.3 is given in the next section. Note that for $s = 2$, $R = \mathbb{Z}$, and polynomials f_1 and f_2 that do not have a nonconstant common factor this result was proved in [Poonen 2003] in a slightly more general form (there the limit is taken over boxes whose sides all increase to infinity; here we only deal with boxes which are cubes). Poonen’s result was inspired by [Ekedahl 1991], where a similar result has been established. Arnold [2009] considers the set of pairs of relatively prime integers as a subset of \mathbb{Z}^2 and proves that its density can be computed by using sets of the form nG , where G is any polygon (so our case corresponds to G being the square $|x| \leq 1, |y| \leq 1$). He calls subsets of \mathbb{Z}^2 (or, more generally, of \mathbb{Z}^n) which have this property *uniformly distributed*. In a subsequent paper we will discuss uniform distribution of sets of the type $S(T)$.

We end this section with an application of our theorems.

3A. The probability that k random elements generate the group \mathbb{Z}^n . In his work on generic properties of one-relator groups Ilya Kapovich was led to the following question: what is the probability that several randomly chosen elements generate the group \mathbb{Z}^n . Even though there is a fairly simple heuristic argument which leads to an answer, neither Kapovich nor we have been able to find a reference containing a proof. The techniques developed in this paper allow us, in particular, to give a rigorous answer to Kapovich’s question. The key observation is contained in the following lemma.

Lemma 3.4. *Let V be an n -dimensional vector space over the finite field \mathbb{F}_q . The number $\alpha_{m,n} = \alpha_{m,n}(q)$ of m -tuples of elements in V that span V is equal to $\prod_{i=0}^{n-1} (q^m - q^i)$.*

Proof. For $m < n$ the formula is obviously true as it yields 0 and there are no m -tuples which span V . The number $\alpha_{n,n}$ equals the number of bases of V , which is well known to be equal to $|\text{GL}_n(\mathbb{F}_q)| = \prod_{i=0}^{n-1} (q^n - q^i)$. This establishes the result for $m = n$. Note now that v_1, \dots, v_m span V if and only if the images of v_2, \dots, v_m in $V/\langle v_1 \rangle$ span $V/\langle v_1 \rangle$. Given $v \in V$, we count the number of m -tuples

which span V and start with $v_1 = v$. If $v = 0$ this number is clearly $\alpha_{m-1,n}$. If $v \neq 0$, then there are $\alpha_{m-1,n-1}$ $(m - 1)$ -tuples which span $V/\langle v \rangle$ and each such tuple lifts to q^{m-1} $(m - 1)$ -tuples from V . Thus we get $q^{m-1}\alpha_{m-1,n-1}$ m -tuples which span V and start at v . Since there are $q^n - 1$ nonzero elements in V , we get the following recursive formula:

$$\alpha_{m,n} = \alpha_{m-1,n} + q^{m-1}(q^n - 1)\alpha_{m-1,n-1}.$$

The recursive formula and a straightforward induction on $m + n$ finish the proof. \square

Theorem 3.5. *Let R be an order in a number field. Define*

$$\zeta_R(s) = \prod_{\mathfrak{p} \in m\text{-Spec}(R)} (1 - |R/\mathfrak{p}|^{-s})^{-1}.$$

For any $k \geq n$ the density of the set of k -tuples that generate the R -module R^n is equal to

$$\prod_{m=k-n+1}^k \zeta_R(m)^{-1}.$$

Proof. Consider R^n as an R -algebra with trivial multiplication and let A be obtained from R^n by the construction of adjunction of unity (in the category of R -algebras). By Lemma 2.3 we see that the density of the set of k -tuples which generate the R -module R^n is the same as the density $\text{den}_k(A)$ of the set of k -tuples which generate the R -algebra A . By Lemmas 2.3 and 2.4, we have $g_k(\mathfrak{p}, A) = N(\mathfrak{p})^k \alpha_{k,n}(N(\mathfrak{p}))$. By Theorem 3.2 and Lemma 3.4 we obtain the formula

$$\text{den}_k(A) = \prod_{\mathfrak{p} \in m\text{-Spec } R} \frac{\prod_{i=0}^{n-1} (N(\mathfrak{p})^k - N(\mathfrak{p})^i)}{N(\mathfrak{p})^{nk}} = \prod_{m=k-n+1}^k \zeta_R(m)^{-1}. \quad \square$$

The answer to Ilya Kapovich’s question is therefore given by the following corollary.

Corollary 3.6. *The probability that k randomly chosen elements generate the group \mathbb{Z}^n is equal to $\prod_{m=k-n+1}^k \zeta(m)^{-1}$, where ζ is the Riemann zeta function.*

This corollary can also be derived directly from Theorem 3.3.

4. Proof of Theorem 3.3

Let us start by recalling some of the notation set down in the previous section. R is an order in a number field K . The degree of K over \mathbb{Q} is d and O_K is the ring of integers of K (that is, the integral closure of R in K). We fix an integral basis w_1, \dots, w_d of R over \mathbb{Z} . Any element r of R can be uniquely written as $r = \sum_{i=1}^d r_i w_i$ with $r_i \in \mathbb{Z}$. For a positive integer N we denote by $R(N)$ the set of all $r \in R$ such that $|r_i| \leq N$ for all i . Clearly $|R(N)| = (2N + 1)^d$. The norm

map from K to \mathbb{Q} is denoted by $N_{K/\mathbb{Q}}$. For an ideal I of R we set $N_{K/\mathbb{Q}}(I)$ for the nonnegative integer which is the greatest common divisor of the norms of all elements in I . We write $N(I)$ for the cardinality of R/I . If \mathfrak{p} is a maximal ideal of R then we write $\mathbb{F}_{\mathfrak{p}}$ for the field R/\mathfrak{p} .

The following lemma is well known but for the readers convenience we include a short proof.

Lemma 4.1. *Let F be a finite field with q elements and let $f(x_1, \dots, x_n)$ be a nonzero polynomial in $F[x_1, \dots, x_n]$. Then the number of solutions of the equation $f(x_1, \dots, x_n) = 0$ in F^n does not exceed $(\deg f)q^{n-1}$.*

Proof. We proceed by induction on n . For $n = 1$ this is just the statement that a polynomial f in one variable over a field has at most $\deg f$ roots. Suppose now that the result holds for polynomials in less than n variables and let $f(x_1, \dots, x_n) = \sum_{i=0}^d f_i(x_1, \dots, x_{n-1})x_n^i$ be a polynomial in n variables with $f_d \neq 0$. By the inductive assumption, the number of solutions to $f_d = 0$ in F^n does not exceed $(\deg f_d) \cdot q^{n-2} \cdot q = (\deg f_d)q^{n-1}$. For each $(a_1, \dots, a_{n-1}) \in F^{n-1}$ such that $f_d(a_1, \dots, a_{n-1}) \neq 0$ there are at most d solutions to $f(a_1, \dots, a_{n-1}, x_n) = 0$. Thus we have at most $(\deg f_d)q^{n-1} + dq^{n-1} \leq (\deg f)q^{n-1}$ solutions to $f = 0$. \square

Proposition 4.2. *Let $f(x_1, \dots, x_n) \in \mathbb{Z}[x_1, \dots, x_n]$ be a nonzero polynomial. Set*

$$Z(f, N) = \{(x_1, \dots, x_n) \in \mathbb{Z}^n : |x_i| \leq N \text{ and } f(x_1, \dots, x_n) = 0\}.$$

Then $|Z(f, N)| \leq (\deg f)(2N + 1)^{n-1}$.

Proof. We proceed by induction on n . For $n = 1$ the result is straightforward. Now assume the results for polynomials in less than n variables and consider a polynomial $f(x_1, \dots, x_n) = \sum_{i=0}^d f_i(x_1, \dots, x_{n-1})x_n^i$ in n variables with $f_d \neq 0$. By the inductive assumption, the number of elements in $Z(f, N)$ for which $f_d = 0$ does not exceed $(\deg f_d) \cdot (2N + 1)^{n-2} \cdot (2N + 1)$. For each (a_1, \dots, a_{n-1}) such that $f_d(a_1, \dots, a_{n-1}) \neq 0$ there are at most d solutions to $f(a_1, \dots, a_{n-1}, x_n) = 0$. Thus we have at most $(\deg f_d)(2N + 1)^{n-1} + d(2N + 1)^{n-1} \leq (\deg f)(2N + 1)^{n-1}$ elements in $Z(f, N)$. \square

Corollary 4.3. *Let $f \in R[x_1, \dots, x_n]$ be a polynomial of positive degree $\deg f > 0$. For a nonzero ideal J of R define*

$$I(f, J, N) = \{(x_1, \dots, x_n) \in R(N)^n : J \subseteq f(x_1, \dots, x_n)R\}.$$

Then $|I(f, J, N)| \leq \delta(J)d(\deg f)(2N + 1)^{dn-1}$, where $\delta(J)$ is the number of integral divisors of the norm $N_{K/\mathbb{Q}}(J)$.

Proof. Write $x_i = \sum_{j=1}^d y_{i,j} w_j$. If $J \subseteq f(x_1, \dots, x_n)R$ then $N_{K/\mathbb{Q}}(f(x_1, \dots, x_n))$ divides $N_{K/\mathbb{Q}}(J)$. There is a polynomial $g(y_{i,j}) \in \mathbb{Z}[y_{1,1}, y_{1,2}, \dots, y_{n,d}]$ of degree $(\deg f)d$ in dn variables such that

$$N_{K/\mathbb{Q}}(f(x_1, \dots, x_n)) = g(y_{i,j}).$$

The result follows now from Proposition 4.2 applied to each of the polynomials $g - k$, where k varies over all divisors of $N_{K/\mathbb{Q}}(J)$. \square

Theorem 4.4. *Let $f \in R[x_1, \dots, x_n]$ be a polynomial of positive degree. For each maximal ideal \mathfrak{p} of R let $f_{\mathfrak{p}}$ be the number of solutions to $f = 0$ over the finite field $\mathbb{F}_{\mathfrak{p}}$. Then the series $\sum_{\mathfrak{p} \in m\text{-Spec } R} f_{\mathfrak{p}}/N(\mathfrak{p})^n$ diverges.*

Proof. Replacing R by O_K changes only a finite number of terms in the sum $\sum_{\mathfrak{p}} f_{\mathfrak{p}}/N(\mathfrak{p})^n$. It suffices then to prove the theorem under the additional assumption that $R = O_K$.

Let L be a number field containing K , with ring of integers S , and such that f has an absolutely irreducible divisor $g \in S[x_1, \dots, x_n]$ of positive degree (so $f/g \in L[x_1, \dots, x_n]$). It is known that the reduction of g modulo all but a finite number of prime ideals of S is absolutely irreducible (see [Schmidt 1976, V.2]). For a maximal ideal P of S let g_P be the number of solutions to $g = 0$ in $(S/P)^n$. By [Schmidt 1976, V, Theorem 5A], we have

$$g_P \geq \frac{1}{2} |S/P|^{n-1} = \frac{1}{2} N(P)^{n-1}$$

provided the reduction of g modulo P is absolutely irreducible and $N(P)$ is sufficiently large, which holds for all but a finite number of maximal ideals of S .

Let Φ be the set of maximal ideals of S which have inertia degree one over R and let Ψ be the set of all prime ideals of R which lie under the ideals of Φ . Let $P \in \Phi$ be a prime ideal of S over $\mathfrak{p} \in \Psi$. Then $S/P = \mathbb{F}_{\mathfrak{p}}$. It follows that $f_{\mathfrak{p}} \geq g_P$ except possibly for a finite number of P for which f/g is not P -integral. Since each maximal ideal of R lies under at most $[L : K]$ prime ideals of S we get that

$$\sum_{\mathfrak{p} \in m\text{-Spec } R} \frac{f_{\mathfrak{p}}}{N(\mathfrak{p})^n} \geq \sum_{\mathfrak{p} \in \Psi} \frac{f_{\mathfrak{p}}}{N(\mathfrak{p})^n} \geq \frac{1}{[L : K]} \sum_{\substack{P \in \Phi \\ N(P) \gg 0}} \frac{g_P}{N(P)^n} \geq \frac{1}{2[L : K]} \sum_{\substack{P \in \Phi \\ N(P) \gg 0}} \frac{1}{N(P)}.$$

It is well known that the set Φ has Dirichlet density equal to 1 [Narkiewicz 1990, 7.2, Corollary 3]; in particular $\sum_{P \in \Phi} 1/N(P)$ diverges. \square

Corollary 4.5. *Under the assumptions of Theorem 3.3, if the polynomials in T have a common divisor of positive degree in $K[x_1, \dots, x_n]$ then both sides of (3) are 0. In particular, Theorem 3.3 is true in this case.*

Proof. Let $f \in R[x_1, \dots, x_n]$ be a polynomial of positive degree which divides all f_i in the ring $K[x_1, \dots, x_n]$. There is a nonzero a in R such that af_i/f is in $R[x_1, \dots, x_n]$ for all i . It follows that $S_N(T) \subseteq I(f, aR, N)$. By Corollary 4.3,

$$\frac{|S_N|}{(2N + 1)^{dn}} \leq \frac{|I(f, aR, N)|}{(2N + 1)^{dn}} \leq \frac{\delta(aR)d(\deg f)}{2N + 1},$$

so the left-hand side of (3) is 0.

For any maximal ideal \mathfrak{p} of R which does not divide a we have $t_{\mathfrak{p}} \geq f_{\mathfrak{p}}$. It follows from Theorem 4.4 that $\sum_{\mathfrak{p} \in \text{m-Spec } R} t_{\mathfrak{p}}/N(\mathfrak{p})^n$ diverges. This is equivalent to the right-hand side of (3) being 0. \square

Lemma 4.6. *Let \mathfrak{p} be a maximal ideal of R with $N(\mathfrak{p}) = p^s$, where p is the characteristic of $\mathbb{F}_{\mathfrak{p}}$. Then any element of $\mathbb{F}_{\mathfrak{p}}$ lifts to at most $(2N + 1)^{d-s}(1 + 2N/p)^s$ elements in $R(N)$.*

Proof. We may assume (after renumbering, if necessary) that w_1, \dots, w_s is a basis of $\mathbb{F}_{\mathfrak{p}}$ over \mathbb{F}_p . Consider a residue class $a \in \mathbb{F}_{\mathfrak{p}}$. To get an element $\sum_{i=1}^d y_i w_i \in R(N)$ in the given residue class a we may choose arbitrarily integers y_{s+1}, \dots, y_d in $[-N, N]$ and then the residue classes of y_1, \dots, y_s modulo p are uniquely determined. Thus each $y_i, i \leq s$, can be chosen in at most $1 + 2N/p$ ways. \square

Lemma 4.7. *Let $f \in R[x_1, \dots, x_{n-1}]$, $g = g_0x_n^k + \dots + g_k \in R[x_1, \dots, x_n]$, where $g_0, \dots, g_k \in R[x_1, \dots, x_{n-1}]$ and $g_0 \neq 0$. Consider the set*

$$D(N) = \left\{ (x_1, \dots, x_n) \in R(N)^n : f(x_1, \dots, x_{n-1}) \neq 0 \text{ and there exists a maximal ideal } \mathfrak{p} \text{ with } N(\mathfrak{p}) > N \text{ and such that } f(x_1, \dots, x_{n-1}) \in \mathfrak{p}, \right. \\ \left. g(x_1, \dots, x_n) \in \mathfrak{p}, \text{ and } g_0(x_1, \dots, x_{n-1}) \notin \mathfrak{p} \right\}.$$

Then there is a constant c such that $|D(N)| \leq c(2N + 1)^{dn-1}$ for all N .

Proof. There are positive integers w and c_1 such that

$$|N_{K/\mathbb{Q}} f(x_1, \dots, x_{n-1})| \leq c_1 N^w,$$

for any $N \geq 1$ and any $x_i \in R(N), i = 1, \dots, n - 1$. If $N > c_1$ and $f(x_1, \dots, x_{n-1})$ is nonzero, then $f(x_1, \dots, x_{n-1})$ belongs to at most w maximal ideals \mathfrak{p} such that $N(\mathfrak{p}) > N$. In fact, if there were more than w maximal ideals in R with norm exceeding N which contain $f(x_1, \dots, x_{n-1})$ then $f(x_1, \dots, x_{n-1})$ would belong to at least $w + 1$ maximal ideals of O_K of norm exceeding N and this would imply that $|N_{K/\mathbb{Q}} f(x_1, \dots, x_{n-1})| > N^{w+1}$, which is not possible. Let

$$G(N, \mathfrak{p}) = \left\{ (x_1, \dots, x_{n-1}) \in R(N)^{n-1} : f(x_1, \dots, x_{n-1}) \in \mathfrak{p} - \{0\} \right\}.$$

Thus, if $N > c_1$ and $(x_1, \dots, x_{n-1}) \in R(N)^{n-1}$, then there are at most w maximal ideals \mathfrak{p} such that $N(\mathfrak{p}) > N$ and $(x_1, \dots, x_{n-1}) \in G(N, \mathfrak{p})$. Let $N > c_1$. Fix a

point $(x_1, \dots, x_{n-1}) \in R(N)^{n-1}$ and let \mathfrak{p} be a maximal ideal such that $N(\mathfrak{p}) > N$ and $(x_1, \dots, x_{n-1}) \in G(N, \mathfrak{p})$. We want to find an upper bound for the number of $x_n \in R(N)$ such that $g(x_1, \dots, x_n) \in \mathfrak{p}$ and $g_0(x_1, \dots, x_{n-1}) \notin \mathfrak{p}$. All such x_n split into at most k residue classes modulo \mathfrak{p} (which correspond to the roots of $g(x_1, \dots, x_{n-1}, x) = 0$ in $\mathbb{F}_{\mathfrak{p}}$). Let $N(\mathfrak{p}) = p^s$, where $p = \text{char } F_{\mathfrak{p}}$. By Lemma 4.6, the number of $x_n \in R(N)$ which belong to a given residue class modulo \mathfrak{p} is at most

$$(2N+1)^{d-s} \max(2, 2(2N+1)/p)^s \leq \max(2^s(2N+1)^{d-s}, 2^s(2N+1)^s/p^s) \\ \leq 3 \cdot 2^s \cdot (2N+1)^{d-1}$$

(we have used the inequalities $1 + 2N/p \leq \max(2, 2(2N+1)/p)$ and $p^s > N \geq (2N+1)/3$). It follows that there are at most $w \cdot k \cdot 3 \cdot 2^s(2N+1)^{d-1}$ values of $x_n \in R(N)$ such that $(x_1, \dots, x_n) \in D(N)$. Hence, if $N > c_1$, then

$$|D(N)| \leq (2N+1)^{d(n-1)} \cdot w \cdot k \cdot 3 \cdot 2^s \cdot (2N+1)^{d-1} \leq c(2N+1)^{dn-1},$$

where $c = 3 \cdot 2^s \cdot w \cdot k$. We can increase c if necessary so that the inequality $|D(N)| \leq c(2N+1)^{dn-1}$ holds for all N . \square

Lemma 4.8. *Let f be a nonzero polynomial in $R[x_1, \dots, x_{n-1}]$ and let*

$$g = g_0x_n^k + \dots + g_k \in R[x_1, \dots, x_n],$$

where $g_0, \dots, g_k \in R[x_1, \dots, x_{n-1}]$, $g_0 \neq 0$. For a maximal ideal \mathfrak{p} of R consider the set

$$D_{\mathfrak{p}}(N) = \{(x_1, \dots, x_n) \in R(N)^n : f(x_1, \dots, x_{n-1}) \in \mathfrak{p}, \\ g(x_1, \dots, x_n) \in \mathfrak{p}, \text{ and } g_0(x_1, \dots, x_{n-1}) \notin \mathfrak{p}\}.$$

Then, if $N(\mathfrak{p}) \leq N$ and the reduction of f modulo \mathfrak{p} is not zero, we have

$$|D_{\mathfrak{p}}(N)| \leq 2^{nd} (\deg f) k (2N+1)^{nd} / N(\mathfrak{p})^2.$$

Proof. The image $Z_{\mathfrak{p}}$ of $D_{\mathfrak{p}}(N)$ in $\mathbb{F}_{\mathfrak{p}}^n$ consists of (some) solutions to $f=0=g$ in $\mathbb{F}_{\mathfrak{p}}^n$ (we use the same notation for a polynomial and its reduction modulo \mathfrak{p}). Now $f=0$ has at most $(\deg f) N(\mathfrak{p})^{n-2}$ solutions in $\mathbb{F}_{\mathfrak{p}}^{n-1}$ (Lemma 4.1) and each such solution extends to at most k solutions of $g=0$, $g_0 \neq 0$ in $\mathbb{F}_{\mathfrak{p}}^n$. Thus $|Z_{\mathfrak{p}}| \leq (\deg f) k N(\mathfrak{p})^{n-2}$. Each element of $Z_{\mathfrak{p}}$ lifts to no more than $[(2N+1)^{d-s}(1+2N/p)^s]^n$ elements of $D_{\mathfrak{p}}(N)$ by Lemma 4.6, where $N(\mathfrak{p}) = p^s$. Thus

$$|D_{\mathfrak{p}}(N)| \leq (\deg f) k N(\mathfrak{p})^{n-2} [(2N+1)^{d-s}(1+2N/p)^s]^n \\ \leq (\deg f) k N(\mathfrak{p})^{n-2} (2N+1)^{n(d-s)} [2^s(2N+1)^s/p^s]^n \\ \leq 2^{nd} (\deg f) k N(\mathfrak{p})^{n-2} (2N+1)^{nd} / N(\mathfrak{p})^n \\ = 2^{nd} (\deg f) k (2N+1)^{nd} / N(\mathfrak{p})^2. \quad \square$$

Proposition 4.9. *Let $f, g \in R[x_1, \dots, x_n]$ be polynomials which are relatively prime as polynomials in $K[x_1, \dots, x_n]$. Define*

$$W_M = W_M(f, g) = \{\mathbf{r} = (r_1, \dots, r_n) \in R^n : \text{there is a maximal ideal } \mathfrak{p} \text{ of } R, \\ \text{with } N(\mathfrak{p}) > M \text{ and such that } f(\mathbf{r}) \in \mathfrak{p} \text{ and } g(\mathbf{r}) \in \mathfrak{p}\}.$$

There is a constant $c > 0$ such that

$$|W_M \cap R(N)^n| \leq c \frac{(2N + 1)^{nd}}{M}$$

for any integers $N > M \geq 1$.

Proof. We use induction on the number n of variables. Note that if f and g are polynomials in n variables for which the result holds, then it also holds for f and g considered as polynomials in $n + 1$ variables. When $n = 0$ the result is clear. Suppose the result is true for less than $n \geq 1$ variables. Consider two relatively prime (in $K[x_1, \dots, x_n]$) polynomials $f, g \in R[x_1, \dots, x_n]$.

The first step is to establish the proposition under the additional assumption that f is irreducible in $K[x_1, \dots, x_n]$ and does not depend on x_n (that is, f is in $R[x_1, \dots, x_{n-1}]$). Let $g = g_0x_n^k + \dots + g_k$, where $g_0, \dots, g_k \in R[x_1, \dots, x_{n-1}]$, $g_0 \neq 0$. We fix f and proceed by induction on the degree k of g in x_n . If $k = 0$ then $g \in R[x_1, \dots, x_{n-1}]$ and the result follows by our inductive assumption that the proposition holds for polynomials in $n - 1$ variables. Suppose that $k > 0$ and the result holds for all polynomials g whose degree in x_n is less than k (and which are relatively prime to f). We may write $ag = \prod h_i$ for some nonzero a in R and polynomials $h_i \in R[x_1, \dots, x_n]$ which are irreducible in $K[x_1, \dots, x_n]$. Note that $W_M(f, g) \subseteq \bigcup W_M(f, h_i)$. Thus, if we show the proposition for each pair f, h_i , then it will also hold for the pair f, g . In other words, we may assume that g is irreducible in $K[x_1, \dots, x_n]$. If $f \mid g_0$ in $K[x_1, \dots, x_{n-1}]$, then there is a nonzero $u \in R$ such that $f \mid ug_0$ in $R[x_1, \dots, x_{n-1}]$. It follows that

$$W_M(f, g) \subseteq W_M(f, u(g - g_0x_n^k))$$

for all M . Since $u(g - g_0x_n^k)$ has degree in x_n smaller than k , the result holds for $f, u(g - g_0x_n^k)$ by our inductive assumption and therefore it also holds for the pair f, g . Thus we may assume that f does not divide g_0 in $K[x_1, \dots, x_{n-1}]$. Since f is irreducible, f and g_0 are relatively prime in $K[x_1, \dots, x_{n-1}]$. For $N > M$ we have

$$W_M(f, g) \cap R(N)^n \subseteq (W_M(f, g_0) \cap R(N)^n) \cup Z(f, N) \cup D(N) \cup \bigcup_{\mathfrak{p}: M < N(\mathfrak{p}) \leq N} D_{\mathfrak{p}}(N),$$

where

$$Z(f, N) = \{\mathbf{r} = (r_1, \dots, r_n) \in R(N)^n : f(\mathbf{r}) = 0\},$$

$$D(N) = \{ \mathbf{r} \in R(N)^n : \text{there is a maximal ideal } \mathfrak{p} \text{ such that } N(\mathfrak{p}) > N, \\ f(\mathbf{r}) \in \mathfrak{p} - \{0\}, g(\mathbf{r}) \in \mathfrak{p}, \text{ and } g_0(\mathbf{r}) \notin \mathfrak{p} \},$$

$$D_{\mathfrak{p}}(N) = \{ \mathbf{r} \in R(N)^n : f(\mathbf{r}) \in \mathfrak{p}, g(\mathbf{r}) \in \mathfrak{p}, g_0(\mathbf{r}) \notin \mathfrak{p} \}.$$

By our inductive assumption that the proposition holds for polynomials in $n - 1$ variables, there is $c_1 > 0$ such that $|W_M(f, g_0) \cap R(N)^n| \leq c_1(2N + 1)^{dn} / M$ for any integers $N > M \geq 1$. Note that if $f(\mathbf{r}) = 0$ then $(f - 1)(\mathbf{r})R = R$. It follows by Corollary 4.3 applied to the polynomial $f - 1$ and the ideal $J = R$ that

$$|Z(f, N)| \leq \delta(R)d(\deg f)(2N + 1)^{dn-1} \leq c_2 \frac{(2N + 1)^{dn}}{M}$$

for some $c_2 > 0$ and all $N > M \geq 1$. Lemma 4.7 assures the existence of $c_3 > 0$ such that $|D(N)| \leq c_3(2N + 1)^{dn-1} \leq c_3(2N + 1)^{dn} / M$. Finally, by Lemma 4.8, there are constants $c_4 > 0, c_5 > 0$ such that

$$\left| \bigcup_{\mathfrak{p}: M < N(\mathfrak{p}) \leq N} D_{\mathfrak{p}}(N) \right| \leq \sum_{\mathfrak{p}: M < N(\mathfrak{p}) \leq N} |D_{\mathfrak{p}}(N)| \leq \sum_{\mathfrak{p}: M < N(\mathfrak{p}) \leq N} 2^{nd}(\deg f)d \frac{(2N + 1)^{nd}}{N(\mathfrak{p})^2} \\ \leq c_4(2N + 1)^{nd} \sum_{\mathfrak{p}: M < N(\mathfrak{p})} N(\mathfrak{p})^{-2} \leq c_4(2N + 1)^{nd}d \sum_{m > M} \frac{1}{m^2} \leq c_5 \frac{(2N + 1)^{nd}}{M}.$$

It follows that $|W_M(f, g) \cap R(N)^n| \leq c(2N + 1)^{nd} / M$, where $c = c_1 + c_2 + c_3 + c_5$. This completes our first step, that is, establishes the proposition under the additional assumption that f is irreducible in $K[x_1, \dots, x_n]$ and does not depend on x_n .

Our second step is to prove the proposition when both f and g are irreducible in $K[x_1, \dots, x_n]$. Consider f and g as polynomials in x_n with coefficients in $R[x_1, \dots, x_{n-1}]$. If one of these polynomials does not depend on x_n , the proposition holds by our first step. Suppose that the degrees with respect to x_n of both f and g are positive. Let $r = \text{Res}(f, g)$ be the resultant of f and g , so r is a nonzero polynomial in $R[x_1, \dots, x_{n-1}]$. Recall that $r = af + bg$ for some polynomials $a, b \in R[x_1, \dots, x_n]$ (see [Cox et al. 2005, §3.1] for a nice account of properties of resultants). It follows that $W_M(f, g) \subseteq W_M(f, r) \cap W_M(g, r)$. Since f and g are irreducible, g and r have no common factor in $K[x_1, \dots, x_n]$ (otherwise g would not depend on x_n). We may write $ar = \prod r_i$, where $r_i \in R[x_1, \dots, x_{n-1}]$ are irreducible in $K[x_1, \dots, x_{n-1}]$ and $a \in R$ is nonzero. Clearly $W_M(f, g) \subseteq W_M(r, g) \subseteq \bigcup W_M(r_i, g)$. Since the proposition holds for each pair r_i, g by the first step, it also holds for the pair f, g .

Finally, without any additional assumptions, we may write $af = \prod f_i, bg = \prod g_i$, where $f_i, g_j \in R[x_1, \dots, x_n]$ are irreducible in $K[x_1, \dots, x_n]$ and $a, b \in R - \{0\}$. Clearly $W_M(f, g) \subseteq \bigcup W_M(f_i, g_j)$. Since the result holds for each pair f_i, g_j , it also holds for f, g . □

Corollary 4.10. *Let $T = \{f_1, \dots, f_s\}$ be a finite set of polynomials in $R[x_1, \dots, x_n]$ which do not have any common nonconstant divisor in $K[x_1, \dots, x_n]$. Define*

$$W_M = W_M(T) = \{\mathbf{r} = (r_1, \dots, r_n) \in R^n : \text{there is a maximal ideal } \mathfrak{p} \text{ of } R \\ \text{with } N(\mathfrak{p}) > M \text{ and such that } f(\mathbf{r}) \in \mathfrak{p} \text{ for every } f \in T\}.$$

There is a constant $c > 0$ such that $|W_M \cap R(N)^n| \leq c(2N + 1)^{nd}/M$ for any integers $N > M \geq 1$.

Proof. We may write $d_i f_i = \prod f_{i,j}$, where $f_{i,j} \in R[x_1, \dots, x_n]$ are irreducible in $K[x_1, \dots, x_n]$ and $d_i \in R$ are nonzero. Then

$$W_M \subseteq \bigcup W_M(f, g),$$

where the union is over all pairs f, g such that f and g are among the polynomials $f_{i,j}$ and are relatively prime. Thus the result follows by Proposition 4.9. \square

Corollary 4.10 is the main ingredient in our proof of Theorem 3.3. In fact, the proof now reduces to a fairly straightforward application of the inclusion-exclusion formula and the Chinese remainder theorem. For the benefit of the reader we provide a detailed argument.

Lemma 4.11. *Let I be a nonzero ideal of R . If m is a positive integer such that $mR \subseteq I$ then*

$$\frac{(2N - m)^d}{N(I)} \leq |(a + I) \cap R(N)| \leq \frac{(2N + m)^d}{N(I)}$$

for any $a \in R$ and any N such that $2N \geq m$.

Proof. The ideal I is a union of $m^d/N(I)$ cosets of mR . Thus any coset of I is also a union of $m^d/N(I)$ cosets of mR . Any coset H of mR is of the form

$$\sum_{i=1}^d a_i w_i + mR,$$

where $0 \leq a_i < m$. The elements of $H \cap R(N)$ are exactly those of the form $\sum_{i=1}^d (a_i + mb_i)w_i$ with $|a_i + mb_i| \leq N$. Thus $(-N - a_i)/m \leq b_i \leq (N - a_i)/m$. Recall now that an interval of length l has at least $l - 1$ and at most $l + 1$ integers in it. It follows that $(2N/m - 1)^d \leq |H \cap R(N)| \leq (2N/m + 1)^d$. Since $a + I$ is a disjoint union of $m^d/N(I)$ cosets of mR , the result follows. \square

Lemma 4.12. *Let I be a nonzero ideal of R . If V is a subset of $(R/I)^n$ and $V(N)$ is the set of elements of $R(N)^n$ whose image in $(R/I)^n$ belongs to V then*

$$\lim_{N \rightarrow \infty} \frac{|V(N)|}{(2N + 1)^{nd}} = \frac{|V|}{N(I)^n}.$$

Proof. Since both sides of the equality are additive for disjoint unions, it suffices to prove the lemma for sets V which contain only one element. In this case, there are cosets $a_1 + I, \dots, a_n + I$ of I such that

$$V(N) = ((a_1 + I) \cap R(N)) \times \dots \times ((a_n + I) \cap R(N)).$$

There is a positive integer m such that $mR \subseteq I$. By Lemma 4.11, we have

$$\frac{(2N - m)^{dn}}{N(I)^n} \leq |V(N)| \leq \frac{(2N + m)^{dn}}{N(I)^n}$$

provided $2N \geq m$. Dividing by $(2N + 1)^{dn}$ and passing to the limit when $N \rightarrow \infty$, we get the result. □

Proof of Theorem 3.3. If the polynomials in T have a common divisor in $K[x_1, \dots, x_n]$ the theorem holds by Corollary 4.5. Thus we may assume that elements of T do not have any common nonconstant divisor in $K[x_1, \dots, x_n]$. For a prime ideal \mathfrak{p} of R define

$$D_{\mathfrak{p}} = \{\mathbf{r} = (r_1, \dots, r_n) \in R^n : f(\mathbf{r}) \in \mathfrak{p} \text{ for every } f \in T\}.$$

Let Φ be a finite set of maximal ideals of R . For any subset Ψ of Φ we denote by $I(\Psi)$ the intersection of all the ideals in Ψ . Note that

$$D_{\Psi} := \bigcap_{\mathfrak{p} \in \Psi} D_{\mathfrak{p}} = \{\mathbf{r} = (r_1, \dots, r_n) \in R^n : f(\mathbf{r}) \in I(\Psi) \text{ for every } f \in T\}.$$

Let V_{Ψ} be the image of D_{Ψ} in $(R/I(\Psi))^n$. Thus V_{Ψ} is simply the set of all common zeros in $(R/I(\Psi))^n$ of the polynomials in T . By the Chinese remainder theorem, we have $R/I(\Psi) \cong \prod_{\mathfrak{p} \in \Psi} R/\mathfrak{p}$ and under this identification we have $V_{\Psi} = \prod_{\mathfrak{p} \in \Psi} V_{\mathfrak{p}}$. It follows that $|V_{\Psi}| = \prod_{\mathfrak{p} \in \Psi} t_{\mathfrak{p}}$. Applying Lemma 4.12 to the set V_{Ψ} and observing that $V_{\Psi}(N) = D_{\Psi} \cap R(N)^n$ we get

$$\lim_{N \rightarrow \infty} \frac{|D_{\Psi} \cap R(N)^n|}{(2N + 1)^{nd}} = \frac{\prod_{\mathfrak{p} \in \Psi} t_{\mathfrak{p}}}{N(I(\Psi))^n} = \prod_{\mathfrak{p} \in \Psi} \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n}.$$

Let W_{Φ} be the complement of the union $\bigcup_{\mathfrak{p} \in \Phi} D_{\mathfrak{p}}$ in R^n . The inclusion-exclusion principle yields the following formula:

$$|W_{\Phi} \cap R(N)^n| = \sum_{\Psi \subseteq \Phi} (-1)^{|\Psi|} |D_{\Psi} \cap R(N)^n|$$

(where $D_{\emptyset} = R^n$), from which we immediately conclude that

$$\lim_{N \rightarrow \infty} \frac{|W_{\Phi} \cap R(N)^n|}{(2N + 1)^{nd}} = \sum_{\Psi \subseteq \Phi} (-1)^{|\Psi|} \prod_{\mathfrak{p} \in \Psi} \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n} = \prod_{\mathfrak{p} \in \Phi} \left(1 - \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n}\right).$$

Suppose now that Φ is the set of all prime ideals of norm $\leq M$. Note that

$$S(T) \subseteq W_\Phi \subseteq S(T) \cup W_M(T),$$

where $W_M(T)$ is defined in Corollary 4.10 and $S(T)$ in Theorem 3.3. Thus

$$|W_\Phi \cap R(N)^n| - |W_M(T) \cap R(N)^n| \leq |S(T) \cap R(N)^n| \leq |W_\Phi \cap R(N)^n|.$$

Note that Corollary 4.10 implies that

$$0 \leq \liminf_{N \rightarrow \infty} \frac{|W_M(T) \cap R(N)^n|}{(2N + 1)^{dn}} \leq \limsup_{N \rightarrow \infty} \frac{|W_M(T) \cap R(N)^n|}{(2N + 1)^{dn}} \leq \frac{c}{M}.$$

This yields

$$\begin{aligned} \prod_{\mathfrak{p}:N(\mathfrak{p}) \leq M} \left(1 - \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n}\right) - \frac{c}{M} &\leq \liminf_{N \rightarrow \infty} \frac{|S(T) \cap R(N)^n|}{(2N + 1)^{dn}} \\ &\leq \limsup_{N \rightarrow \infty} \frac{|S(T) \cap R(N)^n|}{(2N + 1)^{dn}} \leq \prod_{\mathfrak{p}:N(\mathfrak{p}) \leq M} \left(1 - \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n}\right). \end{aligned}$$

Letting M go to infinity we see that

$$\lim_{N \rightarrow \infty} \frac{|S(T) \cap R(N)^n|}{(2N + 1)^{dn}} = \prod_{\mathfrak{p} \in \mathfrak{m}\text{-Spec } R} \left(1 - \frac{t_{\mathfrak{p}}}{N(\mathfrak{p})^n}\right). \quad \square$$

5. The smallest number of generators

Let us return to our discussion of generators of algebras. We first show an application of Theorem 3.2. Let A be an algebra over a commutative ring R , which is finitely generated as an R -module.

Definition 5.1. We denote by $r = r(A) = r(A, R)$ the smallest number of elements which are needed to generate A as an R -algebra.

For a prime ideal \mathfrak{p} of R define

$$r_{\mathfrak{p}} = r_{\mathfrak{p}}(A) = r(A \otimes_R \kappa(\mathfrak{p}), \kappa(\mathfrak{p})),$$

where $\kappa(\mathfrak{p}) = R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}$ is the field of fractions of R/\mathfrak{p} .

Note that $r_{\mathfrak{p}}$ is the smallest number of generators of $A_{\mathfrak{p}}$ as an $R_{\mathfrak{p}}$ -algebra by Lemma 2.6. Clearly $r_{\mathfrak{p}} \leq r$ for every $\mathfrak{p} \in \text{Spec } R$ and $r_{\mathfrak{p}} \leq r_{\mathfrak{q}}$ whenever $\mathfrak{p} \subseteq \mathfrak{q}$ by Corollary 2.11. The first main result of this section is the following theorem.

Theorem 5.2. *Let R be an order in a number field K and let A be an R -algebra which is free as an R -module. Suppose that $k \geq r_{\mathfrak{p}}$ for all prime ideals \mathfrak{p} of R and $k \geq 1 + r_0$. Then $\text{den}_k(A) > 0$. In particular, $k \geq r$.*

Our proof of Theorem 5.2 will use the following nice result, often called the Loomis–Whitney inequality [Loomis and Whitney 1949].

Lemma 5.3. *Let T be a finite set, D a subset of T^s , and let D_i be the projection of D to T^{s-1} along the i -th coordinate. Then $|D|^{s-1} \leq \prod_{i=1}^s |D_i|$.*

As a corollary we get the following lemma.

Lemma 5.4. *Let \mathbb{F} be a finite field with q elements. Suppose that an m -dimensional \mathbb{F} -algebra A can be generated by r elements as an \mathbb{F} -algebra. For $k \geq r$ let ng_k be the number of k -tuples in A^k which do not generate A as an \mathbb{F} -algebra. Then $\text{ng}_k \leq m^{2k/r} q^{mk-k/r}$ for any $k > r$.*

Proof. Let $D(k) \subseteq A^k$ be the set of all k -tuples which do not generate A as an \mathbb{F} -algebra. For each i the projection $D(k)_i \subseteq A^{k-1}$ of $D(k)$ along the i -th coordinate is contained in $D(k-1)$. By the Loomis–Whitney inequality (Lemma 5.3) we have

$$\text{ng}_k^{k-1} \leq \text{ng}_{k-1}^k.$$

A straightforward induction yields now the inequality

$$\text{ng}_k \leq \text{ng}_r^{k/r}.$$

The set $D(r)$ is contained in the set of all zeros of some nonzero polynomial of degree $\leq m^2$ in rm variables by Lemma 2.17. It follows that $|D(r)| = \text{ng}_r \leq m^2 q^{mr-1}$ by Lemma 4.1. Consequently,

$$\text{ng}_k \leq \text{ng}_r^{k/r} \leq m^{2k/r} q^{km-k/r}. \quad \square$$

Proof of Theorem 5.2. Recall that by Theorem 3.2 we have

$$\text{den}_k(A) = \prod_{\mathfrak{p} \in \text{m-Spec } R} \frac{g_k(\mathfrak{p}, A)}{N(\mathfrak{p})^{mk}},$$

where m is the rank of the free R -module A . Since $k \geq r_{\mathfrak{p}}$, we see that $g_k(\mathfrak{p}, A) > 0$ for all maximal ideals \mathfrak{p} . It suffices therefore to show that

$$\prod \frac{g_k(\mathfrak{p}, A)}{N(\mathfrak{p})^{mk}} > 0,$$

where the product is over all maximal ideals with sufficiently large norm. Since the set of all prime ideals \mathfrak{p} of R such that $r_{\mathfrak{p}} = r_0$ is open and contains the zero ideal, we have $r_{\mathfrak{p}} = r_0$ for all but a finite number of maximal ideals \mathfrak{p} . Since $k \geq r_0 + 1$, Lemma 5.4 implies that $g_k(\mathfrak{p}, A) \geq N(\mathfrak{p})^{km} - m^{2k/r_0} N(\mathfrak{p})^{km-k/r_0}$ for every $\mathfrak{p} \in \text{m-Spec } R$ such that $r_0 = r_{\mathfrak{p}}$. It follows that

$$\frac{g_k(\mathfrak{p}, A)}{N(\mathfrak{p})^{mk}} \geq 1 - \frac{m^{2k/r_0}}{N(\mathfrak{p})^{k/r_0}}$$

for all but a finite number of maximal ideals \mathfrak{p} . It suffices therefore to show that

$$\prod \left(1 - \frac{m^{2k/r_0}}{N(\mathfrak{p})^{k/r_0}} \right) > 0,$$

where the product is over all maximal ideals with sufficiently large norm. This in turn is equivalent to showing that the series

$$\sum_{\mathfrak{p} \in \text{m-Spec } R} \frac{m^{2k/r_0}}{N(\mathfrak{p})^{k/r_0}}$$

converges, which is indeed true since $k/r_0 > 1$. □

As an immediate corollary of Theorem 5.2 we get the following.

Theorem 5.5. *Let R be an order in a number field K and let A be an R -algebra which is free as an R -module. If $r_0 < r_{\mathfrak{p}}$ for some maximal ideal \mathfrak{p} of R then $r = \max\{r_{\mathfrak{p}} : \mathfrak{p} \in \text{m-Spec } R\}$. If $r_0 = r_{\mathfrak{p}}$ for all maximal ideals \mathfrak{p} then $r_0 \leq r \leq 1 + r_0$.*

A special case of Theorem 5.5 when $R = \mathbb{Z}$ was shown to us by H. W. Lenstra (private communication, 2007). His proof of this result is purely algebraic and does not provide any way to handle the ambiguity for r when $r_0 = r_{\mathfrak{p}}$ for all maximal ideals \mathfrak{p} . It is known that in this case both $r = r_0$ and $r = r_0 + 1$ are possible. For example, there are infinitely many number fields in which the ring of integers A considered as a \mathbb{Z} -algebra has $r_{\mathfrak{p}} = 1$ for all prime ideals \mathfrak{p} but $r = 2$. As an explicit example one can take the ring of integers in the cubic field $\mathbb{Q}(\sqrt[3]{198})$ [Pleasant's 1974, p. 167]. Later, we will see examples where $\text{den}_{r_0}(A) > 0$, hence $r = r_0$, even though we are unable to find generators.

Question 5.6. Let R be an order in a number field. Suppose that A is an R -algebra which is finitely generated and projective as an R -module. The right-hand side of the formula in Theorem 3.2 makes perfect sense for A and we will continue to denote it by $\text{den}_k A$. Is it true that if $\text{den}_k A > 0$ then A can be generated by k elements as an R -algebra? We believe that the answer is positive. Perhaps there is a notion of density in this case which makes Theorem 3.2 valid?

We have the following generalization of the original result of Lenstra.

Theorem 5.7. *Let R be a commutative ring of dimension ≤ 1 such that $\text{m-Spec } R$ is Noetherian and let A be an R -algebra finitely generated as an R -module. Let h be the smallest nonnegative integer such that $h \geq r_{\mathfrak{p}}$ for all but a finite number of maximal ideals \mathfrak{p} of R . Suppose that $k \geq r_{\mathfrak{p}}$ for all maximal ideals \mathfrak{p} and $k \geq 1 + h$. Then A can be generated by k elements as an R -algebra.*

Proof. Since $\text{m-Spec } R$ is Noetherian, it has a finite number of irreducible components. Note that if an irreducible component of $\text{m-Spec } R$ is finite then it consists of a single maximal ideal. Otherwise it contains infinitely many maximal ideals

and the intersection of all these ideals is a prime ideal which we call the generic ideal of the component. Let T be the set of all prime ideals which are generic ideals of some infinite irreducible component of $\mathfrak{m}\text{-Spec } R$. Thus T is a finite set of minimal prime ideals of R (it can be empty). Note that if $\mathfrak{p} \in T$ then $r_{\mathfrak{p}} \leq r_{\mathfrak{q}}$ for any maximal ideal \mathfrak{q} containing \mathfrak{p} and the equality holds for all but a finite number of such maximal ideals by Corollary 2.11. It follows that $h = \max\{r_{\mathfrak{p}} : \mathfrak{p} \in T\}$. For each prime $\mathfrak{p} \in T$ choose a maximal ideal $\mathfrak{q} \supseteq \mathfrak{p}$ such that $r_{\mathfrak{p}} = r_{\mathfrak{q}}$ and denote this set of chosen maximal ideals by M .

We call a sequence a_1, \dots, a_m of elements of A M -generic if it generates A at \mathfrak{p} for every $\mathfrak{p} \in M$. Note that an M -generic sequence generates A at \mathfrak{p} for all but a finite number of maximal ideals \mathfrak{p} . We claim that there is an M -generic sequence of length h . Indeed, for each $\mathfrak{q} \in M$ there are h elements in A which generate A at \mathfrak{q} . By the Chinese remainder theorem for modules, we may find elements a_1, \dots, a_h in A which generate A at \mathfrak{q} for all $\mathfrak{q} \in M$. Thus a_1, \dots, a_h is M -generic.

We will now show that for every $i \leq h$ there is an M -generic sequence b_1, \dots, b_h such that for every maximal ideal \mathfrak{q} the elements b_1, \dots, b_i can be completed to a set of k elements which generate A at \mathfrak{q} . Our argument is by induction on i . It is clearly true for $i = 0$ (any M -generic sequence of length h works). Suppose that b_1, \dots, b_h is a generic sequence which works for some i . We seek a generic sequence working for $i + 1$ which is of the form $b_1, \dots, b_i, b, b_{i+2}, \dots, b_h$ for some $b \in A$. Note that if b is such that $b - b_{i+1} \in \mathfrak{q}A$ for all $\mathfrak{q} \in M$ then $b_1, \dots, b_i, b, b_{i+2}, \dots, b_h$ is M -generic. Also, there is a finite set W of maximal ideals, disjoint from M , such that for any maximal ideal $\mathfrak{q} \notin W$ and any $b \in A$, the sequence $b_1, \dots, b_i, b, b_{i+1}, \dots, b_h$ generates A at \mathfrak{q} . Since $k > h$, for any $\mathfrak{q} \notin W$ and any $b \in A$, the elements b_1, \dots, b_i, b can be completed to a set of k elements which generate A at \mathfrak{q} . So in our choice of b we only need to worry about maximal ideals in W . For every $\mathfrak{q} \in W$ there is $b_{\mathfrak{q}} \in A$ such that $b_1, \dots, b_i, b_{\mathfrak{q}}$ extends to a set of k elements which generate A at \mathfrak{q} . By the Chinese remainder theorem for modules, we may choose $b \in A$ such that $b - b_{i+1} \in \mathfrak{q}A$ for all $\mathfrak{q} \in M$ and $b - b_{\mathfrak{q}} \in \mathfrak{q}A$ for all $\mathfrak{q} \in W$. For any such b the sequence $b_1, \dots, b_i, b, b_{i+2}, \dots, b_h$ has the required properties for $i + 1$.

Let a_1, \dots, a_h be an M -generic sequence good for $i = h$. Thus, for any maximal ideal \mathfrak{q} outside some finite set U the elements a_1, \dots, a_h generate A at \mathfrak{q} . For each $\mathfrak{q} \in U$, there are elements $a_{h+1}(\mathfrak{q}), \dots, a_k(\mathfrak{q})$ in A such that $a_1, \dots, a_h, a_{h+1}(\mathfrak{q}), \dots, a_k(\mathfrak{q})$ generate A at \mathfrak{q} . By the Chinese remainder theorem for modules, there are elements a_{h+1}, \dots, a_k in A such that $a_i - a_i(\mathfrak{q}) \in \mathfrak{q}A$ for all $\mathfrak{q} \in U$ and all $i = h + 1, \dots, k$. Thus the elements a_1, \dots, a_k generate A at \mathfrak{q} for every maximal ideal \mathfrak{q} , hence they generate A as an R -algebra by Lemma 2.6. \square

The reader familiar with the results of Forster and Swan on the number of generators of modules over Noetherian commutative rings should recognize the

similarities between Theorem 5.7 and Swan’s theorem [Matsumura 1986, Theorem 5.8]. Unlike the result of Swan, Theorem 5.7 only treats the case of rings of dimension ≤ 1 . So far we have not been able to get similar results for rings of higher dimension but we believe that the following conjectural generalization should be true. In order to state it we need to recall briefly some notions (see [Matsumura 1986, p. 35–37] for more details). We denote by $j\text{-Spec } R$ the subspace of $\text{Spec } R$ which consists of those prime ideals which are intersections of some set of maximal ideals of R . We assume that $m\text{-Spec } R$ is a Noetherian space. It turns out that this is equivalent to $j\text{-Spec } R$ being Noetherian, and then both spaces have the same combinatorial dimension. When $\mathfrak{p} \in j\text{-Spec } R$, we write $j\text{-dim } \mathfrak{p}$ for the combinatorial dimension of the closure of $\{\mathfrak{p}\}$ in $j\text{-Spec } R$. For $\mathfrak{p} \in j\text{-Spec } R$ define

$$b(\mathfrak{p}, A) = \begin{cases} 0 & \text{if } A_{\mathfrak{p}} = 0, \\ j\text{-dim } \mathfrak{p} + r_{\mathfrak{p}}(A) & \text{if } A_{\mathfrak{p}} \neq 0. \end{cases}$$

Conjecture 5.8. *Suppose that R is a commutative ring such that $m\text{-Spec } R$ is a Noetherian space. Let A be an R -algebra finitely generated as an R -module. If $\sup\{b(\mathfrak{p}, A) : \mathfrak{p} \in m\text{-Spec } R\} = n < \infty$ then A can be generated as an R -algebra by n elements.*

6. Generators of matrix algebras over finite fields

It is clear from the results of Section 3 that the key step towards understanding the smallest number of generators of an algebra over a commutative ring is to handle the case of algebras over fields. Among the finite-dimensional algebras over fields the best understood class is the class of separable algebras. It was proved in [Mazur and Petrenko 2009] that any separable algebra over an infinite field is two-generated. This is no longer true over finite fields. In this case, separable algebras coincide with finite products of matrix algebras.

By Proposition 2.12, understanding the structure of generators of a semisimple F -algebra reduces to algebras of the form A^m , where A is a simple F -algebra. We have the following result:

Theorem 6.1. *Let F be a field, A a finite-dimensional simple F -algebra, and k, m, n positive integers. Then k elements of A^m , say $a_1 = (a_{11}, \dots, a_{1m}), \dots, a_k = (a_{k1}, \dots, a_{km})$, generate A^m as an F -algebra if and only if the following two conditions are satisfied:*

- (1) *For any $i = 1, \dots, m$, the elements a_{1i}, \dots, a_{ki} generate A as an F -algebra.*
- (2) *There does not exist a pair of different indices i, j for which there is an automorphism Ψ of the F -algebra A such that*

$$a_{1i} = \Psi(a_{1j}), \dots, a_{ki} = \Psi(a_{kj}).$$

Proof. Let B denote the subalgebra of A^m generated by a_1, \dots, a_k . Recall that there is a unique (up to isomorphism) simple A -module M and it is faithful. Let M_i be the pull-back of M via the projection $\pi_i : B \rightarrow A$ on the i -th coordinate. Thus M_i is a B -module which coincides with M as an F -vector space, and for $b \in B$ and $m \in M_i = M$ we have $bm = \pi_i(b)m$. Since π_i is surjective by (1), each M_i is a simple B module. We claim that these B -modules are pairwise nonisomorphic. Indeed, suppose that for some $i \neq j$ the B -modules M_i and M_j are isomorphic and let $\Phi : M_i \rightarrow M_j$ be an isomorphism of these B -modules. For any $a \in A$ there is $b \in B$ such that $\pi_i(b) = a$. Set $\Psi(a) = \pi_j(b)$. We claim that Ψ is well-defined and an automorphism of the F -algebra A . Indeed, if $b_1 \in B$ is another element such that $\pi_i(b_1) = a$ then for any $m \in M_i$ we have $bm = b_1m$. Applying Φ to this equality, we see that $b\Phi(m) = b_1\Phi(m)$ for any $m \in M_i$. Since Φ is an isomorphism, we conclude that $bn = b_1n$ for any $n \in M_j$, that is, $\pi_j(b)m = \pi_j(b_1)m$ for every $m \in M$. Since M is a faithful A -module, we conclude that $\pi_j(b) = \pi_j(b_1)$. This shows that Ψ is well-defined. It is now straightforward to see that Ψ respects addition and multiplication and that it is F -linear. It follows that Ψ is an isomorphism of F -algebras. This however is in contradiction with our assumption (2). It follows that M_i and M_j are not isomorphic as B -modules for $i \neq j$. Note that $\bigoplus_{i=1}^m M_i$ is a semisimple, faithful B -module. It follows that B is semisimple and every simple B -module is isomorphic to one of the M_i 's. By Wedderburn–Artin theory, B is isomorphic to the product $\prod_{i=1}^m B_i$, where $B_i = M_{n_i}(D_i)$, $D_i = \text{End}_B(M_i)$, and $n_i \dim_F(D_i) = \dim_F(M_i) = \dim_F M$. Note that $D_i = \text{End}_B(M_i) = \text{End}_A(M)$ and therefore A is isomorphic to B_i for each i , again by Wedderburn–Artin theory. This proves that $\dim_F A^m = \dim_F B$, and consequently $A^m = B$. \square

As a simple corollary we get the following.

Proposition 6.2. *Let A be a simple finite-dimensional algebra over a field F . For any $k > 0$ the group $\text{Aut}_F(A)$ of F -algebra automorphisms of A acts freely on the set $\text{Gen}_k(A, F)$. The algebra A^m can be generated by k elements as an F -algebra if and only if there are at least m different orbits of the action of $\text{Aut}_F(A)$ on $\text{Gen}_k(A, F)$.*

Proof. The action of $\text{Aut}_F(A)$ on $\text{Gen}_k(A, F)$ is the restriction of the coordinate-wise action of $\text{Aut}_F(A)$ on A^k . If $\Psi \in \text{Aut}_F(A)$ fixes an element of $\text{Gen}_k(A, F)$, then it fixes each member of a set of generators of A as an F -algebra, so Ψ is the identity. This explains why the action is free. Theorem 6.1 says that elements $a_1 = (a_{11}, \dots, a_{1m}), \dots, a_k = (a_{k1}, \dots, a_{km})$ generate A^m as an F -algebra if and only if the elements $(a_{11}, \dots, a_{k1}), \dots, (a_{1m}, \dots, a_{km})$ belong to different orbits of the action of $\text{Aut}_F(A)$ on $\text{Gen}_k(A, F)$. \square

Suppose now that $F = \mathbb{F}_q$ is a finite field with q elements. Then simple finite-dimensional \mathbb{F}_q -algebras are exactly algebras of the form $M_n(\mathbb{F}_{q^s})$ for some positive

integers n, s . Now, by the Skolem–Noether theorem, the group of automorphisms of the \mathbb{F}_q -algebra $M_n(\mathbb{F}_{q^s})$ is the semidirect product of the group $\text{PGL}_n(\mathbb{F}_{q^s})$ and the Galois group $\text{Gal}(\mathbb{F}_{q^s}/\mathbb{F}_q)$. Thus we get the following.

Theorem 6.3. *Let $A = M_n(\mathbb{F}_{q^s})$. Then A^m can be generated by k elements as an \mathbb{F}_q -algebra if and only if*

$$m \leq \frac{|\text{Gen}_k(A, \mathbb{F}_q)|}{s|\text{PGL}_n(\mathbb{F}_{q^s})|}.$$

Furthermore, $|\text{Gen}_k(A^m, \mathbb{F}_q)| = \prod_{i=0}^{m-1} (|\text{Gen}_k(A, \mathbb{F}_q)| - i \cdot s \cdot |\text{PGL}_n(\mathbb{F}_{q^s})|)$.

Proof. As we noted above, $\text{Aut}_{\mathbb{F}_q}(A)$ has $s|\text{PGL}_n(\mathbb{F}_{q^s})|$ elements. Since $\text{Aut}_{\mathbb{F}_q}(A)$ acts freely on $\text{Gen}_k(A, \mathbb{F}_q)$, the number of orbits of this action is equal to

$$\frac{|\text{Gen}_k(A, \mathbb{F}_q)|}{s|\text{PGL}_n(\mathbb{F}_{q^s})|}.$$

The first part of the theorem is now an immediate consequence of Proposition 6.2.

To prove the second part note that according to Proposition 6.2 the elements of $\text{Gen}_k(A^m, \mathbb{F}_q)$ are in bijective correspondence with sequences of length m of elements from $\text{Gen}_k(A, \mathbb{F}_q)$, with no two elements in the same orbit of $\text{Aut}_{\mathbb{F}_q}(A)$. In order to count these sequences, let o be the number of orbits of the action of $\text{Aut}_{\mathbb{F}_q}(A)$ on $\text{Gen}_k(A, \mathbb{F}_q)$ and let t be the size of each orbit. We can choose a sequence of m different orbits O_1, \dots, O_m in $m! \binom{o}{m}$ ways and the number of sequences g_1, \dots, g_m such that $g_i \in O_i$ for $i = 1, \dots, m$ is t^m . Thus

$$|\text{Gen}_k(A^m, \mathbb{F}_q)| = m! \binom{o}{m} t^m = \prod_{i=0}^{m-1} (ot - it).$$

The second part of the theorem follows now immediately from the equalities $ot = |\text{Gen}_k(A, \mathbb{F}_q)|$ and $t = s|\text{PGL}_n(\mathbb{F}_{q^s})|$. □

For a simple separable algebra A over any field F the sets $\text{Gen}_k(A, F)$ are nonempty for any $k \geq 2$. In other words, we have the following.

Theorem 6.4. *Let A be a simple separable algebra over a field F . Then A can be generated by two elements as an F -algebra.*

Proof. For infinite fields F the result has been proved in [Mazur and Petrenko 2009]. When $F = \mathbb{F}_q$ is a finite field with q elements then A is isomorphic to $M_n(\mathbb{F}_{q^s})$ for some positive integers n and s . Let u be a generator of the multiplicative group of \mathbb{F}_{q^s} , so in particular $\mathbb{F}_{q^s} = \mathbb{F}_q[u]$. For $1 \leq i, j \leq n$ let E_{ij} denote the matrix whose (i, j) entry is 1 and all other entries are 0. Let $A = uE_{11}$ and $B = E_{1n} + \sum_{i=1}^{n-1} E_{i+1,i}$. Then $u^k E_{ij} = B^{i-1} A^k B^{n+1-j}$ for all $1 \leq i, j \leq n$ and all $k \geq 0$. It follows that A and B generate the \mathbb{F}_q -algebra $M_n(\mathbb{F}_{q^s})$. □

7. The numbers $|\text{Gen}_k(\mathbf{M}_n(\mathbb{F}_q), \mathbb{F}_q)|$

In this section we will study the numbers $|\text{Gen}_k(\mathbf{M}_n(\mathbb{F}_q), \mathbb{F}_q)|$. In particular, we will compute them when $n \leq 3$. To simplify the notation, we make the following definition.

Definition 7.1. Let m and n be positive integers and let q be a prime power. We introduce the following notation:

- (i) $\mathbf{G}_{m,n}(\mathbb{F}_q) = \text{Gen}_m(\mathbf{M}_n(\mathbb{F}_q), \mathbb{F}_q)$.
- (ii) $\mathfrak{g}_{m,n}(q) = |\mathbf{G}_{m,n}(\mathbb{F}_q)|$.
- (iii) $\text{gen}_{m,n}(q) = \frac{\mathfrak{g}_{m,n}(q)}{|\text{PGL}_n(\mathbb{F}_q)|}$.

Note that by Theorem 6.3, the number $\text{gen}_{m,n}(q)$ is equal to the largest $k \in \mathbb{Z}$ such that $r(\mathbf{M}_n(\mathbb{F}_q)^k, \mathbb{F}_q) \leq m$. Thus our notation agrees with that introduced in Definition 1.3.

When $n = 1$, an m -tuple generates \mathbb{F}_q if and only if it contains a nonzero element. It follows that $\mathfrak{g}_{m,1}(q) = q^m - 1$. From now on in this section we assume that $n \geq 2$, unless stated otherwise.

Our attempt at computing the numbers $\mathfrak{g}_{m,n}(q)$ is based on the following simple observation: a set of matrices does not generate the whole algebra $\mathbf{M}_n(\mathbb{F}_q)$ if and only if there is a maximal subalgebra of $\mathbf{M}_n(\mathbb{F}_q)$ that contains this set. Thus the following is true:

$$\mathbf{G}_{m,n}(\mathbb{F}_q) = \mathbf{M}_n(\mathbb{F}_q)^m - \bigcup \{ \mathcal{A}^m : \mathcal{A} \text{ is a maximal subalgebra of } \mathbf{M}_n(\mathbb{F}_q) \}. \tag{4}$$

Let \mathcal{D} be the subalgebra of scalar matrices of $\mathbf{M}_n(\mathbb{F}_q)$. Since any subalgebra of $\mathbf{M}_n(\mathbb{F}_q)$ contains \mathcal{D} , we can subtract \mathcal{D}^m in the above formula and get that $\mathbf{G}_{m,n}(\mathbb{F}_q)$ is equal to

$$\mathbf{M}_n(\mathbb{F}_q)^m - \mathcal{D}^m - \bigcup \{ \mathcal{A}^m - \mathcal{D}^m : \mathcal{A} \text{ is a maximal subalgebra of } \mathbf{M}_n(\mathbb{F}_q) \}.$$

Since $|\mathbf{M}_n(\mathbb{F}_q)| = q^{n^2}$ and $|\mathcal{D}| = q$, the inclusion-exclusion formula yields

$$\mathfrak{g}_{m,n}(q) = q^{mn^2} - q^m + \sum (-1)^k |(\mathcal{A}_{i_1}^m - \mathcal{D}^m) \cap \dots \cap (\mathcal{A}_{i_k}^m - \mathcal{D}^m)|,$$

where the sum is taken over all nonempty subsets $\{ \mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_k} \}$ of the set of all maximal subalgebras of $\mathbf{M}_n(\mathbb{F}_q)$. Since \mathcal{D} is contained in every subalgebra of $\mathbf{M}_n(\mathbb{F}_q)$, we have

$$(\mathcal{A}_{i_1}^m - \mathcal{D}^m) \cap \dots \cap (\mathcal{A}_{i_k}^m - \mathcal{D}^m) = \mathcal{A}_{i_1}^m \cap \dots \cap \mathcal{A}_{i_k}^m - \mathcal{D}^m = (\mathcal{A}_{i_1} \cap \dots \cap \mathcal{A}_{i_k})^m - \mathcal{D}^m,$$

and therefore

$$\mathfrak{g}_{m,n}(q) = q^{mn^2} - q^m + \sum (-1)^k (|\mathcal{A}_{i_1} \cap \dots \cap \mathcal{A}_{i_k}|^m - q^m), \tag{5}$$

where the sum is taken over all nonempty subsets $\{\mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_k}\}$ of the set of all maximal subalgebras of $M_n(\mathbb{F}_q)$.

In order to evaluate the right-hand side of (5), it is necessary to have a description of all maximal subalgebras of $M_n(\mathbb{F}_q)$. It is quite easy to produce one type of maximal subalgebras of $M_n(\mathbb{F}_q)$. In fact, we have the following result.

Lemma 7.2. *For a proper nontrivial vector subspace U of \mathbb{F}_q^n let \mathcal{A}_U be the set of all matrices from $M_n(\mathbb{F}_q)$ that leave U invariant. Then \mathcal{A}_U is a maximal subalgebra of $M_n(\mathbb{F}_q)$. Moreover, each \mathcal{A}_U is uniquely determined by U , that is, if $\mathcal{A}_U = \mathcal{A}_{U'}$ then $U = U'$.*

Proof. First note that the center of \mathcal{A}_U consists of scalar matrices. In fact, if a matrix A is in the center of \mathcal{A}_U then it acts as a scalar λ on U . The matrix $B = A - \lambda I$ annihilates U and is in the center of \mathcal{A}_U . Suppose that $Bv \neq 0$ for some v . Then there is a projection Π onto U such that $\Pi(Bv) \neq 0$. Since $\Pi \in \mathcal{A}_U$, we have $0 \neq \Pi Bv = B\Pi v = 0$, a contradiction. Thus $B = 0$ and A is a scalar matrix.

Now note that if $U \neq U'$ then there is $A \in \mathcal{A}_U - \mathcal{A}_{U'}$. In fact, if $U' \subsetneq U$ then such an A clearly exists since \mathcal{A}_U is transitive on U . If there exists $v \in U' - U$ then for any w there is an $A \in \mathcal{A}_U$ such that $Av = w$. Taking $w \notin U'$ yields the required A . This, in particular, proves the second assertion.

Take any matrix A not in \mathcal{A}_U and let \mathcal{A}' be the algebra generated by A and \mathcal{A}_U . Note that \mathcal{A}' cannot fix any nontrivial subspace V of \mathbb{F}_q^n . In fact, if $V \neq U$ then, as we have seen above, \mathcal{A}_U is not contained in \mathcal{A}_V and A does not take U into V . Thus \mathbb{F}_q^n is a simple and faithful \mathcal{A}' -module. It follows that \mathcal{A}' is a simple central \mathbb{F}_q -algebra with a simple module of dimension n over \mathbb{F}_q , hence it must be isomorphic to $M_n(\mathbb{F}_q)$. It follows that \mathcal{A}_U is maximal. \square

The following lemma describes a second type of maximal subalgebras of $M_n(\mathbb{F}_q)$.

Lemma 7.3. *Let s be a prime divisor of n and let $m = n/s$. Any \mathbb{F}_q -subalgebra of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$ is maximal. Any two such subalgebras are conjugate in $M_n(\mathbb{F}_q)$ and their number is equal to*

$$s^{-1} \prod_{s \nmid i, 1 \leq i < n} (q^n - q^i).$$

Proof. Let \mathcal{A} be a \mathbb{F}_q -subalgebra of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$. Thus \mathbb{F}_q^n is an \mathcal{A} -module of dimension m over the center of \mathcal{A} (which is isomorphic to \mathbb{F}_{q^s}). It follows that \mathbb{F}_q^n is a simple \mathcal{A} -module. Suppose that \mathcal{A}' is a \mathbb{F}_q -subalgebra of $M_n(\mathbb{F}_q)$ containing \mathcal{A} . Then \mathbb{F}_q^n is a simple and faithful \mathcal{A}' -module. It follows that \mathcal{A}' is simple, hence it is isomorphic to $M_k(\mathbb{F}_{q^r})$, where $kr = n$ and r is the dimension of the center of \mathcal{A}' over \mathbb{F}_q . Clearly, the center of \mathcal{A}' is contained in the center of \mathcal{A} . It follows that $r | s$, and therefore $r = 1$ or $r = s$ (recall that s is

a prime). In the former case we get $\mathcal{A}' = M_n(\mathbb{F}_q)$ and in the latter case we have $\mathcal{A}' = \mathcal{A}$. This shows that \mathcal{A} is maximal.

For the existence of an \mathbb{F}_q -subalgebra of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$ consider the (unique up to isomorphism) simple $M_m(\mathbb{F}_{q^s})$ -module V . It has dimension m as a vector space over \mathbb{F}_{q^s} , so as a \mathbb{F}_q -vector space it is isomorphic to \mathbb{F}_q^n . Thus the action of $M_m(\mathbb{F}_{q^s})$ on V induces an \mathbb{F}_q -algebra embedding of $M_m(\mathbb{F}_{q^s})$ into $M_n(\mathbb{F}_q)$.

Fix now a \mathbb{F}_q -subalgebra \mathcal{A} of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$. By the Noether–Skolem theorem, any \mathbb{F}_q -algebra homomorphism of \mathcal{A} into $M_n(\mathbb{F}_q)$ is given by conjugation with some invertible element of $M_n(\mathbb{F}_q)$. This means that the group $GL_n(\mathbb{F}_q)$ acts transitively on the set of subalgebras of $M_n(\mathbb{F}_q)$ which are isomorphic to $M_m(\mathbb{F}_{q^s})$. Since \mathcal{A} is maximal, the subgroup C of elements which act trivially on \mathcal{A} coincides with the multiplicative group of the center of \mathcal{A} . The quotient of the stabilizer of \mathcal{A} by C is, again by the Noether–Skolem theorem, isomorphic to the group of all automorphisms of the \mathbb{F}_q -subalgebra \mathcal{A} . We have seen earlier that the group of \mathbb{F}_q -algebra automorphisms of $M_m(\mathbb{F}_{q^s})$ has $s|PGL_m(\mathbb{F}_{q^s})|$ elements (see the discussion directly before Theorem 6.3). Therefore the stabilizer of \mathcal{A} has $|C| \cdot s \cdot |PGL_m(\mathbb{F}_{q^s})| = s|GL_m(\mathbb{F}_{q^s})|$ elements. Consequently, the number of \mathbb{F}_q -subalgebras \mathcal{A} of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$ is equal to

$$\frac{|GL_n(\mathbb{F}_q)|}{s|GL_m(\mathbb{F}_{q^s})|} = s^{-1} \prod_{s \nmid i, 1 \leq i < n} (q^n - q^i). \quad \square$$

It turns out that the maximal subalgebras described in Lemmas 7.2 and 7.3 exhaust all possible maximal subalgebras. In other words, we have the following result.

Proposition 7.4. *Let \mathcal{A} be a maximal \mathbb{F}_q -subalgebra of $M_n(\mathbb{F}_q)$. Then either $\mathcal{A} = \mathcal{A}_U$ for some subspace U of \mathbb{F}_q^n or \mathcal{A} is isomorphic to $M_m(\mathbb{F}_{q^s})$ for some prime divisor s of $n = ms$.*

Proof. Suppose that \mathcal{A} fixes some proper nontrivial subspace U of \mathbb{F}_q^n . Then \mathcal{A} is contained in \mathcal{A}_U , hence $\mathcal{A} = \mathcal{A}_U$. If no proper nontrivial subspace of \mathbb{F}_q^n is fixed by \mathcal{A} then \mathbb{F}_q^n is a simple and faithful \mathcal{A} -module. It follows that \mathcal{A} is simple and therefore it is isomorphic to $M_k(\mathbb{F}_{q^r})$, where $kr = n$. Let s be a prime divisor of r . The center of \mathcal{A} contains a subfield F isomorphic to \mathbb{F}_{q^s} . The centralizer of F in $M_n(\mathbb{F}_q)$ consists exactly of those linear transformations of \mathbb{F}_q^n which are F -linear. Thus it is a subalgebra of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$, where $ms = n$. On the other hand, this subalgebra contains \mathcal{A} , hence it must be equal to \mathcal{A} . \square

In order to carry out our strategy to compute the numbers $g_{m,n}(q)$ we need to understand the intersections of maximal subalgebras of $M_n(\mathbb{F}_q)$. This appears to be a very challenging combinatorial problem and so far we have only succeeded in completing the computations for $n \leq 3$. One of the complications in the general

case is that the maximal subalgebras are of two different types. This difficulty disappears when n is a prime by the following observation.

Lemma 7.5. *Let n be a prime number. If \mathcal{A} is a maximal subalgebra of $M_n(\mathbb{F}_q)$ isomorphic to \mathbb{F}_{q^n} , then its intersection with any other maximal subalgebra is equal to \mathcal{D} , the algebra of scalar matrices.*

Proof. Since n is prime, \mathbb{F}_{q^n} has only two subfields, itself and \mathbb{F}_q . In other words, \mathcal{A} has only two subalgebras, \mathcal{A} and \mathcal{D} . Since the intersection cannot be equal to \mathcal{A} , it is equal to \mathcal{D} . □

For the rest of this section we assume that n is a prime number. Thus Lemma 7.5 tells us that if the set $\{\mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_k}\}$ of maximal subalgebras of $M_n(\mathbb{F}_q)$ includes a subalgebra isomorphic to \mathbb{F}_{q^n} , and $k \geq 2$, then the intersection of the subalgebras in this set is equal to \mathcal{D} , and so the corresponding term in (5), $|\mathcal{A}_{i_1} \cap \dots \cap \mathcal{A}_{i_k}|^m - q^m$, is equal to 0. It follows that we can rewrite (5) in the following way:

$$g_{m,n}(q) = q^{mn^2} - q^m - \sum_{\mathcal{A} \cong \mathbb{F}_{q^n}} (|\mathcal{A}|^m - q^m) + \sum (-1)^k (|\mathcal{A}_{U_1} \cap \dots \cap \mathcal{A}_{U_k}|^m - q^m),$$

where the second sum is over all nonempty sets $\{U_1, \dots, U_k\}$ of nontrivial proper subspaces of \mathbb{F}_q^n . By Lemma 7.3, the first sum consists of $n^{-1} \prod_{i=1}^{n-1} (q^n - q^i)$ terms, each term being $q^{mn} - q^m$. Thus we get the following formula:

$$g_{m,n}(q) = q^{mn^2} - q^m - n^{-1}(q^{mn} - q^m) \prod_{i=1}^{n-1} (q^n - q^i) + \sum (-1)^k (|\mathcal{A}_{U_1} \cap \dots \cap \mathcal{A}_{U_k}|^m - q^m), \quad (6)$$

where the sum is over all nonempty sets $\{U_1, \dots, U_k\}$ of nontrivial proper subspaces of \mathbb{F}_q^n .

Let \mathcal{F} be the set of all subalgebras of $M_n(\mathbb{F}_q)$ which are intersections of some of the maximal algebras of the form \mathcal{A}_U . For each $\mathcal{A} \in \mathcal{F}$, define the degree $d(\mathcal{A})$ of \mathcal{A} by

$$d(\mathcal{A}) = \sum (-1)^k, \quad (7)$$

where the sum is over all sets $\{U_1, \dots, U_k\}$ of nontrivial proper subspaces of \mathbb{F}_q^n such that $\mathcal{A}_{U_1} \cap \dots \cap \mathcal{A}_{U_k} = \mathcal{A}$. Thus (6) can be stated as

$$g_{m,n}(q) = q^{mn^2} - q^m - n^{-1}(q^{mn} - q^m) \prod_{i=1}^{n-1} (q^n - q^i) + \sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m). \quad (8)$$

The following simple lemma will be useful for our analysis of elements of \mathcal{F} .

Lemma 7.6. *Let F be a field, V be a vector space over F , and let $v_1, \dots, v_k \in V$ be a minimal linearly dependent collection of vectors (so any $k - 1$ of them are*

linearly independent). Then any linear endomorphism of V that scales v_1, \dots, v_k is a scalar operator when restricted to the linear span of v_1, \dots, v_k .

Proof. Let f be a linear endomorphism of V such that $f(v_i) = \alpha_i v_i$ for some $\alpha_i \in F$ and $i = 1, \dots, k$. The assumptions of the lemma imply that

$$v_k = \beta_1 v_1 + \dots + \beta_{k-1} v_{k-1}$$

for some nonzero $\beta_1, \dots, \beta_{k-1} \in F$. By expressing $f(v_k)$ in two ways, as $\beta_1 \alpha_1 v_1 + \dots + \beta_{k-1} \alpha_{k-1} v_{k-1}$ and as $\alpha_k v_k$, we obtain $\beta_i \alpha_i = \beta_i \alpha_k$ for all i . Since none of the β_i 's is 0, we have $\alpha_i = \alpha_k$ for $i = 1, \dots, k$. \square

7A. The case $n = 2$. In this subsection we evaluate (8) in the case $n = 2$. Any element $\mathcal{A} \in \mathcal{F}$ is of the form $\mathcal{A}_{U_1} \cap \dots \cap \mathcal{A}_{U_k}$, where $k \geq 1$ and U_1, \dots, U_k are distinct lines in \mathbb{F}_q^2 . Note that by Lemma 7.6, $\mathcal{A} = \mathcal{D}$ if $k \geq 3$ and in this case \mathcal{A} does not contribute anything to (8). It follows that if \mathcal{A} is an element of \mathcal{F} different from \mathcal{D} , then it can be expressed as the intersection of maximal subalgebras in a unique way and it is either of the form \mathcal{A}_U or of the form $\mathcal{A}_{U_1} \cap \mathcal{A}_{U_2}$. In the former case, we have $|\mathcal{A}| = q^3$ and $d(\mathcal{A}) = -1$. In the latter case, $|\mathcal{A}| = q^2$ and $d(\mathcal{A}) = 1$. Since the number of lines in \mathbb{F}_q^2 is $q + 1$, (8) takes the following form:

$$\begin{aligned} \mathfrak{g}_{m,2}(q) &= q^{4m} - q^m - 2^{-1}(q^{2m} - q^m)(q^2 - q) \\ &\quad - (q + 1)(q^{3m} - q^m) + 2^{-1}(q + 1)q(q^{2m} - q^m), \end{aligned}$$

which simplifies to

$$\mathfrak{g}_{m,2}(q) = q^{2m+1}(q^{m-1} - 1)(q^m - 1). \tag{9}$$

7B. The case $n = 3$. In this subsection we evaluate (8) for $n = 3$. This is substantially more difficult than the case $n = 2$, but we are still able to analyze all elements of \mathcal{F} . The following combinatorial lemma will help us evaluate the degree of some of the algebras in \mathcal{F} .

Lemma 7.7. *Let X be a finite set. Consider a family \mathcal{S} of subsets of X such that if $Y \in \mathcal{S}$ and $Y \subseteq Y' \subseteq X$, then Y' also belongs to \mathcal{S} . Suppose furthermore that one of the following two conditions is true.*

- (1) *There is $x \in X$ such that $X' - \{x\} \in \mathcal{S}$ for any $X' \in \mathcal{S}$.*
- (2) *There are $x, y \in X$ such that if $X' \in \mathcal{S}$ and $X' - \{x\} \notin \mathcal{S}$ then*
 - (a) *$X' - \{y\} \in \mathcal{S}$ and*
 - (b) *$(X' \cup \{y\}) - \{x\} \notin \mathcal{S}$.*

Then $\sum_{Y \in \mathcal{S}} (-1)^{|Y|} = 0$.

Proof. Let \mathcal{S}_0 be the family of those subsets from \mathcal{S} that do not contain x and let \mathcal{S}_1 be the family of those subsets that contain x . The map $t : Y \mapsto Y \cup \{x\}$ is an injection from \mathcal{S}_0 to \mathcal{S}_1 . Let $\mathcal{S}_2 = \mathcal{S}_1 - t(\mathcal{S}_0)$. We have

$$\sum_{Y \in \mathcal{S}} (-1)^{|Y|} = \sum_{Y \in \mathcal{S}_0} (-1)^{|Y|} + \sum_{Y \in t(\mathcal{S}_0)} (-1)^{|Y|} + \sum_{Y \in \mathcal{S}_2} (-1)^{|Y|}.$$

Since $|t(Y)| = 1 + |Y|$, the first two sums on the right annihilate each other, and so

$$\sum_{Y \in \mathcal{S}} (-1)^{|Y|} = \sum_{Y \in \mathcal{S}_2} (-1)^{|Y|}.$$

Condition (1) exactly means that \mathcal{S}_2 is empty, hence $\sum_{Y \in \mathcal{S}} (-1)^{|Y|} = 0$. If condition (2) holds, we write \mathcal{S}_2 as a disjoint union $\mathcal{S}_2 = \mathcal{S}_{20} \cup \mathcal{S}_{21}$, where \mathcal{S}_{20} consists of those elements of \mathcal{S}_2 which do not contain y . By (b), the map $s : Y \mapsto Y \cup \{y\}$ maps \mathcal{S}_{20} into \mathcal{S}_{21} and (a) implies that s is onto. Thus $s : \mathcal{S}_{20} \rightarrow \mathcal{S}_{21}$ is a bijection and

$$\sum_{Y \in \mathcal{S}_2} (-1)^{|Y|} = \sum_{Y \in \mathcal{S}_{20}} (-1)^{|Y|} + \sum_{Y \in \mathcal{S}_{21}} (-1)^{|Y|} = \sum_{Y \in \mathcal{S}_{20}} ((-1)^{|Y|} + (-1)^{|s(Y)|}) = 0. \quad \square$$

We apply Lemma 7.7 as follows. Given $\mathcal{A} \in \mathcal{F}$, the set $X = X_{\mathcal{A}}$ will consist of all proper nontrivial subspaces of \mathbb{F}_q^3 fixed by \mathcal{A} and the family $\mathcal{S} = \mathcal{S}_{\mathcal{A}}$ will consist of all subsets $\{U_1, \dots, U_k\}$ of X such that $\mathcal{A}_{U_1} \cap \dots \cap \mathcal{A}_{U_k} = \mathcal{A}$. If conditions (1) or (2) hold for $\mathcal{S}_{\mathcal{A}}$, then Lemma 7.7 tells us that $d(\mathcal{A}) = 0$.

Before we start the analysis of elements in \mathcal{F} let us recall that the dot product $v \cdot w = v_1 w_1 + v_2 w_2 + v_3 w_3$ is a nondegenerate symmetric bilinear form on \mathbb{F}_q^3 . The adjoint operator with respect to this bilinear form is the transposition. It follows that if $\mathcal{A}_{U_1} \cap \dots \cap \mathcal{A}_{U_k} = \mathcal{A} \in \mathcal{F}$ then

$$\mathcal{A}_{U_1^\perp} \cap \dots \cap \mathcal{A}_{U_k^\perp} = \mathcal{A}^t := \{A^t : A \in \mathcal{A}\} \in \mathcal{F},$$

where A^t is the transpose of A and U^\perp is the subspace orthogonal to U with respect to the dot product. We will often call \mathcal{A}^t the dual of \mathcal{A} . It is clear that \mathcal{A} and \mathcal{A}^t have the same number of elements and the same degree.

Definition 7.8. Let $\mathcal{A} \in \mathcal{F}$. Then

$$L_{\mathcal{A}} = \{U : \dim U = 1 \text{ and } \mathcal{A} \subseteq \mathcal{A}_U\}$$

is the set of all lines fixed by \mathcal{A} and

$$P_{\mathcal{A}} = \{U : \dim U = 2 \text{ and } \mathcal{A} \subseteq \mathcal{A}_U\}$$

is the set of all planes fixed by \mathcal{A} .

Note that $L_{\mathcal{A}'} = \{\pi^\perp : \pi \in P_{\mathcal{A}}\}$ and $P_{\mathcal{A}'} = \{l^\perp : l \in L_{\mathcal{A}}\}$. Also, $X_{\mathcal{A}} = L_{\mathcal{A}} \cup P_{\mathcal{A}}$.

Consider an algebra $\mathcal{A} \in \mathcal{F}$, $\mathcal{A} \neq \mathcal{D}$. Then \mathcal{A} falls into exactly one of the following cases.

Case I: $L_{\mathcal{A}}$ contains three lines in general position. Recall that we say that three lines in \mathbb{F}_q^3 are in general position if they are not contained in any plane. Dually, three planes are in general position if they do not share any common line. Let $l_1, l_2, l_3 \in L_{\mathcal{A}}$ be three lines in general position. Let π_i be the plane spanned by l_j and l_k , where $\{i, j, k\} = \{1, 2, 3\}$.

Subcase Ia: $L_{\mathcal{A}} = \{l_1, l_2, l_3\}$. In this case $P_{\mathcal{A}} = \{\pi_1, \pi_2, \pi_3\}$. The algebra \mathcal{A} is conjugate to the algebra of all diagonal matrices. In particular, $|\mathcal{A}| = q^3$. Furthermore, $X_{\mathcal{A}} = L_{\mathcal{A}} \cup P_{\mathcal{A}}$ and a subset of $X_{\mathcal{A}}$ belongs to $\mathcal{S}_{\mathcal{A}}$ if and only if it contains one of the following sets: $\{l_1, l_2, l_3\}$, $\{\pi_1, \pi_2, \pi_3\}$, $\{l_1, l_2, \pi_1, \pi_2\}$, $\{l_1, l_3, \pi_1, \pi_3\}$, or $\{l_2, l_3, \pi_2, \pi_3\}$. Thus $\mathcal{S}_{\mathcal{A}}$ has two members of cardinality 3, nine members of cardinality 4, six members of cardinality 5 and one element of cardinality 6. Therefore, $d(\mathcal{A}) = -2 + 9 - 6 + 1 = 2$.

Note that the algebras in this subcase are in bijective correspondence with sets of three lines in general position. Recall that \mathbb{F}_q^3 has $q^2 + q + 1$ lines, and each plane has $q + 1$ lines. It follows that the number of ordered triples of lines in general position is $(q^2 + q + 1)(q^2 + q)q^2$. Thus, the number of algebras in this subcase is $q^3(q + 1)(q^2 + q + 1)/6$. Consequently, the algebras in this subcase contribute the quantity

$$3^{-1}q^{m+3}(q + 1)(q^2 + q + 1)(q^{2m} - 1)$$

to the sum $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Subcase Ib: $L_{\mathcal{A}} \supseteq \{l_1, l_2, l_3, l_4\}$, where l_4 is a line not contained in any of the planes π_1, π_2, π_3 . In this case, by Lemma 7.6, we have $\mathcal{A} = \mathcal{D}$, and \mathcal{A} does not contribute anything to the sum $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

It remains to consider the case when $L_{\mathcal{A}}$ contains a line l_4 which is contained in one of the planes π_1, π_2, π_3 . Changing the numbering if necessary, we may assume that l_4 belongs to π_1 . If there is a line l_5 (different from l_1, \dots, l_4) which is contained in π_2 , the planes through l_4, l_1 and through l_5, l_2 intersect along a line l_6 which does not belong to any of the planes π_1, π_2, π_3 . Thus we are in Subcase Ib. The same argument shows that there is no line in $L_{\mathcal{A}}$ different from l_1, \dots, l_4 and contained in π_3 . Since \mathcal{A} fixes three different lines in π_1 , it acts as a scalar on π_1 by Lemma 7.6. In particular, $L_{\mathcal{A}}$ contains all the lines in π_1 . We will write π for π_1 and l for l_1 . We see that all the remaining algebras in Case I fall in the following subcase.

Subcase Ic: $L_{\mathcal{A}} = \{l\} \cup \{\text{all lines in } \pi\}$. It is easy to see that in this case $P_{\mathcal{A}} = \{\pi\} \cup \{\text{all planes through } l\}$. We will show that $d(\mathcal{A}) = 0$ by applying Lemma 7.7

to $X = X_{\mathcal{A}}$, $\mathcal{P} = \mathcal{P}_{\mathcal{A}}$. We need to verify that $x = l$, $y = \pi$ satisfy condition (2). Suppose that $X' \in \mathcal{S}_{\mathcal{A}}$ and $X' - \{l\} \notin \mathcal{S}_{\mathcal{A}}$. We claim that X' contains at most one plane through l . For suppose otherwise, that there are two planes containing l in X' . Their intersection is l . Thus a matrix fixes all elements of $X' - \{l\}$ if and only if it fixes all elements of X' , that is, $X' - \{l\} \in \mathcal{S}_{\mathcal{A}}$, a contradiction. This proves that indeed X' contains at most one plane different from π . We claim that X' contains at least two lines contained in π . Otherwise, there would be at most one such line in X' , so X' would be a subset of a set of the form $\{l, l', \pi, \pi'\}$ for some line l' contained in π and some plane π' containing l . Thus \mathcal{A} would contain the algebra $\mathcal{A}' = \mathcal{A}_l \cap \mathcal{A}_{l'} \cap \mathcal{A}_{\pi} \cap \mathcal{A}_{\pi'}$. This is, however, not possible, since \mathcal{A}' has an element which is not a scalar on π and all elements of \mathcal{A} act as scalars on π . Indeed, if the line $l'' = \pi \cap \pi'$ is different from l' then \mathcal{A}' equals $\mathcal{A}_l \cap \mathcal{A}_{l'} \cap \mathcal{A}_{l''}$ and contains the matrix which is the identity on l and l' and is 0 on l'' . If $l'' = l'$ then $\mathcal{A}' = \mathcal{A}_l \cap \mathcal{A}_{l'} \cap \mathcal{A}_{\pi}$ contains the algebra $\mathcal{A}_l \cap \mathcal{A}_{l'} \cap \mathcal{A}_{l'_1}$ for any line l'_1 in π which is different from l' .

Thus there are two lines in X' which are contained in π . These two lines span π , so $X' - \{\pi\} \in \mathcal{S}_{\mathcal{A}}$. Also, $(X' \cup \{\pi\}) - \{l\}$ and $X' - \{l\}$ are fixed by the same set of matrices, so $(X' \cup \{\pi\}) - \{l\} \notin \mathcal{S}_{\mathcal{A}}$. This verifies condition (2) of Lemma 7.7, so $d(\mathcal{A}) = 0$. Consequently, the algebras of Subcase Ic do not contribute anything to the sum $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Note that if $P_{\mathcal{A}}$ contains three planes in general position, then the three lines obtained by intersecting pairs of these planes are in general position and belong to $L_{\mathcal{A}}$. Thus from now on we assume that $L_{\mathcal{A}}$ does not contain three lines in general position and that $P_{\mathcal{A}}$ does not contain three planes in general position. If $L_{\mathcal{A}}$ contains more than two elements, then all of the lines in $L_{\mathcal{A}}$ must be contained in some plane π and then, by Lemma 7.6, $L_{\mathcal{A}} = \{\text{all lines in } \pi\}$. Similarly, by duality, if $P_{\mathcal{A}}$ contains more than two elements, then all the planes in $P_{\mathcal{A}}$ share a common line l and $P_{\mathcal{A}} = \{\text{all planes which contain } l\}$. This leads to the following two cases.

Case II: There is a plane π such that $L_{\mathcal{A}} = \{\text{all lines in } \pi\}$. By Lemma 7.6, every element of \mathcal{A} acts as a scalar on π . In particular, $\pi \in P_{\mathcal{A}}$. Note that all the planes in $P_{\mathcal{A}}$ must share a common line l (if $P_{\mathcal{A}} = \{\pi\}$, pick any line in π for l). In fact, suppose that there are $\pi_1, \pi_2 \in P_{\mathcal{A}}$ such that the lines $\pi \cap \pi_1$ and $\pi \cap \pi_2$ are different. Then the line $\pi_1 \cap \pi_2$ belongs to $L_{\mathcal{A}}$ and is not contained in π , which is not possible. Thus, $P_{\mathcal{A}} \subseteq \{\text{all planes which contain } l\}$. We claim that any $X \in \mathcal{S}_{\mathcal{A}}$ contains at least two lines in π different from l . In fact, if the lines in X are contained in $\{l, l_1\}$ then consider a plane π_1 which does not contain l but contains l_1 . There is a matrix A which is 0 on l and is the identity on π_1 and this matrix fixes every plane passing through l . Thus A fixes all elements of X , yet A is not a scalar on π . This means

that $A \notin \mathcal{A}$, and consequently $X \notin S_{\mathcal{A}}$, a contradiction. Now any two lines in X span π . It follows that any matrix which fixes all elements of $X - \{\pi\}$ also fixes π , that is, $X - \{\pi\} \in S_{\mathcal{A}}$. This means that the family $\mathcal{S}_{\mathcal{A}}$ of subsets of $X_{\mathcal{A}}$ satisfies the assumptions of Lemma 7.7, condition (1), with $x = \pi$. It follows that $d(\mathcal{A}) = 0$ and the algebras in this case do not contribute anything to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Case III: There is a line l such that $P_{\mathcal{A}} = \{\text{all planes through } l\}$. Any algebra in this case is dual to an algebra in Case II, hence it has degree 0. Thus algebras in this case do not contribute anything to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

It remains to analyze algebras \mathcal{A} such that both $L_{\mathcal{A}}$ and $P_{\mathcal{A}}$ have at most two elements.

Case IV: $|L_{\mathcal{A}}| = 2 = |P_{\mathcal{A}}|$. We may assume that $L_{\mathcal{A}} = \{l, l'\}$ and $P_{\mathcal{A}} = \{\pi, \pi'\}$, where π' is spanned by l, l' and $\pi \cap \pi' = l$. It is easy to see that the family $\mathcal{S}_{\mathcal{A}}$ has three elements: $\{l, l', \pi\}$, $\{l', \pi, \pi'\}$, and $\{l, l', \pi, \pi'\}$. Thus $d(\mathcal{A}) = -2 + 1 = -1$. Choosing nonzero vectors $v_1 \in l', v_2 \in l$, and $v_3 \in \pi - l$ we get a basis of \mathbb{F}_q^3 and $A \in \mathcal{A}$ if and only if the matrix of the linear transformation given by A , expressed in the basis v_1, v_2, v_3 , has the form

$$\begin{pmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix}.$$

In other words, \mathcal{A} is conjugate to the algebra of all the matrices of the form

$$\begin{pmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix}.$$

In particular, $|\mathcal{A}| = q^4$. To count the number of algebras in Case IV, note that these algebras are in bijective correspondence with triples l, l', π , where π is a plane and l and l' are lines such that $l \subset \pi$ and $l' \not\subset \pi$. There are $q^2 + q + 1$ choices for π and for each π we have $q + 1$ choices of l and q^2 choices of l' . Thus the number of algebras in Case IV is $q^2(q + 1)(q^2 + q + 1)$. Consequently, the algebras in this case contribute

$$-q^{m+2}(q + 1)(q^2 + q + 1)(q^{3m} - 1)$$

to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Case V: $|L_{\mathcal{A}}| = 2$ and $|P_{\mathcal{A}}| = 1$. Thus $L_{\mathcal{A}} = \{l, l'\}$ and $P_{\mathcal{A}} = \{\pi\}$, where π is spanned by l, l' . It is straightforward to see that $\mathcal{S}_{\mathcal{A}}$ has two elements: $\{l, l'\}$ and $\{l, l', \pi\}$. It follows that $d(\mathcal{A}) = 0$ and therefore algebras in this case contribute nothing to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Case V¹: $|L_{\mathcal{A}}| = 1$ and $|P_{\mathcal{A}}| = 2$. Algebras in this case are dual to algebras in Case V, so they have degree 0 and contribute nothing to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Case VI: $L_{\mathcal{A}} = \{l\}$ and $P_{\mathcal{A}} = \{\pi\}$, where $l \not\subset \pi$. It is clear that $\mathcal{S}_{\mathcal{A}}$ has exactly one element: $\{l, \pi\}$. Thus $d(\mathcal{A}) = 1$. Choosing a basis v_1 of l and v_2, v_3 of π we easily see that \mathcal{A} is conjugate to the algebra of all the matrices of the form

$$\begin{pmatrix} * & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix}.$$

In particular, $|\mathcal{A}| = q^5$. To count the number of algebras in Case VI, note that these algebras are in bijective correspondence with pairs l, π , where π is a plane and l is a line not contained in π . There are $q^2 + q + 1$ choices for π and for each π we have q^2 choices of l . Thus the number of algebras in Case VI is $q^2(q^2 + q + 1)$. Consequently, the algebras in this case contribute

$$q^{m+2}(q^2 + q + 1)(q^{4m} - 1)$$

to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Case VII: $L_{\mathcal{A}} = \{l\}$ and $P_{\mathcal{A}} = \{\pi\}$, where $l \subset \pi$. It is clear that $\mathcal{S}_{\mathcal{A}}$ has exactly one element: $\{l, \pi\}$. Thus $d(\mathcal{A}) = 1$. Choosing a basis v_1 of l , v_1, v_2 of π , and a vector $v_3 \notin \pi$, we easily see that \mathcal{A} is conjugate to the algebra of all the matrices of the form

$$\begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix}.$$

In particular, $|\mathcal{A}| = q^6$. To count the number of algebras in Case VII, note that these algebras are in bijective correspondence with pairs l, π , where π is a plane and l is a line contained in π . There are $q^2 + q + 1$ choices for π and for each π we have $q + 1$ choices of l . Thus the number of algebras in Case VII is $(q + 1)(q^2 + q + 1)$. Consequently, the algebras in this case contribute

$$q^m(q + 1)(q^2 + q + 1)(q^{5m} - 1)$$

to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Case VIII: $\mathcal{A} = \mathcal{A}_l$ for some line l . The family $\mathcal{S}_{\mathcal{A}}$ has exactly one element: $\{l\}$, so $d(\mathcal{A}) = -1$. It is easy to see that \mathcal{A} is conjugate to the algebra of all the matrices of the form

$$\begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix}.$$

In particular, $|\mathcal{A}| = q^7$. Algebras in this case are in bijection with lines, so we have $q^2 + q + 1$ such algebras. Thus the algebras in this case contribute

$$-q^m(q^2 + q + 1)(q^{6m} - 1)$$

to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

The following case is the last to consider.

Case VIII¹: $\mathcal{A} = \mathcal{A}_\pi$ for some plane π . This case consists of algebras dual to algebras of Case VIII, so they also contribute

$$-q^m(q^2 + q + 1)(q^{6m} - 1)$$

to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$.

Putting together all the contributions to $\sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m)$ we arrive at the formula

$$\begin{aligned} \sum_{\mathcal{A} \in \mathcal{F}} d(\mathcal{A})(|\mathcal{A}|^m - q^m) &= 3^{-1}q^{m+3}(q+1)(q^2+q+1)(q^{2m}-1) \\ &\quad - q^{m+2}(q+1)(q^2+q+1)(q^{3m}-1) + q^{m+2}(q^2+q+1)(q^{4m}-1) \\ &\quad + q^m(q+1)(q^2+q+1)(q^{5m}-1) - 2q^m(q^2+q+1)(q^{6m}-1). \end{aligned}$$

After inserting this into (8) and simplifying we arrive at the following formula for $g_{m,3}(q)$:

$$\begin{aligned} g_{m,3}(q) &= q^{3m+4}(q^{m-1}-1)(q^{m-1}+1)(q^m-1) \\ &\quad \times (q^{3m-2} + q^{2m-2} - q^m - 2q^{m-1} - q^{m-2} + q + 1). \end{aligned} \tag{10}$$

7C. Lower bound for $g_{m,n}(q)$. So far we have been unable to obtain exact formulas for $g_{m,n}(q)$ for any $n \geq 4$. We have however the following lower bound.

Proposition 7.9. *Let m and n be positive integers and let q be a power of a prime number. Then*

$$g_{m,n}(q) \geq q^{mn^2} - 2^{(n+6)/2}q^{n^2m-(m-1)(n-1)}. \tag{11}$$

Proof. By (4), we have the following inequality:

$$g_{m,n}(q) \geq q^{mn^2} - \sum |\mathcal{A}|^m,$$

where the sum is taken over all maximal subalgebras \mathcal{A} of $M_n(\mathbb{F}_q)$. We use the description of maximal subalgebras given by Proposition 7.4. Let $1 \leq k < n$. The number of k -dimensional subspaces of \mathbb{F}_q^n is

$$\prod_{i=0}^{k-1} (q^n - q^i) \prod_{i=0}^{k-1} (q^k - q^i)^{-1}.$$

For any such subspace V the algebra \mathcal{A}_V has $q^{n^2-nk+k^2}$ elements. Let

$$S_k = \sum |\mathcal{A}_V|^m,$$

where the sum is taken over all k -dimensional subspaces V of \mathbb{F}_q^n . It follows that

$$S_k = q^{(n^2-nk+k^2)m} \prod_{i=0}^{k-1} (q^n - q^i) \prod_{i=0}^{k-1} (q^k - q^i)^{-1}.$$

Using the inequality $\frac{q^n - q^i}{q^k - q^i} \leq q^{n-k} \frac{q}{q-1}$ we get

$$S_k \leq q^{n^2m - (m-1)k(n-k)} \left(\frac{q}{q-1}\right)^k.$$

Note that $S_k = S_{n-k}$ (by duality). Since $\frac{q}{q-1} \leq 2$ and $k(n-k) \geq n-1$, we have

$$\Sigma_a := \sum_{k=1}^{n-1} S_k \leq 2 \sum_{k=1}^{\lfloor n/2 \rfloor} S_k \leq 2^{(n+4)/2} q^{n^2m - (m-1)(n-1)}.$$

For a prime divisor s of n define T_s as the sum $\sum |\mathcal{A}|^m$, where the sum is over all subalgebras of $M_n(\mathbb{F}_q)$ isomorphic to $M_{n/s}(\mathbb{F}_{q^s})$. By Lemma 7.3, we have

$$\begin{aligned} T_s &= q^{(n^2/s)m} \cdot s^{-1} \cdot \prod_{\substack{s \nmid i \\ 1 \leq i < n}} (q^n - q^i) \leq s^{-1} \cdot q^{(n^2/s)m} \cdot q^{n(n-n/s)} \\ &\leq s^{-1} \cdot q^{n^2(m+1)/2}. \end{aligned}$$

Let $\Sigma_b = \sum T_s$, where the sum is over all prime divisors s of n . It is easy to see that the sum $\sum s^{-1}$ of all reciprocals of prime divisors of n does not exceed $2^{(n+4)/2}$. Furthermore, $q^{n^2(m+1)/2} \leq q^{n^2m - (m-1)(n-1)}$. It follows that

$$\Sigma_b \leq 2^{(n+4)/2} q^{n^2m - (m-1)(n-1)}.$$

By Proposition 7.4 we have $\Sigma_a + \Sigma_b = \sum |\mathcal{A}|^m$, where the sum is taken over all maximal subalgebras \mathcal{A} of $M_n(\mathbb{F}_q)$. Thus,

$$\mathfrak{g}_{m,n}(q) \geq q^{mn^2} - 2^{(n+6)/2} q^{n^2m - (m-1)(n-1)}. \quad \square$$

As an immediate consequence of Proposition 7.9 we get the following corollary.

Corollary 7.10. *Let $m, n \geq 2$. The probability that m matrices in $M_n(\mathbb{F}_q)$, chosen under the uniform distribution, generate the \mathbb{F}_q -algebra $M_n(\mathbb{F}_q)$ tends to 1 as $q + m + n \rightarrow \infty$.*

Corollary 7.10 proves and vastly generalizes the conjectural formula [Petrenko and Sidki 2007, (17), p. 27].

8. Finite products of matrix algebras over rings of algebraic integers

Let R be the ring of integers in a number field K . In this final section we apply the techniques developed in our paper to investigate generators of R -algebras A which are products of a finite number of matrix algebras over R . Thus we have

$$A \cong \prod_{i=1}^s M_{n_i}(R)^{m_i},$$

where $1 \leq n_1 < n_2 < \dots < n_s$ and m_i are positive integers. As we have seen in Example 2.13, the algebra A is k -generated if and only if all the algebras $M_{n_i}(R)^{m_i}$ are k -generated. Thus, we may and will focus on the case when $A \cong M_n(R)^m$ for some positive integers n, m . We have the following theorem.

Theorem 8.1. *Let R be the ring of integers in a number field K . Suppose that either $n \geq 3$ or $k \geq 3$ and let $A = M_n(R)^m$ for some positive integer m . Then the following conditions are equivalent.*

- (i) *The R -algebra A admits k generators.*
- (ii) *For every maximal ideal \mathfrak{p} of R the R/\mathfrak{p} -algebra $M_n(R/\mathfrak{p})^m$ admits k generators.*
- (iii) *The density $\text{den}_k(A)$ is positive.*

Furthermore, the following formulas, in which ζ_K denotes the Dedekind zeta function of K , hold for every $k \geq 2$:

- (a) $\text{den}_2(M_2(R)^m) = 0$ for every m ;
- (b) $\text{den}_k(M_2(R)) = \frac{1}{\zeta_K(k-1)\zeta_K(k)}$;
- (c) $\text{den}_k(M_3(R)) = \frac{1}{\zeta_K(2k-2)\zeta_K(k)} \prod_{\mathfrak{p} \in \text{m-Spec } R} \left(1 + \frac{\phi_k(\mathbf{N}(\mathfrak{p}))}{\mathbf{N}(\mathfrak{p})^{3k-2}}\right)$, where $\phi_k(x) = x^{2k-2} - x^k - 2x^{k-1} - x^{k-2} + x + 1$.

Proof. The implications (i) \Rightarrow (ii) and (iii) \Rightarrow (i) are clear. When $k \geq 3$, the implication (ii) \Rightarrow (iii) is an immediate consequence of Theorem 5.2 and the fact that the K -algebra $M_n(K)^m$ is 2-generated [Mazur and Petrenko 2009]. Suppose now that $k = 2$, $n \geq 3$, and (ii) holds. Consider a maximal ideal \mathfrak{p} of R and let $q = \mathbf{N}(\mathfrak{p})$. By Theorem 6.3, the number $g_2(\mathfrak{p}, A)$ of pairs of elements which generate $M_n(R/\mathfrak{p})^m$ is given by

$$g_2(\mathfrak{p}, A) = \prod_{i=0}^{m-1} (g_{2,n}(q) - i \cdot |\text{PGL}_n(\mathbb{F}_q)|).$$

By (ii), we have $g_2(\mathfrak{p}, A) > 0$. Note that $|\mathrm{PGL}_n(\mathbb{F}_q)| \leq q^{n^2-1} \leq q^{2n^2-n+1}$. Furthermore, we have $g_{2,n}(q) \geq q^{2n^2} - 2^n q^{2n^2-n+1}$ by Proposition 7.9. Hence

$$g_2(\mathfrak{p}, A) \geq (\mathbf{N}(\mathfrak{p})^{2n^2} - (2^n + m) \mathbf{N}(\mathfrak{p})^{2n^2-n+1})^m,$$

provided $\mathbf{N}(\mathfrak{p}) > 2^n + m$. By Theorem 3.2, we have

$$\mathrm{den}_2(A) = \prod_{\mathfrak{p} \in m\text{-Spec } R} \frac{g_2(\mathfrak{p}, A)}{\mathbf{N}(\mathfrak{p})^{2mn^2}}.$$

Since all the factors in the product on the right are positive and all but a finite number of them satisfy the inequality

$$\frac{g_2(\mathfrak{p}, A)}{\mathbf{N}(\mathfrak{p})^{2mn^2}} \geq \left(1 - \frac{m + 2^n}{\mathbf{N}(\mathfrak{p})^{n-1}}\right)^m,$$

the product converges to a positive number. In other words, $\mathrm{den}_2(A) > 0$. This completes the proof of the implication (ii) \Rightarrow (iii).

In order to establish formulas (b) and (c) note that

$$\mathrm{den}_k(M_n(R)) = \prod_{\mathfrak{p} \in m\text{-Spec } R} \frac{g_{k,n}(\mathbf{N}(\mathfrak{p}))}{\mathbf{N}(\mathfrak{p})^{kn^2}}$$

by Theorem 3.2. Formulas (b) and (c) follow now from (9) and (10), respectively. To justify (a) note that $g_2(\mathfrak{p}, M_2(R)^m) \leq g_{2,2}(\mathbf{N}(\mathfrak{p}))^m$ for every maximal ideal \mathfrak{p} . It follows that $\mathrm{den}_2(M_2(R)^m) \leq \mathrm{den}_2(M_2(R))^m$. Since by (b) with $k = 2$ we have $\mathrm{den}_2(M_2(R)) = 0$, the equality in (a) follows. \square

Recall now that by Theorem 6.3, the R/\mathfrak{p} -algebra $M_n(R/\mathfrak{p})^m$ is k -generated if and only if $m \leq \mathrm{gen}_{k,n}(\mathbf{N}(\mathfrak{q}))$, where $\mathrm{gen}_{k,n}(q) = g_{k,n}(q)/|\mathrm{PGL}_n(\mathbb{F}_q)|$. Using (10) we get the following theorem.

Theorem 8.2. *Let R be the ring of integers in a number field and let \mathfrak{p} be a maximal ideal of R with smallest norm. Define polynomials $f_k(x)$ by $f_1(x) = 0$ and*

$$f_k(x) = \frac{x^{3k+1}(x^{k-1} - 1)(x^{k-1} + 1)(x^k - 1)}{(x^2 + x + 1)(x - 1)^2(x + 1)} \times (x^{3k-2} + x^{2k-2} - x^k - 2x^{k-1} - x^{k-2} + x + 1) \quad (12)$$

for any $k \geq 2$. Let $k \geq 2$ and m be positive integers. Then the following conditions are equivalent:

- (i) $r(M_3(R)^m, R) = k$;
- (ii) $f_{k-1}(\mathbf{N}(\mathfrak{p})) < m \leq f_k(\mathbf{N}(\mathfrak{p}))$.

In particular, the \mathbb{Z} -algebra $M_3(\mathbb{Z})^m$ is 2-generated if and only if $m \leq 768$.

Proof. By (10), we have $\text{gen}_{k,3}(q) = f_k(q)$ for any $k \geq 2$. By Theorem 8.1, the R -algebra $M_3(R)^m$ is k -generated if and only if $m \leq f_k(N(\mathfrak{q}))$ for every maximal ideal \mathfrak{q} of R . It is easy to see that $f_k(x)$ is increasing on $[2, \infty)$. It follows that $M_3(R)^m$ is k -generated if and only if $m \leq f_k(N(\mathfrak{p}))$. This establishes the equivalence of (i) and (ii). The last claim follows now from the fact that $f_2(2) = 768$. \square

Even though in (9) we established a formula for $\text{gen}_{k,2}(q)$, getting an analog of Theorem 8.2 for products of copies of $M_2(R)$ is more complicated. The difficulty is that the density $\text{den}_2(M_2(R)^m)$ is 0 and we have to find a way to deal with the ambiguity in Theorem 5.5 when $k = 2$. So far we can overcome this difficulty only when R has a maximal ideal of norm 2. We have the following theorem.

Theorem 8.3. *Let R be the ring of integers in a number field and let \mathfrak{p} be a maximal ideal of R with smallest norm. Define polynomials $h_k(x)$ by $h_1(x) = 0$ and*

$$h_k(x) = \frac{x^{2k}(x^{k-1} - 1)(x^k - 1)}{(x - 1)(x + 1)} \quad (13)$$

for any $k \geq 2$. Let $k > 3$ and m be positive integers. Then the following conditions are equivalent:

- (i) $r(M_2(R)^m, R) = k$;
- (ii) $h_{k-1}(N(\mathfrak{p})) < m \leq h_k(N(\mathfrak{p}))$.

Furthermore, there exists an integer t such that $16 \leq t \leq h_2(N(\mathfrak{p}))$, $M_2(R)^m$ is 2-generated if and only if $m \leq t$, and $r(M_2(R)^m, R) = 3$ if and only if $t < m \leq h_3(N(\mathfrak{p}))$. In particular, if $N(\mathfrak{p}) = 2$, then $t = 16$, so in this case (i) and (ii) are equivalent for all $k \geq 2$.

Proof. By (9), we have $\text{gen}_{k,2}(q) = h_k(q)$ for any $k \geq 2$. Suppose that $k \geq 3$. By Theorem 8.1, the R -algebra $M_2(R)^m$ is k -generated if and only if $m \leq h_k(N(\mathfrak{q}))$ for every maximal ideal \mathfrak{q} of R . It is easy to see that $h_k(x)$ is increasing on $[2, \infty)$. It follows that when $k \geq 3$ then $M_2(R)^m$ is k -generated if and only if $m \leq h_k(N(\mathfrak{p}))$. This, in particular, justifies the equivalence of (i) and (ii) when $k > 3$. It also implies the existence of t having all the required properties except possibly the estimate $t \geq 16$. In order to show that $t \geq 16$, we need to establish that $M_2(R)^{16}$ is 2-generated as an R -algebra. It suffices to prove that $M_2(\mathbb{Z})^{16}$ admits two generators as a \mathbb{Z} -algebra. This will be done in Proposition 8.9. Finally, the equality $t = 16$ when $N(\mathfrak{p}) = 2$ follows from the fact that $h_2(2) = 16$. \square

In order to improve on Theorem 8.3 and extend it to matrix algebras of size $n \geq 3$ the following two questions need to be answered.

Question 8.4. Is it true that $t = h_2(N(\mathfrak{p}))$?

Question 8.5. Given positive integers k and n , is $\text{gen}_{k,n}(q)$ an increasing function of q ?

It order to complete our proof of Theorem 8.3 we have to show that $M_2(\mathbb{Z})^{16}$ admits two generators. For that we need several observations, which seem of independent interest.

Proposition 8.6. *Let S be a commutative ring. Two matrices $A, B \in M_2(S)$ generate $M_2(S)$ as an S -algebra if and only if $\det(AB - BA)$ is invertible in S .*

Proof. First we prove the result under the additional assumption that S is a field. Let $N = AB - BA$ and let T be the subalgebra generated by A and B . If T is a proper subalgebra then its dimension is at most 3. It follows that $T/J(T)$ is a semisimple algebra of dimension ≤ 3 (recall that $J(T)$ denotes the Jacobson radical of T). Thus $T/J(T)$ is abelian and therefore $N \in J(T)$. Since the Jacobson radical is nilpotent, N is nilpotent, hence $\det N = 0$.

Conversely, suppose that $\det N = 0$. Recall that any 2×2 matrix X satisfies the identity $X^2 = t_X X - d_X I$, where t_X is the trace and d_X is the determinant of X . Since the trace of N is 0, we have $N^2 = 0$. If $N = 0$ then T is commutative, and hence a proper subalgebra of $M_2(S)$. If $N \neq 0$, the null-space of N is one-dimensional. Using the identity $A^2 = t_A A - d_A I$ we easily see that $AN + NA = t_A N$. It follows that the null-space of N is A -invariant. Similarly, the null-space of N is B -invariant. It follows that the null-space of N is T -invariant, hence T is a proper subalgebra of $M_2(S)$. This completes our proof in the case when S is a field.

If S is any commutative ring then, by Lemma 2.6, the matrices A and B generate $M_2(S)$ if and only if for any maximal ideal M of S the S/M -algebra $M_2(S/M)$ is generated by the images of A and B . By the just established field case of the result, this is equivalent to the condition that $\det(AB - BA) \notin M$ for all maximal ideals M , which in turn is equivalent to claiming that $\det(AB - BA)$ is invertible in S . □

The following observation is due to H. W. Lenstra.

Lemma 8.7. *Let $A, B \in M_2(\mathbb{Z})$ be two matrices with all entries in $\{0, 1\}$. Then A, B generate $M_2(\mathbb{Z})$ if and only if their reductions modulo 2 generate $M_2(\mathbb{F}_2)$.*

Proof. By Proposition 8.6, we need to prove that $\det(AB - BA)$ is odd if and only if it is ± 1 . The “if” part is clear. Suppose then that

$$A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \text{ and } B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}$$

are such that $a_i, b_j \in \{0, 1\}$ and $\det(AB - BA)$ is odd. Note that

$$AB - BA = \begin{pmatrix} a_2b_3 - a_3b_2 & a_2(b_4 - b_1) + b_2(a_1 - a_4) \\ a_3(b_1 - b_4) + b_3(a_4 - a_1) & a_3b_2 - a_2b_3 \end{pmatrix}.$$

The diagonal entries of this matrix are in $\{0, \pm 1\}$ and the off-diagonal entries are in the set $\{0, \pm 1, \pm 2\}$. If $a_2b_3 - a_3b_2 = 0$ then the off-diagonal entries must be

odd and hence are ± 1 . It follows that $\det(AB - BA) = \pm 1$. The same conclusion holds if one of the off-diagonal entries is 0. Suppose now that $a_2b_3 - a_3b_2 = \pm 1$ and the off-diagonal entries are not 0. Then one of the off-diagonal entries, say $a_3(b_1 - b_4) + b_3(a_4 - a_1)$, must be even and nonzero (the other possibility is handled in the same way). This can only happen if $a_3 = b_3 = 1$ and $b_1 - b_4 = a_4 - a_1 = \pm 1$. It follows that one of a_2, b_2 is 0 and the other is 1. Thus $\det(AB - BA) = -1 - (\mp 1)(\pm 2) = 1$. \square

Lemma 8.8. *Let $A, B, A', B' \in M_2(\mathbb{Z})$ be matrices with all entries in $\{0, 1\}$ such that each pair A, B and A', B' generates $M_2(\mathbb{Z})$. If there is an odd prime p such that the reductions modulo p of (A, B) and (A', B') are conjugate in $M_2(\mathbb{F}_p)$ then the pairs (A, B) and (A', B') are conjugate in $M_2(\mathbb{Z})$.*

Proof. For a pair of 2×2 matrices X and Y define

$$\text{conj}(X, Y) = (\text{tr}(X), \det(X), \text{tr}(Y), \det(Y), \text{tr}(XY)).$$

It follows from [Mazur and Petrenko 2009, Theorem 2] that for any principal ideal domain R and any two pairs (X, Y) and (X', Y') of elements in $M_2(R)$ which generate $M_2(R)$ as an R -algebra we have $\text{conj}(X, Y) = \text{conj}(X', Y')$ if and only if $X' = CXC^{-1}$ and $Y' = CYC^{-1}$ for some invertible matrix $C \in M_2(R)$ (in [Mazur and Petrenko 2009] the fifth component of conj is $\det(X + Y)$ but it is equivalent to the version above by the following identity for 2×2 matrices:

$$\text{tr}(X) \text{tr}(Y) - \text{tr}(XY) + \det(X) + \det(Y) - \det(X + Y) = 0.)$$

Under the assumptions of the lemma, the traces of A, B, A', B' are in $\{0, 1, 2\}$ and the determinants of these matrices are in $\{-1, 0, 1\}$. Our assumption that $\text{conj}(A, B) \equiv \text{conj}(A', B') \pmod{p}$ implies then that $\text{tr} A = \text{tr} A'$, $\det A = \det A'$, $\text{tr} B = \text{tr} B'$, and $\det B = \det B'$. It remains to prove that $\text{tr}(AB) = \text{tr}(A'B')$. Let

$$A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}, \quad A' = \begin{pmatrix} a'_1 & a'_2 \\ a'_3 & a'_4 \end{pmatrix}, \quad B' = \begin{pmatrix} b'_1 & b'_2 \\ b'_3 & b'_4 \end{pmatrix}.$$

Then $\text{tr}(AB) = a_1b_1 + a_2b_3 + a_3b_2 + a_4b_4$ and $\text{tr}(A'B') = a'_1b'_1 + a'_2b'_3 + a'_3b'_2 + a'_4b'_4$. Both these numbers belong to $\{0, 1, 2, 3, 4\}$. Suppose that these numbers are different. Since they are congruent modulo p , we see that $p = 3$ and one of these numbers is in $\{0, 4\}$. If $\text{tr}(AB) = 4$ then all the entries a_i and b_j must be 1 so $A = B$, which is not possible. Thus we may assume that $\text{tr}(AB) = 0$ and then $\text{tr}(A'B') = 3$. If $\text{tr}(A) = 0$ then $\text{tr}(A') = 0$, so $a'_1 = a'_4 = 0$ and therefore $\text{tr}(A'B') \leq 2$, a contradiction. Thus $\text{tr}(A) \neq 0$ and in the same way we show that $\text{tr}(B) \neq 0$. If $\text{tr}(A) = 2$ then $a_1 = a_4 = 1$ so $b_1 = b_4 = 0$ and $\text{tr}(B) = 0$, which we have just proved impossible. This shows that $\text{tr}(A) = 1$ and a similar argument yields $\text{tr}(B) = 1$. Thus $\text{tr}(A') = 1 = \text{tr}(B')$. It follows that one of a'_1 and a'_4 is

0. We may assume that $a'_1 = 0$ (the same argument works when $a'_4 = 0$). Then $a'_2 = a'_3 = a'_4 = b'_2 = b'_3 = b'_4 = 1$ and consequently $b'_1 = 0$ and $A' = B'$, a contradiction. \square

We have now the following curious proposition.

Proposition 8.9. *Let x and y be two elements of $M_2(\mathbb{Z})^k$ such that every component of x and y is a matrix whose all entries are in $\{0, 1\}$. Suppose that x, y , considered as elements of $M_2(\mathbb{F}_2)^k$, generate the algebra $M_2(\mathbb{F}_2)^k$. Then x, y generate $M_2(\mathbb{Z})^k$ as a ring. In particular, the ring $M_2(\mathbb{Z})^{16}$ admits two generators.*

Proof. Let $x = (X_1, \dots, X_k), y = (Y_1, \dots, Y_k)$. By Lemma 8.7, each pair (X_i, Y_i) generates $M_2(\mathbb{Z})$. According to Lemma 2.6 and Theorem 6.1, it suffices to prove that for any prime p and any $1 \leq i < j \leq k$, the pairs (X_i, Y_i) and (X_j, Y_j) are not conjugate modulo p . For $p = 2$ this follows from our assumptions and Theorem 6.1. Consequently, the pairs (X_i, Y_i) and (X_j, Y_j) are not conjugate in $M_2(\mathbb{Z})$ whenever $i \neq j$. By Lemma 8.8, the pairs (X_i, Y_i) and (X_j, Y_j) are not conjugate modulo p for any odd prime p . This proves the first part of the proposition.

Since $\text{gen}_{2,2}(2) = 16$ by (9), the algebra $M_2(\mathbb{F}_2)^{16}$ is two-generated. It follows from the first part of the proposition that $M_2(\mathbb{Z})^{16}$ admits two generators. \square

Remark 8.10. We would like to point out that one should not expect any analogs of Proposition 8.9 for matrix rings of size larger than 2. For example, consider the matrices

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Considered as matrices over the field \mathbb{F}_3 with three elements they have a common eigenvector $(1, -1, 1)^t$. Thus these matrices do not generate $M_3(\mathbb{F}_3)$, hence they do not generate $M_3(\mathbb{Z})$. Consider now these matrices as matrices over \mathbb{F}_2 . If they do not generate $M_3(\mathbb{F}_2)$, then they are contained in a maximal subalgebra of $M_3(\mathbb{F}_2)$. By Proposition 7.4, the maximal subalgebra is either a field or it fixes a nontrivial proper subspace. Since $A^2 = A$, the former case is not possible. In the latter case, A and B have a common eigenvector either in their action on column vectors or in their action on row vectors. It is however a straightforward verification to see that no such common eigenvector exists. Thus A and B generate the algebra $M_3(\mathbb{F}_2)$. In fact, in the same way one can see that they generate $M_3(\mathbb{F}_p)$ for any prime p different from 3. With a bit more work, one can see that the subalgebra of $M_3(\mathbb{Z})$ generated by A and B has index 9. Note that by (10), there are 129024 ordered pairs of 3×3 matrices with entries in $\{0, 1\}$, which considered as elements of $M_3(\mathbb{F}_2)$ generate the algebra $M_3(\mathbb{F}_2)$. Tsvetomira Radeva, at our request, performed computations using Java and GAP and found that among them exactly 9132 pairs

do not generate $M_3(\mathbb{Z})$. The computations are based on a result of [Paz 1984] and use the LLL algorithm [Lenstra et al. 1982; Pohst 1987].

We end with the following curious observation. In Theorem 8.1 we defined a family of polynomials $\phi_k(x)$, $k \geq 2$. The polynomial $x^{3k-2} + \phi_k(x)$ is a factor of the polynomial f_k defined in Theorem 8.2. Define polynomials $\psi_k(x)$ as follows:

$$\psi_k(x) = \begin{cases} \frac{x^{3k-2} + \phi_k(x)}{x-1} & \text{if } k \equiv 0, 4 \pmod{6}, \\ \frac{x^{3k-2} + \phi_k(x)}{x^2-1} & \text{if } k \equiv 1, 3 \pmod{6}, \\ \frac{x^{3k-2} + \phi_k(x)}{x^3-1} & \text{if } k \equiv 2 \pmod{6}, \\ \frac{x^{3k-2} + \phi_k(x)}{(x+1)(x^3-1)} & \text{if } k \equiv 5 \pmod{6}. \end{cases} \quad (14)$$

Computations with Maxima show that the polynomials ϕ_k and ψ_k are irreducible for $k \leq 250$. While the polynomials ϕ_k have only six nonzero coefficients, the polynomials ψ_k have complicated structure. For example,

$$\begin{aligned} \psi_{12}(x) = & x^{33} + x^{32} + x^{31} + x^{30} + x^{29} + x^{28} + x^{27} + x^{26} + x^{25} + x^{24} + x^{23} + x^{22} \\ & + 2x^{21} + 2x^{20} + 2x^{19} + 2x^{18} + 2x^{17} + 2x^{16} + 2x^{15} + 2x^{14} + 2x^{13} + 2x^{12} \\ & + x^{11} - x^{10} - 2x^9 - 2x^8 - 2x^7 - 2x^6 - 2x^5 - 2x^4 - 2x^3 - 2x^2 - 2x - 1. \end{aligned}$$

Nevertheless, it seems that all the coefficients of ψ_k are in the set $\{-2, -1, 0, 1, 2\}$. Even though we do not have at present any conceptual reason for it, we propose the following intriguing conjecture.

Conjecture 8.11. *The polynomials ϕ_k and ψ_k are irreducible.*

Acknowledgments

It is our pleasure to thank Max Alekseyev, Nigel Boston, Evgeny Gordon, Rostislav Grigorchuk, Ilya Kapovich, Martin Kassabov, Hendrik Lenstra, Pieter Moree, Tsvetomira Radeva, Peter Sarnak, Said Sidki, John Tate, and Paula Tretkoff. The first author was supported by NSF Grant DMS-0456185. The third author thanks the Max Planck Institute for Mathematics for the warm hospitality, unique research opportunities, and financial support during his visit in July–August of 2009.

References

- [Arnold 2009] V. I. Arnold, “Uniform distribution of indivisible vectors in the space of integers”, *Izv. Ross. Akad. Nauk Ser. Mat.* **73**:1 (2009), 21–30. In Russian; translated in *Izv. Math.* **73**:1 (2009), 21–29. MR 2010h:60023

- [Cox et al. 2005] D. A. Cox, J. Little, and D. O’Shea, *Using algebraic geometry*, 2nd ed., Graduate Texts in Mathematics **185**, Springer, New York, 2005. MR 2005i:13037 Zbl 1079.13017
- [Ekedahl 1991] T. Ekedahl, “An infinite version of the Chinese remainder theorem”, *Comment. Math. Univ. St. Paul.* **40**:1 (1991), 53–59. MR 92h:11027 Zbl 0749.11004
- [Hall 1936] P. Hall, “The Eulerian functions of a group”, *Q. J. Math* **7** (1936), 134–151. Zbl 0014.10402
- [Kravchenko and Petrenko 2006] R. V. Kravchenko and B. V. Petrenko, “Some formulas for the smallest number of generators for finite direct sums of matrix algebras”, preprint, 2006. arXiv math/0611674
- [Lenstra et al. 1982] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, “Factoring polynomials with rational coefficients”, *Math. Ann.* **261**:4 (1982), 515–534. MR 84a:12002 Zbl 0488.12001
- [Loomis and Whitney 1949] L. H. Loomis and H. Whitney, “An inequality related to the isoperimetric inequality”, *Bull. Amer. Math. Soc* **55** (1949), 961–962. MR 11,166d Zbl 0035.38302
- [Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics **8**, Cambridge University Press, 1986. MR 88h:13001 Zbl 0603.13001
- [Mazur and Petrenko 2009] M. Mazur and B. V. Petrenko, “Separable algebras over infinite fields are 2-generated and finitely presented”, *Arch. Math. (Basel)* **93**:6 (2009), 521–529. MR 2011b:16066 Zbl 1190.16026
- [Narkiewicz 1990] W. Narkiewicz, *Elementary and analytic theory of algebraic numbers*, 2nd ed., Springer, Berlin, 1990. MR 91h:11107 Zbl 0717.11045
- [Paz 1984] A. Paz, “An application of the Cayley-Hamilton theorem to matrix polynomials in several variables”, *Linear and Multilinear Algebra* **15**:2 (1984), 161–170. MR 85h:15019 Zbl 0536.15007
- [Petrenko and Sidki 2007] B. V. Petrenko and S. N. Sidki, “On pairs of matrices generating matrix rings and their presentations”, *J. Algebra* **310**:1 (2007), 15–40. MR 2008j:16088 Zbl 1122.16017
- [Pleasants 1974] P. A. B. Pleasants, “The number of generators of the integers of a number field”, *Mathematika* **21** (1974), 160–167. MR 52 #3122 Zbl 0328.12008
- [Pohst 1987] M. Pohst, “A modification of the LLL reduction algorithm”, *J. Symbolic Comput.* **4**:1 (1987), 123–127. MR 89c:11183 Zbl 0629.10001
- [Poonen 2003] B. Poonen, “Squarefree values of multivariable polynomials”, *Duke Math. J.* **118**:2 (2003), 353–373. MR 2004d:11094 Zbl 1047.11021
- [Schmidt 1976] W. M. Schmidt, *Equations over finite fields: an elementary approach*, Lecture Notes in Mathematics **536**, Springer, Berlin, 1976. MR 55 #2744 Zbl 0329.12001

Communicated by Hendrik W. Lenstra

Received 2010-05-09

Revised 2011-01-08

Accepted 2011-02-06

rkchenko@gmail.com

*Laboratoire de Mathématique d’Orsay, Université Paris-Sud,
91405 Orsay Cedex, France*

mazur@math.binghamton.edu

*Department of Mathematical Sciences, Binghamton
University, Binghamton, NY 13902-6000, United States*

bpetren@brockport.edu

*Department of Mathematics, SUNY College at Brockport,
350 New Campus Drive, Brockport, NY 14420, United States*

Moving lemma for additive higher Chow groups

Amalendu Krishna and Jinhyun Park

We study additive higher Chow groups with several modulus conditions. Apart from exhibiting the validity of all known results for the additive Chow groups with these modulus conditions, we prove the moving lemma for them: for a smooth projective variety X and a finite collection \mathcal{W} of its locally closed algebraic subsets, every additive higher Chow cycle is congruent to an admissible cycle intersecting properly all members of \mathcal{W} times faces. This is the additive analogue of the moving lemma for the higher Chow groups studied by S. Bloch and M. Levine.

As an application, we prove that any morphism from a quasiprojective variety to a smooth projective variety induces a pull-back map of additive higher Chow groups. More important applications of this moving lemma are derived in two separate papers by the authors.

1. Introduction

Working with algebraic cycles, formal finite sums of closed subvarieties of a variety, often requires some forms of moving results, as differential geometry often requires Sard's lemma. A classical example is Chow's moving lemma [1956], which moves algebraic cycles under rational equivalence. A modern version for higher Chow groups [Bloch 1994; Levine 1998] shows that, for a smooth quasiprojective variety X and a finite set of locally closed subvarieties of X , one can move (modulo boundaries) admissible cycles to other admissible cycles that intersect a given finite set of subvarieties in the right codimensions. Any such result on moving of cycles is generally referred to as a moving lemma. Such moving results have played a very crucial role in the development and application of the theory of higher Chow groups. For instance, one major application was the construction of a triangulated category of mixed motives over k [Hanamura 2004].

The primary goal of this paper is to prove this latter kind of moving lemma for additive higher Chow groups of a smooth and projective variety, which will

MSC2000: primary 14C25; secondary 19E15.

Keywords: Chow group, algebraic cycle, moving lemma.

serve as an important technical tool in the study of additive higher Chow groups. Already, this moving lemma for additive Chow groups has been crucially used in [Krishna and Park 2011; 2012] for proving some important results about additive higher Chow groups. We expect that this will have many more applications in the future study of additive Chow groups and the infinitesimal K -theory of smooth varieties.

Additive Chow groups of 0-cycles on a field were first introduced in [Bloch and Esnault 2003b] in an attempt to describe the K -theory and motivic cohomology of the ring of dual numbers via algebraic cycles. Bloch and Esnault [2003a] later defined these groups by putting a *modulus* condition on additive Chow cycles in the hope of describing the K -groups of any given truncated polynomial ring over a field. The additive higher Chow groups of any given variety were defined in the most general form in [Park 2009] and were later studied in more detail in [Krishna and Levine 2008], where many nice properties of these groups were established.

The most crucial part of existing definitions of additive higher Chow groups, which makes them distinct from the higher Chow groups, is the *modulus* condition on the admissible additive cycles. This condition also brings an extra subtlety which does not appear in the theory of higher Chow groups. As conjectured in [Krishna and Levine 2008; Park 2009], additive higher Chow groups are expected to complement higher Chow groups for nonreduced schemes so as to obtain the right motivic cohomology groups. In particular, for a smooth projective variety X , one expects an Atiyah–Hirzebruch spectral sequence

$$\mathrm{TH}^{-q}(X, -p - q; m) \Rightarrow K_{-p-q}^{\mathrm{nil}}(X; m), \quad (1-1)$$

where $K^{\mathrm{nil}}(X; m)$ is the homotopy fiber of the restriction map

$$K(X \times \mathrm{Spec}(k[t])) \rightarrow K(X \times \mathrm{Spec}(k[t]/t^{m+1})).$$

Since these statements are still conjectural, it is not clear if the modulus conditions used to study additive higher Chow groups of varieties in the literature are the right ones to give the correct motivic cohomology, for example, ones which would satisfy (1-1). One goal of this paper is to exhibit that the modulus condition (which we call M_{sup} in this paper) used in [Krishna and Levine 2008] may not be the best possible one.

We study the theory of additive Chow groups based on two other modulus conditions in this paper: $M = M_{\mathrm{sum}}$ is based on the modulus condition used in [Bloch and Esnault 2003a; Rülling 2007], and $M = M_{\mathrm{ssup}}$ is a new modulus condition introduced in this paper. Although this new modulus condition M_{ssup} may appear to be mildly stronger than the one used in [Krishna and Levine 2008; Park 2009], it turns out that the resulting additive Chow groups have all the properties known for the additive Chow groups of [Bloch and Esnault 2003a; Krishna and Levine

2008; Park 2009]. In addition, we prove many other crucial structural properties of additive higher Chow groups based on the modulus conditions M_{sum} and M_{ssup} . More important properties are discussed in [Krishna and Park 2011; 2012].

As in the case of higher Chow groups, any theory of additive motivic cohomology which would compute the K -theory as in (1-1) is expected to have a form of moving lemma to make it more amenable to deeper study. The central result of the paper is the following moving lemma:

Theorem 4.1. *For a smooth projective variety X and a finite collection \mathcal{W} of its locally closed algebraic subsets, every additive higher Chow cycle is congruent to an admissible cycle intersecting properly all members of \mathcal{W} times faces. In other words, the inclusion of complexes*

$$\text{TZ}_{\mathcal{W}}^q(X, \cdot ; m) \hookrightarrow \text{TZ}^q(X, \cdot ; m)$$

is a quasiisomorphism.

This is the additive analogue of the moving lemma for the higher Chow groups studied by S. Bloch and M. Levine.

It is known that the moving lemma for all smooth quasiprojective varieties indirectly implies other properties such as \mathbb{A}^1 -homotopy invariance and localization sequences. But these clearly fail for the additive Chow groups. This suggests that the above moving lemma may not be valid for some smooth quasiprojective varieties. A concrete quasiprojective example, where the standard arguments fail, is given in Example 8.2.

Our proof of the above result is broadly speaking based on the techniques of [Bloch 1986; Levine 1998] where the analogous result for the higher Chow groups is proven. However, the main difficulty with the techniques of both these works is that their arguments are mostly intersection theoretic and are not equipped to handle the more delicate modulus condition of additive Chow cycles. So these arguments cannot be directly transported to the additive world. This has made people believe that the additive Chow group may not satisfy the moving lemma.

We achieve the goal by our new *containment*-type argument (see Proposition 2.4) and construction of the additive version of a chain homotopy variety in Section 5. Using these results and Proposition 5.2, we show that we can keep track of the modulus condition whenever we need to move an additive cycle. On the log-additive higher Chow groups of [Krishna and Levine 2008], one can prove the moving lemma for any general smooth quasiprojective varieties using our main theorem.

As the first application of the moving lemma, we establish the contravariant functoriality property of the additive higher Chow groups in the most general form:

Theorem 7.1. *For a morphism $f : X \rightarrow Y$ of quasiprojective varieties over a field k , where Y is smooth and projective, there is a pull-back map*

$$f^* : \mathrm{TH}^q(Y, n; m) \rightarrow \mathrm{TH}^q(X, n; m),$$

and this satisfies the expected composition law.

If X is also smooth and projective, the pull-back map on the additive Chow groups was constructed in [Krishna and Levine 2008] using the action of the higher Chow groups on the additive ones. However, the contravariant functoriality in this general form as above is new, and it is based on a crucial use of the moving lemma (Theorem 4.1) as in the case of general pull-back maps of higher Chow groups [Bloch 1986, Theorem 4.1], and another use of our containment argument to establish the Gysin chain map for regular embeddings. Even in the special case of X being smooth and projective, our proof is different and more direct than the one in [Krishna and Levine 2008].

We give applications of the results in this paper elsewhere. In [Krishna and Park 2011] we investigate the structure of differential graded algebras on the additive higher Chow groups of smooth projective varieties. When $X = \mathrm{Spec}(k)$, this was done in [Rülling 2007]. Higher-dimensional varieties X require involved calculations and arguments as well as the moving lemma and the containment lemma of this paper. As another application of the moving lemma, we showed in [Krishna and Park 2011] that there is an additive analogue of Bloch's *normalized cycle complex* and it is quasiisomorphic to the additive cycle complex. This fact is used to propose and study a motivic cyclic homology theory by constructing a mixed complex in the sense of A. Connes (see [Loday 1998]) from additive higher Chow complexes.

In [Krishna and Park 2012] we apply the moving lemma to construct a triangulated category $\mathcal{DM}(k; m)$ of mixed motives over $k[t]/(t^{m+1})$. This category extends the category of [Hanamura 2004], and some “augmented motives” in the category compute the usual higher Chow groups and the additive higher Chow groups at the same time, as desired originally in [Bloch and Esnault 2003a, §4].

We now outline the structure of this paper. In Section 2, we define our basic objects, the additive higher Chow groups with various modulus conditions. We also prove some preliminary results used repeatedly in the paper. In Section 3, we prove basic properties of these additive Chow groups. In particular, we demonstrate, for the additive higher Chow groups based on the modulus condition M_{ssup} , all those results which are known for the additive higher Chow groups of [Bloch and Esnault 2003a; Krishna and Levine 2008; Park 2009] with slightly different modulus conditions M_{sum} and M_{sup} . Section 4 gives the proofs of further preliminary results needed to prove our moving lemma for the additive higher Chow groups. The subsequent Sections 5 and 6 are devoted to our main result, the moving lemma for additive higher Chow groups. In Section 7, we apply the moving lemma to prove the general contravariant functoriality theorem, Theorem 7.1. In Section 8,

we append some calculations of the additive higher Chow groups we found in the process of working on the problem. This suggests some kind of “pseudo”- \mathbb{A}^1 -homotopy properties of additive higher Chow groups.

Throughout this paper, a *k-scheme*, or a *scheme over k*, is always a separated scheme of finite type over a perfect field *k*. A *k-variety* is an integral *k-scheme*.

2. Additive higher Chow groups

In this section, we define additive higher Chow groups from a more unified perspective than those in the literature by Bloch and Esnault, Rülling, Krishna and Levine, and Park, treating the modulus conditions as “variables”. We also prove some elementary results that are needed to study and compare additive Chow groups based on various modulus conditions.

We begin by fixing some notations which will be used throughout this paper. We write **Sch** / *k*, **Sm** / *k*, and **SmProj** / *k* for the categories of *k-schemes*, smooth quasiprojective varieties, and smooth projective varieties, respectively. We shall let **Sch'** / *k* denote the category of *k-schemes* with only proper maps. $D^-(\mathbf{Ab})$ is the derived category of bounded-above complexes of abelian groups. Recall from [Krishna and Levine 2008; Park 2009] that for a normal variety *X* over *k*, and a finite set of Weil divisors $\{Y_1, \dots, Y_s\}$ on *X*, the supremum of these divisors, denoted by $\sup_{1 \leq i \leq s} Y_i$, is the Weil divisor defined to be

$$\sup_{1 \leq i \leq s} Y_i = \sum_{Y \in \text{Pdiv}(X)} (\max_{1 \leq i \leq s} \text{ord}_Y(Y_i)) [Y], \tag{2-1}$$

where $\text{Pdiv}(X)$ is the set of all prime Weil divisors of *X*. One observes that the set of all Cartier divisors on a normal scheme *X* is contained in the set of all Weil divisors, and the supremum of a collection of Cartier divisors may not remain a Cartier divisor in general, unless *X* is factorial. We shall need some elementary results about Cartier and Weil divisors on normal varieties:

Lemma 2.1. *Let X be a normal variety and let D_1 and D_2 be effective Cartier divisors on X such that $D_1 \geq D_2$ as Weil divisors. Let $Y \subset X$ be a closed subset which intersects D_1 and D_2 properly. Let $f : Y^N \rightarrow X$ be the composite of the inclusion and the normalization of Y_{red} . Then $f^*(D_1) \geq f^*(D_2)$.*

Proof. For any effective Cartier divisor *D* on *X*, let \mathcal{F}_D denote the sheaf of ideals defining *D* as a locally principal closed subscheme of *X*. We first claim that $D_1 \geq D_2$ if and only if $\mathcal{F}_{D_1} \subset \mathcal{F}_{D_2}$. We only need to show the “only if” part, as the other implication is obvious. Now, $D_1 \geq D_2$ implies that $D = D_1 - D_2$ is effective as a Cartier divisor since the group of Cartier divisors forms a subgroup of Weil divisors on a normal scheme. Since $\mathcal{F}_{D_1} \subset \mathcal{F}_{D_2}$ is a local question, we can assume that $X = \text{Spec}(A)$ is a local normal integral scheme and $\mathcal{F}_{D_i} = (a_i)$. Put $a = a_1/a_2$

as an element of the function field of X . We need to show that $a \in A$. Since A is normal, it suffices to show that $a \in A_{\mathfrak{p}}$ for every height-one prime ideal \mathfrak{p} of A . But this is precisely the meaning of $D_1 \geq D_2$. This proves the claim.

Since D_i intersect Y properly, we see that $f^*(D_i)$ is a locally principal closed subscheme of Y^N for $i = 1, 2$. The lemma now follows directly from the above claim. \square

The following is a refinement of [Krishna and Levine 2008, Lemma 3.2].

Lemma 2.2. *Let $f : Y \rightarrow X$ be a surjective map of normal integral k -schemes. Let D be a Cartier divisor on X such that $f^*(D) \geq 0$ on Y . Then $D \geq 0$ on X .*

Proof. As is implicit in the proof of Lemma 2.1, we can localize at the generic points of $\text{Supp}(D)$ and assume that $X = \text{Spec}(A)$, where A is a discrete valuation ring which is essentially of finite type over k . The divisor D is then given by a rational function $a \in K$, where K is the field of fractions of A . Choosing a uniformization parameter π of A , we can write a uniquely as $a = u\pi^n$, where $u \in A^\times$ and $n \in \mathbb{Z}$.

Since f is surjective, there is a closed point $y \in Y$ such that $f(y)$ is the closed point of X . Since Y is integral, the surjectivity of f also implies that the generic point of Y (which is also the generic point of $\text{Spec}(\mathbb{O}_{Y,y})$) must go to the generic point of X under f . Hence the map $\text{Spec}(\mathbb{O}_{Y,y}) \rightarrow X$ is surjective. This implies in particular that the image of π in $\mathbb{O}_{Y,y}$ is a nonzero element of the maximal ideal \mathfrak{m} of the local ring $\mathbb{O}_{Y,y}$. On the other hand, $f^*(D) \geq 0$ implies that as a rational function on Y , a actually lies in $\mathbb{O}_{Y,y}$. Since $u \in \mathbb{O}_{Y,y}^\times$ and $\pi \in \mathfrak{m}$, this can happen only when $n \geq 0$. That is, D is effective. \square

We will assume that a k -scheme X is equidimensional to define the additive Chow groups, although one can easily remove this condition by writing the additive Chow cycles in terms of their dimensions rather than their codimensions. Throughout this paper, for any such scheme X , we shall denote the normalization of X_{red} by X^N . Thus X^N is the disjoint union of the normalizations of all the irreducible components of X_{red} .

Set $\mathbb{A}^1 := \text{Spec } k[t]$, $\mathbb{G}_m := \text{Spec } k[t, t^{-1}]$, $\mathbb{P}^1 := \text{Proj } k[Y_0, Y_1]$, and let $y := Y_1/Y_0$ be the standard coordinate function on \mathbb{P}^1 . We set $\square^n := (\mathbb{P}^1 \setminus \{1\})^n$. For $n \geq 1$, let $B_n = \mathbb{G}_m \times \square^{n-1}$, $\tilde{B}_n = \mathbb{A}^1 \times \square^{n-1}$, $\bar{B}_n = \mathbb{A}^1 \times (\mathbb{P}^1)^{n-1} \supset \tilde{B}_n$, and $\hat{B}_n = \mathbb{P}^1 \times (\mathbb{P}^1)^{n-1} \supset \bar{B}_n$. We use the coordinate system (t, y_1, \dots, y_{n-1}) on \hat{B}_n , with $y_i := y \circ q_i$, where $q_i : \hat{B}_n \rightarrow \mathbb{P}^1$ is the projection onto the i -th \mathbb{P}^1 .

Let $F_{n,i}^1$, for $i = 1, \dots, n - 1$, be the Cartier divisor on \hat{B}_n defined by $\{y_i = 1\}$ and $F_{n,0}$ the Cartier divisor defined by $\{t = 0\}$. Notice that the divisor $F_{n,0}$ is in fact contained in $\bar{B}_n \subset \hat{B}_n$. Let F_n^1 denote the Cartier divisor $\sum_{i=1}^{n-1} F_{n,i}^1$ on \hat{B}_n .

A *face* of B_n is a subscheme F defined by equations of the form

$$y_{i_1} = \epsilon_1, \quad \dots, \quad y_{i_s} = \epsilon_s \quad (\epsilon_j \in \{0, \infty\}).$$

For $\epsilon = 0, \infty$, and $i = 1, \dots, n - 1$, let $\iota_{n,i,\epsilon} : B_{n-1} \rightarrow B_n$ be the inclusion

$$\iota_{n,i,\epsilon}(t, y_1, \dots, y_{n-2}) = (t, y_1, \dots, y_{i-1}, \epsilon, y_i, \dots, y_{n-2}). \quad (2-2)$$

We now define the modulus conditions that we shall consider for defining our additive higher Chow groups.

2A. Modulus conditions.

Definition 2.3. Let X be a k -scheme as above and let V be an integral closed subscheme of $X \times B_n$. Let \bar{V} denote the closure of V in $X \times \widehat{B}_n$ and let

$$v : \bar{V}^N \rightarrow X \times \widehat{B}_n$$

denote the induced map from the normalization of \bar{V} . We fix an integer $m \geq 1$.

- (1) We say that V satisfies the modulus m condition M_{sum} (or the *sum*-modulus condition) on $X \times B_n$ if as Weil divisors on \bar{V}^N ,

$$(m + 1)[v^*(F_{n,0})] \leq [v^*(F_n^1)].$$

This condition was used in [Bloch and Esnault 2003a; Rülling 2007] to study additive Chow groups of 0-cycles on fields.

- (2) We say that V satisfies the modulus m condition M_{sup} (or the *sup*-modulus condition) on $X \times B_n$ if as Weil divisors on \bar{V}^N ,

$$(m + 1)[v^*(F_{n,0})] \leq \sup_{1 \leq i \leq n-1} [v^*(F_{n,i}^1)].$$

This condition was used by in [Krishna and Levine 2008; Park 2009] to define their additive higher Chow groups.

- (3) We say that V satisfies the modulus m condition M_{ssup} (or the *strong sup*-modulus condition) on $X \times B_n$ if there exists an integer $1 \leq i \leq n - 1$ such that

$$(m + 1)[v^*(F_{n,0})] \leq [v^*(F_{n,i}^1)]$$

as Weil divisors on \bar{V}^N .

Since the modulus conditions are defined for a given fixed integer m , we shall often simply say that V satisfies a modulus condition M without mentioning the integer m . Notice that since V is contained in $X \times B_n$, its closure \bar{V} intersects all the Cartier divisors $F_{n,0}$ and $F_{n,i}^1$ ($1 \leq i \leq n - 1$) properly in $X \times \widehat{B}_n$. In particular,

their pull-backs of $F_{n,0}$ and $F_{n,i}^1$ are all effective Cartier divisors on \bar{V}^N . Notice also that

$$M_{\text{ssup}} \Rightarrow M_{\text{sup}} \Rightarrow M_{\text{sum}}. \tag{2-3}$$

The following restriction property of the modulus conditions M_{sum} and M_{ssup} will be used repeatedly in this paper.

Proposition 2.4 (containment lemma). *Let X be a k -scheme, and let $W \subset V$ be irreducible closed subvarieties of $X \times \widehat{B}_n$. If V satisfies M_{sum} , then so does W ; if V satisfies M_{ssup} , then so does W .*

Proof. Let \bar{V} and \bar{W} be the Zariski closures of V and W in $X \times \widehat{B}_n$ and let $\bar{W} \xrightarrow{j} \bar{V}$ be the closed embedding. Let $\nu_1 : \bar{V}^N \rightarrow \bar{V} \hookrightarrow X \times \widehat{B}_n$ be the normalization of \bar{V} :

$$\begin{array}{ccccc}
 W_1^N & \xrightarrow{\bar{g}} & W_1 & \xrightarrow{\bar{j}} & \bar{V}^N \\
 \downarrow f^N & & \downarrow f & & \downarrow \nu_1 \\
 \bar{W}^N & \xrightarrow{g} & \bar{W} & \xrightarrow{j} & \bar{V} \\
 & \searrow \nu_2 & & & \downarrow \nu_1 \\
 & & & & X \times \widehat{B}_n.
 \end{array} \tag{2-4}$$

Let W_1 be $\bar{W} \times_{\bar{V}} \bar{V}^N$, and let f and \bar{j} be the natural projections. Let g and \bar{g} be the normalizations. The map ν_2 is defined so that the lower triangle commutes. By the universal property of normalization, we have a finite surjective morphism $f^N : W_1^N \rightarrow \bar{W}^N$ of normal integral k -schemes that makes the above diagram commutative.

Since $V \cap F_{n,0} = \emptyset$ and $W \neq \emptyset$, we see that $F_{n,0}$ and $F_{n,i}^1$ intersect \bar{W} properly. Now, if V satisfies the modulus condition M_{ssup} , then Lemma 2.1 implies that there is an integer $1 \leq i \leq n - 1$ such that $\bar{g}^* \circ \bar{j}^* [\nu_1^*(F_{n,i}^1 - (m + 1)F_{n,0})] \geq 0$ on W_1^N . In particular, by commutativity, we get $(f^N)^* [\nu_2^*(F_{n,i}^1 - (m + 1)F_{n,0})] \geq 0$ on W_1^N . Since f^N is a finite and surjective map of normal varieties, from Lemma 2.2 we have $[\nu_2^*(F_{n,i}^1 - (m + 1)F_{n,0})] \geq 0$ on \bar{W}^N , that is, W satisfies M_{ssup} too.

The case of M_{sum} follows exactly the same way using F_n^1 instead of $F_{n,i}^1$, noting that F_n^1 is also an effective Cartier divisor. \square

As one can see from the above proposition, although the modulus condition M_{sup} lies between the other two modulus conditions M_{sum} and M_{ssup} , the additive higher Chow groups based on the latter modulus conditions have better structural properties.

In this paper, we study the additive higher Chow groups based on the modulus conditions M_{sum} and M_{ssup} . We shall show in the next section that the additive

Chow groups based on our new modulus condition M_{ssup} satisfy all the properties known to be satisfied by the additive higher Chow groups of Krishna and Levine, Park, Bloch and Esnault, and Rülling.

2B. Additive cycle complex. We define the additive cycle complex based on the above modulus conditions.

Definition 2.5. Let M be a modulus condition—either M_{sum} or M_{ssup} . Let X be a k -scheme, and let r and m be integers with $m \geq 1$.

- (0) $\underline{\text{TZ}}_r(X, 1; m)_M$ is the free abelian group on integral closed subschemes Z of $X \times \mathbb{G}_m$ of dimension r .

For $n > 1$, $\underline{\text{TZ}}_r(X, n; m)_M$ is the free abelian group on integral closed subschemes Z of $X \times B_n$ of dimension $r + n - 1$ such that:

- (1) (Good position) For each face F of B_n , Z intersects $X \times F$ properly:

$$\dim(Z \cap (X \times F)) \leq r + \dim(F) - 1, \text{ and}$$

- (2) (Modulus condition) Z satisfies the modulus m condition M on $X \times B_n$.

As our scheme X is equidimensional of dimension d over k , we write for $q \geq 0$

$$\underline{\text{TZ}}^q(X, n; m)_M = \underline{\text{TZ}}_{d+1-q}(X, n; m)_M.$$

We now observe that the good-position condition on Z implies that the cycle $(\text{id}_X \times \iota_{n,i,\epsilon})^*(Z)$, that we denote by $\partial_i^\epsilon(Z)$, is well-defined and each component satisfies the good-position condition. Moreover, letting $Y = X \times F$ for $F = \iota_{n,i,\epsilon}(B_{n-1})$ in Proposition 2.4, we first of all see that \bar{Y} intersects $X \times F_{n,0}$ and $X \times F_n^1$ properly in $X \times \widehat{B}_n$, and each component of $(\text{id}_X \times \iota_{n,i,\epsilon})^*(Z)$ satisfies the modulus condition M on $X \times B_{n-1}$. We thus conclude that if $Z \subset X \times B_n$ satisfies the above conditions (1) and (2), then every component of $\iota_{n,i,\epsilon}^*(Z)$ also satisfies these conditions on $X \times B_{n-1}$. In particular, we have the cubical abelian group $n \mapsto \underline{\text{TZ}}^q(X, n; m)_M$.

Definition 2.6. The additive cycle complex $\text{TZ}^q(X, \cdot; m)_M$ of X in codimension q and with modulus m condition M is the nondegenerate complex associated to the cubical abelian group $n \mapsto \underline{\text{TZ}}^q(X, n; m)_M$, that is,

$$\text{TZ}^q(X, n; m)_M := \frac{\underline{\text{TZ}}^q(X, n; m)_M}{\underline{\text{TZ}}^q(X, n; m)_{M, \text{degn}}},$$

where the group of degenerate cycles $\underline{\text{TZ}}^q(X, n; m)_{M, \text{degn}}$ is generated by the pull-backs of the cycles under the projections $X \times B_n \rightarrow X \times B_{n-1}$ given by

$$(x, t, y_1, \dots, y_{n-1}) \mapsto (x, t, y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_{n-1}).$$

The boundary map of this complex at level n is given by

$$\partial = \sum_{i=1}^{n-1} (-1)^i (\partial_i^\infty - \partial_i^0),$$

which satisfies $\partial^2 = 0$. The homology

$$\mathrm{TH}^q(X, n; m)_M := H_n(\mathrm{TZ}^q(X, \cdot; m)_M), \quad n \geq 1,$$

is the *additive higher Chow group* of X with modulus m condition M .

From now on, we shall drop the subscript M from the notations and it will be understood that the additive cycle complex or the additive higher Chow group in question is based on the modulus condition M , where M could be either M_{sum} or M_{ssup} . The reader should however always bear in mind that these two are different objects.

A few comments are in order. We could also have defined our additive cycle complex by taking $\mathrm{TZ}_r(X, n; m)$ to be the free abelian group generated by integral closed subschemes of $X \times \widetilde{B}_n$ which have the good-intersection property with respect to the faces of \widetilde{B}_n , and which satisfy the modulus condition on $X \times \widetilde{B}_n$ (see [Krishna and Levine 2008; Park 2009]). However, the following easy consequence of the modulus condition shows that this does not change the cycle complex.

Lemma 2.7. *Let M be a modulus condition in Definition 2.3.*

Then, there is a canonical bijection between the set of irreducible closed subvarieties $V \subset X \times B_n$ satisfying M and the set of irreducible closed subvarieties $W \subset X \times \widetilde{B}_n$ satisfying M .

Here, the correspondence is actually given by the identity map. In other words, any closed subvariety satisfying M on $X \times \widetilde{B}_n$ is in fact a closed subvariety of the smaller space $X \times B_n$.

Proof. First of all, since for any integral closed subscheme V of $X \times \widehat{B}_n$, the pull-back $v^*(F_{n,0})$ on V^N is contained in the open subset $v^{-1}(X \times \widetilde{B}_n)$, we can replace \widehat{B}_n by \widetilde{B}_n in the definition of the modulus conditions.

Now, if Σ and $\widetilde{\Sigma}$ are the two sets in the statement, then the modulus condition forces that if $V \in \Sigma$, then V is the same as its closure in $X \times \widetilde{B}_n$. Conversely, if $V \in \widetilde{\Sigma}$, then the modulus condition again forces V to be contained in $X \times B_n$. \square

Let $\mathrm{TZ}^q(X, \cdot; m)_{\mathrm{sup}}$ be the additive cycle complex as defined in [Krishna and Levine 2008; Park 2009]. This complex is based on the modulus condition M_{sup} above. It follows from (2-3) that there are natural inclusions of cycle complexes

$$\mathrm{TZ}^q(X, \cdot; m)_{\mathrm{ssup}} \hookrightarrow \mathrm{TZ}^q(X, \cdot; m)_{\mathrm{sup}} \hookrightarrow \mathrm{TZ}^q(X, \cdot; m)_{\mathrm{sum}} \tag{2-5}$$

and hence there are natural maps

$$\mathrm{TH}^q(X, \cdot; m)_{\mathrm{ssup}} \rightarrow \mathrm{TH}^q(X, \cdot; m)_{\mathrm{sup}} \rightarrow \mathrm{TH}^q(X, \cdot; m)_{\mathrm{sum}}. \tag{2-6}$$

One drawback of the cycle complex based on M_{sup} is that the underlying modulus condition for a cycle is not necessarily preserved when it is restricted to a face of B_n . This forces one to put an extra *induction* condition in the definition of $\text{TZ}^q(X, \cdot; m)_{\text{sup}}$ that requires for cycles to be admissible, not only must the cycles themselves satisfy M_{sup} on $X \times B_n$, but also all their intersections with various faces must satisfy M_{sup} . In particular, as n gets large, this condition gets more serious, and it might be a very tedious job to find admissible cycles. On the other hand, the definition of our cycle complexes shows that this extra *induction* condition is superfluous for complexes based on M_{sum} or M_{ssup} . Based on this discussion and all the results of this paper, one is led to guess that even though the modulus condition M_{ssup} may appear mildly stronger (and M_{sum} weaker) than the modulus condition M_{sup} , the following conjecture could be true.

Conjecture 2.8. *For a smooth projective variety X over k , the natural inclusions of cycle complexes $\text{TZ}^q(X, \cdot; m)_{\text{ssup}} \hookrightarrow \text{TZ}^q(X, \cdot; m)_{\text{sup}} \hookrightarrow \text{TZ}^q(X, \cdot; m)_{\text{sum}}$ are quasiisomorphisms.*

In Section 3, combined with previously known results, we check that when $S = \text{Spec}(k)$, for groups of 0-cycles, part of the conjecture holds, but we do not yet know how much of this conjecture holds true in general.

3. Basic properties of $\text{TH}^q(X, \cdot; m)$

In this section, our aim is to demonstrate that the additive higher Chow groups defined above for M_{sum} and M_{ssup} have all the properties (except Theorem 3.6 which we do not know for M_{sum}) which are known to be true for the additive Chow groups for M_{sup} of [Krishna and Levine 2008; Park 2009]. Since most of the arguments in the proofs can be given either by quoting these references verbatim or by straightforward modifications of the same, we only give the sketches of the proofs with minimal explanations whenever deemed necessary. We begin with the following structural properties of our additive Chow groups.

Theorem 3.1. *Let $f : Y \rightarrow X$ be a morphism of k -schemes.*

- (1) *If f is projective, there is a natural map of cycle complexes*

$$f_* : \text{TZ}_r(Y, \cdot; m) \rightarrow \text{TZ}_r(X, \cdot; m)$$

that induces the analogous push-forward map on the homology.

- (2) *If f is flat, there is a natural map of cycle complexes*

$$f^* : \text{TZ}_r(X, \cdot; m) \rightarrow \text{TZ}_r(Y, \cdot; m)$$

that induces the analogous pull-back map on the homology. These pull-back and push-forward maps satisfy the obvious functorial properties.

(3) If X is smooth and projective, there is a product

$$\cap_X : \mathrm{CH}^r(X, p) \otimes \mathrm{TH}_s(X, q; m) \rightarrow \mathrm{TH}_{s-r}(X, p + q; m),$$

natural with respect to flat pull-back, that satisfies the projection formula

$$f_*(f^*(a) \cap_X b) = a \cap_Y f_*(b)$$

for $f : X \rightarrow Y$ a morphism of smooth projective varieties. If f is flat in addition, we have an additional projection formula:

$$f_*(a \cap_X f^*(b)) = f_*(a) \cap_Y b.$$

(4) If X is smooth and quasiprojective, there is a product

$$\cap_X : \mathrm{CH}^r(X) \otimes \mathrm{TH}_s(X, q; m) \rightarrow \mathrm{TH}_{s-r}(X, q; m),$$

natural with respect to flat pull-back, that satisfies the projection formula

$$f_*(f^*(a) \cap_X b) = a \cap_Y f_*(b)$$

for $f : X \rightarrow Y$ a projective morphism of smooth quasiprojective varieties. If f is flat in addition, we have an additional projection formula

$$f_*(a \cap_X f^*(b)) = f_*(a) \cap_Y b.$$

Furthermore, all products are associative.

Proof. This follows from the arguments in [Krishna and Levine 2008]. Granting the flat pull-back and the projective push-forward, the theorem is a direct consequence of Lemmas 4.7 and 4.9 of that article, whose proofs are independent of the choice of the modulus conditions of Definition 2.3, as the interested reader may verify. The proofs of the flat pull-back and projective push-forward maps on the level of cycle complexes also follow in the same way as in [Krishna and Levine 2008] using our Lemma 2.2. \square

Theorem 3.2 (projective bundle and blow-up formulae). *Let X be a smooth quasi-projective variety and let V be a vector bundle on X of rank $r + 1$. Let $p : \mathbb{P}(V) \rightarrow X$ be the associated projective bundle over X . Let $\eta \in \mathrm{CH}^1(\mathbb{P}(V))$ be the class of the tautological line bundle $\mathcal{O}(1)$. Then for any $q, n \geq 1$ and $m \geq 2$, the map*

$$\theta : \bigoplus_{i=0}^r \mathrm{TH}^{q-i}(X, n; m) \rightarrow \mathrm{TH}^q(\mathbb{P}(V), n; m)$$

given by

$$(a_0, \dots, a_r) \mapsto \sum_{i=0}^r \eta^i \cap_{\mathbb{P}(V)} p^*(a_i)$$

is an isomorphism.

Suppose that $i : Z \rightarrow X$ is a closed immersion of smooth projective varieties and $\mu : X_Z \rightarrow X$ is the blow-up of X along Z with $i_E : E \rightarrow X_Z$ the exceptional divisor with morphism $q : E \rightarrow Z$. Then the sequence

$$0 \rightarrow \mathrm{TH}^s(X, n; m) \xrightarrow{(i^*, \mu^*)} \mathrm{TH}^s(Z, n; m) \oplus \mathrm{TH}^s(X_Z, n; m) \xrightarrow{q^* - i_E^*} \mathrm{TH}^s(E, n; m) \rightarrow 0$$

is split exact.

Proof. In view of Theorem 3.1, the proof of the theorem is exactly the same as the proofs of [Krishna and Levine 2008, Theorems 5.6 and 5.8]. The basic point is that there is a similar decomposition of the motives of the projective bundle and the blow-up in the triangulated category Mot_k of motives over k [ibid., Section 2]. On the other hand, Theorem 3.1 implies that for each integer $p \geq 1$, the assignment $(X, n) \mapsto \mathrm{TH}^n(X, p; m)$ is a functor from Mot_k to the category of graded abelian groups for any modulus M . We refer the reader to [ibid., Section 5] for details. \square

Recall from [ibid., Section 2.4] that $K^b(\mathbb{Z} \mathbf{SmProj} / k)$ is the homotopy category of the bounded complexes in the additive category $\mathbb{Z} \mathbf{SmProj} / k$ generated by \mathbf{SmProj} / k . We denote the complex concentrated in degree 0 associated to an $X \in \mathbf{SmProj} / k$ by $[X]$. Sending X to $[X]$ defines the functor

$$[-] : \mathbf{SmProj} / k \rightarrow K^b(\mathbb{Z} \mathbf{SmProj} / k).$$

Let $i : Z \rightarrow X$ be a closed immersion in \mathbf{SmProj} / k , $\mu : X_Z \rightarrow X$ the blow-up of X along Z , and $i_E : E \rightarrow X_Z$ the exceptional divisor with structure morphism $q : E \rightarrow Z$. Let $C(\mu)$ be the complex

$$[E] \xrightarrow{(i_E, -q)} [X_Z] \oplus [Z] \xrightarrow{\mu+i} [X]$$

with $[X]$ in degree 0. The category $\mathcal{D}_{\mathrm{hom}}(k)$ is the localization of the triangulated category $K^b(\mathbb{Z} \mathbf{SmProj} / k)$ with respect to the thick subcategory generated by the complexes $C(\mu)$.

Theorem 3.3. *Assume that k admits resolution of singularities. Then the functor $\mathrm{TZ}_r(-; m) : \mathbf{SmProj} / k \rightarrow D^-(\mathbf{Ab})$ extends to a functor*

$$\mathrm{TZ}_r^{\mathrm{log}}(-; m) : \mathbf{Sch} / k \rightarrow D^-(\mathbf{Ab})$$

together with a natural transformation of functors $\mathrm{TZ}_r^{\mathrm{log}}(-; m) \rightarrow \mathrm{TZ}_r(-; m)$ satisfying:

(1) *Let $\mu : Y \rightarrow X$ be a proper morphism in \mathbf{Sch} / k , $i : Z \rightarrow X$ a closed immersion. Suppose that $\mu : \mu^{-1}(X \setminus Z) \rightarrow X \setminus Z$ is an isomorphism. Set $E := \mu^{-1}(Z)$ with maps $i_E : E \rightarrow Y$, $q : E \rightarrow Z$. There is a canonical extension of the sequence in*

$D^-(\mathbf{Ab})$:

$$\mathrm{TZ}_r^{\mathrm{log}}(E; m) \xrightarrow{(i_{E^*}, -q_*)} \mathrm{TZ}_r^{\mathrm{log}}(Y; m) \oplus \mathrm{TZ}_r^{\mathrm{log}}(Z; m) \xrightarrow{\mu_* + i_*} \mathrm{TZ}_r^{\mathrm{log}}(X; m)$$

to a distinguished triangle in $D^-(\mathbf{Ab})$.

(2) Let $i : Z \rightarrow X$ be a closed immersion in \mathbf{Sch}/k , $j : U \rightarrow X$ the open complement. Then there is a canonical distinguished triangle in $D^-(\mathbf{Ab})$:

$$\mathrm{TZ}_r^{\mathrm{log}}(Z; m) \xrightarrow{i_*} \mathrm{TZ}_r^{\mathrm{log}}(X; m) \xrightarrow{j^*} \mathrm{TZ}_r^{\mathrm{log}}(U; m) \rightarrow \mathrm{TZ}_r^{\mathrm{log}}(Z; m)[1],$$

which is natural with respect to proper morphisms of pairs $(X, U) \rightarrow (X', U')$.

(3) For any $X \in \mathbf{Sch}/k$, the natural map $\mathrm{TH}_r^{\mathrm{log}}(X, n; m) \rightarrow \mathrm{TH}_{r+p}^{\mathrm{log}}(X \times \mathbb{A}^p, n; m)$ is an isomorphism.

Proof. The proof of this theorem is exactly the same as the proof of [Krishna and Levine 2008, Corollary 6.2]. By [ibid., Lemma 2.8], the motive functor

$$m_{\mathrm{hom}} : \mathbf{SmProj}/k \rightarrow \mathcal{D}_{\mathrm{hom}}(k)$$

is a category of homological descent in the sense of [Guillén and Navarro Aznar 2002]. Theorem 3.2 immediately implies that $\mathrm{TZ}_r(-; m) : \mathbf{SmProj}/k \rightarrow D^-(\mathbf{Ab})$ extends to a functor $\mathrm{TZ}_r(-; m) : \mathcal{D}_{\mathrm{hom}}(k) \rightarrow D^-(\mathbf{Ab})$. On the other hand, the functor m_{hom} extends to a functor $M_{\mathrm{hom}} : \mathbf{Sch}'/k \rightarrow \mathcal{D}_{\mathrm{hom}}(k)$ by [Krishna and Levine 2008, Theorem 2.9]. The functor $\mathrm{TZ}_r^{\mathrm{log}}(-; m)$ is the composite $\mathrm{TZ}_r(-; m) \circ M_{\mathrm{hom}}$. All the desired properties of $\mathrm{TZ}_r^{\mathrm{log}}(-; m)$ follow from the similar properties of M_{hom} as shown in the same reference. \square

Next we study the question of the existence of the regulator maps from our additive higher Chow groups to the modules of absolute Kähler differentials. First we prove the following result of [Bloch and Esnault 2003a; Rülling 2007] on 0-cycles for the modulus condition M_{ssup} .

Theorem 3.4. *Assume that $\mathrm{char}(k) \neq 2$ and let $\mathbb{W}_m \Omega_k^\bullet$ denote the generalized de Rham–Witt complex of Hesselholt and Madsen (see [Rülling 2007]). Then there is a natural isomorphism*

$$R_{0,m}^n : \mathrm{TH}^n(k, n; m) \rightarrow \mathbb{W}_m \Omega_k^{n-1}.$$

Proof. This is already known for M_{sum} . For the modulus condition M_{ssup} , we first note that the map $R_{0,m}^n$ is the composite map

$$\mathrm{TH}^n(k, n; m)_{\mathrm{ssup}} \rightarrow \mathrm{TH}^n(k, n; m)_{\mathrm{sum}} \xrightarrow{\theta} \mathbb{W}_m \Omega_k^{n-1},$$

where θ is constructed in [Rülling 2007] and this coincides with the regulator map of Bloch and Esnault for $m = 1$. Furthermore for $m = 1$, Bloch and Esnault define

the inverse map $\Omega_k^{n-1} \rightarrow \mathrm{TH}^n(k, n; 1)_{\mathrm{sum}}$ using a presentation of Ω_k^{n-1} . The reader can easily check from the proof of [Bloch and Esnault 2003a, Proposition 6.3] that the inverse map is actually defined from Ω_k^{n-1} to $\mathrm{TH}^n(k, n; 1)_{\mathrm{ssup}}$. This completes the proof when $m = 1$.

For $m \geq 2$, Rülling’s proof for $\mathrm{TH}^n(k, n; 1)_{\mathrm{sum}}$ has these main steps:

- (1) The existence of map $R_{0,m}^n$
- (2) The isomorphism of $R_{0,m}^1$.
- (3) The existence of transfer maps on the additive higher Chow groups for finite extensions of fields.
- (4) Showing that pro-group $\{\mathrm{TH}^n(k, n; m)\}_{n,m \geq 1}$ is an example of a restricted Witt complex; see [Rülling 2007, Remark 4.22].

We have already shown (1) for our $\mathrm{TH}^n(k, n; m)_{\mathrm{ssup}}$. The proof of (3) is a simple consequence of Theorem 3.1. The surjectivity part of (2) follows from the result of Rülling and the isomorphism $\underline{\mathrm{TZ}}^n(k, n; m)_{\mathrm{ssup}} = \underline{\mathrm{TZ}}^n(k, n; m)_{\mathrm{sum}}$. To prove injectivity, we follow the proof of [Rülling 2007, Corollary 4.6.1] and observe that if there is a cycle $\zeta \in \underline{\mathrm{TZ}}^1(k, 1; m)$ such that $R_{0,m}^1(\zeta) = 0$, then ζ is the boundary of a curve C which is an admissible cycle with the modulus condition M_{sum} . But then C is an admissible cycle also with the modulus condition M_{ssup} since one has $M_{\mathrm{ssup}} = M_{\mathrm{sup}} = M_{\mathrm{sum}}$ when $n = 2$ by definition. This proves (2). Note that this does not need any assumptions on the characteristic of the ground field.

For the proof of (4), one checks that Lemma 4.17 of [ibid.] works without change.

Rülling showed that these four ingredients and the universality of the de Rham–Witt complex imply that there is a map

$$\mathbb{W}_m \Omega_k^{n-1} \xrightarrow{S_{0,m}^n} \mathrm{TH}^n(k, n; m)$$

which is surjective. On the other hand, one checks from the construction of the map $R_{0,m}^n$ in [ibid.] that $R_{0,m}^n \circ S_{0,m}^n$ is the identity. □

The following result is an immediate consequence of the results of Rülling and Theorem 3.4. This gives evidence for Conjecture 2.8.

Corollary 3.5. *For every $n, m \geq 1$, the natural maps*

$$\mathrm{TH}^n(k, n; m)_{\mathrm{ssup}} \rightarrow \mathrm{TH}^n(k, n; m)_{\mathrm{sup}} \rightarrow \mathrm{TH}^n(k, n; m)_{\mathrm{sum}}$$

are isomorphisms.

We finally turn to the regulator maps for 1-cycles as considered in [Park 2009].

Theorem 3.6. *Suppose that k is of characteristic zero and assume the modulus condition to be M_{ssup} . Then there is a natural nontrivial regulator map*

$$R_{1,m}^n : \text{TH}^{n-1}(k, n; m)_{\text{ssup}} \rightarrow \Omega_k^{n-3}. \tag{3-1}$$

This map is surjective if k is, moreover, algebraically closed.

Proof. Let $R_{1,m}^n$ be the composite map

$$\text{TH}^n(k, n; m)_{\text{ssup}} \rightarrow \text{TH}^n(k, n; m)_{\text{sup}} \xrightarrow{\theta} \Omega_k^{n-3},$$

where θ is constructed in [Park 2009]. For the nontriviality of $R_{1,m}^n$, Park constructs a 1-cycle Γ (see [Park 2007, Proposition 1.9] and [Krishna and Levine 2008, 7.11]) and shows (see [Park 2007, Lemmas 1.7 and 1.9]) that each component of Γ in fact satisfies the modulus condition M_{ssup} . Hence $R_{1,m}^n$ is nontrivial. If $k = \bar{k}$, then the proof of the surjectivity in [Krishna and Levine 2008, §7] follows from the following:

- (1) An action of k^\times on $\text{TH}^n(k, n; m)$,
- (2) Suitable k^\times -equivariance of $R_{1,m}^3$ up to a scalar,
- (3) The surjectivity of $R_{1,m}^3$, and
- (4) The cap product $\text{CH}^n(k, n) \otimes_{\mathbb{Z}} \text{TH}^2(k, 3; m) \rightarrow \text{TH}^{n+2}(k, n+3; m)$.

The action of k^\times on our additive higher Chow groups is given as in [Park 2007; Krishna and Levine 2008] by

$$a * (x, t_1, \dots, t_{n-1}) = (x/a, t_1, \dots, t_{n-1}). \tag{3-2}$$

This action extends to an action of k^\times on \widehat{B}_n . The proof of (2) now follows from the k^\times -equivariance of the natural map $\text{TZ}_r(k, n; m)_{\text{ssup}} \rightarrow \text{TZ}_r(k, n; m)_{\text{sup}}$ and the results of [Krishna and Levine 2008]. The proof of (3) is a direct consequence of (1), (2), and the fact that k is algebraically closed field of characteristic zero. Finally, (4) is already shown in Theorem 3.1. □

We do not yet know if this theorem holds for M_{sum} because the regulator map $R_{1,m}^n$ in [Park 2007] is not immediately defined on the set of all M_{sum} -admissible 1-cycles. In fact, this was one main obstruction that led to the introduction of the M_{sup} modulus condition in that work. See Section 8A for a related discussion on how one may potentially get around this issue.

4. Preliminaries for moving lemma

The underlying additive cycle complexes and additive higher Chow groups in all the results in the rest of this paper will be based on the modulus condition M_{sum} or

M_{ssup} , unless one of these is specifically mentioned. Our next three sections will be devoted to proving our first main result of this paper:

Theorem 4.1. *Let X be a smooth projective variety over a perfect field k . Let \mathfrak{W} be a finite collection of locally closed subsets of X . Then, the inclusion of additive higher Chow cycle complexes (see below for definitions)*

$$\text{TZ}_{\mathfrak{W}}^q(X, \cdot; m) \hookrightarrow \text{TZ}^q(X, \cdot; m)$$

is a quasiisomorphism. In other words, every admissible additive higher Chow cycle is congruent to another admissible cycle intersecting properly all given finitely many locally closed subsets of X times faces.

In this section, we set up our notations and machinery that are needed to prove this theorem, and prove some preliminary steps. Let X be a smooth projective variety over k and we fix an integer $m \geq 1$. Let \mathfrak{W} be a finite collection of locally closed algebraic subsets of X . If a member of \mathfrak{W} is not irreducible, we always replace it by all of its irreducible components so that we assume all members of \mathfrak{W} are irreducible. For a locally closed subset $Y \subset X$, recall that the codimension $\text{codim}_X Y$ is defined to be the minimum of $\text{codim}_X Z$ for all irreducible components Z of Y .

Definition 4.2. We define $\underline{\text{TZ}}_{\mathfrak{W}}^q(X, n; m)$ to be the subgroup of $\underline{\text{TZ}}^q(X, n; m)$ generated by integral closed subschemes $Z \subset X \times B_n$ such that

- (1) Z is in $\underline{\text{TZ}}^q(X, n; m)$ and
- (2) $\text{codim}_{W \times F}(Z \cap (W \times F)) \geq q$ for all $W \in \mathfrak{W}$ and all faces F of B_n .

It is easy to see that $\underline{\text{TZ}}_{\mathfrak{W}}^q(X, \cdot; m)$ forms a cubical subgroup of $\underline{\text{TZ}}^q(X, \cdot; m)$, giving us the subcomplex

$$\text{TZ}_{\mathfrak{W}}^q(X, \cdot; m) = \frac{\underline{\text{TZ}}_{\mathfrak{W}}^q(X, \cdot; m)}{\underline{\text{TZ}}_{\mathfrak{W}}^q(X, \cdot; m)_{\text{degn}}} \subset \text{TZ}^q(X, \cdot; m).$$

Let $\text{TH}_{\mathfrak{W}}^q(X, \cdot; m)$ denote the homology of the complex $\text{TZ}_{\mathfrak{W}}^q(X, \cdot; m)$. Then the above inclusion induces a natural map of homology,

$$\text{TH}_{\mathfrak{W}}^q(X, \cdot; m) \rightarrow \text{TH}^q(X, \cdot; m). \tag{4-1}$$

More generally, if $e : \mathfrak{W} \rightarrow \mathbb{Z}_{\geq 0}$ is a set-theoretic function, then one can define subcomplexes $\underline{\text{TZ}}_{\mathfrak{W}, e}^q(X, \cdot; m)$ replacing condition (2) above by

- (2e) $\text{codim}_{W \times F}(Z \cap (W \times F)) \geq q - e(W)$.

In this generality, the subcomplex $\underline{\text{TZ}}_{\mathfrak{W}}^q(X, \cdot; m)$ is the same as $\underline{\text{TZ}}_{\mathfrak{W}, 0}^q(X, \cdot; m)$.

Remark 4.3. Let Φ be the set of all set-theoretic functions $e : \mathcal{W} \rightarrow \mathbb{Z}_{\geq 0}$. Give a partial ordering on Φ by declaring $e' \geq e$ if $e'(W) \geq e(W)$ for all $W \in \mathcal{W}$. If two functions $e, e' \in \Phi$ satisfy $e' \geq e$, then for any irreducible admissible subvariety $Z \in \text{TZ}_{\mathcal{W},e}^q(X, n; m)$, we have

$$\text{codim}_{W \times F}(Z \cap (W \times F)) \geq q - e(W) \geq q - e'(W) \tag{4-2}$$

for all $W \in \mathcal{W}$ and all faces $F \subset B_n$. Thus, we have

$$\text{TZ}_{\mathcal{W},e}^q(X, n; m) \subset \text{TZ}_{\mathcal{W},e'}^q(X, n; m) \quad \text{for } e \leq e'. \tag{4-3}$$

Note that if $e \in \Phi$ satisfies $e \geq q$ where q is considered as a constant function in Φ , then automatically

$$\text{TZ}_{\mathcal{W},q}^q(X, n; m) = \text{TZ}_{\mathcal{W},e}^q(X, n; m) = \text{TZ}^q(X, n; m). \tag{4-4}$$

Since $0 \leq e$ for all $e \in \Phi$, for each triple e, e', e'' such that $e \leq e' \leq q \leq e''$, we have

$$\begin{aligned} \text{TZ}_{\mathcal{W}}^q(X, n; m) &\subset \text{TZ}_{\mathcal{W},e}^q(X, n; m) \subset \text{TZ}_{\mathcal{W},e'}^q(X, n; m) \\ &\subset \text{TZ}_{\mathcal{W},q}^q(X, n; m) = \text{TZ}_{\mathcal{W},e''}^q(X, n; m) = \text{TZ}^q(X, n; m). \end{aligned}$$

All these (in)equalities are equivariant with respect to the boundary maps.

Remark 4.4. The main theorem is equivalent to saying that the inclusion

$$\text{TZ}_{\mathcal{W}}^q(X, n; m) \subset \text{TZ}^q(X, n; m)$$

induces an isomorphism $\text{TH}_{\mathcal{W}}^q(X, n; m) \simeq \text{TH}^q(X, n; m)$ for the given modulus condition M .

Our remaining objective in this section is to prove an additive analogue of the spreading argument, which originates from Bloch’s arguments. We begin with the following results.

Lemma 4.5. *Let $f : X \rightarrow Y$ be a dominant morphism of integral normal varieties and let η denote the generic point of Y . Consider the fiber diagram*

$$\begin{array}{ccc} X_{\eta} & \xrightarrow{j_{\eta}} & X \\ \downarrow & & \downarrow f \\ \{\eta\} & \longrightarrow & Y. \end{array} \tag{4-5}$$

Let D be a Weil divisor on X such that $j_{\eta}^(D)$ is effective. Then there is a nonempty open subset $U \subset Y$ such that if $j : f^{-1}(U) \rightarrow X$ denotes the open inclusion, then $j^*(D)$ is also effective.*

Proof. Let $D = \sum n_i D_i$. Then $j_\eta^*(D)$ is effective if and only if for every i with $n_i < 0$, one has $D_i \cap X_\eta = \emptyset$. Since D is a finite sum, it suffices to show that if D is a prime divisor on X such that $D \cap X_\eta = \emptyset$, then there is a nonempty open subset $U \subset Y$ such that $D \cap f^{-1}(U) = \emptyset$.

Our hypothesis implies that $\overline{f(D)}$ is a proper closed subset of Y . Thus $U = Y \setminus \overline{f(D)}$ is the desired open subset of Y . \square

Lemma 4.6. *Let X be a quasiprojective k -variety and let \mathcal{W} be a finite collection of locally closed subsets of X . Let K be a finite field extension of k . Let X_K be the base extension $X_K = X \times_{\text{Spec}(k)} \text{Spec}(K)$, and let ${}^{\circ}\mathcal{W}_K$ be the set of the base extensions of the varieties in \mathcal{W} . Then there are natural maps*

$$p^* : \frac{\text{TZ}^q(X, \cdot; m)}{\text{TZ}_{\mathcal{W}}^q(X, \cdot; m)} \rightarrow \frac{\text{TZ}^q(X_K, \cdot; m)}{\text{TZ}_{\mathcal{W}_K}^q(X_K, \cdot; m)},$$

$$p_* : \frac{\text{TZ}^q(X_K, \cdot; m)}{\text{TZ}_{\mathcal{W}_K}^q(X_K, \cdot; m)} \rightarrow \frac{\text{TZ}^q(X, \cdot; m)}{\text{TZ}_{\mathcal{W}}^q(X, \cdot; m)}$$

such that $p_* \circ p^* = [K : k] \cdot \text{id}$.

Proof. By Theorem 3.1, one also has the flat pull-back and finite push-forward maps $\text{TZ}_{\mathcal{W}'}^q(X, \cdot; m) \rightarrow \text{TZ}_{\mathcal{W}'_K}^q(X_K, \cdot; m)$ and $\text{TZ}_{\mathcal{W}'_K}^q(X_K, \cdot; m) \rightarrow \text{TZ}_{\mathcal{W}'}^q(X, \cdot; m)$ for any \mathcal{W}' . Taking for \mathcal{W}' the collection $\{X\}$ and then \mathcal{W} , and then taking the quotient of the two, we get the desired maps. The last property of the composite map is obvious from the construction of the pull-back and the push-forward maps on the additive cycle complexes; see [Krishna and Levine 2008]. \square

Proposition 4.7 (spreading lemma). *Let $k \subset K$ be a purely transcendental extension. For a smooth projective variety X over k and any finite collection \mathcal{W} of locally closed algebraic subsets of X , let X_K and \mathcal{W}_K be the base extensions as before. Let $p_K : X_K \rightarrow X_k$ be the natural map. Then, the pull-back map*

$$p_K^* : \frac{\text{TZ}^q(X, \cdot; m)}{\text{TZ}_{\mathcal{W}}^q(X, \cdot; m)} \rightarrow \frac{\text{TZ}^q(X_K, \cdot; m)}{\text{TZ}_{\mathcal{W}_K}^q(X_K, \cdot; m)}$$

is injective on homology.

Proof. First, suppose the proposition holds for all infinite fields, and let k be a finite field. Let Z be a cycle on the left quotient group whose pull-back via $k \rightarrow K$ dies. Then, for two different primes ℓ_1 and ℓ_2 and for pro- ℓ_i extensions $k \rightarrow k_i$, the images of Z under the respective pull-backs are zero. Hence, by the norm argument in Lemma 4.6, there exist integers N_i such that $\ell_i^{N_i} Z = 0$ in the left group. This implies that $Z = 0$, thus the proposition holds for the finite field k . Hence, we can assume that k is infinite.

Since the additive Chow group of X_K is an inductive limit of the additive Chow groups of X_L , where $L \subset K$ range over purely transcendental extension of k of

finite transcendence degree over k , we can assume that the transcendence degree of K over k is finite.

Now let $Z \in \text{TZ}^q(X, n; m)$ be a cycle such that $\partial Z \in \text{TZ}_{\mathbb{W}}^q(X, n-1; m)$ where there are admissible cycles $B_K \in \text{TZ}^q(X_K, n+1; m)$ and $V_K \in \text{TZ}_{\mathbb{W}_K}^q(X_K, n; m)$ satisfying $Z_K = \partial(B_K) + V_K$.

We first consider the natural inclusion of complexes

$$\text{TZ}^q(X, \cdot; m) \hookrightarrow z^q(X \times \mathbb{A}_k^1, \cdot - 1).$$

Since K is the function field of some affine space \mathbb{A}_k^r , we can use the specialization argument for Bloch's cycle complexes [1986, Lemma 2.3] to find an open subset $U' \subset \mathbb{A}_k^r$ and cycles

$$B_{U'} \in z^q(X \times U' \times \mathbb{A}_k^1, n), \quad V_{U'} \in z_{\mathbb{W} \times U' \times \mathbb{A}_k^1}^q(X \times U' \times \mathbb{A}_k^1, n-1)$$

such that B_K and V_K are the restrictions of $B_{U'}$ and $V_{U'}$ to the generic point of U' and $Z \times U' = \partial(B_{U'}) + V_{U'}$, respectively. In particular, all components of $B_{U'}$ and $V_{U'}$ intersect all faces of $X \times U' \times B_{n+1}$ and $X \times U' \times B_n$ properly. To make $B_{U'}$ and $V_{U'}$ admissible additive cycles, we modify them using our Lemma 4.5.

To check the modulus condition for our cycles, let η denote the generic point $\text{Spec}(K)$ of U' . Let $\widehat{B}_{U'}^N$ and $\widehat{V}_{U'}^N$ denote the normalizations of the closures of $B_{U'}$ and $V_{U'}$ in $X \times U' \times \widehat{B}_{n+1}$ and $X \times U' \times \widehat{B}_n$, respectively.

We first prove the admissibility under the modulus condition M_{ssup} which is a priori more difficult than M_{sum} . The admissibility of B_K and V_K implies that there are integers $1 \leq i \leq n$ and $1 \leq i' \leq n-1$ such that in (4-5), the Weil divisors $j_{\eta}^*(F_{n+1,i}^1 - (m+1)F_{n+1,0})$ and $j_{\eta}^*(F_{n,i'}^1 - (m+1)F_{n,0})$ are effective on $\widehat{B}_{U',\eta}^N$ and $\widehat{V}_{U',\eta}^N$, respectively. Since X and \widehat{B}_n are projective, the maps $\widehat{B}_{U'}^N, \widehat{V}_{U'}^N \rightarrow U'$ are projective. These maps are dominant since B_K and V_K are nonzero-cycles. Thus we can apply Lemma 4.5 to find an open subset $U \subset U'$ such that

$$j_U^*(F_{n+1,i}^1 - (m+1)F_{n+1,0})$$

and $j_U^*(F_{n,i'}^1 - (m+1)F_{n,0})$ are also effective. The same argument applies for the modulus condition M_{sum} as well. We just have to replace the Cartier divisors $F_{n+1,i}^1$ and $F_{n,i'}^1$ by $F_{n+1,i}^1$ and $F_{n,i'}^1$, respectively. Lemma 4.5 applies in this case, too.

Replacing U' by U , we see that

$$\begin{aligned} B_U \in \text{TZ}^q(X \times U, n+1; m), \quad V_U \in \text{TZ}_{\mathbb{W} \times U}^q(X \times U, n; m), \\ Z \times U = \partial(B_U) + V_U. \end{aligned} \tag{4-6}$$

Next, (4-6) implies that for a k -rational point $u \in U(k)$ (which exists because k is infinite) such that the restrictions of B_U and V_U to $X \times \{u\}$ give well-defined cycles in $z^q(X \times \mathbb{A}_k^1, n)$ and $z_{\mathbb{W} \times \mathbb{A}_k^1}^q(X \times \mathbb{A}_k^1, n-1)$, one has $Z = \partial(i_u^*(B_U)) + i_u^*(V_U)$, where $i_u : X \times \{u\} \rightarrow X \times U$ is the closed immersion.

We now only need to show that $i_u^*(B_U)$ and $i_u^*(V_U)$ satisfy the modulus condition on $X \times \{u\}$. But this follows directly from (4-6) and the containment lemma, Proposition 2.4. \square

5. Moving lemma for projective spaces

We follow the strategy of Bloch and Levine to prove the moving lemma for the additive higher Chow groups. This involves proving the moving lemma first for the projective spaces and then deducing the same for general smooth projective varieties using the techniques of linear projections. This section is devoted to the proof of the moving lemma for the projective spaces. We use the following technique from [Bloch 1986, Lemma 1.1] a few times to prove the proper-intersection properties of moved cycles with the prescribed algebraic sets.

Lemma 5.1. *Let X be an algebraic k -scheme and G a connected algebraic k -group acting on X . Let $A, B \subset X$ be closed subsets, and assume that the fibers of the map*

$$G \times A \rightarrow X \quad (g, a) \mapsto g \cdot a$$

all have the same dimension and that this map is dominant. Then, there exists a nonempty open subset $U \subset G$ such that for all extension fields L of k and for all $g \in U(L)$, the intersection $g(A_L) \cap B_L$ is proper in X_L .

Proposition 5.2 (admissibility of projective image). *Let $f : X \rightarrow Y$ be a projective morphism of quasiprojective varieties over a field k . Let $Z \in \mathbb{TZ}'(X, n; m)$ be an irreducible admissible cycle and let $V = f(Z)$. Then $V \in \mathbb{TZ}^s(Y, n; m)$, where s is the codimension of V in $Y \times B_n$.*

Proof. We prove this in several steps.

Claim 1. *V intersects all codimension-one faces F of B_n properly in B_n .*

Consider $F = F_{n,i}^\epsilon = \iota_{n,i,\epsilon}(B_{n-1})$ for some $i \in \{1, 2, \dots, n-1\}$, $\epsilon \in \{0, \infty\}$, and consider the diagram

$$\begin{array}{ccc} X \times B_{n-1} & \xrightarrow{\iota_{n,i,\epsilon}} & X \times B_n \\ \downarrow f_{n-1} & & \downarrow f_n \\ Y \times B_{n-1} & \xrightarrow{\iota_{n,i,\epsilon}} & Y \times B_n. \end{array}$$

Since F is a divisor in B_n , that V intersects $Y \times F$ properly is equivalent to that $Y \times F \not\supset V$. Towards contradiction, suppose that $V \subset Y \times F$. Then,

$$Z \subset f_n^{-1}(f_n(Z)) = f_n^{-1}(V) \subset f_n^{-1}(Y \times F) = \iota_{n,i,\epsilon}(f_{n-1}^{-1}(Y \times B_{n-1})) = X \times F.$$

By assumption, Z intersects $X \times F$ properly so that we must have $Z \not\subset X \times F$. This contradiction proves the claim.

Claim 2. V intersects all lower-dimensional faces of B_n properly.

By the admissibility assumption, all cycles $\partial_i^\epsilon(Z) = Z \cap (X \times F_{n,i}^\epsilon)$ are admissible. Moreover, it is easy to see that $\partial_i^\epsilon(V) = f_{n-1}(\partial_i^\epsilon(Z))$. Thus we can replace Z by $\partial_i^\epsilon(Z)$ and apply the same argument as above; inductively we see that V has the good-intersection property.

Claim 3. For each face F of B_n , including the case $F = B_n$, the cycle $V \cap (Y \times F)$ has the modulus condition.

For any face $F = \iota(B_i) \subset B_n$, where $\iota : B_i \hookrightarrow B_n$ is a face map, and for the projections $f_i : X \times B_i \rightarrow Y \times B_i$, note that $V \cap (Y \times F) = f_n(Z \cap (X \times F)) = f_i(Z|_{X \times F})$. But the admissibility of Z implies that $Z|_{X \times F}$ is also admissible (see Proposition 2.4). Hence, replacing $Z|_{X \times F}$ by Z , we only need to prove it for $F = B_n$, that is, we just need to show that V satisfies the modulus condition. Consider the diagram

$$\begin{array}{ccc} X \times B_n & \longrightarrow & X \times \widehat{B}_n \\ \downarrow f_n=f & & \downarrow \bar{f}_n=\bar{f} \\ Y \times B_n & \longrightarrow & Y \times \widehat{B}_n. \end{array}$$

Subclaim. Let \bar{V} be the closure of V in $Y \times \widehat{B}_n$ and let \bar{Z} be the closure of Z in $X \times \widehat{B}_n$. Then $\bar{V} = \bar{f}(\bar{Z})$.

Since $Z \subset f^{-1}(V) \subset \bar{f}^{-1}(\bar{V})$ and V is closed, we have $\bar{Z} \subset \bar{f}^{-1}(\bar{V})$. Hence, $\bar{f}(\bar{Z}) \subset \bar{V}$. For the other inclusion, note that $V = f(Z) \subset \bar{f}(\bar{Z})$ and $\bar{f}(\bar{Z})$ is closed because \bar{f} is projective. Hence $\bar{V} \subset \bar{f}(\bar{Z})$. This proves this subclaim.

To prove the modulus condition for V , we take the normalizations $\nu_{\bar{Z}} : \bar{Z}^N \rightarrow \bar{Z}$ and $\nu_{\bar{V}} : \bar{V}^N \rightarrow \bar{V}$ of \bar{Z} and \bar{V} , and consider the following diagram:

$$\begin{array}{ccccc} \bar{Z}^N & \xrightarrow{\nu_{\bar{Z}}} & \bar{Z} & \xrightarrow{\iota_1} & X \times \widehat{B}_n \\ \downarrow f_{\bar{Z}}^N & & \downarrow f_{\bar{Z}}=\bar{f}|_{\bar{Z}} & & \downarrow \bar{f} \\ \bar{V}^N & \xrightarrow{\nu_{\bar{V}}} & \bar{V} & \xrightarrow{\iota_2} & Y \times \widehat{B}_n, \end{array}$$

where ι_1 and ι_2 are the inclusions, and $f_{\bar{Z}}^N$ is given by the universal property of the normalization $\nu_{\bar{V}}$ for dominant morphisms. Note that $f_{\bar{Z}}^N$ is automatically projective and surjective because $f_{\bar{Z}}$ is so. Let $q_{\bar{Z}} := \iota_1 \circ \nu_{\bar{Z}}$ and $q_{\bar{V}} := \iota_2 \circ \nu_{\bar{V}}$.

Suppose Z satisfies the modulus condition M_{ssup} and consider on \widehat{B}_n the Cartier divisors $D_i := F_{n,i}^1 - (m+1)F_{n,0}$ for $1 \leq i \leq n-1$. That the cycle Z has the modulus condition means that $[q_{\bar{Z}}^* \circ \bar{f}^*(D_i)] \geq 0$ for an index i . By the commutativity of the above diagram, this means that the Cartier divisor $f_{\bar{Z}}^{N*}[q_{\bar{V}}^*(D_i)] \geq 0$. By Lemma 2.2, this implies that $[q_{\bar{Z}}^*(D_i)] \geq 0$, which is the modulus condition for V .

If Z satisfies the modulus condition M_{sum} , we use the same argument by replacing $F_{n,i}^1$ with F_n^1 . This finishes the proof of the proposition. \square

Remark 5.3. In Proposition 5.2, if X is projective, $Y = \text{Spec}(k)$, and $n = 1$, then V is always a single point. To see this, let $Z \subset X \times B_1 = X \times \mathbb{G}_m$ be an admissible irreducible closed subvariety. Let $V = p(Z)$, where $p : X \times \mathbb{G}_m \rightarrow \mathbb{G}_m$ is the projection.

Since X is complete, p is a closed map. Hence, $V = p(Z)$ is an irreducible closed subvariety of \mathbb{G}_m . But the only closed subvarieties of \mathbb{G}_m are finite subsets or all of \mathbb{G}_m . On the other hand, if \bar{Z} is the closure of Z in $X \times \mathbb{A}^1$, then the modulus condition implies that $\bar{Z} \cap |X \times \{t = 0\}| = \emptyset$. This implies that V must be a proper subset and hence a finite subset. Since V is irreducible, consequently V must be a nonzero single point.

Hence $Z = W \times \{*\}$ for a closed subvariety $W \subset X$, and a closed point $\{*\} \in \mathbb{G}_m$. Conversely, any such variety is admissible. This classifies all admissible cycles Z when X is projective and $n = 1$.

For $n > 1$, all we can say is that Z is contained in $X \times V$, where V is admissible in $\text{TZ}_s(k, n; m)$ for a suitable s .

5A. Homotopy variety. Now we want to construct the “homotopy variety”. First, we need the following simple result:

Lemma 5.4. *Let $\text{SL}_{r+1,k}$ be the $(r + 1) \times (r + 1)$ special linear group over k , and let η be the generic point of the k -variety $\text{SL}_{r+1,k}$. Let K be its function field (this is a purely transcendental extension of k). Let $\text{SL}_{r+1,K} := \text{SL}_{r+1,k} \otimes_k K$ be base change. Then, there is a morphism of K -varieties $\phi : \square_K^1 \rightarrow \text{SL}_{r+1,K}$ such that $\phi(0)$ is the identity element, and $\phi(\infty)$ is the generic point η considered as a K -rational point.*

Proof. By a general result on the special linear groups, every element of $\text{SL}_{r+1,K}$ is generated by the transvections $E_{ij}(a)$, $i \neq j$, $a \in K$, that are $(r + 1) \times (r + 1)$ matrices where the diagonal entries are 1, the (i, j) -entry is a and all other entries are zero.

For each pair (i, j) , the collection $\{E_{ij}(a) \mid a \in K\}$ forms a one-parameter subgroup of $\text{SL}_{r+1,K}$ isomorphic to $\mathbb{G}_{a,K}$. Thus, for each fixed $b \in K$, define $\phi_{ij}^b : \mathbb{A}_K^1 \rightarrow \text{SL}_{r+1,K}$ by $\phi_{ij}^b(y) := E_{ij}(by)$.

Express the K -rational point η of $\text{SL}_{r+1,K}$ as the (ordered) product

$$\eta = \prod_{l=1}^p E_{i_l j_l}(a_l), \quad \text{for some } i_l, j_l \in \{1, 2, \dots, r + 1\} \text{ and } a_l \in K,$$

and define $\phi' : \mathbb{A}_K^1 \rightarrow \mathrm{SL}_{r+1,K}$ by $\phi' = \prod_{l=1}^p \phi_{i_l j_l}^{a_l}$. By definition, we have $\phi'(0) = \mathrm{Id}$ and $\phi'(1) = \eta$. Composing with the automorphism $\sigma : \mathbb{P}_K^1 \rightarrow \mathbb{P}_K^1$ given by

$$y \mapsto \frac{y}{y-1},$$

which isomorphically maps \square_K^1 to \mathbb{A}_K^1 , we obtain $\phi = \phi' \circ \sigma : \square_K^1 \rightarrow \mathrm{SL}_{r+1,K}$. This ϕ satisfies the desired properties. \square

Recall that one consequence of Lemma 2.7 is that the additive cycle complex with modulus m can also be defined as a complex whose level- n term is the free abelian group of the integral closed subschemes $Z \subset X \times \widetilde{B}_n$ which have the good-intersection property with all faces, and which satisfy the appropriate modulus condition on $X \times \widetilde{B}_n$. The following lemma uses this particular definition of the additive cycle complex.

Lemma 5.5. *Let K be the function field of $\mathrm{SL}_{r+1,k}$, and $\phi : \square_K^1 \rightarrow \mathrm{SL}_{r+1,K}$ be as in the previous lemma. Let $\mathrm{SL}_{r+1,K}$ act on \mathbb{P}_K^r naturally. Consider the composition $H_n = p_{K/k} \circ \mathrm{pr}'_K \circ \mu_\phi$ of morphisms*

$$\mathbb{P}^r \times \mathbb{A}^1 \times \square_K^n \xrightarrow{\mu_\phi} \mathbb{P}^r \times \mathbb{A}^1 \times \square_K^n \xrightarrow{\mathrm{pr}'_K} \mathbb{P}^r \times \mathbb{A}^1 \times \square_K^{n-1} \xrightarrow{p_{K/k}} \mathbb{P}^r \times \mathbb{A}^1 \times \square_k^{n-1},$$

where

$$\begin{cases} \mu_\phi(x, t, y_1, \dots, y_n) := (\phi(y_1)x, t, y_1, \dots, y_n), \\ \mathrm{pr}'_K(x, t, y_1, \dots, y_{n-1}) := (x, t, y_2, \dots, y_{n-1}), \\ p_{K/k} : \text{base change.} \end{cases}$$

Then for any $Z \in \mathrm{TZ}^q(\mathbb{P}_k^r, n; m)$, the cycle $H_n^*(Z) = \mu_\phi^* \circ \mathrm{pr}'_K^*(Z_K)$ is admissible, hence it is in $\mathrm{TZ}^q(\mathbb{P}_K^r, n+1; m)$. Similarly, H_n^* carries $\mathrm{TZ}_W^q(\mathbb{P}_k^r, n; m)$ to $\mathrm{TZ}_{W_K}^q(\mathbb{P}_K^r, n+1; m)$.

Proof. It is enough to prove the second assertion, that for any irreducible admissible Z in $\mathrm{TZ}_W^q(\mathbb{P}^r, n; m)$, the variety $Z' := H_n^*(Z)$, that we informally call the “homotopy variety” of Z , satisfies the admissibility conditions of Definition 2.5.

Claim 1. *The variety Z' intersects $W \times F_K$ properly for all $W \in \mathcal{W}$ and for each face F of B_{n+1} .*

Proof. This follows from the arguments of [Bloch 1986, Lemma (2.2)] and [Levine 1998, Lemma 3.5.11] without any modification. We provide its proof for the sake of completeness. We may assume that \mathcal{W} contains only one nonempty algebraic set W . There are two cases to consider.

Case 1. Suppose F_K comes from $F = \mathbb{A}^1 \times \{0\} \times F'$ for some face $F' \subset \square^{n-1}$. In this case, $Z' \cap (W \times F_K)$ is nothing but $Z_K \cap (W \times \mathbb{A}^1 \times F'_K)$ because $\phi(0) = \mathrm{Id} \in \mathrm{SL}_{r+1,K}$. So, proper intersection is obvious in this case.

Case 2. Suppose F_K does not come from faces of the form in Case 1. We apply Lemma 5.1 with $G = \mathrm{SL}_{r+1,k}$, $X = \mathbb{P}^r \times F$, $A = W \times F$, and $B = \mathrm{pr}'_k{}^*(Z) \cap (\mathbb{P}^r \times F)$, where G acts on X by acting trivially on F and acting naturally on \mathbb{P}^r . By Lemma 5.1, there is a nonempty open subset $U \subset \mathrm{SL}_{r+1}$ such that for all $g \in U$, the intersection $g(A) \cap B$ is proper. By shrinking U if necessary, we may assume that U is invariant under taking the multiplicative inverses. Take $g = \eta^{-1} \in U$, the inverse of the generic point. Thus, after base extension to K , the intersection of $\eta^{-1}(W_K \times F_K)$ with $\mathrm{pr}'_K{}^*(Z_K) \cap (\mathbb{P}^r \times F_K)$ is proper, which means that $\eta(\mathrm{pr}'_K{}^*(Z_K) \cap (\mathbb{P}^r \times F_K))$ intersects properly with $W_K \times F_K$. But the intersection $\mathrm{pr}'_K{}^*(Z_K) \cap (\mathbb{P}^r \times F_K)$ is proper, as Z was admissible. Hence, $\eta(\mathrm{pr}'_K{}^*(Z_K))$ intersects with $W_K \times F_K$ properly. Since F is not of the form $\mathbb{A}^1 \times \{0\} \times F'$, F_K intersects the first component \square_K^1 at $\{\infty\}$ nontrivially. In particular, $\eta(\mathrm{pr}'_K{}^*(Z_K))$ is the same as $\mu_\phi^*(\mathrm{pr}'_K{}^*(Z_K)) = Z'$ by Lemma 5.4. We conclude that Z' intersects with $W_K \times F_K$ properly. This proves the claim and hence Z' has the good-intersection property. Thus we only need to show the modulus condition for Z' to complete the proof of the lemma.

Claim 2. Z' satisfies the modulus condition on $\mathbb{P}^r \times \widetilde{B}_{n+1,K}$.

Proof. We prove this using our containment lemma. In the following, we casually drop the automorphism $\tau : \mathbb{P}^r \times \mathbb{A}^1 \times \square^n \rightarrow \mathbb{P}^r \times \mathbb{A}^1 \times \square^n$ that maps (x, t, y_1, \dots, y_n) to $(x, t, y_2, \dots, y_n, y_1)$ from our notations for simplicity.

Take $V = p(Z)$, where $p : \mathbb{P}^r \times \widetilde{B}_n \rightarrow \widetilde{B}_n$ is the projection. Because $Z \subset p^{-1}(p(Z)) = \mathbb{P}^r \times V$, we have

$$Z' = \mu_\phi^*(Z \times \square_K^1) \subset \mu_\phi^*(\mathbb{P}^r \times V \times \square_K^1) = \mathbb{P}^r \times V \times \square_K^1 =: Z_1, \text{ say.} \quad (5-1)$$

Now, Proposition 5.2 implies that V is an irreducible admissible closed subvariety of \widetilde{B}_n . The flat pull-back property in turn implies that $p^*([V]) = \mathbb{P}^r \times V$ is an irreducible admissible closed subvariety of $\mathbb{P}^r \times \widetilde{B}_n$. In particular, the modulus condition holds for $\mathbb{P}^r \times V$. If \overline{V} is the closure of V in \widehat{B}_n , then commutativity of the diagram

$$\begin{array}{ccccc} \overline{Z}_1^N = \mathbb{P}^r \times \overline{V}^N \times \mathbb{P}_K^1 & \longrightarrow & \mathbb{P}^r \times \widehat{B}_{n+1,K} & \longrightarrow & \widehat{B}_{n+1,K} \\ \downarrow & & \downarrow & & \downarrow \\ \mathbb{P}^r \times \overline{V}^N & \longrightarrow & \mathbb{P}^r \times \widehat{B}_n & \longrightarrow & \widehat{B}_n \end{array}$$

now implies that Z_1 satisfies the modulus condition on $\mathbb{P}^r \times \widetilde{B}_{n+1,K}$ even though it is a degenerate additive cycle. Furthermore, the admissibility of Z and the fact that μ_ϕ is an automorphism imply that \overline{Z}' intersects the Cartier divisors F_{n+1}^1 and $F_{n+1,0}$ properly. Thus we can use (5-1) and apply Proposition 2.4 to conclude that Z' satisfies the modulus condition. This completes the proof of the lemma. \square

Lemma 5.6. *The collection $H_\bullet^* : \mathrm{TZ}^q(\mathbb{P}_k^r, \cdot; m) \rightarrow \mathrm{TZ}^q(\mathbb{P}_K^r, \cdot + 1; m)$ is a chain homotopy satisfying $\partial H^* + H^* \partial = Z_K - \eta(Z_K)$. The same is true for $\mathrm{TZ}_{\mathcal{W}}^q$.*

Proof. It is enough to prove the second assertion. This is straightforward: let $Z \in \mathrm{TZ}_{\mathcal{W}}^q(\mathbb{P}_k^r, n; m)$. Then

$$\begin{aligned} H^* \partial Z &= H^* \sum_{i=1}^{n-1} (-1)^i (\partial_i^\infty - \partial_i^0) Z = \sum_{i=1}^{n-1} (-1)^i (\mu_\phi^*(\mathrm{pr}'_K)^* p_{K/k}^*) (\partial_i^\infty - \partial_i^0) Z \\ &= \sum_{i=1}^{n-1} (-1)^i (\partial_{i+1}^\infty - \partial_{i+1}^0) \mu_\phi^*(\mathrm{pr}'_K)^* Z_K = - \sum_{i=2}^n (-1)^i (\partial_i^\infty - \partial_i^0) H^* Z, \\ \partial H^* Z &= \sum_{i=1}^n (-1)^i (\partial_i^\infty - \partial_i^0) H^* Z = \sum_{i=1}^n (-1)^i (\partial_i^\infty - \partial_i^0) H^* Z \\ &= (-1) (\partial_1^\infty - \partial_1^0) H^* Z + \sum_{i=2}^n (-1)^i (\partial_i^\infty - \partial_i^0) H^* Z. \end{aligned}$$

Hence, $(\partial H^* + H^* \partial) Z = (\partial_1^0 - \partial_1^\infty) H^* Z = Z_K - \eta(Z_K)$. \square

5B. Proof of the moving lemma for projective spaces. We are now ready to finish the proof of Theorem 4.1 for \mathbb{P}^r .

By Lemma 5.6, the base extension

$$p_{K/k}^* : \frac{\mathrm{TZ}^q(\mathbb{P}_k^r, \cdot; m)}{\mathrm{TZ}_{\mathcal{W}}^q(\mathbb{P}_k^r, \cdot; m)} \rightarrow \frac{\mathrm{TZ}^q(\mathbb{P}_K^r, \cdot; m)}{\mathrm{TZ}_{\mathcal{W}_K}^q(\mathbb{P}_K^r, \cdot; m)}$$

is homotopic to the map $\eta p_{K/k}^*$. Note for each admissible cycle $Z \in \mathrm{TZ}^q(\mathbb{P}_k^r, n; m)$, the cycle $\eta(Z_K)$ lies in $\mathrm{TZ}_{\mathcal{W}}^q(\mathbb{P}_K^r, n; m)$. Part of the proof of Claim 1 of Lemma 5.5 is similar to the proof of this assertion:

We may assume that \mathcal{W} has only one nonempty algebraic set, say W . Let F be a face of B_n . In Lemma 5.1, take $G = \mathrm{SL}_{r+1}$ and $X = \mathbb{P}^r \times F$ where G acts on \mathbb{P}^r naturally and B_n trivially. Let $A = W \times F$ and $B = Z \cap (\mathbb{P}^r \times F)$. Since SL_{r+1} acts transitively on \mathbb{P}^r , the map $G \times A \rightarrow X$ is surjective. Hence, by Lemma 5.1, there is a nonempty open subset $U \subset G$ such that for all $g \in U$, the intersection $g(A) \cap B$ is proper in X . By shrinking U further, we may assume that U is closed under taking multiplicative inverses. Taking $g = \eta^{-1}$, the inverse of the generic point, we see that after base extension to K , the intersection of $\eta^{-1}(W \times F)$ with $Z_K \cap (\mathbb{P}^r \times F_K)$ is proper, which means $\eta(Z_K \cap (\mathbb{P}^r \times F_K))$ intersects $W_K \times F_K$ properly. Since Z_K intersects with $\mathbb{P}^r \times F_K$ properly by the assumption, we conclude that $\eta(Z_K)$ intersects $W_K \times F_K$ properly. Thus, $\eta(Z_K) \in \mathrm{TZ}_{\mathcal{W}}^q(\mathbb{P}_K^r, n; m)$. Hence, the induced map on the quotient

$$\eta p_{K/k}^* : \frac{\mathrm{TZ}^q(\mathbb{P}_k^r, \cdot; m)}{\mathrm{TZ}_{\mathcal{W}}^q(\mathbb{P}_k^r, \cdot; m)} \rightarrow \frac{\mathrm{TZ}^q(\mathbb{P}_K^r, \cdot; m)}{\mathrm{TZ}_{\mathcal{W}_K}^q(\mathbb{P}_K^r, \cdot; m)}$$

is zero. Hence the base extension $p_{K/k}^*$ induces a zero map on homology since it is homotopic to the zero map.

On the other hand, by the spreading lemma, Proposition 4.7, the chain map $p_{K/k}^*$ is injective on homology, so the quotient complex $\mathrm{TZ}^q(\mathbb{P}_k^r, \cdot; m) / \mathrm{TZ}_{\mathcal{W}}^q(\mathbb{P}_k^r, \cdot; m)$ must be acyclic. This proves Theorem 4.1 for the projective spaces. \square

6. Generic projections and moving lemma for projective varieties

6A. Generic projections. This section begins with a review of some facts about linear projections. In combination with the moving lemma for \mathbb{P}^r , that we saw in the previous section, we prove the moving lemma for general smooth projective varieties.

Lemma 6.1. *Consider two integers $N > r > 0$. Then for each linear subvariety $L \subset \mathbb{P}^N$ of dimension $N - r - 1$, there exists a linear projection morphism $\pi_L : \mathbb{P}^N \setminus L \rightarrow \mathbb{P}^r$.*

Proof. Fix the coordinates $x = (x_0; \dots; x_N)$ of \mathbb{P}^N . A linear subvariety L is given by $(r+1)$ homogeneous linear equations in x whose corresponding $(N+1) \times (r+1)$ matrix A has the full rank $r+1$. Take the reduced row echelon form of A whose rows are the linear homogeneous functions $P_0(x), \dots, P_r(x)$ in x .

For $x \in \mathbb{P}^N \setminus L$, define $\pi_L(x) := (P_0(x); \dots; P_r(x))$. Since $x \notin L$, we have some $P_i(x) \neq 0$ so that the map π_L is well-defined. By elementary facts about reduced row echelon forms and row equivalences, the subvariety L uniquely decides this map π_L in this process. □

Let X be a smooth projective k -variety. Let $r = \dim X$. Suppose that we have an embedding $X \hookrightarrow \mathbb{P}^N$ for some $N > r$. Consider $\pi_L : \mathbb{P}^N \setminus L \rightarrow \mathbb{P}^r$. Whenever $L \cap X = \emptyset$, we have a finite morphism $\pi_{L,X} := \pi_L|_X : X \rightarrow \mathbb{P}^r$. Such L 's form a nonempty open subset $\text{Gr}(N-r-1, N)_X$ of the Grassmannian $\text{Gr}(N-r-1, N)$. Such a map π_L is automatically flat since X is smooth [Hartshorne 1977, Exercise III-10.9, p. 276]. In particular, the pull-back $\pi_{L,X}^*$ and push-forward $\pi_{L,X*}$ are defined by Theorem 3.1.

For any closed integral admissible cycle Z on $X \times B_n$, define $\tilde{L}(Z)$ to be

$$\tilde{L}(Z) := \pi_{L,X}^*(\pi_{L,X*}([Z])) - [Z].$$

Extending this map linearly, this defines a morphism of complexes

$$\tilde{L} : \text{TZ}^q(X, \cdot; m) \rightarrow \text{TZ}^q(X, \cdot; m).$$

6B. Chow's moving lemma. Recall that for two locally closed subsets A and B of pure codimension a and b , the *excess* of the intersection of A and B on X is defined to be

$$e(A, B) := \max\{a + b - \text{codim}_X(A \cap B), 0\}.$$

That the intersection $A \cap B$ is proper on X means $e(A, B) = 0$. If A and B are cycles, then we define $e(A, B) := e(\text{Supp}(A), \text{Supp}(B))$. The excess measures how far an intersection is from being proper.

Lemma 6.2 [Krishna and Levine 2008, Lemma 1.12]. *Let $X \subset \mathbb{P}^N$ be a smooth closed projective k -subvariety of dimension r . Let Z and W be cycles on X . Then*

there is a nonempty open subscheme $U_{Z,W} \subset \text{Gr}(N-r-1, N)_X$ such that for each field extension $K \supset k$ and each K -point L of $U_{Z,W}$, we have

$$e(\tilde{L}(Z), W) \leq \max\{e(Z, W) - 1, 0\}.$$

For its proof, see [Roberts 1972, Main Lemma, p. 93], or [Levine 1998, Lemma 3.5.4, p. 96] for a slightly different but equivalent version. The point of the projection business is the following lemma:

Lemma 6.3. *Let X be a smooth projective k -variety, and let \mathcal{W} be a finite set of locally closed algebraic subsets of X . Let $m, N \geq 1$, and $q \geq 0$ be integers. Let $e : \mathcal{W} \rightarrow \mathbb{Z}_{\geq 0}$ be a set-theoretic function. Define $e-1 : \mathcal{W} \rightarrow \mathbb{Z}_{\geq 0}$ by*

$$(e-1)(W) := \max\{e(W) - 1, 0\}.$$

Let K be the function field of $\text{Gr}(N-r-1, N)$, and let $L_{\text{gen}} \in \text{Gr}(N-r-1, N)_X(K)$ be the generic point. Then, the map

$$\tilde{L}_{\text{gen}} : \text{TZ}^q(X, \cdot; m) \rightarrow \text{TZ}^q(X_K, \cdot; m)$$

maps $\text{TZ}_{\mathcal{W},e}^q(X, \cdot; m)$ to $\text{TZ}_{\mathcal{W}_K,e-1}^q(X_K, \cdot; m)$.

Proof. The arguments of [Krishna and Levine 2008, Lemma 1.13, p. 84] or [Levine 1998, §3.5.6, p. 97] work in this additive context without change. The central idea is to use a variation of Chow's moving lemma as in Lemma 6.2. \square

6C. Proof of the moving lemma.

Proof of Theorem 4.1. Let L_{gen} be the generic point of the Grassmannian $\text{Gr}(N-r-1, N)$ as in Lemma 6.3. Then, for each function $e : \mathcal{W} \rightarrow \mathbb{Z}_{\geq 0}$, the morphism

$$\tilde{L}_{\text{gen}} = \pi_{L_{\text{gen}}}^* \circ \pi_{L_{\text{gen}}*} - p_{K/k}^* : \frac{\text{TZ}_{\mathcal{W},e}^q(X, \cdot; m)}{\text{TZ}_{\mathcal{W},e-1}^q(X, \cdot; m)} \rightarrow \frac{\text{TZ}_{\mathcal{W}_K,e}^q(X_K, \cdot; m)}{\text{TZ}_{\mathcal{W}_K,e-1}^q(X_K, \cdot; m)}$$

is zero. Hence $\pi_{L_{\text{gen}}}^* \circ \pi_{L_{\text{gen}}*}$ is equal to the base extension morphism $p_{K/k}^*$ on the quotient complex.

On the other hand, $\pi_{L_{\text{gen}}}^* \circ \pi_{L_{\text{gen}}*}$ is written in detail as

$$\frac{\text{TZ}_{\mathcal{W},e}^q(X, \cdot; m)}{\text{TZ}_{\mathcal{W},e-1}^q(X, \cdot; m)} \xrightarrow{\pi_{L_{\text{gen}}*}} \frac{\text{TZ}_{\mathcal{W}'_K,e'}^q(\mathbb{P}_K^r, \cdot; m)}{\text{TZ}_{\mathcal{W}'_K,e'-1}^q(\mathbb{P}_K^r, \cdot; m)} \xrightarrow{\pi_{L_{\text{gen}}}^*} \frac{\text{TZ}_{\mathcal{W}_K,e}^q(X_K, \cdot; m)}{\text{TZ}_{\mathcal{W}_K,e-1}^q(X_K, \cdot; m)},$$

where \mathcal{W}' and e' are defined as follows: for each $W \in \mathcal{W}$, the constructible subset $\pi_{L_{\text{gen}}}(W)$ can be written as

$$\pi_{L_{\text{gen}}}(W) = W'_1 \cup \dots \cup W'_{i_W}$$

for some $i_W \in \mathbb{N}$ and locally closed irreducible sets W'_j in \mathbb{P}_K^r . Let $d_j = \text{codim}_{\mathbb{P}_K^r}(W'_j) - \text{codim}_X(W)$. Let $\mathcal{W}' = \{W'_j \mid W \in \mathcal{W}\}$. Define $e' : \mathcal{W}' \rightarrow \mathbb{Z}_{\geq 0}$ by the rule

$e'(W'_j) := e(W) + d_j$. We have already shown in Section 5B that the moving lemma is true for all projective spaces. In particular, for all functions $e' : W' \rightarrow \mathbb{Z}_{\geq 0}$, the complex in the middle

$$\frac{\mathrm{TZ}_{W'_K, e'}^q(\mathbb{P}_K^r, \cdot; m)}{\mathrm{TZ}_{W'_K, e'-1}^q(\mathbb{P}_K^r, \cdot; m)}$$

is acyclic (see Remark 4.4). Hence, the base extension map

$$P_{K/k}^* : \frac{\mathrm{TZ}_{W, e}^q(X, \cdot; m)}{\mathrm{TZ}_{W, e-1}^q(X, \cdot; m)} \rightarrow \frac{\mathrm{TZ}_{W_K, e}^q(X_K, \cdot; m)}{\mathrm{TZ}_{W_K, e-1}^q(X_K, \cdot; m)}$$

is zero on homology. Consequently, by induction, the base extension map

$$P_{K/k}^* : \frac{\mathrm{TZ}^q(X, \cdot; m)}{\mathrm{TZ}_W^q(X, \cdot; m)} \rightarrow \frac{\mathrm{TZ}^q(X_K, \cdot; m)}{\mathrm{TZ}_{W_K}^q(X_K, \cdot; m)}$$

is zero on homology. On the other hand, this map is also injective on homology by Proposition 4.7. This happens only when

$$\frac{\mathrm{TZ}^q(X, \cdot; m)}{\mathrm{TZ}_W^q(X, \cdot; m)}$$

is acyclic, i.e., the inclusion $\mathrm{TZ}_W^q(X, \cdot; m) \rightarrow \mathrm{TZ}^q(X, \cdot; m)$ is a quasiisomorphism. □

7. Application to contravariant functoriality

In this section, we prove the following general contravariance property of the additive higher Chow groups as an application of the moving lemma.

Theorem 7.1. *Let $f : X \rightarrow Y$ be a morphism of quasiprojective varieties over k , where Y is smooth and projective. Then there is a pull-back map*

$$f^* : \mathrm{TH}^q(Y, n; m) \rightarrow \mathrm{TH}^q(X, n; m)$$

such that for a composition $X \xrightarrow{f} Y \xrightarrow{g} Z$ with Y and Z smooth and projective, we have

$$(g \circ f)^* = f^* \circ g^* : \mathrm{TH}^q(Z, n; m) \rightarrow \mathrm{TH}^q(X, n; m).$$

Before proving this functoriality, we mention one more consequence of our containment lemma (Proposition 2.4).

Corollary 7.2. *Let $X \xrightarrow{i} Y$ be a regular closed embedding of quasiprojective but not necessarily smooth varieties over k . Then there is a Gysin chain map of additive cycle complexes*

$$i^* : \mathrm{TZ}_{\{X\}}^q(Y, \cdot; m) \rightarrow \mathrm{TZ}^q(X, \cdot; m).$$

Proof. Let $\iota : Z \subset \widehat{Y} \times B_n$ be a closed irreducible admissible subvariety in the group $\mathrm{TZ}_{\{X\}}^q(Y, n; m)$. By assumption, Z intersects all faces $X \times F$ properly. Hence the abstract intersection product of cycles $(X \times B_n) \cdot Z = [\iota^*(X \times B_n)] \in z^q(X \times B_n)$ is well-defined. Moreover, the intersection formula for the regular embedding implies that this intersection product commutes with the boundary maps [Fulton 1998, §2.3 and §6.3]. We want this cycle to be $i^*(Z)$. Thus we only need to show that each component of $Z \cap (X \times B_n)$ satisfies the modulus condition in order for i^* to be a map of additive cycle complexes. Since $X \times \widehat{B}_n$ clearly intersects F_n^1 and $F_{n,0}$ properly on $Y \times \widehat{B}_n$, this modulus condition follows directly from Proposition 2.4, for Z has the modulus condition. \square

Proof of Theorem 7.1. We do this by imitating the proof of Theorem 4.1 in [Bloch 1986]. So, let $f : X \rightarrow Y$ be a map as in Theorem 7.1. Such a morphism can be factored as the composition

$$X \xrightarrow{\mathrm{gr}_f} X \times Y \xrightarrow{\mathrm{pr}_2} Y,$$

where gr_f is the graph of f and pr_2 is the projection. Notice that pr_2 is a flat map and moreover, the smoothness of Y implies that gr_f is a regular closed embedding. Let $\Gamma_f \subset X \times Y$ denote the image of gr_f which is necessarily closed.

For $0 \leq i \leq \dim X$, let Y_i be the Zariski closure of the collection of all points $y \in Y$ such that $\dim f^{-1}(y) \geq i$. We use the convention that $\dim \emptyset = -1$. Let \mathcal{W} be the collection of the irreducible components of all Y_i . Then \mathcal{W} is a finite collection.

Claim. *Let $Z \in \mathrm{TZ}_{\mathcal{W}}^q(Y, n; m)$ be an irreducible admissible closed subvariety of $Y \times B_n$. Then $(\mathrm{pr}_2 \times \mathrm{Id}_{B_n})^{-1}(Z) = X \times Z$ in $X \times Y \times B_n$ is an admissible closed subset that intersects $\Gamma_f \times F$ properly in $X \times Y \times B_n$ for all faces $F \subset B_n$. This gives a chain map*

$$\mathrm{pr}_2^* : \mathrm{TZ}_{\mathcal{W}}^q(Y, \cdot; m) \rightarrow \mathrm{TZ}_{\{\Gamma_f\}}^q(X \times Y, \cdot, m).$$

That $(\mathrm{pr}_2 \times \mathrm{Id}_{B_n})^{-1}(Z) = X \times Z$ is admissible is obvious by [Krishna and Levine 2008, §3.4]. Since Z intersects $W \times F$ properly for all $W \in \mathcal{W}$ and faces $F \subset B_n$, we have $\dim \widetilde{Z}_i \leq \dim Y_i + \dim F - q$, where $\widetilde{Z}_i := Z \cap (Y_i \times F)$.

Now, $(X \times Z) \cap (\Gamma_f \times F) = \bigcup_i X \times \widetilde{Z}_i$, and for each i we have $\dim(X \times \widetilde{Z}_i) = \dim X + \dim \widetilde{Z}_i \leq \dim X + \dim F - q = \dim(\Gamma_f \times F) - q$. We conclude that $\mathrm{codim}_{\Gamma_f \times F}(X \times Z) \cap (\Gamma_f \times F) \geq q$, thus obtaining the desired map

$$\mathrm{pr}_2^* : \mathrm{TZ}_{\mathcal{W}}^q(Y, n; m) \rightarrow \mathrm{TZ}_{\{\Gamma_f\}}^q(X \times Y, n; m)$$

for each $n \geq 1$. That this gives a chain map is obvious since f^* clearly commutes with the boundary maps. This proves the claim.

The pull-back map f^* is now given by composing pr_2^* with the Gysin map gr^* of Corollary 7.2 and then using the moving lemma, Theorem 4.1. The composition law can be checked directly from the construction of f^* . This completes the proof of Theorem 7.1. \square

8. Remarks and computations

8A. Moving modulus conditions. We saw that M_{sum} and M_{ssup} seem to have much better structural behavior than the modulus condition M_{sup} of [Krishna and Levine 2008; Park 2009], and this makes the former better suited for being a motivic cohomology. On the other hand, in the main theorem of [Park 2009], the regulators on 1-cycles were defined with the modulus condition M_{sup} . Although we have seen that this regulator map does exist and has good properties with the modulus condition M_{ssup} , its construction doesn't automatically generalize to the groups with M_{sum} . So, one may ask the following.

Question 8.1. *Given an M_{sum} -admissible cycle ξ with $\partial\xi = 0$, can one find an M_{sup} -admissible cycle η and an M_{sum} -admissible cycle Γ such that $\xi = \eta + \partial\Gamma$?*

A positive answer to this question will immediately solve one part of Conjecture 2.8. This is a kind of *deeper moving lemma* than we have proved in this paper. This moving lemma allows one to move the modulus as well as the proper intersection property when we move a cycle. On the other hand, the moving lemma of this paper does not allow changing the modulus conditions. We expect the answer to the above question to be much harder.

8B. Examples.

Example 8.2. We give a simple example where the homotopy used in [Bloch 1986; Levine 1998] doesn't preserve the modulus conditions for additive higher Chow groups of quasiprojective varieties.

Take $X = \mathbb{A}_k^1$ and $n = 1$, so we are interested in admissible cycles in $X \times \tilde{B}_1 = X \times \mathbb{A}_k^1$. Admissible closed subvarieties $Z \subset X \times \mathbb{A}_k^1$ are given by the condition $Z \cap (X \times \{0\}) = \emptyset$. Let $\mathbb{G}_{a,k} = \mathbb{A}_k^1$ act on X by translation, and take its function field $K = k(s)$, s transcendental over k . Take the line $\phi : \square_K^1 \rightarrow \mathbb{G}_{a,K}$ defined by $y \mapsto sy/(y - 1)$ that sends 0 to 0 and ∞ to the k -generic point s of $\mathbb{G}_{a,k}$, which is K -rational in $\mathbb{G}_{a,K}$.

Take Z given by the ideal $(xt+1) \subset k[x, t]$, which is in $\text{TZ}^1(\mathbb{A}^1, 1; m)$. Then, Z_K is given by $(xt+1) \subset K[x, t]$ and $\text{pr}'^* Z_K$ is given by $(xt+1) \subset K[x, t, y/(y - 1)]$. Pulling back through μ_ϕ , we get $(x + sy/(y - 1))t + 1 = 0$. This is the equation for our homotopy variety Z' . Rewriting it as $1 - y = t((y - 1)x + sy)$, we see that it doesn't satisfy any of the given modulus conditions M_{sum} , M_{sup} , M_{ssup} . For

instance, for a given $m \geq 1$, we need $1 - y$ to be divisible by at least t^{1+m} where $m \geq 1$, which is obviously false in this case. Hence $Z' \notin \text{TZ}^1(\mathbb{A}_K^1, 2; m)$.

Example 8.3. Recall from Remark 5.3 that if X is projective, then admissible cycles in $X \times \tilde{B}_1 = X \times \mathbb{A}^1$ have a very simple description: an admissible irreducible closed subvariety Z should be of the form $Y \times \{*\} \subset X \times \mathbb{A}^1$ for some closed subvariety $Y \subset X$, and a closed point $\{*\} \neq \{0\}$ of \mathbb{A}^1 . This variety obviously satisfies all of the modulus conditions.

Note that the admissible variety Z in Example 8.2 is not of the form $Y \times \{*\}$: this happens because $X = \mathbb{A}_k^1$ is not complete.

These two examples seem to suggest that one should possibly modify the definition of the additive higher Chow groups of a quasiprojective variety in such a way that it takes into account the behavior at infinity in any compactification of the underlying variety.

8C. A computation. We finish the paper with a calculation of some additive higher Chow groups, which the authors completed while working on this paper. The following extends [Bloch and Esnault 2003a, Theorem 6.4, p. 153] to affine spaces.

Theorem 8.4. *Assume that $\frac{1}{6} \in k$. Let M be a modulus condition M_{sum} , M_{sup} , or M_{ssup} . Let $X = \mathbb{A}_k^r$, and let $m = 1$. Then, the additive higher Chow groups of zero-dimensional cycles of X are the absolute Kähler differentials of k :*

$$\text{TH}^{r+n}(X, n; 1) \simeq \Omega_{k/\mathbb{Z}}^{n-1}.$$

Remark 8.5. Note that, although it looks similar, this theorem does not imply that additive higher Chow groups have \mathbb{A}^1 -homotopy invariance. For the structure morphism $\mathbb{A}_k^r \rightarrow \text{Spec}(k)$, the pull-backs of 0-cycles on $\text{Spec}(k) \times \tilde{B}_n$ to $X \times \tilde{B}_n$ are r -cycles, not 0-cycles.

Proof. The proof is very similar to that of [Bloch and Esnault 2003a, Theorem 6.4, p. 153]. For a closed point $p \in X \times \tilde{B}_n$ that does not intersect the faces and the divisor $\{t = 0\}$, we define a homomorphism by setting

$$\psi(p) := \text{Tr}_{k(p)/k} \left(\frac{1}{t} \frac{dy_1}{y_1} \wedge \cdots \wedge \frac{dy_{n-1}}{y_{n-1}} \right) (p) \in \Omega_{k/\mathbb{Z}}^{n-1}.$$

In other words, we ignore the coordinate of X . This defines a homomorphism $\psi : \text{TZ}^{r+n}(X, n; 1) \rightarrow \Omega_{k/\mathbb{Z}}^{n-1}$.

Claim 1. *The composition*

$$\psi \circ \partial : \text{TZ}^{r+n}(X, n + 1; 1) \xrightarrow{\partial} \text{TZ}^{r+n}(X, n; 1) \xrightarrow{\psi} \Omega_{k/\mathbb{Z}}^{n-1}$$

is zero.

Proof. This follows from [Bloch and Esnault 2003a, Proposition 6.2, p. 150]. \square

Claim 2. Any two closed admissible points $p, p' \in X \times \widetilde{B}_n$ for which only the coordinates of X differ are equivalent as additive higher Chow cycles.

Proof. Note that the points p, p' are not assumed to be k -rational. Under the natural projections $\pi_? : X \times \widetilde{B}_n \rightarrow ?$, where $? = X, \mathbb{A}^1$ and the i -th projection $\pi_i : X \times \widetilde{B}_n \rightarrow \square$, if one has $\pi_X(p) = a \in X$, $\pi_{\mathbb{A}^1}(p) = b \in \mathbb{A}^1$, and $\pi_i(p) = s_i \in \square$, for not necessarily k -rational closed points $a \in X, 0 \neq b \in \mathbb{A}^1, 0, \infty \neq s_i \in \square$, then one writes $p = (a, b, s_1, \dots, s_{n-1})$. Similarly, under the assumptions of Claim 2, one can write p' as $p' = (a', b, s_1, \dots, s_{n-1})$, where a' is another closed point of X . Consider a parametrized line given in terms of the above notation,

$$C = \left\{ \left(a \frac{y}{y-1} + a' \left(1 - \frac{y}{y-1} \right), b, y, s_1, \dots, s_{n-1} \right) \in X \times \widetilde{B}_{n+1} \mid y \in \square^1 \right\},$$

which is a closed 1-dimensional subvariety of $X \times \widetilde{B}_{n+1}$. This 1-cycle satisfies all the modulus conditions $M_{\text{sum}}, M_{\text{sup}}$, and M_{ssup} having $b \neq 0$, and it intersects all faces properly having constant y_i -coordinate values s_i . Thus C is admissible.

By direct calculations, $\partial_1^0(C) = p', \partial_1^\infty(C) = p$, and $\partial_i^\epsilon(C) = 0$ for $i \geq 2$ and $\epsilon \in \{0, \infty\}$. Hence, $\partial(C) = p' - p$ proving Claim 2.

Given Claim 2, by [Bloch and Esnault 2003a, Proposition 6.3] and the rest of the arguments of [Bloch and Esnault 2003a, Theorem 6.4] for which $\frac{1}{6} \in k$ is used, the theorem follows. □

We remark that the same arguments work for any variety X as long as we can prove Claim 2. In particular, for any connected union of affine spaces, irreducible or not, we can conclude the same results.

Acknowledgements

The authors would like to thank Spencer Bloch, H el ene Esnault, and Marc Levine for their invaluable comments on the work. The authors feel very grateful to the anonymous referees who provided various suggestions that greatly improved many parts of this article.

Park would like to thank TIFR and KAIST for their hospitality and reduced teaching loads, and Juya for support throughout this work.

For this work, Park was partially supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0063180).

References

[Bloch 1986] S. Bloch, ‘‘Algebraic cycles and higher K -theory’’, *Adv. in Math.* **61**:3 (1986), 267–304. MR 88f:18010 Zbl 0608.14004

[Bloch 1994] S. Bloch, ‘‘The moving lemma for higher Chow groups’’, *J. Algebraic Geom.* **3**:3 (1994), 537–568. MR 96c:14007 Zbl 0830.14003

- [Bloch and Esnault 2003a] S. Bloch and H. Esnault, “The additive dilogarithm”, pp. 131–155 in *Documenta Mathematica, Kazuya Kato’s Fiftieth Birthday*, edited by S. e. a. Bloch, Documenta Mathematica, Bielefeld, Germany, 2003. Extra Vol. MR 2005e:19006 Zbl 1052.11048
- [Bloch and Esnault 2003b] S. Bloch and H. Esnault, “An additive version of higher Chow groups”, *Ann. Sci. École Norm. Sup. (4)* **36**:3 (2003), 463–477. MR 2004c:14035 Zbl 1100.14014
- [Chow 1956] W.-L. Chow, “On equivalence classes of cycles in an algebraic variety”, *Ann. of Math.* **64** (1956), 450–479. MR 18,509a Zbl 0073.37304
- [Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete **2**, Springer, Berlin, 1998. MR 99d:14003 Zbl 0885.14002
- [Guillén and Navarro Aznar 2002] F. Guillén and V. Navarro Aznar, “Un critère d’extension des foncteurs définis sur les schémas lisses”, *Publ. Math. Inst. Hautes Études Sci.* **95** (2002), 1–91.
- [Hanamura 2004] M. Hanamura, “Mixed motives and algebraic cycles, II”, *Invent. Math.* **158**:1 (2004), 105–179. MR 2005g:14021 Zbl 1068.14022
- [Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001
- [Krishna and Levine 2008] A. Krishna and M. Levine, “Additive higher Chow groups of schemes”, *J. Reine Angew. Math.* **619** (2008), 75–140. MR 2009d:14005 Zbl 1158.14009
- [Krishna and Park 2011] A. Krishna and J. Park, “DGA-structure on additive higher Chow groups”, preprint, 2011, available at <http://mathsci.kaist.ac.kr/~jinhyun/papers/CDGA.pdf>.
- [Krishna and Park 2012] A. Krishna and J. Park, “Mixed motives over $k[t]/(t^{m+1})$ ”, *J. Inst. Math. Jussieu* **11**:3 (2012), 611–657.
- [Levine 1998] M. Levine, *Mixed motives*, Mathematical Surveys and Monographs **57**, American Mathematical Society, Providence, RI, 1998. MR 99i:14025 Zbl 0902.14003
- [Loday 1998] J.-L. Loday, *Cyclic homology*, 2nd ed., Grundlehren der Math. Wissenschaften **301**, Springer, Berlin, 1998. MR 98h:16014 Zbl 0885.18007
- [Park 2007] J. Park, “Algebraic cycles and additive dilogarithm”, *Int. Math. Res. Not.* **2007**:18 (2007). MR 2008i:14015
- [Park 2009] J. Park, “Regulators on additive higher Chow groups”, *Amer. J. Math.* **131**:1 (2009), 257–276. MR 2009k:14012 Zbl 1176.14001
- [Roberts 1972] J. Roberts, “Chow’s moving lemma”, Appendix 2 to “Motives”, by S. L. Kleiman, pp. 53–82 in *Algebraic geometry* (Oslo, 1970), edited by F. Oort, Wolters-Noordhoff, Groningen, 1972. MR 52 #3152 Zbl 0285.14005
- [Rülling 2007] K. Rülling, “The generalized de Rham-Witt complex over a field is a complex of zero-cycles”, *J. Algebraic Geom.* **16**:1 (2007), 109–169. MR 2007j:14006 Zbl 1122.14006

Communicated by Hélène Esnault

Received 2010-05-30

Revised 2011-01-09

Accepted 2011-02-06

amal@math.tifr.res.in

*School of Mathematics,
Tata Institute of Fundamental Research,
Homi Bhabha Road, Colaba, Mumbai-400005, India*

jinhyun@mathsci.kaist.ac.kr

*Department of Mathematical Sciences,
Korea Advanced Institute of Science and Technology,
291 Daehak-ro, Yuseong-gu, Daejeon 305-701, South Korea*

Fusion rules for abelian extensions of Hopf algebras

Christopher Goff

We investigate the representation theory and fusion rules of a class of cocentral abelian (quasi-)Hopf extensions of Hopf algebras which includes twisted (generalized) quantum doubles of finite groups, and a certain quasi-Hopf algebra of Schauenburg associated to group-theoretical fusion categories. We then present a nontrivial example with noncommutative fusion rules.

1. Introduction

We present here a “ground-up” approach to attaining the fusion rules for a class of cocentral abelian extensions of Hopf algebras. Moreover, by not requiring strict coassociativity of the coproduct in the extension, our results are applicable not only to cocentral abelian (Hopf) extensions of Hopf algebras, but also to certain quasi-Hopf extensions as well. One such example, from [Schauenburg 2002], arises in the study of group-theoretical fusion categories (see also [Natale 2005]). (For a definition of group-theoretical fusion categories and basic properties, see [Etingof et al. 2005].) Another family of examples includes the twisted quantum double of a finite group, introduced in [Dijkgraaf et al. 1991], and the generalization which is defined in [Goff and Mason 2010].

In Section 2, we review definitions and notation, largely following [Kashina et al. 2002; Witherspoon 2004]. (For more information on extensions of Hopf algebras, consult [Andruskiewitsch 1996; Montgomery 1993], or, for quasi-Hopf extensions, [Masuoka 2002].) Then, Section 3 contains explicit formulas for irreducible characters and central idempotents for such extensions, as well as the inner product for which the irreducible characters form an orthonormal set. In Section 4, we write down the character of the tensor product representation and combine it with the inner product to deduce the fusion coefficients. The main result containing the fusion coefficients for irreducible representations, Theorem 4.5, is anticipated in [Witherspoon 2004] but is presented in this note without reference to Hochschild

MSC2000: primary 16S40, 18D10; secondary 16W30.

Keywords: fusion rules, Hopf algebras, quasi-Hopf, abelian extensions, group-theoretical fusion categories.

cohomology *per se*. Corollary 4.6 points out the connection to the K_0 -ring of a group-theoretical fusion category. Then, in Section 5, we apply these formulas to a generalized twisted quantum double of a finite group [Goff and Mason 2010]. Indeed, our Section 5 supersedes [Goff and Mason 2010, Section 3]. Finally, Section 6 contains a nontrivial example of a cocentral abelian extension having noncommutative fusion rules.

2. Cocentral abelian extensions

We follow closely the notation of [Kashina et al. 2002] with a few exceptions. First, our action is a right action, consistent with [Andruskiewitsch and Natale 2003]. Second, our modules will be right modules rather than left.

Let L and G be finite groups and let \mathbb{F} be an algebraically closed field of characteristic not dividing $|G||L|$. An abelian extension H is of the form

$$0 \rightarrow (\mathbb{F}G)^* \rightarrow H \rightarrow \mathbb{F}L \rightarrow 0,$$

where $H = (\mathbb{F}G)^* \#_{\sigma}^{\tau} \mathbb{F}L$, $\sigma : \mathbb{F}L \otimes \mathbb{F}L \rightarrow (\mathbb{F}G)^*$ is a group 2-cocycle, and $\tau : (\mathbb{F}G)^* \rightarrow \mathbb{F}L \otimes \mathbb{F}L$ is the dual of a group 2-cocycle. The condition on \mathbb{F} assures that H is semisimple and cosemisimple. We specialize to a *cocentral* abelian extension, meaning that $(\mathbb{F}L)^* \subseteq Z(H^*)$, and thus that the coaction $\mathbb{F}L \rightarrow \mathbb{F}L \otimes (\mathbb{F}G)^*$ inherent in the extension is trivial. The cocentrality also has consequences for the tensor product structure on irreducible modules, as we will see in Section 4.

There is a right action of $\mathbb{F}L$ on $(\mathbb{F}G)^*$ which induces an action on $\mathbb{F}G$ via $(f \leftarrow \ell)(g) := f(g \leftarrow \ell^{-1})$ for $g \in G, \ell \in L$, and extended linearly. Since L acts as automorphisms of $(\mathbb{F}G)^*$, L permutes the idempotents of the dual basis. Thus, the action can be viewed as an action of L on G , also by automorphisms. For the basis $\{p_g \mid g \in G\}$, we have $p_g \leftarrow \ell = p_{g \leftarrow \ell}$. Moreover, let L_g be the stabilizer of g in L and $\mathcal{O}(g)$ the orbit of g under the action of L . That is,

$$L_g = \{\ell \in L \mid g \leftarrow \ell = g\} \quad \text{and} \quad \mathcal{O}(g) = \{g \leftarrow \ell \mid \ell \in L\}.$$

Let T_g be a complete set of right coset representatives for L_g in L . That is, $L = \bigcup_{y \in T_g} L_g y$. Note that $\mathcal{O}(g) = \{g \leftarrow y \mid y \in T_g\}$.

We can write σ and τ in terms of the dual basis via

$$\sigma(x, y) = \sum_{g \in G} \sigma_g(x, y) p_g \quad \text{and} \quad \tau(x) = \sum_{g, h \in G} \tau_{g, h}(x) (p_g \otimes p_h),$$

where $\sigma_g(x, y), \tau_{g, h}(x) \in \mathbb{F}$. There are many identities satisfied by σ and τ , such as

$$\sigma_{g \leftarrow z}(x, y) \sigma_g(z, xy) = \sigma_g(z, x) \sigma_g(zx, y) \tag{1}$$

and

$$\tau_{g,h}(x)\tau_{g\leftarrow x,h\leftarrow x}(y)\sigma_g(x,y)\sigma_h(x,y) = \tau_{g,h}(xy)\sigma_{gh}(x,y), \tag{2}$$

for all $g, h \in G, x, y, z \in L$.

Writing $p_g \# x$ as $p_g \bar{x}$, we can write multiplication in H as

$$p_k \bar{z} p_h \bar{y} = \delta_{k\leftarrow z,h} \sigma_k(z,y) p_k \bar{z} \bar{y},$$

for all $h, k \in G, y, z \in L$. We also occasionally write p_g for $p_g \bar{1}$ and \bar{x} for $\sum_g p_g \bar{x}$, whence $\bar{x} p_g = p_{g\leftarrow x^{-1}} \bar{x}$. The unit element is $\bar{1}$.

For a cocentral abelian extension, the comultiplication is

$$\Delta(p_g \bar{x}) = \sum_{h \in G} \tau_{h,h^{-1}g}(x) p_h \bar{x} \otimes p_{h^{-1}g} \bar{x},$$

for all $g \in G, x \in L$. The counit ϵ satisfies $\epsilon(p_g \bar{x}) = \delta_{g,1}$. Finally, the antipode S is given by

$$S(p_g \bar{x}) = \sigma_{g^{-1}\leftarrow x}(x^{-1}, x)^{-1} \tau_{g^{-1},g}(x)^{-1} p_{g^{-1}\leftarrow x} \overline{x^{-1}}.$$

Remark 2.1. For H to be a Hopf algebra, Δ must be coassociative, which implies a certain condition on τ . We require only quasioassociativity, which implies the existence of other structures, and a related condition on τ . We omit these details here, as all of our examples are proved elsewhere [Dijkgraaf et al. 1991; Natale 2005; Andruskiewitsch 1996] to be either coassociative or quasioassociative.

3. Modules and characters

Irreducible modules for H are induced from irreducible modules for the group algebra of L_g , but twisted by the 2-cocycle σ_g . Select one g from each orbit under the action of L , then select T_g , a set of right coset representatives. Let

$$H_g := (\mathbb{F}G) \#_{\sigma} \mathbb{F}L_g.$$

If V is a right projective σ_g -representation space for L_g , then $V \otimes p_g$ is a right H_g -module via

$$(v \otimes p_g) \cdot (p_h \bar{x}) = \delta_{g,h} (v \cdot x \otimes p_g)$$

for all $v \in V, h \in G, x \in L_g$.

The irreducible modules for H are induced from these. Let $\widehat{V} = (V \otimes p_g) \otimes_{H_g} H$, which is then a right H -module under right multiplication by H . In other words,

$$\widehat{V} = \sum_{y \in T_g} (v \otimes p_g) \otimes \bar{y},$$

with action given by

$$\begin{aligned}
 [(v \otimes p_g) \otimes \bar{y}] \cdot p_h \bar{x} &= (v \otimes p_g) \sigma_{h \leftarrow y^{-1}}(y, x) p_{h \leftarrow y^{-1}} \overline{yx} \\
 &= (v \otimes p_g) \delta_{g, h \leftarrow y^{-1}} \sigma_g(y, x) p_g \overline{wy'} \\
 &= (v \otimes p_g) \delta_{g, h \leftarrow y^{-1}} \sigma_g(y, x) \sigma_g(w, y')^{-1} (p_g \bar{w}) (\bar{y}') \\
 &= \delta_{g, h \leftarrow y^{-1}} \frac{\sigma_g(y, x)}{\sigma_g(w, y')} [(v \cdot w \otimes p_g) \otimes \bar{y}'],
 \end{aligned}$$

where $w \in L_g, y' \in T_g$ are chosen so that $wy' = yx$.

We introduce the notation $V_{(g,\varphi)}$ to represent the H -module induced from the projective σ_g -representation of L_g that has character φ , and we let $\rho_{(g,\varphi)}$ be the representation of $V_{(g,\varphi)}$, and $\chi_{(g,\varphi)}$ its character. Then one calculates

$$\chi_{(g,\varphi)}(p_h \bar{x}) = \delta_{g \leftarrow y, h} \delta_{yxy^{-1} \in L_g} \frac{\sigma_g(y, x)}{\sigma_g(yxy^{-1}, y)} \varphi(yxy^{-1}), \tag{3}$$

where y is the unique element of T_g that maps g to h . We reiterate that $V_{(g,\varphi)}$ is irreducible if and only if φ is.

Remark 3.1. This can be seen as

$$\chi_{(g,\varphi)}(p_h \bar{x}) = \delta_{g \leftarrow y, h} \delta_{x \in L_h} \varphi^{(y)}(x), \tag{4}$$

where $\varphi^{(y)}$ is a projective representation of $L_h = L_g^y$ (conjugate to φ) with cocycle $\sigma_{g \leftarrow y} = \sigma_h$. See [Costache 2009, Lemma 59] for a similar calculation.

Before writing down the central idempotents, we first note that the character χ_{reg} of the regular representation ρ_{reg} on H satisfies $\chi_{\text{reg}}(p_h \bar{x}) = \delta_{x,1} |\mathbb{C}(h)| |L_h| = \delta_{x,1} |L|$, and that, from the semisimplicity of H ,

$$\rho_{\text{reg}} = \bigoplus_{(h,\psi)} \chi_{(h,\psi)}(1_H) \rho_{(h,\psi)},$$

where h ranges over the orbits and ψ ranges over the irreducible projective σ_h -representations of L_h . Let $z_{(g,\varphi)}$ denote the central idempotent corresponding to the representation $\rho_{(g,\varphi)}$. Then $\rho_{(h,\psi)}(z_{(g,\varphi)}) = \delta_{g,h} \delta_{\varphi,\psi} (\dim \varphi) |L : L_g| \text{id}$.

Set $z_{(g,\varphi)} = \sum_{c \in G, d \in L} \alpha_{c,d} p_c \bar{d}$. We find the $\alpha_{c,d}$ by determining the value of the regular character on $S(p_{a^{-1}} \bar{b}) z_{(g,\varphi)}$ two ways. First,

$$\begin{aligned}
 \chi_{\text{reg}}(S(p_{a^{-1}} \bar{b}) z_{(g,\varphi)}) &= \sum_{c \in G, d \in L} \sigma_{a \leftarrow b}(b^{-1}, b)^{-1} \tau_{a,a^{-1}}(b)^{-1} \alpha_{c,d} \chi_{\text{reg}}(p_{a \leftarrow b} \overline{b^{-1}} p_c \bar{d}) \\
 &= \tau_{a,a^{-1}}(b)^{-1} \alpha_{a,b} |L|.
 \end{aligned}$$

On the other hand, we have

$$\rho_{\text{reg}}(S(p_{a^{-1}} \bar{b}) z_{(g,\varphi)}) = (\dim \varphi) |L : L_g| \rho_{(g,\varphi)}(S(p_{a^{-1}} \bar{b})),$$

which means

$$\chi_{\text{reg}}(S(p_{a^{-1}}\bar{b})z_{(g,\varphi)}) = (\dim \varphi)|L : L_g|\sigma_{a \leftarrow b}(b^{-1}, b)^{-1}\tau_{a,a^{-1}}(b)^{-1}\chi_{(g,\varphi)}(p_{a \leftarrow b}\overline{b^{-1}})$$

Solving for $\alpha_{a,b}$, we obtain

$$z_{(g,\varphi)} = \frac{(\dim \varphi)}{|L_g|} \sum_{a \in G, b \in L} \frac{1}{\sigma_{a \leftarrow b}(b^{-1}, b)} \chi_{(g,\varphi)}(p_{a \leftarrow b}\overline{b^{-1}})(p_a\bar{b}).$$

Simplifying somewhat using the delta functions within $\chi_{(g,\varphi)}$, we have:

Lemma 3.2. *The central idempotent of H corresponding to $V_{(g,\varphi)}$ is*

$$z_{(g,\varphi)} = \frac{(\dim \varphi)}{|L_g|} \sum_{a \in G} \sum_{b \in L_a} \frac{1}{\sigma_a(b^{-1}, b)} \chi_{(g,\varphi)}(p_a\overline{b^{-1}})(p_a\bar{b}). \quad \square$$

Note that the first sum could be over $a \in \mathbb{O}(g)$, as $\chi = 0$ otherwise.

Proposition 3.3. *Letting*

$$\langle \alpha, \beta \rangle = \frac{1}{|L|} \sum_{a \in G} \sum_{b \in L_a} \frac{1}{\sigma_a(b^{-1}, b)} \alpha(p_a\overline{b^{-1}})\beta(p_a\bar{b}), \quad (5)$$

where α, β are characters of H , defines an inner product on the space of characters of H . The irreducible characters form an orthonormal basis with respect to this inner product.

We give three proofs to demonstrate the consistency with the character theory of projective representations of finite groups, and to demonstrate the relationship between certain conjugates of projective representations.

First proof. Clearly, (5) is linear in each component. The symmetry of (5) follows from (1) because $b \in L_a$. Using Lemma 3.2, we have

$$\begin{aligned} \langle \chi_{(g,\varphi)}, \chi_{(h,\psi)} \rangle &= \frac{1}{|L|} \sum_{a \in G} \sum_{b \in L_a} \frac{1}{\sigma_a(b^{-1}, b)} \chi_{(g,\varphi)}(p_a\overline{b^{-1}})\chi_{(h,\psi)}(p_a\bar{b}) \\ &= \frac{1}{|L|} \chi_{(h,\psi)} \left(\frac{|L_g|}{\dim \varphi} z_{(g,\varphi)} \right) \\ &= \left(\frac{1}{|L|} \frac{|L_g|}{\dim \varphi} \right) (\dim \varphi)|L : L_g| \cdot \delta_{g,h} \delta_{\varphi,\psi} = \delta_{g,h} \delta_{\varphi,\psi}. \quad \square \end{aligned}$$

Second proof. From (3), we obtain that $a \in \mathbb{O}(g) \cap \mathbb{O}(h)$ and thus $g = h$ or else the inner product is zero. Thus

$$\begin{aligned}
& \langle \chi_{(g,\varphi)}, \chi_{(h,\psi)} \rangle \\
&= \frac{\delta_{g,h}}{|L|} \sum_{\substack{a \in \mathbb{O}(g) \\ [a=g \leftarrow y]}} \sum_{b \in L_a} \frac{\sigma_g(y, b^{-1}) \sigma_g(y, b)}{\sigma_a(b^{-1}, b) \sigma_g(yb^{-1}y^{-1}, y) \sigma_g(yby^{-1}, y)} \varphi(yb^{-1}y^{-1}) \psi(yby^{-1}) \\
&= \frac{\delta_{g,h}}{|L|} \sum_{\substack{a \in \mathbb{O}(g) \\ [a=g \leftarrow y]}} \sum_{b \in L_a} \frac{1}{\sigma_g(yb^{-1}y^{-1}, yby^{-1})} \varphi(yb^{-1}y^{-1}) \psi(yby^{-1})
\end{aligned}$$

by repeated application of (1). Hence

$$\langle \chi_{(g,\varphi)}, \chi_{(h,\psi)} \rangle = \frac{\delta_{g,h}}{|L_g|} \sum_{c \in L_g} \frac{1}{\sigma_g(c^{-1}, c)} \varphi(c^{-1}) \psi(c) = \delta_{g,h} \langle \varphi, \psi \rangle_{L_g} = \delta_{g,h} \delta_{\varphi, \psi}.$$

Here, $\langle \cdot, \cdot \rangle_{L_g}$ denotes the usual inner product for projective σ_g -representations of L_g . See [Nauwelaerts and Van Oystaeyen 1991, Proposition 2.8], for instance. \square

Third proof. Using Remark 3.1,

$$\begin{aligned}
\langle \chi_{(g,\varphi)}, \chi_{(h,\psi)} \rangle &= \frac{\delta_{g,h}}{|L|} \sum_{\substack{a \in \mathbb{O}(g) \\ [a=g \leftarrow y]}} \sum_{b \in L_a} \frac{1}{\sigma_a(b^{-1}, b)} \varphi^{(y)}(b^{-1}) \psi^{(y)}(b) \\
&= \frac{\delta_{g,h}}{|L|} \sum_{a \in \mathbb{O}(g)} |L_a| \langle \varphi^{(y)}, \psi^{(y)} \rangle_{L_a} = \delta_{g,h} \delta_{\varphi, \psi}.
\end{aligned}$$

It is clear that $\varphi^{(y)} = \psi^{(y)}$ if and only if $\varphi = \psi$. \square

4. Fusion rules

The character of the tensor product representation (via Δ) is

$$\begin{aligned}
& \chi_{(g,\varphi) \otimes (h,\psi)}(p_a \bar{b}) \\
&= \sum_{\substack{f \in G \\ [f \in \mathbb{O}(g), f^{-1}a \in \mathbb{O}(h)] \\ [f=g \leftarrow y, f^{-1}a=h \leftarrow w]}} \delta_{b \in L_f \cap L_{f^{-1}a}} \varphi(yby^{-1}) \psi(wbw^{-1}) \frac{\tau_{f, f^{-1}a}(b) \sigma_g(y, b) \sigma_g(w, b)}{\sigma_g(yby^{-1}, y) \sigma_g(wbw^{-1}, b)} \\
&= \sum_{\substack{f \in G \\ [f \in \mathbb{O}(g), f^{-1}a \in \mathbb{O}(h)] \\ [f=g \leftarrow y, f^{-1}a=h \leftarrow w]}} \delta_{b \in L_f \cap L_{f^{-1}a}} \tau_{f, f^{-1}a}(b) \varphi^{(y)}(b) \psi^{(w)}(b) \\
&= \sum_{\substack{f \in G \\ [f \in \mathbb{O}(g), f^{-1}a \in \mathbb{O}(h)] \\ [f=g \leftarrow y, f^{-1}a=h \leftarrow w]}} \delta_{b \in L_f \cap L_{f^{-1}a}} [\varphi^{(y)} \otimes \psi^{(w)} \tau_{f, f^{-1}a}](b),
\end{aligned}$$

where $[\varphi^{(y)} \otimes \psi^{(w)} \tau_{f, f^{-1}a}]$ is a projective representation (of $L_f \cap L_{f^{-1}a} \leq L_a$) with cocycle σ_a . As explained in [Witherspoon 2004, (4.7)], the cocentrality of the extension, and the fact that the coproduct Δ is an algebra map, together imply that σ_a is cohomologous to $\sigma_f \cdot \sigma_{f^{-1}a}$ on $L_f \cap L_{f^{-1}a}$ via $\tau_{f, f^{-1}a}$. This is the content of Equation (2), which depends on the assumption of cocentrality.

Remark 4.1. If $h = 1$, then $\chi_{(g, \varphi) \otimes (1, \psi)} = \chi_{(g, \varphi \otimes \psi \downarrow_{L_g})}$. If $g = 1$ the result is similar. Hence, the irreducible representations induced from $1 \in G$ are in the center of the fusion algebra and their tensor products with other modules can be reduced to a calculation in the appropriate stabilizer. This generalizes a similar result in [Goff and Mason 2010].

We need two lemmas before calculating the fusion coefficients.

Lemma 4.2. *Let $a, f \in G, y \in L$.*

(1) *Let α and β be projective σ_f -representations of L_f . Then*

$$\langle \alpha, \beta \rangle_{L_f} = \langle \alpha^{(y)}, \beta^{(y)} \rangle_{L_f^y}.$$

Note that $\alpha^{(y)}$ and $\beta^{(y)}$ are $\sigma_{f \leftarrow y}$ -representations of $L_{f \leftarrow y} = L_f^y$.

(2) *Let α be a σ_f -representation of L_f and let β be a $\sigma_{f^{-1}a}$ -representation of $L_{f^{-1}a}$. Then*

$$[\alpha \otimes \beta \tau_{f, f^{-1}a}]^{(y)} = [\alpha^{(y)} \otimes \beta^{(y)} \tau_{f \leftarrow y, f^{-1}a \leftarrow y}]$$

as $\sigma_{a \leftarrow y}$ -representations of $L_f^y \cap L_{f^{-1}a}^y \leq L_a^y$.

Proof. The proof is straightforward, using (4), (1), and (2). □

We need a way to calculate products of L -orbits in $\mathbb{C}G$. The following formula appears in [Witherspoon 2004, Proof of Theorem 4.8], where the author relies on standard trace map properties of the L -algebra $\mathbb{Z}G$, citing general results of [Thévenaz 1995]. Our proof is specific to group actions on sets. Recall that if L acts on G , then L also acts on $G \times G$ diagonally: $(g_1, g_2) \leftarrow \ell = (g_1 \leftarrow \ell, g_2 \leftarrow \ell)$ for $\ell \in L, g_1, g_2 \in G$.

Lemma 4.3. *Let $g, h \in G$. Then*

$$\mathbb{O}(g)\mathbb{O}(h) = \sum_{x \in D} |L_{(g \leftarrow x)h} : L_{g \leftarrow x} \cap L_h| \mathbb{O}((g \leftarrow x)h),$$

where D is a complete set of $L_g \backslash L / L_h$ double coset representatives.

Proof. Consider the orbits of the diagonal action of L on $G \times G$. Evidently, $y \in L_g x L_h$ if and only if $\mathbb{O}_L((g \leftarrow x, h)) = \mathbb{O}_L((g \leftarrow y, h))$. Now pick $x \in D$ and consider the image of $\mathbb{O}_L((g \leftarrow x, h))$ in G under the product map. Clearly, the product $(g \leftarrow x)h$ is fixed by $L_{(g \leftarrow x)h}$ but also each component is fixed by

$L_g^x \cap L_h \leq L_{(g \leftarrow x)h}$. So, the number of distinct ordered pairs $(g \leftarrow xw, h \leftarrow w)$ such that $(g \leftarrow xw)(h \leftarrow w) = (g \leftarrow x)h$ is $|L_{(g \leftarrow x)h} : L_g^x \cap L_h|$. Since L acts by automorphisms, this is also the number of times $\mathbb{O}((g \leftarrow x)h)$ appears in this term of the sum. \square

Remark 4.4. The right hand side in Lemma 4.3 cannot generally be interpreted as a summation over distinct orbits. There may be $y \notin L_g x L_h$ for which $\mathbb{O}((g \leftarrow x)h) = \mathbb{O}((g \leftarrow y)h)$.

Anticipated in [Witherspoon 2004, Theorem 4.8], the following theorem gives the fusion coefficients for irreducible representations of H .

Theorem 4.5. *Let $g, h, k \in G$ and let φ be a σ_g -representation of L_g , ψ a σ_h -representation of L_h , and γ a σ_k -representation of L_k and consider the corresponding induced modules of H . Then*

$$\langle \chi(k, \gamma), \chi_{(g, \varphi) \otimes (h, \psi)} \rangle = \sum_{\substack{x \in D \\ (g \leftarrow x)h \in \mathbb{O}(k) \\ [(g \leftarrow xw')(h \leftarrow w') = k]}} \langle \gamma, [\varphi^{(xw')} \otimes \psi^{(w')} \tau_{g \leftarrow xw', h \leftarrow w'}] \rangle_{L_g^{xw'} \cap L_h^{w'}}$$

where D is a set of those $L_g \backslash L / L_h$ double coset representatives x satisfying

$$(g \leftarrow x)h \in \mathbb{O}(k),$$

and the inner product on $L_g^{xw'} \cap L_h^{w'} \leq L_k$ is of projective σ_k -representations.

Proof. Using the inner product (5), we have

$$\begin{aligned} & \langle \chi(k, \gamma), \chi_{(g, \varphi) \otimes (h, \psi)} \rangle \\ &= \frac{1}{|L|} \sum_{\substack{a \in \mathbb{O}(k) \\ [a = k \leftarrow z]}} \sum_{\substack{f \in \mathbb{O}(g) \\ f^{-1}a \in \mathbb{O}(h) \\ [f = g \leftarrow y] \\ [f^{-1}a = h \leftarrow w]}} \sum_{b \in L_a \cap L_f} \gamma(zb^{-1}z^{-1})\varphi(yby^{-1})\psi(wbw^{-1}) \\ & \quad \cdot \frac{\tau_{f, f^{-1}a}(b)\sigma_k(z, b^{-1})\sigma_g(y, b)\sigma_h(w, b)}{\sigma_a(b^{-1}, b)\sigma_k(zb^{-1}z^{-1}, z)\sigma_g(yby^{-1}, y)\sigma_h(wbw^{-1}, w)} \\ &= \frac{1}{|L|} \sum_{\substack{a \in \mathbb{O}(k) \\ [a = k \leftarrow z]}} \sum_{\substack{f \in \mathbb{O}(g) \\ f^{-1}a \in \mathbb{O}(h) \\ [f = g \leftarrow y] \\ [f^{-1}a = h \leftarrow w]}} \sum_{b \in L_a \cap L_f} \frac{\tau_{f, f^{-1}a}(b)}{\sigma_a(b^{-1}, b)} \gamma^{(z)}(b^{-1})\varphi^{(y)}(b)\psi^{(w)}(b) \\ &= \frac{1}{|L|} \sum_{\substack{a \in \mathbb{O}(k) \\ [a = k \leftarrow z]}} \sum_{\substack{f \in \mathbb{O}(g) \\ f^{-1}a \in \mathbb{O}(h) \\ [f = g \leftarrow y] \\ [f^{-1}a = h \leftarrow w]}} |L_f \cap L_{f^{-1}a}| \langle \gamma^{(z)}, [\varphi^{(y)} \otimes \psi^{(w)} \tau_{f, f^{-1}a}] \rangle_{L_f \cap L_{f^{-1}a}}. \end{aligned}$$

By Lemma 4.2 this is equal to

$$\begin{aligned}
 &= \frac{1}{|L|} \sum_{\substack{a \in \mathbb{O}(k) \\ [a=k \leftarrow z]}} \sum_{\substack{f \in \mathbb{O}(g) \\ f^{-1}a \in \mathbb{O}(h) \\ [f=g \leftarrow y] \\ [f^{-1}a=h \leftarrow w]}} |L_f \cap L_{f^{-1}a}| \cdot \langle \gamma, [\varphi^{(yz^{-1})} \otimes \psi^{(wz^{-1})} \tau_{f \leftarrow z^{-1}, f^{-1}a \leftarrow z^{-1}}] \rangle_{L_f^{z^{-1}} \cap L_{f^{-1}a}^{z^{-1}}} \\
 &= \frac{1}{|L_k|} \sum_{\substack{f \in \mathbb{O}(g) \\ f^{-1}k \in \mathbb{O}(h) \\ [f=g \leftarrow y'] \\ [f^{-1}k=h \leftarrow w']}} |L_f \cap L_{f^{-1}k}| \langle \gamma, [\varphi^{(y')} \otimes \psi^{(w')} \tau_{f, f^{-1}k}] \rangle_{L_f \cap L_{f^{-1}k}},
 \end{aligned}$$

and by Lemma 4.3 this can further be written as

$$\begin{aligned}
 &= \frac{1}{|L_k|} \sum_{\substack{x \in D \\ (g \leftarrow x)h \in \mathbb{O}(k) \\ [(g \leftarrow xw')(h \leftarrow w')=k]}} |L_g^{xw'} \cap L_h^{w'}| |L_{(g \leftarrow x)h} : L_g^x \cap L_h| \cdot \langle \gamma, [\varphi^{(xw')} \otimes \psi^{(w')} \tau_{f, f^{-1}k}] \rangle_{L_g^{xw'} \cap L_h^{w'}} \\
 &= \sum_{\substack{x \in D \\ (g \leftarrow x)h \in \mathbb{O}(k) \\ [(g \leftarrow xw')(h \leftarrow w')=k]}} \langle \gamma, [\varphi^{(xw')} \otimes \psi^{(w')} \tau_{g \leftarrow xw', h \leftarrow w'}] \rangle_{L_g^{xw'} \cap L_h^{w'}},
 \end{aligned}$$

where D is a set of $L_g \setminus L/L_h$ double coset representatives with $(g \leftarrow x)h \in \mathbb{O}(k)$. Thus, the fusion rules for H modules can be determined from the fusion rules for projective σ_k -representations restricted to certain subgroups of L_k . \square

As stated before, the theorem holds for certain quasi-Hopf extensions, including the examples in the following corollary and the next section.

Corollary 4.6. *The fusion rules in Theorem 4.5 describe the K_0 -ring for the group-theoretical module category $\mathcal{C}(G \rtimes L, \omega, L, 1)$, where $\omega \in H^3(G \rtimes L, \mathbb{F}^*)$ is the 3-cocycle associated to $[\sigma, \tau]$ in the relevant Kac exact sequence. See [Schauenburg 2002; Natale 2003; Masuoka 2002] for further cohomological details.*

Proof. Indeed, the theorem holds whenever the structure maps and (1) and (2) hold, even if H is a quasi-Hopf algebra (with coassociator Φ), because the fusion rules for H do not depend on the associativity constraint (determined by Φ) in the category of right H -modules, $\text{Mod-}H$. Thus these fusion rules hold for a certain quasi-Hopf algebra of Schauenburg, denoted (A^{op}, Φ) by Natale [2005], in the case when $A = (\mathbb{F}G)^* \#_{\sigma}^{\tau} \mathbb{F}L$, and the left action \triangleright of G on L is trivial; i.e., when $GL = G \rtimes L$. (In this case, the structure maps and cocycles are exactly as in Section 2.) Natale, in the proof of her Theorem 4.4, cites [Schauenburg 2002] to demonstrate that $(A^{\text{op}}, \Phi)\text{-Mod}$ is tensor-equivalent to $\mathcal{C}(G \rtimes L, \omega, L, 1)$, where $\omega \in H^3(G \rtimes L, \mathbb{F}^*)$ is the 3-cocycle associated to $[\sigma, \tau]$ in the Kac exact sequence. \square

5. Example: generalized twisted quantum doubles of finite groups

Other examples of abelian extensions satisfying the structure maps of Section 2 (and hence having fusion rules determined by Theorem 4.5) include twisted quantum doubles of finite groups [Dijkgraaf et al. 1991] and generalized twisted doubles of finite groups [Goff and Mason 2010]. We expand on the latter, but using right modules here. As mentioned earlier, this section supersedes [Goff and Mason 2010, Section 3].

Let G be a finite group, N a normal subgroup, and $\bar{G} := G/N$. We use the bar notation for elements in \bar{G} , i.e., if $g \in G$ then $\bar{g} = gN \in \bar{G}$. Then G acts naturally on \bar{G} via conjugation, namely $\bar{g} \leftarrow x := x^{-1}\bar{g}x = \bar{g}^x = \overline{g^x} = \bar{x}^{-1}\bar{g}\bar{x}$, for all $x \in G, \bar{g} \in \bar{G}$.

In addition, let $\omega \in H^3(\bar{G}, \mathbb{F}^*)$, and let $\omega' := \text{Infl}_G^{\bar{G}} \omega$. In analogy with σ and τ , define $\theta : \mathbb{F}G \otimes \mathbb{F}G \rightarrow \mathbb{F}\bar{G}^*$ and $\gamma : \mathbb{F}\bar{G}^* \rightarrow \mathbb{F}G \otimes \mathbb{F}G$ via

$$\theta = \sum_{\bar{g} \in \bar{G}} \theta_{\bar{g}} \quad \text{and} \quad \gamma = \sum_{x, y \in \bar{G}} \gamma_0(x, y),$$

where

$$\theta_{\bar{g}}(x, y) = \frac{\omega(\bar{g}, \bar{x}, \bar{y})\omega(\bar{x}, \bar{y}, \bar{g}^{xy})}{\omega(\bar{x}, \bar{g}^x, \bar{y})}, \quad \gamma_{\bar{g}}(x, y) = \frac{\omega(\bar{x}, \bar{y}, \bar{g})\omega(\bar{g}, \bar{x}^g, \bar{y}^g)}{\omega(\bar{x}, \bar{g}, \bar{y}^g)}.$$

Notice that $\theta_{\bar{g}}$ and $\gamma_{\bar{g}}$ could be thought of as functions from $\mathbb{F}\bar{G} \otimes \mathbb{F}\bar{G}$ to \mathbb{F}^* since they pass to the quotient \bar{G} . The generalized twisted double is then $D^\omega(G, \bar{G}) = (\mathbb{F}\bar{G})^* \#_{\theta}^{\gamma} (\mathbb{F}G)$. The maps θ and γ satisfy (1) and (2), *mutatis mutandis* [Dijkgraaf et al. 1991].

The irreducible (right) modules of $D^\omega(G, \bar{G})$ are induced from irreducible projective representations of centralizers. In particular, the character of the irreducible projective $\theta_{\bar{g}}$ -representation φ of $C_G(\bar{g})$ is given by

$$\begin{aligned} \widehat{\chi}_{(\bar{g}, \varphi)}(e(\bar{h}) \bowtie x) &= \delta_{\bar{g}^y, \bar{h}} \delta_{yxy^{-1} \in C_G(\bar{g})} \frac{\theta_{\bar{g}}(y, x)}{\theta_{\bar{g}}(yxy^{-1}, y)} \varphi(yxy^{-1}) \\ &= \delta_{\bar{g}^y, \bar{h}} \delta_{x \in C_G(\bar{h})} \varphi^{(y)}(x). \end{aligned}$$

Consistent with (5), the inner product on characters is given by

$$\langle \alpha, \beta \rangle = \frac{1}{|G|} \sum_{\bar{k} \in \bar{G}} \sum_{x \in C_G(\bar{k})} \frac{1}{\theta_{\bar{k}}(\bar{x}^{-1}, \bar{x})} \alpha(e(\bar{k}) \bowtie x^{-1}) \beta(e(\bar{k}) \bowtie x),$$

and thus the fusion coefficients are given by

$$\langle \widehat{\chi}_{(\bar{k}, \lambda)}, \widehat{\chi}_{(\bar{g}, \varphi) \otimes (\bar{h}, \psi)} \rangle = \sum_{\substack{x \in D \\ [\bar{g}^{xw'} \bar{h}^{w'} = \bar{k}]} } \langle \lambda, [\varphi^{(xw')} \otimes \psi^{(w')} \gamma_0(\bar{g}^{xw'}, \bar{h}^{w'})] \rangle_{C_G(\bar{g}^{xw'}) \cap C_G(\bar{h}^{w'})}$$

where D is a set of $C_G(\bar{g}) \backslash G / C_G(\bar{h})$ double coset representatives with $\bar{g}^x \bar{h} \in \mathbb{O}(\bar{k})$, and the inner product on $C_G(\bar{g}^{xw'}) \cap C_G(\bar{h}^{w'}) \leq C_G(\bar{k})$ is of $\theta_{\bar{k}}$ -representations.

6. Example: noncommutative fusion rules

Noncommutative fusion rules for cocentral abelian extensions are not rare: choose $L = 1, \sigma$ and τ trivial, and any nonabelian G , for instance. Also, see [Kosaki et al. 1997; Nikshych 1998; Zhu 2001]. We give here an example with σ, τ trivial, but nontrivial action of L . Let G be the dihedral group of order 18, and let

$$L \leq \text{Aut } D_9 \cong \mathbb{Z}_9 \rtimes \mathbb{Z}_2^*$$

Namely, $L = \langle 3 \rangle \times \langle 4 \rangle \cong \mathbb{Z}_3 \times \mathbb{Z}_3$. [The first factor of L is with respect to addition, the second, multiplication.] If we let $G = \langle x, y \mid x^2 = y^9 = e, yx = xy^{-1} \rangle$, then

$$(x^c y^d) \leftarrow (a, b) := x^c y^{ac+bd}.$$

We choose L -orbit representatives $S = \{e, y^3, y^6, y, y^2, x, xy, xy^2\}$ with their respective stabilizers. Since L is abelian, $\chi_{(g,\varphi)}^{(\ell)} = \chi_{(g,\varphi)}$ for all $\ell \in L$. Note that in the decomposition of the product of orbits, we have

$$\mathbb{O}(x)\mathbb{O}(y) = 3\mathbb{O}(xy) \quad \text{and} \quad \mathbb{O}(y)\mathbb{O}(x) = 3\mathbb{O}(xy^2),$$

which suffices to guarantee noncommutative fusion rules.

Theorem 6.1. *Let $M(g, \alpha)$ denote the irreducible representation of $(\mathbb{F}G)^*\#(\mathbb{F}L)$ induced from α , an irreducible representation of L_g for $g \in S$. The first five rules are commutative.*

- i. $M(s, \alpha) \otimes M(t, \beta) = M(st, \alpha \otimes \beta)$ for $s, t \in \langle y^3 \rangle$.
- ii. $M(s, \alpha) \otimes M(g, \beta) = M(g, \alpha \downarrow_{L_g} \otimes \beta)$ for $s \in \langle y^3 \rangle, g \in \{y, y^2, x, xy, xy^2\}$.
- iii. $M(g, \alpha) \otimes M(g, \beta) = 3M(h, \alpha \otimes \beta)$ if $\{g, h\} = \{y, y^2\}$.
- iv. $M(y, \alpha) \otimes M(y^2, \beta) = \bigoplus_{s \in \langle y^3 \rangle} \bigoplus_{\gamma \downarrow_{L_y} = \alpha \otimes \beta} M(s, \gamma)$.
- v. $M(g, \alpha) \otimes M(g, \beta) = \bigoplus_{s \in \langle y^3 \rangle} \bigoplus_{\gamma \downarrow_{L_g} = \alpha \otimes \beta} M(s, \gamma)$ for $g \in \{x, xy, xy^2\}$.

The rest of the list holds for all $\alpha, \beta, \delta, \epsilon, \zeta, \eta, \mu, \nu$.

- vi. $M(y, \alpha) \otimes M(x, \beta) = \bigoplus_{\text{all } \gamma} M(xy^2, \gamma) = M(x, \delta) \otimes M(y^2, \epsilon)$
 $= M(y^2, \zeta) \otimes M(xy, \eta) = M(xy, \mu) \otimes M(y, \nu)$.

$$\begin{aligned} \text{vii. } M(y, \alpha) \otimes M(xy, \beta) &= \bigoplus_{\text{all } \gamma} M(x, \gamma) = M(xy, \delta) \otimes M(y^2, \epsilon) \\ &= M(y^2, \zeta) \otimes M(xy^2, \eta) = M(xy^2, \mu) \otimes M(y, \nu). \end{aligned}$$

$$\begin{aligned} \text{viii. } M(y, \alpha) \otimes M(xy^2, \beta) &= \bigoplus_{\text{all } \gamma} M(xy, \gamma) = M(xy^2, \delta) \otimes M(y^2, \epsilon) \\ &= M(y^2, \zeta) \otimes M(x, \eta) = M(x, \mu) \otimes M(y, \nu). \end{aligned}$$

$$\begin{aligned} \text{ix. } M(x, \alpha) \otimes M(xy, \beta) &= \bigoplus_{\text{all } \gamma} M(y, \gamma) = M(xy, \delta) \otimes M(xy^2, \epsilon) \\ &= M(xy^2, \zeta) \otimes M(x, \eta). \end{aligned}$$

$$\begin{aligned} \text{x. } M(x, \alpha) \otimes M(xy^2, \beta) &= \bigoplus_{\text{all } \gamma} M(y^2, \gamma) = M(xy^2, \delta) \otimes M(xy, \epsilon) \\ &= M(xy, \zeta) \otimes M(x, \eta). \end{aligned}$$

Proof. Straightforward. □

Acknowledgement

The author acknowledges the gracious comments of the referee, which led to major improvements in the paper.

References

- [Andruskiewitsch 1996] N. Andruskiewitsch, “Notes on extensions of Hopf algebras”, *Canad. J. Math.* **48**:1 (1996), 3–42. MR 97c:16046 Zbl 0857.16033
- [Andruskiewitsch and Natale 2003] N. Andruskiewitsch and S. Natale, “Braided Hopf algebras arising from matched pairs of groups”, *J. Pure Appl. Algebra* **182** (2003), 119–149. MR 2004d:16064 Zbl 1024.16018
- [Costache 2009] T.-L. Costache, “On irreducible projective representations of finite groups”, *Surv. Math. Appl.* **4** (2009), 191–214. MR 2011h:20022 Zbl 1202.20015
- [Dijkgraaf et al. 1991] R. Dijkgraaf, V. Pasquier, and P. Roche, “Quasi Hopf algebras, group cohomology and orbifold models”, pp. 60–72 in *Proceedings of the fourth meeting on theoretical physics: Recent advances in field theory* (Annecy-le-Vieux, 1990), edited by P. Binétruy et al., Nuclear Phys. B Proc. Suppl. **18B**, Elsevier, Amsterdam, 1991. MR 92m:81238 Zbl 0957.81670
- [Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, “On fusion categories”, *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR 2006m:16051 Zbl 1125.16025
- [Goff and Mason 2010] C. Goff and G. Mason, “Generalized twisted quantum doubles and the McKay correspondence”, *J. Algebra* **324**:11 (2010), 3007–3016. MR 2011i:16046 Zbl 1228.16029
- [Kashina et al. 2002] Y. Kashina, G. Mason, and S. Montgomery, “Computing the Frobenius–Schur indicator for Abelian extensions of Hopf algebras”, *J. Algebra* **251** (2002), 888–913. MR 2003f:16061 Zbl 1012.16040

- [Kosaki et al. 1997] H. Kosaki, A. Munemasa, and S. Yamagami, “On fusion algebras associated to finite group actions”, *Pacific J. Math.* **177**:2 (1997), 269–290. MR 98i:46064 Zbl 0882.46030
- [Masuoka 2002] A. Masuoka, “Hopf algebra extensions and cohomology”, pp. 167–209 in *New directions in Hopf algebras*, edited by S. Montgomery and H.-J. Schneider, Math. Sci. Res. Inst. Publ. **43**, Cambridge Univ. Press, 2002. MR 2003d:16050 Zbl 1011.16024
- [Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, American Mathematical Society, Providence, RI, 1993. MR 94i:16019 Zbl 0793.16029
- [Natale 2003] S. Natale, “On group theoretical Hopf algebras and exact factorizations of finite groups”, *J. Algebra* **270**:1 (2003), 199–211. MR 2004k:16102 Zbl 1040.16027
- [Natale 2005] S. Natale, “Frobenius–Schur indicators for a class of fusion categories”, *Pacific J. Math.* **221**:2 (2005), 353–377. MR 2007j:16070 Zbl 1108.16035
- [Nauwelaerts and Van Oystaeyen 1991] E. Nauwelaerts and F. Van Oystaeyen, “Module characters and projective representations of finite groups”, *Proc. London Math. Soc. (3)* **62**:1 (1991), 151–166. MR 92f:20015 Zbl 0810.20012
- [Nikshych 1998] D. Nikshych, “ K_0 -rings and twisting of finite-dimensional semisimple Hopf algebras”, *Comm. Algebra* **26**:1 (1998), 321–342. MR 99d:16045a Zbl 0912.16018
- [Schauenburg 2002] P. Schauenburg, “Hopf bimodules, coquasibialgebras, and an exact sequence of Kac”, *Adv. Math.* **165**:2 (2002), 194–263. MR 2003e:16052 Zbl 1006.16054
- [Thévenaz 1995] J. Thévenaz, *G-algebras and modular representation theory*, Oxford University Press, New York, 1995. MR 96j:20017 Zbl 0837.20015
- [Witherspoon 2004] S. J. Witherspoon, “Products in Hochschild cohomology and Grothendieck rings of group crossed products”, *Adv. Math.* **185**:1 (2004), 136–158. MR 2005j:16010 Zbl 1063.16012
- [Zhu 2001] Y. Zhu, “Hecke algebras and representation ring of Hopf algebras”, pp. 219–227 in *First international congress of Chinese mathematicians* (Beijing, 1998), AMS/IP Stud. Adv. Math. **20**, Amer. Math. Soc., Providence, RI, 2001. MR 2002c:20011 Zbl 1064.20011

Communicated by Susan Montgomery

Received 2010-08-19

Revised 2011-02-28

Accepted 2011-04-10

cgoff@pacific.edu

Mathematics Department, University of the Pacific,
3601 Pacific Avenue, Stockton, CA 95211, United States

Uniformly rigid spaces

Christian Kappen

We define a new category of nonarchimedean analytic spaces over a complete discretely valued field, which we call *uniformly rigid*. It extends the category of rigid spaces, and it can be described in terms of bounded functions on products of open and closed polydiscs. We relate uniformly rigid spaces to their associated classical rigid spaces, and we transfer various constructions and results from rigid geometry to the uniformly rigid setting. In particular, we prove an analog of Kiehl's patching theorem for coherent ideals, and we define the uniformly rigid generic fiber of a formal scheme of formally finite type. This uniformly rigid generic fiber is more intimately linked to its model than the classical rigid generic fiber obtained via Berthelot's construction.

1. Introduction	341
2. Uniformly rigid spaces	346
3. Coherent modules on uniformly rigid spaces	375
4. Comparison with the theories of Berkovich and Huber	383
Acknowledgements	386
References	386

1. Introduction

Let K be a nonarchimedean field, and let R be its valuation ring, equipped with the valuation topology. Grothendieck had suggested that rigid spaces over K should be viewed as generic fibers of formal schemes of *topologically finite* (tf) type over R , that is, of formal schemes which are locally isomorphic to formal spectra of quotients of strictly convergent power series rings in finitely many variables

$$R\langle T_1, \dots, T_n \rangle.$$

He envisaged that rigid spaces should, in a suitable sense, be obtained from these formal schemes by tensoring over R with K . In accordance with this point of view,

MSC2010: primary 14G22; secondary 14K15.

Keywords: semiaffinoid, uniformly rigid, formally finite type, rigid geometry, formal geometry, Berthelot construction.

there is a generic fiber functor

$$\text{rig} : \left(\begin{array}{l} \text{formal } R\text{-schemes} \\ \text{of locally ff type} \end{array} \right) \rightarrow (\text{rigid } K\text{-spaces})$$

characterized by the property that it maps affine objects to affinoid spaces such that, on the level of functions, it corresponds to the extension of scalars functor $\cdot \otimes_R K$. This functor was more closely studied first by Raynaud and later by Bosch and Lütkebohmert; they proved that it induces an equivalence between the category of quasiparacompact and quasiseparated rigid K -spaces and the category of quasiparacompact admissible formal R -schemes, localized with respect to the class of admissible blowups [Raynaud 1974; Bosch and Lütkebohmert 1993a; Bosch 2005, Theorem 2.8/3].

From now on, let us assume that the absolute value on K is discrete, so that R is noetherian. Berthelot has extended the generic fiber functor to the class of formal R -schemes of locally *formally finite* (ff) type, which are locally isomorphic to formal spectra of topological quotients of mixed formal power series rings in finitely many variables

$$R[[S_1, \dots, S_m]]\langle T_1, \dots, T_n \rangle,$$

where an ideal of definition is generated by the maximal ideal of R and by the S_i ; see [Rapoport and Zink 1996, Section 5.5; Berthelot 1996, 0.2; de Jong 1995, 7.1–7.2]. This extension of rig is characterized by the property that it maps admissible blowups to isomorphisms, where a blowup is called admissible if it is defined by an ideal that locally contains a power of a uniformizer of R ; see [Temkin 2008, 2.1]. The extended rig functor no longer maps affine formal schemes to affinoid spaces; for example, the generic fiber of the affine formal R -scheme $\text{Spf } R[[S]]$ is the open rigid unit disc over K , which is not quasicompact.

While Raynaud's generic fiber functor is precisely described in terms of admissible blowups, Berthelot's extended generic fiber functor is less accessible: for example, let us consider an unbounded function f on the open rigid unit disc \mathbb{D}_K^1 . The resulting morphism φ from \mathbb{D}_K^1 to the rigid projective line does not extend to models of ff type; indeed, the domain of a model of φ cannot be quasicompact, for otherwise f would be bounded. In particular, there exists no admissible blowup of $\text{Spf } R[[S]]$ admitting an extension of φ , and the schematic closure of the graph of φ in the fibered product of $\text{Spf } R[[S]]$ and \mathbb{P}_R^1 over $\text{Spf } R$ does not exist. This phenomenon presents a serious obstacle if one tries for example to develop a theory of Néron models of ff type.

The main object of this article is to present a new category of nonarchimedean analytic spaces, the category of *uniformly rigid spaces*, which are better adapted to formal schemes of locally ff type than Tate's rigid analytic spaces. Intuitively

speaking, uniformly rigid spaces and their morphisms are described in terms of *bounded* functions on finite products of open and closed unit discs. Like rigid K -spaces, uniformly rigid K -spaces are locally ringed G -topological K -spaces, where the letter G indicates that the underlying set of physical points is not equipped with a topology, but with a Grothendieck topology. Let us give a brief overview of our definitions and results.

We say that a K -algebra is *semiaffinoid* if it is obtained from an R -algebra of ff type via the extension of scalars functor $\cdot \otimes_R K$. In other words, semiaffinoid K -algebras are quotients of K -algebras of the form

$$(R\llbracket S_1, \dots, S_m \rrbracket \langle T_1, \dots, T_n \rangle) \otimes_R K.$$

We define the category of semiaffinoid K -spaces as the opposite of the category of semiaffinoid K -algebras, where a morphism of semiaffinoid K -algebras is simply a K -algebra homomorphism. Semiaffinoid K -spaces play the role of “building blocks” for uniformly rigid K -spaces, such that we effectively implement Grothendieck’s original point of view in the ff type situation. Semiaffinoid K -algebras can be studied via the universal properties of the free semiaffinoid K -algebras, which we establish in Theorem 2.13.

We define a G -topology on the category of semiaffinoid K -spaces equipped with its physical points functor by considering compositions of admissible blowups, completion morphisms and open immersions on flat affine models of ff type; see Definitions 2.22 and 2.31. These formal-geometric constructions define semiaffinoid subdomains, which may be regarded as nested rational subdomains involving strict or nonstrict inequalities in semiaffinoid functions. In contrast to the classical rigid case, we cannot avoid nested constructions; this is essentially due to the fact that admissible blowups defined on open formal subschemes need not extend; see Remark 2.23. Just like in rigid geometry, the disconnected covering of the closed semiaffinoid unit disc $\text{sSp } K \langle S \rangle$ by the open semiaffinoid unit disc $\text{sSp } K \llbracket S \rrbracket$ and the semiaffinoid unit circle $\text{sSp } K \langle S, S^{-1} \rangle$ is *not* admissible in the uniformly rigid G -topology; see Example 2.42. In particular, contrary to the rigid-analytic situation, finite coverings of semiaffinoid spaces by semiaffinoid subdomains need not be admissible.

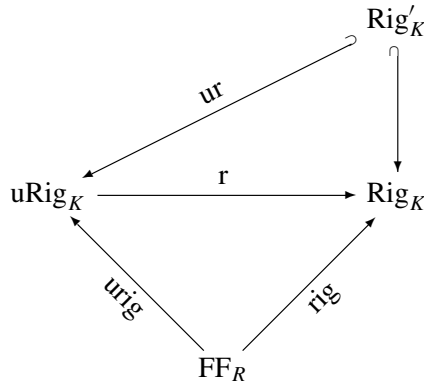
Using methods from formal geometry, we prove a uniformly rigid acyclicity theorem, which in particular implies the following:

Theorem 1.1 (2.41). *The presheaf of semiaffinoid functions is a sheaf for the uniformly rigid G -topology.*

The resulting functor from the category of semiaffinoid K -spaces to the category of locally G -ringed K -spaces is fully faithful; hence global uniformly rigid K -spaces can be defined; see Definition 2.46. They can be constructed by means of

standard gluing techniques; this is possible because uniformly rigid spaces satisfy the properties (G_0) – (G_2) listed in [Bosch et al. 1984, p. 339]. It follows that the category of uniformly rigid K -spaces admits fibered products and that there is a natural generic fiber functor urig from the category of formal R -schemes of locally ff type to the category of uniformly rigid K -spaces. The final picture can be described as follows:

Theorem 1.2 (Section 2D1). *Let Rig_K , uRig_K and FF_R denote the categories of rigid K -spaces, of uniformly rigid K -spaces and of formal R -schemes of locally ff type respectively. Let moreover $\text{Rig}'_K \subseteq \text{Rig}_K$ be the full subcategory of rigid spaces that are quasiparacompact and quasiseparated. There is a diagram of functors*



commuting up to isomorphism, where

- (i) the functor r is defined by applying the functor rig locally to models of ff type, where
- (ii) the functor ur is defined by applying urig to a global Raynaud-type model of locally tf type

and where the following holds:

- (i) The functor ur is a full embedding.
- (ii) The functor r is faithful, yet not fully faithful.
- (iii) For each $X \in \text{uRig}_K$, there is a comparison morphism $\text{comp}_X : X^r \rightarrow X$ that is final among all morphisms of locally G -ringed K -spaces from rigid K -spaces to X ; it is a bijection on physical points, and it induces isomorphisms of stalks.

For $X \in \text{uRig}_K$, we say that X^r is the underlying rigid space of X . Conversely, for $Y \in \text{Rig}'_K$ we say that Y^{ur} is the Raynaud-type uniformly rigid structure on Y . Via the comparison morphisms, uniformly rigid spaces and their underlying rigid spaces are locally indistinguishable; we may thus view a uniformly rigid space as a rigid space equipped with an additional global uniform structure which is encoded in terms of a coarser G -topology and a smaller sheaf of analytic functions. Let us

point out that the open rigid unit disc carries two canonical uniform structures, the one given by a Raynaud model of locally ff type and the one given by the canonical affine model $\mathrm{Spf} R[[S]]$ of ff type. The corresponding uniformly rigid spaces are distinct, since one is not quasicompact while the other one is quasicompact. The fact that r is not fully faithful is seen by the example of an *unbounded* function f on the rigid open unit disc which we considered above: The rigid-analytic morphism φ defined by f does not extend to a morphism of uniformly rigid spaces from $(\mathrm{Spf} R[[S]])^{\mathrm{urig}}$ to $(\mathbb{P}_K^{1,\mathrm{an}})^{\mathrm{ur}}$.

In Section 3, we study coherent modules on uniformly rigid K -spaces. We prove the existence of schematic closures of coherent submodules; see Theorem 3.5. Using the resulting models of coherent ideals, we prove the following analog of Kiehl's theorem A in rigid geometry [Kiehl 1967]:

Theorem 1.3 (3.6). *Coherent ideals on semiaffinoid spaces are associated to their ideals of global functions.*

In particular, closed uniformly rigid subspaces are well-behaved; see Proposition 3.11. Using fibered products and closed uniformly rigid subspaces, we define the notion of separatedness for uniformly rigid K -spaces, and we define the graph of a morphism $f : Y \rightarrow X$ of uniformly rigid K -spaces whose target is separated; see Section 3A1. Using Theorem 3.5, we show that if \mathfrak{X} and \mathfrak{Y} are flat formal R -schemes of locally ff type such that $\mathfrak{X}^{\mathrm{urig}}$ is separated and if $f : \mathfrak{Y}^{\mathrm{urig}} \rightarrow \mathfrak{X}^{\mathrm{urig}}$ is a morphism of uniformly rigid generic fibers, then the schematic closure of the graph of f in $\mathfrak{Y} \times \mathfrak{X}$ exists. As we have noted above, the corresponding statement is false if urig is replaced by Berthelot's generic fiber functor rig .

Semiaffinoid algebras and some associated locally G -ringed K -spaces have already been studied in [Lipshitz and Robinson 2000], where the terminology *quasi-affinoid* is used. The approach in that book includes the situation where R is not discrete and where the machinery of locally noetherian formal geometry is not available. However, no global theory is developed there, and the connection to formal geometry is not discussed. The proof of Theorem 2.13 in the case of a possibly nondiscrete valuation, given in [Lipshitz and Robinson 2000, I.5.2.3], is technically more involved, and it relies upon methods different from ours. The definition of the G -topology in [ibid., III.2.3.2] is less explicit than our definition, so that a deep quantifier elimination theorem [ibid., II, Theorem 4.2] is needed in order to prove an acyclicity theorem. Our approach avoids quantifier elimination.

It is unclear how to reflect uniformly rigid structures on the level of Berkovich's analytic spaces or on the level of Huber's analytic adic spaces; see Section 4. Semiaffinoid K -algebras are equipped with unique K -Banach algebra structures, so that one may consider their valuation spectra. For instance, the spectrum $M(R[[S]] \otimes_R K)$ is the closure of the Berkovich open unit disc within the Berkovich closed unit

disc; it is obtained by adding the Gauss point. However, inclusions of semiaffinoid subdomains need not induce injective maps of valuation spectra, so the formation of the valuation spectrum does not globalize. This corresponds to the fact that in the ff type situation, the functor $\cdot \otimes_R K$ does not commute with complete localization. Nonetheless, we suggest that a uniformly rigid K -space X should be viewed as a compactification of its underlying rigid space X^r . This point of view might be useful in order to obtain a better understanding of the quasicompactifications considered in [Strauch 2008, 3.1] and in [Huber 2007, 3]; it should be further developed within the framework of topos theory. We propose the study of the uniformly rigid topos as a topic for future research.

The author has used uniformly rigid spaces in his doctoral thesis [Kappen 2009], in order to lay the foundations for a theory of formal Néron models of locally ff type. The search for such a theory was strongly motivated by work of C.-L. Chai [2003], who had suggested that Néron models of ff type could be used to study the base change conductor of an abelian variety with potentially multiplicative reduction over a local field. Chai and the author are currently working on further developing the methods of [Chai 2003] within the framework of uniformly rigid spaces.

2. Uniformly rigid spaces

Let R be a discrete valuation ring with residue field k and fraction field K , and let $\pi \in R$ be a uniformizer.

2A. Formal schemes of formally finite type. A morphism of locally noetherian formal schemes is said to be of locally *formally finite* (ff) type if the induced morphism of smallest subschemes of definition is of locally finite type. Equivalently, any induced morphism of subschemes of definition is of locally finite type. A morphism of locally noetherian formal schemes is called of ff type if it is of locally ff type and quasicompact. If A is a noetherian adic ring and if B is a noetherian adic topological A -algebra, then $\mathrm{Spf} B$ is of ff type over $\mathrm{Spf} A$ if and only if B is a topological quotient of a mixed formal power series ring $A[[S_1, \dots, S_m]]\langle T_1, \dots, T_n \rangle$, where $A[[S_1, \dots, S_m]]$ carries the $\mathfrak{a} + (S_1, \dots, S_m)$ -adic topology for any ideal of definition \mathfrak{a} of A [Berkovich 1996, Lemma 1.2]. In this case, we say that the topological A -algebra B is of ff type. Morphisms of locally ff type are preserved under composition, base change and formal completion.

We say that an R -algebra is of *formally finite* (ff) type if it admits a ring topology such that it becomes a topological R -algebra of ff type in the above sense, where R carries the π -adic topology. Equivalently, an R -algebra is of ff type if it admits a presentation as a quotient of a mixed formal power series ring, as above. If S

and T are finite systems of variables and if $\varphi : R[[S]]\langle T \rangle \rightarrow A$ is a surjection, then the φ -image of (S, T) will be called a formal generating system for A .

Lemma 2.1. *If A is a topological R -algebra of ff type, then the biggest ideal of definition of A coincides with the Jacobson radical of A . Moreover, any R -homomorphism of topological R -algebras of ff type is continuous.*

Proof. Let \mathfrak{a} denote the biggest ideal of definition of A ; then \mathfrak{a} is contained in every maximal ideal of A since A is \mathfrak{a} -adically complete. On the other hand, A/\mathfrak{a} is a Jacobson ring since it is of finite type over the residue field k of R ; it follows that \mathfrak{a} coincides with the Jacobson radical of A , as claimed. In particular, the topology on A is determined by the ring structure of A . Let now $A \rightarrow B$ be a homomorphism of R -algebras of ff type; by what we have seen so far, it suffices to see that φ is continuous for the Jacobson-adic topologies. However, for any maximal ideal $\mathfrak{n} \subseteq B$, the preimage $\mathfrak{m} := \mathfrak{n} \cap A$ of \mathfrak{n} in A is maximal, since $k \subseteq A/\mathfrak{m} \subseteq B/\mathfrak{n}$, where B/\mathfrak{n} is a finite field extension of k because the quotient $B/\text{jac } B$ is of finite type over k . □

In particular, the topology on A can be recovered from the ring structure on A , and the category of R -algebras of ff type is canonically equivalent to the category of topological R -algebras of ff type. Lemma 2.1 implies that the category of R -algebras of ff type admits amalgamated sums $\hat{\otimes}$.

2B. Semiaffinoid algebras. We define semiaffinoid K -algebras as the generic fibers of R -algebras of ff type, and we define the category of semiaffinoid K -spaces as the dual of the category of semiaffinoid K -algebras:

Definition 2.2. Let A be a K -algebra.

- (i) An R -model of A is an R -subalgebra $\underline{A} \subseteq A$ such that the natural homomorphism $\underline{A} \otimes_R K \rightarrow A$ is an isomorphism.
- (ii) The K -algebra A is called *semiaffinoid* if it admits an R -model of ff type.
- (iii) A homomorphism of semiaffinoid K -algebras is a homomorphism of underlying K -algebras.
- (iv) The category of semiaffinoid K -spaces is the dual of the category of semiaffinoid K -algebras. If A is a semiaffinoid K -algebra, we write $\text{sSp } A$ to denote the corresponding semiaffinoid K -space, and if $\varphi : \text{sSp } B \rightarrow \text{sSp } A$ is a morphism of semiaffinoid K -spaces, we write φ^* to denote the corresponding K -algebra homomorphism.

By Definition 2.2(i) above, any R -model of a K -algebra is flat over R .

There exists no general analog of the Noether normalization theorem for semiaffinoid K -algebras [Lipshitz and Robinson 2000, I.2.3.5]. However, if A is a semiaffinoid K -algebra admitting a local R -model of ff type, then there exist finitely

many variables S_1, \dots, S_m and a finite K -monomorphism

$$R[[S_1, \dots, S_m]] \otimes_R K \hookrightarrow A.$$

Indeed, if \underline{A} is a local R -model of ff type for A with maximal ideal \mathfrak{m} and if s_0, \dots, s_m is a system of parameters for \underline{A} such that $s_0 = \pi$, then there exists a unique continuous R -homomorphism $\varphi: R[[S_1, \dots, S_m]] \rightarrow \underline{A}$ sending S_i to s_i , for $1 \leq i \leq m$. If \mathfrak{r} denotes the maximal ideal of $R[[S_1, \dots, S_m]]$, then $\underline{A}/\mathfrak{r}\underline{A}$ is k -finite because $\underline{A}/\mathfrak{m}\underline{A}$ is k -finite and because \mathfrak{r} is \mathfrak{m} -primary. By the formal version of Nakayama’s Lemma, cf [Eisenbud 1995, Example 7.2], it follows that φ is finite; here we use that $R[[S_1, \dots, S_m]]$ is \mathfrak{r} -adically complete and that \underline{A} is \mathfrak{r} -adically separated. Since $R[[S_1, \dots, S_m]]$ and \underline{A} have the same dimension, it follows that φ is finite, so we obtain the desired finite monomorphism by extending scalars from R to K .

2B1. The specialization map. The following statement may be compared with [de Jong 1995, 7.1.9]:

Lemma 2.3. *Let A be a semiaffinoid K -algebra, and let $\underline{A} \subseteq A$ be an R -model of ff type. If \mathfrak{m} is a maximal ideal in A , then*

$$\text{sp}_{\underline{A}}(\mathfrak{m}) := \sqrt{(\underline{A} \cap \mathfrak{m}) + \pi \underline{A}}$$

is a maximal ideal in \underline{A} , and A/\mathfrak{m} is a finite extension of K .

Proof. Let us write $\mathfrak{p} := \mathfrak{m} \cap \underline{A}$; then $(\underline{A}/\mathfrak{p})_{\pi} = A/\mathfrak{m}$ is a field, and by the Artin–Tate theorem [Grothendieck 1964, 0.16.3.3] it follows that $\underline{A}/\mathfrak{p}$ is a semilocal ring of dimension ≤ 1 . Moreover, $\underline{A}/\mathfrak{p}$ is of ff type over R and, hence, π -adically complete. Since $\underline{A}/\mathfrak{p} \subseteq A/\mathfrak{m}$ is R -flat and since $(\underline{A}/\mathfrak{p})_{\pi}$ is local, it thus follows from Hensel’s Lemma that $(\underline{A}/\mathfrak{p})/\pi(\underline{A}/\mathfrak{p})$ is local as well [Bourbaki 1998, III.4.6 Proposition 8]. Since $\mathfrak{p}A = \mathfrak{m}$, the class of π in $\underline{A}/\mathfrak{p}$ is nonzero, and so the local noetherian ring $(\underline{A}/\mathfrak{p})/\pi(\underline{A}/\mathfrak{p})$ is zero-dimensional. Thus its quotient modulo its nilradical is a field, and it follows that the radical of $\mathfrak{p} + \pi \underline{A}$ is maximal in \underline{A} , as desired.

To prove that A/\mathfrak{m} is finite over K , it suffices to show that $\underline{A}/\mathfrak{p}$ is finite over R . Since R is π -adically complete and since $\underline{A}/\mathfrak{p}$ is π -adically separated, it thus suffices to show that $\underline{A}/(\mathfrak{p} + \pi \underline{A})$ is finite over k [Eisenbud 1995, Example 7.2]. The ring $\underline{A}/(\mathfrak{p} + \pi \underline{A})$ is noetherian; hence its nilradical is nilpotent, and it thereby suffices to see that the quotient of \underline{A} modulo the maximal ideal $\sqrt{\mathfrak{p} + \pi \underline{A}}$ is k -finite. Since \underline{A} is of ff type over R , since maximal ideals are open and since field extensions of finite type are finite, the desired statement follows. \square

Definition 2.4. If A is a semiaffinoid K -algebra, we call $|X| := \text{Max } A$ the set of physical points of its corresponding semiaffinoid K -space X . We will often write X instead of $|X|$ if no confusion is likely to result.

Remark 2.5. Lemma 2.3 implies that a morphism $\varphi : \text{sSp } A \rightarrow \text{sSp } B$ induces a map on sets of physical points such that for R -models of ff type \underline{A} and \underline{B} with $\varphi^*(B) \subseteq \underline{A}$, the specialization maps $\text{sp}_{\underline{A}}$ and $\text{sp}_{\underline{B}}$ are compatible with respect to φ and the induced morphism $\underline{\varphi} : \text{Spf } \underline{A} \rightarrow \text{Spf } \underline{B}$. This functoriality implies that $\text{sp}_{\underline{A}}$ is surjective onto the set of maximal ideals in \underline{A} . Indeed, let $\mathfrak{r} \subseteq \underline{A}$ be a maximal ideal, and let $\underline{A}|_{\mathfrak{r}}$ denote the \mathfrak{r} -adic completion of A ; then $\text{Max}(\underline{A}|_{\mathfrak{r}} \otimes_R K)$ is nonempty, and any element in this set maps to an element in $\text{Max}(A)$ that maps to \mathfrak{r} under $\text{sp}_{\underline{A}}$. Let us moreover remark that for $x \in X = \text{sSp } A$ with specialization $\mathfrak{n} \subseteq \underline{A}$, the valuation ring of the residue field of A in x coincides with the integral closure of $\underline{A}_{\mathfrak{n}}$ in that residue field, so that the intersection of $\underline{A}_{\mathfrak{n}}$ with the valuation ideal is precisely $\mathfrak{n}\underline{A}_{\mathfrak{n}}$.

2B2. Power-boundedness and topological quasiniipotency. Let X be a semi-affinoid K -space with corresponding semiaffinoid K -algebra A . By Lemma 2.3, A/\mathfrak{m} is K -finite for $\mathfrak{m} \subseteq A$ maximal; hence the discrete valuation on K extends uniquely to A/\mathfrak{m} , so we can define $|f(x)| \in \mathbb{R}_{\geq 0}$ for any $f \in A$, $x \in X$.

Definition 2.6. An element $f \in A$ is called *power-bounded* if $|f(x)| \leq 1$ for all $x \in X$. It is called *topologically quasiniipotent* if $|f(x)| < 1$ for all $x \in X$. We let $\mathring{A} \subseteq A$ denote the R -subalgebra of power-bounded functions, and we let $\check{A} \subseteq \mathring{A}$ denote the ideal of topologically quasiniipotent functions.

For example, $S \in A = R[[S]] \otimes_R K$ is topologically quasiniipotent, while the supremum of the absolute values $|S(x)|$, with x ranging over X , is equal to 1. Thus we see that the classical maximum principle fails for semiaffinoid K -algebras. However, the maximum principle holds if we let x vary in the Berkovich spectrum $M(A)$ of A , where A is equipped with its unique K -Banach algebra topology; see Section 4. Indeed, this follows trivially from the fact that $M(A)$ is compact.

Remark 2.7. If A is a nonreduced semiaffinoid K -algebra, then \mathring{A} cannot be of ff type over R : If $f \in A$ is a nonzero nilpotent function, then $f \in \mathring{A}$ is infinitely π -divisible in \mathring{A} , but R -algebras of ff type are π -adically separated.

Remark 2.8. If $\underline{A} \subseteq A$ is an R -model of ff type, then $\underline{A} \subseteq \mathring{A}$, and $\check{A} \cap \underline{A} \subseteq \underline{A}$ is the biggest ideal of definition. Indeed, by Lemma 2.1 and its proof, the biggest ideal of definition of \underline{A} is given by the Jacobson radical, and hence it suffices to observe that for any $f \in \underline{A}$ and any $x \in \text{sSp } A$ with specialization $\mathfrak{n} \subseteq \underline{A}$, we have $|f(x)| \leq 1$, where $|f(x)| < 1$ if and only if $f \in \mathfrak{n}$. This however is clear from the final statement in Remark 2.5.

For the notion of *normality* for formal R -schemes of locally ff type, we refer to the discussion in [Conrad 1999, 1.2], which is based on the fact that R -algebras of ff type are *excellent*. This excellence result is a consequence of [Valabrega 1975, Proposition 7] if R has equal characteristic, and it follows from [Valabrega

1976, Theorem 9] if R has mixed characteristic. In the following, excellence of R -algebras of ff type will be used without further comments.

The following result is fundamental:

Proposition 2.9. *Let A be a semiaffinoid K -algebra. If A admits a normal R -model of ff type, then this model coincides with \mathring{A} .*

Proof. Let \underline{A} be a normal R -model of ff type for A . By [de Jong 1995, 7.1.8], we may view A as a subring of the ring of global functions on $(\mathrm{Spf} \underline{A})^{\mathrm{rig}}$, and by [de Jong 1995, 7.4.1; 1998], \underline{A} coincides with the ring of power-bounded global functions under this identification. \square

Corollary 2.10. *Let A be a semiaffinoid K -algebra, and let $\underline{A} \subseteq A$ be an R -model of ff type; then the inclusion $\underline{A} \subseteq \mathring{A}$ is integral. If moreover A is reduced, then this inclusion is finite.*

Proof. Let $\varphi: \underline{A} \rightarrow \underline{B}$ denote the normalization of \underline{A} . Then φ is finite since \underline{A} is excellent, and hence \underline{B} is of ff type over R . Extension of scalars yields an induced homomorphism of semiaffinoid K -algebras $\varphi: A \rightarrow B$. Since φ factors through an injective R -homomorphism $\underline{A}/\mathrm{rad}(\underline{A}) \hookrightarrow \underline{B}$, since K is R -flat and since $\mathrm{rad}(A) = \mathrm{rad}(\underline{A})A$, we see that φ factors through an injective K -homomorphism $A/\mathrm{rad}(A) \hookrightarrow B$. By Proposition 2.9, \underline{B} coincides with the ring of power-bounded functions in B . Let us consider a power-bounded function f in A ; then $\varphi(f) \in \underline{B}$. Since φ is finite, there exists an integral equation $P(T) \in \underline{A}[T]$ for $\varphi(f)$ over \underline{A} . By the factorization of φ mentioned above, we conclude that $P(f) \in A$ is nilpotent. If $s \in \mathbb{N}$ is an integer such that $P(f)^s = 0$; then $P(T)^s \in \underline{A}[T]$ is an integral equation for f over \underline{A} . Finally, if A is reduced, then φ is injective, and hence \mathring{A} is an \underline{A} -submodule of the finite \underline{A} -module \underline{B} . Since \underline{A} is noetherian, it follows that \mathring{A} is a finite \underline{A} -module. \square

We immediately obtain the following:

Corollary 2.11. *The ring of power-bounded functions in a reduced semiaffinoid K -algebra is a canonical R -model of ff type containing any other R -model of ff type.*

We conclude that any R -model of ff type can be enlarged so that it contains any given finite set of power-bounded functions:

Corollary 2.12. *Let \underline{A} be an R -model of ff type in a semiaffinoid K -algebra A , and let $M \subseteq A$ be a finite set of power-bounded functions. Then the \underline{A} -subalgebra $\underline{A}[M]$ generated by M over \underline{A} is finite over \underline{A} and, hence, an R -model of ff type for A .*

Proof. The ring extension $\underline{A} \subseteq \underline{A}[M]$ is finite since it is generated by finitely many integral elements. \square

2B3. Free semiaffinoid algebras. Using the results of Section 2B2, we can now establish the universal properties of free semiaffinoid K -algebras; these are semiaffinoid K -algebras of the form $R[[S]]\langle T \rangle \otimes_R K$, for finite systems of variables S and T :

Theorem 2.13. *Let m and n be natural numbers. The semiaffinoid K -algebra $R[[S_1, \dots, S_m]]\langle T_1, \dots, T_n \rangle \otimes_R K$, together with the pair of tuples of functions $((S_1, \dots, S_m), (T_1, \dots, T_n))$, is initial among all semiaffinoid K -algebras A equipped with a pair $((f_1, \dots, f_m), (g_1, \dots, g_n))$ satisfying the property that the g_j are power-bounded and that the f_i are topologically quasিনিপotent.*

Proof. Let us write S and T to denote the systems of the S_i and the T_j . By Corollary 2.12, A admits an R -model of ff type \underline{A} containing the f_i and the g_j . By Remark 2.8, the f_i are topologically nilpotent in \underline{A} ; hence there exists a unique R -homomorphism $\varphi: R[[S]]\langle T \rangle \rightarrow \underline{A}$ sending S_i to f_i and T_j to g_j for all i and j , and so $\varphi := \varphi \otimes_R K$ is a K -homomorphism with the desired properties. It remains to show that these properties determine φ uniquely. Let $\varphi': R[[S]]\langle T \rangle \otimes_R K \rightarrow A$ be any K -homomorphism sending S_i to f_i and T_j to g_j for all i and j , and let us set $\underline{A}' := \varphi'(R[[S]]\langle T \rangle)$ which is of ff type over R . If $\varphi = \varphi'$, then $\underline{A}' \subseteq \underline{A}$. On the other hand, to show that $\varphi = \varphi'$, it suffices to see that, after possibly enlarging \underline{A} , we have $\underline{A}' \subseteq \underline{A}$, in virtue of the universal property of $R[[S]]\langle T \rangle$. If A is reduced, Corollary 2.10 says that we may set \underline{A} equal to the ring of power-bounded functions in A ; in this case the inclusion $\underline{A}' \subseteq \underline{A}$ is obvious. In the general case, we let N denote the nilradical of A ; then, by what we have shown so far,

$$\underline{A}' / (\underline{A}' \cap N) \subseteq \underline{A} / (\underline{A} \cap N) \tag{*}$$

within \mathring{A}/N . The ideal $\underline{A}' \cap N$ is finitely generated since \underline{A}' is noetherian; after enlarging \underline{A} using Corollary 2.12, we may thus assume that \underline{A} contains a generating system n_1, \dots, n_r of $\underline{A}' \cap N$. The inclusion (*) shows that every element $a' \in \underline{A}'$ is the sum of an element $a \in \underline{A}$ and a linear combination $\sum_{i=1}^r a'_i n_i$ with coefficients $a'_i \in \underline{A}'$. Let us write the coefficients a'_i in the analogous way, and let us iterate the procedure. Using the fact that the n_i lie in \underline{A} , the only summands possibly not lying in \underline{A} after s -fold iteration are multiples of products of the n_i involving s factors. Since the n_i are nilpotent, these summands are zero for s big enough; hence $\underline{A}' \subseteq \underline{A}$, as desired. \square

With the universal property of the free semiaffinoid K -algebras at hand, we can now describe the category of semiaffinoid K -algebras in terms of the category of R -models of ff type. Let us recall that a formal blowup in the sense of [Temkin 2008, 2.1] is called *admissible* if it can be defined by a π -adically open coherent ideal.

Corollary 2.14. *Let $\varphi: A \rightarrow B$ be a homomorphism of semiaffinoid K -algebras.*

- (i) Let $\underline{A}_1, \underline{A}_2$ be R -models of ff type for A . If \underline{A}_2 contains a formal generating system of \underline{A}_1 , then \underline{A}_1 is contained in \underline{A}_2 .
- (ii) An inclusion of R -models of ff type for A corresponds to a finite admissible blowup of associated formal spectra.
- (iii) Let $\underline{A} \subseteq A, \underline{B} \subseteq B$ be R -models of ff type such that there exists a formal generating system of \underline{A} mapping to \underline{B} via φ . Then $\varphi(\underline{A}) \subseteq \underline{B}$.
- (iv) Let \underline{A} be an R -model of ff type for A . There exists an R -model of ff type \underline{B} for B such that $\varphi(\underline{A}) \subseteq \underline{B}$. Moreover, if \underline{B}' is any R -model of ff type for B , we can choose \underline{B} such that $\underline{B}' \subseteq \underline{B}$.

Proof. To prove the first statement, let us fix a formal generating system (f, g) of \underline{A}_1 that is contained in \underline{A}_2 . The components of f are topologically quasinilpotent in A ; since \underline{A}_2 is an R -model of ff type for A , they are topologically nilpotent in \underline{A}_2 . Let $\alpha : R[[S]]\langle T \rangle \rightarrow \underline{A}_1$ and $\beta : R[[S]]\langle T \rangle \rightarrow \underline{A}_2$ be the associated R -homomorphisms, where α is surjective because (f, g) formally generates \underline{A}_1 . By Theorem 2.13, $\alpha \otimes_R K$ and $\beta \otimes_R K$ coincide as homomorphisms from $R[[S]]\langle T \rangle \otimes_R K$ to A , so we conclude that $\underline{A}_1 \subseteq \underline{A}_2$: given $a \in \underline{A}_1$, we choose an α -preimage a' of a ; then $a = \alpha(a') = \beta(a') \in \underline{A}_2$.

To prove the second claim, let $\underline{A}_1 \subseteq \underline{A}_2$ be an inclusion of R -models of ff type for A , and let $M \subseteq \underline{A}_2$ be a finite set whose elements are the components of a formal generating system for \underline{A}_2 over R . Then by Corollary 2.12, $\underline{A}_1[M] \subseteq \underline{A}_2$ is an R -model of ff type for A which is finite over \underline{A}_1 . By statement (i), $\underline{A}_2 = \underline{A}_1[M]$ and hence \underline{A}_2 is finite over \underline{A}_1 . Arguing exactly as in the proof of [Bosch and Lütkebohmert 1993a, 4.5], we see that $\underline{A}_1 \subseteq \underline{A}_2$ corresponds to an admissible formal blowup.

To prove part (iii), let us choose a formal generating system (f, g) of \underline{A} such that the components of $\varphi(f)$ and $\varphi(g)$ are contained in \underline{B} . The components of $\varphi(f)$ are topologically nilpotent in \underline{B} since they are topologically quasinilpotent in B . Let $\alpha : R[[S]]\langle T \rangle \rightarrow \underline{A}$ and $\beta : R[[S]]\langle T \rangle \rightarrow \underline{B}$ be the R -homomorphisms defined by (f, g) and $(\varphi(f), \varphi(g))$ respectively; then α is surjective, and Theorem 2.13 shows that $\beta \otimes_R K$ coincides with $\varphi \circ (\alpha \otimes_R K)$. As is the proof of statement (i), we conclude that $\varphi(\underline{A}) \subseteq \underline{B}$.

To prove statement (iv), let us choose a formal generating system (f, g) of \underline{A} . The components of $\varphi(f)$ are topologically quasinilpotent, and the components of $\varphi(g)$ are power-bounded in B . According to Corollary 2.12, there exists an R -model \underline{B} of ff type for B containing \underline{B}' and the components of $\varphi(f)$ and $\varphi(g)$; by statement (iii), $\varphi(\underline{A}) \subseteq \underline{B}$, as desired. \square

We can now show that R -models of ff type for affinoid K -algebras are automatically of tf type:

Corollary 2.15. *Let A be an affinoid K -algebra, and let $\underline{A} \subseteq A$ be an R -model of ff type. Then \underline{A} is of tf type over R .*

Proof. Let \underline{A}' be an R -model of tf type for A , and let \underline{A}'' be an R -model of ff type for A containing both \underline{A} and \underline{A}' ; such an \underline{A}'' exists by Corollary 2.14(iv) applied to the identity on A . By Corollary 2.14(ii), \underline{A}'' is finite over \underline{A}' and, hence, an R -algebra of tf type. After replacing \underline{A}' by \underline{A}'' , we may thus assume that $\underline{A} \subseteq \underline{A}'$. Again by Corollary 2.14(ii), this inclusion is finite. We now mimic the proof of the classical Artin–Tate lemma: Let a_1, \dots, a_m be a system of topological generators of \underline{A}' over R , and for each i let $P_i \in \underline{A}[T]$ be an integral equation for a_i over \underline{A} . Let b_1, \dots, b_n be the coefficients of the P_i in some ordering. Since the R -algebra \underline{A} is of ff type, it is π -adically complete; hence there exists a unique R -homomorphism $R\langle T_1, \dots, T_n \rangle \rightarrow \underline{A}$ sending T_j to b_j for $1 \leq j \leq n$. Let $\underline{B} \subseteq \underline{A}$ denote its image; then \underline{B} is an R -algebra of tf type. Since the a_i topologically generate \underline{A}' over R , they also topologically generate \underline{A}' over \underline{B} . The a_i are, by construction, integral over \underline{B} ; hence \underline{A}' is in fact finite over \underline{B} . Since \underline{B} is noetherian, the \underline{B} -submodule \underline{A} of \underline{A}' is finite as well, and it follows that \underline{A} is of tf type as a \underline{B} -algebra. We conclude that \underline{A} is of tf type over R . \square

2B4. Amalgamated sums.

Proposition 2.16. *The category of semiaffinoid K -algebras admits amalgamated sums. More precisely speaking, if $\varphi_1: A \rightarrow B_1$ and $\varphi_2: A \rightarrow B_2$ are homomorphisms of semiaffinoid K -algebras, then the colimit of the resulting diagram is represented by $(\underline{B}_1 \hat{\otimes}_{\underline{A}} \underline{B}_2) \otimes_R K$, where \underline{A} and the \underline{B}_i are R -models of ff type for A and the B_i respectively such that $\varphi(\underline{A}) \subseteq \underline{B}_1, \underline{B}_2$.*

Proof. By Corollary 2.14(iv), we may choose R -models \underline{A} , \underline{B}_1 and \underline{B}_2 as in the statement of the proposition. Let C be a semiaffinoid K -algebra, and for $i = 1, 2$ let $\tau_i: B_i \rightarrow C$ be a K -homomorphism such that $\tau_1 \circ \varphi_1 = \tau_2 \circ \varphi_2$. By Corollary 2.14(iv), there exists an R -model \underline{C} of ff type for C such that $\tau_i(\underline{B}_i) \subseteq \underline{C}$ for $i = 1, 2$; we let $\underline{\tau}_i: \underline{B}_i \rightarrow \underline{C}$ denote the induced R -homomorphism. Then $\underline{\tau}_1 \circ \underline{\varphi}_1 = \underline{\tau}_2 \circ \underline{\varphi}_2$, since the same holds after inverting π and since π is not a zero divisor in \underline{A} . By the universal property of the complete tensor product in the category of R -algebras of ff type, there exists a unique R -homomorphism $\underline{\tau}: \underline{B}_1 \hat{\otimes}_{\underline{A}} \underline{B}_2 \rightarrow \underline{C}$ such that $\underline{\tau}_i = \underline{\tau} \circ \underline{\sigma}_i$ for $i = 1, 2$, where $\underline{\sigma}_i: \underline{B}_i \rightarrow \underline{B}_1 \hat{\otimes}_{\underline{A}} \underline{B}_2$ is the i th coprojection. Setting $\tau := \underline{\tau} \otimes_R K$ and $\sigma_i := \underline{\sigma}_i \otimes_R K$, we obtain $\tau_i = \tau \circ \sigma_i$ for $i = 1, 2$. We must show that τ is uniquely determined by this property. Let

$$\tau': (\underline{B}_1 \hat{\otimes}_{\underline{A}} \underline{B}_2) \otimes_R K \rightarrow C$$

be any K -homomorphism satisfying $\tau_i = \tau' \circ \sigma_i$ for $i = 1, 2$. By Corollary 2.14(iv), there exists an R -model \underline{C}' of ff type for C containing \underline{C} such that τ' restricts to an R -morphism $\underline{\tau}': \underline{B}_1 \hat{\otimes}_{\underline{A}} \underline{B}_2 \rightarrow \underline{C}'$; then $\tau' = \underline{\tau}' \otimes_R K$. It suffices to show that

τ' coincides with τ composed with the inclusion $\iota : C \subseteq C'$. For $i = 1, 2$, the compositions $\tau' \circ \sigma_i$ and $\iota \circ \tau \circ \sigma_i$ coincide after inverting π , hence they coincide because π is not a zero divisor in B_i , for $i = 1, 2$. The universal property of $(B_1 \hat{\otimes}_A B_2, \sigma_1, \sigma_2)$ implies that $\tau' = \iota \circ \tau$, as desired. \square

Passing to the opposite category, we see that the category of semiaffinoid K -spaces has fibered products.

2B5. *The Nullstellensatz.*

Proposition 2.17. *Semiaffinoid K -algebras are Jacobson rings.*

Proof. Any quotient of a semiaffinoid K -algebra is again semiaffinoid; hence it suffices to show that if A is a semiaffinoid K -algebra and if $f \in A$ is a semiaffinoid function such that $f(x) = 0$ for all $x \in \text{sSp } A$, then f is nilpotent. We may divide A by its nilradical and thereby assume that A is reduced. Let \underline{A} be an R -model of ff type for A , and let $X = (\text{Spf } \underline{A})^{\text{rig}}$ denote the rigid-analytic generic fiber of $\text{Spf } \underline{A}$. Since A is excellent, being a localization of the excellent ring \underline{A} , and since rigid K -spaces are excellent [Conrad 1999, 1.1], it follows from [de Jong 1995, Lemma 7.1.9] that the space X is reduced and that we may view A as a subring of $\Gamma(X, \mathbb{O}_X)$ such that the value of f in a point $x \in X$ agrees with the value of f in the corresponding maximal ideal of A . Since $f(x) = 0$ for all $x \in X$, we see that $f = 0$ as a function on X and, hence, in A . \square

2C. *Semiaffinoid spaces.*

2C1. *The rigid space associated to a semiaffinoid K -space.* Let $X = \text{sSp } A$ be a semiaffinoid K -space. An affine flat formal model of ff type for X is an affine flat formal R -scheme of ff type \mathfrak{X} together with an identification of $\Gamma(\mathfrak{X}, \mathbb{O}_{\mathfrak{X}})$ with an R -model of ff type for A . By Definition 2.2, every semiaffinoid K -space admits an affine flat model of ff type. There is an obvious generic fiber functor urig from the category of affine flat formal R -schemes of ff type to the category of semiaffinoid K -spaces, given by

$$(\text{Spf } \underline{A})^{\text{urig}} := \text{sSp}(A \otimes_R K).$$

Let \mathfrak{X} be a flat affine R -model of ff type for X . Berthelot’s construction yields a rigid K -space $X^r := \mathfrak{X}^{\text{rig}}$ together with a K -homomorphism

$$\varphi : A \rightarrow \Gamma(X^r, \mathbb{O}_{X^r});$$

see [de Jong 1995, 7.1.8]. By our discussion in Section 2B1 and by [ibid., 7.1.9] the homomorphism φ induces a bijection $|X^r| \rightarrow |X|$ and local homomorphisms $A_{\mathfrak{m}} \rightarrow \mathbb{O}_{X^r, x}$ which are isomorphisms on maximal-adic completions, where x is a point of X^r and where $\mathfrak{m} \in \text{Max } A$ is the image of x under the above bijection. We say that X^r is the *rigid space associated to X* via Berthelot’s construction.

It is independent of the choice of \mathfrak{X} , the pair (X^r, φ) being characterized by the following universal property:

Proposition 2.18. *Let Y be a rigid K -space, and let $\psi: A \rightarrow \Gamma(Y, \mathbb{O}_Y)$ be a K -algebra homomorphism. There exists a unique morphism of rigid K -spaces $\sigma: Y \rightarrow X^r$ such that $\psi = \Gamma(\sigma^\sharp) \circ \varphi$.*

Proof. Uniqueness of σ follows from the above-mentioned fact that φ induces a bijection of points and isomorphisms of completed stalks; we may thus assume that Y is affinoid, $Y = \text{Sp } B$. Let $\underline{A} \subseteq A$ be the R -model of ff type corresponding to \mathfrak{X} . By Corollary 2.14(iv) and Corollary 2.15, ψ restricts to an R -homomorphism $\underline{\psi}: \underline{A} \rightarrow \underline{B}$, where \underline{B} is a suitable R -model of tf type for B ; now $\sigma := (\text{Spf } \underline{\psi})^{\text{rig}}$ has the required properties. \square

If $\underline{\tau}: \mathfrak{Y} \rightarrow \mathfrak{X}$ is a morphism of affine flat formal R -schemes of ff type and if $\underline{\tau}^{\text{urig}}$ denotes the induced morphism of associated semiaffinoid K -spaces, we easily see that the unique morphism $(\underline{\tau}^{\text{urig}})^r$ provided by Proposition 2.18 is given by $\underline{\tau}^{\text{rig}}$.

2C2. Semiaffinoid subdomains. Closed subspaces of semiaffinoid K -spaces are easily defined in the usual way:

Definition 2.19. A morphism of semiaffinoid K -spaces is called a *closed immersion* if it corresponds to a surjective homomorphism of semiaffinoid K -algebras. A *closed semiaffinoid subspace* of a semiaffinoid K -space is an equivalence class of closed immersions, where two closed immersions of uniformly rigid K -spaces $i_1: Y_1 \rightarrow X$ and $i_2: Y_2 \rightarrow X$ are called equivalent if there exists an isomorphism $\varphi: Y_1 \xrightarrow{\sim} Y_2$ such that $i_1 = i_2 \circ \varphi$.

If A is a semiaffinoid K -algebra and if $I \subseteq A$ is an ideal, then the natural closed immersion $\text{sSp } A/I \rightarrow \text{sSp } A$ is clearly injective onto the set of maximal ideals containing I . Moreover, if $A \rightarrow C$ is a homomorphism of semiaffinoid K -algebras, then $A/I \hat{\otimes}_A C = C/IC$, because this quotient already represents the amalgamated sum of C and A/I over A in the category of all K -algebras. In particular, closed immersions of semiaffinoid K -spaces are stable under the formation of fibered products.

To define a reasonable structure of G -topological K -space on the set of physical points of a semiaffinoid K -space X , it is natural to consider subsets U of X that canonically inherit a structure of semiaffinoid K -space:

Definition 2.20. A subset U in a semiaffinoid K -space X is called *representable* if there exists a morphism of semiaffinoid K -spaces to X whose image lies in U and which is final with this property. Such a morphism is said to represent all semiaffinoid morphisms to X with image in U .

Remark 2.21. Here we differ from the terminology used in the author's PhD thesis; there the representable subsets are called *semiaffinoid presubdomains* [Kappen 2009, Section 1.3.3].

Clearly X and \emptyset are representable subsets of X . Copying the proof of [Bosch et al. 1984, 7.2.2/1], we see that a morphism representing a subset $U \subseteq X$ is injective with image U and that it induces isomorphisms of infinitesimal neighborhoods of points. Using the existence of fibered products in the category of semiaffinoid K -spaces, we see that representable subsets are preserved under pull-back with respect to morphisms of semiaffinoid spaces. The universal property of representable subsets yields a presheaf \mathbb{O}_X in semiaffinoid K -algebras on the category of representable subsets in X .

In the category of affinoid K -spaces, the representable subsets are called affinoid subdomains [Bosch et al. 1984, 7.2.2/2], and they play a predominant role in the foundations of rigid geometry. In the uniformly rigid setting, we are unable to handle general representable subsets; for instance, we do not know whether representable subsets induce admissible open subsets via the functor r which is induced by Berthelot's construction. We will thus only consider representable subsets of a specific kind, which we call semiaffinoid subdomains:

Definition 2.22. A subset U of a semiaffinoid K -space X is called a semiaffinoid subdomain if there is an affine flat R -model of ff type \mathfrak{X} for X and a finite composition of open immersions, completion morphisms and admissible blowups $\varphi : \mathfrak{U} \rightarrow \mathfrak{X}$ such that \mathfrak{U} is affine and such that U is equal to the image of φ^{urig} . We say that φ represents U as a semiaffinoid subdomain in X . We say that U is an elementary semiaffinoid subdomain in X if φ can be chosen as an open immersion into an admissible blowup, and we say that U is a retrocompact semiaffinoid subdomain in X if φ can be chosen as a composition of open immersions and admissible blowups; such a φ is said to represent U as an elementary or as a retrocompact semiaffinoid subdomain in X respectively.

In Corollary 2.25, we will see that semiaffinoid subdomains are actually representable in the sense of Definition 2.20.

Open immersions of formal R -schemes of ff type induce retrocompact open immersions of rigid generic fibers [de Jong 1995, 7.2.2 and 7.2.4(d)]. Moreover, completion morphisms induce (possibly nonretrocompact) open immersions of rigid generic fibers [de Jong 1995, 7.2.5], and admissible blowups induce isomorphisms of rigid generic fibers; see [Nicaise 2009, 2.19]. Hence a semiaffinoid subdomain $U \subseteq X$ is admissibly open in X^r . In particular, the K -homomorphism $\varphi^{\text{urig},*}$ corresponding to φ^{urig} is flat, since flatness is seen on the level of completions of stalks. Semiaffinoid subdomains may be regarded as nested rational subdomains defined in terms of strict or nonstrict inequalities involving semiaffinoid functions.

For example, the blowup of $\mathfrak{X} = \text{Spf } R[[S]]$ in the ideal (π, S) is covered by the affine open formal subschemes $\mathfrak{X}_1 = \text{Spf } (R[[S]]\langle V \rangle / (\pi V - S)) \cong \text{Spf } R\langle V \rangle$ and $\mathfrak{X}_2 = \text{Spf } (R[[S]]\langle W \rangle / (SW - \pi))$; the completion of \mathfrak{X}_1 along the ideal (π, V) represents the open disc with radius $|\pi|$ within the open unit disc, while the completion of \mathfrak{X}_2 along (π, W) defines the open annulus $|\pi| < |S| < 1$.

Remark 2.23. It is necessary to consider iterations as in Definition 2.22 because if \mathfrak{U} is an open subset of a flat formal R -scheme of ff type \mathfrak{X} , then an admissible blowup of \mathfrak{U} needs not extend to an admissible blowup of \mathfrak{X} ; see [Kappen 2009, Example 1.1.3.12].

In order to understand semiaffinoid subdomains, it will be useful to interpret strict transforms with respect to admissible blowups as pullbacks:

Lemma 2.24. *Let $\mathfrak{Y} \rightarrow \mathfrak{X}$ be a morphism of flat formal R -schemes of locally ff type, let $\mathfrak{X}' \rightarrow \mathfrak{X}$ be an admissible blowup, and let $\mathfrak{Y}' \rightarrow \mathfrak{Y}$ denote the induced admissible blowup of \mathfrak{Y} , that is, the strict transform of \mathfrak{Y} . The resulting square*

$$\begin{array}{ccc} \mathfrak{Y}' & \longrightarrow & \mathfrak{X}' \\ \downarrow & & \downarrow \\ \mathfrak{Y} & \longrightarrow & \mathfrak{X} \end{array}$$

is cartesian in the category of flat formal R -schemes of locally ff type.

Proof. The universal property of the fibered product in the category of flat formal R -schemes of locally ff type is readily verified using the universal property of admissible blowups, the fact that the functor rig maps admissible blowups to isomorphisms and the fact that rig is faithful on the category of flat formal R -schemes of locally ff type. □

In the following, we write \times' to denote the fibered product in the category of flat formal R -schemes of locally ff type. It is obtained from the usual fibered product by dividing out the coherent ideal of π -torsion; in particular, fibered products of affine flat formal R -schemes of ff type in the category of flat formal R -schemes of locally ff type are again affine.

As we have just observed, admissible blowups of flat formal R -schemes are preserved under pullback in the category of flat formal R -schemes of locally ff type. The same is true for open immersions and completion morphisms, since they are flat and since they are preserved under pullback in the category of all formal R -schemes of locally ff type.

Corollary 2.25. *Let X be a semiaffinoid K -space, let $U \subseteq X$ be a semiaffinoid subdomain, and let $Y \rightarrow X$ be a morphism of semiaffinoid K -spaces.*

- (i) *The preimage of U in Y is a semiaffinoid subdomain in Y .*

- (ii) If $\mathfrak{U} \rightarrow \mathfrak{X}$ represents U as a semiaffinoid subdomain in X and if $\mathfrak{Y} \rightarrow \mathfrak{X}$ is a model of $Y \rightarrow X$, then the projection $\mathfrak{U} \times'_{\mathfrak{X}} \mathfrak{Y} \rightarrow \mathfrak{Y}$ represents the preimage of U as a semiaffinoid subdomain in Y .
- (iii) If φ represents U as a semiaffinoid subdomain in X , then φ^{urig} represents all semiaffinoid morphisms to X with image in U . In particular, semiaffinoid subdomains are representable in the sense of Definition 2.20.

The analogous statements hold if we consider retrocompact or elementary semiaffinoid subdomains and their retrocompact or elementary representations.

Proof. Statement (ii) implies statement (i) in view of Corollary 2.14(iv). To show (ii), let us consider a factorization

$$\mathfrak{U} \xrightarrow{\varphi_n} \mathfrak{U}_n \xrightarrow{\varphi_{n-1}} \dots \xrightarrow{\varphi_1} \mathfrak{U}_1 \xrightarrow{\varphi_0} \mathfrak{X} \tag{\dagger}$$

of $\mathfrak{U} \rightarrow \mathfrak{X}$, where the φ_i are admissible blowups, open immersions or completion morphisms. By the remarks preceding this Corollary, we see that the projection $\mathfrak{U} \times'_{\mathfrak{X}} \mathfrak{Y} \rightarrow \mathfrak{Y}$ defines a semiaffinoid subdomain in Y . Passing to associated rigid spaces, we see that this semiaffinoid subdomain coincides with the preimage of U in Y . To prove (iii), let us write φ to denote $\mathfrak{U} \rightarrow \mathfrak{X}$, and let us assume that the image of $Y \rightarrow X$ lies in U ; we must show that $Y \rightarrow X$ factors uniquely through φ^{urig} . Since φ^{urig} induces an injection of physical points and isomorphisms of completed stalks, uniqueness follows from Krull’s Intersection theorem. Let us show that the desired factorization exists. Again, Corollary 2.14(iv) shows that $Y \rightarrow X$ admits a model $\mathfrak{Y} \rightarrow \mathfrak{X}$ with target \mathfrak{X} . Let us consider the pullback

$$\mathfrak{Y}_{n+1} \xrightarrow{\psi_n} \mathfrak{Y}_n \xrightarrow{\psi_{n-1}} \dots \xrightarrow{\psi_1} \mathfrak{Y}_1 \xrightarrow{\varphi_0} \mathfrak{Y}$$

of (\dagger) under $\mathfrak{Y} \rightarrow \mathfrak{X}$ in the category of flat formal R -schemes of locally ff type; then \mathfrak{Y}_{n+1} is affine, and all ψ_i that are open immersions or completion morphisms are isomorphisms: Indeed, $Y \rightarrow X$ factors through U , specialization maps are surjective onto the sets of closed points of flat formal R -schemes of locally ff type, and the closed points lie very dense in formal R -schemes of this type. Hence, the composition $\mathfrak{Y}_{n+1} \rightarrow \mathfrak{Y}$ is a composition of admissible blowups; by [Temkin 2008, 2.1.6], it is an admissible blowup. Since \mathfrak{Y}_{n+1} is affine, [Grothendieck 1961b, 3.4.2] shows that $\mathfrak{Y}_{n+1} \rightarrow \mathfrak{Y}$ is a finite admissible blowup. After applying urig , we thus obtain the desired factorization of $Y \rightarrow X$. \square

By Corollary 2.25(iii), every semiaffinoid subdomain may be viewed as a semiaffinoid K -space in a natural way.

Question 2.26. One may ask whether every representable subset of a semiaffinoid K -space is in fact a semiaffinoid subdomain. Unfortunately, we do not know the answer.

Corollary 2.27. *Let X be a semiaffinoid K -space, let $U \subseteq X$ be a semiaffinoid subdomain, and let \mathfrak{X} be a flat affine R -model of ff type for X . Then there exists a representation of U as a semiaffinoid subdomain in X with target \mathfrak{X} .*

Proof. Let $\mathfrak{U}' \rightarrow \mathfrak{X}'$ be a representation of U as a semiaffinoid subdomain in X , let us write $X = \text{sSp } A$, and let $\underline{A}, \underline{A}' \subseteq A$ be the R -models of ff type of A corresponding to \mathfrak{X} and \mathfrak{X}' respectively. By Corollary 2.14(iv) applied to the identity on A , there exists an R -model of ff type \underline{A}'' of A containing both \underline{A} and \underline{A}' . By Corollary 2.14(ii), the inclusions $\underline{A} \subseteq \underline{A}''$ and $\underline{A}' \subseteq \underline{A}''$ correspond to finite admissible blowups $\mathfrak{X}'' \rightarrow \mathfrak{X}$ and $\mathfrak{X}'' \rightarrow \mathfrak{X}'$. By Corollary 2.25(ii), the strict transform $\mathfrak{U}'' \rightarrow \mathfrak{X}''$ of $\mathfrak{U}' \rightarrow \mathfrak{X}'$ under $\mathfrak{X}'' \rightarrow \mathfrak{X}'$ represents U as a semiaffinoid subdomain in X . Composing this representation with the admissible blowup $\mathfrak{X}'' \rightarrow \mathfrak{X}$, we obtain a representation $\mathfrak{U}'' \rightarrow \mathfrak{X}$ of U as a semiaffinoid subdomain in X with target \mathfrak{X} , as desired. \square

Remark 2.28. One can easily show that if $U \subseteq X$ is a semiaffinoid subdomain and if $\mathfrak{Y} \rightarrow \mathfrak{X}$ is a model of the inclusion of U into X , then there exists a finite admissible blowup \mathfrak{Y}' of \mathfrak{Y} such that the composition $\mathfrak{Y}' \rightarrow \mathfrak{X}$ represents U as a semiaffinoid subdomain in X ; this fact will not be needed in the following.

Corollary 2.29. *Let X be a semiaffinoid K -space.*

- (i) *Let $U \subseteq X$ be a semiaffinoid subdomain, and let V be a subset of U . Then V is a semiaffinoid subdomain in U if and only if it is a semiaffinoid subdomain in X .*
- (ii) *The set of semiaffinoid subdomain in X is stable under the formation of finite intersections.*

Proof. If V is semiaffinoid in X , then $V = V \cap U$ is semiaffinoid in U by Corollary 2.25(i). Conversely, assume that V is semiaffinoid in U , and let $\mathfrak{U} \rightarrow \mathfrak{X}$ be a representation of U as a semiaffinoid subdomain in X . By Corollary 2.27, there exists a representation $\mathfrak{V} \rightarrow \mathfrak{U}$ of V as a semiaffinoid subdomain in U ; the composition $\mathfrak{V} \rightarrow \mathfrak{U} \rightarrow \mathfrak{X}$ represents V as a semiaffinoid subdomain in X . This settles the first statement. To show (ii), let us consider two semiaffinoid subdomains U and V in X . By Corollary 2.25(i), $U \cap V$ is a semiaffinoid subdomain in U ; by part (i), $U \cap V$ is thus a semiaffinoid subdomain in X . \square

These results obviously remain true if we only consider retrocompact semiaffinoid subdomains instead of general semiaffinoid subdomains. Similarly, elementary semiaffinoid subdomains are preserved under pullback with respect to morphisms of semiaffinoid spaces. However, if U is an elementary semiaffinoid subdomain in a semiaffinoid K -space X and if V is an elementary semiaffinoid

subdomain in U , then V needs not be elementary in X . Likewise, if U is a semi-affinoid subdomain in X and if V is a retrocompact semiaffinoid subdomain in U , then V needs not be retrocompact in X .

We conclude this section by identifying retrocompact semiaffinoid subdomains in affinoid K -spaces:

Lemma 2.30. *Let A be an affinoid K -algebra; then a retrocompact semiaffinoid subdomain U in $s\mathrm{Sp} A$ is an affinoid subdomain in $\mathrm{Sp} A$.*

Proof. Let $\varphi: \mathfrak{Y} \rightarrow \mathfrak{X}$ be a morphism defining U as a retrocompact semiaffinoid subdomain in X . By Corollary 2.15, \mathfrak{X} is of tf type over R . Since φ is adic, \mathfrak{Y} is of tf type over R as well, such that φ^{rig} is a morphism of affinoid K -spaces. By Corollary 2.25(iii), φ represents all semiaffinoid maps with image in U ; in particular it represents all affinoid maps with image in U . Hence, U is an affinoid subdomain in $\mathrm{Sp} A$. □

Conversely, it is clear that for any affinoid K -algebra A , the rational subdomains in $\mathrm{Sp} A$ define semiaffinoid subdomains in $s\mathrm{Sp} A$. Let $U \subseteq \mathrm{Sp} A$ be a general affinoid subdomain in $\mathrm{Sp} A$. By the theorem of Gerritzen and Grauert [Bosch et al. 1984, 7.3.5/1], U is a finite union of rational subdomains. Let \mathfrak{X} be any affine flat formal R -model of tf type for $\mathrm{Sp} A$. By [Bosch and Lütkebohmert 1993a, Lemma 4.4], there exist an admissible formal blowup $\mathfrak{X}' \rightarrow \mathfrak{X}$ of \mathfrak{X} and an open formal subscheme $\mathfrak{U} \subseteq \mathfrak{X}'$ such that $U = \mathfrak{U}^{\mathrm{rig}}$. However, we do not know whether \mathfrak{U} is affine, so we do not know whether a general affinoid subdomain U in $\mathrm{Sp} A$ is a semiaffinoid subdomain or even a representable subset in $s\mathrm{Sp} A$. Nonetheless, we will see that affinoid subdomains in $\mathrm{Sp} A$ are admissible open in the uniformly rigid G -topology on $s\mathrm{Sp} A$; see Proposition 2.34.

2C3. G -topologies on semiaffinoid spaces. We first define an auxiliary G -topology $\mathcal{T}_{\mathrm{aux}}$ on the category of semiaffinoid K -spaces equipped with the physical points functor; see [Bosch et al. 1984, 9.1.2]. The $\mathcal{T}_{\mathrm{aux}}$ -admissible subsets of a semiaffinoid K -space are the semiaffinoid subdomains of that space. If I is a rooted tree and if $i \in I$ is a vertex, we let $\mathrm{ch}(i)$ denote the set of children of i .

Definition 2.31. Let X be a semiaffinoid K -space, and let $(X_i)_{i \in I}$ be a finite family of semiaffinoid subdomains in X .

- (i) We say that $(X_i)_{i \in I}$ is an elementary covering of X if there exist an affine flat R -model of ff type \mathfrak{X} for X , an admissible blowup $\mathfrak{X}' \rightarrow \mathfrak{X}$ and an affine open covering $(\mathfrak{X}_i)_{i \in I}$ of \mathfrak{X}' such that for each $i \in I$, $\mathfrak{X}_i \subseteq \mathfrak{X}' \rightarrow \mathfrak{X}$ represents X_i as a semiaffinoid subdomain in X .
- (ii) We say that $(X_i)_{i \in I}$ is a treelike covering of X if there exists a rooted tree structure on I such that $X_r = X$, where r is the root of I , and such that

- $(X_j)_{j \in \text{ch}(i)}$ is an elementary covering of X_i for all $i \in I$ which are not leaves. A rooted tree structure on I with these properties is called suitable for $(X_i)_{i \in I}$.
- (iii) We say that $(X_i)_{i \in I}$ is a leaflike covering if it extends to a treelike covering $(X_i)_{i \in J}$, $J \supseteq I$, where J admits a suitable rooted tree structure such that I is identified with the set of leaves of J .
- (iv) We say that $(X_i)_{i \in I}$ is \mathcal{T}_{aux} -admissible if it admits a leaflike refinement.

If $(X_i)_{i \in I}$ is an elementary, treelike or leaflike covering of X , then by definition all X_i are retrocompact in X . For trivial reasons, condition (iv) in Definition 2.31 can be checked after refinement.

Arguing as in the proof of Corollary 2.27, we see that an elementary covering can be represented with respect to any flat affine R -model of ff type \mathfrak{X} of X . It follows that any treelike covering $(X_i)_{i \in I}$ together with a suitable rooted tree structure on I admits a model with respect to \mathfrak{X} ; that is, we have

- (i) for each $i \in I$, an affine flat R -model of ff type \mathfrak{X}_i for X_i such that $\mathfrak{X}_r = \mathfrak{X}$, where r denotes the root of I ,
- (ii) for each inner $i \in I$, an admissible blowup $\mathfrak{X}'_i \rightarrow \mathfrak{X}_i$ and
- (iii) for each inner $i \in I$ and for each child j of i , an open immersion $\mathfrak{X}_j \hookrightarrow \mathfrak{X}'_i$ such that $\mathfrak{X}_j \subseteq \mathfrak{X}'_i \rightarrow \mathfrak{X}_i$ represents X_j as a semiaffinoid subdomain in X_i .

Arguing as in the proof of Corollary 2.25, we see that elementary, treelike and leaflike coverings, suitable rooted tree structures and models in the above sense are preserved under pullback with respect to morphisms $Y \rightarrow X$ of semiaffinoid K -spaces and their models $\mathfrak{Y} \rightarrow \mathfrak{X}$, where \mathfrak{Y} and \mathfrak{X} are flat affine models of ff type for Y and X respectively.

Lemma 2.32. *Let X be a semiaffinoid K -space, let $(U_i)_{i \in I}$ be a covering of X by semiaffinoid subdomains, and for each $i \in I$, let $(V_{ij})_{j \in J_i}$ be a covering of U_i . If all of these coverings are leaflike or \mathcal{T}_{aux} -admissible, then the same holds for the covering $(V_{ij})_{i \in I, j \in J_i}$ of X .*

Proof. Let us first consider the case where the given coverings are leaflike. Let us choose a treelike covering $(U_i)_{i \in I'}$ of U extending $(U_i)_{i \in I}$ together with a suitable rooted tree structure on I' such that $I \subseteq I'$ is the set of leaves. Similarly, for each $i \in I$ we choose a treelike covering $(V_{ij})_{j \in J'_i}$ extending $(V_{ij})_{j \in J_i}$ together with a suitable rooted tree structure on J'_i such that $J_i \subseteq J'_i$ is identified with the set of leaves for all $i \in I$. For each $i \in I$, we glue the rooted tree J'_i to the rooted tree I' by identifying the root of J'_i with the leaf i of I' . We obtain a rooted tree J' whose set of leaves is identified with the disjoint union of the sets J_i , $i \in I$. For each $i \in I$, $U_i = V_{ir_i}$, where r_i is the root of J'_i ; hence we obtain a covering $(V_j)_{j \in J'}$ such that the given rooted tree structure on J' is suitable for $(V_j)_{j \in J'}$; indeed, this can be

checked locally on the rooted tree J' . We conclude that the composite covering $(V_{ij})_{i \in I, j \in J_i}$ of X is leaflike. The statement for \mathcal{T}_{aux} -admissible coverings now follows by passing to leaflike refinements. \square

Combining Lemma 2.32 and the fact that \mathcal{T}_{aux} -admissible coverings are stable under pullback, we see that the semiaffinoid subdomains and the \mathcal{T}_{aux} -admissible coverings define a G-topology on the category of semiaffinoid K -spaces equipped with the physical points functor. The following proposition suggests that \mathcal{T}_{aux} should be viewed as an analog of the weak G-topology in rigid geometry. We first define:

Definition 2.33. A retrocompact covering of a semiaffinoid K -space X is a finite family of retrocompact semiaffinoid subdomains of X that covers X on the level of physical points.

If I is a rooted tree, we write $\text{lv}(I)$ to denote the set of leaves of that tree, and we write $v(I)$ denote the volume of the tree, that is, its number of vertices.

Proposition 2.34. *Retrocompact coverings of semiaffinoid spaces are \mathcal{T}_{aux} -admissible.*

Proof. Let X be a semiaffinoid K -space, and let $(X_i)_{i \in I}$ be a finite family of retrocompact semiaffinoid subdomains in X covering X on the level of sets; we have to show that $(X_i)_{i \in I}$ is \mathcal{T}_{aux} -admissible. For each $i \in I$, we choose a retrocompact representation φ_i of X_i in X , such that the targets of the φ_i all coincide with a fixed flat affine target \mathfrak{X} . For each $i \in I$, we choose a factorization

$$\varphi_i = \beta_{i1} \circ \psi_{i1} \circ \cdots \circ \beta_{in_i} \circ \psi_{in_i},$$

where the ψ_{ij} are open immersions and the β_{ij} are admissible blowups,

$$\mathfrak{X}_{ij} \xrightarrow{\psi_{ij}} \mathfrak{X}'_{ij} \xrightarrow{\beta_{ij}} \mathfrak{X}_{i,j-1},$$

with $\mathfrak{X}_{i0} = \mathfrak{X}$. Let v denote the sum of the n_i ; we say that v is the total length of the given retrocompact representation. Let $\mathfrak{X}' \rightarrow \mathfrak{X}$ be an admissible blowup dominating all $\beta_{i1} : \mathfrak{X}'_{i1} \rightarrow \mathfrak{X}$, and let $\mathfrak{U}_i \subseteq \mathfrak{X}'$ denote the preimage of $\mathfrak{X}_{i1} \subseteq \mathfrak{X}'_{i1}$. The $\mathfrak{X}'_{i1}^{\text{rig}}$ cover $\mathfrak{X}^{\text{rig}}$, the specialization map $\text{sp}_{\mathfrak{X}'}$ is surjective onto the closed points of \mathfrak{X}' , and the closed points in \mathfrak{X}' lie very dense; hence \mathfrak{X}' is covered by the \mathfrak{U}_i . For each $i \in I$, we consider the pullback ψ'_i of

$$\beta_{i2} \circ \psi_{i2} \circ \cdots \circ \beta_{in_i} \circ \psi_{in_i}$$

under $\mathfrak{U}_i \subseteq \mathfrak{X}' \rightarrow \mathfrak{X}'_{i1}$, and moreover for each $j \in I$ different from i we consider the pullback φ'_{ij} of

$$\varphi_j = \beta_{j1} \circ \psi_{j1} \circ \cdots \circ \beta_{jn_j} \circ \psi_{jn_j}$$

under $\mathfrak{U}_i \subseteq \mathfrak{X}' \rightarrow \mathfrak{X}$, both in the category of flat formal R -schemes of ff type. For each $i \in I$, we choose a finite affine covering of \mathfrak{U}_i . For each constituent \mathfrak{V}_{is} of this covering with semiaffinoid generic fiber V_{is} , we choose finite affine coverings of $(\psi'_i)^{-1}(\mathfrak{V}_{is})$ and of $(\varphi'_{ij})^{-1}(\mathfrak{V}_{is})$, for $j \in I \setminus \{i\}$. We obtain a retrocompact covering of V_{is} , together with retrocompact representations as above of total length $v - 1$. If we let i and s vary, the resulting retrocompact covering of X refines $(X_i)_{i \in I}$. Since the V_{is} , for varying i and s , form an elementary covering of X , it suffices to see that the given retrocompact covering of V_{is} is \mathcal{T}_{aux} -admissible, which now follows by induction on v ; the case $v = 1$ is trivial. \square

Definition 2.35. Let $\mathcal{T}_{\text{urig}}$ denote the finest G-topology on the category of semiaffinoid K -spaces which is slightly finer than \mathcal{T}_{aux} in the sense of [Bosch et al. 1984, 9.1.2/1].

The G-topology $\mathcal{T}_{\text{urig}}$ is called the uniformly rigid G-topology. It exists by [Bosch et al. 1984, 9.2.1/2], and it is saturated in the sense that it satisfies conditions (G₀)–(G₂) in [Bosch et al. 1984, 9.1.2], saying that $\mathcal{T}_{\text{urig}}$ -admissibility of subsets can be checked locally with respect to $\mathcal{T}_{\text{urig}}$ -admissible coverings and that admissibility of a covering by $\mathcal{T}_{\text{urig}}$ -admissible subsets can be checked after refinement.

As a corollary of [BGR] 9.1.2/3, we obtain the following explicit description of the uniformly rigid G-topology on a semiaffinoid K -space:

Proposition 2.36. *Let X be a semiaffinoid K -space.*

- (i) *A subset $U \subseteq X$ is $\mathcal{T}_{\text{urig}}$ -admissible if and only if it admits a covering $(U_i)_{i \in I}$ by semiaffinoid subdomains such that for any morphism $\varphi: Y \rightarrow X$ of semiaffinoid K -spaces with $\varphi(Y) \subseteq U$, the induced covering of Y has a leaflike refinement.*
- (ii) *A covering $(U_i)_{i \in I}$ of a $\mathcal{T}_{\text{urig}}$ -admissible subset U in X by $\mathcal{T}_{\text{urig}}$ -admissible subsets is $\mathcal{T}_{\text{urig}}$ -admissible if and only if for any morphism $\varphi: Y \rightarrow X$ of semiaffinoid K -spaces with $\varphi(Y) \subseteq U$, the induced covering of Y has a leaflike refinement.*

Corollary 2.37. *Let X be a semiaffinoid K -space.*

- (i) *For any semiaffinoid subdomain U of X , the uniformly rigid G-topology on X restricts to the uniformly rigid G-topology on U .*
- (ii) *If $U \subseteq X$ is a finite union of retrocompact semiaffinoid subdomains in X , then U is $\mathcal{T}_{\text{urig}}$ -admissible, and every finite covering of U by retrocompact semiaffinoid subdomains in X is $\mathcal{T}_{\text{urig}}$ -admissible.*

Proof. By Corollary 2.29(i), the semiaffinoid subdomains in U are the semiaffinoid subdomains in X contained in U , and by Corollary 2.25(iii) the semiaffinoid

morphisms to X with image in U correspond to the semiaffinoid morphisms to U . Hence, statement (i) follows from parts (i) and (ii) of Proposition 2.36.

To prove the second statement, let $(U_i)_{i \in I}$ be a finite family of retrocompact semiaffinoid subdomains of X such that U is the union of the U_i . Let Y be any semiaffinoid K -space, and let $\varphi: Y \rightarrow X$ be any semiaffinoid morphism whose image lies in U . Then $(\varphi^{-1}(U_i))_{i \in I}$ is a retrocompact covering of Y ; by Proposition 2.34, it admits a leaflike refinement. By Proposition 2.36(i), we conclude that U is a $\mathcal{T}_{\text{urig}}$ -admissible subset of X , and by Proposition 2.36(ii) we see that the covering $(U_i)_{i \in I}$ of U is $\mathcal{T}_{\text{urig}}$ -admissible. \square

In particular, Corollary 2.37(ii) and the theorem of Gerritzen and Grauert [Bosch et al. 1984, 7.3.5/1] show that if A is an affinoid K -algebra and if $U \subseteq \text{Sp } A$ is an affinoid subdomain, then $U \subseteq \text{sSp } A$ is $\mathcal{T}_{\text{urig}}$ -admissible.

Remark 2.38 (quasicompactness). Proposition 2.36(ii) shows that semiaffinoid K -spaces are quasicompact in $\mathcal{T}_{\text{urig}}$, see [Bosch et al. 1984, p. 337]. By the maximum principle for affinoid K -algebras; it follows that $\text{sSp}(R[[S]] \otimes_R K)$ has no $\mathcal{T}_{\text{urig}}$ -admissible covering by semiaffinoid subdomains whose rings of functions are affinoid. In particular, the covering of $\text{sSp}(R[[S]] \otimes_R K)$ provided by Berthelot's construction is not $\mathcal{T}_{\text{urig}}$ -admissible.

Remark 2.39 (bases for $\mathcal{T}_{\text{urig}}$). Proposition 2.36 implies that the semiaffinoid subdomains form a basis for the uniformly rigid G -topology on a semiaffinoid K -space [Bosch et al. 1984, p. 338]. The retrocompact semiaffinoid subdomains in $\text{sSp}(K\langle S \rangle)$ do not form a basis for $\mathcal{T}_{\text{urig}}$: Indeed, $\text{sSp}(R[[S]] \otimes_R K)$ is a semiaffinoid subdomain in $\text{sSp}(K\langle S \rangle)$; by Lemma 2.30 and Remark 2.38, it does not admit a $\mathcal{T}_{\text{urig}}$ -admissible covering by retrocompact semiaffinoid subdomains in $\text{sSp}(K\langle S \rangle)$. Thus, even though the K -algebra $K\langle S \rangle$ is affinoid, the uniformly rigid G -topology on $\text{sSp}(K\langle S \rangle)$ turns out to be strictly coarser than the rigid G -topology on $\text{Sp}(K\langle S \rangle)$. We do not know whether this discrepancy already appears on the level of admissible subsets.

We conclude our discussion of the uniformly rigid G -topology $\mathcal{T}_{\text{urig}}$ by showing that it is finer than the Zariski topology \mathcal{T}_{Zar} which, on a semiaffinoid K -space X , is generated by the nonvanishing loci $D(f)$ of semiaffinoid functions f on X :

Proposition 2.40. *The uniformly rigid G -topology $\mathcal{T}_{\text{urig}}$ is finer than the Zariski topology \mathcal{T}_{Zar} .*

Proof. Let $X = \text{sSp } A$ be a semiaffinoid K -space, let $U \subseteq X$ be a Zariski-open subset, and let $f_1, \dots, f_n \in A$ be semiaffinoid functions such that U is the union of the Zariski-open subsets $D(f_i) = \{x \in \text{Max } A; f_i(x) \neq 0\}$. Let Y be a nonempty semiaffinoid K -space, and let $\varphi: Y \rightarrow X$ be a semiaffinoid morphism whose image is contained in U . For each i , the preimage $\varphi^{-1}(D(f_i))$ is the set of points $y \in Y$

where $\varphi^* f_i \neq 0$. Since Y is covered by the $\varphi^{-1}(D(f_i))$, the $\varphi^* f_i$ generate the unit ideal in B . That is, there exist elements b_1, \dots, b_n in B such that $b_1 \varphi^* f_1 + \dots + b_n \varphi^* f_n = 1$. Let us set $\gamma := (\max_i |b_i|_{\text{sup}})^{-1}$; this number is well-defined since the b_i are bounded functions on Y without a common zero. By the strict triangle inequality, $\max_i |\varphi^* f_i(y)| \geq \gamma$ for all $y \in Y$. For each i , let $Y_i \subseteq Y$ denote the set of points $y \in Y$ where $|\varphi^* f_i(y)| \geq \gamma$; then $(Y_i)_{1 \leq i \leq n}$ is a retrocompact covering of Y refining $(\varphi^{-1}(D(f_i)))_{1 \leq i \leq n}$. By Proposition 2.34, retrocompact coverings are $\mathcal{T}_{\text{urig}}$ -admissible; hence $U \subseteq X$ is $\mathcal{T}_{\text{urig}}$ -admissible. If $(U_j)_{j \in J}$ is a Zariski-covering of U , we may pass to a refinement and assume that for all $j \in J$, $U_j = D(g_j) \subseteq X$ for some semiaffinoid function g_j on X ; we can then argue along the same lines to prove that $(U_j)_{j \in J}$ is a $\mathcal{T}_{\text{urig}}$ -admissible covering of U . \square

The above argument works even though the maximum principle might fail on Y . Let us point out that our proof shows the following: If f_1, \dots, f_n are semiaffinoid functions on X , if

$$U = \bigcup_{i=1}^n D(f_i)$$

is the associated Zariski-open subset of X , and if we set

$$U_{\geq \varepsilon} = \bigcup_{i=1}^n \{x \in X ; |f_i(x)| \geq \varepsilon\}$$

for $\varepsilon \in \sqrt{|\overline{K^*}|}$, then the resulting covering $(U_{\geq \varepsilon})_\varepsilon$ of U by finite unions of retrocompact semiaffinoid subdomains of X is $\mathcal{T}_{\text{urig}}$ -admissible. In particular, Zariski-open subsets in semiaffinoid spaces need not be quasicompact in the uniformly rigid G-topology. As a consequence, the sheaf of uniformly rigid functions on a semiaffinoid K -space, to be defined in the following section, may have unbounded sections on Zariski-open subsets.

2C4. The acyclicity theorem. Let X be a semiaffinoid K -space. We show that the presheaf \mathbb{O}_X that we introduced after Definition 2.20 is a sheaf for \mathcal{T}_{aux} and, hence, extends uniquely to a sheaf for $\mathcal{T}_{\text{urig}}$. More generally, we show that every \mathbb{O}_X -module associated to a finite module over the ring of global functions on X is *acyclic* for any $\mathcal{T}_{\text{urig}}$ -admissible covering of X in the sense of [Bosch et al. 1984] p. 324. Adopting methods from [Lütkebohmert 1990], we derive our acyclicity theorem from results in formal geometry; we also use ideas from [Lipshitz and Robinson 2000, III.3.2].

Let us recall from [Bosch et al. 1984, p. 324] that if \mathcal{F} is a presheaf in \mathbb{O}_X -modules on \mathcal{T}_{aux} , a covering $(X_i)_{i \in I}$ of X by semiaffinoid subdomains is called \mathcal{F} -*acyclic* if the associated augmented Čech complex is acyclic. The covering $(X_i)_{i \in I}$

is called *universally \mathcal{F} -acyclic* if $(X_i \cap U)_{i \in I}$ is $\mathcal{F}|_U$ -acyclic for any semiaffinoid subdomain $U \subseteq X$.

Theorem 2.41. *For a semiaffinoid K -space X , \mathcal{T}_{aux} -admissible coverings are \mathbb{O}_X -acyclic.*

Proof. Let us first consider an elementary covering $(X_i)_{i \in I}$. Let us choose a formal representation $(\mathfrak{X}, \beta: \mathfrak{X}' \rightarrow \mathfrak{X}, (\mathfrak{X}_i)_{i \in I})$ of $(X_i)_{i \in I}$, where β is an admissible blowup and where $(\mathfrak{X}_i)_{i \in I}$ is a finite affine covering of \mathfrak{X}' such that $\mathfrak{X}_i \subseteq \mathfrak{X}' \rightarrow \mathfrak{X}$ represents X_i in X . By the ff type transcription of [Lütkebohmert 1990, 2.1], $\beta^\# \otimes_R K$ is an isomorphism; hence β induces a natural identification of augmented Čech complexes

$$C_{\text{aug}}^\bullet((X_i)_{i \in I}, \mathbb{O}_X) \cong C_{\text{aug}}^\bullet((\mathfrak{X}_i)_{i \in I}, \mathbb{O}_{\mathfrak{X}'} \otimes_R K).$$

We have to show that the complex on the right hand side is acyclic. Since $\mathbb{O}_{\mathfrak{X}'} \otimes_R K$ is a sheaf on \mathfrak{X}' , it suffices to show that

$$\check{H}^q((\mathfrak{X}_i)_{i \in I}, \mathbb{O}_{\mathfrak{X}'} \otimes_R K) = 0$$

for all $q \geq 1$. Since I is finite, we have an identification

$$\check{H}^q((\mathfrak{X}_i)_{i \in I}, \mathbb{O}_{\mathfrak{X}'} \otimes_R K) = \check{H}^q((\mathfrak{X}_i)_{i \in I}, \mathbb{O}_{\mathfrak{X}'}) \otimes_R K.$$

By the comparison theorem [Grothendieck 1961b, 4.1.5 and 4.1.7] and by the vanishing theorem [Grothendieck 1961b, 1.3.1], the higher cohomology groups of a coherent sheaf on an affine noetherian formal scheme vanish. Since the \mathfrak{X}_i are affine, Leray’s theorem implies that

$$\check{H}^q((\mathfrak{X}_i)_{i \in I}, \mathbb{O}_{\mathfrak{X}'}) = H^q(\mathfrak{X}', \mathbb{O}_{\mathfrak{X}'}).$$

By [Grothendieck 1961b, 1.4.11], $H^q(\mathfrak{X}', \mathbb{O}_{\mathfrak{X}'}) = \Gamma(\mathfrak{X}, R^q \beta_* \mathbb{O}_{\mathfrak{X}'})$, and by the ff type transcription of [Lütkebohmert 1990, 2.1] this module is π -torsion. We have thus finished the proof in the case where $(X_i)_{i \in I}$ is an elementary covering.

Let us turn towards the general case. By definition, every \mathcal{T}_{aux} -admissible covering of X has a leaflike refinement; by [Bosch et al. 1984, 8.1.4/3] it is enough to show that the leaflike coverings of X are universally \mathbb{O}_X -acyclic. Since leaflike coverings are preserved with respect to pullback under morphisms of semiaffinoid K -spaces, it suffices to show that any leaflike covering $(X_i)_{i \in I}$ of X is \mathbb{O}_X -acyclic.

Let $(X_j)_{j \in J}$ be a treelike covering of X extending $(X_i)_{i \in I}$, and let us choose a suitable rooted tree structure on J such that $I \subseteq J$ is identified with the set of leaves of J . We argue by induction on the volume of J . If J has only one vertex, the covering $(X_i)_{i \in I}$ is trivial and, hence, \mathbb{O}_X -acyclic. Let us assume that J has more than one vertex. Let $\iota \in I$ be a leaf of J such that the length $l(\iota)$ of the path from ι to the root is maximal in $\{l(i) ; i \in I\}$. Let $\iota' := \text{par}(\iota)$ denote the parent of ι .

By maximality of $l(\iota)$, all siblings $i \in \text{ch}(\iota')$ of ι are leaves of J . Let $J' := J \setminus \text{ch}(\iota')$ be the rooted subtree of J that is obtained by removing the siblings of ι (including ι itself). Then

- (i) the set of leaves of J' is $I' := (I \setminus \text{ch}(\iota')) \cup \{\iota'\}$,
- (ii) $(X_j)_{j \in J'}$ is a treelike covering of X , and
- (iii) $v(J') < v(J)$.

By our induction hypothesis, the covering $(X_i)_{i \in I'}$ is \mathbb{O}_X -acyclic. Since $(X_i)_{i \in I'}$ is a refinement of $(X_i)_{i \in I'}$, [Bosch et al. 1984, 8.1.4/3] says that it suffices to prove that for any $r \geq 0$ and any tuple $(i_0, \dots, i_r) \in (I')^{r+1}$, the covering $(X_i \cap X_{i_0 \dots i_r})_{i \in I}$ of $X_{i_0 \dots i_r}$ is \mathbb{O}_X -acyclic, where $X_{i_0 \dots i_r}$ denotes the intersection $X_{i_0} \cap \dots \cap X_{i_r}$. Let us assume that there exists some $0 \leq s \leq r$ such that $i_s \neq \iota'$. Then $i_s \in I$. Since $X_{i_0 \dots i_r} \subseteq X_{i_s}$, we see that the trivial covering of $X_{i_0 \dots i_r}$ refines $(X_i \cap X_{i_0 \dots i_r})_{i \in I}$. Since trivial coverings restrict to trivial coverings and since trivial coverings are acyclic, we deduce from [Bosch et al. 1984, 8.1.4/3] that $(X_i \cap X_{i_0 \dots i_r})_{i \in I}$ is acyclic. It remains to consider the case where all i_s , $0 \leq s \leq r$, coincide with ι' . That is, it remains to see that the covering $(X_i \cap X_{\iota'})_{i \in I}$ of $X_{\iota'}$ is \mathbb{O}_X -acyclic. It admits the elementary covering $(X_i)_{i \in \text{ch}(\iota')}$ as a refinement. Since elementary coverings restrict to elementary coverings and since elementary coverings are \mathbb{O}_X -acyclic by what we have shown so far, we conclude by [Bosch et al. 1984, 8.1.4/3] that $(X_i \cap X_{\iota'})_{i \in I}$ is \mathbb{O}_X -acyclic, as desired. \square

By [Bosch et al. 1984, 9.2.3/1], \mathbb{O}_X extends uniquely to a sheaf for $\mathcal{T}_{\text{urig}}$ which we again denote by \mathbb{O}_X and which we call the structural sheaf or the sheaf of uniformly rigid functions. We can now easily discuss a fundamental example of a nonadmissible finite covering of a semiaffinoid K -space by semiaffinoid subdomains:

Example 2.42. The canonical covering of the semiaffinoid closed unit disc

$$\text{sSp}(K\langle T \rangle)$$

by the semiaffinoid open unit disc $\text{sSp}(R[[T]] \otimes_R K)$ and the semiaffinoid unit circle $\text{sSp}(K\langle T, T^{-1} \rangle)$ is not $\mathcal{T}_{\text{urig}}$ -admissible and, hence, not \mathcal{T}_{aux} -admissible. Indeed, the two covering sets are nonempty and disjoint, while the ring of functions $K\langle T \rangle$ on the closed semiaffinoid unit disc has no nontrivial idempotents.

If X is a semiaffinoid K -space with ring of global functions A and if M is a finite A -module, the presheaf $M \otimes \mathbb{O}_X$ sending a semiaffinoid subdomain U in X to $M \otimes_A \mathbb{O}_X(U)$ is an \mathbb{O}_X -module, which we call the \mathbb{O}_X -module *associated* to M . A presheaf \mathcal{F} equipped with an \mathbb{O}_X -module structure is called *associated* if it is isomorphic to $M \otimes \mathbb{O}_X$ for some finite A -module M . We sometimes abbreviate $\tilde{M} := M \otimes \mathbb{O}_X$.

Corollary 2.43. *Let X be a semiaffinoid K -space, and let \mathcal{F} be an associated \mathcal{O}_X -module. Then every \mathcal{T}_{aux} -admissible covering $(X_i)_{i \in I}$ of X is \mathcal{F} -acyclic.*

Proof. By [Bosch et al. 1984, 8.1.4/3], we may assume that I is finite. Using Theorem 2.41, the proof is now literally the same as the proof of [Bosch et al. 1984, 8.2.1/5]. \square

In particular, $M \otimes \mathcal{O}_X$ is a \mathcal{T}_{aux} -sheaf. By [Bosch et al. 1984, 9.2.3/1], $M \otimes \mathcal{O}_X$ extends uniquely to a $\mathcal{T}_{\text{urig}}$ -sheaf on X which we again denote by $M \otimes \mathcal{O}_X$ or by \tilde{M} and which we call the sheaf associated to M .

Remark 2.44. If $U \subseteq X$ is a representable subset that is $\mathcal{T}_{\text{urig}}$ -admissible, then $\mathcal{O}_X(U) = \mathcal{O}_U(U)$. Indeed, U admits a $\mathcal{T}_{\text{urig}, X}$ -admissible covering by semiaffinoid subdomains in X ; since morphisms of semiaffinoid spaces are continuous for $\mathcal{T}_{\text{urig}}$, this covering is also $\mathcal{T}_{\text{urig}, U}$ -admissible, so the statement follows from the fact that both \mathcal{O}_X and \mathcal{O}_U are $\mathcal{T}_{\text{urig}}$ -sheaves. However, it is not clear whether $\mathcal{T}_{\text{urig}, X}$ restricts to $\mathcal{T}_{\text{urig}, U}$; for example, we do not know whether a semiaffinoid subdomain of U is $\mathcal{T}_{\text{urig}, X}$ -admissible. Of course, this does not affect our theory since we do not deal with general representable subsets.

The category of abelian sheaves on $(X, \mathcal{T}_{\text{urig}}|_X)$ has enough injective objects, so the functor $\Gamma(X, \cdot)$ from the category of abelian sheaves on X to the category of abelian groups has a right derived functor $H^\bullet(X, \cdot)$. By the acyclicity theorem and its Corollary 2.43, this right derived functor can, for associated \mathcal{O}_X -modules, be calculated in terms of Čech cohomology:

Corollary 2.45. *Let X be a semiaffinoid K -space, and let \mathcal{F} be an associated \mathcal{O}_X -module. Then the natural homomorphism $\check{H}^q(U, \mathcal{F}) \rightarrow H^q(U, \mathcal{F})$ is an isomorphism for all $\mathcal{T}_{\text{urig}}$ -admissible subsets $U \subseteq X$. In particular, $H^q(U, \mathcal{F}) = 0$ for all $q > 0$ and all semiaffinoid subdomains $U \subseteq X$.*

Proof. The system S of semiaffinoid subdomains in X satisfies the following properties:

- (i) S is stable under the formation of intersections,
- (ii) every $\mathcal{T}_{\text{urig}}$ -admissible covering $(U_i)_{i \in I}$ of a $\mathcal{T}_{\text{urig}}$ -admissible subset $U \subseteq X$ admits a $\mathcal{T}_{\text{urig}}$ -admissible refinement by sets in S , and
- (iii) $\check{H}^q(U, \mathcal{F})$ vanishes for all $q > 0$ and all $U \in S$;

hence the statement follows by means of the standard Čech spectral sequence argument. \square

Transcribing the proof of [Bosch et al. 1984, 7.3.2/1], we see that if A is a semiaffinoid K -algebra with associated semiaffinoid K -space X and if $\mathfrak{m} \subseteq A$ is a maximal ideal corresponding to a point $x \in X$, then the stalk $\mathcal{O}_{X, x}$ is local with

maximal ideal $\mathfrak{m}_{\mathbb{O}_{X,x}}$ which coincides with the ideal of germs of functions vanishing in x . The arguments in the proof of [ibid., 7.3.2/3] are also seen to work in our situation, showing that the natural homomorphisms $A/\mathfrak{m}^{n+1} \rightarrow \mathbb{O}_{X,x}/\mathfrak{m}^{n+1}\mathbb{O}_{X,x}$ are isomorphisms for all $n \in \mathbb{N}$. The rings $\mathbb{O}_{X,x}$ are noetherian, which can for example be seen by imitating the proof of [ibid., 7.3.2/7].

Transcribing the discussion at the beginning of [ibid., 9.3.1], we see that the uniformly rigid G -topology and the sheaf of uniformly rigid functions define a functor from the category of semiaffinoid K -spaces to the category of locally ringed G -topological K -spaces. The proof of [ibid., 9.3.1/2] carries over verbatim to the semiaffinoid situation, showing that this functor is fully faithful. We call a locally ringed G -topological K -space *semiaffinoid* if it lies in the essential image of this functor.

2D. Uniformly rigid spaces. We are now able to define the category of uniformly rigid K -spaces:

Definition 2.46. Let X be a locally ringed G -topological K -space.

- (i) An admissible *semiaffinoid covering* of X is an admissible covering $(X_i)_{i \in I}$ of X such that for each $i \in I$, $(X_i, \mathbb{O}_X|_{X_i})$ is a semiaffinoid K -space.
- (ii) The space X is called *uniformly rigid* if it satisfies conditions (G_0) – (G_1) in [Bosch et al. 1984, 9.1.2] and if it admits an admissible semiaffinoid covering.
- (iii) An admissible open subset U of a uniformly rigid K -space X is called an *open semiaffinoid subspace* of X if $(U, \mathbb{O}_X|_U)$ is a semiaffinoid K -space.

Remark 2.47. In the author’s PhD thesis, open semiaffinoid subspaces were simply called semiaffinoid subspaces [Kappen 2009, Section 1.3.9]

The category uRig_K of uniformly rigid K -spaces is a full subcategory of the category of locally G -topological K -spaces, and it contains the category of semiaffinoid K -spaces as a full subcategory.

Remark 2.48. We do not know whether an open semiaffinoid subspace U of a semiaffinoid K -space X is necessarily a semiaffinoid subdomain in X . However, one easily verifies that U is a representable subset in X . Moreover, one sees that U is locally a semiaffinoid subdomain in X ; see Lemma 2.52 for a precise statement. In rigid geometry, the open affinoid subvarieties [Bosch et al. 1984, p. 357] of an affinoid space are precisely the affinoid subdomains, which means that there is no need to distinguish between the two notions in the affinoid setting.

Remark 2.49. Let $X = \text{sSp } A$ be a semiaffinoid K -space, and let $U = \text{sSp } B$ be an open semiaffinoid subspace of X ; then the restriction homomorphism $A \rightarrow B$ is flat. Indeed, for every maximal ideal $\mathfrak{n} \subseteq B$ with corresponding point $x \in U$ and preimage $\mathfrak{m} \subseteq A$, the induced homomorphism $A_{\mathfrak{m}} \rightarrow B_{\mathfrak{n}}$ induces an isomorphism of

maximal-adic completions; by the flatness criterion [Bourbaki 1998, III.5.2, Theorem 1], we conclude that $A \rightarrow B_{\mathfrak{n}}$ is flat for all maximal ideals \mathfrak{n} in B , which implies that $A \rightarrow B$ is flat.

Lemma 2.50. *The open semiaffinoid subspaces of a uniformly rigid K -space X form a basis for the G -topology on X .*

Proof. Let $(X_i)_{i \in I}$ be an admissible semiaffinoid covering of X , and let $U \subseteq X$ be an admissible open subset. Then $(X_i \cap U)_{i \in I}$ is an admissible covering of U . For each $i \in I$, $X_i \cap U$ is admissible open in X_i and, hence, admits an admissible covering by semiaffinoid subdomains of X_i . Hence, U has an admissible semiaffinoid covering. \square

It follows that if X is a uniformly rigid K -space and if $U \subseteq X$ is an admissible open subset, then $(U, \mathbb{O}_X|_U)$ is a uniformly rigid K -space, again.

It is now clear that the gluing theorem [Bosch et al. 1984, 9.3.2/1] and its proof carry over verbatim to the uniformly rigid setting. Similarly, a morphism of uniformly rigid spaces can be defined locally on the domain; this is the uniformly rigid version of [ibid., 9.3.3/1], and again the proof is obtained by literal transcription. Furthermore, a uniformly rigid K -space is determined by its functorial points with values in semiaffinoid K -spaces.

We can also copy the proof of [ibid., 9.3.3/2] to see that if X is a semiaffinoid K -space and if Y is a uniformly rigid K -space, then the set of morphisms from Y to X is naturally identified with the set of K -algebra homomorphisms from $\mathbb{O}_X(X)$ to $\mathbb{O}_Y(Y)$.

Let \mathfrak{X} be an affine formal R -scheme of ff type with semiaffinoid generic fiber X . The associated specialization map $\mathrm{sp}_{\mathfrak{X}}$ which we discussed in Section 2B1 is naturally enhanced to a morphism of G -ringed R -spaces $\mathrm{sp}_{\mathfrak{X}}: X \rightarrow \mathfrak{X}$. Morphisms of uniformly rigid K -spaces being defined locally on the domain, we see that $\mathrm{sp}_{\mathfrak{X}}$ is *final* among all morphisms of G -ringed R -spaces from uniformly rigid K -spaces to \mathfrak{X} . Using this universal property, we can invoke gluing techniques to construct the *uniformly rigid generic* fiber $\mathfrak{X}^{\mathrm{urig}}$ of a general formal R -scheme of locally ff type \mathfrak{X} , together with a functorial specialization map $\mathrm{sp}_{\mathfrak{X}}: \mathfrak{X}^{\mathrm{urig}} \rightarrow \mathfrak{X}$ which is universal among all morphisms of G -ringed R -spaces from uniformly rigid K -spaces to \mathfrak{X} ; this process does not involve Berthelot's construction. It is easily seen that urig is *faithful* on the category of *flat* formal R -schemes of locally ff type. A formal R -model of a uniformly rigid K -space X is a formal R -scheme \mathfrak{X} of locally ff type together with an isomorphism $X \cong \mathfrak{X}^{\mathrm{urig}}$. The map $\mathrm{sp}_{\mathfrak{X}}$ is surjective onto the closed points of \mathfrak{X} whenever \mathfrak{X} is flat over R . This follows from Remark 2.5, together with the remark that the underlying topological space of \mathfrak{X} is a Jacobson space [Grothendieck and Dieudonné 1971, 0.2.8 and 6.4], so that the condition on a point in \mathfrak{X} of being closed is local.

Question 2.51. Under what conditions does a uniformly rigid K -space admit a formal R -model?

By Proposition 2.16, the category of semiaffinoid K -spaces has *fibered products*; following the method outlined in [Bosch et al. 1984, 9.3.5], we see that the category of uniformly rigid K -spaces has fibered products as well and that these are constructed by gluing semiaffinoid fibered products of open semiaffinoid subspaces. It is clear from this description that the *urig*-functor preserves fibered products.

Open semiaffinoid subspaces of semiaffinoid spaces can be described in the style of the Gerritzen–Grauert theorem [ibid., 7.3.5/3]:

Lemma 2.52. *Let X be a semiaffinoid K -space, and let $U \subseteq X$ be an open semiaffinoid subspace. Then U admits a leaflike covering $(U_i)_{i \in I}$ such that each U_i is a semiaffinoid subdomain in X .*

Proof. By Lemma 2.50, U admits an admissible covering $(V_j)_{j \in J}$ by semiaffinoid subdomains V_j of X ; by Proposition 2.36, this covering is refined by a leaflike covering $(U_i)_{i \in I}$ of U . Via pullback, the V_j are semiaffinoid subdomains of U . Let $\varphi: I \rightarrow J$ denote a refinement map. By Corollary 2.29(i), for each $i \in I$ the set U_i is a semiaffinoid subdomain in $V_{\varphi(i)}$ and, hence, in X , as desired. \square

A morphism of uniformly rigid K -spaces is called *flat* in a point of its domain if it induces a flat homomorphism of stalks in this point, and it is called flat if it is flat in all points. Clearly a morphism of semiaffinoid K -spaces is flat in this sense if and only if the underlying homomorphism of rings of global sections is flat.

2D1. Comparison with rigid geometry. In Section 2C1, we have defined the rigid space X^r associated to a semiaffinoid K -space $X = \text{sSp } A$ together with a universal K -homomorphism $A \rightarrow \Gamma(X^r, \mathbb{C}_{X^r})$ which induces a bijection $X^r \rightarrow X$ of physical points and isomorphisms of completed stalks. We will show that this universal homomorphism extends to a morphism $\text{comp}_X: X^r \rightarrow X$ of locally G -ringed K -spaces which is final among all morphisms from rigid K -spaces to X . To do so, we first show that the above bijection is continuous, that is, that the rigid G -topology \mathcal{T}_{rig} is finer than $\mathcal{T}_{\text{urig}}$. We will need the following elementary fact from rigid geometry; the proof is left as an exercise to the reader:

Lemma 2.53. *Let X be an affinoid K -space, and let $U \subseteq X$ be a subset admitting a covering $(U_i)_{i \in I}$ by admissible open subsets $U_i \subseteq X$ such that for any affinoid K -space Y and any morphism $\varphi: Y \rightarrow X$ with image in U , the induced covering $(\varphi^{-1}(U_i))_{i \in I}$ of Y has a refinement which is a finite covering by affinoid subdomains. Then $U \subseteq X$ is admissible.*

Proposition 2.54. *The rigid G -topology \mathcal{T}_{rig} on X is finer than the uniformly rigid G -topology $\mathcal{T}_{\text{urig}}$.*

Proof. It is clear that \mathcal{T}_{aux} -admissible subsets and \mathcal{T}_{aux} -admissible coverings are \mathcal{T}_{rig} -admissible. Let $U \subseteq X$ be a \mathcal{T}_{rig} -admissible subset. To check that U is \mathcal{T}_{rig} -admissible, we may work locally on X^r . Let $V' \subseteq X^r$ be an affinoid subspace; by Proposition 2.18, the open immersion $V' \hookrightarrow X^r$ corresponds to a morphism $V \rightarrow X$, where V denotes the semiaffinoid K -space associated to V' such that $V' = V^r$. After pulling U back under this morphism, we may thus assume that the K -algebra of global functions on X is affinoid. Let $(U_i)_{i \in I}$ be a covering of U by semiaffinoid subdomains in X such that condition (i) of Proposition 2.36 is satisfied. Let Y be an affinoid K -space, and let $\varphi: Y \rightarrow X^r$ be a morphism of rigid spaces that factors through U . By Proposition 2.18, we may also view φ as a morphism of semiaffinoid K -spaces. By assumption, the covering $(\varphi^{-1}(U_i))_{i \in I}$ of Y has a leaflike refinement; by Lemma 2.30, this refinement is affinoid. It now follows from Lemma 2.53 that $U \subseteq X$ is \mathcal{T}_{rig} -admissible.

Let now $(U_i)_{i \in I}$ be a \mathcal{T}_{rig} -admissible covering of U by \mathcal{T}_{rig} -admissible subsets U_i . We have seen that U and the U_i are \mathcal{T}_{rig} -admissible; we claim that the covering $(U_i)_{i \in I}$ is \mathcal{T}_{rig} -admissible as well. Again, we may work locally on X^r and thereby assume that the K -algebra of functions on X is affinoid. Let Y be an affinoid K -space, and let $\varphi: Y \rightarrow X^r$ be a morphism of affinoid K -spaces, which we may also view as a morphism of semiaffinoid K -spaces. Since $(U_i)_{i \in I}$ is \mathcal{T}_{rig} -admissible, we see by Proposition 2.36(ii) that $(\varphi^{-1}(U_i))_{i \in I}$ has a leaflike and, hence, affinoid refinement. It follows that $(U_i)_{i \in I}$ is \mathcal{T}_{rig} -admissible. \square

If $U \subseteq X$ is a semiaffinoid subdomain, then the morphism $U^r \rightarrow X^r$ provided by Proposition 2.18 is an open immersion onto the preimage of U under the continuous bijection $\text{comp}_X: X^r \rightarrow X$; hence comp_X extends to a morphism of G -ringed K -spaces with respect to \mathcal{T}_{aux} , which then again extends uniquely to a morphism of G -ringed K -spaces with respect to $\mathcal{T}_{\text{urig}}$. One easily verifies that comp_X is local.

Proposition 2.55. *The morphism comp_X is final among all morphisms from rigid K -spaces to X .*

Proof. Let Y be a rigid K -space, and let $\psi: Y \rightarrow X$ be a morphism of locally G -ringed K -spaces. By Proposition 2.18, there is a unique morphism $\psi^r: Y \rightarrow X^r$ such that ψ and $\text{comp}_X \circ \psi^r$ coincide on global sections. Since the points and the completed stalks of X are recovered from the K -algebra of global sections of X , it follows that ψ and $\text{comp}_X \circ \psi^r$ coincide. \square

Let X be any uniformly rigid K -space. Since the open semiaffinoid subspaces of X form a basis for the G -topology on X , we can use standard gluing arguments to show that the comparison morphisms attached to these open semiaffinoid subspaces glue to a universal comparison morphism

$$\text{comp}_X: X^r \rightarrow X$$

from a rigid K -space to X .

Remark 2.56. The functor $X \mapsto X^r$ is faithful, yet not fully faithful. For example, it is easily seen that an unbounded function on the rigid open unit disc induces a morphism to the rigid projective line over K which is not induced by a morphism from the semiaffinoid open unit disc $s\mathrm{Sp}(R[[S]] \otimes_R K)$ to the uniformly rigid projective line over K . Likewise, the functor r forgets the distinction between the semiaffinoid open unit disc just mentioned and the uniformly rigid open unit disc that is the generic fiber of a quasiparacompact formal R -model of locally tf type for the rigid open unit disc. One can prove that $X \mapsto X^r$ is fully faithful on the full subcategory of reduced semiaffinoid K -spaces.

Remark 2.57. The functor $X \mapsto X^r$ preserves fibered products. Indeed, this may be checked in the semiaffinoid situation, where it follows from the fact that fibered products of semiaffinoid spaces are uniformly rigid generic fibers of fibered products of affine flat formal R -models, together with the fact that Berthelot’s generic fiber functor preserves fibered products [de Jong 1995, 7.2.4(g)]. In particular, $X \mapsto X^r$ preserves group structures.

Remark 2.58. We have seen that comp_X induces isomorphisms of completed stalks. Examining Berthelot’s construction, one easily sees that comp_X in fact already induces isomorphisms of noncompleted stalks; the proof of this statement is left as an exercise.

We have seen that every uniformly rigid K -space X has an underlying classical rigid K -space X^r such that X and X^r share all local properties. That is, a uniformly rigid K -space can be seen as a rigid K -space equipped with an additional global uniform structure. Every quasiparacompact and quasiseparated rigid K -space carries a canonical uniformly rigid structure, which may be called the Raynaud-type uniform structure: let \mathcal{C} temporarily denote the category of quasiparacompact flat formal R -schemes of locally tf type, and let $\mathcal{C}_{\mathrm{Bl}}$ denote its localization with respect to the class of admissible formal blowups. It follows easily from the definitions that the functor $\mathrm{urig}|_{\mathcal{C}}: \mathcal{C} \rightarrow \mathrm{uRig}_K$ factors through a functor $\mathrm{ur}': \mathcal{C}_{\mathrm{Bl}} \rightarrow \mathrm{uRig}_K$. By [Bosch 2005, Theorem 2.8/3], the functor rig induces an equivalence $\mathrm{rig}_{\mathrm{Bl}}$ between $\mathcal{C}_{\mathrm{Bl}}$ and the category Rig'_K of quasiparacompact and quasiseparated rigid K -spaces. The functor $\mathrm{rig}_{\mathrm{Bl}}$ will be called the *Raynaud equivalence*. Composing ur' with a quasiinverse of $\mathrm{rig}_{\mathrm{Bl}}$, we obtain a functor $\mathrm{ur}: \mathrm{Rig}'_K \rightarrow \mathrm{uRig}_K$; if Y is in Rig'_K , we say that $Y^{\mathrm{ur}} := \mathrm{ur}(Y)$ is the *uniformly rigid K -space associated to Y* . Of course, it depends on the choice of a quasiinverse of the Raynaud equivalence.

Proposition 2.59. *The composite functor $r \circ \mathrm{ur}$ is quasiisomorphic to the identity on Rig'_K .*

Proof. Let $\text{rig}_{\text{Bl}}^{-1}$ denote the chosen inverse of the Raynaud equivalence. Let Y be an object of Rig'_K ; then $\text{rig}_{\text{Bl}}^{-1}(Y)$ is a quasiparacompact flat formal R -model of locally tf type for Y , and $Y^{\text{ur}} = \text{rig}_{\text{Bl}}^{-1}(Y)^{\text{urig}}$, which implies that $(Y^{\text{ur}})^r = \text{rig}_{\text{Bl}}^{-1}(Y)^{\text{rig}}$, functorially in Y . That is, $r \circ \text{ur} = \text{rig} \circ \text{rig}_{\text{Bl}}^{-1}$, which is isomorphic to the identity functor. \square

In particular, after choosing an isomorphism $r \circ \text{ur} \cong \text{id}$, the comparison morphisms $\text{comp}_{Y^{\text{ur}}}$ induce functorial comparison morphisms

$$\text{comp}_Y : Y \cong (Y^{\text{ur}})^r \rightarrow Y^{\text{ur}}$$

for all quasiparacompact and quasiseparated rigid K -spaces Y .

Corollary 2.60. *For $Y \in \text{Rig}'_K$, the morphism comp_Y is the initial morphism from Y to a uniformly rigid K -space.*

Proof. Let X be a uniformly rigid K -space, and let $\psi : Y \rightarrow X$ be a morphism of locally G -ringed K -spaces. The morphism comp_Y is a bijection on points, and it induces isomorphisms of stalks; hence the morphism $Y^{\text{ur}} \rightarrow X$ that we seek is unique if it exists. If Y is affinoid and X is semiaffinoid, there is nothing to show. Let $(X_i)_{i \in I}$ be an admissible semiaffinoid covering of X , and let $(Y_j)_{j \in J}$ be an admissible affinoid covering of Y refining $(\psi^{-1}(X_i))_{i \in I}$. It suffices to see that $(Y_j^{\text{ur}})_{j \in J}$ is an admissible covering of Y^{ur} . By [Bosch 2005, Lemma 2.8/4], there exists a flat quasiparacompact R -model of locally tf type \mathfrak{Y} for Y such that $(Y_j)_{j \in J}$ is induced by an open covering of \mathfrak{Y} . Since $\mathfrak{Y}^{\text{urig}} = Y^{\text{ur}}$, it follows that $(Y_j^{\text{ur}})_{j \in J}$ is an admissible covering of Y^{ur} , as desired. \square

Corollary 2.61. *The functor ur is fully faithful.*

Proof. Let X and Y be objects in Rig'_K . By Proposition 2.59, by the global variant of Proposition 2.55 and by Corollary 2.60, we have functorial bijections

$$\text{Hom}(Y, X) \cong \text{Hom}(Y, (X^{\text{ur}})^r) \cong \text{Hom}(Y, X^{\text{ur}}) \cong \text{Hom}(Y^{\text{ur}}, X^{\text{ur}}). \quad \square$$

Of course, if X is any uniformly rigid K -space, then the comparison morphism

$$\text{comp}_X : X^r \rightarrow X$$

is *not* initial all morphisms from X^r to uniformly rigid K -spaces. For example, if X is the semiaffinoid open unit disc $\text{sSp}(R[[S]] \otimes_R K)$, then the natural morphism comp_{X^r} from the rigid open unit disc X^r to its uniform rigidification $(X^r)^{\text{ur}}$ does not extend to a morphism $X \rightarrow (X^r)^{\text{ur}}$. Indeed, such a morphism would have to be the identity on points, but X is quasicompact, while $(X^r)^{\text{ur}}$ is not quasicompact.

The functor $Y \mapsto Y^{\text{ur}}$ does *not* respect arbitrary open immersions. For example, if $Y' \subseteq Y$ is the inclusion of the open rigid unit disc into the closed rigid unit disc, the morphism $(Y')^{\text{ur}} \rightarrow Y^{\text{ur}}$ is not an open immersion: its image is the semiaffinoid

open unit disc, while $(Y')^{\text{ur}}$ is not quasicompact. However, it follows from [Bosch and Lütkebohmert 1993b, 5.7] that ur preserves open immersions of *quasicompact* rigid K -spaces.

Quasiseparated rigid K -spaces are obtained from affinoid K -spaces by gluing along *quasicompact* admissible open subspaces, it thus follows that ur preserves fibered products. Indeed, this can now be checked in an affinoid situation, where the statement is clear from the construction of semiaffinoid fibered products. In particular, $Y \mapsto Y^{\text{ur}}$ preserves group structures.

3. Coherent modules on uniformly rigid spaces

Let X be a G -ringed K -space, and let \mathcal{F} be an \mathbb{O}_X -module. Let us recall some standard definitions concerning the coherence property [Bosch 2005, 1.14/2] :

- (i) \mathcal{F} is called of *finite type* if there exists an admissible covering $(X_i)_{i \in I}$ of X together with exact sequences

$$\mathbb{O}_X^{s_i}|_{X_i} \rightarrow \mathcal{F}|_{X_i} \rightarrow 0.$$

- (ii) \mathcal{F} is called *coherent* if \mathcal{F} is of finite type and if for any admissible open subspace $U \subseteq X$, the kernel of any morphism $\mathbb{O}_X^s|_U \rightarrow \mathcal{F}|_U$ is of finite type.

If X is a semiaffinoid K -space with ring of functions A , then the functor $M \mapsto \tilde{M}$ on the category of finite A -modules is well-behaved, as it is shown by the following lemma. The proof of Lemma 3.1 is identical to the proof of [Bosch 2005, 1.14/1]; one uses the fact that the restriction homomorphisms associated to semiaffinoid subdomains are flat:

Lemma 3.1. *The functor $M \mapsto \tilde{M}$ from the category of finite A -modules to the category of \mathbb{O}_X -modules is fully faithful, and it commutes with the formation of kernels, images, cokernels and tensor products. Moreover, a sequence of finite A -modules*

$$0 \rightarrow M' \rightarrow M \rightarrow M'' \rightarrow 0$$

is exact if and only if the associated sequence

$$0 \rightarrow \tilde{M}' \rightarrow \tilde{M} \rightarrow \tilde{M}'' \rightarrow 0$$

of \mathbb{O}_X -modules is exact.

For a semiaffinoid K -space $X = \text{sSp } A$, we have $\mathbb{O}_X^r = A^r \otimes \mathbb{O}_X$. Since A is *noetherian*, it follows from Lemma 3.1 that kernels and cokernels of morphisms of type $\mathbb{O}_X^r \rightarrow \mathbb{O}_X^s$ are associated. We thus conclude that an \mathbb{O}_X -module on a uniformly rigid K -space X is coherent if and only if there exists an admissible semiaffinoid covering $(X_i)_{i \in I}$ of X such that $\mathcal{F}|_{X_i}$ is associated for all $i \in I$.

In particular, the structural sheaf \mathcal{O}_X of any uniformly rigid K -space X is coherent. Moreover, it follows from Lemma 3.1 that kernels and cokernels of morphisms of coherent \mathcal{O}_X -modules are coherent.

Lemma 3.2. *Let $\varphi: Y \rightarrow X$ be a morphism of uniformly rigid K -spaces, and let \mathcal{F} be a coherent \mathcal{O}_X -module. Then $\varphi^*\mathcal{F}$ is a coherent \mathcal{O}_Y -module.*

Proof. Indeed, we may assume that X and Y are semiaffinoid, with $X = \text{sSp } A$ and $Y = \text{sSp } B$, and that \mathcal{F} is associated to a finite A -module M . Then $\varphi^*\mathcal{F}$ is associated to $M \otimes_A B$, where B is an A -algebra via φ^* . \square

Definition 3.3. Let X be a uniformly rigid K -space. An \mathcal{O}_X -module \mathcal{F} is called *strictly coherent* if for any open semiaffinoid subspace $U \subseteq X$, the restriction $\mathcal{F}|_U$ is an associated module.

For example, the structural sheaf of a uniformly rigid K -space is strictly coherent. Since we do not know whether an open semiaffinoid subspace of a semiaffinoid K -space is a semiaffinoid subdomain, it is not a priori clear whether any associated module on a semiaffinoid K -space is strictly coherent. In Corollary 3.6, however, we will show that this is indeed the case.

Let X be a uniformly rigid K -space. We will be interested in coherent \mathcal{O}_X -modules \mathcal{F} with the property that there exists an injective \mathcal{O}_X -homomorphism $\mathcal{F} \hookrightarrow \mathcal{O}_X^r$ for some $r \in \mathbb{N}$. This property is clearly satisfied by coherent ideals, and it is preserved under pullback with respect to flat morphisms of uniformly rigid spaces. We will study integral models of such \mathcal{F} , and we will show that any such \mathcal{F} is strictly coherent.

If \mathfrak{X} is a formal R -scheme of locally ff type and if $\underline{\mathcal{F}}$ is a coherent $\mathcal{O}_{\mathfrak{X}}$ -module, we obtain a coherent \mathcal{O}_X -module $\underline{\mathcal{F}}^{\text{urig}}$ on $\mathfrak{X}^{\text{urig}}$ which we call the *uniformly rigid generic fiber* of $\underline{\mathcal{F}}$. If X is a uniformly rigid K -space, if \mathcal{F} is a coherent \mathcal{O}_X -module and if \mathfrak{X} is a flat formal R -model of locally ff type for X , then an R -model of \mathcal{F} on \mathfrak{X} is a coherent $\mathcal{O}_{\mathfrak{X}}$ -module $\underline{\mathcal{F}}$ together with an isomorphism $\underline{\mathcal{F}}^{\text{urig}} \cong \mathcal{F}$ that is compatible with the given identification $\mathfrak{X}^{\text{urig}} \cong X$. Sometimes we will not mention the isomorphism $\underline{\mathcal{F}}^{\text{urig}} \cong \mathcal{F}$ explicitly. Clearly

$$\text{sp}_{\mathfrak{X},*}(\underline{\mathcal{F}}) = \underline{\mathcal{F}} \otimes_R K,$$

and urig factors naturally through the functor $\underline{\mathcal{F}} \mapsto \underline{\mathcal{F}} \otimes_R K$. Let us abbreviate $\underline{\mathcal{F}}_K := \underline{\mathcal{F}} \otimes_R K$.

For any $r \in \mathbb{N}$, the coherent \mathcal{O}_X -module \mathcal{O}_X^r admits the natural model $\mathcal{O}_{\mathfrak{X}}^r$ on every flat formal R -model of locally ff type \mathfrak{X} for X . We will show that coherent submodules $\mathcal{F} \subseteq \mathcal{O}_X^r$ inherit this property by taking schematic closures. Let us first consider the affine situation:

Lemma 3.4. *Let A be an R -algebra, let M be an A -module, and let $N \subseteq M \otimes_R K$ be an $A \otimes_R K$ -submodule. Then there exists a unique A -submodule $\underline{N} \subseteq M$ such*

that the natural homomorphism $\underline{N} \otimes_R K \rightarrow \underline{M} \otimes_R K$ is an isomorphism onto N and such that $\underline{M}/\underline{N}$ is R -flat.

Proof. Let us abbreviate $M := \underline{M} \otimes_R K$, and let us set

$$\underline{N} := \ker(\underline{M} \rightarrow M/N);$$

then \underline{N} is an \underline{A} -submodule of \underline{M} . For any $n \in N$, there exists an $s \in \mathbb{N}$ such that $\pi^s n$ lies in the image of \underline{M} in M ; the natural K -homomorphism $\underline{N} \otimes_R K \rightarrow N$ is thus bijective. As an \underline{A} -submodule of M/N , the quotient $\underline{M}/\underline{N}$ is free of π -torsion and, hence, R -flat.

If $\underline{N}' \subseteq \underline{M}$ is another \underline{A} -submodule whose image in M generates N as an $\underline{A} \otimes_R K$ -module, then \underline{N}' lies in the kernel \underline{N} of $\underline{M} \rightarrow M/N$. If in addition $\underline{M}/\underline{N}'$ is flat over R , then the natural homomorphism $\underline{M}/\underline{N}' \rightarrow \underline{M}/\underline{N}' \otimes_R K = M/N$ is injective, which proves that \underline{N}' coincides with this kernel. \square

Theorem 3.5. *Let X be a uniformly rigid K -space, let $\mathcal{F}' \subseteq \mathcal{F}$ be an inclusion of coherent \mathbb{O}_X -modules, and let \mathfrak{X} be an R -model of locally ff type for X such that \mathcal{F} admits be an R -model $\underline{\mathcal{F}}$ on \mathfrak{X} . Then there exists a unique coherent $\mathbb{O}_{\mathfrak{X}}$ -submodule $\underline{\mathcal{F}}' \subseteq \underline{\mathcal{F}}$ such that $\underline{\mathcal{F}}/\underline{\mathcal{F}}'$ is R -flat and such that the given isomorphism $\underline{\mathcal{F}}^{\text{urig}} \cong \underline{\mathcal{F}}$ identifies $(\underline{\mathcal{F}}')^{\text{urig}}$ with $\underline{\mathcal{F}}'$.*

Proof. We may work locally on \mathfrak{X} and thereby assume that \mathfrak{X} is affine. Uniqueness of $\underline{\mathcal{F}}'$ is a consequence of Lemma 3.4. Since \mathcal{F}' is coherent, there exists a treelike covering $(X_i)_{i \in I}$ of X such that $\mathcal{F}'|_{X_i}$ is associated for all $i \in \text{lv}(I)$. Let us choose a model of this covering, that is,

- (i) for each $i \in I$, an affine flat R -model of ff type \mathfrak{X}_i for X_i ,
- (ii) for each inner $i \in I$ an admissible blowup $\beta_i : \mathfrak{X}'_i \rightarrow \mathfrak{X}_i$ and
- (iii) for each inner $i \in I$ and for each child j of i an open immersion $\varphi_j : \mathfrak{X}_j \hookrightarrow \mathfrak{X}'_i$ such that $\mathfrak{X}_j \subseteq \mathfrak{X}'_i \rightarrow \mathfrak{X}_i$ represents X_j in X_i .

For each $i \in I$, we let $\underline{\mathcal{F}}|_{\mathfrak{X}_i}$ denote the pullback of $\underline{\mathcal{F}}$ to \mathfrak{X}_i , and for each inner vertex $i \in I$, we let $\underline{\mathcal{F}}|_{\mathfrak{X}'_i}$ denote the pullback of $\underline{\mathcal{F}}$ to \mathfrak{X}'_i . Let i be an inner vertex of I , and let us *assume* that for each child j of i , we are given a coherent submodule

$$\underline{\mathcal{F}}'_j \subseteq \underline{\mathcal{F}}|_{\mathfrak{X}_j}$$

such that $\underline{\mathcal{F}}|_{\mathfrak{X}_j}/\underline{\mathcal{F}}'_j$ is R -flat and such that $(\underline{\mathcal{F}}'_j)^{\text{urig}} = \mathcal{F}'|_{X_j}$. By Lemma 3.4, this assumption is satisfied if all children of i are leaves in I . Using the uniqueness assertion in Lemma 3.4, we see that the $\underline{\mathcal{F}}'_j$ glue to a unique coherent submodule

$$\underline{\mathcal{G}}_i \subseteq \underline{\mathcal{F}}|_{\mathfrak{X}'_i}.$$

The quotient $\underline{\mathcal{F}}|_{\mathfrak{X}'_i}/\underline{\mathcal{G}}_i$ is R -flat; by [Grothendieck 1961b, 3.4.2], $\beta_{i*}(\underline{\mathcal{F}}|_{\mathfrak{X}'_i}/\underline{\mathcal{G}}_i)$ thus is a coherent R -flat $\mathbb{O}_{\mathfrak{X}_i}$ -module. By definition, $\underline{\mathcal{F}}|_{\mathfrak{X}'_i} = \beta_i^* \underline{\mathcal{F}}|_{\mathfrak{X}_i}$, so we have a natural

homomorphism of coherent $\mathbb{O}_{\mathfrak{X}_j}$ -modules

$$\underline{\mathcal{F}}|_{\mathfrak{X}_i} \rightarrow \beta_{i*}\underline{\mathcal{F}}|_{\mathfrak{X}'_i} \rightarrow \beta_{i*}(\underline{\mathcal{F}}|_{\mathfrak{X}'_i}/\underline{\mathcal{G}}_i).$$

Let $\underline{\mathcal{F}}'_i$ denote its kernel; the resulting exact sequence of coherent $\mathbb{O}_{\mathfrak{X}_i}$ -modules

$$0 \rightarrow \underline{\mathcal{F}}'_i \rightarrow \underline{\mathcal{F}}|_{\mathfrak{X}_i} \rightarrow \beta_{i*}(\underline{\mathcal{F}}|_{\mathfrak{X}'_i}/\underline{\mathcal{G}}_i)$$

shows that $\underline{\mathcal{F}}|_{\mathfrak{X}_i}/\underline{\mathcal{F}}'_i$ is R -flat. We claim that the coherent X_i -module $\mathcal{F}'|_{X_i} = \underline{\mathcal{G}}_i^{\text{urig}}$ is associated to $\underline{\mathcal{F}}'_i$. To prove this, it suffices to show that the morphism

$$(\beta_i^*\underline{\mathcal{F}}'_i)_K \rightarrow (\beta_i^*\beta_{i*}\underline{\mathcal{G}}_i)_K \rightarrow \underline{\mathcal{G}}_{i,K} \tag{\ddagger}$$

induced by the natural morphism $\underline{\mathcal{F}}'_i \rightarrow \beta_{i*}\underline{\mathcal{G}}_i$ is an isomorphism. By the ff type variant of [Lütkebohmert 1990, 2.1], the second morphism in (\ddagger) is an isomorphism, so we must show that the first morphism is an isomorphism as well. Let \underline{X}_i be the spectrum of the ring of global functions on \mathfrak{X}_i , and let $b_i: \underline{X}'_i \rightarrow \underline{X}_i$ be the admissible blowup such that $\beta_i = b_i^\wedge$, where we use a wedge to denote the formal completion with respect to an ideal of definition of \mathfrak{X} . Let $\underline{F}_i, \underline{F}'_i$ and \underline{G}_i denote the algebraizations of $\underline{\mathcal{F}}|_{\mathfrak{X}_i}, \underline{\mathcal{F}}'_i$ and $\underline{\mathcal{G}}_i$ respectively, which exist by [Grothendieck 1961b, 5.1.4]; then

$$\underline{\mathcal{F}}|_{\mathfrak{X}_i} = (b_i^*\underline{F}_i)^\wedge.$$

By [Grothendieck 1961b, 4.1.5],

$$\beta_{i*}(\underline{\mathcal{F}}|_{\mathfrak{X}'_i}/\underline{\mathcal{G}}_i) = (b_{i*}((b_i^*\underline{F}_i)/\underline{G}_i))^\wedge,$$

so we have a short exact sequence

$$0 \rightarrow \underline{F}'_i \rightarrow \underline{F}_i \rightarrow b_{i*}((b_i^*\underline{F}_i)/\underline{G}_i)$$

which under $\cdot \otimes_R K$ induces a short exact sequence

$$0 \rightarrow \underline{F}'_{i,K} \rightarrow \underline{F}_{i,K} \rightarrow (b_{i,K})_*((b_{i,K}^*\underline{F}_{i,K})/\underline{G}_{i,K}).$$

Since $b_{i,K}$ is an isomorphism and, hence, flat, we obtain an induced short exact sequence

$$0 \rightarrow b_{i,K}^*\underline{F}'_{i,K} \rightarrow b_{i,K}^*\underline{F}_{i,K} \rightarrow b_{i,K}^*(b_{i,K})_*((b_{i,K}^*\underline{F}_{i,K})/\underline{G}_{i,K});$$

since $b_{i,K}^*(b_{i,K})_*$ is naturally isomorphic to the identity functor, this shows that $b_{i,K}^*\underline{F}'_{i,K} = \underline{G}_{i,K}$. Hence, the natural morphism

$$b_i^*\underline{F}'_j \rightarrow b_i^*b_{i*}\underline{G}_j$$

becomes an isomorphism under $\cdot \otimes_R K$. That is, its kernel and cokernel are π -torsion. It follows that kernel and cokernel of the completed morphism

$$\beta_i^* \underline{\mathcal{F}}'_i \rightarrow \beta_i^* \beta_{i*} \underline{\mathcal{G}}_i$$

are π -torsion as well, which yields our claim.

Let us now prove the statement of the proposition by induction on the volume $v(I)$ of I . We may assume that I has more than one vertex. Let j be a leaf of I whose path to the root has maximal length, and let i be the parent of j . Then all children of i are leaves of I , so the assumption in the argument above is satisfied. By what we have shown so far, $\underline{\mathcal{F}}'|_{X_i}$ is associated to a unique coherent $\mathbb{O}_{\mathfrak{X}_i}$ -submodule $\underline{\mathcal{F}}'_i \subseteq \underline{\mathcal{F}}|_{\mathfrak{X}_i}$ such that $\underline{\mathcal{F}}|_{\mathfrak{X}_i}/\underline{\mathcal{F}}'_i$ is R -flat. We may thus replace $\text{subt}(i)$ by $\{i\}$. By induction on $v(I)$, the desired statement follows. \square

Corollary 3.6. *We conclude:*

- (i) *A coherent submodule of an associated module on a semiaffinoid K -space is associated.*
- (ii) *Coherent submodules and coherent quotients of strictly coherent modules are strictly coherent.*
- (iii) *An associated module on a semiaffinoid K -space is strictly coherent.*

Proof. Let us first show (i). Let $X = \text{sSp } A$ be a semiaffinoid K -space, let $\underline{A} \subseteq A$ be an R -model of ff type, and let $\underline{\mathcal{F}}'$ be a coherent submodule of an associated module \underline{M} . Since \underline{M} admits a model \underline{M} over $\text{Spf } \underline{A}$, Theorem 3.5 implies that $\underline{\mathcal{F}}' \cong (\underline{\mathcal{F}}')^{\text{urig}}$ for a coherent module $\underline{\mathcal{F}}'$ on $\text{Spf } \underline{A}$. Since coherent modules on affine formal schemes are associated, it follows that $\underline{\mathcal{F}}'$ is associated.

Let us prove statement (ii). Let X be a uniformly rigid K -space, let $\underline{\mathcal{F}}$ be a strictly coherent \mathbb{O}_X -module and let $\underline{\mathcal{F}}' \subseteq \underline{\mathcal{F}}$ be a coherent submodule. For every open semiaffinoid subspace $U \subseteq X$, the restriction $\underline{\mathcal{F}}'|_U$ is a coherent submodule of $\underline{\mathcal{F}}|_U$, and $\underline{\mathcal{F}}|_U$ is associated by assumption on $\underline{\mathcal{F}}$. It follows from (i) that $\underline{\mathcal{F}}'|_U$ is associated; hence $\underline{\mathcal{F}}'$ is strictly coherent. Let now $\underline{\mathcal{F}}''$ be a coherent quotient of $\underline{\mathcal{F}}$. Then the kernel $\underline{\mathcal{F}}'$ of the projection $\underline{\mathcal{F}} \rightarrow \underline{\mathcal{F}}''$ is a coherent submodule of $\underline{\mathcal{F}}$ and, hence, strictly coherent by what we have seen so far. Let $U \subseteq X$ be an open semiaffinoid subspace; then we have a short exact sequence

$$0 \rightarrow \underline{\mathcal{F}}'|_U \rightarrow \underline{\mathcal{F}}|_U \rightarrow \underline{\mathcal{F}}''|_U \rightarrow 0$$

where the first two modules are associated. It follows from Lemma 3.1 that $\underline{\mathcal{F}}''|_U$ is associated as well.

Finally, statement (iii) follows from statement (ii) because by Lemma 3.1, an associated module is a quotient of a finite power of the structural sheaf. \square

If \mathfrak{X} is a flat formal R -scheme of locally ff type and if $\underline{\mathcal{F}}$ is a coherent $\mathbb{O}_{\mathfrak{X}}$ -module, we do not know in general whether $\underline{\mathcal{F}}^{\text{urig}}$ is strictly coherent. In particular, we

unfortunately do not know whether the analog of Kiehl’s theorem [Kiehl 1967, 1.2] holds in general, that is to say whether every coherent module on a semiaffinoid K -space is associated. Let us point out that the analogous question for *quasicoherent* modules on *rigid* spaces was open for a long time; it was finally settled in the negative by O. Gabber [Conrad 2006, Example 2.1.6].

Conjecture 3.7. The general uniformly rigid analog of Kiehl’s theorem does not hold.

Remark 3.8. The general uniformly rigid analog of Kiehl’s theorem is equivalent to the following statement: let \mathfrak{X} be an admissible blowup of a flat affine formal R -scheme of ff type, and let \mathcal{F} be a coherent sheaf on $X = \mathfrak{X}^{\text{unig}}$ that admits flat models $\underline{\mathcal{F}}_i$ locally with respect to an affine open covering $(\mathfrak{X}_i)_{i \in I}$ of \mathfrak{X} ; then \mathcal{F} admits a model on \mathfrak{X} . Indeed, this equivalence follows by arguing as in the proof of [Lütkebohmert 1990, Theorem 2.3]. However, it seems impossible in general to modify the models $\underline{\mathcal{F}}_i$ such that they glue to a model of \mathcal{F} on \mathfrak{X} : Let us assume that $I = \{1, 2\}$. After multiplying $\underline{\mathcal{F}}_1$ by a suitable power of π , we may assume that $\underline{\mathcal{F}}_1$ is contained in $\underline{\mathcal{F}}_2$ on the intersection \mathfrak{X}_{12} of \mathfrak{X}_1 and \mathfrak{X}_2 . Let $n \in \mathbb{N}$ be big enough such that $\pi^n \underline{\mathcal{F}}_2 \subseteq \underline{\mathcal{F}}_1$ on \mathfrak{X}_{12} ; then $\underline{\mathcal{G}} := \underline{\mathcal{F}}_1|_{\mathfrak{X}_{12}}/\pi^n \underline{\mathcal{F}}_2|_{\mathfrak{X}_{12}}$ is a coherent subsheaf of $(\underline{\mathcal{F}}_2/\pi^n \underline{\mathcal{F}}_2)|_{\mathfrak{X}_{12}}$; see the proof of [Lütkebohmert 1990, Lemma 2.2]. If \mathfrak{X} is of tf type over R , then the closed formal subscheme of \mathfrak{X} cut out by π^n is a scheme, and by chasing denominators [Grothendieck and Dieudonné 1960, 9.4.7] one can extend $\underline{\mathcal{G}}$ to a coherent subsheaf, again denoted by $\underline{\mathcal{G}}$, on all of \mathfrak{X}_2 . Let $\underline{\mathcal{F}}'_2$ denote the preimage of $\underline{\mathcal{G}}$ under the projection $\underline{\mathcal{F}}_2 \rightarrow \underline{\mathcal{F}}_2/\pi^n \underline{\mathcal{F}}_2$; then $\underline{\mathcal{F}}'_2$ is a model of \mathcal{F} on \mathfrak{X}_2 which glues to $\underline{\mathcal{F}}_1$, and we obtain a model of \mathcal{F} on all of \mathfrak{X} . In our situation, however, \mathfrak{X}_2 might not be of tf type, and hence the closed formal subscheme of \mathfrak{X}_2 cut out by π^n might not be a scheme. On a formal scheme though it is in general not possible to extend coherent subsheaves because of convergence problems. Thus, Lütkebohmert’s proof of Kiehl’s theorem fails in the uniformly rigid situation. Similar problems occur if one tries to carry over Kiehl’s original proof.

3A. Closed uniformly rigid subspaces.

Definition 3.9. A morphism of uniformly rigid K -spaces $\varphi: Y \rightarrow X$ is called a *closed immersion* if there exists an admissible semiaffinoid covering $(X_i)_{i \in I}$ of X such that for each $i \in I$, the restriction $\varphi^{-1}(X_i) \rightarrow X_i$ of φ is a closed immersion of semiaffinoid K -spaces in the sense of Definition 2.19.

We easily see that closed immersions are injective on the level of physical points.

Lemma 3.10. *Let $\varphi: Y \rightarrow X$ be a closed immersion of uniformly rigid K -spaces. Then $\varphi^\sharp: \mathbb{O}_X \rightarrow \varphi_* \mathbb{O}_Y$ is an epimorphism of sheaves. Moreover, the \mathbb{O}_X -modules $\varphi_* \mathbb{O}_Y$ and $\ker \varphi^\sharp$ are strictly coherent.*

Proof. The \mathbb{O}_X -module \mathbb{O}_X is strictly coherent. By Corollary 3.6(ii), it thus suffices to show that φ^\sharp is an epimorphism and that both $\ker \varphi^\sharp$ and $\varphi_*\mathbb{O}_Y$ are coherent. Considering an admissible semiaffinoid covering $(X_i)_{i \in I}$ of X such that for all $i \in I$, the restriction $\varphi^{-1}(X_i) \rightarrow X_i$ of φ is a closed immersion of semiaffinoid K -spaces, we reduce to the case where both X and Y are semiaffinoid and where φ corresponds to a surjective homomorphism of semiaffinoid K -algebras. Now the desired statements follow from Lemma 3.1. \square

Proposition 3.11. *Let $\varphi: Y \rightarrow X$ be a morphism of uniformly rigid K -spaces. Then the following are equivalent:*

- (i) φ is a closed immersion.
- (ii) For each open semiaffinoid subspace $U \subseteq X$, the restriction $\varphi^{-1}(U) \rightarrow U$ is a closed immersion of semiaffinoid K -spaces in the sense of Definition 2.19.

Proof. The implication (ii) \Rightarrow (i) is trivial, the open semiaffinoid subspaces forming a basis for the G -topology on X . Let us assume that (i) holds, let \mathcal{I} denote the kernel of φ^\sharp , and let $U \subseteq X$ be an open semiaffinoid subspace; then φ induces a short exact sequence

$$0 \rightarrow \mathcal{I}|_U \rightarrow \mathbb{O}_U \rightarrow \varphi_*\mathbb{O}_Y|_U \rightarrow 0.$$

Let A denote the ring of functions on U . By Lemma 3.10, \mathcal{I} and $\varphi_*\mathbb{O}_Y$ are strictly coherent; hence the above short exact sequence is associated to a short exact sequence of A -modules

$$0 \rightarrow I \rightarrow A \rightarrow B \rightarrow 0.$$

Since morphisms from uniformly rigid K -spaces to semiaffinoid K -spaces correspond to K -homomorphisms of rings of global functions, we can now mimic the proof of [Bosch et al. 1984, 9.4.4/1] to see that the restriction $\varphi^{-1}(U) \rightarrow U$ of φ is associated to the projection $A \rightarrow B$: it suffices to see that the natural morphism $\varphi^{-1}(U) \rightarrow \text{sSp } B$ is an isomorphism. This can be checked locally on $\text{sSp } B$ with respect to the preimage under $\text{sSp } B \rightarrow U$ of a leaflike refinement of $(U \cap X_i)_{i \in I}$, where $(X_i)_{i \in I}$ is an admissible semiaffinoid covering of X satisfying the conditions of Definition 3.9. \square

Remark 3.12. The proof of [Bosch et al. 1984, 9.4.4/1] resorts to [ibid., 8.2.1/4]. However, as our argument above shows, this is in fact unnecessary — which is to our advantage, because the statement of 8.2.1/4 fails to hold in the semiaffinoid situation: Example 2.42 yields a bijective morphism of semiaffinoid K -spaces which induces isomorphisms of stalks and which is not an isomorphism.

In particular, a morphism of semiaffinoid K -spaces is a closed immersion in the sense of Definition 3.9 if and only if it is a closed immersion of semiaffinoid K -spaces in the sense of Definition 2.19. We can now define a *closed uniformly*

rigid subspace as an equivalence class of closed immersions, in the usual way. By standard gluing arguments, we see that the closed uniformly rigid subspaces of a uniformly rigid K -space X correspond to the coherent \mathbb{C}_X -ideals. We easily see that closed immersions of uniformly rigid K -spaces are preserved under base change.

It is clear that closed immersions of formal R -schemes of locally ff type induce closed immersions on uniformly rigid generic fibers. Conversely, given a uniformly rigid K -space X together with an R -model of locally ff type \mathfrak{X} and a closed uniformly rigid subspace $V \subseteq X$, there exists a unique R -flat closed formal subscheme $\mathfrak{V} \subseteq \mathfrak{X}$ such that the given isomorphism $\mathfrak{X}^{\text{urig}} \cong X$ identifies $\mathfrak{V}^{\text{urig}}$ with V . Indeed, this is an immediate consequence of Theorem 3.5. We say that \mathfrak{V} is the *schematic closure* of V in \mathfrak{X} .

The comparison functors studied in Section 2D1 preserve closed immersions. This can be verified in the semiaffinoid and affinoid situations respectively. In the case of the functor ur , there is nothing to show. In the case of the functor r , the statement follows by looking at schematic closures and using the fact that Berthelot’s construction preserves closed immersions [de Jong 1995, 7.2.4(e)].

3A1. Separated uniformly rigid spaces. As usual, a morphism $\varphi: Y \rightarrow X$ of uniformly rigid K -spaces is called *separated* if its *diagonal morphism*

$$\Delta_\varphi: Y \rightarrow Y \times_X Y$$

is a closed immersion. A uniformly rigid K -space X is called *separated* if its structural morphism $X \rightarrow \text{sSp } K$ is separated. If X is a uniformly rigid K -space, we let Δ_X denote the diagonal of its structural morphism.

Semiaffinoid K -spaces are visibly separated. Moreover, uniformly rigid generic fibers of separated morphisms of formal R -schemes of locally ff type are separated, since functor urig preserves fibered products and closed immersions. Similarly, the comparison functors studied in Section 2D1 preserve the separatedness property.

Lemma 3.13. *Let X be a separated uniformly rigid K -space. The intersection of two open semiaffinoid subspaces in X is an open semiaffinoid subspace in X .*

Proof. Let U and V be open semiaffinoid subspaces in X . We easily see, using points with values in finite field extensions of K , that $U \cap V$ is the Δ_X -preimage of $U \times_{\text{sSp } K} V$ which is an open semiaffinoid subspace of $X \times_{\text{sSp } K} X$. Since Δ_X is a closed immersion by assumption on X , it follows from Proposition 3.11 that $U \cap V$ is an open semiaffinoid subspace of X . □

Corollary 3.14. *Let X be a separated uniformly rigid K -space, and let \mathcal{F} be a coherent \mathbb{C}_X -module. Then the natural morphism*

$$\check{H}^q(X, \mathcal{F}) \xrightarrow{\sim} H^q(X, \mathcal{F})$$

is an isomorphism for all $q \geq 0$.

Proof. Let S denote the set of open semiaffinoid subspaces U in X with the property that $\mathcal{F}|_U$ is associated. By Lemma 3.13, this set is stable under the formation of intersections. It is clearly a basis for the G-topology on X , and $\check{H}^q(U, \mathcal{F}) = 0$ for any U in S and any $q \geq 0$ by Corollary 2.43. We conclude by the usual Čech spectral sequence argument. \square

If X is a separated uniformly rigid K -space and if $\varphi: Y \rightarrow X$ is a morphism of uniformly rigid K -spaces, then the graph $\Gamma_\varphi: Y \rightarrow Y \times X$ of φ is a closed immersion since it is obtained from Δ_X via pullback. In particular, if \mathfrak{X} and \mathfrak{Y} are R -models of locally ff type for X and Y respectively, the schematic closure of Γ_φ in $\mathfrak{Y} \times \mathfrak{X}$ is well-defined. Here fibered products without indication of the base are understood over $\text{sSp } K$ or $\text{Spf } R$ respectively.

4. Comparison with the theories of Berkovich and Huber

The category of formal R -schemes of locally ff type is a full subcategory of Huber’s category of adic spaces [Huber 1996]. If \mathfrak{X} is a formal R -scheme of locally ff type, viewed as an adic space, then by [Huber 1996, 1.2.2] the fibered product $\mathfrak{X} \times_{\text{Spa}(R,R)} \text{Spa}(K, R)$ is the adic space associated to the rigid generic fiber $\mathfrak{X}^{\text{rig}}$ of \mathfrak{X} . That is, the uniform structure induced by \mathfrak{X} is lost. In fact, we do not see a way to view the category of uniformly rigid spaces as a full subcategory of Huber’s category of adic spaces. The main obstacle lies in the fact that if \underline{A} is an R -algebra of ff type, equipped with its natural Jacobson-adic topology, and if $A = \underline{A} \otimes_R K$, then the pair (A, \underline{A}) is in general *not* an f-adic ring in the sense of [Huber 1996]. For example, for $\underline{A} = R[[S]]$ there exists no ring topology on A such that \underline{A} is open in this topology: There is a unique such group topology, but multiplication by π^{-1} in A is not continuous because there is no $n \in \mathbb{N}$ such that $\pi^{-1} S^n \in R[[S]] \otimes_R K$ is contained in $R[[S]]$.

The situation is different if we consider the π -adic topology on R -algebras of ff type. If \underline{A}^π denotes the ring \underline{A} equipped with its π -adic topology, then the pair (A, \underline{A}^π) is an f-adic ring in the sense of Huber. The induced topology on A is in fact a K -Banach algebra topology; if, for $f \in A$ nonzero, we set $v_{\underline{A}}(f) := \max\{n \in \mathbb{N}; \pi^{-n} f \in \underline{A}\}$, then $|f|_{\underline{A}} := |\pi|^{v_{\underline{A}}(f)}$ defines a K -Banach algebra norm on A which induces the topology defined by \underline{A}^π . If $\underline{A} = R[[S]]\langle T \rangle$ is a mixed formal power series ring in finitely many variables, then $|\cdot|_{R[[S]]\langle T \rangle}$ is the Gauss norm, and it coincides with the supremum seminorm taken over all points in $\text{Max } A$. Using [Bosch et al. 1984, 3.7.5/2], one proves that all K -Banach algebra structures on A are equivalent; in particular, the valuation spectrum $M(A)$ in the sense of [Berkovich 1990, 1.2] is well defined. One shows that reduced semiaffinoid K -algebras are Banach function algebras, and one verifies that the supremum seminorm, taken over all points in $\text{Max } A$ or, equivalently, over all

points in $M(A)$, takes values in $\sqrt{|K|}$. For a more detailed discussion, including proofs, we refer to [Kappen 2009, Section 1.2.5].

The topological space $M(A)$ may be viewed as a compactification of the rigid space $(\text{sSp } A)^r$. To illustrate this idea in terms of an example, let us first explain how the specialization map extends to valuation spectra. If A is a semiaffinoid K -algebra and if \underline{A} is an R -model of ff type for A , there exists a natural specialization map

$$\text{sp}_{\underline{A}} : M(A) \rightarrow \text{Spec}(\underline{A}/\pi \underline{A})$$

extending the specialization map which we discussed in Section 2B1: let x be a point in $M(A)$, represented by a character $\chi_x : A \rightarrow \mathcal{K}$ with values in some valued field extension \mathcal{K} of K ; then $\text{sp}_{\underline{A}}(x) := \ker(\tilde{\chi}_x : \underline{A}/\pi \underline{A} \rightarrow \tilde{\mathcal{K}})$, where $\tilde{\mathcal{K}}$ is the residue field of \mathcal{K} and where $\tilde{\chi}_x$ is the reduction of χ_x .

Lemma 4.1. *The map $\text{sp}_{\underline{A}}$ is surjective onto $\text{Spec}(\underline{A}/\pi \underline{A})$. If $\underline{A}/\pi \underline{A}$ is a domain, the residue norm $|\cdot|_{\underline{A}}$ is multiplicative and, hence, defines a point in $M(A)$. This point specializes to the generic point of $\text{Spec}(\underline{A}/\pi \underline{A})$, and it is the only point in $M(A)$ with this property.*

Proof. Surjectivity of $\text{sp}_{\underline{A}}$ follows from [Grothendieck 1961a, 7.1.7]. If $\underline{A}/\pi \underline{A}$ is a domain, then $|\cdot|_{\underline{A}} \in M(A)$ clearly specializes to $\pi \underline{A}$. Moreover, the local ring $\underline{A}_{\pi \underline{A}}$ is then a discrete valuation ring, such that every character χ of a point $x \in M(A)$ specializing to the generic point of $\text{Spec } \underline{A}/\pi \underline{A}$ is equivalent to the character given by the natural homomorphism from A to the fraction field of the π -adic completion of $\underline{A}_{\pi \underline{A}}$. It follows that x equals $|\cdot|_{\underline{A}}$. \square

One can easily verify that when $\underline{A}/\pi \underline{A}$ is a domain, then $\{|\cdot|_{\underline{A}}\}$ is the Shilov boundary of $M(A)$ [Kappen 2009, 1.2.5.12]; we will not use this fact in the following. Let us now discuss the example of the open unit disc $\text{sSp}(R[[S]] \otimes_R K)$:

Example 4.2. The set $M(R[[S]] \otimes_R K)$ is naturally identified with the closure of the Berkovich open unit disc within $M(K\langle S \rangle)$, which is obtained by adding the Gauss point.

Proof. To understand the continuous map $i : M(R[[S]] \otimes_R K) \rightarrow M(K\langle S \rangle)$ induced by the natural isometry $K\langle S \rangle \hookrightarrow R[[S]] \otimes_R K$, we distinguish the points in $M(R[[S]] \otimes_R K)$ with respect to their specializations to the scheme $\text{Spec } k[[S]]$. Applying Lemma 4.1 to $\underline{A} = R[[S]]$, we see that the unique point above the generic point of $\text{Spec } k[[S]]$ is the Gauss point $|\cdot|_{\text{Gauss}}$, which maps to the Gauss point in $M(K\langle S \rangle)$ via i . If $x \in M(R[[S]] \otimes_R K)$ is a point specializing to the special point of $\text{Spec } k[[S]]$, then for any character χ_x representing x , the induced R -homomorphism $\hat{\chi}_x : R[[S]] \rightarrow \hat{\mathcal{K}}$ is continuous for the (π, S) -adic topology on $R[[S]]$ and the valuation topology on $\hat{\mathcal{K}}$. In particular, χ_x is determined by the χ_x -image of the variable S . We conclude that the map i is injective and that it maps

the complement of the Gauss point onto the Berkovich open unit disc. The image of i is the continuous image of a compact set and, hence, compact. Since $M(K\langle S \rangle)$ is Hausdorff, it follows that the image of i is closed in $M(K\langle S \rangle)$. \square

Remark 4.3. Given a complete nontrivially valued nonarchimedean field K with valuation ring R , one may wonder whether the points of the rigid open unit disc over K lie dense in $M(R[[S]] \otimes_R K)$; this question is called the one-dimensional nonarchimedean Corona problem. It is yet unanswered; see the introduction of [Deninger 2010] for a brief survey including other versions of nonarchimedean Corona problems. If K is discretely valued (which is the overall assumption in this paper), our discussion of Example 4.2 above shows that the Corona question has a positive answer: indeed, let $Z \subseteq M(R[[S]] \otimes_R K)$ be the closure of the set of classical points; then the image of Z under the natural map i to the K -analytic space $M(K\langle S \rangle)$ is closed. Working locally on $M(K\langle S \rangle)$, we see that $i(Z)$ contains the Berkovich open unit disc and, hence, its closure. We have seen in Example 4.2 that i is injective onto that closure; thus it follows that $Z = M(R[[S]] \otimes_R K)$. The one-dimensional nonarchimedean Corona problem is significantly more challenging when K is not discretely valued: then the ring $R[[S]] \otimes_R K$ is not noetherian, it has maximal ideals of infinite height [van der Put 1974, Corollary 4.9], and it contains functions with infinitely many zeros on the rigid open unit disc.

It is natural to ask whether one can associate a topological space to a uniformly rigid K -space such that, in the semiaffinoid case, one recovers the construction $\text{sSp } A \mapsto M(A)$ which we described above. However, the formation of $M(A)$ does not behave well with respect to localization; see the following example. This is not surprising: the Banach K -algebra structure on A restricts to the π -adic topology on an R -model of ff type \underline{A} for A , and complete localization of \underline{A} with respect to the π -adic topology does in general not agree with complete localization with respect to the topology defined by the Jacobson radical. Similarly, the extended specialization map $\text{sp}_{\underline{A}}$ maps onto the algebraization $\text{Spec}(\underline{A}/\pi \underline{A})$ of the special fiber $\text{Spf}(\underline{A}/\pi \underline{A})$ of $\text{Spf } \underline{A}$ whose formation, again, does in general not commute with localization.

Example 4.4. If $\underline{A} = R\langle X, Y \rangle[[Z]]/(XY - Z)$, equipped with the Jacobson-adic topology, and if $\underline{B} = \underline{A}_{\{X-Y\}}$, then the induced map $M(\underline{B}) \rightarrow M(\underline{A})$ is not injective.

Proof. Let us write $\mathfrak{X} := \text{Spf } \underline{A}$, and let $\mathfrak{X}_0 := \text{Spec } k[X, Y]/(XY)$ denote the smallest subscheme of definition of \mathfrak{X} . Since \mathfrak{X} is formally smooth over R , its special fiber \mathfrak{X}_k is formally smooth over k . The underlying topological space $|\mathfrak{X}_k| = |\mathfrak{X}_0|$ is connected; hence the ring $\underline{A}/\pi \underline{A}$ is a domain. By Lemma 4.1, there exists a unique point $|\cdot|_{\underline{A}}$ of $M(\underline{A})$ specializing to the generic point of the algebraization $\mathfrak{X}_k^\pi := \text{Spec}(\underline{A}/\pi \underline{A})$ of the special fiber $\mathfrak{X}_k = \text{Spf}(\underline{A}/\pi \underline{A})$ of \mathfrak{X} . On the other

hand, let us consider the open formal subscheme $\mathfrak{U} := \text{Spf } \underline{B}$ of \mathfrak{X} . Its underlying smallest subscheme of definition \mathfrak{U}_0 is

$$\mathfrak{U}_0 = \text{Spec } (k[X, Y]/(XY))_{X=Y} = \text{Spec}(k[X, X^{-1}]) \amalg \text{Spec}(k[Y, Y^{-1}]),$$

so \mathfrak{U} has exactly two connected components. We conclude that \underline{B} is a nontrivial direct sum $\underline{B}_1 \oplus \underline{B}_2$ of flat R -algebras of ff type. Since \mathfrak{U} is formally R -smooth, we see that $\underline{B}_i/\pi \underline{B}_i$ is a domain for $i = 1, 2$. We obtain an induced nontrivial decomposition $B = B_1 \oplus B_2$ and, hence, a nontrivial decomposition $M(B) = M(B_1) \amalg M(B_2)$. By the proof of the statement in Example 4.2, there exist unique elements $|\cdot|_{B_i} \in M(B_i)$, $i = 1, 2$, specializing to the respective generic point of $\mathfrak{U}_{i,k}^\pi := \text{Spec } \underline{B}_i/\pi \underline{B}_i$. To prove that the natural map $M(B) \rightarrow M(A)$ is not injective, it suffices to see that it maps the elements $|\cdot|_{B_1}, |\cdot|_{B_2}$ in $M(B)$ to $|\cdot|_A$. By functoriality of the specialization map, it thus suffices to observe that the natural morphism $\mathfrak{U}_{i,k}^\pi \rightarrow \mathfrak{X}_k^\pi$ maps the generic point to the generic point. However, this is clear because $\underline{A}/\pi \underline{A} \rightarrow \underline{B}_i/\pi \underline{B}_i$ is injective. Indeed, it is a flat homomorphism of domains, where flatness follows from the fact that $\mathfrak{U}_{i,k} \rightarrow \mathfrak{X}_k$ is an open immersion of formal schemes. \square

In the light of Example 4.4, it is unclear how to define a global analog of $M(A)$. Nonetheless, we think that a quasicompact uniformly rigid K -space X should be viewed as a compactification of its underlying rigid K -space X^r . This should be made more precise by studying the topos of X .

Acknowledgements

The present paper contains parts of the first chapter of the author's dissertation [Kappen 2009]. He would like to express his gratitude to his thesis advisor Siegfried Bosch. He also thanks Brian Conrad, Ofer Gabber, Ulrich Görtz, Philipp Hartwig, Simon Hüskens, Christian Wahle and the referee for helpful discussions and comments, and he would like to thank the Massachusetts Institute of Technology for its hospitality.

This work was financially supported by the Studienstiftung (German National Academic Foundation), by the Graduiertenkolleg Analytische Topologie und Metageometrie of the University of Münster and by the Hamburger Stiftung für Internationale Forschungs- und Studienvorhaben; the author would like to extend his gratitude to these institutions.

References

- [Berkovich 1990] V. G. Berkovich, *Spectral theory and analytic geometry over non-archimedean fields*, Mathematical Surveys and Monographs **33**, American Mathematical Society, Providence, RI, 1990. MR 91k:32038 Zbl 0715.14013

- [Berkovich 1996] V. G. Berkovich, “Vanishing cycles for formal schemes, II”, *Invent. Math.* **125**:2 (1996), 367–390. MR 98k:14031 Zbl 0852.14002
- [Berthelot 1996] P. Berthelot, “Cohomologie rigide et cohomologie rigide à supports propres”, preprint, Université de Rennes 1, 1996, Available at http://perso.univ-rennes1.fr/pierre.berthelot/publis/Cohomologie_Rigide_I.p%20df.
- [Bosch 2005] S. Bosch, “Lectures on formal and rigid geometry”, preprint, Universität Münster, 2005, Available at <http://www.math.uni-muenster.de/sfb/about/publ/heft378.pdf>.
- [Bosch and Lütkebohmert 1993a] S. Bosch and W. Lütkebohmert, “Formal and rigid geometry, I: Rigid spaces”, *Math. Ann.* **295**:2 (1993), 291–317. MR 94a:11090 Zbl 0808.14017
- [Bosch and Lütkebohmert 1993b] S. Bosch and W. Lütkebohmert, “Formal and rigid geometry, II: Flattening techniques”, *Math. Ann.* **296**:3 (1993), 403–429. MR 94e:11070 Zbl 0808.14018
- [Bosch et al. 1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis: A systematic approach to rigid analytic geometry*, Grundlehren der Mathematischen Wissenschaften **261**, Springer, Berlin, 1984. MR 86b:32031 Zbl 0539.14017
- [Bourbaki 1998] N. Bourbaki, *Commutative algebra: Chapters 1–7*, Elements of Mathematics, Springer, Berlin, 1998. MR 2001g:13001 Zbl 0902.13001
- [Chai 2003] C.-L. Chai, “A bisection of the Artin conductor”, preprint, University of Pennsylvania, 2003, Available at http://www.math.upenn.edu/~chai/papers_pdf/bAcond_v21.pdf.
- [Conrad 1999] B. Conrad, “Irreducible components of rigid spaces”, *Ann. Inst. Fourier (Grenoble)* **49**:2 (1999), 473–541. MR 2001c:14045 Zbl 0928.32011
- [Conrad 2006] B. Conrad, “Relative ampleness in rigid geometry”, *Ann. Inst. Fourier (Grenoble)* **56**:4 (2006), 1049–1126. MR 2007h:14029 Zbl 1125.14009
- [Deninger 2010] C. Deninger, “Invariant functions on p -divisible groups and the p -adic corona problem”, *Tokyo J. Math.* **33**:2 (2010), 393–406. MR 2779265 Zbl 05869972
- [Eisenbud 1995] D. Eisenbud, *Commutative algebra, with a view toward algebraic geometry*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. MR 97a:13001 Zbl 0819.13001
- [Grothendieck 1961a] A. Grothendieck, “Éléments de géométrie algébrique, II: Étude globale élémentaire de quelques classes de morphismes”, *Inst. Hautes Études Sci. Publ. Math.* **8** (1961), 5–222. MR 36 #177b Zbl 0118.36206
- [Grothendieck 1961b] A. Grothendieck, “Éléments de géométrie algébrique, III: Étude cohomologique des faisceaux cohérents, I”, *Inst. Hautes Études Sci. Publ. Math.* **11** (1961), 5–167. MR 36 #177c Zbl 0118.36206
- [Grothendieck 1964] A. Grothendieck, “Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, I”, *Inst. Hautes Études Sci. Publ. Math.* **20** (1964), 5–259. MR 30 #3885 Zbl 0136.15901
- [Grothendieck and Dieudonné 1960] A. Grothendieck and J. Dieudonné, “Éléments de géométrie algébrique, I: Le langage des schémas”, *Inst. Hautes Études Sci. Publ. Math.* **4** (1960), 1–228. MR 36 #177a Zbl 0118.36206
- [Grothendieck and Dieudonné 1971] A. Grothendieck and J. Dieudonné, *Éléments de géométrie algébrique, I*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen **166**, Springer, Berlin, 1971. Zbl 0203.23301
- [Huber 1996] R. Huber, *Étale cohomology of rigid analytic varieties and adic spaces*, Aspects of Mathematics **E30**, Friedr. Vieweg, Braunschweig, 1996. MR 2001c:14046 Zbl 0868.14010
- [Huber 2007] R. Huber, “A finiteness result for the compactly supported cohomology of rigid analytic varieties, II”, *Ann. Inst. Fourier (Grenoble)* **57**:3 (2007), 973–1017. MR 2008i:14037 Zbl 1146.14015

- [de Jong 1995] A. J. de Jong, “Crystalline Dieudonné module theory via formal and rigid geometry”, *Inst. Hautes Études Sci. Publ. Math.* **82** (1995), 5–96. Erratum in **87** (1998), 175. MR 97f:14047 Zbl 0864.14009
- [de Jong 1998] J. de Jong, “Erratum to: ‘Crystalline Dieudonné module theory via formal and rigid geometry’”, *Inst. Hautes Études Sci. Publ. Math.* **87** (1998), 175. MR 99i:14050
- [Kappen 2009] C. Kappen, *Uniformly rigid spaces and Néron models of formally finite type*, Thesis, Universität Münster, 2009. Zbl 1219.14002 arXiv 1003.1022
- [Kiehl 1967] R. Kiehl, “Theorem A und Theorem B in der nichtarchimedischen Funktionentheorie”, *Invent. Math.* **2** (1967), 256–273. MR 35 #1834 Zbl 0202.20201
- [Lipshitz and Robinson 2000] L. Lipshitz and Z. Robinson, *Rings of separated power series and quasi-affinoid geometry*, Astérisque **264**, Société Math. de France, Paris, 2000. MR 2001g:32017 Zbl 0957.32011
- [Lütkebohmert 1990] W. Lütkebohmert, “Formal-algebraic and rigid-analytic geometry”, *Math. Ann.* **286**:1-3 (1990), 341–371. MR 90m:14015 Zbl 0716.32022
- [Nicaise 2009] J. Nicaise, “A trace formula for rigid varieties, and motivic Weil generating series for formal schemes”, *Math. Ann.* **343**:2 (2009), 285–349. MR 2010b:14043 Zbl 1177.14050
- [van der Put 1974] M. van der Put, “The non-archimedean corona problem”, pp. 287–317 in *Table ronde d’analyse non archimédienne* (Paris, 1972), Bull. Soc. Math. France, Mém. **39–40**, Soc. Math. France, Paris, 1974. Supplement to Bull. Soc. Math. France, Mém. 102. MR 51 #939 Zbl 0303.46048
- [Rapoport and Zink 1996] M. Rapoport and T. Zink, *Period spaces for p -divisible groups*, Annals of Mathematics Studies **141**, Princeton University Press, 1996. MR 97f:14023 Zbl 0873.14039
- [Raynaud 1974] M. Raynaud, “Géométrie analytique rigide d’après Tate, Kiehl, ...”, pp. 319–327 in *Table ronde d’analyse non archimédienne* (Paris, 1972), Bull. Soc. Math. France, Mém. **39–40**, Soc. Math. France, Paris, 1974. Supplement to Bull. Soc. Math. France, Mém., 102. MR 57 #10012 Zbl 0299.14003
- [Strauch 2008] M. Strauch, “Deformation spaces of one-dimensional formal modules and their cohomology”, *Adv. Math.* **217**:3 (2008), 889–951. MR 2009a:22014 Zbl 1140.22017
- [Temkin 2008] M. Temkin, “Desingularization of quasi-excellent schemes in characteristic zero”, *Adv. Math.* **219**:2 (2008), 488–522. MR 2009h:14027 Zbl 1146.14009
- [Valabrega 1975] P. Valabrega, “On the excellent property for power series rings over polynomial rings”, *J. Math. Kyoto Univ.* **15**:2 (1975), 387–395. MR 51 #12852 Zbl 0306.13011
- [Valabrega 1976] P. Valabrega, “A few theorems on completion of excellent rings”, *Nagoya Math. J.* **61** (1976), 127–133. MR 53 #10790 Zbl 0319.13008

Communicated by Brian Conrad

Received 2010-09-06

Revised 2011-02-22

Accepted 2011-03-25

christian.kappen@uni-due.de

*Institut für Experimentelle Mathematik, Universität
Duisburg-Essen, Ellernstrasse 29, D-45326 Essen, Germany
<http://esaga.uni-due.de/christian.kappen>*

On a conjecture of Kontsevich and Soibelman

Lê Quy Thuong

Dedicated to Professor Hà Huy Vui on the occasion of his sixtieth birthday

We consider a conjecture of Kontsevich and Soibelman which is regarded as a foundation of their theory of motivic Donaldson–Thomas invariants for noncommutative $3d$ Calabi–Yau varieties. We will show that, in some certain cases, the answer to this conjecture is positive.

1. Introduction

Kontsevich and Soibelman [2008] introduce and give discussions on the motivic Donaldson–Thomas invariants which are defined for noncommutative $3d$ Calabi–Yau varieties and take values in certain Grothendieck groups of algebraic varieties. One of the main objectives of Kontsevich and Soibelman’s paper is to define the motivic Hall algebra which generates Toën’s notion [2006] of the derived Hall algebra. For \mathcal{C} an ind-constructible triangulated A_∞ -category over a field κ , the motivic Hall algebra $H(\mathcal{C})$ is constructed to become a graded associative algebra, which admits for each strict sector V an element A_V^{Hall} invertible in the completed motivic Hall algebra and satisfying the factorization property. It is believed that, in the case of $3d$ Calabi–Yau category, there is a homomorphism Φ of the motivic Hall algebra into the motivic quantum torus defined in terms of the motivic Milnor fiber of the potential. Then the motivic Donaldson–Thomas invariants appear as the collection of the images of A_V^{Hall} under the homomorphism Φ .

In fact, the following conjecture plays a central role in the existence of Φ . Assume that the characteristic of κ is zero. Let F be a formal series on the affine space $\mathbb{A}_\kappa^d = \mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{A}_\kappa^{d_2} \times_\kappa \mathbb{A}_\kappa^{d_3}$, depending in a constructible way on finitely many extra parameters, such that $F(0, 0, 0) = 0$ and F has degree zero with respect to the diagonal action of the multiplicative group $\mathbb{G}_{m,\kappa}$ with the weights $(1, -1, 0)$. In particular, $F(x, 0, 0)$ is the zero function on $\mathbb{A}_\kappa^{d_1}$. We denote by $X_0(F)$ the set of the zeros of F on \mathbb{A}_κ^d . Consider the natural inclusions

MSC2010: primary 14B05; secondary 14B07, 14J17, 32S05, 32S30, 32S55.

Keywords: arc spaces, motivic Milnor fiber, motivic zeta function, Newton polyhedron.

$$i_1 : \mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{G}_{m,\kappa} \rightarrow X_0(F) \times_\kappa \mathbb{G}_{m,\kappa} \quad \text{and} \quad i_0 : \{0\} \times_\kappa \mathbb{G}_{m,\kappa} \rightarrow X_0(F) \times_\kappa \mathbb{G}_{m,\kappa}.$$

Consider the motivic Milnor fiber \mathcal{S}_F of F in the ring $\mathcal{M}_{X_0(F) \times_\kappa \mathbb{G}_{m,\kappa}}^{\mathbb{G}_{m,\kappa}}$, the localization of the relative Grothendieck ring defined in [Guibert et al. 2005; 2006]. Denote by h the function on $\mathbb{A}_\kappa^{d_3}$ defined by $h(z) = F(0, 0, z)$. We write $\mathcal{S}_{h,0}$ for the pullback $i_0^* \mathcal{S}_h$. We denote by the integral $\int_{\mathbb{A}_\kappa^{d_1}}$ the pushforward of the canonical morphism $\pi : \mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{G}_{m,\kappa} \rightarrow \text{Spec}(k) \times_\kappa \mathbb{G}_{m,\kappa}$.

Conjecture 1.1 [Kontsevich and Soibelman 2008]. *With the previous notations and hypotheses, the following formula holds in $\mathcal{M}_{\mathbb{G}_{m,\kappa}}^{\mathbb{G}_{m,\kappa}}$:*

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_F = \mathbb{L}^{d_1} \mathcal{S}_{h,0}.$$

In this paper, we consider the conjecture in some special (actually quite general) cases, namely, when F is a composition of a polynomial in two variables and a pair of two regular functions (Theorem 5.1), or F has the form

$$F(x, y, z) = g(x, y, z) + h(z)^\ell$$

with ℓ sufficiently large (function of Steenbrink type, Theorem 5.6) under some additional conditions of nondegeneracy with respect to its Newton polyhedron (this would be the general case for the conjecture if we did not assume ℓ sufficiently large). For these cases, we use previous results of Guibert, Loeser and Merle [Guibert et al. 2006; 2009] for the motivic Milnor fiber of composite functions or functions of Steenbrink type. We also use in an important way, via Proposition 4.8, the explicit computation of the motivic Milnor fiber of a regular function via its Newton polyhedron (suggested by [Guibert 2002]). These lead to the positive answer to the conjecture in the cases considered.

2. Motivic zeta function and motivic Milnor fiber

Let us recall some basic notations in the theory of motivic integration which will be used in this paper. For references, we follow [Denef and Loeser 1998, 1999a; 2001; Guibert 2002; Guibert et al. 2005; 2006].

2A. Let κ be a field of characteristic zero. For a variety X over κ , we denote by $\mathcal{L}_m(X)$ the space of m -arcs on X , and by $\mathcal{L}(X)$ a limit of the projective system of spaces $\mathcal{L}_m(X)$ and (canonical) morphisms $\mathcal{L}_l(X) \rightarrow \mathcal{L}_m(X)$ ($l \geq m$). In this paper, we use the notation π_m for the canonical morphism $\mathcal{L}(X) \rightarrow \mathcal{L}_m(X)$. The $\mathbb{G}_{m,\kappa}$ -action on $\mathcal{L}_m(X)$ and $\mathcal{L}(X)$ is given by $a \cdot \varphi(t) = \varphi(at)$. The notation \mathcal{M}_X can be found in [Guibert et al. 2006]. As in [Guibert et al. 2005], we denote by $\mathcal{M}_{X \times_\kappa \mathbb{G}_{m,\kappa}}^{\mathbb{G}_{m,\kappa}}$ the localization at \mathbb{L} of the relative Grothendieck ring of $\mathbb{G}_{m,\kappa}$ -equivariant

morphisms $Y \rightarrow X \times_{\kappa} \mathbb{G}_{m,\kappa}$ endowed with a monomial $\mathbb{G}_{m,\kappa}$ -action, where \mathbb{L} is the class of the line bundle $\mathbb{A}^1_{X \times_{\kappa} \mathbb{G}_{m,\kappa}}$.

From now on, the group scheme $\mathbb{G}_{m,\kappa} = \text{Spec}(\kappa[t, t^{-1}])$ will be written simply as \mathbb{G} .

2B. Motivic zeta function and motivic Milnor fiber. Let X be a smooth variety over κ of pure dimension n , and let $g : X \rightarrow \mathbb{A}^1_{\kappa}$ be a function on X , with zero locus $X_0(g)$. For $m \geq 1$, we define

$$\mathcal{X}_m(g) := \{\varphi \in \mathcal{L}_m(X) \mid \text{ord}_t g(\varphi) = m\}.$$

Note that this variety is invariant by the \mathbb{G} -action on $\mathcal{L}_m(X)$. Furthermore, g induces a morphism $g_m : \mathcal{X}_m(g) \rightarrow \mathbb{G}$, assigning to a point φ in $\mathcal{L}_m(X)$ the coefficient $\text{ac}(g(\varphi))$ of t^m in $g(\varphi(t))$, which we also denote by $\text{ac}(g)(\varphi)$. This morphism is a diagonally monomial of weight m with respect to the \mathbb{G} -action on $\mathcal{X}_m(g)$ since $g(s \cdot \varphi) = s^m g_m(\varphi)$. We thus consider the class $[\mathcal{X}_m(g)]$ of $\mathcal{X}_m(g)$ in $\mathcal{M}^{\mathbb{G}}_{X_0(g) \times_{\kappa} \mathbb{G}}$. We can now consider the *motivic zeta function*

$$Z_g(T) := \sum_{m \geq 1} [\mathcal{X}_m(g)] \mathbb{L}^{-mn} T^m$$

in $\mathcal{M}^{\mathbb{G}}_{X_0(g) \times_{\kappa} \mathbb{G}}[[T]]$. Note that $Z_g = 0$ if $g = 0$ on X .

By using a log-resolution of $X_0(g)$, Denef and Loeser [1998; 2001] proved that $Z_g(T)$ is a rational series in $\mathcal{M}^{\mathbb{G}}_{X_0(g) \times_{\kappa} \mathbb{G}}[[T]]_{sr}$ (see next paragraph) and they also showed that one can consider the limit $\lim_{T \rightarrow \infty} Z_g(T)$ in $\mathcal{M}^{\mathbb{G}}_{X_0(g) \times_{\kappa} \mathbb{G}}$. Then the *motivic Milnor fiber* of g is defined as

$$\mathcal{G}_g := - \lim_{T \rightarrow \infty} Z_g(T).$$

2C. Rational series and their limits. Let A be one of the rings

$$\mathbb{Z}[\mathbb{L}, \mathbb{L}^{-1}], \quad \mathbb{Z}[\mathbb{L}, \mathbb{L}^{-1}, (1/(1 - \mathbb{L}^{-i}))_{i>0}], \quad \mathcal{M}^{\mathbb{G}}_{S \times_{\kappa} \mathbb{G}}.$$

We denote by $A[[T]]_{sr}$ the A -submodule of $A[[T]]$ generated by 1 and by finite products of terms $p_{e,i}(T) = \mathbb{L}^e T^i / (1 - \mathbb{L}^e T^i)$ with e in \mathbb{Z} and i in $\mathbb{N}_{>0}$. There is a unique A -linear morphism

$$\lim_{T \rightarrow \infty} : A[[T]]_{sr} \rightarrow A$$

such that

$$\lim_{T \rightarrow \infty} \left(\prod_{i \in I} p_{e_i, j_i}(T) \right) = (-1)^{|I|}$$

for every family $((e_i, j_i))_{i \in I}$ in $\mathbb{Z} \times \mathbb{N}_{>0}$ with I finite (possibly empty).

We will use the notation

$$\mathbb{R}_{\geq 0}^I := \{a = (a_1, \dots, a_n) \in \mathbb{R}_{\geq 0}^n \mid a_i = 0 \text{ for } i \notin I\},$$

$$\mathbb{R}_{> 0}^I := \{a = (a_1, \dots, a_n) \in \mathbb{R}_{\geq 0}^n \mid a_i = 0 \iff i \notin I\},$$

for I a subset of $\{1, \dots, n\}$. The sets $\mathbb{Z}_{\geq 0}^I$, $\mathbb{Z}_{> 0}^I$ and $\mathbb{N}_{> 0}^I$ are defined similarly.

Let Δ be a rational polyhedral convex cone in $\mathbb{R}_{> 0}^I$ and let $\overline{\Delta}$ denote its closure in $\mathbb{R}_{\geq 0}^I$ with I a finite set. Let l and l' be two integer linear forms on \mathbb{Z}^I positive on $\overline{\Delta} \setminus \{0\}$. Consider the series

$$S_{\Delta, l, l'}(T) := \sum_{k \in \Delta \cap \mathbb{N}_{> 0}^I} \mathbb{L}^{-l'(k)} T^{l(k)}$$

in $\mathbb{Z}[\mathbb{L}, \mathbb{L}^{-1}][[T]]$.

Lemma 2.1 [Guibert 2002]. *With previous notations and hypotheses, assuming that Δ is open in its linear span $\overline{\Delta}$, the series $S_{\Delta, l, l'}(T)$ lies in $\mathbb{Z}[\mathbb{L}, \mathbb{L}^{-1}][[T]]_{sr}$ and*

$$\lim_{T \rightarrow \infty} S_{\Delta, l, l'}(T) = (-1)^{\dim(\Delta)}.$$

3. The Newton polyhedron of a regular function

3A. Newton polyhedra. Let $g(x) = \sum_{\alpha \in \mathbb{N}^n} a_\alpha x^\alpha$ be a polynomial in n variables $x = (x_1, \dots, x_n)$ such that $g(0) = 0$. We denote by $\text{supp}(g)$ the set of exponents α in \mathbb{N}^n with $a_\alpha \neq 0$. The Newton polyhedron Γ of g is the convex hull of $\text{supp}(g) + \mathbb{R}_{\geq 0}^n$. For a compact face γ of Γ , we denote by g_γ the quasihomogenous polynomial

$$g_\gamma(x) = \sum_{\alpha \in \gamma} a_\alpha x^\alpha.$$

We say g is *nondegenerate* with respect to its Newton polyhedron Γ if, for every compact face γ of Γ , the *face function* g_γ is smooth on \mathbb{G}^n .

To the Newton polyhedron Γ we associate a function l_Γ which assigns to a vector a in $\mathbb{R}_{\geq 0}^n$ the value $\inf_{b \in \Gamma} \langle a, b \rangle$, with $\langle a, b \rangle$ being the standard inner product of a and b . For a in $\mathbb{R}_{\geq 0}^n$, we denote by γ_a the face of Γ on which the restriction of the function $\langle a, \cdot \rangle$ on Γ attains its minimum, i.e., $b \in \Gamma$ is in γ_a if and only if

$$\langle a, b \rangle = l_\Gamma(a) = \min_{b \in \Gamma} \langle a, b \rangle.$$

For $a = 0$ in $\mathbb{R}_{\geq 0}^n$, $\gamma_a = \Gamma$. If $a \neq 0$, γ_a is a proper face of Γ . Furthermore, γ_a is a compact face of Γ if and only if a is in $\mathbb{R}_{> 0}^n$. For any face γ of the Newton polyhedron Γ , we denote by $\sigma(\gamma)$ the cone $\{a \in \mathbb{R}_{\geq 0}^n \mid \gamma_a = \gamma\}$. Its closure is given by $\overline{\sigma}(\gamma) = \{a \in \mathbb{R}_{\geq 0}^n \mid \gamma_a \supset \gamma\}$.

A *fan* \mathcal{F} is a finite set of rational polyhedral cones such that every face of a cone of \mathcal{F} is also a cone of \mathcal{F} , and the intersection of two arbitrary cones of \mathcal{F} is the common face of them.

3B. Partition of $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ with respect to g . Write $n = n_1 + n_2$ with $n_1 \geq 0$, $n_2 \geq 0$. Let g be a function on \mathbb{A}_k^n that is nondegenerate with respect to the Newton polyhedron Γ of g . Let γ be a compact face of Γ . A proper face ϵ of Γ is said to *lean* on γ if there exists a subset I of $\{1, \dots, n\}$ such that

$$\epsilon = \gamma + \mathbb{R}_{\geq 0}^I = \{a + b \mid a \in \gamma, b \in \mathbb{R}_{\geq 0}^I\}.$$

Note that $\dim(\epsilon) = \dim(\gamma) + |I|$. Clearly, the face ϵ is noncompact when I is nonempty. The following lemmas are trivial.

Lemma 3.1. *If $\gamma + \mathbb{R}_{\geq 0}^I$ is a face leaning on a compact face γ of Γ , then for every J subset of I , $\gamma + \mathbb{R}_{\geq 0}^J$ is also a face of Γ leaning on γ .*

Notice that if $I = \emptyset$, the face $\gamma + \mathbb{R}_{\geq 0}^I$ reduces to the compact face γ . If $\epsilon = \gamma + \mathbb{R}_{\geq 0}^I$, we denote $\sigma_{\gamma, I} := \sigma(\epsilon)$. It is clear that $\dim(\sigma_{\gamma, I}) = n - |I| - \dim(\gamma)$.

Lemma 3.2. *If $\sigma_{\gamma, I}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$, then for every subset J of I , $\sigma_{\gamma, J}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$. Moreover, $\sigma_{\gamma, I}$ is a face of $\sigma_{\gamma, J}$.*

Lemma 3.3. *Assume that γ is a compact face and $\epsilon = \gamma + \mathbb{R}^I$ is a face of Γ . Then $\sigma_{\gamma, I}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ if and only if I is a subset of $\{1, \dots, n_1\}$.*

Fix a compact face γ of Γ . Let M be a maximal element (in the inclusion relation) of the family of the subsets of $\{1, \dots, n_1\}$ such that $\gamma + \mathbb{R}_{\geq 0}^M$ is a face of Γ (thus, by Lemma 3.3, $\sigma_{\gamma, M}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$). Then, for every subset I of M , $\gamma + \mathbb{R}_{\geq 0}^I$ is a face of Γ due to Lemma 3.1, and $\sigma_{\gamma, I}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ by Lemma 3.2. We thus have proved the following result.

Proposition 3.4. *There exists a canonical fan in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ with respect to g partitioning it into the cones $\sigma_{\gamma, I}$, where I runs over the subsets of M , M runs over the maximal subsets of $\{1, \dots, n_1\}$ such that $\gamma + \mathbb{R}_{\geq 0}^M$ is a face of Γ , and γ runs over the compact faces of Γ .*

Example 3.5. Consider a function $g(x_1, \dots, x_n)$ with Γ_g having a unique vertex P . Then the k -dimensional faces of Γ leaning on P have the form

$$P + \mathbb{R}_{\geq 0}^I$$

with I a subset of $\{1, \dots, n\}$ and $|I| = k$, for $k = 0, \dots, n - 1$. We deduce from Lemma 3.3 that the canonical partition of $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ with respect to g is given by the cones $\sigma_{P, I}$, with I a subset of $\{1, \dots, n_1\}$.

Remark 3.6. In the case $n_1 = 0$, we reduce to the work in [Guibert 2002]. More clearly, for each compact face γ of Γ , all the maximal subsets M of $\{1, \dots, n\}$, of which $\gamma + \mathbb{R}_{\geq 0}^M$ is a face of Γ and $\sigma_{\gamma, M} \subset \mathbb{R}_{> 0}^{n_2}$, are empty.

4. Computation of $i_1^* \mathcal{P}_g$ and $\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{P}_g$

Consider a regular function g on \mathbb{A}_κ^n . We assume that g is nondegenerate with respect to its Newton polyhedron Γ . Denote by i_1 the natural inclusion $\mathbb{A}_\kappa^{n_1} \hookrightarrow \mathbb{A}_\kappa^n$ or $\mathbb{A}_\kappa^{n_1} \times_\kappa \mathbb{G} \hookrightarrow \mathbb{A}_\kappa^n \times_\kappa \mathbb{G}$.

4A. The motivic zeta function $Z_g(T)$. We identify the arc space $\mathcal{L}(\mathbb{A}_\kappa^n)$ with the space of formal power series $\kappa[[t]]^n$ via the system of coordinates x_1, \dots, x_n . For every arc $\varphi \in \mathcal{L}(\mathbb{A}_\kappa^n)$, we note $\text{ord}_t x(\varphi) = (\text{ord}_t x_1(\varphi), \dots, \text{ord}_t x_n(\varphi))$. For every $m \in \mathbb{N}_{>0}$ and $a \in \mathbb{N}^n$ we set

$$\mathcal{X}_{a,m}(g) = \mathcal{X}_m(g) \cap \pi_m(\mathcal{X}_a),$$

where the spaces $\mathcal{X}_m(g)$ and \mathcal{X}_a are defined as follows:

$$\begin{aligned} \mathcal{X}_m(g) &= \{\varphi \in \mathcal{L}_m(\mathbb{A}_\kappa^n) \mid \text{ord}_t g(\varphi) = m\}, \\ \mathcal{X}_a &= \{\varphi \in \mathcal{L}(\mathbb{A}_\kappa^n) \mid \text{ord}_t x(\varphi) = a\}. \end{aligned}$$

It is clear that $\mathcal{X}_{a,m}(g)$ is a variety over $X_0(g) \times_\kappa \mathbb{G}$ in which the morphism to $X_0(g)$ is induced by the canonical morphism $\mathcal{L}_m(\mathbb{A}_\kappa^n) \rightarrow \mathbb{A}_\kappa^n$ and the morphism to \mathbb{G} is the morphism $\text{ac}(g)$. Note that $\mathcal{X}_{a,m}(g)$ is invariant by the \mathbb{G} -action on $\mathcal{L}_m(\mathbb{A}_\kappa^n)$.

For every $a \in \mathbb{N}^n$ and $\varphi \in \mathcal{X}_a$, $\text{ord}_t g(\varphi) \geq l_\Gamma(a)$ by the definition of l_Γ . Furthermore, $\mathcal{X}_m(g)$ can be expressed as a disjoint union $\bigcup_{a \in \mathbb{N}^n} \mathcal{X}_{a,m}(g)$ of the subspaces $\mathcal{X}_{a,m}(g)$ for a in \mathbb{N}^n . Then the motivic zeta function $Z_g(T)$ of g can be written in the following form:

$$\begin{aligned} Z_g(T) &= \sum_{a \in \mathbb{N}^n} \sum_{m \geq l_\Gamma(a)} [\mathcal{X}_{a,m}(g)] \mathbb{L}^{-nm} T^m \\ &= \sum_{a \in \mathbb{N}^n} \left([\mathcal{X}_{a,l_\Gamma(a)}(g)] \mathbb{L}^{-nl_\Gamma(a)} T^{l_\Gamma(a)} + \sum_{m \geq l_\Gamma(a)+1} [\mathcal{X}_{a,m}(g)] \mathbb{L}^{-nm} T^m \right) \\ &=: Z^0(T) + Z^1(T). \end{aligned}$$

There is a canonical partition of $\mathbb{R}_{\geq 0}^n$ into the rational polyhedral cones $\sigma(\gamma)$ with γ running over the proper faces of Γ , so we deduce that

$$\begin{aligned} Z^0(T) &= \sum_{\gamma} \sum_{a \in \sigma(\gamma)} [\mathcal{X}_{a,l_\Gamma(a)}(g)] \mathbb{L}^{-nl_\Gamma(a)} T^{l_\Gamma(a)}, \\ Z^1(T) &= \sum_{\gamma} \sum_{a \in \sigma(\gamma)} \sum_{k \geq 1} [\mathcal{X}_{a,l_\Gamma(a)+k}(g)] \mathbb{L}^{-n(l_\Gamma(a)+k)} T^{l_\Gamma(a)+k}, \end{aligned}$$

where the sum \sum_{γ} runs over the proper faces γ of Γ .

4B. Computation of $i_1^*Z_g(T)$. Assume that g satisfies the additional condition that $\mathbb{A}_k^{n_1}$ is naturally included in $X_0(g)$ via the morphism i_1 . To compute $i_1^*Z_g(T)$, we consider the canonical fan in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ with respect to g . Denote by Γ_c the set of compact faces of Γ and by \mathfrak{M}_γ the set of maximal subsets M of $\{1, \dots, n_1\}$ such that $\gamma + \mathbb{R}_{\geq 0}^M$ is a face of Γ . By Proposition 3.4, we can partition $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ into the cones $\sigma_{\gamma, I}$, with I a subset of M , M in \mathfrak{M}_γ and γ in Γ_c . Assume that $\mathfrak{M}_\gamma = \{M_1, \dots, M_p\}$. We denote by \mathfrak{S}_γ the family of subsets of one of the sets M_1, \dots, M_p . Then we have

$$i_1^*Z^0(T) = i_1^* \left(\sum_{\gamma \in \Gamma_c} \sum_{I \in \mathfrak{S}_\gamma} \sum_{a \in \sigma_{\gamma, I}} [\mathcal{X}_{a, l_\Gamma(a)}(g)] \mathbb{L}^{-nl_\Gamma(a)} T^{l_\Gamma(a)} \right),$$

$$i_1^*Z^1(T) = i_1^* \left(\sum_{\gamma \in \Gamma_c} \sum_{I \in \mathfrak{S}_\gamma} \sum_{a \in \sigma_{\gamma, I}} \sum_{k \geq 1} [\mathcal{X}_{a, l_\Gamma(a)+k}(g)] \mathbb{L}^{-n(l_\Gamma(a)+k)} T^{l_\Gamma(a)+k} \right).$$

4C. Class of $\mathcal{X}_{a, m}(g)$. For a compact face γ of Γ , consider the variety $X_\gamma := \mathbb{G}^n \setminus g_\gamma^{-1}(0)$ endowed with a \mathbb{G} -action as follows: if $\gamma = \gamma_a$, $a = (a_1, \dots, a_n)$ then we set

$$s \cdot (\xi_1, \dots, \xi_n) = (s^{a_1} \xi_1, \dots, s^{a_n} \xi_n).$$

For each compact γ and I in \mathfrak{S}_γ , consider the morphism

$$g_{\gamma, I} : X_\gamma = \mathbb{G}^n \setminus g_\gamma^{-1}(0) \rightarrow X_0(g) \times_\kappa \mathbb{G}$$

given by

$$g_{\gamma, I}(\xi_1, \dots, \xi_n) = ((\hat{\xi}_1, \dots, \hat{\xi}_n), g_\gamma(\xi_1, \dots, \xi_n)),$$

where $\hat{\xi}_i$ is defined by

$$\hat{\xi}_i = \begin{cases} \xi_i & \text{if } i \in I, \\ 0 & \text{otherwise.} \end{cases}$$

The first projection $X_\gamma \rightarrow X_0(g)$ is \mathbb{G} -equivariant in an obvious manner, and for $\gamma = \gamma_a$, the second $X_\gamma \rightarrow \mathbb{G}$ is diagonally monomial of weight $l_\Gamma(a)$ with respect to the \mathbb{G} -action since $g_\gamma(s \cdot (\xi_1, \dots, \xi_n)) = s^{l_\Gamma(a)} g_\gamma(\xi_1, \dots, \xi_n)$ for any s in \mathbb{G} . This defines a class $[g_{\gamma, I} : X_\gamma \rightarrow X_0(g) \times_\kappa \mathbb{G}]$ in $\mathcal{M}_{X_0(g) \times_\kappa \mathbb{G}}^\mathbb{G}$, which we denote by $\Phi_{\gamma, I}$. Notice that $\Phi_{\gamma, I}$ does not depend on the action thanks to the construction of the Grothendieck group (cf. [Guibert et al. 2005, 2006]).

We denote by $\Psi_{\gamma, I}$ the class in $\mathcal{M}_{X_0(g) \times_\kappa \mathbb{G}}^\mathbb{G}$ of the morphism

$$g_\gamma^{-1}(0) \times_\kappa \mathbb{G} \rightarrow X_0(g) \times_\kappa \mathbb{G},$$

which maps $((\xi_1, \dots, \xi_n), t)$ to $((\hat{\xi}_1, \dots, \hat{\xi}_n), t^{l_\Gamma(a)})$ for $\gamma = \gamma_a$, with the \mathbb{G} -action on $g_\gamma^{-1}(0)$ given by $s \cdot (\xi_1, \dots, \xi_n) = (s^{a_1} \xi_1, \dots, s^{a_n} \xi_n)$, the \mathbb{G} -action on \mathbb{G} given by the multiplicative translation, $g_\gamma^{-1}(0) \times_\kappa \mathbb{G} \rightarrow X_0(g) \times_\kappa \mathbb{G}$ being \mathbb{G} -equivariant, and

$g_\gamma^{-1}(0) \times_\kappa \mathbb{G} \rightarrow \mathbb{G}$ being diagonally monomial of weight $l_\Gamma(a)$ with respect to the \mathbb{G} -action.

Lemma 4.1. *The following formulas hold in $\mathcal{M}_{X_0(g) \times_\kappa \mathbb{G}}^\mathbb{G}$ for every a in $\sigma_{\gamma,I}$:*

- (i) *If there is a nonempty subset I of $\{1, \dots, n\}$ such that $a_i > m$ for any $i \in I$ and $g|_{\mathbb{A}_\kappa^{I^c}} = 0$, then $[\mathcal{X}_{a,m}(g)] = 0$.*

If $a_i \leq l_\Gamma(a)$ for any $i = 1, \dots, n$, we have

- (ii) $[\mathcal{X}_{a,l_\Gamma(a)}(g)] = \Phi_{\gamma,I} \mathbb{L}^{nl_\Gamma(a)-s(a)}$,
- (iii) $[\mathcal{X}_{a,l_\Gamma(a)+k}(g)] = \Psi_{\gamma,I} \mathbb{L}^{n(l_\Gamma(a)+k)-s(a)}$ for $k \geq 1$.

Here, $\mathbb{A}_\kappa^{I^c} := \{(x_1, \dots, x_n) \in \mathbb{A}_\kappa^n \mid x_i = 0 \ \forall i \in I\}$, and $s(a) := \sum_{i=1}^n a_i$.

Proof. Item (i) follows from the definition of $\mathcal{X}_{a,m}(g)$ and from the hypothesis on g . Indeed, every element of $\pi_m(\mathcal{X}_a)$ has the form $\varphi(t) = (x_1(t), \dots, x_n(t))$, where $x_j(t)$ is a polynomial of degree $\leq m$ in a variable t for any $j = 1, \dots, n$, and $x_i(t)$ is the zero polynomial if i is in I . Then $g(\varphi(t)) = 0$ and $\text{ord}_t g(\varphi) = \infty$, which means that $\mathcal{X}_{a,m}(g) = \emptyset$.

Items (ii) and (iii) may be deduced from the proofs of [Guibert 2002, Lemmas 2.1.1, 2.1.2] and from the isomorphism $\mathcal{M}_{X_0(g)}^\mu \cong \mathcal{M}_{X_0(g) \times_\kappa \mathbb{G}}^\mathbb{G}$ (cf. [Guibert et al. 2006, Proposition 2.6]). In [Guibert 2002, Section 2.1] (in particular, Lemmas 2.1.1 and 2.1.2), Guibert only considers functions of the form $\sum_{\alpha \in \mathbb{N}_{>0}^n} f_\alpha x^\alpha$. Observe that his condition that $\alpha \in \mathbb{N}_{>0}^n$ is equivalent to the condition that $a_i \leq l_\Gamma(a)$ for any $i = 1, \dots, n$. Finally, notice that the hypothesis of nondegeneracy with respect to Γ is in fact the main tool for the proofs.

There is also a way to prove (ii) directly as follows. An element $\varphi(t)$ of $\mathcal{X}_{a,l_\Gamma(a)}(g)$ has the form $\varphi(t) = (x_1(t), \dots, x_n(t))$, where $x_i(t) = \sum_{m=a_i}^{l_\Gamma(a)} c_{i,m} t^m$ with $c_{i,a_i} \neq 0$ for $i = 1, \dots, n$. Note that the coefficient of $t^{l_\Gamma(a)}$ in $g(\varphi(t))$ is equal to

$$\begin{aligned} \frac{1}{l_\Gamma(a)!} \cdot \frac{d^{l_\Gamma(a)} g(\varphi(t))}{dt^{l_\Gamma(a)}} \Big|_{t=0} &= \frac{1}{l_\Gamma(a)!} \cdot \frac{d^{l_\Gamma(a)} g_\gamma(\varphi(t))}{dt^{l_\Gamma(a)}} \Big|_{t=0} \\ &= g_\gamma(c_{1,a_1}, \dots, c_{n,a_n}), \end{aligned}$$

which is nonzero for every a in $\sigma_{\gamma,I}$ and $(c_{1,a_1}, \dots, c_{n,a_n})$ in X_γ . One deduces from this that $\mathcal{X}_{a,l_\Gamma(a)}(g)$ is isomorphic to $X_\gamma \times_\kappa \mathbb{A}_\kappa^{nl_\Gamma(a)-s(a)}$ via the map

$$\varphi(t) \mapsto ((c_{i,a_i})_{1 \leq i \leq n}, (c_{i,m})_{1 \leq i \leq n, a_i+1 \leq m \leq l_\Gamma(a)}).$$

Here the action of \mathbb{G} on \mathbb{A}_κ^1 is trivial. For any s in \mathbb{G} , the arc $\varphi(st)$ is mapped to

$$((s^{a_i} c_{i,a_i})_{1 \leq i \leq n}, (c_{i,m})_{1 \leq i \leq n, a_i+1 \leq m \leq l_\Gamma(a)}),$$

which is by definition equal to

$$s \cdot ((c_{i,a_i})_{1 \leq i \leq n}, (c_{i,m})_{1 \leq i \leq n, a_i+1 \leq m \leq l_\Gamma(a)}).$$

This means that the \mathbb{G} -action is compatible with the isomorphism; that is, the isomorphism is \mathbb{G} -equivariant. Then item (ii) follows. \square

Remark 4.2. We do not know yet how to compute $[\mathcal{Z}_{a, l_\Gamma(a)+k}(g)]$ for $k \geq 0$ without the assumptions as in Lemma 4.1.

Remark 4.3. Lemma 4.1 and Remark 4.2 explain the reason why in the rest of this paper we will always assume that no vertex of the Newton polyhedron Γ of g lies in a coordinate plane; that is, $a_i \leq l_\Gamma(a)$ for any $i = 1, \dots, n$. In this case, $l_\Gamma(a)$ is expressed as $\sum_{i=1}^n \alpha_i a_i$ with $\alpha_i > 0$ for any $i = 1, \dots, n$. By Lemma 4.1, this hypothesis guarantees that, for every compact face γ of Γ , with I in \mathfrak{S}_γ , all the terms of the sum $\sum_{a \in \sigma_{\gamma, I}} [\mathcal{Z}_{a, l_\Gamma(a)}(g)]$ are nonzero if $\Phi_{\gamma, I}$ is nonzero, and all the terms of the sum $\sum_{a \in \sigma_{\gamma, I}} [\mathcal{Z}_{a, l_\Gamma(a)+k}(g)]$ (where $k > 0$) are nonzero if $\Psi_{\gamma, I}$ is nonzero. In our work, we want to consider sums of this type that can be reduced to the case of Lemma 2.1.

4D. An explicit formula for $i_1^* \mathcal{S}_g$. Assume that g is a regular function on \mathbb{A}_k^n that is nondegenerate with respect to its Newton polyhedron Γ , that no vertex of Γ lies in a coordinate m -plane ($m = 1, \dots, n - 1$), and that $X_0(g)$ contains $\mathbb{A}_k^{n_1} \times_k \{0\}$. One then deduces from Remark 4.3 and Lemma 4.1 that

$$i_1^* Z^0(T) = \sum_{\gamma \in \Gamma_c} \sum_{I \in \mathfrak{S}_\gamma} \sum_{a \in \sigma_{\gamma, I}} i_1^* \Phi_{\gamma, I} \mathbb{L}^{-s(a)} T^{l_\Gamma(a)},$$

and that

$$\begin{aligned} i_1^* Z^1(T) &= i_1^* \left(\sum_{\gamma \in \Gamma_c} \sum_{I \in \mathfrak{S}_\gamma} \sum_{a \in \sigma_{\gamma, I}} \Psi_{\gamma, I} \mathbb{L}^{-s(a)} T^{l_\Gamma(a)} \sum_{k \geq 1} \mathbb{L}^{-k} T^k \right) \\ &= \frac{\mathbb{L}^{-1} T}{1 - \mathbb{L}^{-1} T} \sum_{\gamma \in \Gamma_c} \sum_{I \in \mathfrak{S}_\gamma} \sum_{a \in \sigma_{\gamma, I}} i_1^* \Psi_{\gamma, I} \mathbb{L}^{-s(a)} T^{l_\Gamma(a)}. \end{aligned}$$

Proposition 4.4. *With the previous notation and hypotheses, the following formula holds in $\mathcal{M}_{\mathbb{A}_k^{n_1} \times_k \mathbb{G}}^{\mathbb{G}}$:*

$$i_1^* \mathcal{S}_g = \sum_{\gamma \in \Gamma_c} (-1)^{n+1-\dim(\gamma)} \sum_{I \in \mathfrak{S}_\gamma} (-1)^{|I|} [\mathbb{A}_k^{n_1} \times_{X_0(g)} (\Phi_{\gamma, I} - \Psi_{\gamma, I})].$$

Proof. The positivity of the sum function s on $\overline{\sigma_{\gamma, I}} \setminus \{0\}$ is evident, and that of the function l_Γ on $\overline{\sigma_{\gamma, I}} \setminus \{0\}$ follows straightforward from Remark 4.3. Applying Lemma 2.1, notice that $\dim(\sigma_{\gamma, I}) = n - |I| - \dim(\gamma)$; we have

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{a \in \sigma_{\gamma, I}} \Phi_{\gamma, I} \mathbb{L}^{-s(a)} T^{l_\Gamma(a)} &= \Phi_{\gamma, I} \lim_{T \rightarrow \infty} \sum_{a \in \sigma_{\gamma, I}} \mathbb{L}^{-s(a)} T^{l_\Gamma(a)} \\ &= (-1)^{n-|I|-\dim(\gamma)} \Phi_{\gamma, I} \end{aligned}$$

and

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{a \in \sigma_{\gamma, I}} \Psi_{\gamma, I} \mathbb{L}^{-s(a)} T^{l_{\Gamma}(a)} &= \Psi_{\gamma, I} \lim_{T \rightarrow \infty} \sum_{a \in \sigma_{\gamma, I}} \mathbb{L}^{-s(a)} T^{l_{\Gamma}(a)} \\ &= (-1)^{n-|I|-\dim(\gamma)} \Psi_{\gamma, I}. \end{aligned}$$

It follows that

$$\lim_{T \rightarrow \infty} i_1^* Z^0(T) = \sum_{\gamma \in \Gamma_c} (-1)^{n-\dim(\gamma)} \sum_{I \in \mathfrak{S}_{\gamma}} (-1)^{|I|} i_1^* \Phi_{\gamma, I},$$

and

$$\lim_{T \rightarrow \infty} i_1^* Z^1(T) = \sum_{\gamma \in \Gamma_c} (-1)^{n+1-\dim(\gamma)} \sum_{I \in \mathfrak{S}_{\gamma}} (-1)^{|I|} i_1^* \Psi_{\gamma, I}.$$

Then the proposition is proved. □

Example 4.5 (cf. Example 3.5). In the case where Γ_g has a unique compact face P , the classes $\Psi_{P, I}$ vanish. If we assume that $\alpha_i > 0$ for every $i = 1, \dots, n$, we have

$$i_1^* \mathcal{G}_g = (-1)^{n+1} \sum_{I \subset \{1, \dots, n_1\}} (-1)^{|I|} [\mathbb{A}_{\kappa}^{n_1} \times_{X_0(g)} \Phi_{P, I}].$$

Corollary 4.6 [Guibert 2002]. Assume that g is given by $g(x) = \sum_{\alpha \in \mathbb{N}_{>0}^n} a_{\alpha} x^{\alpha}$ in $\kappa[x]$ with $g(0) = 0$. If g is nondegenerate with respect to Γ , then

$$\mathcal{G}_{g, 0} = (-1)^{n-1} \sum_{\gamma \in \Gamma_c} (-1)^{\dim(\gamma)} [\{0\} \times_{X_0(g)} (\Phi_{\gamma, I} - \Psi_{\gamma, I})]$$

holds in $\mathcal{M}_{\mathbb{G}}^{\mathbb{G}}$.

Proof. (See Remark 3.6) Apply Proposition 4.4 to the case $n_1 = 0$. Here the natural inclusion $i_1 : \mathbb{A}_{\kappa}^{n_1} \hookrightarrow \mathbb{A}_{\kappa}^n$ reduces to the inclusion $i_0 : \{0\} \hookrightarrow \mathbb{A}_{\kappa}^n$. Moreover, in this case, by Lemma 3.3, for every compact face γ of Γ , we have $\mathfrak{S}_{\gamma} = \{\emptyset\}$. Thus this corollary follows. Observe that this formula was already obtained by Guibert (cf. [2002, Proposition 2.1.6]). □

4E. Consider the function $g(x) = \sum_{\alpha \in H \cap \mathbb{N}^n} a_{\alpha} x^{\alpha}$ on \mathbb{A}_{κ}^n , where H is the hyperplane in $\mathbb{R}_{\geq 0}^n$ defined by the equation

$$\alpha_1 + \dots + \alpha_{n_1} = \alpha_{n_1+1} + \dots + \alpha_p,$$

for some fixed p such that $n_1 < p \leq n$. Here, as well as in Corollary 4.6, we use the notation x^{α} for $x_1^{\alpha_1} \dots x_n^{\alpha_n}$, where $\alpha = (\alpha_1, \dots, \alpha_n)$. Because $\text{supp}(g)$ lies on the hyperplane H , the compact faces of Γ are contained in H . Moreover, for the same reason, for each compact γ , there exist noncompact faces of Γ leaning on γ . Note that, in this case, $\mathbb{A}_{\kappa}^{n_1}$ is naturally viewed as a subset of $X_0(g)$.

Lemma 4.7. *Assume that $g(x) = \sum_{\alpha \in H \cap \mathbb{N}^n} a_\alpha x^\alpha$ is nondegenerate with respect to Γ . Then, for every compact face γ of Γ , we have $|\mathfrak{M}_\gamma| = 1$, and the unique element of \mathfrak{M}_γ is nonempty.*

Proof. Let γ be a compact face of Γ . Assume that $\gamma + \mathbb{R}_{\geq 0}^I$ is a face of Γ . Then, by Lemma 3.3, the cone $\sigma_{\gamma, I}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$ if and only if I is contained in $\{1, \dots, n_1\}$. Furthermore, we claim that if $\gamma + \mathbb{R}_{\geq 0}^I$ and $\gamma + \mathbb{R}_{\geq 0}^J$ are faces leaning on γ such that the corresponding cones $\sigma_{\gamma, I}$ and $\sigma_{\gamma, J}$ are both contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$, then so is $\gamma + \mathbb{R}_{\geq 0}^{I \cup J}$. Indeed, since $(\alpha_1, \dots, \alpha_n)$ is in H , one deduces that if I and J are contained in $\{1, \dots, n_1\}$, the intersection of $\gamma + \mathbb{R}_{\geq 0}^{I \cup J}$ with the interior of Γ is empty. This, together with the fact that $\gamma + \mathbb{R}_{\geq 0}^I$ and $\gamma + \mathbb{R}_{\geq 0}^J$ are faces of Γ , shows that $\gamma + \mathbb{R}_{\geq 0}^{I \cup J}$ is a face of Γ leaning on γ such that $\sigma_{\gamma, I \cup J}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$.

As a consequence of the above claim, for each compact face γ of Γ , there exists a unique maximal subset M of $\{1, \dots, n_1\}$ such that $\gamma + \mathbb{R}_{\geq 0}^M$ is a face of Γ that leans on γ , and $\sigma_{\gamma, M}$ is contained in $\mathbb{R}_{\geq 0}^{n_1} \times \mathbb{R}_{> 0}^{n_2}$. The nonemptiness of the set M follows from the fact that $\text{supp}(g)$ lies on the hyperplane H . □

Proposition 4.8. *Assume that $g(x) = \sum_{\alpha \in H \cap \mathbb{N}_{> 0}^n} a_\alpha x^\alpha$ is nondegenerate with respect to Γ . Then $\int_{\mathbb{A}_k^{d_1}} i_1^* \mathcal{S}_g$ vanishes in $\mathcal{M}_G^{\mathbb{G}}$.*

Proof. Let γ be a compact face of Γ . By Lemma 4.7, the set \mathfrak{M}_γ has a unique element and this element is nonempty. Assume $\mathfrak{M}_\gamma = \{M\}$ with $|M| \geq 1$. Note that $A_\gamma = \int_{\mathbb{A}_k^{d_1}} i_1^* \Phi_{\gamma, I}$ and $B_\gamma = \int_{\mathbb{A}_k^{d_1}} i_1^* \Psi_{\gamma, I}$ depend only on γ , not on I contained in M . Because $\sum_{j=0}^m (-1)^j \binom{m}{j} = 0$ for $m \geq 1$, one deduces that

$$\sum_{I \subset M} (-1)^{|I|} (A_\gamma - B_\gamma) = 0.$$

The hypothesis on g that $\alpha \in H \cap \mathbb{N}_{> 0}^n$ means no vertex of the Newton polyhedron Γ of g lies in a coordinate plane. By Proposition 4.4, the image $\int_{\mathbb{A}_k^{d_1}} i_1^* \mathcal{S}_g$ of \mathcal{S}_g vanishes in $\mathcal{M}_G^{\mathbb{G}}$. □

5. The Kontsevich–Soibelman conjecture

In this section, we will show that under certain assumptions, Conjecture 1.1 is true.

5A. Composition with a polynomial in two variables. We consider Conjecture 1.1 of Kontsevich and Soibelman in the case where F has the form $F(x, y, z) = f(g_1(x, y), g_2(z))$, where f is a polynomial in two variables with $f(0, y)$ nonzero of positive degree, g_1 is a function on $\mathbb{A}_k^{d_1} \times_k \mathbb{A}_k^{d_2}$ such that $g_1(tx, t^{-1}y) = g_1(x, y)$ and $g_1(0, 0) = 0$, and g_2 is a regular function on $\mathbb{A}_k^{d_3}$. Let $\mathbf{g} = g_1 \times g_2$ and $X_0(\mathbf{g}) = \{(x, y, z) \mid g_1(x, y) = g_2(z) = 0\}$. In particular, $X_0(\mathbf{g})$ contains $\mathbb{A}_k^{d_1} \times \{0\}$.

We denote by i_1 the inclusion of $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{G}$ into $X_0(f \circ \mathbf{g}) \times_\kappa \mathbb{G}$. Recall that, in this case, $h(z) = f(0, g_2(z))$.

Theorem 5.1. *Assume that f is a polynomial in two variables with $f(0, y)$ nonzero of positive degree. Let g_1 be a regular function on $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{A}_\kappa^{d_2}$ nondegenerate with respect to its Newton polyhedron Γ_{g_1} such that $g_1(0, 0) = 0$, no vertex of Γ_{g_1} lies in a coordinate plane, and $g_1(tx, t^{-1}y) = g_1(x, y)$ for every t in \mathbb{G} . Let g_2 be a regular function on $\mathbb{A}_\kappa^{d_3}$. Then, the formula*

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_{f \circ \mathbf{g}} = \mathbb{L}^{d_1} \mathcal{S}_{h,0}$$

holds in $\mathcal{M}_{\mathbb{G}}^{\mathbb{G}}$. In other words, in this case, Conjecture 1.1 is true.

Proof. In [Guibert et al. 2009], Guibert, Loeser and Merle consider the motivic Milnor fiber of a composition of the form $f(g_1, g_2)$ where g_1 and g_2 have no variable in common and f is a polynomial in $\kappa[x, y]$ such that $f(0, y)$ is nonzero of positive degree. To describe it, they used the generalized convolution operators Ψ_Q defined in [Guibert et al. 2005] and the tree of contact $\tau(f, 0)$ constructed in terms of Puiseux expansions by Guibert [2002]. Here 0 is the origin of \mathbb{A}_κ^d , with $d = d_1 + d_2 + d_3$. To any rupture vertex v of $\tau(f, 0)$ one attaches a weighted homogeneous polynomial Q_v in $\kappa[X, Y]$. The virtual objects A_v are defined inductively in terms of the tree of contact $\tau(f, 0)$ and A_{v_0} , where v_0 is the first (extended) rupture vertex of the tree and A_{v_0} depends only on g . Let i be the inclusion of $X_0(\mathbf{g}) \times_\kappa \mathbb{G}$ into $X_0(f \circ \mathbf{g}) \times_\kappa \mathbb{G}$. Let m_0 be the order of 0 as a root of $f(0, y)$. By the main theorem of [Guibert et al. 2009], the formula

$$i^* \mathcal{S}_{f \circ \mathbf{g}} = \mathcal{S}_{g_2}^{m_0}([X_0(g_1)]) - \sum_v \Psi_{Q_v}(A_v)$$

holds in $\mathcal{M}_{X_0(\mathbf{g}) \times_\kappa \mathbb{G}}^{\mathbb{G}}$, where Ψ_{Q_v} denotes the convolution defined in the same paper and the sum runs over the augmented set of rupture vertices of the tree $\tau(f, 0)$. The i_1 in the theorem is the inclusion of $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{G}$ into $X_0(f \circ \mathbf{g}) \times_\kappa \mathbb{G}$, but by abuse of notation, we also use i_1 for the inclusion $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{G} \hookrightarrow X_0(\mathbf{g}) \times_\kappa \mathbb{G}$. Thus i_1 and $i \circ i_1$ are in fact the same thing. Applying the operator $\int_{\mathbb{A}_\kappa^{d_1}} i_1^*$ to both sides of the previous formula, we have

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_{f \circ \mathbf{g}} = \int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_{g_2}^{m_0}([X_0(g_1)]) - \sum_v \int_{\mathbb{A}_\kappa^{d_1}} i_1^* \Psi_{Q_v}(A_v).$$

We claim that, with the previous notation and hypotheses, the formula

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_{g_2}^{m_0}([X_0(g_1)]) = \mathbb{L}^{d_1} \mathcal{S}_{h,0}$$

holds in $\mathcal{M}_{\mathbb{G}}^{\mathbb{G}}$. Indeed, as in the proof of [Guibert et al. 2006, Theorem 5.18], one can check that

$$i_1^* \mathcal{S}_{g_2}^{m_0}([X_0(g_1)]) = [g_1^{-1}(0)] \boxtimes \mathcal{S}_{g_2}^{m_0}.$$

By the hypotheses on g_1 and the fact that $i_1(\mathbb{A}_k^{d_1}) \cap g_2^{-1}(0) = \{0\}$, we have

$$i_1^*[g_1^{-1}(0)] = [\mathbb{A}_k^{d_1}] = \mathbb{L}^{d_1} \quad \text{and} \quad i_1^* \mathcal{S}_{g_2}^{m_0} = i_0^* \mathcal{S}_{g_2}^{m_0} = \mathcal{S}_{g_2}^{m_0,0}.$$

One deduces that

$$i_1^* \mathcal{S}_{g_2}^{m_0}([X_0(g_1)]) = i_1^*([g_1^{-1}(0)] \boxtimes \mathcal{S}_{g_2}^{m_0}) = \mathbb{L}^{d_1} \mathcal{S}_{g_2}^{m_0,0}.$$

By the definitions of h and m_0 , we have $\mathcal{S}_{g_2}^{m_0,0} = \mathcal{S}_{h,0}$, and the claim then follows. So, in order to finish the proof of Theorem 5.1, it suffices to prove that $\int_{\mathbb{A}_k^{d_1}} i_1^* \Psi_Q(A_v) = 0$ for every (extended) rupture vertex v of $\tau(f, 0)$.

Let v_0 be the first (extended) rupture vertex of the tree of contact $\tau(f, p)$. As in [Guibert et al. 2009], the virtual object A_{v_0} in

$$\mathcal{M}_{X_0(\mathbf{g}) \times_{\kappa} (\mathbb{A}_k^1 \times_{\kappa} \mathbb{G})}^{\mathbb{G}}$$

is defined by $A_{v_0} := \mathcal{S}'_{g_2} \boxtimes \mathcal{S}_{g_1}$, where \mathcal{S}'_{g_2} is an element in $\mathcal{M}_{X_0(g_2) \times_{\kappa} \mathbb{A}_k^1}^{\mathbb{G}}$ that is the “disjoint sum” of \mathcal{S}_{g_2} in $\mathcal{M}_{X_0(g_2) \times_{\kappa} \mathbb{G}}^{\mathbb{G}}$ and $X_0(g_2)$ in $\mathcal{M}_{X_0(g_2)}$.

Lemma 5.2. *Assume that g_1 is a regular function on $\mathbb{A}_k^{d_1} \times_{\kappa} \mathbb{A}_k^{d_2}$ nondegenerate with respect to its Newton polyhedron Γ_{g_1} such that $g_1(0, 0) = 0$, no vertex of Γ_{g_1} lies in a coordinate plane, and $g_1(tx, t^{-1}y) = g_1(x, y)$ for every t in \mathbb{G} . Let g_2 be a regular function on $\mathbb{A}_k^{d_3}$. Then $\int_{\mathbb{A}_k^{d_1}} i_1^* \Psi_Q(A_{v_0})$ vanishes in $\mathcal{M}_{\mathbb{G}}^{\mathbb{G}}$ for every quasihomogeneous polynomial Q .*

Proof. The assumptions on g_1 mean that we can write g_1 in the form

$$g_1(x, y) = \sum_{(\alpha, \beta) \in H \cap \mathbb{N}_{>0}^{d_1+d_2}} a_{\alpha\beta} x_1^{\alpha_1} \cdots x_{d_1}^{\alpha_{d_1}} y_1^{\beta_1} \cdots y_{d_2}^{\beta_{d_2}},$$

where H is given by $\alpha_1 + \cdots + \alpha_{d_1} = \beta_1 + \cdots + \beta_{d_2}$. By Proposition 4.8, $\int_{\mathbb{A}_k^{d_1}} i_1^* \mathcal{S}_{g_1}$ vanishes in $\mathcal{M}_{\mathbb{G}}^{\mathbb{G}}$, hence $\int_{\mathbb{A}_k^{d_1}} i_1^* A_{v_0}$ vanishes in $\mathcal{M}_{\mathbb{A}_k^1 \times_{\kappa} \mathbb{G}}^{\mathbb{G}}$. Here i_1 is once again abused to denote the natural inclusion $\mathbb{A}_k^{d_1} \times_{\kappa} \mathbb{A}_k^1 \times_{\kappa} \mathbb{G} \hookrightarrow X_0(\mathbf{g}) \times_{\kappa} \mathbb{A}_k^1 \times_{\kappa} \mathbb{G}$. Because the diagram

$$\begin{array}{ccc} \mathcal{M}_{X_0(\mathbf{g}) \times_{\kappa} \mathbb{A}_k^1 \times_{\kappa} \mathbb{G}}^{\mathbb{G}} & \xrightarrow{\Psi_Q} & \mathcal{M}_{X_0(\mathbf{g}) \times_{\kappa} \mathbb{G}}^{\mathbb{G}} \\ \int_{\mathbb{A}_k^{d_1}} i_1^* \downarrow & & \int_{\mathbb{A}_k^{d_1}} i_1^* \downarrow \\ \mathcal{M}_{\mathbb{A}_k^1 \times_{\kappa} \mathbb{G}}^{\mathbb{G}} & \xrightarrow{\Psi_Q} & \mathcal{M}_{\mathbb{G}}^{\mathbb{G}} \end{array}$$

commutes, the lemma follows. □

Let v be an arbitrary rupture vertex of the tree of contact $\tau(f, 0)$ and $a(v)$ the predecessor of v in the augmented set of rupture vertices. Then the polynomial Q_v is a factor of $Q_{a(v)}$. Suppose that $Q_v(X, 1)$ has m_v disjoint zeroes in \mathbb{A}_κ^1 .

Lemma 5.3. *The equality $A_v = m_v A_{a(v)}$ holds in $\mathcal{M}_{X_0(\mathfrak{g}) \times_\kappa \mathbb{A}_\kappa^1 \times_\kappa \mathbb{G}}^\mathbb{G}$.*

Proof. We first notice that $Q_v^{-1}(0)$ is a smooth subvariety in $\mathbb{G} \times_\kappa \mathbb{G}$ equivariant under a diagonal \mathbb{G} -action, and that the second projection pr_2 of the product $\mathbb{A}_\kappa^1 \times_\kappa \mathbb{G}$ induces a homogeneous fibration $Q_v^{-1}(0) \rightarrow \mathbb{G}$. We denote by B_v the restriction of $A_{a(v)}$ above $Q_v^{-1}(0)$. Then, by [Guibert et al. 2009], the element A_v in $\mathcal{M}_{X_0(\mathfrak{g}) \times_\kappa \mathbb{A}_\kappa^1 \times_\kappa \mathbb{G}}^\mathbb{G}$ is defined as the external product of the class of $\text{id} : \mathbb{A}_\kappa^1 \rightarrow \mathbb{A}_\kappa^1$ by the induced map $\text{pr}_2 : B_v \rightarrow \mathbb{G}$, which is diagonally monomial when restricted to $X_0(\mathfrak{g}) \times_\kappa \mathbb{G} \times_\kappa \mathbb{G}$.

Consider the fibration $\text{pr}_2 : B_v \rightarrow \mathbb{G}$ defined by the composition of $B_v \rightarrow Q_v^{-1}(0)$ and $\text{pr}_2 : Q_v^{-1}(0) \rightarrow \mathbb{G}$. Then each fiber of $\text{pr}_2 : B_v \rightarrow \mathbb{G}$ is a disjoint union of m_v copies of a fiber of $A_{a(v)} \rightarrow \mathbb{A}_\kappa^1 \times_\kappa \mathbb{G}$ over one point (a, b) in $\mathbb{A}_\kappa^1 \times_\kappa \mathbb{G}$. It follows that $A_v = m_v A_{a(v)}$ in $\mathcal{M}_{X_0(\mathfrak{g}) \times_\kappa (\mathbb{A}_\kappa^1 \times_\kappa \mathbb{G})}^\mathbb{G}$. □

It follows from Lemma 5.2 and Lemma 5.3 that $\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \Psi Q_v(A_v) = 0$ for every (extended) rupture vertex v of $\tau(f, 0)$, completing the proof of Theorem 5.1. □

Remark 5.4. In the case $f(x, y) = x + y$, the result can also be obtained directly from the motivic Thom–Sebastiani theorem [Denef and Loeser 1999b; 2001].

5B. In the next proposition, we prove the conjecture of Kontsevich and Soibelman under some other conditions on $F = g$, namely assuming F is nondegenerate with respect to its Newton polyhedron Γ and no vertex of Γ lies in a coordinate plane.

Proposition 5.5. *Let g be a regular function on $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{A}_\kappa^{d_2} \times_\kappa \mathbb{A}_\kappa^{d_3}$ such that $g(0, 0, z) = 0$ for every z in $\mathbb{A}_\kappa^{d_3}$, $g(tx, t^{-1}y, z) = g(x, y, z)$ for every t in \mathbb{G} , and (x, y, z) in $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{A}_\kappa^{d_2} \times_\kappa \mathbb{A}_\kappa^{d_3}$. If g is nondegenerate with respect to its Newton polyhedron Γ and no vertex of Γ lies in a coordinate plane, then $\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_g$ vanishes in $\mathcal{M}_\mathbb{G}^\mathbb{G}$. In other words, Conjecture 1.1 is true in this case.*

Proof. Write the function g in the form

$$g(x, y, z) = \sum_{(a,b,c) \in H \cap \mathbb{N}_{>0}^d} g_{a,b,c} x^a y^b z^c,$$

where $d = d_1 + d_2 + d_3$ and H is given by the equation $a_1 + \dots + a_{d_1} = b_1 + \dots + b_{d_2}$. By Proposition 4.8, $\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_g$ vanishes in $\mathcal{M}_\mathbb{G}^\mathbb{G}$. Notice that in this case $h(z) = F(0, 0, z) = g(0, 0, z) = 0$, hence $\mathcal{S}_{h,0}$ also vanishes in $\mathcal{M}_\mathbb{G}^\mathbb{G}$. □

5C. Functions of Steenbrink type. We consider now the case that

$$F(x, y, z) = g(x, y, z) + h(z)^\ell,$$

where g is as in Proposition 5.5, $h(z)$ is regular on $\mathbb{A}_\kappa^{d_3}$ such that $h(0) = 0$, and ℓ is a large enough natural number. By composition with the projection, we will view h as a function on \mathbb{A}_κ^d .

Theorem 5.6. *Let $F(x, y, z) = g(x, y, z) + h(z)^\ell$, where g is as in Proposition 5.5, $h(z)$ is regular on $\mathbb{A}_\kappa^{d_3}$ such that $h(0) = 0$, and ℓ is a natural number. There exists a positive real number N such that, if $\ell > N$, the following formula holds in $\mathcal{M}_\mathbb{G}$:*

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_F = \mathbb{L}^{d_1} \mathcal{S}_{h^\ell, 0}.$$

Proof. Let us denote by i and j the inclusions of $(X_0(g) \cap X_0(h)) \times_\kappa \mathbb{G}$ into $X_0(g) \times_\kappa \mathbb{G}$ and $X_0(F) \times_\kappa \mathbb{G}$, respectively. The existence of N is shown by [Guibert et al. 2006, Theorem 5.7]. Also by this theorem, for $\ell > N$, we have

$$j^* \mathcal{S}_F - i^* \mathcal{S}_g = \mathcal{S}_{h^\ell}([X_0(g)]) - \Psi_\Sigma(\mathcal{S}_{h^\ell}(\mathcal{S}_g)),$$

where Ψ_Σ is the convolution defined in [Guibert et al. 2006]. Then we get

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_F - \int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_g = \int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_{h^\ell}([X_0(g)]) - \int_{\mathbb{A}_\kappa^{d_1}} i_1^* \Psi_\Sigma(\mathcal{S}_{h^\ell}(\mathcal{S}_g)).$$

Now, by Proposition 5.5, $\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_g = 0$. An analogue to the proof of Lemma 5.2 shows that $\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \Psi_\Sigma(\mathcal{S}_{h^\ell}(\mathcal{S}_g))$ vanishes. One deduces that

$$\int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_F = \int_{\mathbb{A}_\kappa^{d_1}} i_1^* \mathcal{S}_{h^\ell}([X_0(g)]).$$

Define a function g' on $\mathbb{A}_\kappa^{d_1} \times_\kappa \mathbb{A}_\kappa^{d_2}$ by setting $g'(x, y) = g(x, y, 0)$. Then we have that $g'(0, 0) = 0$ and $g'(tx, t^{-1}y) = g'(x, y)$ for any t in \mathbb{G} . Furthermore, we have an identity in $\mathcal{M}_{X_0(g)}$ as follows

$$[X_0(g)] = [X_0(g')] + [\{(x, y, z) \in \mathbb{A}_\kappa^{d_1+d_2} \times (\mathbb{A}_\kappa^{d_3} \setminus \{0\}) \mid g(x, y, z) = 0\}].$$

As in the proof of Theorem 5.1, since h^ℓ and g' have no variable in common, we have

$$i_1^* \mathcal{S}_{h^\ell}([X_0(g')]) = \mathbb{L}^{d_1} \mathcal{S}_{h^\ell, 0}$$

in $\mathcal{M}_{\mathbb{A}_\kappa^{d_1} \times \mathbb{G}}$. It remains to notice that

$$i_1^* \mathcal{S}_{h^\ell}([\{(x, y, z) \in \mathbb{A}_\kappa^{d_1+d_2} \times (\mathbb{A}_\kappa^{d_3} \setminus \{0\}) \mid g(x, y, z) = 0\}]) = 0,$$

because the intersection

$$i_1(\mathbb{A}_\kappa^{d_1}) \cap \{(x, y, z) \in \mathbb{A}_\kappa^{d_1+d_2} \times (\mathbb{A}_\kappa^{d_3} \setminus \{0\}) \mid g(x, y, z) = 0\}$$

is empty. Thus, $\int_{\mathbb{A}_k^{d_1}} i_1^* \mathcal{P}_F = \mathbb{L}^{d_1} \mathcal{P}_{h^\ell, 0}$ in $\mathcal{M}_{\mathbb{G}}^{\mathbb{G}}$, as needed. \square

Acknowledgments

This work was suggested by François Loeser, my advisor, who proposed that I consider the conjecture first in the case of composition $f(g_1, g_2)$ and encouraged me in each step of proof. I am deeply grateful to him for his suggestions of methods to approach the problem and for his help in preparing the manuscript. I would like to thank the referee for his contributions to the paper, which made it more readable.

References

- [Denef and Loeser 1998] J. Denef and F. Loeser, “Motivic Igusa zeta functions”, *J. Algebraic Geom.* **7**:3 (1998), 505–537. MR 99j:14021 Zbl 0943.14010
- [Denef and Loeser 1999a] J. Denef and F. Loeser, “Germs of arcs on singular algebraic varieties and motivic integration”, *Invent. Math.* **135**:1 (1999), 201–232. MR 99k:14002 Zbl 0928.14004
- [Denef and Loeser 1999b] J. Denef and F. Loeser, “Motivic exponential integrals and a motivic Thom-Sebastiani theorem”, *Duke Math. J.* **99** (1999), 285–309. MR 2000k:14006 Zbl 0966.14015
- [Denef and Loeser 2001] J. Denef and F. Loeser, “Geometry on arc spaces of algebraic varieties”, pp. 327–348 in *European Congress of Mathematics* (Barcelona, 2000), vol. I, edited by C. Casacuberta et al., Progr. Math. **201**, Birkhäuser, Basel, 2001. MR 2004c:14037 Zbl 1079.14003
- [Guibert 2002] G. Guibert, “Espaces d’arcs et invariants d’Alexander”, *Comment. Math. Helv.* **77**:4 (2002), 783–820. MR 2003k:14021 Zbl 1046.14008
- [Guibert et al. 2005] G. Guibert, F. Loeser, and M. Merle, “Nearby cycles and composition with a nondegenerate polynomial”, *Int. Math. Res. Not.* **2005**:31 (2005), 1873–1888. MR 2006f:11145 Zbl 1093.14032
- [Guibert et al. 2006] G. Guibert, F. Loeser, and M. Merle, “Iterated vanishing cycles, convolution, and a motivic analogue of a conjecture of Steenbrink”, *Duke Math. J.* **132**:3 (2006), 409–457. MR 2007e:14011 Zbl 1173.14301
- [Guibert et al. 2009] G. Guibert, F. Loeser, and M. Merle, “Composition with a two variable function”, *Math. Res. Lett.* **16** (2009), 439–448. MR 2010j:14021 Zbl 1187.14046
- [Kontsevich and Soibelman 2008] M. Kontsevich and Y. Soibelman, “Stability structures, motivic Donaldson-Thomas invariants and cluster transformations”, preprint, 2008. arXiv 0811.2435v1
- [Toën 2006] B. Toën, “Derived Hall algebras”, *Duke Math. J.* **135**:3 (2006), 587–615. MR 2007h:18021 Zbl 1117.18011

Communicated by Ehud Hrushovski

Received 2010-10-01 Revised 2010-12-06 Accepted 2011-01-19

leqthuong@math.jussieu.fr *École Normale Supérieure, Département de Mathématiques
et Applications, UMR 8553 CNRS, 45 rue d’Ulm,
75230 Paris cedex 05, France*

Current address: *Institut de Mathématiques de Jussieu, UMR 7586 CNRS,
4 place Jussieu, 75005 Paris, France*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in ANT are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use L^AT_EX but submissions in other varieties of T_EX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibT_EX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 6 No. 2 2012

Arithmetic of singular Enriques surfaces KLAUS HULEK and MATTHIAS SCHÜTT	195
An upper bound on the Abbes–Saito filtration for finite flat group schemes and applications YICHAO TIAN	231
On the smallest number of generators and the probability of generating an algebra ROSTYSLAV V. KRAVCHENKO, MARCIN MAZUR and BOGDAN V. PETRENKO	243
Moving lemma for additive higher Chow groups AMALENDU KRISHNA and JINHYUN PARK	293
Fusion rules for abelian extensions of Hopf algebras CHRISTOPHER GOFF	327
Uniformly rigid spaces CHRISTIAN KAPPEN	341
On a conjecture of Kontsevich and Soibelman LÊ QUY THUONG	389



1937-0652(2012)6:2;1-D