Algebra &
Number
Theory

Algebra & Number Theory

www.jant.org

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

Stable reduction of $X_0(p^3)$

Ken McMurdy and Robert Coleman
with an appendix by Everett W. Howe

This paper is dedicated to Siegfried Bosch, whose foundational work in rigid analysis was invaluable in our development of the theory of semistable coverings.

We determine the stable models of the modular curves $X_0(p^3)$ for primes $p \geq 13$. An essential ingredient is the close relationship between the deformation theories of elliptic curves and formal groups, which was established in the Woods Hole notes of 1964. This enables us to apply results of Hopkins and Gross in our analysis of the supersingular locus.

1. Introduction

Let n be an integer and p a prime. It is known that if $n \geq 3$ and $p \geq 5$, or if $n \geq 1$ and $p \geq 11$, the modular curve $X_0(p^n)$ does not have a model with good reduction over the ring of integers of any complete subfield of $\mathbb{C}_p$. By a model for a scheme C over a complete local field K, we mean a scheme S over the ring of integers $\mathbb{O}_K$ of K such that $C \cong S \otimes_{\mathbb{O}_K} K$. When a curve C over K does not have a model with good reduction over $\mathbb{O}_K$, it may have the "next best thing", that is, a stable model. The stable model is unique up to isomorphism if it exists, which it does over the ring of integers in some finite extension of K, as long as the genus of the curve is at least 2. Moreover, if $\mathscr{C}$ is a stable model for C over $\mathbb{O}_K$, and $K \subseteq L \subseteq \mathbb{C}_p$, then $\mathscr{C} \otimes_{\mathbb{O}_K} \mathbb{O}_L$ is a stable model for $C \otimes_K L$ over $\mathbb{O}_L$. The special fiber of any stable model for C is called the stable reduction.

Here is a brief summary of prior results regarding the stable models of modular curves at prime power levels. Deligne and Rapoport [1973, §VI.6] found models for $X_0(p)$ and $X_1(p)$ over $\mathbb{Z}_p$ and $\mathbb{Z}_p[\mu_p]$ that become stable over the quadratic unramified extension. Edixhoven [1990, Theorem 2.1.2] found stable models for $X_0(p^2)$ over the ring of integers, R, in the Galois extension of $\mathbb{Q}_p^{\mathrm{unr}}$ of degree $(p^2 - 1)/2$. Bouw and Wewers [2004, Theorem 4.1 and Corollary 3.4] found stable models of $X_0(p)$ and $X(p)$ over $\mathbb{Z}_p$ and R by completely different means. Krir [1996, Théorème 1] proved that the Jacobian of $X_0(p^n)$ has a semistable

model over the ring of integers of an explicit Galois extension L_n of $\mathbb{Q}_p^{\text{unr}}$ of degree $p^{2(n-2)}(p^2 - 1)$ for $n \geq 2$, which implies that $X_0(p^n)$ has a stable model over the ring of integers of L_n by [Deligne and Mumford 1969, Theorem 2.4]. Also, stable models for $X_0(125)$ and $X_0(81)$ were computed explicitly in [McMurdy 2004, §2; 2008, §3], and [2008, §5] gave a conjectural stable reduction of $X_0(p^4)$. The main result of this paper is the construction of a stable model for $X_0(p^3)$, when $p \geq 13$, over the ring of integers of some finite extension of $\mathbb{Q}_p$ that is made explicit in [CM 2006].

We introduce the notion of a semistable covering of a smooth complete curve over a complete nonarchimedean field in Section 2C. We prove that any curve over a complete stable subfield of $\mathbb{C}_p$ has a semistable covering if and only if it has a semistable model, and moreover we can determine the corresponding reduction from the covering (see Theorem 2.36). Finding a semistable covering is often easier in practice than finding a semistable model directly, and this is what we do for $X_0(p^3)$ in Sections 6–9.

Overview. Our approach is rigid analytic, in that we construct a stable model of $X_0(p^3)$ by actually constructing a stable covering by wide opens (an equivalent rigid analytic notion which was introduced in [Coleman and McCallum 1988, §1]). A covering $\mathscr{C}^o$ of the ordinary locus can be obtained by extending the ordinary affinoids $\mathbf{X}_{ab}^{\pm}$ defined in [Coleman 2005, §1] to wide open neighborhoods $W_{ab}^{\pm}$. The supersingular locus essentially breaks up into the union of finitely many deformation spaces of height 2 formal groups with level structure [Lubin et al. 1964]. We use results from [Hopkins and Gross 1994] and [de Shalit 1994] to produce a covering $\mathscr{C}^s$ of this region. Finally, we show that the genus of the covering $\mathscr{C}^o \cup \mathscr{C}^s$ is at least the genus of $X_0(p^3)$, and therefore that the overall covering is stable. This argument is laid out as follows.

First, in Section 2, we recall or prove the general rigid analytic results that are necessary for a stable covering argument. These results are proved not only over complete subfields of $\mathbb{C}_p$, but over more general complete nonarchimedean-valued fields. For example, Proposition 2.34 is the aforementioned result that the genus of any stable covering must equal the genus of the curve. We also revise and extend results of Bosch, and of Bosch and Lütkebohmert, on the rigid geometry of algebraic curves. A rigid analytic version of the Riemann existence theorem is proved in Appendix A.

In Section 3, we recall and fix notation for some results specifically pertaining to $X_0(p^n)$ and its rigid subspaces. This is done from the moduli-theoretic point of view, which is that points of $X_0(p^n)$ correspond to pairs (E, C), where E is a generalized elliptic curve and C is a cyclic subgroup of order p^n. There is a detailed discussion in Section 3A of the theory of the canonical subgroup of an elliptic curve

and its connection with the geometry of $X_0(p)$ [Buzzard 2003, §3]. Section 3B is where we define wide open neighborhoods, $W_{ab}^{\pm} \supseteq \mathbf{X}_{ab}^{\pm}$, of the irreducible affinoids that make up the ordinary locus of $X_0(p^n)$.

All of the necessary results regarding deformations of formal groups are given in Section 4. First we precisely state the relationship between deformations of elliptic curves and formal groups, which we call *Woods Hole theory* [Lubin et al. 1964, §6]. This is then used in Section 4A (along with the result of Howe in Appendix B) to prove that all of the connected components of the supersingular locus of $X_0(p^n)$ are (nearly) isomorphic. Because of this fact, we are able to focus on those regions $W_A(p^n)$ that correspond to a supersingular elliptic curve $A/\mathbb{F}_p$ for which $j(A) \neq 0, 1728$. Specifically, this enables us to directly apply results of de Shalit [1994, §3] for the forgetful map from $X_0(p)$ to the j-line. The other important consequence of Woods Hole theory is that it gives us a natural action of

$$\mathrm{Aut}(\hat{A}) \cong (\mathrm{End}(A) \otimes \mathbb{Z}_p)^*$$

on $W_A(p^n)$. In Section 4B we recall results from [Hopkins and Gross 1994] that describe this action in great detail, and we derive the specific consequences that we need for our analysis of $X_0(p^3)$.

Once the groundwork has been laid, the remaining sections are devoted to constructing stable coverings of $X_0(p^2)$ and $X_0(p^3)$. In Section 5 we construct a stable covering for $X_0(p^2)$ over an explicit Galois extension of $\mathbb{Q}_p$ of degree $12(p^2 - 1)$, essentially showing that the wide open subspaces defined in Section 3 are sufficient. To be more precise, the stable covering consists of

$$\{W_{20}, W_{11}^+, W_{11}^-, W_{02}\} \cup \{W_A(p^2) : A \text{ is supersingular}\}.$$

This reproves Edixhoven's [1990] result from the point of view of this paper. It also gives a moduli-theoretic interpretation to the wide opens and underlying affinoids in the stable covering.

As in the stable covering of $X_0(p^2)$, the ordinary region of $X_0(p^3)$ is covered by six wide opens: $W_{30}, W_{21}^{\pm}, W_{12}^{\pm}$, and W_{03}. Unlike $W_A(p^2)$, however, $W_A(p^3)$ must itself be covered by smaller wide opens, since its reduction contains multiple irreducible components. First of all, the reduction of $W_A(p^3)$ contains two isomorphic lifts of some supersingular component of $X_0(p^2)$, with each meeting exactly three of the ordinary components. These two "old" components are connected through a central genus-0 component that we call the *bridging component*. To complete the picture, the bridging component then meets (in distinct points) a certain number of isomorphic copies of the curve $y^2 = x^p - x$. A partial picture of the stable reduction of $X_0(p^3)$, including one complete supersingular region (corresponding to a fixed supersingular curve A) and the six ordinary components, is given in Figure 1. The
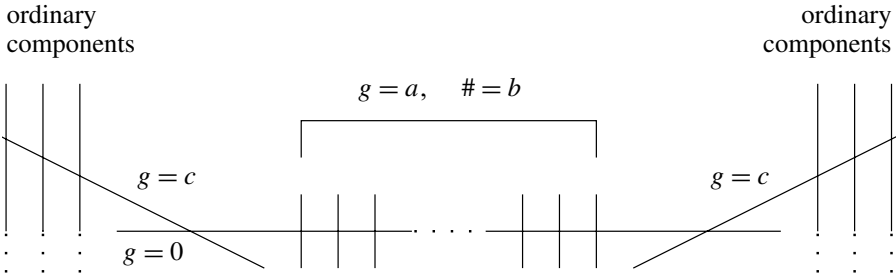
Figure 1. Partial picture of the stable reduction of $X_0(p^3)$.

number and genera of the components, as marked on the graph, are as follows:

$$(a, b, c) = \begin{cases} ((p-1)/2, 2(p+1)/3, (p-5)/6) & \text{if } j(A) = 0, \\ ((p-1)/2, p+1, (p-3)/4) & \text{if } j(A) = 1728, \\ ((p-1)/2, 2(p+1), (p-1)/2) & \text{otherwise.} \end{cases}$$

Complete graphs with intersection multiplicities are given in Section 9A. As a consequence of these results, it follows that the new part of $J_0(p^3)$ has potential good reduction isogenous to the product of $(p^2 - 1)/6$ copies of the Jacobian of $y^2 = x^p - x$.

It should be noted that the field of definition of our stable covering ultimately depends on the field of definition of certain elliptic curves that have "fake CM". In [CM 2006] we proved results about these fake CM curves that then made it possible for us to define the stable model over the ring of integers of an explicit finite extension of $\mathbb{Q}_p$ and compute the associated Weil group action, assuming the results of this paper. In [CM 2006] we also dealt with the $p \leq 11$ cases explicitly and computed the stable reduction of $X_0(Np^3)$ for $(N, p) = 1$. We expect that our methods will extend to $X_0(Np^n)$, and will have applications to modular forms as in [CM 2006, Remark 6.10]. We understand that Wewers also has a different approach with applications to local Langlands.

2. Rigid analytic foundation

We fix some notation for the p-adic analysis and more general nonarchimedean analysis. Throughout this section, unless otherwise stated, we let K be a complete nonarchimedean valued field with absolute value $|\cdot|$. We denote the ring of integers of K by R_K, its maximal ideal by m_K, and the residue field by $\mathbb{F}_K$. Let p be the characteristic of $\mathbb{F}_K$ (which we allow to be 0). Let $\mathbf{C}$ be the completion of an algebraic closure of K, and denote its ring of integers, maximal ideal, and residue field by $\mathbf{R}$, $m_{\mathbf{R}}$, and $\overline{\mathbb{F}}$. Note that $\overline{\mathbb{F}}$ is then an algebraic closure of $\mathbb{F}_K$. Whenever

$\mathbb{F}_K$ is perfect and has positive characteristic, we let $W(\mathbb{F}) \subseteq \mathbf{R}$ denote the ring of Witt vectors for any field $\mathbb{F} \subseteq \overline{\mathbb{F}}$. The value group of K will be denoted $|K^*|$, and we let

$$\mathfrak{R} := \mathfrak{R}_K = \{x \in \mathbf{R} : x^n \in |K^*| \text{ for some } n \in \mathbb{N}\}$$

(equivalently, $\mathfrak{R} := |\mathbf{C}^*|$). Then if $S \subseteq \mathbf{R}$, we let $\mathfrak{R}S = \mathfrak{R} \cap S$.

Occasionally, for technical reasons, we will need to assume that K is a stable field [Bosch et al. 1984, Definition 3.6.1/1]. By [1984, Proposition 3.6.2/6], this is the case if and only if $e(L/K)f(L/K) = [L:K]$ for all finite extensions L/K, where $e(L/K) = |L^*|/|K^*|$ and $f(L/K) = [\mathbb{F}_L : \mathbb{F}_K]$ are the ramification index and residue degree of L over K. There are also two special cases that we will consider for certain results. First, for a fixed prime p, let $\mathbb{C}_p$ be the completion of a fixed algebraic closure of $\mathbb{Q}_p$, let $\mathbb{R}_p \subseteq \mathbb{C}_p$ be its ring of integers, and let $m_{\mathbb{R}_p}$ be the maximal ideal of $\mathbb{R}_p$. Let υ denote the unique valuation on $\mathbb{C}_p$ with $\upsilon(p) = 1$, and $|\cdot|$ the absolute value given by $|0| = 0$ and $|x| = p^{-\upsilon(x)}$ for $x \neq 0$. In this case $\mathfrak{R} = |\mathbb{C}_p^*| = p^{\mathbb{Q}}$. Also $\mathbb{C}_p$ is stable, as is the completion of any tamely ramified extension of a finite extension of $\mathbb{Q}_p$. The second specific nonarchimedean valued field that will be considered is $\overline{\mathbb{F}}_p((T))$, for which the corresponding field $\mathbf{C}$ will be denoted Ω_p. Both $\overline{\mathbb{F}}_p((T))$ and Ω_p are stable, and in this case we have $\mathfrak{R} = |T|^{\mathbb{Q}}$.

Hypothesis T. The field $\mathbf{C}$ is isomorphic to either $\mathbb{C}_p$ or Ω_p.

In fact, for our purposes, this hypothesis can be relaxed to "$\mathbf{C}$ is an immediate extension[1] of $\mathbb{C}_p$ or Ω_p".

Remark 2.1. Suppose K satisfies Hypothesis T. Then if A is an Abelian variety over K and $P \in A(\mathbf{C})$, then 0 is in the closure of $\{nP : n \in \mathbb{N}\}$; see the proof of Lemma 2.19.

Now, for $r \in \mathfrak{R}$, we let $B_K^d[r]$ and $B_K^d(r)$ denote the closed and open d-dimensional polydisks over K of radius r around 0, that is, the rigid spaces over K whose $\mathbf{C}$-valued points are $\{(x_1, \ldots, x_d) \in \mathbf{C}^d : |x_i| \leq r\}$ and $\{(x_1, \ldots, x_d) \in \mathbf{C}^d : |x_i| < r\}$, respectively. In particular, let $B_K[r] := B_K^1[r]$ and $B_K(r) := B_K^1(r)$ denote the *closed disk* and *open disk* of radius r around 0. If $r, s \in \mathfrak{R}$ and $r \leq s$, let $A_K[r,s]$ and $A_K(r,s)$ be the rigid spaces over K whose $\mathbf{C}$-valued points are $\{x \in \mathbf{C} : r \leq |x| \leq s\}$ and $\{x \in \mathbf{C} : r < |x| < s\}$, which we call *closed annuli* and *open annuli*. The *semiopen annuli*, $A_K[r,s)$ and $A_K(r,s]$, are similarly defined. The width of such an annulus is defined to be $\log_p(s/r)$ or $\ln(s/r)$ if $p = 0$. Note that all closed or open disks over K, and all closed or open annuli over K of the same width, are potentially isomorphic. Here and throughout the paper, we use the adverb "potentially" in various contexts to mean "after finite base extension." A closed

---

[1] In the classical theory, an extension of valued fields is said to be immediate if the corresponding value groups and residue fields are isomorphic. This notion was introduced by Krull.

annulus of width 0 will be called a *circle*, and we will also denote the circle, $A_K[s, s]$, by $C_K[s]$.

If X is a rigid space over K and $f \in A(X) := \mathbb{O}_X(X)$, let $|f|_{\sup}$ denote the sup of $|f(x)|$ over all $x \in X(\mathbf{C})$. Then set

$$A^o(X) = \{f \in A(X) : |f|_{\sup} \le 1\},$$
$$A^+(X) = \mathrm{cl}\{f \in A(X) : |f|_{\sup} < 1\}, \text{ and}$$
$$\overline{A(X)} = A^o(X)/A^+(X),$$

where cl is the closure in $A^o(X)$. We define the reduction $\overline{X}$ of X to be the affine scheme Spec $\overline{A(X)}$. Suppose now that $X = \mathrm{Sp}\,(A)$ is an affinoid. Then $|f|_{\sup}$ is just the usual spectral seminorm of f, which we also denote by $\|f\|_X$ when X is re-duced and $|\cdot|_{\sup}$ is a norm. There is a canonical reduction map $\mathrm{Red}: X(\mathbf{C}) \to \overline{X}(\overline{\mathbb{F}})$, which we denote by $x \mapsto \bar{x}$. If X is reduced and $\widetilde{Y}$ is any subscheme of $\overline{X}$, then $Y := \mathrm{Red}^{-1}\widetilde{Y}$ is the rigid space admissibly covered by affinoid subdomains Z of X such that $\overline{Z}$ maps into $\widetilde{Y}$. As a special case, when $\widetilde{Y} \subseteq \overline{X}$ is an open affine, Y is the unique subaffinoid of X such that $Y(\mathbf{C}) = \{x \in X(\mathbf{C}) : \bar{x} \in \widetilde{Y}(\overline{\mathbb{F}})\}$, and we call Y a Zariski subaffinoid of X. When $\overline{X}$ is a reduced affine curve, we let $\overline{X}^c$ denote the unique complete curve that contains $\overline{X}$ as an affine open and is nonsingular at all other points (which we call the *points at infinity*).

If X is a rigid space over K, and $L \supseteq K$ is a complete subfield of $\mathbf{C}$, we write $P \in X(L)$ to mean that P is an L-valued point of X. An unspecified $P \in X$ should be read as $P \in X(\mathbf{C})$. We use the notations X_L and $X_{\mathbb{F}_L}$ for the extensions of X and $\overline{X}$ by scalars.

2A. *Annuli.*

Proposition 2.2. *Let $f : \mathscr{A}_1 \to \mathscr{A}_2$ be a degree d unramified surjection of annuli over K (open or closed). Then the width of $\mathscr{A}_1$ is $1/d$ times the width of $\mathscr{A}_2$.*

Proof. Extend scalars to $\mathbf{C}$, and choose isomorphisms $\psi_i : \mathscr{A}_i \to A_{\mathbf{C}}(r_i, 1)$ for some $r_i \in \mathscr{R}$ with $r_i < 1$. Let T be the natural parameter on $A_{\mathbf{C}}(r_1, 1)$. Then viewing $\tilde{f} := \psi_2 \circ f \circ \psi_1^{-1}$ as an invertible function on $A_{\mathbf{C}}(r_1, 1)$, we may write $\tilde{f}$ as either $cT^d(1 + g(T))$ or $cT^{-d}(1 + g(T))$, where $g(T) \in A^+(A_{\mathbf{C}}(r_1, 1))$. In the first case, for any $t \in A_{\mathbf{C}}(r_1, 1)$, we clearly have $|\tilde{f}(t)| = |c| \cdot |t|^d$. So by surjectivity of f, this implies that $|c| = 1$ and $r_1^d = r_2$. Thus, $\log_p(1/r_1) = (1/d)\log_p(1/r_2)$. The second case is very similar. $\qquad\square$

Definition 2.3. For any $r \in \mathbb{R}_+^* \setminus \mathscr{R}$, we let

$$K_r = \left\{\sum_{n \in \mathbb{Z}} a_n T^n : a_n \in K, \ \lim_{|n| \to \infty} |a_n| r^n = 0\right\}.$$

Then K_r is a field, and $f \mapsto \max\{|a_n|r^n\}$ is a valuation[2] if $f(T) = \sum_{n \in \mathbb{Z}} a_n T^n$. If $r_1, \ldots, r_n \in \mathbb{R}_+^*$ have linearly independent images in the $\mathbb{Q}$-vector space $\mathbb{R}_+^*/\mathfrak{R}$, we let $K_{r_1,\ldots,r_n} := (K_{r_1,\ldots,r_{n-1}})_{r_n}$ and $K_\varnothing = K$.

Then $K_{r_1,\ldots,r_n} \cong K_{r_1} \hat{\otimes}_K \ldots \hat{\otimes}_K K_{r_n}$, and its value group is generated by $\mathfrak{R}$ and $\{r_1, \ldots, r_n\}$ [Berkovich 1990, pp. 21–22]. If m is a positive integer, the map $f(T) \mapsto f(T^m)$ gives an injection from K_{r^m} into K_r for any r.

Lemma 2.4. *The group* $\mathrm{Aut}_{\mathrm{cont}}(K_{r_1,\ldots,r_n}/K)$ *contains a subgroup* H, *for which* $h \mapsto h^d$ *is a bijection whenever* $p \nmid d$, *and whose fixed field is* K.

Proof. It suffices to do the case $n = 1$. Let $r = r_1$. Suppose $\alpha \in K$ such that $|\alpha| < r$. If $f \in K_r$ and $f(T) = \sum_{n \in \mathbb{Z}} a_n T^n$, we set

$$f^{\sigma_\alpha}(T) = \sum_{n>0} a_{-n} \left(\frac{T^{-1}}{1 - \alpha T^{-1}} \right)^n + \sum_{n \geq 0} a_n (T - \alpha)^n.$$

Then $\sigma_\alpha \in \mathrm{Aut}_{\mathrm{cont}}(K_r/K)$, and if $a_n = 0$ for large $|n|$, then $f^{\sigma_\alpha}(T)$ is the image of the rational function $f(T - \alpha)$ in K_r. It follows, by continuity, that the map $\alpha \mapsto \sigma_\alpha$ is an injective homomorphism from the subgroup $B_r := \{\alpha \in K : |\alpha| < r\}$ of K^+ into $\mathrm{Aut}_{\mathrm{cont}}(K_r/K)$. Since $p \nmid d$, $\alpha \mapsto d\alpha$ is a bijection on B_r.

Now, if $f^{\sigma_\alpha}(T) = \sum_{n \in \mathbb{Z}} b_n T^n$, then

$$b_n = \sum_{m \geq n} \left((-1)^{m-n} \binom{m}{n} + \binom{-n-1}{-m-1} \right) a_m \alpha^{m-n},$$

where we set $\binom{a}{b} = 0$ if $a < 0$ or $b < 0$. Suppose $f^{\sigma_\alpha} = f$ for all $\alpha \in B_r$. Then in the formula above, we must have $b_n = a_n$ for all $\alpha \in B_r$. This can only happen if $a_n = 0$ for all $n \neq 0$. Therefore $f \in K$, and we may take H to be the image of B_r in $\mathrm{Aut}_{\mathrm{cont}}(K_r/K)$. $\square$

Lemma 2.5. *Let* X *be a reduced affinoid over* K, *and let* $f : X \to A_K[a, b]$ *be finite, flat, and of degree* d, *where* $p \nmid d$ *and* $a, b \in \mathfrak{R}$ *with* $a < 1 < b$. *Let* T *be the natural parameter on* $A_K[a, b]$. *Suppose there exists a function* G *on* $X[1] := f^{-1}C_K[1]$ *such that* $\|G^d - f^*T\|_{X[1]} < 1$. *Then there exist* $a_1, b_1 \in \mathfrak{R}[a, b]$ *with* $a_1 < 1 < b_1$, *and a function* S *on* $f^{-1}A_K[a_1, b_1]$, *such that* $S^d = f^*T$.

Proof. Setting $s = \overline{G}$ and $t = \overline{T}$, we have $\mathbb{O}(\overline{X[1]}) = \mathbb{F}_K[s, s^{-1}]$ and $\mathbb{O}(\overline{C[1]}) = \mathbb{F}_K[t, t^{-1}]$, and $\bar{f} : \overline{X[1]} \to \overline{C[1]}$ is given by $t = s^d$. Let $V = A_K[a, 1]$ and $U = f^{-1}(V)$. Then $\overline{C_K[1]}$ is an affine open in $\overline{V}$. Therefore, identifying $\mathbb{O}(\overline{U})$ with its image in $\mathbb{O}(\overline{X[1]})$, we have

$$\mathbb{O}(\overline{X[1]}) = \mathbb{O}(\overline{U}) \otimes_{\mathbb{O}(\overline{V})} \mathbb{O}(\overline{C_K[1]}).$$

---

[2]Some authors call this an absolute value.

Thus s is in the image of $\mathbb{O}(\overline{U})$. So we may lift s to a function $S_0 \in A^o(U)$ such that

$$\|S_0^d - f^*T\|_{X[1]} < 1.$$

Now, choose $a_1 \in \mathcal{R}[a, 1)$ such that $|S_0^d - f^*T| < |f^*T|$ on $U_1 := f^{-1}A_K[a_1, 1]$. Let $p(x) = x^d - (f^*T/S_0^d)$, considered as a polynomial over $A^o(U_1)$. Then $x_0 := 1$ satisfies $|p(x_0)| < 1$ and $|p'(x_0)| = 1$ over all of U_1. Therefore, by the usual Hensel's lemma argument, there exists a unique $x \in A^o(U_1)$ with $p(x) = 0$ and $\|x - 1\|_{U_1} < 1$. Letting $S_1 = S_0 x$, we have an $S_1 \in A(U_1)$ whose restriction to $X[1]$ is a lift of s, and for which $S_1^d = f^*T$.

By precisely the same argument, there is a function $S_2 \in A(U_2)$ that reduces to s on $X[1]$ and satisfies $S_2^d = f^*T$, where $U_2 = f^{-1}A_K[1, b_1]$ for some $b_1 \in \mathcal{R}(1, b]$. Also, since X is reduced, $(S_1/S_2)^d = 1$ on $X[1]$ (with $p \nmid d$), and $\|S_i - G\|_{X[1]} < 1$, we must have $S_1 = S_2$ on $X[1]$. Therefore, S_1 and S_2 patch to a function S on $f^{-1}A_K[a_1, b_1]$ with $S^d = f^*T$. □

Theorem 2.6. *Suppose $a < b \in \mathcal{R}$. Any finite connected étale cover over K of the annulus $A_K[a, b]$ (respectively $A_K(a, b)$) of degree d, where $d < p$ if $p \neq 0$, is an annulus isomorphic over K to $A_K[a^{1/d}c, b^{1/d}c]$ (respectively $A_K(a^{1/d}c, b^{1/d}c)$) for some $c \in |K^*|^{1/d}$.*

Proof. We will first prove the statement for closed annuli.

Let W be a connected rigid space over K, and let $f : W \to A_K[a, b]$ be finite and étale of degree $d < p$ (if $p \neq 0$). Initially, we also assume that $a, b \in |K^*|$. For each $r \in |K^*| \cap [a, b]$, let W_r be the inverse image in W of the circle $C_K[r]$. Then the connected components of W_r, which we denote by $\{V_{r1}, \ldots, V_{rm_r}\}$, are affinoids over K, with each V_{ri} finite and étale of degree d_{ri} over $C_K[r]$, such that $\sum d_{ri} = d$. As $d < p$ or $p = 0$, each $\overline{V}_{ri}$ must be finite and étale of degree d_{ri} over $\overline{C_K[r]} \cong \mathbf{G}_m$. Thus, there must exist an isomorphism $\sigma_{ri} : V_{ri} \to C_K[r^{1/d_{ri}}]$ such that $f \circ \sigma_{ri}^{-1}$ reduces to $x \mapsto x^{d_{ri}}$ on $\mathbf{G}_m$ (with respect to the standard parameters). Moreover, this implies by [Lemma 2.5](#) that for each $r \in |K^*| \cap (a, b)$ there exist $\alpha_r, \beta_r \in \mathcal{R}[a, b]$ with $\alpha_r < r < \beta_r$, and an embedding

$$F_r : \coprod_{i=1}^{m_r} A_K(\alpha_r^{1/d_{ri}}, \beta_r^{1/d_{ri}}) \hookrightarrow W$$

such that $\mathrm{Im}(F_r) = f^{-1}A_K(\alpha_r, \beta_r)$. In fact, F_r^{-1} can be defined on the i-th component of $f^{-1}A_K(\alpha_r, \beta_r)$ by a parameter S_{ri} such that $S_{ri}^{d_{ri}} = f^*T$ (where T is the natural parameter on $A_K(\alpha_r, \beta_r)$). Similarly, we have embeddings F_a and F_b, each of a disjoint union of semiopen annuli into W, with images $f^{-1}A_K[a, \beta_a)$ and $f^{-1}A_K(\alpha_b, b]$.

Suppose further that $[a, b] = [a, \beta_a) \cup (\alpha_b, b] \cup \bigcup_{r \in |K^*| \cap (a,b)} (\alpha_r, \beta_r)$. Then by compactness of $[a, b]$, we may choose a finite set $\{r_1, \ldots, r_n\} \subset |K^*| \cap (a, b)$ such

that $[a, b]$ is covered by $[a, \beta_a)$, $(\alpha_b, b]$ and the intervals $(\alpha_{r_i}, \beta_{r_i})$ for $1 \leq i \leq n$. Whenever two of these intervals overlap, it is clear from the properties of F_r that the inverse images in W of the corresponding subannuli of $A_K[a, b]$ must have the same number of connected components. Therefore, as W is connected, it follows that $m_r = 1$ for all $r \in |K^*| \cap [a, b]$. Thus, F_r is an isomorphism of $A_K(\alpha_r^{1/d}, \beta_r^{1/d})$ onto $f^{-1} A_K(\alpha_r, \beta_r)$, given by a parameter S_r with $S_r^d = f^*T$ (for $r \in |K^*| \cap (a, b)$, and similarly for $r = a$ or b). We claim that F_a, F_b, and the F_{r_i} can be used to construct an isomorphism of $A_K[a^{1/d}, b^{1/d}]$ onto W. Indeed, whenever $(\alpha_{r_i}, \beta_{r_i}) \cap (\alpha_{r_j}, \beta_{r_j}) = (\alpha_{r_j}, \beta_{r_i})$, we have a parameter S_{r_i} on $f^{-1} A_K(\alpha_{r_i}, \beta_{r_i})$ such that $S_{r_i}^d = f^*T$, and likewise for r_j. After adjusting by a d-th root of unity in K if necessary, S_{r_i} and S_{r_j} agree on $f^{-1} A_K(\alpha_{r_j}, \beta_{r_i})$. Therefore the two parameters patch to a parameter S_{ij} that identifies $f^{-1} A_K(\alpha_{r_i}, \beta_{r_j})$ with $A_K(\alpha_{r_i}^{1/d}, \beta_{r_j}^{1/d})$. After finitely many such patching steps, we have constructed a parameter S on W over K such that S^d equals f^*T and thus defines an isomorphism from W onto $A_K[a^{1/d}, b^{1/d}]$.

More generally, without making the above two suppositions, for each $r \in [a, b]$ take M_r to be a finite Galois extension of K such that $r \in |M_r^*|$ if $r \in \mathcal{R}$ and K_r (defined as above) otherwise. Then we may choose $\alpha_r, \beta_r \in \mathcal{R}_{M_r}[a, b]$ and an embedding F_r that is defined over M_r, precisely as was done over K. Now, we know that $[a, b]$ is covered by $[a, \beta_a)$, $(\alpha_b, b]$, and $\{(\alpha_r, \beta_r) : r \in (a, b)\}$. So by compactness, there exists a finite set $t_1, \ldots, t_m \in (a, b)$ such that $[a, b]$ is covered by $[a, \beta_a)$, $(\alpha_b, b]$, and $\{(\alpha_{t_i}, \beta_{t_i}) : 1 \leq i \leq m\}$. Choose a finite Galois extension L of K so that the images of the t_i in $\mathbb{R}_+^*/|L^*|$ generate a torsion-free abelian group. Then choose $r_1, \ldots, r_n \in \mathbb{R}_+^*$ so that their images form a basis for this group. Then the argument above can be applied to produce a parameter S on W, which is defined over $L_{r_1, \ldots, r_n}$ such that $S^d = f^*T$.

Now, if $\sigma \in \mathrm{Aut}_{\mathrm{cont}}(L_{r_1, \ldots, r_n}/L)$, the map $\sigma \mapsto \zeta(\sigma) := S^\sigma/S$ is a 1-cocycle with values in $\mu_d(A(W_{L_{r_1, \ldots, r_n}}))$. Since W is connected, this equals $\mu_d(L_{r_1, \ldots, r_n})$, which is $\mu_d(L)$. It follows from [Lemma 2.4](#) that $\zeta(\sigma) = 1$ for all σ in a subgroup whose fixed field is L. Thus S is defined over L. Then, for $\sigma \in \mathrm{Gal}(L/K)$, $S^\sigma = h(\sigma)S$, where h is a 1-cocycle. So by Hilbert's Theorem 90 there exists $\gamma \in L^*$ such that $h(\sigma) = \gamma^\sigma/\gamma$. Then $H := S/\gamma$ is defined over K and $H^d = \alpha T$ for some $\alpha \in K^*$. Therefore H defines an isomorphism of W onto $A_K[a^{1/d}c, b^{1/d}c]$, where $c = |\alpha|^{1/d}$.

To deal with open annuli $A_K(a, b)$, choose sequences $\{a_n\}$ and $\{b_n\}$ in $\mathcal{R}_K(a, b)$ such that $a_n < b_n$, $a_n \to a$ and $b_n \to b$. For large n, $W_{[a_n, b_n]} := f^{-1} A_K[a_n, b_n]$ is connected, and it is finite étale over $A_K[a_n, b_n]$ of degree d. Therefore, it is isomorphic to $A_K[a_n^{1/d} c_n, b_n^{1/d} c_n]$ by what we have proven. The theorem follows when we let n go to infinity. $\qquad\square$

Remark 2.7. (i) When K is algebraically closed, there exists $a = c_0 < \cdots < c_{n+1} = b$ in $\mathfrak{R}$ such that $f^{-1} A(c_i, c_{i+1})$ is a disjoint union of open annuli [Lütkebohmert 1993, Lemma 2.3]. One could then use Hilbert's Theorem 90 and Lemma 2.5, as in the proof above, to give another proof of the theorem.

(ii) One can obtain the same conclusion about W, for any finite étale surjection f whose Galois closure has degree prime to p when $p \neq 0$.

If X is a reduced affinoid over K and $P \in \overline{X}(\mathbb{F}_K)$, we let $R_X(P)$ denote the *residue class* of P. When the context makes it clear, we will drop the subscript X. This is the open rigid subspace of X whose $\mathbf{C}$-valued points reduce to P, or equivalently, the subspace $\mathrm{Red}^{-1} P$, where P is naturally identified with a subscheme of $\overline{X}$. Alternatively, suppose $f_1, \ldots, f_m \in A^o(X)$ are such that $\bar{f}_1, \ldots, \bar{f}_m$ generate the maximal ideal of P. Then $R(P)$ is admissibly covered by the increasing sequence of affinoids whose $\mathbf{C}$ valued points are

$$\{x \in X(\mathbf{C}) : |f_i(x)| \leq r_n, \ 1 \leq i \leq m\},$$

where $r_n \in \mathfrak{R}$, $r_n < r_{n+1}$ and $\lim_{n \to \infty} r_n = 1$. If x is a point of X such that $\bar{x} = P$ (which always exists by [Tate 1971, Theorem 6.4]), this is naturally isomorphic to the formal fiber $X_+(x)$ of Bosch (by [1977a, Satz 6.1]).

Proposition 2.8. *Let K be a stable field. Suppose X is a reduced pure d-dimensional affinoid over K, $\|A(X)\| = |K|$ (equivalently, $A^o(X) \otimes_{R_K} \mathbb{F}_K$ is reduced), and $P \in \overline{X}(\mathbb{F}_K)$. Then $\overline{A(R(P))} \cong \hat{\mathbb{O}}_{\overline{X}, P}$.*

Proof. Let $I(P)$ be the closure of $m_K A^o(R(P))$ in $A^o(R(P))$. Bosch [1977a, p. 44] proved that $A^o(R(P))/I(P) \cong \hat{\mathbb{O}}_{\overline{X}, P}$ when there exists a surjective map $\phi : T_n \to A(X)$ such that $\mathring{\phi}$ is surjective.[3] That such a map exists when K is stable and $\|A(X)\|_X = |K|$ follows from [Bosch et al. 1984, Corollary 6.4.3/6]. It is clear that $I(P) \subseteq A^+(R(P)) \subseteq \mathrm{rad}(I(P))$. Since $\overline{X}$ is reduced, so is $\hat{\mathbb{O}}_{\overline{X}, P}$, and hence $I(P) = A^+(R(P))$. The proposition follows. $\qquad\square$

Definition 2.9. Let P be a point on a curve C over a field k. We say that P is an *ordinary double point* over k if $\hat{\mathbb{O}}_{C, P} \cong k[\![u, v]\!]/(uv)$.

Hypothesis B. R_K contains a bald subring [Bosch et al. 1984, Definition 1.7.2/1] with the same residue field.

K satisfies Hypothesis B if it is discretely valued, if its residue field is perfect, or if its residue field lifts to a subfield. In particular, this is the case if K satisfies Hypothesis T. We do not know if all complete, nonarchimedean-valued fields K satisfy Hypothesis B.

---

[3] As the example at the end of [Bosch et al. 1984, §6.4] implies, ϕ need not be distinguished; see [Bosch et al. 1984, Definition 6.4.3/2].

Proposition 2.10. *Let X be a reduced, irreducible affinoid over a stable field K satisfying Hypothesis B. Suppose that $\overline{X}$ is a reduced curve and $P \in \overline{X}(\mathbb{F}_K)$. Then P is an ordinary double point over $\mathbb{F}_K$ if and only if the residue class $R(P)$ is isomorphic to $A_K(r, 1)$ for some $r \in |K^*|$.*

This was proven in [BL 1985, Proposition 2.3] when K is algebraically closed, and we adapt their proof to our case here.

Lemma 2.11. *Let I be a bald subring of R_K, and $\{r_1, r_2, \dots\}$ a zero sequence in R_K. Then there exists a bald subring of R_K containing I and r_n for all $n \geq 1$.*

Proof. The proof is almost identical to that of [Bosch et al. 1984, Corollary 1.7.2/5]; just replace the I in the proof of 1.7.2/4 with this I. $\square$

Lemma 2.12. *Let X be a reduced, one-dimensional affinoid, with reduced reduction, over a stable field K satisfying Hypothesis B. Suppose that $f, g \in C := A(\overline{X})$ generate a maximal ideal $\mathcal{M} = (f, g)$, such that $C/\mathcal{M} \cong \mathbb{F}_K$ and $fg \in \mathcal{M}^3$. Let*

$$U = \begin{cases} X & \text{if } fg = 0, \\ \{x \in X : f(\bar{x}) = g(\bar{x}) = 0\} & \text{otherwise.} \end{cases}$$

Then there exist $F, G \in A^o(U)$ and $c \in R_K$ such that

$$\overline{F} - f \in \mathcal{M}^2 \overline{A(U)}, \quad \overline{G} - g \in \mathcal{M}^2 \overline{A(U)}, \quad \text{and} \quad FG = c,$$

where we use Proposition 2.8 to identify $\overline{A(U)}$ with $\hat{\mathcal{O}}_{\overline{X}, \overline{U}}$.

Proof. Suppose that $f_1, g_1 \in A^o(X)$ are such that $f = \bar{f}_1$ and $g = \bar{g}_1$, and that $\alpha : X \to B_K[1]$ is a finite morphism. That $\overline{X}$ is reduced implies $\|A(X)\|_X = |K|$. So by [Bosch et al. 1984, Corollary 6.4.1/4], $\alpha^* : A^o(B_K[1]) \to A^o(X)$ is finite. Now suppose $\alpha^*(A^o(B_K[1])) = R_K\langle T \rangle$. As C is torsion-free (because $\overline{X}$ is flat over $\mathbf{A}^1$) and finitely generated over $\mathbb{F}_K[T]$, it is free. Choose $h_1, \dots, h_n \in A^o(X)$ so that $\bar{h}_1, \dots, \bar{h}_n$ is a basis for C over $\mathbb{F}_K[T]$. Then $h_1, \dots, h_n$ is a basis for $A^o(X)$ over $R_K\langle T \rangle$. Thus $B := \{h_i T^j : 1 \leq i \leq n, \ j \in \mathbb{N}_0\}$ is an orthonormal Schauder basis [ibid., Definition 2.7.2/1] for $A(X)$. As $C = \mathcal{M} \oplus \mathbb{F}_K$, the ring C has a basis over $\mathbb{F}_K$ of the form $\{1, \bar{\alpha}_i f, \bar{\beta}_j g : i, j \in \mathbb{N}\}$ for some α_i, β_j in a subring of $A^o(X)$ finitely generated over a bald subring of R_K with residue field $\mathbb{F}_K$. It follows from Lemma 2.11 that the change of basis matrix from B to $\{1, \alpha_i f_1, \beta_j g_1 : i, j \in \mathbb{N}\}$ has entries in a bald subring of R_K [ibid., Definition 1.7.2/1]. Hence by the lifting theorem of [ibid., Theorem 2.7.3/2], this is also an orthonormal Schauder basis. Hence $A^o(X) = R_K + M$, where $M = (f_1, f_2)$.

We have

$$f_1 g_1 = \pi c_1 + f_1(\pi a_1 + b_1) + g_1(\pi a_2 + b_2),$$

with $c_1 \in R_K$, $a_i \in A^o(X)$, $b_i \in M^2$ (and $b_i = 0$ if $fg = 0$) for some $\pi \in R_K$, $|\pi| < 1$. Let $I = \pi R_K + f_1 A^o(X) + g_1 A^o(X) = \pi A^o(X) + M$, and let $J = \pi A^o(X)$

if $fg = 0$ and I otherwise. Let $f_2 = f_1 - (\pi a_2 + b_2)$, and $g_2 = g_1 - (\pi a_1 + b_1)$. Then

$$\begin{aligned}
f_2 g_2 &= \pi c_1 + (\pi a_1 + b_1)(\pi a_2 + b_2) \\
&\equiv \pi c_1 + \pi^2 c_2' \quad \text{mod } (\pi A^o(X) + M)^3 \\
&\equiv \pi c_1 + \pi^2 c_2' \quad \text{mod } \pi^2 M \quad \text{if } fg = 0,
\end{aligned}$$

for some $c_2' \in R_K$. Now $I^n = \pi^n R_K + f_1 I^{n-1} + g_1 I^{n-1}$, so this implies

$$f_2 g_2 = \pi c_1 + \pi^2 c_2 + f_1 r_{2,1} + g_1 r_{2,2},$$

where $c_2 \in R_K$, $r_{2,i} \in J^2$. Let $k_n = 2^{n-2} + 1$ for $n \geq 2$ and $k_1 = 1$. Suppose

$$f_n g_n = \pi c_1 + \pi^2 c_2 + \pi^4 c_3 + \cdots + \pi^{2k_{n-1}} c_n + f_1 r_{n,1} + g_1 r_{n,2}$$

for some $r_{n,i} \in J^{k_n}$. Set $f_{n+1} = f_n - r_{n,2}$ and $g_{n+1} = g_n - r_{n,1}$. Then

$$\begin{aligned}
f_{n+1} g_{n+1} &= \pi c_1 + \pi^2 c_2 + \pi^4 c_3 + \cdots + \pi^{2k_{n-1}} c_n + r_{n,1} r_{n,2} \\
&= \pi c_1 + \pi^2 c_2 + \pi^4 c_3 + \cdots + \pi^{2k_{n-1}} c_n + \pi^{2k_n} c_{n+1} + f_1 r_{n+1,1} + g_1 r_{n+1,2},
\end{aligned}$$

where $r_{n+1,i} \in J^{k_{n+1}}$.

Finally, let $r_{1,1} = \pi a_1 + b_1$ and $r_{1,2} = \pi a_2 + b_2$. Set $F = f_1 - \sum_{n \geq 1} r_{n,2}$ and $G = g_1 - \sum_{n \geq 1} r_{n,1}$. Then these are elements of $A^o(U)$ that satisfy

$$FG = c := \pi c_1 + \pi^2 c_2 + \sum_{n \geq 3} \pi^{2^{n-2}+2} c_n. \qquad \square$$

Proof of Proposition 2.10. Suppose $P \in \overline{X}(\mathbb{F}_K)$ is an ordinary double point. We can apply Lemma 2.12 to conclude that there exist $F, G \in A^o(R(P))$ and $c \in R_K$ such that $(\overline{F}, \overline{G}) = \mathcal{M}_P$ and $FG = c$. Thus we have a morphism $R(P) \to A_K(|c|, 1)$. That this is an isomorphism follows, as in the proof of [BL 1985, Proposition 2.3].

Conversely, suppose that $R(P)$ is isomorphic to the annulus $A_K(r, 1)$ for some $r \in |K^*|$ with $r < 1$. Then $A^o(R(P)) \cong R_K[[T, cT^{-1}]]$, where $c \in K$ with $|c| = r$. So applying Proposition 2.8 we have

$$\hat{\mathcal{O}}_{\overline{X}, P} \cong \overline{A(R(P))} \cong \mathbb{F}_K[[x, y]]/(xy),$$

and hence P is an ordinary double point of $\overline{X}$. $\qquad \square$

For a rigid space W over K, set

$$D^i(W/K) = \text{Ker}(d : \Omega^i_{W/K}(W) \to \Omega^{i+1}_{W/K}(W))/d\Omega^{i-1}_{W/K}(W),$$

where if $A(W)$ is the ring of rigid functions on W, then $\Omega^i_{W/K}(W)$ is the $A(W)$ module of rigid i-forms on W.

Lemma 2.13. *Suppose $W = A_K(r, s)$ or $B_K(r) \setminus \{0\}$, where $r, s \in |L^*|$ for some finite extension L/K. Then*

$$D^0(W/K) \cong D^1(W/K) \cong K.$$

Proof. If $r, s \in |K^*|$, the lemma is clear. For in this case, we may choose $a, b \in K$ with $|a| = r$ and $|b| = s$, and let $x = T/b$ and $y = a/T$, where T is the natural parameter on $\mathbf{A}^1_K$. Then $A_K(W)$ is equal to the set of functions represented by

$$\sum_{n=0}^{\infty} a_n x^n + \sum_{n=1}^{\infty} b_n y^n,$$

where $a_n, b_n \in K$, $a_n t^n \to 0$ and $b_n t^n \to 0$ as $n \to \infty$, for $|t| < 1$. There is a natural continuous linear map ρ_K from $\Omega^1_{W/K}(W)$ onto K such that $\rho_K(dv/v) = 1$ for any parameter v on W_K such that $|v(u)| > |v(w)|$ if $|u| > |w|$ and $u, w \in A_K(r, s)(\mathbf{C})$. Moreover, for any $\omega \in \Omega^1_{W/K}(W)$, $\omega \in dA_K(W)$ if and only if $\rho_K(\omega) = 0$.

More generally, suppose L is a finite Galois extension of K with Galois group G, and that $r, s \in |L^*|$. Then G acts on $\Omega^1(A_L(r, s))$ such that $\Omega^1(A_L(r, s))^G = \Omega^1(A_K(r, s))$ and $\rho_L(\omega^\sigma) = \rho_L(\omega)^\sigma$. Also, if $r, s \in |K^*|$, then $\rho_L|_{\Omega^1(A_K)} = \rho_K$. Suppose $\omega \in \Omega^1(A_K(r, s))$ and $\rho_L(\omega) = 0$. Then Hilbert's additive Theorem 90 gives $\omega \in dA(A_K(r, s))$. Thus we have an injective K-linear map $D^1(W/K) \to L^G = K$. If $\omega \in \Omega^1(A_L(r, s))$, $\rho_L(\omega) = 1$ and $v = \sum_{\sigma \in G} \omega^\sigma$, then $v \in \Omega^1(A_K(r, s))$ and $\rho_L v = [L : K]$. So this map is an isomorphism. $\square$

From the proof we see that for any open annulus W over K, there are two residue maps from Ω^1_W onto K. In particular, they are $\mathrm{res}_{r,s} \circ f^*$ and $-\mathrm{res}_{r,s} \circ f^*$, where $f : A_L(r, s) \to W_L$ is an isomorphism and $\mathrm{res}_{r,s} = \rho_L|_{\Omega^1(A_K)}$ for any extension L of K such that $r, s \in |L^*|$. By an oriented annulus over K, we mean a pair (W, ρ), where W is an open annulus and ρ is a choice of one of the residue maps.

An *end* of a rigid space W over K is an element of the inverse limit of the set of connected components of $W \setminus Z$, where Z ranges over finite unions of subaffinoids of W defined over K (ordered by containment). We let $\mathscr{E}(W)$ denote the set of ends of W, and we let $e(W) = |\mathscr{E}(W)|$ (which may be infinite). For example, $e(W) = 2$ whenever W is an open annulus. If W is admissibly covered by a countable number of affinoids, and f is a real-valued function of $W(\mathbf{C})$, it makes sense to compute the limit of f at an end $e \in \mathscr{E}(W)$. In particular, we define $\lim_{x \to e} f(x) = \lim_{n \to \infty} f(x_n)$, where $\{x_n\}$ is any sequence in $W(\mathbf{C})$ such that for any Z as above, x_n is contained in a connected component of $W \setminus Z$ that maps to e for sufficiently large n (provided this limit exists and is independent of the sequence).

The following result is used in the proof of [CM 2006, Theorem 5.2].

Proposition 2.14. *Suppose K is discretely valued and U is a rigid space over K with two ends such that for some finite extension L of K, U_L is isomorphic to the open annulus $A_L(|u|, 1)$, where $u \in K^*$ and $|u| < 1$. Then $U \cong A_K(|u|, 1)$.*

Proof. We may suppose that $e(L/K) > 1$ and L is a Galois extension of K with Galois group G. Let $M = A(U_L)$. Then G naturally acts on M, and $M^G = A(U)$. Let $R = R_L$, and let $\mathbb{F}_L$ and $\mathbb{F}_K$ denote the residue fields of L and K. Let π be a uniformizing parameter on L.

Let a and b denote the ends of U, and suppose $F \in M$ is an isomorphism from U_L onto $A_L(|u|, 1)$ such that $\lim_{x \to a} |F(x)| = 1$. Then we may use F to identify M with $L\{\{T, u/T\}\}$, and the group M_a of orientation-preserving automorphisms of $A_L(|u|, 1)$ with (under composition)

$$\left\{ T\left(\sum_{i=0}^{\infty} a_i T^i + \sum_{i=1}^{\infty} b_i (u/T)^i \right) : a_i, b_i \in R \text{ and } a_0 \in R^* \right\}.$$

The group G preserves M_a. For $\sigma \in G$, set $\sigma(T) = G_\sigma(T)$. For

$$h(T) = \sum_{i=0}^{\infty} a_i T^i + \sum_{i=1}^{\infty} b_i (u/T)^i \in L\{\{T, u/T\}\},$$

set

$$h^\sigma(T) = \sum_{i=0}^{\infty} a_i^\sigma T^i + \sum_{i=1}^{\infty} b_i^\sigma (u/T)^i.$$

Then

$$G_\tau^\sigma \circ G_\sigma = G_{\sigma\tau}. \tag{1}$$

We will show that there exists $F \in M_a$ such that

$$F^\sigma \circ F^{-1} = G_\sigma. \tag{2}$$

This will imply that $F^{-1}(T) \in A(U)$, and as $F^{-1}(T)$ is a parameter on U, it will then follow that U is an annulus over K.

So first let I be the ideal in $C := R[\![T, u/T]\!]$ generated by π, T and u/T, and suppose that

$$G_\sigma(T) \equiv a(\sigma)T \mod TI, \quad \text{where } a(\sigma) \in R^*.$$

Then, from (1), we have $a(\sigma)^\tau a(\tau) \equiv a(\sigma\tau) \mod \pi$. Using Hilbert's Theorem 90 applied to $\mathbb{F}_L/\mathbb{F}_K$, one can show there exists a $c \in R^*$ such that $c^\sigma/c \equiv a(\sigma) \mod \pi$. Let $h(T) = cT$. Then we have

$$(h^{-\sigma} \circ G_\sigma \circ h)(T) \equiv T \mod TI.$$

Now, suppose $G_\sigma(T) = T(1 + h_\sigma(T))$, where $h_\sigma(T) \in I^k$.

Lemma 2.15. *Suppose* $h(T) := \sum_{i=1}^{\infty} B_{-i}(u/T)^i + \sum_{i=0}^{\infty} B_i T^i$ *is in* C. *Then* $h(T) \in I^k$ *if and only if* $B_i \equiv 0 \bmod \pi^{k-|i|} R$.

Proof. Let S_k be the R-module of series whose coefficients satisfy the bounds above. The lemma is clearly true for $k = 0$. Suppose it is true for k. Let $\mathcal{T}$ be the continuous involution of the R-algebra C that takes T to u/T. Then I and S_k are preserved by $\mathcal{T}$. As $\pi^{k-i} T^i \in I^k$ for $0 \leq i \leq k$, and $T^{k+1} R[[T]] \subseteq I^{k+1}$, it follows that $S_{k+1} \subseteq I^{k+1}$. We have

$$\pi(\pi^{k-i} T^i) = \pi^{k+1-i} T^i \quad \text{and} \quad T(\pi^{k-i} T^i) = \pi^{k+1-(i+1)} T^{i+1},$$

and because $\upsilon(u) \geq 2\upsilon(\pi)$,

$$T(\pi^{k-i}(u/T)^i) = u\pi^{k-i}(u/T)^{i-1} \in \begin{cases} \pi^{k+1-(i-1)}(u/T)^{i-1} R & \text{if } i > 0, \\ \pi^{k+1-1} T R & \text{if } i = 0. \end{cases}$$

Thus $I^{k+1} \subseteq S_{k+1}$. $\qquad\square$

Now suppose

$$h_\sigma(T) = \sum_{i=1}^{\infty} B_{-i}(\sigma)(u/T)^i + \sum_{i=0}^{\infty} B_i(\sigma) T^i.$$

Then, since

$$T(1 + h_\tau(T))\big(1 + h_\sigma^\tau(T(1 + h_\tau(T)))\big) \equiv T(1 + h_\sigma^\tau(T) + h_\tau(T)) \mod TI^{2k},$$

it follows that

$$G_\sigma^\tau \circ G_\tau(T) \equiv$$
$$T\left(1 + \sum_{i=1}^{2k}(B_{-i}(\tau) + B_{-i}(\sigma)^\tau)(u/T)^i + \sum_{i=0}^{2k}(B_i(\tau) + B_i(\sigma)^\tau)T^i\right) \mod TI^{2k}.$$

Therefore, by [Lemma 2.15](#) we have

$$B_i(\sigma\tau) \equiv B_i(\tau) + B_i(\sigma)^\tau \mod \pi^{2k-|i|}.$$

Finally, using Hilbert's Theorem 90 again, we can find $C_i \in \pi^{k-|i|} R \cap R$ such that

$$C_i^\tau - C_i \equiv B_i(\tau) \mod \pi^{2k-|i|} \quad \text{for } -2k \leq i \leq 2k.$$

So let

$$H(T) = T\left(1 + \sum_{i=1}^{2k} C_{-i}(u/T)^i + \sum_{i=0}^{k^2-1} C_i T^i\right).$$

Then $H \in TI^k$ and $H^\sigma \circ H^{-1} \equiv G_\sigma \mod TI^{2k}$. Thus we can find a convergent sequence $F_k \in M_a$ such that $F_k^\sigma \circ F_k^{-1} \to G_\sigma$ in M_a. The limit, $F \in M_a$, must satisfy (2). $\qquad\square$

Remark 2.16. Suppose K is discretely valued and U is a rigid space with one end over K, such that U_L is isomorphic to the open disk $B_L(1)$ for some finite extension L of K. Then it follows from a similar argument that $U \cong B_K(1)$.

2B. *Wide open spaces.* In [Coleman 1989, §III] we defined wide open spaces over $\mathbb{C}_p$. Now we need to use them over more general fields. Suppose W is a one-dimensional smooth rigid space over K. Then W is a wide open space, or wide open, over K if it contains affinoid subdomains X and Y such that

  (i) $W \setminus X$ is a disjoint union of finitely many open annuli,

 (ii) X is relatively compact in Y, and

(iii) $Y \cap V$ is a semiopen annulus for all connected components V of $W \setminus X$.

We call X an underlying affinoid of W. From the definition, it is immediate that there is a natural bijection between $\mathscr{E}(W)$, the set of ends of W, and $\mathrm{CC}(W \setminus X)$, the set of connected components of $W \setminus X$. And X is connected to each element of $\mathrm{CC}(W \setminus X)$. So $e(W)$ is finite in this case. We call the connected component of $W \setminus X$ that corresponds to an element e of $\mathscr{E}(W)$ an *annulus at e*.

Remark 2.17. It is not immediate that the intrinsic definition of a wide open space given above is equivalent to the one given in [Coleman 1989, §III] when $K = \mathbb{C}_p$. However, this will follow in one direction from Theorem 2.18 and in the other from Theorem 2.40.

Theorem 2.18. *Let W be a wide open over K with underlying affinoid X. Then W may be completed to a proper algebraic curve C over K by gluing open disks onto the connected components of $W \setminus X$.*

Proof. More specifically, let $\mathscr{S}$ be the set of connected components of $W \setminus X$. For each open annulus $V \in \mathscr{S}$, let $\alpha_V : V \to B_V$ be an embedding of V into an open disk over K such that $B_V \setminus \alpha_V(V')$ is connected for any concentric annulus $V' \subseteq V$ that is connected to X. We will show that

$$C := \left(W \cup \coprod_{V \in \mathscr{S}} B_V \right) \Big/ \{\alpha_V(V) = V\}_{V \in \mathscr{S}}$$

is isomorphic to a complete algebraic curve.

It is clear that C is smooth of dimension one. Therefore, to establish the claim, by the Riemann existence theorem (Theorem A.2), we need only show that C is proper [Bosch et al. 1984, Definition 9.6.2/2]. The number of connected components of W is finite and equals the number of connected components of X, and so we may assume without loss of generality that W is connected. In this case X is contained in a residue class $R(P)$ of Y (where P is the image of $\overline{X}$ in $\overline{Y}$). Choose an $f \in A^o(Y)$ such that P is an isolated zero of $\bar{f}$. This can be done by first passing to a finite extension L of K so that $\overline{Y}_L$ is reduced and so that there is such a function

$g \in A^o(Y_L)$. Then let f be the norm of g. Now by [BL 1985, Lemma 2.4], if $\alpha \in \mathfrak{R}$, and α is less than and sufficiently close to 1, then $\{x \in R(P) : |f(x)| \geq \alpha\}$ is the set of $\mathbf{C}$-valued points in a subdomain U_α of W which, after a finite extension (the field in that lemma is algebraically closed), becomes isomorphic to a finite union of semiopen annuli. In fact, for α sufficiently close to 1, U_α must decompose as $\coprod_{V \in \mathscr{S}} A_{\alpha,V}$, where each $A_{\alpha,V}$ is a concentric semiopen annulus in V. Thus $B_{1,V} := B_V \setminus \alpha_V(V \cap Y)$ and $B_{\alpha,V} := B_{1,V} \cup \alpha_V(A_{\alpha,V})$ are closed disks. Also, we may define X_α to be the rigid subdomain of W whose $\mathbf{C}$-valued points are $\{x \in R(P) : |f(x)| \leq \alpha\}$.

Then, $\mathscr{U} := \{X_\beta, B_{\beta,V} : V \in \mathscr{S}\}$ and $\mathscr{V} := \{Y, B_{\alpha,V} : V \in \mathscr{S}\}$, for any $\beta \in \mathfrak{R}$ with $\alpha < \beta < 1$, are two finite admissible affinoid coverings of C such that each element of $\mathscr{U}$ is relatively compact in an element of $\mathscr{V}$. So C is proper if it is separated. To verify that C is separated, we must show that the diagonal map $\Delta : C \to C \times_K C$ is a closed immersion. This can be checked locally using the admissible affinoid cover of $C \times_K C$ given by $\{Z \times_K Z' : Z, Z' \in \mathscr{U}\}$. Indeed, for every $Z, Z' \in \mathscr{U}$, $\Delta^{-1}(Z \times_K Z') = Z \cap Z'$ is an affinoid and $\Delta^* : A(Z \times_K Z') \to A(Z \cap Z')$ is surjective. This is obvious when $Z = Z'$. Otherwise, when $Z \cap Z' \neq \varnothing$ we must have $\{Z, Z'\} = \{X_\beta, B_{\beta,V}\}$ for some $V \in \mathscr{S}$. So in this case, $Z \cap Z'$ is a circle over K, and in particular the concentric circle in $V \cap Y$ defined by $|f(x)| = \beta$. To obtain surjectivity, first make a finite base extension L of K so that $\overline{(X_\beta)_L}$ and $\overline{(B_{\beta,V})_L}$ are reduced. Then $\mathbb{O}(\overline{(X_\beta)_L})$ is isomorphic to a subring of

$$\mathbb{F}_L[t_1, \ldots, t_N]/(t_i t_j)_{i \neq j}$$

that contains a power of the ideal $(t_1, \ldots, t_N)$. Also, if t_i is the particular parameter corresponding to V, then $\mathbb{O}(\overline{(B_{\beta,V})_L})$ can be identified via the gluing map with $\mathbb{F}_L[t_i^{-1}]$. So Δ^* is surjective, as $\mathbb{O}(\overline{Z \cap Z'}) = \mathbb{F}_L[t_i, t_i^{-1}]$. Thus, C is separated over K [Bosch et al. 1984, Definition 9.6.1/1], and hence proper [Bosch et al. 1984, Definition 9.6.2/2]. Therefore, C is an algebraic curve by the Riemann existence theorem. $\square$

When a wide open W is completed to a curve C as above, the underlying affinoid X is the complement in C of a finite union of open disks. As we will now show, this results in a close connection between the reductions of C and the canonical reduction of X. Of particular interest will be the case when (W, X) is basic (defined below), in which case, provided K is stable and assuming Hypothesis T, we show that X is the minimal underlying affinoid and C has a model that reduces to $\overline{X}^c$.

Lemma 2.19. *Assume Hypothesis T. Let C be a smooth complete curve over K, and let Z be a nonempty subset of $C(\overline{K})$ that is Galois stable over K and open in the canonical topology* [Bosch et al. 1984, §7.2.1]. *If Q is a point in $C(K)$, there*

exists a function f on C, defined over K, with a pole only at Q and zeroes only in Z.

Proof. We can assume $g = g(C) > 0$ and $Q \notin Z$. Identify C with its image in its Jacobian J by $x \mapsto (x) - (Q)$. Then $U := [g]_J Z = Z[+]_J \cdots [+]_J Z$ is open in $J(\overline{K})$. Let $P \neq 0 \in U$. We claim that there is a sequence $m_1, m_2, \ldots$ of positive integers such that $[m_n]_J P \to 0$.

By [BL 1984, Theorems 5.1, 6.6, 7.4 and 7.5] (see [Cherry 1994, Theorems 2.1 and 2.2] also), there is a finite extension L of K,[4] a commutative rigid analytic group $\hat{J}$, and formal analytic groups J^{fm} and B over L [Bosch 1977b, Definition 1.4] (see also [Bosch 1976, introduction] and [Cherry 1994, p. 397]), such that B is proper and we have an injective composition $(J^{\mathrm{fm}})^{\mathrm{rg}} \to \hat{J} \to J_L^{\mathrm{rg}}$ such that the image of $(J^{\mathrm{fm}})^{\mathrm{rg}}$ in J_L^{rg} is a maximal connected subgroup with a formal analytic structure.[5] Moreover, there is a diagram with exact rows and columns

$$
\begin{array}{ccccccccc}
& & & & 0 & & & & \\
& & & & \downarrow & & & & \\
& & & & \mathbb{Z}^t & & & & \\
& & & & \downarrow & & & & \\
0 & \rightarrow & (\mathbf{G}_m^{\mathrm{rg}})^t & \longrightarrow & \hat{J} & \longrightarrow & B^{\mathrm{rg}} & \rightarrow & 0 \\
& & & & \downarrow & & & & \\
& & & & J_L^{\mathrm{rg}} & & & & \\
& & & & \downarrow & & & & \\
& & & & 0 & & & &
\end{array}
$$

where $t \in \mathbb{N}$ (the toric rank) and the image of $\mathbb{Z}^t$ is a discrete closed subgroup. This induces an exact sequence $0 \to (\mathbf{G}_m^{\mathrm{fm}})^t \to J^{\mathrm{fm}} \to B \to 0$, of formal analytic groups[6] and implies that $\hat{J}(L)/J^{\mathrm{fm}}(L)$ is isomorphic to $(\mathbf{G}_m^{\mathrm{rg}}(L)/\mathbf{G}_m^{\mathrm{fm}}(L))^t$ and the reduction of J^{fm} over the residue field of L is semiabelian.[7]

So $J(L)/J^{\mathrm{fm}}(L)$ is isomorphic to $(\mathbf{G}_m^{\mathrm{rg}}(L)/\mathbf{G}_m^{\mathrm{fm}}(L))^t / \Gamma$, where Γ is the injective image of $\mathbb{Z}^t \to \hat{J}(L)/J^{\mathrm{fm}}(L) \to (\mathbf{G}_m^{\mathrm{rg}}(L)/\mathbf{G}_m^{\mathrm{fm}}(L))^t$. Assuming Hypothesis T for L, $\mathbf{G}_m^{\mathrm{rg}}(L)/\mathbf{G}_m^{\mathrm{fm}}(L) = L^*/R_L^*$ is isomorphic to a subgroup of $\mathbb{Q}$, and hence it follows that $J(L)/J^{\mathrm{fm}}(L)$ is torsion. As all elements on a semiabelian variety over a finite field are torsion, some multiple $[k]_J P$ of P lies in the image of the kernel of reduction of J^{fm}, and then $[p^n k]_J P \to 0$.

Now, since U is open and $[m_n]_J P \to 0$, there is a positive integer m such that $-[m-1]_J P = P[-]_J[m]_J P \in U$. Thus $0 \in [mg]_J Z$. More specifically, there is

---

[4]While the field is assumed to be algebraically closed in [BL 1984], it is explained on [BL 1984, p. 257] how to show that $\hat{J}$ may be defined over a finite extension.

[5]If Y is a scheme or formal analytic space, Y^{rg} will denote the associated rigid space.

[6]$\mathbf{G}_m^{\mathrm{fm}}$ denotes the formal completion of $\mathbf{G}_m$ along its reduction.

[7]Formal analytic spaces have canonical reductions.

a principal divisor D of the form

$$(m-1)\sum_{i=1}^{g}(P_i) + \sum_{i=1}^{g}(R_i) - mg(Q),$$

where P_i and $R_i \in Z$. If g is a function over L with this divisor, we can take $f = \prod_\sigma g^\sigma$, where σ ranges over embeddings of L/K into $\mathbf{C}/K$. □

Lemma 2.20. *Suppose C is a complete curve over K and U is an open disk in C. Then $Y := C \setminus U$ is a nonempty open in the canonical topology.*

Proof. Let P be any point in Y, which is nonempty since U is not proper and so cannot equal C. By Riemann–Roch, we can choose a meromorphic function g on C with a pole only at P. Then because $g|_U$ is holomorphic and finite to one, $g(U)$ is an open disk in $\mathbf{A}^1$. Let E be an open disk around infinity in $\mathbf{P}^1 \setminus g(U)$. Then $g^{-1}(E)$ is an open neighborhood of P in Y, in the canonical topology. □

Proposition 2.21. *Assume Hypothesis T. Suppose C is a smooth complete curve over K. Let L be a finite Galois extension of K, and let T be a finite, nonempty, Galois stable subset of $C(L)$. Suppose $\mathfrak{D} = \{D_t : t \in T\}$ is a Galois stable collection of disjoint open disks over L in C, such that $D_t \cap T = \{t\}$ for all $t \in T$. Then if $U = \bigcup \mathfrak{D}$, then $X := C \setminus U$ is a one-dimensional affinoid over K, and the image of the ring of algebraic functions, $\mathbb{O}_C(C \setminus T)$, is dense in $A(X)$.*

Proof. X is nonempty, since U is not proper, and X is open in the canonical topology by Lemma 2.20. Therefore, Lemma 2.19 implies that for each Galois orbit $S \subseteq T$ there exists a function $f_S \in \mathbb{O}_C(C \setminus S)$, defined over K, that has a pole at each $s \in S$ and zeroes only on X. Set $\mathfrak{D}_0 = \mathfrak{D}$, $X_0 = X$, and $U_0 = U$. Then for each $n \geq 1$, choose a Galois stable collection $\mathfrak{D}_n$ of $|T|$ open disks over L in C, such that $\mathfrak{D}_{n+1} \subseteq \mathfrak{D}_n$ for all $n \geq 0$ and $\bigcap_n \mathfrak{D}_n = T$. Set $U_n = \bigcup \mathfrak{D}_n$ and $X_n = C \setminus U_n$. Let D_{tn} be the disk in $\mathfrak{D}_n$ that contains any particular $t \in T$, and for any Galois orbit $S \subseteq T$, set $M_{Sn} = \inf\{|f_S(x)| : x \in \bigcup_{s \in S} D_{sn}\}$. (Note that this infimum is positive and does not belong to the set, since $|g|_{\sup}$ exists and is not equal to $|g(x)|$ for any $x \in D$ when g is a rigid function on an open disk D that vanishes at only finitely many points.) We claim

$$X_n = Z_n := \{x \in C : |f_S(x)| \leq M_{Sn} \text{ for all Galois orbits } S \subseteq T\}.$$

It is clear that $Z_n \subseteq X_n$ since Z_n cannot intersect D_{tn} for any $t \in T$. For the other direction, note that f_S is defined over K and has poles only on S, and so $f_S : C \to \mathbf{P}^1$ has degree $|S| d_S$ where $d_S := -\text{ord}_s f_S$ for any $s \in S$. Moreover, since f_S has no zeroes on $\mathfrak{D}$, $f_S|_{D_{sn}}$ is a d_S to 1 map onto the disk $\mathbf{P}^1 \setminus B_K[M_{Sn}]$. It follows that $M_{Sn} \in \mathfrak{R}$ and $\|f_S\|_{X_n} = M_{Sn}$. Thus, $X_n \subseteq Z_n$. So $X_n = Z_n$, and in particular X_n is an affinoid.

For each n and S, we may choose $e_{Sn} \in \mathbb{N}$ and $a_{Sn} \in K^*$ such that $|a_{Sn}| = M_{Sn}^{e_{Sn}}$. Then, using the notation of [Bosch et al. 1984, §7.2.3], we have

$$Z_n = Z_{n+1}(f_S^{e_{Sn}}/a_{Sn} : S \text{ is a Galois orbit in } T).$$

It follows from [Bosch et al. 1984, Proposition 7.2.3/1] that the image of $A(X_{n+1})$ is dense in $A(X_n)$. Suppose $g \in A(X)$ and $\epsilon > 0$. Then there exist functions $h_n \in A(X_n)$ such that $\|h_1 - g\|_X < \epsilon$ and $\|h_{n+1} - h_n\|_{X_n} < \epsilon/n$ for $n \geq 1$. It follows that the sequence h_n converges to an element $h_\epsilon \in A(C \setminus T)$ such that $\|h_\epsilon - g\|_X < \epsilon$. The proposition follows from the fact that $\mathbb{O}_C(C \setminus T)$ is dense in $A(C \setminus T)$. □

Corollary 2.22. *Assume Hypothesis T. Let W be a wide open over K. Then the image of $A(W)$ is dense in $A(X)$ for each underlying affinoid X.*

Proof. Glue open disks B_V to W to make a complete curve C as in the proof of Theorem 2.18. For each $V \in \mathscr{S}$, choose a point $t_V \in B_V \setminus W$ defined over K. Then let $T = \{t_V : V \in \mathscr{S}\}$ and follow the procedure above, noting that the map from $\mathbb{O}_C(C \setminus T)$ to $A(X)$ factors through $A(W)$. □

Corollary 2.23. *With the same hypotheses and notation as Proposition 2.21, set $A_0 = \{f \in \mathbb{O}_C(C \setminus T) : \|f\|_X \leq 1\}$. If $A_0 \otimes \mathbb{F}_K$ is reduced, then $\mathrm{Spec}\, A_0 \otimes \mathbb{F}_K \cong \overline{X}$.*

A *basic wide open pair* over K is a pair (W, X) where W is a connected wide open over K and X is an underlying affinoid. In addition, we require that $\|A(X)\|_X = |K|$, that X have irreducible reduction with at worst ordinary double points as singularities, and that the components of $W \setminus X$ be isomorphic to annuli of the form $A_K(1, s)$. If (W, X) is a basic wide open pair for some X, we say that W is a *basic wide open*. By Proposition 2.21 and Corollary 2.23, basic wide open pairs can be constructed by taking $(W, X) = (C \setminus \bigcup_{i=1}^n D_i, C \setminus \bigcup_{i=1}^n U_i)$. Here C is a connected smooth complete curve over K that has a model $\mathscr{C}$ over R_K whose reduction is irreducible and has at worst ordinary double points as singularities, $\{U_1, \ldots, U_n\}$ is a finite collection of distinct residue classes of smooth points, and each D_i is an affinoid disk in U_i. The converse, that all basic wide open pairs can be constructed in this manner, follows, when K is stable and assuming Hypothesis T, from Theorem 2.27 (and thus the two notions are equivalent in this case).

Lemma 2.24. *Assume Hypothesis T. Suppose $f : X \to Y$ is a map between smooth one-dimensional affinoids over K, and $\overline{X}$ is irreducible.*

  (i) *If $\bar{f} : \overline{X} \to \overline{Y}$ is a surjection, then f is a surjection.*

  (ii) *If $\bar{f}(\overline{X}) \subseteq \overline{Y}$ is an open affine and $X(\mathbf{C}) \to Y(\mathbf{C})$ is an injection, then $\bar{f}$ is an injection.*

Proof. For both parts, we may extend scalars to **C**. To prove (i), suppose $\bar{f}$ is a surjection but that there exists a $y \in Y$ that is not in the image of f. Let $\lambda \in A^o(Y)$ be a function that vanishes only at y and such that $\|\lambda\|_Y = 1$.[8] On the one hand, if we let $L = f^*(\lambda)$, the fact that $\bar{f}$ is a surjection implies that $\|L\|_X = 1$. On the other hand, since L does not vanish on X, L^{-1} exists and we may choose $c \in \mathbf{C}$ such that $|c| = \|L^{-1}\|_X$. Now the fact that $\bar{f}$ is a surjection implies that $|c| > 1$. Thus, if we let $M = c^{-1}L^{-1}$, we have $\bar{L}, \bar{M} \neq 0$ but $\bar{L}\bar{M} = 0$. So $\bar{X}$ must be reducible.

For (ii), let $Y' \subseteq Y$ be the Zariski subaffinoid for which $\bar{f}(\bar{X}) = \bar{Y'}$. Suppose there are distinct points $x_1, x_2 \in \bar{X}$ for which $\bar{f}(x_1) = \bar{f}(x_2)$. Let $X' = X \setminus R(x_2)$. Then f restricts to a map $f' : X' \to Y'$ that reduces to a surjection. Thus, by (i), f' is a surjection. But this is a contradiction since $f(R(x_2)) \subseteq Y'$ and f is an injection. Therefore, $\bar{f}$ must also be an injection. $\square$

Lemma 2.25. *Suppose $h : B \to Y$ is an analytic map from an open annulus or open disk into a reduced affinoid. Then the image of B is contained in a residue class of Y.*

Proof. This is clear when Y is an affinoid disk. The general case follows. $\square$

Remark 2.26. The same statement is true with B a connected wide open in place of an open annulus.

Theorem 2.27. *Suppose K is stable and satisfies Hypothesis T. Let (W, X) be a basic wide open pair over K. Attach disks B_V to W to obtain a complete curve C, as in the proof of Theorem 2.18. Then C has a model over R_K whose reduction is $\bar{X}^c$. Also, if x is a point at ∞ in $\bar{X}^c(\bar{\mathbb{F}})$, then $x \in \bar{X}^c(\mathbb{F}_K)$ and $\{P \in C(\mathbf{C}) : \bar{P} = x\}$ is equal to $B_V(\mathbf{C})$ for some $V \in \mathcal{G} = \mathrm{CC}(W \setminus X)$.*

Proof. Choose a finite, Galois stable set of points $Y \subset X(L)$, for some finite extension L of K, that reduce to distinct smooth points in $\bar{X}_L(\mathbb{F}_L)$. The set $\{R(\bar{y}) : y \in Y\}$ of residue classes of X_L is a Galois stable set of open disks in C over L. Therefore, by Proposition 2.21, $Z := C \setminus \bigcup_{y \in Y} R(\bar{y})$ is an affinoid over K. Moreover, $X_1 := X \cap Z$ is a formal subdomain of X [BL 1985, p. 351], whose reduction is $\bar{X} \setminus \{\bar{y} : y \in Y\}$. We will show that X_1 is also a formal subdomain of Z, and hence $\mathscr{C} := \{X, Z\}$ is a formal covering of C.

To do this, let $Z_{\mathcal{T}} := Z \setminus \bigcup_{V \in \mathcal{T}} B_V$ for any $\mathcal{T} \subseteq \mathcal{G}$. This is an affinoid over K by Proposition 2.21. We claim that $Z_{\mathcal{T}}$ has irreducible reduction, and that B_V is a residue class of $Z_{\mathcal{T}}$ for each $V \in \mathcal{G} \setminus \mathcal{T}$. This is clearly true for $\mathcal{T} = \mathcal{G}$, because $Z_{\mathcal{G}} = X \setminus \bigcup_{y \in Y} R(\bar{y})$ is a Zariski subaffinoid of X and $\mathcal{G} \setminus \mathcal{T}$ is empty. Suppose it holds for some $\mathcal{T}$, and suppose also that $V \in \mathcal{T}$. Let $\mathcal{T'} = \mathcal{T} \setminus \{V\}$, so

---

[8]This can be done by embedding Y in a smooth, complete curve [Van der Put 1980, Theorem 1.1] and applying Lemma 2.19.

that $Z_{\mathcal{T}'} = Z_{\mathcal{T}} \coprod B_V$. By Lemma 2.25, applied to the inclusion of B_V into $Z_{\mathcal{T}'}$, B_V is contained in the residue class $R(\bar{t}_V)$. If $B_V \neq R(\bar{t}_V)$, the map $\overline{Z}_{\mathcal{T}} \to \overline{Z}_{\mathcal{T}'}$ is a surjection. But this is impossible by Lemma 2.24 since $Z_{\mathcal{T}}$ has irreducible reduction and $Z_{\mathcal{T}} \neq Z_{\mathcal{T}'}$. Therefore, B_V is a residue class of $Z_{\mathcal{T}'}$, $Z_{\mathcal{T}}$ is a Zariski subaffinoid of $Z_{\mathcal{T}'}$, and in particular $Z_{\mathcal{T}'}$ has irreducible reduction. The claim now follows for all $\mathcal{T}$ by induction. Taking $\mathcal{T} = \varnothing$, we see that Z has irreducible reduction, and that each disk B_V is a residue class of Z. Thus, X_1 is a formal subdomain of Z, and $\mathcal{C}$ is a formal covering of C. Moreover, by Proposition 2.8, the reduction of Z is the disjoint union of $\overline{X}_1$ and finitely many smooth points. Thus C has semistable reduction with respect to $\mathcal{C}$ [BL 1985, Definition 1.5]. So using the argument of [BL 1985, p. 377], it follows that C has a model with reduction $\overline{X}^c$. Moreover, the residue classes of the points at infinity on $\overline{X}^c$ are precisely the disks B_V over K. $\qquad\square$

It may be proven that over $\mathbf{C}$, all wide opens that are not disks or annuli have minimal underlying affinoids. In fact, one can show that if W is a wide open over K that is not a disk or an annulus, W has an affinoid subdomain that becomes the minimal underlying affinoid of W_L, where L is a finite extension of K; see Remark 2.41. However, this fact is not used in this paper.

Lemma 2.28. *Suppose K is stable and assume Hypothesis T. If (W, Z) is a basic wide open pair over K, and W is not a disk or annulus, then Z is a minimal underlying affinoid of W.*

Proof. Suppose there are e ends. Glue in disks, as above, to get a smooth complete curve C, so that $C \setminus Z$ is the union of e open disks $U_1 \ldots U_e$. Then by Theorem 2.27, C will have a model $\mathcal{C}$ with reduction isomorphic to the completion of $\overline{Z}$.

We can and will extend scalars to $\mathbf{C}$. Suppose V is any underlying affinoid of W and A is a component of $W \setminus V$. Then $A \cap U_i \neq \varnothing$ for some i. Set $U = U_i$. We claim that A is contained in U.

Identify A with $A_{\mathbf{C}}(r, s)$ so that $A_{\mathbf{C}}(t, s)$ is connected to V for any $t \in \mathcal{R}(r, s)$. It follows from [BL 1985, Proposition 5.4(c)] that every circle in A that intersects a residue class of C is contained in that class. Hence $A \cap U$ contains $C_{\mathbf{C}}[t]$ for any $t \in \mathcal{R}(r, s)$ with $C_{\mathbf{C}}[t] \cap U \neq \varnothing$. In fact, $A \cap U \supset A_{\mathbf{C}}(r, t]$ for any such t. Let $q = LUB\{t \in \mathcal{R}(r, s) : A_{\mathbf{C}}(r, t) \subseteq U\}$. Suppose that $q < s$, and let

$$v = LUB\{t \in \mathcal{R}(q, s) : C_{\mathbf{C}}[t] \cap Z \neq \varnothing\} = GLB\{t \in \mathcal{R}(q, s) : C_{\mathbf{C}}[t] \cap Z = \varnothing\}.$$

The number v exists since U is disconnected from U_j for $j \neq i$. For $u \in \mathcal{R}[q, v)$, $C_{\mathbf{C}}[u] \subseteq Z$ (again by [BL 1985, Proposition 5.4(c)]). Let $w = q$ if $q \in \mathcal{R}$, and $w \in \mathcal{R}(q, u)$ otherwise, and set $Y = A_{\mathbf{C}}[w, u]$. We have a rigid morphism $Y \to Z$. Since $\overline{Y}$ is either a line or two lines crossing at a point, and $\overline{Z}$ is irreducible, not isomorphic to $\mathbf{A}^1$ or $\mathbf{G}_m$, with only ordinary double points as singularities, it follows that the

map $\overline{Y} \to \overline{Z}$ is constant. This means $A \setminus U$ is contained in a residue class R of Z. Thus $\{U, R\}$ is a disjoint admissible cover of A. This is impossible as A is connected.

From the contradiction, we know that $q = s$, and thus $A \subseteq U$. Now, since each component of $W \setminus V$ is contained in $W \setminus Z$, we have shown that $Z \subseteq V$. $\qquad\square$

The final two results of this section provide useful criteria for determining when a rigid space is a wide open.

Theorem 2.29. *Suppose X is a smooth, one-dimensional affinoid over a stable field K satisfying Hypothesis B, and x is a point of degree one on $\overline{X}$. Then, if $U = R_X(x)$, there is a finite extension L of K such that U_L is a connected wide open over L. Moreover, the number of ends of U_L equals the number of branches of $\overline{X}_L$ through x.*

This is a consequence of the following lemma. Recall that $\mathbb{F}_K \subseteq \overline{\mathbb{F}}$.

Lemma 2.30. *Suppose X is a pure, one-dimensional reduced affinoid over a stable field K satisfying Hypothesis B, with reduced reduction, and $x \in \overline{X}(\mathbb{F}_K)$ is a degree one point. Choose any $f \in A^o(X)$ such that $\bar{f}$ has an isolated zero at x, that is, such that x is the only zero in a Zariski neighborhood. For $r \in \mathcal{R}(0, 1)$, let $V(r)$ be the subspace of X such that*

$$V(r)(\mathbf{C}) = \{y \in R(x)(\mathbf{C}) : r < |f(y)| < 1\}.$$

Then for r sufficiently close to 1, there is a finite extension L of K such that $V_L(r)$ is a disjoint union of $m := |(n^{-1}x)(\overline{\mathbb{F}})|$ open annuli, where $n : Y \to X_{\overline{\mathbb{F}}}$ is the normalization of $X_{\overline{\mathbb{F}}} := \overline{X} \otimes_{\mathbb{F}_K} \overline{\mathbb{F}}$.

Proof. Without loss of generality, we may assume that x is the only zero of $\bar{f}$ (otherwise replace X with a suitable Zariski subaffinoid). Let $Z := Z_r$ be the subaffinoid of X whose $\mathbf{C}$-valued points are $\{y \in X(\mathbf{C}) : |f(y)| \geq r\}$. Let X_x be the curve obtained from Y by identifying $n^{-1}(x')$ to a point for each $x' \in X_{\overline{\mathbb{F}}}(\overline{\mathbb{F}}) \setminus \{x\}$ (thus, X_x is the minimal finite surjective cover of $X_{\overline{\mathbb{F}}}$ that is smooth at all points above x). It is proven in the remark after [BL 1985, Lemma 2.4] that for $r \in \mathcal{R}(0, 1)$ sufficiently close to 1, the reduction of $Z_{\mathbf{C}}$ is isomorphic to the union of X_x and m lines, each crossing a single point above x normally.[9]

There is a finite extension M of K such that $\overline{Z}_M$ is reduced, so $\overline{Z}_{\mathbf{C}} \cong (\overline{Z}_M)_{\overline{\mathbb{F}}}$. Thus, there is a finite extension L of K such that $\overline{Z}_L$ is isomorphic to the union of a finite surjective cover of $\overline{X}_{\mathbb{F}_L}$ that is smooth at all points above x, and m lines each crossing a single point above x normally. Now apply Proposition 2.10. $\qquad\square$

---

[9]This is a minor correction of the statement in [BL 1985].

Proposition 2.31. *Assume* Hypothesis T. *Suppose* $f : U \to W$ *is a finite surjective morphism over* K *of a smooth, one-dimensional rigid space onto a wide open, with finitely many branch points all defined over* K. *If* f *has degree strictly less than* p, *then* U *is a wide open over* K.

Proof. First we claim that an underlying affinoid $X \subseteq W$ can be chosen so that it contains the set $\mathcal{B}$ of branch points of f. Indeed, let X_1 be *any* underlying affinoid of W. Glue disks B_V onto W for each annulus $V \in W \setminus X_1$, as in the proof of Theorem 2.18, to obtain a complete curve C. Then for each V, choose an open disk D_V over K such that $B_V \setminus \alpha_V(V) \subset D_V \subseteq B_V$ and $D_V \cap \mathcal{B} = \varnothing$. The rigid subspace $X := C \setminus \bigcup D_V$ is, by Proposition 2.21, an affinoid that is disjoint from $\mathcal{B}$ and easily shown to be underlying in W.

Now suppose X is relatively compact in some affinoid $Y \subseteq W$. Then $f^{-1}(X)$ and $f^{-1}(Y)$ are affinoids in U. Moreover, as f is finite, and the image of $\overline{X}$ in $\overline{Y}$ is finite, it follows that the image of $\overline{f^{-1}(X)}$ in $\overline{f^{-1}(Y)}$ is finite. So $f^{-1}(X)$ is relatively compact in $f^{-1}(Y)$. All that remains is to check that $U \setminus f^{-1}(X)$ is the disjoint union of open annuli, and for this Theorem 2.6 suffices. $\square$

2C. *Semistable coverings.* For a wide open W over K, let

$$H^i_{DR}(W/K) = D^i(W/K).$$

Using Lemma 2.13, the arguments in the proof of [Coleman 1989, Theorem 4.2] generalize and allow us to conclude that $H^i_{DR}(W/K)$ is finite-dimensional over K. We define the genus of W, which we denote by $g(W)$, to be

$$\tfrac{1}{2}(\dim_K H^1_{DR}(W/K) - e(W) + 1).$$

Then $2g(W)$ can be interpreted as the dimension of the first compactly supported de Rham cohomology group of W. For example, in Corollary 2.33, we show that

$$2g(W) = \dim_K(\ker(H^1_{DR}(W/K) \to H^1_{DR}((W \setminus X)/K))),$$

where X is any underlying affinoid of W. We also show in Proposition 2.32 that if a wide open W is completed to a projective curve C by attaching disks at the ends, as in Theorem 2.18, then $g(W) = g(C)$. As an immediate corollary of this and of Theorem 2.27, if (W, X) is a basic wide open pair over a complete, stable field K satisfying Hypothesis T, and X has good reduction $\overline{X}$, then $(\overline{X})^c$ will also have genus $g(W)$.

Proposition 2.32. *Let* W *be a connected wide open over* K. *Suppose* C *is a smooth, complete curve (over* K) *obtained by attaching disks at the ends of* W, *as in* Theorem 2.18. *Then* $g(W) = g(C)$.

Proof. The main idea is to view W and the attached disks as an admissible covering of C, and then to apply the (generalized) Mayer–Vietoris sequence of de Rham cohomology (over K). So first suppose $D_1, \ldots, D_n$ are the disks, and set $D_0 = W$. Then Mayer–Vietoris gives us the exact sequence

$$0 \to H^0_{DR}(C) \to \bigoplus_i H^0_{DR}(D_i) \to \bigoplus_{i \neq j} H^0_{DR}(D_i \cap D_j) \to$$
$$H^1_{DR}(C) \to \bigoplus_i H^1_{DR}(D_i) \to \bigoplus_{i \neq j} H^1_{DR}(D_i \cap D_j) \to H^2_{DR}(C) \to 0.$$

Using Lemma 2.13, the definition above of $g(W)$, and the fact that $H^1_{DR}(D_i) = 0$ for $i > 0$, we count dimensions to obtain

$$1 - (e(W) + 1) + e(W) - 2g(C) + (2g(W) + e(W) - 1) - e(W) + 1 = 0.$$

From this we conclude that $g(C) = g(W)$. $\qquad\square$

Corollary 2.33. *Suppose W is a wide open over K, and X is an underlying affinoid of W. Then*

$$2g(W) = \dim_K (\ker(H^1_{DR}(W/K) \to H^1_{DR}((W \setminus X)/K))).$$

Proof. Suppose C is a smooth complete curve obtained by gluing disks to the ends of W. Then arguing from Mayer–Vietoris exactly as in the above proof, we have the exact sequence

$$0 \to H^1_{DR}(C) \to H^1_{DR}(W) \to H^1_{DR}(W \setminus X) \to K \to 0.$$

Now apply Proposition 2.32. $\qquad\square$

Let C be a wide open or a smooth proper curve over K. Let $\mathcal{C}$ be a finite set of basic wide open pairs (U, U^u) over K such that $\mathcal{C}^w := \{U, (U, U^u) \in \mathcal{C}\}$ is an admissible covering of C. Then we call $\mathcal{C}$ a *semistable covering* over K if the following conditions hold:

(i) If $U, V \in \mathcal{C}^w$ and $U \neq V$, the intersection of U and V is a disjoint union of connected components of $U \setminus U^u$ (by definition, annuli of the form $A_K(1, s)$).

(ii) If U, V and W are three distinct elements of $\mathcal{C}^w$, their intersection is empty.

We say that a semistable covering $\mathcal{C}$ is *stable* if none of the elements of $\mathcal{C}^w$ are disks or annuli. Having a semistable covering is not immediately equivalent to having a *semistable reduction* in the sense of [BL 1985, Definition 1.5]. When the context is clear, we will abuse notation by dropping the superscript w and writing $U \in \mathcal{C}$ to mean $U \in \mathcal{C}^w$.

Proposition 2.34. *Suppose $\mathcal{C}$ is a semistable covering of a smooth proper curve C over K. Let $\Gamma_{\mathcal{C}}$ be the unoriented graph without loops whose vertices correspond to the elements of $\mathcal{C}$ and whose edges with endpoints corresponding to distinct $U, V \in \mathcal{C}$ correspond to the connected components of $U \cap V$. Then*

$$g(C) = \sum_{U \in \mathcal{C}} g(U) + \mathrm{Betti}(\Gamma_{\mathcal{C}}).$$

Proof. Again, we begin with the Mayer–Vietoris sequence (of de Rham cohomology over K) associated to this covering.

$$\cdots \to \bigoplus_{U,V \in \mathcal{C}} H^{i-1}_{DR}(U \cap V) \to H^i_{DR}(C) \to \bigoplus_{U \in \mathcal{C}} H^i_{DR}(U) \to \cdots$$

It is immediate that

$$H^0_{DR}(C) \cong H^2_{DR}(C) \cong K, \quad \bigoplus_{U \in \mathcal{C}} H^2_{DR}(U) = 0, \quad \text{and} \quad \bigoplus_{U \in \mathcal{C}} H^0_{DR}(U) \cong K^{\#\mathcal{C}}.$$

Also, by applying Lemma 2.13 and condition (i) from above, we see that

$$\bigoplus_{U,V \in \mathcal{C}} H^0_{DR}(U \cap V) \cong \bigoplus_{U,V \in \mathcal{C}} H^1_{DR}(U \cap V) \cong K^{\#\mathcal{E}},$$

where $\mathcal{E}$ is the edge set of $\Gamma_{\mathcal{C}}$. Now to prove the proposition, we simply count dimensions over K and compute the dimension of $H^1_{DR}(C)$ using the exact sequence. We have

$$2g(C) = \sum_{U \in \mathcal{C}} (2g(U) + e(U) - 1) - \#\mathcal{C} + 2 = 2\Big(\sum_{U \in \mathcal{C}} g(U) + \#\mathcal{E} - \#\mathcal{C} + 1 \Big)$$

$$= 2\Big(\sum_{U \in \mathcal{C}} g(U) + \mathrm{Betti}(\Gamma_{\mathcal{C}}) \Big). \qquad \square$$

Definition 2.35. A *semistable model* $\mathcal{B}$ of a curve C over K is a flat, proper scheme over R_K whose generic fiber is C, such that all of the singular points of the special fiber of $\mathcal{B}$ have degree 1 and are ordinary double points. We say that $\mathcal{B}$ is *stable* if it is the final object in the category of semistable models over K.[10]

See [BL 1985] and [Van der Put 1984] for a rigid analytic treatment of the theory of stable models of curves over complete nonarchimedean fields, and in particular, for a rigid analytic proof of the generalization to arbitrary complete

---

[10]This weakens the definition of the semistable model in [Mumford 1977] since it allows smooth rational components that meet the other components in only one point. Requiring the singular points to have degree 1 means that $X_0(p)$ usually does not have a stable model over $\mathbb{Q}_p$, but does over $W(\mathbb{F}_{p^2}) \otimes \mathbb{Q}_p$.

nonarchimedean fields [Van der Put 1984, Corollary 3.3] (see also [BL 1985][11]) of the existence theorem of Deligne and Mumford. Moreover, we will use the results of [BL 1984; 1985] to prove the following theorem, which relates stable coverings to stable models. This result generalizes [Coleman 2003, Proposition 2.1], and the proof we give is more complete than the one given there.

Theorem 2.36. *Let C be a smooth complete curve over a stable field K satisfying Hypothesis B.*

(i) *If C has a semistable model over R_K whose reduction has at least two components, then C has an associated semistable covering over K.*

(ii) *If K satisfies Hypothesis T, and C has a semistable covering over K, then C has an associated semistable model over R_K whose reduction has at least two components.[12]*

Stable coverings are precisely those that correspond to stable models whose reductions have at least two components.

Proof. Suppose $\mathscr{C}$ is a semistable model for C over R_K, and let $\mathscr{I}_{\mathscr{C}}$ be the set of irreducible components in the reduction of $\mathscr{C}$. For each $\Gamma \in \mathscr{I}_{\mathscr{C}}$, let

$$\Gamma^o = \Gamma \setminus \bigcup_{\Gamma' \in \mathscr{I}_{\mathscr{C}}, \Gamma' \neq \Gamma} \Gamma'.$$

Assume, without loss of generality, that $\overline{\mathscr{C}}$ is connected.

For each affine open $U \subseteq \overline{\mathscr{C}}$, there is a natural affinoid subdomain of C^{rg}, which we denote by $\mathrm{Red}^{-1} U$, whose points are all the points of C^{rg} that reduce to points of U. To see this, let $\mathrm{Spec}\, S$ be any affine open subscheme of $\mathscr{C}$ that reduces to U and $\hat{S} = \varprojlim_n S/\pi^n S$ for some $\pi \in R_K$ with $0 < |\pi| < 1$. Then $\hat{S}$ is an admissible R_K-algebra in the sense of [BL 1993, p. 293], as can be seen from [BL 1993, Lemma 1.2]. Then $\hat{S} \otimes_{R_K} K$ is an affinoid algebra over K [BL 1993, §4] that up to canonical isomorphism does not depend on the choices. We refer to the affinoid $\mathrm{Sp}(\hat{S} \otimes_{R_K} K)$ as $\mathrm{Red}^{-1} U$. Because U is reduced, $\overline{\mathrm{Red}^{-1} U} \cong U$. More generally, suppose V is the union of finitely many subschemes W of $\overline{\mathscr{C}}$, with each contained in some affine open U_W. Then we let $\mathrm{Red}^{-1} V$ be the open rigid subspace that is the union in C^{rg} of the subspaces $\mathrm{Red}^{-1} W \subseteq \mathrm{Red}^{-1} U_W$, as was defined in the beginning of Section 2. This subspace is independent of the choices of U_W.

If $\Gamma \in \mathscr{I}_{\mathscr{C}}$, let $W_{\Gamma} = \mathrm{Red}^{-1} \Gamma$ and $X_{\Gamma} = \mathrm{Red}^{-1} \Gamma^o$. We claim that $\{(W_{\Gamma}, X_{\Gamma}) : \Gamma \in \mathscr{I}_{\mathscr{C}}\}$ is a semistable covering. First, W_{Γ} is a smooth, one-dimensional rigid

---

[11]Bosch and Lütkebohmert [BL 1985, p. 377], while proving the theorem of Deligne and Mumford, remark that their argument does not require the field to be discretely valued.

[12]In fact, when K satisfies Hypothesis T we have a natural one-to-one correspondence between semistable coverings and semistable models whose reductions have at least two components. It would be interesting to know if this is true more generally.

space over K, and X_Γ is an affinoid subdomain, such that $W_\Gamma \setminus X_\Gamma$ is a disjoint union of a finite number of annuli of the form $A_K(1, s)$ by Proposition 2.10. Also, X_Γ has absolutely irreducible reduction with at worst ordinary double points as singularities. Moreover, if $\Gamma, \Gamma', \Gamma'' \in \mathcal{I}_\mathscr{C}$, then $W_\Gamma \cap W_{\Gamma'}$ is a union of connected components of $W_\Gamma \setminus X_\Gamma$ if $\Gamma \neq \Gamma'$, and $W_\Gamma \cap W_{\Gamma'} \cap W_{\Gamma''} = \varnothing$ if Γ, Γ' and Γ'' are all distinct.

What remains to be shown for (i) is that (W_Γ, X_Γ) is a basic wide open pair for each $\Gamma \in \mathcal{I}_\mathscr{C}$, and for that, all we have to show is that there exists an affinoid subdomain Y of W_Γ such that X_Γ is relatively compact in Y and $Y \cap V$ is a semiopen annulus for each connected component of $W_\Gamma \setminus X_\Gamma$. (That W_Γ is connected will follow from the absolute irreducibility of $\overline{X}_\Gamma$.) For this, let $\mathcal{S}_\Gamma$ be the set of singular points in $\overline{\mathscr{C}}$ where Γ intersects some other component. Blow up $\mathscr{C}$ at every point in $\mathcal{S}_\Gamma$ to obtain a new model $\mathscr{C}_\Gamma$ that is defined over K and becomes semistable over an, at worst, quadratic extension L. Let $\hat{\Gamma}$ be the proper transform of Γ in $\mathscr{C}_\Gamma$, and let $\mathcal{I}_{\mathscr{C}_\Gamma}$ be the set of irreducible components in the reduction of $\mathscr{C}_\Gamma$. Set

$$\widetilde{Y}_\Gamma = \overline{\mathscr{C}}_\Gamma \setminus \bigcup_{\substack{\Gamma' \in \mathcal{I}_{\mathscr{C}_\Gamma}, \\ \Gamma' \cap \hat{\Gamma} = \varnothing}} \Gamma',$$

and let $Y_\Gamma = \mathrm{Red}^{-1}(\widetilde{Y}_\Gamma)$. It is clear that $(X_\Gamma)_L \subseteq Y_\Gamma \subseteq (W_\Gamma)_L$ and Y_Γ is naturally defined over K. Although $\widetilde{Y}_\Gamma$ is not an affine open in $\overline{\mathscr{C}}_\Gamma$, Y_Γ is the reduction inverse of an affine open in the model obtained from $\mathscr{C}_\Gamma$ by blowing down $\hat{\Gamma}$. This affine open will consist of $|\mathcal{S}_\Gamma|$ lines intersecting in a single singular point that contains the reduction of X_Γ. Thus, not only is Y_Γ also an affinoid subdomain of W_Γ, but X_Γ is relatively compact in Y_Γ. Finally, by applying Proposition 2.10 again, we see that the intersection of Y_Γ with each component of $W_\Gamma \setminus X_\Gamma$ is a semiopen annulus. Therefore, we are done with (i).

To prove (ii), suppose $\mathscr{C}$ is a semistable covering of C. Then by Theorem 2.27, there is a natural one-to-one correspondence between $(\overline{U^u})^c \setminus \overline{U^u}$, $\mathrm{CC}(U \setminus U^u)$, and $\mathscr{E}(U)$, for each $U \in \mathscr{C}$. If $e \in \mathscr{E}(U)$, let $x(e)$ denote the corresponding point on $(\overline{U^u})^c$ and let $A(e)$ denote the corresponding connected component of $U \setminus U^u$ (an annulus). If $e \in \mathscr{E}(U)$ and $f \in \mathscr{E}(V)$ for $U, V \in \mathscr{C}$, we say that $e \sim f$ whenever $A(e) = A(f)$ (equivalently, $A(e) \cap A(f) \neq \varnothing$). Let $\mathscr{E}$ denote the quotient of $\coprod_{U \in \mathscr{C}} \mathscr{E}(U)$ by this equivalence relation. We define $\overline{\mathscr{C}}$ to be the curve over $\mathbb{F}_K$ obtained from $\coprod_{U \in \mathscr{C}} (\overline{U^u})^c$ by identifying the points $x(e)$ and $x(f)$ whenever $e \sim f$. The reduction maps from $U(\mathbf{C}) \to (\overline{U})^c(\overline{\mathbb{F}})$ for each $U \in \mathscr{C}$, which are guaranteed by Theorem 2.27, patch together to form a natural Galois equivariant reduction map from $C(\mathbf{C}) \to \overline{\mathscr{C}}(\overline{\mathbb{F}})$. We will show that in fact there is a model over R_K whose reduction is $\overline{\mathscr{C}}$.

Let T be a finite Galois stable set of points of $C(\overline{K})$ that, by the above reduction map, injects into the smooth locus of $\overline{\mathscr{C}}$, and such that $T \cap U \neq \varnothing$ for each $U \in \mathscr{C}$. Since each $t \in T$ lies on a unique U^u, the residue class, $R(\overline{t}) := R_{U^u}(\overline{t})$, is well defined and can be viewed as an open disk in C over $\overline{K}$. Moreover, as $\mathscr{C}$ is defined over K, $R(T) := \bigcup_{t \in T} R(\overline{t})$ is Galois stable over K. So by Proposition 2.21, $Z_T := C \setminus R(T)$ is an affinoid over K. We want to show that $\overline{Z}_T = \overline{\mathscr{C}} \setminus \overline{T}$, where $\overline{T} = \{\overline{t} : t \in T\}$.

For each $U \in \mathscr{C}$, we let $U_T = U^u \setminus R(T)$, a Zariski subaffinoid of U^u. Then the affinoid Z_T is the disjoint union of $\coprod_{U \in \mathscr{C}} U_T$ and $\coprod_{e \in \mathscr{E}} A(e)$. Now fix U and consider the natural inclusion map $U_T \hookrightarrow Z_T$. Since the reduction of U_T is irreducible, it follows that $\overline{U}_T$ maps to a point or onto an affine open of some irreducible component Γ_U of $\overline{Z}_T$. As U_T and Z_T are both connected to $R(T)$, the first case is not possible. Therefore, by Lemma 2.24(ii), $\overline{U}_T$ must inject into some such Γ_U. Let U_T' be the subaffinoid of Z_T that lies above the image of $\overline{U}_T$. By Lemma 2.24(i), the inclusion of U_T into U_T' is surjective, and therefore an equality. As the U_T don't intersect, and $U_T = U_T'$ for each $U \in \mathscr{C}$, the Γ_U must be distinct components.

Now suppose $e \in \mathscr{E}$. By applying Lemma 2.25 to the inclusion of $A(e)$ into Z_T, we see that $A(e)$ must be contained in a residue class, $R(y_e) := R_{Z_T}(y_e)$, for some point $y_e \in \overline{Z}_T(\mathbb{F}_K)$. Thus there can be no irreducible components of $\overline{Z}_T$ other than $\{\Gamma_U : U \in \mathscr{C}\}$. Moreover, it is clear that $\bigcup_{e \in \mathscr{E}} A(e) = \bigcup_{e \in \mathscr{E}} R(y_e)$. So using the fact that residue classes of an affinoid are connected,[13] it follows that the y_e are distinct and hence $A(e) = R(y_e)$ for each $e \in \mathscr{E}$. From connectivity, we also have that $y_e \in \Gamma_U \cap \Gamma_V$ whenever $U \neq V$ and $A(e) \subseteq U \cap V$, and by Proposition 2.10 this must be a normal crossing. Therefore, as claimed, we have shown that $\overline{Z}_T = \overline{\mathscr{C}} \setminus \overline{T}$, and we use equality here to emphasize that the canonical reduction map on the $Z_T(\mathbb{C})$ is compatible with the previously defined reduction map on $C(\mathbb{C})$.

To finish the proof, let T_1 and T_2 be two finite Galois stable sets of points of $C(\overline{K})$ satisfying the above conditions on T, and such that $\overline{T}_1 \cap \overline{T}_2 = \varnothing$. Then $Z := Z_{T_1} \cap Z_{T_2}$ is equal to $Z_{T_1 \cup T_2}$. Therefore, Z is a formal subdomain of both Z_{T_1} and Z_{T_2} by the compatibility of reduction maps. So C has semistable reduction $\overline{Z}_{T_1} \cup \overline{Z}_{T_2} = \overline{\mathscr{C}}$ with respect to the formal covering $\{Z_{T_1}, Z_{T_2}\}$; see [BL 1985, Definition 1.5]. Then, by the same argument as used in the proof of Theorem 2.27, C has a semistable model over R_K whose reduction is isomorphic to $\overline{\mathscr{C}}$. $\qquad\square$

Remark 2.37. As a consequence of Theorem 2.36 we have the result that whenever K is stable and satisfies Hypothesis B, every semistable curve over R_K can be constructed by gluing together wide opens taken out of curves with *good reduction* over R_K. Crossings of distinct irreducible components are created by gluing two

---

[13]If R is a residue class, $A^o(R)$ is a local ring.

annuli at the ends of two distinct wide opens, while self-intersections within a
component are created by gluing two annuli at distinct ends of a single wide open.

Lemma 2.38. *Suppose D is a closed disk and U is either an open disk or open
annulus in a smooth complete curve C, all defined over K, such that $D \cap U \neq \varnothing$.
Then either $D \subseteq U$, $U \subseteq D$, $D \cup U$ is an open disk, or $D \cup U = C \cong \mathbf{P}^1$ and U is
an open disk.*

Proof. We can assume $K = \mathbf{C}$ and $g(C) > 0$. When U is an open disk the lemma
follows from [BL 1985, Proposition 5.4(a)]. So suppose U is an open annulus, with
$U \nsubseteq D$ and $D \nsubseteq U$. We first show that every concentric circle R of U that intersects
D must be contained in D. Indeed, applying [BL 1985, Proposition 5.4(c)] to the
height 1 annulus R and the disk D, and using $D \nsubseteq R$, we can conclude that R is
contained in some closed disk E. Then by [BL 1985, Proposition 5.4(a)], we have
$D \subseteq E$ or $E \subseteq D$. Either way, it follows that $R \subseteq D$.

Now choose a parametrization $\psi : A_K(r, s) \xrightarrow{\cong} U$. By the preceding argument,
we can then choose $t \in \mathcal{R}(r, s)$ such that $Y_t := \psi(C_K[t]) \subseteq D$. Then $C \setminus Y_t$ and
$U \setminus Y_t$ have two connected components each. Since U is connected, $U \setminus Y_t \nsubseteq C \setminus D$.
Thus there exists $u \in \mathcal{R}(r, s)$ such that $u \neq t$ and $Y_u \subseteq D$. We can assume that
$u < t$ and Y_u is contained in the connected component Z of $C \setminus Y_t$ that lies inside D.
Because $A_K[u, t]$ is connected, it follows that $\psi(A_K[u, t]) \subseteq Z$.

Now choose a $P \in Z \setminus U$, and let $\phi : B_K[1] \xrightarrow{\cong} D$ be any parametrization such
that $\phi(0) = P$. We may assume that $\phi(C_K[v]) = Y_v$ whenever $Y_v \subseteq D$. Thus,
$\phi(A_K(r, 1)) = U \cap D$ and $s > 1$. Finally, we let $V = D \cup U$. Then V is a wide
open with one end and $\phi^{-1} B_K[t]$ is an underlying affinoid for $t \in \mathcal{R}(r, 1]$. Hence
$g(V) = 0$, and so by the Riemann existence theorem, V is isomorphic to $\mathbf{P}^1$ minus
a closed disk (in particular, an open disk). $\qquad\square$

If $\mathscr{C}$ is a semistable covering of C, we define a *residue class* of $\mathscr{C}$ to be either a
residue class of U^u or a component of $U \setminus U^u$ for some $U \in \mathscr{C}$.

Corollary 2.39. *Suppose $\mathscr{C}$ is a stable covering of C. Then every closed disk D in
C is contained in a residue class of $\mathscr{C}$.*

Proof. Extend scalars to $\mathbf{C}$. The curve C is not isomorphic to $\mathbf{P}^1$ as $\mathbf{P}^1$ does not have
a stable covering. Let R be a residue class of $\mathscr{C}$ such that $D \cap R \neq \varnothing$, and suppose
$D \nsubseteq R$. First suppose R is an open disk. If necessary, refine $\mathscr{C}$ to a semistable
covering $\mathscr{C}'$ for which all underlying affinoids have smooth reduction, none are
closed disks, and R is a residue class of U^u for some $U \in \mathscr{C}'$. By Lemma 2.38, we
have $R \subseteq D$.

This latter containment implies that $U^u \cap D$ is a nonempty affinoid with good
reduction. Every such affinoid is a Zariski subaffinoid of a closed disk E_1 in D,
because D is a closed disk. Since U^u has good reduction, E_1 is a disk, and $C \ncong \mathbf{P}^1$,
it follows that U^u is a Zariski subaffinoid of E_1.

Set $U_1 = U$. Then U_1^u is not equal to E_1 since none of the underlying affinoids in $\mathscr{C}'$ are disks. Therefore, there exists a residue disk R_1 of E_1 such that $A_1 := R_1 \cap U_1$ is a component of $U_1 \setminus U_1^u$ (an open annulus). Now $E_2 := R_1 \setminus A_1$ is a closed disk. Let U_2 be the other element of $\mathscr{C}'$ containing A_1. By the same argument as above, and the fact that both U_2^u and E_2 are connected to A_1, it follows that U_2^u must be a Zariski subaffinoid of E_2. Again (for $i = 2$ now), we must have $U_i^u \neq E_i$. Proceeding in this manner, we eventually exhaust the underlying affinoids of $\mathscr{C}'$ or find a $V \in \mathscr{C}'$ such that V^u is a closed disk. Thus, we have a contradiction.

Now suppose R is an annulus. If an annulus at one end of R is contained in D, and U^u is connected to R at that end for some $U \in \mathscr{C}$, then D intersects every residue class of U^u. In particular, it intersects an open disk. Now apply the above argument. □

Theorem 2.40. *Suppose C is a smooth complete curve over a stable field K satisfying* [Hypothesis B](), *and $\mathscr{D}$ is a finite (possibly empty) collection of disjoint closed disks in C all defined over K. Then there exists a semistable covering $\mathscr{C}$ of C over some finite extension of K such that*

(1) *$D = U_D^u$ for some $U_D \in \mathscr{C}$ for each $D \in \mathscr{D}$, and*

(2) *$\mathscr{C} \setminus \{(U_D, D) : D \in \mathscr{D}\}$ is a semistable covering of $W := C \setminus \bigcup_{D \in \mathscr{D}} D$.*

Proof. If $\mathscr{D}$ is empty, or if $|\mathscr{D}| = 1$ and $g(C) = 0$, the theorem follows directly from [Theorem 2.36]() and [[Deligne and Mumford 1969](); [Van der Put 1984]()].

Otherwise, suppose we have a semistable covering $\mathscr{C}$ of C that is *compatible* with $\mathscr{D}$, in the sense that each $D \in \mathscr{D}$ is either contained in a residue class of $\mathscr{C}$ or equal to U^u for some $U \in \mathscr{C}$. Then we can refine $\mathscr{C}$ to obtain a covering that satisfies the conclusions of the theorem. Indeed, suppose $D \in \mathscr{D}$ and $D \neq U^u$ for any $U \in \mathscr{C}$. Then D is contained in a residue class R of $\mathscr{C}$, and there are three possibilities to consider. First, D could be contained in a residue disk R of U^u for some $U \in \mathscr{C}$. In this case we refine our covering to

$$\mathscr{C}_D := \mathscr{C} \setminus \{(U, U^u)\} \cup \{(U \setminus D, U^u \setminus R), (R, D)\}.$$

The second possibility is that D is contained in a residue annulus R of some U^u. Applying [Lemma 2.38]() from above, there must then be a concentric circle T in R, and a residue disk S in T, such that $D \subseteq S$. In this case, we let

$$\mathscr{C}_D := \mathscr{C} \setminus \{(U, U^u)\} \cup \{(U \setminus T, U^u \setminus R), (R \setminus D, T \setminus S), (S, D)\}.$$

Finally, the residue class R that contains D may be a connected component of $U \cap V$ for two distinct $U, V \in \mathscr{C}$. Again there must be a concentric circle T in R and a residue disk S in T such that $D \subseteq S$. Let R_U and R_V be the connected components

of $R \setminus T$ that are connected to U and V, respectively. Let $\hat{U} = (U \setminus R) \cup R_U$ and $\hat{V} = (V \setminus R) \cup R_V$. Then we may take as our refined cover

$$\mathscr{C}_D := \mathscr{C} \setminus \{(U, U^u), (V, V^u)\} \cup \{(\hat{U}, U^u), (\hat{V}, V^u), (R \setminus D, T \setminus S), (S, D)\}.$$

After applying this procedure finitely many times, we are done.

The only issue remaining is that of finding a compatible covering as a starting point. If $\mathscr{C}$ is a stable covering, then it is compatible with $\mathscr{D}$ by Corollary 2.39. So when $g \geq 2$ we are done by Theorem 2.36. If $g = 0$, and $D_1, D_2 \in \mathscr{D}$ with $D_1 \neq D_2$, then $\mathscr{C} := \{(C \setminus D_1, D_2), (C \setminus D_2, D_1)\}$ is compatible with $\mathscr{D}$. If $g(C) = 1$, $D \in \mathscr{D}$, and U is the largest open disk in C containing D, then $\mathscr{C} := \{(C \setminus D, C \setminus D), (U, D)\}$ is compatible with $\mathscr{D}$. So in each case we are able to construct the desired covering of C. □

Remark 2.41. If $g(C) \geq 2$, or $g(C) = 0$ and $|\mathscr{D}| \geq 3$, or $g(C) = 1$ and $|\mathscr{D}| \geq 1$, there exists a final object $\mathscr{C}_{\mathscr{D}}$ in the category of such coverings. In these cases, $C \setminus \bigcup_{D \in \mathscr{D}} U_D$ is the minimal underlying affinoid of $W := C \setminus \bigcup_{D \in \mathscr{D}} D$.

Corollary 2.42. *Let f be a meromorphic function with finitely many zeroes and poles on a wide open W over a stable field K satisfying Hypothesis B. Then there is a semistable covering $\mathscr{C}$ of W over a finite extension of K such that for each $U \in \mathscr{C}$, U^u has good reduction and all the zeroes and poles of f are contained in $\bigcup_{U \in \mathscr{C}} U^u$.*

Proof. Glue in disks to get a complete curve C. Let $\mathscr{D}$ be the union of $CC(C \setminus W)$ with a finite collection of disjoint closed disks in W that contain the support of f. Apply the theorem to get a semistable covering $\mathscr{C}_1$ of C over some finite extension of K, and then throw out those $U \in \mathscr{C}_1$ for which $U^u \in CC(C \setminus W)$. This yields a semistable covering $\mathscr{C}_2$ of W such that all the zeroes and poles of f are contained in $\bigcup_{U \in \mathscr{C}_2} U^u$. Let $\mathscr{S}$ be the collection of singular residue classes in U^u for all $U \in \mathscr{C}_2$. For each $R \in \mathscr{S}$, choose a concentric circle $A_R \subset R$ (such an R is an open annulus). Then

$$\mathscr{C} := \left\{ \left(U \setminus \bigcup_{R \in \mathscr{S}} A_R, U^u \setminus \bigcup_{R \in \mathscr{S}} R\right) : U \in \mathscr{C}_2 \right\} \cup \{(R, A_R) : R \in \mathscr{S}\}$$

satisfies the requirements of the corollary. □

Our final result of this section is a lemma that will play a key role in the proof of our main theorem, Theorem 9.2.

Lemma 2.43. *Suppose W is a connected wide open over a stable field K satisfying Hypothesis B, with minimal underlying affinoid W^u, and let $X \subset W$ be an affinoid subdomain with smooth irreducible (connected) reduction such that $g(W) = g(\overline{X}^c) > 0$. If X is connected to all but at most one component of $W \setminus W^u$, then W is a basic wide open and X is a Zariski subaffinoid of W^u.*

Proof. First glue disks to W to obtain a smooth connected complete curve C over K. Then by [Coleman 2005, Theorem A1][14], there exists a semistable model $\mathcal{T}$ of C over a finite extension E of K, and a subset S of the set of components of $\overline{\mathcal{T}}$ such that $X_E = X(\mathcal{T}, S)$. Moreover, there exists an $s \in S$ such that $X(\mathcal{T}, s)$ is a Zariski subaffinoid of X_E.

Let $\mathcal{C} := \mathcal{C}_{\mathcal{T}}$ be the semistable covering of C_E associated to $\mathcal{T}$ by Theorem 2.36. This implies by Proposition 2.34 that $\text{Betti}(\Gamma_{\mathcal{C}}) = 0$ and $g(z) = 0$ for all $z \in S$ different from s. It follows that C_E has good reduction isomorphic to $\overline{X}_E^c$, that X_E is a Zariski subaffinoid of C_E, and that each affinoid disk in $C \setminus W$ is contained in a residue class of C_E. Furthermore, the statement that X is connected to all but at most one end of W implies that the elements of $C \setminus W$ lie in distinct residue classes of C_E, and that the complement of these residue classes is the minimal underlying affinoid of W_E that equals W_E^u.

We now know that $A^o(W_E^u) \cong A(W_E^u) \cap A^o(X_E)$, under restriction ρ, and that $A(W_E^u) \cong A(W^u) \otimes_K E$. Also, $A^o(W^u) = A^o(W_E^u) \cap A(W^u)$ and $A^o(X_E) = A^o(X) \otimes_{R_K} R_E$ because X has good reduction.

It follows that there is some nonzero element $m \in R_K$ with $|m| < 1$, such that $m A^o(W_E^u) \subset A^o(W^u) \otimes_{R_K} R_E$. So if $c \in A^o(W_E^u)$, then $\rho(c) = \sum_i r_i b_i$, where $r_1, \ldots, r_n$ is a basis for R_E over R_K and $b_i \in A^o(X)$. It follows that $m b_i = \rho(a_i)$ for some $a_i \in \rho(A^o(W^u))$. Thus $a_i/m \in A^o(W_E^u)$, and so $a_i/m \in A^o(W^u)$. Therefore $A^o(W_E^u) = A^o(W^u) \otimes_{R_K} R_E$. This implies that W^u has good reduction, which completes the proof. $\square$

2D. *Riemann–Hurwitz for wide opens.* For this entire section we assume that K is a stable field satisfying Hypothesis B. Let $\mathcal{A}$ be an *oriented* annulus over K. Suppose f is a function on $\mathcal{A}$, and ω a differential, each with no zeroes or poles in $\mathcal{A}(\mathbf{C})$. Then we define $\text{ord}_{\mathcal{A}} f = \text{res}_{\mathcal{A}} (df/f)$, which is an integer (see the proof of [Coleman 1989, Lemma 2.1]), and $\text{ord}_{\mathcal{A}} \omega = \text{ord}_{\mathcal{A}} (\omega/dz)$ for any $z \in A(\mathcal{A})^*$ with $\text{ord}_{\mathcal{A}} z = 1$ (which is independent of the choice of z). Using this definition, we can also define ord_e at any end e of a wide open W/K. Indeed, suppose ν is either a meromorphic function or differential on W, with finitely many zeroes and poles in $W(\mathbf{C})$. Over some finite extension L of K, W_L will have an underlying affinoid X_L containing the support of ν. Let $\mathcal{A}$ be the component of $W_L \setminus X_L$ corresponding to a fixed $e \in \mathcal{E}(W)$, and let $\psi : A_K(r, s) \to \mathcal{A}$ be an isomorphism such that $\psi(A_K(t, s))$ is connected to X whenever $r < t < s$. Then we define the *inherited orientation* on $\mathcal{A}$ by $\text{res}_{\mathcal{A}} = \text{res}_{r,s} \circ \psi^*$, and we set $\text{ord}_e \nu = \text{ord}_{\mathcal{A}} \nu$.

---

[14]The proof of this result was based on [Coleman 2003, Proposition 2.1], which is now a special case of Theorem 2.36.

Let $\mathrm{Div}(W) := \mathbb{Z}^{W(\mathbf{C}) \cup \mathscr{E}(W)}$, and for any $D \in \mathrm{Div}(W)$ let

$$\deg D = \sum_{P \in W(\mathbf{C})} D(P) + \sum_{e \in \mathscr{E}(W)} D(e).$$

Then for ν as above, set $(\nu) = (\nu)_{\mathrm{fin}} + (\nu)_{\mathrm{inf}}$, where

$$(\nu)_{\mathrm{fin}} = \sum_{P \in W(\mathbf{C})} \mathrm{ord}_P \nu \quad \text{and} \quad (\nu)_{\mathrm{inf}} = \sum_{e \in \mathscr{E}(W)} \mathrm{ord}_e \nu.$$

Lemma 2.44. *Suppose f is a meromorphic function and ω is a meromorphic differential on $B(1) := B_{\mathbf{C}}(1)$, each with finitely many zeroes and poles, and each supported on $B[r] := B_{\mathbf{C}}[r]$ for some $r < 1$. Let $\mathscr{A} = A_{\mathbf{C}}(r, 1)$, oriented so that $\mathrm{res}_{\mathscr{A}} = \mathrm{res}_{r,1}$ (so not the inherited orientation from $B(1)$ as a wide open). Then*

$$\mathrm{ord}_{\mathscr{A}} f = \sum_{P \in B(1)} \mathrm{ord}_P f, \quad \mathrm{ord}_{\mathscr{A}} \omega = \sum_{P \in B(1)} \mathrm{ord}_P \omega, \quad \mathrm{res}_{\mathscr{A}} \omega = \sum_{P \in B(1)} \mathrm{res}_P \omega.$$

Proof. Let z be the natural parameter on $B(1)$. For the first equation, suppose f is supported on $\{P_1, \ldots, P_n\}$ with $\mathrm{ord}_{P_i} f = e_i$ and $z(P_i) = \alpha_i$. By the Weierstrass preparation theorem, we may write $f(z) = \prod_{i=1}^n (z - \alpha_i)^{e_i} \cdot u(z)$, where u is a unit. Then

$$\mathrm{ord}_{\mathscr{A}} f = \sum_{i=1}^n \mathrm{res}_{\mathscr{A}} \left(\frac{e_i \, \mathrm{d}z}{z - \alpha_i} \right) = \sum_{i=1}^n e_i = \sum_{P \in B(1)} \mathrm{ord}_P f.$$

The other two equations follow from essentially the same argument. $\qquad\square$

Theorem 2.45. *Let f be a rigid function and ω a differential on W, each with finitely many poles and zeroes in $W(\mathbf{C})$. Then*

(i) $\deg(f) = 0$,

(ii) $\deg(\omega) = 2g(W) - 2$, *and*

(iii) $\sum_{P \in W(\mathbf{C})} \mathrm{res}_P \omega + \sum_{e \in \mathscr{E}(W)} \mathrm{res}_e \omega = 0$.

Proof. Attach disks at the ends of W to obtain a smooth projective curve C. For any rational function g on C, it follows from Lemma 2.44 that $\deg(g|_W) = 0$.

For more general f, suppose first that (W, X) is a basic wide open pair, X has good reduction, and $(f)_{\mathrm{fin}}$ is supported on X. In this case, there exists a $g \in \mathbb{O}_C$ and a Zariski subaffinoid Y of X such that f/g is regular on Y and $|(f/g) - 1|_Y < 1$ (in particular, we could choose Y so that f and g have no poles or zeroes on Y). It follows that there is a wide open V, with $Y \subset V \subseteq W$, such that (V, Y) is a basic wide open pair and $|(f/g) - 1|_V < 1$. Hence, $(f|_V) = (g|_V)$. Now, we have a natural map $\beta : \mathrm{Div}(W) \to \mathrm{Div}(V)$. Indeed, the elements of $\mathscr{E}(V)$ are in one-to-one correspondence with the connected components of $V \setminus Y$, which in turn are in one-to-one correspondence (by intersection) with the connected components of $W \setminus Y = (W \setminus X) \cup (X \setminus Y)$. Thus, as $e \in \mathscr{E}(W)$ corresponds to a unique component of

$W \setminus X$, it then corresponds to a unique end of V that we take to be $\beta(e)$. Similarly, if $P \in W(\mathbf{C})$, we let $\beta(P) = P$ if $P \in V(\mathbf{C})$, and the element of $\mathscr{E}(V)$ corresponding to the component of $X \setminus Y$ that contains P otherwise. Extend this map by linearity. Since $\deg \beta(D) = \deg D$ and $(f|_V) = \beta(f)$, we have

$$\deg(f) = \deg(f|_V) = \deg(g|_V) = 0.$$

To complete the proof that $\deg(f) = 0$, let $\mathscr{C}$ be a semistable covering of W such that U^u has good reduction and $(f|_U)_{\text{fin}}$ is supported on U^u for each $U \in \mathscr{C}$ (which exists by Corollary 2.42). Then

$$(f) = \sum_{U \in \mathscr{C}} (f|_U),$$

where we regard both sides as elements of

$$\{D \in \mathbb{Z}^{W(\mathbf{C}) \cup \bigcup_{U \in \mathscr{C}} \mathscr{E}(U)} : D(a) = -D(b) \text{ if } a \in \mathscr{E}(U), b \in \mathscr{E}(V), U \neq V, U_a = V_b\}.$$

Therefore,

$$\deg(f) = \sum_{U \in \mathscr{C}} \deg(f|_U) = 0.$$

Statements (ii) and (iii) are clearly true whenever $\omega = \eta|_W$ for $\eta \in \Omega^1_C$, by Lemma 2.44. Moreover, the general case of (ii) then follows from (i) and the fact that $(f\omega) = (f) + (\omega)$. Finally, the general case of (iii) will follow once we know it for basic wide opens, by an argument similar to that above.

So suppose (W, X) is a basic wide open pair, with X and C as above. For a reduced affinoid X over K and $\omega \in \Omega^1_{X/K}$, we set

$$|\omega|_X = \inf\{|a| : a \in K, \omega \in aA^o(X) \mathrm{d}A^o(X)\}.$$

Using Riemann–Roch, we can find $\eta \in \Omega^1_C$ such that $\omega - \eta$ has no poles on X and $|\omega - \eta|_X < \epsilon$. Note that $\omega - \eta$ extends to a regular differential on a wide open neighborhood V of X in W. Then statement (iii) for W follows from the general fact that if (V, X) is a basic wide open pair, $\omega \in \Omega^1_{V/K}$, $|\omega|_X < \epsilon$ and $e \in \mathscr{E}(V)$, then $|\text{res}_e \omega| < \epsilon$. Indeed, let $T : U \to A(1, \infty)$ be a parameter on the component U of $V \setminus X$ corresponding to e, such that $|T(x)| \to 1$ as $x \to X$. Suppose on U

$$\omega = \sum_{n=-\infty}^{\infty} a_n T^n \mathrm{d}T.$$

Then $|\omega|_X = \max\{|a_n| : -\infty < n < \infty\}$. So $|\text{res}_e \omega| = |a_{-1}| < \epsilon$. $\qquad \square$

Suppose $f : W \to V$ is a finite map. As f is finite, f naturally maps $\mathscr{E}(W)$ to $\mathscr{E}(V)$. For $a \in W(\mathbf{C}) \cup \mathscr{E}(W)$, let $\delta_f(a) = \text{ord}_a f^* \mathrm{d}T$, where T is a parameter at $b := f(a)$ such that $\text{ord}_b T = 1$. When a and b are ends, there exist annuli $\mathscr{A}$

and $\mathcal{B}$ at a and b such that f restricts to a finite étale map from $\mathcal{A}$ onto $\mathcal{B}$. Let $e_f(a)$ be the degree of this map. Otherwise, at a point in $W(\mathbf{C})$, let $e_f(a)$ denote the usual ramification index.

Lemma 2.46. *With notation as above, if ω is a differential with finitely many zeroes and poles on W, then*

$$\mathrm{ord}_a \, f^* \omega = e_f(a) \, \mathrm{ord}_b \, \omega + \delta_f(a).$$

Proof. First suppose $a \in \mathscr{E}(W)$, and let $\mathcal{A}$ and $\mathcal{B}$ be annuli at a and b such that ω is regular and nonvanishing on $\mathcal{B}$. Choose parameters S and T on $\mathcal{A}$ and $\mathcal{B}$ respectively, such that $\mathrm{ord}_a \, S = \mathrm{ord}_b \, T = 1$. Then $f^* T|_{\mathcal{A}} = S^e g(S)$ and $\omega|_{\mathcal{B}} = T^d h(T) \mathrm{d} T$, where g is a unit on $\mathcal{A}$ with $\mathrm{ord}_a \, g = 0$, h is unit on $\mathcal{B}$ with $\mathrm{ord}_b \, h = 0$, $e = e_f(a)$, and $d = \mathrm{ord}_b \, \omega$. So

$$(f^* \omega)|_{\mathcal{A}} = (S^e g(S))^d h(S^e g(S)) f^* \mathrm{d} T,$$

from which the lemma follows.

The proof when $a \in W(\mathbf{C})$ is very similar. Let $\mathcal{A}$ and $\mathcal{B}$ be the stalks at a and b. Then after choosing uniformizers S and T, respectively, the map $f : \mathcal{A} \to \mathcal{B}$ is given by a homomorphism between formal power series rings over $\mathbf{C}$. Thus, we have $\omega = T^d h(T) \mathrm{d} T$ and $f^* T = S^e g(S)$, where $e = e_f(a)$, $d = \mathrm{ord}_b \, \omega$, and g and h are formal power series with nonzero constant terms. The lemma follows from the computation above and the fact that $h(S^e g(S))$ will again have nonzero constant term. □

Corollary 2.47. *Suppose that $|e_f(a)| = 1$ in K, or that $e_f(a) \neq 0$ in K and $a \in W(\mathbf{C})$. Then $\delta_f(a) = e_f(a) - 1$.*

Proof. Keeping the same notation as above, we compute $\delta_f(a)$ directly from the definition

$$\delta_f(a) = \mathrm{ord}_a \, \mathrm{d} T = \mathrm{ord}_a \, \mathrm{d}(S^e g(S)) = \mathrm{ord}_a (eg(S) + Sg'(S)) + e - 1.$$

When $a \in W(\mathbf{C})$ and $e_f(a) \neq 0$ in K, this equals $e - 1$, since the power series $eg(S) + Sg'(S)$ must have nonzero constant term. On the other had, if $a \in \mathscr{E}(W)$ and $|e_f(a)| = 1$, it is straightforward to show that $eg(S) + Sg'(S)$ has constant absolute value on $\mathcal{A}$. So either way we are done. □

Theorem 2.48. *Let $f : W \to V$ be a finite map of wide opens of degree d. Then*

$$2g(W) - 2 = d(2g(V) - 2) + \sum_{a \in W(\mathbf{C}) \cup \mathscr{E}(W)} \delta_f(a).$$

Furthermore, under the hypotheses of [Corollary 2.47](#), this is

$$d(2g(V) - 2) + \sum_{a \in W(\mathbf{C}) \cup \mathscr{E}(W)} (e_f(a) - 1).$$

Proof. Let ω be a nonzero meromorphic differential on V. Then the degree of $f^*\omega$ must be $2g(W) - 2$ by Theorem 2.45. On the other hand, we obtain the right side of the equation if we compute $\deg(f^*\omega)$ using Lemma 2.46, Corollary 2.47, and

$$\sum_{\substack{x \in W(\mathbf{C}) \\ f(x) = y}} e_f(x) = \sum_{\substack{a \in \mathscr{E}(W) \\ f(a) = b}} e_f(a) = d. \qquad \square$$

Proposition 2.49. *Suppose (W, W^u) and (V, V^u) are basic wide open pairs, and $f : W \to V$ is a finite map such that $f(W^u) = V^u$. Let X and Y be the completions of $\overline{W^u}$ and $\overline{V^u}$, and let $\bar{f} : X \to Y$ be the induced map. If $\bar{f}$ is separable, then*

$$\delta_f(a) = \mathrm{length}(\Omega_{X/Y})_a.$$

Proof. First, we can lift $\bar{f}$ to a map $g : C_X \to C_Y$ between complete liftings of X and Y. There must exist wide open neighborhoods $W' \subseteq W$ and $V' \subseteq V$ of W^u and V^u, and embeddings $\phi_X : W' \to C_X$ and $\phi_Y : V' \to C_Y$, such that $f(W') = V'$, $\overline{\phi_X|_{W^u}}$ and $\overline{\phi_Y|_{V^u}}$ are the natural inclusions, and

$$\overline{g \circ \phi_X|_{W^u}} = \overline{\phi_Y \circ f|_{W^u}}.$$

Now, Hartshorne's version [1977, Corollary 2.4] of Hurwitz's theorem implies the proposition for

$$h := \phi_Y^{-1} \circ g \circ \phi_X : W' \to V'.$$

The proposition follows because $\delta_f(a) = \delta_f(a') = \delta_g(a'')$, where a' is the component of $W' \setminus W^u$ corresponding to a and $a'' = \phi_X(a')$. $\qquad \square$

3. $X_0(p^n)$ and its subspaces

Now that the rigid analytic foundation has been laid, we turn our focus specifically to the curve $X_0(p^n)$, which we will always think of in moduli-theoretic terms. More precisely, we think of $X_0(p^n)$ as the rigid analytic curve over $\mathbb{Q}_p$ whose points over $\mathbb{C}_p$ are in a one-to-one correspondence with (isomorphism classes of) pairs (E, C), where $E/\mathbb{C}_p$ is a generalized elliptic curve and C is a cyclic subgroup of order p^n. We implicitly make use of this correspondence when we speak loosely of "the point (E, C)". There are various natural maps from $X_0(p^n)$ to $X_0(p^m)$ that can be defined by way of this moduli-theoretic interpretation of points, and we begin this section by fixing notation for these fundamental maps.

Definition 3.1. First let

$$\pi_f, \pi_v : \coprod_{n \geq 1} X_0(p^n) \to \coprod_{n \geq 0} X_0(p^n)$$

be the maps given by $\pi_f(E, C) = (E, pC)$ and $\pi_v(E, C) = (E/C[p], C/C[p])$, where $C[p^i]$ is the kernel of multiplication by p^i in C. Then by letting $\pi_{ab} =$

$\pi_f^b \circ \pi_v^a$, we get maps

$$\pi_{ab} : \coprod_{n \geq a+b} X_0(p^n) \to \coprod_{n \geq 0} X_0(p^n).$$

Remark 3.2. This definition is identical to the definition in [Coleman 2005, §1]. We also note that over $\mathbb{C}$, π_{ab} corresponds to the map on the upper half plane that takes z to $p^a z$.

Another map crucial to this paper is the Atkin–Lehner involution,

$$w : \coprod_{n \geq 0} X_0(p^n) \to \coprod_{n \geq 0} X_0(p^n),$$

which is defined by the formula

$$w_n(E, C) = (E/C, E[p^n]/C), \quad \text{where } w_n := w|_{X_0(p^n)}.$$

The Atkin–Lehner involution is compatible with the level-lowering maps in the sense that $\pi_f \circ w = w \circ \pi_v$ (or equivalently, $w \circ \pi_f = \pi_v \circ w$, since w is an involution).

3A. *Canonical subgroups and supersingular annuli.* We now introduce some natural rigid subspaces of $X_0(p^n)$ over finite extensions of $\mathbb{Q}_p$ using the theory of the canonical subgroup, which we now review and extend [Buzzard 2003, §3].[15] If E is an elliptic curve over $\mathbb{C}_p$, we let $h(E)$ denote the minimum of 1 and the valuation of a lifting of the Hasse invariant of the reduction of a nonsingular model of E mod p, if it exists, and 0 otherwise[16] (this is denoted by $v(E)$ in [Buzzard 2003]). Katz [1973, §3] constructed a rigid analytic section s_1 of $\pi_f : X_0(p) \to X(1)$ over the wide open W_1 whose $\mathbb{C}_p$-valued points are represented by generalized elliptic curves E such that $h(E) < p/(p+1)$, when $p \geq 5$. Both W_1 and s_1 are defined over $\mathbb{Q}_p$. Changing notation slightly from [Buzzard 2003], we let $K_1(E) \subseteq E$ denote the subgroup of order p for which $s_1(E) = (E, K_1(E))$, and we call $K_1(E)$ the canonical subgroup of order p.

Using [Buzzard 2003, Theorem 3.3], we can also define canonical subgroups of higher order. For $n \geq 1$, we generalize W_1 by taking W_n to be the wide open in $X(1)$ where $h(E) < p^{2-n}/(p+1)$ (the complement of finitely many affinoid disks, one in each supersingular residue class). For $E \in W_n$ we then define $K_n(E)$ inductively, as in [ibid., Definition 3.4], as the preimage of $K_{n-1}(E/K_1(E))$ under the natural projection from $E \to E/K_1(E)$. This is a cyclic subgroup when $E \in W_n$

---

[15]Although Buzzard works over a complete discrete valuation ring, all of his results can be extended to complete local rings.

[16]As pointed out in [BL 1985, Remark 6.4], the good reduction of E is well defined if it exists.

by [ibid., Theorem 3.3], and we call it the canonical subgroup of order p^n.[17] Thus, when E has supersingular reduction, either $h(E) \geq p/(p+1)$ (and E is too super-singular in the language of [ibid.]), or there is a largest $n \geq 1$ for which $K_n(E)$ can be defined. In the first case, we define the *canonical subgroup* of E, written $K(E)$, to be the trivial subgroup, and in the second we let $K(E) = K_n(E)$ for this largest n. Whenever $E/\mathbb{C}_p$ has ordinary reduction (by this we mean ordinary good or multiplicative[18]) we let $K(E)$ be the p-power torsion of E that is contained in the kernel of reduction, which does not depend on the good or multiplicative model.

It is important to note that s_1 also generalizes, in the sense that the map defined by $s_n(E) = (E, K_n(E))$ is also a rigid analytic section of $\pi_{0n} : X_0(p^n) \to X(1)$ over the wide open W_n. To see this, first regard $X_0(p^n)$ for $n > 1$ as the normalization of the fiber product of $X_0(p^{n-1})$ with itself over $X_0(p^{n-2})$ via the maps π_f and π_v. More specifically, let

$$\psi_n : X_0(p^n) \to X_0(p^{n-1}) \times_{\pi_f, \pi_v} X_0(p^{n-1})$$

be the isomorphism described by $\psi_n = (\pi_v, \pi_f)$ (after normalization of the right side). Now assume that s_{n-1} is rigid analytic. With [Buzzard 2003, Theorem 3.3], it is straightforward to verify that over W_n, we have

$$\pi_f \circ s_{n-1} \circ \pi_{1n-2} \circ s_{n-1} = \pi_v \circ s_{n-1}.$$

Thus we may define a rigid analytic map from W_n to $X_0(p^n)$ by

$$s_n := \psi_n^{-1} \circ (s_{n-1} \circ \pi_{1n-2} \circ s_{n-1}, s_{n-1}).$$

Again by the same theorem, we see that this map does indeed take E to $(E, K_n(E))$. So by induction we are done. Note that both W_n and s_n are defined over $\mathbb{Q}_p$.

Another way to focus on rigid subspaces of $X_0(p^n)$ is to fix the isomorphism class of the reduction of E. In particular, we make the following definition.

Definition 3.3. For a fixed elliptic curve A over a finite field $\mathbb{F}$, we let $W_A(p^n)$ represent the rigid subspace of $X_0(p^n)$ (over $\mathbb{Q}_p \otimes W(\mathbb{F})$) whose points over $\mathbb{C}_p$ are represented by pairs (E, C) with $\bar{E} \cong A$.

Of course, $W_A(1)$ (for any A) is just a residue disk of the j-line. When A is a supersingular curve, it is well known that $W_A(p)$ is isomorphic, over $\mathbb{Q}_{p^2} := W(\mathbb{F}_{p^2}) \otimes \mathbb{Q}_p$, to an open annulus of width $i(A) := |\mathrm{Aut}(A)|/2$. This means that one can choose a parameter x_A on $W_A(p)$ over $\mathbb{Q}_{p^2}$ that identifies it with the (open)

---

[17]Thinking of the kernel of reduction of E as a disk, the set of points of order p^n are not always equidistant from the identity. When $E \in W_n$, $K_n(E)$ is the union over $i \leq n$ of those points of order p^i that are closest to the identity.

[18]Equivalently, $j(E)$ is not congruent to a supersingular j-invariant modulo $m_{\mathbb{R}_p}$.

annulus $A_{\mathbb{Q}_{p^2}}(p^{-i(A)}, 1)$. In fact, we can and will always do this in such a way that $v(x_A(E, C)) = i(A)h(E)$ when $C = K_1(E)$, and $i(A)(1 - h(E/C))$ otherwise (this is justified in [Buzzard 2003, Theorem 3.3 and §4]).

Now, inside the annulus, $W_A(p)$, there are three concentric circles that will be essential for our analysis of $X_0(p^2)$ and $X_0(p^3)$. First there is the *too-supersingular circle*, which we denote by $\mathbf{TS}_A$, whose points correspond to pairs (E, C), where the canonical subgroup of E is trivial. Equivalently, these are the points with $h(E) \geq p/(p+1)$. Next there is the *self-dual circle*, denoted by $\mathbf{SD}_A$, whose points correspond to pairs (E, C), where the subscheme C of order p is potentially self-dual, that is, isomorphic to its Cartier dual after finite base extension. Equivalently, $\mathbf{SD}_A$ consists of those points that satisfy $h(E) = 1/2$ and $C = K_1(E)$. When $A/\mathbb{F}_p$, $\mathbf{SD}_A$ can also be described as the unique circle in $W_A(p)$ that is fixed by the involution w_1, and hence we call it the *Atkin–Lehner circle*. Finally, we must also consider what might be called the *anti-Atkin–Lehner circle*. It is the subspace $\mathbf{C}_A \subseteq W_A(p)$ whose points correspond to pairs (E, C') for which there exists a C such that $(E, C) \in \mathbf{SD}_A$ but $C' \neq C$. We let $\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$ be the map that corresponds to replacing the cyclic subgroup C' with $K_1(E)$. Then τ_f is rigid analytic since it is the restriction of $s_1 \circ \pi_f$.

Remark 3.4. The fact that the regions above are circles follows from Buzzard's discussion of rigid subspaces of $X_1(p)$ [2003, §4]. Using a parameter x_A chosen as above, the circles $\mathbf{TS}_A$, $\mathbf{SD}_A$ and $\mathbf{C}_A$ are those where $v(x_A)/i(A)$ equals $p/(p+1)$, $1/2$, and $1 - 1/(2p)$, respectively.

From above, whenever $A/\mathbb{F}_p$ is supersingular, $W_A(p)$ is an annulus preserved by the Atkin–Lehner involution w_1, and which is mapped onto the residue disk $W_A(1)$ via π_f. In our analysis of the stable models of $X_0(p^2)$ and $X_0(p^3)$, we will need to work with fairly explicit approximations for the restrictions of π_f and w_1 to these subspaces:

Theorem 3.5. *Let* $\mathbb{Z}_{p^2} := W(\mathbb{F}_{p^2})$ *and let* $A/\mathbb{F}_p$ *be a supersingular curve with* $j(A) \neq 0, 1728$. *Then there are parameters* s *and* t *over* $\mathbb{Z}_{p^2}$ *that identify* $W_A(1)$ *with the disk* $B_{\mathbb{Q}_{p^2}}(1)$ *and* $W_A(p)$ *with the annulus* $A_{\mathbb{Q}_{p^2}}(p^{-1}, 1)$, *and there are series* $F(T), G(T) \in T\mathbb{Z}_{p^2}[\![T]\!]$ *such that*

(i) $w_1^*(t) = \kappa/t$ *for some* $\kappa \in \mathbb{Z}_{p^2}$ *with* $v(\kappa) = 1$, *and*

(ii) $\pi_f^* s = F(t) + G(\kappa/t)$, *where* $F'(0) \equiv 1 \pmod{p}$ *and* $G(T) \equiv (F(T))^p \pmod{p}$.

Proof. One only has to translate results in [de Shalit 1994, §2, §3]. Our t and κ are de Shalit's y and π. Then our $\pi_f^* s$ is de Shalit's $\psi(y) - \beta_0$. The theorem follows from [de Shalit 1994, (4) of §2, Lemma 1 and Corollaries 2–4 of §3]. $\qquad\square$

Note that the parameter t from Theorem 3.5 is a suitable choice for x_A. This follows from condition (ii), which guarantees that π_f has degree $p+1$ on the circle where $v(t) = p/(p+1)$ and has degree 1 or p on all other concentric circles.

3B. *Neighborhoods of the ordinary locus.* The finitely many subspaces $W_A(p^n)$ (defined above), where A runs over supersingular curves over $\mathbb{F}_{p^2}$, cover the supersingular locus of $X_0(p^n)$ over $\mathbb{Q}_{p^2}$, that is, the subspace whose points over $\mathbb{C}_p$ correspond to pairs (E, C), where E has supersingular reduction. Furthermore, these subspaces become connected wide opens over $\mathbb{C}_p$ by Theorem 2.29. We will now describe a finite collection $W_{ab}^{\pm} \subseteq X_0(p^n)$ of subspaces that cover the ordinary locus. These will, in fact, be shown to be basic wide opens when $n \leq 3$, and we do expect this to hold more generally. Essentially, we extend the irreducible affinoids, $\mathbf{X}_{ab}^{\pm}$ (introduced in [Coleman 2005][19]), to wide open neighborhoods, by considering points (E, C) that are *nearly* ordinary in the sense that either $K(E)$ or $K(E/C)$ is large.

More precisely, for $a \geq b \geq 0$ with $a + b = n$, we start by letting

$$W_{ab} = \{(E, C) : |K(E)| \geq p^n, |K(E) \cap C| = p^a\}.$$

For $b > a \geq 0$ with $a + b = n$, we then define $W_{ab} = w_n(W_{ba})$. Now we show that the pairing on $K_a(E)$, which was defined in [Coleman 2005] for points $(E, C) \in W_{ab}$ where E has ordinary reduction, carries over to all points in W_{ab}. Let (E, C) be a point of W_{ab} with $a \geq b$, and let $A, B \in K_a(E)$. Then by the definition of W_{ab}, we can choose $P \in C$ and $Q \in K_n(E)$ such that $p^b P = A$ and $p^b Q = B$. Now set $\mathscr{P}_{E,C}(A, B) = e_n(P, Q)$, where $e_n(\cdot, \cdot)$ denotes the Weil pairing on $E[p^n]$. This gives a well-defined pairing of $K_a(E)$ with itself onto μ_{p^b}. Furthermore, if $p > 2$, there are exactly two isomorphism classes of pairings on $\mathbb{Z}/p^a\mathbb{Z}$ onto μ_{p^b} whenever $b > 0$. Let $e^+(\cdot, \cdot)$ and $e^-(\cdot, \cdot)$ be representatives for these classes. Then, essentially repeating the argument from [Coleman 2005] for the ordinary affinoids, $\mathbf{X}_{ab}^{\pm}$, there is a rigid subspace $W_{ab}^{\pm}$ of $X_0(p^n)$ defined over $\mathbb{Q}_p(\sqrt{(-1)^{(p-1)/2}})$ whose $\mathbb{C}_p$-valued points are

$$\{(E, C) \in W_{ab} \mid (K_a(E), \mathscr{P}_{E,C}) \cong (\mathbb{Z}/p^a\mathbb{Z}, e^{\pm})\}.$$

Set $W_{n0}^+ = W_{n0}^- = W_{n0}$, and for $b > a \geq 0$, set $W_{ab}^{\beta} = w_n(W_{ba}^{(\frac{-1}{p})\beta})$.

Thus, $\mathbf{X}_{ab}^{\pm}$ is just the affinoid whose points are those $(E, C) \in W_{ab}^{\pm}$ for which E has ordinary or multiplicative reduction. It is *not* immediate that $W_{ab}^{\pm}$ is a basic wide open with $\mathbf{X}_{ab}^{\pm}$ as a minimal underlying affinoid. We will show that this is the case, however, when $n \leq 3$, and we do expect it to hold for arbitrary n as well.

---

[19]When $a < b$, the X_{ab}^{β} here is the same as $X_{ab}^{(\frac{-1}{p})\beta}$ from [Coleman 2005].

The affinoid $\mathbf{X}_{ab}^{\pm}$ is well understood from results of [Coleman 2005]. In particular, we have the following result, proven but not made explicit therein.

Proposition 3.6. *The affinoid $\mathbf{X}_{ab}^{\pm}$ with $a \geq b > 0$ is defined and has good reduction over $\mathbb{Q}_p(\mu_{p^b})$.*

Proof. It was proven in [Coleman 2005, §0] that $\mathbf{X}_{ab}^{\pm}$ is an affinoid defined over the quadratic subfield of $\mathbb{Q}_p(\mu_p)$. For $\zeta \in \mu_{p^b}$, we can define an embedding a_ζ of $\mathbf{X}_{ab}^{\pm}$ onto an affinoid in $X_1(p^b)^{\mathrm{bal}}$ by taking $a_\zeta(E, C)$ to be the point that is represented by the balanced $\Gamma_1(p^b)$-structure [Katz and Mazur 1985, (3.3)]

$$P, E \; \underset{\check{\alpha}}{\overset{\alpha}{\rightleftarrows}} \; E/C, P'.$$

Here we have $P \in K_b(E)$, $\mathcal{P}_{E,C}(P, P) = \zeta$, and $P' = \alpha(Q)$ for some $Q \in E[p^b]$ such that $(P, Q) = \zeta$. This image affinoid reduces to $\mathrm{Ig}(p^b)$ by (the extension to level 1 of) [Katz and Mazur 1985, p. 450].[20] $\qquad \square$

Corollary 3.7. *The affinoid $\mathbf{X}_{ab}^{\pm}$ with $a + b = n$ is defined and has good reduction over $\mathbb{Q}_p(\mu_{p^{\lfloor n/2 \rfloor}})$.*

Proof. When $a \geq b$, this follows immediately from Proposition 3.6. Otherwise, apply w_n first. $\qquad \square$

4. Formal groups

In the previous section we defined a finite collection $W_A(p^n)$ of connected wide opens that cover the supersingular locus of $X_0(p^n)$. Unfortunately, $W_A(p^n)$ is only basic when $n \leq 2$. Therefore, to arrive at a stable covering of $X_0(p^n)$, it is necessary to use smaller subspaces of $W_A(p^n)$. One approach is to use canonical subgroup considerations, as in Section 3A. Another is to use the interpretation from [Lubin et al. 1964] of an elliptic curve, over a complete local ring R with residue characteristic p, as a lifting of some formal group in characteristic p. In particular, this will enable us to use explicit formulas of Hopkins and Gross, which we recall in Section 4B.

Theorem 4.1 (Woods Hole theory). *Suppose R is the ring of integers in a complete subfield of $\mathbb{C}_p$, with residue field $\mathbb{F}$. The category of elliptic curves over R is equivalent to the category of triples (F, A, α), where F is a formal group over R, A is an elliptic curve over $\mathbb{F}$, and $\alpha : \overline{F} \to \hat{A}$ is an isomorphism. A morphism between two triples, (F, A, α) and (F', A', β), is a pair (σ, τ), where $\sigma : F \to F'$*

---

[20]$\mathrm{Ig}(p^b)$ is the Igusa curve in characteristic p that classifies pairs (E, ψ), where E is an elliptic curve and $\psi : \mu_{p^b} \hookrightarrow E$ (studied in [Igusa 1968]).

and $\tau : A \to A'$ are homomorphisms such that the following diagram commutes.

$$
\begin{array}{ccc}
\bar{F} & \xrightarrow{\bar{\sigma}} & \bar{F}' \\
\alpha \downarrow & & \downarrow \beta \\
\hat{A} & \xrightarrow{\hat{\tau}} & \hat{A}'
\end{array}
$$

Proof. If E is an elliptic curve over R, let $\mathscr{F}_R(E) = (\hat{E}, \bar{E}, \iota)$, where $\iota : \bar{\hat{E}} \to \hat{\bar{E}}$ is the natural isomorphism. This is a functor, compatible with changing R, from the first category to the second. We claim this is an equivalence of categories. The analogous statement is proven when R is a local Artinian ring with residue field of characteristic p in [Lubin et al. 1964, §6]. Then on [Lubin et al. 1964, p. 7], it is explained that by "passing to the limit... one sees that it continues to hold over a complete local Noetherian ring". Thus the theorem is true when R is the ring of integers in a complete discretely valued subfield of $\mathbb{C}_p$.

To obtain it more generally, we apply [Lubin et al. 1964, Theorem 1]. This theorem implies that given an elliptic curve A over an algebraic extension of $\mathbb{F}_p$, the collection of liftings of $\hat{A}$ to $\mathbb{R}_p$ is naturally the set of points in a wide open disk $\mathscr{D}$. On the other hand, the set of liftings of A to $\mathbb{R}_p$ is the set of points in a residue disk $\mathscr{R}$ of $X(1)$ and $\mathscr{F}$ yields a degree one rigid analytic map from $\mathscr{R}$ to $\mathscr{D}$ with dense image. Hence it is an isomorphism. $\qquad\square$

In light of this theorem, we may think of points $(E, C) \in W_A(p^n)$ as triples (F, C, α), where F is a formal group, $C \subseteq F$ is a cyclic subgroup of order p^n, and $\alpha : \bar{F} \to \hat{A}$ is an isomorphism. We then refer to such a triple as a Woods Hole representation of (E, C). There are two specific ways in which we apply this theory. First of all, from the fact that all supersingular elliptic curves are p-prime isogenous, we are able to show that all supersingular regions $W_A(p^n)$ for a fixed p and n are *nearly* isomorphic. Along with the result in Appendix B, this enables us to do all of our calculations under the simplifying assumption that $A/\mathbb{F}_p$ and $j(A) \neq 0, 1728$. Secondly, we use extensively the natural action of the p-adic group $\mathrm{Aut}(\hat{A})$ on $W_A(p)$, which was studied in detail in [Hopkins and Gross 1994].

4A. *All supersingular regions are (nearly) isomorphic.*

Proposition 4.2. *Let A and $A'/\mathbb{F}_{p^2}$ be two supersingular elliptic curves, with $j(A)$ not equal to 0 or 1728. Let $\mathbb{F}/\mathbb{F}_{p^2}$ be a finite extension over which A and A' are p-prime isogenous (which always exists). Then the wide open $W_{A'}(p^n)$ is isomorphic over $W(\mathbb{F}) \otimes \mathbb{Q}_p$ to the quotient of $W_A(p^n)$ by a faithful action of $\mathrm{Aut}(A')/\{\pm 1\}$.*

Proof. Suppose $\iota : A \to A'$ is an isogeny of degree prime to p over $\mathbb{F}$. Since $(\deg \iota, p) = 1$, the induced map $\hat{\iota} : \hat{A} \to \hat{A}'$ is an isomorphism of formal groups.

So in Woods Hole terms we may define a map $\psi_\iota : W_A(p^n) \to W_{A'}(p^n)$ by taking

$$\psi_\iota(F, C, \alpha) = (F, C, \hat{\iota} \circ \alpha).$$

To show that the map is, in fact, well defined, suppose that the triples (F_1, C_1, α_1) and (F_2, C_2, α_2) represent the same point of $W_A(p^n)$. This means that there are isomorphisms $\gamma : F_1 \to F_2$ (mapping C_1 to C_2) and $\tau : A \to A$ such that $\alpha_2 \circ \overline{\gamma} = \hat{\tau} \circ \alpha_1$. Because $j(A) \neq 0, 1728$, we know that $\tau = \pm 1$. Therefore $\hat{\tau}$ commutes with all isogenies. In particular, composing with $\hat{\iota}$ on both sides, we have

$$\hat{\iota} \circ \alpha_2 \circ \overline{\gamma} = \hat{\tau} \circ \hat{\iota} \circ \alpha_1.$$

Therefore $(F_1, C_1, \hat{\iota} \circ \alpha_1)$ and $(F_2, C_2, \hat{\iota} \circ \alpha_2)$ are Woods Hole representations of the same point in $W_{A'}(p^n)$, and ψ_ι is well defined.

To show that ψ_ι is onto, choose any point of $W_{A'}(p^n)$ and let (F, C, β) be one of its Woods Hole representations (so $\beta : \overline{F} \to \hat{A}'$ is an isomorphism). Since $\hat{\iota}$ is an isomorphism, we can choose a point of $W_A(p^n)$ by taking $(F, C, \hat{\iota}^{-1} \circ \beta)$, and this point maps onto our chosen point of $W_{A'}(p^n)$ by definition. (Note, however, that this does *not* define a map from $W_{A'}(p^n)$ to $W_A(p^n)$, since our original choice of triple was noncanonical.)

Finally, suppose that two points of $W_A(p^n)$, represented by (F_1, C_1, α_1) and (F_2, C_2, α_2), have the same image in $W_{A'}(p^n)$. Then there must be isomorphisms $\gamma : F_1 \to F_2$ (taking C_1 to C_2) and $\tau : A' \to A'$ such that

$$\hat{\iota} \circ \alpha_2 \circ \overline{\gamma} = \hat{\tau} \circ \hat{\iota} \circ \alpha_1 \quad \text{and} \quad \alpha_2 \circ \overline{\gamma} = \widehat{\mathrm{id}} \circ (\hat{\iota}^{-1} \circ \hat{\tau} \circ \hat{\iota}) \circ \alpha_1.$$

In particular, $\tau \mapsto ((F, C, \alpha) \mapsto (F, C, \hat{\iota}^{-1}\tau\hat{\iota} \circ \alpha))$ gives a faithful action of $\mathrm{Aut}(A')/\{\pm 1\}$ on the fibers of ψ. $\square$

Remark 4.3. Suppose now that $\mathbb{F} \supseteq \mathbb{F}_{p^2}$ is a field over which all supersingular curves are p-prime isogenous. It follows, then, that all of the regions $W_A(p^n)$ are nearly isomorphic over $W(\mathbb{F}) \otimes \mathbb{Q}_p$. We showed in [CM 2006, Theorem 5.5] that this $\mathbb{F}$ can always be taken to be $\mathbb{F}_{p^{24}}$.

4B. *Woods Hole action and Gross–Hopkins theory.* The other way we use Woods Hole theory is to define a continuous action of a p-adic group on $W_A(p^n)$. In particular, when A is a supersingular elliptic curve, it is well known [Tate 1966, Main Theorem] that

$$B := \mathrm{End}(\hat{A}) \cong \mathbb{Z}_p[i, j, k],$$

where i^2 is a quadratic nonresidue, $j^2 = -p$, and $ij = -ji = k$. Furthermore, we may take j to be the Frobenius endomorphism whenever A is defined over $\mathbb{F}_p$. Then $B^* = \mathrm{Aut}(\hat{A})$ acts on $W_A(p^n)$ by

$$\rho(F, C, \alpha) = (F, C, \rho \circ \alpha) \quad \text{for } \rho \in B^*.$$

Remark 4.4. The subgroup $\mathbb{Z}_p^* \subseteq B^*$ acts trivially on $W_A(p^n)$. Indeed, for $\rho \in \mathbb{Z}_p^*$, just take $\sigma = \rho^{-1}$ and $\tau = \mathrm{id}$ in Theorem 4.1. Not only does this define an isomorphism between (F, α) and $(F, \rho \circ \alpha)$, but in fact the isomorphism leaves invariant the subgroups of F of order p^n.

Hopkins and Gross studied the analogous action for deformation spaces of finite height formal groups, and explicitly computed the action in the height 2 case in [1994, §25]. To better understand their results (and translate them into our setting), we now offer a brief review of their theory under suitable simplifying assumptions. First, let K be a finite unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F} \supseteq \mathbb{F}_{p^2}$, and let $F_0/\mathbb{F}$ be a fixed height 2 formal group. They show that there is a rigid space over K, denoted by X_K, whose L-valued points for any finite extension L of K correspond to liftings of F_0 to a formal group over $\mathbb{O}_L$. Here two liftings are equivalent (say, (G_1, γ_1) and (G_2, γ_2) with $\gamma_i : \overline{G}_i \xrightarrow{\sim} F_0$) if there is an isomorphism between them that induces the identity on F_0. Then $\mathrm{Aut}(F_0)$ acts (rigid analytically) on X_K in the same manner as above, and Hopkins and Gross make this action completely explicit with their crystalline period mapping

$$\Phi : X_K \to \mathbf{P}_K^1,$$

which can be understood as follows. Again, it is well known that $B := \mathrm{End}(F_0)$ is isomorphic to the maximal order of some quaternion algebra over $\mathbb{Q}_p$, and hence $B \otimes K$ is (noncanonically) isomorphic to $M_{2 \times 2}(K)$. Since the image of B^* in $M_{2 \times 2}(K)$ must take lines to lines, we thus obtain an action of B^* on $\mathbf{P}_K^1$. Hopkins and Gross define the (rigid analytic) map Φ and decompose $B \otimes K$ in such a way that $\Phi(\rho x) = \Phi(x)^\rho$ for all $\rho \in B^*$, that is, Φ is B^*-equivariant. So the beauty of their theory is that the action of B^* on X_K can be concretely expressed in terms of linear algebra.

Indeed, suppose $A/\mathbb{F}_p$ with $j(A) \neq 0, 1728$. Then X_K and $W_A(1)$ are naturally isomorphic over the unramified quadratic extension K of $\mathbb{Q}_p$, and we may decompose B as $R \oplus Rj$, where

$$R = \mathbb{Z}_p[i] \cong \mathbb{O}_K$$

and j is the Frobenius endomorphism of A (as above). Then by [Hopkins and Gross 1994, §25], $\rho = \alpha + j\beta \in B^*$ (with $\alpha, \beta \in \mathbb{Z}_p[i]$) acts on $\mathbf{P}^1(K) = \Phi(X_K)$ via multiplication on the *right* by the matrix

$$\begin{bmatrix} \alpha & -p\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix}. \tag{3}$$

Of course, this formula only completely defines the action of B^* on B^*-stable subspaces of $W_A(1)$ on which Φ is an injection (for example, the canonical liftings of [Gross 1986, 2.1]). Hopkins and Gross [1994, 25.12] specify an affinoid

disk $Y \subseteq W_A(1)$ for which this is the case, and which maps via Φ onto the disk $v(t) \geq 1/p$, where t is the parameter on $\mathbf{P}^1$ corresponding to the row vector $[1, t]$. This parameter is distinct from the parameter on $W_A(p)$ from Theorem 3.5. However, from the explicit action of B^* on $\Phi(Y)$ and the fact that this t vanishes at some canonical lifting (which is necessarily too supersingular), there is significant compatibility between the two. In particular, it is clear that the canonical section of $\pi_f : W_A(p) \to W_A(1)$ exists over the annulus in $Y \subseteq W_A(1)$ described by $1/p < v(t) < p/(p+1)$ and preserves valuations with respect to the two parameters. As B^* acts equivariantly with respect to π_f, the upshot of all this is that B^* acts on the subannulus of $W_A(p)$ that is identified via $\Phi \circ \pi_f$ with the annulus $1/p < v(t) < p/(p+1)$ according to

$$\rho(t) = \frac{-p\bar{\beta} + \bar{\alpha}t}{\alpha + \beta t}, \quad \text{where } \rho = \alpha + j\beta. \tag{4}$$

In particular, we are most interested in the action of B^* on the Atkin–Lehner circle (equivalently, where $v(t) = 1/2$). The following proposition and remark summarize the specific results (still assuming that $A/\mathbb{F}_p$ and $j(A) \neq 0, 1728$) which we will need for our explicit analysis of $X_0(p^3)$.

Definition 4.5. For $\rho = \alpha + j\beta$ (as above), let $\rho' = \bar{\alpha} + j\bar{\beta}$, and let B' be the set of all $\rho \in B^*$ such that $\rho\rho' \in \mathbb{Z}_p^*$. Alternatively, B' is just the set of all $\rho \in B^*$ with $\rho = a + bi + dk$.

Proposition 4.6. *Let $j(A) \neq 0, 1728$. For any $\rho \in B^*$, let $w_\rho := \rho \circ w_1$. Then w_ρ is an automorphism of $\mathbf{SD}_A$ with two fixed points and is an involution exactly when $\rho \in B'$.*

Proof. If $(E_2, K(E_2)) = w_1(E_1, K(E_1))$ are two points of $\mathbf{SD}_A$, this means that there is a degree p isogeny $f : E_1 \to E_2$ with kernel $K(E_1)$. Since A is supersingular with $\mathrm{Aut}(A) = \pm 1$, f can only induce $\pm j$ in $\mathrm{End}(A)$ (and hence in $B = \mathrm{End}(\hat{A})$). Now, $j \notin B^*$, but the full group, $(B \otimes K)^\times$, acts equivariantly on $\Phi(X_K)$ by [Hopkins and Gross 1994, 23.11]. So this means that on $\mathbf{SD}_A$ we may identify w_1 with $\pm j$ (and the sign is irrelevant).

To determine when w_ρ is an involution, we first verify that $\rho \circ w_1 = w_1 \circ \rho'$ (equivalently, $\rho j = j\rho'$). This shows that w_ρ^2 acts like $\rho\rho'$, and only $\mathbb{Z}_p^* \subseteq B^*$ acts trivially from (4). So w_ρ is an involution exactly when $\rho \in B'$. In particular, w_ρ is given by

$$w_\rho(t) = \frac{-p\bar{\alpha} - p\bar{\beta}t}{-p\beta + \alpha t}, \tag{5}$$

and the explicit formula shows that in any case w_ρ has two fixed points. $\qquad\square$

Remark 4.7. To better understand how $\rho \in B^*$ and w_ρ act on $\overline{\mathbf{SD}}_A$, we could choose the parameter $u = t/\sqrt{-p}$ that identifies $\mathbf{SD}_A$ with $C[1]$. Then, by reducing

equations (4) and (5) from above, on $\overline{\mathbf{SD}}_A \cong \mathbf{G}_m$ we have

$$\rho\bar{u} = \bar{a}a^{-1}\bar{u} = \zeta\bar{u} \quad \text{and} \quad w_\rho\bar{u} = \frac{\bar{a}a^{-1}}{\bar{u}} = \frac{\zeta}{\bar{u}} \quad \text{for some } \zeta \in \mu_{p+1} \subseteq \mathbb{F}_{p^2}^*.$$

So on $\overline{\mathbf{SD}}_A$, the w_ρ reduce to $p+1$ distinct involutions with $2(p+1)$ distinct fixed points (in a $\mu_{2(p+1)}$ orbit). Furthermore, each of these involutions of $\overline{\mathbf{SD}}_A$ lifts to an involution of $\mathbf{SD}_A$.

Another way to think of the fixed points of the automorphisms $\{w_\rho\}$ is that they correspond to elliptic curves whose formal groups have complex multiplication by the ring of integers in a ramified quadratic extension of $\mathbb{Q}_p$ (see Proposition 4.9 below). This point of view becomes crucial when we determine the field of definition of our stable model, because it ties our construction to the arithmetic theory of CM elliptic curves. To this end, we make the following definition.

Definition 4.8. For K a complete subfield of $\mathbb{C}_p$, an elliptic curve E/K has fake CM if $\mathrm{End}_K \hat{E} \neq \mathbb{Z}_p$ and potential fake CM if $\mathrm{End}_{\mathbb{C}_p} \hat{E} \neq \mathbb{Z}_p$.

Proposition 4.9. *Let (E, C) be any point of $\mathbf{SD}_A$. Then the following statements are equivalent.*

(i) *(E, C) is fixed by w_ρ for some $\rho \in B'$.*

(ii) *(E, C) is fixed by w_ρ for some $\rho \in B^*$.*

(iii) *E has potential fake CM by $\mathbb{Z}_p[\pi]$, where $\pi \in \mathrm{End}(\hat{E})$ and $C = \ker \pi$.*

Proof. We will show that (ii) is equivalent to both (i) and (iii) with a Woods Hole argument. So before we begin we must reinterpret condition (ii) in the language of Theorem 4.1. Let (F, α, C) be a Woods Hole representation of (E, C), and let $\iota_C : F \to F/C$ be the natural map. Then (E, C) is a fixed point of w_ρ if and only if there is an isomorphism $\sigma : F/C \to F$ that makes the following diagram commute.

$$
\begin{array}{ccccccc}
\overline{F} & \xrightarrow{\bar{\iota}_C} & \overline{F/C} & \xrightarrow{\mathrm{id}} & \overline{F/C} & \xrightarrow{\bar{\sigma}} & \overline{F} \\
{\scriptstyle\alpha}\downarrow & & {\scriptstyle\beta}\downarrow & & {\scriptstyle\rho\circ\beta}\downarrow & & {\scriptstyle\alpha}\downarrow \\
\hat{A} & \xrightarrow{j} & \hat{A} & \xrightarrow{\rho} & \hat{A} & \xrightarrow{\mathrm{id}} & \hat{A}
\end{array}
$$

Note that the first commuting square represents the isogeny (of elliptic curves) $E \to E/C$. The pair $(F/C, \rho \circ \beta)$ then corresponds to the elliptic curve $\rho(E/C)$.

Now, to show (iii) implies (ii), suppose first that we are given $\pi \in \mathrm{End}(F)$ with $\ker(\pi) = C$. Then π must factor as $\sigma \circ \iota_C$ for some isomorphism $\sigma : F/C \to F$, and we may take $\rho = \alpha \circ \bar{\sigma} \circ \beta^{-1} \in B^*$ in the diagram above. Conversely, if (E, C) is a fixed point of w_ρ for some $\rho = a + bi + cj + dk \in B^*$, and hence we

have a commutative diagram as above, $\text{End}(F)$ must contain both $\pi := \sigma \circ \iota_C$ and $\pi_0 := \pi + pc$. Using the diagram to compute inside $\text{End}(\hat{A})$, we then have

$$\alpha \circ \bar{\pi}_0 \circ \alpha^{-1} = \rho j - cj^2 = (\rho - cj)j.$$

Note that $\rho - cj \in B'$. So $\pi_0^2 \in p\mathbb{Z}_p^*$, which means that $\mathbb{Z}_p[\pi_0] = \mathbb{Z}_p[\pi]$ is already the maximal order in a ramified quadratic extension of $\mathbb{Q}_p$ (and hence all of $\text{End}(F)$). Thus we have shown that (ii) implies (iii). We also get for free, however, that (ii) implies (i), since (E, C) is now also fixed by w_{ρ_0}, where $\rho_0 := \rho - cj \in B'$. □

Corollary 4.10. *If $(E, C) \in \mathbf{SD}_A$ is fixed by w_{ρ_0} for some $\rho_0 \in B'$, then w_ρ fixes (E, C) precisely when $\rho = a\rho_0 + bj$ for $a \in \mathbb{Z}_p^*$ and $b \in \mathbb{Z}_p$.*

Remark 4.11. With notation as above, suppose that $(E, C) \in \mathbf{SD}_A$ is fixed by w_ρ and H is one of the noncanonical subgroups of E of order p (so $(E, H) \in \mathbf{C}_A$). Since $\pi_0^2 \in p\mathbb{Z}_p^*$ and $\ker(\pi_0) = C$, we must have $\pi_0(H) = C$. This implies that $a + b\pi_0 \in (\mathbb{Z}_p[\pi_0])^* \cong \text{Aut}(F)$ fixes the noncanonical subgroups of order p when $p \mid b$, and transitively permutes them otherwise.

Remark 4.12. We showed in [CM 2006, Remark 3.11] that the points that satisfy the conditions of Proposition 4.9 are precisely the canonical liftings of $\hat{A}$ in the sense of [Gross 1986, 2.1], where K is one of the ramified quadratic extensions of $\mathbb{Q}_p$ and $\hat{A}$ is given the structure of a formal $\mathbb{O}_K$-module.

5. Stable reduction of $X_0(p^2)$

At this point we have done enough groundwork to prove a rigid analytic reformulation of Edixhoven's result [1990, Theorem 2.1.2] on the stable reduction of $X_0(p^2)$. Most of the work is in computing the reduction of $\mathbf{Y}_A$, the underlying affinoid of $W_A(p^2)$. This is done by first embedding $\mathbf{Y}_A$ into the product of two circles (specifically $\mathbf{TS}_A \times \mathbf{TS}_A$) and then applying the explicit formula of Theorem 3.5. After that, we use results from Section 2 to show that the wide opens in

$$\{W_{20}, W_{11}^+, W_{11}^-, W_{02}\} \cup \{W_A(p^2) : A \text{ is supersingular}\}$$

intersect properly and comprise a stable covering of $X_0(p^2)$.

Lemma 5.1. *Let $\mathbf{Y}_A = \pi_v^{-1}(\mathbf{TS}_A)$. If $A/\mathbb{F}_p$, $\mathbf{Y}_A$ is naturally isomorphic to*

$$S := \{(x, y) \in \mathbf{TS}_A \times \mathbf{TS}_A \mid x \neq y, \ \pi_f(x) = \pi_f(y)\}.$$

Proof. If $(x, y) \in S$, then $x = (E, C_1)$ and $y = (E, C_2)$ for E some too supersingular curve, and C_1 and C_2 are two distinct subgroups of order p. So we can define a map $\psi : S \to \mathbf{Y}_A$ by taking (x, y) to $(E/C_1, p^{-1}C_2/C_1)$. It is immediate that this takes values in $\mathbf{Y}_A$ since $\pi_v \circ \psi(x, y)$ is then

$$(E/E[p], p^{-1}C_2/E[p]) \cong (E, C_2).$$

Also, we can define a map going the other way, say ϕ, by taking $(E, C) \in \mathbf{Y}_A$ to the pair $(x, y) \in S$ with $x = (E/pC, E[p]/pC)$ and $y = (E/pC, C/pC)$. This takes values in S precisely because $\pi_v(E, C) = (E/pC, C/pC) \in \mathbf{TS}_A$, and it is straightforward to check that $\psi \circ \phi$ and $\phi \circ \psi$ are the respective identities. $\square$

Proposition 5.2. *Let A be as in Theorem 3.5. Then if K is any extension of $W(\mathbb{F}_{p^2}) \otimes \mathbb{Q}_p$ such that $(p+1) \mid e(K)$, $\overline{\mathbf{Y}}_A := \overline{(\mathbf{Y}_A)}_K$ is a smooth, affine curve of genus $(p-1)/2$ with 4 points at infinity (equation given below).*

Proof. Let x and y be parameters on $\mathbf{TS}_A$ that are specializations of the parameter t on $W_A(p)$ from Theorem 3.5. Then by Lemma 5.1, $\mathbf{Y}_A$ can be described by the equation

$$F(x) + G(\kappa/x) = F(y) + G(\kappa/y),$$

where $v(x) = v(y) = p/(p+1)$. Now choose any $\alpha \in K$ with $v(\alpha) = 1/(p+1)$ and substitute $u = \alpha^p/x$ and $v = \alpha^p/y$ into the above equation for $\mathbf{Y}_A$ (so that $v(u) = v(v) = 0$). Dividing through by α^p, we obtain an equation for $\mathbf{Y}_A$ in u and v that has integral coefficients and satisfies the congruence

$$u^{-1} - v^{-1} \equiv (v^p - u^p)(\kappa/\alpha^{p+1})^p \pmod{\alpha}.$$

Now let $b = (\kappa/\alpha^{p+1})^p$ (a unit), and we obtain

$$1 \equiv buv(v - u)^{p-1} \pmod{\alpha}$$

as an equation for $\overline{\mathbf{Y}}_A$.

Strictly speaking, the above curve has three infinite points, with projective coordinates $(0:1:0)$, $(1:0:0)$, and $(1:1:0)$. However, while the first two are nonsingular, the third splits into two points in the normalization. The genus can easily be computed by applying Riemann–Hurwitz to the equation

$$s^{p+1} = \frac{b}{4}(r^2 - 1),$$

where $s = 1/(v - u)$ and $r = (v + u)/(v - u)$. $\square$

Theorem 5.3. *Let $p \geq 13$ be a prime, and K an extension of $W(\mathbb{F}_{p^{24}}) \otimes \mathbb{Q}_p(\mu_p)$ with $(p+1) \mid e(K)$. The following is a semistable covering of $X_0(p^2)$ over K:*

$$\mathcal{C}_0(p^2) := \{W_{20}, W_{11}^+, W_{11}^-, W_{02}\} \cup \{W_A(p^2) : A \text{ is supersingular}\}.$$

The affinoids $\mathbf{X}_{ab}^{\pm}$ and $\mathbf{Y}_A$ are minimal underlying affinoids in $W_{ab}^{\pm}$ and $W_A(p^2)$.

Proof. The wide opens $W_{ab}^{\pm}$, which cover the ordinary locus, are disjoint from each other, and we have

$$\mathbf{Y}_A = W_A(p^2) \setminus \bigcup W_{ab}^{\pm}.$$

By Proposition 3.6, all four ordinary affinoids have good reduction over K. Therefore, it suffices to show that each $\mathbf{Y}_A$ has good reduction, and that $W_A(p^2) \cap W_{ab}^{\pm}$ is always an annulus.

First we demonstrate that the wide open intersections are annuli over K (where we still assume $(p+1) \mid e$). In the case of W_{20} this is immediate, as $W_{20} \cap W_A(p^2)$ maps isomorphically onto the annulus

$$x_A^{-1}(A(p^{-i(A)/(p+1)}, 1)) \subseteq W_A(p)$$

over K, via π_f. Similarly, $W_{11}^{\pm} \cap W_A(p^2)$ maps onto the same annulus via π_f, but with degree $(p-1)/2$. Then Theorem 2.6 implies that this too is an annulus over K. Finally, $W_{02} \cap W_A(p^2)$ must be an annulus since it is isomorphic to the region $W_{20} \cap W_{A^\sigma}(p^2)$, by the Atkin–Lehner involution w_2.

Next we consider the reductions of the affinoids $\mathbf{Y}_A$. Since $p \geq 13$, Theorem B.1 guarantees us a supersingular elliptic curve A_0 for which Proposition 5.2 directly applies. Then for any other supersingular curve A we use Proposition 4.2. In particular, we choose a surjection ψ_l that maps $W_{A_0}(p^2)$ onto $W_A(p^2)$ with degree $i(A)$. If $i(A) = 1$, the two regions are isomorphic and we are done. In any case, ψ_l necessarily takes $\mathbf{Y}_{A_0}$ to $\mathbf{Y}_A$ and is étale. Therefore $\overline{\mathbf{Y}}_A$ is isomorphic to the quotient of $\overline{\mathbf{Y}}_{A_0}$ by an automorphism of degree $i(A)$ (which fixes the four infinite points). Hence $\mathbf{Y}_A$ has good reduction and we are done. $\square$

Corollary 5.4. *For any supersingular curve A, the reduction of $\mathbf{Y}_A$ must have* (*with the correct choice of parameters*) *the equation*

$$y^{(p+1)/i(A)} = x^2 - 1,$$

and genus $(p+1)/(2i(A)) - 1$.

Proof. After a change of coordinates, the reduction of $\mathbf{Y}_{A_0}$ has the equation $y^{p+1} = x^2 - 1$, with two of the four infinite points moved to $(\pm 1, 0)$ and two still at infinity. Now, any automorphism of order $i(A)$ that acts on this curve and fixes these four points must fix x and take y to ζy, where $\zeta^{i(a)} = 1$. $\square$

Remark 5.5. Let K be as in Theorem 5.3, with $e_p(K) = (p^2 - 1)/2$. By computing the widths of the annuli in the stable covering (see Section 9A for more details), one finds intersection multiplicities of $i(A)$ where $\mathbf{X}_{11}^{\pm}$ meets $\mathbf{Y}_A$ and of $i(A) \cdot (p-1)/2$ where $\mathbf{X}_{20}$ and $\mathbf{X}_{02}$ meet $\mathbf{Y}_A$.

The following implies [Coleman 2005, Theorem 3.1].

Corollary 5.6. *The point (E, C) is not in $S := W_{20} \cup W_{11}^+ \cup W_{11}^- \cup W_{02}$ if and only if $pC = K(E)$ and $E[p]/pC = K(E/C)$, or equivalently $K(E/pC) = 0$.*

Proof. (E, C) is not in S if and only if it is in some $\mathbf{Y}_A$, which by definition means that E/pC has trivial canonical subgroup. This is equivalent to $pC = K(E)$ and $E[p]/pC = K(E/C)$ by [Buzzard 2003, Theorem 3.3(vi)]. □

Corollary 5.7. *The Hecke correspondence T_ℓ takes a divisor supported on S to a divisor supported on S if and only if $\ell \neq p$.*

Proof. This follows from the fact that

$$T_\ell(E, C) = \sum_{\substack{\deg \alpha = \ell \\ |\alpha C| = |C|}} (\alpha E, \alpha C). \qquad \square$$

Remark 5.8. Using the fact that $X(p) \cong X_0(p^2) \times_{X_0(p)} X_1(p)$, Jared Weinstein and the second author have used the results of this section to determine a stable model of $X(p)$.

6. Outline of $X_0(p^3)$ analysis

At this point we would like to construct a stable covering for $X_0(p^3)$ in much the same way as was just done for $X_0(p^2)$. By analogy, the natural starting point would be the covering consisting of

$$\{W_{30}, W_{21}^+, W_{21}^-, W_{12}^+, W_{12}^-, W_{03}\} \cup \{W_A(p^3) : A \text{ is supersingular}\}.$$

This is not stable, however, because $W_A(p^3)$ is not a basic wide open. This can actually be seen immediately from the fact that each $W_A(p^3)$ at least contains the affinoids $\mathbf{E}_{1A} := \pi_f^{-1}(\mathbf{Y}_A)$ and $\mathbf{E}_{2A} := \pi_v^{-1}(\mathbf{Y}_{A^\sigma})$ (which are nontrivial from Section 5). So our covering for $X_0(p^3)$ must at least be refined to take these regions into account. In fact, things are much more complicated.

For simplicity, suppose that $A/\mathbb{F}_p$ with $j(A) \neq 0, 1728$ (other $W_A(p^3)$ can be handled by Proposition 4.2). Since π_{11} maps $W_A(p^3)$ onto the width 1 annulus, $W_A(p)$, this gives us a convenient way to keep track of where various subspaces are in relation to each other. For example, it follows from Section 5 that the above affinoids, $\mathbf{E}_{1A}$ and $\mathbf{E}_{2A}$, lie over the circles described by $v(x_A) = p/(p+1)$ and $v(x_A) = 1/(p+1)$ respectively (with parameter x_A as in Section 3). The former is the too-supersingular circle, and the latter is what was called the nearly too-supersingular circle in [Coleman 2005, §3]. Lying in between these two circles is the Atkin–Lehner circle, $\mathbf{SD}_A$, where $v(x_A) = 1/2$. So lying "in between" $\mathbf{E}_{1,A}$ and $\mathbf{E}_{2,A}$ in some sense is the affinoid $\mathbf{Z}_A := \pi_{11}^{-1}(\mathbf{SD}_A)$. It turns out that this affinoid is where all of the new complication arises at the p^3 level. We now give a brief summary of the analysis of $\mathbf{Z}_A$ that will follow in Sections 7 and 8.

Much of our analysis of $\mathbf{Z}_A$ is explicit (see Section 8), and is based on an embedding into the product of two circles as in Lemma 5.1. More specifically, let

$\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$ be as in Section 3. Then $\mathbf{Z}_A$ can be identified with

$$S := \{(x, y) \in \mathbf{C}_A \times \mathbf{C}_A \mid \tau_f(x) = w_1 \circ \tau_f(y)\}.$$

Since $\pi_f \circ \tau_f = \pi_f$, this identification along with de Shalit's result (Theorem 3.5) gives us a way to explicitly compute the reduction of $\mathbf{Z}_A$ as

$$X^{p+1} + X^{-(p+1)} = Z^p.$$

So $\overline{\mathbf{Z}}_A$ has $2(p+1)$ cuspidal singular points, and its normalization is a copy of the affine line whose completion is what we will call a "bridging component". Basically, we want to show that the $2(p+1)$ singular residue classes of $\mathbf{Z}_A$ are basic wide open subspaces, with underlying affinoids that reduce to $y^2 = x^p - x$.

To motivate and explain this, consider the identity $\pi_{11} \circ w_3 = w_1 \circ \pi_{11}$, relating the Atkin–Lehner involutions on $X_0(p^3)$ and $X_0(p)$. It follows immediately that w_3 preserves $\mathbf{Z}_A$, as well as $\widetilde{\mathscr{D}} := \pi_{11}^{-1}(\mathscr{D})$, where $\mathscr{D}$ is either of the residue disks of $\mathbf{SD}_A$ preserved by w_1. Furthermore, a moduli-theoretic argument shows that w_3 has $2p$ fixed points that lie $p : 1$ over the w_1 fixed points in $\mathbf{SD}_A$. So $\widetilde{\mathscr{D}} \subseteq \mathbf{Z}_A$ is a wide open with one end upon which the involution w_3 acts with p fixed points. We show that $\widetilde{\mathscr{D}}$ is in fact isomorphic to the complement of an affinoid disk near infinity in a hyperelliptic curve that reduces to $y^2 = x^p - x$ (w_3 is the hyperelliptic involution). Such an argument, however, would only account for two of the singular residue classes of $\mathbf{Z}_A$. To handle all of them, we use the action of $B^* = \mathrm{Aut}(\hat{A})$ to generalize the pair (w_1, w_3) to a pair $(w_\rho, \widetilde{w}_\rho)$, as was done in Proposition 4.6. Thus we are able to handle all $2(p+1)$ residue classes because of Remark 4.7.

Once we have actually constructed all of the nontrivial components in the stable reduction of $X_0(p^3)$, the argument is reduced to showing that nothing else interesting can happen. We do this in Section 9, with a total genus calculation playing a key role. Again we first use the fact that all supersingular regions are (nearly) isomorphic along with the result of Appendix B, so that calculations only need to be done for a supersingular curve with $A/\mathbb{F}_p$ and $j(A) \neq 0, 1728$. The remaining cases of $p \leq 11$ were handled explicitly in [CM 2006, §6], which we hope makes our construction more understandable, and which completes Theorem 9.2.

7. The bridging component

Fix a supersingular elliptic curve $A/\mathbb{F}_p$ with $j(A) \neq 0, 1728$. In this section we begin our analysis of the affinoid $\mathbf{Z}_A := \pi_{11}^{-1}(\mathbf{SD}_A) \subseteq W_A(p^3)$. In particular, we show by a moduli-theoretic argument that $\mathbf{Z}_A$ can be embedded into $\mathbf{C}_A \times \mathbf{C}_A$. Using the embedding, we then construct a family of involutions on $\mathbf{Z}_A$. These involutions are compatible (with respect to π_{11}) with the involutions of $\mathbf{SD}_A$ that were introduced in Proposition 4.6.

Proposition 7.1. *Let* $\mathbf{C}_A$ *and* $\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$ *be as in Section 3. There is a natural isomorphism* ψ *from*

$$S := \{(x, y) \in \mathbf{C}_A \times \mathbf{C}_A \mid \tau_f(x) = w_1 \circ \tau_f(y)\}$$

to $\mathbf{Z}_A$, *such that* $w_3(\psi(x, y)) = \psi(y, x)$ *and* $\pi_{11}(\psi(x, y)) = \tau_f(x)$.

Proof. Suppose $(x, y) \in S$. Then there exists an $(E, C) \in \mathbf{SD}_A$ such that $x = (E, H)$ for some $H \neq C$. The p noncanonical subgroups of E/C are precisely the subgroups D/C, where $D \subseteq E$ is a cyclic subgroup of order p^2 with $pD = C$ [Buzzard 2003, 3.3]. Therefore, since $\tau_f(x) = w_1(\tau_f(y)) = (E, C)$, there is a unique D such that $y = (E/C, D/C)$. Hence we can define a map $\psi : S \to W_A(p^3)$ by

$$\psi(x, y) = (E/H, (p^{-1}D)/H).$$

Note that $(p^{-1}D)/H$, and hence ψ, is well defined since $pD = C$ and H span $E[p]$. The key fact to check is that $\psi(x, y)$ lies in $\mathbf{Z}_A$, that is, $\pi_{11}(\psi(x, y)) \in \mathbf{SD}_A$.

$$\begin{aligned} \pi_{11}(E/H, (p^{-1}D)/H) &= (E/\langle H, pD \rangle, D/\langle H, pD \rangle) \\ &= (E/E[p], D/E[p]) \\ &\equiv (E, pD) = (E, C) \in \mathbf{SD}_A. \end{aligned}$$

This calculation shows that $\psi(x, y) \in \mathbf{Z}_A$, that $\pi_{11}(\psi(x, y)) = \tau_f(x)$, and more. Once a point $(E, C) \in \mathbf{SD}_A$ is fixed, there are p independent choices for both H and D. Therefore we have produced p^2 points of $\mathbf{Z}_A$ that are in the image of ψ and in the π_{11}-fiber over that particular $(E, C) \in \mathbf{SD}_A$. Since the total degree of $\pi_{11} : X_0(p^3) \to X_0(p)$ is only p^2, we can conclude that ψ maps *onto* $\mathbf{Z}_A$, and hence is an isomorphism. We now describe its inverse. For an arbitrary $(E, K) \in \mathbf{Z}_A$, let $x(E, K) = (E/p^2K, E[p]/p^2K)$, $y(E, K) = (E/pK, K/pK)$, and $\phi(E, K) = (x(E, K), y(E, K))$. To show that $\phi = \psi^{-1}$, it suffices to check that $\phi \circ \psi$ is the identity on S. We have

$$\begin{aligned} x(E/H, (p^{-1}D)/H) &= (E/\langle H, C \rangle, p^{-1}H/\langle H, C \rangle) \\ &= (E/E[p], p^{-1}H/E[p]) \equiv (E, H) \end{aligned}$$

and

$$\begin{aligned} y(E/H, (p^{-1}D)/H) &= (E/\langle H, D \rangle, p^{-1}D/\langle H, D \rangle) \\ &= (E/\langle E[p], D \rangle, p^{-1}D/\langle E[p], D \rangle) \\ &\equiv (E/pD, D/pD) = (E/C, D/C). \end{aligned}$$

Now that we have determined ψ^{-1}, we can verify the claim regarding w_3 by applying $\psi^{-1} \circ w_3 \circ \psi$ to the pair (x, y), where $x = (E, H)$ and $y = (E/C, D/C)$.

We have

$$
\begin{aligned}
w_3 \circ \psi(x, y) &= w_3(E/H, (p^{-1}D)/H) \\
&= (E/\langle H, p^{-1}D \rangle, p^{-3}H/\langle H, p^{-1}D \rangle) \\
&= (E/\langle E[p], p^{-1}D \rangle, p^{-3}H/\langle E[p], p^{-1}D \rangle) \\
&\equiv (E/D, p^{-2}H/D),
\end{aligned}
$$

$$
\begin{aligned}
x(E/D, p^{-2}H/D) &= (E/\langle D, H \rangle, p^{-1}D/\langle D, H \rangle) \\
&= (E/\langle E[p], D \rangle, p^{-1}D/\langle E[p], D \rangle) \\
&\equiv (E/pD, D/pD) = (E/C, D/C) = y,
\end{aligned}
$$

$$
\begin{aligned}
y(E/D, p^{-2}H/D) &= (E/\langle D, p^{-1}H \rangle, p^{-2}H/\langle D, p^{-1}H \rangle) \\
&= (E/E[p^2], p^{-2}H/E[p^2]) \equiv (E, H) = x. \qquad \square
\end{aligned}
$$

Proposition 7.2. *For each $\rho \in B^*$, we can define an automorphism $\tilde{w}_\rho$ of $\mathbf{Z}_A$ (identified with S) by*

$$
\tilde{w}_\rho(x, y) = (\rho y, \rho' x).
$$

Furthermore, $\tilde{w}_\rho$ is compatible with w_ρ, in the sense that $\pi_{11} \circ \tilde{w}_\rho = w_\rho \circ \pi_{11}$, and is an involution of $\mathbf{Z}_A$ whenever $\rho \in B'$.

Proof. The action of B^* on $W_A(p)$ preserves circles. So at least this defines a map from $\mathbf{C}_A \times \mathbf{C}_A$ to itself. To verify that it preserves the subspace S we need to check that

$$
\tau_f(x) = w_1 \circ \tau_f(y) \Rightarrow \tau_f(\rho y) = w_1 \circ \tau_f(\rho' x).
$$

But τ_f commutes with B^*. So this follows from the identity $\rho w_1 = w_1 \rho'$, which was shown in the proof of [Proposition 4.6](#).

By [Remark 4.4](#), the inverse of $\tilde{w}_\rho$ is given by $\tilde{w}_\xi$ for any $\xi \in B^*$ with $\xi \rho' \in \mathbb{Z}_p^*$. In particular, $\tilde{w}_\rho$ is an involution exactly when $\rho \in B'$. Finally, the compatibility relation follows easily from the fact that $\pi_{11}(x, y) = \tau_f(x)$. $\qquad \square$

Corollary 7.3. *Every fixed point of $\tilde{w}_\rho$ lies (via π_{11}) over a fixed point of w_ρ. If $\mathcal{D}_\rho \subseteq \mathbf{SD}_A$ is one of the two residue disks that are preserved by w_ρ, then $\widetilde{\mathcal{D}}_\rho := \pi_{11}^{-1}(\mathcal{D}_\rho)$ is invariant under $\tilde{w}_\rho$.*

Proof. These are immediate consequences of $\pi_{11} \circ \tilde{w}_\rho = w_\rho \circ \pi_{11}$. $\qquad \square$

Remark 7.4. Let 1_B be the multiplicative identity in B. Then w_{1_B} is $w_1|_{\mathbf{SD}_A}$ and $\tilde{w}_{1_B}$ is $w_3|_{\mathbf{Z}_A}$.

Recall from [Proposition 4.9](#) that the fixed points of w_ρ correspond to pairs (E, C), where E has fake CM by $\mathbb{Z}_p[\pi]$ and $\ker(\pi) = C$. The points of $\mathbf{Z}_A$ that lie over such a fixed point then correspond to pairs $(E/H, p^{-1}D/H)$, where H and D are as in the proof of [Proposition 7.1](#). In particular, $H \subseteq E$ is a noncanonical

subgroup of order p, and $D \subseteq E$ is cyclic of order p^2 such that $pD = C$. Combining these facts with Corollary 7.3 gives us a convenient way to describe (and count) the fixed points of $\widetilde{w}_\rho$.

Proposition 7.5. *Let (E, C) be a fixed point of w_ρ for some $\rho \in B^*$, such that* $\operatorname{End}(\hat{E}) = \mathbb{Z}_p[\pi]$ *with* $\ker(\pi) = C$. *If $\rho \in B'(1 + pjB)$, there are p fixed points of $\widetilde{w}_\rho$ lying over (E, C), specifically those pairs $(E/H, p^{-1}D/H)$ with $\pi(D) = H$. Otherwise, $\widetilde{w}_\rho$ has no fixed points.*

Proof. Fix a Woods Hole triple (F, α, C) corresponding to (E, C). Then E/C is equivalent to some pair $(F/C, \beta)$, such that the diagram from the proof of Proposition 4.9 commutes. Note that an explicit isomorphism from $\rho(E/C)$ to E is then given by the pair (σ, id). To determine the $\widetilde{w}_\rho$ fixed points, it will be useful to similarly describe the isomorphism from $\rho'(E)$ to E/C, which exists by $\rho \circ w_1 = w_1 \circ \rho'$. This can be done by replacing $\rho \circ j$ with $j \circ \rho'$ in the diagram, and repeating the first isogeny, to obtain the following.

$$
\begin{array}{ccccccc}
\overline{F} & \xrightarrow{\bar{\iota}_C} & \overline{F/C} & \xrightarrow{\bar{\sigma}} & \overline{F} & \xrightarrow{\bar{\iota}_C} & \overline{F/C} \\
{\scriptstyle \rho' \circ \alpha}\downarrow & & {\scriptstyle \rho \circ \beta}\downarrow & & {\scriptstyle \alpha}\downarrow & & {\scriptstyle \beta}\downarrow \\
\hat{A} & \xrightarrow{j} & \hat{A} & \xrightarrow{\text{id}} & \hat{A} & \xrightarrow{j} & \hat{A}
\end{array}
$$

Since $j^2 = -p$, this diagram shows that an isomorphism from $\rho'(E)$ to E/C is given by the pair (γ, id), where $\gamma = -p^{-1}\iota_C \circ \sigma \circ \iota_C$.

Now, choose a point lying over (E, C) by taking $x = (E, H) = (F, \alpha, H)$ and $y = (E/C, D/C) = (F/C, \beta, D/C)$. We must determine when

$$\widetilde{w}_\rho(x, y) = (\rho y, \rho' x) = (x, y).$$

Since an isomorphism from $\rho(E/C)$ to E is given by (σ, id), the condition $\rho y = x$ is equivalent to $\sigma(D/C) = H$. Similarly, the condition $\rho' x = y$ is equivalent to $\gamma(H) = D/C$. Putting these in terms of π, the first condition is $\pi(D) = H$ and the second is $\pi(D) = -(\pi^2/p)(H)$. By Remark 4.11, these two conditions are equivalent when $\rho \in B'(1 + pjB)$, and incompatible otherwise. $\qquad\square$

Remark 7.6. If (E, C) is *any* point lying over a fixed point of w_ρ via π_{11}, it is a fake Heegner point in the sense that E has fake CM and $\operatorname{End}(\widehat{E/C})$ is isomorphic to $\operatorname{End}(\hat{E})$. In fact, one can show in this case that $\operatorname{End}(\hat{E}) \cong \mathbb{Z}_p[\lambda]$ for some λ such that $\ker(\lambda) = C$.

8. Explicit analysis

In this section, we use Proposition 7.1 and Theorem 3.5 to explicitly compute the reduction of $\mathbf{Z}_A$ (for $A/\mathbb{F}_p$ and $j(A) \neq 0, 1728$), in much the same way that the

reduction of $\mathbf{Y}_A$ was computed in the proof of Proposition 5.2. We obtain

$$X^{p+1} + X^{-(p+1)} = Z^p.$$

Moreover, the residue classes of $\mathbf{Z}_A$ that have singular reduction on this model are shown to coincide with those regions $\widetilde{\mathscr{D}}_\rho$ that were described in Corollary 7.3. From the previous section we know that $\widetilde{\mathscr{D}}_\rho$ is acted on by the involution $\widetilde{w}_\rho$, with p fixed points. In addition, from the explicit equation for $\overline{\mathbf{Z}}_A$, we are able to deduce that $\widetilde{\mathscr{D}}_\rho$ is a connected wide open with one end, and that $\widetilde{\mathscr{D}}_\rho/\widetilde{w}_\rho$ is a disk. Putting all of this information together (and a little more), we are able to show in Section 8B that $\widetilde{\mathscr{D}}_\rho$ is a basic wide open whose underlying affinoid reduces to $y^2 = x^p - x$.

8A. *Reduction of $\mathbf{Z}_A$.* Recall that Proposition 7.1 identifies $\mathbf{Z}_A$ with the subspace of $\mathbf{C}_A \times \mathbf{C}_A$ defined by $\tau_f(x) = w_1 \circ \tau_f(y)$. From this embedding we can obtain an explicit equation for $\mathbf{Z}_A$, provided we can derive approximation formulas for w_1 on $\mathbf{SD}_A$ and $\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$. Such formulas follow readily from Theorem 3.5. However, while the formula in this theorem is given over $\mathbb{Q}_p \otimes W(\mathbb{F}_{p^2})$, we will ultimately need to work over a finite base extension. This extension can be generated by fixing a square root $\sqrt{\kappa}$ of κ in $\mathbb{C}_p$ (where κ is as in Theorem 3.5) and a $\beta \in \mathbb{C}_p$ satisfying

$$\beta^{p^2} \equiv \kappa \pmod{p^{3/2-1/2p^2}}. \tag{6}$$

Remark 8.1. For example, if $g(x) = x^{p^2} - \sqrt{\kappa}x$, and γ is a root of $g(g(x))/g(x)$, one may take $\beta = \gamma^{2(p^2-1)}$. Then, by Lubin–Tate theory, applied to the Lubin–Tate formal group over $F := \mathbb{Q}_p(\sqrt{\kappa}) \otimes_{\mathbb{Z}_p} W(\mathbb{F}_{p^2})$, with endomorphism $g(x)$, $F(\beta)$ is Galois over F with Galois group $C_p \times C_p$.

Proposition 8.2. *Over $R := \mathbb{Z}_p[\sqrt{\kappa}, \beta] \otimes W(\mathbb{F}_{p^2})$, the reduction of $\mathbf{Z}_A$ has the equation*

$$X^{p+1} + X^{-(p+1)} = Z^p.$$

Hence, over R, its reduction is a reduced, connected, affine curve of genus zero with only one branch through each singular point.

Proof. First we derive an approximation for $\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$ in terms of the parameter t from Theorem 3.5. For any $P_1 \in \mathbf{SD}_A$ and $P_2 \in \mathbf{C}_A$, we note that $P_1 = \tau_f(P_2)$ if and only if $\pi_f(P_1) = \pi_f(P_2)$. Thus, an approximation for τ_f should follow from approximations for π_f on $\mathbf{SD}_A$ and $\mathbf{C}_A$. Now, we know from [Buzzard 2003, 3.3] that $\mathbf{SD}_A$ and $\mathbf{C}_A$ are the circles described by $v(t) = 1/2$ and $v(t) = 1 - 1/2p$. In particular, we must have $v(t(P_1)) = 1/2$ and $v(t(P_2)) = $

$1 - 1/2p$. Therefore, from Theorem 3.5 we can approximate π_f on $\mathbf{SD}_A$ and $\mathbf{C}_A$:

$$s(\pi_f(P_1)) \equiv t(P_1) \qquad (\text{mod } p),$$
$$s(\pi_f(P_2)) \equiv t(P_2) + (\kappa/t(P_2))^p \quad (\text{mod } p).$$

Hence an approximation for $\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$ is given by

$$t(\tau_f(P)) \equiv t(P) + (\kappa/t(P))^p \quad (\text{mod } p).$$

To describe the reduction of $\mathbf{Z}_A$ via Proposition 7.1, we now choose parameters that identify $\mathbf{C}_A$ and $\mathbf{SD}_A$ with the unit circle, $C[1]$. For such a parameter on $\mathbf{SD}_A$ we let $U := t/\sqrt{\kappa}$, and on $\mathbf{C}_A$ we let $X := t/\alpha$, where

$$\alpha = (\beta^{(p^2+1)/2}/\sqrt{\kappa})^{p(2p-1)}$$

(note that $\upsilon(\alpha) = 1 - 1/2p$). In terms of these new parameters, the Atkin–Lehner involution is just given by $w_1^* U = 1/U$. Also, using the defining congruence for β, the approximation formula above for τ_f becomes

$$\tau_f^* U \equiv \alpha X/\sqrt{\kappa} + X^{-p} \quad (\text{mod } \sqrt{p}).$$

Now let Y and V be analogous parameters on copies of $\mathbf{C}_A$ and $\mathbf{SD}_A$, so that the equation $\tau_f(P) = w_1(\tau_f(Q))$ on $\mathbf{C}_A \times \mathbf{C}_A$ (which defined the subspace $S \cong \mathbf{Z}_A$) becomes $\tau_f^* U = 1/\tau_f^* V$. Then on S the parameters X and Y satisfy the congruence relations

$$(\alpha X/\sqrt{\kappa} + X^{-p})(\alpha Y/\sqrt{\kappa} + Y^{-p}) \equiv 1 \quad (\text{mod } \sqrt{p}),$$
$$\alpha X^{p+1}/\sqrt{\kappa} + \alpha Y^{p+1}/\sqrt{\kappa} + 1 \equiv X^p Y^p \quad (\text{mod } \sqrt{p}). \qquad (7)$$

Finally, we define a new parameter Z on $\mathbf{C}_A \times \mathbf{C}_A$ by $XY = \beta^{(p-1)/2} Z + 1$. Then $\mathbf{Z}_A$ is determined over $R \otimes \mathbb{Q}_p$ by $|X| \leq 1$ and $|Z| \leq 1$. The congruence

$$X^{p+1} + X^{-(p+1)} \equiv Z^p \quad (\text{mod } m_R),$$

where m_R is the maximal ideal of R, follows from (7). $\qquad \square$

Proposition 8.3. *The involutions $\widetilde{w}_\rho$ on $\mathbf{Z}_A$ reduce to the involutions on $\overline{\mathbf{Z}}_A$ given by*

$$t_\zeta : (X, Z) \mapsto (\zeta/X, Z),$$

where ζ varies over all $(p+1)$-st roots of unity. The $\widetilde{\mathcal{D}}_\rho^i$ coincide with the singular residue classes of $\mathbf{Z}_A$, which are described by $X^{2p+2} \equiv 1$.

Proof. We use the compatibility relation in the proof of Proposition 7.2, namely $\pi_{11} \circ \widetilde{w}_\rho = w_\rho \circ \pi_{11}$. Recall from Proposition 7.1 that $\pi_{11}(x, y) = \tau_f(x)$ (with

notation consistent with that of the previous proposition). So from the proof of the previous proposition, an explicit formula for π_{11} as a map from $\overline{\mathbf{Z}}_A$ to $\overline{\mathbf{SD}}_A$ is

$$U = \pi_{11}(X, Z) = X^{-p}.$$

Now, we know from Remark 4.7 that on $\overline{\mathbf{SD}}_A$ the involutions w_ρ reduce to those of the form $U \to \zeta/U$ (where ζ is any $(p+1)$-st root of unity). So fix a ρ and corresponding ζ. Choose any point (X_0, Z_0) on $\overline{\mathbf{Z}}_A$, and let $(X_1, Z_1) = \widetilde{w}_\rho(X_0, Z_0)$. We can compute both sides of the compatibility relation above:

$$w_\rho \circ \pi_{11}(X_0, Z_0) = w_\rho(X_0^{-p}) = \zeta X_0^p,$$
$$\pi_{11} \circ \widetilde{w}_\rho(X_0, Z_0) = \pi_{11}(X_1, Z_1) = X_1^{-p}.$$

Since $\zeta = \zeta^{-p}$, we must have $X_1 = \zeta/X_0$ and subsequently $Z_1 = Z_0$. In other words, we have shown that on $\overline{\mathbf{Z}}_A$ we have $\widetilde{w}_\rho(X, Z) = (\zeta/X, Z)$.

Keeping the same notation, the points of $\overline{\mathbf{SD}}_A$ that are fixed by w_ρ are the two described by $U^2 = \zeta$, and by definition $\widetilde{\mathscr{D}}_\rho^1$ and $\widetilde{\mathscr{D}}_\rho^2$ are π_{11}^{-1} of the corresponding residue classes. Since $\pi_{11}: \overline{\mathbf{Z}}_A \to \overline{\mathbf{SD}}_A$ is given by $U = X^{-p}$, this is equivalent to saying that $\widetilde{\mathscr{D}}_\rho^i$ are the classes of $\mathbf{Z}_A$ described by $X^2 \equiv \zeta$. Letting ζ vary over all $(p+1)$-st roots of unity, we obtain all the residue classes described by $X^{2p+2} \equiv 1$, and these are easily verified to be the singular ones. $\qquad\square$

Proposition 8.4. *For any $\rho \in B'$, the residue classes of the affinoid quotient $\mathbf{Z}_A/\widetilde{w}_\rho$, which are the images of the $\widetilde{\mathscr{D}}_\rho^i$, are disks over $\mathbb{Z}_p[\sqrt{\kappa}, \beta] \otimes W(\mathbb{F}_{p^2})$.*

Proof. Let ζ be the $(p+1)$-st root of unity such that $\widetilde{w}_\rho$ reduces to t_ζ on $\overline{\mathbf{Z}}_A$. Let $f_\zeta(x)$ be the unique polynomial of degree $p+1$ such that

$$f_\zeta(X + \zeta/X) = X^{p+1} + X^{-(p+1)}.$$

Then $f_\zeta(x) = z^p$ is an equation for the reduction of $\mathbf{Z}_A/\widetilde{w}_\rho$. Also

$$f_\zeta'(X + \zeta/X) = \frac{X^{2p+2}-1}{X^p(X^2-\zeta)},$$

and the right side doesn't vanish at ϵ if $\epsilon^2 = \zeta$. Thus $f_\zeta'(2\epsilon) \neq 0 \bmod p$, and the two residue classes of $\mathbf{Z}_A/\widetilde{w}_\rho$ described by $X = \pm\epsilon$ are disks. $\qquad\square$

From Theorem 2.29 and Proposition 2.31, we now conclude that (over a suitable field extension) $\widetilde{\mathscr{D}}_\rho^i$ is a connected wide open with one end. Furthermore, using Theorem 2.48 and the fact that there are p branch points in the degree 2 quotient of $\widetilde{\mathscr{D}}_\rho^i$ by $\widetilde{w}_\rho$, we compute the genus of $\widetilde{\mathscr{D}}_\rho^i$ to be $(p-1)/2$. To summarize, we have the following corollary.

Corollary 8.5. *Let L be a complete stable subfield of $\mathbb{C}_p$ containing R, over which the fixed points of $\widetilde{w}_\rho$ are defined. Over L, the rigid spaces $\widetilde{\mathscr{D}}_\rho^i$ for $i = 1$ or 2 are connected wide opens with one end of genus $(p-1)/2$.*

8B. *The new components.* We now show that over a suitable base extension, the $2(p+1)$ residue classes $\widetilde{\mathcal{D}}_\rho^i \subseteq \mathbf{Z}_A$ are basic wide opens, and we compute the reductions of their underlying affinoids. The main idea is to construct an automorphism of order p on each $\widetilde{\mathcal{D}}_\rho^i$ that transitively permutes the p fixed points of the involution $\tilde{w}_\rho$. This induces an automorphism on the quotient $\widetilde{\mathcal{D}}_\rho^i/\tilde{w}_\rho$, a disk by Corollary 8.5, which then must be conjugate to a translation.

First we define automorphisms of order p on the disk $\tau_f^{-1}(\mathcal{D}) \subseteq \mathbf{C}_A$, where $\tau_f : \mathbf{C}_A \to \mathbf{SD}_A$ is as in Section 3, and $\mathcal{D}$ is either of the two residue disks of $\mathbf{SD}_A$ fixed by w_ρ. Recall that points of $\mathbf{SD}_A$ correspond to pairs (E, C) where $h(E) = 1/2$ and C is canonical. One of the key facts that we use in our construction is that over the residue disk $\mathcal{D}$ one can analytically choose a generator up to sign for each of these canonical subgroups. This amounts to choosing a section σ of the forgetful map from $X_1(p)$ to $X_0(p)$ over $\mathcal{D}$, given by

$$\sigma : (E, K_1(E)) \mapsto (E, P_\sigma(E)),$$

where $P_\sigma(E)$ is a pair consisting of a generator of $K_1(E)$ and its inverse. Such a section exists because this map is an étale map of annuli over $\mathbf{SD}_A$ (over any extension of $\mathbb{Q}_p$ whose ramification index is divisible by $2(p-1)$). In fact, the group $(\mathbb{Z}/p\mathbb{Z})^*/\{\pm 1\}$ acts simply transitively on the set of the sections over $\mathcal{D}$. Once σ is chosen, automorphisms of $\tau_f^{-1}(\mathcal{D})/\mathcal{D}$ can be constructed by looking closely at the Weil pairing.

Lemma 8.6. *For any $\zeta \in \mu_p^*$ and σ (as above), we can define an analytic automorphism of $\tau_f^{-1}(\mathcal{D})/\mathcal{D}$ by $S_{\sigma,\zeta}(E, H) = (E, \langle R \rangle)$, where $R \in E[p]$ is chosen so that $e_p(P, R) = \zeta$ and $R - P \in H$ for some $P \in P_\sigma(E)$. Also,*

(i) $S_{\sigma,\zeta}^i(E, H) = (E, \langle R + (i-1)P \rangle)$ *for $i \in \mathbb{Z}$;*

(ii) $f_{\sigma,\zeta} : \mathbb{Z}/p\mathbb{Z} \to \mathrm{Aut}_{an}(\tau_f^{-1}(\mathcal{D})/\mathcal{D})$, *defined by $f_{\sigma,\zeta}(i) = S_{\sigma,\zeta}^i$ for $i \neq 0$ and the identity otherwise, is an injective homomorphism;*

(iii) $S_{a\sigma,\zeta^b} = S_{\sigma,\zeta}^{a^2/b}$ *for any $a, b \in (\mathbb{Z}/p\mathbb{Z})^*$;*

(iv) $S_{\sigma,\zeta}^\tau = S_{\sigma^\tau,\zeta^\tau}$ *for any $\tau \in \mathrm{Aut}_{\mathrm{cont}}(\mathbb{C}_p)$ that preserves $\mathcal{D}$.*

Proof. Fix σ and $\zeta \in \mu_p^*$. For a given pair (E, H) and choice of $P \in P_\sigma(E)$, there is a unique $R \in E[p]$ that satisfies the two conditions. Note also that reversing the sign of P just reverses the sign of R. Since $\langle R \rangle = \langle -R \rangle$ is neither H nor the canonical subgroup, it follows that $S_{\sigma,\zeta}$ is at least a well-defined automorphism of $\tau_f^{-1}(\mathcal{D})/\mathcal{D}$ with no fixed points.

Fix $P \in P_\sigma(E)$. It is easy to verify (i) by induction, and then (ii) follows immediately. To prove (iii), we note that by definition $S_{a\sigma,\zeta^b}(E, H)$ is the pair $(E, \langle Q \rangle)$ where $e_p(aP, Q) = \zeta^b$ and $Q - aP \in H$. A simple Weil pairing calculation shows

that Q is just $aP + (b/a)(R - P)$. So we verify (iii) by checking that

$$e_p(aP + (b/a)(R - P), R + (a^2/b - 1)P) = 1.$$

Finally, property (iv) follows from Galois properties of the Weil pairing and the fact that $\mathscr{D}$ is connected. $\qquad\square$

Proposition 8.7. *Let L be a finite extension of $\mathbb{Q}_p(\sqrt{\kappa}, \beta)$ in $\mathbb{C}_p$, with $\sqrt{\kappa}$ and β as in Equation (6), over which the fixed points of $\widetilde{w}_\rho$ are defined. Then $\widetilde{\mathscr{D}}_\rho^i$ is a basic wide open over a quadratic extension of L, whose underlying affinoid has good reduction, which can be described by $y^2 = x^p - x$.*

Proof. As usual, let $\mathscr{D}$ be either of the residue disks of $\mathbf{SD}_A$ fixed by the involution w_ρ, and let $\widetilde{\mathscr{D}}$ be the wide open lying over $\mathscr{D}$ via π_{11}. Then the embedding of $\mathbf{Z}_A = \pi_{11}^{-1}(\mathbf{SD}_A)$ into $\mathbf{C}_A \times \mathbf{C}_A$ embeds $\widetilde{\mathscr{D}}$ into $\tau_f^{-1}(\mathscr{D}) \times \rho'\tau_f^{-1}(\mathscr{D})$. Therefore, by the previous lemma, we can lift any automorphism $S := S_{\sigma,\zeta}$ on $\tau_f^{-1}(\mathscr{D})$ (for a fixed σ and ζ) to an automorphism $\widetilde{S} := \widetilde{S}_{\sigma,\zeta}$ of $\widetilde{\mathscr{D}}$ by taking

$$\widetilde{S}(x, y) = (S(x), \rho'S(\rho y)).$$

One easily checks that $\widetilde{S}$ also has order p, since $\widetilde{S}^i(x, y) = (S^i(x), \rho'S^i(\rho y))$. Furthermore, $\widetilde{S}$ commutes with $\widetilde{w}_\rho$:

$$\begin{aligned}
\widetilde{S}\widetilde{w}_\rho(x, y) &= \widetilde{S}(\rho y, \rho'x) \\
&= (S(\rho y), \rho'S\rho(\rho'x)) = (S(\rho y), \rho'S(x)), \\
\widetilde{w}_\rho\widetilde{S}(x, y) &= \widetilde{w}_\rho(S(x), \rho'S(\rho y)) \\
&= (\rho\rho'S(\rho y), \rho'S(x)) = (S(\rho y), \rho'S(x)).
\end{aligned}$$

It follows that $\widetilde{S}$ passes to an automorphism of $\widetilde{\mathscr{D}}/\widetilde{w}_\rho$ with order p and no fixed points, which acts transitively on the images of the p fixed points of $\widetilde{w}_\rho$. This is the key idea in the proof of the proposition.

To finish the argument, recall from Corollary 8.5 that $\widetilde{\mathscr{D}}$ is a connected wide open with one end. The involution $\widetilde{w}_\rho$ acts on it with p fixed points, and the quotient space, say $U := \widetilde{\mathscr{D}}/\widetilde{w}_\rho$, is a disk, by Proposition 8.4. It follows that, over a quadratic extension of L, $\widetilde{\mathscr{D}}$ can be described by

$$y_0^2 = (x_0 - \alpha_1) \cdots (x_0 - \alpha_p),$$

where x_0 is a parameter for U and the α_i are the x_0 coordinates of the p fixed points. Without loss of generality, we choose x_0 so that U is identified with the disk, $v(x_0) > 0$. Because $\widetilde{S}$ passes to an automorphism of a disk of order p and no fixed points, it must reduce to a translation, in the sense that there exists an $a \in R_p$ with $v(a) > 0$ such that for all $x_0 \in U$ we have

$$v(\widetilde{S}(x_0) - (x_0 + a)) > v(a).$$

Therefore, after possible reordering, the x_0 coordinates of the fixed points must satisfy

$$\beta_i := \frac{\alpha_i - \alpha_1}{a} \equiv i \pmod{m_p}.$$

So if we make the changes of variables $x = (x_0 - \alpha_1)/a$ and $y = y_0/a^{p/2}$, we identify $\widetilde{\mathscr{D}}$ with the wide open

$$y^2 = (x - \beta_1)(x - \beta_2) \cdots (x - \beta_p), \quad \text{where } v(x) > -v(a),$$

whose minimal underlying affinoid, determined by $v(x) \geq 0$, reduces as claimed. $\qquad \Box$

Remark 8.8. The results of Section 8 were proven for $A/\mathbb{F}_p$ with $j(A) \neq 0, 1728$, but similar results now follow for any other supersingular A', by Proposition 4.2. Since $\mathbf{Z}_{A'}$ is an étale quotient of $\mathbf{Z}_A$ of degree $i(A)$, $\overline{\mathbf{Z}}_{A'}$ is a genus 0 curve with $2(p+1)/i(A')$ singular points, corresponding to basic wide opens that are isomorphic to those described in Proposition 8.7. Note, however, that one might need to replace the field L from Corollary 8.5 by a finite unramified extension in order to define the surjection from $W_A(p^3)$ onto $W_{A'}(p^3)$ and describe the underlying affinoids. In general, the reduction of the bridging component has the equation

$$X^{(p+1)/i(A)} + X^{-(p+1)/i(A)} = Z^p.$$

Lemma 8.9. *The Hecke correspondence T_ℓ takes a divisor supported on $\bigcup_A \mathbf{Z}_A$ to a divisor supported on $\bigcup_A \mathbf{Z}_A$ for all primes $\ell \neq p$.*

Proof. A point (E, C) lies on some $\mathbf{Z}_A$ if and only if C is cyclic of order p^3 and pC/p^2C is self-dual. If $f : E \to F$ is an isogeny such that $\ker(f) \cap C = 0$, the same is true for $(F, f(C))$. $\qquad \Box$

Remark 8.10. The analogous statement, for the union of all of the underlying affinoids in Proposition 8.7 (corresponding to new components), follows from the results of [CM 2006, §8].

9. Stable reduction of $X_0(p^3)$

In this section we give the stable covering of $X_0(p^3)$. In particular, we give a covering by *basic* wide opens, whose intersections are annuli as in Proposition 2.34. We already defined some of these wide opens, namely the $W_{ab}^{\pm}$, in Section 3. They cover the ordinary locus, and will be shown to be basic with the $\mathbf{X}_{ab}^{\pm}$ as underlying affinoids. From our analysis of $\mathbf{Z}_A$ in Section 8, we now know that $W_A(p^3)$ is not a basic wide open. So our next priority is to specify some new wide open subspaces that cover each $W_A(p^3)$ and that can ultimately be shown to be basic.

Now let A be any supersingular elliptic curve mod p (no restriction). Identify $W_{A^\sigma}(p)$, where σ is the Frobenius automorphism, with the annulus $A(p^{-i(A)}, 1)$, as explained in Section 3A. Then we can define three subspaces of $W_A(p^3)$:

$$V_1(A) := \pi_{11}^{-1} A(p^{-i(A)}, p^{-i(A)/2}),$$
$$V_2(A) := \pi_{11}^{-1} A(p^{-i(A)/2}, 1),$$
$$U(A) := \pi_{11}^{-1} A(p^{-pi(A)/(p+1)}, p^{-i(A)/(p+1)}).$$

First we want to show that these subspaces are wide opens (over $\mathbb{C}_p$). Since $V_1(A)$ is a union of residue classes of the affinoid

$$\pi_{11}^{-1}(X_{01} \cup A(p^{-i(A)}, p^{-i(A)/2}]),$$

and since it is connected, it is in fact one residue class and therefore a wide open, by Theorem 2.29. The same argument applies to $V_2(A)$ and $U(A)$, the latter being a residue class of

$$\pi_{11}^{-1} A[p^{-pi(A)/(p+1)}, p^{-i(A)/(p+1)}].$$

Remark 9.1. The points of $A(p^{-pi(A)/(p+1)}, p^{-i(A)/(p+1)})$ are pairs (E, C), where C is the canonical subgroup of E and $E[p]/C$ is the canonical subgroup of E/C.

Two of these supersingular wide opens will in fact be shown to be basic. More specifically, $V_1(A)$ is a wide open neighborhood of the affinoid

$$\mathbf{E}_{1A} := \pi_{11}^{-1} C[p^{-pi(A)/(p+1)}],$$

which will be shown to be an underlying affinoid with good reduction. Points of $\mathbf{E}_{1A}$ are pairs (E, C), such that $E/p^2 C$ is too supersingular. Alternatively, $\mathbf{E}_{1A}$ can be described as $\pi_f^{-1} \mathbf{Y}_A$, which is a key point because it implies that $\mathbf{E}_{1A}$ is nontrivial. Similarly, $V_2(A)$ is a neighborhood of

$$\mathbf{E}_{2A} := \pi_{11}^{-1} C[p^{-i(A)/(p+1)}].$$

Points of $\mathbf{E}_{2A}$ are pairs (E, C) with E/pC too supersingular, and $\mathbf{E}_{2A}$ maps onto $\mathbf{Y}_{A^\sigma}$ via π_v. $U(A)$ is not basic, because its underlying affinoid $\mathbf{Z}_A$ has the $\widetilde{\mathscr{D}}_\rho^i$ as (bad) residue classes. However, the $\widetilde{\mathscr{D}}_\rho^i$ were shown to be basic in Proposition 8.7. So this problem can essentially be solved by removing the underlying affinoids of the $\widetilde{\mathscr{D}}_\rho^i$ from $U(A)$ (obtaining a basic wide open) and then including the $\widetilde{\mathscr{D}}_\rho^i$ in the overall covering. To be more precise, let $\mathscr{S}(A)$ denote the set of singular residue classes of $\mathbf{Z}_A$, and for each $S \in \mathscr{S}(A)$ let $\mathbf{X}_S$ be the underlying affinoid of S. Let $\hat{U}(A)$ denote the wide open given by

$$\hat{U}(A) := U(A) \setminus \bigcup_{S \in \mathscr{S}(A)} \mathbf{X}_S.$$

Theorem 9.2. *Let $p \geq 13$ be a prime. The covering $\mathscr{C}_0(p^3)$ of $X_0(p^3)$, which is made up of*

$$\{W_{ab}^{\pm} \mid a, b \geq 0, a + b = 3\}$$

and the union over all supersingular curves A of

$$\{V_1(A), V_2(A), \hat{U}(A)\} \cup \mathscr{S}(A),$$

is stable (over $\mathbb{C}_p$).

Proof. We know that the elements of $\mathscr{C}_0(p^3)$ are wide opens, and that $(S, \mathbf{X}_S)$ is a basic wide open pair for each $S \in \mathscr{S}(A)$. It is also easy to verify that condition (ii) of Proposition 2.34 holds, by simply listing for each wide open the other members of the covering that intersect it nontrivially. In particular, the $W_{ab}^{\pm}$ are disjoint from each other, and each $W_{ab}^{\pm}$ intersects $W_A(p^3)$ only at $V_2(A)$ when $a > b$, and only at $V_1(A)$ otherwise. Similarly, while $V_1(A)$, $V_2(A)$, and the residue classes $S \in \mathscr{S}(A)$ are pairwise disjoint, each of these wide opens intersects $\hat{U}(A)$ nontrivially. This completely describes all adjacency relations of wide opens in the covering, and it follows immediately that every triple intersection is empty. The bulk of what we still have to show is that whenever two wide opens in the cover do intersect, the intersection is the disjoint union of annuli. Then we have to show that each wide open is basic, with an underlying affinoid that has good reduction.

We start by showing that

$$U_{ab}^{\pm}(A) := W_{ab}^{\pm} \cap W_A(p^3)$$

is a wide open annulus in all cases. For U_{30} and $U_{21}^{\pm}$ it suffices to consider the map π_{02} from $X_0(p^3)$ to $X_0(p)$. The restriction of π_{02} to U_{30} is an isomorphism onto the annulus

$$B := A(p^{-i(A)/(p(p+1))}, 1) \cong A(1, p^{i(A)/(p(p+1))})$$

(considered as a subspace of $W_A(p)$, which has been identified with $A(p^{-i(A)}, 1)$ as in Section 3A). So U_{30} is an annulus right away. U_{21}^+ and U_{21}^- also map onto B via π_{02}, but each with degree $(p-1)/2$. To see that $U_{21}^{\pm}$ is at least connected, we look at how π_{02} reduces when restricted to a map between the affinoid regions $\mathbf{X}_{21}^{\pm}$ and $\mathbf{X}_{10}$. The latter is an isomorphic copy of the ordinary locus of $X(1)$, and by [Coleman 2005, p. 5] the reduction of $\mathbf{X}_{21}^{\pm}$ is isomorphic to the ordinary locus of $\mathrm{Ig}(p)$. Furthermore, by these identifications, π_{02} reduces to the forgetful map from $\mathrm{Ig}(p)$ to $X(1)$, which is totally ramified at the supersingular points. This implies that one of the ends of B totally ramifies in the restriction of π_{02} to $U_{21}^{\pm}$. Hence $U_{21}^{\pm}$ must be connected. Now it follows directly from Theorem 2.6 that $U_{21}^{\pm}$ is an annulus. Similar arguments can be made for $U_{12}^{\pm}$ and U_{03} using π_{20}. Alternatively one can use the fact that the Atkin–Lehner involution, w_3, switches W_{ab} with W_{ba}

and $W_A(p^3)$ with $W_{A^\sigma}(p^3)$. Note that from this argument we also deduce that each $(W_{ab}^\pm, \mathbf{X}_{ab}^\pm)$ is a basic wide open pair.

Among the remaining intersections of wide opens in the covering, we also have $S \cap \hat{U}(A)$ for each $S \in \mathscr{S}(A)$. It is immediate, however, that this is an annulus, since S is a basic wide open with one end, and by definition $S \cap \hat{U}(A)$ is the complement in S of its underlying affinoid $\mathbf{X}_S$. So all that remains to be proven is that $V_i(A) \cap \hat{U}(A)$ is the disjoint union of annuli (in fact, one annulus), and that $(V_1(A), \mathbf{E}_{1A})$, $(V_2(A), \mathbf{E}_{2A})$, and $(\hat{U}(A), \mathbf{Z}_A)$ are basic wide open pairs. This essentially comes down to a genus computation and Proposition 2.34.

First shrink each $\hat{U}(A)$ to a basic wide open neighborhood $\hat{U}'(A)$ of $\mathbf{Z}_A$, and call the resulting covering $\mathscr{C}_1(p^3)$. Although we do not know that $\mathscr{C}_1$ is semistable (and in fact it isn't), Proposition 2.34 can still be applied as the wide opens in the covering intersect properly in the disjoint union of annuli. Moreover, we know that the intersection of $\hat{U}'(A)$ with $V_i(A)$ is just one annulus, because $\overline{\mathbf{Z}}_A$ has only two points at infinity (see Theorem 2.29). So the Betti number of the graph associated to $\mathscr{C}_1(p^3)$ is exactly $5(s_p - 1)$, where s_p is the number of supersingular j-invariants (mod p). To apply Proposition 2.34, we need to know the genera of the wide opens in $\mathscr{C}_1(p^3)$. The genus of W_{ab} is 0 when $ab = 0$ and $g(\mathrm{Ig}(p))$ otherwise, by [Coleman 2005, §1]. The genus of $\hat{U}'(A)$ is 0 and the genus of each $S \in \mathscr{S}(A)$ is $(p-1)/2$, by Proposition 8.2 and Corollary 8.5. The only genera that aren't immediately available are those of $V_1(A)$ and $V_2(A)$. We can, however, provide a lower bound for these genera. Recall that $\mathbf{E}_{1A}$ maps onto $\mathbf{Y}_A$ via π_f, and $\mathbf{E}_{2A}$ maps onto $\mathbf{Y}_{A^\sigma}$ via π_v. So by a Riemann–Hurwitz argument we know that $g(V_i(A)) \geq g(\overline{\mathbf{Y}}_A)$ (which we know from Corollary 5.4).

We now compute a lower bound for the genus of $X_0(p^3)$, using the above and Proposition 2.34. For brevity we only discuss the case $p = 12k + 5$. Then $s_p = k + 1$ and from [Igusa 1968, p. 103] we have $g(\mathrm{Ig}(p)) = 3k^2 - k$. There are k supersingular regions with $j(A) \neq 0, 1728$, each of which contributes two wide opens $V_1(A)$ and $V_2(A)$ of genus at least $g(\mathbf{Y}_A) = 6k + 2$, and $24k + 12$ residue classes $S \in \mathscr{S}(A)$ with genus $6k + 2$. In addition, we have one supersingular region corresponding to $j(A) = 0$ that contributes two wide opens $V_1(A)$ and $V_2(A)$ of genus at least $g(\mathbf{Y}_A) = 2k$, and $8k + 4$ residue classes $S \in \mathscr{S}(A)$ of genus $6k + 2$. Summing up the Betti number and genera as in Proposition 2.34, we have

$$g(X_0(p^3)) \leq 5k + 4(3k^2 - k) + 2(2k) + (8k + 4)(6k + 2)$$
$$+ k(2(6k + 2) + (24k + 12)(6k + 2))$$
$$\leq 144k^3 + 192k^2 + 73k + 8.$$

This is now easily shown to be the *actual* genus of $X_0(p^3)$ using the well-known genus formula [Shimura 1971, Propositions 1.40 and 1.43]. Thus the inequalities

above are actually equalities. Furthermore, since $g(V_i(A)) \geq g(\overline{\mathbf{E}}_{iA}) \geq g(\overline{\mathbf{Y}}_A)$, Lemma 2.43 implies that $V_1(A)$ and $V_2(A)$ are basic wide opens such that $\mathbf{E}_{1A}$ and $\mathbf{E}_{2A}$ are Zariski subaffinoids of the underlying affinoids. Then, since the reductions of these affinoids each have at least four points at infinity, and since $V_i(A)$ has only four ends, it follows that $\mathbf{E}_{1A}$ and $\mathbf{E}_{2A}$ are the underlying affinoids (with good reduction). Therefore $V_i(A) \cap \hat{U}(A)$ must be an annulus, and we have shown that $\mathscr{C}_0(p^3)$ is a stable covering. □

Remark 9.3. Since $\mathbf{E}_{1A} = \pi_f^{-1}(\mathbf{Y}_A)$, and since $\mathbf{E}_{1A}$ has good reduction with $g(\overline{\mathbf{E}}_{1A}) = g(\overline{\mathbf{Y}}_A)$, it follows that $\pi_f : \overline{\mathbf{E}}_{1A} \to \overline{\mathbf{Y}}_A$ is purely inseparable and factors as Frobenius followed by an isomorphism. Hence, $\overline{\mathbf{E}}_{1A} \cong \overline{\mathbf{Y}}_A^\sigma$, and similarly $\overline{\mathbf{E}}_{2A} \cong \overline{\mathbf{Y}}_{A^\sigma}^\sigma$.

9A. *Graphs and intersection data.* From Theorem 9.2, it is now straightforward to generate graphs for the stable reduction of $X_0(p^3)$ according to the four classes of p (mod 12), and we include these graphs below in Figures 2–5. To make the graphs more understandable, a brief description of how the various components are organized and labeled is in order. First of all, recall from Section 3B that there are
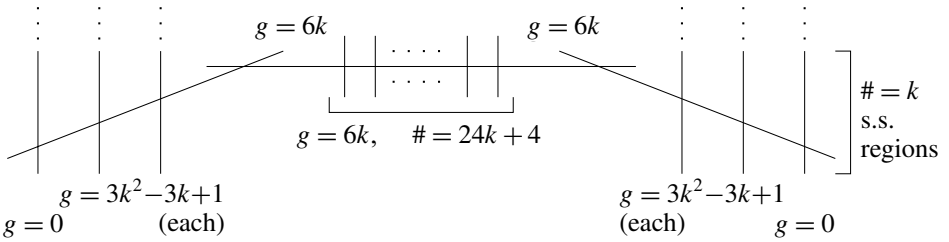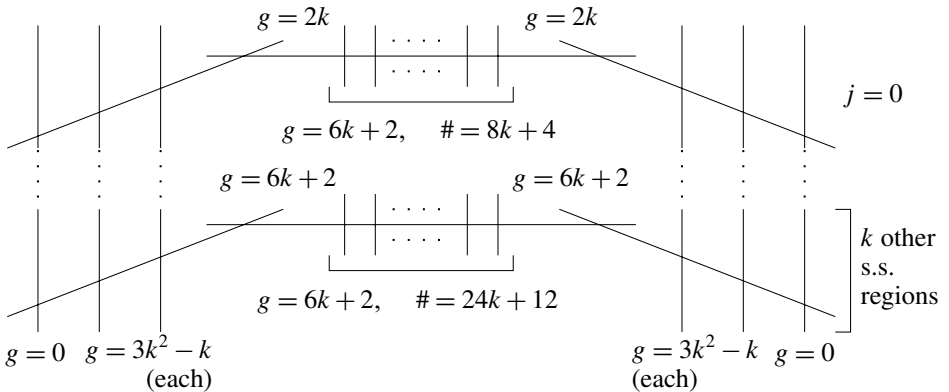

Figure 2. Graph of $X_0(p^3)$ when $p = 12k + 1$.

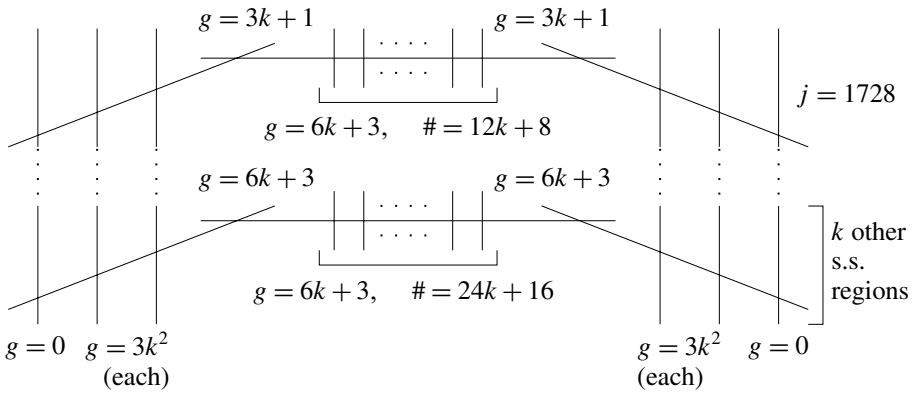**Figure 3.** Graph of $X_0(p^3)$ when $p = 12k + 5$.

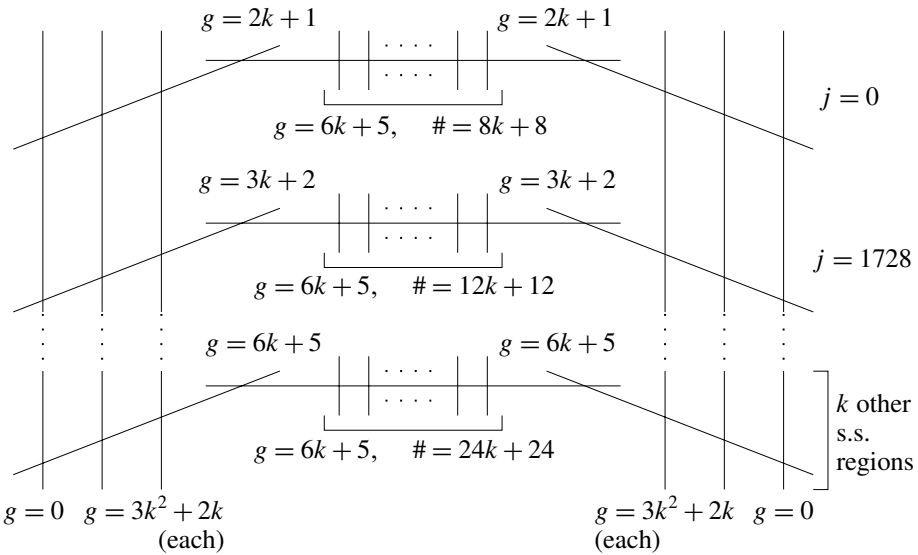**Figure 4.** Graph of $X_0(p^3)$ when $p = 12k + 7$.

Figure 5. Graph of $X_0(p^3)$ when $p = 12k + 11$.

six ordinary components in every case, namely those corresponding to $\mathbf{X}_{30}$, $\mathbf{X}_{21}^{\pm}$, $\mathbf{X}_{12}^{\pm}$, and $\mathbf{X}_{03}$. These are always presented as vertical components and labeled explicitly with their genera. In addition to the six ordinary components, we have one connected, acyclic configuration of components for each supersingular elliptic curve A. This configuration is always presented as a horizontal chain of three components, corresponding to $\mathbf{E}_{2A}$, $\mathbf{Z}_A$, and $\mathbf{E}_{1A}$ (in that order), along with a number of unmarked vertical components intersecting the middle component. We explicitly label the genera of the reductions of $\mathbf{E}_{1A}$ and $\mathbf{E}_{2A}$, but not the central "bridging component", as it always has genus 0. Below the central horizontal component, we

list the number of copies of $y^2 = x^p - x$ that intersect it, as well as the genus of each copy. Finally, we point out for clarification that the components corresponding to $\mathbf{X}_{30}$ and $\mathbf{X}_{21}^{\pm}$ meet each supersingular region in exactly one point in the reduction of $\mathbf{E}_{2A}$, while the same can be said for the other three ordinary components and $\mathbf{E}_{1A}$. In particular, one is reading the graph properly if the Betti number (equivalently the toric rank of the Jacobian) appears to be $5(ss - 1)$, where ss is the number of supersingular j-invariants. This fact generalizes, as we show in the following theorem.

Theorem 9.4. *The toric rank of $J_0(Np^n)$ for $(N, p) = 1$ and $n \geq 0$ is given by $(s(N) - 1)(2n - 1)$, where $s(N)$ is the number of supersingular points on $X_0(N)$ mod p.*

Proof. For $N = 1$ and $n \leq 1$ this follows from [Deligne and Rapoport 1973, §VI.6]. After inverting isogenies, we have the exact sequence

$$0 \to J_0(p^{n-1}) \to J_0(p^n) \times J_0(p^n) \to J_0(p^{n+1}) \to J_0(p^{n+1})^{\text{new}} \to 0.$$

It follows from [Katz and Mazur 1985, Theorem 14.7.2] that $J_0(p^n)^{\text{new}}$ has potential good reduction for $n > 1$. Thus, by induction, the theorem is true for all $J_0(p^n)$. The result for more general N follows from essentially the same argument. □

To go with the stable reduction graphs, we include the intersection multiplicities in Table 1. These numbers have been obtained via a rigid analytic reformulation. In particular, suppose that X and Y are components of a curve with semistable reduction over some extension $K/\mathbb{Q}_p$, and that they intersect in an ordinary double point P. Then $R(P)$ is an annulus (by Proposition 2.10), say with width $w(P)$. In this case, the intersection multiplicity of X and Y at P can be found by

$$M_K(P) = e_p(K) \cdot w(P).$$

Note that while intersection multiplicity depends on K, the width makes sense even over $\mathbb{C}_p$, which in some sense makes width a more natural invariant from the purely geometric perspective.

P	(X_{30}, E_{2A}), (X_{03}, E_{1A})	$(X_{21}^{\pm}, E_{2A})$, $(X_{12}^{\pm}, E_{1A})$	$(\mathbf{Z}_A, E_{2A})$, $(\mathbf{Z}_A, E_{1A})$	$(\mathbf{X}_S, \mathbf{Z}_A)$
$w(P)$	$\dfrac{i(A)}{p(p+1)}$	$\dfrac{2 \cdot i(A)}{p(p^2-1)}$	$\dfrac{(p-1) \cdot i(A)}{2p^2(p+1)}$	$\dfrac{1}{4p^2} \, *$
$M_K(P)$	$p(p-1) \cdot i(A)$	$2p \cdot i(A)$	$\dfrac{(p-1)^2 \cdot i(A)}{2}$	$\dfrac{p^2-1}{4} \, *$

Table 1. Intersection multiplicity data for $X_0(p^3)$.

For our calculations on $X_0(p^3)$, we take $e_p(K) = p^2(p^2 - 1)$, since this is the ramification index over $\mathbb{Q}_p$ for the field of Krir (see [CM 2006, §5] for details). First we treat those singular points where $\mathbf{E}_{iA}$ meets either an ordinary component or the bridging component. The reduction inverse of any such singular point is an annulus in the supersingular locus that surjects via the forgetful map onto some subannulus of $W_A(p)$. Using Hasse invariant and canonical subgroup considerations, we can determine this subannulus and in particular its width. Then we apply Proposition 2.2. For example, the ordinary component corresponding to $\mathbf{X}_{30}$ intersects the one corresponding to (each) $\mathbf{E}_{2A}$ in a unique singular point. As we saw in the proof of Theorem 9.2, the corresponding annulus maps via π_{02} (the forgetful map) onto the subannulus of $W_A(p)$ described by

$$0 < v(x_A) < \frac{i(A)}{p(p+1)} \quad \text{with degree } 1.$$

The ordinary components corresponding to $\mathbf{X}_{21}^{\pm}$ also meet the reduction of $\mathbf{E}_{2A}$ in exactly one singular point (each). The corresponding two annuli surject onto this same subannulus, but with degree $(p-1)/2$. Using this line of reasoning, we arrive at most of the data in Table 1. Note that any two components intersect in at most one point, and so we may designate a singular point in the stable reduction unambiguously by listing a pair of intersecting components.

The only intersection multiplicities that do *not* follow readily from the above reasoning come from singular points where a copy of $y^2 = x^p - x$ (denoted $\mathbf{X}_S$ for $S \in \mathcal{S}(A)$ as in the theorem) intersects a bridging component. At such a singular point, the corresponding annulus maps via π_{11} onto an annulus that is the complement of an affinoid disk inside a residue disk of $\mathbf{SD}_A$. Unfortunately, it is not at all clear what the width of this image annulus is. We have some theoretical evidence and some computational evidence [McMurdy 2004, Remark on p. 27] that suggest that the width of the original annulus, that is, the annulus of intersection, is $1/4p^2$. Therefore, we have included this in Table 1 with an asterisk to indicate that it is our current best guess.

Appendix A: Riemann existence theorem

The p-adic Riemann existence theorem is well known, but not apparently in the literature.[21] Here we recall and adapt the proof of the existence of global meromorphic functions given in [Grauert and Remmert 1977, pp. 208–209], and then use results from [Kiehl 1967] and [Köpf 1974] to deduce the final result.

---

[21]This is possibly because it follows the same lines of reasoning as those used in "the" complex case; see [Springer 1957] for a history of the complex proofs and for a proof that has no obvious p-adic analogue.

Theorem A.1. *Suppose X is a proper[22] one-dimensional smooth rigid space over a complete local field K or a compact Riemann surface, $\mathscr{F} \neq 0$ is a locally free sheaf on X, and D is a divisor of positive degree. Then*

$$\lim_{n \to \infty} \dim_K \mathscr{F}(nD)(X) = \infty,$$

where $K = \mathbb{C}$ if X is a Riemann surface.

Proof. Let $E \leq E'$ be divisors on X, and let $\mathscr{T} = \mathscr{F}(E')/\mathscr{F}(E)$. Then

$$0 \to \mathscr{F}(E)(X) \to \mathscr{F}(E')(X) \to \mathscr{T}(X) \to H^1(X, \mathscr{F}(E)) \to H^1(X, \mathscr{F}(E')) \to 0$$

is exact. Moreover, if r is the rank of $\mathscr{F}$, we have

$$\dim_K \mathscr{T}(X) = r \deg(E' - E).$$

Now, for any coherent sheaf $\mathscr{S}$ on X, let

$$\chi(\mathscr{S}) = \dim_K H^0(X, \mathscr{S}) - \dim_K H^1(X, \mathscr{S}).$$

We deduce that

$$\chi(\mathscr{F}(D)) - \chi(\mathscr{F}) = r \deg D.$$

The theorem follows. □

Theorem A.2 (*p-adic Riemann existence theorem*). *Let X be a smooth proper rigid space of dimension one over a complete local field K. Then X is isomorphic to the analytification of a complete algebraic curve over K.*

Proof. By the previous theorem, there exists a nonconstant map $f : X \to \mathbf{P}^1_K$ that must be finite since X is proper of dimension one. By Kiehl's direct image theorem [1967, Theorem 3.3], it follows that $f_*\mathcal{O}_X$ is a coherent sheaf of analytic algebras on $\mathbf{P}^1_K$. Then, we know from [Köpf 1974, Sätze 4.11 and 5.1] that $f_*\mathcal{O}_X \cong g_*\mathcal{O}_Y$, where g is a finite morphism from some algebraic curve Y onto $\mathbf{P}^1_K$.

To complete the proof, let $\mathscr{C}$ be an admissible open covering of $\mathbf{P}^1_K$ by affinoids. Then $f^{-1}(\mathscr{C})$ and $g^{-1}(\mathscr{C})$ are admissible open coverings of X and Y by affinoids. Moreover, for each $U \in \mathscr{C}$, we have

$$A(f^{-1}U) = f_*\mathcal{O}_X(U) \cong g_*\mathcal{O}_Y(U) = A(g^{-1}U).$$

Thus $f^{-1}U \cong g^{-1}U$ for each $U \in \mathscr{C}$, and these isomorphisms are compatible, which implies that $X \cong Y$. □

---

[22]See [Bosch et al. 1984, 9.6.2] for definition.

Appendix B: Supersingular curves

by Everett W. Howe

Theorem B.1. *For $p \geq 13$ there is a supersingular elliptic curve E defined over $\mathbb{F}_p$ with $j(E) \neq 0, 1728$.*

Proof. Note that there is always at least one supersingular curve over $\mathbb{F}_p$, because the number of curves of trace 0 is given by the Kronecker class number $H(-4p)$, which is positive [Schoof 1987]. So if p is a prime for which neither $j = 0$ nor $j = 1728$ is supersingular, then there exists a supersingular curve over $\mathbb{F}_p$ with $j \neq 0, 1728$.

If p is a prime for which $j = 0$ is supersingular, then p is inert in the field $\mathbb{Q}(\sqrt{-3})$. But then the elliptic curve over $\mathbb{Q}$ with $j = 2^4 \cdot 3^3 \cdot 5^3 = 54000$ (which has CM by the order $\mathbb{Z}[\sqrt{-3}]$) reduces to a supersingular curve over $\mathbb{F}_p$. (If an elliptic curve over $\mathbb{F}_p$ is not supersingular then its endomorphism ring tensored with $\mathbb{Q}$ is an imaginary quadratic field in which p splits.) Note that 54000 is neither 0 nor 1728 modulo p for $p > 11$.

If p is a prime for which $j = 1728$ is supersingular, then p is inert in the field $\mathbb{Q}(i)$. Then the elliptic curve over $\mathbb{Q}$ with $j = 2^3 \cdot 3^3 \cdot 11^3 = 287496$ (which has CM by $\mathbb{Z}[2i]$) reduces to a supersingular curve over $\mathbb{F}_p$, and 287496 is neither 0 nor 1728 modulo p when $p > 11$. □

Appendix C: Concordance with [CM 2006]

Some of the references in [CM 2006] are no longer correct due to some shuffling of the material in this paper. This problem can be resolved by noting the following:

- The reference to §2 on page 265 should be to Section 2C.
- Theorem 2.6 is referred to as Lemma 3.3 on page 295, and as Lemma 2.3 on page 278.
- Proposition 2.14 is referred to as Proposition 3.14 on page 279.
- Proposition 2.34 is referred to as Proposition 2.5 on pages 267 and 278.
- Definition 2.35 and Theorem 2.36 are referred to as Definition 2.6 and Proposition 2.7 on pages 279, 292 and 293.
- Proposition 3.6 is referred to as Lemma 3.6 on page 278.
- Proposition 4.6 is referred to as Corollary 4.6 on page 270.
- Remark 4.7 and Proposition 4.9 are referred to as Remark 4.8 and Proposition 4.10 on page 272.
- Proposition 7.5 and Remark 7.6 are referred to as Proposition 7.4 and Remark 7.5 on pages 267, 275, 277 and 281.
- Theorem B.1 is cited as "results of E. Howe in §10" on page 262.

Index of important notation

w_n, Atkin–Lehner involution on $X_0(p^n)$

$K_n(E)$, canonical subgroup of E of order p^n　　　　　　　　

$K(E)$, (maximal) canonical subgroup of E

$h(E)$, valuation of Hasse invariant of E (almost)

s_n, rigid analytic section of π_{0n} over W_n

$W_A(p^n)$, wide open subspace of $X_0(p^n)$ where $\overline{E} \cong A$

x_A, parameter on $W_A(p)$

$i(A) := |\mathrm{Aut}(A)|/2$

$\mathbf{TS}_A$ and $\mathbf{SD}_A$, too-supersingular and self-dual circles inside $W_A(p)$

$\mathbf{C}_A$ and τ_f, special circle of $W_A(p)$ and map to $\mathbf{SD}_A$

$\mathbf{X}_{ab}^{\pm}$, ordinary affinoids　　　　　　　　

$W_{ab}^{\pm}$, wide open neighborhood of $\mathbf{X}_{ab}^{\pm}$

$\mathrm{Ig}(p^n)$, level p^n Igusa curve

(F, A, α), Woods Hole representation of an elliptic curve　　　　　　　　

$\hat{A}$, formal group of A

B, quaternionic order over $\mathbb{Z}_p$ isomorphic to $\mathrm{End}(\hat{A})$　　　　　　　　

Φ, Gross–Hopkins period map

B', special subset of B^*

w_ρ, generalized Atkin–Lehner involution of $\mathbf{SD}_A$ for $\rho \in B'$

$\mathbf{Y}_A$, nontrivial affinoid in $W_A(p^2)$　　　　　　　　

$\mathscr{C}_0(p^2)$, stable covering of $X_0(p^2)$

$\mathbf{E}_{1,A}$ and $\mathbf{E}_{2,A}$, two pullbacks of $\mathbf{Y}_A$ to $X_0(p^3)$　　　　　　　　

$\mathbf{Z}_A := \pi_{11}^{-1}(\mathbf{SD}_A)$, affinoid in $W_A(p^3)$ that corresponds to
　　the "bridging component"

$\widetilde{w}_\rho$, generalized Atkin–Lehner involution of $\mathbf{Z}_A$ for $\rho \in B'$　　　　　　　　

$\mathscr{D}_\rho^i$ and $\widetilde{\mathscr{D}}_\rho^i$, residue classes of $\mathbf{SD}_A$ and $\mathbf{Z}_A$ invariant under w_ρ and $\widetilde{w}_\rho$

$S_{\sigma,\zeta}$, $\tilde{S}_{\sigma,\zeta}$, order p automorphisms of $\tau_f^{-1}(\mathscr{D}_\rho^i)$ and $\widetilde{\mathscr{D}}_\rho^i$　　　　　　　　

$V_i(A)$ and $U(A)$, wide open neighborhoods of $\mathbf{E}_{i,A}$ and $\mathbf{Z}_A$　　　　　　　　

$\mathscr{S}(A)$, singular residue classes of $\mathbf{Z}_A$

$\mathbf{X}_S$, underlying affinoid of $S \in \mathscr{S}(A)$

$\hat{U}(A)$, basic wide open refinement of $U(A)$

$\mathscr{C}_0(p^3)$, stable covering of $X_0(p^3)$

$M_K(P)$, intersection multiplicity at an ordinary double point　　　　　　　　

$w(P)$, width of the annulus that lifts an ordinary double point

Acknowledgements

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

[BL 1984] S. Bosch and W. Lütkebohmert, "Stable reduction and uniformization of abelian varieties, II", *Invent. Math.* **78**:2 (1984), 257–297. MR 86j:14040b Zbl 0554.14015

[BL 1985] S. Bosch and W. Lütkebohmert, "Stable reduction and uniformization of abelian varieties, I", *Math. Ann.* **270**:3 (1985), 349–379. MR 86j:14040a Zbl 0554.14012

[BL 1993] S. Bosch and W. Lütkebohmert, "Formal and rigid geometry, I: Rigid spaces", *Math. Ann.* **295**:2 (1993), 291–317. MR 94a:11090 Zbl 0808.1401

[Bosch 1976] S. Bosch, "Rigid analytische Gruppen mit guter Reduktion", *Math. Ann.* **223**:3 (1976), 193–205. MR 57 #6033 Zbl 0355.14019

[Bosch 1977a] S. Bosch, "Eine bemerkenswerte Eigenschaft der formellen Fasern affinoider Räume", *Math. Ann.* **229**:1 (1977), 25–45. MR 56 #5952 Zbl 0385.32008

[Bosch 1977b] S. Bosch, "Zur Kohomologietheorie rigid analytischer Räume", *Manuscripta Math.* **20**:1 (1977), 1–27. MR 58 #22683 Zbl 0343.14004

[Bosch et al. 1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis*, Grundlehren der Mathematischen Wissenschaften **261**, Springer, Berlin, 1984. MR 86b:32031 Zbl 0539.14017

[Bouw and Wewers 2004] I. I. Bouw and S. Wewers, "Stable reduction of modular curves", pp. 1–22 in *Modular curves and abelian varieties* (Barcelona, 2002), edited by J. Cremona et al., Progr. Math. **224**, Birkhäuser, Basel, 2004. MR 2005b:11079 Zbl 1147.11316

[Buzzard 2003] K. Buzzard, "Analytic continuation of overconvergent eigenforms", *J. Amer. Math. Soc.* **16**:1 (2003), 29–55. MR 2004c:11063 Zbl 1076.11029

[Cherry 1994] W. Cherry, "Non-Archimedean analytic curves in abelian varieties", *Math. Ann.* **300**:3 (1994), 393–404. MR 96i:14021 Zbl 0808.14019

[CM 2006] R. Coleman and K. McMurdy, "Fake CM and the stable model of $X_0(Np^3)$", *Doc. Math.* Extra Vol. (2006), 261–300. MR 2008j:11068 Zbl 1155.11030

[Coleman 1989] R. F. Coleman, "Reciprocity laws on curves", *Compositio Math.* **72**:2 (1989), 205–235. MR 91c:14028 Zbl 0706.14013

[Coleman 2003] R. F. Coleman, "Stable maps of curves", *Doc. Math.* Extra Vol. (2003), 217–225. MR 2005b:14057 Zbl 1100.14515

[Coleman 2005] R. F. Coleman, "On the components of $X_0(p^n)$", *J. Number Theory* **110**:1 (2005), 3–21. MR 2005k:11118 Zbl 1108.14022

[Coleman and McCallum 1988] R. Coleman and W. McCallum, "Stable reduction of Fermat curves and Jacobi sum Hecke characters", *J. Reine Angew. Math.* **385** (1988), 41–101. MR 89h:11026 Zbl 0654.12003

[Deligne and Mumford 1969] P. Deligne and D. Mumford, "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* 36 (1969), 75–109. MR 41 #6850 Zbl 0181.48803

[Deligne and Rapoport 1973] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques", pp. 143–316 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **Vol. 349**, Springer, Berlin, 1973. MR 49 #2762 Zbl 0281.14010

[Edixhoven 1990] B. Edixhoven, "Minimal resolution and stable reduction of $X_0(N)$", *Ann. Inst. Fourier* (*Grenoble*) **40**:1 (1990), 31–67. MR 92f:11080 Zbl 0679.14009

[Grauert and Remmert 1977] H. Grauert and R. Remmert, *Theorie der Steinschen Räume*, Grundlehren der Mathematischen Wissenschaften **227**, Springer, Berlin, 1977. MR 80j:32001 Zbl 0379.32001

[Gross 1986] B. H. Gross, "On canonical and quasicanonical liftings", *Invent. Math.* **84**:2 (1986), 321–326. MR 87g:14051

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Hopkins and Gross 1994] M. J. Hopkins and B. H. Gross, "Equivariant vector bundles on the Lubin–Tate moduli space", pp. 23–88 in *Topology and representation theory* (Evanston, IL, 1992), edited by E. M. Friedlander and M. E. Mahowald, Contemp. Math. **158**, Amer. Math. Soc., Providence, RI, 1994. MR 95b:14033 Zbl 0807.14037

[Igusa 1968] J.-i. Igusa, "On the algebraic theory of elliptic modular functions", *J. Math. Soc. Japan* **20** (1968), 96–106. MR 39 #1457 Zbl 0164.21101

[Katz 1973] N. M. Katz, "p-adic properties of modular schemes and modular forms", pp. 69–190 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Mathematics **350**, Springer, Berlin, 1973. MR 56 #5434 Zbl 0271.10033

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies **108**, Princeton University Press, 1985. MR 86i:11024 Zbl 0576.14026

[Kiehl 1967] R. Kiehl, "Der Endlichkeitssatz für eigentliche Abbildungen in der nichtarchimedischen Funktionentheorie", *Invent. Math.* **2** (1967), 191–214. MR 35 #1833

[Köpf 1974] U. Köpf, *Über eigentliche Familien algebraischer Varietäten über affinoiden Räumen*, Schr. Math. Inst. Univ. Münster (2) **7**, Universität Münster, 1974. MR 54 #10657 Zbl 0275.14006

[Krir 1996] M. Krir, "Degré d'une extension de $\mathbf{Q}_p^{\mathrm{nr}}$ sur laquelle $J_0(N)$ est semi-stable", *Ann. Inst. Fourier* (*Grenoble*) **46**:2 (1996), 279–291. MR 98g:11072 Zbl 0853.11042

[Lubin et al. 1964] J. Lubin, J.-P. Serre, and J. Tate, "Elliptic curves and formal groups", Lecture notes from the summer institute on algebraic geometry (Woods Hole, MA), 1964, Available at www.ma.utexas.edu/users/voloch/lst.html.

[Lütkebohmert 1993] W. Lütkebohmert, "Riemann's existence problem for a p-adic field", *Invent. Math.* **111**:2 (1993), 309–330. MR 94d:32048 Zbl 0780.32005

[McMurdy 2004] K. McMurdy, "Stable model of $X_0(125)$", *LMS J. Comput. Math.* **7** (2004), 21–36. MR 2005b:11082 Zbl 1134.11329

[McMurdy 2008] K. McMurdy, "Stable reduction of $X_0(81)$", pp. 91–109 in *Computational arithmetic geometry*, Contemp. Math. **463**, Amer. Math. Soc., Providence, RI, 2008. MR 2010d:11065 Zbl 1161.11015

[Mumford 1977] D. Mumford, "Stability of projective varieties", *Enseignement Math.* (2) **23**:1-2 (1977), 39–110. MR 56 #8568 Zbl 0363.14003

[Schoof 1987] R. Schoof, "Nonsingular plane cubic curves over finite fields", *J. Combin. Theory Ser. A* **46**:2 (1987), 183–211. MR 88k:14013 Zbl 0632.14021

[de Shalit 1994] E. de Shalit, "Kronecker's polynomial, supersingular elliptic curves, and *p*-adic periods of modular curves", pp. 135–148 in *p-adic monodromy and the Birch and Swinnerton–Dyer conjecture* (Boston, 1991), edited by B. Mazur and G. Stevens, Contemp. Math. **165**, Amer. Math. Soc., Providence, RI, 1994. MR 95f:11042 Zbl 0863.14015

[Shimura 1971] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Iwanami Shoten, Tokyo, 1971. MR 47 #3318 Zbl 0221.10029

[Springer 1957] G. Springer, *Introduction to Riemann surfaces*, Addison-Wesley, Reading, MA, 1957. MR 19,1169g Zbl 0078.06602

[Tate 1966] J. Tate, "Endomorphisms of abelian varieties over finite fields", *Invent. Math.* **2** (1966), 134–144. MR 34 #5829 Zbl 0147.20303

[Tate 1971] J. Tate, "Rigid analytic spaces", *Invent. Math.* **12** (1971), 257–289. MR 46 #5323 Zbl 0212.25601

[Van der Put 1980] M. van der Put, "The class group of a one-dimensional affinoid space", *Ann. Inst. Fourier* (*Grenoble*) **30**:4 (1980), 155–164. MR 82h:14018 Zbl 0426.14014

[Van der Put 1984] M. van der Put, "Stable reductions of algebraic curves", *Nederl. Akad. Wetensch. Indag. Math.* **46**:4 (1984), 461–478. MR 86a:14023 Zbl 0588.14021

kmcmurdy@ramapo.edu   *Department of Mathematics (TAS),*
                             Ramapo College of New Jersey, 505 Ramapo Valley Rd.,
                             Mahwah, NJ 07430, United States
                             http://phobos.ramapo.edu/~kmcmurdy

coleman@math.berkeley.edu   *Department of Mathematics, University of California,*
                             Berkeley, CA 94720, United States
                             http://math.berkeley.edu/~coleman/

however@alumni.caltech.edu   *Center for Communications Research, 4320 Westerra Court,*
                             San Diego, CA 92121-1969, United States
                             http://alumnus.caltech.edu/~however/

Cyclotomic function fields, Artin–Frobenius automorphisms, and list error correction with optimal rate

Venkatesan Guruswami

Algebraic error-correcting codes that achieve the optimal trade-off between rate and fraction of errors corrected (in the model of list decoding) were recently constructed by a careful "folding" of the Reed–Solomon code. The "low-degree" nature of this folding operation was crucial to the list decoding algorithm. We show how such folding schemes useful for list decoding arise out of the Artin–Frobenius automorphism at primes in Galois extensions. Using this approach, we construct new folded algebraic-geometric codes for list decoding based on cyclotomic function fields with a cyclic Galois group. Such function fields are obtained by adjoining torsion points of the Carlitz action of an irreducible $M \in \mathbb{F}_q[T]$. The Reed–Solomon case corresponds to the simplest such extension (corresponding to the case $M = T$). In the general case, we need to descend to the fixed field of a suitable Galois subgroup in order to ensure the existence of many degree 1 places that can be used for encoding.

Our methods shed new light on algebraic codes and their list decoding, and lead to new codes with optimal trade-off between rate and error correction radius. Quantitatively, these codes provide list decoding (and list recovery/soft decoding) guarantees similar to folded Reed–Solomon codes but with an alphabet size that is only polylogarithmic in the block length. In comparison, for folded RS codes, the alphabet size is a large polynomial in the block length. This has applications to fully explicit (with no brute-force search) binary concatenated codes for list decoding up to the Zyablov radius.

1. Introduction

1A. *Background, context, and motivation.* Error-correcting codes enable reliable transmission of information over a noisy communication channel (as well as reliable data storage and retrieval from a storage medium). The idea behind error-correcting codes is to *encode* the message to be transmitted (or stored) into a longer, redundant string called a *codeword*, which is then communicated over the noisy channel. This is accompanied by a *decoding* procedure that recovers the correct message even when several symbols in the transmitted codeword are corrupted. In this work, we focus on the worst-case model of errors; here, the channel noise has a single parameter $\rho \in (0, 1)$. We do not assume anything about how the errors are distributed beyond an upper bound of ρ on the total fraction of positions where errors may be caused.

The principal trade-off in this theory is between the redundancy and the fraction ρ of errors that can be corrected. Formally, a code is given by an injective encoding function $E : \Sigma^k \to \Sigma^n$. The *block length* of the code equals n, and Σ is its *alphabet*. The redundancy is measured by the *rate R* of the code, defined as the ratio k/n of the number of information symbols to the number of codeword symbols. The larger the rate, the less redundant the code. We are interested in an asymptotically good family of codes, that is, an infinite family of codes of increasing block lengths whose rates are lower bounded by R. The goal is to correct a fraction ρ of errors with as high a rate R as possible for the code family. It is simple to see that this rate R cannot exceed $1 - \rho$. Indeed, the channel could corrupt the last ρ fraction of symbols, and the first $(1-\rho)n$ symbols should thus contain enough information to recover the Rn message symbols, implying $R \leqslant 1 - \rho$.

Quite remarkably, this simplistic upper bound can in fact be met, via a natural family of algebraic codes together with efficient decoding algorithms. Specifically, recent progress in algebraic coding theory [Parvaresh and Vardy 2005; Guruswami and Rudra 2008] has led to the construction of explicit codes over large alphabets that achieve the optimal rate versus error correction radius trade-off — namely, they admit efficient *list decoding* algorithms to correct close to the optimal fraction $1 - R$ of errors with rate R. List decoding is an error correction model where the

decoder is allowed to output a small list of messages which must include the correct message. Allowing such a list is essential in order to correct more than a fraction $(1-R)/2$ of errors with rate R. In practice, having more than one codeword on the list is a rare event, and in case of multiple candidates, one can also return the closest codeword (as the one with the highest likelihood). Also, for many applications of codes, pinning down the message to a small list suffices, and some application-specific context information can be used to identify the correct message from the list. See, for instance, [Guruswami 2007, Chapter 1] or [Guruswami 2004] for more detailed background on list decoding.

We now return to the mathematics of optimal rate codes for list decoding. The algebraic codes constructed in [Guruswami and Rudra 2008] are *folded* Reed–Solomon (RS) codes, where the RS encoding $(f(1), f(\gamma), \ldots, f(\gamma^{n-1}))$ of a low-degree polynomial $f \in \mathbb{F}_q[T]$ is viewed as a codeword of length $N = n/m$ over the alphabet $\mathbb{F}_q^m$ by identifying successive blocks of m symbols. Here γ is a primitive element of the field $\mathbb{F}_q$.

Simplifying matters somewhat, the principal algebraic engine behind the list decoding algorithm in [Guruswami and Rudra 2008] was the identity $f(\gamma T) \equiv f(T)^q$ (mod $(T^{q-1} - \gamma)$), and the fact that $(T^{q-1} - \gamma)$ is irreducible over $\mathbb{F}_q$. This gave a low-degree algebraic relation between $f(T)$ and $f(\gamma T)$ in the residue field $\mathbb{F}_q[T]/(T^{q-1}-\gamma)$. This together with an algebraic relation found by a certain "interpolation step" during decoding enabled us to find the list of all relevant message polynomials $f(T)$ efficiently. Essentially, this gave two algebraically independent low-degree polynomial relations between the residues of $f(T)$ and $f(\gamma T)$ in the extension field $\mathbb{F}_q[T]/(T^{q-1} - \gamma)$. Solving these gives the list of possible values for $f(T) \mod (T^{q-1}-\gamma)$, which also suffices to identify the message polynomial $f(T)$, as its degree is less than $q - 1$.

One of the motivations of this work is to gain a deeper understanding of the general algebraic principles underlying the above folding, with the hope of extending it to more general algebraic-geometric (AG) codes — an interesting algebraic question in its own right, but also important for potentially improving the alphabet size of the codes, as well as the decoding complexity and output list size of the decoding algorithm. (The large complexity and list size of the folded RS decoding algorithm in [Guruswami and Rudra 2008] are a direct consequence of the large degree q in the identity relating $f(\gamma T)$ and $f(T)$.)

The precursor to the folded RS codes were the Parvaresh–Vardy codes [2005]. Here the encoding of a message polynomial $f(T)$ consists of the evaluations of f at distinct elements of $\mathbb{F}_q$ together with the evaluations of a few other algebraically related polynomials $f_1(T), \ldots, f_m(T)$ (for some parameter $m \geqslant 1$) at these points. The algebraic relations between f_i and f are used at the decoder together with a multivariate polynomial relation between $f(T), f_1(T), \ldots, f_m(T)$ to solve for

$f(T)$. An extension of the Parvaresh–Vardy codes [2005] to arbitrary AG codes was achieved in [Guruswami and Patthak 2008]. But in these codes, there is a substantial loss in rate since the encoding includes the evaluations of additional function(s) explicitly picked to satisfy a low-degree relation over some residue field. The crucial insight in the construction of folded RS codes was the fact that this additional function could just be the closely related function $f(\gamma T)$ — the image of $f(T)$ under the automorphism $T \mapsto \gamma T$ of $\mathbb{F}_q(T)$. It is a priori not clear for which algebraic function fields one can have a similar algebraic phenomenon and thereby deduce constructions of folded list-decodable codes analogous to folded RS codes.

1B. *Summary of our contributions.* We explain how folding schemes conducive to list decoding (such as the above relation between $f(\gamma T)$ and $f(T)$) arise out of the *Artin–Frobenius automorphism* at primes in Galois extensions. We then use this approach to construct new list-decodable folded AG codes based on *cyclotomic function fields* with a cyclic Galois group. Cyclotomic function fields [Carlitz 1938; Hayes 1974] are obtained by adjoining torsion points of the Carlitz action of an irreducible $M \in \mathbb{F}_q[T]$. The RS case corresponds to the simplest such extension (corresponding to the case $M = T$). In the general case, we need to descend to the fixed field of a suitable Galois subgroup in order to ensure the existence of many degree 1 places that can be used for encoding. We establish some key algebraic lemmas that characterize the desired subfield in terms of the appropriate generator μ in the algebraic closure of $\mathbb{F}_q(T)$ and its minimal polynomial over $\mathbb{F}_q(T)$. We then tackle the computational algebra challenge of computing a representation of the subfield and its rational places, and the message space, that is conducive for efficient encoding and decoding of the associated AG code.

Our constructions lead to some substantial quantitative improvements in the alphabet size, which we discuss in Section 1D. We also make some simplifications in the list decoding algorithm and avoid the need of a zero-increasing basis at each code place (Lemma 6.2). This, together with several other ideas, lets us implement the list decoding algorithm in polynomial time assuming *only* the natural representation of the code needed for efficient encoding, namely a basis for the message space. Computing such a basis remains an interesting challenge in computational function field theory. Our description and analysis of the list decoding algorithm in this work is *self-contained*, though it builds strongly on the framework of the algorithms in [Sudan 1997; Parvaresh and Vardy 2005; Guruswami and Patthak 2008; Guruswami and Rudra 2008].

1C. *Galois extensions and Artin automorphisms in list decoding.* We will now discuss how and why Artin–Frobenius automorphisms arise in the seemingly distant world of list decoding, and why we make the choice of cyclotomic function

fields for the underlying function field. In order to generalize the folding operation from the RS case, it is natural to look for function fields whose automorphisms we understand reasonably well. Galois extensions are a natural subclass of function fields to consider, with the hope that some automorphism in the Galois group will give a low-degree relation over some residue field. Unfortunately, the explicit constructions of good AG code families are typically based on a tower of function fields [Garcia and Stichtenoth 1995; 1996], where each step is Galois, but the whole extension is not. (Stichtenoth [2006] recently showed the existence of a Galois extension with the optimal trade-off between genus and number of rational places, but this extension is not, and cannot be, cyclic, as we require.)

In Galois extensions K/F, for each place A' in the extension field K, there is a special and important automorphism called the Artin–Frobenius automorphism (see, for example, [Marcus 1977, Chapter 4]) that simply powers the residue of any (regular) function at that place. The exponent or degree of this map is the norm of the place A of F lying below A'. Since the degree dictates the complexity of decoding, we would like this norm to be small. On the other hand, the residue field at A' needs to be large enough so that the message functions can be uniquely identified by their residue modulo A'. The most appealing way to realize this is if the place A is inert, that is, has a unique A' lying above it. However, this condition can only hold if the Galois group is cyclic, a rather strong restriction. For example, it is known [Frey et al. 1992] that even abelian extensions must be *asymptotically bad*.

In order to construct AG codes, we also need to have a good control of how certain primes split in the extension. For cyclotomic function fields, and of course their better-known number-theoretic counterparts $\mathbb{Q}(\omega)$ obtained by adjoining a root of unity ω, this theory is well-developed. As mentioned earlier, the cyclotomic function field we use itself has very few rational places. So we need to descend to an appropriate subfield where many degree 1 places of $\mathbb{F}_q(T)$ split completely, and develop some underlying theory concerning the structure of this subfield that can be exploited for efficient computation with them.

The Artin–Frobenius automorphism[1] is of course a well-known and fundamental notion in algebraic number theory, playing a role in the Chebotarev density theorem and Dirichlet's theorem on infinitude of primes in arithmetic progressions, as well as quadratic and more general reciprocity laws. We find it rather intriguing that this notion ends up playing an important role in algorithmic coding theory as well.

---

[1]Following [Rosen 2002], we will henceforth refer to the Artin–Frobenius automorphisms as simply Artin automorphisms. Many texts refer to these as Frobenius automorphisms. Since the latter term is most commonly associated with automorphism $x \mapsto x^q$ of $\mathbb{F}_{q^m}$, we use the term Artin automorphism to refer to the general notion that applies to all Galois extensions. The association of a place with its Artin–Frobenius automorphism is called the Artin map.

1D. *Long codes achieving list decoding capacity and explicit binary concatenated codes.* Quantitatively, our cyclotomic function field codes achieve list decoding (and list recovery[2]) guarantees similar to folded RS codes, but with an alphabet size that is only *polylogarithmic* in the block length. In comparison, for folded RS codes, the alphabet size is a large polynomial in the block length. We note that Guruswami and Rudra [2008] also present capacity-achieving codes of rate R for list decoding a fraction $(1 - R - \varepsilon)$ of errors with alphabet size $|\Sigma| = 2^{(1/\varepsilon)^{O(1)}}$, a fixed constant depending only on ε. But these codes do not have the strong "list recovery" (or more generally, soft decoding) property of folded RS codes.

Our codes inherit the powerful list recovery property of folded RS codes, which makes them very useful as outer codes in constructions of concatenated codes.[3] In fact, due to their small alphabet size, they are even better in this role. Indeed, they can serve as outer codes for a family of concatenated codes list-decodable up to the so-called Zyablov radius, *with no brute-force search* for the inner codes. This is the first such construction for list decoding. It is similar to the "Justesen-style" explicit constructions for rate versus distance from [Justesen 1972; Shen 1993], except even easier, as one can use the ensemble of *all linear codes* instead of the succinct Wozencraft ensemble at the inner level of the concatenated scheme.

1E. *Related work.* Codes based on cyclotomic function fields have been considered previously in the literature. Some specific (nonasymptotic) constructions of function fields with many rational places over small fields $\mathbb{F}_q$ ($q \leqslant 5$) appear in [Niederreiter and Xing 1996; 1997]. Cyclotomic codes based on the action of polynomials T^a for small a appear in [Quebbemann 1988], but decoding algorithms are not discussed for these codes, nor are these extensions cyclic as we require. Our approach is more general and works based on the action of an arbitrary irreducible polynomial. Exploiting the Artin automorphism of cyclotomic fields for an algorithmic purpose is also new to this work.

Independent of our work, Huang and Narayanan [2008] have considered AG codes constructed from Galois extensions, and observed how automorphisms of large order can be used for folding such codes. To our knowledge, the only instantiation of this approach that improves on folded RS codes is the one based on cyclotomic function fields from our work. As an alternate approach, they also

---

[2]List recovery is a generalization of list decoding where for each position a set of possible symbols is provided as input to the decoder, and the goal is to find all codewords that agree with some element of the input sets for at least a certain fraction of positions; see Remark 6.11.

[3]In binary concatenated codes, the message is first encoded by an "outer" code over a large alphabet Σ, and then each outer codeword symbol is encoded by an "inner" binary code $C_{in} : \Sigma \rightarrow \{0, 1\}^b$. Despite its simplicity, code concatenation remains the preeminent method for constructing good codes over small alphabets such as binary codes.

propose a decoding method that works with folding via automorphisms of small order. This involves computing several coefficients of the power series expansion of the message function at a low-degree place. Unfortunately, piecing together these coefficients into a function could lead to an exponential list size bound. The authors suggest a heuristic assumption under which they can show that for a *random* received word, the expected list size and running time are polynomially bounded.

2. Background on cyclotomic function fields

We assume familiarity with basic background on global fields and their extensions such as valuations and places, Galois extensions, decomposition of primes, ramification, Artin–Frobenius automorphism, etc. In this section, we will focus on background material concerning cyclotomic function fields. These are the function-field analog of the classic cyclotomic number fields from algebraic number theory. This theory was developed by Hayes [1974], building upon ideas due to Carlitz [1938]. The objective was to develop an explicit class field theory classifying all abelian extensions of the rational function field $\mathbb{F}_q(T)$, analogous to classic results for $\mathbb{Q}$ and imaginary quadratic extensions of $\mathbb{Q}$. The common idea in these results is to allow a ring of "integers" in the ground field to act on part of its algebraic closure, and obtain abelian extensions by adjoining torsion points of this action. We will now describe these extensions of $\mathbb{F}_q(T)$.

Let T be an indeterminate over the finite field $\mathbb{F}_q$. Let $R_T = \mathbb{F}_q[T]$ denote the polynomial ring, and $F = \mathbb{F}_q(T)$ the field of rational functions. Let F^{ac} be a fixed algebraic closure of F. Let $\mathrm{End}_{\mathbb{F}_q}(F^{\mathrm{ac}})$ be the ring of $\mathbb{F}_q$-endomorphisms of F^{ac}, thought of as a $\mathbb{F}_q$-vector space. We consider two special elements of $\mathrm{End}_{\mathbb{F}_q}(F^{\mathrm{ac}})$:

 (i) the Frobenius automorphism τ defined by $\tau(z) = z^q$ for all $z \in F^{\mathrm{ac}}$, and

 (ii) the map μ_T defined by $\mu_T(z) = Tz$ for all $z \in F^{\mathrm{ac}}$.

The substitution $T \to \tau + \mu_T$ yields a ring homomorphism from R_T to $\mathrm{End}_{\mathbb{F}_q}(F^{\mathrm{ac}})$ given by

$$f(T) \mapsto f(\tau + \mu_T).$$

Using this, we can define the *Carlitz action* of R_T on F^{ac} as follows: for $M \in R_T$,

$$C_M(z) = M(\tau + \mu_T)(z) \quad \text{for all } z \in F^{\mathrm{ac}}.$$

This action endows F^{ac} with the structure of an R_T-module, which is called the Carlitz module. For a nonzero polynomial $M \in R_T$, define the set

$$\Lambda_M = \{z \in F^{\mathrm{ac}} \mid C_M(z) = 0\},$$

to consist of the M-torsion points of F^{ac}, that is, the elements annihilated by the Carlitz action of M (this is also the set of zeroes of the polynomial $C_M(Z) \in R_T[Z]$).

Since R_T is commutative, Λ_M is in fact an R_T-submodule of F^{ac}. It is in fact a cyclic R_T-module, naturally isomorphic to $R_T/(M)$.

The cyclotomic function field $F(\Lambda_M)$ is obtained by adjoining the set Λ_M of M-torsion points to F.[4] The following result summarizes some fundamental facts about cyclotomic function fields, stated for the special case when M is irreducible (we will only use such extensions). Proofs can also be found in graduate texts [Rosen 2002, Chapter 12; Villa Salvador 2006, Chapter 12]. In what follows, we will often use the convention that an irreducible polynomial $P \in R_T$ is identified with the place of F that is the zero of P, and also denote this place by P. Recall that these are all the places of F, with the exception of the place P_∞, which is the unique pole of T.

For a place P, we denote by $\mathbb{O}_P$ the ring of regular functions at P (that is, the valuation ring corresponding to the place P). Thus $\mathbb{O}_P/P$ is the residue field at P.

Proposition 2.1 [Hayes 1974]. *Let $M \in R_T$ be a nonzero degree d monic polynomial that is irreducible over $\mathbb{F}_q$. Let $K = F(\Lambda_M)$.*

(i) *$C_M(Z)$ is a separable polynomial in Z of degree q^d over R_T, of the form $\sum_{i=0}^{d}[M,i]Z^{q^i}$ where the degree of $[M,i]$ as a polynomial in T is $q^i(d-i)$, and further $[M,0] = M$.*
The polynomial $\psi_M(Z) = C_M(Z)/Z$ is irreducible in $R_T[Z]$. The field K is equal to the splitting field of $\psi_M(Z)$, and is generated by any nonzero element $\lambda \in \Lambda_M$, that is, $K = F(\lambda)$.

(ii) *K/F is a Galois extension of degree $(q^d - 1)$ and $\mathrm{Gal}(K/F)$ is isomorphic to $(R_T/(M))^*$, the cyclic multiplicative group of units of the field $R_T/(M)$. The Galois automorphism σ_N associated with $\bar{N} \in (R_T/(M))^*$ is given by $\sigma_N(\lambda) = C_N(\lambda)$.*
The Galois automorphisms commute with the Carlitz action: for any $\sigma \in \mathrm{Gal}(K/F)$ and $A \in R_T$, $\sigma(C_A(x)) = C_A(\sigma(x))$ for all $x \in K$.

(iii) *If $P \in R_T$ is a monic irreducible polynomial different from M, then the Artin automorphism at the place P is equal to σ_P.*

(iv) *The integral closure of R_T in $F(\lambda)$ equals $R_T[\lambda]$.*

(v) *The genus g_M of $F(\Lambda_M)$ satisfies $2g_M - 2 = d(q^d - 2) - (q/q - 1)(q^d - 1)$.*

The splitting behavior of primes in the extension $F(\Lambda_M)/F$ will be crucial for our construction. We record this as a separate proposition below.

---

[4]It is instructive to compare this with the more familiar setting of cyclotomic number fields. There, one lets $\mathbb{Z}$ act on the multiplicative group $(\mathbb{Q}^{\mathrm{ac}})^*$ with the endomorphism corresponding to $n \in \mathbb{Z}$ sending $\zeta \mapsto \zeta^n$ for $\zeta \in \mathbb{Q}^{\mathrm{ac}}$. The n-torsion points now equal $\{\zeta \in \mathbb{Q}^{\mathrm{ac}} \mid \zeta^n = 1\}$, that is, the n-th roots of unity. Adjoining these gives the various cyclotomic number fields.

Proposition 2.2. *Let $M \in R_T$, $M \neq 0$, be a monic, irreducible polynomial of degree d.*

(i) *Ramification at M: the place M is totally ramified in the extension $F(\Lambda_M)/F$. If $\lambda \in \Lambda_M$ is a root of $C_M(z)/z$ and $\tilde{M}$ is the unique place of $F(\Lambda_M)$ lying above M, then λ is a $\tilde{M}$-prime element, that is, $v_{\tilde{M}}(\lambda) = 1$.*

(ii) *Ramification at P_∞: the infinite place P_∞ of F, that is, the pole of T, splits into $(q^d - 1)/(q - 1)$ places of degree 1 in $F(\Lambda_M)/F$, each with ramification index $(q - 1)$. Its decomposition group equals $\mathbb{F}_q^*$.*

(iii) *Splitting at other places: if $P \in R_T$ is a monic, irreducible polynomial different from M, then P is unramified in $F(\Lambda_M)/F$, and splits into $(q^d - 1)/f$ primes of degree $f \deg P$ where f is the order of P modulo M (that is, the smallest positive integer e such that $P^e \equiv 1 \pmod{M}$).*

3. Reed–Solomon codes as cyclotomic function field codes

We now discuss how RS codes arise out of the simplest cyclotomic extension $F(\Lambda_T)/F$. This serves both as a warm-up for our later results, and as a method to illustrate that one can view the folding employed in [Guruswami and Rudra 2008] as arising naturally from the Artin automorphism at a certain prime in the extension $F(\Lambda_T)/F$.

We have $\Lambda_T = \{u \in F^{\mathrm{ac}} \mid u^q + Tu = 0\}$. Pick a nonzero $\lambda \in \Lambda_T$. By Proposition 2.2, the only ramified places in $F(\Lambda_T)/F$ are T and the pole P_∞ of T. Both of these are totally ramified and have a unique place above them in $F(\Lambda_T)$. Denote by Q_∞ the place above P_∞ in $F(\Lambda_T)$.

We have $\lambda^{q-1} = -T$, so λ has a pole of order one at Q_∞, and no poles elsewhere. The place $T + 1$ splits completely into $n = q - 1$ places of degree 1 in $F(\Lambda_T)$. The evaluation of λ at these places corresponds to the roots of $x^{q-1} = 1$, that is, to nonzero elements of $\mathbb{F}_q$. Thus the places above $T + 1$ can be described as $P_1, P_\gamma, \ldots, P_{\gamma^{q-2}}$, where γ is a primitive element of $\mathbb{F}_q$ and $\lambda(P_{\gamma^i}) = \gamma^i$ for $i = 0, 1, \ldots, q - 2$.

For $k < q - 1$, define $\mathcal{M}_k = \left\{ \sum_{i=0}^{k-1} \beta_i \lambda^i \mid \beta_i \in \mathbb{F}_q \right\}$. $\mathcal{M}_k$ has q^k elements, each with at most $(k - 1)$ poles at Q_∞ and no poles elsewhere. Consider the $\mathbb{F}_q$-linear map $E_{\mathrm{RS}} : \mathcal{M}_k \to \mathbb{F}_q^n$ defined as

$$E_{\mathrm{RS}}(f) = \big(f(P_1), f(P_\gamma), \ldots, f(P_{\gamma^{q-2}}) \big).$$

Clearly this just defines an $[n, k]_q$ RS code, consisting of evaluations of polynomials of degree $< k$ at elements of $\mathbb{F}_q^*$.

Consider the place $T + \gamma$ of F. The condition $(T + \gamma)^f \equiv 1 \pmod{T}$ is satisfied if and only if $\gamma^f = 1$, which happens if and only if $(q - 1) \mid f$. Therefore, the place

$T + \gamma$ remains inert in $F(\Lambda_T)/F$. Let A denote the unique place above $T + \gamma$ in $F(\Lambda_T)$. The degree of A equals $q - 1$.

The Artin automorphism at A, σ_A, is given by $\sigma_A(\lambda) = C_{T+\gamma}(\lambda) = C_\gamma(\lambda) = \gamma\lambda$. Note that this implies $f(P_{\gamma^{i+1}}) = \sigma_A(f)(P_{\gamma^i})$ for $0 \leqslant i < q - 2$. By the property of the Artin automorphism, we have $\sigma_A(f) \equiv f^q \pmod{A}$ for all $f \in R_T[\lambda]$. Note that this is same as the condition $f(\gamma\lambda) \equiv f(\lambda)^q \pmod{(\lambda^{q-1} - \gamma)}$ treating f as a polynomial in λ. This corresponds to the algebraic relation between $f(X)$ and $f(\gamma X)$ in the ring $\mathbb{F}_q[X]$ that was used by Guruswami and Rudra [2008] in their decoding algorithm, specifically in the task of finding all $f(X)$ of degree less than k satisfying $Q(X, f(X), f(\gamma X)) = 0$ for a given $Q \in \mathbb{F}_q[X, Y, Z]$. In the cyclotomic language, this corresponds to finding all $f \in R_T[\lambda]$ with fewer than k poles at Q_∞ satisfying $Q(f, \sigma_A(f)) = 0$ for $Q \in R_T[\lambda](Y, Z)$. Since $\deg A = q - 1 \geqslant k$, f is determined by its residue at A, and we know $\sigma_A(f) \equiv f^q \pmod{A}$. Therefore, we can find all such f by finding the roots of the univariate polynomial $Q(Y, Y^q)$ mod A over the residue field $\mathbb{O}_A/A$.

4. Subfield construction from cyclic cyclotomic function fields

In this section, we will construct the function field construction that will be used for our AG codes, and establish the key algebraic facts concerning it. The approach will be to take the cyclotomic field $K = F(\Lambda_M)$, where M is an irreducible of degree $d > 1$, and get a code over $\mathbb{F}_q$. But the only places of degree 1 in $F(\Lambda_M)$ are the ones above the pole P_∞ of T. There are only $(q^d - 1)/(q - 1)$ such places above P_∞, which is much smaller than the genus. So we descend to a subfield where many degree 1 places split completely. This is done by taking a subgroup H of $(\mathbb{F}_q[T]/(M))^*$ with many degree 1 polynomials and considering the fixed field $E = K^H$. For every irreducible $N \in R_T$ such that $\bar{N} = N \mod M \in H$, the place N splits completely in the extension E/F (this follows from the fact that C_N is the Artin automorphism at the place N). This technique has also been used in works mentioned earlier [Quebbemann 1988; Niederreiter and Xing 1996; 1997], though our approach is more general and works with any irreducible M. The study of algorithms for cyclotomic codes and the role played by the Artin automorphism in their list decoding is also novel to our work.

4A. *Table of parameters.* Since there is an unavoidable surfeit of notation and parameters used in this section and Section 5, we summarize them for easy reference in the Appendix.

4B. *Function field construction.* Let $\mathbb{F}_r$ be a subfield of $\mathbb{F}_q$. Let $M \in \mathbb{F}_r[T]$ be a monic polynomial that is irreducible over $\mathbb{F}_q$ (note that we require $M(T)$ to have coefficients in the smaller field $\mathbb{F}_r$, but demand irreducibility in the ring $\mathbb{F}_q[T]$).

The following lemma follows from the general characterization of when binomials $T^m - \alpha$ are irreducible in $\mathbb{F}_q[T]$ [Lidl and Niederreiter 1986, Chapter 3].

Lemma 4.1. *Let $d \geqslant 1$ be an odd integer such that every prime factor of d divides $(r - 1)$ and $\gcd(d, (q - 1)/(r - 1)) = 1$. Let γ be a primitive element of $\mathbb{F}_r$. Then $T^d - \gamma \in \mathbb{F}_r[T]$ is irreducible in $\mathbb{F}_q[T]$.*

A simple choice for which the above conditions are met is $r = 2^a$, $q = r^2$, and $d = r - 1$ (we will need a more complicated choice for our list decoding result in Theorem 7.1). For the sake of generality as well as clarity of exposition, we will develop the theory without making specific choices for the parameters, a somewhat intricate task we will undertake in Section 7.

For the rest of this section, fix $M(T) = T^d - \gamma$ as guaranteed by Lemma 4.1. We continue with the notation $F = \mathbb{F}_q(T)$, $R_T = \mathbb{F}_q[T]$, and $K = F(\Lambda_M)$. Fix a generator $\lambda \in \Lambda_M$ of K/F so that $K = F(\lambda)$.

Let G be the Galois group of K/F, which is isomorphic to the cyclic multiplicative group $(\mathbb{F}_q[T]/(M))^*$. Let $H \subset G$ be the subgroup $\mathbb{F}_q^* \cdot (\mathbb{F}_r[T]/(M))^*$. The cardinality of H is $(r^d - 1) \cdot (q - 1)/(r - 1)$. Note that since G is cyclic, there is a unique subgroup H of this size. Indeed, if $\Gamma \in G$ is an arbitrary generator of G, then $H = \{1, \Gamma^b, \Gamma^{2b}, \ldots, \Gamma^{q^d - 1 - b}\}$, where

$$b = \frac{|G|}{|H|} = \frac{q^d - 1}{r^d - 1} \cdot \frac{r - 1}{q - 1}. \tag{4-1}$$

Let $A \in R_T$ be an arbitrary polynomial such that $A \mod M$ is a generator of $(\mathbb{F}_q[T]/(M))^*$. We can then take Γ so that $\Gamma(\lambda) = C_A(\lambda)$. We fix a choice of A in the sequel and assume that A is precomputed and known. In Section 5C we will pick such an A of appropriately large degree D. The effective version of Dirichlet's theorem for irreducible polynomials in arithmetic progressions guarantees the existence of such polynomials A for large enough degree [Rosen 2002, Theorem 4.8].

Note that by Proposition 2.1(ii), the Galois action commutes with the Carlitz action and therefore $\Gamma^j(\lambda) = C_{A^j}(\lambda)$ for all $j \geqslant 1$. Thus knowing the polynomial A lets us compute the action of the automorphisms of H on any desired element of $K = F(\lambda)$.

Let $E \subset K$ be the subfield of K fixed by the subgroup H, that is,

$$E = \{x \in K \mid \sigma(x) = x \text{ for all } \sigma \in H\}.$$

The field E will be the one used to construct our codes. We first record some basic properties of the extension E/F, and how certain places decompose in this extension.

Proposition 4.2. *Let* $E = F(\Lambda_M)^H$.

(i) *E/F is a Galois extension of degree $[E : F] = b$.*

(ii) *The place M is the only ramified place in E/F, and it is totally ramified with a unique place M' above it in E.*

(iii) *The infinite place P_∞ of F, that is, the pole of T, splits completely into b degree 1 places in E.*

(iv) *The genus g_E of E equals $d(b-1)/2 + 1$.*

(v) *For each $\beta \in \mathbb{F}_r$, the place $T - \beta$ of F splits completely into b degree 1 places in E.*

(vi) *If $A \in R_T$ is irreducible of degree $\ell \geqslant 1$ and $A \mod M$ is a primitive element of $R_T/(M)$, then the place A is inert in E/F. The Artin automorphism σ_A at A satisfies*

$$\sigma_A(x) \equiv x^{q^\ell} \pmod{A'} \tag{4-2}$$

for all $x \in \mathbb{O}_{A'}$, where A' is the unique place of E lying above A.

Proof. By Galois theory, $[E : F] = |G|/|H| = b$. Since G is abelian, E/F is Galois with Galois group isomorphic to G/H. Since $E \subset K$, and M is totally ramified in K, it must also be totally ramified in E. The only other place ramified in K is P_∞, and since H contains the decomposition group $\mathbb{F}_q^*$ of P_∞, P_∞ must split completely in E/F.

The genus of E is easily computed, since E/F is a tamely ramified extension [Stichtenoth 1993, Sec. III.5]. Since only the place M of degree d is ramified, we have $2g_E - 2 = d(b-1)$.

Since $H \supset \mathbb{F}_r[T]$, for $\beta \in \mathbb{F}_r$, the Artin automorphism $\sigma_{T-\beta}$ of the place $T - \beta$ in K/F belongs to H. The Artin automorphism of $T - \beta$ in the extension E/F is the restriction of $\sigma_{T-\beta}$ to E, which is trivial since H fixes E. It follows that $T - \beta$ splits completely in E.

For an irreducible polynomial $A \in R_T$ which has order $q^d - 1$ modulo M, by Proposition 2.2(iii), the place A remains inert in the extension K/F, and therefore also in the subextension E/F. Since the degree of the place A equals ℓ, (4-2) follows from the definition of the Artin automorphism at A. □

4C. *A generator for E and its properties.* We would like to represent elements of E and to be able to evaluate them at the places above $T - \beta$. To this end, we will exhibit a $\mu \in F^{\mathrm{ac}}$ such that $E = F(\mu)$ along with a defining equation for μ (which will then aid in the evaluations of μ at the requisite places).

Theorem 4.3. *Let λ be an arbitrary nonzero element of Λ_M (so that $K = F(\lambda)$). Define*

$$\mu \stackrel{\mathrm{def}}{=} \prod_{\sigma \in H} \sigma(\lambda) = C_{A^b}(\lambda) C_{A^{2b}}(\lambda) \cdots C_{A^{q^d-1}}(\lambda). \tag{4-3}$$

Then the fixed field $E = K^H$ equals the extension field $F(\mu)$. The minimal polynomial $h \in R_T[Z]$ of μ over F is given by

$$h(Z) = \prod_{j=0}^{b-1} (Z - \Gamma^j(\mu)).$$

Further, the polynomial $h(Z)$ can be computed in $q^{O(d)}$ time.

Proof (sketch). By definition, μ is fixed by each $\pi \in H$ and so $\mu \in E$. Therefore $F(\mu) \subseteq E$. To show $E = F(\mu)$, we will argue that $[F(\mu) : F] = b$, which in turn follows if we show that $h(Z)$ has coefficients in F and is irreducible over F. It is easy to see that the coefficients of h are fixed by Γ and hence by all of $\mathrm{Gal}(K/F)$, and so must belong to F. Since λ and all its Galois conjugates $C_{A^i}(\lambda)$ are integral over F, each $\Gamma^j(\mu)$ is integral over F, and thus so is each coefficient of h. But since we already know they belong to F, the coefficients must in fact lie in R_T.

The irreducibility of h over R_T can be shown using Eisenstein's criterion with respect to M. Indeed, except the leading coefficient, every other coefficient of h is divisible by λ, and since $\lambda \in \tilde{M}$ (by [Proposition 2.2](#)), these coefficients belong to the ideal $F \cap \tilde{M} = M$. The constant term of h equals $\prod_{0 \leqslant i < q^d - 1} C_{A^i}(\lambda)$, which is also the constant term of

$$C_M(Z)/Z = \prod_{0 \leqslant i < q^d - 1} (Z - C_{A^i}(\lambda)).$$

The latter equals M by [Proposition 2.1(i)](#). Thus the constant term of h is not divisible by M^2. By Eisenstein's criterion, h must be irreducible over F.

Finally, we address how the coefficients of $h(Z)$ can be computed efficiently. Note that for $j = 0, 1, \ldots, b - 1$,

$$\Gamma^j(\mu) = \prod_{\substack{0 \leqslant i < q^d - 1 \\ i \bmod b = j}} \Gamma^i(\lambda) = \prod_{\substack{0 \leqslant i < q^d - 1 \\ i \bmod b = j}} C_{A^i}(\lambda). \tag{4-4}$$

Using this, we can compute $\Gamma^j(\mu)$ for $0 \leqslant j \leqslant b - 1$ as a formal polynomial in λ with coefficients from R_T. We can divide this polynomial by the monic polynomial $C_M(\lambda)/\lambda$ (formally, over the polynomial ring $R_T[\lambda]$) and represent $\Gamma^j(\mu)$ as a polynomial of degree less than $(q^d - 1)$ in λ. Using this representation, we can compute the polynomials

$$h^{(i)}(Z) = \prod_{j=0}^{i} (Z - \Gamma^j(\mu)) \quad \text{for } 1 \leqslant i \leqslant b - 1$$

iteratively, as an element of $R_T[\lambda][Z]$, with all coefficients having degree less than $(q^d - 1)$ in λ. When $i = b - 1$, we would have computed $h(Z)$ — we know at the end all the coefficients will have degree 0 in λ and belong to R_T. □

Using $\prod_{j=0}^{b-1} \Gamma^j(\mu) = M$ from the above argument, and $v_{M'}(\Gamma^j(\mu)) = v_{M'}(\mu)$, we conclude that $v_{M'}(\mu) = 1$, that is, μ (as well as each of its Galois conjugates $\Gamma^j(\mu)$) is M'-prime. We record this fact below. It will useful to establish that the integral closure of R_T in E equals $R_T[\mu]$ (Proposition 5.1), a fact we will use en route characterizing the message space in Theorem 5.2.

Lemma 4.4. *The element μ has a simple zero at M', that is, $v_{M'}(\mu) = 1$.*

With the minimal polynomial $h(Z)$ of μ at our disposal, we turn to computing the evaluations of μ at the b places above $T - \beta$; call them $P_j^{(\beta)}$ for $j = 0, 1, \ldots, b - 1$, for each $\beta \in \mathbb{F}_r$. (Recall that the place $T - \beta$ splits completely in E/F by Proposition 4.2(v).) The following lemma identifies the set of evaluations of μ at these places. This method is related to Kummer's theorem on splitting of primes [Stichtenoth 1993, Section III.3].

Lemma 4.5. *Consider the polynomial $\bar{h}^{(\beta)}(Z) \in \mathbb{F}_q[Z]$ obtained by evaluating the coefficients of $h(Z)$, which are polynomials in T, at β. Then*

$$\bar{h}^{(\beta)}(Z) = \prod_{j=0}^{b-1}(Z - \mu(P_j^{(\beta)})).$$

In particular, the set of evaluations of μ at the places above $(T - \beta)$ equals the roots of $\bar{h}^{(\beta)}$ in $\mathbb{F}_q$, and can be computed in $b^{O(1)}$ time given $h \in R_T[Z]$.

Proof. We know $h(Z) = \prod_{j=0}^{b-1}(Z - \Gamma^j(\mu))$. Therefore

$$\bar{h}^{(\beta)}(Z) = \prod_{j=0}^{b-1}(Z - \Gamma^j(\mu)(P_0^{(\beta)})) = \prod_{j=0}^{b-1}\big(Z - \mu\big(\Gamma^{-j}(P_0^{(\beta)})\big)\big) = \prod_{j=0}^{b-1}(Z - \mu(P_j^{(\beta)})),$$

where the last step uses the fact that $\Gamma^{-j}(P_0^{(\beta)})$ for $j = 0, 1, \ldots, b-1$ is precisely the set of places above $T - \beta$. $\qquad\square$

5. Code construction from cyclotomic function fields

We will now describe the AG codes based on the function field E. A tempting choice for the message space is perhaps $\{\sum_{i=0}^{b-1} a_i(T)\mu^i\} \subset R_T[\mu]$, where $a_i(T)$ are polynomials of some bounded degree. This is certainly a $\mathbb{F}_q$-linear space and messages in this space have no poles outside the places lying above P_∞. However, the valuations of μ at these places are complicated — one needs the Newton polygon method to estimate them [Villa Salvador 2006, Section 12.4] — and since μ has both zeroes and poles among these places, it is hard to get good bounds on the total pole order of such messages at each of the places above P_∞.

5A. *Message space.* Let M' be the unique totally ramified place M' in E lying above M; $\deg M' = \deg M = d$. We will use as message space elements of $R_T[\mu]$ that have no more than a certain number ℓ of poles at the place M' and no poles elsewhere. These can equivalently be thought of (via a natural correspondence) as elements of E that have bounded (depending on ℓ) pole order at each place above P_∞, and no poles elsewhere, and we can develop our codes and algorithms in this equivalent setting. Since the literature on AG codes typically focuses on one-point codes where the messages have poles at a unique place, we work with functions with poles restricted to M'.

Formally, for an integer $\ell \geqslant 1$, let $\mathcal{L}(\ell M')$ be the space of functions in E that have no poles outside M' and at most ℓ poles at M'. $\mathcal{L}(\ell M')$ is an $\mathbb{F}_q$-vector space, and by the Riemann–Roch theorem, $\dim \mathcal{L}(\ell M') \geqslant \ell d - g + 1$, where $g = d(b-1)/2 + 1$ is the genus of E. We will assume that $\ell \geqslant b$, in which case $\dim \mathcal{L}(\ell M') = \ell d - g + 1$.

We will represent the code by a basis of $\mathcal{L}(\ell M')$ over $\mathbb{F}_q$. Of course, we first need to understand how to represent a single function in $\mathcal{L}(\ell M')$. Theorem 5.2 below suggests a representation for elements of $\mathcal{L}(\ell M')$ that we can use. Its proof uses the following claim, which can be established using Lemma 4.4 and an argument similar to the one used to prove that the integral closure of R_T in $K = F(\lambda)$ equals $R_T[\lambda]$ [Rosen 2002, Proposition 12.9].

Proposition 5.1. *The integral closure of R_T in E equals*

$$R_T[\mu] = \left\{ \sum_{i=0}^{b-1} a_i \mu^i \mid a_i \in R_T \right\}.$$

Theorem 5.2. *A function f in E with poles only at M' has a unique representation of the form*

$$f = \frac{\sum_{i=0}^{b-1} a_i \mu^i}{M^e}, \tag{5-1}$$

where $e \geqslant 0$ is an integer, each $a_i \in R_T$, and not all the a_i's are divisible by M (as polynomials in T).

Proof. If f has poles only at M', there must be a smallest integer $e \geqslant 0$ such that $M^e f$ has no poles outside the places above P_∞. This means that $M^e f$ must belong to the integral closure (*ring of integers*) of R_T in E, that is, the minimal polynomial of $M^e f$ over R_T is monic. By Proposition 5.1, we have $M^e f \in R_T[\mu]$ and so we can write $f = M^{-e} \sum_{i=0}^{b-1} a_i \mu^i$ as claimed. The uniqueness of the representation follows since $\{1, \mu, \ldots, \mu^{b-1}\}$ forms a basis of E over F. $\square$

5B. *Succinctness of representation.* In order to be able to efficiently compute with the representation (5-1) of functions in $\mathcal{L}(\ell M')$, we need the guarantee that the representation will be *succinct*, that is, of size polynomial in the code length.

We show that this will be the case by obtaining an upper bound on the degree of the coefficients $a_i \in R_T$ in Lemma 5.3 below. This is not as straightforward as one might hope, and we thank G. Anderson and D. Thakur for help with its proof. For the choice of parameters we will make (in Theorems 6.10 and 7.1), this upper bound will be polynomially bounded in the code length. Therefore, the assumed representation of the basis functions is of polynomial size.

Lemma 5.3. *Suppose $f \in \mathcal{L}(\ell M')$ is given by $f = M^{-e} \sum_{i=0}^{b-1} a_i \mu^i$ for $a_i \in R_T$ (not all divisible by M) and $e \geqslant 0$. Then the degree of each a_i is at most $\ell + q^d b$.*

Proof. Let $g = M^e f = \sum_{i=0}^{b-1} a_i \mu^i$. We know that g has at most eb poles at each place of E that lies above P_∞ (since f has no poles at these places). Using the fact that f has at most ℓ poles at M', and the uniqueness of the representation $f = M^{-e} \sum_{i=0}^{b-1} a_i \mu^i$, it is easy to argue that $eb \leqslant \ell + b$. So, g has at most $\ell + b$ poles at each place of E lying above P_∞.

Let $\sigma = \sigma_A$; we know that σ is a generator of $\mathrm{Gal}(E/F)$. For $j = 0, 1, \ldots, b-1$, we have $\sigma^j(g) = \sum_{i=0}^{b-1} a_i \sigma^j(\mu^i)$. Let $\boldsymbol{a} = (a_0, a_1, \ldots, a_{b-1})^T$ be the (column) vector of coefficients, and let $\boldsymbol{g} = (g, \sigma(g), \ldots, \sigma^{b-1}(g))^T$. Denoting by Φ the $b \times b$ matrix with $\Phi_{ji} = \sigma^j(\mu^i)$ for $0 \leqslant i, j \leqslant b-1$, we have the system of equations $\Phi \boldsymbol{a} = \boldsymbol{b}$.

We can thus determine the coefficients a_i by solving this linear system. By Cramér's rule, $a_i = \det \Phi_i / \det \Phi$, where Φ_i is obtained by replacing the i-th column of Φ by the column vector $\boldsymbol{g}$. The square of the denominator $\det \Phi$ is the discriminant of the field extension E/F, and belongs to R_T. Thus the degree of a_i is at most the pole order of $\det \Phi_i$ at an arbitrary place, say $\tilde{P}$, above P_∞. By the definition (4-3) of μ, and the fact that λ and its conjugates have at most one pole at the places above P_∞ in $F(\Lambda_M)$, it follows that μ has at most $(q^d - 1)/b$ poles at $\tilde{P}$. The same holds for all its conjugates $\sigma^j(\mu)$. The function g and its conjugates $\sigma^j(g)$ have at most $\ell + b$ poles at $\tilde{P}$. All in all, this yields a crude upper bound of

$$\frac{q^d - 1}{b} \frac{(b-1)b}{2} + \ell + b \leqslant \ell + q^d b$$

for the pole order of $\det \Phi_i$ at $\tilde{P}$, and hence also the degree of the polynomial $a_i \in R_T$. $\qquad \square$

5C. *Rational places for encoding and their ordering.* So far, the polynomial $A \in R_T$ was any monic irreducible polynomial that was a primitive element modulo M, so that its Artin automorphism σ_A generates $\mathrm{Gal}(E/F)$. We will now pick A to have degree D satisfying

$$D > \frac{\ell d}{b} \quad \text{and} \quad D > 3d, \tag{5-2}$$

where the latter condition (in fact even $D > 2d + o(d)$ suffices) ensures that there are at least $q^D/2Dq^d$ irreducible polynomials of degree D with any desired residue modulo M. This follows from the effective version of Dirichlet's theorem for polynomials; see for instance [Rosen 2002, Theorem 4.8].

For D satisfying the above conditions, an irreducible polynomial A of degree D that is primitive modulo M can be found by a Las Vegas algorithm in $(Dq^d)^{O(1)}$ time by picking a random polynomial and checking that it works, or deterministically by brute force in $q^{O(d+D)}$ time. Both of these bounds are within the decoding time claimed in Theorem 6.10, and will be polynomial in the block length for our parameter choices in Theorem 7.1. By Proposition 2.1, A remains inert in E/F, and let us denote by A' the unique place of E that lies over A. The degree of A' equals Db.

For each $\beta \in \mathbb{F}_r$, fix an arbitrary place $P_0^{(\beta)}$ lying above $T - \beta$ in E. For $j = 0, 1, \ldots, b - 1$, define

$$P_j^{(\beta)} = \sigma_A^{-j}(P_0^{(\beta)}) . \tag{5-3}$$

Since $\mathrm{Gal}(E/F)$ acts transitively on the set of primes above a prime, and σ_A generates $\mathrm{Gal}(E/F)$, these constitute all the places above $T - \beta$. Lemma 4.5 already tells us the *set* of evaluations of μ at these places, but not which evaluation corresponds to which point. We have $\mu(\sigma_A^{-j}(P_0^{(\beta)})) = \sigma_A^j(\mu)(P_0^{(\beta)})$; hence, to compute evaluations of μ at all these b places according to the ordering (5-3), it suffices to know

(i) the value at $\mu(P_0^{(\beta)})$, which we can find by simply picking one of the roots from Lemma 4.5 arbitrarily, and

(ii) a representation of $\sigma_A(\mu)$ as an element of $R_T[\mu]$ (since $\sigma_A(\mu)$ is integral over R_T, it belongs to $R_T[\mu]$ by virtue of Proposition 5.1). Note that $T(P_0^{(\beta)}) = \beta$, so once we know $\mu(P_0^{(\beta)})$, we can evaluate any element of $R_T[\mu]$ at $P_0^{(\beta)}$.

We now show that $\sigma_A(\mu) \in R_T[\mu]$ can be computed efficiently.

Lemma 5.4. (i) *The values of $\sigma_A^j(\mu)$ for $0 \leqslant j \leqslant b - 1$ as elements of $R_T[\mu]$ can be computed in $q^{O(d)}$ time.*

(ii) *The values $\mu(P_j^{(\beta)})$ for $\beta \in \mathbb{F}_r$ and $j = 0, 1, \ldots, b - 1$ can be computed in $q^{O(d)}$ time. Knowing these values, we can compute any function in the message space $\mathcal{L}(\ell M')$ represented in the form (5-1) at the places $P_j^{(\beta)}$ in $\mathrm{poly}(\ell, q^d)$ time.*

Proof. Part (ii) follows from (i) and the discussion above. To prove (i), note that once we compute $\sigma_A(\mu)$, we can recursively compute $\sigma_A^j(\mu)$ for $j \geqslant 2$, using the relation $h(\mu) = 0$ to replace μ^b and higher powers of μ in terms of $1, \mu, \ldots, \mu^{b-1}$.

By definition (4-3), we have $\mu = \prod_{0 \leqslant i < (q^d-1)/b} C_{A^{ib} \bmod M}(\lambda)$. Thus one can compute an expression

$$\mu = \sum_{i=0}^{q^d-2} e_i \lambda^i \in R_T[\lambda]$$

with coefficients $e_i \in R_T$ in $q^{O(d)}$ time. By successive multiplication in the ring $R_T[\lambda]$ (using the relation $C_M(\lambda) = 0$ to express λ^{q^d-1} and higher powers in terms of $1, \lambda, \ldots, \lambda^{q^d-2}$), we can compute, for $l = 0, 1, \ldots, b-1$, expressions

$$\mu^l = \sum_{i=0}^{q^d-2} e_{il} \lambda^i$$

with $e_{il} \in R_T$ in $q^{O(d)}$ time.

We have

$$\sigma_A(\mu) = \sum_{i=0}^{q^d-2} e_i \sigma_A(\lambda)^i = \sum_{i=0}^{q^d-2} e_i C_{A \bmod M}(\lambda)^i.$$

So one can likewise compute an expression $\sigma_A(\mu) = \sum_{i=0}^{q^d-2} f_i \lambda^i$ with $f_i \in R_T$ in $q^{O(d)}$ time. The task now is to rewrite this expression for $\sigma_A(\mu)$ as an element of $R_T[\mu]$, of the form $\sum_{l=0}^{b-1} a_l \mu^l$, for unknowns $a_l \in R_T$ that are to be determined. We will argue that this can be accomplished by solving a linear system.

Indeed, using the expressions $\mu^l = \sum_{i=0}^{q^d-2} e_{il} \lambda^i$, the coefficients a_l satisfy the following system of linear equations over R_T:

$$\sum_{l=0}^{b-1} e_{il} a_l = f_i \quad \text{for } i = 0, 1, \ldots, q^d - 2. \tag{5-4}$$

Since the representation $\sigma_A(\mu) = \sum_{l=0}^{b-1} a_l \mu^l$ is unique, the system has a unique solution. By Cramér's rule, the degree of each a_l is at most $q^{O(d)}$. Therefore, we can express the system (5-4) as a linear system of size $q^{O(d)}$ over $\mathbb{F}_q$ in unknowns the coefficients of all the polynomials $a_l \in R_T$. By solving this system in $q^{O(d)}$ time, we can compute the representation of $\sigma_A(\mu)$ as an element of $R_T[\mu]$. $\square$

5D. *The basic cyclotomic algebraic-geometric code.* The basic AG code $\mathscr{C}^0$ based on subfield E of the cyclotomic function field $F(\Lambda_M)$ is defined as

$$\mathscr{C}^0 = \left\{ \left(f(P_j^{(\beta)}) \right)_{\beta \in F_r, 0 \leqslant j < b} \mid f \in \mathscr{L}(\ell M') \right\}, \tag{5-5}$$

where the ordering of the places $P_j^{(\beta)}$ above $T - \beta$ is as in (5-3). We record the standard parameters of the above AG code, which follows from Riemann–Roch, the genus of E from Proposition 4.2, and that a nonzero $f \in \mathscr{L}(\ell M')$ can have at most $\ell \deg M' = \ell d$ zeroes.

Lemma 5.5. *Suppose $\ell \geqslant b$. Then $\mathscr{C}^0$ is an $\mathbb{F}_q$-linear code of block length $n = rb$, dimension $k = \ell d - d(b-1)/2$, and distance at least $n - \ell d$.*

Lemma 5.4(ii) implies the following.

Lemma 5.6 (Efficient encoding). *Given a basis for the message space $\mathscr{L}(\ell M')$ represented in the form (5-1), the generator matrix of the cyclotomic code $\mathscr{C}^0$ can be computed in $\mathrm{poly}(\ell, q^d, q^D)$ time.*

5E. *The folded cyclotomic code.* Let $m \geqslant 1$ be an integer. For convenience, we assume $m \mid b$ (though this is not really necessary). Analogously to the construction of folded RS codes [Guruswami and Rudra 2008], the folded cyclotomic code $\mathscr{C}$ is obtained from $\mathscr{C}^0$ by bundling together successive m-tuples of symbols into a single symbol to give a code of length $N = n/m$ over $\mathbb{F}_q^m$. Formally,

$$\mathscr{C} = \left\{ \left(f(P_{m\iota}^{(\beta)}), f(P_{m\iota+1}^{(\beta)}), \ldots, f(P_{m\iota+m-1}^{(\beta)}) \right)_{\beta \in F_r, 0 \leqslant \iota < b/m} \;\middle|\; f \in \mathscr{L}(\ell M') \right\}. \quad (5\text{-}6)$$

We will index the N positions of codewords in $\mathscr{C}$ by pairs (β, ι) for $\beta \in \mathbb{F}_r$ and $\iota \in \{0, 1, \ldots, (b/m) - 1\}$.

The generator matrix of unfolded code $\mathscr{C}^0$, which can be computed given a basis for $\mathscr{L}(\ell M')$ according to Lemma 5.6, obviously suffices for encoding. Later we will argue that the *same representation* also suffices for polynomial time list decoding.

5F. *Folding and Artin–Frobenius automorphism.* The unique place A' that lies above A has degree $D' \overset{\text{def}}{=} Db$. The residue field at A', denoted by $K_{A'}$, is isomorphic to $\mathbb{F}_{q^{D'}}$. By our choice $Db > \ell d$, this immediately implies that a message in $\mathscr{L}(\ell M')$ is uniquely determined by its evaluation at A'.

Lemma 5.7. *The map $\mathrm{ev}_{A'} : \mathscr{L}(\ell M') \to K_{A'}$ given by $\mathrm{ev}_{A'}(f) = f(A')$ is one-one.*

The key algebraic property of our folding is the following.

Lemma 5.8. *For every $f \in \mathscr{L}(\ell M')$:*

(i) *For every $\beta \in \mathbb{F}_r$ and $0 \leqslant j < b-1$, $\sigma_A(f)(P_j^{(\beta)}) = f(P_{j+1}^{(\beta)})$.*

(ii) *$\sigma_A(f)(A') = f(A')^{q^D}$.*

Proof. Part (i) follows since we ordered the places above $T - \beta$ such that $P_{j+1}^{(\beta)} = \sigma_A^{-1}(P_j^{(\beta)})$.

Part (ii) follows from the property of the Artin automorphism at A, since the norm of the place A equals $q^{\deg A} = q^D$. (A nice discussion of the Artin–Frobenius automorphism, albeit in the setting of number fields, appears in [Marcus 1977, Chapter 4].) $\qquad\square$

6. List decoding algorithm

We now turn to list decoding the folded cyclotomic code $\mathscr{C}$ defined in (5-6). The underlying approach is similar to that of the algorithm for list decoding folded RS codes [Guruswami and Rudra 2008] and AG generalizations of Parvaresh–Vardy codes [2005; Guruswami and Patthak 2008]. We will therefore not repeat the entire rationale and motivation behind the algorithm development. But our technical presentation and analysis is self-contained. In fact, our presentation here does offer some simplifications over previous descriptions of AG list decoding algorithms from [Guruswami and Sudan 1999; 2001; Guruswami and Patthak 2008]. A principal strength of the new description is that it *avoids the use of zero-increasing bases* at each code place $P_j^{(\beta)}$. This simplifies the algorithm as well as the representation of the code needed for decoding.

The list decoding problem for $\mathscr{C}$ up to e errors corresponds to solving the following function reconstruction problem. Recall that the length of the code is $N = n/m = rb/m$, and the codeword positions are indexed by $\mathbb{F}_r \times \{0, 1, \ldots, (b/m)-1\}$.

Input: Collection $\mathscr{T}$ of N tuples $\left(y_{m\iota}^{(\beta)}, y_{m\iota+1}^{(\beta)}, \ldots, y_{m\iota+m-1}^{(\beta)}\right) \in \mathbb{F}_q^m$ for $\beta \in \mathbb{F}_r$ and $0 \leqslant \iota < b/m$.

Output: A list of all $f \in \mathscr{L}(\ell M')$ whose encoding according to $\mathscr{C}$ agrees with the (β, ι)-th tuple for at least $N - e$ codeword positions.

6A. *Algorithm description.* We describe the algorithm at a high level below and later justify how the individual steps can be implemented efficiently, and under what condition the decoding will succeed. We stress that regardless of complexity considerations, even the *combinatorial* list-decodability property "proved" by the algorithm is nontrivial.

Algorithm List-Decode($\mathscr{C}$)

Parameters: • An integer parameter s, $2 \leqslant s \leqslant m$, for s-variate interpolation;
   • an integer parameter $w \geqslant 1$ that governs the zero order (multiplicity) guaranteed by interpolation; and
   • an integer parameter $\Delta \geqslant 1$ that is the total degree of the interpolated s-variate polynomial.

Step 1 (Interpolation): Find a nonzero polynomial $Q(Z_1, Z_2, \ldots, Z_s)$ of total degree at most Δ with coefficients in $\mathscr{L}(\ell M')$ such that for each $\beta \in \mathbb{F}_r$, $0 \leqslant \iota < b/m$, and $j' \in \{0, 1, \ldots, m-s\}$, the shifted polynomial

$$Q\left(Z_1 + y_{m\iota+j'}^{(\beta)}, Z_2 + y_{m\iota+j'+1}^{(\beta)}, \ldots, Z_s + y_{m\iota+j'+s-1}^{(\beta)}\right) \tag{6-1}$$

has the property that the coefficient of the monomial $Z_i^{n_1} Z_2^{n_2} \cdots Z_s^{n_s}$ vanishes at $P_{m\iota+j'}^{(\beta)}$ whenever its total degree $n_1 + n_2 + \cdots + n_s < w$.

Step 2 (Root-finding): Find a list of all $f \in \mathscr{L}(\ell M')$ satisfying

$$Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) = 0.$$

Output those whose encoding according to $\mathscr{C}$ agrees with at least $N - e$ of the m-tuples in $\mathscr{T}$.

6B. *Analysis of error correction radius.*

Lemma 6.1. *If $k(\Delta + 1)^s \geqslant N(m - s + 1)(w + s - 1)^s$ (where, as we recall, $k = \ell d - d(b - 1)/2$ is the dimension of $\mathscr{L}(\ell M')$), then a nonzero polynomial Q with the stated properties exists. If we know the evaluations of the functions in a basis $\{\phi_1, \phi_2, \ldots, \phi_k\}$ of $\mathscr{L}(\ell M')$ at the places $P_j^{(\beta)}$, then such a Q can be found by solving a homogeneous system of linear equations over $\mathbb{F}_q$ with at most $Nm(w + s)^s$ equations and unknowns.*

Proof. The proof is standard and follows by counting degrees of freedom versus number of constraints. One can express the desired polynomial as

$$\sum_{n_1, n_2, \ldots, n_s} q_{(n_1, \ldots, n_s)} Z_1^{n_1} \cdots Z_s^{n_s},$$

with unknowns $q_{(n_1, \ldots, n_s)} \in \mathbb{F}_q$. The number of coefficients is $k\binom{\Delta + s}{s} > k(\Delta + 1)^s / s!$. One can express for each place $P_{m\iota + j'}^{(\beta)}$ the required condition at that place by $\binom{w + s - 1}{s}$ linear conditions (this quantity is the number of monomials of total degree less than w), for a total of

$$N(m - s + 1)\binom{w + s - 1}{s} < N(m - s + 1)\frac{(w + s - 1)^s}{s!}$$

constraints. When the number of unknowns exceeds the number of constraints, a nonzero solution must exist. A solution can also be found efficiently once the linear system is set up, which can clearly be done if we know the evaluations of ϕ_i's at the code places (that is, a *generator matrix* of the code). $\square$

Lemma 6.2. *Let Q be the polynomial found in* [Step 1](). *If the encoding of some f as per $\mathscr{C}$ agrees with $(y_{m\iota}^{(\beta)}, y_{m\iota+1}^{(\beta)}, \cdots, y_{m\iota+m-1}^{(\beta)})$ for some position (β, ι), then $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f))$ has at least w zeroes at each of the $(m - s + 1)$ places $P_{m\iota+j'}^{(\beta)}$ for $j' = 0, 1, \ldots, m - s$.*

Proof. The proof differs slightly from earlier proofs of similar statements (for example, [Guruswami and Patthak 2008, Lemma 6.6]) in that it avoids the use of zero-increasing bases and is thus simpler. We will prove the claim for $j' = 0$, and the same proof works for any $j' \leqslant m - s$. Note that agreement on the m-tuple at position (b, ι) implies that

$$f(P_{m\iota}^{(\beta)}) = y_{m\iota}^{(\beta)}, \ f(P_{m\iota+1}^{(\beta)}) = y_{m\iota+1}^{(\beta)}, \ldots, f(P_{m\iota+s-1}^{(\beta)}) = y_{m\iota+s-1}^{(\beta)}.$$

By Lemma 5.8(i), this implies

$$f(P_{m\iota}^{(\beta)}) = y_{m\iota}^{(\beta)}, \sigma_A(f)(P_{m\iota}^{(\beta)}) = y_{m\iota+1}^{(\beta)}, \ldots, \sigma_{A^{s-1}}(f)(P_{m\iota}^{(\beta)}) = y_{m\iota+s-1}^{(\beta)}.$$

Denote by Q^* the shifted polynomial (6-1) for the triple $(\beta, \iota, 0)$. We have

$$\begin{aligned}
Q\big(f, \sigma_A(f), &\ldots, \sigma_{A^{s-1}}(f)\big) \\
&= Q^*\big(f - y_{m\iota}^{(\beta)}, \sigma_A(f) - y_{m\iota+1}^{(\beta)}, \cdots, \sigma_A^{s-1}(f) - y_{m\iota+s-1}^{(\beta)}\big) \\
&= \sum_{\substack{n_1, n_2, \ldots, n_s \\ w \leqslant n_1 + \cdots + n_s \leqslant \Delta}} q_{(n_1, \ldots, n_s)}^* \big(f - f(P_{m\iota}^{(\beta)})\big)^{n_1} \big(\sigma_A(f) - \sigma_A(f)(P_{m\iota}^{(\beta)})\big)^{n_2} \\
&\qquad\qquad\qquad\qquad \cdots \big(\sigma_{A^{s-1}}(f) - \sigma_{A^{s-1}}(f)(P_{m\iota}^{(\beta)})\big)^{n_s}.
\end{aligned}$$

for some coefficients $q_{(n_1, \ldots, n_s)}^* \in \mathbb{F}_q$. Since each term of the function in the last expression has valuation at least w at $P_{m\iota}^{(\beta)}$, so does $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f))$. $\square$

Lemma 6.3. *If the encoding of $f \in \mathcal{L}(\ell M')$ has at least $N - e$ agreements with the input tuples $\mathcal{T}$, and $(N - e)(m - s + 1)w > d\ell(\Delta + 1)$, then*

$$Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) = 0.$$

Proof. Since f has no poles outside M', neither do $\sigma_{A^i}(f)$ for $1 \leqslant i < s$. Moreover, $v_{M'}(\sigma_A(f)) = v_{\sigma_A^{-1}(M')}(f) = v_{M'}(f)$ (since M' is the unique place above M and is thus fixed by every Galois automorphism). Since $f \in \mathcal{L}(\ell M')$, this implies $\sigma_{A^i}(f) \in \mathcal{L}(\ell M')$ for every i. Since each coefficient of Q also belongs to $\mathcal{L}(\ell M')$, we conclude that $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) \in \mathcal{L}((\ell + \ell\Delta)M')$. On the other hand, by Lemma 6.2, $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f))$ has at least $(N-e)(m-s+1)w$ zeroes. If $(N - e)(m - s + 1)w > \ell(\Delta + 1)d$, then $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f))$ has more zeroes than poles and must thus equal 0. $\square$

Putting together the above lemmas, we can conclude the following about the list decoding radius guaranteed by the algorithm. Note that we have not yet discussed how Step 2 may be implemented, or why it implies a reasonable bound on the output list size. We will do this in Section 6C.

Theorem 6.4. *For every s, $2 \leqslant s \leqslant m$, and any $\zeta > 0$, for the choice $w = \lceil s/\zeta \rceil$ and a suitable choice of the parameter Δ, the algorithm* List-Decode($\mathcal{C}$) *successfully list decodes up to e errors whenever*

$$e < (N - 1) - (1 + \zeta)\left(\frac{k}{m - s + 1}\right)^{1 - 1/s} N^{1/s}\left(1 + \frac{d(b - 1)}{2k}\right). \qquad (6\text{-}2)$$

Proof. Picking $w = \lceil s/\zeta \rceil$ and

$$\Delta + 1 = \left\lceil \left(\frac{N(m - s + 1)}{k}\right)^{1/s} (w + s - 1) \right\rceil,$$

the requirement of Lemma 6.1 is met. By Lemma 5.5, the dimension k satisfies $\ell d = k + d(b - 1)/2$. A straightforward computation reveals that for this choice,

the bound (6-2) implies the decoding condition $(N-e)(m-s+1)w > \ell d(\Delta+1)$, under which Lemma 6.3 guarantees successful decoding. $\qquad\square$

Remark 6.5. The error correction radius above is nontrivial only when $s \geqslant 2$. We will see later how to pick parameters so that the error fraction approaches $1 - R^{1-1/s}$. For AG codes, even $s = 1$ led to a nontrivial guarantee of about $1 - \sqrt{R}$ in [Guruswami and Sudan 1999], and for folded RS codes the error fraction with s-variate interpolation was $1 - R^{s/(s+1)}$. The weaker bound we get is due to restricting the pole order of coefficients of Q to at most ℓ, the number of poles allowed for messages. This is similar to the algorithm in [Guruswami and Patthak 2008, Section 5]. Since we let grow s anyway, this does not hurt us. It also avoids some difficult technical complications that would arise otherwise (discussed in, for example, [Guruswami and Patthak 2008]), and allows us to implement the interpolation step just using the natural generator matrix of the code.

6C. *Root-finding using the Artin automorphism.* So far we have not discussed how Step 2 of decoding can be performed, and why in particular it implies a reasonably small upper bound on the number of solutions $f \in \mathscr{L}(\ell M')$ that it may find in the worst case. We address this now. This is where the properties of the Artin automorphism σ_A will play a crucial role. Recall that $K_{A'} = \mathbb{O}_{A'}/A'$ denotes the residue field at the place A' of E lying above A, and that we picked A so that $D = \deg A$ obeyed $Db > \ell d$.

Lemma 6.6. *Suppose* $f \in \mathbb{O}_{A'}$ *satisfies*

$$Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) = 0$$

for some $Q \in \mathbb{O}_{A'}[Z_1, Z_2, \ldots, Z_s]$. *Let* $\overline{Q} \in K_{A'}[Z_1, Z_2, \ldots, Z_s]$ *be the polynomial obtained by reducing the coefficients of Q modulo A'. Then* $f(A') \in K_{A'}$ *obeys*

$$\overline{Q}\big(f(A'), f(A')^{q^D}, f(A')^{q^{2D}}, \ldots, f(A')^{q^{D(s-1)}}\big) = 0. \tag{6-3}$$

Proof. If $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) = 0$, then surely

$$\overline{Q}\big(f(A'), \sigma_A(f)(A'), \ldots, \sigma_{A^{s-1}}(f)(A')\big) = 0.$$

The claim (6-3) now follows immediately from Lemma 5.8(ii). $\qquad\square$

Lemma 6.7. *If $Q(Z_1, \ldots, Z_s)$ is a nonzero polynomial of total degree at most $\Delta < q^D$ all of whose coefficients belong to $\mathscr{L}(\ell M')$, then the polynomial $\Phi \in K_{A'}[Y]$ defined as*

$$\Phi(Y) \overset{\text{def}}{=} \overline{Q}\big(Y, Y^{q^D}, \ldots, Y^{q^{D(s-1)}}\big)$$

is a nonzero polynomial of degree at most $\Delta \cdot q^{D(s-1)}$.

Proof. If $\psi \in \mathscr{L}(\ell M')$ is nonzero, then $\psi(A') \neq 0$. (Otherwise, the degree of the zero divisor of ψ will be at least $\deg A' = bD > \ell d$, and thus exceed the degree of the pole divisor of ψ.) It follows that if $Q \neq 0$, then $\overline{Q}(Z_1, \ldots, Z_s)$ obtained by reducing coefficients of Q modulo A' is also nonzero.[5] Since the degree of $\overline{Q}$ in each Z_i is at most $\Delta < q^D$, it is easy to see that $\Phi(Y) = \overline{Q}(Y, Y^{q^D}, \ldots, Y^{q^{D(s-1)}})$ is also nonzero. The degree of Φ is at most $q^{D(s-1)}$ times the total degree of $\overline{Q}$, which is at most Δ. $\qquad\square$

By the above two lemmas, we see that one can compute the set of residues $f(A')$ of all f satisfying $Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) = 0$ by computing the roots in $K_{A'}$ of $\Phi(Y)$. Since $\mathrm{ev}_{A'}$ is injective on $\mathscr{L}(\ell M')$ (Lemma 5.7), this also lets us recover the message $f \in \mathscr{L}(\ell M')$.

Lemma 6.8. *Given a nonzero polynomial $Q(Z_1, \ldots, Z_s)$ with coefficients from $\mathscr{L}(\ell M')$ and degree $\Delta < q^D$, the set of functions*

$$\mathscr{S} = \{f \in \mathscr{L}(\ell M') \mid Q(f, \sigma_A(f), \ldots, \sigma_{A^{s-1}}(f)) = 0\}$$

has cardinality at most q^{Ds}.

Moreover, knowing the evaluations of a basis $\mathscr{B} = \{\phi_1, \phi_2, \ldots, \phi_k\}$ of $\mathscr{L}(\ell M')$ at the place A', one can compute the coefficients expressing each $f \in \mathscr{S}$ in the basis $\mathscr{B}$ in $q^{O(Ds)}$ time.

Proof. As argued above, any desired $f \in \mathscr{L}(\ell M')$ has the property that $\Phi(f(A')) = 0$, so the evaluations of functions in $\mathscr{S}$ take at most $\mathrm{degree}(\Phi) \leqslant \Delta q^{D(s-1)} \leqslant q^{Ds}$ values. Since $\mathrm{ev}_{A'}$ is injective on $\mathscr{S}$, this implies $|\mathscr{S}| \leqslant q^{Ds}$. The second part follows since we can compute the roots of Φ in $K_{A'}$ in time $\mathrm{poly}(q^{Ds}, \log|K_{A'}|) \leqslant q^{O(Ds)}$. Knowing $f(A')$, we can recover f (in terms of the basis $\mathscr{B}$) by solving a linear system if we know the evaluations of the functions in the basis $\mathscr{B}$ at A'. The next section discusses a convenient representation for computations in $K_{A'}$. $\qquad\square$

6C.1. *Representation of the residue field $K_{A'}$.* The following gives a convenient representation for elements of $K_{A'}$ which can be used in computations involving this field.

Lemma 6.9. *The elements $\{1, \mu(A), \ldots, \mu(A)^{b-1}\}$ form a basis for $K_{A'}$ over the field $R_T/(A) \simeq \mathbb{F}_{q^D}$. In other words, elements of $K_{A'}$ can be expressed in a unique way as*

$$\sum_{i=0}^{b-1} b_i(T) \mu(A)^i,$$

where each $b_i \in R_T$ has degree less than D.

---

[5]This is simplicity we gain by restricting the coefficients of Q to also belong to $\mathscr{L}(\ell M')$.

Proof. Since A is inert in E/F, the minimal polynomial $h(Z)$ of μ over F has the property that $\bar{h}(Z)$, obtained by reducing the coefficients of h modulo A, is irreducible over the residue field $R_T/(A)$. Thus $\mu(A)$ generates $K_{A'}$ over $R_T/(A)$, and in fact the minimal polynomial of $\mu(A)$ with respect to K_A equals $\bar{h}(Z)$. Note that the coefficients of $\bar{h}$, which belong to $R_T/(A)$, have a natural representation as a polynomial in R_T of degree less than $\deg A = D$. $\qquad\square$

We note that given the representation of the basis $\mathcal{B} = \{\phi_1, \phi_2, \dots, \phi_k\}$ in the form guaranteed by Theorem 5.2, one can trivially compute the evaluations of $\phi_i(A')$ in the above form. There is no need to explicitly compute $\mu(A) \in \mathcal{O}_A/A$. Therefore, the decoding algorithm requires no additional preprocessed information beyond a basis for the message space $\mathcal{L}(\ell M')$ — the rest can all be computed efficiently from the basis alone.

6D. *Wrap-up.* We are now ready to state our final decoding claim.

Theorem 6.10. *For any s, $2 \leqslant s \leqslant m$, and $\zeta > 0$, the folded cyclotomic code $\mathcal{C} \subseteq (\mathbb{F}_q^m)^N$ defined in (5-6) can be list decoded in time $(Nm)^{O(1)}(s/\zeta)^{O(s)} + q^{O(Ds)}$ from a fraction ρ of errors*

$$\rho = 1 - (1+\zeta)\Big(\frac{R_0 m}{m - s + 1}\Big)^{1 - 1/s}\Big(1 + \frac{d}{2R_0 r}\Big), \tag{6-4}$$

where $R_0 = k/n$ is the rate of the code. The size of the output list is at most q^{Ds}. The decoding algorithm assumes polynomial amount of preprocessed information consisting of basis functions $\{\phi_1, \dots, \phi_k\}$ for the message space $\mathcal{L}(\ell M')$ represented in the form (5-1). (This is the same *representation used for encoding, and it is succinct by Lemma 5.3.)*

Proof. We first note that bound on fraction of errors follows from Theorem 6.4, and the fact that $k = R_0 n = R_0 Nm = R_0 br$. By Lemma 6.1 and its proof, in Step 1 of the algorithm we can find a nonzero polynomial Q (of degree less than q^D) such that for any $f \in \mathcal{L}(\ell M')$ that needs to be output by the list decoder, we must have $Q(f, \sigma_A(f), \dots, \sigma_{A^{s-1}}(f)) = 0$. We can evaluate the basis functions ϕ_i at $P_j^{(\beta)}$ in $(\ell q^d)^{O(1)}$ time by Lemma 5.4, and with this information, the running time of this interpolation step can be bounded by $(Nm)^{O(1)}(w + s)^{O(s)} = (Nm)^{O(1)}(s/\zeta)^{O(s)}$ (since $w = O(s/\zeta)$). We can also efficiently compute the evaluations of ϕ_i at A' in the representation suggested by Lemma 6.9. Therefore, by Lemma 6.8, we can then find a list of the at most q^{Ds} functions f satisfying $Q(f, \sigma_A(f), \dots, \sigma_{A^{s-1}}(f)) = 0$ in $q^{O(Ds)}$ time. $\qquad\square$

Remark 6.11 (List recovery). A similar claim holds for the more general *list recovery* problem, where for each position we are given as input a set of up to l elements of $\mathbb{F}_q^m$, and the goal is to find all codewords which agree with some element of the input sets for at least a fraction $(1 - \rho)$ of positions. In this case, $1 - \rho$ only needs

to be only a factor $l^{1/s}$ larger than the bound (6-4). By picking $s \gg l$, the effect of l can be made negligible. This feature is very useful in concatenation schemes; see Section 7A and [Guruswami and Rudra 2008] for further details.

7. Long codes with optimal rate for list decoding

We now describe the parameter choices which lead to capacity-achieving list-decodable codes, that is, codes of rate R_0 that can correct a fraction $1 - R_0 - \varepsilon$ of errors (for any desired $0 < R_0 < 1$), and whose alphabet size is polylogarithmic in the block length; the formal statement appears in Theorem 7.1 below. (Recall that for folded RS codes, the alphabet size is a large polynomial in the block length.) Using concatenation and expander-based ideas, [Guruswami and Rudra 2008] also presents capacity-achieving codes over a fixed alphabet size (that depends on the distance ε to capacity alone). The advantage of our codes is that they inherit strong list recovery properties similar to the folded RS codes (Remark 6.11). This is very useful in concatenation schemes, and indeed our codes can be used as outer codes for an explicit family of binary concatenated codes list-decodable up to the Zyablov radius, *with no brute-force search* for the inner code (see Section 7A below).

We now describe our main result on how to obtain the desired codes from the construction $\mathscr{C}$ and Theorem 6.10. The underlying parameter choices to achieve this require a fair bit of care.

Theorem 7.1 (Main theorem). *For every R_0, $0 < R_0 < 1$, and every constant $\varepsilon > 0$, the following holds for infinitely many integers $\mathbf{q}$ which are powers of two. There is a code of rate at least R_0 over an alphabet of size $\mathbf{q}$ with block length*

$$N \geqslant 2^{\mathbf{q}^{\Omega(\varepsilon^2 / \log(1/R_0))}}$$

that can be list decoded up to a fraction $1 - R_0 - \varepsilon$ of errors in time bounded by $(N \log(1/R_0)/\varepsilon^2)^{O(1/(R_0\varepsilon)^2)}$.

Proof. Suppose R_0, $0 < R_0 < 1$, and $\varepsilon > 0$ are given. Let $c = 2\lfloor \frac{10}{R_0\varepsilon} \rfloor + 1$, and let $\phi(c)$ denote the Euler's totient function of c.

Let $u \geqslant 1$ be an arbitrary integer; we will get a family of codes by varying u. The code we construct will be a folded cyclotomic code $\mathscr{C}$ defined in (5-6). Let $x = \phi(c)u$. Note that $2^x \equiv 1 \pmod{c}$. We first pick q, r, d as follows: $r = 2^x$, $q = r^2$, and $d = (2^x - 1)/c$. For this choice, $d | r - 1$ and $(q - 1)/(r - 1) = r + 1$ is coprime to d, as required in Lemma 4.1. So we can take $M(T) = T^d - \gamma \in \mathbb{F}_r[T]$ for γ primitive in $\mathbb{F}_r$ as the irreducible polynomial over $\mathbb{F}_q$.

For the above choice, $d/r < 1/c \leqslant \varepsilon R_0/20$, so that $d/2R_0r < \varepsilon/10$. By picking

$$s = \Theta(\varepsilon^{-1} \log(1/R_0)), \quad m = \Theta(s/\varepsilon), \quad \text{and} \quad \zeta = \varepsilon/20,$$

we can ensure that the decoding radius ρ guaranteed in (6-4) by Theorem 6.10 is at least $1 - (1 + \varepsilon) R_0$.

The degree b of the extension E/F, introduced in (4-1), is given by

$$b = (r^d + 1)/(r + 1).$$

The length of the unfolded cyclotomic code $\mathscr{C}^0$ (defined in (5-5)) equals $n = rb > r^d/2$. We need to ensure that the rate of $\mathscr{C}^0$, which is equal to the rate of the folded cyclotomic code $\mathscr{C}$, is at least R_0. To this end, we will pick

$$\ell = \left\lceil \frac{b}{2} + \frac{R_0 r b}{d} \right\rceil. \tag{7-1}$$

It is easily checked that for our choice of parameters, $\ell \geqslant b$. By Lemma 5.5, the rate of $\mathscr{C}^0$ equals $d(\ell - (b-1)/2)/rb$, which is at least R_0 for the above choice of ℓ.

We next pick the value of D, the degree of the irreducible A, which is the key quantity governing the list size and decoding complexity. To satisfy the condition (5-2), we need $D > \max\{\ell d/b, 3d\}$. For the ℓ chosen above, this condition is surely met if $D > 3r$. We can thus pick

$$D = \Theta(r) = \Theta(dc) = \Theta(d/(R_0 \varepsilon)) .$$

The running time of the list decoding algorithm is dominated by the $q^{O(Ds)}$ term, and for the above choice of parameters can be bounded by $q^{O(d/(R_0\varepsilon)^2)}$. The block length of the code N satisfies

$$N = \frac{n}{m} > \frac{r^d}{2m} = \frac{q^{d/2}}{2m} = \Omega\left(\frac{\varepsilon^2 q^{d/2}}{\log(1/R_0)} \right).$$

As a function of N, the decoding complexity is therefore bounded by

$$(N \log(1/R_0)/\varepsilon^2)^{O(1/(R_0\varepsilon)^2)}.$$

The alphabet size of the folded cyclotomic code is $\mathbf{q} = q^m$, and we can bound the block length N from below as a function of $\mathbf{q}$ as:

$$
\begin{aligned}
N &\geqslant \frac{q^{d/2}}{2m} \geqslant \frac{q^{\Omega(r/c)}}{2m} \geqslant \frac{q^{\Omega(\varepsilon R_0 \sqrt{q})}}{2m} \\
&\geqslant 2^{\sqrt{q}} \quad \text{(for large enough } q \text{ compared to } 1/R_0, 1/\varepsilon) \\
&= 2^{\mathbf{q}^{1/(2m)}} \geqslant 2^{\mathbf{q}^{\Omega(\varepsilon^2/\log(1/R_0))}}.
\end{aligned}
$$

This establishes the claimed lower bound on block length, and completes the proof of the theorem. □

7A. *Concatenated codes list-decodable up to Zyablov radius.* Using the strong list recovery property of folded RS codes, a polynomial time construction of binary codes list-decodable up to the Zyablov radius was given in [Guruswami and Rudra 2008, Theorem 5.3]. The construction used folded RS codes as outer codes in a concatenation scheme, and involved an undesirable brute-force search to find a binary inner code that achieves list decoding capacity. The time to construct the code grew faster than $N^{\Omega(1/\varepsilon)}$, where ε is the distance of the decoding radius to the Zyablov radius. This result as well as our result below hold not only for binary codes but also codes over any fixed alphabet; for sake of clarity, we state results only for binary codes.

As the folded cyclotomic codes from Theorem 7.1 are much longer than the alphabet size, by using them as outer codes, it is possible to achieve a similar result without having to search for an inner code, by using as inner codes *all possible binary linear codes* of a certain rate!

Theorem 7.2. *Let $0 < R_0, r < 1$ and $\varepsilon > 0$. Let $\mathscr{C}$ be a folded cyclotomic code guaranteed by Theorem 7.1 with rate at least R_0 and a large enough block length N. Let $\mathscr{C}^*$ be a binary code obtained by concatenating $\mathscr{C}$ with all possible binary linear maps of rate r (each one used a roughly equal number of times). Then $\mathscr{C}^*$ is a binary linear code of rate at least $R_0 \cdot r$ that can be list decoded from a fraction $(1 - R_0)H^{-1}(1 - r) - \varepsilon$ of errors in $N^{(1/\varepsilon)^{O(1)}}$ time.*

We briefly discuss the idea behind proving the above claim. As the alphabet size of folded cyclotomic codes is polylogarithmic in N, each outer codeword symbol can be expressed using $O_\varepsilon(\log \log N)$ bits. Hence the total number of such inner codes S will be at most $2^{O_\varepsilon((\log \log N)^2)} \ll N$ for large enough N. The N outer codeword positions will be partitioned into S (roughly) equal parts in an arbitrary way, and each inner code used to encode all the outer codeword symbols in one of the parts. Most of the inner codes achieve list decoding capacity — if their rate is r, they can list decode $H^{-1}(1 - r) - \varepsilon$ fraction of errors with constant sized lists (of size $2^{O(1/\varepsilon)}$). This suffices for analyzing the standard algorithm for decoding concatenated codes (namely, list decode the inner codes to produce a small set of candidate symbols for each position, and then list recover the outer code based on these sets). Arguing as in [Guruswami and Rudra 2008, Theorem 5.3], we can thus prove Theorem 7.2.

Appendix: List of parameters

Since the construction of the cyclotomic function field and the associated error-correcting code used a large number of parameters, we summarize them below for easy reference.

We begin by recalling the parameters concerning the function field construction.

q	size of the ground finite field
r	size of the subfield $\mathbb{F}_r \subset \mathbb{F}_q$
F	the field $\mathbb{F}_q(T)$ of rational functions
R_T	the ring of polynomials $\mathbb{F}_q[T]$
P_∞	the place of F that is the unique pole of T
M	polynomial $T^d - \gamma \in \mathbb{F}_r[T]$, irreducible over $\mathbb{F}_q$
d	degree of the irreducible polynomial M
C_M	the Carlitz action corresponding to M
Λ_M	the M-torsion points in F^{ac} under the action C_M
K	the cyclotomic function field $F(\Lambda_M)$
λ	nonzero element of Λ_M that generates K over F; $K = F(\lambda)$
G	the Galois group of K/F, naturally isomorphic to $(R_T/(M))^*$
H	the subgroup $\mathbb{F}_q^* \cdot \mathbb{F}_r[T]$ of G
E	the fixed field K^H of H
μ	primitive element for E/F; $E = F(\mu)$
b	the degree $[E:F]$ of the extension E/F
g	the genus of E/F, equals $d(b-1)/2 + 1$

The construction of the code $\mathcal{C}^0$ from lrefeqbasic-cycl and its folded version $\mathcal{C}$ from lrefeqcode-def used further parameters, listed here:

M'	the unique place of E lying above M
ℓ	maximum pole order at M' of message functions; $\ell \geqslant b$
$\mathcal{L}(\ell M')$	$\mathbb{F}_q$-linear space of messages of the codes
n	block length of $\mathcal{C}^0$, $n = br$
k	dimension of the $\mathbb{F}_q$-linear code $\mathcal{C}$, $k = \ell d - g + 1$
m	folding parameter
N	block length of folded code $\mathcal{C}$, $N = n/m$
$P_j^{(\beta)}$	the rational places lying above $T - \beta$ in E, for $\beta \in \mathbb{F}_r$ and $0 \leqslant j < b$
A	an irreducible polynomial (place of F) that remains inert in E/F
D	the degree of the polynomial A; satisfies $Db > \ell d$
σ_A	the Artin automorphism of the extension E/F at A
A'	the unique place of E lying above A

Acknowledgments

Many thanks to Dinesh Thakur for several illuminating discussions about Carlitz–Hayes theory and cyclotomic function fields. I thank D. Thakur and Greg Anderson for helping me with the proof of Lemma 5.3. I am grateful to Andrew Granville for pointing me to the effective version of Dirichlet's theorem for polynomials in arithmetic progressions discussed in [Rosen 2002].

References

[Carlitz 1938] L. Carlitz, "A class of polynomials", *Trans. Amer. Math. Soc.* **43**:2 (1938), 167–182. MR 1501937 Zbl 0018.19806

[Frey et al. 1992] G. Frey, M. Perret, and H. Stichtenoth, "On the different of abelian extensions of global fields", pp. 26–32 in *Coding theory and algebraic geometry* (Luminy, 1991), edited by H. Stichtenoth and M. A. Tsfasman, Lecture Notes in Math. **1518**, Springer, Berlin, 1992. MR 93h:11129 Zbl 0776.11067

[Garcia and Stichtenoth 1995] A. Garcia and H. Stichtenoth, "A tower of Artin–Schreier extensions of function fields attaining the Drinfeld–Vlăduţ bound", *Invent. Math.* **121**:1 (1995), 211–222. MR 96d:11074 Zbl 0822.11078

[Garcia and Stichtenoth 1996] A. Garcia and H. Stichtenoth, "On the asymptotic behaviour of some towers of function fields over finite fields", *J. Number Theory* **61**:2 (1996), 248–273. MR 97i:11067 Zbl 0893.11047

[Guruswami 2004] V. Guruswami, *List decoding of error-correcting codes*, Lecture Notes in Computer Science **3282**, Springer, Berlin, 2004. Zbl 1075.94001

[Guruswami 2007] V. Guruswami, "Algorithmic Results in List Decoding", *Foundations and Trends in Theoretical Comp. Sci.* **2**:2 (2007). Zbl 05318520

[Guruswami and Patthak 2008] V. Guruswami and A. C. Patthak, "Correlated algebraic-geometric codes: improved list decoding over bounded alphabets", *Math. Comp.* **77**:261 (2008), 447–473. MR 2009a:94054 Zbl 1150.94013

[Guruswami and Rudra 2008] V. Guruswami and A. Rudra, "Explicit codes achieving list decoding capacity: error-correction with optimal redundancy", *IEEE Trans. Inform. Theory* **54**:1 (2008), 135–150. MR 2010b:94096

[Guruswami and Sudan 1999] V. Guruswami and M. Sudan, "Improved decoding of Reed–Solomon and algebraic-geometry codes", *IEEE Trans. Inform. Theory* **45**:6 (1999), 1757–1767. MR 2000j:94033 Zbl 0958.94036

[Guruswami and Sudan 2001] V. Guruswami and M. Sudan, "On representations of algebraic-geometry codes", *IEEE Trans. Inform. Theory* **47**:4 (2001), 1610–1613. MR 2002b:94046 Zbl 1002.94041

[Hayes 1974] D. R. Hayes, "Explicit class field theory for rational function fields", *Trans. Amer. Math. Soc.* **189** (1974), 77–91. MR 48 #8444 Zbl 0292.12018

[Huang and Narayanan 2008] M.-D. Huang and A. K. Narayanan, "Folded algebraic geometric codes from Galois extensions", 2008. Personal communication.

[Justesen 1972] J. Justesen, "A class of constructive asymptotically good algebraic codes", *IEEE Trans. Information Theory* **IT-18** (1972), 652–656. MR 52 #5190 Zbl 0256.94008

[Lidl and Niederreiter 1986] R. Lidl and H. Niederreiter, *Introduction to finite fields and their applications*, Cambridge University Press, 1986. MR 88c:11073 Zbl 0629.12016

[Marcus 1977] D. A. Marcus, *Number fields*, Springer, New York, 1977. MR 56 #15601 Zbl 0383. 12001

[Niederreiter and Xing 1996] H. Niederreiter and C. Xing, "Explicit global function fields over the binary field with many rational places", *Acta Arith.* **75**:4 (1996), 383–396. MR 97d:11177 Zbl 0877.11065

[Niederreiter and Xing 1997] H. Niederreiter and C. Xing, "Cyclotomic function fields, Hilbert class fields, and global function fields with many rational places", *Acta Arith.* **79**:1 (1997), 59–76. MR 97m:11141 Zbl 0891.11057

[Parvaresh and Vardy 2005] F. Parvaresh and A. Vardy, "Correcting errors beyond the Guruswami– Sudan radius in polynomial time", pp. 285–294 in *Proceedings of the 43nd Annual Symposium on Foundations of Computer Science* (Pittsburgh, PA, 2005), IEEE, 2005.

[Quebbemann 1988] H.-G. Quebbemann, "Cyclotomic Goppa codes", *IEEE Trans. Inform. Theory* **34**:5, part 2 (1988), 1317–1320. MR 90b:11135 Zbl 0665.94014

[Rosen 2002] M. Rosen, *Number theory in function fields*, Graduate Texts in Mathematics **210**, Springer, New York, 2002. MR 2003d:11171 Zbl 1043.11079

[Shen 1993] B.-Z. Shen, "A Justesen construction of binary concatenated codes that asymptoti- cally meet the Zyablov bound for low rate", *IEEE Trans. Inform. Theory* **39**:1 (1993), 239–242. MR 93m:94046 Zbl 0766.94022

[Stichtenoth 1993] H. Stichtenoth, *Algebraic function fields and codes*, Springer, Berlin, 1993. MR 94k:14016 Zbl 0816.14011

[Stichtenoth 2006] H. Stichtenoth, "Transitive and self-dual codes attaining the Tsfasman–Vlăduţ– Zink bound", *IEEE Trans. Inform. Theory* **52**:5 (2006), 2218–2224. MR 2006m:94126

[Sudan 1997] M. Sudan, "Decoding of Reed Solomon codes beyond the error-correction bound", *J. Complexity* **13**:1 (1997), 180–193. MR 98f:94024 Zbl 0872.68026

[Villa Salvador 2006] G. D. Villa Salvador, *Topics in the theory of algebraic function fields*, Birk- häuser, Boston, 2006. MR 2007i:11002 Zbl 1154.11001

guruswami@cmu.edu   *Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, United States*
http://www.cs.cmu.edu/~venkatg

Algebraic properties of generic tropical varieties

Tim Römer and Kirsten Schmitz

We show that the algebraic invariants multiplicity and depth of the quotient ring S/I of a polynomial ring S and a graded ideal $I \subset S$ are closely connected to the fan structure of the generic tropical variety of I in the constant coefficient case. Generically the multiplicity of S/I is shown to correspond directly to a natural definition of multiplicity of cones of tropical varieties. Moreover, we can recover information on the depth of S/I from the fan structure of the generic tropical variety of I if the depth is known to be greater than 0. In particular, in this case we can see if S/I is Cohen–Macaulay or almost-Cohen–Macaulay from the generic tropical variety of I.

1. Introduction

As a very new area of mathematics, tropical geometry has received a lot of attention, from various points of view, in the last few years; see [Develin and Sturmfels 2004; Katz et al. 2008; Mikhalkin 2006; Speyer and Sturmfels 2004; Gathmann 2006; Itenberg et al. 2007] for review articles. One approach to tropical geometry, which provides an effective tool for studying questions in algebraic geometry, is to associate a combinatorial object to a projective algebraic variety; see for example [Draisma 2008; Gathmann and Markwig 2008]. More precisely, the tropical variety $T(X)$ of an algebraic variety X is the real-valued image of X under some valuation map [Draisma 2008; Jensen et al. 2008; Speyer and Sturmfels 2004]. In certain settings, $T(X)$ has the structure of a polyhedral complex [Bieri and Groves 1984; Jensen 2007], and there is a practical characterization in terms of initial ideals given in [Speyer and Sturmfels 2004; Draisma 2008, Theorem 4.2]. If the valuation on the ground field is trivial, $T(X)$ is a subfan of the Gröbner fan of the ideal I defining X. We only consider this constant coefficient case, and we define the tropical variety as a fan associated to I instead of X. In this situation the ideal I need not be a radical ideal. So let K be an infinite field and $K[x_1, \ldots, x_n]$ be the

polynomial ring in n variables over K. In this setting, the tropical variety $T(I)$ of a graded ideal $I \subset K[x_1, \ldots, x_n]$ is defined to be the subfan of the Gröbner fan of I that consists of all cones such that the corresponding initial ideal does not contain a monomial.

The tropical variety of an ideal depends on the choice of coordinates in the following sense. For $g \in \mathrm{GL}_n(K)$, the image $g(I)$ of a graded ideal $I \subset S = K[x_1, \ldots, x_n]$ is also a graded ideal, and all important algebraic invariants of S/I, such as the dimension, multiplicity and depth, are preserved under g. In fact, $g(I)$ can be considered as the ideal I given in different coordinates. In general, for $I \subset K[x_1, \ldots, x_n]$ and $g \in \mathrm{GL}_n(K)$, we have

$$T(I) \neq T(g(I)).$$

We can, however, find a nonempty Zariski-open set $U \subset \mathrm{GL}_n(K)$ such that $T(g(I))$ is the same fan $\mathrm{gT}(I)$ for every $g \in U$ [Römer and Schmitz 2009, Corollary 6.7]. The fact that U is dense in $\mathrm{GL}_n(K)$ justifies the name *generic tropical variety of I* for this fan. In Corollary 7.4 of the same reference it was shown that $\mathrm{gT}(I)$ as a set depends only on the dimension of S/I. More precisely, if S/I is m-dimensional, the underlying set of $\mathrm{gT}(I)$ is always the m-skeleton of a particular complete fan $\mathcal{W}_n$ in $\mathbb{R}^n$. We show that under certain conditions, we can recover information on the depth of S/I in addition to the dimension from the fan structure of $\mathrm{gT}(I)$ induced by the Gröbner fan. As one of the main results, we can completely describe generic tropical varieties as fans if S/I is Cohen–Macaulay or almost-Cohen–Macaulay. With this we can determine if S/I is Cohen–Macaulay or almost-Cohen–Macaulay from the fan structure of the generic tropical variety of I if we know the depth of S/I to be greater than 0. Moreover, we show that the multiplicities associated with the maximal cones of $\mathrm{gT}(I)$ as done in [Dickenstein et al. 2007] correspond directly to the multiplicity of S/I.

Our paper is organized as follows. In Section 2 we introduce basic results and necessary notation. In Section 3 we show that for an m-dimensional ring S/I for a graded ideal I, the generic tropical variety is always a subfan of the m-skeleton of $\mathcal{W}_n$ by showing that the fan structure induced by $\mathcal{W}_n$ is the coarsest possible on the underlying set. This will be important in all following sections. Sections 4 and 5 are devoted to the depth of S/I. In Section 4 we show that $\mathrm{gT}(I)$ is equal to the m-skeleton of $\mathcal{W}_n$ if and only if S/I is Cohen–Macaulay or almost-Cohen–Macaulay, where $\dim(S/I) = m$. In Section 5 we show that we can recover the depth of S/I from $\mathrm{gT}(I)$ if we know it to be greater than 0 and less than $\dim(S/I) - 1$. We also give more structural results depending on $\mathrm{depth}(S/I)$ on $\mathrm{gT}(I)$ as a fan for a special class of ideals. We show in Section 6 that the multiplicities defined on the maximal cones of $T(I)$ as in [Dickenstein et al. 2007] generically behave in a nice way. These multiplicities coincide with the multiplicity of S/I.

2. Preliminaries

Let K be an algebraically closed field of characteristic 0 and let $S = K[x_1, \ldots, x_n]$ be the polynomial ring in n variables over K. The ω-*weight* $\mathrm{wt}_\omega(cx^\nu)$ of some term $cx^\nu = cx_1^{\nu_1} \cdots x_n^{\nu_n} \in S$ is defined as $\mathrm{wt}_\omega(cx^\nu) = \omega \cdot \nu$ for any $\omega \in \mathbb{R}^n$. For a homogeneous polynomial $f \in S$ with $f = \sum_{\nu \in \mathbb{N}^n} a_\nu x^\nu$ and $\omega \in \mathbb{R}^n$, the *initial polynomial* $\mathrm{in}_\omega(f)$ of f consists of all terms of f whose ω-weight $\omega \cdot \nu$ is minimal. We use multiplicative term orders $\succ$ on the monomials of S and define $\mathrm{in}_\succ(f)$ to be the term cx^ν of f for which $cx^\nu \succ dx^\mu$ for every other term dx^μ of f. For $\omega \in \mathbb{R}^n$ and a term order $\succ$ we can consider the refinement $\succ_\omega$. This is the term order that first compares terms by their ω-weight and uses $\succ$ to break ties. Note that while initial polynomials with respect to ω are defined by taking terms of minimal ω-weight, the symbol $\succ$ suggests that $\mathrm{in}_\succ(f)$ is the "largest" term of f. The reason for considering this counterintuitive setup is that in Gröbner basis theory one usually considers the largest terms as initial terms, while in tropical geometry it is convenient to work with the minimal ω-weight.

In particular, we will repeatedly need a (degree) reverse lexicographic term order. With respect to an ordering $x_{j_1} \succ x_{j_2} \succ \cdots \succ x_{j_n}$ of the variables, this monomial order is defined as follows. For $\nu, \mu \in (\mathbb{N}_0)^n$ we have $x^\nu \succ x^\mu$ if either $\sum_i \nu_i > \sum_i \mu_i$ or $\sum_i \nu_i = \sum_i \mu_i$ and there exists $k \in \{1, \ldots, n\}$ such that $\nu_{j_i} = \mu_{j_i}$ for $i > k$ and $\nu_{j_k} < \mu_{j_k}$. Since we only consider graded ideals, we will simply call this order a *reverse lexicographic order*. If no ordering of the variables is specified, we mean the reverse lexicographic order with respect to $x_1 \succ x_2 \succ \cdots \succ x_n$.

We consider graded ideals $I \subset S$ and always assume $I \neq (0)$ if not stated otherwise. The dimension $\dim(S/I)$ refers to the Krull dimension of the ring S/I. Since we assume $I \neq (0)$, we always have $\dim(S/I) < n$. The *initial ideal of $I \subset S$ with respect to $\omega \in \mathbb{R}^n$* is defined as

$$\mathrm{in}_\omega(I) = (\mathrm{in}_\omega(f) : f \in I).$$

For $I \subset S = K[x_1, \ldots, x_n]$, we define the *tropical variety* of I by

$$T(I) = \{\omega \in \mathbb{R}^n : \mathrm{in}_\omega(I) \text{ does not contain a monomial}\}.$$

This is a special case, called the *constant coefficient case*, of the usual definition of a tropical variety as the image of a projective variety under a valuation map; see for example [Draisma 2008; Speyer and Sturmfels 2004]. In this case, K is considered to have a trivial valuation [Draisma 2008, Theorem 4.2]. Then the tropical variety $T(I)$ is a subfan of the *Gröbner fan* $\mathrm{GF}(I)$ of I as observed in [Bogart et al. 2007]. Recall that the Gröbner fan is a complete fan in $\mathbb{R}^n$, where $\omega, \omega' \in \mathbb{R}^n$ are in the same relatively open cone if $\mathrm{in}_\omega(I) = \mathrm{in}_{\omega'}(I)$; see for example [Mora and Robbiano 1988; Sturmfels 1996]. Sometimes we denote the ideal $\mathrm{in}_\omega(I)$ for a relatively open cone $\mathring{C}$ of $\mathrm{GF}(I)$ and $\omega \in \mathring{C}$ by $\mathrm{in}_C(I)$.

We study the structure of the tropical variety under a generic coordinate transformation in the following sense. For $g \in \mathrm{GL}_n(K)$, we regard the K-algebra automorphism induced by

$$K[x_1, \ldots, x_n] \to K[x_1, \ldots, x_n],$$

$$x_i \mapsto \sum_{j=1}^{n} g_{ji} x_j.$$

In the sequel, we identify g with this automorphism and call both of them g. Note that this definition differs from [Römer and Schmitz 2009, Definition 2.4] by a transposition of the matrix g. However, this does not affect the results proved in that paper. We consider $\mathrm{GL}_n(K)$ equipped with the Zariski-topology. If S/I is 0-dimensional, then for every $g \in \mathrm{GL}_n(K)$, the tropical variety $T(g(I))$ is empty [Römer and Schmitz 2009, Lemma 2.5]. We will therefore always assume that $\dim(S/I) > 0$. In Corollary 6.9 of the same work it was shown that for a graded ideal $I \subset S = K[x_1, \ldots, x_n]$ with $\dim(S/I) > 0$, there exists a Zariski-open set $\varnothing \neq U \subset \mathrm{GL}_n(K)$ such that $T(g(I))$ is the same fan for every $g \in U$. This fan is denoted by $\mathrm{gT}(I)$ and called the *generic tropical variety* of I. If $g \in U$, then $g(I)$ is called a *generic coordinate transformation* of I. Moreover, by [Römer and Schmitz 2009, Theorem 3.1] we know that there is also a *generic Gröbner fan* $\mathrm{gGF}(I)$ such that $\mathrm{GF}(g(I)) = \mathrm{gGF}(I)$ as a fan for every $g \in U$. The monomial initial ideal $\mathrm{in}_\succ(g(I))$ with respect to a term order $\succ$ is exactly the *generic initial ideal* $\mathrm{gin}_\succ(I)$ for $g \in U$. These generic initial ideals correspond to the maximal cones of $\mathrm{gGF}(I)$. In the following, we will fix a nonempty Zariski-open subset $U \subset \mathrm{GL}_n(K)$ such that

$$\mathrm{GF}(g(I)) = \mathrm{gGF}(I) \quad \text{and} \quad T(g(I)) = \mathrm{gT}(I) \quad \text{for every } g \in U, \quad (2\text{-}1)$$

and refer to it simply as U.

The generic tropical variety as a set is always equal to some skeleton of a particular complete fan $\mathcal{W}_n$ in $\mathbb{R}^n$. We recall that this fan is defined by the maximal cones $C_i = \{\omega \in \mathbb{R}^n : \omega_i = \min_k\{\omega_k\}\}$ for $i = 1, \ldots, n$. Note that to define a fan in $\mathbb{R}^n$ or to show that two fans in $\mathbb{R}^n$ are the same, it suffices to do this for the maximal cones. This is because every cone in a fan is a face of a maximal cone, so all cones in a fan are determined by the maximal cones. Every m-dimensional cone C_A in $\mathcal{W}_n$ for $m \in \{1, \ldots, n\}$ has the form $C_A = \{\omega \in \mathbb{R}^n : \omega_i = \min_k\{\omega_k\} \text{ for } i \in A\}$, where $A \subset \{1, \ldots, n\}$ with $|A| = n - m + 1$. On the other hand, every nonempty $A \subset \{1, \ldots, n\}$ defines a cone of $\mathcal{W}_n$ in this way, which we will denote by C_A. We let $\mathcal{W}_n^m$ be the m-skeleton of $\mathcal{W}_n$, that is, the fan consisting of all cones of $\mathcal{W}_n$ of dimension less than or equal to m. In [Römer and Schmitz 2009, Corollary 7.4], it was shown that for a graded ideal $I \subset S = K[x_1, \ldots, x_n]$ with $\dim(S/I) = m$, the generic tropical variety $\mathrm{gT}(I)$ coincides with $\mathcal{W}_n^m$ as a set.

For a fan $\mathscr{F}$ in $\mathbb{R}^n$, we denote by $|\mathscr{F}|$ the set $\bigcup_{C \in \mathscr{F}} C$ (without its fan structure), where the union is taken over all cones of $\mathscr{F}$. The notation C for a cone of $\mathscr{F}$ always refers to a closed cone. By $\mathring{C}$ we denote the relative interior of C. We say that a fan $\mathscr{E}$ in $\mathbb{R}^n$ *refines* a fan $\mathscr{F}$ in $\mathbb{R}^n$ if $|\mathscr{E}| = |\mathscr{F}|$ and for every relatively open cone $\mathring{C}$ of $\mathscr{E}$ there exists a relatively open cone $\mathring{D}$ of $\mathscr{F}$ with $\mathring{C} \subset \mathring{D}$. In Proposition 3.5 we show that it suffices to check this condition for the maximal cones of $\mathscr{E}$.

3. Fan structures on the set $|\mathscr{W}_n^m|$

In this section we assume $0 < m < n$. The aim is to show that $\mathscr{W}_n^m$ is the coarsest fan structure on the set $|\mathscr{W}_n^m|$. By this we mean that every fan $\mathscr{F}$ in $\mathbb{R}^n$ with $|\mathscr{F}| = |\mathscr{W}_n^m|$ refines $\mathscr{W}_n^m$ as a fan. We first prove that any fan on $|\mathscr{W}_n^m|$ is pure, by proving this statement for any subset of $\mathbb{R}^n$ that permits a pure fan structure of dimension at most $n - 1$. We repeatedly need the following lemma.

Lemma 3.1. *Let $\mathscr{F}$ be a fan in $\mathbb{R}^n$ and C a cone of $\mathscr{F}$. Let $\omega \in \mathring{C}$ and $(\omega_i)_{i \in \mathbb{N}}$ be a sequence such that $\omega_i \in |\mathscr{F}| \backslash C$ and $\lim_{i \to \infty} \omega_i = \omega$. Then there exists a cone D in $\mathscr{F}$ containing a subsequence of $(\omega_i)_{i \in \mathbb{N}}$ such that C is a proper face of D.*

Proof. Since $\omega_i \in \mathscr{F}$, there exists some other cone $C_i \neq C$ such that $\omega_i \in \mathring{C}_i$. But $\mathscr{F}$ has only finitely many cones, so there exists a subsequence $(\omega_{j_i})_{j_i \in \mathbb{N}}$ of $(\omega_i)_{i \in \mathbb{N}}$ such that $\omega_{j_i} \in D$ for one particular cone D of $\mathscr{F}$. By the choice of ω_i, we have $D \neq C$. Now $\lim_{i \to \infty} \omega_{j_i} = \omega$ and D is closed, so $\omega \in D$. Because $\mathring{C} \cap \mathring{D} = \varnothing$, we have $\omega \in \partial D$. By assumption, C and D intersect in a face of both of them. Since $\omega \in \mathring{C}$ is in this intersection, this face is C. Hence, $C \subsetneq D$ as a face. $\square$

With this we can show in the following proposition that any fan structure on $|\mathscr{W}_n^m|$ is pure.

Proposition 3.2. *Let $\mathscr{E}$ be a pure m-dimensional fan in $\mathbb{R}^n$ and $\mathscr{F}$ an arbitrary fan in $\mathbb{R}^n$ with $|\mathscr{F}| = |\mathscr{E}|$. Then $\mathscr{F}$ is also a pure m-dimensional fan.*

Proof. Let C be any cone in $\mathscr{F}$. Assume that $\dim C < m$ and let $\omega \in \mathring{C}$. Since for any open neighborhood $W(\omega) \subset \mathbb{R}^n$ of ω we have $\dim W(\omega) \cap C < m$, there always exists $v \in (W(\omega) \cap |\mathscr{F}|) \backslash C$. So if we choose a sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ with $\varepsilon_n > 0$ for every $n \in \mathbb{N}$ and $\lim_{n \to \infty} \varepsilon_n = 0$, there exists $v_n \in |\mathscr{F}| \backslash C$ with $|v_n - \omega| < \varepsilon_n$. By Lemma 3.1, we obtain a cone D of $\mathscr{F}$ such that $C \subsetneq D$. Since $\dim D > \dim C$, either the proof is complete if $\dim D = m$, or we can apply the same procedure to D instead of C. Either way, we obtain an m-dimensional cone of which C is a face after finitely many steps. $\square$

This immediately implies the following corollary, which is a generalization of the fact that the tropical variety of a prime ideal P with $\dim(S/P) = m$ that does not contain a monomial is a pure m-dimensional fan [Bieri and Groves 1984].

Corollary 3.3. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m$. Then $\mathrm{gT}(I)$ is a pure m-dimensional fan.*

To prove that a fan $\mathscr{E} \subset \mathbb{R}^n$ refines another fan $\mathscr{F} \subset \mathbb{R}^n$, it suffices to consider the maximal cones of $\mathscr{E}$. This will be the result of the next two statements.

Lemma 3.4. *Let D, C be cones in $\mathbb{R}^n$ such that $D \subset C$ and $D \cap \mathring{C} \neq \varnothing$. Then $\mathring{D} \subset \mathring{C}$.*

Proof. Let $p \in \mathring{D}$, and for some $\varepsilon > 0$ let $W = \{u \in \mathring{D} : |u - p| < \varepsilon\} \subset \mathring{D}$ be a relatively open neighborhood of p in $\mathring{D}$. If $p \in \partial C$, then there exists a face F of C with $p \in F$. Let $H = \{\omega \in \mathbb{R}^n : a \cdot \omega = 0\}$ be a defining hyperplane of F, so $F = H \cap C$, and let $C \subset H^- = \{\omega \in \mathbb{R}^n : a \cdot \omega \leq 0\}$. Since $W \subset \mathring{D} \subset C$, we know that $a \cdot u \leq 0$ for every $u \in W$. In addition, we have $a \cdot p = 0$, because $p \in F$. Assume there exists $u \in W$ such that $a \cdot u < 0$. Then we can choose $0 < \lambda < 1$ very small such that $p + \lambda(p - u) \in \mathring{D}$. Moreover, $|(p + \lambda(p - u)) - p| = |\lambda(p - u)| < \varepsilon$, so $p + \lambda(p - u) \in W$. But $(p + \lambda(p - u)) \cdot a = -\lambda u \cdot a > 0$, which is a contradiction to $p + \lambda(p - u) \in C$. Hence, $a \cdot u = 0$ for every $u \in W$. Thus $W \subset H$, and since W is relatively open in D, we also have $\mathrm{aff}(D) \subset H$. But then

$$D \subset \mathrm{aff}(D) \cap C \subset H \cap C = F,$$

which is a contradiction to $D \cap \mathring{C} \neq \varnothing$. Hence, $p \notin \partial C$ and we get that $\mathring{D} \subset \mathring{C}$. $\square$

Proposition 3.5. *Let $\mathscr{E}, \mathscr{F} \subset \mathbb{R}^n$ be two fans. Then $\mathscr{E}$ refines $\mathscr{F}$ as a fan if and only if $|\mathscr{E}| = |\mathscr{F}|$ and for every maximal cone $C \subset \mathscr{E}$ there exists a cone $D \subset \mathscr{F}$ such that $\mathring{C} \subset \mathring{D}$.*

Proof. One implication follows directly from the definition of refinement. For the other one, we have to show that for any cone $K \subset \mathscr{E}$, there exists a cone $L \subset \mathscr{F}$ such that $\mathring{K} \subset \mathring{L}$. If K is maximal, this is true by assumption. Let $K \subset \mathscr{E}$ be not maximal. Then there exists a maximal cone $C \subset \mathscr{E}$ such that K is a face of C. Moreover, we know that $\mathring{C} \subset \mathring{D}$ for some cone $D \in \mathscr{F}$. So $K \subset D$. Assume that such a cone L does not exist. If $K \cap \mathring{D} \neq \varnothing$, this would imply $\mathring{K} \subset \mathring{D}$ by Lemma 3.4 and we could set $L = D$. Hence, $K \cap \mathring{D} = \varnothing$. Then $K \subset \partial D$ and by [Bruns and Gubeladze 2009, Lemma 1.5], it follows that $K \subset E$ for a proper face E of D. Since $\dim E < \dim D$, we can use a suitable induction to obtain a sequence of cones in $\mathscr{F}$ of strictly decreasing dimension such that K does not intersect the relative interior of each cone. The last cone in this sequence has to be the lineality space A of $\mathscr{F}$. So by this induction, we get $K \subset \partial A$, which is a contradiction to $\partial A = \varnothing$. Hence, there has to exist a cone $L \subset \mathscr{F}$ such that $\mathring{K} \subset \mathring{L}$. $\square$

The proof of the next auxiliary result is elementary, so we omit it.

Lemma 3.6. *Let $C \subset \mathbb{R}^n$ be a cone and let* $\dim C = m$. *Also, let* $D_1, \ldots, D_s \subset \mathbb{R}^n$ *be cones such that*

$$C \subset \bigcup_{i=1}^{s} D_i,$$

where $\dim D_1 = m$ *and* $\dim D_2, \ldots, \dim D_s < m$. *Then* $C \subset D_1$.

With these prerequisites, we can show that for $m < n$, the fan structure of $\mathcal{W}_n^m$ is actually the coarsest possible on the set $|\mathcal{W}_n^m|$ in the sense that every other fan $\mathcal{F} \subset \mathbb{R}^n$ with $|\mathcal{F}| = |\mathcal{W}_n^m|$ refines $\mathcal{W}_n^m$ as a fan. In particular, this will imply that $\mathrm{gT}(I)$ refines $\mathcal{W}_n^m$ as a fan for a graded ideal $I \subset S = K[x_1, \ldots, x_n]$ with $\dim(S/I) = m$.

Proposition 3.7. *Let* $m < n$ *and* $\mathcal{F} \subset \mathbb{R}^n$ *be a fan with* $|\mathcal{F}| = |\mathcal{W}_n^m|$. *Then for every relatively open cone* $\mathring{C}$ *of* $\mathcal{F}$, *there exists a relatively open cone* $\mathring{C}_A$ *of* $\mathcal{W}_n^m$ *such that* $\mathring{C} \subset \mathring{C}_A$.

Proof. $\mathcal{W}_n^m = \dot{\bigcup}_{|A| \geq n-m+1} \mathring{C}_A$ is the disjoint union of all relatively open cones of $\mathcal{W}_n$ whose defining set $A \subset \{1, \ldots, n\}$ has at least $n - m + 1$ elements. By Proposition 3.5 it suffices to prove the condition for the maximal cones of $\mathcal{F}$. Let C be a maximal cone of $\mathcal{F}$. Then $\dim C = m$, as $\mathcal{F}$ is pure by Proposition 3.2. Since $\dim \bigcup_{|A|>n-m+1} \mathring{C}_A = m - 1 < m$, there exists an $\omega \in \mathring{C}$ that is contained in the interior of some maximal cone $\mathring{C}_{A_1}$ of $\mathcal{W}_n^m$ with $|A_1| = n - m + 1$. Assume there exists $v \in \mathring{C}$ such that $v \in \mathring{C}_{A_2}$ for a different maximal cone $\mathring{C}_{A_2}$ of $\mathcal{W}_n^m$. Then $|A_1 \cap A_2| < n - m + 1$. We have to consider two cases:

- If $A_1 \cap A_2 \neq \varnothing$, the minimal coordinates of $\omega + v$ are attained exactly at the indices in $A_1 \cap A_2$. But $|A_1 \cap A_2| < n - m + 1$, so $\omega + v \notin |\mathcal{W}_n^m|$. This is a contradiction to $\omega + v \in \mathring{C} \subset |\mathcal{W}_n^m|$.

- Assume that $A_1 \cap A_2 = \varnothing$. Since $\dim C = m$, we can change the coordinates of ω that are not contained in A_1 independently from each other by adding or subtracting small real numbers without leaving $\mathring{C}$. The same is true for the coordinates of v that are not in A_2. Hence, we can change every coordinate of $\omega + v$ by a small amount without leaving $\mathring{C}$, since $A_1 \cap A_2 = \varnothing$. But then we can assume that the minimum of the coordinates of $\omega + v$ is attained only once. Again we have $\omega + v \notin |\mathcal{W}_n^m|$, contradicting $\omega + v \in \mathring{C} \subset |\mathcal{W}_n^m|$.

Hence, no element of $\mathring{C}$ can be contained in the relative interior of any maximal cone of $\mathcal{W}_n^m$ other than C_{A_1}. But then

$$\mathring{C} \subset \mathring{C}_{A_1} \cup \left(\bigcup_{|A|>n-m+1} \mathring{C}_A \right).$$

Taking the topological closure, this implies $C \subset C_{A_1}$ by Lemma 3.6. Since both cones have the same dimension, we also have $\mathring{C} \subset \mathring{C}_{A_1}$ by Lemma 3.4. $\square$

Proposition 3.7 implies as a corollary that the generic tropical variety always refines $\mathcal{W}_n^m$ as a fan.

Corollary 3.8. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m > 0$. Then for every relatively open cone $\mathring{C}$ of $\mathrm{gT}(I)$, there exists a relatively open cone $\mathring{D}$ of $\mathcal{W}_n^m$ such that $\mathring{C} \subset \mathring{D}$.*

Proof. Since $|\mathrm{gT}(I)| = |\mathcal{W}_n^m|$ by [Römer and Schmitz 2009, Corollary 7.4] and $\mathcal{W}_n^m$ is pure m-dimensional, it follows from Proposition 3.2 that $\mathrm{gT}(I)$ is a pure m-dimensional fan. The claim now is a consequence of Proposition 3.7.   □

4. Generic tropical varieties of Cohen–Macaulay and almost-Cohen–Macaulay rings

In addition to the dimension of S/I, it is also possible to recover information on the depth of S/I from the generic tropical variety of I. We will show that for an ideal I with $\dim(S/I) = m$ and $\mathrm{depth}(S/I) > 0$, the generic tropical variety is $\mathcal{W}_n^m$ as a fan if and only if $\mathrm{depth}(S/I) = \dim(S/I)$ or $\mathrm{depth}(S/I) = \dim(S/I) - 1$. Thus we can read off whether S/I is Cohen–Macaulay or almost-Cohen–Macaulay from the fan $\mathrm{gT}(I)$.

To define the depth of S/I, recall that a system of linear forms $l_1, \ldots, l_t \in S/I$ is called a *regular sequence for S/I* if l_i is not a zero-divisor on $(S/I)/(l_1, \ldots, l_{i-1})$ for $i = 1, \ldots, t$.

Definition 4.1. For a graded ideal $I \subset S = K[x_1, \ldots, x_n]$, we define the *depth of S/I* to be

$$\mathrm{depth}(S/I) = \max \left\{ \begin{array}{l} t \in \mathbb{N} : \text{ there is a regular sequence} \\ \quad \text{of linear forms } l_1, \ldots, l_t \in S/I \end{array} \right\}.$$

The depth is bounded from above by the dimension of S/I; see for example [Bruns and Herzog 1993, Proposition 1.2.12]. Also, we know that $\mathrm{depth}(S/I) \geq \mathrm{depth}(S/\mathrm{gin}_\succ(I))$ for any term order $\succ$. Equality holds if $\succ$ is a reverse lexicographic order with respect to some ordering of the variables. These two statements follow from [Bruns and Conca 2004, Corollary 3.5 and Remark 3.6] together with the Auslander–Buchsbaum formula.

In general it is not possible to see the depth of S/I in the fan $T(I)$, as the following example shows.

Example 4.2. For $1 \leq k \leq n$, consider the ideal

$$I = (x_1(x_1 + x_2), x_2(x_1 + x_2), \ldots, x_k(x_1 + x_2)) \subset S = K[x_1, \ldots, x_n].$$

Then $\dim(S/I) = n - 1$ and $\mathrm{depth}(S/I) = n - k$. But the tropical variety $T(I)$ always consists of only one cone $T(I) = \{\omega \in \mathbb{R}^n : \omega_1 = \omega_2\}$ which is independent

of k. So we have obtained a collection of ideals of every possible depth from 0 to $n-1$ such that the tropical variety is always the same.

The connection of depth(S/I) with gT(I) is established by the following proposition, taken as a reformulation of [Herzog and Srinivasan 1998, Lemma 3.1] and relying on [Eliahou and Kervaire 1990]. Since a generic initial ideal J is a monomial ideal, there exists a system of monomial generators of J. The unique smallest system of monomial generators with respect to inclusion will be called a *minimal system of generators*, and its elements are *minimal generators* of J.

Proposition 4.3. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m$, and $\succ$ be any term order with $x_1 \succ \cdots \succ x_n$. Let $\operatorname{depth}(S/\operatorname{gin}_\succ(I)) = t$. Then:*

- *Every minimal generator of $\operatorname{gin}_\succ(I)$ is divisible by one of $x_1, \ldots, x_{n-m}$.*
- *x_{n-m}^d is one of the minimal generators of $\operatorname{gin}_\succ(I)$ for some $d \in \mathbb{N}$.*
- *The minimal generators of $\operatorname{gin}_\succ(I)$ are elements of $K[x_1, \ldots, x_{n-t}]$.*
- *There exists a minimal generator of $\operatorname{gin}_\succ(I)$ that is divisible by x_{n-t}.*

In particular, if $\succ$ is the reverse lexicographic order, these statements are true for $t = \operatorname{depth}(S/I)$, since then $\operatorname{depth}(S/I) = \operatorname{depth}(S/\operatorname{gin}_\succ(I))$.

Recall that by [Römer and Schmitz 2009, Corollary 7.4], the condition for $\omega \in \mathbb{R}^n$ to be in gT(I) is that the minimum of its coordinates be attained at least $n-m+1$ times. So Proposition 4.3 already shows that the cases where $\operatorname{depth}(S/I) = m$ and $\operatorname{depth}(S/I) = m - 1$ are special. We use the following standard definition.

Definition 4.4. Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal. If $\operatorname{depth}(S/I) = \dim(S/I)$, then S/I will be called *Cohen–Macaulay*. If $\operatorname{depth}(S/I) = \dim(S/I) - 1$, then S/I is called *almost-Cohen–Macaulay*. In what follows, we say I is *Cohen–Macaulay* or *almost-Cohen–Macaulay* if S/I has the corresponding property.

In this case, the refinement $\succ_\omega$ of every $\omega \in \operatorname{gT}(I)$ with respect to an appropriate reverse lexicographic order $\succ$ yields the same generic initial ideal as with respect to $\succ$. In the following statement, the set U denotes the Zariski-open subset of $\operatorname{GL}_n(K)$ as defined in (2-1).

Lemma 4.5. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded Cohen–Macaulay or almost-Cohen–Macaulay ideal, $\succ$ be the reverse lexicographic order, and $\omega \in \mathcal{W}_n^m \subset \mathbb{R}^n$ with $\omega_1 = \omega_2 = \cdots = \omega_{n-m+1} \leq \omega_{n-m+2}, \ldots, \omega_n$. Moreover, let $\succ_\omega$ be the refinement of ω with respect to $\succ$. Then the reduced Gröbner bases of $g(I)$ with respect to $\succ$ and $\succ_\omega$ are the same for $g \in U$. In particular, $\operatorname{gin}_{\succ_\omega}(I) = \operatorname{gin}_\succ(I)$.*

Proof. Since for a given degree t any term containing none of $x_{n-m+2}, \ldots, x_n$ is smaller than any term divisible by one of them with respect to $\succ_\omega$, the term orders $\succ$ and $\succ_\omega$ coincide up to the term x_{n-m+1}^t. By Proposition 4.3, the minimal generators

of $\text{gin}_{\succ}(I)$ are monomials in $K[x_1, \ldots, x_{n-m+1}]$. Then for $g \in U$, the leading terms of the reduced Gröbner basis $\mathcal{G}(g)$ of $g(I)$ are terms in $K[x_1, \ldots, x_{n-m+1}]$. Since the leading terms of two elements of $\mathcal{G}(g)$ are the same with respect to $\succ$ and $\succ_\omega$, every S-pair with respect to $\succ_\omega$ is the same as with respect to $\succ$. As $\mathcal{G}(g)$ is a Gröbner basis with respect to $\succ$, every such S-pair reduces to 0. So the set $\mathcal{G}(g)$ is a Gröbner basis with respect to $\succ_\omega$ as well. Hence, $\text{gin}_{\succ_\omega}(I) = \text{gin}_{\succ}(I)$.   □

We can now formulate the reverse statement of Proposition 3.7 for Cohen–Macaulay and almost-Cohen–Macaulay ideals.

Proposition 4.6. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded Cohen–Macaulay or almost-Cohen–Macaulay ideal with $\dim(S/I) = m$. Then for every relatively open cone $\mathring{C}_A \subset \mathcal{W}_n^m$ there exists a relatively open cone $\mathring{C}$ of $\text{gT}(I)$ with $\mathring{C}_A \subset \mathring{C}$.*

Proof. Let $A \subset \{1, \ldots, n\}$ with $|A| \geq n - m + 1$, so $\mathring{C}_A$ is an open cone of $\mathcal{W}_n^m$. We need to show that for $\omega, \omega' \in \mathring{C}_A$, we have $\text{in}_\omega(g(I)) = \text{in}_{\omega'}(g(I))$ for every $g \in U$. Without loss of generality, we may assume $\{1, \ldots, n-m+1\} \subset A$. Let $\succ$ denote the reverse lexicographic order. By Lemma 4.5, the reduced Gröbner basis $\mathcal{G}(g) = \{h_1(g), \ldots, h_s(g)\}$ of $g(I)$ with respect to $\succ$ is also a reduced Gröbner basis with respect to $\succ_\omega$ and $\succ_{\omega'}$ for $g \in U$. So $\{\text{in}_\omega(h_1(g)), \ldots, \text{in}_\omega(h_s(g))\}$ and $\{\text{in}_{\omega'}(h_1(g)), \ldots, \text{in}_{\omega'}(h_s(g))\}$ are Gröbner bases of $\text{in}_\omega(g(I))$ and $\text{in}_{\omega'}(g(I))$. However, all the leading terms of the $h_i(g)$ are elements of $K[x_1, \ldots, x_{n-m+1}]$. Hence, $\text{in}_\omega(h_i(g))$ and $\text{in}_{\omega'}(h_i(g))$ exactly consist of those terms of $h_i(g)$ that contain only variables x_j for which ω_j and ω'_j respectively are minimal. But these variables are the same for ω and ω' by assumption, so we obtain $\text{in}_\omega(g(I)) = \text{in}_{\omega'}(g(I))$. This shows that all $\omega \in \mathring{C}_A$ are contained in the same open cone $\mathring{C}$ of $T(g(I)) = \text{gT}(I)$ for $g \in U$.   □

In the case of a Cohen–Macaulay or almost-Cohen–Macaulay ideal I such that $\dim(S/I) = m$, the generic tropical variety is equal to $\mathcal{W}_n^m$ as a fan. This generalizes the result [Römer and Schmitz 2009, Corollary 7.4] for this class of ideals.

Corollary 4.7. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded Cohen–Macaulay or almost-Cohen–Macaulay ideal with $\dim(S/I) = m$. Then $\text{gT}(I) = \mathcal{W}_n^m$ as a fan.*

Proof. Let $\mathring{C}$ be a relatively open cone of $\text{gT}(I)$. By Corollary 3.8, there exists a cone D of $\mathcal{W}_n^m$ such that $\mathring{C} \subset \mathring{D}$. On the other hand, by Proposition 4.6 there exists a cone E of $\text{gT}(I)$ with $\mathring{D} \subset \mathring{E}$. But then $\mathring{C} \subset \mathring{E}$ are two cones of $\text{gT}(I)$ with $\mathring{C} \cap \mathring{E} \neq \varnothing$. This implies $\mathring{C} = \mathring{E}$ and thus $\mathring{C} = \mathring{D}$. This shows that every maximal cone of $\text{gT}(I)$ is equal to some maximal cone of $\mathcal{W}_n^m$. By the same argument, it follows that every maximal cone from $\mathcal{W}_n^m$ is equal to some maximal cone of $\text{gT}(I)$, so the two fans are the same.   □

To show that Corollary 4.7 is wrong for every ideal that is not Cohen–Macaulay or almost-Cohen–Macaulay, we need the following auxiliary result.

Lemma 4.8. *Let $c \in \mathbb{N}$ and $\omega \in \mathbb{R}^n$ be such that $0 = \omega_1 = \cdots = \omega_{n-m+1}$ and $c\omega_i < \omega_{i+1}$ for $i = n-m+1, \ldots, n-1$. Let $\succ$ be the reverse lexicographic order. Then $\succ$ and $\succ_\omega$ are the same term orders for the monomials of any degree up to c.*

Proof. Let $t \leq c$ and x^ν, x^μ be two monomials of degree t. We write $x^\nu = y_1 z_1$ and $x^\mu = y_2 z_2$, where $y_1, y_2 \in K[x_1, \ldots, x_{n-m+1}]$ and $z_1, z_2 \in K[x_{n-m+2}, \ldots, x_n]$. If $z_1 = z_2$, it is clear from the definition that $x^\nu \succ x^\mu$ if and only if $x^\nu \succ_\omega x^\mu$. Otherwise, let $k \geq n-m+2$ be the largest index such that $\nu_k \neq \mu_k$. Without loss of generality, we may assume that no variable x_j divides x^ν or x^μ for $j > k$ and that $\nu_k < \mu_k$, so $x^\nu \succ x^\mu$. For the ω-weight of x^ν and x^μ we obtain the upper bound

$$\mathrm{wt}_\omega(x^\nu) \leq \mathrm{wt}_\omega(x_{k-1}^{t-\nu_k} x_k^{\nu_k}) = \omega_{k-1}(t-\nu_k) + \omega_k \nu_k$$

and the lower bound

$$\mathrm{wt}_\omega(x^\mu) \geq \mathrm{wt}_\omega(x_1^{t-\mu_k} x_k^{\mu_k}) = \omega_k \mu_k.$$

So it is enough to show that $\omega_{k-1}(t-\nu_k) + \omega_k \nu_k < \omega_k \mu_k$. We have

$$\begin{aligned}
c(\omega_k \mu_k - (\omega_{k-1}(t-\nu_k) + \omega_k \nu_k)) &= c\omega_k(\mu_k - \nu_k) - c\omega_{k-1}(t-\nu_k) \\
&> c\omega_k(\mu_k - \nu_k) - \omega_k(t-\nu_k) \\
&= \omega_k(c(\mu_k - \nu_k) - (t-\nu_k)) \geq 0.
\end{aligned}$$

The last inequality is true, since $t - \nu_k \leq c$ and $c(\mu_k - \nu_k) > c$, as we know $\mu_k > \nu_k$. It follows that $\mathrm{wt}_\omega(x^\nu) < \mathrm{wt}_\omega(x^\mu)$, so $x^\nu \succ_\omega x^\mu$. Hence, $\succ$ and $\succ_\omega$ coincide up to degree c. $\qquad\square$

We can now completely characterize when $\mathrm{gT}(I)$ is equal to a skeleton of the generic tropical fan for ideals of $\mathrm{depth}(S/I) > 0$. If $\dim(S/I) = 0$, we know that $\mathrm{gT}(I)$ is empty, since every graded ideal with $\dim(S/I) = 0$ contains a monomial. In the cases $\dim(S/I) = 1$ and $\dim(S/I) = 2$, the fan $\mathrm{gT}(I)$ is equal to $\mathcal{W}_n^1$ and $\mathcal{W}_n^2$ respectively by [Römer and Schmitz 2009, Examples 8.3 and 8.4]. Note that in these cases, every ideal of $\mathrm{depth}(S/I) > 0$ is Cohen–Macaulay or almost Cohen–Macaulay. For ideals with arbitrary dimension $\dim(S/I) > 0$, we have:

Theorem 4.9. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m > 0$ and $\mathrm{depth}(S/I) > 0$. Then S/I is Cohen–Macaulay or almost-Cohen–Macaulay if and only if $\mathrm{gT}(I) = \mathcal{W}_n^m$ as a fan.*

Proof. We show that if $t = \mathrm{depth}(S/I) < m-1$, then $\mathrm{gT}(I) \neq \mathcal{W}_n^m$ as a fan. For this, let $\succ$ be the reverse lexicographic order with $x_1 \succ \cdots \succ x_{n-t} \succ x_{n-t+1} \succ \cdots \succ x_n$ and $\succ'$ be the reverse lexicographic order with $x_1 \succ' \cdots \succ' x_{n-t+1} \succ' x_{n-t} \succ' \cdots \succ' x_n$. Let c be the maximal degree of the minimal generators of $\mathrm{gin}_\succ(I)$ and $\mathrm{gin}_{\succ'}(I)$. For the purpose of this proof, for $a, b \in \mathbb{R}_+$ we write $a \ll b$ if $ac < b$.

Choose $\omega, \upsilon \in \mathbb{R}^n$ such that $0 = \omega_1 = \cdots = \omega_{n-m+1} \ll \omega_{n-m+2} \ll \cdots \ll \omega_n$ and

$$0 = \upsilon_1 = \cdots = \upsilon_{n-m+1} \ll \upsilon_{n-m+2} \ll \cdots \ll \upsilon_{n-t+1} \ll \upsilon_{n-t} \ll \cdots \ll \upsilon_n.$$

By Lemma 4.8, we know that $\succ$ and $\succ_\omega$ are the same term orders up to degree c. Since $\mathrm{gin}_\succ(I)$ is generated by monomials of degree at most c, for a fixed $g \in U$ (as defined in (2-1)), all elements of the reduced Gröbner basis $\mathcal{G}$ of $g(I)$ with respect to $\succ$ have degree at most c. The leading term of every element of $\mathcal{G}$ is the same with respect to $\succ$ and $\succ_\omega$. Thus every S-pair of elements of $\mathcal{G}$ reduces to zero with respect to $\succ_\omega$ as well. So $\mathcal{G}$ is also a Gröbner basis of $\mathrm{gin}_{\succ_\omega}(I)$. This implies $\mathrm{gin}_{\succ_\omega}(I) = \mathrm{gin}_\succ(I)$.

By Lemma 4.8 and the same argument as before, we can show that $\mathrm{gin}_{\succ_\upsilon}(I) = \mathrm{gin}_{\succ'}(I)$. Since $\mathrm{depth}(S/\mathrm{gin}_\succ(I)) = \mathrm{depth}(S/\mathrm{gin}_{\succ'}(I)) = t$, we know that x_{n-t} divides one of the minimal generators of $\mathrm{gin}_\succ(I)$ but does not divide one of the minimal generators of $\mathrm{gin}_{\succ'}(I)$ by Proposition 4.3. So $\mathrm{gin}_{\succ_\omega}(I) = \mathrm{gin}_\succ(I) \neq \mathrm{gin}_{\succ'}(I) = \mathrm{gin}_{\succ_\upsilon}(I)$, and it follows that $\mathrm{in}_\omega(g(I)) \neq \mathrm{in}_\upsilon(g(I))$ for $g \in U$. Hence, ω and υ are not in the same relatively open cone of $\mathrm{gT}(I)$, but they are in the same relatively open cone of $\mathcal{W}_n^m$. This implies $\mathrm{gT}(I) \neq \mathcal{W}_n^m$ as a fan.

The converse of this statement, that $\mathrm{depth}(S/I) \geq m - 1$ implies $\mathrm{gT}(I) = \mathcal{W}_n^m$ as a fan, has already been proved in Corollary 4.7. $\square$

In particular, this theorem gives a negative answer to the question posed in the introduction of [Römer and Schmitz 2009] of whether the generic tropical variety of I as a fan only depends on the dimension of S/I. If $\mathrm{depth}(S/I) = 0$, it is not possible to obtain a statement like Theorem 4.9; see Remark 5.12.

5. Generic tropical varieties and depth

In this section we will consider a certain class of ideals I such that $\dim(S/I) - 1 > \mathrm{depth}(S/I) > 0$ for which we can recover the depth from the generic tropical varieties. These ideals have the property that the rings $S/\mathrm{gin}_\succ(I)$ for all generic initial ideals of I have the same depth as S/I itself. This makes it possible to use Proposition 4.3 on all of these. We express this property by considering the *generic depth* of S/I.

Definition 5.1. For a graded ideal $I \subset S = K[x_1, \ldots, x_n]$, we call

$$\mathrm{gdepth}(S/I) = \min\{\mathrm{depth}(S/\mathrm{gin}_\succ(I))\},$$

where the minimum is taken over all possible generic initial ideals of I, the *generic depth of S/I*. If $\mathrm{depth}(S/I) = \mathrm{gdepth}(S/I)$, then I is called a *maximal-gdepth ideal*.

Note that since $\mathrm{depth}(S/\mathrm{gin}_\succ(I)) \leq \mathrm{depth}(S/I)$ for any generic initial ideal of I, the ideal I is a maximal-gdepth ideal if and only if $\mathrm{depth}(S/I) = \mathrm{depth}(S/\mathrm{gin}_\succ(I))$

for every generic initial ideal of I. A trivial example of a maximal-gdepth ideal is a principal ideal like $I = (x_1^d + \cdots + x_n^d)$ in $S = K[x_1, \ldots, x_n]$ for an integer $d \geq 1$, because every gin is also a principal ideal generated by a polynomial of degree d. Next we describe two interesting classes of maximal-gdepth ideals.

Example 5.2. The first example is the class of strongly stable ideals I. These are by definition monomial ideals, so the tropical variety $T(I)$ is of course empty. We are, however, interested in the generic tropical variety $\mathrm{gT}(I)$. This fan is not empty if $\dim(S/I) > 0$, and it contains algebraic information on I, as we show.

Recall that a monomial ideal $I \subset K[x_1, \ldots, x_n]$ is called strongly stable with respect to some ordering $x_{i_1} > \cdots > x_{i_n}$ of the variables $x_1, \ldots, x_n$ if for every monomial $u \in I$ we also have $x_{i_j} u x_{i_k}^{-1} \in I$ for every x_{i_k} that divides u and every $j < k$. Every ideal $\mathrm{gin}_\succ(I)$ is a strongly stable ideal with respect to the ordering of the variables given by $\succ$, since $\mathrm{char}(K) = 0$. Moreover, if I is strongly stable with respect to $x_{i_1} > \cdots > x_{i_n}$ and $\succ$ is a term order with $x_{i_1} \succ \cdots \succ x_{i_n}$, then $\mathrm{gin}_\succ(I) = I$. We now explain that strongly stable ideals are maximal-gdepth ideals.

Let $I \subset K[x_1, \ldots, x_n]$ be a graded strongly stable ideal with respect to $x_1 > \cdots > x_n$. Let $\succ$ be any term order with $x_{i_1} \succ \cdots \succ x_{i_n}$ for $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$. In addition, let $\succ'$ be the reverse lexicographic order with $x_{i_1} \succ' \cdots \succ' x_{i_n}$. Consider the image of I under the K-algebra isomorphism ϕ that maps x_j to x_{i_j}. Then $\phi(I)$ is a strongly stable ideal with respect to term orders with $x_{i_1} \succ \cdots \succ x_{i_n}$, so in particular with respect to $\succ$ and to $\succ'$. So we know that $\mathrm{gin}_\succ(\phi(I)) = \mathrm{gin}_{\succ'}(\phi(I)) = \phi(I)$. Let $\varnothing \neq U_1 \subset \mathrm{GL}_n(K)$ be Zariski-open such that $\mathrm{in}_\succ(g(I)) = \mathrm{gin}_\succ(I)$ for every $g \in U_1$, and $\varnothing \neq U_2 \subset \mathrm{GL}_n(K)$ Zariski-open such that $\mathrm{in}_\succ(h(\phi(I))) = \mathrm{gin}_\succ(\phi(I))$ for every $h \in U_2$. Note that also the set $\varnothing \neq U_2' = \{h \circ \phi \in \mathrm{GL}_n(K) : h \in U_2\}$ is Zariski-open, as it can be defined by the polynomials obtained by permuting the polynomials defining U_2 according to ϕ. Hence, $U_1 \cap U_2' \neq \varnothing$. For $k \in U_1 \cap U_2'$ we have $\mathrm{in}_\succ(k(I)) = \mathrm{gin}_\succ(I) = \mathrm{gin}_\succ(\phi(I))$. In addition, $\mathrm{gin}_{\succ'}(I) = \mathrm{gin}_{\succ'}(\phi(I))$ by the same argument. Hence, $\mathrm{gin}_\succ(I) = \mathrm{gin}_{\succ'}(I)$. Since for any reverse lexicographic order $\succ$ we have $\mathrm{depth}(S/\mathrm{gin}_\succ(I)) = \mathrm{depth}(S/I)$, this implies that strongly stable ideals are maximal-gdepth ideals.

A concrete example for such an ideal is $(x_1^2, x_1 x_2, x_1 x_3, x_1 x_4) \subseteq K[x_1, \ldots, x_5]$. Applying a K-algebra automorphism in $\mathrm{GL}_n(K)$ provides examples of maximal-gdepth ideals that are not monomial ideals.

Example 5.3. The second example class is the class of ideals such that I and every $\mathrm{gin}_\succ(I)$ is generated by polynomials of the same degree. We can see that these ideals are also maximal-gdepth as follows. Let $S = K[x_1, \ldots, x_n]$ with the standard $\mathbb{Z}$-grading and note that S/I is a graded S-module. We denote by $\beta_{i,j}$ the graded Betti number $\beta_{i,j} = \beta_{i,j}(S/I) = \dim_K(\mathrm{Tor}_i^S(S/I, K))_j$. For $d \in \mathbb{N}$, let $S(-d)$ be the graded module S with the grading given by $S(-d)_j = S_{j-d}$. Recall

that for some $d \in \mathbb{N}$, we say that I has a d-linear resolution if and only if $\beta_{i,i+j} = 0$ for $j \neq d - 1$, $i \geq 1$. This is equivalent to the fact that the minimal graded free resolution of S/I has the form

$$0 \to S(-d - p + 1)^{\beta_p} \to \cdots \to S(-d - 1)^{\beta_2} \to S(-d)^{\beta_1} \to S \to S/I \to 0,$$

where $\beta_i = \beta_{i,i+d}$ is the i-th total Betti number and p is the projective dimension $p = \text{projdim}(S/I)$ of S/I. Let $I \subset S$ be a graded ideal such that I and $\text{gin}_\succ(I)$ are generated by polynomials of degree d for every term order $\succ$. Let $J = \text{gin}_\succ(I)$ be a given generic initial ideal if I. We now show that $\text{depth}(S/I) = \text{depth}(S/J)$. As J is strongly stable and generated in one degree, the minimal graded free resolution (as constructed in [Eliahou and Kervaire 1990]) is linear. Since $\beta_{i,j}(S/I) \leq \beta_{i,j}(S/J)$ for every i, j (see for example [Bruns and Conca 2004, Proposition 3.3]), this implies that S/I has a linear resolution as well. So I and J have linear resolutions, which in turn means that their total Betti numbers depend only on the Hilbert series of S/I and S/J respectively. But $H_{S/I}(t) = H_{S/J}(t)$, so the Betti numbers and in particular the projective dimensions of S/I and S/J are the same. By the Auslander–Buchsbaum formula $\text{depth}(S/I) + \text{projdim}(S/I) = n$, it follows that $\text{depth}(S/I) = \text{depth}(S/J)$. This is true for every generic initial ideal of I. Hence, I is a maximal-gdepth ideal.

Recall that the Castelnuovo–Mumford regularity $\text{reg}(S/I)$ of S/I is defined to be the maximal $j \in \mathbb{Z}$ such that $\beta_{i,i+j} \neq 0$ for some $i \geq 0$. It is well-known that the ideal $I_{\geq t}$ (generated by all homogeneous components I_s of I with $s \geq t$) has a linear resolution if $t \geq \text{reg}(S/I) + 1$; see [Eisenbud and Goto 1984]. Observe further that by the construction of gin, we know $\text{gin}_\succ(I)_{\geq t} = \text{gin}_\succ(I_{\geq t})$. Since there exist only finitely many different generic initial ideals for I, we can find a t such that all $\text{gin}_\succ(I_{\geq t})$ have a linear resolution. In particular, they are generated in degree t. Then also $I_{\geq t}$ is generated in degree t (and has a linear resolution). So every high truncation $I_{\geq t}$ of an arbitrary graded ideal I is a maximal-gdepth ideal. Unfortunately we cannot decide in general which t one has to take.

In this section we give a structural result on generic tropical varieties of maximal-gdepth ideals. We will see that these as fans are closely related to the following refinement of $\mathcal{W}_n^m$.

Definition 5.4. Let $\mathcal{W}_n^m$ be the m-skeleton of the standard tropical fan in $\mathbb{R}^n$, and let $0 < t < m - 1$. The refinement of $\mathcal{W}_n^m$ containing all open cones

$$\{\omega \in \mathbb{R}^n : \omega_{i_1} = \cdots = \omega_{i_{n-m+1}} < \omega_{i_{n-m+2}}, \ldots, \omega_{i_{n-t}} < \omega_{i_{n-t+1}}, \ldots, \omega_{i_n}\}$$

for any permutation $(i_1, \ldots, i_n)$ of $\{1, \ldots, n\}$ as maximal open cones will be called the t-refinement of $\mathcal{W}_n^m$ and denoted by $\mathcal{W}_n^{m,t}$.

Lemma 5.5. *Let* $I \subset S = K[x_1, \ldots, x_n]$ *be a maximal-*gdepth *ideal such that* $\dim(S/I) = m < n$ *and* $\operatorname{depth}(S/I) = t$ *for some* $0 < t < m - 1$. *Let* $\omega \in \mathring{C}$ *for some maximal cone* C *of* $\operatorname{gT}(I)$ *and let* $\omega_{i_1} \le \cdots \le \omega_{i_n}$ *for* $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$. *Then* $\omega_{i_{n-t}} < \omega_{i_{n-t+1}}$.

Proof. Without loss of generality we have

$$\omega_1 = \cdots = \omega_{n-m+1} < \omega_{n-m+2} \le \cdots \le \omega_n,$$

since $\omega \in |\mathcal{W}_n^m|$. Let us assume that $\omega_{n-t} = \omega_{n-t+1}$. For $\varepsilon > 0$ we define $u^\varepsilon \in \mathbb{R}^n$ by

$$u_i^\varepsilon = \begin{cases} \omega_i - \varepsilon & \text{for } i < n - t, \\ \omega_i & \text{for } i = n - t, \\ \omega_i + \varepsilon & \text{for } i = n - t + 1, \\ \omega_i + 2\varepsilon & \text{for } i > n - t + 1. \end{cases}$$

In the same way we define $v_i^\varepsilon \in \mathbb{R}^n$ by

$$v_i^\varepsilon = \begin{cases} \omega_i - \varepsilon & \text{for } i < n - t, \\ \omega_i & \text{for } i = n - t + 1, \\ \omega_i + \varepsilon & \text{for } i = n - t, \\ \omega_i + 2\varepsilon & \text{for } i > n - t + 1. \end{cases}$$

Note that u^ε and v^ε are contained in $\operatorname{gT}(I)$ for every choice of ε. Let $\succ$ be any term order. Then $x_i \succ_{u^\varepsilon} x_{n-t} \succ_{u^\varepsilon} x_{n-t+1} \succ_{u^\varepsilon} x_j$ and $x_i \succ_{v^\varepsilon} x_{n-t+1} \succ_{v^\varepsilon} x_{n-t} \succ_{v^\varepsilon} x_j$ for $i < n - t$, $j > n - t + 1$. Since

$$\operatorname{depth}(S/I) = \operatorname{depth}(S/\operatorname{gin}_{\succ_{u^\varepsilon}}(I)) = \operatorname{depth}(S/\operatorname{gin}_{\succ_{v^\varepsilon}}(I)),$$

by [Proposition 4.3](#) the monomial ideal $\operatorname{gin}_{\succ_{u^\varepsilon}}(I)$ contains a minimal monomial generator divisible by x_{n-t}, but none that is divisible by x_{n-t+1}. On the other hand, $\operatorname{gin}_{\succ_{v^\varepsilon}}(I)$ contains a minimal generator divisible by x_{n-t+1}, but none that is divisible by x_{n-t}. Hence, the reduced Gröbner bases of $\operatorname{in}_{u^\varepsilon}(g(I))$ and $\operatorname{in}_{v^\varepsilon}(g(I))$ are different with respect to the same term order $\succ$. Since the reduced Gröbner basis of an ideal is unique with respect to a given term order, this implies $\operatorname{in}_{u^\varepsilon}(g(I)) \ne \operatorname{in}_{v^\varepsilon}(g(I))$, so u^ε and v^ε are in different cones of $\operatorname{gT}(I)$. So in every neighborhood of ω in $\operatorname{gT}(I)$, there are elements that are in different cones of $\operatorname{gT}(I)$. This is a contradiction to the fact that $\omega \in \mathring{C}$ and C is maximal. $\square$

We can now show that for maximal-gdepth ideals, $\operatorname{gT}(I)$ refines $\mathcal{W}_n^{m,t}$ as a fan.

Proposition 5.6. *Let* $I \subset S = K[x_1, \ldots, x_n]$ *be a maximal-*gdepth *ideal with* $\dim(S/I) = m < n$ *and* $0 < \operatorname{depth}(S/I) = t < m - 1$. *Let* C *be a maximal cone of* $\operatorname{gT}(I)$. *Then there exists a cone* $D \subset \mathcal{W}_n^{m,t}$ *such that* $\mathring{C} \subset \mathring{D}$.

Proof. By Lemma 5.5, we know that $\mathring{C}$ does not intersect the $(m-1)$-skeleton X of $\mathcal{W}_n^{m,t}$. So $\mathring{C}$ must be contained in the union $\mathcal{W}_n^{m,t} \setminus X$ of the open maximal cones of $\mathcal{W}_n^{m,t}$. Since it is convex, $\mathring{C}$ is connected and thus contained in one connected component of $\mathcal{W}_n^{m,t} \setminus X$. But the connected components of $\mathcal{W}_n^{m,t} \setminus X$ are the open maximal cones themselves, so $\mathring{C}$ must be contained in some maximal open cone $\mathring{D}$ of $\mathcal{W}_n^{m,t}$. □

This shows that the fan $\mathrm{gT}(I)$ is always finer than the fan $\mathcal{W}_n^{m,t}$ for maximal-gdepth ideals. We can now give a complementary result by showing that every maximal cone contains a t-dimensional orthant of $\mathbb{R}^n$. For this we will need the following basic observation from Gröbner basis theory.

Lemma 5.7. *Let $I \subset [x_1, \ldots, x_n]$ be a graded ideal and let $\omega, \omega' \in \mathbb{R}^n$. Let $\succ$ be a term order and $\mathcal{G}$ be the reduced Gröbner basis of I with respect to $\succ_\omega$. If $\mathrm{in}_\omega(f) = \mathrm{in}_{\omega'}(f)$ for every $f \in \mathcal{G}$, then $\mathrm{in}_\omega(I) = \mathrm{in}_{\omega'}(I)$.*

Proof. Since $\{\mathrm{in}_\omega(f) : f \in \mathcal{G}\}$ is a reduced Gröbner basis for $\mathrm{in}_\omega(I)$ with respect to $\succ$ (see for example [Maclagan and Thomas 2007, Lemma 2.4.2]), it follows that $\mathrm{in}_\omega(I) = (\mathrm{in}_{\omega'}(f) : f \in \mathcal{G}) \subset \mathrm{in}_{\omega'}(I)$. This implies that $\mathrm{in}_{\succ_\omega}(I) \subset \mathrm{in}_{\succ_{\omega'}}(I)$. As there cannot be a proper inclusion of two initial ideals (see [Maclagan and Thomas 2007, Corollary 2.2.3]), this means $\mathrm{in}_{\succ_\omega}(I) = \mathrm{in}_{\succ_{\omega'}}(I)$. Therefore we have $\mathrm{in}_\omega(I) = \mathrm{in}_{\omega'}(I)$, because $\{\mathrm{in}_\omega(f) : f \in \mathcal{G}\}$ is also a reduced Gröbner basis of $g(I)$ with respect to $\succ_{\omega'}$. □

Proposition 5.8. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a maximal-gdepth ideal with $\dim(S/I) = m$, let $0 < \mathrm{depth}(S/I) = t < m-1$, and let c be the maximal total degree of a minimal generator of a generic initial ideal of I. Let $\omega \in \mathrm{gT}(I)$ with*

$$0 = \omega_1 = \cdots = \omega_{n-m+1} < \omega_{n-m+2} \leq \cdots \leq \omega_{n-t} < \omega_{n-t+1}, \ldots, \omega_n$$

such that $\omega_{n-t}c < \omega_j$ for $j > n-t$ and $\omega \in \mathring{C}$ for some maximal cone C of $\mathrm{gT}(I)$. Then

$$\omega + \mathrm{cone}(e_{n-t+1}, \ldots, e_n) \subset \mathring{C},$$

where e_i denotes the i-th standard basis vector of $\mathbb{R}^n$.

Proof. Let $\succ$ be any term order. Then for the refinement $\succ_\omega$ of ω by Proposition 4.3, the generic initial ideal $\mathrm{gin}_{\succ_\omega}(I)$ is minimally generated in $K[x_1, \ldots, x_{n-t}]$, since I is maximal-gdepth and $\mathrm{depth}(S/\mathrm{gin}_{\succ_\omega}(I)) = \mathrm{depth}(S/I) = t$. Hence, for $g \in U$ (where U is defined as in (2-1)) there exists a reduced Gröbner basis $\mathcal{G}$ of $g(I)$ with respect to $\succ_\omega$ such that $\mathrm{in}_{\succ_\omega}(f) \in K[x_1, \ldots, x_{n-t}]$ for every $f \in \mathcal{G}$.

We show that $\mathrm{in}_{\omega'}(f) \in K[x_1, \ldots, x_{n-t}]$ for every $\omega' \in \omega + \mathrm{cone}(e_{n-t+1}, \ldots, e_n)$ and every $f \in \mathcal{G}$. To see this, we need to show that every term of f that contains one of the $x_{n-t+1}, \ldots, x_n$ has larger ω'-weight than any term of f in $K[x_1, \ldots, x_{n-t}]$. By the choice of c, the ω'-weight of a term of f in $K[x_1, \ldots, x_{n-t}]$ is bounded

from above by $c\omega_{n-t}$. But any of the $x_{n-t+1}, \ldots, x_n$ has weight strictly larger than $c\omega_{n-t}$. Since we already know that f contains a term in $K[x_1, \ldots, x_{n-t}]$, we have $\mathrm{in}_{\omega'}(f) \in K[x_1, \ldots, x_{n-t}]$. By the choice of ω', all terms in $K[x_1, \ldots, x_{n-t}]$ have the same ω-weight and ω'-weight. So $\mathrm{in}_{\omega'}(f) = \mathrm{in}_{\omega}(f)$ for $f \in \mathcal{G}$. Then by Lemma 5.7, it follows that $\mathrm{in}_{\omega}(g(I)) = \mathrm{in}_{\omega'}(g(I))$ for $g \in U$. Hence, $\omega' \in \mathring{C}$ for every $\omega' \in \omega + \mathrm{cone}(e_{n-t+1}, \ldots, e_n)$. $\square$

For maximal-gdepth ideals I, it is therefore possible to obtain $\mathrm{depth}(S/I)$ from the generic tropical variety of I, as shown in the following theorem.

Theorem 5.9. *Let* $I \subset S = K[x_1, \ldots, x_n]$ *be a maximal-gdepth ideal such that* $\dim(S/I) = m$ *and* $0 < \mathrm{depth}(S/I) < m - 1$. *Then*

$$\mathrm{depth}(S/I) = \min\{t \in \mathbb{N} : \mathrm{gT}(I) \text{ refines } \mathcal{W}_n^{m,t}\}.$$

Proof. Let $T = \mathrm{depth}(S/I)$. By Proposition 5.6, we already know that $\mathrm{gT}(I)$ refines $\mathcal{W}_n^{m,T}$ as a fan. On the other hand let $t < T$. Then we can choose $\omega \in \mathbb{R}^n$ such that

$$\omega_1 = \cdots = \omega_{n-m+1} < \omega_{n-m+2}, \ldots, \omega_{n-t} < \omega_{n-t+1}, \ldots, \omega_n$$

with $\omega \in \mathring{C}$ for some maximal cone C of $\mathrm{gT}(I)$ and $\omega_{n-t}c < \omega_j$ for $j > n - t$, where c is chosen as in Proposition 5.8. Define $\omega_i' = \omega_i$ for $i \neq n - t$ and choose $\omega_{n-t}' > \omega_{n-t+1}$. Since $t < T$, by Proposition 5.8 we know that $\omega' \in \mathring{C}$ as well. But by definition of $\mathcal{W}_n^{m,t}$, we know that ω and ω' are in different open cones of $\mathcal{W}_n^{m,t}$. So $\mathrm{gT}(I)$ cannot refine $\mathcal{W}_n^{m,t}$ as a fan. $\square$

Remark 5.10. Note that it is also possible to recover $\mathrm{depth}(S/I)$ from $\mathrm{gT}(I)$ for arbitrary graded ideals $I \subset K[x_1, \ldots, x_n]$ with $\dim(S/I) = m$ and $0 < \mathrm{depth}(S/I) < m - 1$ in the following way. Let $\mathrm{depth}(S/I) = t$ and let $\succ$ be the reverse lexicographic term order. Let c be the maximal degree of a minimal generator of any generic initial ideal with respect to a reverse lexicographic term order. Choose $\omega \in \mathring{C}$ for some maximal cone C of $\mathrm{gT}(I)$ as in Lemma 4.8. We now show that for this particular choice of ω, we have

$$\omega + \mathrm{cone}(e_{n-t+1}, \ldots, e_n) \subset \mathring{C},$$

but

$$\omega + \mathrm{cone}(e_{n-t}, \ldots, e_n) \not\subset \mathring{C}.$$

Since $\succ$ and $\succ_{\omega}$ coincide up to degree c, this implies $\mathrm{gin}_{\succ_{\omega}}(I) = \mathrm{gin}_{\succ}(I)$. In particular, $\mathrm{depth}(S/\mathrm{gin}_{\succ_{\omega}}(I)) = \mathrm{depth}(S/I)$. By the same proof as in Proposition 5.8, we obtain that $\omega + \mathrm{cone}(e_{n-t+1}, \ldots, e_n) \subset \mathring{C}$.

Assume that $\omega + \mathrm{cone}(e_{n-t}, \ldots, e_n) \subset \mathring{C}$. Let $\succ'$ be the reverse lexicographic term order with $x_1 \succ' \cdots \succ' x_{n-t-1} \succ' x_{n-t+1} \succ' \cdots \succ' x_n \succ' x_{n-t}$. Then we define $\omega' \in \mathbb{R}^n$ by $\omega_i' = \omega_i$ for $i \neq n - t$ and $\omega_{n-t}' > \omega_n c$. By assumption, we know that

$\omega' \in \mathring{C}$. Since $\succ_{\omega'}$ and $\succ'$ coincide up to degree c by Lemma 4.8, we know that $\mathrm{gin}_{\succ_{\omega'}}(I) = \mathrm{gin}_{\succ'}(I)$. As in the proof of Theorem 4.9, we get that

$$\mathrm{gin}_{\succ_\omega}(I) = \mathrm{gin}_\succ(I) \neq \mathrm{gin}_{\succ'}(I) = \mathrm{gin}_{\succ_{\omega'}}(I).$$

This implies $\mathrm{in}_\omega(g(I)) \neq \mathrm{in}_{\omega'}(g(I))$ for $g \in U$, which is a contradiction to $\omega, \omega' \in \mathring{C}$. Hence, $\omega + \mathrm{cone}(e_{n-t}, \ldots, e_n) \not\subset \mathring{C}$.

To obtain $\mathrm{depth}(S/I)$ from $\mathrm{gT}(I)$, we can therefore determine ω as described above. Then we have

$$\mathrm{depth}(S/I) = \min\{t : \omega + \mathrm{cone}(e_{n-t+1}, \ldots, e_n) \subset \mathring{C}\}$$

for this particular choice of ω.

As we saw in Proposition 5.6, the generic tropical variety of a maximal-gdepth ideal with $\dim(S/I) = m$ and $\mathrm{depth}(S/I) = t$ with $0 < t < m - 1$ always refines $\mathscr{W}_n^{m,t}$. It is also true that any of the fans $\mathscr{W}_n^{m,t}$ is the generic tropical variety of some ideal, which we will see by focusing on a class of strongly stable ideals generated in degree 2.

Proposition 5.11. *Let* $0 < m < n$ *and* $0 < t < m - 1$. *The ideal*

$$I = (x_1, \ldots, x_{n-m-1}, x_{n-m}^2, x_{n-m}x_{n-m+1}, \ldots, x_{n-m}x_{n-t}) \subset S = K[x_1, \ldots, x_n]$$

*is a maximal-*gdepth *ideal with* $\dim(S/I) = m$, $\mathrm{depth}(S/I) = t$ *and* $\mathrm{gT}(I) = \mathscr{W}_n^{m,t}$ *as a fan.*

Proof. We first show that $\dim(S/I) = m$ and $\mathrm{depth}(S/I) = t$. Since I is strongly stable with respect to the reverse lexicographic order $\succ$, we have $\mathrm{gin}_\succ(I) = I$. So I has only one minimal prime by [Eisenbud 1995, Corollary 15.25], which is $(x_1, \ldots, x_{n-m})$. Thus, $\dim(S/I) = m$. To see that $\mathrm{depth}(S/I) = t$, we note again that $\mathrm{gin}_\succ(I) = I$. By Proposition 4.3, it follows that $\mathrm{depth}(S/I) = n - (n - t) = t$. In particular, I is a maximal-gdepth ideal (see Example 5.2).

By Proposition 5.6, we know that every maximal cone $\mathring{C}$ of $\mathrm{gT}(I)$ is contained in some maximal cone $\mathring{D}$ of $\mathscr{W}_n^{m,t}$. So it remains to show that for every $\omega, \omega' \in \mathring{D}$ for some maximal cone D of $\mathscr{W}_n^{m,t}$, we have $\mathrm{in}_\omega(g(I)) = \mathrm{in}_{\omega'}(g(I))$ for $g \in U$. Let D be the maximal cone of $\mathscr{W}_n^{m,t}$ given by

$$\mathring{D} = \{\omega \in \mathbb{R}^n : \omega_{i_1} = \cdots = \omega_{i_{n-m+1}} < \omega_{i_{n-m+2}}, \ldots, \omega_{i_{n-t}} < \omega_{i_{n-t+1}}, \ldots, \omega_{i_n}\}$$

for some permutation $(i_1, \ldots, i_n)$ of $\{1, \ldots, n\}$. Let $\omega \in \mathring{D}$ be fixed, $\succ_\omega$ be the refinement of ω with respect to the reverse lexicographic order $\succ$ with $x_{i_{n-m+2}} \succ \cdots \succ x_{i_n} \succ x_{i_1} \succ \cdots \succ x_{i_{n-m+1}}$, and $\mathscr{G}$ be the reduced Gröbner basis of $g(I)$ with respect to $\succ_\omega$ for a fixed $g \in U$. Note that $x_{i_1} \succ_\omega \cdots \succ_\omega x_{i_{n-m+1}}$ and $x_{i_k} \succ_\omega x_{i_j}$ for $k \in \{n-m+2, \ldots, n-t\}$, $j \in \{n-t+1, \ldots, n\}$. Let $(q_1, \ldots, q_n)$ be the permutation

on $\{1, \ldots, n\}$ such that $x_{q_1} \succ_\omega x_{q_2} \succ_\omega \cdots \succ_\omega x_{q_n}$. As in Example 5.2, we know $\mathrm{gin}_{\succ_\omega}(I) = \phi(I)$ for the K-algebra isomorphism ϕ induced by $\phi(x_j) = x_{q_j}$. So

$$\mathrm{gin}_{\succ_\omega}(I) = (x_{i_1}, \ldots, x_{i_{n-m-1}}, x_{i_{n-m}}^2, x_{i_{n-m}} x_{i_{n-m+1}}, \ldots, x_{i_{n-m}} x_{i_{n-t}}).$$

Hence, $\mathrm{in}_{\succ_\omega}(f) = x_{i_j}$ for some $j \in \{1, \ldots, n-m-1\}$ or $\mathrm{in}_{\succ_\omega}(f) = x_{i_{n-m}} x_{i_k}$ for some $k \in \{n-m, \ldots, n-t\}$ for every $f \in \mathcal{G}$. Let $\omega' \in \mathring{D}$. We now show that $\mathrm{in}_\omega(f) = \mathrm{in}_{\omega'}(f)$ for every $f \in \mathcal{G}$.

If $\mathrm{in}_{\succ_\omega}(f) = x_{i_j}$ for some $j \in \{1, \ldots, n-m-1\}$, then by comparing weights, $\mathrm{in}_\omega(f)$ is exactly the sum of all linear terms $a_{i_k} x_{i_k}$ that appear in f, with $a_{i_k} \in K$, $k \in \{1, \ldots, n-m-1\}$. But the same is true for $\mathrm{in}_{\omega'}(f)$, since ω and ω' have the same minimal coordinates. So in this case $\mathrm{in}_\omega(f) = \mathrm{in}_{\omega'}(f)$.

If $\mathrm{in}_{\succ_\omega}(f) = x_{i_{n-m}} x_{i_k}$ for some $k \in \{n-m, \ldots, n-t\}$, we have to distinguish two subcases.

Case 1. If $k = n-m$ or $k = n-m+1$, then $\mathrm{in}_\omega(f)$ is the sum of all monomials in $K[x_{n-m}, x_{n-m+1}]$ that appear in f. The same is true for $\mathrm{in}_{\omega'}(f)$ by the same argument as before, so $\mathrm{in}_\omega(f) = \mathrm{in}_{\omega'}(f)$.

Case 2. For $k > n-m+1$, we need to show that certain terms cannot appear in f. First note that no term that is divisible by any of $x_{i_1}, \ldots, x_{i_{n-m-1}}$ can appear in f, since f is part of a reduced Gröbner basis with respect to $\succ_\omega$, and such a term would be divisible by a leading term of another element of $\mathcal{G}$. For the same reason, f cannot contain the monomial $x_{i_{n-m}} x_{i_s}$ for $s \in \{n-m+2, \ldots, n-t\}\setminus\{k\}$. Note that $x_{i_{n-m+1}}^2$ cannot appear in f either, since then $\mathrm{wt}_\omega(x_{i_{n-m+1}}^2) < \mathrm{wt}_\omega(x_{i_{n-m}} x_{i_k})$. Furthermore, assume that f contains the monomial $x_{i_{n-m+1}} x_{i_s}$ for some index $s \in \{n-m+2, \ldots, n-t\}\setminus\{k\}$. Then for $v \in \mathbb{R}^n$ with $v_{i_1} = \cdots = v_{i_{n-m+1}} < v_{i_s} < v_{i_j}$ for $j \in \{n-m+2, \ldots, n\}\setminus\{k\}$, we know $\mathrm{in}_v(f) = x_{i_{n-m+1}} x_{i_s}$ is a monomial, since every other possible term of f has greater v-weight. This is a contradiction to $v \in \mathrm{gT}(I)$. This that implies $x_{i_{n-m+1}} x_{i_s}$ for $s \in \{n-m+2, \ldots, n-t\}\setminus\{k\}$ does not appear in f either.

With this we can determine the initial forms $\mathrm{in}_\omega(f)$ and $\mathrm{in}_{\omega'}(f)$. As we have $\mathrm{wt}_\omega(x_{i_{n-m}} x_{i_k}) = \mathrm{wt}_\omega(x_{i_{n-m+1}} x_{i_k})$, these two terms have to appear in $\mathrm{in}_\omega(f)$, if $x_{i_{n-m+1}} x_{i_k}$ is a term of f. Assume there exists another term in $\mathrm{in}_\omega(f)$; then it would have to be of the form $x_{i_{n-m}} x_{i_r}$ or $x_{i_{n-m+1}} x_{i_r}$ for some $r \in \{n-t+1, \ldots, n\}$, or of the form $x_{i_a} x_{i_b}$ for some $a, b \in \{n-m+2, \ldots, n-t\}$. The former can't occur, since $\mathrm{wt}_\omega(x_{i_{n-m}} x_{i_r}) = \mathrm{wt}_\omega(x_{i_{n-m+1}} x_{i_r}) > \mathrm{wt}_\omega(x_{i_{n-m}} x_{i_k})$. Assume that $x_{i_a} x_{i_b}$ appears in $\mathrm{in}_\omega(f)$ for some $a, b \in \{n-m+2, \ldots, n-t\}$; then of course $\mathrm{wt}_\omega(x_{i_{n-m}} x_{i_k}) = \mathrm{wt}_\omega(x_{i_a} x_{i_b})$. But we know that $x_{i_a} x_{i_b} \succ x_{i_{n-m}} x_{i_k}$, by the choice of $\succ$. This is a contradiction to $\mathrm{in}_{\succ_\omega}(f) = x_{i_{n-m}} x_{i_k}$, so $\mathrm{in}_\omega(f)$ only contains the monomials $x_{i_{n-m}} x_{i_k}$ and $x_{i_{n-m+1}} x_{i_k}$.

The same is true for $\mathrm{in}_{\omega'}(f)$, as we will see. We show that $\mathrm{in}_{\succ_{\omega'}}(f) = x_{i_{n-m}} x_{i_k}$ as well. By the same argument as above, it follows that only the terms $x_{i_{n-m}} x_{i_k}$ and $x_{i_{n-m+1}} x_{i_k}$ appear in $\mathrm{in}_{\omega'}(f)$, and thus $\mathrm{in}_\omega(f) = \mathrm{in}_{\omega'}(f)$. Since $\mathrm{wt}_{\omega'}(x_{i_{n-m}} x_{i_r}) = \mathrm{wt}_{\omega'}(x_{i_{n-m+1}} x_{i_k}) > \mathrm{wt}_{\omega'}(x_{i_{n-m}} x_{i_k})$ for $r \in \{n-t+1, \ldots, n\}$, terms of this form cannot occur as the leading term. Assume that $\mathrm{in}_{\succ_{\omega'}}(f) = x_{i_a} x_{i_b}$ for some $a, b \in \{n-m+2, \ldots, n-t\}$. Then $x_{i_a} x_{i_b} \in \mathrm{gin}_{\succ_{\omega'}}(I)$. But we know that $\dim(S/\mathrm{gin}_{\succ_{\omega'}}(I)) = \dim(S/I) = m$ and $x_{i_1} \succ_{\omega'} \cdots \succ_{\omega'} x_{i_{n-m}} \succ_{\omega'} x_{i_j}$ for $j > n-m$. By Proposition 4.3, this implies that $\mathrm{gin}_{\succ_{\omega'}}(I)$ cannot contain a monomial that is not divisible by one of $x_{i_1}, \ldots, x_{n-m}$, which is a contradiction to $x_{i_a} x_{i_b} \in \mathrm{gin}_{\succ_{\omega'}}(I)$.

We have now shown that $\mathrm{in}_\omega(f) = \mathrm{in}_{\omega'}(f)$ for every $f \in \mathcal{G}$. Hence, by Lemma 5.7, we have $\mathrm{in}_\omega(g(I)) = \mathrm{in}_{\omega'}(g(I))$ for $g \in U$. Thus every maximal cone of $\mathcal{W}_n^{m,t}$ is contained in a maximal cone of $\mathrm{gT}(I)$. The claim now follows from this together with Proposition 5.6. $\qquad\square$

This of course raises the question of whether it is always true that $\mathrm{gT}(I) = \mathcal{W}_n^{m,t}$ for strongly stable ideals or even maximal-gdepth ideals $I \subset S = K[x_1, \ldots, x_n]$ with $\dim(S/I) = m$ and $0 < \mathrm{depth}(S/I) = t < m - 1$. Computations with gfan [Jensen 2009] indicate that this is not the case. For example, the ideal $I = (x_1^2, x_1 x_2, x_1 x_3^2, x_1 x_3 x_4) \subset K[x_1, \ldots, x_5]$ is strongly stable with respect to $x_1 > \cdots > x_5$ and has dimension $\dim(K[x_1, \ldots, x_5]/I) = 4$ and $\mathrm{depth}(K[x_1, \ldots, x_5]/I) = 1$ by Proposition 4.3. However, computing $\mathrm{gT}(I)$ with gfan yields that $\mathrm{gT}(I)$ has 60 maximal cones. Thus $\mathrm{gT}(I) \neq \mathcal{W}_5^{4,1}$, which has only 30 maximal cones.

Remark 5.12. If $\mathrm{depth}(S/I) = 0$, we cannot make a statement about the fan structure of the generic tropical variety of I. To see this, we can, for example, consider the ideals

$$I = (x_1, \ldots, x_{n-m-1}, x_{n-m}^2, x_{n-m} x_{n-m+1}, \ldots, x_{n-m} x_n) \subset S = K[x_1, \ldots, x_n]$$

for $0 < m < n$, which are the ideals of Proposition 5.11 for $t = 0$. By the same argument as before, we have $\dim(S/I) = m$ and $\mathrm{depth}(S/I) = 0$. Hence, by Corollary 3.8 we know that $\mathrm{gT}(I)$ as a fan is a refinement of $\mathcal{W}_n^m$. Using the same arguments as in the proof of Proposition 5.11, we can show that $\mathrm{in}_\omega(g(I)) = \mathrm{in}_{\omega'}(g(I))$ for $\omega, \omega' \in \mathring{C}$ for any maximal cone C of $\mathcal{W}_n^m$ for every $g \in U$ (as defined in (2-1)). This shows that $\mathrm{gT}(I)$ is equal to $\mathcal{W}_n^m$ as a fan.

On the other hand, we can find ideals with $\mathrm{depth}(S/I) = 0$ whose generic tropical variety is a proper refinement of $\mathcal{W}_n^m$. For example, for the ideal $I = (x_1^2, x_1 x_2, x_1 x_3^2, x_1 x_3 x_4) \subset K[x_1, \ldots, x_4]$ we have $\dim(K[x_1, \ldots, x_4]/I) = 3$ and $\mathrm{depth}(K[x_1, \ldots, x_4]/I) = 0$ by Proposition 4.3. If we compute $\mathrm{gT}(I)$ with gfan, however, we obtain a fan with 12 maximal cones that refines the fan $\mathcal{W}_4^3$ with only 6 maximal cones.

6. Multiplicities

Let $S = K[x_1, \ldots, x_n]$ as before. For a finitely generated graded S-module M, we denote by $H_M(t)$ the Hilbert series of M. Recall that the Hilbert series of $0 \neq M$ can be written as

$$H_M(t) = \frac{Q_M(t)}{(1-t)^d},$$

where $Q_M(t) \in \mathbb{Z}[t, t^{-1}]$ is a Laurent polynomial with $Q_M(1) \neq 0$ and d is the Krull dimension of M. It is well-known that $Q_M(1) \neq 0$, and this number is called the *multiplicity* $m(M)$ *of* M. As always, we set $m(I) = m(S/I) = Q_{S/I}(1)$ for a graded ideal I. We call this the *multiplicity of* I, although more precisely it is the multiplicity of S/I.

To express the multiplicity of I in terms of the multiplicities of its minimal primes, we use the following formula, known as the associativity formula for multiplicities. Note that all minimal prime ideals of a graded ideal are graded themselves. For a minimal prime ideal P of I, let $\ell((S/I)_P)$ denote the length of the localization of the S-module S/I at P. We then have

$$m(I) = \sum \ell((S/I)_P) m(P),$$

with the sum taken over all minimal primes of I such that $\dim(S/I) = \dim(S/P)$; see [Vasconcelos 1998, Formula (9. 4)].

We define the multiplicity of a maximal cone in $T(I)$ in a slightly more general setting than in [Dickenstein et al. 2007], where the multiplicity of a maximal cone C in $T(P)$ for a prime ideal P is defined as the sum of the multiplicities of all monomial-free minimal primes of the initial ideal $\text{in}_C(P)$ corresponding to C. Note that by [Gräbe 1993, Theorem 1], for every minimal prime Q of $\text{in}_C(P)$, we have $\dim(S/Q) = \dim(S/\text{in}_C(P))$. For an arbitrary ideal $I \subset S = K[x_1, \ldots, x_n]$ this is not true, and in our definition we consider only those prime ideals P of $\text{in}_C(I)$ such that S/P has the same dimension as $S/\text{in}_C(I)$.

Definition 6.1. Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal and C be a maximal cone of $T(I)$. Let $J = \text{in}_C(I)$ be the initial ideal of I corresponding to C. Then the *intrinsic multiplicity* $m(C)$ of C is defined as $m(C) = \sum \ell((S/J)_P)$, where the sum is taken over all minimal primes P of J with $\dim(S/P) = \dim(S/J)$ that do not contain a monomial.

Note that in general $T(I)$ need not be pure, so in general this definition of intrinsic multiplicities will not give rise to a tropical fan as defined in [Gathmann et al. 2009, Definition 2.8]. However, we only need this definition for generic tropical varieties, and these are pure by Proposition 3.2. Even if I is a radical ideal and $T(I)$ a pure fan, the multiplicity of the cones of $T(I)$ need not have anything to do with the multiplicity of I, as the following example shows.

Example 6.2. Let $0 \le k \le n$ and $f_k = x_1 \cdots x_k(x_1 + x_2) \in K[x_1, \ldots, x_n]$. Then $m(f_k) = \deg(f_k) = k + 1$. But we see that for $I = (f_k)$, the tropical variety

$$T(I) = \{\omega \in \mathbb{R}^n : \omega_1 = \omega_2\}$$

consists of only one cone. The corresponding initial ideal is (f_k). By factorization, this has only one monomial-free minimal prime ideal, which is $(x_1 + x_2)$. Since $\ell((S/(f_k))_{(x_1+x_2)}) = 1$, the only cone of $\mathrm{gT}(I)$ has multiplicity 1. So in general it is impossible to obtain the multiplicity of the ideal from the multiplicity of the maximal cones of the tropical variety, at least for ideals that are not prime.

In contrast, we can now prove that generically, the intrinsic multiplicities of the maximal cones in the tropical variety are constant and equal to the multiplicity of the ideal. For this, we first show that for a graded ideal I, the minimal prime ideals of the initial ideals of I that correspond to the maximal cones in $\mathrm{gT}(I)$ contain no monomial. Recall that by U, we denote the Zariski-open subset of $\mathrm{GL}_n(K)$ as defined in (2-1).

Proposition 6.3. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m$. Let C be a maximal cone of $\mathrm{gT}(I)$ and $\omega \in \mathring{C}$. Then no minimal prime P of $\mathrm{in}_\omega(g(I))$ with $\dim(S/P) = m$ contains a monomial for $g \in U$.*

Proof. Since $\mathrm{gT}(I) = \mathscr{W}_n^m$ as a set, we can assume $\omega_1 = \cdots = \omega_{n-m+1} < \omega_j$ for $j > n - m + 1$ without loss of generality. For $g \in U$, let $\mathrm{in}_\omega(g(I)) \subset P$ be a minimal prime ideal with $\dim(S/P) = m$. Assume that P contains a monomial x^ν. Since P is prime, this implies that P contains a variable x_k for some k. We choose $\{i_1, \ldots, i_{n-m}\} \subset \{1, \ldots, n-m+1\} \setminus \{k\}$ and a term order $\succ$ such that

$$x_{i_1} \succ x_{i_2} \succ \cdots \succ x_{i_{n-m}} \succ x_j \text{ for } j \notin \{i_1, \ldots, i_{n-m}\}.$$

Then

$$\mathrm{gin}_{\succ_\omega}(I) = \mathrm{in}_\succ(\mathrm{in}_\omega(g(I))) \subset \mathrm{in}_\succ(P),$$

with $\dim(S/\mathrm{gin}_{\succ_\omega}(I)) = \dim(S/\mathrm{in}_\succ(P)) = m$. Let Q be a minimal prime of $\mathrm{in}_\succ(P)$. Since the dimensions coincide, Q is also a minimal prime of $\mathrm{gin}_{\succ_\omega}(I)$. But $\mathrm{gin}_{\succ_\omega}(I)$ has only one minimal prime, which is $(x_{i_1}, \ldots, x_{i_{n-m}})$ by the choice of the term order $\succ$; see for example [Eisenbud 1995, Corollary 15.25]. Hence, Q does not contain x_k. This is a contradiction to the fact that $x_k \in P$, and therefore $x_k \in \mathrm{in}_\succ(P) \subset Q$. Thus, P cannot contain a monomial. $\square$

Remark 6.4. Note that together with [Römer and Schmitz 2009, Lemma 7.2], where $\mathrm{gT}(I)$ can be replaced by $T(g(I))$ for every $g \in U$, and [Römer and Schmitz 2009, Corollary 3.2], this gives another, simpler proof that generic tropical varieties exist as described in [Römer and Schmitz 2009].

To use the associativity formula for multiplicities to show that $m(C) = m(I)$ in generic tropical varieties, we need to show that generically all minimal primes of $\text{in}_\omega(g(I))$ have multiplicity 1. This we do by showing that they are linear, that is, generated by linear forms.

Lemma 6.5. *Let $P \subset K[x_1, \ldots, x_n]$ be a graded prime ideal with $\dim(S/P) = 1$. Then P is a linear ideal.*

Proof. As $P \neq (x_1, \ldots, x_n)$, we know that $V(P) \neq \{0\}$. Let

$$0 \neq a = (a_1, \ldots, a_n) \in V(P) \subset K^n.$$

Then $V(Q) = K(a_1, \ldots, a_n)$ for the linear ideal $Q = (a_i x_j - a_j x_i : i < j)$. Since $V(Q) \subset V(P)$ and both are prime, this implies $P \subset Q \subset (x_1, \ldots, x_n)$. But $\dim(S/P) = 1$ and $Q \neq (x_1, \ldots, x_n)$, hence $P = Q$ is linear. $\qquad\square$

Lemma 6.6. *For a fixed $t < n$, let $R = K[x_1, \ldots, x_t]$. Let $J \subset S = K[x_1, \ldots, x_n]$ be a graded ideal and $J \subset P \subset S$ be a minimal prime of J with $\dim(S/J) = \dim(S/P) = m$. If $(J \cap R)S = J$, then also $(P \cap R)S = P$.*

Proof. It is clear that $(P \cap R)S \subset P$. As $J \subset P$, we know that $J = (J \cap R)S \subset (P \cap R)S \subset P$. Since P is prime, so are $P \cap R$ and $(P \cap R)S$. But P is a minimal prime of J, and hence $(P \cap R)S = P$. $\qquad\square$

With this we can prove that for $I \subset K[x_1, \ldots, x_n]$, the minimal primes of the initial ideals corresponding to the maximal cones of $\text{gT}(I)$ of the same dimension as S/I have multiplicity 1.

Proposition 6.7. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m$ and $\omega \in \mathring{C}$ for some maximal cone C of $\text{gT}(I)$. Then for every $g \in U$, every minimal prime P of $\text{in}_\omega(g(I))$ with $\dim(S/P) = \dim(S/\text{in}_\omega(g(I)))$ is a linear ideal. In particular, $m(P) = 1$.*

Proof. Without loss of generality we can assume $\omega_1 = \cdots = \omega_{n-m+1} < \omega_j$ for $j > n - m + 1$. Let $g \in U$, and let $\text{in}_\omega(g(I)) \subset P$ be a minimal prime with $\dim(S/P) = m$. Let $\mathcal{G} = \{f_1, \ldots, f_t\}$ be a reduced Gröbner basis of $g(I)$ with respect to $\succ_\omega$ for a term order $\succ$ with $x_1 \succ \cdots \succ x_n$. Then

$$(\text{in}_{\succ_\omega}(f_i) : i = 1, \ldots, t) = \text{in}_\succ(\text{in}_\omega(g(I))) = \text{gin}_{\succ_\omega}(I).$$

Note that $x_1 \succ_\omega \cdots \succ_\omega x_{n-m+1} \succ_\omega x_j$ for $j > n - m + 1$. Let $A \subset \{1, \ldots, t\}$ be the set of all indices i such that $\text{in}_{\succ_\omega}(f_i) \in K[x_1, \ldots, x_{n-m}]$. We define $\tilde{J} = (\text{in}_{\succ_\omega}(f_i) : i \in A)$ to be the ideal generated by all initial forms of elements in $\mathcal{G}$ that are not divisible by $x_{n-m+1}, \ldots, x_n$. Since $\tilde{J} \subset \text{gin}_{\succ_\omega}(I)$, we know that $\dim(S/\tilde{J}) \geq m$. As $\text{gin}_{\succ_\omega}(I)$ is a strongly stable ideal, by [Proposition 4.3](#) there exists $1 \leq k \leq t$ such that $\text{in}_{\succ_\omega}(f_k) = x_{n-m}^d$ for some $d \in \mathbb{N}$. Hence, $x_{n-m}^d \in \tilde{J}$. But $\tilde{J}$ is also a

strongly stable ideal, so again by Proposition 4.3 it follows that $\dim(S/\tilde{J}) \leq m$. Thus, $\dim(S/\tilde{J}) = m$. We set $J = (\mathrm{in}_\omega(f_i) : i \in A)$. Then we have

$$m = \dim(S/\mathrm{in}_\omega(g(I))) \leq \dim(S/J) = \dim(S/\mathrm{in}_\succ(J)) \leq \dim(S/\tilde{J}) = m,$$

where the first inequality holds because $J \subset \mathrm{in}_\omega(g(I))$, and the second because $\tilde{J} \subset \mathrm{in}_\succ(J)$. So $\dim(S/J) = m$. Since $J \subset \mathrm{in}_\omega(g(I)) \subset P$ and all considered rings have the same dimension, P is also a minimal prime ideal of J.

Let R be the polynomial ring $K[x_1, \ldots, x_{n-m+1}]$, so $S = R[x_{n-m+2}, \ldots, x_n]$. For $i \in A$, every term of f_i that has minimal ω-weight has to be a term in R by the choice of ω. So we know that $J = (J \cap R)S$. From Lemma 6.6 it now follows that $P = (P \cap R)S$. Let $\tilde{P} = P \cap R$. Then we have

$$S/P = S/(\tilde{P}S) \cong R/\tilde{P}[x_{n-m+2}, \ldots, x_n].$$

Hence, $\dim(S/P) = \dim(R/\tilde{P}) + (m-1)$, so $R/\tilde{P}$ has dimension $m - (m-1) = 1$ in R. By Lemma 6.5, we know that $\tilde{P}$ is linear. So $P = \tilde{P}S$ is linear as well and in particular, $m(P) = 1$. $\square$

Theorem 6.8. *Let $I \subset S = K[x_1, \ldots, x_n]$ be a graded ideal with $\dim(S/I) = m$. Then for $g \in U$ and any maximal cone C of $T(g(I))$, we have $m(C) = m(I)$, so the intrinsic multiplicity of every maximal cone equals the multiplicity of I.*

Proof. First note that the Hilbert series and thus the multiplicity of I does not change if one passes to any initial ideal of I; see for example [Eisenbud 1995, Theorem 15.26]. Moreover, the Hilbert series is of course not affected by a coordinate change.

By Proposition 6.3, for $g \in U$ and any maximal cone C of $T(g(I)) = gT(I)$, we know that every minimal prime P of $\mathrm{in}_C(g(I))$ with $\dim(S/P) = m$ does not contain a monomial. Moreover, by Proposition 6.7, every such minimal prime P of $\mathrm{in}_C(g(I))$ has multiplicity $m(P) = 1$. Thus with the associativity formula for multiplicities, we get

$$m(C) = \sum \ell((S/\mathrm{in}_C(g(I)))_P) = \sum \ell((S/\mathrm{in}_C(g(I)))_P)m(P)$$
$$= m(\mathrm{in}_C(g(I))) = m(I),$$

as the sum is taken over all minimal primes of $\mathrm{in}_C(g(I))$ with $\dim(S/P) = m$. $\square$

Remark 6.9. The fan $gT(I)$ equipped with the weights $m(C)$ for the maximal cones $C \in gT(I)$ is a tropical fan in the sense of [Gathmann et al. 2009, Definition 2.8]. It can be shown directly by elementary methods that the balancing condition is fulfilled for each cone of dimension $\dim(S/I) - 1$. See [Speyer 2005, Theorem 2.5.1] for a proof in a more general case.

We briefly explain [Dickenstein et al. 2007, Example (1)] in our case. This example states that for an irreducible polynomial $f \in K[x_1, \ldots, x_n]$, the intrinsic multiplicity $m(C)$ of a given cone C of $T(f)$ is exactly the lattice length of the edge corresponding to C in the Newton polytope of f. Here, the lattice length of an edge is defined as the number of integer points on this edge minus 1.

Example 6.10. Let $0 \neq f \in K[x_1, \ldots, x_n]$ be a homogeneous polynomial of degree t. Then every maximal cone of $gT(f)$ has multiplicity t, as $m(g(f)) = \deg g(f) = t$ for every $g \in U$. Let $N(g(f))$ be the Newton polytope of $g(f)$ for $g \in U$. By [Römer and Schmitz 2009, Lemma 8.5], for $g \in U$ we know that

$$N(g(f)) = \operatorname{conv}(te_1, \ldots, te_n),$$

where $e_1, \ldots, e_n$ are the standard basis vectors in $\mathbb{R}^n$. Now a maximal cone C of $gT(f)$ is given by

$$C = \{\omega \in \mathbb{R}^n : \omega_{i_1} = \omega_{i_2} \leq \omega_{i_j} \text{ for } j \neq 1, 2\}$$

for some coordinates i_1, i_2. This corresponds to the edge $\operatorname{conv}(te_{i_1}, te_{i_2})$ of $N(g(f))$ for $g \in U$. This edge has lattice length t, that is, $|\{\mathbb{Z}^n \cap \operatorname{conv}(te_{i_1}, te_{i_2})\}| = t + 1$. So the lattice length coincides with the intrinsic multiplicity $m(C)$.

Acknowledgments

References

[Bieri and Groves 1984] R. Bieri and J. R. J. Groves, "The geometry of the set of characters induced by valuations", *J. Reine Angew. Math.* **347** (1984), 168–195. MR 86c:14001 Zbl 0526.13003

[Bogart et al. 2007] T. Bogart, A. N. Jensen, D. Speyer, B. Sturmfels, and R. R. Thomas, "Computing tropical varieties", *J. Symbolic Comput.* **42**:1-2 (2007), 54–73. MR 2007j:14103 Zbl 1121.14051

[Bruns and Conca 2004] W. Bruns and A. Conca, "Gröbner bases, initial ideals and initial algebras", in *Homological methods in commutative algebra* (Tehran, 2004), edited by L. L. Avramov et al., 2004.

[Bruns and Gubeladze 2009] W. Bruns and J. Gubeladze, *Polytopes, rings, and K-theory*, Springer, Dordrecht, 2009. MR 2010d:19001 Zbl 1168.13001

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR 95h:13020 Zbl 0788.13005

[Develin and Sturmfels 2004] M. Develin and B. Sturmfels, "Tropical convexity", *Doc. Math.* **9** (2004), 1–27. Erratum in **9** (2004), 205–206. MR 2005i:52010 Zbl 1054.52004

[Dickenstein et al. 2007] A. Dickenstein, E. M. Feichtner, and B. Sturmfels, "Tropical discriminants", *J. Amer. Math. Soc.* **20**:4 (2007), 1111–1133. MR 2008j:14095 Zbl 1166.14033

[Draisma 2008] J. Draisma, "A tropical approach to secant dimensions", *J. Pure Appl. Algebra* **212**:2 (2008), 349–363. MR 2008j:14102 Zbl 1126.14059

[Eisenbud 1995] D. Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. MR 97a:13001 Zbl 0819.13001

[Eisenbud and Goto 1984] D. Eisenbud and S. Goto, "Linear free resolutions and minimal multiplicity", *J. Algebra* **88**:1 (1984), 89–133. MR 85f:13023 Zbl 0531.13015

[Eliahou and Kervaire 1990] S. Eliahou and M. Kervaire, "Minimal resolutions of some monomial ideals", *J. Algebra* **129**:1 (1990), 1–25. MR 91b:13019 Zbl 0701.13006

[Gathmann 2006] A. Gathmann, "Tropical algebraic geometry", *Jahresber. Deutsch. Math.-Verein.* **108**:1 (2006), 3–32. MR 2007e:14088 Zbl 1109.14038

[Gathmann and Markwig 2008] A. Gathmann and H. Markwig, "Kontsevich's formula and the WDVV equations in tropical geometry", *Adv. Math.* **217**:2 (2008), 537–560. MR 2370275 Zbl 1131.14057

[Gathmann et al. 2009] A. Gathmann, M. Kerber, and H. Markwig, "Tropical fans and the moduli spaces of tropical curves", *Compos. Math.* **145**:1 (2009), 173–195. MR 2009m:14085 Zbl 1169.51021

[Gräbe 1993] H.-G. Gräbe, "Two remarks on independent sets", *J. Algebraic Combin.* **2**:2 (1993), 137–145. MR 94e:13051

[Herzog and Srinivasan 1998] J. Herzog and H. Srinivasan, "Bounds for multiplicities", *Trans. Amer. Math. Soc.* **350**:7 (1998), 2879–2902. MR 99g:13033 Zbl 0899.13026

[Itenberg et al. 2007] I. Itenberg, G. Mikhalkin, and E. Shustin, *Tropical algebraic geometry*, Oberwolfach Seminars **35**, Birkhäuser, Basel, 2007. MR 2008e:14082 Zbl 1162.14300

[Jensen 2007] A. N. Jensen, *Algorithmic aspects of Gröbner fans and tropical varieties*, Ph.D. thesis, Aarhus University, 2007, Available at http://tinyurl.com/2ehs2ho.

[Jensen 2009] A. N. Jensen, "Gfan, a software system for Gröbner fans and tropical varieties", version 0.4, 2009, Available at http://www.math.tu-berlin.de/~jensen/software/gfan/gfan.html.

[Jensen et al. 2008] A. N. Jensen, H. Markwig, and T. Markwig, "An algorithm for lifting points in a tropical variety", *Collect. Math.* **59**:2 (2008), 129–165. MR 2009a:14077 Zbl 1151.13021

[Katz et al. 2008] E. Katz, H. Markwig, and T. Markwig, "The j-invariant of a plane tropical cubic", *J. Algebra* **320**:10 (2008), 3832–3848. MR 2010b:14122 Zbl 1185.14030

[Maclagan and Thomas 2007] D. Maclagan and R. R. Thomas, "Computational algebra and combinatorics of toric ideals", pp. 1–106 in *Commutative algebra and combinatorics* (Allahabad, 2003), vol. 1, edited by R. V. Gurjar et al., Ramanujan Math. Soc. Lect. Notes Ser. **4**, Ramanujan Math. Soc., Mysore, 2007. MR 2009c:13069

[Mikhalkin 2006] G. Mikhalkin, "Tropical geometry and its applications", pp. 827–852 in *International Congress of Mathematicians* (Madrid, 2006), vol. 2, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2006. MR 2008c:14077 Zbl 1103.14034

[Mora and Robbiano 1988] T. Mora and L. Robbiano, "The Gröbner fan of an ideal", *J. Symbolic Comput.* **6**:2-3 (1988), 183–208. MR 90d:13004 Zbl 0668.13017

[Römer and Schmitz 2009] T. Römer and K. Schmitz, "Generic tropical varieties", preprint, 2009. arXiv 0904.0120

[Speyer 2005] D. E. Speyer, *Tropical geometry*, ProQuest LLC, Ann Arbor, MI, 2005. MR 2623018

[Speyer and Sturmfels 2004] D. Speyer and B. Sturmfels, "The tropical Grassmannian", *Adv. Geom.* **4**:3 (2004), 389–411. MR 2005d:14089 Zbl 1065.14071

[Sturmfels 1996] B. Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series **8**, American Mathematical Society, Providence, RI, 1996. MR 97b:13034 Zbl 0856.13020

[Vasconcelos 1998] W. V. Vasconcelos, *Computational methods in commutative algebra and algebraic geometry*, Algor. Comput. Math. **2**, Springer, Berlin, 1998. MR 99c:13048 Zbl 0896.13021

troemer@uos.de   *Institut für Mathematik, Universität Osnabrück, 49069 Osnabrück, Germany*

kischmit@uos.de   *Institut für Mathematik, Universität Osnabrück, 49069 Osnabrück, Germany*

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 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 LaTeX but submissions in other varieties of TeX, 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 BibTeX 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@mathscipub.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