Algebra & Number Theory

Volume 18
2024
No. 7

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2024 is US $525/year for the electronic version, and $770/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 18:7 (2024)
https://doi.org/10.2140/ant.2024.18.1221

Serre weights for three-dimensional wildly ramified Galois representations

Daniel Le, Bao V. Le Hung, Brandon Levin and Stefano Morra

We formulate and prove the weight part of Serre's conjecture for three-dimensional mod p Galois representations under a genericity condition when the field is unramified at p. This removes the assumption made previously that the representation be tamely ramified at p. We also prove a version of Breuil's lattice conjecture and a mod p multiplicity one result for the cohomology of $U(3)$-arithmetic manifolds. The key input is a study of the geometry of the Emerton–Gee stacks using the local models we introduced previously (2023).

1. Introduction

The goal of this paper is to prove a generalization of the weight part of Serre's conjecture for three-dimensional mod p Galois representations which are *generic* at p. We also prove a generalization of Breuil's lattice conjecture for these representations and the Breuil–Mézard conjecture for generic tamely potentially crystalline deformation rings of parallel weight $(2, 1, 0)$. For a detailed discussion of these conjectures, see [Le et al. 2020], where we establish the tame case of these conjectures.

1.1. *Results.*

1.1.A. *The weight part of Serre's conjecture.* Let p be a prime, and let F/F^+ be a CM extension of a totally real field $F^+ \neq \mathbb{Q}$. Assume that all places in F^+ above p split in F/F^+. Let G be a definite unitary group over F^+ split over F which is isomorphic to $U(3)$ at each infinite place and split at each place above p. A (*global*) *Serre weight* is an irreducible $\overline{\mathbb{F}}_p$-representation V of $G(\mathcal{O}_{F^+,p})$. These are all of the form $\bigotimes_{v|p} V_v$ with V_v an irreducible $\overline{\mathbb{F}}_p$-representation of $G(k_v)$, where k_v is the residue field of

MSC2020: 11F33, 11F80, 22E50.
Keywords: weight part of Serre's conjectures, local models for Galois representations, mod p cohomology of arithmetic manifolds, congruences of automorphic forms.

F^+ at v. For a mod p Galois representation $\bar{r}: G_F \to \mathrm{GL}_3(\overline{\mathbb{F}}_p)$, let $W(\bar{r})$ denote the collection of modular Serre weights for $\bar{r}$. That is, $V \in W(\bar{r})$ if the Hecke eigensystem attached to $\bar{r}$ appears in a space of mod p automorphic forms on G of weight V for some prime to p level.

For each place $v \mid p$, fix a place $\tilde{v}$ of F dividing v which identifies $G(k_v)$ with $\mathrm{GL}_3(k_v)$. Define $\bar{\rho}_v := \bar{r}|_{\mathrm{Gal}(\bar{F}_{\tilde{v}}/F_{\tilde{v}})}$. We can now state the main theorem.

Theorem 1.1.1 (Theorem 5.4.2). *Assume that p is unramified in F and that $\bar{\rho}_v$ is 8-generic for all $v \mid p$. Assume that $\bar{r}$ is modular* (*i.e.*, *$W(\bar{r})$ is nonempty*) *and satisfies the Taylor–Wiles hypotheses.*

Then

$$\bigotimes_{v \mid p} V_v \in W(\bar{r}) \Leftrightarrow V_v \in W^g(\bar{\rho}_v) \text{ for all } v \mid p,$$

where $W^g(\bar{\rho}_v)$ is an explicit set of irreducible $\overline{\mathbb{F}}_p$-representations of $\mathrm{GL}_3(k_v)$ attached to $\bar{\rho}_v$ (*see Definition 1.2.1*).

In particular, this affirms the expectation from local-global compatibility that $W(\bar{r})$ depends only on the restrictions of $\bar{r}$ to places above p.

Remark 1.1.2. This is the first complete description of $W(\bar{r})$ in dimension greater than two for representations $\bar{r}$ that are wildly ramified above p. Some lower bounds were previously obtained in [Gee and Geraghty 2012; Morra and Park 2017; Herzig et al. 2017; Le et al. 2018a; 2018b].

The first obstacle we overcome is the lack of a conjecture. One basic problem is that while tame representations (when restricted to inertia) depend only on discrete data, wildly ramified representations vary in nontrivial moduli. Buzzard–Diamond–Jarvis defined a recipe in terms of crystalline lifts in dimension two. However, after [Le et al. 2018a], it was clear that the crystalline lifts perspective is insufficient in higher dimension. In higher dimensions, Herzig defined a combinatorial/representation theoretic recipe for a collection of weights $W^?(\bar{\rho}_v)$ when $\bar{\rho}_v$ is tame. For possibly wildly ramified $\bar{\rho}$, Gee et al. [2018] make a conjectural conjecture: they define a conjectural set *conditional* on a version of the Breuil–Mézard conjecture. Our first step is to prove a version of the Breuil–Mézard conjecture (Theorem 1.2.2 and Remark 1.2.3) when $n = 3$.

Having established a version of the Breuil–Mézard conjecture when $n = 3$, the weight set from [Gee et al. 2018] turns out (in generic cases) to have a simple geometric interpretation. Let $\mathcal{X}_3$ be the moduli stack of (φ, Γ)-modules recently constructed by Emerton and Gee [2023]. The irreducible components of $\mathcal{X}_3$ are labeled by irreducible mod p representations of $\mathrm{GL}_3(k_v)$ and $W^g(\bar{\rho}_v)$ is defined so that $V_v \in W^g(\bar{\rho}_v)$ if and only if $\bar{\rho}_v$ lies on $\mathcal{C}_{V_v}$. However, this definition of $W^g(\bar{\rho}_v)$ gives very little insight into its structure. We study $W^g(\bar{\rho}_v)$ using the local models developed in [Le et al. 2023b] combined with the explicit calculations of tamely potentially crystalline deformation rings in [Le et al. 2018a; 2020]. We ultimately arrive at an explicit description of all possible weight sets $W^g(\bar{\rho}_v)$ which allows us to then employ the Taylor–Wiles patching method to prove Theorem 1.1.1.

1.1.B. *Breuil's lattice conjecture and mod p multiplicity one.* The weight part of Serre's conjecture can be viewed as a local-global compatibility result in the mod p Langlands program. In this section, we mention two further local-global compatibility results — one mod p and one p-adic. We direct the reader to the introduction of [Le et al. 2020] for further context for the following two results.

In the global setup above assume further that F/F^+ is unramified at all finite places and G is quasisplit at all finite places. Let $r : G_F \to \mathrm{GL}_3(\overline{\mathbb{Q}}_p)$ be a modular Galois representation which is tamely potentially crystalline with Hodge–Tate weights (2, 1, 0) at each place above p and unramified outside p (though our results hold true when r is minimally split ramified; see Section 5.4). Write λ for the Hecke eigensystem corresponding to r. We fix places $\tilde{v}\,|\,v\,|\,p$ of F and F^+ respectively. We let $\widetilde{H}$ be the integral p-adically completed cohomology with infinite level at v, hyperspecial level outside v, and constant coefficients. Set $\rho \overset{\text{def}}{=} r|_{G_{F_{\tilde{v}}}}$ and let $\sigma(\tau)$ be the tame type corresponding to the Weil–Deligne representation associated to ρ under the inertial local Langlands correspondence (so that $\widetilde{H}[\lambda][1/p]$ contains $\sigma(\tau)$ with multiplicity one). Let $\bar{r}$ and $\bar{\rho}$ be the reductions of r and ρ, respectively.

Theorem 1.1.3 (Theorem 5.4.4). *Assume that p is unramified in F^+, r is unramified outside p, $\bar{\rho}$ is 11-generic, and $\bar{r}$ satisfies Taylor–Wiles hypotheses. Then, the lattice*

$$\sigma(\tau) \cap \widetilde{H}[\lambda]$$

depends only on ρ.

We now let $\overline{H}$ be the mod p reduction of $\widetilde{H}$. Thus, $\overline{H}$ is the mod p cohomology with infinite level at v (and hyperspecial level at places outside v with constant coefficients) of a $U(3)$-arithmetic manifold.

Theorem 1.1.4 (Theorem 5.4.3). *Let $\sigma(\tau)^\circ$ be an $\mathcal{O}$-lattice in $\sigma(\tau)$ with irreducible "upper alcove" cosocle. Under the assumptions of Theorem 1.1.3, $\mathrm{Hom}_{\mathrm{GL}_3(\mathcal{O}_{F_{\tilde{v}}})}(\sigma(\tau)^\circ, \overline{H}[\lambda])$ is a one-dimensional $\overline{\mathbb{F}}_p$-vector space.*

See Sections 1.4 and 2.1.F, for the notion of upper and lower alcove for Serre weights for GL_3. The statement of Theorem 1.1.4 is also true when the cosocle is not necessarily upper alcove if one imposes a condition on the shape of $\bar{\rho}$ with respect to τ; see Theorem 5.4.3.

1.2. *Methods.*

1.2.A. *Local methods: geometry of the Emerton–Gee stack and local models.* We begin by recalling the set $W^g(\bar{\rho})$ that appears in Theorem 1.1.1. Let K be a finite unramified extension of $\mathbb{Q}_p$ of degree f, with ring of integers $\mathcal{O}_K$ and residue field k. Let $\mathcal{X}_{K,n}$ be the Noetherian formal algebraic stack over $\mathrm{Spf}\,\mathbb{Z}_p$ defined in [Emerton and Gee 2023, Definition 3.2.1]. It has the property that for any complete local Noetherian $\mathbb{Z}_p$-algebra R, the groupoid $\mathcal{X}_{K,n}(R)$ is equivalent to the groupoid of rank n projective R-modules equipped with a continuous G_K-action; see [Emerton and Gee 2023, §3.6.1]. In particular, $\mathcal{X}_{K,n}(\overline{\mathbb{F}}_p)$ is the groupoid of continuous Galois representations $\bar{\rho} : G_K \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$. As explained in [Le et al. 2023b, §7.4], there is a bijection $\sigma \mapsto \mathcal{C}_\sigma$ between irreducible $\overline{\mathbb{F}}_p$-representations of $\mathrm{GL}_n(k)$ and the irreducible components of the reduced special fiber of $\mathcal{X}_{K,n}$. (This is a relabeling of the bijection of [Emerton and Gee 2023, Theorem 6.5.1].)

Definition 1.2.1. Let $\bar{\rho} \in \mathcal{X}_{K,n}(\overline{\mathbb{F}}_p)$. Define the set of *geometric weights* of $\bar{\rho}$ to be

$$W^g(\bar{\rho}) = \{\sigma \mid \bar{\rho} \in \mathcal{C}_\sigma(\overline{\mathbb{F}}_p)\}.$$

While Definition 1.2.1 is simple, it does not appear to be an easy task to determine the possible sets $W^g(\bar{\rho})$. The irreducible components of $\mathcal{X}_{K,n}$ are described in terms of closures of substacks, but we expect the closure relations and component intersections in $\mathcal{X}_{K,n}$ to be rather complicated.

We now specialize to the case $n = 3$. A key tool in the analysis of the sets $W^g(\bar{\rho})$ in this setting is the description of certain potentially crystalline substacks. For a tame inertial type τ, let $\mathcal{X}^{\eta,\tau} \subset \mathcal{X}_{K,3}$ be the substack parametrizing potentially crystalline representations of type τ and parallel weight $(2, 1, 0)$. Recall that $\sigma(\tau)$ denotes the representation of $\mathrm{GL}_3(\mathcal{O}_K)$ obtained by applying the inertial local Langlands correspondence to τ (it is the inflation of a Deligne–Lusztig representation; see Section 2.1.C). The following is an application of the theory of local models of [Le et al. 2023b]:

Theorem 1.2.2 (Corollary 3.3.3). *If τ is a 4-generic tame inertial type, then $\mathcal{X}^{\eta,\tau}$ is normal and Cohen–Macaulay and its special fiber $\mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ is reduced. Moreover, $\mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ is the scheme-theoretic union*

$$\bigcup_{\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))} \mathcal{C}_\sigma.$$

Remark 1.2.3. This shows that the choice of cycles $\mathcal{Z}_\sigma = \mathcal{C}_\sigma$ solves the Breuil–Mézard equations for the above $\mathcal{X}^{\eta,\tau}$ (see [Emerton and Gee 2023, Conjecture 8.2.2; Le et al. 2023b, Conjecture 8.1.1]).

The equality of the underlying reduced $\mathcal{X}^{\eta,\tau}_{\mathbb{F},\mathrm{red}}$ and the scheme-theoretic union $\bigcup_{\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))} \mathcal{C}_\sigma$ is proved in Theorem 1.3.5 in [Le et al. 2023b] though we reprove it here with a weaker genericity condition (see Remark 1.2.4). The key point is to prove that the special fiber of $\mathcal{X}^{\eta,\tau}$ is in fact reduced. (If we replace η by $\lambda + \eta$ with λ dominant and nonzero or $n = 3$ by $n > 3$, the Breuil–Mézard conjecture predicts that the analogous stacks never have reduced special fiber.) The special fiber of $\mathcal{X}^{\eta,\tau}$ has an open cover with open sets labeled by f-tuples of $(2, 1, 0)$-admissible elements $(\widetilde{w}_j)$ in the extended affine Weyl group of GL_3. The complexity of the geometry of the open sets increases as the lengths of the $\widetilde{w}_j$ decrease. When the length of $\widetilde{w}_j$ is greater than 1 for all j, the reducedness immediately follows from the calculations in [Le et al. 2018a, §5.3]. Otherwise, the calculations of [Le et al. 2018a, §8] give an explicit upper bound on the special fiber which when combined with $\mathcal{X}^{\eta,\tau}_{\mathbb{F},\mathrm{red}} = \bigcup_{\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))} \mathcal{C}_\sigma$ must be an equality, and the reducedness follows.

Remark 1.2.4. An inexplicit genericity condition appears in the main theorems of [Le et al. 2023b] (see §1.2.1 of [loc. cit.]). While we use the models constructed in [loc. cit.], we reprove some of its main theorems in Sections 3.2 and 3.3 with the inexplicit condition replaced by the more typical genericity condition on the gaps between the exponents of the inertial characters in τ. This is possible because of the computations in [Le et al. 2018a; 2020].

Finally, we analyze $W^g(\bar{\rho})$ using local models. The special fibers of the local models embed inside the affine flag variety where irreducible components appear as subvarieties of translated affine Schubert varieties. In Section 4, we introduce a subset $W_{\mathrm{obv}}(\bar{\rho}) \subset W^{g}(\bar{\rho})$ of *obvious weights* for (possibly) wildly ramified $\bar{\rho}$, which has a simple interpretation in terms of the affine flag variety. Obvious weights generalize the notion of ordinary weights that appear in [Gee and Geraghty 2012] and the additional weights appearing in the exceptional cases of [Morra and Park 2017; Herzig et al. 2017; Le et al. 2018b]. The set $W_{\mathrm{obv}}(\bar{\rho})$ gives upper and lower bounds for $W^{g}(\bar{\rho})$. We finally show that, in almost all cases, one can determine $W^{g}(\bar{\rho})$ from $W_{\mathrm{obv}}(\bar{\rho})$ (Theorem 4.2.5). This last part uses a curious piece of numerology from the calculations of [Le et al. 2018a] — points in the special fibers of the local models never lie on exactly three components.

1.2.B. *Global methods: patching.* To prove Theorems 1.1.3 and 1.1.4 we combine the explicit description of the weight sets $W(\bar{r})$, coming from Theorems 1.1.1 and 4.2.5, with the Kisin–Taylor–Wiles methods developed in [Emerton et al. 2015] and employed in [Le et al. 2020, §5]. A crucial ingredient is the analysis of certain intersections of cycles in the special fiber of deformation rings. The local models introduced in [Le et al. 2023b] allow us to algebraize the computations made for the tame case in [Le et al. 2020, §3.6].

We now turn to Theorem 1.1.1. The key input into its proof beyond the Kisin–Taylor–Wiles method is the fact that the local Galois deformation rings of type (η, τ) are domains when τ is 4-generic. This is guaranteed by the fact that the stacks $\mathcal{X}^{\eta,\tau}$ are normal (Theorem 1.2.2). Then the supports of the patched modules of type τ are either empty or the entire potentially crystalline deformation rings of type (η, τ). The proof is then similar to the tame case in [Le et al. 2020] — one propagates modularity between obvious weights and then to shadow weights using carefully chosen types — except for one new wrinkle. From the axioms of a weak patching functor, one cannot deduce the modularity of an obvious weight to get started! Indeed one cannot rule out that $\bar{\rho}_v$ lies on a unique component $\mathcal{C}_{\sigma_v}$ and $W(\bar{r})$ contains exactly one Serre weight $\sigma' \neq \sigma \overset{\mathrm{def}}{=} \bigotimes_{v|p} \sigma_v$ with the property that for any tame inertial type τ, if $\mathrm{JH}(\bar{\sigma}(\tau))$ contains σ', then it also contains σ. We use a patched version of the weight cycling technique introduced in [Emerton et al. 2013] to rule out this pathology. In fact, we axiomatize our setup to make clear the ingredients that our method requires.

1.3. ***Overview.*** Section 2 covers background on tame inertial L-parameters and representation theory in Section 2.1, and Breuil–Kisin modules with tame descent data in Section 2.2, following [Le et al. 2023b]. Section 2.1.F gives a comparison between parametrizations of Serre weights in [Le et al. 2020; 2023b].

Section 3 establishes the main results about the geometry of local deformation rings. We specialize the theory of local models in [Le et al. 2023b] to dimension three. The main results are Theorem 3.3.2 and Corollary 3.3.3, which establish the geometric properties that we need, some of which are specific to dimension three.

In Section 4, we analyze possible sets of geometric weights using the affine flag variety. Theorem 4.2.5 gives a complete explicit description when $\bar{\rho}$ is sufficiently generic.

Section 5 contains our global applications. In Section 5.1, we introduce the axioms of patching functors following [Le et al. 2023b, §6] and prove the weight part of Serre's conjecture *assuming* the modularity of at least one obvious weight (Proposition 5.1.11). The latter condition is then removed in Section 5.2 using modules with an arithmetic action (Theorem 5.2.6). In Section 5.3, we prove results on mod p multiplicity one and Breuil's lattice conjectures for patched modules (Theorems 5.3.1, 5.3.13), generalizing analogous results in [Le et al. 2020] to the wildly ramified case. Finally, Section 5.4 proves our main global theorems.

1.4. *Notation.* For any field K we fix once and for all a separable closure $\overline{K}$ and let $G_K \stackrel{\text{def}}{=} \mathrm{Gal}(\overline{K}/K)$. If K is a nonarchimedean local field, we let $I_K \subset G_K$ denote the inertial subgroup. We fix a prime $p \in \mathbb{Z}_{>0}$. Let $E \subset \overline{\mathbb{Q}}_p$ be a subfield which is finite-dimensional over $\mathbb{Q}_p$. We write $\mathcal{O}$ to denote its ring of integers, fix an uniformizer $\varpi \in \mathcal{O}$ and let $\mathbb{F}$ denote the residue field of E. We will assume throughout that E is sufficiently large.

We consider the group $G \stackrel{\text{def}}{=} \mathrm{GL}_3$ (defined over $\mathbb{Z}$). We write B for the subgroup of upper triangular matrices, $T \subset B$ for the split torus of diagonal matrices and $Z \subset T$ for the center of G. Let $\Phi^+ \subset \Phi$ (resp. $\Phi^{\vee,+} \subset \Phi^\vee$) denote the subset of positive roots (resp. positive coroots) in the set of roots (resp. coroots) for (G, B, T). Let Δ (resp. $\Delta^\vee$) be the set of simple roots (resp. coroots). Let $X^*(T)$ be the group of characters of T which we identify with $\mathbb{Z}^3$ by letting the standard i-th basis element $\varepsilon_i = (0, \dots, 1, \dots, 0)$ (with the 1 in the i-th position) correspond to extracting the i-th diagonal entry of a diagonal matrix. In particular, we let ε'_1 and ε'_2 be $(1, 0, 0)$ and $(0, 0, -1)$ respectively.

We write W (resp. W_a, $\widetilde{W}$) for the Weyl group (resp. the affine Weyl group, the extended affine Weyl group) of G. If $\Lambda_R \subset X^*(T)$ denotes the root lattice for G we then have

$$W_a = \Lambda_R \rtimes W, \qquad \widetilde{W} = X^*(T) \rtimes W$$

and use the notation $t_\nu \in \widetilde{W}$ to denote the image of $\nu \in X^*(T)$. The Weyl groups W, $\widetilde{W}$, and W_a act naturally on $X^*(T)$ and on $X^*(T) \otimes_{\mathbb{Z}} A$ for any ring A by extension of scalars.

Let $\langle\ ,\ \rangle$ denote the duality pairing on $X^*(T) \times X_*(T)$, which extends to a pairing on

$$(X^*(T) \otimes_{\mathbb{Z}} A) \times (X_*(T) \otimes_{\mathbb{Z}} A)$$

for any ring A. We say that a weight $\lambda \in X^*(T)$ is *dominant* if $0 \le \langle \lambda, \alpha^\vee \rangle$ for all $\alpha \in \Delta$. Set $X^0(T)$ to be the subgroup consisting of characters $\lambda \in X^*(T)$ such that $\langle \lambda, \alpha^\vee \rangle = 0$ for all $\alpha \in \Delta$, and $X_1(T)$ to be the subset consisting of characters $\lambda \in X^*(T)$ such that $0 \le \langle \lambda, \alpha^\vee \rangle < p$ for all $\alpha \in \Delta$

We fix an element $\eta \in X^*(T)$ such that $\langle \eta, \alpha^\vee \rangle = 1$ for all positive simple roots α. We define the *p-dot action* as $t_\lambda w \cdot \mu = t_{p\lambda} w(\mu + \eta) - \eta$. By letting w_0 denote the longest element in W define $\widetilde{w}_h \stackrel{\text{def}}{=} w_0 t_{-\eta}$.

Recall that for $(\alpha, n) \in \Phi^+ \times \mathbb{Z}$, we have the p-root hyperplane $H_{\alpha,n} \stackrel{\text{def}}{=} \{\lambda : \langle \lambda + \eta, \alpha^\vee \rangle = np\}$. A p-alcove is a connected component of the complement $X^*(T) \otimes_{\mathbb{Z}} \mathbb{R} \setminus \big(\bigcup_{(\alpha,n)} H_{\alpha,n}\big)$. We say that a p-alcove C is p-restricted (resp. dominant) if $0 < \langle \lambda + \eta, \alpha^\vee \rangle < p$ (resp. $0 < \langle \lambda + \eta, \alpha^\vee \rangle$) for all simple roots $\alpha \in \Delta$ and $\lambda \in C$. If $C_0 \subset X^*(\underline{T}) \otimes_{\mathbb{Z}} \mathbb{R}$ denotes the dominant base alcove (i.e., the alcove defined by the condition $0 < \langle \lambda + \eta, \alpha^\vee \rangle < p$ for all positive roots $\alpha \in \Phi^+$, we let

$$\widetilde{W}^+ \stackrel{\text{def}}{=} \{\widetilde{w} \in \widetilde{W} : \widetilde{w} \cdot C_0 \text{ is dominant}\} \quad \text{and} \quad \widetilde{W}^+_1 \stackrel{\text{def}}{=} \{\widetilde{w} \in \widetilde{W}^+ : \widetilde{w} \cdot C_0 \text{ is } p\text{-restricted}\}.$$

We sometimes refer to C_0 as the *lower alcove* and $C_1 \stackrel{\text{def}}{=} \widetilde{w}_h \cdot C_0$ as the *upper alcove*.

Let now $\mathcal{O}_p$ be a finite étale $\mathbb{Z}_p$-algebra. We have an isomorphism $\mathcal{O}_p \cong \prod_{v \in S_p} \mathcal{O}_v$ where S_p is a finite set and $\mathcal{O}_v$ is the ring of integers of a finite unramified extension F_v^+ of $\mathbb{Q}_p$. Let $G_0 \stackrel{\text{def}}{=} \mathrm{Res}_{\mathcal{O}_p/\mathbb{Z}_p} G_{/\mathcal{O}_p}$ with Borel subgroup $B_0 \stackrel{\text{def}}{=} \mathrm{Res}_{\mathcal{O}_p/\mathbb{Z}_p} B_{/\mathcal{O}_p}$, maximal torus $T_0 \stackrel{\text{def}}{=} \mathrm{Res}_{\mathcal{O}_p/\mathbb{Z}_p} T_{/\mathcal{O}_p}$, and $Z_0 \stackrel{\text{def}}{=} \mathrm{Res}_{\mathcal{O}_p/\mathbb{Z}_p} Z_{/\mathcal{O}_p}$. We assume that $\mathcal{O}$ contains the image of any ring homomorphism $\mathcal{O}_p \to \overline{\mathbb{Z}}_p$ and write $\mathcal{J} \stackrel{\text{def}}{=} \mathrm{Hom}_{\mathbb{Z}_p}(\mathcal{O}_p, \mathcal{O})$. We can and do fix an identification of $\underline{G} \stackrel{\text{def}}{=} (G_0)_{/\mathcal{O}}$ with the split reductive group $G^{\mathcal{J}}_{/\mathcal{O}}$. We similarly define $\underline{B}$, $\underline{T}$, and $\underline{Z}$. Corresponding to $(\underline{G}, \underline{B}, \underline{T})$, we have the set of positive roots $\underline{\Phi}^+ \subset \underline{\Phi}$ and the set of positive coroots $\underline{\Phi}^{\vee,+} \subset \underline{\Phi}^\vee$. The notation $\underline{\Lambda}_R$, $\underline{W}$, $\underline{W}_a$, $\underline{\widetilde{W}}$, $\underline{\widetilde{W}}^+$, $\underline{\widetilde{W}}_1^+$ should be clear as should the natural isomorphisms $X^*(\underline{T}) = X^*(T)^{\mathcal{J}}$ and the like. Given an element $j \in \mathcal{J}$, we use a subscript notation to denote j-components obtained from the isomorphism $\underline{G} \cong G^{\mathcal{J}}_{/\mathcal{O}}$ (so that, for instance, given an element $\widetilde{w} \in \underline{\widetilde{W}}$ we write $\widetilde{w}_j$ to denote its j-th component via the induced identification $\underline{\widetilde{W}} \cong \widetilde{W}^{\mathcal{J}}$). For sake of readability, we abuse notation and still write w_0 to denote the longest element in $\underline{W}$, and fix a choice of an element $\eta \in X^*(\underline{T})$ such that $\langle \eta, \alpha^\vee \rangle = 1$ for all $\alpha \in \underline{\Delta}$. The meaning of w_0, η and $\widetilde{w}_h \stackrel{\text{def}}{=} w_0 t_{-\eta}$ should be clear from the context.

The absolute Frobenius automorphism on $\mathcal{O}_p/p$ lifts canonically to an automorphism φ of $\mathcal{O}_p$. We define an automorphism π of the identified groups $X^*(\underline{T})$ and $X_*(\underline{T}^\vee)$ by the formula $\pi(\lambda)_\sigma = \lambda_{\sigma \circ \varphi^{-1}}$ for all $\lambda \in X^*(\underline{T})$ and $\sigma : \mathcal{O}_p \to \mathcal{O}$. We assume that, in this case, the element $\eta \in X^*(\underline{T})$ we fixed is π-invariant. We similarly define an automorphism π of $\underline{W}$ and $\underline{\widetilde{W}}$.

Let F_p^+ be $\mathcal{O}_p[1/p]$ so that F_p^+ is isomorphic to the (finite) product $\prod_{v \in S_p} F_v^+$ where $F_v^+ \stackrel{\text{def}}{=} \mathcal{O}_v[1/p]$ for each $v \in S_p$. Let

$$\underline{G}^\vee_{/\mathbb{Z}} \stackrel{\text{def}}{=} \prod_{F_p^+ \to E} G^\vee_{/\mathbb{Z}}$$

be the dual group of $\underline{G}$ so that the Langlands dual group of G_0 is ${}^L\underline{G}_{/\mathbb{Z}} \stackrel{\text{def}}{=} \underline{G}^\vee \rtimes \mathrm{Gal}(E/\mathbb{Q}_p)$, where $\mathrm{Gal}(E/\mathbb{Q}_p)$ acts on the set of homomorphisms $F_p^+ \to E$ by postcomposition.

We now specialize to the case where $S_p = \{v\}$ is a singleton. Hence $F_p^+ = K$ is an unramified extension of degree f with ring of integers $\mathcal{O}_K$ and residue field k. Let $W(k)$ be ring of Witt vectors of k, which is also the ring of integers of K.

We denote the arithmetic Frobenius automorphism on $W(k)$ by φ; it acts as raising to p-th power on the residue field.

Recall that we fixed a separable closure $\overline{K}$ of K. We choose $\pi \in \overline{K}$ such that $\pi^{p^f-1} = -p$ and let $\omega_K : G_K \to \mathcal{O}_K^\times$ be the character defined by $g(\pi) = \omega_K(g)\pi$, which is independent of the choice of π. We fix an embedding $\sigma_0 : K \hookrightarrow E$ and define $\sigma_j = \sigma_0 \circ \varphi^{-j}$, which identifies $\mathcal{J} = \mathrm{Hom}(k, \mathbb{F}) = \mathrm{Hom}_{\mathbb{Q}_p}(K, E)$ with $\mathbb{Z}/f\mathbb{Z}$. We write $\omega_f : G_K \to \mathcal{O}^\times$ for the character $\sigma_0 \circ \omega_K$.

Let ε denote the p-adic cyclotomic character. If W is a de Rham representation of G_K over E, then for each $\kappa \in \mathrm{Hom}_{\mathbb{Q}_p}(K, E)$, we write $\mathrm{HT}_\kappa(W)$ for the multiset of Hodge–Tate weights labeled by embedding κ normalized so that the p-adic cyclotomic character ε has Hodge–Tate weight $\{1\}$ for every κ. For $\mu = (\mu_j)_{j \in \mathcal{J}} \in X^*(\underline{T})$, we say that a 3-dimensional representation W has Hodge–Tate weights μ if

$$\mathrm{HT}_{\sigma_j}(W) = \{\mu_{1,j}, \mu_{2,j}, \mu_{3,j}\}.$$

Our convention is the opposite of that of [Emerton and Gee 2023; Caraiani et al. 2016], but agrees with that of [Gee et al. 2018].

We say that a 3-dimensional potentially semistable representation $\rho : G_K \to \mathrm{GL}_n(E)$ has type (μ, τ) if ρ has Hodge–Tate weights μ and the restriction to I_K of the Weil–Deligne representation attached to ρ (via the *covariant* functor $\rho \mapsto \mathrm{WD}(\rho)$) is isomorphic to the inertial type τ. Note that this differs from the conventions of [Gee et al. 2018] via a shift by η.

Let Γ be a group. If V is a finite length Γ-representation, we let $\mathrm{JH}(V)$ be the (finite) set of Jordan–Hölder factors of V. If V° is a finite $\mathcal{O}$-module with a Γ-action, we write $\overline{V}^\circ$ for the Γ-representation $V^\circ \otimes_{\mathcal{O}} \mathbb{F}$ over $\mathbb{F}$.

If X is an ind-scheme defined over $\mathcal{O}$, we write $X_E \overset{\mathrm{def}}{=} X \times_{\mathrm{Spec}\,\mathcal{O}} \mathrm{Spec}\, E$ and $X_{\mathbb{F}} \overset{\mathrm{def}}{=} X \times_{\mathrm{Spec}\,\mathcal{O}} \mathrm{Spec}\, \mathbb{F}$ to denote its generic and special fiber, respectively. If M is any $\mathcal{O}$-module we write $\overline{M}$ to denote $M \otimes_{\mathcal{O}} \mathbb{F}$.

If P is a statement, the symbol $\delta_P \in \{0, 1\}$ takes value 1 if P is true, and 0 if P is false.

1.4.A. *Tables.* In Tables 1–3 we write α, β and γ for the elements of $\widetilde{W}$ corresponding to (12), (23) and $w_0 t_{(1,0,-1)}$, respectively. Moreover, the image of $\underline{1} \overset{\mathrm{def}}{=} (1, 1, 1) \in X^*(T)$ in $\widetilde{W}$ is denoted as $t_{\underline{1}}$. We identify the elements above with matrices in $\mathrm{GL}_3(\mathbb{Z}(\!(v)\!))$ via the embedding $\widetilde{W} \hookrightarrow \mathrm{GL}_3(\mathbb{Z}(\!(v)\!))$ defined by

$$\alpha \mapsto \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \beta \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \gamma \mapsto \begin{pmatrix} 0 & 0 & v^{-1} \\ 0 & 1 & 0 \\ v & 0 & 0 \end{pmatrix} \quad \text{and} \quad t_{\underline{1}} \mapsto \begin{pmatrix} v & 0 & 0 \\ 0 & v & 0 \\ 0 & 0 & v \end{pmatrix}.$$

2. Background

2.1. ***Affine Weyl group, tame inertial types and Deligne–Lusztig representations.*** Throughout this section, we assume that $S_p = \{v\}$. Thus $\mathcal{O}_p = \mathcal{O}_K$ is the ring of integers of a finite unramified extension K of $\mathbb{Q}_p$ and $G_0 = \mathrm{Res}_{\mathcal{O}_K/\mathbb{Z}_p} G_{/\mathcal{O}_K}$. We drop subscripts v from notation and we identify $\mathcal{J} = \mathrm{Hom}_{\mathbb{Q}_p}(K, E)$ with $\mathbb{Z}/f\mathbb{Z}$ via $\sigma_j \overset{\mathrm{def}}{=} \sigma_0 \circ \varphi^{-j} \mapsto j$.

2.1.A. *Admissibility.* We follow [Le et al. 2023b, §2.1–§2.4], specializing to the case of $n = 3$. We denote by $\leq$ the Bruhat order on $\underline{\widetilde{W}} \cong X^*(\underline{T}) \rtimes \underline{W}$ associated to the choice of the dominant base alcove $\underline{C}_0$ and set

$$\mathrm{Adm}(\eta) = \{\widetilde{w} \in \underline{\widetilde{W}} \mid \widetilde{w} \leq t_{s(\eta)} \text{ for some } s \in \underline{W}\}.$$

We will also consider the partially ordered group $\underline{\widetilde{W}}^\vee$ which is identified with $\underline{\widetilde{W}}$ as a group, but whose Bruhat order is defined by the antidominant base alcove (and still denoted as $\leq$). Then $\mathrm{Adm}^\vee(\eta)$ is defined as above, using now the antidominant order. We have an order reversing bijection $\widetilde{w} \mapsto \widetilde{w}^*$ between $\underline{\widetilde{W}}$ and $\underline{\widetilde{W}}^\vee$ defined as $(\widetilde{w}^*)_j \overset{\mathrm{def}}{=} w_j^{-1} t_{\nu_j}$ if $\widetilde{w}_j = t_{\nu_j} w_j$.

2.1.B. *Tame inertial types and Deligne–Lusztig representations.* An inertial type (for K) is the $\mathrm{GL}_3(E)$-conjugacy class of a homomorphism $\tau : I_K \to \mathrm{GL}_3(E)$ with open kernel and which extends to the Weil group of G_K. An inertial type is *tame* if it factors through the tame quotient of I_K. We will sometimes identify a tame inertial type with a fixed choice of a representative in its class.

Given $s = (s_0, \dots, s_{f-1}) \in \underline{W}$ and $\mu \in X^*(\underline{T}) \cap \underline{C}_0$, we have an associated integer $r \in \{1, 2, 3\}$ (which is the order of the element $s_0 s_1 \cdots s_{f-1} \in W$), integers $\boldsymbol{a}'^{(j')} \in \mathbb{Z}^3$ for $0 \le j' \le fr-1$ and a tame inertial type $\tau(s, \mu+\eta)$ defined as $\tau(s, \mu+\eta) \stackrel{\text{def}}{=} \sum_{i=1}^{3} (\omega_{fr})^{\boldsymbol{a}_i'^{(0)}}$ (see [Le et al. 2023b, Example 2.4.1, equations (5.2), (5.1)] for the details of this construction). We say that (s, μ) is *the lowest alcove presentation* for the tame inertial type $\tau(s, \mu+\eta)$ and that $\tau(s, \mu+\eta)$ is *N-generic* if μ is N-deep in alcove $\underline{C}_0$. We say that a tame inertial type τ has *a lowest alcove presentation* if there exists a pair (s, μ) as above such that $\tau \cong \tau(s, \mu+\eta)$ (in which case we will say that (s, μ) is a lowest alcove presentation for τ), and that τ is N-generic if τ has a lowest alcove presentation (s, μ) such that μ is N-deep in alcove $\underline{C}_0$. We remark that different choices of pairs (s, μ) as above can give rise to isomorphic tame inertial types (see [Le et al. 2019, Proposition 2.2.15]). If τ is a tame inertial type of the form $\tau = \tau(s, \mu+\eta)$, we write $\widetilde{w}(\tau)$ for the element $t_{\mu+\eta}s \in \underline{\widetilde{W}}$. (In particular, when writing $\widetilde{w}(\tau)$ we use an implicit lowest alcove presentation for τ).

Repeating the above with E replaced by $\mathbb{F}$, we obtain the notion of inertial $\mathbb{F}$-types and lowest alcove presentations for tame inertial $\mathbb{F}$-types. We use the notation $\bar{\tau}$ to denote a tame inertial $\mathbb{F}$-type $\bar{\tau} : I_K \to \mathrm{GL}_3(\mathbb{F})$. We say that a tame inertial $\mathbb{F}$-type is N-generic if it admits a lowest alcove presentation (s, μ) such that μ is N-deep in $\underline{C}_0$.

If μ is 1-deep in $\underline{C}_0$, then for each $0 \le j' \le fr-1$ there is a unique element $s'_{\mathrm{or}, j'} \in W$ such that $(s'_{\mathrm{or}, j'})^{-1}(\boldsymbol{a}'^{(j')})$ is dominant. (In the terminology of [Le et al. 2018a], see Definition 2.6 of [loc. cit.], the fr-tuple $(s'_{\mathrm{or}, j'})_{0 \le j' \le fr-1}$ is the *orientation* of $(\boldsymbol{a}'^{(j')})_{0 \le j' \le fr-1}$.)

To a pair $(s, \mu) \in \underline{W} \times X^*(\underline{T})$, we can also associate a virtual $G_0(\mathbb{F}_p)$-representation over E which we denote $R_s(\mu)$ (cf. [Gee et al. 2018, Definition 9.2.2], where $R_s(\mu)$ is denoted as $R(s, \mu)$). In particular, $R_1(\mu)$ is a principal series representation. If $\mu - \eta$ is 1-deep in $\underline{C}_0$ then $R_s(\mu)$ is an irreducible representation. In analogy with the terminology for tame inertial type, if $\mu - \eta$ is N-deep in alcove $\underline{C}_0$ for $N \ge 0$, we call $(s, \mu-\eta)$ an N-generic *lowest alcove presentation* for $R_s(\mu)$, and say that $R_s(\mu)$ is *N-generic*.

2.1.C. *Inertial local Langlands correspondence.* Given a tame inertial type $\tau : I_K \to \mathrm{GL}_3(E)$, [Caraiani et al. 2016, Theorem 3.7] gives an irreducible smooth E-valued representation $\sigma(\tau)$ of $G_0(\mathbb{F}_p) = \mathrm{GL}_3(k)$ over E satisfying results towards the inertial local Langlands correspondence (see [loc. cit.] for the properties satisfied by $\sigma(\tau)$). (By inflation, we will consider $\sigma(\tau)$ as a smooth representation of $G_0(\mathbb{Z}_p)$ without further comment.) This representation need not be uniquely determined by τ and in what follows $\sigma(\tau)$ will denote either a particular choice that we have made or any choice that satisfies the properties of [Caraiani et al. 2016, Theorem 3.7] (see also [Le et al. 2023b, Theorem 2.5.3] and the discussion following it).

When $\tau = \tau(s, \mu+\eta)$ is a tame inertial type such that $\mu \in \underline{C}_0$ is 1-deep, the representation $\sigma(\tau)$ can be taken to be $R_s(\mu+\eta)$ thanks to [Le et al. 2019, Corollary 2.3.5].

2.1.D. *Serre weights.* We finally recall the notion of Serre weights for $G_0(\mathbb{F}_p)$, and the notion of *lowest alcove presentations* for them, following [Le et al. 2023b, §2.2]. A *Serre weight* for $G_0(\mathbb{F}_p)$ is the isomorphism class of an (absolutely) irreducible representation of $G_0(\mathbb{F}_p)$ over $\mathbb{F}$. (We will sometimes refer to a representative for the isomorphism class as a Serre weight.)

Given $\lambda \in X_1(\underline{T})$, we write $F(\lambda)$ for the Serre weight with highest weight λ; the assignment $\lambda \mapsto F(\lambda)$ induces a bijection between $X_1(\underline{T})/(p-\pi)X^0(\underline{T})$ and the set of Serre weights (see [Gee et al. 2018, Lemma 9.2.4]). We say that $F(\lambda)$ is N-deep if λ is (this does not depend on the choice of λ).

Recall from [Le et al. 2023b, §2.2] the equivalence relation on $\underline{\widetilde{W}} \times X^*(\underline{T})$ defined by

$$(\widetilde{w}, \omega) \sim (t_\nu \widetilde{w}, \omega - \nu)$$

for all $\nu \in X^0(\underline{T})$. For (an equivalence class of) a pair $(\widetilde{w}_1, \omega - \eta) \in \underline{\widetilde{W}}_1^+ \times (X^*(\underline{T}) \cap \underline{C}_0)/\sim$ the Serre weight $F_{(\widetilde{w}_1,\omega)} \stackrel{\text{def}}{=} F(\pi^{-1}(\widetilde{w}_1)\cdot(\omega-\eta))$ is well defined, i.e., is independent of the representative of the equivalence class of $(\widetilde{w}_1, \omega)$. The equivalence class of $(\widetilde{w}_1, \omega)$ is called a *lowest alcove presentation* for the Serre weight $F_{(\widetilde{w}_1,\omega)}$. The Serre weight $F_{(\widetilde{w}_1,\omega)}$ is N-deep if and only if $\omega - \eta$ is N-deep in alcove $\underline{C}_0$. As above, we sometimes implicitly choose a representative for a lowest alcove presentation to make an a priori sense of an expression, though it is a posteriori independent of this choice.

2.1.E. *Compatibility for lowest alcove presentations.* Recall that we have a canonical isomorphism $\underline{\widetilde{W}}/\underline{W}_a \cong X^*(\underline{Z})$ where $\underline{W}_a \cong \Lambda_{\underline{R}} \rtimes \underline{W}$ is the affine Weyl group of $\underline{G}$. Given an algebraic character $\zeta \in X^*(\underline{Z})$, we say that an element $\widetilde{w} \in \underline{\widetilde{W}}$ is *ζ-compatible* if it corresponds to ζ via the isomorphism $\underline{\widetilde{W}}/\underline{W}_a \cong X^*(\underline{Z})$. In particular, a lowest alcove presentation (s, μ) for a tame inertial type (resp. a lowest alcove presentation $(s, \mu - \eta)$ for a Deligne–Lusztig representation) is ζ-compatible if the element $t_{\mu+\eta}s \in \underline{\widetilde{W}}$ (resp. $t_\mu s \in \underline{\widetilde{W}}$) is ζ-compatible. Similarly, a lowest alcove presentation $(\widetilde{w}_1, \omega)$ for Serre weight is *ζ-compatible* if the element $t_{\omega-\eta}\widetilde{w}_1 \in \underline{\widetilde{W}}$ is ζ-compatible.

2.1.F. *A comparison to* [Le et al. 2020]. In [Le et al. 2020], the parametrization of Serre weights is slightly different from the one in [Le et al. 2023b]. Here, we give a dictionary between the two.

Define a map

$$\underline{\widetilde{W}} \times X^*(\underline{T}) \to \underline{\widetilde{W}}/\underline{W}_a \cong X^*(\underline{Z}), \quad (\widetilde{w}, \omega) \mapsto t_{\omega-\eta}\widetilde{w}\underline{W}_a \tag{2-1}$$

and write $(\underline{\widetilde{W}} \times X^*(\underline{T}))^\zeta$ for the preimage of $\zeta \in X^*(\underline{Z})$ (presentations *compatible* with ζ). The map (2-1) is constant on equivalence classes, and we write $(\underline{\widetilde{W}} \times X^*(\underline{T}))^\zeta/\sim$ for the set of equivalence classes in the preimage of ζ. The equivalence relation restricts to one on $\underline{\widetilde{W}}_1 \times X^*(\underline{T})$ or $\underline{\widetilde{W}}_1 \times (X^*(\underline{T}) \cap \underline{C}_0 + \eta)$, and we use similar notation, e.g., $(\underline{\widetilde{W}}_1 \times (X^*(\underline{T}) \cap \underline{C}_0 + \eta))^\zeta/\sim$, for these subsets.

We let $\underline{\Lambda}_W$ and $\widetilde{W}^{\mathrm{der}}$ be $X^*(\underline{T})/X^0(\underline{T})$ and $\widetilde{W}/X^0(\underline{T})$, respectively. Recall from [Le et al. 2020, §2.1] the set

$$\mathcal{P}^{\mathrm{der}} = \{(\omega, \widetilde{w}) \in \underline{\Lambda}_W \times \underline{\widetilde{W}}_1^{\mathrm{der},+} \mid t_\omega \pi(\widetilde{w}) \in \underline{W}_a\}.$$

Letting $\mathcal{A}$ be the set of p-restricted alcoves in $X^*(\underline{T}) \otimes_{\mathbb{Z}} \mathbb{R}$, the map

$$\beta : \mathcal{P}^{\mathrm{der}} \to \underline{\Lambda}_W \times \mathcal{A}, \quad (\omega, \widetilde{w}) \mapsto (\omega, \pi(\widetilde{w}) \cdot \underline{C}_0)$$

is a bijection by [Le et al. 2020, Lemma 2.1.1].

For $\lambda \in X^*(\underline{T})$, the map

$$(\underline{\widetilde{W}}_1 \times X^*(\underline{T}))^{\lambda-\eta|_{\underline{Z}}}/\sim \xrightarrow{\iota_\lambda} \mathcal{P}^{\mathrm{der}}, \quad (\widetilde{w}, \omega) \mapsto (\omega - \lambda, \pi^{-1}(\widetilde{w}))$$

is a bijection. (Here $\omega - \lambda$ also denotes its image in $\underline{\Lambda}_W$.)

Then $\beta\circ\iota_\lambda : (\underline{\widetilde{W}}_1 \times X^*(\underline{T}))^{\lambda-\eta|_{\underline{Z}}}/\sim \to \underline{\Lambda}_W \times \mathcal{A}$ is a bijection which induces a bijection

$$(\underline{\widetilde{W}}_1 \times X^*(\underline{T}) \cap \underline{C}_0+\eta)^{\lambda-\eta|_{\underline{Z}}}/\sim \to \underline{\Lambda}_W^{\lambda} \times \mathcal{A}, \quad (\widetilde{w},\omega)\mapsto(\omega-\lambda,\widetilde{w}\cdot\underline{C}_0) \tag{2-2}$$

when $\lambda-\eta\in\underline{C}_0$ and $\underline{\Lambda}_W^{\lambda}$ is defined to be the set of $\omega'\in\underline{\Lambda}_W$ satisfying $\omega'+\lambda-\eta\in\underline{C}_0$, see [Le et al. 2020, §2.1]. By the definition of $\mathfrak{Tr}_\lambda$ in [Le et al. 2020, §2.1], for $(\widetilde{w},\omega)\in(\underline{\widetilde{W}}_1 \times X^*(\underline{T}) \cap \underline{C}_0+\eta)^{\lambda-\eta|_{\underline{Z}}}/\sim$, we have

$$F_{(\widetilde{w},\omega)} = F(\mathfrak{Tr}_\lambda(\omega-\lambda,\widetilde{w}\cdot\underline{C}_0)). \tag{2-3}$$

2.1.G. *Reduction of Deligne–Lusztig representations.* For $i\in\{1,2\}$, let ε_i denote the image of ε_i' via the surjection $X^*(\underline{T})\twoheadrightarrow\underline{\Lambda}_W$.

Proposition 2.1.1. *Let $\lambda-\eta$ and $\mu-\eta$ be 0-deep and 1-deep in $\underline{C}_0$, respectively, such that $\mu+\eta-\lambda\in\underline{\Lambda}_R$. If $\sigma\in\mathrm{JH}(\overline{R_s(\mu)})$ is a 0-deep Serre weight, then σ is contained in $F\big(\mathfrak{Tr}_\lambda(t_{\mu-\lambda}s(\Sigma))\big)$, where $\Sigma=(\Sigma_0)^f\subseteq \underline{\Lambda}_W^{\lambda+\eta}\times\mathcal{A}$ and*

$$\Sigma_0 \stackrel{\text{def}}{=} \big\{(\varepsilon_1+\varepsilon_2,0),(\varepsilon_1-\varepsilon_2,0),(\varepsilon_2-\varepsilon_1,0),(0,1),(\varepsilon_1,1),(\varepsilon_2,1),(0,0),(\varepsilon_1,0),(\varepsilon_2,0)\big\}.$$

If $\mu-\eta$ is furthermore 2-deep, then $\mathrm{JH}(\overline{R_s(\mu)})$ is $F\big(\mathfrak{Tr}_\lambda(t_{\mu-\lambda}s(\Sigma))\big)$.

Proof. [Herzig 2009, Appendix, Theorem 3.4] gives the identity

$$\overline{R_s(\mu)} = \sum_{\widetilde{w}\in\underline{\widetilde{W}}_1^+/X^0(\underline{T})} \overline{W}\big(\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta)\big)$$

at the level of characters (in our situation, $\gamma'^{\,\mathrm{Fr}-1}_{w_1,w_2}$ is 1 if $w_1=w_2$ and is 0 otherwise). That $\mu-\eta$ is 1-deep in $\underline{C}_0$ implies that $\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta)$ is -1-deep in a p-restricted alcove. This implies that for each $\widetilde{w}\in\underline{\widetilde{W}}_1^+$, $\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta)+\eta$ is dominant so that $\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta)$ is dominant or $W(\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta))$ is zero. If $\sigma\in\mathrm{JH}(\overline{R_s(\mu)})$, then $\sigma\in\mathrm{JH}\big(\overline{W}(\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta))\big)$ for some $\widetilde{w}\in\underline{\widetilde{W}}_1^+$. Herzig [2006, Proposition 4.9] gives a decomposition of such p-restricted Weyl modules. The proof of [Le et al. 2020, Proposition 2.3.4] shows that $\sigma\in F\big(\mathfrak{Tr}_\lambda(t_{\mu-\lambda}s(\Sigma))\big)$. If $\mu-\eta$ is 2-deep, then $\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta)$ is 0-deep in a p-restricted alcove and hence dominant. The description of $\mathrm{JH}(\overline{R_s(\mu)})$ again follows from the proof of [Le et al. 2020, Proposition 2.3.4]. $\square$

If $\mu-\eta$ is 1-deep in $\underline{C}_0$, we let the subset $\mathrm{JH}_{\mathrm{out}}(\overline{R_s(\mu)})\subset\mathrm{JH}(\overline{R_s(\mu)})$ be the Serre weights of the form $F\big(\widetilde{w}\cdot(t_\mu s(\widetilde{w}_h\widetilde{w})^{-1}(0)-\eta)\big)$ for some $\widetilde{w}\in\underline{\widetilde{W}}_1^+$. We call the elements of $\mathrm{JH}_{\mathrm{out}}(\overline{R_s(\mu)})$ the *outer weights* (of $\mathrm{JH}(\overline{R_s(\mu)})$). In the notation of Proposition 2.1.1, $\mathrm{JH}_{\mathrm{out}}(\overline{R_s(\mu)})$ is the subset $F\big(\mathfrak{Tr}_\lambda(t_{\mu-\lambda}s(\Sigma_{\mathrm{out}}))\big)$, where $\Sigma_{\mathrm{out}}=(\Sigma_{\mathrm{out},0})^f\subseteq\underline{\Lambda}_W^{\lambda+\eta}\times\mathcal{A}$ and

$$\Sigma_{\mathrm{out},0} \stackrel{\text{def}}{=} \big\{(\varepsilon_1+\varepsilon_2,0),(\varepsilon_1-\varepsilon_2,0),(\varepsilon_2-\varepsilon_1,0)(0,1),(\varepsilon_1,1),(\varepsilon_2,1)\big\}.$$

If $\lambda-\eta$ and $\mu-\eta$ are 0-deep and 2-deep in $\underline{C}_0$ respectively, and $\mu+\eta-\lambda\in\underline{\Lambda}_R$, we define $W^?(\bar{\tau}(s,\mu+\eta))$ to be the set of Serre weights $F\big(\mathfrak{Tr}_\lambda\big(t_{\mu+\eta-\lambda}s(r(\Sigma))\big)\big)$, where $r(\Sigma)$ is defined by swapping the digits of $\underline{a}\in\mathcal{A}$ in the elements $(\varepsilon,\underline{a})\in\Sigma$.

2.1.H. *The covering order.*

Definition 2.1.2. We say that a 3-deep Serre weight σ_0 *covers* σ if

$$\sigma \in \bigcap_{\substack{R \text{ 1-generic} \\ \sigma_0 \in \mathrm{JH}(\bar{R})}} \mathrm{JH}(\bar{R})$$

(where R runs over 1-generic Deligne–Lusztig representations).

Lemma 2.1.3. *A Serre weight σ_1 covers σ_2 if and only if $\sigma_2 \uparrow \sigma_1$.*

Proof. This follows from [Le et al. 2023b, Proposition 2.3.12(4)] where the slightly weaker genericity hypotheses come from Section 2.1.G. □

2.1.I. *L-parameters.* We now assume that S_p has arbitrary finite cardinality. An L-parameter (over E) is a $\underline{G}^\vee(E)$-conjugacy class of a continuous homomorphism $\rho : G_{\mathbb{Q}_p} \to {}^L\underline{G}(E)$ which is compatible with the projection to $\mathrm{Gal}(E/\mathbb{Q}_p)$ (such homomorphism is called *L-homomorphism*). An inertial L-parameter is a $\underline{G}^\vee(E)$-conjugacy class of a homomorphism $\tau : I_{\mathbb{Q}_p} \to \underline{G}^\vee(E)$ with open kernel, and which admits an extension to an L-homomorphism. An inertial L-parameter is *tame* if some (equivalently, any) representative in its equivalence class factors through the tame quotient of $I_{\mathbb{Q}_p}$.

Fixing isomorphisms $\overline{F_v^+} \xrightarrow{\sim} \overline{\mathbb{Q}}_p$ for all $v \in S_p$, we have a bijection between L-parameters (resp. tame inertial L-parameters) and collections of the form $(\rho_v)_{v\in S_p}$ (resp. of the form $(\tau_v)_{v\in S_p}$) where for all $v \in S_p$ the element $\rho_v : G_{F_v^+} \to \mathrm{GL}_3(E)$ is a continuous Galois representation (resp. the element $\tau_v : I_{F_v^+} \to \mathrm{GL}_3(E)$ is a tame inertial type for F_v^+).

We have similar notions when E is replaced by $\mathbb{F}$. Again we will often abuse terminology, and identify an L-parameter (resp. a tame inertial L-parameter) with a fixed choice of a representative in its class. This shall cause no confusion, and nothing in what follows will depend on this choice.

The definitions and results of Sections 2.1.B–2.1.H generalize in the evident way for tame inertial L-parameters and L-homomorphism. (In the case of the inertial local Langlands correspondence of Section 2.1.C, given a tame inertial L-parameter τ corresponding to the collection of tame inertial types $(\tau_v)_{v\in S_p}$, we let $\sigma(\tau)$ be the irreducible smooth E-valued representation of $G_0(\mathbb{Z}_p)$ given by $\bigotimes_{v\in S_p} \sigma(\tau_v)$.)

2.2. *Breuil–Kisin modules.* We recall some background on Breuil–Kisin modules with tame descent data. We refer the reader to [Le et al. 2020, §3.1–3.2; 2023b, §5.1] for further detail, with the caveat that we are following the conventions of the latter on the labeling of embeddings for tame inertial types and Breuil–Kisin modules (see [loc. cit., Remark 5.1.2]).

Let $\tau = \tau(s, \mu+\eta)$ be a tame inertial type with lowest alcove presentation (s, μ) which we fix throughout this section (recall that μ is 1-deep in $\underline{C}_0$). Recall that $r \in \{1, 2, 3\}$ is the order of $s_0 s_1 s_2 \cdots s_{f-1} \in W$. We write K'/K for the unramified extension of degree r contained in $\overline{K}$ and set $f' \stackrel{\text{def}}{=} fr$, $e' \stackrel{\text{def}}{=} p^{f'} - 1$. We identify $\mathrm{Hom}_{\mathbb{Q}_p}(K', E)$ with $\mathbb{Z}/f'\mathbb{Z}$ via $\sigma_{j'} \stackrel{\text{def}}{=} \sigma_0' \circ \varphi^{-j'} \mapsto j'$, where $\sigma_0' : K' \hookrightarrow E$ is a fixed choice for an embedding extending $\sigma_0 : K \hookrightarrow E$. In this way, restriction of embeddings corresponds to reduction modulo f in the above identifications.

Let $\pi' \in \overline{K}$ be an e'-th root of $-p$, let $L' \overset{\text{def}}{=} K'(\pi')$ and $\Delta' \overset{\text{def}}{=} \mathrm{Gal}(L'/K') \subset \Delta \overset{\text{def}}{=} \mathrm{Gal}(L'/K)$. We define the $\mathcal{O}_{K'}^{\times}$-valued character $\omega_{K'}(g) \overset{\text{def}}{=} g(\pi')/\pi'$ for $g \in \Delta'$ (this does not depend on the choice of π'). Given an $\mathcal{O}$-algebra R, we set $\mathfrak{S}_{L',R} \overset{\text{def}}{=} (W(k') \otimes_{\mathbb{Z}_p} R)[\![u']\!]$. The latter is endowed with an endomorphism $\varphi : \mathfrak{S}_{L',R} \to \mathfrak{S}_{L',R}$ acting as Frobenius on $W(k')$, trivially on R, and sending u' to $(u')^p$. It is endowed moreover with an action of Δ as follows: for any g in Δ', $g(u') = (g(\pi')/\pi')u' = \omega_{K'}(g)u'$ and g acts trivially on the coefficients; if $\sigma^f \in \mathrm{Gal}(L'/K)$ is the lift of p^f-Frobenius on $W(k')$ which fixes π', then σ^f is a generator for $\mathrm{Gal}(K'/K)$, acting in natural way on $W(k')$ and trivially on both u' and R. Set $v = (u')^{e'}$,

$$\mathfrak{S}_R \overset{\text{def}}{=} (\mathfrak{S}_{L',R})^{\Delta=1} = (W(k) \otimes_{\mathbb{Z}_p} R)[\![v]\!]$$

and $E(v) \overset{\text{def}}{=} v + p = (u')^{e'} + p$.

Let $Y^{[0,2]}(R)$ be the groupoid of Breuil–Kisin modules of rank 3 over $\mathfrak{S}_{L',R}$, height in [0, 2] and descent data of type τ (see [Caraiani and Levin 2018, §3; Le et al. 2020, Definition 3.1.3; 2023b, Definition 5.1.3]):

Definition 2.2.1. An object of $Y^{[0,2],\tau}(R)$ is the datum of

(1) a finitely generated projective $\mathfrak{S}_{L',R}$-module $\mathfrak{M}$ which is locally free of rank 3;

(2) an injective $\mathfrak{S}_{L',R}$-linear map $\phi_{\mathfrak{M}} : \varphi^*(\mathfrak{M}) \to \mathfrak{M}$ whose cokernel is annihilated by $E(v)^2$; and

(3) a semilinear action of Δ on $\mathfrak{M}$ which commutes with $\phi_{\mathfrak{M}}$, and such that, for each $j' \in \mathrm{Hom}_{\mathbb{Q}_p}(K', E)$,

$$(\mathfrak{M} \otimes_{W(k'),\sigma_{j'}} R) \mod u' \cong \tau^{\vee} \otimes_{\mathcal{O}} R$$

as Δ'-representations.

Note that $\mathfrak{M}^{(j')} \overset{\text{def}}{=} \mathfrak{M} \otimes_{W(k'),\sigma_{j'}} R$ is an $R[\![u']\!]$-submodule of $\mathfrak{M}$ in a standard way, endowed with a semilinear action of Δ' and the Frobenius $\phi_{\mathfrak{M}}$ induces Δ'-equivariant morphisms $\phi_{\mathfrak{M}}^{(j')} : \varphi^*(\mathfrak{M}^{(j'-1)}) \to \mathfrak{M}^{(j')}$. In particular, by letting τ' denote the tame inertial type for K' obtained from τ via the identification $I_{K'} = I_K$ induced by the inclusion $K' \subseteq \overline{K}$, the semilinear action of Δ induces an isomorphism $\iota_{\mathfrak{M}} : (\sigma^f)^*(\mathfrak{M}) \cong \mathfrak{M}$ (see [Le et al. 2018a, §6.1]) as elements of $Y^{[0,2],\tau'}(R)$.

Let $\mathfrak{M} \in Y^{[0,2],\tau}(R)$. Recall that an *eigenbasis* of $\mathfrak{M}$ is a collection of bases $\beta^{(j')} = (f_1^{(j')}, f_2^{(j')}, f_3^{(j')})$ for each $\mathfrak{M}^{(j')}$ such that Δ' acts on $f_i^{(j')}$ via the character $\omega_{f'}^{-a_i'^{(0)}}$ (see Section 2.1.B for the definition of $\boldsymbol{a}'^{(0)} \in \mathbb{Z}^3$) and such that $\iota_{\mathfrak{M}}((\sigma^f)^*(\beta^{(j')})) = \beta^{(j'+f)}$ for all $j' \in \mathrm{Hom}_{\mathbb{Q}_p}(K', E)$. Given an eigenbasis β for $\mathfrak{M}$, we let $C_{\mathfrak{M},\beta}^{(j')}$ be the matrix of $\phi_{\mathfrak{M}}^{(j')} : \varphi^*(\mathfrak{M}^{(j'-1)}) \to \mathfrak{M}^{(j')}$ with respect to the bases $\varphi^*(\beta^{(j'-1)})$ and $\beta^{(j')}$ and set

$$A_{\mathfrak{M},\beta}^{(j')} \overset{\text{def}}{=} \mathrm{Ad}((s'_{\mathrm{or},j'})^{-1}(u')^{-\boldsymbol{a}'^{(j')}})(C_{\mathfrak{M},\beta}^{(j')})$$

for $j' \in \mathrm{Hom}_{\mathbb{Q}_p}(K', E)$. It is an element of $\mathrm{GL}_3(R(\!(v+p)\!))$ with coefficients in $R[\![v+p]\!]$, is upper triangular modulo v and only depends on the restriction of j' to K (see [Le et al. 2023b, §5.1]).

Let $\mathcal{I}(\mathbb{F})$ denote the Iwahori subgroup of $\mathrm{GL}_3(\mathbb{F}(\!(v)\!))$ relative to the Borel of upper triangular matrices. We define the *shape* of a mod p Breuil–Kisin module $\mathfrak{M} \in Y^{[0,2],\tau}(\mathbb{F})$ to be the element $\tilde{z} = (\tilde{z}_j) \in \underline{\widetilde{W}}^{\vee}$

such that for any eigenbasis β and any $j \in \mathcal{J}$, the matrix $A^{(j)}_{\mathfrak{M},\beta}$ lies in $\mathcal{I}(\mathbb{F})\tilde{z}_j\,\mathcal{I}(\mathbb{F})$. This notion doesn't depend on the choice of eigenbasis, see [Le et al. 2023b, Proposition 5.1.8] (but it *does* depend on the lowest alcove presentation of τ; see [ibid., Remark 5.1.5]).

3. Local models in mixed characteristic and the Emerton–Gee stack

We assume throughout this section that $S_p = \{v\}$ so $\mathcal{O}_p = \mathcal{O}_K$ is the ring of integer of a finite unramified extension K of $\mathbb{Q}_p$. We identify $\mathcal{J} = \mathrm{Hom}_{\mathbb{Q}_p}(K, E)$ with $\mathbb{Z}/f\mathbb{Z}$ via $\sigma_j \overset{\text{def}}{=} \sigma_0 \circ \varphi^{-j} \mapsto j$.

3.1. ***Local models in mixed characteristic.*** We now define the mixed characteristic local models which are relevant to our paper. We follow closely [Le et al. 2023b, §4] and the notation therein.

For any Noetherian $\mathcal{O}$-algebra R, define

$$L\mathcal{G}_{\mathcal{O}}(R) \overset{\text{def}}{=} \{A \in \mathrm{GL}_3(R(\!(v+p)\!)),\ A \text{ is upper triangular modulo } v\};$$
$$L^+\mathcal{G}_{\mathcal{O}}(R) \overset{\text{def}}{=} \{A \in \mathrm{GL}_3(R[\![v+p]\!]),\ A \text{ is upper triangular modulo } v\};$$
$$L^{[0,2]}\mathcal{G}_{\mathcal{O}}(R) \overset{\text{def}}{=} \left\{ A \in L\mathcal{G}_{\mathcal{O}}(R),\ \begin{matrix} A,\ (v+p)^2A^{-1} \text{ are elements of } \mathrm{Mat}_3(R[\![v+p]\!]) \\ \text{and are upper triangular modulo } v \end{matrix} \right\}.$$

The fpqc quotients $L^+\mathcal{G}_{\mathcal{O}} \backslash L^{[0,2]}\mathcal{G}_{\mathcal{O}} \hookrightarrow L^+\mathcal{G}_{\mathcal{O}} \backslash L\mathcal{G}_{\mathcal{O}}$ induced from inclusions $L^+\mathcal{G}_{\mathcal{O}}(R) \subseteq L^{[0,2]}\mathcal{G}_{\mathcal{O}}(R) \subseteq L\mathcal{G}_{\mathcal{O}}(R)$ are representable by a projective scheme $\mathrm{Gr}^{[0,2]}_{\mathcal{G},\mathcal{O}}$ and an ind-projective ind-scheme $\mathrm{Gr}_{\mathcal{G},\mathcal{O}}$, respectively.

For any $\tilde{z} = (\tilde{z}_j)_{j\in\mathcal{J}} \in \underline{\widetilde{W}}^\vee$ and any Noetherian $\mathcal{O}$-algebra R, define

$$\widetilde{U}(\tilde{z})(R) = \big(\widetilde{U}(\tilde{z}_j)(R)\big)_{j\in\mathcal{J}} \subseteq (L\mathcal{G}_{\mathcal{O}}(R))^{\mathcal{J}}$$

to be the set of f-tuples of matrices $(A^{(j)})_{j\in\mathcal{J}} \in (L\mathcal{G}_{\mathcal{O}}(R))^{\mathcal{J}}$ such that for all $1 \le i, k \le 3$ and $j \in \mathcal{J}$,

- $A^{(j)}_{ik} \in v^{\delta_{i>k}} R[v+p, 1/(v+p)]$;
- $\deg_{v+p}(A^{(j)}_{ik}) \le \nu_{j,k} - \delta_{i<z_j(k)}$; and
- $\deg_{v+p}(A^{(j)}_{z_j(k)k}) = \nu_{j,k}$ and the coefficient of the leading term is a unit of R,

where we have written $\tilde{z} = zt_\nu$ and $\nu = (\nu_{j,1}, \nu_{j,2}, \nu_{j,3})_{j\in\mathcal{J}}$ (and recall from Section 1.4 the notation for the Kronecker deltas $\delta_{i>k}, \delta_{i<z_j(k)}$). We set $\widetilde{U}^{[0,2]}(\tilde{z})(R) \overset{\text{def}}{=} \widetilde{U}(\tilde{z})(R) \cap (L^{[0,2]}\mathcal{G}(R))^{\mathcal{J}}$. Note that both $\widetilde{U}^{[0,2]}(\tilde{z})$ and $\widetilde{U}(\tilde{z})$ are endowed with a $\underline{T}^\vee_{\mathcal{O}}$-action induced by left multiplication of matrices. It follows from [Le et al. 2023b, Lemmas 3.2.2 and 3.2.7] that the natural map $\widetilde{U}(\tilde{z}) \to \mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},\mathcal{O}}$ (resp. $\widetilde{U}^{[0,2]}(\tilde{z}) \to \mathrm{Gr}^{[0,2],\,\mathcal{J}}_{\mathcal{G},\mathcal{O}}$) factors as a $\underline{T}^\vee_{\mathcal{O}}$-torsor map followed by an open immersion. (We have written $\mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},\mathcal{O}}$ for the product, over $\mathrm{Spec}\,\mathcal{O}$, of f-copies of $\mathrm{Gr}_{\mathcal{G},\mathcal{O}}$ indexed over elements $j \in \mathcal{J}$, and $\mathrm{Gr}^{[0,2],\mathcal{J}}_{\mathcal{G},\mathcal{O}}$ is defined similarly.)

We now compare the objects above with groupoids of Breuil–Kisin modules with tame descent. Let $(s, \mu) \in \underline{W} \times X^*(\underline{T})$ be a lowest alcove presentation for the tame inertial type $\tau \overset{\text{def}}{=} \tau(s, \mu+\eta)$. We have the *twisted shifted* conjugation action of $\underline{T}^\vee_{\mathcal{O}}$ on $\widetilde{U}^{[0,2]}(\tilde{z})$ given by

$$A^{(j)} \mapsto t_j A^{(j)} \mathrm{Ad}(s_j^{-1})(t_{j-1}),$$

which is exactly the restriction to $\underline{T}^\vee_{\mathcal{O}}$ of the (s,μ)-twisted φ-conjugation in [Le et al. 2023b, §5.2]. By [Le et al. 2023b, Corollary 5.2.3] the quotient of $\widetilde{U}^{[0,2]}(\tilde{z})_{\mathbb{F}}$ by this action is isomorphic to an open substack $Y^{[0,2],\tau}_{\mathbb{F}}(\tilde{z})$ of $Y^{[0,2],\tau}_{\mathbb{F}}$. We denote by $Y^{[0,2],\tau}(\tilde{z})$ the open substack of $Y^{[0,2],\tau}$ induced by $Y^{[0,2],\tau}_{\mathbb{F}}(\tilde{z})$ (see [Le et al. 2023b, Definition 5.2.4]).

By [Le et al. 2023b, Theorem 5.3.1], whenever μ is 3-deep in $\underline{C}_0$, we have a morphism of p-adic formal algebraic stacks over $\mathcal{O}$,

$$\widetilde{U}^{[0,2]}(\tilde{z})^{\wedge_p} \to Y^{[0,2],\tau}(\tilde{z}) \hookrightarrow Y^{[0,2],\tau},$$

where the left map is a $\underline{T}^\vee_{\mathcal{O}}$-torsor for the twisted shifted conjugation action on the source and the second map is an open immersion.

We finally consider Breuil–Kisin modules with height bounded by the cocharacter $\eta \in X_*(\underline{T}^\vee)$. The fiber $\mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},E}$ of $\mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},\mathcal{O}}$ over E is the affine Grassmannian of GL_3 and we let $M_{\mathcal{J}}(\leq\eta)$ be the Zariski closure in $\mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},\mathcal{O}}$ of the open affine Schubert cell associated to $(v+p)^\eta$ in $\mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},E}$. Let $\widetilde{U}(\tilde{z},\leq\eta)$ be the pull back of $\widetilde{U}^{[0,2]}(\tilde{z})$ along the closed immersion $M_{\mathcal{J}}(\leq\eta) \hookrightarrow \mathrm{Gr}^{\mathcal{J}}_{\mathcal{G},\mathcal{O}}$.

Let $Y^{\leq\eta,\tau}$ denote the closed p-adic formal substack of $Y^{[0,2],\tau}$ appearing in [Caraiani and Levin 2018, §5] (and denoted $Y^{\eta,\tau}$ there). The computations of [Le et al. 2018a, §4] give the following:

Proposition 3.1.1. *Let $\tilde{z} \in \mathrm{Adm}^\vee(\eta)$. Then $\widetilde{U}(\tilde{z},\leq\eta)$ is an affine scheme over $\mathcal{O}$, with presentations $\bigotimes_{j=0}^{f-1}\mathcal{O}(\widetilde{U}(\tilde{z}_j,\leq\eta_j))$, where the $\mathcal{O}$-algebras $\mathcal{O}(\widetilde{U}(\tilde{z}_j,\leq\eta_j))$ are given in Table 1. Moreover, let $(s,\mu) \in \underline{W} \times (X^*(\underline{T}) \cap \underline{C}_0)$ be a lowest alcove presentation of $\tau \stackrel{\text{def}}{=} \tau(s,\mu+\eta)$, with μ being 3-deep in $\underline{C}_0$ and denote by $Y^{\leq\eta,\tau}(\tilde{z})$ the pull back of $Y^{[0,2],\tau}(\tilde{z})$ along the closed immersion $Y^{\leq\eta,\tau} \hookrightarrow Y^{[0,2],\tau}$.*

Then we have a morphism of p-adic formal algebraic stacks over $\mathcal{O}$

$$\widetilde{U}(\tilde{z},\leq\eta)^{\wedge_p} \to Y^{\leq\eta,\tau}(\tilde{z}) \hookrightarrow Y^{\leq\eta,\tau}. \tag{3-1}$$

where the left map is a $\underline{T}^\vee_{\mathcal{O}}$-torsor for the twisted shifted conjugation action on the source and the second map is an open immersion.

3.2. *Monodromy condition.* We introduce closed subspaces of $Y^{\leq\eta,\tau}$, $Y^{\leq\eta,\tau}(\tilde{z})$ and $\widetilde{U}(\tilde{z},\leq\eta)^{\wedge_p}$, and compare them with potentially crystalline substacks of the Emerton–Gee stack (introduced below).

Recall the element $\eta \in X^*(\underline{T})$ we fixed in Section 1.4. Let $\tau \stackrel{\text{def}}{=} \tau(s,\mu+\eta)$ be a tame inertial type with lowest alcove presentation (s,μ), where μ is 1-deep in alcove $\underline{C}_0$. By [Le et al. 2023b, Proposition 7.1.6] the datum of a p-adically complete, topologically finite type flat $\mathcal{O}$-algebra R, and a morphism $f : \mathrm{Spf}\, R \to Y^{[0,2],\tau}$ (corresponding to an element $\mathfrak{M} \in Y^{[0,2],\tau}(R)$) defines a p-saturated ideal $I_{\mathfrak{M},\nabla_\infty} \subset R$ which is compatible with flat base change. (In the terminology of [loc. cit.], and in the case where $\mathfrak{M}$ is free over $\mathfrak{S}_{L',R}$, the morphism $f : \mathrm{Spf}\, R \to Y^{[0,2],\tau}$ factors through $\mathrm{Spf}\, R \to \mathrm{Spf}(R/I_{\mathfrak{M},\nabla_\infty})$ if and only if $\mathfrak{M}$ satisfies the *monodromy condition* [Le et al. 2023b, Definition 7.1.2].) This gives rise to a $\mathcal{O}$-flat closed substack $Y^{[0,2],\tau,\nabla_\infty} \hookrightarrow Y^{[0,2],\tau}$ (see [Le et al. 2023b, §7.2]) characterized by the property that for any $\mathfrak{M} \in Y^{[0,2],\tau}(R)$ corresponding to $\mathrm{Spf}\, R \to Y^{[0,2],\tau}$, the pullback of the substack $Y^{[0,2],\tau,\nabla_\infty}$

$\tilde{z}_j t_{-\underline{1}}$	$A^{(j)}$	$\mathcal{O}(\widetilde{U}(\tilde{z}_j, \le\eta_j))$
$\alpha\beta\alpha\gamma$	$\begin{pmatrix} (v+p)^2c^*_{11} & 0 & 0 \\ v(v+p)c_{21} & (v+p)c^*_{22} & 0 \\ v(c_{31}+(v+p)c'_{31}) & vc_{32} & c^*_{33} \end{pmatrix}$	$\mathcal{O}[c^*_{11}, c_{21}, c_{31}, c'_{31}, c^*_{22}, c_{32}, c^*_{33}]$
$\beta\gamma\alpha\gamma$	$\begin{pmatrix} (v+p)c^*_{11} & (v+p)c_{12} & 0 \\ 0 & (v+p)^2c^*_{22} & 0 \\ vc_{31} & v(c_{32}+(v+p)c'_{32}) & c^*_{33} \end{pmatrix}$	$\mathcal{O}[c^*_{11}, c_{12}, c^*_{22}, c_{31}, c_{32}, c'_{32}, c^*_{33}]$
$\beta\alpha\gamma$	$\begin{pmatrix} (v+p)c_{11} & (v+p)c^*_{12} & 0 \\ v(v+p)c^*_{21} & (v+p)c_{22} & 0 \\ v(c_{31}+(v+p)c'_{31}) & vc_{32} & c^*_{33} \end{pmatrix}$	$\mathcal{O}[c_{11}, c^*_{12}, c^*_{21}, c_{22}, c_{31}, c'_{31}, c_{32}, c^*_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = (c_{11}c_{22} + pc^*_{12}c^*_{21})$
$\alpha\beta\gamma$	$\begin{pmatrix} (v+p)^2c^*_{11} & 0 & 0 \\ v(c_{21}+(v+p)c'_{21}) & c_{22} & c^*_{23} \\ v(c_{21}c_{33}(c^*_{23})^{-1}+(v+p)c'_{31}) & vc^*_{32} & c_{33} \end{pmatrix}$	$\mathcal{O}[c^*_{11}, c_{21}, c'_{21}, c_{22}, c^*_{23}, c'_{31}, c^*_{32}, c_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = (c_{22}c_{33} + pc^*_{32}c^*_{23})$
$\alpha\beta\alpha$	$\begin{pmatrix} c_{11} & c_{11}c_{32}(c^*_{31})^{-1} & c_{13}+(v+p)c^*_{13} \\ 0 & (v+p)c^*_{22} & (v+p)c_{23} \\ vc^*_{31} & vc_{32} & c_{33}+(v+p)c'_{33} \end{pmatrix}$	$\mathcal{O}[c_{11}, c_{12}, c_{13}, c^*_{13}, c^*_{22}, c_{23}, c^*_{31}, c_{32}, c_{33}, c'_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = \begin{pmatrix} c_{11}c_{32} - c_{12}c^*_{31},\ c_{11}c_{33} + pc_{13}c^*_{31} \\ c_{11}c'_{33} - c_{13}c^*_{31} + pc^*_{13}c^*_{31} \end{pmatrix}$
Table 1 (continued)		

Table 1. We list the $\mathcal{O}$-algebras $\mathcal{O}(\widetilde{U}(\tilde{z}_j, \le\eta_j))$ appearing in Proposition 3.1.1. Note that $\mathcal{O}(\widetilde{U}(\tilde{z}_j, \le\eta_j)) \cong \mathcal{O}(\widetilde{U}(\delta\tilde{z}_j\delta^{-1}, \le\eta_j))$ using the following change of coordinates in terms of universal matrices $A^{(j)}$: for $? \in \{\varnothing, *, \prime\}$ we have $c^?_{ik} \mapsto c^?_{(i+1)(k+1)}$, where, for $1 \le i, k \le 3$, the integers $(i+1), (k+1) \in \{1, 2, 3\}$ are taken modulo 3.

is $\mathrm{Spf}(R/I_{\mathfrak{M},\nabla_\infty})$. We finally define $Y^{\le\eta,\tau,\nabla_\infty}$ as the pullback of $Y^{[0,2],\tau,\nabla_\infty} \hookrightarrow Y^{[0,2],\tau}$ along the closed immersion $Y^{\le\eta,\tau} \hookrightarrow Y^{[0,2],\tau}$.

Let $\mathcal{X}_{K,3}$ be the Noetherian formal algebraic stack over $\mathrm{Spf}\,\mathcal{O}$ defined in [Emerton and Gee 2023, Definition 3.2.1]. It has the property that for any complete local Noetherian $\mathcal{O}$-algebra R with finite residue field, the groupoid $\mathcal{X}_{K,3}(R)$ is equivalent to the groupoid of rank 3 projective R-modules equipped with a continuous G_K-action, see [Emerton and Gee 2023, §3.6.1]. (In particular, we will consider closed points of $\mathcal{X}_{K,3}(\mathbb{F})$ as continuous Galois representations $\bar{\rho}: G_K \to \mathrm{GL}_3(\mathbb{F})$, and conversely.) Moreover, by [Emerton and Gee 2023, Theorem 4.8.12], there is a unique $\mathcal{O}$-flat closed formal substack $\mathcal{X}^{\eta,\tau}$ of $\mathcal{X}_{K,3}$ which parametrizes, over finite flat $\mathcal{O}$-algebras, those G_K-representations which after inverting p are potentially crystalline with Hodge-Tate weight η and inertial type τ. We define $\mathcal{X}^{\le\eta,\tau}$ in the same way, except that the condition on Hodge-Tate weights becomes $\le\eta$. In particular, $\mathcal{X}^{\le\eta,\tau}$ is the scheme theoretic union of the substacks $\mathcal{X}^{\lambda,\tau}$ for λ dominant and $\lambda \le \eta$.

The substacks $\mathcal{X}^{\eta,\tau}$, $\mathcal{X}^{\le\eta,\tau}$ have the following fundamental properties:

$\tilde{z}_j t_{-\underline{1}}$	$A^{(j)}$	$\mathcal{O}(\widetilde{U}(\tilde{z}_j, \leq\eta_j))$
$\alpha\beta$	$\begin{pmatrix} c_{31}c_{12}(c^*_{32})^{-1} & c_{12} & c_{13}+(v+p)c^*_{13} \\ vc^*_{21} & c_{22} & c_{23}+(v+p)c'_{23} \\ vc_{31} & vc^*_{32} & \big(c_{31}c_{23}(c^*_{21})^{-1}+(v+p)c'_{33}\big) \end{pmatrix}$	$\mathcal{O}[c_{12}, c_{13}, c^*_{13}, c^*_{21}, c_{22}, c_{23}, c'_{23}, c_{31}, c^*_{32}, c'_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = \begin{pmatrix} c_{22}c_{31}+pc^*_{21}c^*_{32},\ c_{12}c_{23}-c_{22}c_{13} \\ c^*_{21}c^*_{32}c_{13}-pc^*_{21}c^*_{32}c^*_{13}-c'_{33}c^*_{21}c_{12}=0 \end{pmatrix}$
$\beta\alpha$	$\begin{pmatrix} c_{11} & \big((c^*_{31})^{-1}c_{11}c_{32}+(v+p)c^*_{12}\big) & c_{13} \\ 0 & (v+p)c'_{22} & (v+p)c^*_{23} \\ c^*_{31}v & c_{32}v & c_{33}+(v+p)c'_{33} \end{pmatrix}$	$\mathcal{O}[c_{11}, c^*_{12}, c_{13}, c'_{22}, c^*_{23}, c^*_{31}, c_{32}, c_{33}, c'_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = \begin{pmatrix} c_{11}c_{33}+pc^*_{31}c_{13} \\ c'_{22}(c_{11}c'_{33}-c_{13}c^*_{31})-pc^*_{23}c^*_{12}c^*_{31} \end{pmatrix}$
α	$\begin{pmatrix} c_{11} & c_{12}+(v+p)c^*_{12} & c_{13} \\ c^*_{21}v & c_{22}+(v+p)c'_{22} & c_{23} \\ c_{31}v & c_{32}v & (c_{33}+(v+p)c^*_{33}) \end{pmatrix}$	$\mathcal{O}[c_{11}, c_{12}, c^*_{12}, c_{13}, c^*_{21}, c_{22}, c'_{22}, c_{23}, c_{31}, c_{32}, c_{33}, c^*_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = \begin{pmatrix} c_{11}c_{22}+pc_{12}c^*_{21},\ c_{11}c_{23}+pc_{13}c^*_{21},\ c_{12}c_{23}-c_{13}c_{22}, \\ c_{11}c_{32}-c_{31}c_{12},\ c_{11}c_{33}+pc_{31}c_{13},\ c_{12}c_{33}+pc_{32}c_{13}, \\ pc^*_{21}c_{32}+c_{22}c_{31},\ c^*_{21}c_{33}-c_{23}c_{31},\ c_{22}c_{33}+pc_{32}c_{23}, \\ c_{11}c'_{22}c^*_{33}+c_{13}c^*_{21}c_{32}-c_{13}c'_{22}c_{31} \\ \qquad -c_{12}c^*_{21}c^*_{33}+pc^*_{21}c^*_{12}c^*_{33} \end{pmatrix}$
id	$\begin{pmatrix} c_{11}+c^*_{11}(v+p) & c_{12} & c_{13} \\ vc_{21} & c_{22}+c^*_{22}(v+p) & c_{23} \\ vc_{31} & vc_{32} & c_{33}+c^*_{33}(v+p) \end{pmatrix}$	$\mathcal{O}[c_{11}, c^*_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c^*_{22}, c_{23}, c_{31}, c_{32}, c_{33}, c^*_{33}]/I_{\tilde{z}_j}$ $I_{\tilde{z}_j} = \begin{pmatrix} c_{11}c_{22}+pc_{12}c_{21},\ c_{11}c_{23}+pc_{13}c_{21},\ c_{12}c_{23}-c_{13}c_{22}, \\ c_{11}c_{32}-c_{31}c_{12},\ c_{11}c_{33}+pc_{31}c_{13},\ c_{12}c_{33}+pc_{32}c_{13}, \\ pc_{21}c_{32}+c_{22}c_{31},\ c_{21}c_{33}-c_{23}c_{31},\ c_{22}c_{33}+pc_{32}c_{23}, \\ c_{11}c^*_{22}c^*_{33}+c_{22}c^*_{33}c^*_{11}+c_{33}c^*_{11}c^*_{22} \\ \qquad -c^*_{11}c_{23}c_{32}-c^*_{22}c_{13}c_{31}-c^*_{33}c_{12}c_{21}+c_{21}c_{13}c_{32} \end{pmatrix}$

Table 1 (continued). The remaining relevant $\mathcal{O}$-algebras for Proposition 3.1.1.

Theorem 3.2.1 [Emerton and Gee 2023, Theorem 4.8.12, Proposition 4.8.10]. *Let $\mathcal{X}^{?,\tau}$ denote either the substack $\mathcal{X}^{\eta,\tau}$ or $\mathcal{X}^{\leq\eta,\tau}$. Then*:

(1) *The stack $\mathcal{X}^{?,\tau}$ is a p-adic formal algebraic stack, flat and topologically of finite type over* $\operatorname{Spf}\mathcal{O}$.

(2) *$\mathcal{X}^{\eta,\tau}$ is equidimensional of dimension $1+3f$, while for λ dominant and $\lambda<\eta$, $\mathcal{X}^{\lambda,\tau}$ is equidimensional of dimension $<1+3f$.*

(3) *Let $\bar{\rho}\in\mathcal{X}^{\eta,\tau}(\mathbb{F})$ corresponding to a mod p representation of G_K. Then the potentially crystalline deformation ring $R^{\eta,\tau}_{\bar{\rho}}$ is a versal ring to $\mathcal{X}^{\eta,\tau}$ at $\bar{\rho}$.*

(4) *For any smooth map* $\operatorname{Spf} R\to\mathcal{X}^{?,\tau}$ *with R being a p-adically complete, topologically of finite type $\mathcal{O}$-algebra, the ring R is reduced and $R[1/p]$ is regular.*

Using Proposition 3.1.1, we can finally relate the objects introduced so far:

Theorem 3.2.2. *Let $\tilde{z}\in\operatorname{Adm}^\vee(\eta)$ and assume that the character μ (appearing in the lowest alcove presentation (s,μ) of τ) is 4-deep. We have a commutative diagram of p-adic formal algebraic stacks* *over* $\mathrm{Spf}\,\mathcal{O}$:

$$\begin{tikzcd}
\widetilde{U}(\tilde{z}, \leq\eta)^{\wedge_p} \arrow[r] & Y^{\leq\eta,\tau}(\tilde{z}) \arrow[r, hook] & Y^{\leq\eta,\tau} \\
\widetilde{U}(\tilde{z}, \leq\eta, \nabla_{\tau,\infty}) \arrow[r] \arrow[u, hook] & Y^{\leq\eta,\tau,\nabla_\infty}(\tilde{z}) \arrow[r, hook] \arrow[u, hook] & Y^{\leq\eta,\tau,\nabla_\infty} \arrow[u, hook] \\
\widetilde{\mathcal{X}}^{\leq\eta,\tau}(\tilde{z}) \arrow[r] \arrow[u, "\cong"] & \mathcal{X}^{\leq\eta,\tau}(\tilde{z}) \arrow[r, hook] \arrow[u, "\cong"] & \mathcal{X}^{\leq\eta,\tau} \arrow[u, "\cong"] \\
\widetilde{\mathcal{X}}^{\eta,\tau}(\tilde{z}) \arrow[r] \arrow[u, hook] & \mathcal{X}^{\eta,\tau}(\tilde{z}) \arrow[r, hook] \arrow[u, hook] & \mathcal{X}^{\eta,\tau} \arrow[u, hook]
\end{tikzcd}$$

where

- *all the stacks appearing in the left column and in the central column are* ***defined*** *so that all the squares in the diagram are cartesian*;
- *the hooked horizontal arrows are open immersion*;
- *the left horizontal arrows are* $\underline{T}^\vee_{\mathcal{O}}$*-torsor for the twisted shifted conjugation action on the source* (*induced by the twisted shifted conjugation action on* $\widetilde{U}(\tilde{z}, \leq\eta)^{\wedge_p}$);
- *the vertical hooked arrows are closed immersions and the vertical arrows decorated with "$\cong$" are isomorphisms.*

In particular, $\widetilde{U}(\tilde{z}, \leq\eta, \nabla_{\tau,\infty})$ *is an affine p-adic formal scheme over* $\mathrm{Spf}\,\mathcal{O}$, *topologically of finite type. Furthermore, if* $\ell(\tilde{z}_j) \geq 2$ *for all* j, *then* $\widetilde{U}(\tilde{z}, \leq\eta, \nabla_{\tau,\infty})$ *is the p-adically completed tensor product over* $\mathcal{O}$ *of the rings in Table 2.*

Proof. This is [Le et al. 2023b, Proposition 7.2.3]. The last assertion follows from the computations in [Le et al. 2018a, §5.3] noting that the whole discussion there applies to the p-adic completion (as opposed to completions at closed points), and that the computations of [loc. cit.] can be performed with less stringent genericity assumptions (see the proof of Theorem 3.3.2 for the precise genericity). □

Remark 3.2.3. Note that $\mathcal{X}^{\eta,\tau} \subseteq \mathcal{X}^{\leq\eta,\tau}$ can be characterized as the union of the $(1+3f)$-dimensional irreducible components (which is the maximal possible dimension). In particular, by letting $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})$ denote the maximal reduced $\mathcal{O}$-flat $(1+6f)$-dimensional closed p-adic formal subscheme of $\widetilde{U}(\tilde{z}, \leq\eta, \nabla_{\tau,\infty})$, Theorem 3.2.2 gives an identification of $\widetilde{\mathcal{X}}^{\eta,\tau}(\tilde{z})$ with $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})$.

3.3. ***Special fibers.*** Let Fl denote the affine flag variety over $\mathbb{F}$ for $\mathrm{GL}_{3/\mathbb{F}}$ (with respect to the Iwahori relative to the upper triangular Borel), identified with the special fiber of $\mathrm{Gr}_{\mathcal{G},\mathcal{O}}$. As in [Le et al. 2023b, (4.7)], we define the closed sub-ind scheme $\mathrm{Fl}^{\nabla_0} \hookrightarrow \mathrm{Fl}$.

$\tilde{z}_j t_{-\underline{1}}$	$\mathcal{O}(\widetilde{U}(\tilde{z}_j, \eta, \nabla_{(s,\mu)}))$
$\alpha\beta\alpha\gamma$	$\mathcal{O}[c_{11}^*, c_{21}, c_{31}', c_{22}^*, c_{32}, c_{33}^*]$
$\beta\gamma\alpha\gamma$	$\mathcal{O}[c_{11}^*, c_{12}, c_{22}^*, c_{31}, c_{32}', c_{33}^*]$
$\beta\alpha\gamma$	$\mathcal{O}[c_{11}, c_{12}^*, c_{21}^*, c_{22}, c_{31}', c_{32}, c_{33}^*]/I_{\tilde{z}_j,\nabla_{(s,\mu)}}$ $I_{\tilde{z}_j,\nabla_{(s,\mu)}} = (c_{11}c_{22} + pc_{12}^*c_{21}^*)$
$\alpha\beta\gamma$	$\mathcal{O}[c_{11}^*, c_{21}', c_{22}, c_{23}^*, c_{31}', c_{32}^*, c_{33}]/I_{\tilde{z}_j,\nabla_{s_j(\mu_j+\eta_j)}}$ $I_{\tilde{z}_j,\nabla_{(s,\mu)}} = (c_{22}c_{33} + pc_{32}^*c_{23}^*)$
$\alpha\beta\alpha$	$\mathcal{O}[c_{32}, c_{23}, c_{33}', c_{31}^*, c_{22}^*, c_{13}^*]/I_{\tilde{z}_j,\nabla_{(s,\mu)}}$ $I_{\tilde{z}_j,\nabla_{s_j(\mu_j+\eta_j)}} = \big(c_{11}\big((\boldsymbol{a}-\boldsymbol{b})c_{23}c_{32} - (\boldsymbol{a}-\boldsymbol{c})c_{22}^*c_{33}'\big) + p(e' - \boldsymbol{a} + \boldsymbol{c})c_{31}^*c_{22}^*c_{13}^*\big)$
$\alpha\beta$	$\mathcal{O}[\![c_{31}, c_{22}, c_{12}, c_{23}', c_{33}', c_{21}^*, c_{13}^*, c_{32}^*]\!]/I_{\tilde{z}_j,\nabla_{s_j(\mu_j+\eta_j)}}$ $I_{\tilde{z}_j,\nabla_{(s,\mu)}} = \begin{pmatrix} c_{12}\big((\boldsymbol{a}-\boldsymbol{b})c_{31}c_{23}' + (\boldsymbol{b}-\boldsymbol{c})c_{21}^*c_{33}'\big) - p(e' - \boldsymbol{a} + \boldsymbol{c})c_{21}^*c_{32}^*c_{13}^*, \\ c_{22}c_{31} + pc_{21}^*c_{32}^* \end{pmatrix}$
$\beta\alpha$	$\mathcal{O}[\![c_{11}, c_{22}', c_{13}, c_{32}, c_{33}', c_{31}^*, c_{12}^*, c_{23}^*]\!]/I_{\tilde{z}_j,\nabla_{s_j(\mu_j+\eta_j)}}$ $I_{\tilde{z}_j,\nabla_{(s,\mu)}} = \begin{pmatrix} c_{11}\big((\boldsymbol{a}-\boldsymbol{b})c_{32}c_{23}^* - (\boldsymbol{a}-\boldsymbol{c})c_{22}'c_{33}'\big) - p(e' - \boldsymbol{a} + \boldsymbol{c})c_{12}^*c_{23}^*c_{31}^*, \\ c_{22}'(c_{11}c_{33}' - c_{13}c_{31}^*) - pc_{23}^*c_{12}^*c_{31}^* \end{pmatrix}$

Table 2. We list the $\mathcal{O}$-algebras $\mathcal{O}(\widetilde{U}(\tilde{z}_j, \eta_j, \nabla_{(s,\mu)}))$ appearing in Theorem 3.2.2. The triple $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is $(s'_{\mathrm{or},j'})^{-1}(\boldsymbol{a}'^{(j')})$. Note that

$$\mathcal{O}(\widetilde{U}(\tilde{z}_j, \eta, \nabla_{(s,\mu)})) \cong \mathcal{O}(\widetilde{U}(\delta\tilde{z}_j\delta^{-1}, \eta_j, \nabla_{(s,\mu)}))$$

using the following change of coordinates: for $? \in \{\varnothing, *, \prime\}$ we have $c_{ik}^? \mapsto c_{(i+1)(k+1)}^?$ (where, for $1 \le i, k \le 3$, the integers $(i+1), (k+1) \in \{1, 2, 3\}$ are taken modulo 3) and moreover $\boldsymbol{a} \mapsto \boldsymbol{b}$, $\boldsymbol{b} \mapsto \boldsymbol{c}$, $\boldsymbol{c} \mapsto \boldsymbol{a} - e'$.

3.3.A. *Labeling components of* $(\mathrm{Fl}^{\nabla_0})^{\mathcal{J}}$.

Definition 3.3.1. Let $(\widetilde{w}_1, \omega) \in \underline{\widetilde{W}}_1^+ \times (X^*(\underline{T}) \cap \underline{C}_0 + \eta)$ and write $S_{\mathbb{F}}^\circ(\widetilde{w}_1^* w_0^*)$ for the open affine Schubert cell associated to $\widetilde{w}_1^* w_0^* \in \underline{\widetilde{W}}^\vee$. We define $C_{(\widetilde{w}_1,\omega)}$ to be the Zariski closure in $(\mathrm{Fl}^{\nabla_0})^{\mathcal{J}}$ of $S_{\mathbb{F}}^\circ(\widetilde{w}_1^* w_0^*) t_{\omega^*} \cap (\mathrm{Fl}^{\nabla_0})^{\mathcal{J}}$. It is an irreducible subvariety of $(\mathrm{Fl}^{\nabla_0})^{\mathcal{J}}$ of dimension $3f$ when $\omega - \eta$ is 2-deep (see [Le et al. 2023b, Proposition 4.3.3]). It does not depend on the equivalence class of the pair $(\widetilde{w}_1, \omega)$ defined in Section 2.1.D.

3.3.B. $\underline{T}_{\mathbb{F}}^\vee$*-torsors.* Replacing the Iwahori with the pro-v Iwahori in the construction of Fl yields a $\underline{T}_{\mathbb{F}}^\vee$-torsor $\widetilde{\mathrm{Fl}}^{\mathcal{J}} \to \mathrm{Fl}^{\mathcal{J}}$. We use $\tilde{\cdot}$ to denote the pullback via this $\underline{T}_{\mathbb{F}}^\vee$-torsor of objects introduced so far (e.g., $\widetilde{C}_{(\widetilde{w}_1,\omega)} \subset (\widetilde{\mathrm{Fl}}^{\nabla_0})^{\mathcal{J}}$). Let $\widetilde{\mathrm{Fl}}^{[0,2]} \subset \widetilde{\mathrm{Fl}}$ be the pullback of the special fiber of $\mathrm{Gr}_{\mathcal{G},\mathcal{O}}^{[0,2]}$.

On the other hand, $\widetilde{\mathrm{Fl}}^{\mathcal{J}}$ is also endowed with a $\underline{T}^{\vee}_{\mathbb{F}}$-action by twisted shifted conjugation induced, at level of matrices, by $(A^{(j)})_{j\in\mathcal{J}} \mapsto (t_j A^{(j)} t_{j-1}^{-1})_{j\in\mathcal{J}}$ for $(t_j)_{j\in\mathcal{J}} \in \underline{T}^{\vee}_{\mathbb{F}}$.

3.3.C. *Labeling by Serre weights.* Recall from [Le et al. 2023b, Lemma 2.2.3] the bijection $(\widetilde{w}_1, \omega) \mapsto (F_{(\widetilde{w}_1,\omega)}, t_{\omega-\eta}\widetilde{w}_1 \underline{W}_a/\underline{W}_a)$ between lowest alcove presentations $(\widetilde{w}_1, \omega)$ of 0-deep Serre weights and 0-deep Serre weights σ with the choice of an algebraic central character $\zeta \in X^*(\underline{Z})$ lifting the central character of σ. If $(\widetilde{w}_1, \omega)$ maps to (σ, ζ) under this bijection, then we set $C^{\zeta}_{\sigma} \stackrel{\text{def}}{=} C_{(\widetilde{w}_1,\omega)}$. If the algebraic central character $\zeta \in X^*(\underline{Z})$ is understood, we will simply write C_{σ}.

3.3.D. *The local model diagram in characteristic p.* As explained in [Le et al. 2023b, §7.4] there is a bijection $\sigma \mapsto \mathcal{C}_{\sigma}$ between Serre weights and the top dimensional (namely, $3f$-dimensional) irreducible components of $\mathcal{X}_{K,3}$. (This is a relabeling of the bijection of [Emerton and Gee 2023, Theorem 6.5.1].) The main result of this section describes sufficiently generic $\mathcal{C}_{\sigma}$ in terms of the coordinate charts of Theorem 3.2.2.

Recall that $\Phi\text{-Mod}^{\text{ét},n}_{K,\mathbb{F}}$ denotes the fppf stack over $\mathbb{F}$ whose R-points, for a finite type $\mathbb{F}$-algebra R, parametrize projective rank n étale $(\varphi, \mathcal{O}_{\mathcal{E},K}\otimes_{\mathbb{F}_p} R)$-modules (recall that $\mathcal{O}_{\mathcal{E},K}$ denotes the p-adic completion of $(W(k)[\![v]\!])[1/v]$). We have a morphism $(\mathcal{X}_{K,3})_{\mathbb{F}} \to \Phi\text{-Mod}^{\text{ét},n}_{K,\mathbb{F}}$ corresponding to "restriction to G_{K_∞}" (see [Emerton and Gee 2023, §3.2]).

Theorem 3.3.2. *Let (s,μ) be a ζ-compatible 4-generic lowest alcove presentation of a tame inertial type τ. Let $\tilde{z} \in \mathrm{Adm}^{\vee}(\eta)$ and $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$.*

We have a commutative diagram

$$\begin{tikzcd}
 & & & \widetilde{\mathcal{C}}^{\zeta}_{\sigma} \arrow[d, hook] \\
\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}} \arrow[r, hook] \arrow[urrr, bend left, "\circ" marking] & \widetilde{U}(\tilde{z},\eta,\nabla_{\tau,\infty})_{\mathbb{F}} \arrow[r, hook] & \widetilde{U}(\tilde{z},\leq\eta)_{\mathbb{F}} \arrow[r] \arrow[ddd] & \widetilde{\mathrm{Fl}}^{[0,2]}_{\mathcal{J}}\cdot s^* t_{\mu^*+\eta^*} \arrow[ddd, "T^{\vee,\mathcal{J}}_{\mathbb{F}}"] \\
\widetilde{\mathcal{C}}_{\sigma}(\tilde{z}) \arrow[r, hook] \arrow[u, "\cong"] \arrow[d, "T^{\vee,\mathcal{J}}_{\mathbb{F}}"'] & \widetilde{\mathcal{X}}^{\eta,\tau}(\tilde{z})_{\mathbb{F}} \arrow[u, "\text{Remark 3.2.3}", "\cong"'] \arrow[d, "T^{\vee,\mathcal{J}}_{\mathbb{F}}"] & & \\
\mathcal{C}_{\sigma}(\tilde{z}) \arrow[r, hook] \arrow[d, hook, "\circ" marking] & \mathcal{X}^{\eta,\tau}(\tilde{z})_{\mathbb{F}} \arrow[d, hook, "\circ" marking] & & \\
\mathcal{C}_{\sigma} \arrow[r, hook] & \mathcal{X}^{\eta,\tau}_{\mathbb{F}} \arrow[r, hook] \arrow[d, hook] & Y^{\leq\eta,\tau}_{\mathbb{F}} \arrow[r, hook, "\iota_{(s,\mu)}"] & \left[\widetilde{\mathrm{Fl}}^{[0,2]}_{\mathcal{J}}\cdot s^* t_{\mu^*+\eta^*}/T^{\vee,\mathcal{J}}_{\mathbb{F}}\text{-sh.cnj}\right] \arrow[d, hook, "\iota_0"] \\
 & (\mathcal{X}_{K,3})_{\mathbb{F}} \arrow[rr] & & \Phi\text{-Mod}^{\text{ét},n}_{K,\mathbb{F}}
\end{tikzcd} \tag{3-2}$$

where

- *All the squares are cartesian (this defines the previously undefined objects $\mathcal{C}_{\sigma}(\tilde{z})$, $\widetilde{\mathcal{C}}_{\sigma}(\tilde{z})$ and $\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}}$).*

• *All the hooked arrows decorated with a circle are open immersions; all the hooked undecorated arrows are monomorphisms and, except* ι_0, *are moreover closed immersions; all the arrow decorated with* $T_{\mathbb{F}}^{\vee,\mathcal{J}}$ *are* $T_{\mathbb{F}}^{\vee,\mathcal{J}}$*-torsors.*

• $\widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}} \to Y_{\mathbb{F}}^{\leq \eta,\tau}$ *is an open immersion followed by a* $T_{\mathbb{F}}^{\vee,\mathcal{J}}$*-torsor map. The map* $\widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}} \to \widetilde{\mathrm{Fl}}_{\mathcal{J}}^{[0,h]} \cdot s^* t_{\mu^*+\eta^*}$ *is given by the formula*

$$(A^{(j)})_j \mapsto (A^{(j)})_j s^* t_{\mu^*+\eta^*}$$

and is a locally closed immersion.

• *The map* $\iota_{(s,\mu)}$ *is given, fpqc locally, by* $\mathfrak{M} \mapsto (A^{(j)}_{\mathfrak{M},\beta})_j s^* t_{\mu^*+\eta^*}$, *for any choice of eigenbasis* β *for* $\mathfrak{M} \in Y_{\mathbb{F}}^{\leq \eta,\tau}(R)$.

• *The map* ι_0 *is defined, fpqc locally, by sending the class of a tuple* $(A^{(j)})_{j\in\mathcal{J}}$ *to the free rank n étale* φ*-module with Frobenius given by* $(A^{(j)})_{j\in\mathcal{J}}$ *in the standard basis.*

• *If* $\sigma = F(\mathfrak{Tr}_{\mu+2\eta}(s(\varepsilon), a))$, *then the closed immersion* $\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}} \hookrightarrow \widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$ *corresponds to the ideal* $\sum_j \widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j} \mathcal{O}(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}})$ *with* $\widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j}$ *in Table 3 (or the unit ideal if* $\widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j}$, *up to symmetry, does not appear in Table 3 for some* j*); in particular* $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$ *is reduced.*

• *The bottom horizontal map identifies* $\mathcal{X}_{\mathbb{F}}^{\eta,\tau}$ *with the reduced union of* $[\widetilde{C}^{\zeta}_{\sigma'}/T_{\mathbb{F}}^{\vee,\mathcal{J}}\text{-sh.cnj}]$ *for* $\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau))$.

Proof. Theorem 3.2.2 (together with Remark 3.2.3) and [Le et al. 2023b, Proposition 5.4.7, Theorem 7.4.2], imply the existence of the portion of diagram (3-2) which excludes the leftmost vertical column, the top triangle, and the identification of $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$ with entries of Table 3. (In the notation of [Le et al. 2023b, Proposition 5.4.7] the monomorphism ι_0 would be denoted as $\iota_{s^* t_{\mu^*+\eta^*}}$, the morphism $\iota_{(s,\mu)}$ would be the diagonal arrow.) Furthermore, all stated properties of this portion of the diagram are already known to hold, except possibly for the last item. We now explain how to fill in the missing parts with all the desired properties except for the last item.

(1) We first deal with the case $\ell(\tilde{z}_j) \geq 2$ for all $0 \leq j \leq f-1$. In this situation, the computations in [Le et al. 2018a, §5.3.1] show that $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}} \hookrightarrow \widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}}$ identifies with the scheme given by Table 3. Indeed, we note that:

- The computations in [Le et al. 2018a, §5.3.1] of various completions of $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$ are in fact valid for $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$.
- The computations are performed with an unnecessary strong genericity condition: indeed, by using the "(1,3)-entry" of the leading term in the monodromy condition, one recovers the last displayed equation at page 59 of [loc. cit.] with $n-3$ replaced by $n-1$.

Choosing the closed subscheme $\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}}$ according to Table 3, we have now constructed the top horizontal arrow of diagram (3-2). This uniquely induces the leftmost vertical column of the diagram for *some* choice of irreducible component $\mathcal{C}_\kappa$ of $\mathcal{X}^{\eta,\tau}$. We need to show that

$\tilde{z}_j t_{-\underline{1}}$	$\widetilde{U}(\tilde{z}_j, \eta_j, \nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$	$(\varepsilon_j, a_j) \in \Sigma_0$	$\widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j}$
$\alpha\beta\alpha\gamma$	$\begin{pmatrix} v^2c_{11}^* & 0 & 0 \\ v^2c_{21} & vc_{22}^* & 0 \\ v^2d_{31} & vc_{32} & c_{33}^* \end{pmatrix}$	$(\varepsilon_1+\varepsilon_2, 0)$	(0)
$\beta\gamma\alpha\gamma$	$\begin{pmatrix} vc_{11}^* & vc_{12} & 0 \\ 0 & v^2c_{22}^* & 0 \\ vc_{31} & vc_{32}+v^2d_{32} & c_{33}^* \end{pmatrix}$ $(-1-b+c)c_{32}c_{11}^* - (-1-a+c)c_{12}c_{31} = 0$	$(\varepsilon_2, 1)$	(0)
$\beta\alpha\gamma$	$\begin{pmatrix} vc_{11} & vc_{12}^* & 0 \\ v^2c_{21}^* & vc_{22} & 0 \\ v(c_{31}+vd_{31}) & vc_{32} & c_{33}^* \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}} = 0;$ $(-1-a+c)c_{12}^*c_{31} - (-1-b+c)c_{32}c_{11} = 0$	$(\varepsilon_1+\varepsilon_2, 0)$ $(\varepsilon_2, 1)$	(c_{11}) (c_{22})
$\alpha\beta\gamma$	$\begin{pmatrix} v^2c_{11} & 0 & 0 \\ v(c_{21}+vd_{21}) & c_{22} & c_{23}^* \\ v(c_{21}c_{33}(c_{23}^*)^{-1}+vd_{31}) & vc_{32}^* & c_{33} \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}} = 0;$ $(-1-a+c)c_{21}c_{32}^* + (b-c)d_{31}c_{22} = 0$	$(\varepsilon_1+\varepsilon_2, 0)$ $(\varepsilon_1, 1)$	(c_{22}) (c_{33})
$\alpha\beta\alpha$	$\begin{pmatrix} c_{11} & c_{11}c_{32}(c_{31}^*)^{-1} & d_{33}c_{11}(c_{31}^*)^{-1}+vc_{13}^* \\ 0 & vc_{22}^* & vc_{23} \\ vc_{31}^* & vc_{32} & vd_{33} \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}} = 0;$ $c_{11}((a-b)c_{23}c_{32} - (a-c)c_{33}^*d_{33}) = 0$	$(0, 0)$ $(0, 1)$	(c_{11}) $((a-b)c_{23}c_{32} - (a-c)c_{33}^*d_{33})$

Table 3 (continued)

Table 3. The table records data relevant to Theorem 3.3.2. The first column records the components of $\tilde{z}$. The second column records the coordinates of $(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}})$ in terms of the universal matrix $A^{(j)}$ and the relations between its coefficients. Recall that in the statement of Theorem 3.3.2 the Serre weight σ is parametrized by $(\mu + \eta - \lambda + s(\varepsilon), \underline{a}) \in \underline{\Lambda}_R^{\lambda} \times \mathcal{A}$. The ideal corresponding to the closed immersion $\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}} \hookrightarrow \widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$ is of the form $\sum_{j=0}^{f-1} \widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j}$, where each $\widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j}$ is a minimal prime ideal of $\mathcal{O}(\widetilde{U}(\tilde{z}_j, \eta_j, \nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}})$. The elements $(\varepsilon_j, a_j) \in \Sigma_0$ are specified in the third column and the ideal $\widetilde{\mathfrak{P}}_{(\varepsilon_j,a_j),\tilde{z}_j}$ specified in the fourth column records. The structure constants that feature in the presentation are given by $(a, b, c) \in \mathbb{F}_p^3$ with $(a, b, c) \equiv s_j^{-1}(\mu_j + \eta_j) \mod p$.

$\tilde{z}_j t_{-\underline{1}}$	$\widetilde{U}(\tilde{z}_j, \eta_j, \nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$	$(\varepsilon_j, a_j) \in \Sigma_0$	$\widetilde{\mathfrak{P}}_{(\varepsilon_j, a_i), \tilde{z}_j}$
$\alpha\beta$	$\begin{pmatrix} c_{13}c_{12}(c^*_{32})^{-1} & c_{12} & c_{13}+vc^*_{13} \\ vc^*_{21} & c_{22} & c_{23}+vd_{23} \\ vc_{31} & vc^*_{32} & c_{31}c_{23}(c^*_{21})^{-1}+vd_{33} \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}}=0;$ $c_{12}((a-b)c_{31}d_{23}-(b-c)d_{33}c^*_{21})=0;$ $(-1-a+c)c_{23}c^*_{32}=(-1-a+b)c_{22}d_{33}=0$	$(\varepsilon_1, 1)$	(c_{12}, c_{31})
		$(\varepsilon_1-\varepsilon_2, 0)$	(c_{31}, d_{33})
		$(0, 0)$	(c_{12}, c_{22})
		$(0, 1)$	$(c_{22}, (b-c)d_{33}c^*_{21}+(a-b)c_{31}d_{23})$
$\beta\alpha$	$\begin{pmatrix} c_{11} & (c^*_{31})^{-1}c_{11}c_{32}+vc^*_{12} & c_{13} \\ 0 & vd_{22} & vc^*_{23} \\ vc^*_{31} & vc_{32} & c_{33}+vd_{33} \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}}=0;$ $c_{11}((a-b)c_{32}c^*_{23}-(a-c)d_{22}d_{33})=0;$ $(1+a-c)c_{33}c^*_{23}c^*_{12}=c_{13}((a-b)c_{32}c^*_{23}-(a-c)d_{22}d_{33})$	$(\varepsilon_2, 1)$	(d_{22}, c_{11})
		$(\varepsilon_2-\varepsilon_1, 0)$	(d_{22}, c_{32})
		$(0, 0)$	(c_{11}, c_{13})
		$(0, 1)$	$\big((a-b)c_{32}c^*_{23}-(a-c)d_{22}d_{33},$ $c_{13}c^*_{31}-c_{11}d_{33}\big)=0$
α	$\begin{pmatrix} c_{11} & c_{12}+vc^*_{12} & c_{13} \\ vc^*_{21} & c_{22}+vd_{22} & c_{23} \\ vc_{31} & vc_{32} & (c^*_{21})^{-1}c_{31}c_{23}+vc^*_{33}=0 \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}}=0;$ $(a-b)c_{12}c^*_{33}-(a-c)c_{13}\widetilde{c}_{32}=0;$ $(e-a+c)c_{23}\widetilde{c}_{32}-(e-a+b)c_{22}c^*_{33}=0;$ $(c-1-a)c_{31}c_{23}c^*_{12}-(c-1-a)c_{31}c_{13}d_{22}$ $+(c-1-b)c_{32}c_{13}c^*_{21}+c_{12}c^*_{33}c^*_{21}-c_{11}d_{22}c^*_{33}=0$	$(\varepsilon_1, 1)$	(c_{11}, c_{13}, c_{31})
		$(\varepsilon_2, 0)$	$(c_{11}, c_{31}, c_{32}c^*_{21}-d_{22}c_{31})$
		$(\varepsilon_2, 1)$	$\big(c_{11}, c_{32}c^*_{21}-d_{22}c_{31},$ $(a-b)c_{13}d_{22}+(-1-a+c)c_{23}c^*_{12}\big)$
		$(\varepsilon_2-\varepsilon_1, 0)$	$(c_{23}, d_{22}, c_{32}c^*_{21}-d_{22}c_{31})$
		$(0, 0)$	(c_{11}, c_{13}, c_{23})
		$(0, 1)$	$\big(c_{11}c^*_{33}-c_{13}c_{31}, c_{23},$ $(a-b)c_{31}d_{22}+(c-b)(c_{32}c^*_{21}-d_{22}c_{31})\big)$
id	$\begin{pmatrix} c_{11}+vc^*_{11} & c_{12} & c_{13} \\ vc_{21} & c_{22}+vc^*_{22} & c_{23} \\ vc_{31} & vc_{32} & c_{33}+vc^*_{33} \end{pmatrix}$ $I_{\tilde{z}_j,\mathbb{F}}=0;$ $(c-1-a)c^*_{22}c_{33}+(b-1-a)c_{22}c^*_{33}-(c-1-a)c_{23}c_{32}=0;$ $(a-b)c^*_{33}c_{11}+(c-1-b)c_{33}c^*_{11}-(a-b)c_{13}c_{31}=0;$ $(b-c)c^*_{11}c_{22}+(a-c)c_{11}c^*_{22}-(b-c)c_{12}c_{21}=0$	$(\varepsilon_1, 0)$	$(c_{ii},\ i=1,2,3,\ c_{21}, c_{31}, c_{23})$
		$(\varepsilon_1, 1)$	$\big(c_{31}, c_{33}, c_{11},$ $(-1-a+c)c_{32}c_{13}-(-1-a+b)c_{12}c^*_{33},$ $c_{21}c_{13}-c_{23}c^*_{11}\big)$
		$(\varepsilon_2, 0)$	$(c_{ii},\ i=1,2,3,\ c_{12}, c_{31}, c_{32})$
		$(\varepsilon_2, 1)$	$\big(c_{12}, c_{22}, c_{11},$ $(a-b)c_{21}c_{13}-(-1-b+c)c_{23}c^*_{11},$ $c_{21}c_{32}-c_{31}c^*_{22}\big)$
		$(0, 0)$	$(c_{ii},\ i=1,2,3,\ c_{13}, c_{23}, c_{12})$
		$(0, 1)$	$\big(c_{23}, c_{33}, c_{22},$ $(a-b)c_{21}c_{32}-(a-c)c_{31}c^*_{22},$ $c_{32}c_{13}-c_{12}c^*_{33}\big)$

Table 3 (continued). Further data relevant to Theorem 3.3.2.

(a) the composite of the top horizontal arrows identifies $\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}}$ with an nonempty open subscheme of $\widetilde{\mathcal{C}}^{\zeta}_{\sigma}$; and

(b) $\mathcal{C}_{\kappa} = \mathcal{C}_{\sigma}$.

For item (a), we use [Le et al. 2023b, Propositions 4.3.4, 4.3.5]: one checks that according to Table 3, the image of $\widetilde{\mathfrak{P}}_{\sigma,\tilde{z}}$ along the top horizontal arrows is, in the notation of [loc. cit.], a $T^{\vee,\mathcal{J}}_{\mathbb{F}}$-torsor over

$$S^{\nabla_0}_{\mathbb{F}}(\widetilde{w}_1, \widetilde{w}_2, \tilde{s})$$

for suitable choices of $\widetilde{w}_1$, $\widetilde{w}_2$, and for $\tilde{s}$ taken to be $t_{\mu+\eta}s$, and this is exactly an open subscheme of $\widetilde{\mathcal{C}}^{\zeta}_{\sigma}$.

For item (b), we use the just established item (a), and then use the same argument as in the third paragraph in the proof of [Le et al. 2023b, Theorem 7.4.2] to recognize that $\mathcal{C}_{\kappa}$ is actually $\mathcal{C}_{\sigma}$.

(2) We deal with the general case. The computations in [Le et al. 2018a, §8], namely Propositions 8.3, 8.11 and 8.13, as well as those of[1] [loc. cit., §5.3.2, §5.3.3], show that the closed embedding of

$$\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}} \hookrightarrow \widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}}$$

factors through $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathrm{table},\mathbb{F}}$, where we temporarily write $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathrm{table},\mathbb{F}}$ for the scheme defined in the second column of Table 3.

We have to prove that this closed immersion is actually an isomorphism. Let $n_{\tilde{z}}$ be

$$\#\bigl(W^?\bigl(\bar{\tau}(sz^*, \mu + s(\nu^*) + \eta)\bigr) \cap \mathrm{JH}(\overline{R_s(\mu+\eta)})\bigr).$$

Note that the arguments of [Le et al. 2018a, §8] show that $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathrm{table},\mathbb{F}}$ is *reduced*, and that its number of irreducible component is $n_{\tilde{z}}$. We will show that there are at least $n_{\tilde{z}}$ irreducible components of $\mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ which intersect the open substack $\mathcal{X}^{\eta,\tau}(\tilde{z})_{\mathbb{F}}$.

Now, from the previously established cases of diagram (3-2), we see that $\mathcal{X}^{\eta,\tau}$ must contain all the $\mathcal{C}_{\sigma'}$ which occurs in diagram (3-2) for $\tilde{z}'$ such that $\ell(\tilde{z}'_j) \geq 3$ for all $0 \leq j \leq f-1$. In particular, $\mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ contains all $\mathcal{C}_{\sigma'}$ such that $\sigma' \in \mathrm{JH}(\overline{R_s(\mu+\eta)})$. Note that by definition,

$$\mathcal{X}^{\eta,\tau}(\tilde{z})_{\mathbb{F}} = \mathcal{X}^{\eta,\tau}_{\mathbb{F}} \cap \bigl[\widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}} s^* t_{\mu^*+\eta^*} / T^{\vee,\mathcal{J}}_{\mathbb{F}}\text{-sh.cnj}\bigr].$$

We are thus reduced to showing that there are at least $n_{\tilde{z}}$ choices of σ' as above such that

$$\mathcal{C}_{\sigma'} \cap \bigl[\widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}} s^* t_{\mu^*+\eta^*} / T^{\vee,\mathcal{J}}_{\mathbb{F}}\text{-sh.cnj}\bigr] \neq \varnothing.$$

[1]The computations of [Le et al. 2018a, §5.3.2, §5.3.3] are also performed with unnecessary strong genericity conditions: again using the "(1,3)-entries" of the leading term in the monodromy condition recovers, in Sections 5.3.2 and 5.3.3 of [Le et al. 2018a], respectively, the equations

$$c_{11}((a-b)c_{32}c^*_{23} - (a-c)c'_{22}c'_{33}) - p(e-a+c)c^*_{12}c^*_{23}c^*_{31} + O(p^{n-1}),$$
$$c_{12}((a-b)c_{31}c'_{23} + (b-c)c^*_{21}c'_{33}) - p(e-a+c)c^*_{21}c^*_{32}c^*_{13} + O(p^{n-1})$$

(see the last displayed equations on pages 60 and 61 of [Le et al. 2018a]) obtaining the "monodromy equations" claimed in [loc. cit.] as soon as μ is 4-deep in alcove $\underline{C}_0$.

But this last condition is equivalent to

$$\widetilde{C}^{\zeta}_{\sigma'} \cap \widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}} s^* t_{\mu^*+\eta^*} \neq \varnothing,$$

and in turn equivalent to $\tilde{z}s^*t_{\mu^*+\eta^*} \in \widetilde{C}^{\zeta}_{\sigma'}$. To summarize, we need to show the combinatorial statement that the number of $\widetilde{C}^{\zeta}_{\sigma'}$ which contain $\tilde{z}s^*t_{\mu^*+\eta^*}$ is exactly $n_{\tilde{z}}$. But this is the same combinatorial statement as [Le et al. 2023b, Theorem 4.7.6], and we observe that the conclusion of that Theorem holds in our current setup: this follows from the invariance property [Le et al. 2023b, Proposition 4.3.5] of $\widetilde{C}^{\zeta}_{\sigma'}$, as well as the fact that $\tilde{z}'s^*t_{\mu^*+\eta^*} \in \widetilde{C}^{\zeta}_{\sigma'}$ whenever $\widetilde{C}^{\zeta}_{\sigma'}$ occurs in diagram (3-2) for $\tilde{z}'$ such that $\ell(\tilde{z}'_j) \geq 3$ for $0 \leq j \leq f-1$.

At this point, we have shown that $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$ identifies with $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathrm{table},\mathbb{F}}$, and thus we establish the top horizontal arrow of diagram (3-2). The rest of the proof now is exactly the same as in the previous case.

Finally, it remains to check the last item in the theorem. But the reducedness follow from the reducedness of each $\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}}$, while the identification of the irreducible components was already established in the arguments above. □

Corollary 3.3.3. *Let $\tilde{z} \in \mathrm{Adm}^{\vee}(\eta)$ and assume that the character μ (appearing in the lowest alcove presentation (s, μ) of τ) is 4-deep. Then*

- *$\mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ is reduced;*
- *$\mathcal{O}(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty}))$ is a normal domain; and*
- *for any $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ the ring $R^{\eta,\tau}_{\bar{\rho}}$ is either 0 or a normal domain.*

Remark 3.3.4. It can be showed that both $\mathcal{O}(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty}))$ and $R^{\eta,\tau}_{\bar{\rho}}$ are Cohen–Macaulay. This can be done by either explicit inspection of the schemes occurring in Table 3 as in [Le et al. 2018a, §8], or by using the cyclicity of patched modules proven in Theorem 5.3.1 below.

Proof. The fact that $\mathcal{O}\big(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})\big)$ is a normal domain follows from the fact that its special fiber is reduced, as in the proof of [Le et al. 2018a, Proposition 8.5]. The statement for $R^{\eta,\tau}_{\bar{\rho}}$ follows in the same way, noting that $R^{\eta,\tau}_{\bar{\rho},\mathbb{F}}$, being (equisingular to) a completion of the excellent reduced ring $\mathcal{O}(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})_{\mathbb{F}})$, is reduced. □

We can finally introduce the notion of Serre weights attached to a continuous Galois representation $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$.

Definition 3.3.5. Let $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ be a continuous Galois representation. We define $W^g(\bar{\rho})$ to be the set of Serre weights σ such that $\bar{\rho} \in \mathcal{C}_{\sigma}(\mathbb{F})$ (cf. [Le et al. 2023b, Definition 9.1.2]).

If $\bar{\rho}|_{I_K}$ is *tame* so that $\bar{\rho}|_{I_K}$ is isomorphic to a tame inertial $\mathbb{F}$-type $\tau(w, \mu+\eta)$, then we define $W^?(\bar{\rho})$ to be $W^?(\bar{\tau}(w, \mu+\eta))$ (see Section 2.1.G for the latter). We also say that $\bar{\rho}$ is *N-generic* if $\tau(w, \mu+\eta)$ is.

Finally, we say that a Serre weight is *generic* if it is a Jordan–Hölder constituent of a 4-generic Deligne–Lusztig representation. (By equation (2-3) and Section 2.1.G, a Serre weight is generic if and only if it admits a lowest alcove presentation $(\widetilde{w}_1, \omega)$ such that $(\omega, \widetilde{w}_1 \cdot \underline{C}_0) \in t_{\mu+\eta}s(\Sigma)$ for some $\mu \in \underline{C}_0 \cap X^*(\underline{T})$ that is 4-deep.) A generic Serre weight is necessarily 2-deep by [Le et al. 2023b, Proposition 2.3.7]. We let $W^g_{\mathrm{gen}}(\bar{\rho})$ and $W^?_{\mathrm{gen}}(\bar{\rho})$ denote the subsets of generic Serre weights of $W^g(\bar{\rho})$ and $W^?(\bar{\rho})$, respectively.

Remark 3.3.6. When restricted to the present setting, the subset $W^g_{\mathrm{gen}}(\bar{\rho}) \subseteq W^g(\bar{\rho})$ defined in [Le et al. 2023b, Definition 9.1.2] (consisting of 8-deep Serre weights) is a subset of the set defined above.

Corollary 3.3.7. *Let $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ be semisimple and 4-generic. Then $W^g_{\mathrm{gen}}(\bar{\rho}) = W^?_{\mathrm{gen}}(\bar{\rho})$.*

Proof. Let σ be a generic Serre weight so that $\sigma \in \mathrm{JH}\big(\overline{R_s(\mu+\eta)}\big)$ for some 4-deep $\mu \in \underline{C}_0$ and $s \in \underline{W}$. By Theorem 3.3.2, if $\bar{\rho} \notin \mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ then $\sigma \notin W^g(\bar{\rho})$. Else $\bar{\rho}$, being semisimple, corresponds to a point in $T^{\vee,\mathcal{J}}_{\mathbb{F}}\tilde{z} \in \widetilde{U}(\tilde{z}, \leq \eta)_{\mathbb{F}}$ in the diagram (3-2). The proof of Theorem 3.3.2 (precisely, the end of the third paragraph in the proof of item (2) there) shows that $\bar{\rho} \in \widetilde{\mathcal{C}}_\sigma$ if and only if $\sigma \in W^?(\bar{\rho})$. □

Proposition 3.3.8. *If $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ is a Galois representation such that $\bar{\rho}^{\mathrm{ss}}|_{I_K}$ is 6-generic, then every Serre weight in $W^g(\bar{\rho})$ is generic, i.e., $W^g_{\mathrm{gen}}(\bar{\rho}) = W^g(\bar{\rho})$.*

Proof. Since the proof uses methods which are now well-known (see, e.g., [Gee et al. 2017, §3]) but orthogonal to those of this section, we will be brief. Let $\bar{\rho}$ be as in the statement of the proposition. Then in particular, $p > 6$. Suppose that $F(\lambda) \in W^g(\bar{\rho})$. Let $\bar{\rho}' : G_K \to \mathrm{GL}_3(\mathbb{F})$ be a maximally nonsplit Galois representation lying only on $\mathcal{C}_{F(\lambda)}$. Then $\bar{\rho}'$ has an ordinary crystalline lift of weight $\lambda + \eta$ by [Emerton and Gee 2023, Lemma 5.5.4]. Then by [Emerton and Gee 2014, Corollary A.7] there is an automorphic globalization $\bar{r}' : G_{F^+} \to \mathcal{G}_3(\mathbb{F})$ (where $\mathcal{G}_3$ is the algebraic group defined in [Clozel et al. 2008, §2.1] with $n = 3$) of $\bar{\rho}'$ which is potentially diagonalizably automorphic in the sense of [Le et al. 2019, Theorem 4.3.1]. The proof of [Le et al. 2019, Theorem 4.3.8] implies that $\mathrm{Hom}_{G(\mathcal{O}_{F^+_p})}\big(\bigotimes_{v|p} \overline{W(\lambda)}^\vee, S(U)\big)^{\mathrm{ord}} \neq 0$ in the notation of [loc. cit.]. Since the natural map $\mathrm{Hom}_{G(\mathcal{O}_{F^+_p})}\big(\bigotimes_{v|p} F(\lambda)^\vee, S(U)\big)^{\mathrm{ord}} \to \mathrm{Hom}_{G(\mathcal{O}_{F^+_p})}\big(\bigotimes_{v|p} \overline{W(\lambda)}^\vee, S(U)\big)^{\mathrm{ord}}$ is an isomorphism by [Gee and Geraghty 2012, Lemma 6.1.3], we conclude that $\mathrm{Hom}_{G(\mathcal{O}_{F^+_p})}\big(\bigotimes_{v|p} F(\lambda)^\vee, S(U)\big)$ is nonzero. This implies that $\mathrm{Hom}_{G(\mathcal{O}_{F^+_p})}\big(\bigotimes_{v|p} \bar{R}_1(\lambda)^\vee, S(U)\big) \neq 0$ by [Herzig 2011, Lemma 2.3]. In particular, $\bar{\rho}'$ has a potentially semistable lift of type (η, τ) for $\tau = \tau(1, \lambda)$. Since $W^g(\bar{\rho}') = \{F(\lambda)\}$, we conclude that $\mathcal{C}_{F(\lambda)}$ is a subset of the substack of $\mathcal{X}_{K,3}$ corresponding to potentially semistable representations of type (η, τ). In particular, since $\bar{\rho} \in \mathcal{C}_{F(\lambda)}(\mathbb{F})$, $\bar{\rho}$ has a potentially semistable lift of type (η, τ). By [Enns 2019, Lemma 5], $\bar{\rho}^{\mathrm{ss}}$ has a semistable lift of type (η, τ). Then $R_1(\lambda)$ is 1-generic by [Enns 2019, Proposition 7] (the proof of [Enns 2019, Theorem 8] shows that $R_1(\lambda)$ is 2-generic in the sense of [Enns 2019, Definition 2] where $\delta = 6$ and $n + 1 = 4$ here so that $R_1(\lambda)$ is 1-generic by [Le et al. 2019, Remark 2.2.8]). In particular, any potentially semistable lift of type (η, τ) is potentially crystalline. By the proof of [Le et al. 2019, Proposition 3.3.2], τ is 4-generic (the proof shows that *any* lowest alcove presentation of τ is 3-generic, but one compatible with a 6-generic lowest alcove presentation of $\bar{\rho}^{\mathrm{ss}}_{I_K}$ is 4-generic). We conclude that $R_1(\lambda)$ is 4-generic and that $F(\lambda)$ is generic. □

Corollary 3.3.9. *Let $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ be semisimple and 6-generic. Then $W^g(\bar{\rho}) = W^?(\bar{\rho})$.*

Proof. This follows from Corollary 3.3.7, Proposition 3.3.8, and the fact that $W^?_{\mathrm{gen}}(\bar\rho) = W^?(\bar\rho)$ in this case. $\square$

Remark 3.3.10. The notions and the results of this section hold true, *mutatis mutandis*, when the set S_p has arbitrary finite cardinality, and τ, $\bar\rho$ are a tame inertial L-parameter and a continuous $\mathbb{F}$-valued L-homomorphism respectively. In this case $\widetilde{U}(\tilde z)$, $\widetilde{U}^{[0,2]}(\tilde z)$, etc. are fibered products, over $\operatorname{Spec}\mathcal{O}$ and over the elements $v \in S_p$, of objects of the form $\widetilde{U}(\tilde z_v)$, $\widetilde{U}^{[0,2]}(\tilde z)$, etc. for $\tilde z_v \in (\widetilde{W}^\vee)^{\mathrm{Hom}_{\mathbb{Q}_p}(F_v^+,E)}$; the twisted shifted $\underline{T}^\vee_{\mathcal{O}}$-conjugation action is now induced by $A^{(j)} \mapsto t_j A^{(j)} s_j^{-1}(t_{\pi^{-1}(j)})$ for $j \in \mathrm{Hom}_{\mathbb{Q}_p}(F_p, E)$. Analogously, the algebraic stacks $Y^{[0,2],\tau}$, $\mathcal{X}_{K,3}$, $\mathcal{X}^{\eta,\tau}$, etc. are fibered products, over $\operatorname{Spf}\mathcal{O}$ and over the elements $v \in S_p$, of $Y^{[0,2],\tau_v}$, $\mathcal{X}_{F_v^+,3}$, $\mathcal{X}^{\eta_v,\tau_v}$, etc. The results of this section hold true in this more general setting.

4. Geometric Serre weights

The irreducible components of $\mathcal{X}_{K,n}$ from [Emerton and Gee 2023, Definition 3.2.1] give rise to a partition of $\mathcal{X}_{K,n}$ with locally closed parts

$$\bigcap_{\sigma\in W^+} \mathcal{C}_\sigma \setminus \bigcup_{\sigma\notin W^+} \mathcal{C}_\sigma$$

indexed by sets W^+ of Serre weights. It is of interest to determine the geometric properties of these pieces, e.g., when they are nonempty. In principle, one can directly study

$$\bigcap_{\sigma\in W^+} \mathcal{C}_\sigma \setminus \bigcup_{\substack{\sigma\notin W^+\\ \sigma \text{ generic}}} \mathcal{C}_\sigma \tag{4-1}$$

for a set W^+ of *generic* Serre weights using the relationship between $\mathcal{X}_{K,n}$ and Fl^{∇_0}, but this seems to be complicated even when $n = 4$. In this section, we determine when (4-1) is nonempty in generic cases when $n = 3$ using a notion of *obvious weights* for wildly ramified representations.

4.1. ***Intersections of generic irreducible components in*** $\mathbf{Fl}^{\nabla_0}$**.** We first study the geometry of Fl^{∇_0}. The set $\mathcal{J}$ will be a singleton, and so we will omit it from the notation. For $n \in \mathbb{N}$, let $\mathcal{C}_{n\text{-deep}}$ be the set of $\omega \in X^*(T)$ such that $\omega - \eta$ is n-deep in C_0. Recall from Section 3.3.A that given $(\widetilde{w}, \omega) \in \widetilde{W}_1 \times \mathcal{C}_{2\text{-deep}}$, we have the irreducible subvariety $C_{(\widetilde{w},\omega)}$ of Fl^{∇_0}. We define $\mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ as the union of the $C_{(\widetilde{w},\omega)}$ with $(\widetilde{w}, \omega) \in \widetilde{W}_1 \times \mathcal{C}_{2\text{-deep}}$ (in particular, these $C_{(\widetilde{w},\omega)}$ are its irreducible components). The action of $T^\vee_{\mathbb{F}}$ (resp. $\mathbb{G}_m$) on Fl^{∇_0} induced by right multiplication (resp. loop rotation $t \cdot v = t^{-1}v$) preserves $\mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ and its irreducible components. We let $\widetilde{T}^\vee_{\mathbb{F}}$ be the extended torus $T^\vee_{\mathbb{F}} \times \mathbb{G}_m$. We write $\mathrm{Fl}^{\nabla_0,T^\vee}_{2\text{-deep}}$ for the set of $T^\vee_{\mathbb{F}}$-fixed points (or equivalently $\widetilde{T}^\vee_{\mathbb{F}}$-fixed points) of $\mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ under right translation.

Definition 4.1.1. For $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}(\mathbb{F})$, let $W^g_{2\text{-deep}}(x^*)$ be the set

$$\{(\widetilde{w}, \omega) \in (\widetilde{W}_1 \times \mathcal{C}_{2\text{-deep}}) \mid x^* \in C_{(\widetilde{w},\omega)}(\mathbb{F})\}/\sim.$$

Recall that the equivalence relation is given by $(\widetilde{w}, \omega) \sim (t_\nu \widetilde{w}, \omega - \nu)$ for any $\nu \in X^0(T)$.

The main result of this section classifies the sets $W^g_{\text{2-deep}}(x^*)$ for $x^* \in \mathrm{Fl}^{\nabla_0}_{\text{2-deep}}(\mathbb{F})$. Combining the proof of Theorem 3.3.2 (namely, the combinatorial statement in the proof of item (2) there) and Corollary 3.3.7 (see also [Le et al. 2023b, Proposition 2.6.2]), we obtain the following description of $W^g_{\text{2-deep}}(x^*)$ for $x^* \in \mathrm{Fl}^{\nabla_0,T^\vee}_{\text{2-deep}}(\mathbb{F})$

Theorem 4.1.2. *For $x^* \in \mathrm{Fl}^{\nabla_0,T^\vee}_{\text{2-deep}}$, $W^g_{\text{2-deep}}(x^*)$ is the set*

$$\{(\widetilde{w}, x\widetilde{w}_2^{-1}(0)) \in \widetilde{W}_1^+ \times \mathcal{C}_{\text{2-deep}} \mid \widetilde{w}_2 \uparrow \widetilde{w}\}/\sim .$$

A first step towards understanding $W^g_{\text{2-deep}}(x^*)$ will be the determination of the subset $W_{\mathrm{obv}}(x^*) \subset W^g_{\text{2-deep}}(x^*)$ of *obvious weights*. It is defined as follows (see [Le et al. 2022]).

Definition 4.1.3. (1) Let $y \in \widetilde{W}$ and $\widetilde{w} \in \widetilde{W}_1$ be such that $y\widetilde{w}^{-1}(0) \in \mathcal{C}_{\text{2-deep}}$. We define $C_{(\widetilde{w},y\widetilde{w}^{-1}(0))}(y^*)$ to be the intersection $C_{(\widetilde{w},y\widetilde{w}^{-1}(0))} \cap T^\vee_{\mathbb{F}} \setminus \widetilde{U}(y^*)_{\mathbb{F}}$ in $\mathrm{Fl}^{\nabla_0}_{\text{2-deep}}$.

(2) Let $x^* \in \mathrm{Fl}^{\nabla_0}_{\text{2-deep}}(\mathbb{F})$. Then define $\mathrm{SP}(x^*)$ to be the set

$$\{(y, (\widetilde{w}, y\widetilde{w}^{-1}(0))) \in \widetilde{W} \times (\widetilde{W}_1 \times \mathcal{C}_{\text{2-deep}})/\sim \mid x^* \in C_{(\widetilde{w},y\widetilde{w}^{-1}(0))}(y^*)(\mathbb{F})\}$$

and $S(x^*)$ and $W_{\mathrm{obv}}(x^*)$ be the images of $\mathrm{SP}(x^*)$ under the projections of $\widetilde{W} \times (\widetilde{W}_1 \times \mathcal{C}_{\text{2-deep}})/\sim$ to $\widetilde{W}$ and $(\widetilde{W}_1 \times \mathcal{C}_{\text{2-deep}})/\sim$, respectively.

We call elements in $S(x^*)$ specializations of x^*. The set $\mathrm{SP}(x^*)$ is the set of specialization pairs consisting of a specialization and an obvious weight.

Lemma 4.1.4. *For $x^* \in \mathrm{Fl}^{\nabla_0,T^\vee}_{\text{2-deep}}(\mathbb{F})$ and $\widetilde{w} \in \widetilde{W}_1$ such that $x\widetilde{w}^{-1}(0) \in \mathcal{C}_{\text{2-deep}}$, $x^* \in C_{(\widetilde{w},x\widetilde{w}^{-1}(0))}$.*

Proof. For any $\widetilde{w} \in \widetilde{W}_1$, since $x^* \in S^\circ_{\mathbb{F}}(\widetilde{w}^* w_0^*)(x\widetilde{w}^{-1}w_0)^*$, $x^* \in S^{\nabla_0}_{\mathbb{F}}(\widetilde{w}, e, x\widetilde{w}^{-1}w_0) = C_{(\widetilde{w},x\widetilde{w}^{-1}(0))}$ (see [Le et al. 2023b, (4.9) and Proposition 4.3.5]). □

Remark 4.1.5. (1) Let $x^* \in \mathrm{Fl}^{\nabla_0,T^\vee}_{\text{2-deep}}$. Since x^* is the unique $T^\vee_{\mathbb{F}}$-fixed point of $C_{(\widetilde{w},x\widetilde{w}^{-1}(0))}(x^*)$,

$$\mathrm{SP}(x^*) = \{(x, (\widetilde{w}, x\widetilde{w}^{-1}(0))) \in \widetilde{W} \times (\widetilde{W}_1 \times \mathcal{C}_{\text{2-deep}})/\sim\}$$

by Lemma 4.1.4. In particular, $S(x^*) = \{x\}$. Moreover, for all tame inertial $\mathbb{F}$-types $\bar{\tau}$, $(\widetilde{w}, \omega) \in W_{\mathrm{obv}}(\widetilde{w}(\bar{\tau})^*)$ if and only if $F_{(\widetilde{w},\omega)} \in W_{\mathrm{obv}}(\bar{\tau})$ in the sense of [Le et al. 2023b, Definition 2.6.3].

(2) For $x^* \in \mathrm{Fl}^{\nabla_0}_{\text{2-deep}}(\mathbb{F})$, clearly, $W_{\mathrm{obv}}(x^*) \subset W^g_{\text{2-deep}}(x^*)$.

To determine $W^g_{\text{2-deep}}(x^*)$, we first determine $\mathrm{SP}(x^*)$. The idea is that $W_{\mathrm{obv}}(x^*)$ gives a lower bound for $W^g_{\text{2-deep}}(x^*)$ (Remark 4.1.5(2)) while $S(x^*)$ gives an upper bound by Lemma 4.1.7(2) and Theorem 4.1.2, and $\mathrm{SP}(x^*)$ combines these invariants into a more uniformly behaved set (see Corollary 4.1.10).

The following results are key to our analysis of $\mathrm{SP}(x^*)$. For $x^* \in \mathrm{Fl}^{\nabla_0}_{\text{2-deep}}$, let $\theta_{x^*} : \mathrm{SP}(x^*) \to W$ be the map that takes $(y, (\widetilde{w}, \omega))$ to the image of $y\widetilde{w}^{-1}$ in W.

Proposition 4.1.6. *For $x^* \in \mathrm{Fl}^{\nabla_0}_{\text{2-deep}}$, $\theta_{x^*} : \mathrm{SP}(x^*) \to W$ is injective.*

Proof. This is [Le et al. 2022, Proposition 3.6.4]. (It can also be proven by direct computation in the case of GL_3.) □

Let $\mathcal{I} \overset{\text{def}}{=} L^+\mathcal{G}_{\mathbb{F}}$ be the Iwahori group scheme. For $\widetilde{w}_\tau \in \widetilde{W}$ and $x^* \in \mathrm{Fl}$, let $\widetilde{w}(x^*, \widetilde{w}_\tau) \in \widetilde{W}$ be the unique element such that $x^* \in \mathcal{I}\backslash\mathcal{I}\widetilde{w}(x^*, \widetilde{w}_\tau)^*\mathcal{I}\widetilde{w}_\tau^*$.

Lemma 4.1.7. *Suppose that $y \in S(x^*)$. Then*

(1) $\widetilde{w}(y^*, \widetilde{w}_\tau) \le \widetilde{w}(x^*, \widetilde{w}_\tau)$; *and*

(2) $W^g_{2\text{-deep}}(x^*) \subset W^g_{2\text{-deep}}(y^*)$.

Proof. That $y \in S(x^*)$ implies that $x^* \in T^\vee_{\mathbb{F}} \backslash \widetilde{U}(y^*)$ or equivalently that y^* is in the $\widetilde{T}^\vee_{\mathbb{F}}$-orbit closure of x^*. For (1), y^* is in the ($\widetilde{T}^\vee_{\mathbb{F}}$-orbit) closure of $\mathcal{I}\backslash\mathcal{I}\widetilde{w}(x^*, \widetilde{w}_\tau)^*\mathcal{I}\widetilde{w}_\tau^*$ which implies the desired inequality. For (2), if $x^* \in C_{(\widetilde{w},\omega)}$, then $y^* \in C_{(\widetilde{w},\omega)}$ since $C_{(\widetilde{w},\omega)}$ is $\widetilde{T}^\vee_{\mathbb{F}}$-stable and closed. ☐

The following result provides a method to start with an element of $\mathrm{SP}(x^*)$ and produce another using a simple reflection in W.

Proposition 4.1.8. *Let $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ and $s \in W$ be a simple reflection. Suppose that for some $y \in \widetilde{W}$,*

(1) $y\widetilde{w}^{-1}st_{-\eta}\widetilde{w}(0) - \eta$, $y\widetilde{s w}^{-1}(0) - \eta$, *and* $yt_{-(sw)^{-1}(\eta)}(0) - \eta$ *are* 2*-deep*; *and*

(2) $\big(y, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*)$,

where $\widetilde{sw} \in \widetilde{W}_1$ is the unique element up to $X^0(T)$ such that $\widetilde{sw}\widetilde{w}^{-1}s^{-1} \in X^*(T)$. Then either $\big(y\widetilde{w}^{-1}s\widetilde{w}, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*)$ or $\big(y, (\widetilde{sw}, y\widetilde{sw}^{-1}(0))\big)$ is in $\mathrm{SP}(x^*)$.

Proof. Let $\widetilde{w}_\tau$ be $y(\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w})^{-1}$ so that $\widetilde{w}(y^*, \widetilde{w}_\tau) = \widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w}$. In Galois-theoretic language, this corresponds to the choice of the inertial type τ in [Le et al. 2018a, Proposition 7.16(3)]. We will see that there are only two possibilities for $\widetilde{w}(x^*, \widetilde{w}_\tau)$. First, $\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w} \le \widetilde{w}(x^*, \widetilde{w}_\tau)$ by Lemma 4.1.7(1).

Let $M(\le\eta)_{\mathbb{F}} \subset \mathrm{Fl}$ be the reduced closure of $\cup_{w\in W}\mathcal{I}\backslash\mathcal{I}t_{w^{-1}(\eta)}\mathcal{I}$ (this is compatible with the notation in Section 3.1). Let $M(\eta, \nabla_{\widetilde{w}_\tau(0)})_{\mathbb{F}}\widetilde{w}_\tau^*$ be the intersection $M(\eta)_{\mathbb{F}}\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0}$. For $\tilde{z} \in \widetilde{W}^\vee$, let $M(\eta, \nabla_{\widetilde{w}_\tau(0)})_{\mathbb{F}}(\tilde{z})$ denote the intersection $M(\eta, \nabla_{\widetilde{w}_\tau(0)})_{\mathbb{F}} \cap (T^\vee_{\mathbb{F}} \backslash \widetilde{U}(\tilde{z}))$ which is isomorphic to

$$U(\tilde{z}, \eta, \nabla_{\widetilde{w}_\tau(0)})_{\mathbb{F}} \overset{\text{def}}{=} T^\vee_{\mathbb{F}} \backslash \widetilde{U}(\tilde{z}, \eta, \nabla_{\widetilde{w}_\tau(0)})_{\mathbb{F}}.$$

Now, x^* lies in $C_{(\widetilde{w}, y\widetilde{w}^{-1}(0))}$ which is the closure of

$$M(\eta, \nabla_{\widetilde{w}_\tau(0)})_{\mathbb{F}}(t^*_{w^{-1}(\eta)})\widetilde{w}_\tau^* = \mathcal{I}\backslash\mathcal{I}t_{w^{-1}(\eta)}\mathcal{I}\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0},$$

hence $\widetilde{w}(x^*, \widetilde{w}_\tau) \le t_{w^{-1}(\eta)}$. Combining this with the last paragraph, we have

$$\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w} \le \widetilde{w}(x^*, \widetilde{w}_\tau) \le t_{w^{-1}(\eta)}.$$

Since $\ell(\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w}) = 3 = \ell(t_{w^{-1}(\eta)}) - 1$ (this is a consequence of a more general result in [Le et al. 2022], but can be checked directly using [Le et al. 2018a, Table 1]), we see that $\widetilde{w}(x^*, \widetilde{w}_\tau) = t_{w^{-1}(\eta)}$ or $\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w}$.

If $\widetilde{w}(x^*, \widetilde{w}_\tau) = t_{w^{-1}(\eta)}$, then $\big(y\widetilde{w}^{-1}s\widetilde{w}, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*)$ (this is represented by the red and blue parts in Figure 1). We claim that if $\widetilde{w}(x^*, \widetilde{w}_\tau) = \widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w}$, then $x^* \in C_{(\widetilde{sw}, y\widetilde{sw}^{-1}(0))}(y^*)$ (this is

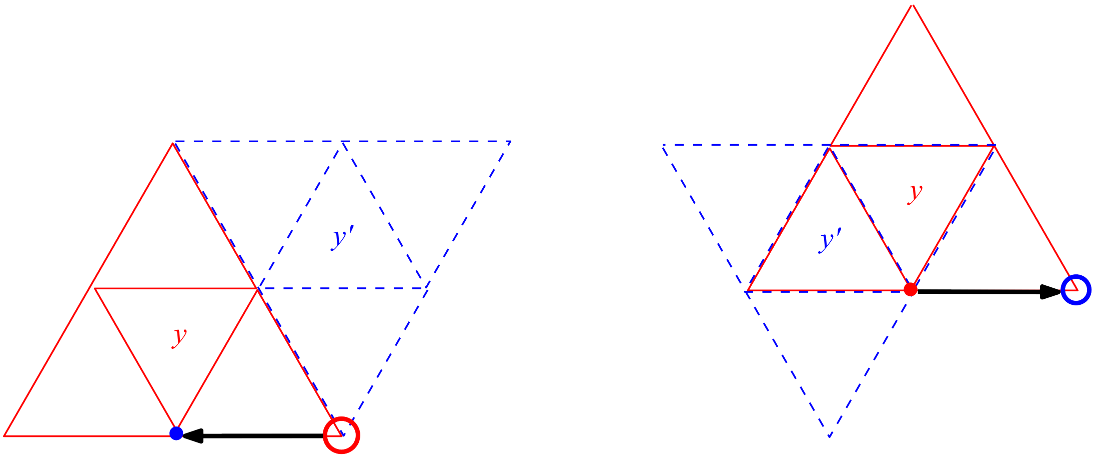

Figure 1. We illustrate the dichotomy given by the last paragraph of the proof of Proposition 4.1.8. We represent the data $(y, (\widetilde{w}, \omega)) \in \mathrm{SP}(x^*)$ by the alcove labeled by y and the dot (resp. circle) at $\omega \in X^*(T)/X^0(T)$ when $\widetilde{w} \cdot C_0$ is lower (resp. upper) alcove (thus the left picture is the case where $\widetilde{w} \cdot C_0$ is upper alcove, and the right picture the case where $\widetilde{w} \cdot C_0$ is lower alcove). The starting pair $\big(y, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*)$ is given by the red triangle (with vertexes labeled by the set $W_{\mathrm{obv}}(y^*)$) and the source of the arrow (labeled by the obvious weight $(\widetilde{w}, y\widetilde{w}^{-1}(0)) \in W_{\mathrm{obv}}(x^*)$). The dotted triangle represents the possible new specialization, while the tip of the arrow represents the new obvious weight.

represented by the arrows in Figure 1). It suffices to show that

$$\mathcal{I}\backslash\mathcal{I}(\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w})^*\mathcal{I}\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0} \subset \mathcal{I}\backslash\mathcal{I}t^*_{(sw)^{-1}(\eta)}\mathcal{I}t^*_{-(sw)^{-1}(\eta)}y^* \cap \mathrm{Fl}^{\nabla_0}. \tag{4-2}$$

Using (1), [Le et al. 2023b, Theorem 4.2.4] shows that both

$$\mathcal{I}\backslash\mathcal{I}(\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w})^*\mathcal{I}\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0} \quad \text{and} \quad \mathcal{I}\backslash\mathcal{I}(w_0s\widetilde{sw})^*\mathcal{I}(\widetilde{sw}^{-1}\widetilde{w}_h^{-1})^*\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0} \tag{4-3}$$

are isomorphic to $\mathbb{A}^2_{\mathbb{F}}$. The equality $\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{w} = \widetilde{sw}^{-1}\widetilde{w}_h^{-1}w_0s\widetilde{sw}$ implies that the latter space in (4-3) is contained in the former by the proof of [Le et al. 2023b, Proposition 4.3.4] (one can directly check that $(\widetilde{sw}^{-1}\widetilde{w}_h^{-1})(w_0s)\widetilde{sw}$ is a reduced factorization) so that the spaces in (4-3) are equal. Finally, observe that

$$\mathcal{I}\backslash\mathcal{I}(w_0s\widetilde{sw})^*\mathcal{I}(\widetilde{sw}^{-1}\widetilde{w}_h^{-1})^*\widetilde{w}_\tau^* \subset \mathcal{I}\backslash\mathcal{I}(w_0\widetilde{sw})^*\mathcal{I}(w_0sw_0^{-1})^*(\widetilde{sw}^{-1}\widetilde{w}_h^{-1})^*\widetilde{w}_\tau^*$$

and

$$\begin{aligned}\mathcal{I}\backslash\mathcal{I}(w_0\widetilde{sw})^*\mathcal{I}(w_0sw_0^{-1})^*(\widetilde{sw}^{-1}\widetilde{w}_h^{-1})^*\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0}& \\ &= \mathcal{I}\backslash\mathcal{I}(\widetilde{sw}^{-1}\widetilde{w}_h^{-1}w_0\widetilde{sw})^*\mathcal{I}(\widetilde{w}_h\widetilde{sw})^*(w_0sw_0^{-1})^*(\widetilde{sw}^{-1}\widetilde{w}_h^{-1})^*\widetilde{w}_\tau^* \cap \mathrm{Fl}^{\nabla_0} \\ &= \mathcal{I}\backslash\mathcal{I}t^*_{(sw)^{-1}(\eta)}\mathcal{I}t^*_{-(sw)^{-1}(\eta)}y^* \cap \mathrm{Fl}^{\nabla_0}\end{aligned}$$

by (the proof of) [Le et al. 2023b, Proposition 4.3.4]. Putting this all together yields (4-2). $\square$

Proposition 4.1.9. *For $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$, $\mathrm{SP}(x^*)$ is nonempty.*

Proof. By hypothesis, we have $x^* \in C_{(\widetilde{w},\omega)}$ for some $(\widetilde{w},\omega) \in \widetilde{W}_1 \times \mathcal{C}_{2\text{-deep}}$. It follows from the definition of $C_{(\widetilde{w},\omega)}$ that it is a closed subscheme of $\overline{S^\circ_{\mathbb{F}}(\widetilde{w}^* w_0^*)}t_{\omega^*} \cap \mathrm{Fl}^{\nabla_0}$. Thus $C_{(\widetilde{w},\omega)} \subset \cup_{\tilde{z}\le \widetilde{w}^* w_0^*} S^\circ_{\mathbb{F}}(\tilde{z})t_{\omega^*}$. Note that $S^\circ_{\mathbb{F}}(\tilde{z}) \subset T^\vee_{\mathbb{F}} \setminus \widetilde{U}(\tilde{z})_{\mathbb{F}} t_{\omega^*} = T^\vee_{\mathbb{F}} \setminus \widetilde{U}(\tilde{z}t_{\omega^*})_{\mathbb{F}}$, for example, see [Le et al. 2023b, Proposition 4.2.13]. It follows that $C_{(\widetilde{w},\omega)}$ has an open cover $\bigcup_{\tilde{z}\le \widetilde{w}^* w_0^*} C_{(\widetilde{w},\omega)}(\tilde{z}t_{\omega^*})$.

If $\widetilde{w}\cdot C_0 = C_0$, then any $\tilde{z} \le \widetilde{w}^* w_0^*$ satisfies $\tilde{z}^* \in W\widetilde{w}$. Choosing such a $\tilde{z}$ with $x^* \in C_{(\widetilde{w},\omega)}(\tilde{z}t_{\omega^*})$ gives $(t_\omega \tilde{z}^*, (\widetilde{w},\omega)) \in \mathrm{SP}(x^*)$ and we are done.

Suppose now that $\widetilde{w}\cdot C_0$ is the upper p-restricted alcove. For $\tilde{z} \le \widetilde{w}^* w_0^*$, either $\tilde{z}^* \in W\widetilde{w}^*$, or $\tilde{z}^* \in W\widetilde{w}'$, where $\widetilde{w}'$ is the unique element in $\widetilde{W}_1$ such that $\widetilde{w}' < \widetilde{w}$. In particular $\widetilde{w}'\cdot C_0 = C_0$. There are two cases:

- If $x^* \in C_{(\widetilde{w},\omega)}(\tilde{z}t_{\omega^*})$ for some $\tilde{z}^* \in W\widetilde{w}^*$, we get $(t_\omega \tilde{z}^*, (\widetilde{w},\omega)) \in \mathrm{SP}(x^*)$ as above.
- Otherwise, $x^* \in \big(\bigcup_{\tilde{z}^* \in W\widetilde{w}'} S^\circ_{\mathbb{F}}(\tilde{z})t_{\omega^*}\big) \cap \mathrm{Fl}^{\nabla_0} = C_{(\widetilde{w}',\omega)}$. Repeating our arguments with $(\widetilde{w},\omega)$ replaced by $(\widetilde{w}',\omega)$, we are also done in this case. □

Corollary 4.1.10. *Let $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$. If there exists $y_0 \in S(x^*)$ with $y_0(0) \in \mathcal{C}_{6\text{-deep}}$, then $\theta_{x^*} : \mathrm{SP}(x^*) \to W$ is bijective.*

Proof. By Proposition 4.1.6, it suffices to show that $\theta_{x^*} : \mathrm{SP}(x^*) \to W$ is surjective. By Proposition 4.1.9, $\mathrm{SP}(x^*)$ is nonempty. If $\big(y, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*)$ and $s \in W$ is a simple reflection, then either

$$\big(y\widetilde{w}^{-1}s\widetilde{w}, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*) \quad \text{or} \quad \big(y, (s\widetilde{w}, ys\widetilde{w}^{-1}(0))\big) \in \mathrm{SP}(x^*)$$

by Proposition 4.1.8 so that $\theta_{x^*}\big(y, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big)s$ is in the image of θ_{x^*}. (The hypothesis that $y_0(0) \in \mathcal{C}_{6\text{-deep}}$ guarantees that Proposition 4.1.8(1) applies.) Since simple reflections generate W, the result follows. □

Lemma 4.1.11. *Suppose that $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ such that there exists $y_0 \in S(x^*)$ with $y_0(0) \in \mathcal{C}_{6\text{-deep}}$. Then there exists $\lambda - \eta \in C_0$ and $w \in W$ such that the image of $W_{\mathrm{obv}}(x^*)$ under (2-2) is one of the sets*

(1) $w\{(0,0)\}$;

(2) $w\{(\varepsilon_1-\varepsilon_2, 1)\}$;

(3) $w\{(0,0), (\varepsilon_1-\varepsilon_2, 1)\}$;

(4) $w\{(0,0), (\varepsilon_1-\varepsilon_2, 1), (\varepsilon_2-\varepsilon_1, 1)\}$;

(5) $w\{(0,1), (\varepsilon_1, 0), (\varepsilon_2, 0), (\varepsilon_1+\varepsilon_2, 1)\}$; *and*

(6) $w\big\{(0,0), (\varepsilon_1-\varepsilon_2, 1), (\varepsilon_2-\varepsilon_1, 1), (\varepsilon_1, 0), (\varepsilon_2, 0), (\varepsilon_1+\varepsilon_2, 1)\big\}$.

Moreover, every possibility arises. Finally, with respect to the six above alternatives for $W_{\mathrm{obv}}(x^*)$ the image of $W^g_{2\text{-deep}}(x^*)$ under (2-2) is contained in

(1′) $w\{(0,0), (0,1)\}$;

(2′) $w\{(\varepsilon_1-\varepsilon_2, 1)\}$;

(3′) $w\{(0,0), (0,1), (\varepsilon_1-\varepsilon_2, 1)\}$;

(4′) $w\{(0,0),(0,1),(\varepsilon_1-\varepsilon_2,1),(\varepsilon_2-\varepsilon_1,1)\}$;

(5′) $w\{(0,1),(\varepsilon_1,0),(\varepsilon_1,1),(\varepsilon_2,0),(\varepsilon_2,1),(\varepsilon_1+\varepsilon_2,1)\}$; *and*

(6′) $w\big\{(0,0),(0,1),(\varepsilon_1-\varepsilon_2,1),(\varepsilon_2-\varepsilon_1,1),(\varepsilon_1,0),(\varepsilon_1,1),(\varepsilon_2,0),(\varepsilon_2,1),(\varepsilon_1+\varepsilon_2,1)\big\}$.

Remark 4.1.12. The sets in the second part of Lemma 4.1.11 are the minimal sets containing the corresponding sets in the first part closed under changing a 0 in the second argument to a 1. Since the set in the second part are obtained by taking intersections $\bigcap_{y\in S(x^*)} W^g_{2\text{-deep}}(y^*)$ which are closed under this operation, these sets are a natural upper bound for $W^g_{2\text{-deep}}(x^*)$.

Proof. We will illustrate the proof with various figures, all of which follow the same graphic conventions as in Figure 1.

Recall that we have canonical isomorphisms $\widetilde{W}/W_a \xrightarrow{\sim} X^*(Z)$ and $\pi_0(\mathrm{Fl}) \xrightarrow{\sim} X^*(Z)$. In this proof we will choose various $\lambda \in X^*(T)$ with the property that the image of t_λ in $X^*(Z)$ is the same as the image of x^*, and use (2-2) to identify $\big(\widetilde{W}_1 \times (X^*(T)\cap C_0+\eta)^{\lambda-\eta|_Z}\big)/\sim$ and $\Lambda^\lambda_W \times \mathcal{A}$, since the latter set is more convenient to work with here.

By Corollary 4.1.10, one obtains the elements of $W_{\mathrm{obv}}(x^*)$ by repeatedly applying the process described in Proposition 4.1.8 which we call a *simple walk*. We use the following two basic facts repeatedly.

(i) If $(\varepsilon,a)\in W_{\mathrm{obv}}(x^*)$, then either $W_{\mathrm{obv}}(x^*)=\{(\varepsilon,a)\}$ or there is a simple walk from some $(y,(\varepsilon,a))$ giving another element $(\varepsilon',a')\in W_{\mathrm{obv}}(x^*)$ in which case $a\neq a'$ and $\varepsilon-\varepsilon'\in W\{\varepsilon_1,\varepsilon_2\}$.

(ii) $W^g_{2\text{-deep}}(x^*)\subset\bigcap_{y\in S(x^*)} W^g_{2\text{-deep}}(y^*)$ by Lemma 4.1.7(2).

The analysis can be divided into a number of cases.

- Suppose that there is no element of the form $(\varepsilon,1)$ in $W_{\mathrm{obv}}(x^*)$. Then $W_{\mathrm{obv}}(x^*)$ consists of a single element by (i), which after changing λ, we assume to be $\{(0,0)\}$. Then by Corollary 4.1.10, $S(x^*)=\{t_\lambda w \mid w\in W\}$. Then $W^g_{2\text{-deep}}(x^*)\subset\bigcap_{y\in S(x^*)} W^g_{2\text{-deep}}(y^*)=\{(0,0),(0,1)\}$ by (ii); see Figure 2, left. This gives (1) and (1′).

- Suppose now that $(\varepsilon,1)\in W_{\mathrm{obv}}(x^*)$ for some $\varepsilon\in\Lambda_W$. Say $(y_1,(\varepsilon,1))\in \mathrm{SP}(x^*)$. If a simple walk produces $(y_2,(\varepsilon,1))\in\mathrm{SP}(x^*)$ (i.e., a new element of $S(x^*)$ rather than $W_{\mathrm{obv}}(x^*)$), then there exist $\lambda-\eta\in C_0$ and $w\in W$ such that $y_1=t_\lambda w$ and $y_2=t_\lambda w w_0$. Then

$$W^g_{2\text{-deep}}(y_1^*)\cap W^g_{2\text{-deep}}(y_2^*)=w\{(0,0),(0,1),(\varepsilon_1-\varepsilon_2,1),(\varepsilon_2-\varepsilon_1,1)\}.$$

Fact (i) precludes $w(0,1)$ from being in $W_{\mathrm{obv}}(x^*)$. Moreover, if $w(\varepsilon_1-\varepsilon_2,1)$ and $w(\varepsilon_2-\varepsilon_1,1)$ are in $W_{\mathrm{obv}}(x^*)$, then so is $w(0,0)$. Changing w if necessary, we assume that ε is $w(\varepsilon_1-\varepsilon_2,1)$. Then $W_{\mathrm{obv}}(x^*)$ is one of cases (2), (3), or (4).

  - In case (2), $S(x^*)$ has six elements by Corollary 4.1.10, and furthermore

$$W^g_{2\text{-deep}}(x^*)\subset\bigcap_{y\in S(x^*)} W^g_{2\text{-deep}}(y^*)=\{(\varepsilon,1)\}$$

  by (ii) — this is (2′); see Figure 2, middle.

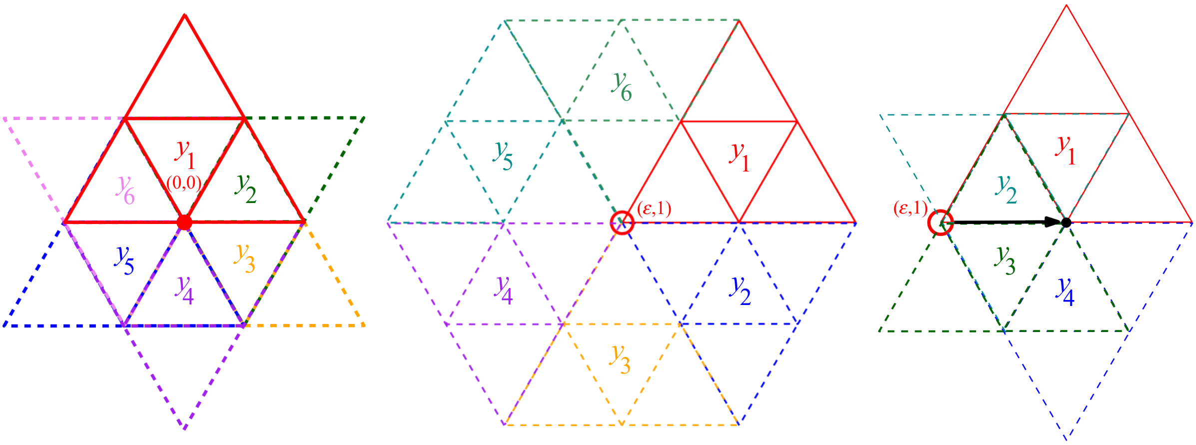

Figure 2. Left: Case where $W_{\text{obv}}(x^*) = \{(0,0)\}$ and $\#S(x^*) = \#W$. The set $W_{\text{obv}}(x^*)$ is pictured by the red dot and the set $S(x^*)$ by the 6 triangles. The simple walks only produce new specializations. Middle: Case where $W_{\text{obv}}(x^*) = \{(\varepsilon,1)\}$ and $\#S(x^*) = \#W$. The set $W_{\text{obv}}(x^*)$ is pictured by the red circle and the set $S(x^*)$ by the 6 triangles. Again, the simple walks only produce new specializations. Right: Case where $\#S(x^*) = 4$. We have the four specializations given by the triangles, and the starting obvious weight $(\varepsilon,1)$ by the red circle. The simple walk producing the new obvious weight is pictured by the black thickened arrow.

- In case (3), $S(x^*)$ has four elements, and $\bigcap_{y\in S(x^*)} W^g_{2\text{-deep}}(y^*)$ is given by (3′); see Figure 2, right.
- In case (4), given by Figure 3, left, $S(x^*) = \{y_1, y_2\}$ and $W^g_{2\text{-deep}}(y_1^*) \cap W^g_{2\text{-deep}}(y_2^*)$ is given by (4′).

• Finally, we suppose that $(\varepsilon,1) \in W_{\text{obv}}(x^*)$ for some $\varepsilon \in \Lambda_W$ *and* a simple walk starting with $(\varepsilon',1)$ for *any* $\varepsilon' \in \Lambda^\lambda_W$ such that $(\varepsilon',1) \in W_{\text{obv}}(x^*)$ *always* yields a new element of $W_{\text{obv}}(x^*)$. Let $(y_1,(\varepsilon,1)) \in \mathrm{SP}(x^*)$. After possibly changing λ, there exists $w \in W$ such that $W^g_{2\text{-deep}}(y_1^*)$ is

$$w\big\{(0,0),(0,1),(\varepsilon_1-\varepsilon_2,1),(\varepsilon_2-\varepsilon_1,1),(\varepsilon_1,0),(\varepsilon_1,1),(\varepsilon_2,0),(\varepsilon_2,1),(\varepsilon_1+\varepsilon_2,1)\big\}$$

and $\varepsilon = \varepsilon_1+\varepsilon_2$. By assumption, we have that $(y_1, w(\varepsilon_1,0))$ and $(y_1, w(\varepsilon_2,0)) \in \mathrm{SP}(x^*)$.

- If a simple walk starting with $(y_1, w(\varepsilon_1,0))$ yields $(y_2, w(\varepsilon_1,0)) \in \mathrm{SP}(x^*)$ for some $y_2 \in \widetilde{W}$, then $W^g_{2\text{-deep}}(x^*) \subset W^g_{2\text{-deep}}(y_1^*) \cap W^g_{2\text{-deep}}(y_2^*)$ which is (5′) (see Figure 3, middle).

  We claim that $w(\varepsilon_1,1) \notin W_{\text{obv}}(\bar\rho)$. If $w(\varepsilon_1,1) \in W_{\text{obv}}(\bar\rho)$ then, as argued before with $w(\varepsilon_1+\varepsilon_2,1)$, there would necessarily be two elements in $W_{\text{obv}}(x^*) \subset W^g_{2\text{-deep}}(y_1^*) \cap W^g_{2\text{-deep}}(y_2^*)$ which correspond to two adjacent vertices in Figure 3, middle. However, there are no such elements in (5′). Similarly, $w(\varepsilon_2,1) \notin W_{\text{obv}}(x^*)$. A simple walk from $(y_2, w(\varepsilon_1,0)) \in \mathrm{SP}(x^*)$ yields two elements in $\mathrm{SP}(x^*)$ — $(y_1, w(\varepsilon_1,0))$ and $(y_3, w(\varepsilon',a))$. If $y_2 \neq y_3$, then $w(\varepsilon_2,0) \notin W^g_{2\text{-deep}}(y_3^*)$, which contradicts Lemma 4.1.7(2). We conclude that $w(\varepsilon',a) = w(0,1) \in W_{\text{obv}}(\bar\rho)$. This gives the set in (5).

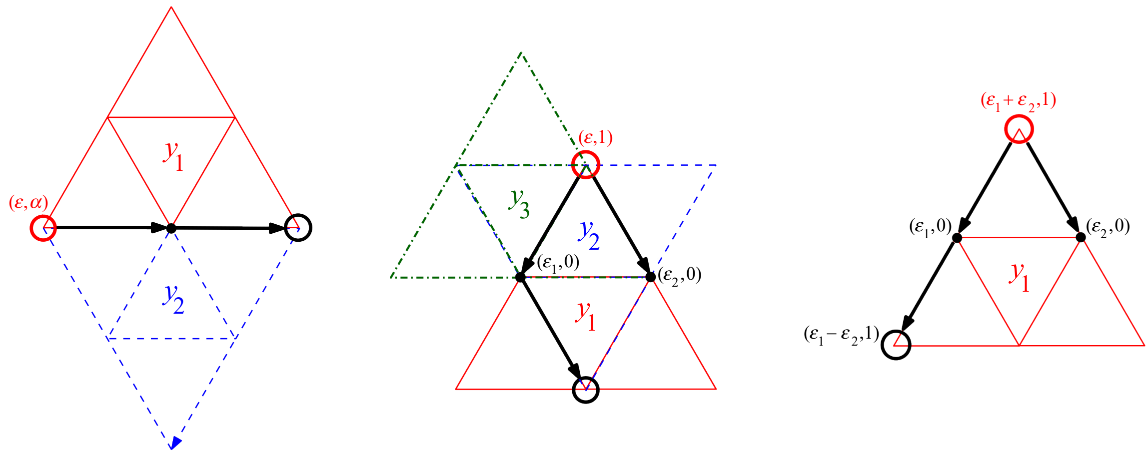

Figure 3. Left: Case where $\#S(x^*)=2$. We have the two specializations given by the red and blue triangles, and the starting obvious weight $(\varepsilon, 1)$ by the red circle. The simple walks producing the new obvious weights are pictured by the black thickened arrows. Middle: We picture (for $w=1$) the simple walks pertaining the third bullet point in the analysis of the proof of Lemma 4.1.11. The arrows from $(\varepsilon, 1) \in W_{\mathrm{obv}}(x^*)$ to $(\varepsilon_1, 0)$, $(\varepsilon_2, 0)$ exist by the assumption that any simple walk from $(\varepsilon, 1)$ always yields a new element of $W_{\mathrm{obv}}(x^*)$. In this middle picture, we consider the case where a simple walk from $(y_1, (\varepsilon_1, 0))$ yields $(y_2, (\varepsilon_1, 0)) \in \mathrm{SP}(x^*)$ with $y_2 \neq y_1$ (such y_2 is uniquely determined.) Since $(y_3, (\varepsilon_2, 0)) \notin \mathrm{SP}(x^*)$, we have the arrow from $(\varepsilon_1, 0)$ to $(0, 1)$. Right: Again, we consider the third bullet point in the analysis of the proof of Lemma 4.1.11, but now we consider the case when a simple walk from $(y_1, (\varepsilon_1, 0))$ yields $(y_1, (\varepsilon_1-\varepsilon_2, 1)) \in \mathrm{SP}(x^*)$ (hence, yields the arrow from $(\varepsilon_1, 0)$ to $(\varepsilon_1-\varepsilon_2, 1)$; compare with the middle figure.

– If the process in Proposition 4.1.8 from $(y_1, (\varepsilon_1, 0)) \in \mathrm{SP}(x^*)$ yields $(y_1, w(\varepsilon_1-\varepsilon_2, 1)) \in \mathrm{SP}(x^*)$ instead of $(y_2, w(\varepsilon_1, 0))$, then $S(x^*)=\{y_1\}$ by (ii) since y_1 is the unique $y \in \widetilde{W}$ such that $W^g_{2\text{-deep}}(y^*)$ contains $w(\varepsilon_1-\varepsilon_2, 1)$, $w(\varepsilon_1, 0)$, $w(\varepsilon_2, 0)$, and $w(\varepsilon_1+\varepsilon_2, 1)$; see Figure 3, right.

We conclude from Corollary 4.1.10 that $W_{\mathrm{obv}}(x^*)=W_{\mathrm{obv}}(y_1^*)$ which is given by (6). Then the upper bounds for $W^g_{2\text{-deep}}(x^*)$ in cases (5) and (6) again follow from (ii).

We now show that every possibility arises.

- One checks that case (6) arises when $x^*=\mathcal{I}\setminus\mathcal{I}(t_\lambda w)^*$.
- If we let $\gamma=t_{(1,0,-1)}w_0$ and $x^* \in \mathcal{I}\setminus\mathcal{I}\gamma^*\mathcal{I}(t_\lambda w)^*(\mathbb{F})$ but is not equal to $\mathcal{I}\setminus\mathcal{I}(t_\lambda w\gamma)^*$, then one can check that $t_\lambda w$, $t_\lambda w\gamma \in S(x^*)$ so that $W_{\mathrm{obv}}(x^*) \subset W^g_{2\text{-deep}}(t_\lambda w) \cap W^g_{2\text{-deep}}(t_\lambda w\gamma)$ by Lemma 4.1.7. This rules out (6). One can furthermore check that the image of $W_{\mathrm{obv}}(x^*)$ under (2-2) contains (5) so that out of the six possibilities it must equal (5).
- Next, if s_1 and $s_2 \in W$ denote the simple reflections and x^* is generic in

$$\mathcal{I}\setminus\mathcal{I}s_1s_2\mathcal{I}s_2s_1(t_\lambda w)^* \cap \mathcal{I}\setminus\mathcal{I}s_2s_1\mathcal{I}s_1s_2(t_\lambda w)^*,$$

then $t_\lambda w,\ t_\lambda w w_0 \in S(x^*)$ so that

$$W_{\mathrm{obv}}(x^*) \subset W^g_{2\text{-deep}}(t_\lambda w) \cap W^g_{2\text{-deep}}(t_\lambda w w_0)$$

by Lemma 4.1.7. This rules out (6). One can furthermore check that the image of $W_{\mathrm{obv}}(x^*)$ under (2-2) contains (4) so that out of the six possibilities it must equal (4).

- If x^* is generic in $\mathcal{I} \setminus \mathcal{I}s_2s_1\mathcal{I}s_1s_2(t_\lambda w)^*$, then $t_\lambda w, t_\lambda w s_2, t_\lambda w s_2 s_1, t_\lambda w w_0 \in S(x^*)$. Then by similar arguments as above $W_{\mathrm{obv}}(x^*)$ is contained in (3). One can furthermore check that $W_{\mathrm{obv}}(x^*)$ contains (3).
- If x^* is generic in $\mathcal{I} \setminus \mathcal{I}w_0\mathcal{I}(t_\lambda w)^*$, then $W^g_{2\text{-deep}}(x^*)$ only has one element, namely (1).
- If x^* is generic in an upper alcove component, then $W^g_{2\text{-deep}}(x^*)$ only has one element corresponding to (2). □

Theorem 4.1.13. *Suppose that $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ such that there exists $y_0 \in S(x^*)$ with $y_0(0) \in \mathcal{C}_{6\text{-deep}}$. Then there exist $\lambda - \eta \in C_0$ and $w \in W$ such that under (2-2) the image of $W_{\mathrm{obv}}(x^*)$ is as in Lemma 4.1.11 and the image of $W^g_{2\text{-deep}}(x^*)$ is correspondingly one of*

(1) $w\{(0,0)\}$ *or* $w\{(0,0),(0,1)\}$;

(2) $w\{(\varepsilon_1-\varepsilon_2,1)\}$;

(3) $w\{(0,0),(\varepsilon_1-\varepsilon_2,1)\}$;

(4) $w\{(0,0),(\varepsilon_1-\varepsilon_2,1),(\varepsilon_2-\varepsilon_1,1)\}$;

(5) $w\{(0,1),(\varepsilon_1,0),(\varepsilon_1,1),(\varepsilon_2,0),(\varepsilon_2,1),(\varepsilon_1+\varepsilon_2,1)\}$; *and*

(6) $w\big\{(0,0),(0,1),(\varepsilon_1-\varepsilon_2,1),(\varepsilon_2-\varepsilon_1,1),(\varepsilon_1,0),(\varepsilon_1,1),(\varepsilon_2,0),(\varepsilon_2,1),(\varepsilon_1+\varepsilon_2,1)\big\}$.

Moreover, every possibility arises.

Proof. We explain how the bounds on $W_{\mathrm{obv}}(x^*)$ and Table 3 can be used to determine $W^g_{2\text{-deep}}(x^*)$. Let $x^* \in \mathrm{Fl}^{\nabla_0}_{2\text{-deep}}$ be as in the statement of the theorem, and let λ and w be as in Lemma 4.1.11. Define $\Sigma^g(x^*)$ to be the image of $W^g_{2\text{-deep}}(x^*)$ under (2-2). Table 3 (with s_j in the notation there taken to be w in this proof) implies that the number of irreducible components of the completion of $\widetilde{U}(\tilde{z},\eta,\nabla_{w^{-1}(\mu+\eta)})_{\mathbb{F}}$ at an $\mathbb{F}$-point is never three (and that each irreducible component is smooth). Theorem 3.3.2 then implies that $\#\Sigma^g(x^*) \cap t_\nu s(\Sigma_0) \neq 3$ for all $t_\nu s \in W_a$. (The relevant type τ in Theorem 3.3.2 is 4-generic since $\widetilde{w}(y_0^*, \widetilde{w}(\tau)) \le \widetilde{w}(x^*, \widetilde{w}(\tau)) \in \mathrm{Adm}(\eta)$ by Lemma 4.1.7(1) and $y_0(0) \in \mathcal{C}_{6\text{-deep}}$.) This is the key fact that we will use in our analysis of $\Sigma^g(x^*)$.

The upper and lower bounds, say $\Sigma^{\mathrm{ub}}(x^*)$ and $\Sigma^{\mathrm{lb}}(x^*)$, respectively, for $\Sigma^g(x^*)$ from Lemma 4.1.11 give upper and lower bounds for $\Sigma^g(x^*) \cap t_\nu s(\Sigma_0)$ for each $t_\nu s \in W_a$. For each $(\varepsilon, a) \in \Sigma^{\mathrm{ub}}(x^*) \setminus \Sigma^{\mathrm{lb}}(x^*)$ in cases (3)–(6), one can choose $t_\nu s \in W_a$ such that

- $(\Sigma^{\mathrm{ub}}(x^*) \setminus \Sigma^{\mathrm{lb}}(x^*)) \cap t_\nu s(\Sigma_0) = \{(\varepsilon, a)\}$; and
- $\#\Sigma^{\mathrm{ub}}(x^*) \cap t_\nu s(\Sigma_0) = 3$ or $\#\Sigma^{\mathrm{lb}}(x^*) \cap t_\nu s(\Sigma_0) = 3$.

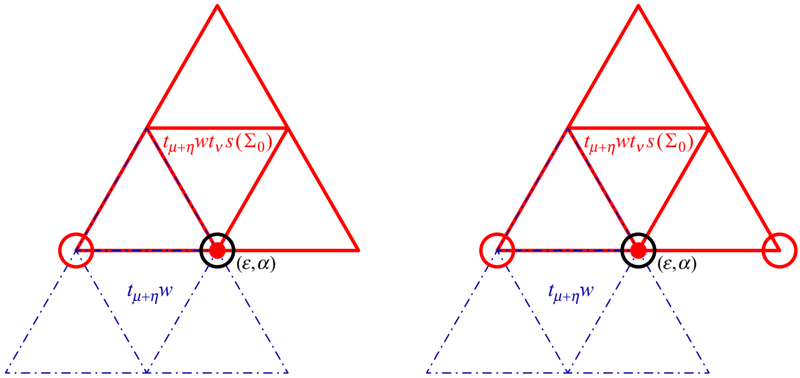

Figure 4. Choice of $t_\nu s$ in cases (3) and (4). The red circles and the dot represent $\Sigma^{\mathrm{lb}}(x^*)$, the black circle the element in $\Sigma^{\mathrm{ub}}(x^*) \setminus \Sigma^{\mathrm{lb}}(x^*)$. We have fixed an element $t_{\mu+\eta}w \in \widetilde{W}$ so that $s^{-1}t_{\nu^*}w^{-1}t_{\mu^*+\eta^*} \in S(x^*)$.

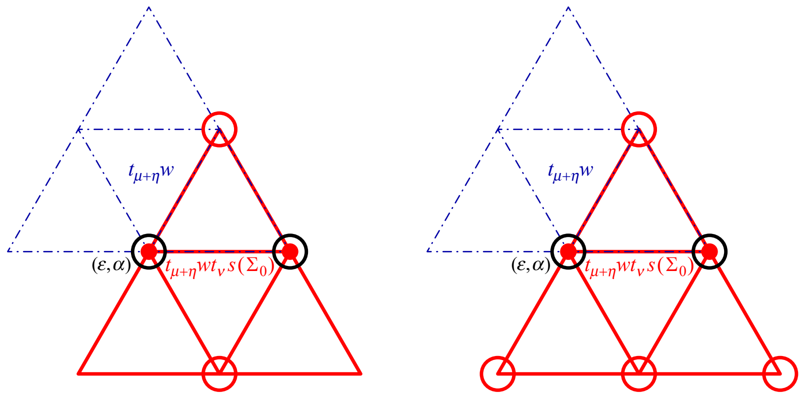

Figure 5. Choice of $t_\nu s$ in cases (5) and (6).The red circles and dots represent $\Sigma^{\mathrm{lb}}(x^*)$, the black circles the element in $\Sigma^{\mathrm{ub}}(x^*) \setminus \Sigma^{\mathrm{lb}}(x^*)$. We have fixed an element $t_{\mu+\eta}w \in \widetilde{W}$ so that $s^{-1}t_{\nu^*}w^{-1}t_{\mu^*+\eta^*} \in S(x^*)$.

Then the fact above implies that $\#\Sigma^{\mathrm{lb}}(x^*) \cap t_\nu s(\Sigma_0) = 3$ if and only if $(\varepsilon, a) \in \Sigma^g(x^*)$. From this, one checks that $\Sigma^g(x^*)$ is as claimed. We now illustrate in Figures 4 and 5 the choices of τ in cases (3)–(6).

Finally, we show that every possibility arises. Since Lemma 4.1.11 showed that every one of the six possibilities arises, we only need to show that the two possibilities in case (1) arise. The case $w\{(0,0)\}$ arises when x^* is generic on a lower alcove component so that $W^g_{2\text{-deep}}(x^*)$ only has one element corresponding to this component. The case $w\{(0,0), (0,1)\}$ arises when x^* is generic in the intersection of the two components corresponding to $w\{(0,0), (0,1)\}$. Indeed, this intersection is two-dimensional so as long as x^* is not in cases (5) or (6) which are of dimensions one and zero, respectively, the case $w\{(0,0), (0,1)\}$ must apply. □

Remark 4.1.14. The notions in this section extend to the case of products: if $x^* = (x_i^*)_{i\in\mathcal{J}} \in (\mathrm{Fl}^{\nabla_0}_{2\text{-deep}})^{\mathcal{J}}$, we let $W^g_{2\text{-deep}}(x^*)$ and $W_{\mathrm{obv}}(x^*)$ be the subsets $\prod_{i\in\mathcal{J}} W^g_{2\text{-deep}}(x_i^*)$ and $\prod_{i\in\mathcal{J}} W_{\mathrm{obv}}(x_i^*)$ of

$$\big(\underline{\widetilde{W}}_1 \times (X^*(\underline{T}) \cap C_0 + \eta)\big)/\sim,$$

respectively. The natural analogues of Theorems 4.1.2 and 4.1.13 generalize to this setting.

4.2. ***Classification of geometric weight sets.*** Let $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ be a continuous Galois representation. If τ is a 4-generic tame inertial type and $\bar{\rho}$ arises as an $\mathbb{F}$-point of $\mathcal{X}^{\eta,\tau}_{\mathbb{F}}$ then Theorem 3.2.2 attaches a Breuil–Kisin module $\mathfrak{M} \in Y^{[0,2],\tau}(\mathbb{F})$ to $\bar{\rho}$. In this scenario, we define the *shape* $\widetilde{w}^*(\bar{\rho}, \tau) \in \widetilde{W}^{\vee,\mathcal{J}}$ of $\bar{\rho}$ with respect to τ to be the shape of $\mathfrak{M} \in Y^{[0,2],\tau}(\mathbb{F})$ (see Section 2.2). Let $W^g_{2\text{-deep}}(\bar{\rho})$ be the set of Serre weights $\{\sigma \mid \sigma \text{ is 2-deep and } \bar{\rho} \in \mathcal{C}_\sigma(\mathbb{F})\}$.

Definition 4.2.1. Let $\mathrm{SP}(\bar{\rho})$ be the set of pairs $(\bar{\rho}_1, \sigma)$ with $\bar{\rho}_1$ a tame inertial $\mathbb{F}$-type and $\sigma \in W^g_{2\text{-deep}}(\bar{\rho})$ such that there exists a 4-generic tame inertial type τ with

- $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$ and
- $\widetilde{w}^*(\bar{\rho}, \tau) = \widetilde{w}^*(\widetilde{\bar{\rho}}_1, \tau) \in t_{\underline{W}^\vee \eta}$

for any, or equivalently all, extensions $\widetilde{\bar{\rho}}_1 : G_K \to \mathrm{GL}_3(\mathbb{F})$ of $\bar{\rho}_1$ (in particular $\sigma \in W_{\mathrm{obv}}(\bar{\rho}_1)$). Let $W_{\mathrm{obv}}(\bar{\rho}) \subset W^g_{2\text{-deep}}(\bar{\rho})$ be the image of $\mathrm{SP}(\bar{\rho})$ under the projection to $W^g_{2\text{-deep}}(\bar{\rho})$. Let $S(\bar{\rho})$ be the image of $\mathrm{SP}(\bar{\rho})$ under the projection to the set of tame inertial $\mathbb{F}$-types. We call an element of $S(\bar{\rho})$ a *specialization* of $\bar{\rho}$.

Definition 4.2.2. We say that a Galois representation $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ is m-generic if the tame inertial $\mathbb{F}$-type $\bar{\rho}^{\mathrm{ss}}|_{I_K}$ is m-generic and $\bar{\rho}$ has an m-generic specialization.

Remark 4.2.3. (1) If $m \geq 6$, $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ is semisimple, and $\bar{\rho}|_{I_K}$ is m-generic, then $\bar{\rho}$ is m-generic in the sense of Definition 4.2.2 since $\bar{\rho}|_{I_K} \in S(\bar{\rho})$.

(2) It is shown (in greater generality under a suitable genericity assumption) in [Le et al. 2022] that $\bar{\rho}^{\mathrm{ss}}|_{I_K} \in S(\bar{\rho})$ so that the requirement that $\bar{\rho}$ has an m-generic specialization in Definition 4.2.2 is superfluous when m is sufficiently large.

We now recall the setting of Theorem 3.3.2. In particular, we have a pair (σ, ζ) which corresponds to a lowest alcove presentation $(\widetilde{w}, \omega)$ of a Serre weight σ. Given the auxiliary choice of an appropriate tame inertial type τ (and letting (s, μ) be the compatible lowest alcove presentation), we have the diagram

$$\begin{tikzcd}
 & \widetilde{C}^{\zeta}_{\sigma} = \widetilde{C}_{(\widetilde{w},\omega)} \arrow[r] \arrow[d, hook] & C_{(\widetilde{w},\omega)} \arrow[d, hook] \\
 & \widetilde{\mathrm{Fl}}^{[0,2]}_{\mathcal{J}} \cdot s^* t_{\mu^*+\eta^*} \arrow[r] \arrow[d, "T^{\vee,\mathcal{J}}_{\mathbb{F}}"] & \mathrm{Fl}_{\mathcal{J}} \\
\mathcal{C}_{\sigma} \arrow[r, hook] \arrow[d, hook] & {[\widetilde{\mathrm{Fl}}^{[0,2]}_{\mathcal{J}} \cdot s^* t_{\mu^*+\eta^*} / T^{\vee,\mathcal{J}}_{\mathbb{F}}\text{-sh.cnj}]} \arrow[d, hook, "\iota_0"] & \\
(\mathcal{X}_{K,3})_{\mathbb{F}} \arrow[r, hook] & \Phi\text{-}\mathrm{Mod}^{\text{ét},n}_{K,\mathbb{F}} &
\end{tikzcd} \tag{4-4}$$

In particular the composite of the middle column gives a map $\widetilde{C}^{\zeta}_{\sigma} \to \Phi\text{-}\mathrm{Mod}^{\text{ét},n}_{K,\mathbb{F}}$ which does not depend on τ and which factors through $\mathcal{X}_{K,3}$. Note also that $\widetilde{C}^{\zeta}_{\sigma}$ is a subvariety of $(\widetilde{\mathrm{Fl}}^{\nabla_0})^{\mathcal{J}}$, and that the rightmost vertical arrow factors through $(\mathrm{Fl}^{\nabla_0})^{\mathcal{J}}$.

The following proposition relates the Galois theoretic notions in Section 3.3 with the geometric notions in Section 4.1 (or rather, its product version as in Remark 4.1.14):

Proposition 4.2.4. *Let $\bar{\rho} \in \mathcal{X}_{K,3}(\mathbb{F})$ be 6-generic and $(\widetilde{w}, \omega)$ be a lowest alcove presentation of an element of $W^g_{\mathrm{gen}}(\bar{\rho})$ compatible with a 6-generic lowest alcove presentation of $\bar{\rho}^{\mathrm{ss}}$. If $\tilde{x}^* \in (\widetilde{\mathrm{Fl}}^{\nabla_0})^{\mathcal{J}}(\mathbb{F})$ has images $x^* \in (\mathrm{Fl}^{\nabla_0}_{2\text{-deep}})^{\mathcal{J}}(\mathbb{F})$ and $\bar{\rho} \in \mathcal{X}_{K,3}(\mathbb{F})$ then $(\widetilde{w}, \omega)$ is in $W_{\mathrm{obv}}(x^*)$ (resp. $W^g_{2\text{-deep}}(x^*)$) if and only if $F_{(\widetilde{w},\omega)} \in W_{\mathrm{obv}}(\bar{\rho})$ (resp. $W^g_{2\text{-deep}}(\bar{\rho})$).*

Proof. This follows from Theorem 3.3.2, applied to suitably chosen auxiliary 4-generic types τ containing $F_{(\widetilde{w},\omega)}$. ☐

Theorem 4.2.5. *If $\bar{\rho} : G_K \to \mathrm{GL}_3(\mathbb{F})$ is 6-generic, then there exist $\lambda \in X^*(\underline{T})$ and $w \in \underline{W}$ such that $W_{\mathrm{obv}}(\bar{\rho})$ is $F\big(\mathfrak{Tr}_{\lambda}\big(\prod_{j\in\mathcal{J}} w_j \Sigma_{\mathrm{obv},j}(\bar{\rho})\big)\big)$ where for each $j \in \mathcal{J}$, $\Sigma_{\mathrm{obv},j}(\bar{\rho})$ is one of the sets*

(1) $\{(0,0)\}$;

(2) $\{(\varepsilon_1 - \varepsilon_2, 1)\}$;

(3) $\{(0,0), (\varepsilon_1 - \varepsilon_2, 1)\}$;

(4) $\{(0,0), (\varepsilon_1 - \varepsilon_2, 1), (\varepsilon_2 - \varepsilon_1, 1)\}$;

(5) $\{(0,1), (\varepsilon_1, 0), (\varepsilon_2, 0), (\varepsilon_1 + \varepsilon_2, 1)\}$; *and*

(6) $\big\{(0,0), (\varepsilon_1 - \varepsilon_2, 1), (\varepsilon_2 - \varepsilon_1, 1), (\varepsilon_1, 0), (\varepsilon_2, 0), (\varepsilon_1 + \varepsilon_2, 1)\big\}$.

Furthermore, $W^g_{\mathrm{gen}}(\bar{\rho}) = F\big(\mathfrak{Tr}_{\lambda}\big(\prod_{j\in\mathcal{J}} w_j \Sigma^g_j(\bar{\rho})\big)\big)$, where with respect to the six above alternatives for $\Sigma_{\mathrm{obv},j}(\bar{\rho})$, $\Sigma^g_j(\bar{\rho})$ is

(1′) $\{(0,0)\}$ *or* $\{(0,0), (0,1)\}$;

(2′) $\{(\varepsilon_1 - \varepsilon_2, 1)\}$;

(3′) $\{(0,0), (\varepsilon_1 - \varepsilon_2, 1)\}$;

(4′) $\{(0,0), (\varepsilon_1 - \varepsilon_2, 1), (\varepsilon_2 - \varepsilon_1, 1)\}$;

(5′) $\{(0,1), (\varepsilon_1, 0), (\varepsilon_1, 1), (\varepsilon_2, 0), (\varepsilon_2, 1), (\varepsilon_1 + \varepsilon_2, 1)\}$; *or*

(6′) $\big\{(0,0), (0,1), (\varepsilon_1 - \varepsilon_2, 1), (\varepsilon_2 - \varepsilon_1, 1), (\varepsilon_1, 0), (\varepsilon_1, 1), (\varepsilon_2, 0), (\varepsilon_2, 1), (\varepsilon_1 + \varepsilon_2, 1)\big\}$.

Moreover, every possibility arises. If $\bar{\rho}$ is furthermore 8-generic, then $W^g(\bar{\rho}) = F\big(\mathfrak{Tr}_{\lambda}\big(\prod_{j\in\mathcal{J}} w_j \Sigma^g_j(\bar{\rho})\big)\big)$ with Σ^g_j as above.

Proof. This follows from Proposition 4.2.4, Theorem 4.1.13, and Proposition 3.3.8. ☐

Remark 4.2.6. By the proof of Theorem 4.2.5 (and Lemma 4.1.11), if $\bar{\rho}$ is 6-generic and has an m-generic specialization, then every specialization is $(m-4)$-generic.

Theorem 4.2.7. *Let τ and τ' be 4-generic tame inertial types and $\bar{\rho} \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$ and $\bar{\rho}' \in \mathcal{X}^{\eta,\tau'}(\mathbb{F})$. Suppose further that $\tilde{x}^*, \tilde{x}'^* \in (\widetilde{\mathrm{Fl}}^{\nabla_0})^{\mathcal{J}}(\mathbb{F})$ have images $x^*, x'^* \in (\mathrm{Fl}^{\nabla_0}_{2\text{-deep}})^{\mathcal{J}}(\mathbb{F})$ and $\bar{\rho}, \bar{\rho}' \in \mathcal{X}_3(\mathbb{F})$, respectively, such that x'^* is in the closure of the $\widetilde{T}^\vee_{\mathbb{F}}$-orbit of x^*. Then $\bar{\rho}' \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$ and $\widetilde{w}^*(\bar{\rho}', \tau) \leq \widetilde{w}^*(\bar{\rho}, \tau)$. Moreover, $W^g_{\mathrm{gen}}(\bar{\rho}) \subset W^g_{\mathrm{gen}}(\bar{\rho}')$.*

Proof. Since $\bar{\rho} \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$, $x^* \in C_\sigma$ for some $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$ by Theorem 3.3.2. Since C_σ is closed and $\widetilde{T}^\vee_{\mathbb{F}}$-stable, $x'^* \in C_\sigma$. We conclude that $\bar{\rho}' \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$ again using Theorem 3.3.2. Similarly, x'^* is in the closure of $\mathcal{I} \setminus \mathcal{I}\widetilde{w}^*(\bar{\rho}, \tau)\mathcal{I}\widetilde{w}^*(\tau)$ which implies the desired inequality.

If $\bar{\rho} \in \mathcal{C}_\sigma$ for a generic weight σ, then the same argument above shows that $\bar{\rho}' \in \mathcal{C}_\sigma$. □

Corollary 4.2.8. *Let τ and τ' be 4-generic tame inertial types and $\bar{\rho} \in \mathcal{X}^{\eta,\tau}(\mathbb{F}) \cap \mathcal{X}^{\eta,\tau'}(\mathbb{F})$. Suppose that $\bar{\rho}_1 \in \mathcal{X}^{\eta,\tau'}(\mathbb{F})$ such that $\widetilde{w}^*(\bar{\rho}_1, \tau') = \widetilde{w}^*(\bar{\rho}, \tau')$. Then $\bar{\rho}_1 \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$ and $\widetilde{w}^*(\bar{\rho}_1, \tau) \leq \widetilde{w}^*(\bar{\rho}, \tau)$.*

Proof. This follows from Theorem 4.2.7 since there is a contracting $\widetilde{T}^\vee_{\mathbb{F}}$-cocharacter for each translated Schubert cell (see [Le et al. 2023b, Lemma 3.4.7]). □

5. Results for patching functors

We start in Section 5.1 by recalling the formalism of weak (minimal) patching functors and we prove abstract versions of Serre weight conjectures assuming the modularity of an obvious weight (see Propositions 5.1.10 and 5.1.11 below). This assumption is removed in Section 5.2 if the weak patching functor comes from an *arithmetic module*. In Section 5.3, we prove results on cyclicity of patching functors arising from arithmetic modules and we finally give global applications of the above results in Section 5.4.

5.1. ***Patching functors and Serre weights.*** We recall the setup and the basic definitions for weak minimal patching functors. Recall from Section 1.4 that we write the finite étale $\mathbb{Z}_p$-algebra $\mathcal{O}_p$ as the product $\prod_{v \in S_p} \mathcal{O}_v$, where S_p is a finite set and for each $v \in S_p$, $\mathcal{O}_v$ is the ring of integers in a finite unramified extension F_v^+ of $\mathbb{Q}_p$, and that ${}^L\underline{G}$ denotes the Langlands dual group of $G_0 \stackrel{\mathrm{def}}{=} \mathrm{Res}_{\mathcal{O}_p/\mathbb{Z}_p}(\mathrm{GL}_{3/\mathcal{O}_p})$. Following Section 2.1.I, an $\mathbb{F}$-valued L-homomorphism $\bar{\rho} : G_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ (resp. a tame inertial L-parameter $\tau : I_{\mathbb{Q}_p} \to \underline{G}^\vee(E)$) is identified with a collection $(\bar{\rho}_v)_{v \in S_p}$ of continuous homomorphisms $\bar{\rho}_v : G_{F_v^+} \to \mathrm{GL}_3(\mathbb{F})$ (resp. with a collection $(\tau_v)_{v \in S_p}$ of tame inertial types $\tau_v : I_{F_v^+} \to \mathrm{GL}_3(E)$).

Let $\bar{\rho}$ be an L-homomorphism over $\mathbb{F}$ with corresponding collection $(\bar{\rho}_v)_{v \in S_p}$. We write R_∞ for the $\mathcal{O}$-algebra $R_{\bar{\rho}} \widehat{\otimes}_{\mathcal{O}} R^p$, where

$$R_{\bar{\rho}} \stackrel{\mathrm{def}}{=} \widehat{\bigotimes_{v \in S_p, \mathcal{O}}} R^{\square}_{\bar{\rho}_v}$$

and R^p is a (nonzero) complete local Noetherian equidimensional flat $\mathcal{O}$-algebra with residue field $\mathbb{F}$ such that each irreducible component of Spec R^p and of Spec $\bar{R}^p$ is geometrically irreducible (we remind the reader that $\bar{M}$ denotes $M \otimes_{\mathcal{O}} \mathbb{F}$ for any $\mathcal{O}$-module M). We suppress the dependence on R^p below. For a Weil–Deligne inertial L-parameter τ, let $R_\infty(\tau)$ be $R_\infty \otimes_{R_{\bar{\rho}}} R^{\eta,\tau}_{\bar{\rho}}$, where

$$R^{\eta,\tau}_{\bar{\rho}} \stackrel{\mathrm{def}}{=} \widehat{\bigotimes_{v \in S_p, \mathcal{O}}} R^{\eta_v,\tau_v}_{\bar{\rho}_v}.$$

Let X_∞, $X_\infty(\tau)$, and $\overline{X}_\infty(\tau)$ be Spec R_∞, Spec $R_\infty(\tau)$, and Spec $\overline{R}_\infty(\tau)$ respectively. Let $\mathrm{Mod}(X_\infty)$ be the category of coherent sheaves over X_∞, and let $\mathrm{Rep}_{\mathcal{O}}(\mathrm{GL}_3(\mathcal{O}_p))$ denote the category of topological $\mathcal{O}[\mathrm{GL}_3(\mathcal{O}_p)]$-modules which are finitely generated over $\mathcal{O}$.

Recall from Section 2.1.C that given a tame inertial L-parameter τ we have an irreducible smooth E-representation $\sigma(\tau)$ attached to it. If $\sigma^\circ(\tau) \subseteq \sigma(\tau)$ is an $\mathcal{O}$-lattice, we write $\bar{\sigma}^\circ(\tau)$ for $\sigma^\circ(\tau) \otimes_{\mathcal{O}} \mathbb{F}$ in what follows.

Definition 5.1.1. A *weak patching functor* for an L-homomorphism $\bar{\rho} : W_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ is a nonzero covariant exact functor $M_\infty : \mathrm{Rep}_{\mathcal{O}}(\mathrm{GL}_3(\mathcal{O}_p)) \to \mathrm{Mod}(X_\infty)$ satisfying the following: if τ is an inertial L-parameter and $\sigma^\circ(\tau)$ is an $\mathcal{O}$-lattice in $\sigma(\tau)$ then:

(1) $M_\infty(\sigma^\circ(\tau))$ is either zero or a maximal Cohen–Macaulay sheaf on $X_\infty(\tau)$.

(2) For all $\sigma \in \mathrm{JH}(\bar{\sigma}^\circ(\tau))$, $M_\infty(\sigma)$ is a maximal Cohen–Macaulay sheaf on $\overline{X}_\infty(\tau)$ (or is 0).

(3) Suppose σ° is an $\mathcal{O}$-lattice in a principal series representation $R_1(\mu)$. Then $M_\infty(\sigma^\circ)$ is supported on the potentially *semistable* locus of type $(\eta, \tau(1, \mu))$ in X_∞.

We say that a weak patching functor M_∞ is *minimal* if R^p is formally smooth over $\mathcal{O}$ and whenever τ is an inertial L-parameter, $M_\infty(\sigma^\circ(\tau))[p^{-1}]$, which is locally free over (the regular scheme) Spec $R_\infty(\tau)[p^{-1}]$, has rank at most one on each connected component.

Remark 5.1.2. The above definition of weak patching functor is slightly weaker than that in [Le et al. 2023b, Definition 6.2.1] and closer in spirit to that of [Le et al. 2019, Definition 4.2.1]: the purpose of the third item is to eliminate nonregular Serre weights.

Let d be the (common) dimension of $\overline{X}_\infty(\tau)$ for any inertial L-parameter τ. If M is an $\overline{R}_\infty$-module whose action factors through $\overline{R}_\infty(\tau)$ for some inertial L-parameter τ, let $Z(M)$ be the associated d-dimensional cycle. Note that $Z(M_\infty(-))$ is additive in exact sequences.

We now fix an L-homomorphism $\bar{\rho} : W_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ and a weak patching functor M_∞. Let $W(\bar{\rho})$ be the set of Serre weights σ such that $M_\infty(\sigma) \neq 0$.

Proposition 5.1.3. *If $\bar{\rho} : W_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ is an L-homomorphism with 6-generic semisimplification, then $W_{\mathrm{gen}}(\bar{\rho}) = W(\bar{\rho})$.*

Proof. We adapt the argument in Proposition 3.3.8: If $F(\lambda) \in W(\bar{\rho})$, then $M_\infty(F(\lambda)) \neq 0$ so that $M_\infty(R_1(\lambda)) \neq 0$. This implies that $\bar{\rho}$, and hence $\bar{\rho}^{\mathrm{ss}}$, has a potentially semistable lift of type $(\eta, \tau(1, \lambda))$. The rest of the argument is the same. □

Proposition 5.1.4. (1) *If $\bar{\rho}^{\mathrm{ss}}|_{I_K}$ is 7-generic, then $W(\bar{\rho}) \subset W^?(\bar{\rho}^{\mathrm{ss}}|_{I_K})$.*

(2) *If $\bar{\rho}^{\mathrm{ss}}|_{I_K}$ is 7-generic, then for any 4-generic $\bar{\rho}_1 \in S(\bar{\rho})$, $W(\bar{\rho}) \subset W^?(\bar{\rho}_1)$. (Note that $S(\bar{\rho})$ consists of tame inertial $\mathbb{F}$-types so that $W^?(\bar{\rho}_1)$ is defined.)*

Proof. Let σ be in $W(\bar{\rho})$. By the proof of Propositions 5.1.3 and 3.3.8, $R_1(\lambda)$ is 5-generic if $\sigma = F(\lambda)$ so that σ is 3-deep. Let $\bar{\rho}_1$ be a 4-generic element of $S(\bar{\rho})$ or $\bar{\rho}^{\mathrm{ss}}|_{I_K}$. Suppose that $\sigma \notin W^?(\bar{\rho}_1)$. By (the proof of) [Le et al. 2020, Lemma 2.3.13] and Proposition 2.1.1, we can find a 1-generic type τ such that $\sigma \in \mathrm{JH}(\sigma(\tau))$ and $\mathrm{JH}(\sigma(\tau)) \cap W^?(\bar{\rho}_1) = \varnothing$. That $\sigma \in W(\bar{\rho})$ implies that $\bar{\rho}$, and thus $\bar{\rho}^{\mathrm{ss}}$, has a potentially crystalline lift of type τ as in the proof of Proposition 5.1.3. Proposition 3.3.2 in [Le et al. 2019] implies that τ is 4-generic. Lemma 5 in [Enns 2019] or Corollary 4.2.8 then implies that $\widetilde{\bar{\rho}}_1 \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$ for any extension $\widetilde{\bar{\rho}}_1$ of $\bar{\rho}_1$ to an L-homomorphism. Theorem 3.3.2 implies that $\mathrm{JH}(\bar{\sigma}(\tau)) \cap W^g_{\mathrm{gen}}(\widetilde{\bar{\rho}}_1) \neq \varnothing$. On the other hand,

$$\mathrm{JH}(\bar{\sigma}(\tau)) \cap W^g_{\mathrm{gen}}(\widetilde{\bar{\rho}}_1) = \mathrm{JH}(\bar{\sigma}(\tau)) \cap W^?_{\mathrm{gen}}(\bar{\rho}_1) = \mathrm{JH}(\bar{\sigma}(\tau)) \cap W^?(\bar{\rho}_1) = \varnothing$$

by Corollary 3.3.7. This is a contradiction. □

For a Serre weight σ, let $\mathfrak{p}(\sigma)$ be the prime ideal or unit ideal in $R_{\bar{\rho}}$ corresponding to the pullback of the stack $\mathcal{C}_\sigma$ to $\operatorname{Spec} R_{\bar{\rho}}$. For an inertial L-parameter τ, let $I(\tau)$ be the kernel of the surjection $R_{\bar{\rho}} \twoheadrightarrow R^{\eta,\tau}_{\bar{\rho}}$. Observe that if $I(\tau) \subset \mathfrak{p}(\sigma) \neq 1$, then $\mathfrak{p}(\sigma)$ induces a minimal prime of $\bar{R}^{\eta,\tau}_{\bar{\rho}}$, and all minimal primes arise this way.

Lemma 5.1.5. *Suppose that τ is an inertial L-parameter corresponding to a collection of 4-generic tame inertial types $(\tau_v)_{v \in S_p}$. Then any minimal prime ideal of $R_\infty(\tau)$ is of the form $I(\tau)R_\infty + \mathfrak{p}R_\infty$ for some minimal prime ideal $\mathfrak{p} \subset R^p$.*

If M is a nonzero finitely generated maximal Cohen–Macaulay $R_\infty(\tau)$-module, then $\operatorname{Ann}_{R_{\bar{\rho}}}(\bar{M}) = I(\tau) + (\varpi)$.

Proof. Since $R^{\eta,\tau}_{\bar{\rho}}$ is geometrically irreducible (its special fiber is reduced after arbitrary finite extension of $\mathbb{F}$ and hence is normal; see the proof of [Le et al. 2020, Lemma 3.5.4]), the first part follows from [Barnet-Lamb et al. 2011, Lemma 3.3(5)]. Similarly, any minimal prime of $\bar{R}_\infty(\tau)$ is of the form $\mathfrak{p}(\sigma)\bar{R}_\infty + \bar{\mathfrak{p}}\bar{R}_\infty$, where $\mathfrak{p}(\sigma)$ corresponds to a minimal prime of $\bar{R}^{\eta,\tau}_{\bar{\rho}}$, and $\bar{\mathfrak{p}}$ is a minimal prime of $\bar{R}^p$.

If M is a nonzero finitely generated maximal Cohen–Macaulay $R_\infty(\tau)$-module, then $Z(\bar{M})$ is at least the reduction of the cycle in $\operatorname{Spec} R_\infty(\tau)[1/p]$ corresponding to a minimal prime of $R_\infty(\tau)$. In particular, for any prime $\mathfrak{p}(\sigma)$ of $R_{\bar{\rho}}$ inducing a minimal prime of $\bar{R}^{\eta,\tau}_{\bar{\rho}}$, $\operatorname{Ann}_{R_\infty(\tau)}(\bar{M})$ is contained in a prime induced by $\mathfrak{p}(\sigma)\bar{R}_\infty + \bar{\mathfrak{p}}\bar{R}_\infty$ for some minimal prime $\bar{\mathfrak{p}}$ of $\bar{R}^p$. Since $R_\infty/(\mathfrak{p}(\sigma)R_\infty + \bar{\mathfrak{p}}R_\infty) \cong R_{\bar{\rho}}/\mathfrak{p}(\sigma) \mathbin{\widehat{\otimes}} R^p/\bar{\mathfrak{p}}$, $(\mathfrak{p}(\sigma)R_\infty + \bar{\mathfrak{p}}R_\infty) \cap R_{\bar{\rho}} = \mathfrak{p}(\sigma)$ by Lemma 5.1.6. We conclude that $\operatorname{Ann}_{R_{\bar{\rho}}}(\bar{M}) \subset \mathfrak{p}(\sigma)$ for each minimal prime ideal $\mathfrak{p}(\sigma)\bar{R}_\infty(\tau)$ of $\bar{R}_\infty(\tau)$. Since $\bar{R}_\infty(\tau)$ is reduced, $\operatorname{Ann}_{R_{\bar{\rho}}}(\bar{M}) \subset I(\tau) + (\varpi)$. The reverse inclusion is clear. □

Lemma 5.1.6. *Let $\mathbb{F}$ be a field. If R and S are complete Noetherian local $\mathbb{F}$-algebras with residue field $\mathbb{F}$, then the natural map $R \to R \mathbin{\widehat{\otimes}}_{\mathbb{F}} S$, $r \mapsto r \mathbin{\widehat{\otimes}} 1$ is an injection.*

Proof. Let $\mathfrak{m}_S \subset S$ be the maximal ideal. The composition $R \to R \mathbin{\widehat{\otimes}}_{\mathbb{F}} S \to R \mathbin{\widehat{\otimes}}_{\mathbb{F}} (S/\mathfrak{m}_S) \cong R \otimes_{\mathbb{F}} \mathbb{F}$ is the isomorphism given by $r \mapsto r \otimes 1$. The result follows. □

For the rest of the section, we assume that $\bar{\rho}$ is 8-generic. In particular, every element of $S(\bar{\rho})$ is 4-generic by Remark 4.2.6.

Lemma 5.1.7. *If* σ_1 *is a Serre weight and* $\mathrm{Ann}_{R_{\bar\rho}} M_\infty(\sigma_1) \subset \mathfrak{p}(\sigma_2) \subsetneq R_{\bar\rho}$ *for a Serre weight* σ_2, *then* $\sigma_2 \uparrow \sigma_1$.

Proof. Since $M_\infty(\sigma_1)$ is nonzero by assumption, Proposition 5.1.4 implies that $\sigma_1 \in W^?(\bar\rho_1)$ for any specialization $\bar\rho_1$ of $\bar\rho$. Then σ_1 is 6-deep because $\bar\rho$ is 8-generic. If $\sigma_1 \in \mathrm{JH}(\bar\sigma(\tau))$ for a tame inertial type τ (necessarily 4-generic), then

$$\bigcap_{\sigma\in\mathrm{JH}(\bar\sigma(\tau))} \mathfrak{p}(\sigma) = (\varpi) + I(\tau) \subset \mathrm{Ann}_{R_{\bar\rho}} M_\infty(\sigma_1) \subset \mathfrak{p}(\sigma_2) \subsetneq R_{\bar\rho}$$

by Theorem 3.3.2. This implies that $\sigma_2 \in \mathrm{JH}(\bar\sigma(\tau))$. We conclude that σ_1 covers σ_2 (Definition 2.1.2). The result now follows from Lemma 2.1.3. □

Lemma 5.1.8. *If* $W_{\mathrm{obv}}(\bar\rho) \cap W(\bar\rho)$ *is nonempty*, *then* $W_{\mathrm{obv}}(\bar\rho) \subset W(\bar\rho)$.

Proof. Let $\sigma_0 \in W_{\mathrm{obv}}(\bar\rho)$. We claim that there is an $n \in \mathbb{N}$ and sequences of tame inertial types $(\tau_i)_{i=1}^n$, specializations $(\bar\rho_i)_{i=1}^n$ (elements in $S(\bar\rho)$), and (not necessarily distinct) Serre weights $(\sigma_i)_{i=1}^n$ such that

- $\{\sigma_i\}_{i=0}^n = W_{\mathrm{obv}}(\bar\rho)$;
- $\bar\rho_i \in S(\bar\rho)$ for all $i = 1, \dots, n$;
- $W^?(\bar\rho_i) \cap \mathrm{JH}(\bar\sigma(\tau_i)) = W_{\mathrm{obv}}(\bar\rho_i) \cap \mathrm{JH}(\bar\sigma(\tau_i)) = \{\sigma_{i-1}, \sigma_i\}$ for all $i = 1, \dots, n$.

Indeed, the proof of Corollary 4.1.10 gives a sequence of elements $\big(y, (\widetilde{w}, y\widetilde{w}^{-1}(0))\big)$ in $\mathrm{SP}(x^*)$. We define the sequences by taking the specializations $\bar\rho_i$ corresponding to the elements y, taking the Serre weights σ_i corresponding to the elements $F_{(\widetilde{w}, y\widetilde{w}^{-1}(0))}$, and taking the tame inertial types τ_i to be $\tau\big(u, y(\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0 s\widetilde{w})^{-1}(0)\big)$ where u is the image of $y(\widetilde{w}^{-1}\widetilde{w}_h^{-1}w_0 s\widetilde{w})^{-1}$ in W (see also the proof of Proposition 4.1.8). We will use these sequences to prove the result by induction.

Suppose that $\sigma_{i-1} \in W(\bar\rho)$ for some $1 \le i \le n$. Then $M_\infty(\bar\sigma^\circ(\tau_i))$ is nonzero. Since $M_\infty(\bar\sigma^\circ(\tau_i))$ is a nonzero finitely generated maximal Cohen–Macaulay $\bar R_\infty(\tau_i)$-module, Proposition 5.1.4 and Lemma 5.1.5 give

$$\prod_{\sigma'\in\mathrm{JH}(\bar\sigma(\tau_i))\cap W^?(\bar\rho_i)} \mathrm{Ann}_{R_{\bar\rho}} M_\infty(\sigma') \subset \mathrm{Ann}_{R_{\bar\rho}} M_\infty(\bar\sigma^\circ(\tau_i)) = I(\tau_i) + (\varpi) \subset \mathfrak{p}(\sigma_i).$$

Then $\mathrm{Ann}_{R_{\bar\rho}} M_\infty(\sigma_{i-1}) \subset \mathfrak{p}(\sigma_i)$ or $\mathrm{Ann}_{R_{\bar\rho}} M_\infty(\sigma_i) \subset \mathfrak{p}(\sigma_i)$. The former contradicts Lemma 5.1.7 and so $\sigma_i \in W(\bar\rho)$. □

Lemma 5.1.9. *If* $\bar\rho$ *is semisimple and* 8-*generic*, *and* $\sigma \in W^?(\bar\rho)$, *then there exists a tame inertial* L-*homomorphism* τ *such that*

(1) $\sigma \in \mathrm{JH}(\bar\sigma(\tau))$; *and*

(2) $\sigma' \in W^?(\bar\rho) \cap \mathrm{JH}(\bar\sigma(\tau))$ *implies that* $\sigma' \uparrow \sigma$.

Proof. This follows from [Le et al. 2020, Lemma 3.5.9]. □

In fact, τ is unique. We say that τ is minimal with respect to $\bar\rho$ and σ.

Proposition 5.1.10. *Let $\bar{\rho} : W_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ be an 8-generic L-homomorphism and let M_∞ be a weak patching functor. If $W_{\mathrm{obv}}(\bar{\rho}) \cap W(\bar{\rho})$ is nonempty, then $\mathrm{Ann}_{R_{\bar{\rho}}} M_\infty(\sigma) \subset \mathfrak{p}(\sigma)$ for all Serre weights σ. In particular, $W^g(\bar{\rho}) \subset W(\bar{\rho})$.*

Proof. The inclusion is trivial if $\sigma \notin W^g(\bar{\rho})$. Suppose that $\sigma \in W^g(\bar{\rho})$. Choose an 8-generic $\bar{\rho}' \in S(\bar{\rho})$ (e.g., $\bar{\rho}^{\mathrm{ss}}$) and choose the tame inertial type τ which is minimal with respect to $\bar{\rho}'$ and σ.

For any Serre weight $\sigma' \uparrow \sigma$, $\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau))$. Theorem 4.2.5 implies that $\mathrm{JH}(\bar{\sigma}(\tau)) \cap W_{\mathrm{obv}}(\bar{\rho})$ is nonempty. Lemma 5.1.8 implies that $M_\infty(\bar{\sigma}^\circ(\tau))$ is nonzero for any lattice $\sigma^\circ(\tau) \subset \sigma(\tau)$. Since $M_\infty(\bar{\sigma}^\circ(\tau))$ is a nonzero finitely generated maximal Cohen–Macaulay $\bar{R}_\infty(\tau)$-module, Proposition 5.1.4 and Lemma 5.1.5 give

$$\prod_{\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau)) \cap W^?(\bar{\rho}')} \mathrm{Ann}_{R_{\bar{\rho}}} M_\infty(\sigma') \subset \mathrm{Ann}_{R_{\bar{\rho}}} M_\infty(\bar{\sigma}^\circ(\tau)) = I(\tau) + (\varpi) \subset \mathfrak{p}(\sigma).$$

Then $\mathrm{Ann}_{R_{\bar{\rho}}} M_\infty(\sigma') \subset \mathfrak{p}(\sigma)$ for some $\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau)) \cap W^?(\bar{\rho}')$. Lemma 5.1.7 implies that $\sigma \uparrow \sigma'$. That τ is minimal with respect to $\bar{\rho}'$ and σ implies that $\sigma' \uparrow \sigma$, hence $\sigma = \sigma'$. □

Proposition 5.1.11. *Let $\bar{\rho} : W_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ be an 8-generic L-homomorphism and let M_∞ be a weak minimal patching functor. Assume that $W_{\mathrm{obv}}(\bar{\rho}) \cap W(\bar{\rho})$ is nonempty. Then $Z(M_\infty(\sigma))$ is the irreducible or zero cycle corresponding to the prime or unit ideal $\mathfrak{p}(\sigma) R_\infty$. In particular, $W(\bar{\rho}) = W^g(\bar{\rho})$.*

Proof. Let τ be a 4-generic tame inertial type. Let $\mathcal{C}_\sigma(\bar{\rho})$ be the irreducible or zero cycle corresponding to the ideal $\mathfrak{p}(\sigma) R_\infty$. Then

$$Z(\bar{R}_\infty(\tau)) \geq Z\big(M_\infty(\overline{\sigma^\circ(\tau)})\big) = \sum_{\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))} Z(M_\infty(\sigma)) \geq \sum_{\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))} \mathcal{C}_\sigma(\bar{\rho}),$$

where the first inequality follows from the fact that M_∞ is minimal (see [Le et al. 2018a, Proposition 7.14]) and the second inequality follows from Proposition 5.1.10. However the first and last expression are equal by Theorem 3.3.2, which forces the inequalities to be equalities. We conclude that the result holds for all $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$ for a 4-generic tame inertial type τ. In particular, the result holds for all generic σ.

Finally, suppose σ is nongeneric. Then Proposition 5.1.3 shows that $Z(M_\infty(\sigma)) = 0$ and $\sigma \notin W^g(\bar{\rho})$, by Proposition 3.3.8, so $\mathfrak{p}(\sigma) R_\infty = R_\infty$. □

5.2. *Arithmetic patched modules.* Let R_∞ be as in Section 5.1 and set $F_p \overset{\mathrm{def}}{=} \mathcal{O}_p \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.

Definition 5.2.1. An *arithmetic $R_\infty[\mathrm{GL}_3(F_p)]$-module* for an L-homomorphism $\bar{\rho} : W_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ is a nonzero $\mathcal{O}$-module M_∞ with commuting actions of R_∞ and $\mathrm{GL}_3(F_p)$ satisfying the following axioms:

(1) The $R_\infty[\mathrm{GL}_3(\mathcal{O}_p)]$-action on M_∞ extends to $R_\infty[\![\mathrm{GL}_3(\mathcal{O}_p)]\!]$ making M_∞ a finitely generated $R_\infty[\![\mathrm{GL}_3(\mathcal{O}_p)]\!]$-module.

(2) The functor $\mathrm{Hom}_{\mathcal{O}[\![\mathrm{GL}_3(\mathcal{O}_p)]\!]}(-, M_\infty^\vee)^\vee : \mathrm{Rep}_{\mathcal{O}}(\mathrm{GL}_3(\mathcal{O}_p)) \to \mathrm{Mod}(X_\infty)$, denoted $M_\infty(-)$, is a weak patching functor.

(3) The action of $\mathcal{H}^{\mathrm{GL}_3(F_p)}_{\mathrm{GL}_3(\mathcal{O}_p)}(\sigma(\tau)) \cong \mathcal{H}^{\mathrm{GL}_3(F_p)}_{\mathrm{GL}_3(\mathcal{O}_p)}(\sigma(\tau)^\circ)[1/p]$ on $M_\infty(\sigma(\tau)^\circ)[1/p]$ factors through the composite

$$\mathcal{H}^{\mathrm{GL}_3(F_p)}_{\mathrm{GL}_3(\mathcal{O}_p)}(\sigma(\tau)) \xrightarrow{\eta_\infty} R^{\eta,\tau}_{\bar{\rho}}[1/p] \to R_\infty(\tau)[1/p],$$

where the map η_∞ is the map denoted by η in [Caraiani et al. 2016, Theorem 4.1] except with r_p normalized so that $r_p(\pi) = \mathrm{rec}_p(\pi \otimes |\det|^{(n-1)/2})$.

We say that an arithmetic $R_\infty[\mathrm{GL}_3(F_p)]$-module M_∞ is *minimal* if $M_\infty(-)$ is.

Let I be the preimage of $B_0(\mathbb{F}_p)$ under the reduction map $G_0(\mathbb{Z}_p) \to G_0(\mathbb{F}_p)$. Let I_1 be the (unique) pro-p Sylow subgroup of I. Let $\chi : I/I_1 \to \mathcal{O}^\times$ be a character. Let $\theta(\chi)$ be $\mathrm{ind}_I^{G(\mathcal{O}_p)} \chi$. If χ is regular, i.e., $\chi = \chi^s$ implies $s = 1$ for $s \in W(\mathrm{GL}_3^{S_p})$, then $\theta(\chi)[1/p]$ is absolutely irreducible.

Lemma 5.2.2. *If $\chi : I/I_1 \to \mathcal{O}^\times$ is a regular character, then the $G_0(\mathbb{Z}_p)$-cosocle of $\theta(\chi)$ is isomorphic to the unique Serre weight $\sigma(\chi)$ with $\sigma(\chi)^{I_1} \cong \bar{\chi}$.*

Proof. By Frobenius reciprocity, $\mathrm{Hom}_{G_0(\mathbb{Z}_p)}(\theta(\chi), \sigma) \cong \mathrm{Hom}_I(\chi, \sigma) \cong \mathrm{Hom}_I(\bar{\chi}, \sigma^{I_1})$. Then

$$\mathrm{Hom}_{G(\mathbb{Z}_p)}(\theta(\chi), \sigma) \neq 0$$

if and only if $\sigma = \sigma(\chi)$ in which case it is one-dimensional. □

Let $s \in W(\mathrm{GL}_3^{S_p})$ and χ^s be the character such that $\chi^s(t) = \chi(s^{-1}ts)$ for

$$t \in T_0(\mathbb{Z}_p) \cong T(\mathcal{O}_p) \cong \prod_{v \in S_p} T(\mathcal{O}_v).$$

The representations $\theta(\chi)$ and $\theta(\chi^s)$ are isomorphic.

Lemma 5.2.3. *Let $\chi : I/I_1 \to \mathcal{O}^\times$ be a regular character. Fix $s \in W(\mathrm{GL}_3^{S_p})$, and let $\theta(\chi) \to \overline{\theta(\chi^s)}$ be a nonzero map which is unique up to scalar. Let $I(\chi, s)$ be the image of this map. Then $\sigma(\chi) \in \mathrm{JH}(I(\chi, s))$.*

Proof. The natural surjection $\theta(\chi) \twoheadrightarrow I(\chi, s)$ induces a surjection on $G_0(\mathbb{Z}_p)$-cosocles by Lemma 5.2.2. Thus the cosocle of $I(\chi, s)$ is isomorphic to $\sigma(\chi)$. □

Let $\chi = \bigotimes_{v \in S_p} \chi_v$ be as above and decompose each χ_v as $\chi_{v,1} \otimes \chi_{v,2} \otimes \chi_{v,3}$ in the usual way. For each $v \in S_p$, let τ_v (resp. $\tau_{v,1}$) be the tame inertial type $(\chi_{v,1} \oplus \chi_{v,2} \oplus \chi_{v,3}) \circ \mathrm{Art}_{F_v}^{-1}$ (resp. $\chi_{v,1} \circ \mathrm{Art}_{F_v}^{-1}$). Then letting $\tau \overset{\mathrm{def}}{=} (\tau_v)_{v \in S_p}$ (resp. $\tau_1 \overset{\mathrm{def}}{=} (\tau_{v,1})_{v \in S_p}$) we have that $\sigma(\tau) = \theta(\chi)[1/p]$ (resp. $\sigma(\tau_1)$ is the inflation of χ_1 to $\mathcal{O}_p^\times$). For each $v \in S_p$, let $U_{\tau_v}^{\tau_{1,v}}$ be the endomorphism defined in [Le et al. 2023a, §10.1.2] so that $U_\tau^{\tau_1} \overset{\mathrm{def}}{=} \prod_{v \in S_p} U_{\tau_v}^{\tau_{1,v}}$ is an endomorphism of $\mathrm{ind}_{\mathrm{GL}_3(\mathcal{O}_p)}^{\mathrm{GL}_3(F_p)} \theta(\chi)$.

Lemma 5.2.4. *With χ and τ as above, suppose that τ is 4-generic. If $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$ is not an outer weight, then $\eta_\infty(U_{\tau_1,\tau})$ vanishes on $\mathcal{C}_\sigma(\bar{\rho})$.*

Proof. Up to a unit, for each $v \in S_p$, the image of $\eta_\infty(U_{\tau_{v,1},\tau_v}) \pmod{\varpi} \in \mathbb{F}$ in (the completion of) the second column of Table 1 is a nonempty product of diagonal elements modulo v by [Dotto and Le 2021, Corollary 3.7]. One can check that each of these diagonal elements modulo v is contained in each of the ideals in the final column corresponding to $(0, 0)$, $(\varepsilon_1, 0)$, or $(\varepsilon_2, 0)$. □

Lemma 5.2.5. *Let $\bar{\rho}$ be 8-generic, and let M_∞ be an arithmetic $R_\infty[\mathrm{GL}_3(F_p)]$-module. Then*

$$\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma) \subset \mathcal{C}_\sigma(\bar{\rho}).$$

In particular, $W(\bar{\rho}) \subset W^g(\bar{\rho})$.

Proof. Let σ' be a Serre weight such that $\mathcal{C}_{\sigma'}(\bar{\rho}) \subset \mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma)$. We will show that $\sigma' = \sigma$. Set χ to be the Teichmüller lift of σ^{I_1}, i.e., $\sigma = \sigma(\chi)$. Let τ and τ_1 be defined in terms of χ as before, e.g., $\sigma(\tau) \cong \theta(\chi)[1/p]$. Since σ covers σ' by Lemma 5.1.7, $\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau))$. Since the only weight in $\mathrm{JH}_{\mathrm{out}}(\bar{\sigma}(\tau))$ that σ covers is σ itself, we conclude that $\sigma' = \sigma$ or σ' is not in $\mathrm{JH}_{\mathrm{out}}(\bar{\sigma}(\tau))$.

We have

$$\begin{aligned}\mathcal{C}_{\sigma'}(\bar{\rho}) \subset \mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma) \subset \mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(I(\chi, s))& \\ = \mathrm{Supp}_{R_{\bar{\rho}}} \eta_\infty(U_{\tau_1,\tau}) M_\infty(\theta(\chi)/(\varpi)) &\subset \mathrm{Supp}_{R_{\bar{\rho}}} \eta_\infty(U_{\tau_1,\tau}) \bar{R}^\tau_{\bar{\rho}},\end{aligned}$$

where the second inclusion follows from Lemma 5.2.3 and the equality follows from the fact that $M_\infty(\theta(\chi))/(\eta_\infty(U_{\tau_1,\tau}), \varpi) \cong M_\infty((\theta(\chi^s) \otimes_{\mathcal{O}} \mathbb{F})/I(\chi, s))$, where $s_v = (132)$ for all $v \in S_p$ by [Le et al. 2023a, (10.1.9)] (and using the exactness of $M_\infty(-)$). Then $(\eta_\infty(U_{\tau_1,\tau}), \varpi)_{\mathfrak{p}(\sigma')}/(\varpi)_{\mathfrak{p}(\sigma')} \neq 0$. Since $\bar{R}^\tau_{\bar{\rho}}$ is reduced, Lemma 5.2.4 implies that $\sigma' \in \mathrm{JH}_{\mathrm{out}}(\bar{\sigma}(\tau))$. □

Theorem 5.2.6. *Let $\bar{\rho}$ be 8-generic and M_∞ be an arithmetic $R_\infty[\mathrm{GL}_3(F_p)]$-module. For a Serre weight σ, $\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma) = \mathcal{C}_\sigma(\bar{\rho})$. In particular, $W(\bar{\rho}) = W^g(\bar{\rho})$. If M_∞ is furthermore minimal, then $Z(M_\infty(\sigma))$ is the irreducible or zero cycle corresponding to the prime or unit ideal $\mathfrak{p}(\sigma)R_\infty$.*

Proof. If σ is nongeneric, then $\sigma \notin W(\bar{\rho})$ and $\sigma \notin W^g(\bar{\rho})$ as in the proof of Proposition 5.1.11 and the desired equality holds. Since M_∞ is nonzero, there is a generic $\sigma \in W(\bar{\rho})$. Choose a 4-generic tame inertial type τ such that $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$. Then $M_\infty(\sigma(\tau))$ is nonzero and in fact

$$\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\bar{\sigma}(\tau)) = \bigcup_{\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau))} \mathcal{C}_{\sigma'}(\bar{\rho})$$

by Theorem 3.3.2. On the other hand, we have

$$\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\bar{\sigma}(\tau)) = \bigcup_{\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau))} \mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma').$$

Then Lemma 5.2.5 implies that $\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma') = \mathcal{C}_{\sigma'}(\bar{\rho})$ for all $\sigma' \in \mathrm{JH}(\bar{\sigma}(\tau))$; in particular,

$$\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma) = \mathcal{C}_\sigma(\bar{\rho}).$$

It is easy to see from Section 2.1.G and Theorem 4.2.5 that $W_{\mathrm{obv}}(\bar{\rho}) \cap \mathrm{JH}(\bar{\sigma}(\tau))$ is nonempty. Combined with the above, $W_{\mathrm{obv}}(\bar{\rho}) \cap W(\bar{\rho})$ is nonempty. By Proposition 5.1.10, $W^g(\bar{\rho}) \subset W(\bar{\rho})$. With Lemma 5.2.5, we have $W^g(\bar{\rho}) = W(\bar{\rho})$. By the above parenthetical, $\mathrm{Supp}_{R_{\bar{\rho}}} M_\infty(\sigma) = \mathcal{C}_\sigma(\bar{\rho})$ if $\sigma \in W(\bar{\rho})$ while it holds trivially otherwise.

If M_∞ is minimal, then the last part now follows from Proposition 5.1.11. □

5.3. ***Cyclicity for patching functors.*** In this section, we show that certain patched modules for tame types are locally free of rank one over the corresponding local deformation space. The argument follows closely that of [Le et al. 2020, §5.2].

Recall from Section 2.1.C the irreducible smooth E-representation $\sigma(\tau)$ attached to a tame inertial L-parameter τ. Given $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$ we write $\sigma(\tau)^{\sigma}$ for an $\mathcal{O}$-lattice, unique up to homothety, in $\sigma(\tau)$ with cosocle σ. For an L-parameter $\bar{\rho} : G_{\mathbb{Q}_p} \to {}^{L}\underline{G}(\mathbb{F})$, we write $W^{g}(\bar{\rho}, \tau)$ for the intersection $W^{g}(\bar{\rho}) \cap \mathrm{JH}(\bar{\sigma}(\tau))$. Throughout this section, we fix an L-parameter $\bar{\rho}$ and a weak minimal patching functor M_∞ for $\bar{\rho}$ which comes from an arithmetic $R_\infty[\mathrm{GL}_3(F_p)]$-module. The main result of this section is the following:

Theorem 5.3.1. *Suppose that $\bar{\rho} : G_{\mathbb{Q}_p} \to {}^{L}\underline{G}(\mathbb{F})$ is a 11-generic L-parameter arising from an $\mathbb{F}$-point of $\mathcal{X}^{\eta,\tau}$ for tame inertial L-parameter τ (in particular, τ is 9-generic) and let $\tilde{z} \stackrel{\text{def}}{=} \widetilde{w}^{*}(\bar{\rho}, \tau)$. Let $F(\lambda) \in W^{g}(\bar{\rho}, \tau)$ be a Serre weight such that for all $j \in \mathcal{J}$,*

$$\lambda_{\pi^{-1}(j)} \in X_1(T) \textit{ is in alcove } \widetilde{w}_h \cdot C_0 \textit{ if } \ell(\tilde{z}_j) \leq 1. \tag{5-1}$$

Then $M_\infty(\sigma(\tau)^{F(\lambda)})$ is a free $R_\infty(\tau)$-module of rank 1.

The proof is similar to the case when $\bar{\rho}$ is semisimple ([Le et al. 2020, Theorem 5.1.1] with slightly weaker genericity assumptions), and we will indicate the necessary modifications. First, [Le et al. 2020, Theorem 5.1.1] relies on a structure theorem for lattices in generic Deligne–Lusztig representations of $\underline{G}_0(\mathbb{F}_p)$ [loc. cit., Theorem 4.1.9]. The following proposition improves the genericity hypothesis of that result. We refer the reader to [loc. cit.] for unexplained notation or terminology.

Proposition 5.3.2. *Let R be $R_s(\mu)$ where $\mu - \eta \in \underline{C}_0$ is 9-deep. Then the radical filtration of $\bar{R}^{\sigma}$ is predicted by the extension graph with respect to σ, and the graph distance, the radical distance and the saturation distance from σ all coincide on $\Gamma(\bar{R}^{\sigma})$.*

Proof. As we now explain, the proof of [Le et al. 2020, Proposition 4.3.7] works for 9-generic R using some minor improvements to genericity hypotheses. Replace $R^{\mathrm{expl},\nabla}_{\overline{\mathfrak{M}},\widetilde{w}}$ with a suitable completion of $\big(\mathcal{O}(\widetilde{U}(\widetilde{w}, \eta, \nabla_{\tau,\infty}))\big)$ and the primes $\mathfrak{p}^{\mathrm{expl}}(\sigma)$ with suitable completions of the primes corresponding to $\widetilde{\mathfrak{P}}_{\sigma,\widetilde{w}}$ in Theorem 3.3.2. The results of [Le et al. 2020, §3.5, 3.6] appearing in the proof of [Le et al. 2020, Proposition 4.3.7] hold for 7-generic $\bar{\rho}_S$. Indeed, [Le et al. 2020, Theorem 3.5.2] holds by the same argument using Proposition 5.1.4 in place of [Le et al. 2020, Proposition 3.5.6]. The rest of the results follow from Theorem 3.3.2. In particular, it holds for the $\bar{\rho}_S$ chosen in the proof of [Le et al. 2020, Proposition 4.3.7] since R is 9-generic. (Proposition 3.4.5 in [Le et al. 2020] holds with $(n-3)$ replaced by $(n-2)$. Indeed, [Le et al. 2019, Proposition 3.3.2] holds with $m-n$ replaced by $m-n+1$. The proof shows this stronger result *and* that all lowest alcove presentations of τ are $(m-n)$-generic.) All the subsequent statements appearing in [Le et al. 2020, §4.3] then hold for 9-generic R (note that [Le et al. 2020, Theorem 4.2.16] holds for 8-generic R). □

We prove Theorem 5.3.1 through a series of lemmas. Until the end of the proof of Theorem 5.3.1, fix $\bar{\rho}$ 8-generic, τ a tame inertial L-parameter such that $\bar{\rho} \in \mathcal{X}^{\eta,\tau}(\mathbb{F})$ (in particular 6-generic), $\tilde{z} = \widetilde{w}^*(\bar{\rho}, \tau)$, $\widetilde{w} \overset{\text{def}}{=} \tilde{z}^*$, and λ satisfying (5-1). We write $\bar{\sigma}(\tau)^\sigma$ for $\sigma(\tau)^\sigma \otimes_{\mathcal{O}} \mathbb{F}$ in what follows.

Below, we modify the proofs of [Le et al. 2020, §5.1]. We will refer to the following as the usual modifications: we replace $\tau_{\mathcal{S}}$, $\bar{\rho}_{\mathcal{S}}$, $W^?(\bar{\rho}_{\mathcal{S}})$, $W^?(\bar{\rho}_{\mathcal{S}}, \tau_{\mathcal{S}})$, $\widetilde{w}_i$, and $\widetilde{w}_i^*$ in [loc. cit.] by τ, $\bar{\rho}$, $W^g(\bar{\rho})$, $W^g(\bar{\rho}, \tau)$, $\tilde{z}_i$, and $\widetilde{w}_i$, respectively. (In [Le et al. 2020, §5.1], the set S_p is denoted $\mathcal{S}$.)

Lemma 5.3.3. *Assume that τ is 9-generic (for instance, if $\bar{\rho}$ is 11-generic) and $\ell(\tilde{z}_j) > 1$ for all $j \in \mathcal{J}$. Let V be a quotient of $\bar{\sigma}(\tau)^\sigma$. Then $M_\infty(V)$ is a cyclic $R_\infty(\tau)$-module.*

Proof. First, the scheme-theoretic support of $M_\infty(\sigma)$ is, by Theorem 5.2.6, (nonempty and) generically reduced and hence reduced, e.g., by the proof of [Le et al. 2020, Lemma 3.6.2]. It is then formally smooth by Table 3, and so $M_\infty(\sigma)$ is free over its scheme-theoretic support by the Auslander–Buchsbaum–Serre theorem and the Auslander–Buchsbaum formula.

Now the proof of [Le et al. 2020, Lemma 5.1.3] applies after the usual modifications. Moreover, in the setup for [Le et al. 2020, Proposition 4.3.7, Lemma 3.6.10], $R^{\text{expl},\nabla}_{\overline{\mathfrak{M}},\widetilde{w}}$ should be replaced by $\big(\mathcal{O}(\widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty}))\big)^\wedge_x$ (for a suitable $x \in \widetilde{U}(\tilde{z}, \eta, \nabla_{\tau,\infty})(\mathbb{F})$) and N by $\log_2 \# W^g(\bar{\rho}, \tau)$. (Lemma 3.6.10 in [Le et al. 2020] holds for $\bar{\rho}$ and τ with $W^?(\bar{\rho}, \tau)$ replaced by $W^g(\bar{\rho}, \tau)$ by Theorem 3.3.2.) $\square$

We now assume that $\bar{\rho}$ is 11-generic (in particular, τ is 9-generic). We fix a semisimple $\bar{\rho}^{\text{sp}} : G_{\mathbb{Q}_p} \to {}^L\underline{G}(\mathbb{F})$ such that $\widetilde{w}^*(\bar{\rho}^{\text{sp}}, \tau) = \widetilde{w}^*(\bar{\rho}, \tau) = \tilde{z}$. By Corollary 4.2.8, if $\bar{\rho} \in \mathcal{X}^{\eta,\tau'}(\mathbb{F})$ for a 4-generic τ', then $\bar{\rho}^{\text{sp}} \in \mathcal{X}^{\eta,\tau'}(\mathbb{F})$ and

$$\widetilde{w}^*(\bar{\rho}^{\text{sp}}, \tau') \leq \widetilde{w}^*(\bar{\rho}, \tau'). \tag{5-2}$$

Let $(s, \mu - \eta)$ be the 7-generic lowest alcove presentation for $\bar{\rho}^{\text{sp}}$ compatible with the implicit 9-generic lowest alcove presentation of τ so that $\bar{\rho}^{\text{sp}}|_{I_{\mathbb{Q}_p}} \cong \bar{\tau}(s, \mu)$.

Now let V be a quotient of $\bar{\sigma}(\tau)^\sigma$ such that there exist subsets $\Sigma_{V,j} \subseteq \widetilde{w}_j^{-1}(\Sigma_0)$ such that

$$\prod_{i \in \mathcal{J}} \Sigma_{V,i} \xrightarrow{\sim} \mathrm{JH}(V), \quad (\omega, a) \mapsto \sigma_{(\omega,a)} \overset{\text{def}}{=} F(\mathfrak{Tr}_\mu(s\omega, a))$$

is a bijection. We will show the cyclicity of $M_\infty(V)$ by inducting on the complexity of the set $W^g(\bar{\rho}, \tau) \cap \mathrm{JH}(V)$.

Let $\Sigma_j^g \subset r(\Sigma_0)$ such that $(\omega, a) \mapsto F(\mathfrak{Tr}_\mu(s\omega, a))$ defines a bijection from $\Sigma_j^g \to W^g(\bar{\rho})$. (The sets Σ_j^g exist by Section 2.1.G, Corollary 3.3.9, Theorem 4.2.5, and Corollary 4.2.8.)

Lemma 5.3.4. *Suppose that for all $j \in \mathcal{J}$ either $\ell(\tilde{z}_j) > 1$ or $\Sigma_{V,j} \subset \{(\varepsilon, 1), (0, 0), (\varepsilon_1, 0), (\varepsilon_2, 0)\}$ for some $\varepsilon \in \{0, \varepsilon_1, \varepsilon_2\}$. Then $M_\infty(V)$ is a cyclic $R_\infty(\tau)$-module.*

Proof. We induct on

$$n \overset{\text{def}}{=} \#\big\{i \in \mathcal{J} : \ell(\tilde{z}_i) \leq 1 \text{ and } \#\overline{\Sigma}_{V,i} = 3\big\}.$$

If $n = 0$ we let τ' be the tame inertial L-parameter corresponding to $\tau'_{\mathcal{S}}$ constructed in the first paragraph of [Le et al. 2020, Lemma 5.1.4] with respect to $\bar{\rho}^{\text{sp}}$. Since $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau))$ by construction, $\mathcal{X}^{\eta,\tau'}(\mathbb{F})$

contains the 11-generic $\bar{\rho}$ so that τ' is 9-generic. Then $\bar{\rho}$ arises from a point in $\widetilde{U}(\tilde{z}', \eta, \nabla_{\tau',\infty})(\mathbb{F})$ by Theorem 3.3.2 since $\sigma \in \mathrm{JH}(\bar{\sigma}(\tau'))$ and $\ell(\tilde{z}'_j) > 1$ by (5-2). The result follows now from Lemma 5.3.3.

Observe that if $\ell(\tilde{z}_j) \leq 1$, then $\lambda_{\pi^{-1}(j)} \in \widetilde{w}_h \cdot C_0$ so that $(\varepsilon, 1)$ is necessarily in Σ_j^g since $F(\lambda) \in W^g(\bar{\rho}) \cap \mathrm{JH}(V)$. The general case then follows verbatim as in the proof of the general case in [loc. cit.] with the usual modifications. Moreover, references to [Le et al. 2020, Theorem 3.6.4, Table 3, Theorem 4.1.9] are replaced by references to Theorem 3.3.2, Table 3, and Proposition 5.3.2 respectively; and references to [Le et al. 2020, Lemmas 3.6.12, 3.6.16 (3.19)] are replaced by references to Lemma 5.3.9 and 5.3.10 after localization at x (in fact [Le et al. 2020, Lemma 3.6.12] is sufficient for the case of $\tilde{z}_j = t_{\underline{1}}$). □

Lemma 5.3.5. *Suppose that $\ell(\tilde{z}_j) > 1$ or*

$$\Sigma_{V,j} \subset \widetilde{w}_j^{-1}(\Sigma_0 \setminus \{(\nu_1, 0), (\nu_2, 1), (\nu_3, 0)\}),$$

where (ν_1, ν_2, ν_3) is $(\varepsilon_1 - \varepsilon_2, \varepsilon_1, \varepsilon_1 + \varepsilon_2)$, $(\varepsilon_2 - \varepsilon_1, \varepsilon_2, \varepsilon_1 + \varepsilon_2)$, or $(\varepsilon_1 - \varepsilon_2, 0, \varepsilon_2 - \varepsilon_1)$. Then $M_\infty(V)$ is a cyclic $R_\infty(\tau)$-module.

Proof. This follows from the proof of [Le et al. 2020, Lemma 5.1.5] with the usual modifications. (In the reduction step in the first paragraph of the proof, one possibly changes τ and so possibly changes $\bar{\rho}^{\mathrm{sp}}$. This only affects this proof.) References to [Le et al. 2020, Theorem 4.1.9] are replaced by Proposition 5.3.2 above, and references to [Le et al. 2020, Lemma 5.1.4] are replaced by references to Lemma 5.3.4.

Finally, we can and do choose V^2 in the final paragraph of the proof of [Le et al. 2020, Lemma 5.1.5] so that if $(\varepsilon', 0) \in V_{i'}^2$ in the notation of [loc. cit.], then $(\varepsilon', 0) \in \Sigma_{i'}^g$. Indeed, $\ell(\tilde{z}_{i'}) \leq 1$ and [Le et al. 2018a, §8] ensure that $\widetilde{U}(\tilde{z}_{i'}, \eta_{i'}, \nabla_{s_{i'}^{-1}(\mu_{i'}+\eta_{i'})})_{\mathbb{F}}$ has at least 5 components, where $\tau = \tau(s, \mu + \eta)$, so that $\#\Sigma_{i'}^g \geq 6$ and contains two of $\widetilde{w}_{i'}^{-1}((0, 0), (\varepsilon_1, 0), (\varepsilon_2, 0))$ by Theorem 4.2.5. Then by Theorem 5.2.6, we can apply [Emerton et al. 2015, Lemma 10.1.13] as described in [Le et al. 2020]. □

Remark 5.3.6. There was a gap in the proof of [Le et al. 2020, Lemma 5.1.5]: in the proof there one needs to possibly change the type τ to an auxiliary type, which may cause a loss of 2 in the genericity. Since we need to apply [Le et al. 2020, Theorem 4.1.9] to this auxiliary type, one needs to increase the genericity assumption by 2 in [Le et al. 2020, Theorem 5.1.1].

Lemma 5.3.7. *Suppose that $\ell(\tilde{z}_j) > 1$ or $\Sigma_{V,j} \subset \widetilde{w}_j^{-1}(\Sigma_0 \setminus \{(\nu, 0)\})$, where ν is $\varepsilon_1 - \varepsilon_2$, $\varepsilon_2 - \varepsilon_1$, or $\varepsilon_1 + \varepsilon_2$. Then $M_\infty(V)$ is a cyclic $R_\infty(\tau)$-module.*

Proof. This follows from the proof of [Le et al. 2020, Lemma 5.1.6] with the usual modifications and using Lemma 5.3.5 and Lemmas 5.3.11 and 5.3.12 below (completed at x). □

Lemma 5.3.8. *With V as described before Lemma 5.3.4, $M_\infty(V)$ is a cyclic $R_\infty(\tau)$-module.*

Proof. The argument in the proof of [Le et al. 2020, Lemma 5.1.7] holds verbatim with the usual modifications and the reference to [Le et al. 2020, Lemma 5.1.6] replaced by a reference to Lemma 5.3.7. □

Proof of Theorem 5.3.1. Theorem 5.3.1 follows from the proof of [Le et al. 2020, Theorem 5.1.1] using Lemma 5.3.8 in place of [Le et al. 2020, Lemma 5.1.7]. □

In the lemmas below we refer the reader to Section 1.4.A for unexplained notation. These lemmas are algebraizations of [Le et al. 2020, Lemmas 3.6.12, 3.6.14, 3.6.16 (3.19), 3.6.16 (3.17) and (3.18)]. Their proof follows verbatim in our setting by replacing $\overline{R}^{\mathrm{expl},\nabla}_{\overline{\mathfrak{M}},\tilde{z}}$ and the ideals $\mathfrak{c}_{(\omega,a)}$ of [loc. cit.] with $\widetilde{U}(\tilde{z},\eta_j,\nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$ (with $\mu_j\in X^*(T)$ 4-deep) and the ideals $\widetilde{\mathfrak{P}}_{(\omega,a),\alpha t_{\underline{1}}}$, respectively. (The second displayed equation in the statement of the Lemma 5.3.12 is not covered by [Le et al. 2020, Lemma 3.6.16(3.17)], but the proof is analogous.) Alternatively, one observes that all the ideal equalities we need to verify can be checked after projecting $\widetilde{U}(\tilde{z}_j,\eta_j,\nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$ to Fl, where there is a contracting $T^\vee_{\mathbb{F}}$-action with unique fixed point $\tilde{z}_j$. Since all the ideals involved are $T^\vee_{\mathbb{F}}$-equivariant one only needs to check the equalities after completion at $\tilde{z}_j$, which are exactly the results of [Le et al. 2020, §3.6.3].

Lemma 5.3.9 [Le et al. 2020, Lemma 3.6.12]. *In $\widetilde{U}(t_{\underline{1}},\eta_j,\nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$, we have the following ideal relation*:

$$\big(\widetilde{\mathfrak{P}}_{(0,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,0),t_{\underline{1}}}\big)+\big(\widetilde{\mathfrak{P}}_{(0,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),t_{\underline{1}}}\big)=\big(\widetilde{\mathfrak{P}}_{(0,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),t_{\underline{1}}}\big).$$

Lemma 5.3.10 [Le et al. 2020, Lemma 3.6.16 (3.19)]. *In $\widetilde{U}(\alpha t_{\underline{1}},\eta_j,\nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$, we have the following ideal relation*:

$$\big(\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\big)+\big(\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2-\varepsilon_1,0),\alpha t_{\underline{1}}}\big)=\big(\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\big).$$

Lemma 5.3.11 [Le et al. 2020, Lemma 3.6.14]. *In $\widetilde{U}(t_{\underline{1}},\eta_j,\nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$, we have the following ideal relation*:

$$\begin{aligned}&\big(\widetilde{\mathfrak{P}}_{(0,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),t_{\underline{1}}}\big)\\&\quad+\big(\widetilde{\mathfrak{P}}_{(0,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,0),t_{\underline{1}}}\big)=\big(\widetilde{\mathfrak{P}}_{(0,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,0),t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),t_{\underline{1}}}\big).\end{aligned}$$

Lemma 5.3.12 [Le et al. 2020, Lemma 3.6.16 (3.17), (3.18)]. *In $\widetilde{U}(\alpha t_{\underline{1}},\eta_j,\nabla_{s_j^{-1}(\mu_j+\eta_j)})_{\mathbb{F}}$, we have the following ideal relations*:

$$\begin{aligned}&\big(\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2-\varepsilon_1,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,1),\alpha t_{\underline{1}}}\big)\\&\qquad+\big(\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,1),\alpha t_{\underline{1}}}\big)=\big(\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\big),\\&\big(\widetilde{\mathfrak{P}}_{(\varepsilon_2,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2-\varepsilon_1,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\big)\\&\qquad+\big(\widetilde{\mathfrak{P}}_{(\varepsilon_2,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_1,1),\alpha t_{\underline{1}}}\big)=\big(\widetilde{\mathfrak{P}}_{(\varepsilon_2,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,0),\alpha t_{\underline{1}}}\big),\\&\big(\widetilde{\mathfrak{P}}_{(\varepsilon_1,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\big)\\&\qquad+\big(\widetilde{\mathfrak{P}}_{(\varepsilon_1,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,1),\alpha t_{\underline{1}}}\big)=\big(\widetilde{\mathfrak{P}}_{(\varepsilon_1,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(0,1),\alpha t_{\underline{1}}}\cap\widetilde{\mathfrak{P}}_{(\varepsilon_2,0),\alpha t_{\underline{1}}}\big).\end{aligned}$$

5.3.A. *Gauges for patching functors.* Let τ be a tame inertial L-parameter τ and $\bar{\rho}:G_{\mathbb{Q}_p}\to{}^L\underline{G}(\mathbb{F})$ be an 11-generic L-homomorphism arising from an $\mathbb{F}$-point $\mathcal{X}^{\eta,\tau}(\mathbb{F})$. Let M_∞ be a weak minimal patching functor coming from an arithmetic $R_\infty[\mathrm{GL}_3(F_p)]$-module.

The scheme $X_\infty(\tau)$ is normal and Cohen–Macaulay by Corollary 3.3.3 and Remark 3.3.4. We let $Z\subset X_\infty(\tau)$ be the locus of points lying on two irreducible components of the special fiber of $X_\infty(\tau)$ (in particular $Z \subset X_\infty(\tau)$ has codimension at least two) and

$$j : U \overset{\text{def}}{=} X_\infty(\tau) \setminus Z \hookrightarrow X_\infty(\tau)$$

be the natural open immersion.

Theorem 5.3.13. *Let σ, $\kappa \in \mathrm{JH}(\bar{\sigma}(\tau))$ and let $\iota : \sigma(\tau)^\kappa \hookrightarrow \sigma(\tau)^\sigma$ be a saturated injection. For any $\theta \in W^g(\bar{\rho})$, let $m(\theta)$ be the multiplicity with which θ appears in the cokernel of ι. Suppose further that σ is as in Theorem 5.3.1. Then the induced injection $M_\infty(\iota) : M_\infty(\sigma(\tau)^\kappa) \hookrightarrow M_\infty(\sigma(\tau)^\sigma)$ has image*

$$j_* j^* \Bigg(\prod_{\theta \in W^g(\bar{\rho})} \mathfrak{p}(\theta)^{m(\theta)} R_\infty(\tau) \Bigg) M_\infty(\sigma(\tau)^\sigma),$$

where $\mathfrak{p}(\theta)$ is the minimal prime ideal of $(R^\tau_{\bar{\rho}})_{\mathbb{F}}$ corresponding to θ via Theorem 3.3.2 and localization at $\bar{\rho} \in \mathcal{X}^{\eta,\tau}_{\mathbb{F}}$.

Proof. The proof follows verbatim the argument of [Le et al. 2020, Lemmas 5.2.1 and 5.2.2, Theorem 5.2.3] after replacing occurrences of $W^?(\bar{\rho}_S)$ there with $W^g(\bar{\rho})$. □

5.4. *Global applications.* In this section, we deduce global applications of Theorems 5.2.6, 5.3.1, and 5.3.13 generalizing results of [Le et al. 2020, §5.3] in the tamely ramified setting. Let F^+ be a totally real field, S_p the set of places of F^+ dividing p, and F/F^+ a CM extension. We assume that all places of S_p are unramified over $\mathbb{Q}$ and split in F. We start with the following modularity lifting result.

Theorem 5.4.1. *Let F/F^+ be a CM extension, and let $r : G_F \to \mathrm{GL}_3(E)$ be a continuous representation such that*

- *r is unramified at all but finitely many places;*
- *r is potentially crystalline at places dividing p of type (η, τ), where τ is a tame inertial type that admits a lowest alcove presentation (s, μ) with μ 4-deep in alcove $\underline{C}_0$;*
- *$r^c \cong r^\vee \varepsilon^{-2}$;*
- *$\zeta_p \notin \bar{F}^{\ker \mathrm{ad} \bar{r}}$ and $\bar{r}(G_{F(\zeta_p)}) \subset \mathrm{GL}_3(\mathbb{F})$ is an adequate subgroup; and*
- *$\bar{r} \cong \bar{r}_\iota(\pi)$ for some π a regular algebraic conjugate self-dual cuspidal (RACSDC) automorphic representation of $\mathrm{GL}_3(\mathbb{A}_F)$ of weight 0 so that $\sigma(\tau)$ is a K-type for π at places dividing p.*

Then r is automorphic, i.e., $r \cong r_\iota(\pi')$ for some π' a RACSDC automorphic representation of $\mathrm{GL}_3(\mathbb{A}_F)$.

Proof. This follows from standard base change and Taylor–Wiles patching arguments using Corollary 3.3.3; see [Le et al. 2023b, Theorem 9.2.1; 2018a, Theorem 7.4], and the addendum in [Le et al. 2020, §6]. □

We now suppose that $F^+ \neq \mathbb{Q}$. Let $\mathcal{O}_{F^+,p} \overset{\text{def}}{=} \mathcal{O}_{F^+} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ be the finite étale $\mathbb{Z}_p$-algebra denoted $\mathcal{O}_p$ in Section 5.1, 5.2. We fix an outer form $G_{/F^+}$ of GL_n which splits over F and is definite at all archimedean places of F^+. There exists $N \in \mathbb{N}$, a reductive model $\mathcal{G}$ of G defined over $\mathcal{O}_{F^+}[1/N]$, and an isomorphism

$$\iota : \mathcal{G}_{/\mathcal{O}_F[1/N]} \xrightarrow{\iota} \mathrm{GL}_{3/\mathcal{O}_F[1/N]} \tag{5-3}$$

(see [Emerton et al. 2013, §7.1]). Given a subset $\mathcal{P}$ of finite complement in the set of finite places of F^+ which split in F and are coprime to pN, let $\mathbb{T}_{\mathcal{P}}$ be the universal Hecke algebra $\mathbb{T}_{\mathcal{P}}$ for places in $\mathcal{P}$ (see [Le et al. 2023b, §9.1]). Given a compact open $U = U^pU_p \leq G(\mathbb{A}_{F^+}^{\infty,p}) \times \mathcal{G}(\mathcal{O}_{F^+,p})$ and a finite $\mathcal{O}$-module W with a continuous U-action, let $S(U, W)$ be the space of algebraic automorphic forms of level U and coefficients W, as in [Le et al. 2023b, (9.2)]. If U is unramified at places in $\mathcal{P}$, then $S(U, W)$ has a natural $\mathbb{T}_{\mathcal{P}}$-action. Let $\mathbb{T}_{\mathcal{P}}(U, W)$ be the quotient of $\mathbb{T}_{\mathcal{P}}$ acting faithfully on $S(U, W)$.

Let $\mathcal{G}_3$ be the group scheme over $\mathbb{Z}$ defined in [Clozel et al. 2008, §2.1]. We consider a continuous Galois representation $\bar{r} : G_{F^+} \to \mathcal{G}_3(\mathbb{F})$ which is *automorphic* in the sense of [Le et al. 2023b], i.e., for which there exists a maximal ideal $\mathfrak{m} \subseteq \mathbb{T}_{\mathcal{P}}(U, W)$, for some level U and coefficients W satisfying

$$\det(1 - \bar{r}(\mathrm{Frob}_w)X) = \sum_{j=0}^{2} (-1)^j (N_{F/\mathbb{Q}}(w))^{\binom{j}{2}} T_w^{(j)} X^j \mod \mathfrak{m}$$

for all $w \in \mathcal{P}$. Note that the collection $(\bar{r}|_{G_{F_v^+}})_{v\in S_p}$ defines an $\mathbb{F}$-valued L-parameter, which will be denoted as $\bar{r}_p$ in what follows. For such $\bar{r}$, we define as in [Le et al. 2023b, Definition 9.1.1] the set $W(\bar{r})$ of *modular Serre weights for* $\bar{r}$.

Theorem 5.4.2. *Let $\bar{r} : G_{F^+} \to \mathcal{G}_3(\mathbb{F})$ be an automorphic Galois representation. Assume further that*

- *$\bar{r}|_{G_F}(G_{F(\zeta_p)})$ is adequate; and*
- *$\bar{r}_p$ is 8-generic.*

Then

$$W(\bar{r}) = W^g(\bar{r}_p).$$

Proof. The proof of [Le et al. 2020, Theorem 5.3.3] applies verbatim after replacing [loc. cit., Theorem 3.5.2] with Theorem 5.2.6 above. □

5.4.A. *Mod p multiplicity one.* We continue using the setup from Section 5.4. We assume further that F/F^+ is unramified at all finite places. We now let S_0 denote the set of finite places of F^+ away from p where $\bar{r}$ ramifies and assume that every place of S_0 splits in F. For each $v \in S_0$, with fixed lift $\tilde{v}$ in F, we let $\tau_{\tilde{v}}$ be the minimally ramified type in the sense of [Clozel et al. 2008, Definition 2.4.14] corresponding to $\bar{r}|_{G_{F_{\tilde{v}}}} : G_{F_{\tilde{v}}} \to \mathrm{GL}_3(\mathbb{F})$ and $\sigma(\tau_v) \stackrel{\text{def}}{=} \sigma(\tau_{\tilde{v}}) \circ \iota_{\tilde{v}}$ be the $\mathcal{G}(\mathcal{O}_{F_v^+})$-representation attached to it (where $\iota_{\tilde{v}}$ is the localization at $\tilde{v}$ of the isomorphism (5-3); $\sigma(\tau_v)$ is independent of the choice of $\tilde{v}|v$; see [Le et al. 2020, §5.3]). Fix an $\mathcal{O}$-lattice W_{S_0} in $\bigotimes_{v\in S_0} \sigma(\tau_v)$. We have the following mod p multiplicity one result.

Theorem 5.4.3. *Let $\bar{r} : G_{F^+} \to \mathcal{G}_3(\mathbb{F})$ be a continuous Galois representation such that $\bar{r}_p$ is 11-generic. Let τ and $F(\lambda) \in W^g(\bar{r}_p, \tau)$ be as in the statement of Theorem 5.3.1. Assume moreover that*

- *$\bar{r} : G_{F^+} \to \mathcal{G}_3(\mathbb{F})$ is automorphic;*
- *$\bar{r}|_{G_F}(G_{F(\zeta_p)})$ is adequate; and*
- *the places at which $\bar{r}$ ramifies split in F.*

Then

$$S\Big(U^p\mathcal{G}(\mathcal{O}_{F^+,p}),\ \Big(\bar{\sigma}(\tau)^{F(\lambda)}\circ\prod_{v\in S_p}\iota_{\tilde{v}}\Big)^{\vee}\otimes_{\mathcal{O}}W_{S_0}\Big)[\mathfrak{m}]$$

is one-dimensional over $\mathbb{F}$, *where* $\mathfrak{m}$ *is the maximal ideal in the Hecke algebra* $\mathbb{T}_{\mathcal{P}}$ *corresponding to* $\bar{r}$.

Proof. The proof of [Le et al. 2020, Theorem 5.3.4] applies verbatim after replacing the reference to [Le et al. 2020, Theorem 5.2.1] by a reference to Theorem 5.3.1 above. □

5.4.B. *Breuil's lattice conjecture.* Now consider an automorphic Galois representation $r: G_F\to \mathrm{GL}_3(E)$ as in Theorem 5.4.1 which is *minimally ramified*, i.e., for any place $\tilde{v}$ of F lying above some $v\in S_0$, the Galois representation $r|_{G_{F_{\tilde{v}}}}$ is minimally ramified in the sense of [Clozel et al. 2008, Definition 2.4.14]. Let λ be the kernel of the system of Hecke eigenvalues $\alpha:\mathbb{T}_{\mathcal{P}}\to\mathcal{O}$ associated to r, i.e., α satisfies

$$\det(1-r^{\vee}(\mathrm{Frob}_w)X)=\sum_{j=0}^{3}(-1)^j(N_{F/\mathbb{Q}}(w))^{\binom{j}{2}}\alpha(T_w^{(j)})X^j$$

for all $w\in\mathcal{P}$. For $U^p\le G(\mathbb{A}_{F^+}^{\infty,p})$ and a finite $\mathcal{O}$-module W with a continuous U^p-action. Let

$$S(U^p,W)\overset{\text{def}}{=}\varinjlim_{U_p}S(U^pU_p,W)\quad\text{and}\quad\widetilde{S}(U^p,W)\overset{\text{def}}{=}\varprojlim_{s\in\mathbb{N}}S(U^p,W/\varpi^s),$$

where U_p runs over the compact open subgroups of $\mathcal{G}(\mathcal{O}_{F^+,p})$. By Theorem 5.4.1, $\widetilde{S}(U^p,W_{S_0})[\lambda]$ is nonzero.

Theorem 5.4.4. *Let* $r: G_F\to\mathrm{GL}_3(E)$ *and* τ *be as in Theorem 5.4.1. Assume furthermore that* r *is minimally ramified and that the places at which* $\bar{r}$ *ramifies are split in* F. *Finally, assume that* $\bar{r}_p$ *is* 11*-generic. Then the lattice* $\sigma(\tau)\cap\widetilde{S}(U^p,W_{S_0})[\lambda]\subset\sigma(\tau)\cap\widetilde{S}(U^p,W_{S_0})[\lambda]\otimes_{\mathcal{O}}E$ *depends only on* r_p.

Proof. The proof of [Le et al. 2020, Theorem 5.3.5] applies verbatim after replacing occurrences of $W^?(\bar{r}_S)$ there with $W^g(\bar{r}_p)$ and the reference to [Le et al. 2020, Theorem 5.2.3] with a reference to Theorem 5.3.13. □

Acknowledgements

The main ideas of this article date back to the summer of 2016, but the formulation of the results were rather awkward due to the lack of the Emerton–Gee stack and its local model theory at that time. For this reason we decided to write up this paper only after the release of [Le et al. 2023b]. We apologize for the long delay. Part of the work has been carried out during visits at the Institut Henri Poinaré (2016), the Institute for Advanced Study (2017), Mathematisches Forschungsinstitut Oberwolfach (2019), University of Arizona, and Northwestern University. We would like to heartily thank these institutions for the outstanding research conditions they provided, and for their support.

Finally, Le was supported by the National Science Foundation under agreements DMS-1128155 and DMS-1703182 and an AMS-Simons travel grant. Le Hung acknowledges support from the National

Science Foundation under grants DMS-1128155, DMS-1802037 and the Alfred P. Sloan Foundation. Levin was partially supported by Simons Foundation/SFARI (no. 585753), National Science Foundation grant DMS-1952556, and the Alfred P. Sloan Foundation. Morra was supported by the Institut Universitaire de France and the ANR-18-CE40-0026 (CLap CLap).

References

[Barnet-Lamb et al. 2011] T. Barnet-Lamb, D. Geraghty, M. Harris, and R. Taylor, "A family of Calabi–Yau varieties and potential automorphy, II", *Publ. Res. Inst. Math. Sci.* **47**:1 (2011), 29–98. MR Zbl

[Caraiani and Levin 2018] A. Caraiani and B. Levin, "Kisin modules with descent data and parahoric local models", *Ann. Sci. Éc. Norm. Supér.* (4) **51**:1 (2018), 181–213. MR Zbl

[Caraiani et al. 2016] A. Caraiani, M. Emerton, T. Gee, D. Geraghty, V. Paškūnas, and S. W. Shin, "Patching and the p-adic local Langlands correspondence", *Camb. J. Math.* **4**:2 (2016), 197–287. MR Zbl

[Clozel et al. 2008] L. Clozel, M. Harris, and R. Taylor, "Automorphy for some l-adic lifts of automorphic mod l Galois representations", *Publ. Math. Inst. Hautes Études Sci.* **108** (2008), 1–181. MR Zbl

[Dotto and Le 2021] A. Dotto and D. Le, "Diagrams in the mod p cohomology of Shimura curves", *Compos. Math.* **157**:8 (2021), 1653–1723. MR Zbl

[Emerton and Gee 2014] M. Emerton and T. Gee, "A geometric perspective on the Breuil–Mézard conjecture", *J. Inst. Math. Jussieu* **13**:1 (2014), 183–223. MR Zbl

[Emerton and Gee 2023] M. Emerton and T. Gee, *Moduli stacks of étale (φ, Γ)-modules and the existence of crystalline lifts*, Annals of Mathematics Studies **215**, Princeton University Press, 2023. MR Zbl

[Emerton et al. 2013] M. Emerton, T. Gee, and F. Herzig, "Weight cycling and Serre-type conjectures for unitary groups", *Duke Math. J.* **162**:9 (2013), 1649–1722. MR Zbl

[Emerton et al. 2015] M. Emerton, T. Gee, and D. Savitt, "Lattices in the cohomology of Shimura curves", *Invent. Math.* **200**:1 (2015), 1–96. MR Zbl

[Enns 2019] J. Enns, "On weight elimination for $\mathrm{GL}_n(\mathbb{Q}_{p^f})$", *Math. Res. Lett.* **26**:1 (2019), 53–66. MR Zbl

[Gee and Geraghty 2012] T. Gee and D. Geraghty, "Companion forms for unitary and symplectic groups", *Duke Math. J.* **161**:2 (2012), 247–303. MR Zbl

[Gee et al. 2017] T. Gee, F. Herzig, T. Liu, and D. Savitt, "Potentially crystalline lifts of certain prescribed types", *Doc. Math.* **22** (2017), 397–422. MR Zbl

[Gee et al. 2018] T. Gee, F. Herzig, and D. Savitt, "General Serre weight conjectures", *J. Eur. Math. Soc.* (*JEMS*) **20**:12 (2018), 2859–2949. MR Zbl

[Herzig 2006] F. Herzig, *The weight in a Serre-type conjecture for tame n-dimensional Galois representations*, Ph.D. thesis, Harvard University, 2006, available at https://www.proquest.com/docview/305338172. Zbl

[Herzig 2009] F. Herzig, "The weight in a Serre-type conjecture for tame n-dimensional Galois representations", *Duke Math. J.* **149**:1 (2009), 37–116. MR Zbl

[Herzig 2011] F. Herzig, "The classification of irreducible admissible mod p representations of a p-adic GL_n", *Invent. Math.* **186**:2 (2011), 373–434. MR Zbl

[Herzig et al. 2017] F. Herzig, D. Le, and S. Morra, "On mod p local-global compatibility for GL_3 in the ordinary case", *Compos. Math.* **153**:11 (2017), 2215–2286. MR Zbl

[Le et al. 2018a] D. Le, B. V. Le Hung, B. Levin, and S. Morra, "Potentially crystalline deformation rings and Serre weight conjectures: shapes and shadows", *Invent. Math.* **212**:1 (2018), 1–107. MR Zbl

[Le et al. 2018b] D. Le, S. Morra, and C. Park, "On mod p local-global compatibility for $\mathrm{GL}_3(\mathbb{Q}_p)$ in the non-ordinary case", *Proc. Lond. Math. Soc.* (3) **117**:4 (2018), 790–848. MR Zbl

[Le et al. 2019] D. Le, B. V. Le Hung, and B. Levin, "Weight elimination in Serre-type conjectures", *Duke Math. J.* **168**:13 (2019), 2433–2506. MR Zbl

[Le et al. 2020] D. Le, B. V. Le Hung, B. Levin, and S. Morra, "Serre weights and Breuil's lattice conjecture in dimension three", *Forum Math. Pi* **8** (2020), art. id. e5. MR Zbl

[Le et al. 2022] D. Le, B. V. L. Hung, B. Levin, and S. Morra, "Extremal weights and a tameness criterion for mod p Galois representations", preprint, 2022. To appear in *J. Eur. Math. Soc.* (*JEMS*). arXiv 2206.06442

[Le et al. 2023a] D. Le, B. V. L. Hung, S. Morra, C. Park, and Z. Qian, "Moduli of Fontaine–Laffaille representations and a mod-p local-global compatibility result", preprint, 2023. To appear in *Mem. Amer. Math. Soc.* arXiv 2109.02720

[Le et al. 2023b] D. Le, B. V. Le Hung, B. Levin, and S. Morra, "Local models for Galois deformation rings and applications", *Invent. Math.* **231**:3 (2023), 1277–1488. MR Zbl

[Morra and Park 2017] S. Morra and C. Park, "Serre weights for three-dimensional ordinary Galois representations", *J. Lond. Math. Soc.* (2) **96**:2 (2017), 394–424. MR Zbl

Communicated by Shou-Wu Zhang
Received 2022-02-09 Revised 2023-04-21 Accepted 2023-09-03

ledt@purdue.edu Department of Mathematics, Purdue University, West Lafayette, IN, United States

lhvietbao@googlemail.com Department of Mathematics, Northwestern University, Evanston, IL, United States

bwlevin@rice.edu Department of Mathematics, Rice University, Houston, TX, United States

stefano.morra@math.univ-paris13.fr Laboratoire Analyse Géométrie Applications, Université Paris 8, Villetaneuse, France

ALGEBRA AND NUMBER THEORY 18:7 (2024)
https://doi.org/10.2140/ant.2024.18.1275

Combining Igusa's conjectures on exponential sums and monodromy with semicontinuity of the minimal exponent

Raf Cluckers and Kien Huu Nguyen

We combine two of Igusa's conjectures with recent semicontinuity results by Mustaţă and Popa to form a new, natural conjecture about bounds for exponential sums. These bounds have a deceivingly simple and general formulation in terms of degrees and dimensions only. We provide evidence consisting partly of adaptations of already known results about Igusa's conjecture on exponential sums, but also some new evidence like for all polynomials in up to 4 variables. We show that, in turn, these bounds imply consequences for Igusa's (strong) monodromy conjecture. The bounds are related to estimates for major arcs appearing in the circle method for local-global principles.

1. Introduction

Let f be a polynomial in n variables over $\mathbb{Z}$ and of degree $d > 1$, and let s be the (complex affine) dimension of the critical locus of the degree d homogeneous part of f. The main objects of our study are the finite exponential sums from (1) and their estimates in terms of n, d, and s as in Conjecture 1 below. For any positive integer N and any complex primitive N-th root of unity ξ, consider the exponential sum

$$\sum_{x\in(\mathbb{Z}/N\mathbb{Z})^n} \xi^{f(x)}. \tag{1}$$

When N runs over the set of prime numbers, the sums from (1) fall under the scope of works by Grothendieck, Deligne, Katz, Laumon, and others, building in particular on the Weil conjectures. We don't pursue new results for N running over the set of prime numbers. Instead we put forward new bounds for these sums uniformly in general N with, roughly, a win of a factor

$$N^{-(n-s)/d}$$

on the trivial bound; see Conjecture 1 below. In this context, Birch [1962] proved and used bounds with exponent

$$(n-s)/2^{d-1}(d-1)$$

MSC2020: primary 11L07; secondary 03C98, 11F23, 11S40, 14E18.
Keywords: Igusa's conjectures on monodromy and exponential sums, local-global principles, circle method, major arcs, minimal exponent, motivic oscillation index, log canonical threshold, Igusa's local zeta functions, motivic and p-adic integration, log resolutions.

instead of our projected and stronger $(n-s)/d$; see (5) below. The bounds in the conjecture look deceivingly simple, but a reduction argument to, say, the case $s=0$, turns out to be surprisingly hard in general, and moreover, the case $s=0$ for nonhomogeneous f is surprisingly hard as well. As evidence for Conjecture 1 we prove an almost generic case (based on the Newton polyhedron of f), as well as the case with up to 4 variables, and, the cases restricted to those N which are cube free and more generally $(d+2)$-th power free.

Conjecture 1 combines two of Igusa's conjectures, namely on exponential sums and on monodromy, and represents an update of these conjectures in line with the recently proved semicontinuity result for the minimal exponent by Mustață and Popa [2020] and the conjectured equality of the minimal exponent with the motivic oscillation index; see [Cluckers et al. 2019] and Section 2.7 below.

1.1. Let us make all this more precise, for f a polynomial over $\mathbb{Z}$ in n variables. For an integer $N>0$ and a complex primitive N-th root of unity ξ, put

$$E_f(N,\xi) := \left| \frac{1}{N^n} \sum_{x\in(\mathbb{Z}/N\mathbb{Z})^n} \xi^{f(x)} \right|, \tag{2}$$

which is simply the complex modulus of the sum in (1) normalized by the number of terms. Write d for the degree of f and f_d for the homogeneous degree d part of f. We assume that $d>1$. Write $s=s(f)$ for the dimension of the critical locus of f_d, namely, of the solution set in $\mathbb{C}^n$ of the equations

$$0=\frac{\partial f_d}{\partial x_1}(x)=\cdots=\frac{\partial f_d}{\partial x_n}(x). \tag{3}$$

Note that $0\le s\le n-1$. Our projected bounds are as follows:

Conjecture 1. *Given f, n, s, and d as above and any $\varepsilon>0$, one has*

$$E_f(N,\xi) \ll N^{-(n-s)/d+\varepsilon}. \tag{4}$$

In this context, note that Birch [1962, Lemma 5.4] obtained the following bound, based on the very same data of f, n, s, and d (and, assuming f to be homogeneous):

$$E_f(N,\xi) \ll N^{-\frac{n-s}{2^{d-1}(d-1)}+\varepsilon}, \tag{5}$$

which he used to estimate major arcs to obtain general local-global principles (see Section 1.9 below).

Remarkably, the weakening of Conjecture 1 with

$$-\frac{n-s}{2(d-1)}+\varepsilon \tag{6}$$

in the exponent of N in (4) instead of $-(n-s)/d+\varepsilon$ has just been shown in [Nguyen 2021], vastly improving Birch's bounds (5). The case of Conjecture 1 with $d=3$ is in line with the resembling (but averaged) bounds (170) of [Hooley 1988]. In the one variable case, similar bounds to (4) have already been studied; see, e.g., [Chalk 1987; Hua 1959] and some generalizations in [Cochrane and Zheng 1999;

2000]. Knowing only n, s and d, the exponent $-(n-s)/d$ is optimal in (4), as witnessed by $f = \sum_{i=1}^{n-s} x_i^d$ and $N = p^d$ for primes p; see also the example in (20).

Remark 1.2. The notation in (4) means that, given f and $\varepsilon > 0$, there is a constant $c = c(f, \varepsilon)$ such that, for all integers $N \geq 1$ and all primitive N-th roots of unity ξ, the value $E_f(N, \xi)$ is no larger than $cN^{-(n-s)/d+\varepsilon}$.

Remark 1.3. The critical case of Conjecture 1 is with N having a single prime divisor. Indeed, by the Chinese remainder theorem, if one writes $N = \prod_i p_i^{e_i}$ for distinct prime numbers p_i and integers $e_i > 0$, then one has

$$E_f(N, \xi) = \prod_i E_f(p_i^{e_i}, \xi_i) \tag{7}$$

for some primitive $p_i^{e_i}$-th roots of unity ξ_i. In detail, if one writes $1/N = \sum_i a_i/p_i^{e_i}$ with $(a_i, p_i) = 1$, then one takes $\xi_i = \xi^{b_i}$ with $b_i = a_i N/p_i^{e_i}$.

1.4. Conjecture 1 simplifies Igusa's original question on exponential sums (recalled in Section 1.6) to bounds involving only n, d, and s. It opens a way to proceed with Igusa's conjecture on exponential sums beyond the case of nonrational singularities that is obtained recently in [Cluckers et al. 2019].

In most of the evidence that we provide below, one can furthermore take $\varepsilon = 0$ and one may wonder to which extent this sharpening of Conjecture 1 holds. Such a sharpening with $\varepsilon = 0$ goes beyond Igusa's conjectures in ways explained in Section 2.5. One may also wonder whether the implied constant c can be taken depending only on f_d and n (and ε, but not on f). If one excludes a finite set S (depending on f or just on f_d) of prime divisors of N, then it seems furthermore possible that the implied constant can be taken depending only on d and n (and ε); see Remark 5.7 for more details.

1.5. In Section 2 we relate the bounds from Conjecture 1 to Igusa's monodromy conjecture. Conjecture 1 implies the strong monodromy conjecture for poles of local zeta functions with real part in the range strictly between $-(n-s)/d$ and zero. More precisely, we show under Conjecture 1 that there are no poles (of a local zeta function of f) with real part in this range except -1 (see Proposition 2.3); from [Mustață and Popa 2020] it follows correspondingly that there are no zeros of the Bernstein–Sato polynomial of f in this range other than -1 (see Proposition 2.4). Note that Conjecture 1 is much stronger than the strong monodromy conjecture in the mentioned range, as the latter implies merely a much weaker variant of Conjecture 1, namely the bounds from (15) instead of (4), where the constant c_p is allowed to depend on p.

1.6. Igusa's original question on exponential sums predicts upper bounds with a noncanonical exponent coming from a choice of log resolution for homogeneous f (with $f = f_d$); see [Igusa 1978]. More precisely, let $h : Y \to X = \mathbb{A}_{\mathbb{Q}}^n$ be a log resolution of $D = f^{-1}(0) := \mathrm{Spec}(\mathbb{Q}[x_1, \ldots, x_n]/(f))$, i.e., Y is an integral smooth scheme, $h : Y \to X$ is a proper map, the restriction $h : Y \backslash h^{-1}(D) \to X \backslash D$ is an isomorphism, and $(h^{-1}(D))_{\mathrm{red}}$ has simple normal crossings as subscheme of Y. Such a log resolution exists by the work of Hironaka [1964]. Write $h^{-1}(D) = \sum_{i \in I} N_i E_i$ and $\mathrm{Div}(h^*(dx_1 \wedge \cdots \wedge dx_n)) = \sum_{i \in I}(\nu_i - 1)E_i$

for irreducible components E_i of $(h^{-1}(D))_{\mathrm{red}}$ and positive integers N_i, ν_i. By blowing up further, one may suppose that $E_j \cap E_i = \varnothing$ whenever $(\nu_i, N_i) = (\nu_j, N_j) = (1, 1)$ and $i \neq j$. Put

$$J = \{i \in I \mid (\nu_i, N_i) \neq (1, 1)\} \quad \text{and} \quad \sigma_0 = \sigma_0(h) := \min_{i \in J} \frac{\nu_i}{N_i}.$$

Note that σ_0 depends on the choice of h in general. Igusa originally conjectured, for any $\sigma < \sigma_0$, and under a few extra conditions that are most likely superfluous (namely, that $\sigma_0 > 2$ and that f is homogeneous), that one has a bound

$$E_f(N, \xi) \leq cN^{-\sigma} \tag{8}$$

for all $N > 0$, all primitive N-th roots of unity ξ and a constant c independent of N, ξ. In the case that $\sigma_0 \leq 1$, the bounds (8) are proved even more generally than in Igusa's original conjecture in [Cluckers et al. 2019]; see Section 3.2 below. Furthermore, precisely (and only) in the case that $\sigma_0 \leq 1$ holds, the value σ_0 is independent of the choice of h, and, is called the log canonical threshold of f.

When one takes a fixed prime number p, Igusa [1978] proves that inequality (8) holds for N of the form $N = p^m$ with $m \geq 1$, primitive N-th root of unity ξ and a constant $c = c_p$ depending on p and $\sigma < \sigma_0$ (but not on m, ξ).

When f satisfies the nondegeneracy condition of Section 4, there is a toric log resolution h of D related to the Newton polyhedron of f at zero. In this case, $\sigma_0(h) = \sigma_f$ with σ_f defined again from the Newton polyhedron (see Section 4 for the definition of σ_f). Denef and Sperber [2001] conjectured that when f is nondegenerate, one can replace inequality (8) in Igusa's conjecture by

$$E_f(p^m, \xi) \leq cm^{\kappa-1} p^{-\sigma_f m}, \tag{9}$$

where κ is an invariant coming from the Newton polyhedron of f at zero (see Section 4) and c is independent of p, m, ξ. Thus, the Denef–Sperber conjecture is a bit stronger than Igusa's conjecture in the case of nondegenerate polynomials, by the more explicit form of the exponents σ_f and κ; it has been proved and generalized in [Denef and Sperber 2001; Cluckers 2008a; 2010; Castryck and Nguyen 2019]. See Proposition 4.1 below.

In [Cluckers and Veys 2016], some of Igusa's original conditions, like homogeneity for f, were dropped with some care, namely by focusing on squareful integers; see Section 2.7 for more details. Igusa's condition that $\sigma_0(h) > 2$ for some h was already dropped before; it was more relevant for his intended application of his conjecture to local-global principles than for the content of the conjecture itself. Additionally, Igusa's noncanonical exponent $\sigma_0(h)$ was replaced by canonical candidates for the exponent: the motivic oscillation index of f, and, (expected to be equal) the minimal exponent of $f - v$ with v a well-chosen critical value of $f : \mathbb{C}^n \to \mathbb{C}$, see [Cluckers and Veys 2016; Cluckers et al. 2019] and Section 2.7 below. Our suggested bounds encompass several issues related to the minimal exponent (and, the motivic oscillation index), by replacing them by the much simpler and natural value $(n - s)/d$, yielding Conjecture 1 as new variant of (8). As an extra upshot, Conjecture 1 makes sense again for all positive integers N, and not only for squareful integers. Although the bounds from Conjecture 1 seem simple and very natural, they appear surprisingly hard to show in general, and even the much weaker bounds with constants depending on p and N running over powers of p as in (15), remain elusive in general up to date, even in the case with $s = 0$.

1.7. From Section 3 on we develop evidence for Conjecture 1. We first rephrase some well-known results as evidence, namely, Igusa's treatment of the smooth homogeneous case (with $f = f_d$ and $s = 0$), the case with degree $d = 2$, the case with $(n-s)/d \leq 1$, the case of at most 3 variables, and, the case with cube free N. We then generalize this further to new evidence for all N which are $(d+2)$-th power free (see Section 3.6). This treatment of the $(d+2)$-th power free case is mainly provided for expository reasons, as it uses some recent results on bounds of [Cluckers et al. 2016] in the context of motivic integration and uniform p-adic integration as in [Cluckers et al. 2018]; it indicates that the case $N = p^e$ with p prime and e small is generally more easy than with e large.

In Section 4, we show Conjecture 1 when f is nondegenerate with respect to its Newton polyhedron at zero, using recent work from [Castryck and Nguyen 2019] and some elementary reasoning on Newton polyhedrons. This shows that Conjecture 1 holds under often generic conditions, including the generic weighted homogeneous case, see Remark 4.5.

In Section 5, we show Conjecture 1 for all polynomial in up to 4 variables. This uses [Cluckers et al. 2019] to reduce to the case with $n = 4$, $d = 3$ and $s = 0$.

In our final Remark 5.7, we explain that throughout the evidence for Conjecture 1 of this paper, up to excluding a finite set S of primes divisors of N (depending on f), the constant c can be taken depending only on d, n and ε.

Let us finally mention the further evidence of [Nguyen 2021] for Conjecture 1, with the weakened exponent $(n-s)/2(d-1)$ in the upper bound of (4) instead of $(n-s)/d$.

1.8. In his vast program from [Igusa 1978], Igusa studies a certain adèlic Poisson summation formula related to f, inspired by Weil's work [1965] on the Hasse principle and Birch's work [1962] on more general local-global principles. Conjecture 1 would imply that Igusa's adèlic Poisson summation formula for f holds under the simple condition

$$n - s > 2d \tag{10}$$

which simplifies (and generalizes) the list of conditions put forward by Igusa [1978], and would drop in particular the condition of homogeneity on f.

1.9. Also for obtaining (or just for streamlining) local-global principles, Conjecture 1 may play a role. When f is homogeneous, the sums $E_f(N, \xi)$ appear for estimating the contribution of the major arcs in the circle method to get a local-global principle for f when

$$n - s > (d-1)2^d, \tag{11}$$

in work by Birch [1962] and in the recent sharpening from [Browning and Prendiville 2017], which both rely on Birch's bounds (5) quoted above. Birch shows that any homogeneous form $f = f_d$ with (11) and having smooth local zeros for each completion of $\mathbb{Q}$ automatically has a nontrivial rational zero. One may hope one day to replace Condition (11) on homogeneous f by (10), which is in line with a conjectured local-global principle from [Browning and Heath-Brown 2017]. Conjecture 1 would put the remaining obstacle completely with the estimation for the minor arcs (where actually already lie the limits of the current strategies). Other possible applications may be for small solutions of congruences as studied in, e.g., [Baker 1983].

Note that Birch's method [1962] also helps to understand the distribution of rational points in the projective hypersurface X associated with a homogeneous polynomial f. More precisely, the singular series

$$\mathfrak{S}(f) = \sum_{N\geq 1} N^{-n} \sum_{a\in(\mathbb{Z}/N\mathbb{Z})^\times} \sum_{x\in(\mathbb{Z}/N\mathbb{Z})^n} \exp\left(\frac{2\pi i a f(x)}{N}\right)$$

which is equal to the product of p-adic densities of f will contribute to the dominant term in the asymptotic formula of the number points of X of bounded height. Conjecture 1 implies that the singular series $\mathfrak{S}(f)$ is absolutely convergent if $n-s>2d$. Thereby Conjecture 1 may be useful for the future research on the distribution of rational points in algebraic hypersurfaces.

1.10. ***Generalization to a ring of integers.*** Before giving precise statements and proofs, we formulate a natural generalization to rings of integers (a generality we will not use later in this paper). For a ring of integers $\mathcal{O}$ of a number field and a polynomial g over $\mathcal{O}$, one can formulate an analogous conjecture with summation sets $(\mathcal{O}/I)^n$ with nonzero ideals I of $\mathcal{O}$ and primitive additive characters $\psi:\mathcal{O}/I\to\mathbb{C}^\times$. More precisely, let g be a polynomial in n variables of degree $d>1$ and with coefficients in $\mathcal{O}$. For any nonzero ideal I of $\mathcal{O}$ and any primitive additive character $\psi:\mathcal{O}/I\to\mathbb{C}^\times$, let $N_I:=[\mathcal{O}:I]$ be the absolute norm of I and consider

$$E_g(I,\psi) := \left|\frac{1}{N_I^n}\sum_{x\in(\mathcal{O}/I)^n}\psi(g(x))\right|. \tag{12}$$

Write s for the dimension of the critical locus of the degree d homogeneous part g_d of g. As a generalization of the above questions, one may wonder whether for each $\varepsilon>0$ (or more strongly with $\varepsilon=0$) one has

$$E_g(I,\psi) \ll N_I^{-\frac{n-s}{d}+\varepsilon}. \tag{13}$$

As above with the Chinese remainder theorem, one can rephrase this using the finite completions of the field of fractions of $\mathcal{O}$. Furthermore, one can study similar sums for the local fields $\mathbb{F}_q((t))$ (with similar methods in the large characteristic case); see, e.g., [Cluckers and Veys 2016, Section 2.6; Cluckers et al. 2019, Section 1.2].

2. Link with the monodromy conjecture

2.1. Fix a prime number p. For each integer $m\geq 0$ let $a_{p,m}$ be the number of solutions in $(\mathbb{Z}/p^m\mathbb{Z})^n$ of the equation $f(x)\equiv 0 \bmod p^m$, and consider the Poincaré series

$$P_{f,p}(T) := \sum_{m\geq 0}\frac{a_{p,m}}{p^{mn}}T^m,$$

in $\mathbb{Z}[\![T]\!]$. Igusa [1975; 1978] showed that $P_{f,p}(T)$ is a rational function in T, using a log resolution of $f^{-1}(0)$. Let T_0 be a complex pole of $P_{f,p}(T)$ and let t_0 be the real part of a complex number s_0 with $p^{-s_0} = T_0$. Let $h : Y \to \mathbb{A}^n_{\mathbb{Q}}$ be a log resolution of $f^{-1}(0)$ and $(N_i, \nu_i)_{i \in I}$ as in Section 3. Igusa [1978] showed that t_0 belongs to the set $\mathcal{P}_h = \{-\nu_i/N_i \mid i \in I\}$. However, $\mathcal{P}_h$ depends on the choice of log resolution h. Igusa [1975, Theorem 2] also showed a strong link between exponential sums and local zeta functions (see also [Denef 1991, Corollary 1.4.5; [Denef and Veys 1995, Proposition 2.7]]), yielding the following corollary.

Corollary 2.2 [Igusa 1975]. *For p, f, T_0 and t_0 as above, if T_0 is furthermore a pole of $(T-p)P_{f,p}(T)$, then*

$$p^{mt_0} \le c'_p E_f(p^m, \xi) \tag{14}$$

for infinitely many m and ξ and a constant c'_p independent of m, ξ.

Proof. Proposition 2.7 of [Denef and Veys 1995] gives finitely many complex numbers T, finitely many characters $\chi : \mathbb{C}^\times \to \mathbb{C}^\times$ of finite order, finitely many integers $b \ge 0$, and finitely many complex numbers c such that for large m, $E_f(p^m, \xi)$ is (the complex modulus of) a finite $\mathbb{C}$-linear combination of the terms

$$\chi(\xi) \cdot T^{-m} m^b \xi^c,$$

where furthermore a term with $T = T_0$ appears nontrivially in this linear combination for each pole T_0 of $(T-p)P_{f,p}(T)$. Now the corollary follows by looking at the dominant terms, namely, with largest occurring real part of T and for such T the largest occurring value for b. □

Denef [1991] formulated a strong variant of Igusa's monodromy conjecture by asking whether t_0 as above is automatically a zero of the Bernstein–Sato polynomial of f. The following result addresses this question in a range of values for t_0, namely strictly between $-(n-s)/d$ and zero, assuming Conjecture 1 for f.

Proposition 2.3 (strong monodromy conjecture, in a range). *Let f, n, s, and d be as in the introduction and suppose that Conjecture 1 holds for f. Let t_0 be coming as above from a pole T_0 of $P_{f,p}(T)$ for a prime number p. Suppose that moreover $t_0 > -(n-s)/d$. Then $t_0 = -1$, and hence, t_0 is a zero of the Bernstein–Sato polynomial of f.*

Proposition 2.3 is a form of the strong monodromy conjecture in the range strictly between $-(n-s)/d$ and zero. We don't pursue the highest generality here, and leave the generalization for other variants of zeta functions like twisted p-adic local zeta (or even motivic) functions to the reader. Proposition 2.4 below gives a related statement for the zeros of the Bernstein–Sato polynomial of f.

Proof of Proposition 2.3. Let p be a prime number. Let t_0 be the real part of a complex number s_0 such that $T_0 := p^{-s_0}$ is a pole of $P_{f,p}(T)$. Suppose that for all $\varepsilon > 0$ there exists $c_p = c_p(f, \varepsilon)$ such that

$$E_f(p^m, \xi) \le c_p \cdot (p^m)^{-\frac{n-s}{d}+\varepsilon} \text{ for all } m > 0 \text{ and all primitive } \xi. \tag{15}$$

By Corollary 2.2 it follows that t_0 either equals -1 or, one has

$$p^{mt_0} \leq c'_p E_f(p^m, \xi) \tag{16}$$

for infinitely many pairs (m, ξ) and a constant c'_p independent of m, ξ. Clearly the bound from (15) holds if Conjecture 1 holds for f. By (15) and (16), if $t_0 > -(n-s)/d$, then $t_0 = -1$. Since f is nonconstant, the value -1 is automatically a zero of the Bernstein–Sato polynomial of f. This completes the proof of the proposition. □

Showing the bounds (15) from the above proof for general f does not seem easy, although they are much weaker (and much less useful adelically) than the bounds from (4), because of the dependence of c_p on p.

In view of the strong monodromy conjecture, Proposition 2.3 should be compared with the following absence of zeros of the Bernstein–Sato polynomial in a similar range, apart from -1. Recall that the zeros of the Bernstein–Sato polynomial are negative rational numbers.

Proposition 2.4. *Let f, n, s, and d be as in the introduction and let r be any zero of the Bernstein–Sato polynomial of f. Then either $r = -1$, or, $r \leq -(n-s)/d$.*

Proof. We write $f = f_0 + \cdots + f_d$ with f_i is the homogeneous part of degree i of f. Item (3) of Theorem E of [Mustaţă and Popa 2020] states that the minimal exponent $\tilde{\alpha}_{f,0}$ of f at zero is at least $(n-s)/d$ if f is homogeneous. Recall that the minimal exponent $\tilde{\alpha}_f$ of f is equal to $\min_{x \in f^{-1}(0)} \tilde{\alpha}_{f,x}$. Moreover, if $\varphi : \mathbb{A}^n \to \mathbb{A}^n$ is a linear change of variables then $\tilde{\alpha}_{f,x} = \tilde{\alpha}_{f\circ\varphi,\varphi^{-1}(x)}$, and, for any constant $\beta \neq 0$ one has $\tilde{\alpha}_{f,x} = \tilde{\alpha}_{\beta f,x}$. Let $g_\lambda(x)$ be $f_d + \sum_{0 \leq i \leq d-1} \lambda^{d-i} f_i$. Then for each $\lambda \neq 0$ we have $g_\lambda(x) = \lambda^d f(x/\lambda)$. Write $X = \mathbb{A}^n \times \mathbb{A}^1$, $T = \mathbb{A}^1$, $\pi : \mathbb{A}^n \times \mathbb{A}^1 \to \mathbb{A}^1$ for the projection, $h(x, \lambda) = g_\lambda(x)$ and $D = h^{-1}(0)$. For each $x \in f^{-1}(0)$, we consider the section $s_x : T \to X$ with $\lambda \mapsto (\lambda x, \lambda)$, then $s_x(\lambda) \in D_\lambda$ since $h(\lambda x, \lambda) = g_\lambda(\lambda x) = \lambda^d f(\lambda x/\lambda) = 0$ if $\lambda \neq 0$ and $h(0, 0) = f_d(0) = 0$. Now we can use item (2) of Theorem E of [Mustaţă and Popa 2020] for X, T, π, D and s_x to see that for each $x \in f^{-1}(0)$ we have

$$\tilde{\alpha}_{f,x} = \tilde{\alpha}_{g_\lambda,\lambda x} \geq \tilde{\alpha}_{g_0,0} = \tilde{\alpha}_{f_d,0} \geq (n-s)/d$$

for all $\lambda \neq 0$ in a small enough neighborhood of 0. Thus,

$$\tilde{\alpha}_f = \min_{x \in f^{-1}(0)} \tilde{\alpha}_{f,x} \geq (n-s)/d.$$

The proposition now follows directly from the definition of the minimal exponent $\tilde{\alpha}_f$ of f as the smallest zero of $b_f(-s)/(s-1)$, where $b_f(s)$ is the Bernstein–Sato polynomial of f. □

2.5. The variant of Conjecture 1 with $\varepsilon = 0$ (or even just the bounds (15) with $\varepsilon = 0$) implies for any pole T_0 of $P_{f,p}(T)$ with corresponding value t_0 the following bound on the order of the pole: If t_0 equals $-(n-s)/d$ and $-(n-s)/d \neq -1$, then the pole T_0 has multiplicity at most one, and, if $t_0 = -1 = -(n-s)/d$, then the pole T_0 has multiplicity at most two, by a similar reasoning as for Corollary 2.2.

Remark 2.6. Conjecture 1 implies the bounds (15) with moreover constants c_p taken independently from p, and, the conjecture in turns would follow from this. By (7), the variant of Conjecture 1 with $\varepsilon = 0$ is equivalent with the bounds (15) with $\varepsilon = 0$ and such that furthermore the products of the constants c_p over any set P of primes is bounded independently of P.

2.7. The minimal exponent of f is defined as the smallest zero of the quotient $b_f(-s)/(s-1)$ with $b_f(s)$ the Bernstein–Sato polynomial of f if such a zero exists, and it is defined as $+\infty$ otherwise. Write $\hat{\alpha}_f$ for the minimum of the minimal exponents of $f - v$ for v running over the (complex) critical values of f. In a more canonical variant of Igusa's original question, one may wonder more technically than Conjecture 1 whether for all $\varepsilon > 0$ one has

$$E_f(N, \xi) \ll N^{-\hat{\alpha}_f+\varepsilon} \text{ for all } \xi \text{ and all squareful integers } N, \tag{17}$$

similarly as the question introduced in [Cluckers and Veys 2016] for the motivic oscillation index (and where the necessity of working with squareful integers N is explained). Recall that an integer N is called squareful if for any prime p dividing N also p^2 divides N. In [Castryck and Nguyen 2019; Chambille and Nguyen 2021; Cluckers 2008a; 2010; 2019; Denef and Sperber 2001], evidence is given for this sharper but more technical question. As mentioned above, $\hat{\alpha}_f$ is hard to compute in general, and $(n-s)/d$ is much more transparent. However, $\hat{\alpha}_f$ is supposedly equal to the motivic oscillation index of f, which in turn is optimal as exponent of N^{-1} in the upper bounds for $E_f(N, \xi)$ for squareful N (see the last section of [Cluckers et al. 2019], or, a reasoning as for Corollary 2.2). Note that by Proposition 2.4, one has

$$\hat{\alpha}_f \geq (n-s)/d, \tag{18}$$

which shows that (17) is indeed a sharper (or equally sharp) bound than (4).

3. Some first evidence

In this section we translate some well-known results into evidence for Conjecture 1, and we show the (new) case of $(d+2)$-th power free N. A key (but hard) case of Conjecture 1 for inhomogeneous f is when f_d is projectively smooth, namely with $s = 0$, since the case of general s can be derived from a sufficiently uniform form of the inhomogeneous case with $s = 0$, see, e.g., how (22) is used below for squarefree N. However, the inhomogeneous case with $s = 0$ seems very hard at the moment. This should not be confused with Igusa's more basic case recalled in Section 3.1, for homogeneous f with $s = 0$.

3.1. When f itself is smooth homogeneous, namely, $f = f_d$ and $s = 0$, then Conjecture 1 with $\varepsilon = 0$ is known by Igusa's bounds [1978] by a straightforward computation and reduction to Deligne's bounds. In detail, if $f = f_d$ and $s = 0$, Igusa [1978] showed (using [Deligne 1974]) that for each prime p there is a constant c_p such that

$$E_f(p^m, \xi) \leq c_p p^{-mn/d} \text{ for all integers } m > 0 \text{ and all choices of } \xi, \tag{19}$$

and, that one can take $c_p = 1$ when p is larger than some value M depending on f. More precisely, one can take $c_p = 1$ when p does not divide d and when the reduction of f modulo p is smooth. Furthermore, Igusa [1978] shows that the exponent $-n/d$ of p^m is optimal in the upper bound of (19) when $m = d$. This easily shows that the exponent $(n-s)/d$ is optimal in Conjecture 1, for example by taking

$$f = (x_1 + \cdots + x_{s+1})^d + x_{s+2}^d + \cdots + x_n^d \tag{20}$$

and $N = p^d$ for all prime numbers p.

3.2. When f is such that

$$(n-s)/d \leq 1, \tag{21}$$

then Conjecture 1 follows from [Cluckers et al. 2019] and its recent solution of Igusa's conjecture for nonrational singularities. Indeed, in [Cluckers et al. 2019] the stronger (and optimal) upper bounds from (17) are shown for all squareful N in the case of nonrational singularities, as well as the case with 1 in the exponent instead of $\hat{\alpha}_f$ in the case of rational singularities. Recall that this is indeed stronger, by (18). The bounds for those integers N that are not squareful are recovered by the treatment of squarefree N below, by writing a general integer as a product of a squareful and a squarefree integer. We mention on the side that $\hat{\alpha}_f \leq 1$ if and only if $f - v = 0$ has nonrational singularities for some critical value $v \in \mathbb{C}$ of f, by [Saito 1993] and that in this case $\hat{\alpha}_f$ equals the minimum of the log canonical thresholds of $f - v$ for v running over the (complex) critical values of f. These results under condition (21) imply that Conjecture 1 holds for all f in three (or less) variables. Indeed, the degree two case is easy by diagonalizing f_2 over $\mathbb{Q}$, and, (21) holds when $n \leq 3 \leq d$. More surprisingly, Igusa's Conjecture (with the motivic oscillation index in the upper bound) is proved recently in [Nguyen and Veys 2022] for all polynomials in 3 variables. Some related results of the special case with $n \leq 2$ are developed in [Fraser and Wright 2020; Lichtin 2013; Veys 2020]. In Section 5 we will prove that Conjecture 1 holds for all polynomials in up to four variables.

3.3. Although it is classical, let us explain the case of $d = 2$ in more detail, by showing that Conjecture 1 holds with $\varepsilon = 0$ for f of degree $d = 2$. In fact, the argument as in the proof of Lemma 25 of [Heath-Brown 1996] is shorter and simpler for the case $d = 2$, but our treatment will be useful later in this paper. First suppose that the degree two part of f is a diagonal form, namely, $f_2(x) = \sum_{i=1}^n a_i x_i^2$ for some $a_i \in \mathbb{Z}$. In this case it is sufficient to show the case with $n = 1$ and $d = 2$ (indeed, $f = h_1(x_1) + \cdots + h_n(x_n)$ for some polynomials h_j in one variable x_j and of degree ≤ 2). But this case follows readily from Hua's bounds, see [Hua 1959] or [Chalk 1987], and is in fact elementary.

For general f with $d = 2$, by diagonalizing f_2 over $\mathbb{Q}$ and taking a suitable integer multiple, we find a matrix $T \in \mathbb{Z}^{n \times n}$ with nonzero determinant so that $f_2(Tx)$ is a diagonal form over $\mathbb{Z}$ in the variables x, namely, $f_2(Tx) = \sum_{i=1}^n a_i x_i^2$ for some $a_i \in \mathbb{Z}$. The map sending x to Tx transforms $\mathbb{Z}_p^n$ into a set of the form $\prod_{j=1}^n p^{e_{p,j}} \mathbb{Z}_p$ for some integers $e_{p,j} \geq 0$ (called a box). By composing with a map of the form

$(x_j)_j \mapsto (b_j x_j)_j$ for some integers b_j it is clear that we may assume that T is already such that $e_{p,j} = e_p$ for all p and all j and some integers $e_p \geq 0$. Hence, the case $d = 2$ follows from Lemma 3.4.

Lemma 3.4. *Let f, n, s, and d be as in the introduction and let $T \in \mathbb{Z}^{n\times n}$ be a matrix with nonzero determinant and such that, for each prime p, the transformation $x \mapsto Tx$ maps $\mathbb{Z}_p^n$ onto $p^{e_p}\mathbb{Z}_p^n$ for some $e_p \geq 0$. Then, Conjecture 1 for each of the polynomials $g_i(x) := f(i + Tx)$ for $i \in \mathbb{Z}^n$ implies Conjecture 1 for f, and, similarly for Conjecture 1 with $\varepsilon = 0$.*

Proof. For each $i \in \mathbb{Z}^n$, write $g_i(x)$ for the polynomial $f(i + Tx)$. For any prime p, let $\mu_{p,n}$ be the Haar measure on $\mathbb{Q}_p^n$, normalized so that $\mathbb{Z}_p^n$ has measure 1. For any integer $m > 0$ and any primitive p^m-th root of unity ξ, we have, by the change of variables formula for p-adic integrals, and with $e = e_p$ and with integrals taken against the measure $\mu = \mu_{p,n}$,

$$E_f(p^m, \xi) = \left| \int_{x\in\mathbb{Z}_p^n} \xi^{f(x) \bmod p^m} \mu \right| \leq \sum_{j=1}^{n} \sum_{i_j=0}^{p^e-1} \left| \int_{x\in i+(p^e\mathbb{Z}_p)^n} \xi^{f(x) \bmod p^m} \mu \right|.$$

For each i we further have

$$\begin{aligned} \left| \int_{x\in i+(p^e\mathbb{Z}_p)^n} \xi^{f(x) \bmod p^m} \mu \right| &= p^{-ne} \left| \int_{x\in\mathbb{Z}_p^n} \xi^{f(i+p^e x) \bmod p^m} \mu \right| \\ &= p^{-ne} \left| \int_{x\in\mathbb{Z}_p^n} \xi^{g_i(x) \bmod p^m} \mu \right| = p^{-ne} E_{g_i}(p^m, \xi). \end{aligned}$$

Since $e_p = 0$ for all but finitely many primes p, we are done. □

3.5. When one restricts to integers N which are squarefree (namely, not divisible by a nontrivial square), then Conjecture 1 with $\varepsilon = 0$ follows from Deligne's bound [1974], as we now explain. The reasoning is classical but also instructive for later use in this paper; a similar induction argument on $s \geq 0$ already appears in [Hooley 1991]. By [Deligne 1974], for each prime number p such that the reduction of f_d modulo p is smooth, one has

$$E_f(p, \xi) \leq (d-1)^n p^{-n/2} \quad \text{for each primitive } p\text{-th root of unity } \xi, \tag{22}$$

where smooth means that the reduction modulo p of the equations (3) has 0 as the only solution over an algebraic closure of $\mathbb{F}_p$. If $s = 0$ then the reduction of f_d modulo p is smooth whenever p is large and thus Conjecture 1 for squarefree N and with $\varepsilon = 0$ follows for f with $s = 0$ (note the different exponent of p in (22) and of N in 1 when $d > 2$). We proceed by induction on s by restricting f to hyperplanes, as follows. The bound (22) for all large p is our base case when $s = 0$. Now suppose that $s > 0$. After a linear coordinate change of $\mathbb{A}_{\mathbb{Z}}^n$, we may suppose that the polynomial $g(\hat{x}) := f(0, \hat{x})$ in the variables $\hat{x} = (x_2, \ldots, x_n)$ is still of degree d and that its degree d homogeneous part g_d has critical locus of dimension $s - 1$. Hence, for large prime p, the reduction of g_d modulo p has also critical locus of dimension $s-1$, in $\mathbb{A}^{n-1}_{\mathbb{F}_p}$. Hence, for large p, one has by induction on s that

$$\left|\frac{1}{p^{n-1}}\sum_{\hat{x}=(x_2,\dots,x_n)\in\mathbb{F}_p^{n-1}}\xi^{f(a,\hat{x})}\right|\leq (d-1)^{n-s}p^{-(n-s)/2} \tag{23}$$

for each $a\in\mathbb{F}_p$ and each primitive p-th root of unity ξ. Indeed, the polynomial $f(a,\hat{x})$ has g_d mod p as its degree d homogeneous part for each $a\in\mathbb{F}_d$. Now, summing over $a\in\mathbb{F}_p$ and dividing by p gives

$$\left|\frac{1}{p^n}\sum_{x\in\mathbb{F}_p^n}\xi^{f(x)}\right|\leq (d-1)^{n-s}p^{-(n-s)/2} \tag{24}$$

for large p (coming from the condition that the reduction of g_d modulo p has critical locus of dimension $s-1$). Conjecture 1 with $\varepsilon=0$ for squarefree integers N thus follows, by comparing the exponents in the upper bounds of (24) and (4), which allows to swallow the constant $(d-1)^{n-s}$ when $d>2$. (Alternatively, one can use the much more general Theorem 5 of [Katz 1999] when $d>2$ and an argument as in Section 3.3 when $d=2$.)

3.6. When one restricts to integers N which are cube free (namely, not divisible by a nontrivial cube), then Conjecture 1 with $\varepsilon=0$ follows exactly in the same way as for squarefree N, but now using both the bounds from [Heath-Brown 1985] and from [Deligne 1974]. Indeed, this similarly gives

$$E_f(p^2,\xi)\leq (d-1)^{n-s}p^{-(n-s)} \tag{25}$$

for large p and all ξ. Together with the squarefree case, this implies the cube free case of Conjecture 1, with $\varepsilon=0$ (note again the different exponent of p^2 in (25) and of N in 1, when $d>2$). In fact, with some more work we can go up to $(d+2)$-th powers instead of just cubes, as follows.

Proposition 3.7. *Conjecture 1 with $\varepsilon=0$ holds when restricted to integers N which are not divisible by a nontrivial $(d+2)$-th power. In detail, let f, n, s, and d be as in the introduction. Then there is a constant $c=c(f_d)$ (depending only on f_d) such that for all integers $N>0$ which are not divisible by a nontrivial $(d+2)$-th power and all primitive N-th roots ξ of 1, one has*

$$E_f(N,\xi)\leq cN^{-(n-s)/d}. \tag{26}$$

We will prove Proposition 3.7 by making a link between $E_f(p^m,\xi)$ and finite field exponential sums, as follows. For any prime p, any $m>0$, any point P in $\mathbb{F}_p^n$ and any ξ, write

$$E_f^P(p^m,\xi):=\left|\frac{1}{p^{mn}}\sum_{x\in P+(p\mathbb{Z}/p^m\mathbb{Z})^n}\xi^{f(x)}\right|. \tag{27}$$

Compared to $E_f(p^m,\xi)$, the summation set for $E_f^P(p^m,\xi)$ has p-adically zoomed in around the point P.

Let us consider the following positive characteristic analogues,

$$E_f(t^m,\psi):=\left|\frac{1}{p^{mn}}\sum_{x\in(\mathbb{F}_p[t]/(t^m))^n}\psi(f(x))\right|, \tag{28}$$

and

$$E_f^P(t^m,\psi) := \left|\frac{1}{p^{mn}} \sum_{x\in P+(t\mathbb{F}_p[t]/(t^m))^n} \psi(f(x))\right|, \tag{29}$$

for any primitive additive character

$$\psi : \mathbb{F}_p[t]/(t^m) \to \mathbb{C}^\times,$$

where primitive means that ψ does not factor through the projection

$$\mathbb{F}_p[t]/(t^m) \to \mathbb{F}_p[t]/(t^{m-1}).$$

The sums of (28), resp. (29), can be rewritten as finite field exponential sums, to which classical bounds like (24) apply. This is done by identifying the summation set with $\mathbb{F}_p^{mn}$, resp. with $\mathbb{F}_p^{(m-1)n}$, namely by sending a polynomial in t to its coefficients, while forgetting the constant terms in the second case.

We first prove the following variant of Proposition 3.7.

Proposition 3.8. *Let f, n, s, and d be as in the introduction. Then there is a constant M (depending only on f_d) such that for all primes p with $p > M$, all integers $m > 0$ with $m \le d+1$, and all primitive additive characters $\psi : \mathbb{F}_p[t]/(t^m) \to \mathbb{C}^\times$ one has*

$$E_f(t^m,\psi) \le p^{-m\cdot\frac{n-s}{d}}. \tag{30}$$

Proof. By a reasoning as for the squarefree case, it is sufficient to treat the case with $s = 0$ for large p, while letting the f_i for $i < d$ vary over homogeneous polynomials in $\mathbb{F}_p[t,x]$ of degree i in x, and while keeping f_d fixed in $\mathbb{Z}[x]$. So, we may assume that $s = 0$, and, by the squarefree case treated above, that $m > 1$. We also may assume that $d \ge 3$ by the above treatment of the case $d = 2$. For each p, let C_p be the set of critical points of the reduction of f modulo p. Since $s = 0$, one has $\#C_p \le c_1$ for some constant c_1 depending only on n and d, see for example the final inequality of [Heath-Brown 1985], or Lemma 5.3 below. Clearly we have

$$E_f(t^m,\psi) = \sum_{P\in C_p} E_f^P(t^m,\psi) \tag{31}$$

for all primes $p > d$, all $m > 1$ and all primitive $\psi : \mathbb{F}_p[t]/(t^m) \to \mathbb{C}^\times$. For $m < d$, note that

$$\frac{1}{p^{mn}}\cdot\#(t\mathbb{F}[\![t]\!]/(t^m))^n = p^{n(m-1)-mn} = p^{-n} < p^{-mn/d}. \tag{32}$$

For $m < d$, we thus find by (31) that

$$E_f(t^m,\psi) \le c_1 p^{-n}, \tag{33}$$

and (30) follows when $m < d$ and p is large enough so that the constant c_1 is eaten by the extra saved power of p coming from (32). We now treat the case that $m = d$. If $p > d$ is such that the reduction of f modulo t is smooth homogeneous of degree d, then $C_p = \{0\}$ and there is nothing left to prove since then $E_f(t^m,\psi) = E_f^{P_0}(t^m,\psi) \le p^{-n} = p^{-mn/d}$, with $P_0 = \{0\}$. If the reduction of f modulo t is not homogeneous of degree d, and, p is larger than d, then there is a constant c_2 (depending only on n and d) such that

$$E_f^P(t^m, \psi) \leq c_2 p^{-n-1/2} \tag{34}$$

for all P in C_p and all primitive ψ. Indeed, this follows from the worst case of (23) applied to $E_f^P(t^m, \psi)$, after rewriting it as a finite field exponential sum as explained just above the proposition. The case $m = d$ for (30) follows, where the constant c_2 is eaten by the extra saved power of p coming from $d \geq 3$ and (34). For $d = m+1$, when we rewrite $E_f^P(t^m, \psi)$ for $P \in C_p$ as a finite field exponential sum over $(m-1)n$ variables running over $\mathbb{F}_p$, we can again apply (23), now in $n(m-1)$ variables and with highest degree part f_d which has singular locus of dimension $n(m-2)$ inside $\mathbb{A}^{n(m-1)}$. Since $d \geq 3$, we again can use the extra saved power of p to obtain (30) and the proposition is proved. □

The transfer principle for bounds from Theorem 3.1 of [Cluckers et al. 2016] can be applied to compare the exponential sums $E_f(p^m, \xi)$ and $E_f(t^m, \psi)$; we will use it in the following basic form. Recall that a Presburger subset A of $\mathbb{N}$ is a Boolean combination of congruence classes and subintervals of $\mathbb{N}$.

Corollary 3.9 [Cluckers et al. 2016]. *Let g be a homogeneous polynomial over $\mathbb{Z}$, of degree $d > 1$ and in n variables. Consider a real number $\sigma > 0$ and a Presburger subset $A \subset \mathbb{N}$. Then the following two statements are equivalent.*

(1) *There exist constants c and M such that for all primes $p > M$ and all polynomials f in $\mathbb{Z}_p[x_1, \ldots, x_n]$ of degree d and with $f_d = g$, one has*

$$E_f(p^m, \xi) \leq c p^{-\sigma m}$$

for all $m \in A$ and all primitive p^m-th roots of unity ξ.

(2) *There exist constants c' and M' such that for all primes $p > M'$ and all polynomials f in $\mathbb{F}_p[[t]][x_1, \ldots, x_n]$ of degree d and such that $f_d = g \mod (p)$ holds in $\mathbb{F}_p[x]$, one has*

$$E_f(t^m, \psi) \leq c' p^{-\sigma m}$$

for all $m \in A$ and all primitive characters $\psi : \mathbb{F}_p[t]/(t^m) \to \mathbb{C}^\times$.

In the corollary, we have extended the notation E_f to more general f, namely with more general coefficients than just in $\mathbb{Z}$, in the obvious way. More generally than Corollary 3.9, Theorem 3.1 of [Cluckers et al. 2016] allows to transfer bounds that hold for motivic families of functions, and, the families in the corollary are a special case of such family.

Proof of Corollary 3.9. Clearly the left hand sides and of the right hand sides of the inequalities come from motivic functions H and G as required in Theorem 3.1 of [Cluckers et al. 2016]. Now the corollary readily follows from the conclusion of Theorem 3.1 of [Cluckers et al. 2016] for such H and G. □

Proof of Proposition 3.7. We show for all large primes p, all integers $m > 0$ with $m \leq d+1$, and all primitive p^m-th roots of 1, that

$$E_f(p^m, \xi) \leq p^{-m \cdot \frac{n-s}{d}}, \tag{35}$$

where moreover the lower bound on p depends only on f_d. The case $d = 2$ is already shown. For $m \neq d \geq 3$ this follows at once from Corollary 3.9 and the corresponding extra power savings when $m \neq d$ in the proof of Proposition 3.8. Indeed, the transfer principle holds uniformly in f as long as f_d and n are fixed, since these bounds (with the extra power savings) from the proof of Proposition 3.8 depend only on f_d and n. Let us now treat the remaining case of $m = d$. It is again enough to treat the case $s = 0$. For a prime $p > d$ such that the reduction of f modulo p is not smooth homogeneous of degree d, we are done similarly by the transfer principle for bounds from [Cluckers et al. 2016] and the corresponding power savings in the proof Proposition 3.8. If $m = d$ and $p > d$ is such that the reduction of f modulo p is smooth homogeneous of degree d, then we have that $P_0 = \{0\}$ is the unique critical point of the reduction of f modulo p, and thus

$$E_f(p^m, \xi) = E_f^{P_0}(p^m, \xi) \leq p^{-n} = p^{-mn/d}.$$

The proof of Proposition 3.7 is thus finished. □

4. The nondegenerate case

In this section we show that Conjecture 1 with $\varepsilon = 0$ holds for nondegenerate polynomials, where the nondegeneracy condition is with respect to the Newton polyhedron of $f - f(0)$ at zero as in [Castryck and Nguyen 2019] (which is slightly different than the nondegeneracy notion of [Kouchnirenko 1976; Varčenko 1976]). The nondegeneracy condition generalizes the situation where f is a sum of monomials in separate variables, like $x_1x_2 + x_3x_4$. In detail, writing $f(x) - f(0) = \sum_{i \in \mathbb{N}^n} \beta_i x^i$ in multi-index notation, let

$$\mathrm{Supp}_f := \{i \in \mathbb{N}^n \mid \beta_i \neq 0\}$$

be the support of $f - f(0)$. Further, let

$$\Delta_0(f) := \mathrm{Conv}(\mathrm{Supp}_f + (\mathbb{R}_{\geq 0})^n)$$

be the convex hull of the sum-set of Supp_f with $(\mathbb{R}_{\geq 0})^n$ where $\mathbb{R}_{\geq 0}$ is $\{x \in \mathbb{R} \mid x \geq 0\}$. The set $\Delta_0(f)$ is called the Newton polyhedron of $f - f(0)$ at zero. Let σ_f be the unique real value such that $(1/\sigma_f, \ldots, 1/\sigma_f)$ is contained in a proper face of $\Delta_0(f)$. Further, let κ denote the maximal codimension in $\mathbb{R}^n$ of τ when τ varies over the faces of $\Delta_0(f)$ containing $(1/\sigma_f, \ldots, 1/\sigma_f)$. For each face τ of the polyhedron $\Delta_0(f)$, consider the polynomial

$$f_\tau := \sum_{i \in \tau} \beta_i x^i.$$

Call f nondegenerate with respect to $\Delta_0(f)$ when for each face τ of $\Delta_0(f)$ and each critical point P of $f_\tau : \mathbb{C}^n \to \mathbb{C}$, at least one coordinate of P equals zero. Recall that a complex point $P \in \mathbb{C}^n$ is called a critical point of f_τ if $\partial f_\tau / \partial x_i(P) = 0$ for all $i = 1, \ldots, n$.

The following proposition slightly extends the main result of [Castryck and Nguyen 2019] as it covers small primes as well. Note that [Castryck and Nguyen 2019, Theorem 1.4.1] gives evidence for Igusa's conjecture on exponential sums in the variant of [Cluckers and Veys 2016, Conjecture 1.2].

Proposition 4.1 [Castryck and Nguyen 2019, Theorem 1.4.1]. *Suppose that f is nondegenerate with respect to $\Delta_0(f)$. Then, there is a constant c such that for all primes p, all integers $m \geq 2$ and all primitive p^m-th roots of unity ξ one has*

$$E_f(p^m, \xi) \leq cp^{-m\sigma_f} m^{\kappa-1}. \tag{36}$$

From Proposition 4.1 we will derive the following evidence for Conjecture 1.

Theorem 4.2. *Let f, n, s, and d be as in the introduction. Suppose that f is nondegenerate with respect to $\Delta_0(f)$. Then Conjecture 1 with $\varepsilon = 0$ holds for f. Namely, there is c such that for all integers $N > 0$ and all primitive N-th roots of unity ξ, one has*

$$E_f(N, \xi) \leq cN^{-\frac{n-s}{d}}.$$

Furthermore, for all large primes p (with "large" depending on f), all integers $m > 0$ and all primitive p^m-th roots of unity ξ one has

$$E_f(p^m, \xi) \leq p^{-m\frac{n-s}{d}}.$$

Proof of Proposition 4.1. In [Castryck and Nguyen 2019] it is shown that one can take a constant c so that (36) holds for all large primes p and all $m \geq 2$. So, there is only left to prove that for each prime p there is a constant c_p (depending on p) such that for each integer $m \geq 2$ one has

$$E_f(p^m, \xi) \leq c_p p^{-m\sigma_f} m^{\kappa-1}. \tag{37}$$

Indeed, (37) is only used for the finitely many remaining primes. First, if f is nondegenerate with respect to $\Delta_0(f)$ we show that $f(0)$ is the only possible critical value of f, by induction on n. If $n = 1$, by the nondegeneracy of f, we get that f has no critical point in $\mathbb{C}^\times$ and we are done. Now suppose that $n > 1$. Let f be a polynomial in n variables which is nondegenerate with respect to $\Delta_0(f)$. Suppose that $u = (u_1, \dots, u_n)$ is a critical point of f. By the nondegeneracy of f there exists $1 \leq j \leq n$ such that $u_j = 0$. Without lost of generality we can suppose that $j = n$. We can write $f = f(0) + \sum_{i=0}^{d} g_i(x_1, \dots, x_{n-1})x_n^i$ for some polynomials g_i with furthermore $g_0(0) = 0$. It is sufficient to show that $f(u_1, \dots, u_n) = f(0)$. Since $u_n = 0$, it suffices to show that $g_0(u_1, \dots, u_{n-1}) = 0$. By the fact that f is nondegenerate with respect to $\Delta_0(f)$ we get that g_0 is nondegenerate with respect to $\Delta_0(g_0)$. It is clear that $(u_1, \dots, u_{n-1})$ is a critical point of g_0. So, we can use the inductive hypothesis to deduce that $g_0(u_1, \dots, u_{n-1}) = g_0(0) = 0$. Now, since f has no other possible critical value than $f(0)$ and since there exists a toric log resolution of $f - f(0)$ whose numerical properties (in particular its discrepancy numbers) are controlled by $\Delta_0(f)$ (see for example [Varčenko 1976]), inequality (37) follows from Igusa's work [1978]. Here, we use the following information on the discrepancy numbers coming from the toric log resolution π of $f - f(0)$, in relation to $\Delta_0(f)$. If E is an irreducible component of the exceptional locus of π and if one writes N_E for the multiplicity of E in the divisor $\pi^*(f - f(0))$ and $\nu_E - 1$ for the multiplicity of E in $\pi^*(dx_1 \wedge \cdots \wedge dx_n)$, then one has $\nu_E/N_E \geq \sigma$. Furthermore, any intersection of $\kappa + 1$ many such E for which the equality $\nu_E/N_E = \sigma$ holds is empty. Since $f(0)$ is the only critical value of f, we are now done by Igusa's work [1978]. $\square$

The proof of Theorem 4.2 relies on Proposition 4.1 and Lemma 4.3. Note that the following Lemmas 4.3 and 4.4 do not require f to be nondegenerate.

Lemma 4.3. *Let f, n, s, and d be as in the introduction.* (*In particular*, *f is allowed to be inhomogeneous and there is no condition on nondegeneracy.*) *Suppose that $d \geq 3$. Then one has $\sigma_f \geq (n-s)/d$, and, equality holds if and only if there is a smooth form g of degree d in $n-s$ variables such that*

$$f(x) - f(0) = g(x_{i_1}, \ldots, x_{i_{n-s}})$$

for some i_j with $1 \leq i_1 < i_2 < \ldots < i_{n-s} \leq n$.

We will first prove Lemma 4.3 in the case that $s = 0$, using the following lemma. We write $\mathrm{Conv}(\mathrm{Supp}_f)$ for the convex hull of Supp_f in $\mathbb{R}^n$.

Lemma 4.4. *Let f, n, s, and d be as in the introduction. Suppose furthermore that $d \geq 3$, $s = 0$ and that $f = f_d$, namely, f is smooth homogeneous of degree at least* 3. *Then*

$$\dim(\mathrm{Conv}(\mathrm{Supp}_f)) = n - 1,$$

and, *the point $(d/n, \ldots, d/n)$ belongs to the interior of* $\mathrm{Conv}(\mathrm{Supp}_f)$. *In particular*, *$\sigma_f = n/d$ and $\kappa = 1$.*

Proof. Since $f = f_d$, it is clear that

$$\dim(\mathrm{Conv}(\mathrm{Supp}_f)) \leq n - 1.$$

Suppose now that either $\dim(\mathrm{Conv}(\mathrm{Supp}_f)) < n - 1$, or, that $(d/n, \ldots, d/n)$ does not belong to the interior of $\mathrm{Conv}(\mathrm{Supp}_f)$. We try to find a contradiction. By our assumptions, there exists a hyperplane $H = \{a \in \mathbb{R}^n \mid k \cdot a = 0\}$ for some $k \in \mathbb{R}^n \setminus \{0\}$ such that the point $(d/n, \ldots, d/n)$ belongs to H and such that Supp_f belongs to the half space $H_+ := \{a \in \mathbb{R}^n \mid k \cdot a \geq 0\}$. Let I be the subset of $\{1, \ldots, n\}$ consisting of i with $k_i > 0$ and let J be the subset of $\{1, \ldots, n\}$ consisting of j with $k_j < 0$. Clearly I and J are disjoint. Since $(d/n, \ldots, d/n)$ belongs to H, it follows that I and J are both nonempty and that

$$\sum_{i \in I} k_i = \sum_{j \in J} |k_j|. \tag{38}$$

Furthermore, the inclusion $\mathrm{Supp}_f \subset H_+$ implies that

$$\sum_{i \in I} k_i a_i \geq \sum_{j \in J} |k_j| a_j \text{ for all } a \in \mathrm{Supp}_f. \tag{39}$$

Consider the set Supp_f^0 consisting of those $a \in \mathrm{Supp}_f$ with moreover $a_i = 1$ for some $i \in I$ and $a_{i'} = 0$ for all $i' \notin J \cup \{i\}$. For $a \in \mathrm{Supp}_f^0$ write $t(a)$ for the unique $i \in I$ with $a_i = 1$ and write

$$I_0 := \{i \in I \mid \exists a \in \mathrm{Supp}_f^0 \text{ with } t(a) = i\}.$$

Clearly we can write

$$f(x) = \sum_{i\in I_0} x_i g_i(x_j)_{j\in J} + \sum_{a\in \mathrm{Supp}_f \setminus \mathrm{Supp}_f^0} \beta_a x^a$$

for some polynomials g_i in the variables $(x_j)_{j\in J}$. Also, the algebraic set

$$\bigcap_{i\notin J}\{x_i = 0\} \bigcap_{i\in I_0}\{g_i = 0\}$$

in $\mathbb{A}^n_{\mathbb{C}}$ has dimension at least $|J| - |I_0|$ and is contained in Crit_f, the critical locus of f. By our smoothness condition $s = 0$, this implies

$$|I_0| \geq |J|. \tag{40}$$

Hence, we can write $I_0 = \{i_1, \ldots, i_\ell\}$ and $J = \{j_1, \ldots, j_m\}$ with $m \leq \ell$ and with

$$k_{i_1} \geq k_{i_2} \geq \cdots \geq k_{i_\ell} \quad \text{and} \quad |k_{j_1}| \geq |k_{j_2}| \geq \cdots \geq |k_{j_m}|. \tag{41}$$

To prove the lemma it is now sufficient to show that

$$k_{i_r} > |k_{j_r}| \quad \text{for all } r \text{ with } 1 \leq r \leq m. \tag{42}$$

Indeed, (42) gives a contradiction with (38). To prove (42), we suppose that there is r_0 with $1 \leq r_0 \leq m$ and with

$$k_{i_{r_0}} \leq |k_{j_{r_0}}| \tag{43}$$

and we need to find a contradiction. If there exists $a \in \mathrm{Supp}_f^0$ such that $a_{j_{r_1}} \geq 1$ for some $r_1 \leq r_0$, then let a be such an element and let t be $t(a)$; otherwise, let a be arbitrary and put $t = 0$. We will now show that $t < r_0$. If $t = 0$ this is clear, so, suppose that $t > 0$. Since $d \geq 3$ and $a \in \mathrm{Supp}_f^0$, we find by (39) and (41) that

$$k_{i_t} = \sum_{i\in I} k_i a_i \geq \sum_{j\in J} a_j |k_j| > |k_{j_{r_1}}| \geq |k_{j_{r_0}}|. \tag{44}$$

Together with (41) and (43), this implies that $t < r_0$ as desired. We can thus write

$$f = \sum_{1\leq \ell \leq r_0 - 1} x_{i_\ell} h_\ell(x_j)_{j\in J} + \sum_{a\in A} \beta_a x^a \tag{45}$$

with

$$A = \left\{ a \in \mathrm{Supp}_f \;\middle|\; \sum_{i\notin\{j_1,\ldots,j_{r_0}\}} a_i \geq 2 \right\}$$

and with some polynomials h_ℓ in the variables $(x_j)_{j\in J}$. It follows that the algebraic set

$$\bigcap_{i\notin\{j_1,\ldots,j_{r_0}\}} \{x_i = 0\} \bigcap_{1\leq \ell\leq r_0-1} \{h_\ell = 0\}$$

has dimension at least 1 and is contained in Crit_f, again a contradiction with our smoothness assumption $s = 0$. So, relation (42) follows and the lemma is proved. □

The case of Lemma 4.3 with $s = 0$ is derived from Lemma 4.4, as follows.

Proof of Lemma 4.3 with $s = 0$. Let f be of degree $d \geq 3$ and with $s = 0$. We need to show that $\sigma_f \geq n/d$, and, that $\sigma_f = n/d$ if and only if $f = f_d$. Since f_d is smooth, Lemma 4.4 implies that $(d/n, \dots, d/n)$ belongs to $\Delta_0(f)$, and hence, $\sigma_f \geq n/d$. Suppose now that $f \neq f_d$. Then there exists $a \in \mathrm{Supp}_f$ with $\sum_{i=1}^n a_i < d$. Hence, by Lemma 4.4 and the definition of $\Delta_0(f)$, there exists $\varepsilon > 0$ such that

$$\{x \in \mathbb{R}^n \mid \|x - (d/n, \dots, d/n)\| \leq \varepsilon\} \subset \Delta_0(f).$$

Therefore it is clear that $\sigma_f > n/d$. This finishes the proof of Lemma 4.3 with $s = 0$. □

Proof of Lemma 4.3 with $s > 0$. To prove the lemma with $s > 0$ we may suppose that

$$\sigma_f \leq (n - s)/d. \tag{46}$$

By the definition of σ_f we have

$$\min_{a \in \mathrm{Conv}(\mathrm{Supp}_f)} \max(a) = 1/\sigma_f, \tag{47}$$

where $\max(a) = \max_{1 \leq i \leq n}\{a_i\}$ and where $\mathrm{Conv}(\mathrm{Supp}_f)$ is the convex hull of Supp_f. We set

$$k := \min_{\max(a) = 1/\sigma_f} \#\{i \mid a_i = 1/\sigma_f\},$$

where the minimum is taken over $a \in \mathrm{Conv}(\mathrm{Supp}_f)$. Let $a \in \mathrm{Conv}(\mathrm{Supp}_f)$ realize this minimum, namely, with $\#\{i \mid a_i = 1/\sigma\} = k$ and with $\max(a) = 1/\sigma_f$. We may suppose that

$$a_1 = \cdots = a_k = 1/\sigma_f \quad \text{and} \quad a_i < 1/\sigma_f \ \text{ if } i > k.$$

Let $b \in \mathrm{Conv}(\mathrm{Supp}_f)$ be such that $\max(b) = 1/\sigma_f$. Then, for each $\lambda \in [0, 1]$, the point $c_\lambda := \lambda a + (1 - \lambda) b$ lies in $\mathrm{Conv}(\mathrm{Supp}_f)$. When λ is sufficiently close to 1, then we have $c_{\lambda,i} < 1/\sigma_f$ for all $i > k$, and, the definition of k implies that $b_i = 1/\sigma_f$ for all $1 \leq i \leq k$. By the same reasoning, for each $b \in \mathrm{Conv}(\mathrm{Supp}_f)$ one has $b_i \geq 1/\sigma_f$ for some i with $1 \leq i \leq k$. The definition of k and (47) also tell us that $k/\sigma_f \leq d$, and thus we find

$$k \leq n - s \tag{48}$$

from (46). For any tuple of complex numbers $C = (c_{i,j})_{1 \leq i,j \leq s}$ we consider the polynomial

$$g_C = f\Bigg(x_1, \dots, x_{n-s}, x_{n-s+1} + \sum_{1 \leq j \leq n-s} c_{1,j} x_j, \dots, x_n + \sum_{1 \leq j \leq n-s} c_{s,j} x_j\Bigg).$$

For a generic choice of C one has $\mathrm{Supp}_f \subset \mathrm{Supp}_{g_C}$. Furthermore, we show that for a generic choice of C the polynomial

$$h_C = f_d\Bigg(x_1, \dots, x_{n-s}, \sum_{1 \leq j \leq n-s} c_{1,j} x_j, \dots, \sum_{1 \leq j \leq n-s} c_{s,j} x_j\Bigg)$$

is smooth homogeneous in $n - s$ variables, where f_d is the degree d homogeneous part of f. For a generic choice of $e_n = (e_{n,i})_{i<n}$ one has

$$\dim(\mathrm{Sing}(f_{d,e_n})) = n - s,$$

where

$$f_{d,e_n}(x_1,\dots,x_{n-1}) := f_d\left(x_1,\dots,x_{n-1},\sum_{i=1}^{n-1} e_{n,i}x_i\right),$$

considered as a polynomial in $n-1$ variables x_i with $i<n$. We repeat this argument to see that for a generic choice $E=(e_{n-s+1},\dots,e_n)$ with $e_j=(e_{j,i})_{i<j}$ one has that

$$\dim(\mathrm{Sing}(f_d|_{V_E})) = n-s,$$

where

$$V_E = \left\{x \;\middle|\; x_n = \sum_{i<n} e_{n,i}x_i,\dots,x_{n-s+1} = \sum_{i\le n-s} e_{n-s+1,i}x_i\right\}.$$

It is clear that the smoothness of $f_d|_{V_E}$ for generic E corresponds to the smoothness of h_C for generic C. Let us fix such a choice of C with all these properties, namely, that h_C is smooth and that $\mathrm{Supp}_f \subset \mathrm{Supp}_{g_C}$. If $a\in \mathrm{Supp}_{g_C}$, it is easy to see that $a_i \ge b_i$ for all i with $1\le i\le n-s$ and for some $b\in \mathrm{Supp}_f$. Hence, $\sigma_{g_C}\le \sigma_f$, by the definition of k and our chosen ordering of the coordinates. On the other hand, from $\mathrm{Supp}_f\subset \mathrm{Supp}_{g_C}$ it follows that $\sigma_{g_C}\ge\sigma_f$, and hence, we have

$$\sigma_{g_C}=\sigma_f.$$

Let π be the coordinate projection from $\mathbb{R}^n$ to $\mathbb{R}^{n-s}$. Then, for any $e=(e_j)_{j=1,\dots,s}$, consider the polynomial

$$g_{C,e}(x_1,\dots,x_{n-s}) := g_C(x_1,\dots,x_{n-s},e_1,\dots,e_s).$$

Then, for generic choice of e, we have

$$\mathrm{Supp}_{g_{C,e}} = \pi(\mathrm{Supp}_{g_C}).$$

Let us fix such a choice of e. It is clear that

$$\sigma_{g_{C,e}} = \sigma_{g_C},$$

by the definition of k and our ordering of the coordinates. Note that the highest degree homogeneous part of $g_{C,e}$ equals h_C, which is smooth. Thus, we can use Lemma 4.3 with $s=0$ (which is already proved) for $g_{C,e}$. So, we find

$$\sigma_{g_{C,e}} = (n-s)/d \quad\text{and}\quad g_{C,e}-g_{C,e}(0)=h_C.$$

Hence,

$$\pi(\mathrm{Supp}_f)\subset \pi(\mathrm{Supp}_{g_C})\subset \{a\in\mathbb{R}^{n-s} \mid a_1+\cdots+a_{n-s}=d\}.$$

This holds if and only if $f-f(0)=f_d=h(x_1,\dots,x_{n-s})$ for some polynomial h, which is smooth homogeneous since $\dim(\mathrm{Crit}_{f_d})=s$. This finishes the proof of the Lemma 4.3. $\square$

We are now ready to prove Theorem 4.2.

Proof of Theorem 4.2. The case that $d=2$ is treated in Section 3.3. Hence, we may suppose that $d \geq 3$. By Proposition 4.1, there exists a constant c_2 such that for all integers $m > 1$, all primes p and all primitive p^m-th roots of unity ξ we have

$$E_f(p^m, \xi) \leq c_2 p^{-m\sigma_f} m^{\kappa-1}. \tag{49}$$

By Lemma 4.3 we have $\sigma_f \geq (n-s)/d$. If $\sigma_f > (n-s)/d$, then we have $(n-s)/d < \frac{1}{2}(n-s)$, from using $d \geq 3$ and $s < n$. Conjecture 1 for this case follows by combining (7) and (49) with the squarefree case from Section 3.5 . If $\sigma_f = (n-s)/d$, we use Lemma 4.3 again to see that $f = g_d + f(0)$ for a smooth form g_d of degree d in $n-s$ variables. Conjecture 1 for this case follows by Igusa's case from Section 3.1. □

Remark 4.5. If f is weighted homogeneous, then the notion of nondegenericity with respect to $\Delta_0(f)$ is generic, but otherwise the genericity is more subtle, by the difference between "critical points" and "singular points". In fact, whether or not the notion of nondegenericity with respect to $\Delta_0(f)$ is generic depends on Supp_f. When Supp_f is contained in a hyperplane which does not contain the origin 0 and has a normal vector with nonnegative coordinates (see [Castryck and Nguyen 2019, Section 2.2]), then the condition of nondegeneracy on the coefficients β_i is generic within this support, that is, for any γ outside a Zariski closed subset of $\mathbb{C}^{\mathrm{Supp}_f}$, the polynomial $\sum_{i \in \mathrm{Supp}_f} \gamma_i x^i$ is nondegenerate with respect to its Newton polyhedron at zero. This hyperplane condition generalizes the case of weighted homogeneous polynomials. However, in the general case, this genericity may be lost since we imposed conditions on critical points of f_τ instead of on singular points as is done more traditionally in [Kouchnirenko 1976], [Varčenko 1976]. Especially for $\tau = \Delta_0(f)$ this makes a difference when the mentioned hyperplane condition is not met. For instance, polynomials of the form $f(x) = ax^3 + by^3 + cxy$ for nonzero a, b, and c are never nondegenerate in our sense, the problem being with $\tau = \Delta_0(f)$.

5. The four variable case

In this final section we prove Conjecture 1 when $n \leq 4$ (Theorem 5.1), and a slightly stronger result when furthermore $d \leq 3$ and $s = 0$ (Proposition 5.2).

Theorem 5.1. *Let f, n, s, and d be as in the introduction and suppose that $n \leq 4$. Then Conjecture 1 holds for f.*

The proof of Theorem 5.1 relies on a concrete lemma inspired by Weierstrass preparation (see Lemma 5.4), properties of $\hat{\alpha}_f$ based on results on minimal exponents from [Mustaţă and Popa 2020], bounds from [Cluckers et al. 2019; Nguyen and Veys 2022], and Igusa's results as summarized in [Denef 1991].

Proof of Theorem 5.1. Suppose that $n \leq 4$. If $(n-s)/d \leq 1$ or $d = 2$, then Conjecture 1 follows by the arguments in Sections 3.2 and 3.3. Hence, we may concentrate on the case that $d = 3$ and $s = 0$, but this follows from Proposition 5.2 below, the squarefree case from Section 3.5, the inequality (18) and the relation (7). □

The rest of this section will focus on the proof of Proposition 5.2, which says slightly more than Theorem 5.1 in the case that $d \le 3$ and $s = 0$.

Proposition 5.2. *Let f, n, s, and d be as in the introduction and suppose that $n \le 4$, $d \le 3$, and $s = 0$. Then for all $\varepsilon > 0$ there is a constant c such that*

$$E_f(p^m, \xi) \le c(p^m)^{-\hat{\alpha}_f+\varepsilon} \text{ for all primes } p, \text{ all } m > 1 \text{ and all } \xi, \tag{50}$$

with $\hat{\alpha}_f$ as in Section 2.7. Furthermore, the value $\hat{\alpha}_f$ is equal to the motivic oscillation index of f as given in [Cluckers et al. 2019]. Hence, $\hat{\alpha}_f$ is the optimal exponent in (50).

The optimality of the exponent $\hat{\alpha}_f$ in (50) means that there is a constant $c_0 > 0$ such that for infinitely many primes p and (some) arbitrarily large m one has

$$c_0(p^m)^{-\hat{\alpha}_f} \le E_f(p^m, \xi) \text{ for some } \xi. \tag{51}$$

The motivic oscillation index of f as given in [Cluckers et al. 2019] (which corresponds to the one from [Cluckers 2008b] but without the negative sign) is the optimal exponent of p^{-m} in (50), see Section 3.4 of [Cluckers et al. 2019]; therefore, the equality of $\hat{\alpha}_f$ with the motivic oscillation index is a useful property and implies (51).

The following auxiliary lemma is well known, see for example the final inequality of [Heath-Brown 1985], where furthermore an explicit upper bound on the number of critical points is obtained.

Lemma 5.3. *Suppose that $g = g_0 + \cdots + g_d$ is a polynomial in $\mathbb{C}[x_1, \ldots, x_n]$ of degree d and with $\dim(\mathrm{Crit}_{g_d}) = 0$, where g_i is the degree i homogeneous part of g, and where Crit_{g_d} is the critical locus of $g_d : \mathbb{C}^n \to \mathbb{C}$, i.e., the scheme associated with the ideal generated by the $\partial g_d/\partial x_i$ for $1 \le i \le n$. Then $\mathrm{Crit}_g(\mathbb{C})$ is a finite set.*

Proof. This is shown by homogenizing g as in the reasoning towards the final inequality of [Heath-Brown 1985], where it is even shown that $\#\mathrm{Crit}_g(\mathbb{C}) \le (d-1)^n$, by an application of Bézout's theorem. □

Lemma 5.4. *Let f in $\in \mathbb{Z}[x_1, x_2, x_3, x_4]$ be of degree $d = 3$ and with $s = 0$. Suppose that one can write $f = f_2 + f_3$ with f_i homogeneous of degree i and with $f_2 \ne 0$. Then there exists a finite field extension K of $\mathbb{Q}$ and a linear transformation $x_i = \sum_{j=1}^4 a_{ij} y_j$ with $(a_{ij})_{1 \le i,j \le 4} \in \mathrm{GL}_4(K)$ such that*

$$f(x_1, x_2, x_3, x_4) = g(y_1, y_2, y_3, y_4) \quad \textit{with} \quad \frac{\partial^2 g}{\partial y_1^2}(0,0,0,0) \ne 0 \quad \textit{and} \quad \frac{\partial^3 g}{\partial y_1^3}(0,0,0,0) = 0.$$

Proof. By a simple calculation we have

$$\frac{\partial^2 g}{\partial y_1^2}(0,0,0,0) = 2f_2(a_{11}, a_{21}, a_{31}, a_{41}) \quad \text{and} \quad \frac{\partial^3 g}{\partial y_1^3}(0,0,0,0) = 6f_3(a_{11}, a_{21}, a_{31}, a_{41}).$$

So, it suffices to show the following relation on zero loci

$$Z(f_3(a_{11}, a_{21}, a_{31}, a_{41})) \not\subseteq Z(f_2(a_{11}, a_{21}, a_{31}, a_{41})) \cup Z(\det(A)) \tag{52}$$

viewed as algebraic subsets of $\mathbb{A}^{16}_{\overline{\mathbb{Q}}}$ and with $A=(a_{ij})_{1\le i,j\le 4}$ and $\overline{\mathbb{Q}}$ the algebraic closure of $\mathbb{Q}$. Since $s=\dim(\mathrm{Sing}(f_3))=0$ one has that f_3 is absolutely irreducible. Therefore if (52) does not hold then

$$f_3(a_{11},a_{21},a_{31},a_{41})\,|\,f_2(a_{11},a_{21},a_{31},a_{41})\quad\text{or}\quad f_3(a_{11},a_{21},a_{31},a_{41})\,|\det(A).$$

It is clear that $f_3\nmid f_2$ since $f_2\neq 0$. Also, the polynomial $\det(A)$ is absolutely irreducible of degree 4. So, (52) must hold. The lemma is proved. □

We first prove the part of Proposition 5.2 for large primes, as follows.

Proposition 5.5. *With notation and assumptions from Proposition 5.2, there exist an integer M and a constant c such that for all $p>M$ all $m>1$, for all primitive p^m-th roots of unity ξ we have*

$$E_f(p^m,\xi)\le cm^3p^{-m\hat{\alpha}_f}. \tag{53}$$

Proof. We may focus on the case $d=3$. By Lemma 5.3, the set $\mathrm{Crit}_f(\overline{\mathbb{Q}})$ is finite. Because of [Denef 1991, Remark 4.5.3], we are done when f has no critical point. Similarly, when f has a critical point a such that $f-f(a)$ has multiplicity 3 at a the proposition follows by Section 3.1 and the fact that $\hat{\alpha}_f=\frac{4}{3}$ in this case. Now we suppose that if a is a critical point of f then $f-f(a)$ has multiplicity 2 at a. By [Denef 1991, Remark 4.5.3], it suffices to show that if $a\in\mathrm{Crit}_f(\overline{\mathbb{Q}})$ then there exist an integer M and a constant c such that for all $p>M$ with $a\in\mathbb{Z}_p^4$, all $m>1$, all primitive p^m-th roots of unity ξ we have

$$E_f^{(a)}(p^m,\xi):=\int_{a+p\mathbb{Z}_p^4}\xi^{f(x)\bmod p^m}\mu\le cm^3p^{-m\hat{\alpha}_f}. \tag{54}$$

Without loss of generality, we can suppose that $a=0$ and $f(0)=0$. Lemma 5.4 gives us a finite field extension K of $\mathbb{Q}$ and a linear transformation $x_i=\sum_{1\le j\le 4}a_{ij}y_j$ with $(a_{ij})_{1\le i,j\le 4}\in\mathrm{GL}_4(K)$ such that if $g(y_1,y_2,y_3,y_4)=f(x_1,x_2,x_3,x_4)$ then

$$g(y_1,y_2,y_3,y_4)=h_2(y_2,y_3,y_4)y_1^2+h_1(y_2,y_3,y_4)y_1+h_0(y_2,y_3,y_4)$$

with polynomials $h_0,h_1,h_2\in K(y_2,y_3,y_4)$ and $h_2(0,0,0)\neq 0,\ h_1(0,0,0)=h_0(0,0,0)=0$. We set

$$z_1=h_2(y_2,y_3,y_4)y_1+\tfrac{1}{2}h_1(y_1,y_3,y_4)$$

and $z_2=y_2, z_3=y_3, z_4=y_4$. At the new coordinates (z_1,z_2,z_3,z_4) we have $g(y_1,y_2,y_3,y_4)=h(z_1,z_2,z_3,z_4)$ with

$$h(z_1,z_2,z_3,z_4)h_2(z_2,z_3,z_4)=z_1^2+r(z_2,z_3,z_4)$$

and

$$r(z_2,z_3,z_4)=h_2(z_2,z_3,z_4)h_0(z_2,z_3,z_4)-\tfrac{1}{4}h_1^2(z_2,z_3,z_4).$$

Using Lemma 5.3 and the argument from [Nguyen and Veys 2022, Proposition 5.9 and Section 6] (to compare between the weights of suitable ℓ-adic cohomology groups related to f and those related to

z_1^2+r), it now follows that there exist an integer M and a constant c such that for all $p > M$ all $m > 1$, all primitive p^m-th roots of unity ξ we have

$$E_f^{(0)}(p^m,\xi) \le cm^3 p^{-m\hat{\alpha}_f}. \tag{55}$$

So we are done. □

By Proposition 5.5, in order to prove Proposition 5.2, it remains to prove the following proposition.

Proposition 5.6. *Let f be in $\mathbb{Z}[x_1,x_2,x_3,x_4]$ of degree $d=3$ and with $s=0$. Then, for each p and $\varepsilon>0$ there exist a constant $c_{p,\varepsilon}$ and an integer m_p such that for all integers $m>m_p$ and all primitive p^m-th roots of unity ξ we have*

$$E_f(p^m,\xi) \le c_{p,\varepsilon} p^{-m(\hat{\alpha}_f-\varepsilon)} \tag{56}$$

Proof. By Lemma 5.3, the set $\mathrm{Crit}_f(\mathbb{Z}_p)$ is finite for each p. If there exists a point $a\in\mathrm{Crit}_f(\mathbb{Z}_p)$ such that the multiplicity of f at a is 3 then we are done as in Proposition 5.5. If $\mathrm{Crit}_f(\mathbb{Z}_p)=\varnothing$ then by a basic argument there exists an integer m_p such that we have $E_f(p^m,\xi)=0$ for all $m\ge m_p$ and all primitive p^m-th roots of unity ξ.

Let us now suppose that $\mathrm{Crit}_f(\mathbb{Z}_p)\neq\varnothing$ and that the multiplicity of $f-f(a)$ at a is 2 for all $a\in\mathrm{Crit}_f(\mathbb{Z}_p)$. By [Igusa 1974; 1978; Denef and Veys 1995], it is sufficient to show that the real part of every nontrivial pole of the Igusa local zeta functions of f at U_a is at most $-\hat{\alpha}_f$, where U_a is a small enough neighborhood of a in $\mathbb{Z}_p^4$. Here, we recall that if L is a finite extension of $\mathbb{Q}_p$, $\mathcal{O}_L$ is the ring of integers of L, V is an open subset of $\mathcal{O}_L^n$, F is an analytic function defined on a neighborhood of V and χ is a multiplicative character of $\mathcal{O}_L^\times$ then the twisted Igusa local zeta function of F at V associated with χ given by

$$Z_\chi(L,V,F,s)=\int_V \chi\big(\mathrm{ac}(F(x))\big)|F(x)|^s\,|dx|,$$

where $|dx|$ is the normalized Haar measure on L^n so that the measure of $\mathcal{O}_L^n$ is 1, $\mathrm{ac}(z)=z\varpi_L^{-\mathrm{ord}_L(z)}$ for nonzero z and for a fixed uniformizing element ϖ_L of $\mathcal{O}_L$ and the usual valuation map $\mathrm{ord}_L : L\to\mathbb{Z}\cup\{+\infty\}$ (see, e.g., [Denef 1991; Veys and Zúñiga Galindo 2008]). We say that s is a nontrivial pole of the Igusa local functions of F at V if s is a pole of $Z_\chi(L,V,F,s)$ when $\chi\neq\chi_{\text{trivial}}$ or s is a pole of $(p^{s+1}-1)Z_\chi(L,V,F,s)$ when $\chi=\chi_{\text{trivial}}$. Let L be a finite extension of $\mathbb{Q}_p$, V be an open subset of $\mathcal{O}_L^n$, F be an analytic function on V and denote by $\mathrm{Pol}_V(F)$ the set of the real parts of the nontrivial poles of the Igusa local zeta functions of F at V. By [Igusa 1974; Igusa 1978; Denef and Veys 1995; Veys and Zúñiga Galindo 2008], $\mathrm{Pol}_V(F)$ is a finite set. To prove the proposition, it thus remains to show for all $a\in\mathrm{Crit}_f(\mathbb{Z}_p)$, all small enough neighborhoods U_a of a in $\mathbb{Z}_p^4$, that

$$\alpha\le-\hat{\alpha}_f \text{ for all } \alpha\in\mathrm{Pol}_{U_a}(f-f(a)). \tag{57}$$

The rest of the proof will show this (57).

Fix $a\in\mathrm{Crit}_f(\mathbb{Z}_p)$; to simplify the notation we suppose that $a=(0,0,0,0)$ and $f(0)=0$. Up to using a transformation as in Lemma 3.4, we may suppose that f_2 is diagonal, where we write $f=f_2+f_3$ with f_i homogeneous of degree i for each i. Suppose that the coefficient of x_1^2 is nonzero in f_2. Using

Weierstrass preparation, we can suppose that for a small enough open (p-adic) neighborhood $U = p^{m_0}\mathbb{Z}_p^4$ of 0, there exist analytic functions u, g_1, g_0 on U such that we have

$$f|_U = u(x)(x_1^2 + 2g_1(x_2, x_3, x_4)x_1 + g_0(x_2, x_3, x_4))$$

and $g_1(0, 0, 0) = g_0(0, 0, 0) = 0 \neq u(0)$, with $x = (x_1, x_2, x_3, x_4)$. By shrinking U we can furthermore suppose that $\chi(\mathrm{ac}(u))$ and $\mathrm{ord}_p(u)$ are constant on U for each multiplicative character $\chi \in \mathfrak{C}$. Moreover, we can suppose that 0 is the only critical point of f in U. With these conditions on U, it is easy to show that $\mathrm{Pol}_U(f) = \mathrm{Pol}_U(g)$, where

$$g(x) = x_1^2 + 2g_1(x_2, x_3, x_4)x_1 + g_0(x_2, x_3, x_4) \tag{58}$$

with g_i as above. By setting $y_1 = x_1 + g_1(x_2, x_3, x_4)$, $y_2 = x_2$, $y_3 = x_3$, $y_4 = x_4$ then at the new coordinates (y_1, y_2, y_3, y_4) we have $g(x_1, x_2, x_3, x_4) = y_1^2 + h_1(y_2, y_3, y_4)$, and hence, by the change of variables formula, we may suppose that $g_1 = 0$ in (58), so that $g(x) = x_1^2 + g_0(x_2, x_3, x_4)$. By using the argument in [Denef 1991, Section 5.1] and by noting that $\mathrm{Pol}_{p^{m_0}\mathbb{Z}_p}(x_1^2) = \{-1/2\}$ and enlarging m_0 if needed, we have $\mathrm{Pol}_U(f) = \mathrm{Pol}_U(g) = -1/2 + \mathrm{Pol}_{U'}(g_0) = \{-1/2 + \alpha | \alpha \in \mathrm{Pol}_{U'}(g_0)\}$, where $U' = p^{m_0}\mathbb{Z}_p^3$.

As in the proof of Proposition 5.5, we can find a finite extension L of $\mathbb{Q}_p$, a Zariski open subset W of $\mathbb{A}_L^4$ and a change of coordinates such that

$$f|_W = \tilde{u}(x)(x_1^2 + \tilde{g}_0(x_2, x_3, x_4)),$$

where $\tilde{u}, \tilde{g}_0$ are regular functions on W and $\tilde{u}$ is nonzero on W. Let ϖ_L be a uniformizing element of $\mathcal{O}_L$. Let m_1 be a large enough integer and $\tilde{U} = \varpi_L^{m_1}\mathcal{O}_L^4$. By the above argument we also have $\mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0) - 1/2 = \mathrm{Pol}_{\tilde{U}}(f) = \mathrm{Pol}_{\tilde{U}}(g) = -1/2 + \mathrm{Pol}_{\tilde{U}'}(g_0)$, where $\tilde{U}' = \varpi_L^{m_1}\mathcal{O}_L^3$. Thus $\mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0) = \mathrm{Pol}_{\tilde{U}'}(g_0)$.

Suppose first that $\tilde{g}_0$ has a nonrational singularity at $0 \in \mathbb{A}^3$, namely, $\tilde{\alpha}_{\tilde{g}_0,0}$ is equal to the log-canonical threshold at 0 of $\tilde{g}_0$. Let π be a log resolution of $\tilde{g}_0$ at 0, let $(\nu_i, N_i)_{i \in I}$ be the numerical data of π so that $1 \geq \tilde{\alpha}_{\tilde{g}_0,0} = \min_{(\nu_i,N_i)\neq(1,1)} \nu_i/N_i$ (see [Denef 1991; Mustaţă 2012]). Thus, by using [Igusa 1974; Igusa 1978] we have $\alpha \leq -\tilde{\alpha}_{\tilde{g}_0,0}$ for all $\alpha \in \mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0)$. Moreover, it follows from [Yano 1978, Corollary 3.17] that $\tilde{\alpha}_{f,0} = 1/2 + \tilde{\alpha}_{\tilde{g}_0,0}$. If there is $\beta \in \mathrm{Pol}_{U'}(g_0)$ such that $\beta > -\tilde{\alpha}_{f,0} + 1/2 = -\tilde{\alpha}_{\tilde{g}_0,0} \geq -1$ then it follows by [Igusa 1974; Igusa 1978; Veys and Zúñiga Galindo 2008] that there is a log-resolution π' of g_0 with a numerical data (ν, N) satisfying $\nu < \tilde{\alpha}_{\tilde{g}_0,0}N \leq N$. Thus we can use [Veys and Zúñiga Galindo 2008, Theorem 2.7] to see that there is $\beta' \in \mathrm{Pol}_{\tilde{U}'}(g_0)$ such that $\beta' > -\tilde{\alpha}_{\tilde{g}_0,0}$. This contradicts the fact that $\mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0) = \mathrm{Pol}_{\tilde{U}'}(g_0)$ and $\alpha \leq -\tilde{\alpha}_{\tilde{g}_0,0}$ for all $\alpha \in \mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0)$. Thus we can conclude that $\beta \leq -\tilde{\alpha}_{f,0} + 1/2$ for all $\beta \in \mathrm{Pol}_{U'}(g_0)$. Therefore $\alpha \leq -\tilde{\alpha}_{f,0}$ for all $\alpha \in \mathrm{Pol}_U(f)$ as desired.

Suppose now that $\tilde{g}_0$ has a rational singularity at $0 \in \mathbb{A}^3$, namely $\tilde{\alpha}_{\tilde{g}_0,0}$ is strictly larger than the log-canonical threshold at 0 of $\tilde{g}_0$, or equivalently, $\tilde{\alpha}_{\tilde{g}_0,0} > 1$. We can use a classical result from [Durfee 1979] to see that either $\tilde{g}_0$ is smooth at 0, or, we can use an analytic isomorphism to transform $\tilde{g}_0$ to one of the following forms:

$$x_2^{d+1} + x_3^2 + x_4^2 \quad \text{with } d \geq 1, \qquad x_2^{d-1} + x_2x_3^2 + x_4^2 \quad \text{with } d \geq 4,$$
$$x_2^4 + x_3^3 + x_4^2, \quad x_2^3 + x_2x_3^3 + x_4^2, \quad x_2^3 + x_3^5 + x_4^2.$$

By using [Yano 1978, Corollary 3.17], we can compute $\tilde{\alpha}_{\tilde{g}_0,0}$ in each case. In fact, $\tilde{\alpha}_{\tilde{g}_0,0} \leq \frac{3}{2}$ if $\tilde{g}_0$ is not smooth at 0. Moreover, it is easy to find a log-resolution π of $\tilde{g}_0$ such that for all numerical data $(\nu, N) \neq (1, 1)$ we have $\nu \geq \tilde{\alpha}_{\tilde{g}_0,0} N$. For example, in the first form, we have $\tilde{\alpha}_{\tilde{g}_0,0} = 1 + 1/(d+1)$ and we can choose π such that its numerical data is $\{(1, 1)\} \cup \{(3+2m, 2+2m) \mid 0 \leq m \leq k-1\} \cup \{(2+2k, 2k+1)\}$ or $\{(1, 1)\} \cup \{(3+2m, 2+2m) \mid 0 \leq m \leq k-1\}$ depending on $d = 2k$ or $d = 2k-1$, respectively. Thus, it follows from [Igusa 1974; 1978] that $\alpha \leq -\tilde{\alpha}_{\tilde{g}_0,0}$ for all $\alpha \in \mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0)$.

We now show that the multiplicity e of g_0 at 0 is at most 2. If this is not true, we can find a log-resolution π' of g_0 such that there is a numerical data $(\nu, N) \neq (1, 1)$ with $\nu \leq N$ (indeed, we can find a log-resolution of g_0 after using a blowing up at 0, thus the numerical data $(3, e)$ appears in this log-resolution). Thus we can use [Igusa 1974; 1978; Veys and Zúñiga Galindo 2008] as above to contradict the facts that $\mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0) = \mathrm{Pol}_{\tilde{U}'}(g_0)$ and $\alpha \leq -\tilde{\alpha}_{\tilde{g}_0,0}$ for all $\alpha \in \mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0)$. This shows that e is at most 2. If $e = 1$ or $\tilde{g}_0$ is smooth at 0 then it is easy to have that $\mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0) = \mathrm{Pol}_{\tilde{U}'}(g_0) = \mathrm{Pol}_{U'}(g_0) = \varnothing$, so our claim follows. If $e = 2$ and $\tilde{g}_0$ is not smooth at 0, we can use Weierstrass preparation again and the above argument to suppose that $g_0(x_2, x_3, x_4) = x_2^2 + h(x_3, x_4)$ for some analytic function h in at most two variables. We also have, as above, that $\mathrm{Pol}_U(f) = -1 + \mathrm{Pol}_{U''}(h)$ and $\mathrm{Pol}_{\tilde{U}'}(\tilde{g}_0) = \mathrm{Pol}_{\tilde{U}'}(g_0) = -\frac{1}{2} + \mathrm{Pol}_{\tilde{U}''}(h)$, where $U'' = p^{m_0}\mathbb{Z}_p^2$ and $\tilde{U}'' = \varpi_L^{m_1}\mathcal{O}_L^2$. If there is $\beta \in \mathrm{Pol}_{U''}(h)$ with $\beta > -\tilde{\alpha}_{f,0} + 1 = -\tilde{\alpha}_{\tilde{g}_0,0} + \frac{1}{2} \geq -1$ then we can repeat the above argument to get a contradiction. Our desired result (57) now follows from $\mathrm{Pol}_U(f) = -1 + \mathrm{Pol}_{U''}(h)$ and $\beta \leq -\tilde{\alpha}_{f,0} + 1$ for all $\beta \in \mathrm{Pol}_{U''}(h)$. □

Remark 5.7. For each of the above cases in which Conjecture 1 is shown in this paper, one moreover sees that, after excluding a finite set S (which depends on f) of prime divisors of N, the implied constant can be taken depending only on d and n (and on ε). The only case where this is not directly clear is for the case with $(n-s)/d \leq 1$, since its treatment in [Cluckers et al. 2019] uses a chosen log resolution which depends on f. However, the complexity of such log resolutions (and of the corresponding proof in [Cluckers et al. 2019]) remains bounded when n and d are fixed. Indeed, one first takes a log resolution of a generic polynomial of degree d in n variables; this then yields a log resolution for polynomials whose coefficients lie in a dense Zariski open subset U of the parameter space. One proceeds similarly for a generic polynomial with parameters in the complement of the dense open U.

Note that the exclusion of a finite list of prime divisors of N is necessary, as can be seen when one replaces a polynomial f by pf for some prime p. It is not clear at the moment whether the finite set S has to depend fully on f in general, or, just on f_d.

Acknowledgements

The authors would like to thank T. Browning, T. Cochrane, J. Denef, N. Katz, E. Kowalski, F. Loeser, M. Mustaţă, and J. Wright for inspiring discussions on the topics of this paper. The authors are partially supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) with ERC Grant Agreement 615722 MOTMELSUM, by the Labex CEMPI (ANR-11-LABX-0007-01), and by KU Leuven IF C16/23/010. Nguyen is partially supported by Fund for Scientific Research - Flanders (Belgium) (F.W.O.) 12X3519N, and by the Excellence Research Chair "FLCarPA: L-functions in positive characteristic and applications" financed by the Normandy Region. Nguyen is also funded by Vingroup Joint Stock Company and supported by Vingroup Innovation Foundation (VinIF) under the project code VINIF.2021.DA00030, and by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 101.04-2019.316.

References

[Baker 1983] R. C. Baker, "Small solutions of congruences", *Mathematika* **30**:2 (1983), 164–188. MR Zbl

[Birch 1962] B. J. Birch, "Forms in many variables", *Proc. Roy. Soc. London Ser. A* **265** (1962), 245–263. MR Zbl

[Browning and Heath-Brown 2017] T. Browning and R. Heath-Brown, "Forms in many variables and differing degrees", *J. Eur. Math. Soc. (JEMS)* **19**:2 (2017), 357–394. MR Zbl

[Browning and Prendiville 2017] T. D. Browning and S. M. Prendiville, "Improvements in Birch's theorem on forms in many variables", *J. Reine Angew. Math.* **731** (2017), 203–234. MR Zbl

[Castryck and Nguyen 2019] W. Castryck and K. H. Nguyen, "New bounds for exponential sums with a non-degenerate phase polynomial", *J. Math. Pures Appl.* (9) **130** (2019), 93–111. MR Zbl

[Chalk 1987] J. H. H. Chalk, "On Hua's estimates for exponential sums", *Mathematika* **34**:2 (1987), 115–123. MR Zbl

[Chambille and Nguyen 2021] S. Chambille and K. H. Nguyen, "Proof of the Cluckers–Veys conjecture on exponential sums for polynomials with log-canonical threshold at most a half", *Int. Math. Res. Not.* **2021**:9 (2021), 6498–6542. MR Zbl

[Cluckers 2008a] R. Cluckers, "Igusa and Denef–Sperber conjectures on nondegenerate p-adic exponential sums", *Duke Math. J.* **141**:1 (2008), 205–216. MR Zbl

[Cluckers 2008b] R. Cluckers, "Igusa's conjecture on exponential sums modulo p and p^2 and the motivic oscillation index", *Int. Math. Res. Not.* **2008**:4 (2008), art. id. rnm118. MR Zbl

[Cluckers 2010] R. Cluckers, "Exponential sums: questions by Denef, Sperber, and Igusa", *Trans. Amer. Math. Soc.* **362**:7 (2010), 3745–3756. MR Zbl

[Cluckers and Veys 2016] R. Cluckers and W. Veys, "Bounds for p-adic exponential sums and log-canonical thresholds", *Amer. J. Math.* **138**:1 (2016), 61–80. MR Zbl

[Cluckers et al. 2016] R. Cluckers, J. Gordon, and I. Halupczok, "Transfer principles for bounds of motivic exponential functions", pp. 111–127 in *Families of automorphic forms and the trace formula*, edited by W. Müller et al., Springer, 2016. MR Zbl

[Cluckers et al. 2018] R. Cluckers, J. Gordon, and I. Halupczok, "Uniform analysis on local fields and applications to orbital integrals", *Trans. Amer. Math. Soc. Ser. B* **5** (2018), 125–166. MR Zbl

[Cluckers et al. 2019] R. Cluckers, M. Mustaţă, and K. H. Nguyen, "Igusa's conjecture for exponential sums: optimal estimates for nonrational singularities", *Forum Math. Pi* **7** (2019), art. id. e3. MR Zbl

[Cochrane and Zheng 1999] T. Cochrane and Z. Zheng, "Pure and mixed exponential sums", *Acta Arith.* **91**:3 (1999), 249–278. MR Zbl

[Cochrane and Zheng 2000] T. Cochrane and Z. Zheng, "Exponential sums with rational function entries", *Acta Arith.* **95**:1 (2000), 67–95. MR Zbl

[Deligne 1974] P. Deligne, "La conjecture de Weil, I", *Inst. Hautes Études Sci. Publ. Math.* **43** (1974), 273–307. MR Zbl

[Denef 1991] J. Denef, "Report on Igusa's local zeta function", exposé 741, pp. 359–386 in *Séminaire Bourbaki*, 1990/1991, Astérisque **201-203**, 1991. MR Zbl

[Denef and Sperber 2001] J. Denef and S. Sperber, "Exponential sums mod p^n and Newton polyhedra", *Bull. Belg. Math. Soc. Simon Stevin* **suppl.** (2001), 55–63. MR Zbl

[Denef and Veys 1995] J. Denef and W. Veys, "On the holomorphy conjecture for Igusa's local zeta function", *Proc. Amer. Math. Soc.* **123**:10 (1995), 2981–2988. MR Zbl

[Durfee 1979] A. H. Durfee, "Fifteen characterizations of rational double points and simple critical points", *Enseign. Math.* (2) **25**:1-2 (1979), 131–163. MR Zbl

[Fraser and Wright 2020] R. Fraser and J. Wright, "The local sum conjecture in two dimensions", *Int. J. Number Theory* **16**:8 (2020), 1667–1699. MR Zbl

[Heath-Brown 1985] D. R. Heath-Brown, "A multiple exponential sum to modulus p^2", *Canad. Math. Bull.* **28**:4 (1985), 394–396. MR Zbl

[Heath-Brown 1996] D. R. Heath-Brown, "A new form of the circle method, and its application to quadratic forms", *J. Reine Angew. Math.* **481** (1996), 149–206. MR Zbl

[Hironaka 1964] H. Hironaka, "Resolution of singularities of an algebraic variety over a field of characteristic zero, I", *Ann. of Math.* (2) **79**:1 (1964), 109–203. MR Zbl

[Hooley 1988] C. Hooley, "On nonary cubic forms", *J. Reine Angew. Math.* **386** (1988), 32–98. MR Zbl

[Hooley 1991] C. Hooley, "On the number of points on a complete intersection over a finite field", *J. Number Theory* **38**:3 (1991), 338–358. MR Zbl

[Hua 1959] L.-K. Hua, *Additive Primzahltheorie*, B. G. Teubner, Leipzig, 1959. MR Zbl

[Igusa 1974] J.-I. Igusa, "Complex powers and asymptotic expansions, I: Functions of certain types", *J. Reine Angew. Math.* **268/269** (1974), 110–130. MR Zbl

[Igusa 1975] J.-I. Igusa, "Complex powers and asymptotic expansions, II: Asymptotic expansions", *J. Reine Angew. Math.* **278/279** (1975), 307–321. MR Zbl

[Igusa 1978] J.-I. Igusa, *Forms of higher degree*, Lectures on Mathematics and Physics **59**, Tata Institute of Fundamental Research, Bombay, 1978. MR Zbl

[Katz 1999] N. M. Katz, "Estimates for "singular" exponential sums", *Internat. Math. Res. Notices* **1999**:16 (1999), 875–899. MR Zbl

[Kouchnirenko 1976] A. G. Kouchnirenko, "Polyèdres de Newton et nombres de Milnor", *Invent. Math.* **32**:1 (1976), 1–31. MR Zbl

[Lichtin 2013] B. Lichtin, "On a conjecture of Igusa", *Mathematika* **59**:2 (2013), 399–425. MR Zbl

[Mustaţă 2012] M. Mustaţă, "IMPANGA lecture notes on log canonical thresholds", pp. 407–442 in *Contributions to algebraic geometry*, edited by P. Pragacz, Eur. Math. Soc., Zürich, 2012. MR Zbl

[Mustaţă and Popa 2020] M. Mustaţă and M. Popa, "Hodge ideals for $\mathbb{Q}$-divisors, V-filtration, and minimal exponent", *Forum Math. Sigma* **8** (2020), art. id. e19. MR Zbl

[Nguyen 2021] K. H. Nguyen, "On a uniform bound for exponential sums modulo p^m for Deligne polynomials", preprint, 2021. arXiv 2111.11898

[Nguyen and Veys 2022] K. H. Nguyen and W. Veys, "On the motivic oscillation index and bound of exponential sums modulo p^m via analytic isomorphisms", *J. Math. Pures Appl.* (9) **157** (2022), 211–242. MR Zbl

[Saito 1993] M. Saito, "On b-function, spectrum and rational singularity", *Math. Ann.* **295**:1 (1993), 51–74. MR Zbl

[Varčenko 1976] A. N. Varčenko, "Newton polyhedra and estimates of oscillatory integrals", *Funkcional. Anal. i Priložen.* **10**:3 (1976), 13–38. In Russian; translated in *Functional Anal. Appl.* **18**:3 (1976), 175–196. MR Zbl

[Veys 2020] W. Veys, "On the log canonical threshold and numerical data of a resolution in dimension 2", *Manuscripta Math.* **163**:1-2 (2020), 1–11. MR Zbl

[Veys and Zúñiga Galindo 2008] W. Veys and W. A. Zúñiga Galindo, "Zeta functions for analytic mappings, log-principalization of ideals, and Newton polyhedra", *Trans. Amer. Math. Soc.* **360**:4 (2008), 2205–2227. MR Zbl

[Weil 1965] A. Weil, "Sur la formule de Siegel dans la théorie des groupes classiques", *Acta Math.* **113** (1965), 1–87. MR Zbl

[Yano 1978] T. Yano, "On the theory of b-functions", *Publ. Res. Inst. Math. Sci.* **14**:1 (1978), 111–202. MR Zbl

Communicated by Antoine Chambert-Loir
Received 2022-02-25 Revised 2023-04-06 Accepted 2023-09-03

raf.cluckers@univ-lille.fr Laboratoire Painlevé, Université Lille 1, CNRS - UMR 8524, Cité Scientifique, Villeneuve d'Ascq, France

Department of Mathematics, KU Leuven, Belgium kien.nguyenhuu@kuleuven.be

Laboratoire de Mathématiques Nicolas Oresme,
Université de Caen Normandie, CNRS - UMR 6139, Caen, France

Department of Mathematics, KU Leuven, Belgium

Thang Long Institute of Mathematics and Applied Sciences, Hanoi, Vietnam

ALGEBRA AND NUMBER THEORY 18:7 (2024)
https://doi.org/10.2140/ant.2024.18.1305

Exceptional characters and prime numbers in sparse sets

Jori Merikoski

We develop a lower bound sieve for primes under the (unlikely) assumption of infinitely many exceptional characters. Compared with the illusory sieve due to Friedlander and Iwaniec which produces asymptotic formulas, we show that less arithmetic information is required to prove nontrivial lower bounds. As an application of our method, assuming the existence of infinitely many exceptional characters we show that there are infinitely many primes of the form a^2+b^8.

1. Introduction

Understanding the distribution of prime numbers along polynomial sequences is one of the basic questions in analytic number theory. For sparse polynomial sequences the problem is solved only in a handful of cases. The most notable are the Friedlander and Iwaniec [1998b] theorem of primes of the form a^2+b^4 and the result of Heath-Brown [2001] of primes of the form a^3+2b^3, which has been generalized to binary cubic forms by Heath-Brown and Moroz [2002] and to general incomplete norm forms by Maynard [2020]. Also, the result of Friedlander and Iwaniec has been extended by Heath-Brown and Li [2017] to primes of the form a^2+p^4 where p is a prime.

Let $\pm D$ be a fundamental discriminant and let $\chi_D(n)=(D/n)$ be the associated primitive real character. We say that χ_D is exceptional if $L(1,\chi_D)$ is very small, say,

$$L(1,\chi_D)=\sum_{n=1}^{\infty}\frac{\chi_D(n)}{n}\leq \log^{-100} D. \tag{1-1}$$

It is conjectured that (for an exponent such as 100) there are at most finitely many exceptional characters, which is closely related to the conjecture that L-functions do not have zeros close to $s=1$ (so-called

MSC2020: primary 11N32; secondary 11N36.
Keywords: prime numbers, Siegel zero, sieve methods, prime values of polynomials, exceptional character.

Siegel zeros). However, assuming that there do exist infinitely many exceptional characters, it is possible to prove very strong results on distribution of prime numbers. For example, Heath-Brown [1983] has shown that the twin prime conjecture follows from such an assumption, and Drappeau and Maynard [2019] have bounded sums of Kloosterman sums along primes. The potential benefit of such results is that for an unconditional proof we are now allowed to assume the nonexistence of exceptional characters, which in turn implies strong regularity in the distribution of primes in arithmetic progressions. Such a bifurcation in the proof has been successfully used to solve problems, for example, in the proof of Linnik's theorem [1944] and in many results in the theory of L-functions.

The state of the art method using exceptional characters is the so-called illusory sieve developed by Friedlander and Iwaniec [2003; 2004; 2005], which is geared towards counting primes in sparse sets. Assuming the existence of infinitely many exceptional characters (with the exponent 100 in (1-1) replaced by 200), Friedlander and Iwaniec [2005] proved that there are infinitely many prime numbers of the form a^2+b^6. For their method it is required to solve the corresponding ternary divisor problem, that is, show an asymptotic formula for $\sum \tau_3(a^2+b^6)$. This essentially comes down to showing that the sequence has an exponent of distribution $\frac{2}{3}-\varepsilon$. Friedlander and Iwaniec [2006] have solved this problem for a^2+b^6 in a form that is narrowly sufficient for the illusory sieve.

Their method fails for sparser polynomial sequences such as a^2+b^8, which has an exponent of distribution $\frac{5}{8}-\varepsilon$. The purpose of this article is to develop a lower bound version of the illusory sieve. That is, instead of aiming for an asymptotic formula for primes of the form a^2+b^8, we just want to prove a lower bound of the correct order of magnitude for the number of primes. Morally speaking, we are able to show a nontrivial lower bound for primes in sequences with a level of distribution greater than $(1+\sqrt{e})/(1+2\sqrt{e}) = 0.61634\ldots$ (see Theorem 16), so that the sequence a^2+b^8 qualifies.

We will state the general version of our lower bound sieve at the end of this article (Theorem 16). For now we state the result for primes of the form a^2+b^8. For any $n \geq 0$ define

$$\kappa_n := \int_0^1 \sqrt{1-t^n}\, dt.$$

Theorem 1. *If there are infinitely many exceptional primitive characters χ, then there are infinitely many prime numbers of the form a^2+b^8. More precisely, if $L(1,\chi_D) \leq \log^{-100} D$, then for $\exp(\log^{10} D) < x < \exp(\log^{16} D)$ we have*

$$\sum_{\substack{a^2+b^8 \leq x \\ a,b>0}} \Lambda(a^2+b^8) \geq (0.189-o(1)) \cdot \frac{4}{\pi}\kappa_8 x^{5/8}$$

and

$$\sum_{\substack{a^2+b^8 \leq x \\ a,b>0}} \Lambda(a^2+b^8) \leq (1+o(1)) \cdot \frac{4}{\pi}\kappa_8 x^{5/8}.$$

Remark. Note that $\kappa_2 = \pi/4$, so that the coefficient is in fact κ_8/κ_2, and $\frac{4}{\pi}\kappa_8 x^{5/8}$ is the expected main term. It turns out that the upper bound result is much easier and for this having an exponent of distribution $\frac{1}{2}$ is sufficient.

1.1. ***Sketch of the argument.*** We present here a nonrigorous sketch of the proof of the lower bound in Theorem 1. Let

$$a_n := 1_{(n,D)=1} \sum_{\substack{n=a^2+b^8 \\ (a,b)=1 \\ a,b>0}} 1,$$

so that our goal is to estimate $\sum_{n\sim x} a_n \Lambda(n)$.

Let $\chi = \chi_D$. Similarly as in [Friedlander and Iwaniec 2005], we define the Dirichlet convolutions

$$\lambda := 1 * \chi \quad \text{and} \quad \lambda' := \chi * \log,$$

so that

$$\lambda * \Lambda = (1 * \chi) * (\mu * \log) = (\chi * \log) * (1 * \mu) = \lambda'. \tag{1-2}$$

Note that $\lambda(n) \geq 0$ and $\lambda'(n) \geq \Lambda(n) \geq 0$ (by using $\lambda' = \lambda * \Lambda$).

The basic idea in arguments using the exceptional characters is as follows. Since

$$L(1,\chi)^{-1} = \sum_n \mu(n)\chi(n)/n = \prod_p \left(1 - \frac{\chi(p)}{p}\right)$$

is large, we expect that $\chi(p) = \mu(p)$ for most primes (in a range depending on D), so that heuristically we have $\chi \approx \mu$ and $\lambda' \approx \Lambda$. Hence, we expect that

$$\sum_{n\sim x} a_n \Lambda(n) \approx \sum_{n\sim x} a_n \lambda'(n). \tag{1-3}$$

Since the modulus of χ is small, morally $\lambda'(n)$ is of same complexity as the divisor function $\tau(n)$, so that we have replaced the original sum by a much simpler sum.

Making the approximation (1-3) rigorous is the difficult part of the argument, especially for sparse sequences a_n. Friedlander and Iwaniec succeeded in this under the assumption that the exponent of distribution is almost $\frac{2}{3}$, which was sufficient to handle primes in the sequence $a^2 + b^6$. In our application a_n has the exponent of distribution $\frac{5}{8} - \varepsilon$. This results in an additional error term compared to [Friedlander and Iwaniec 2005], but we are able to show that the contribution from this is smaller (but of the same order) as the main term.

To bound the error term in (1-3), using $\lambda' = \lambda * \Lambda$ we see that

$$\lambda'(n) - \Lambda(n) = \sum_{\substack{n=km \\ m>1}} \Lambda(k)\lambda(m).$$

Let $z = x^{\varepsilon}$ (in the proof we choose a slightly smaller z for technical reasons). Then

$$\begin{aligned}\sum_{n\sim x} a_n\Lambda(n) &\geq \sum_{n\sim x} a_n\Lambda(n)1_{(n,P(z))=1}\\ &= \sum_{n\sim x} a_n\lambda'(n)1_{(n,P(z))=1} - \sum_{\substack{km\sim x\\ k,m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(km,P(z))=1}\\ &=: S_1 - S_2.\end{aligned}$$

Note that by removing the small prime factors we have guaranteed that $m \geq z$ in the second sum, so that we expect $\lambda(m) \approx (1*\mu)(m) = 0$ for almost all m in S_2. Thus, we expect that S_1 gives us the main term and that $S_2 = o(S_1)$.

Remark. The above decomposition has a close resemblance to the recent work of Granville [2021] using the identity

$$\Lambda(n)1_{(n,P(z))} = 1_{(n,P(z))}\log n - \sum_{\substack{n=\ell m\\ (\ell m,P(z))=1\\ \ell,m\geq z}} \Lambda(\ell).$$

For the main term S_1 we can handle the condition $(n, P(z)) = 1$ by the fundamental lemma of the sieve, so we ignore this detail for the moment. Thus, we have to evaluate

$$\sum_{n\sim x} a_n\lambda'(n) = \sum_{mn\sim x} a_{mn}\chi(m)\log n.$$

We have $m \geq x^{1/2}$ or $n \geq x^{1/2}$, so that we are able to compute S_1 provided that our sequence a_n has an exponent of distribution $\frac{1}{2}$. This is because the modulus of χ is $x^{o(1)}$, so that χ is essentially of the same complexity as the constant function 1. We find that S_1 gives the expected main term, so that we need to bound the error term S_2.

Similarly as in the argument in [Friedlander and Iwaniec 2005], the range $x^{2/3}$ plays a special role. With this in mind, we define $\gamma = \frac{1}{24} + \varepsilon$ so that $\frac{2}{3} - \gamma = \frac{5}{8} - \varepsilon$ is the exponent of distribution. We split S_2 into three parts depending on the size of k

$$\begin{aligned}S_2 = \sum_{\substack{km\sim x\\ k>x^{1/3+\gamma}\\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(km,P(z))=1} &+ \sum_{\substack{km\sim x\\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma}\\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(km,P(z))=1}\\ &+ \sum_{\substack{km\sim x\\ z\leq k\leq x^{1/3-2\gamma}\\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(km,P(z))=1}\\ =: S_{21} + S_{22} + S_{23}.&\end{aligned}$$

By similar arguments as in [Friedlander and Iwaniec 2005], we are able use the lacunarity of $\lambda(m)$ to bound the terms S_{21} and S_{23} suitably in terms of $L(1, \chi)$, using the fact that the exponent of the distribution is $\frac{2}{3}-\gamma$. That is, for S_{21} we write

$$S_{21} \leq (\log x) \sum_{\substack{km\sim x\\ k>x^{1/3+\gamma}\\ m\geq z}} a_{km}\lambda(m)1_{(m,P(z))=1},$$

and for S_{23} we drop $1_{(m,P(z))=1}$ by positivity and write

$$\lambda(m)=\sum_{m=cd}\chi(d),$$

where c or d is $> x^{1/3+\gamma}$. In all cases we get a variable $> x^{1/3+\gamma}$, so that these can be evaluated as Type I sums. This gives

$$S_{21}+S_{23} \ll_C x^{5/8}(\log^{-C} x + L(1,\chi)\log^5 x),$$

which is sufficient by the assumption that χ is an exceptional character.

The novel part in our argument is the treatment of the middle range

$$S_{22}=\sum_{\substack{km\sim x\\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma}\\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(km,P(z))=1}.$$

Note that also in [Friedlander and Iwaniec 2005] a narrow range near $x^{2/3}$ has to be discarded, but the argument there requires $\gamma=o(1)$. Thanks to the restriction $(m,P(z))=1$, it turns out that we are able to handle all parts of S_{22} except when m is a prime number. To see this, if m is not a prime, then $m=m_1m_2$ for some $m_1,m_2\geq z$, and we essentially get (recall that $\lambda(m)\geq 0$)

$$\sum_{\substack{km\sim x\\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma}\\ m\notin\mathbb{P}}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} \leq \sum_{\substack{km_1m_2\leq x\\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma}\\ m_1,m_2\geq z}} a_{km}\Lambda(k)\lambda(m_1)\lambda(m_2)1_{(m_1m_2,P(z))=1},$$

since λ is multiplicative and the part where $(m_1,m_2)>1$ gives a negligible contribution. For the part $km_1>x^{1/2}$ we use $\lambda(m_1)\leq\tau(m_1)\ll 2^{1/\varepsilon}$ and combine variables $\ell=km_1$ to get a bound

$$\ll \sum_{z\leq m_2\ll x^{1/2}} \lambda(m_2)\sum_{\ell\sim x/m_2} a_{\ell m_2},$$

which can be bounded suitably in terms of $L(1,\chi)$ by a similar argument as with S_{21}. The part $km_1\leq x^{1/2}$ is handled similarly, using $\lambda(m_2)\leq\tau(m_2)\ll 2^{1/\epsilon}$ and extracting $L(1,\chi)$ from $\lambda(m_1)$ this time. Thus, the contribution from the composite m is negligible.

Hence, it remains to bound

$$S_{222}:=\sum_{\substack{kp\sim x\\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma}}} a_{km}\Lambda(k)\lambda(p)=\sum_{\substack{kp\sim x\\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma}}} a_{km}\Lambda(k)(1+\chi(p)).$$

Here we are not able to make use of the lacunarity of $\lambda(p)$. However, since S_{222} counts products of two primes of medium sizes, we immediately see that S_{222} should be smaller than the main term by a factor of $O(\gamma)$, so that at least for small enough γ we get a nontrivial lower bound. We use the linear sieve upper bound to the variable p to make this upper bound rigorous and precise, which leads to the constant 0.189 in Theorem 1.

The paper is structured as follows. In Section 2 we carry out the sieve argument and the proof of Theorem 1 assuming a sufficient exponent of distribution for a_n (Propositions 8 and 9). In Section 3 we prove Propositions 8 and 9 by generalizing the arguments in [Friedlander and Iwaniec 2006]. Lastly, in Section 4 we state a general version of the sieve and explain how the method could be improved assuming further arithmetic information.

Remark. Our sieve argument is inspired by Harman's sieve method [2007], although the exact details in this setting turn out to be quite different. The moral of the story is that all sieve arguments should be continuous with respect to the quality of the arithmetic information, which in this case is measured solely by the exponent of distribution. That is, even though we fail to obtain an asymptotic formula after some point $\big($in this case $\frac{2}{3}\big)$, we still expect to be able to produce lower and upper bounds of the correct order of magnitude with slightly less arithmetic information.

1.2. ***Notations.*** For functions f and g with $g \geq 0$, we write $f \ll g$ or $f = O(g)$ if there is a constant C such that $|f| \leq Cg$. The notation $f \asymp g$ means $g \ll f \ll g$. The constant may depend on some parameter, which is indicated in the subscript (e.g., $\ll_\epsilon$). We write $f = o(g)$ if $f/g \to 0$ for large values of the variable. For summation variables we write $n \sim N$ meaning $N < n \leq 2N$.

For two functions f and g with $g \geq 0$, it is convenient for us to denote $f(N) \lll g(N)$ if $f(N) \ll g(N) \log^{O(1)} N$. For parameters such as ε we write $f(N) \lll_\varepsilon g(N)$ to mean $f(N) \ll_\varepsilon g(N) \log^{O_\varepsilon(1)} N$. A typical bound we use is $S(N) = \sum_{n \leq N} \tau_k(n)^K \lll_{k,K} N$, where τ_k is the k-fold divisor function. We say that an arithmetic function f is divisor bounded if $|f(n)| \lll \tau(n)^K$ for some K.

For a statement E we denote by 1_E the characteristic function of that statement. For a set A we use 1_A to denote the characteristic function of A.

We let $e(x) := e^{2\pi i x}$ and $e_q(x) := e(x/q)$ for any integer $q \geq 1$. We denote

$$\lambda := 1 * \chi \quad \text{and} \quad \lambda' := \chi * \log .$$

2. The sieve argument

In this section state the arithmetic information (Propositions 8 and 9) and assuming this we give the proof of Theorem 1 using a sieve argument with exceptional characters. We postpone the proof of Propositions 8 and 9 to Section 3. From here on we let q denote the modulus of the exceptional character $\chi = \chi_q$, to avoid conflating it with the level of distribution which we will denote by D; this also agrees with the notations in [Friedlander and Iwaniec 2005, Section 14]. Throughout this section we denote

$$a_n := 1_{(n,q)=1} \sum_{\substack{n=a^2+b^8\\(a,b)=1\\a,b>0}} 1, \quad \text{and} \quad b_n := 1_{(n,q)=1} \frac{1}{4} \sum_{\substack{n=a^2+b^2\\(a,b)=1\\a,b>0}} b^{-3/4}.$$

In b_n we are counting the representations a^2+b^2 weighted with the probability that b is a perfect fourth power so that heuristically we expect $\sum_{n\sim x} a_n \Lambda(n) = (1+o(1)) \sum_{n\sim x} b_n \Lambda(n)$. Differing from [Friedlander and Iwaniec 2005], it is convenient for us to write certain parts of the argument as a comparison between a_n and b_n. This is inspired by Harman's sieve method [2007], where the idea is to apply the same combinatorial decompositions to the sums over a_n and b_n and then compare, using positivity to drop certain terms entirely.

We let $g(d)$ denote the multiplicative function defined by

$$g(p^k) = 1_{p\nmid q} \frac{\varrho(p^k)}{p^k}\left(1+\frac{1}{p}\right)^{-1}, \tag{2-1}$$

where $\varrho(d)$ denotes the number of solutions to $\nu^2+1 \equiv 0(d)$. Note that for all primes p we have $\varrho(p) = 1+\chi_4(p)$. We also define

$$g_1(p^k) = \frac{\varrho(p^k)}{p^k}\left(1+\frac{1}{p}\right)^{-1}.$$

2.1. *Preliminaries.* We have collected here some standard estimates that will be needed in the sieve argument.

Lemma 2. *Let*

$$G_q := \prod_{p\,|\,q} (1-g_1(p))^{-1}.$$

Then

$$\prod_{p\le z}(1-g(p)) = (1+o(1))\frac{G_q\zeta(2)}{L(1,\chi_4)} \prod_{p\le z}(1-1/p) = (1+o(1))\frac{G_q\zeta(2)}{L(1,\chi_4)e^{\gamma_1}\log z}$$

and

$$\begin{aligned}\sum_{n\le x}\Lambda(n) b_n &= (1+o(1))\frac{G_q\zeta(2)}{L(1,\chi_4)}\sum_{n\le x} b_n = (1+o(1))\frac{4}{\pi}\kappa_8 x^{5/8}\\ &= (1+o(1))e^{\gamma_1}\log z \prod_{p\le z}(1-g(p))\sum_{n\le x} b_n,\end{aligned}$$

where $\gamma_1 = 0.577\ldots$ denotes the Euler–Mascheroni constant.

Proof. The first asymptotic follows from

$$\prod_p \frac{1-g(p)}{1-1/p} = G_q \prod_p (1-\chi_4(p)/p)(1-p^{-2})^{-1} = \frac{G_q\zeta(2)}{L(1,\chi_4)}$$

and Mertens' theorem. To get the second part we apply prime number theorem for Gaussian primes $a+ib$, splitting the sum into boxes $(a,b)\in[z_1, z_1+x/\log^{10}x]\times[z_2, z_2+x/\log^{10}x]$ so that $b^{-3/4}=(1+o(1))z_2^{-3/4}$, noting that the contribution from boxes with $z_1\le x/\log^{10}x$ or $z_2\le x/\log^{10}x$ is trivially $\ll x^{5/8}/\log x$ (by writing $\Lambda(n)\le\log x$). The prime number theorem in small boxes follows splitting the boxes in to smaller polar boxes and applying [Iwaniec and Kowalski 2004, Theorem 5.36], for instance.

Here the condition $(a^2+b^2, q)=1$ implicit in b_n cancels the multiplicative factor G_q, since by an expansion using the Möbius function

$$\begin{aligned}\sum_{n\le x} b_n &= \sum_{\substack{n=a^2+b^2\\(a,b)=1\\a,b>0}} 1_{(n,q)=1}\tfrac{1}{4}b^{-3/4} = \sum_{d\mid q}\mu(d)\sum_{\substack{n=a^2+b^2\\(a,b)=1\\a,b>0\\d\mid n}}\tfrac{1}{4}b^{-3/4}\\ &=(1+o(1))\sum_{d\mid q}\mu(d)g_1(d)\sum_{\substack{n=a^2+b^2\\(a,b)=1\\a,b>0}}\tfrac{1}{4}b^{-3/4}\\ &=(1+o(1))G_q^{-1}\sum_{\substack{n=a^2+b^2\\(a,b)=1\\a,b>0}}\tfrac{1}{4}b^{-3/4}\end{aligned}$$

For the last asymptotic note that by the change of variables $t=u^{1/4}$

$$\tfrac{1}{4}\int_0^1 u^{-3/4}\sqrt{1-u^2}\,dt=\int_0^1\sqrt{1-t^8}\,dt=\kappa_8$$

and $L(1,\chi_4)=\pi/4$. □

We also require the following basic estimate; see [Friedlander and Iwaniec 1998a, Lemma 1], for instance.

Lemma 3. *For every square-free integer n and every $k\ge 2$ there exists some $d\mid n$ such that $d\le n^{1/k}$ and*

$$\tau(n)\le 2^k\tau(d)^k.$$

From this we get the more general version.

Lemma 4. *For every integer n and every $k\ge 2$ there exists some $d\mid n$ such that $d\le n^{1/k}$ and*

$$\tau(n)\le 2^{k^2}\tau(d)^{k^3}.$$

Proof. Write $n=b_1b_2^2\cdots b_{k-1}^{k-1}b_k^k$ with $b_1,\dots,b_{k-1}$ square-free, by letting b_k be the largest integer such that $b_k^k\mid n$, so that n/b_k^k is k-free and splits uniquely into $b_1b_2^2\cdots b_{k-1}^{k-1}$ with b_j square-free. We have

$$\tau(n)\le\tau(b_1)\tau(b_2)^2\cdots\tau(b_k)^k.$$

By Lemma 3 for all $j \le k-1$ there is $d_j \mid b_j$ with $d_j \le b_j^{1/k}$ and $\tau(b_j) \le 2^k \tau(d_j)^k$. Hence, for $d = d_1 \cdots d_{k-1} b_k$ we have

$$d \le (b_1 \cdots b_{k-1})^{1/k} b_k = (b_1 \cdots b_{k-1} b_k^k)^{1/k} \le n^{1/k}$$

and

$$\tau(n) \le (2\tau(d_1) \cdots \tau(d_{k-1})\tau(b_k))^{k^2} \le 2^{k^2} \tau(d)^{k^3}. \qquad \square$$

To bound the final error term we require the linear sieve upper bound for primes; apply [Friedlander and Iwaniec 2010, Theorem 11.12] with $z = D$ and $s = 1$, using $F(1) = 2e^{\gamma}$.

Lemma 5 (linear sieve upper bound for primes). *Let $(c_n)_{n\ge1}$ be a sequence of nonnegative real numbers. For some fixed X_0 depending only on the sequence $(c_n)_{n\ge1}$, define r_d for all square-free $d \ge 1$ by*

$$\sum_{n \equiv 0(d)} c_n = g_0(d) X_0 + r_d,$$

where $g_0(d)$ is a multiplicative function, depending only on the sequence $(a_n)_{n\ge1}$, satisfying $0 \le g_0(p) < 1$ for all primes p. Let $D \ge 2$ (the level of distribution). Suppose that there exists a constant $L > 0$ that for any $2 \le w < D$ we have

$$\prod_{w \le p < D} (1 - g_0(p))^{-1} \le \frac{\log D}{\log w}\left(1 + \frac{L}{\log w}\right).$$

Then

$$\sum_p c_p \le (1 + O(\log^{-1/6} D)) X_0 2e^{\gamma_1} \prod_{p \le D} (1 - g_0(p)) + \sum_{\substack{d \le D \\ d \text{ square free}}} |r_d|.$$

The following lemma gives a basic upper bound for smooth numbers; see [Tenenbaum 2015, Chapter III.5, Theorem 1], for instance.

Lemma 6. *For any $2 \le z \le y$ we have*

$$\sum_{\substack{n \sim y \\ P^+(n) < z}} 1 \ll y e^{-u/2},$$

where $u := \log y / \log z$.

We also need the following simple divisor sum bound.

Lemma 7. *Let $M \gg 1$ and let $Z = M^{c_1/(\log\log M)^{c_2}}$ for some constants $c_1, c_2 > 0$. Then for any $K > 0$*

$$\sum_{m \sim M} \tau(m)^K 1_{(m, P(Z))=1} \ll_{c_1, c_2, K} M.$$

Proof. For some $L = L(K)$ we have by a standard sieve bound

$$\begin{aligned}\sum_{m\sim M}\tau(m)^K 1_{(m,P(Z))=1} \ll \sum_{m\sim M}\tau_L(m)1_{(m,P(Z))=1} &= \sum_{n_1\cdots n_L\sim M} 1_{(n_1,P(Z))=1}\cdots 1_{(n_L,P(Z))=1}\\ &\ll_{c_1,c_2,K} \frac{M(\log\log M)^{c_2}}{\log M}\sum_{n_1,\dots,n_{L-1}\ll M}\frac{1_{(n_1,P(Z))=1}\cdots 1_{(n_{L-1},P(Z))=1}}{n_1\cdots n_{L-1}}\\ &\ll_{c_1,c_2,K} \frac{M(\log\log M)^{c_2}}{\log M}\biggl(\prod_{Z\le p\le M}\Bigl(1+\frac{1}{p}\Bigr)\biggr)^{L-1}\\ &\ll_{c_1,c_2,K} \frac{M(\log\log M)^{c_2}}{\log M}\biggl(\frac{\log M}{\log Z}\biggr)^{L-1} \ll_{c_1,c_2,K} M,\end{aligned}$$

by computing the sum over $n_j = \max\{n_1,\dots,n_L\}$ first. □

2.2. *Arithmetic information.* For the sieve argument we need arithmetic information given by the following two propositions, which state that a_n has an exponent of distribution $\frac{5}{8}-\varepsilon$. We will prove these in Section 3. The first is just a standard sieve axiom on the level of distribution of the sequence a_n, and the second is similar but twisted with the quadratic character χ. For the rest of this section we denote

$$X := \sum_{n\sim x} b_n.$$

Recall that $X \asymp x^{5/8}$ by Lemma 2.

Proposition 8 (type I information). *Let $B > 0$ be a large constant and let $\Delta \in [\log^{-B} x, 1]$. Let $\varepsilon > 0$ be small but fixed. Let $D \le x^{5/8-\varepsilon}$ and N be such that $DN \asymp x$. Let $\alpha(d)$ be divisor bounded coefficients and let $g(d)$ be as in (2-1). Then for any $C > 0$*

$$\sum_{d\sim D}\alpha(d)\sum_{\substack{n\sim x/d\\ n\in(N,N(1+\Delta)]}} a_{dn} = \sum_{d\sim D}\alpha(d)\sum_{\substack{n\sim x/d\\ n\in(N,N(1+\Delta)]}} b_{dn} + O_{B,C}(X\log^{-C} x)$$

and

$$\sum_{d\le D}\alpha(d)\sum_{n\sim x/d} a_{dn} = \sum_{d\le D}\alpha(d)\sum_{n\sim x/d} b_{dn} + O_C(X\log^{-C} x) = X\sum_{d\le D}\alpha(d)g(d) + O_C(X\log^{-C} x). \tag{2-2}$$

Furthermore, for $D \le x^{2/3+\varepsilon}$ we have the last asymptotic

$$\sum_{d\le D}\alpha(d)\sum_{n\sim x/d} b_{dn} = X\sum_{d\le D}\alpha(d)g(d) + O_C(X\log^{-C} x)$$

and for $\Delta = \log^{-B} x$ for any fixed $B > 0$ the bound

$$\sum_{d\le D}\alpha(d)\sum_{n\in(N,N(1+\Delta)]} b_{dn} \ll \Delta X\sum_{d\le D}|\alpha(d)|g(d).$$

Remark. In our set up the last asymptotic actually holds up to $D \le x^{1-\varepsilon}$, but we will not need this.

Proposition 9 (type I_χ information). *Let $B > 0$ be a large constant and let $\Delta \in [\log^{-B} x, 1]$. Let $\exp(\log^{10} q) < x < \exp(\log^{16} q)$. Let $D \le x^{5/8-\varepsilon}$ and N be such that $DN \asymp x$. Let $\alpha(d)$ be divisor bounded coefficients. Then for any $C > 0$*

$$\sum_{d\sim D} \alpha(d) \sum_{\substack{n\sim x/d \\ n\in(N,N(1+\Delta)]}} a_{dn}\chi(n) = \sum_{d\sim D} \alpha(d) \sum_{\substack{n\sim x/d \\ n\in(N,N(1+\Delta)]}} b_{dn}\chi(n) + O_{B,C}(X\log^{-C} x) \ll_{B,C} X\log^{-C} x$$

and

$$\sum_{d\le D} \alpha(d) \sum_{n\sim x/d} a_{dn}\chi(n) = \sum_{d\le D} \alpha(d) \sum_{n\sim x/d} b_{dn}\chi(n) + O_C(X\log^{-C} x) \ll_C X\log^{-C} x.$$

Furthermore, the bounds for the sums with b_{dn} hold up to $D \le x^{2/3+\varepsilon}$.

We will also need the following proposition to bound certain error terms in terms of $L(1,\chi)$. This follows from [Friedlander and Iwaniec 2005, Lemmata 3.7 and 3.9] (as mentioned in [Friedlander and Iwaniec 2005, Section 14], the $g(d)$ defined by (2-1) is easily shown to satisfy the required assumptions).

Proposition 10 (exceptional characters). *Let $\lambda := (1*\chi)$. Then for any $x > z \ge q^9$ we have*

$$\sum_{n\le x} \chi(n)g(n) \ll L(1,\chi) \quad \textit{and} \quad \sum_{z<n\le x} \lambda(n)g(n) \ll L(1,\chi)\log^2 x.$$

2.3. *Initial decomposition.* Let $\varepsilon > 0$ be small and define the parameter $\gamma := \frac{1}{24}+\varepsilon$ so that $\frac{2}{3}-\gamma = \frac{5}{8}-\varepsilon$ is the exponent of distribution of a_n. Using $\lambda' = \lambda * \Lambda$ (see (1-2)) we get

$$\lambda'(n) - \Lambda(n) = \sum_{\substack{n=km \\ m>1}} \Lambda(k)\lambda(m).$$

Hence, for $z := x^{1/(\log\log x)^2}$ we have

$$\begin{aligned}\sum_{n\sim x} a_n\Lambda(n) &= \sum_{n\sim x} a_n\Lambda(n)1_{(n,P(z))=1} + O_C(x^{5/8}/\log^C x) \\ &= \sum_{n\sim x} a_n\lambda'(n)1_{(n,P(z))=1} - \sum_{\substack{km\sim x \\ k,m\ge z}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} + O_C(x^{5/8}/\log^C x) \\ &=: S_1 - S_2 + O_C(x^{5/8}/\log^C x).\end{aligned}$$

Similarly as in [Friedlander and Iwaniec 2005], by the lacunarity of $\lambda(m)$ we expect that $S_2 = o(S_1)$, but this is out of reach. We will show that $S_1 = (1+o(1))\sum_{n\sim x} b_n\Lambda(n)$ and $S_2 \le (0.811+o(1))\cdot\sum_{n\sim x} b_n\Lambda(n)$ which together imply Theorem 1.

Remark. For technical reasons we have chosen z a bit smaller than x^ε (compare with Section 1.1). This has the benefit that evaluating S_1 is a lot easier. On the downside bounding S_2 is slightly more difficult and we require Lemma 4 for this.

2.4. *Sum S_1*. Let $D_1 := x^\varepsilon$ for some small $\varepsilon > 0$. We expand the condition $1_{(n,P(z))=1}$ by using the Möbius function and split the sum to get

$$S_1 = \sum_{n\sim x} a_n \lambda'(n) \sum_{d \mid (n,P(z))} \mu(d) = \sum_{n\sim x} a_n \lambda'(n) \sum_{\substack{d \mid (n,P(z)) \\ d\le D_1}} \mu(d) + \sum_{n\sim x} a_n \lambda'(n) \sum_{\substack{d \mid (n,P(z)) \\ d> D_1}} \mu(d) =: S_1' + R_1.$$

To handle the error term R_1, note that if $d \mid P(z)$ and $d > D_1$, then d has a divisor in $[D_1, 2zD_1]$. Since $z = x^{1/(\log\log x)^2}$, by Lemma 4 (with $k=2$ applied to the variable n/d to get $n = cdn'$ with $\tau(n) \le \tau(c)^{O(1)}$), Proposition 8, and Lemma 6 we get

$$\begin{aligned} R_1 &\ll (\log x) \sum_{\substack{n\sim x \\ \exists d \mid (n,P(z)), d\in[D_1,2zD_1]}} a_n \tau(n)^3 \\ &\ll (\log x) \sum_{\substack{d\in[D_1,2zD_1] \\ d \mid P(z)}} \sum_{c\le (2x)^{1/2}} \tau(cd)^{O(1)} \sum_{n'\sim x/cd} a_{cdn'} \\ &\ll (\log x) x^{5/8} \sum_{\substack{d\in[D_1,2zD_1] \\ d \mid P(z)}} \sum_{c\le (2x)^{1/2}} \tau(cd)^{O(1)} g(cd) \ll_C x^{5/8} \log^{-C} x. \end{aligned} \tag{2-3}$$

To get the last bound use $\tau(cd)^{O(1)} g(cd) \le \tau(cd)^{O(1)}/(cd) \le \tau(c)^{O(1)}\tau(d)^{O(1)}/(cd)$ and apply Lemma 4 to the variable d before using Lemma 6.

For the main term we write

$$\begin{aligned} S_1' &= \sum_{\substack{d \mid P(z) \\ d\le D_1}} \mu(d) \sum_{\substack{n\sim x \\ n\equiv 0(d)}} a_n \lambda'(n) = \sum_{\substack{d \mid P(z) \\ d\le D_1}} \mu(d) \sum_{\substack{mn\sim x \\ mn\equiv 0(d)}} a_{mn} \chi(m) \log n \\ &= \sum_{\substack{d \mid P(z) \\ d\le D_1}} \mu(d) \sum_{\substack{mn\sim x \\ mn\equiv 0(d) \\ n> x^{1/2}}} a_{mn} \chi(m) \log n + \sum_{\substack{d \mid P(z) \\ d\le D_1}} \mu(d) \sum_{\substack{mn\sim x \\ mn\equiv 0(d) \\ n\le x^{1/2}}} a_{mn} \chi(m) \log n =: S_{11} + S_{12}. \end{aligned}$$

We write (denoting $d_1 = (m, d)$)

$$S_{11} = \sum_{\substack{d_1 d_2 \mid P(z) \\ d_1 d_2 \le D_1}} \mu(d_1 d_2) \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \chi(d_1 m) \sum_{\substack{n\sim x/md_1 d_2 \\ d_2 n > x^{1/2}}} a_{d_1 d_2 mn} \log d_2 n$$

We will use Proposition 8 to evaluate this sum but first we need to remove the cross-condition $d_2 n > x^{1/2}$ and the weight $\log d_2 n$ by using a finer-than-dyadic decomposition to the sums over d_2 and n. That is, for $\Delta = \log^{-B} x$ for some large $B > 0$ we split S_{11} into

$$\sum_{\substack{i,j \ll \log^{B+1} x \\ D_2 = (1+\Delta)^i \\ N = (1+\Delta)^j \\ D_2 N (1+\Delta)^2 > x^{1/2}}} \sum_{\substack{d_1 d_2 \mid P(z) \\ d_1 d_2 \le D_1 \\ d_2 \in (D_2, D_2(1+\Delta)]}} \mu(d_1 d_2) \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \chi(d_1 m) \sum_{\substack{n\in (N, N(1+\Delta)] \\ n\sim x/md_1 d_2 \\ d_2 n > x^{1/2}}} a_{d_1 d_2 mn} \log d_2 n.$$

Here we can write

$$\log d_2 n = \log D_2 N + O(\log^{-B} x),$$

where the error term will contribute by Lemma 4 and Proposition 8

$$\ll \log^{-B} x \sum_{n\sim x} \tau_4(n) a_n \ll \log^{-B} x \sum_{n\sim x} \tau(n)^4 a_n \ll \log^{-B} x \sum_{d \ll x^{1/2}} \tau(d)^{O(1)} \sum_{n\sim x/d} a_n \ll_B x^{5/8} \log^{O(1)-B} x$$

so that $S_{11} = S'_{11} + O_B(x \log^{O(1)-B} x)$ with

$$S'_{11} := \sum_{\substack{i,j \ll \log^{B+1} x \\ D_2=(1+\Delta)^i \\ N=(1+\Delta)^j \\ D_2N(1+\Delta)^2 > x^{1/2}}} \log D_2 N \sum_{\substack{d_1d_2 \mid P(z) \\ d_1d_2 \le D_1 \\ d_2 \in (D_2, D_2(1+\Delta)]}} \mu(d_1d_2) \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \chi(d_1 m) \sum_{\substack{n \in (N, N(1+\Delta)] \\ n \sim x/md_1d_2 \\ d_2 n > x^{1/2}}} a_{d_1d_2mn}.$$

The cross-condition $d_2 n > x^{1/2}$ holds trivially and may be dropped except in the diagonal part where

$$(1+\Delta)^{-2} x^{1/2} < D_2 N \le x^{1/2}.$$

The contribution from this diagonal part is bounded by using Proposition 8

$$\begin{aligned}
&\ll (\log x) \sum_{\substack{i,j \ll \log^{B+1} x \\ D_2=(1+\Delta)^i \\ N=(1+\Delta)^j \\ (1+\Delta)^{-2}x^{1/2} < D_2N \le x^{1/2}}} \sum_{\substack{d_1d_2 \mid P(z) \\ d_1d_2 \le D_1 \\ d_2 \in (D_2, D_2(1+\Delta)]}} \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \sum_{\substack{n \in (N, N(1+\Delta)] \\ n \sim x/md_1d_2}} a_{d_1d_2mn} \\
&\ll_C x^{5/8} \log^{-C} x + (\log x) \sum_{\substack{i,j \ll \log^{B+1} x \\ D_2=(1+\Delta)^i \\ N=(1+\Delta)^j \\ (1+\Delta)^{-2}x^{1/2} < D_2N \le x^{1/2}}} \sum_{\substack{d_1d_2 \mid P(z) \\ d_1d_2 \le D_1 \\ d_2 \in (D_2, D_2(1+\Delta)]}} \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \sum_{\substack{n \in (N, N(1+\Delta)] \\ n \sim x/md_1d_2}} b_{d_1d_2mn} \\
&\ll_C x^{5/8} \log^{-C} x + (\log^{O(1)} x) x^{5/8} \Delta^2 \sum_{\substack{i,j \ll \log^{B+1} x \\ D_2=(1+\Delta)^i \\ N=(1+\Delta)^j \\ (1+\Delta)^{-2}x^{1/2} < D_2N \le x^{1/2}}} 1 \ll_B x^{5/8} \log^{O(1)-B} x
\end{aligned}$$

by choosing $C = B$. Hence, the cross-condition $d_2 n > x^{1/2}$ may be dropped and we get $S_{11} = S''_{11} + O_B(x \log^{O(1)-B} x)$ with

$$S''_{11} := \sum_{\substack{i,j \ll \log^{B+1} x \\ D_2=(1+\Delta)^i \\ N=(1+\Delta)^j \\ D_2N(1+\Delta)^2 > x^{1/2}}} \log D_2 N \sum_{\substack{d_1d_2 \mid P(z) \\ d_1d_2 \le D_1 \\ d_2 \in (D_2, D_2(1+\Delta)]}} \mu(d_1d_2) \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \chi(d_1 m) \sum_{\substack{n \in (N, N(1+\Delta)] \\ n \sim x/md_1d_2}} a_{d_1d_2mn}.$$

Applying a similar decomposition to the corresponding sum with $b_{d_1 d_2 mn}$ and using Proposition 8 we get

$$\begin{aligned} S_{11} &= \sum_{\substack{d_1 d_2 \mid P(z) \\ d_1 d_2 \le D_1}} \mu(d_1 d_2) \sum_{\substack{d_1 m \ll x^{1/2} \\ (m,d_2)=1}} \chi(d_1 m) \sum_{\substack{n \sim x/md_1 d_2 \\ d_2 n > x^{1/2}}} b_{d_1 d_2 mn} \log d_2 n + O_C(x^{5/8} \log^{-C} x) \\ &=: M_{11} + O_C(x^{5/8} \log^{-C} x). \end{aligned}$$

Similarly, we get by Proposition 9 (denoting $d_2 = (n, d)$)

$$\begin{aligned} S_{12} &= \sum_{\substack{d_1 d_2 \mid P(z) \\ d_1 d_2 \le D_1}} \mu(d_1 d_2) \chi(d_1) \sum_{\substack{d_2 n \le x^{1/2} \\ (n,d_1)=1}} \log d_2 n \sum_{m \sim x/nd_1 d_2} a_{d_1 d_2 mn} \chi(m) \\ &= \sum_{\substack{d_1 d_2 \mid P(z) \\ d_1 d_2 \le D_1}} \mu(d_1 d_2) \chi(d_1) \sum_{\substack{d_2 n \le x^{1/2} \\ (n,d_1)=1}} \log d_2 n \sum_{m \sim x/nd_1 d_2} b_{d_1 d_2 mn} \chi(m) + O_C(x^{5/8} \log^{-C} x) \\ &=: M_{12} + O_C(x^{5/8} \log^{-C} x) \end{aligned}$$

That is, in the sums S_{11} and S_{12} we have managed to replace a_n by b_n. By reversing the steps to recombine we get

$$M_{11} + M_{12} = \sum_{n \sim x} b_n \lambda'(n) \sum_{\substack{d \mid (n, P(z)) \\ d \le D_1}} \mu(d) =: M_1$$

By a similar argument as in (2-3) we can add the part $d > D_1$ back into the sum and we get

$$M_1 = \sum_{n \sim x} b_n \lambda'(n) 1_{(n,P(z))=1} + O_C(x^{5/8}/\log^C x) \ge \sum_{n \sim x} b_n \Lambda(n) 1_{(n,P(z))=1} + O_C(x^{5/8}/\log^C x)$$

by using $\lambda'(n) \ge \Lambda(n)$. Thus, by Lemma 2 we have

$$S_1 \ge (1 + o(1)) \sum_{n \sim x} b_n \Lambda(n),$$

so that for the lower bound result it suffices to show that

$$S_2 \le (0.811 + o(1)) \cdot \sum_{n \sim x} b_n \Lambda(n).$$

We now proceed to do this, and at the end of this section we will show how to get the upper bound in Theorem 1.

Remark. We have used Lemma 6 to handle the restriction $(n, P(z)) = 1$ instead of applying the fundamental lemma of sieve. Thanks to this we were able to use the trivial lower bound $\lambda'(n) \ge \Lambda(n)$ to simplify the evaluation of the main term.

2.5. ***Sum S_2.*** Recall that $\gamma = \frac{1}{24} + \varepsilon$ and $\frac{2}{3} - \gamma = \frac{5}{8} - \varepsilon$. We split the sum S_2 into three ranges according to the size of k

$$\begin{aligned} S_2 &= \sum_{\substack{km\sim x \\ k,m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} \\ &= \sum_{\substack{km\sim x \\ k>x^{1/3+\gamma} \\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} + \sum_{\substack{km\sim x \\ x^{1/3-2\gamma}<k\leq x^{1/3+\gamma} \\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} \\ &\qquad + \sum_{\substack{km\sim x \\ z\leq k\leq x^{1/3-2\gamma} \\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} \\ &=: S_{21} + S_{22} + S_{23}. \end{aligned}$$

Using the assumption that $L(1,\chi)$ is small, we will show that the contribution from S_{21} and S_{23} is negligible, and that $S_{22} \leq (0.811 + o(1)) \cdot \sum_{n\sim x} b_n\Lambda(n)$.

2.5.1. *Sum S_{21}.* Here we have $k > x^{1/3+\gamma}$, so that by a crude estimate we get

$$S_{21} = \sum_{\substack{km\sim x \\ k\geq x^{1/3+\gamma} \\ m\geq z}} a_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1} \ll (\log x) \sum_{z\leq m\ll x^{2/3-\gamma}} \lambda(m) \sum_{k\sim x/m} a_{km} := S_{21}' = M_{21} + R_{21},$$

where

$$M_{21} := (\log x) X \sum_{z\leq m\ll x^{2/3-\gamma}} \lambda(m)g(m) \quad \text{and} \quad R_{21} := S_{21}' - M_{21}.$$

By Proposition 8 we get

$$R_{21} \ll_C x^{5/8}\log^{-C} x,$$

and by Proposition 10 we have

$$M_{21} \ll x^{5/8} L(1,\chi)\log^3 x.$$

Hence, we have

$$S_{21} \ll_C x^{5/8} L(1,\chi)\log^3 x + x^{5/8}\log^{-C} x.$$

2.5.2. *Sum S_{23}.* Recall that here $m \gg x^{2/3+2\gamma}$. By positivity we may drop the condition $(m, P(z)) = 1$. Writing

$$\lambda(m) = \sum_{cd=m} \chi(d)$$

we split the sum S_{23} into two ranges, $d \leq x^{1/3+\gamma}$ or $d > x^{1/3+\gamma}$. We get $S_{23} \leq S_{231} + S_{232}$, where

$$S_{231} := \sum_{z \leq k \leq x^{1/3-2\gamma}} \Lambda(k) \sum_{c \ll x^{2/3-\gamma}/k} \sum_{\substack{d \sim x/ck \\ d > x^{1/3+\gamma}}} \chi(d) a_{cdk},$$

$$S_{232} := \sum_{z \leq k \leq x^{1/3-2\gamma}} \Lambda(k) \sum_{d \leq x^{1/3+\gamma}} \chi(d) \sum_{c \sim x/dk} a_{cdk}.$$

By Proposition 9 we get (after applying a finer-than-dyadic decomposition similarly as with S_{11} to remove cross-conditions)

$$S_{231} \ll_C x^{5/8} \log^{-C} x.$$

By Propositions 8 and 10 we get (since the contribution from $(k, d) > 1$ is trivially negligible)

$$\begin{aligned} S_{232} &= X \sum_{z \leq k \ll x^{1/3-2\gamma}} \Lambda(k) \sum_{d \leq x^{1/3+\gamma}} \chi(d) g(dk) + O_C(x^{5/8} \log^{-C} x) \\ &\ll_C X \sum_{z \leq k \ll x^{1/3-2\gamma}} \Lambda(k) g(k) \sum_{d \leq x^{1/3+\gamma}} \chi(d) g(d) + x^{5/8} \log^{-C} x \\ &\ll_C x^{5/8} L(1, \chi) \log x + x^{5/8} \log^{-C} x. \end{aligned}$$

Combining the bounds, we have

$$S_{23} \ll_C x^{5/8} L(1, \chi) \log x + x^{5/8} \log^{-C} x.$$

2.5.3. *Sum S_{22}.* We have

$$S_{22} = \sum_{\substack{km \sim x \\ x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}}} a_{km} \Lambda(k) \lambda(m) 1_{(m, P(z))=1}.$$

It turns out that we can handle all parts except when m is a prime, so we write

$$S_{22} = \sum_{\substack{km \sim x \\ x^{1/3-2\gamma} < k \leq x^{1/3+\gamma} \\ m \notin \mathbb{P}}} a_{km} \Lambda(k) \lambda(m) 1_{(m, P(z))=1} + \sum_{\substack{kp \sim x \\ x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}}} a_{kp} \Lambda(k) \lambda(p) =: S_{221} + S_{222}.$$

In S_{221} we have $m = m_1 m_2$ for $m_1, m_2 \geq z$. Since $(m_1 m_2, P(z)) = 1$, the part where $(m_1, m_2) > 1$ trivially contributes at most $\ll z^{-1} x^{5/8} \log^{O(1)} x$ which is negligible. Hence, using $\lambda(m_1 m_2) = \lambda(m_1)\lambda(m_2)$ for $(m_1, m_2) = 1$ we get

$$S_{221} \leq \sum_{\substack{k m_1 m_2 \sim x \\ x^{1/3-2\gamma} < k \leq x^{1/3+\gamma} \\ m_1, m_2 \geq z}} a_{k m_1 m_2} \Lambda(k) \lambda(m_1) \lambda(m_2) 1_{(m_1 m_2, P(z))=1} + O_C(x^{5/8} \log^{-C} x).$$

We split this sum into two parts according to $k m_1 > x^{1/2}$ or $k m_1 \leq x^{1/2}$. In either case we get $m_j \ll x^{1/2}$ for some $j \in \{1, 2\}$. We combine the variables $\ell = k m_{2-j}$ and use $\lambda(m_{2-j}) \leq \tau(m_{2-j})$ to obtain by

Lemma 4

$$S_{221} \leq (\log x) \sum_{z \leq m \ll x^{1/2}} \lambda(m) \sum_{\ell \sim x/m} \tau(\ell) 1_{(\ell, P(z))=1} a_{\ell m} + O_C(x^{5/8} \log^{-C} x)$$
$$\ll_K (\log x) \sum_{z \leq m \ll x^{1/2}} \lambda(m) \sum_{d \leq x^{1/K}} \tau(d)^{O_K(1)} 1_{(d, P(z))=1} \sum_{\ell \sim x/dm} a_{d\ell m} + O_C(x^{5/8} \log^{-C} x).$$

By Proposition 8 we get (once we choose K large enough so that $\frac{1}{2} + 1/K < \frac{2}{3} - \gamma$)

$$S_{221} \ll_K M_{221} + O_C(x^{5/8} \log^{-C} x),$$

where

$$M_{221} = X(\log x) \sum_{z \leq m \ll x^{1/2}} \lambda(m) \sum_{d \leq x^{1/K}} \tau(d)^{O_K(1)} 1_{(d, P(z))=1} g(d) g(m),$$

since the contribution from the part the part $(d, m) > 1$ is negligible by a trivial bound. Thus, by Proposition 10 and Lemma 7 we have

$$M_{221} \ll_C X(\log x) \sum_{d \leq x^{1/K}} \tau(d)^{O_K(1)} g(d) 1_{(d, P(z))=1} \sum_{z \leq m \ll x^{1/2}} \lambda(m) g(m) \ll_C x^{5/8} L(1, \chi) \log^5 x.$$

Combining the above bounds we get

$$S_{221} \ll_C x^{5/8} L(1, \chi) \log^5 x + x^{5/8} \log^{-C} x,$$

so all that remains is to bound the sum S_{222}. The savings here will come from the fact that k is restricted to a fairly narrow range.

2.6. ***Bounding the error term S_{222}.*** We have

$$S_{222} := \sum_{\substack{kp \sim x \\ x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}}} a_{kp} \Lambda(k)(1 + \chi(p)).$$

We will apply the linear sieve upper bound to the nonnegative sequence

$$c_n := a_{kn}(1 + \chi(n))$$

with level of distribution $x^{2/3-\gamma}/k$ (note that by exploiting the cancellation from $\chi(n)$ we save a factor of 2 compared to using the trivial bound $\lambda(p) \leq 2$). For $(d, k) = 1$ define $R(d, k)$ by

$$\sum_{\substack{n \sim x/k \\ n \equiv 0(d)}} a_{kn}(1 + \chi(n)) = g(d) g(k) X + R(d, k).$$

Note that the contribution from sums with $(d, k) > 1$ is negligible by trivial estimates. Then by Lemma 5 with $D_k = x^{2/3-\gamma}/k$ we have

$$S_{222} \leq (1 + o(1)) M_{222} + R_{222},$$

where

$$M_{222} := X \sum_{x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}} \Lambda(k) g(k) 2e^{\gamma_1} \prod_{p \leq D_k} (1 - g(p))$$

and

$$R_{222} = \sum_{\substack{dk \leq x^{2/3-\gamma} \\ (d,k)=1}} \Lambda(k) |R(d,k)| \ll_C x^{5/8} \log^{-C} x$$

by Propositions 8 and 9. Applying Lemma 2 we get

$$M_{222} = (2 + o(1)) \sum_{n \sim x} b_n \Lambda(n) \sum_{x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}} \frac{\Lambda(k) g(k)}{\log(x^{2/3-\gamma}/k)} =: D(\gamma) \sum_{n \sim x} b_n \Lambda(n).$$

By the prime number theorem (for $p \equiv 1(4)$) we have (denoting $k = x^{\alpha}$)

$$\begin{aligned} D(\gamma) &\sim 2 \sum_{x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}} \frac{\Lambda(k) g(k)}{k \log(x^{2/3-\gamma}/k)} \\ &\sim 2 \sum_{x^{1/3-2\gamma} < k \leq x^{1/3+\gamma}} \frac{1}{k \log(x^{2/3-\gamma}/k)} \sim 2 \int_{1/3-2\gamma}^{1/3+\gamma} \frac{d\alpha}{2/3 - \gamma - \alpha} \\ &\sim 2 \log \frac{1 + 3\gamma}{1 - 6\gamma}. \end{aligned}$$

We have $D\left(\frac{1}{24}\right) < 0.811$. Since $\varepsilon > 0$ can be taken to be arbitrarily small, this implies

$$S_{222} \leq (0.811 + o(1)) \cdot \sum_{n \sim x} b_n \Lambda(n),$$

completing the proof of Theorem 1. □

2.7. ***Proof of the upper bound result.*** We now explain how to get the upper bound result in Theorem 1. By Section 2.4 we have by negativity of S_2

$$\begin{aligned} \sum_{n \sim x} a_n \Lambda(n) \leq S_1 + O_C(x^{5/8}/\log Cx) &= \sum_{n \sim x} b_n \lambda'(n) 1_{(n,P(z))=1} + O_C(x^{5/8}/\log Cx) \\ &= \sum_{n \sim x} b_n \Lambda(n) 1_{(n,P(z))=1} + M_2 + O_C(x^{5/8}/\log Cx), \end{aligned}$$

where by reversing the initial decomposition on the b_n-side (Section 2.3)

$$M_2 := \sum_{\substack{km \sim x \\ k,m \geq z}} b_{km} \Lambda(k) \lambda(m) 1_{(m,P(z))=1}$$

which is the same as S_2 but with a_n replaced by b_n. Now M_2 can be bounded similarly as S_2, except that we decompose with $\gamma = 0$ to get $M_2 = M_{21} + M_{23}$ with

$$M_{21} := \sum_{\substack{km\sim x\\ k>x^{1/3}\\ m\geq z}} b_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1}, \quad M_{23} := \sum_{\substack{km\sim x\\ k\leq x^{1/3}\\ m\geq z}} b_{km}\Lambda(k)\lambda(m)1_{(m,P(z))=1}.$$

By similar arguments as above for S_{21}, S_{23} we get

$$M_{21} + M_{23} \ll_C x^{5/8} L(1,\chi)\log^5 x + x^{5/8}/\log^C x,$$

since for b_n we have an exponent of distribution $> \frac{2}{3}$ by Propositions 8 and 9. That is, to prove the upper bound we only needed that a_n has an exponent of distribution $\frac{1}{2}+\varepsilon$ instead of $\frac{5}{8}-\varepsilon$.

3. Type I sums

In this section we will prove Propositions 8 and 9. The arguments are straightforward generalizations of the arguments in [Friedlander and Iwaniec 2006; 2005, Section 14]. Since it does not require much additional effort, we give the arguments in this section for the sequences $a^2 + b^{2k}$ for any $k \geq 1$, which yields the exponent of distribution $\frac{1}{2} + \frac{1}{2k} - \varepsilon$, as claimed in [Friedlander and Iwaniec 2006, below Theorem 4].

For the arguments in this section it is convenient for us to define $\lll$ to mean an inequality modulo logarithmic factors, that is, for two functions f and g with $g \geq 0$ we write $f(N) \lll g(N)$ if $f(N) \ll g(N)\log^{O(1)} N$. For parameters such as ε we write $f(N) \lll_\varepsilon g(N)$ to mean $f(N) \ll_\varepsilon g(N)\log^{O_\varepsilon(1)} N$.

Proposition 8 is a consequence of the following proposition, which we will prove in this section.

Proposition 11. *Let $M, L, D \gg 1$. Let $k \geq 1$ integer and let λ_ℓ be a coefficient such that $|\lambda_\ell| \leq 1_{\ell=n^k}$. Let ψ denote a fixed C^∞-smooth compactly supported function and denote $\psi_M(x) := \psi(x/M)$. Then for any divisor bounded $\alpha(d)$ and any real number $m_0 \lll M$ we have*

$$\sum_{d\sim D}\alpha(d)\Bigg(\sum_{\substack{(\ell,m)=1\\ \ell\sim L\\ \ell^2+m^2\equiv 0(d)}} \lambda_\ell\psi_M(m-m_0) - \int\psi_M(t)\,dt\frac{\varrho(d)}{d}\sum_{\substack{(\ell,d)=1\\ \ell\sim L}}\lambda_\ell\frac{\varphi(\ell)}{\ell}\Bigg)$$
$$\lll_\varepsilon M^\varepsilon(L+M)^{1/2}D^{1/2}L^{1/(2k)}.$$

Proof of Proposition 8 assuming Proposition 11. For the sequence b_n, which counts $n = a^2+b^2$ weighted with $b^{-1+1/k}/k$, we will apply similar arguments as below but with $k=1$, renormalizing the corresponding λ_ℓ appropriately. For a_n which counts $n = a^2+b^8$ we write $m = a$ and $\ell = b^4$, so that we are applying the above proposition with $k = 4$. Similarly as with the treatment of the sum S_{11}, we use a finer-than-dyadic decomposition to remove the cross-condition $m^2+\ell^2 \sim x$ that is, writing $\Delta = \log^{-B} x$ for some large B, we partition the sum into $\ll \Delta^{-2}\log^2 x$ parts where $\ell \in [L_0, L_0(1+\Delta)]$ and $m \in [M_0, M_0(1+\Delta)]$ with $L_0^2 + M_0^2 \sim x$ and $L_0, M_0 \ll \sqrt{x}$. In fact, we need to refine this decomposition so that for m we use a

C^∞-smooth finer-than-dyadic partition of unity. Then the resulting coefficients for m are C^∞-smooth functions of the form $\psi_M(m-M_0)$, where $M=M_0\Delta$ is the width of the window around $M_0 \ll \sqrt{x}$. We can now drop the condition $\ell^2+m^2\sim x$, with an error contribution bounded by $x^{5/8}\log^{-B+O(1)} x$ coming from the edges (where $L_0^2+M_0^2$ is in $[x(1+\Delta)^{-2}, x(1+\Delta)^2]$ or $[2x(1+\Delta)^{-2}, 2x(1+\Delta)^2]$). To see this, note that we have by Proposition 11 using $M_0, L_0 \ll x^{1/2}$

$$\sum_{d\sim D}|\alpha(d)| \sum_{\substack{m\\ \ell\in[L_0,L_0(1+\Delta)]\\ m^2+\ell\equiv 0(d)}} \lambda_\ell \psi_{\Delta M_0}(m-M_0) \ll_C x^{5/8}\log^{-C} x + \Delta^{1+1/k} L_0^{1/k} M_0 \sum_{d\sim D}\frac{|\alpha(d)|\varrho(d)}{d}$$
$$\ll_C x^{5/8}\log^{-C} x + x^{5/8}\log^{-(1+1/k)B+O(1)} x,$$

and that the number of edge cases is $\ll \log^{B+O(1)} x$, so that we save a factor of $\log^{O(1)-B/k} x$, which is sufficient for $B\gg k$.

We can now apply Proposition 11 in each of the parts separately. Note that then we have $L, M \ll x^{1/2}$ and $D \ll x^{5/8-\varepsilon}$, so that the error term is bounded by $x^{5/8-\varepsilon/4}$. To remove the condition $(\ell^2+m^2, q)=1$ implicit in Proposition 8 we may expand using the Möbius function to get

$$\sum_{\substack{\ell^2+m^2\equiv 0(d)\\ (\ell^2+m^2,q)=1}} = \sum_{f\mid q}\mu(f) \sum_{\ell^2+m^2\equiv 0(df)}$$

since $(d,q)=1$, and apply Proposition 11 with level $x^{5/8-\varepsilon}q \ll x^{5/8-\varepsilon/2}$.

Denote $\lambda_\ell^{(1)} = 1_{\ell=n^k}$ and $\lambda_\ell^{(2)} = k^{-1}\ell^{-1+1/k}$. Let $\tilde g(d)$ extend $g(d)$ to $(d,q)>1$, that is,

$$\tilde g(p^k) := \frac{\varrho(p^k)}{p^k}\left(1+\frac{1}{p}\right)^{-1}.$$

We still have to evaluate the main term in Proposition 11 to get (2-2). Recombining the finer-than-dyadic decomposition to a dyadic one for the variable ℓ, this follows we once show that for $j\in\{1,2\}$

$$\sum_{d\sim D}\alpha(d)\int \psi_M(t)dt\,\frac{\varrho(d)}{d}\sum_{\substack{(\ell,d)=1\\ \ell\sim L}}\lambda_\ell^{(j)}\frac{\varphi(\ell)}{\ell} = \sum_{d\sim D}\alpha(d)\tilde g(d)\sum_{\substack{(\ell,m)=1\\ \ell\sim L}}\lambda_\ell^{(2)}\psi_M(m-m_0) + O(x^{5/8-\eta}),$$

which follows easily once we show that

$$\sum_{d\sim D}\alpha(d)\int \psi_M(t)\,dt\,\frac{\varrho(d)}{d}\sum_{\substack{(\ell,d)=1\\ \ell\sim L}}\lambda_\ell^{(j)}\frac{\varphi(\ell)}{\ell}$$
$$= \sum_{d\sim D}\alpha(d)\frac{\varrho(d)}{d}\frac{\varphi(d)}{d}\prod_{p\mid d}(1-p^{-2})^{-1}\frac{1}{\zeta(2)}\sum_{\substack{m\\ \ell\sim L}}\lambda_\ell^{(2)}\psi_M(m-m_0) + O(x^{5/8-\eta}). \quad (3\text{-}1)$$

Define

$$H_d := \prod_{p\nmid d}(1-p^{-2}) = \sum_{(c,d)=1}\frac{\mu(c)}{c^2} = \frac{1}{\zeta(2)}\prod_{p\mid d}(1-p^{-2})^{-1}$$

and note that

$$\sum_{\ell\sim L}\lambda_\ell^{(1)}=(1+L^{-\varepsilon_k})\sum_{\ell\sim L}\lambda_\ell^{(2)}$$

and

$$\int\psi_M(t)dt=\sum_m\psi_M(m-m_0)+O_C(M^{-C}).$$

Then, since $M\lll x^{1/2}$, the claim (3-1) follows once we show

$$\sum_{d\le D}\frac{\alpha(d)\varrho(d)}{d}\Bigg(\sum_{\substack{(\ell,d)=1\\ \ell\sim L}}\lambda_\ell^{(j)}\frac{\varphi(\ell)}{\ell}-\frac{\varphi(d)}{d}H_d\sum_{\ell\sim L}\lambda_\ell^{(j)}\Bigg)\lll 1.$$

To show this, note also that

$$\frac{\varphi(\ell)}{\ell}=\sum_{c\,|\,\ell}\frac{\mu(c)}{c}.$$

Then for $\lambda_\ell=1_{\ell=n^k}$ (and similarly for $\lambda_\ell=k^{-1}\ell^{-1+1/k}$)

$$\sum_{d\le D}\frac{\alpha(d)\varrho(d)}{d}\Bigg(\sum_{\substack{(\ell,d)=1\\ \ell\sim L}}\lambda_\ell\frac{\varphi(\ell)}{\ell}-\frac{\varphi(d)}{d}H_d\sum_{\ell\sim L}\lambda_\ell\Bigg)$$

$$\begin{aligned}
&=\sum_{d\le D}\frac{\alpha(d)\varrho(d)}{d}\sum_{(c,d)=1}\frac{\mu(c)}{c}\Bigg(\sum_{\substack{(\ell,d)=1\\ \ell\sim L/c}}\lambda_{c\ell}-\frac{\varphi(d)}{cd}\sum_{\ell\sim L}\lambda_\ell\Bigg)\\
&=\sum_{d\le D}\frac{\alpha(d)\rho(d)}{d}\sum_{(c,d)=1}\frac{\mu(c)}{c}\sum_{e\,|\,d}\mu(e)\Bigg(\sum_{\ell\sim L/ce}\lambda_{ce\ell}-\frac{1}{ce}\sum_{\ell\sim L}\lambda_\ell\Bigg)\\
&=\sum_{d\le D}\frac{\alpha(d)\rho(d)}{d}\sum_{(c,d)=1}\frac{\mu(c)}{c}\sum_{e\,|\,d}\mu(e)\Bigg(\sum_{n\sim L^{1/k}/ce}1-\frac{1}{ce}\sum_{n\sim L^{1/k}}1\Bigg)\\
&\ll\sum_{d\le D}\frac{|\alpha(d)|\rho(d)}{d}\sum_{e\,|\,d}\Bigg(\sum_{c\ll L^{1/k}/e}\frac{1}{c}+\frac{L^{1/k}}{e}\sum_{c\gg L^{1/k}/e}\frac{1}{c^2}\Bigg)\lll 1
\end{aligned}$$

by writing $\ell=(nce)^k$ since ce is square free. □

Proposition 9 follows by a similar argument from the following (recall that a_n and b_n are supported on $(n,q)=1$).

Proposition 12. *Let $M,L,D\gg 1$. Let $k\ge 1$ integer and let λ_ℓ be a coefficient such that $|\lambda_\ell|\le 1_{\ell=n^k}$. Let ψ denote a fixed C^∞-smooth compactly supported function and denote $\psi_M(x):=\psi(x/M)$. Let χ denote a primitive quadratic Dirichlet character associated to a fundamental discriminant $\pm q$ with $q>1$. Then for any divisor bounded $\alpha(d)$ and any real number $m_0\lll M$ we have for some $\eta>0$*

$$\sum_{d\sim D}\alpha(d)\sum_{\substack{(\ell,m)=1\\ \ell\sim L\\ \ell^2+m^2\equiv 0(d)}}\lambda_\ell\psi_M(m-m_0)\chi(\ell^2+m^2)\lll_\varepsilon q^2M^\varepsilon(L+M)^{1/2}D^{1/2}L^{1/(2k)}+q^{-\eta}ML^{1/k}.$$

For the proof of Propositions 11 and 12 we need the following large sieve inequality; see [Friedlander and Iwaniec 2005, Lemma 14.4] for the proof.

Lemma 13. *Let $q \geq 1$. Then for any complex numbers α_n we have*

$$\sum_{\substack{d \sim D \\ (d,q)=1}} \sum_{\nu^2+1 \equiv 0(d)} \Big| \sum_{n \leq N} \alpha_n e_d(\nu n \bar{q}) \Big| \ll (Dq+N) \sum_{n \leq N} |\alpha_n|^2,$$

where $q\bar{q} \equiv 1(d)$.

We also require the Poisson summation formula.

Lemma 14 (truncated Poisson summation formula). *Let $\psi : \mathbb{R} \to \mathbb{C}$ be a fixed C^∞-smooth compactly supported function with $\|\psi\|_1 \leq 1$ and let $M \gg 1$. Fix a real number m_0. Let $d \geq 1$ be an integer. Then for any $\varepsilon > 0$ we have uniformly in m_0*

$$\sum_{m \equiv a(d)} \psi_M(m-m_0) = \int \sum_{0 \leq |h| \leq M^\varepsilon d/M} \psi_M(td-m_0)e(ht)e_d(-ah)dt + O_{C,\varepsilon}(M^{-C}).$$

Proof. Applying the Poisson summation formula we get

$$\begin{aligned}\sum_{m \equiv a(d)} \psi_M(m-m_0) &= \sum_n \psi_M(nd+a-m_0) = \sum_h \int \psi_M(td+a-m_0)e(ht)\,dt \\ &= \sum_h \int \psi_M(td-m_0)e(ht)e_d(-ha)du.\end{aligned}$$

by the change of variables $t \mapsto t-a/d$. For $|h| > M^\varepsilon d/M$ we can iterate integration by parts to show that the contribution from this part is $\ll_{C,\varepsilon} M^{-C}$. □

We also need the following Weil bound for character sums; see [Kowalski 2021, Theorem 3.1], for instance.

Lemma 15. *Let $q \geq 1$ and let χ be a primitive quadratic character of modulus q. Let $a, b \in \mathbb{Z}$ and $(a,q)=1$. Then*

$$\sum_{m(q)} \chi(am^2+b) \ll_\varepsilon (b,q)^{1/2} q^{1/2+\varepsilon}.$$

3.1. ***Proof of Propositions 11 and 12.*** We first note that there is a gap in the proof given in [Friedlander and Iwaniec 2005, Section 14], namely, the argument around their application of Poisson summation works only if the sum is restricted to $(\ell, q) = 1$. To fix this we must first bound the contribution $\ell = n^k$ which have a large factor whose prime factors divide q. Let $q_0 = q_0(n) = q_0(\ell)$ denote the smallest factor of n such that $(n/q_0, q) = 1$. The parts of the sums in Proposition 12 where $q_0 > q^\eta$ can be bounded trivially. To see this, note that by the divisor boundedness $\alpha(d)$ and Lemma 4 we have

$$\sum_{d\sim D}\alpha(d)\sum_{\substack{(\ell,m)=1\\ \ell\sim L\\ \ell^2+m^2\equiv 0(d)\\ q_0>q^\eta}}\lambda_\ell\psi_M(m-m_0)\chi(\ell^2+m^2)\ll\sum_{\substack{m\asymp m_0\\ n\sim L^{1/k}\\ q_0>q^\eta}}\tau(m^2+n^{2k})^{O(1)}$$
$$\ll\sum_{d\ll m_0^{1/2}}\tau(d)^{O(1)}\sum_{\substack{n\sim L^{1/k}\\ q_0>q^\eta}}\sum_{\substack{m\asymp m_0\\ m^2\equiv -n^{2k}(d)}}1$$
$$\lll M\sum_{d\ll m_0^{1/2}}\frac{\tau(d)^{O(1)}}{d}\sum_{\substack{n\sim L^{1/k}\\ q_0>q^\eta}}1\lll M\sum_{\substack{n\sim L^{1/k}\\ q_0>q^\eta}}1$$

and

$$\sum_{\substack{n\sim L^{1/k}\\ q_0>q^\eta}}1\le\sum_{\substack{q_0>q^\eta\\ p\,|\,q_0\Rightarrow p\,|\,q}}\sum_{n\sim L^{1/k}/q_0}\ll L^{1/k}\sum_{\substack{q_0>q^\eta\\ p\,|\,q_0\Rightarrow p\,|\,q}}q_0^{-1}\le q^{-\eta/2}L^{1/k}\prod_{p\,|\,q}(1-p^{-1/2})^{-1}\ll q^{-\eta/4}L^{1/k}.$$

Hence, we may assume that λ_ℓ is supported on $q_0(\ell)<q^\eta$ for some small $\eta>0$.

Note that since $d\,|\,\ell^2+m^2$, we may add the condition $(d,q)=1$ since otherwise $\chi(\ell^2+m^2)=0$. Expanding the condition $(\ell,m)=1$ using the Möbius function, we get

$$\sum_{\substack{d\sim D\\ (d,q)=1}}\alpha(d)\sum_{\substack{(\ell,m)=1\\ \ell\sim L\\ \ell^2+m^2\equiv 0(d)}}\lambda_\ell\psi_M(m-m_0)\chi(\ell^2+m^2)$$
$$=\sum_{\substack{b\ll LM\\ (b,q)=1}}\mu(b)\sum_{\substack{d\sim D\\ (d,q)=1}}\alpha(d)\sum_{\substack{\ell\sim L/b\\ (\ell,d)=1}}\lambda_{b\ell}\sum_{\substack{m\\ b^2(\ell^2+m^2)\equiv 0(d)}}\psi_{M/b}(m-m_0/b)\chi(\ell^2+m^2).$$

Writing $b_1=(d,b)$ and $b_2=b/b_1$ we get (absorbing $(d,b_2)=1$ into the coefficient $\alpha(d)$ and redefining $\alpha(d)$ as $\alpha(b_1d)$)

$$\sum_{\substack{b_1b_2\ll LM\\ (b_1b_2,q)=1}}\mu(b_1b_2)\sum_{\substack{d\sim D/b_1\\ (d,q)=1}}\alpha(d)\sum_{\substack{\ell\sim L/b\\ (\ell,d)=1}}\lambda_{b\ell}\sum_{\substack{m\\ \ell^2+m^2\equiv 0(d)}}\psi_{M/b}(m-m_0/b)\chi(\ell^2+m^2).$$

Let $q_\ell:=q_0^k$ so that $(q,\ell/q_\ell)=1$. Defining $\nu(d)$ and $\beta(q)$ so that $m\equiv\nu\ell(d)$ and $m\equiv\beta(\ell/q_\ell)(q)$ we get by the Chinese remainder theorem

$$m\equiv\nu\ell q\bar{q}+\beta(\ell/q_\ell)d\bar{d}(dq),$$

where the inverses $\bar{q}$ and $\bar{d}$ are computed modulo d and q, respectively. Using Lemma 14 we get for $H := M^{\varepsilon} b_2 D q / M$

$$\begin{aligned}
&\sum_{\substack{m \\ \ell^2+m^2 \equiv 0(d)}} \psi_{M/b}(m - m_0/b)\chi(\ell^2+m^2) \\
&\quad= \sum_{\substack{\nu(d) \\ \nu^2+1\equiv 0(d)}} \sum_{\beta(q)} \chi(\beta^2+q_\ell^2) \sum_{\substack{m \\ m \equiv \nu \ell q \bar{q} + \beta(\ell/q_\ell) d\bar{d} (dq)}} \psi_{M/b}(m-m_0/b) \\
&\quad= \sum_{\substack{\nu(d) \\ \nu^2+1\equiv 0(d)}} \sum_{\beta(q)} \chi(\beta^2+q_\ell^2) \int \sum_{0 \le |h| \le H} \psi_{M/b}(tdq - m_0/b) e(ht) e_d(-\nu h \ell \bar{q}) e_q(-\beta h (\ell/q_\ell)\bar{d})\, dt \\
&\qquad\qquad + O_{C,\varepsilon}(M^{-C}).
\end{aligned}$$

Making the change of variables and $\beta \mapsto \beta d$ this becomes

$$\int \sum_{0\le |h| \le H} \Bigl(\sum_{\beta(q)} \chi(\beta^2 d^2 + q_\ell^2) e_q(-\beta h (\ell/q_\ell)) \Bigr) \sum_{\nu^2+1 \equiv 0(d)} \psi_M(tbdq - m_0) e(ht) e_d(-\nu h \ell \bar{q})\, dt.$$

From $h = 0$ we get a total contribution

$$\sum_{\substack{b_1 b_2 \ll LM \\ (b_1 b_2, q)=1}} \mu(b_1 b_2) \sum_{\substack{d \sim D/b_1 \\ (d,q)=1}} \alpha(d) \varrho(d) \sum_{\substack{\ell \sim L/b \\ (\ell,d)=1}} \lambda_{b\ell} \frac{M}{b_2 d q} \int \psi(t) dt \sum_{\beta(q)} \chi(\beta^2 d^2 + q_\ell^2) \lll q^{-1/4} M L^{1/k}$$

by using the bound (Lemma 15)

$$\sum_{\beta(q)} \chi(\beta^2 d^2 + q_\ell^2) \ll_\varepsilon (q, q_\ell^2)^{1/2} q^{1/2+\varepsilon}$$

and the fact that $q_\ell = q_0^k \ll q^{\eta k}$ for some small η.

For $h \neq 0$ we can by symmetry restrict to $h < 0$. We first want to remove the cross-condition $\chi(\beta^2 d^2 + q_\ell^2)$ between the variables d and ℓ. To do this we fix the value of q_ℓ modulo q and split ℓ into congruence classes $q_\ell \equiv \gamma(q)$. Hence, we get for some $|c_{h,\ell}(t,q,\beta,\gamma)| \le 1$ and $|c_{h,\ell}(t,q)| \le 1$ that the total contribution from $h \neq 0$ is

$$\begin{aligned}
&\sum_{\gamma(q)} \sum_{\substack{b_1 b_2 \ll LM \\ (b_1 b_2, q)=1}} \mu(b_1 b_2) \int \sum_{\beta(q)} \sum_{\substack{d \sim D/b_1 \\ (d,q)=1}} \alpha(d) \chi(\beta^2 d^2 + \gamma^2) \\
&\qquad\qquad \times \sum_{\nu^2+1\equiv 0(d)} \sum_{\substack{\ell \sim L/b \\ (\ell,d)=1 \\ q_\ell \equiv \gamma(q)}} \lambda_{b\ell} \sum_{1 \le h \le H} c_{h,\ell}(t,q,\beta,\gamma) e_d(\nu h \ell \bar{q}) \psi_M(tbdq - m_0)\, dt \\
&\quad \lll q^2 \sum_{b_1 b_2 \ll LM} \int \sum_{\substack{d \sim D/b_1 \\ (d,q)=1}} |\alpha(d)| \sum_{\nu^2+1 \equiv 0(d)} \Bigl| \sum_{\substack{\ell \sim L/b \\ (\ell,d)=1}} \lambda_{b\ell} \sum_{1\le h \le H} c_{h,\ell}(t,q) e_d(\nu h \ell \bar{q}) \psi_M(tbdq - m_0) \Bigr| dt.
\end{aligned}$$

Note that $\psi_M(tbdq - m_0)$ vanishes outside $|tbdq - m_0| \ll M$. Hence, by $d \sim D/b_1$ and $m_0 \lll M$ the integral over t is supported on a fixed set $T(b_1, b_2)$ with measure bounded by $\lll M/b_2 q D$ so that by taking the maximal t the last expression is bounded by

$$\lll q \sum_{b_1 b_2 \ll LM} \frac{M}{b_2 D} \sum_{\substack{d \sim D/b_1 \\ (d,q)=1}} |\alpha(d)| \sum_{\nu^2+1 \equiv 0(d)} \Bigg| \sum_{\substack{\ell \sim L/b \\ (\ell,d)=1}} \lambda_{b\ell} \sum_{1 \le h \le H} c_{h,\ell} e_d(\nu h \ell \bar{q}) \Bigg|$$

for some coefficients $c_{h,\ell} = c_{h,\ell}(b_1, b_2, q, m_0)$ independent of d with $|c_{h,\ell}| \le 1$. Expanding the condition $(\ell, d) = 1$ this is bounded by

$$\frac{qM}{D} \sum_{b_1 b_2 \ll LM} \frac{1}{b_2} \sum_{c \ll DL} \sum_{\substack{d \sim D/b_1 c \\ (d,q)=1}} |\alpha(cd)| \sum_{\nu^2+1 \equiv 0(d)} \Bigg| \sum_{\ell \sim L/bc} \lambda_{bc\ell} \sum_{1 \le h \le H} c_{h,c\ell} e_d(\nu h c \ell \bar{q}) \Bigg|. \tag{3-2}$$

By Cauchy–Schwarz and Lemma 13 the sum over d is bounded by (denoting $H_1 := H/b_2 = M^\varepsilon Dq/M$)

$$\begin{aligned}
&\lll \frac{D^{1/2}}{(b_1 c)^{1/2}} \Bigg(\sum_{\substack{d \sim D/b_1 c \\ (d,q)=1}} \sum_{\nu^2+1 \equiv 0(d)} \Bigg| \sum_{\ell \sim L/bc} \lambda_{bc\ell} \sum_{1 \le h \le H} c_{h,c\ell} e_d(\nu h c \ell \bar{q}) \Bigg|^2 \Bigg)^{1/2} \\
&\ll \frac{D^{1/2}}{(b_1 c)^{1/2}} (Dq/b_1 c + HL/b)^{1/2} \Bigg(\sum_{1 \le j \ll H_1 L/c} \Bigg| \sum_{\substack{j = \ell h \\ \ell \sim L/bc}} \lambda_{bc\ell} \Bigg|^2 \Bigg)^{1/2} \\
&\ll \frac{1}{bc^{1/2}} (Dq + (DH_1 L)^{1/2}) \Bigg(\sum_{1 \le j \ll H_1 L/c} \Bigg| \sum_{\substack{j = \ell h \\ \ell \sim L/bc}} \lambda_{bc\ell} \Bigg|^2 \Bigg)^{1/2}.
\end{aligned}$$

By Cauchy–Schwarz we get (writing $m = bcj = bj'$ and $B := LM$ so that $1/b = j'/m \ll H_1 L/m$)

$$\begin{aligned}
\sum_{b \ll LM} \frac{\tau(b)}{b} \sum_{c \ll DL} \frac{1}{c^{1/2}} \Bigg(\sum_{1 \le j \ll H_1 L/c} \Bigg| \sum_{\substack{j = \ell h \\ \ell \sim L/bc}} \lambda_{bc\ell} \Bigg|^2 \Bigg)^{1/2} &\lll \Bigg(\sum_{\substack{j' \ll H_1 L \\ b \ll B}} \frac{1}{b} \tau(j') \Bigg| \sum_{\substack{j' = \ell h \\ \ell \sim L/b}} \lambda_{b\ell} \Bigg|^2 \Bigg)^{1/2} \\
&\ll \Bigg(\sum_{m \ll H_1 LB} \frac{H_1 L}{m} \tau(m)^2 \Bigg| \sum_{\substack{m = \ell h \\ \ell \sim L}} \lambda_\ell \Bigg|^2 \Bigg)^{1/2} \\
&\le \Bigg(H_1 L \sum_{n^k \sim L} \sum_{h \ll H_1 B} \frac{\tau(hn^k)^4}{hn^k} \Bigg)^{1/2} \\
&\le \Bigg(H_1 L \sum_{n^k \sim L} \sum_{h \ll H_1 B} \frac{\tau(h)^4 \tau(n)^{4k}}{hn^k} \Bigg)^{1/2} \\
&\lll H_1^{1/2} L^{1/2k}.
\end{aligned}$$

Hence, the final bound for (3-2) is

$$\begin{aligned}&\lll \frac{qM}{D}(Dq+(DH_1L)^{1/2})H_1^{1/2}L^{1/(2k)}\\&=qM^{\varepsilon}(Mq^{1/2}H_1^{1/2}L^{1/(2k)}+MH_1L^{1/2+1/(2k)}D^{-1/2})\\&=M^{\varepsilon}q^2(D^{1/2}M^{1/2}L^{1/(2k)}+D^{1/2}L^{1/2+1/(2k)})\end{aligned}$$

by using $H_1 = M^{\varepsilon}Dq/M$. □

4. A General version of the sieve

From our argument in Section 2 we can infer the following general result. We have not made an effort to minimize the assumptions or optimize the powers of logarithms.

Theorem 16. *Let x be large and let χ_D be a real primitive character associated to a fundamental discriminant $D = x^{o(1)}$ with $D \gg_C \log^C x$. Let a_n and b_n be nonnegative sequences supported on $(n, D) = 1$, and let $g(d)$ be the associated multiplicative function. Suppose that $g(d) \ll \tau(d)^{O(1)}/d$. Assume that g satisfies the assumptions of Lemma 5 and assume that Proposition 10 holds. Suppose that for any $z > x^{\varepsilon}$ we have*

$$\sum_{n\sim x} b_n\Lambda(n) = (1+o(1))\frac{1}{e^{\gamma_1}\log z}\prod_{p\le z}(1-g(p))\sum_{n\sim x} b_n$$

and

$$\sum_{k\sim z}\Lambda(k)g(k) = (1+o(1))\sum_{k\sim z}\frac{\Lambda(k)}{k}.$$

Suppose also that for some $\epsilon > 0$ we have the crude bounds

$$\sum_{n\sim x} a_n\Lambda(n)1_{(n,P(x^{\epsilon}))>1},\quad \sum_{n\sim x} b_n\Lambda(n)1_{(n,P(x^{\epsilon}))>1} = o\Big(\sum_{n\sim x}\Lambda(n)b_n\Big).$$

Suppose that the exponent of distribution is at least $\alpha = \frac{2}{3}-\gamma$ for some $\gamma < \frac{1}{6}$ (in the sense of Propositions 8 and 9). Then

$$\sum_{n\sim x}\Lambda(n)a_n \ge \Big(1-2\log\frac{1+3\gamma}{1-6\gamma} - O(L(1,\chi_D)\log^5 x) - o(1)\Big)\sum_{n\sim x}\Lambda(n)b_n.$$

Assuming that the exponent of distribution is at least $\frac{1}{2}+\varepsilon$ we have

$$\sum_{n\sim x}\Lambda(n)a_n \le (1+O(L(1,\chi_D)\log^5 x)+o(1))\sum_{n\sim x}\Lambda(n)b_n.$$

In particular, if $L(1,\chi_D) \le \log^{-100} D$ and $\exp(\log^{10} D) < x < \exp(\log^{16} D)$, then the lower bound is nontrivial as soon as the exponent of distribution satisfies

$$\alpha > \frac{1+\sqrt{e}}{1+2\sqrt{e}} = 0.61634\ldots.$$

Remark. With much more effort it is possible to get the same result as above with $L(1, \chi)\log x$ in place of $L(1, \chi)\log^5 x$, so that one only needs $L(1, \chi_D) = o(1/\log D)$.

Remark. Unfortunately the above theorem just misses out the next case $a^2 + b^{10}$, which has an exponent of distribution $\frac{3}{5} - \varepsilon$. Similarly as with the linear sieve, further improvements are possible if we make use of well-factorability of the weights [Friedlander and Iwaniec 2010, Chapter 12.7]. For example, the upper bound for the sum S_{222} can be improved if we are able to handle certain Type I/II sums (that is, Type I sums where the modulus is kd with d well-factorable). Note also that in S_{21} and S_{23} the weight factorizes and furthermore there is some smoothness available in the weight. Hence, assuming suitable arithmetic information (of Type I/II or Type I_2) we could handle some parts near the edges of S_{22} by a similar argument as for the sums S_{21} or S_{23}. Unfortunately we do not know how to carry this out for the sequence $a^2 + b^{10}$, but possibly sums of Kloosterman sums methods might be able to handle these sums. It is also unclear if the handling of the sum S_{222} is optimal but we have not found a way to improve this.

Remark. The ideas in this paper can be used also to the problem of primes in short intervals, to improve the result of Friedlander and Iwaniec [2004] which gives primes in intervals of length $x^{39/79} < x^{1/2}$ under the assumption of exceptional characters. The sieve argument is slightly different here since for this problem we can also utilize the available Type I/II and Type I_2 information furnished by the exponential sum estimates used for the problem of largest prime factor on short intervals [Baker and Harman 2009; Fouvry and Iwaniec 1989; Liu and Wu 1999]. The details will appear elsewhere.

Acknowledgements

I am grateful to my supervisor Kaisa Matomäki for helpful comments and encouragement. I also wish to thank Kyle Pratt for comments on an early version of this manuscript. I am grateful to the anonymous referee for helpful comments. During the work the author was funded by UTUGS Graduate School. Part of the article was also completed while I was working on projects funded by the Academy of Finland (project no. 319180) and the Emil Aaltonen foundation.

References

[Baker and Harman 2009] R. Baker and G. Harman, "Numbers with a large prime factor, II", pp. 1–14 in *Analytic number theory*, edited by W. W. L. Chen et al., Cambridge Univ. Press, 2009. MR Zbl

[Drappeau and Maynard 2019] S. Drappeau and J. Maynard, "Sign changes of Kloosterman sums and exceptional characters", *Proc. Amer. Math. Soc.* **147**:1 (2019), 61–75. MR Zbl

[Fouvry and Iwaniec 1989] E. Fouvry and H. Iwaniec, "Exponential sums with monomials", *J. Number Theory* **33**:3 (1989), 311–333. MR Zbl

[Friedlander and Iwaniec 1998a] J. Friedlander and H. Iwaniec, "Asymptotic sieve for primes", *Ann. of Math.* (2) **148**:3 (1998), 1041–1065. MR Zbl

[Friedlander and Iwaniec 1998b] J. Friedlander and H. Iwaniec, "The polynomial $X^2 + Y^4$ captures its primes", *Ann. of Math.* (2) **148**:3 (1998), 945–1040. MR Zbl

[Friedlander and Iwaniec 2003] J. B. Friedlander and H. Iwaniec, "Exceptional characters and prime numbers in arithmetic progressions", *Int. Math. Res. Not.* **2003**:37 (2003), 2033–2050. MR Zbl

[Friedlander and Iwaniec 2004] J. B. Friedlander and H. Iwaniec, "Exceptional characters and prime numbers in short intervals", *Selecta Math.* (*N.S.*) **10**:1 (2004), 61–69. MR Zbl

[Friedlander and Iwaniec 2005] J. B. Friedlander and H. Iwaniec, "The illusory sieve", *Int. J. Number Theory* **1**:4 (2005), 459–494. MR Zbl

[Friedlander and Iwaniec 2006] J. B. Friedlander and H. Iwaniec, "A polynomial divisor problem", *J. Reine Angew. Math.* **601** (2006), 109–137. MR Zbl

[Friedlander and Iwaniec 2010] J. Friedlander and H. Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications **57**, American Mathematical Society, Providence, RI, 2010. MR Zbl

[Granville 2021] A. Granville, "An alternative to Vaughan's identity", *Riv. Math. Univ. Parma* (*N.S.*) **12**:1 (2021), 119–124. MR Zbl

[Harman 2007] G. Harman, *Prime-detecting sieves*, London Mathematical Society Monographs Series **33**, Princeton University Press, 2007. MR Zbl

[Heath-Brown 1983] D. R. Heath-Brown, "Prime twins and Siegel zeros", *Proc. London Math. Soc.* (3) **47**:2 (1983), 193–224. MR Zbl

[Heath-Brown 2001] D. R. Heath-Brown, "Primes represented by x^3+2y^3", *Acta Math.* **186**:1 (2001), 1–84. MR Zbl

[Heath-Brown and Li 2017] D. R. Heath-Brown and X. Li, "Prime values of a^2+p^4", *Invent. Math.* **208**:2 (2017), 441–499. MR Zbl

[Heath-Brown and Moroz 2002] D. R. Heath-Brown and B. Z. Moroz, "Primes represented by binary cubic forms", *Proc. London Math. Soc.* (3) **84**:2 (2002), 257–288. MR Zbl

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications **53**, American Mathematical Society, Providence, RI, 2004. MR Zbl

[Kowalski 2021] E. Kowalski, "Exponential sums over finite fields: elementary methods", 2021, https://people.math.ethz.ch/~kowalski/exponential-sums-elementary.pdf. Zbl

[Linnik 1944] U. V. Linnik, "On the least prime in an arithmetic progression, I: The basic theorem", *Rec. Math. N.S.* **15(57)** (1944), 139–178. MR Zbl

[Liu and Wu 1999] H.-Q. Liu and J. Wu, "Numbers with a large prime factor", *Acta Arith.* **89**:2 (1999), 163–187. MR Zbl

[Maynard 2020] J. Maynard, "Primes represented by incomplete norm forms", *Forum Math. Pi* **8** (2020), art. id e3. MR Zbl

[Tenenbaum 2015] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, French ed., Graduate Studies in Mathematics **163**, American Mathematical Society, Providence, RI, 2015. MR Zbl

Communicated by Andrew Granville
Received 2022-08-31 Revised 2023-06-19 Accepted 2023-09-03

jori.merikoski@maths.ox.ac.uk *Mathematical Institute, University of Oxford, Oxford, United Kingdom*

ALGEBRA AND NUMBER THEORY 18:7 (2024)
https://doi.org/10.2140/ant.2024.18.1333

Polyhedral and tropical geometry of flag positroids

Jonathan Boretsky, Christopher Eur and Lauren Williams

A *flag positroid* of ranks $\boldsymbol{r} := (r_1 < \cdots < r_k)$ on $[n]$ is a flag matroid that can be realized by a real $r_k \times n$ matrix A such that the $r_i \times r_i$ minors of A involving rows $1, 2, \ldots, r_i$ are nonnegative for all $1 \le i \le k$. In this paper we explore the polyhedral and tropical geometry of flag positroids, particularly when $\boldsymbol{r} := (a, a+1, \ldots, b)$ is a sequence of consecutive numbers. In this case we show that the nonnegative tropical flag variety $\mathrm{TrFl}^{\ge 0}_{\boldsymbol{r},n}$ equals the nonnegative flag Dressian $\mathrm{FlDr}^{\ge 0}_{\boldsymbol{r},n}$, and that the points $\boldsymbol{\mu} = (\mu_a, \ldots, \mu_b)$ of $\mathrm{TrFl}^{\ge 0}_{\boldsymbol{r},n} = \mathrm{FlDr}^{\ge 0}_{\boldsymbol{r},n}$ give rise to coherent subdivisions of the flag positroid polytope $P(\underline{\boldsymbol{\mu}})$ into flag positroid polytopes. Our results have applications to Bruhat interval polytopes: for example, we show that a complete flag matroid polytope is a Bruhat interval polytope if and only if its (≤ 2)-dimensional faces are Bruhat interval polytopes. Our results also have applications to realizability questions. We define a *positively oriented flag matroid* to be a sequence of positively oriented matroids $(\chi_1, \ldots, \chi_k)$ which is also an oriented flag matroid. We then prove that every positively oriented flag matroid of ranks $\boldsymbol{r} = (a, a+1, \ldots, b)$ is realizable.

1. Introduction

In recent years there has been a great deal of interest in the tropical Grassmannian [Speyer and Sturmfels 2004; Herrmann et al. 2009; 2014; Cachazo et al. 2019; Bossinger 2021], and matroid polytopes and their subdivisions [Speyer 2008; Ardila et al. 2010; Early 2022], as well as “positive” [Postnikov 2007; Speyer and Williams 2005; 2021; Oh 2008; Ardila et al. 2016; Le and Fraser 2019; Lukowski et al. 2023; Arkani-Hamed et al. 2021b] and “flag” [Tsukerman and Williams 2015; Brandt et al. 2021; Bossinger et al. 2017; Jarra and Lorscheid 2024; Joswig et al. 2023; Boretsky 2022] versions of the above objects. The aim

MSC2020: 05Exx.
Keywords: flag variety, flag positroid, positroid polytope, Bruhat interval polytope, flag Dressian, positively oriented flag matroid, tropical flag variety.

of this paper is to illustrate the beautiful relationships between the nonnegative tropical flag variety, the nonnegative flag Dressian, and flag positroid polytopes and their subdivisions, unifying and generalizing some of the existing results. We will particularly focus on the case of flag varieties (respectively, flag positroids) consisting of subspaces (respectively, matroids) of *consecutive* ranks. This case includes both Grassmannians and complete flag varieties.

For positive integers n and d with $d < n$, we let $[n]$ denote the set $\{1, \dots, n\}$ and we let $\binom{[n]}{d}$ denote the collection of all d-element subsets of $[n]$. Given a subset $S \subseteq [n]$ we let $\boldsymbol{e}_S$ denote the sum of standard basis vectors $\sum_{i \in S} \boldsymbol{e}_i$. For a collection $\mathcal{B} \subset \binom{[n]}{d}$, we let

$$P(\mathcal{B}) = \text{the convex hull of } \{\boldsymbol{e}_B : B \in \mathcal{B}\} \text{ in } \mathbb{R}^n.$$

The collection $\mathcal{B}$ is said to define a *matroid M of rank d on* $[n]$ if every edge of the polytope $P(\mathcal{B})$ is parallel to $\boldsymbol{e}_i - \boldsymbol{e}_j$ for some $i \neq j \in [n]$. In this case, we call $\mathcal{B}$ the set of *bases* of M, and define the *matroid polytope* $P(M)$ of M to be the polytope $P(\mathcal{B})$. When $\mathcal{B}$ indexes the nonvanishing Plücker coordinates of an element A of the Grassmannian $\mathrm{Gr}_{d,n}(\mathbb{C})$, we say that A *realizes* M, and it is well-known that $P(\mathcal{B})$ is the moment map image of the closure of the torus orbit of A in the Grassmannian [Gelfand et al. 1987]. We assume familiarity with the fundamentals of matroid theory as in [Oxley 2011] and [Borovik et al. 2003].

The above definition of matroid in terms of its polytope is due to [Gelfand et al. 1987]. Flag matroids are natural generalizations of matroids that admit the following polytopal definition.

Definition 1.1 [Borovik et al. 2003, Corollary 1.13.5 and Theorem 1.13.6]. Let $\boldsymbol{r} = (r_1, \dots, r_k)$ be a sequence of increasing integers in $[n]$. A *flag matroid* of ranks $\boldsymbol{r}$ on $[n]$ is a sequence $\boldsymbol{M} = (M_1, \dots, M_k)$ of matroids of ranks $(r_1, \dots, r_k)$ on $[n]$ such that all vertices of the polytope

$$P(\boldsymbol{M}) = P(M_1) + \cdots + P(M_k), \text{ the Minkowski sum of matroid polytopes,}$$

are equidistant from the origin. The polytope $P(\boldsymbol{M})$ is called the *flag matroid polytope* of $\boldsymbol{M}$; we sometimes say it is a flag matroid polytope of *rank* $\boldsymbol{r}$.

Flag matroids are exactly the type A objects in the theory of Coxeter matroids [Gelfand and Serganova 1987; Borovik et al. 2003]. Just as a realization of a matroid is a point in a Grassmannian, a realization of a flag matroid is a point in a flag variety. More concretely, a *realization* of a flag matroid of ranks $(r_1, \dots, r_k)$ is an $r_k \times n$ matrix A over a field such that for each $1 \leq i \leq k$, the $r_i \times n$ submatrix of A formed by the first r_i rows of A is a realization of M_i. For an equivalent definition of flag matroids in terms of Plücker relations on partial flag varieties; see [Jarra and Lorscheid 2024, Proposition A].

There is a notion of moment map for any flag variety (indeed for any generalized partial flag variety G/P) [Gelfand and Serganova 1987; Borovik et al. 2003]. When a flag matroid $\boldsymbol{M}$ can be realized by a point A in the flag variety, then its matroid polytope $P(\boldsymbol{M})$ is the moment map image of the closure of the torus orbit of A in the flag variety [Gelfand and Serganova 1987; Borovik et al. 2003, Corollary 1.13.5].

There are natural "positive" analogues of matroids, flag matroids, and their polytopes.

Definition 1.2. Let $\boldsymbol{r} = (r_1, \dots, r_k)$ be a sequence of increasing integers in $[n]$. We say that a flag matroid $(M_1, \dots, M_k)$ of ranks $\boldsymbol{r}$ on $[n]$ is a *flag positroid* if it has a realization by a real matrix A such that the $r_i \times n$ submatrix of A formed by the first r_i rows of A has all nonnegative minors for each $1 \le i \le k$.

We refer to the flag matroid polytope of a flag positroid as a *flag positroid polytope*. It follows from our definition above that flag positroids are realizable.

Setting $k = 1$ in Definition 1.2 gives the well-studied notion of *positroids* and *positroid polytopes* [Postnikov 2007; Oh 2008; Ardila et al. 2016]. Therefore each flag positroid is a sequence of positroids.

In recent years it has been gradually understood that the tropical geometry of the Grassmannian and flag variety, and in particular, the *Dressian* and *flag Dressian*, are intimately connected to (flag) matroid polytopes and their subdivisions [Speyer 2008; Herrmann et al. 2009; Brandt et al. 2021]; see also [Maclagan and Sturmfels 2015, Section 4]. A particularly attractive point of view, which sheds light on the above connections, is the theory of (*flag*) *matroids over hyperfields* [Baker and Bowler 2019; Jarra and Lorscheid 2024]. In this framework, the Dressian and flag Dressian are the Grassmannian and flag variety over the *tropical hyperfield*, while matroids and flag matroids are the points of the Grassmannian and flag variety over the *Krasner hyperfield*.

The tropical geometry of the *positive* Grassmannian and flag variety are particularly nice: the positive tropical Grassmannian equals the positive Dressian, whose cones in turn parametrize subdivisions of the hypersimplex into positroid polytopes [Speyer and Williams 2005; 2021; Lukowski et al. 2023; Arkani-Hamed et al. 2021b]. And the positive tropical complete flag variety equals the positive complete flag Dressian, whose cones parametrize subdivisions of the permutohedron into *Bruhat interval polytopes* [Boretsky 2022; Joswig et al. 2023]. Theorem A below unifies and generalizes the above results.

Definition 1.3. Let $\mathbb{T} = \mathbb{R} \cup \{\infty\}$ be the set underlying the tropical hyperfield, endowed with the topology such that $-\log : \mathbb{R}_{\ge 0} \to \mathbb{T}$ is a homeomorphism. Given a point $w \in \mathbb{T}^{\binom{[n]}{r}}$, we define the *support* of w to be $\underline{w} = \left\{S \in \binom{[n]}{r} : w_S \ne \infty\right\}$. When $\underline{w}$ is the set of bases of a matroid, we identify $\underline{w}$ with that matroid. Let $\mathbb{P}(\mathbb{T}^{\binom{[n]}{r}})$ be the tropical projective space of $\mathbb{T}^{\binom{[n]}{r}}$, which is defined as $(\mathbb{T}^{\binom{[n]}{r}} \setminus \{(\infty, \dots, \infty)\})/\sim$, where $w \sim w'$ if $w = w' + (c, \dots, c)$ for some $c \in \mathbb{R}$.

Our main result is the following.

Theorem A. Suppose $\boldsymbol{r}$ is a sequence of consecutive integers $(a, \dots, b)$ for some $1 \le a \le b \le n$. Then, for $\boldsymbol{\mu} = (\mu_a, \dots, \mu_b) \in \prod_{i=a}^{b} \mathbb{P}(\mathbb{T}^{\binom{[n]}{i}})$, the following statements are equivalent:

(a) $\boldsymbol{\mu} \in \mathrm{TrFl}_{\boldsymbol{r},n}^{\ge 0}$, the nonnegative tropicalization of the flag variety, i.e., the closure of the coordinate-wise valuation of points in $\mathrm{Fl}_{\boldsymbol{r},n}(\mathcal{C}_{\ge 0})$.

(b) $\boldsymbol{\mu} \in \mathrm{FlDr}_{\boldsymbol{r},n}^{\ge 0}$, the nonnegative flag Dressian, i.e., the "solutions" to the positive-tropical Grassmann–Plücker and incidence-Plücker relations.

(c) Every face in the coherent subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ of the polytope $P(\underline{\boldsymbol{\mu}}) = P(\underline{\mu_1}) + \cdots + P(\underline{\mu_k})$ induced by $\boldsymbol{\mu}$ is a flag positroid polytope (of rank $\boldsymbol{r}$).

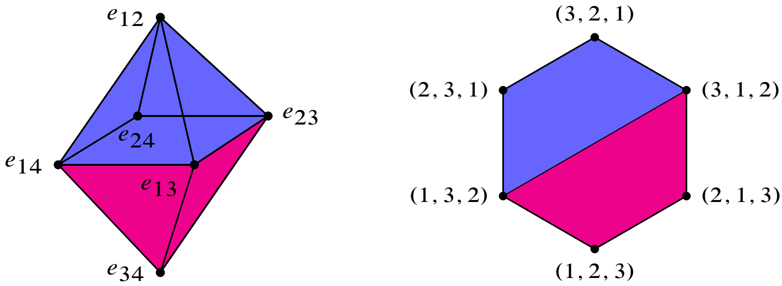

Figure 1. Left-hand side: the coherent subdivision of the hypersimplex into positroid polytopes induced by a point $\boldsymbol{\mu} \in \mathrm{Dr}_{2,4}^{\geq 0}$ such that $\mu_{13} + \mu_{24} = \mu_{23} + \mu_{14} < \mu_{12} + \mu_{34}$. Right-hand side: the coherent subdivision of the permutohedron into Bruhat interval polytopes induced by a point $\boldsymbol{\mu} \in \mathrm{FlDr}_{(1,2,3),3}^{\geq 0}$ such that $\mu_2 + \mu_{13} = \mu_1 + \mu_{23} < \mu_3 + \mu_{12}$.

(d) Every face of dimension at most 2 in the subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ of $P(\underline{\boldsymbol{\mu}})$ is a flag positroid polytope (of rank $\boldsymbol{r}$).

(e) The support $\underline{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ is a flag matroid, and $\boldsymbol{\mu}$ satisfies every three-term positive-tropical incidence relation (respectively, every three-term positive-tropical Grassmann–Plücker relation) when $a < b$ (respectively, $a = b$).

For the definitions of the objects in Theorem A, see Proposition 3.6 for (a), Definition 3.3 for (b), Definition 5.1 for (c), and Definition 3.8 for (e).

We note that if $\boldsymbol{r} = (d)$ is a single integer, Theorem A describes the relationship between the nonnegative tropical Grassmannian, the nonnegative Dressian, and subdivisions of positroid polytopes (e.g., the hypersimplex, if $\boldsymbol{\mu}$ has no coordinates equal to ∞) into positroid polytopes. And when $\boldsymbol{r} = (1, 2, \ldots, n)$, Theorem A describes the relationship between the nonnegative tropical complete flag variety, the nonnegative complete flag Dressian, and subdivisions of Bruhat interval polytopes (e.g., the permutohedron, if $\boldsymbol{\mu}$ has no coordinates equal to ∞) into Bruhat interval polytopes. We illustrate this relationship in the case where $\boldsymbol{\mu}$ has no coordinates equal to ∞ in Figure 1.

We prove the equivalence (a)$\Longleftrightarrow$(b) in Section 3.2, the implications (b)$\Longrightarrow$(c)$\Longrightarrow$(d)$\Longrightarrow$(e) in Section 5.2, and the implication (e)$\Longrightarrow$(b) in Section 6.1.

Theorem A has applications to flag positroid polytopes.

Corollary 1.4. *For a flag matroid $\boldsymbol{M} = (M_a, M_{a+1}, \ldots, M_b)$ of consecutive ranks $\boldsymbol{r} = (a, a+1, \ldots, b)$, its flag matroid polytope $P(\boldsymbol{M})$ is a flag positroid polytope if and only if its (≤ 2)-dimensional faces are flag positroid polytopes (of rank $\boldsymbol{r}$).*

Proof. Let $\boldsymbol{\mu} = (\mu_a, \ldots, \mu_b)$, with $\mu_i \in \{0, \infty\}^{\binom{[n]}{i}}$, where the coordinates of each μ_i are either 0 or ∞ based on whether we have a basis or nonbasis of M_i. This gives rise to the trivial subdivision of the corresponding flag matroid polytope $P(\underline{\boldsymbol{\mu}}) = P(\boldsymbol{M})$. The result now follows from the equivalence of (c) and (d) in Theorem A. □

In the Grassmannian case, that is, the case that $\boldsymbol{r} = (d)$ is a single integer, the flag positroid polytopes of rank $\boldsymbol{r}$ are precisely the positroid polytopes, and in that case the above corollary appeared as [Lukowski et al. 2023, Theorem 3.9].

Also in the Grassmannian case, the objects discussed in Theorem A are closely related to questions of realizability. Note that by definition, every positroid has a realization by a matrix whose Plücker coordinates are nonnegative, so it naturally defines a *positively oriented matroid*, that is, an oriented matroid defined by a chirotope whose values are all 0 and 1. Conversely, every positively oriented matroid can be realized by a positroid: this was first proved in [Ardila et al. 2017] using positroid polytopes, and subsequently in [Speyer and Williams 2021], using the positive tropical Grassmannian. It is natural then to ask if there is an analogous realizability statement in the setting of flag matroids, and if one can characterize when a sequence of positroids forms a flag positroid; indeed, this was part of the motivation for [Benedetti et al. 2022], which studied quotients of uniform positroids. Note however that questions of realizability for flag matroids are rather subtle: for example, a sequence of positroids that form a realizable flag matroid can still fail to be a flag positroid (see Example 4.4). By working with *oriented* flag matroids, we give an answer to this realizability question in Corollary 1.5, in the case of consecutive ranks.

Corollary 1.5. *Suppose $(M_1, \dots, M_k)$ is a sequence of positroids on $[n]$ of consecutive ranks $\boldsymbol{r} = (r_1, \dots, r_k)$. Then, when considered as a sequence of positively oriented matroids, $(M_1, \dots, M_k)$ is a flag positroid if and only if it is an oriented flag matroid.*

We define a *positively oriented flag matroid* to be a sequence of positively oriented matroids $(\chi_1, \dots, \chi_k)$ which is also an oriented flag matroid. Corollary 1.5 then says that every positively oriented flag matroid of consecutive ranks $(r_1, \dots, r_k)$ is realizable.

See Section 4.1 for a review of oriented matroids and oriented flag matroids. Note that because a positroid by definition has a realization over $\mathbb{R}$ with all nonnegative minors, it defines a positively oriented matroid. In Section 4.2, we deduce Corollary 1.5 from the equivalence of (a) and (b) in Theorem A. Another proof using ideas from discrete convex analysis is sketched in Remark 4.7. In both proofs, the consecutive ranks condition is indispensable. We do not know whether the corollary holds if $\boldsymbol{r} = (r_1, \dots, r_k)$ fails to satisfy the consecutive rank condition.

Question 1.6. Suppose M and M' are positroids on $[n]$ such that, when considered as positively oriented matroids, they form an oriented flag matroid (M, M'). Is (M, M') then a flag positroid?

One may attempt to answer the question by appealing to the fact [Kung 1986, Exercise 8.14] that for a flag matroid (M, M'), one can always find a flag matroid $(M_1, \dots, M_k)$ of consecutive ranks such that $M_1 = M$ and $M_k = M'$. However, the analogous statement fails for flag positroids; see Example 4.6 for an example of a flag positroid (M, M') on $[4]$ of ranks $(1, 3)$ such that there is no flag positroid (M, M_2, M') with rank of M_2 equal to 2.

The consecutive rank condition has recently shown up in [Bloch and Karp 2023], which studied the relation between two notions of total positivity for partial flag varieties, “Lusztig positivity” and “Plücker positivity" (see Section 2.1). In particular, the Plücker positive subset of a partial flag variety agrees with the Lusztig positive subset of the partial flag variety precisely when the flag variety consists of linear subspaces of consecutive ranks [Bloch and Karp 2023, Theorem 1.1].

A *generalized Bruhat interval polytope* [Tsukerman and Williams 2015, Definition 7.8 and Lemma 7.9] can be defined as the moment map image of the closure of the torus orbit of a point A in the nonnegative part $(G/P)^{\geq 0}$ (in the sense of Lusztig) of a flag variety G/P. When $\boldsymbol{r}$ is a sequence of consecutive integers, it then follows from [Bloch and Karp 2023] that generalized Bruhat interval polytopes for $\mathrm{Fl}_{\boldsymbol{r};n}^{\geq 0}$ are precisely the flag positroid polytopes of ranks $\boldsymbol{r}$. In the complete flag case, a generalized Bruhat interval polytope is just a *Bruhat interval polytope* [Kodama and Williams 2015], that is, the convex hull of the permutation vectors $(z(1), \dots, z(n))$ for all permutations z lying in some Bruhat interval $[u, v]$.

We can now restate Corollary 1.4 as follows.

Corollary 1.7. *For a flag matroid on* $[n]$ *of consecutive ranks* $\boldsymbol{r}$, *its flag matroid polytope is a generalized Bruhat interval polytope if and only if its* (≤ 2)*-dimensional faces are generalized Bruhat interval polytopes. In particular, for a complete flag matroid on* $[n]$, *its flag matroid polytope is a Bruhat interval polytope if and only if its* (≤ 2)*-dimensional faces are Bruhat interval polytopes.*

The structure of this paper is as follows. In Section 2, we give background on total positivity and Bruhat interval polytopes. In Section 3, we introduce the tropical flag variety, the flag Dressian, and nonnegative analogues of these objects; we also prove the equivalence of (a) and (b) in Theorem A. In Section 4 we discuss positively oriented flag matroids and prove Corollary 1.5. In Section 5 we explain the relation between the flag Dressian and subdivisions of flag matroid polytopes, then prove that (b)$\Longrightarrow$(c)$\Longrightarrow$(d)$\Longrightarrow$(e) in Theorem A. We prove some key results about three-term incidence and Grassmann–Plücker relations in Section 6, which allow us to prove (e)$\Longrightarrow$(b) in Theorem A. Section 7 concerns projections of positive Richardsons to positroids: we characterize the positroid constituents of complete flag positroids, and we characterize when two adjacent-rank positroids form an oriented matroid quotient, or equivalently, can appear as constituents of a complete flag positroid. In Section 8, we make some remarks about the various fan structures for $\mathrm{TrFl}_{\boldsymbol{r};n}^{>0}$; we then discuss fan structures and coherent subdivisions in the case of the Grassmannian and complete flag variety, including a detailed look at the case of $\mathrm{TrFl}_4^{\geq 0}$.

2. Background on total positivity and Bruhat interval polytopes

2.1. ***Background on total positivity.*** Let $n \in \mathbb{Z}_+$ and let $\boldsymbol{r} = \{r_1 < \cdots < r_k\} \subseteq [n]$. For a field $\Bbbk$, let $G = \mathrm{GL}_n(\Bbbk)$, and let $\mathrm{P}_{\boldsymbol{r};n}(\Bbbk)$ denote the parabolic subgroup of G of block upper-triangular matrices with diagonal blocks of sizes $r_1, r_2 - r_1, \dots, r_k - r_{k-1}, n - r_k$. We define the *partial flag variety*

$$\mathrm{Fl}_{\boldsymbol{r};n}(\Bbbk) := \mathrm{GL}_n(\Bbbk)/\,\mathrm{P}_{\boldsymbol{r};n}(\Bbbk).$$

As usual, we identify $\mathrm{Fl}_{\boldsymbol{r};n}(\Bbbk)$ with the variety of partial flags of subspaces in $\Bbbk^n$:

$$\mathrm{Fl}_{\boldsymbol{r};n}(\Bbbk) = \{(V_1 \subset \cdots \subset V_k) : V_i \text{ a linear subspace of } \Bbbk^n \text{ of dimension } r_i \text{ for } i = 1, \dots, k\}.$$

We write $\mathrm{Fl}_n(\Bbbk)$ for the complete flag variety $\mathrm{Fl}_{1,2,\dots,n;n}(\Bbbk)$. Note that $\mathrm{Fl}_n(\Bbbk)$ can be identified with $\mathrm{GL}_n(\Bbbk)/B(\Bbbk)$, where $B(\Bbbk)$ is the subgroup of upper-triangular matrices. There is a natural projection π from $\mathrm{Fl}_n(\Bbbk)$ to any partial flag variety by simply forgetting some of the subspaces.

If A is an $r_k \times n$ matrix such that V_{r_i} is the span of the first r_i rows, we say that A is a *realization* of $V := (V_1 \subset \cdots \subset V_k) \in \mathrm{Fl}_{\boldsymbol{r};n}$. Given any realization A of V and any $1 \le i \le k$, we have the *Plücker coordinates* or *flag minors* $p_I(A)$ where $I \in \binom{[n]}{r_i}$; concretely, $p_I(A)$ is the determinant of the submatrix of A occupying the first r_i rows and columns I. This gives the Plücker embedding of $\mathrm{Fl}_{\boldsymbol{r};n}(\Bbbk)$ into $\mathbb{P}^{\binom{[n]}{r_1}-1} \times \cdots \times \mathbb{P}^{\binom{[n]}{r_k}-1}$ taking V to $((p_I(A))_{I\in\binom{[n]}{r_1}}, \dots, (p_I(A))_{I\in\binom{[n]}{r_k}})$.

We now let $\Bbbk$ be the field $\mathbb{R}$ of real numbers. With this understanding, we will often drop the $\mathbb{R}$ from our notation.

Definition 2.1. We say that a real matrix is *totally positive* if all of its minors are positive. We let $\mathrm{GL}_n^{>0}$ denote the subset of GL_n of totally positive matrices.

There are two natural ways to define positivity for partial flag varieties. The first notion comes from work of Lusztig [1994]. The second notion uses Plücker coordinates, and was initiated in work of Postnikov [2007].

Definition 2.2. We define the (*Lusztig*) *positive part* of $\mathrm{Fl}_{\boldsymbol{r};n}$, denoted by $\mathrm{Fl}_{\boldsymbol{r};n}^{>0}$, as the image of $\mathrm{GL}_n^{>0}$ inside $\mathrm{Fl}_{\boldsymbol{r};n} = \mathrm{GL}_n \,/\, \mathrm{P}_{\boldsymbol{r};n}$. We define the (*Lusztig*) *nonnegative part* of $\mathrm{Fl}_{\boldsymbol{r};n}$, denoted by $\mathrm{Fl}_{\boldsymbol{r};n}^{\ge 0}$, as the closure of $\mathrm{Fl}_{\boldsymbol{r};n}^{>0}$ in the Euclidean topology.

We define the *Plücker positive part* (respectively, *Plücker nonnegative part*) of $\mathrm{Fl}_{\boldsymbol{r};n}$ to be the subset of $\mathrm{Fl}_{\boldsymbol{r};n}$ where all Plücker coordinates are positive (respectively, nonnegative).[1]

It is well-known that the Lusztig positive part of $\mathrm{Fl}_{\boldsymbol{r};n}$ is a subset of the Plücker positive part of $\mathrm{Fl}_{\boldsymbol{r};n}$, and that the two notions agree in the case of the Grassmannian [Talaska and Williams 2013, Corollary 1.2]. The two notions also agree in the case of the complete flag variety [Boretsky 2022, Theorem 5.21]. More generally, we have the following.

Theorem 2.3 [Bloch and Karp 2023, Theorem 1.1]. *The Lusztig positive* (*respectively*, *Lusztig nonnegative*) *part of* $\mathrm{Fl}_{\boldsymbol{r};n}$ *equals the Plücker positive* (*respectively*, *Plücker nonnegative*) *part of* $\mathrm{Fl}_{\boldsymbol{r};n}$ *if and only if the set* $\boldsymbol{r}$ *consists of consecutive integers.*

See [Bloch and Karp 2023, Section 1.4] for more references and a nice discussion of the history. Since in this paper we will be mainly studying the case where $\boldsymbol{r}$ consists of consecutive integers, we will use the two notions interchangeably when there is no ambiguity.

Let B and B^- be the opposite Borel subgroups consisting of upper-triangular and lower-triangular matrices. Let $W = S_n$ be the Weyl group of GL_n. Given $u, v \in W$, the *Richardson variety* is the intersection of opposite Bruhat cells

$$\mathcal{R}_{u,v} := (B\dot{v}B/B) \cap (B^-\dot{u}B/B),$$

[1]The reader who is concerned about the fact that we are working with projective coordinates can replace "all Plücker coordinates are positive" by "all Plücker coordinates are nonzero and have the same sign".

where $\dot{v}$ and $\dot{u}$ denote permutation matrices in GL_n representing v and u. It is well-known that $\mathcal{R}_{u,v}$ is nonempty precisely when $u \le v$ in Bruhat order, and in that case is irreducible of dimension $\ell(v) - \ell(u)$.

For $u, v \in W$ with $u \le v$, let $\mathcal{R}_{u,v}^{>0} := \mathcal{R}_{u,v} \cap \mathrm{Fl}_n^{\ge 0}$ be the positive part of the Richardson variety. Lusztig conjectured and Rietsch [1998] proved that

$$\mathrm{Fl}_n^{\ge 0} = \bigsqcup_{u \le v} \mathcal{R}_{u,v}^{>0} \tag{1}$$

is a cell decomposition of $\mathrm{Fl}_n^{\ge 0}$. Moreover, Rietsch showed that one obtains a cell decomposition of the nonnegative partial flag variety $\mathrm{Fl}_{\boldsymbol{r};n}^{\ge 0}$ by projecting the cell decomposition of $\mathrm{Fl}_n^{\ge 0}$ [Rietsch 1998; 2006, Section 6]. Specifically, if we let $W_{\boldsymbol{r}}$ be the parabolic subgroup of W generated by the simple reflections $\{s_i \mid 1 \le i \le n-1 \text{ and } i \notin \{r_1, \dots, r_k\}\}$, then one obtains a cell decomposition by using the projections $\pi(\mathcal{R}_{u,v}^{>0})$ of the cells $\mathcal{R}_{u,v}^{>0}$ where $u \le v$ and v is a minimal-length coset representative of $W/W_{\boldsymbol{r}}$. (We note moreover that Rietsch's results hold for G a semisimple, simply connected linear algebraic group over $\mathbb{C}$ split over $\mathbb{R}$).

In the case of the Grassmannian, Postnikov [2007] studied the Plücker nonnegative part $\mathrm{Gr}_{d,n}^{\ge 0}$ of the Grassmannian, and gave a decomposition of it into *positroid cells* $S_{\mathcal{B}}^{>0}$ by intersecting $\mathrm{Gr}_{d,n}^{\ge 0}$ with the matroid strata. Concretely, if $\mathcal{B}$ is the collection of bases of an element of $\mathrm{Gr}_{d,n}^{\ge 0}$, then $S_{\mathcal{B}}^{>0} = \{A \in \mathrm{Gr}_{d,n}^{\ge 0} \mid p_I(A) \ne 0 \text{ if and only if } I \in \mathcal{B}\}$. This cell decomposition of $\mathrm{Gr}_{d,n}^{\ge 0}$ agrees with Rietsh's cell decomposition [Talaska and Williams 2013, Corollary 1.2].

2.2. ***Background on (generalized) Bruhat interval polytopes.*** Bruhat interval polytopes were defined in [Kodama and Williams 2015], motivated by the connections to the full Kostant–Toda hierarchy.

Definition 2.4 [Kodama and Williams 2015]. Given two permutations u and v in S_n with $u \le v$ in Bruhat order, the *Bruhat interval polytope* $P_{u,v}$ is defined as

$$P_{u,v} = \mathrm{Conv}\{(x(1), x(2), \dots, x(n)) \mid u \le x \le v\} \subset \mathbb{R}^n. \tag{2}$$

We also define the (*twisted*) *Bruhat interval polytope* $\tilde{P}_{u,v}$ by

$$\tilde{P}_{u,v} = \mathrm{Conv}\{(n+1-x^{-1}(1), n+1-x^{-1}(2), \dots, n+1-x^{-1}(n)) \mid u \le x \le v\} \subset \mathbb{R}^n. \tag{3}$$

While the definition of Bruhat interval polytope in (2) is more natural from a combinatorial point of view, as we'll see shortly, the definition in (3) is more natural from the point of view of the moment map. Note that the set of Bruhat interval polytopes is the same as the set of twisted Bruhat interval polytopes; it is just a difference in labeling.

Remark 2.5. If we choose any point A in the cell $\mathcal{R}_{u,v}^{>0} \subset \mathrm{Fl}_n^{\ge 0}$ (thought of as an $n \times n$ matrix), and let M_i be the matroid represented by the first i rows of A, then $\tilde{P}_{u,v}$ is the Minkowski sum of the matroid polytopes $P(M_1), \dots, P(M_n)$ [Kodama and Williams 2015, Corollary 6.11]. In particular, $\tilde{P}_{u,v}$ is the matroid polytope of the flag matroid $M_1, \dots, M_n$.

Following [Tsukerman and Williams 2015], we can generalize the notion of Bruhat interval polytope as follows; see [loc. cit., Section 7.2] for notation.

Definition 2.6. Choose a generalized partial flag variety $G/P = G/P_J$, let W_J be the associated parabolic subgroup of the Weyl group W, and let $u, v \in W$ with $u \le v$ in Bruhat order and v a minimal-length coset representative of W/W_J. Let π denote the projection from G/B to G/P, and let A be an element of the cell $\pi(\mathcal{R}_{u,v}^{>0})$ of (Lusztig's definition of) $(G/P)^{\ge 0}$.

A *generalized Bruhat interval polytope* $\tilde{P}_{u,v}^{J}$ can be defined in any of the following equivalent ways [Tsukerman and Williams 2015, Definition 7.8, Lemma 7.9, Proposition 7.10, Remark 7.11] and [Borovik et al. 2003, Preface]:

- The moment map image of the closure of the torus orbit of A in G/P (which is a Coxeter matroid polytope).
- The moment map image of the closure of the cell $\pi(\mathcal{R}_{u,v}^{>0})$.
- The moment map image of the closure of the projected Richardson variety $\pi(\mathcal{R}_{u,v})$.
- The convex hull $\mathrm{Conv}\{z \cdot \rho_J \mid u \le z \le v\} \subset \mathfrak{t}_{\mathbb{R}}^*$, where ρ_J is the sum of fundamental weights $\sum_{j \in J} \omega_j$, and $\mathfrak{t}_{\mathbb{R}}^*$ is the dual of the real part of the Lie algebra $\mathfrak{t}$ of the torus $T \subset G$.

Remark 2.7. When $G = \mathrm{GL}_n$ with fundamental weights $\boldsymbol{e}_1, \boldsymbol{e}_1 + \boldsymbol{e}_2, \dots, \boldsymbol{e}_1 + \cdots + \boldsymbol{e}_{n-1}$, each generalized Bruhat interval polytope $\tilde{P}_{u,v}^{J}$ is the flag positroid polytope associated to a matrix A representing a point of $\mathrm{Fl}_{\boldsymbol{r};n}^{\ge 0}$, with $\boldsymbol{r} = (r_1, \dots, r_k)$. In this case the generalized Bruhat interval polytope is precisely the Minkowski sum $P(M_1) + \cdots + P(M_k)$ of the matroid polytopes $P(M_i)$, where M_i is the matroid realized by the first r_i rows of A. In particular, the generalized Bruhat interval polytope $\tilde{P}_{u,v}^{\varnothing} = \tilde{P}_{u,v}$ is the Minkowski sum $P(M_1) + \cdots + P(M_n)$, where M_i is the positroid realized by the first i rows of any matrix representing a point of $A \in \mathcal{R}_{v,w}^{>0}$. We will discuss how to read off the matroids M_i from (u, v) in Section 7.2.

As mentioned in the introduction, when $\boldsymbol{r}$ is a sequence of consecutive ranks, the generalized Bruhat interval polytopes for $\mathrm{Fl}_{\boldsymbol{r};n}^{\ge 0}$ are precisely the flag positroid polytopes of ranks $\boldsymbol{r}$. When $\boldsymbol{r} = (1, 2, \dots, n)$, we recover the notion of Bruhat interval polytope, and when $\boldsymbol{r}$ is a single integer, we recover the notion of positroid polytope.

3. The nonnegative tropicalization

3.1. ***Background on tropical geometry.*** We define the main objects in (a) and (b) of Theorem A, and record some basic properties. For a more comprehensive treatment of tropicalizations and positive-tropicalizations, we refer to [Maclagan and Sturmfels 2015, Chapter 6] and [Speyer and Williams 2005], respectively.

For a point $w = (w_1, \dots, w_m) \in \mathbb{T}^m$, we write $\bar{w}$ for its image in the tropical projective space $\mathbb{P}(\mathbb{T}^m)$. For $a = (a_1, \dots, a_n) \in \mathbb{Z}^m$, write $a \bullet w = a_1 w_1 + \cdots + a_m w_m$.

Definition 3.1. For a real homogeneous polynomial

$$f = \sum_{a\in\mathcal{A}} c_a x^a \in \mathbb{R}[x_1, \dots, x_m], \quad \text{where } \mathcal{A} \text{ is a finite subset of } \mathbb{Z}^m_{\geq 0} \text{ and } 0 \neq c_a \in \mathbb{R},$$

the *extended tropical hypersurface* $V_{\mathrm{trop}}(f)$ and the *nonnegative tropical hypersurface* $V^{\geq 0}_{\mathrm{trop}}(f)$ are subsets of the tropical projective space $\mathbb{P}(\mathbb{T}^m)$ defined by

$$V_{\mathrm{trop}}(f) = \{\bar{w} \in \mathbb{P}(\mathbb{T}^m) \mid \text{the minimum in } \min_{a\in\mathcal{A}}(a \bullet w), \text{ if finite, is achieved at least twice}\},$$

and

$$V^{\geq 0}_{\mathrm{trop}}(f) = \left\{ \bar{w} \in \mathbb{P}(\mathbb{T}^m) \;\middle|\; \begin{array}{l} \text{the minimum in } \min_{a\in\mathcal{A}}(a \bullet w), \text{ if finite, is achieved at least twice,} \\ \text{including at some } a, a' \in \mathcal{A} \text{ such that } c_a \text{ and } c_{a'} \text{ have opposite signs} \end{array} \right\}.$$

We say that a point *satisfies the tropical relation* of f if it is in $V_{\mathrm{trop}}(f)$, and that it *satisfies the positive-tropical relation* of f if it is in $V^{\geq 0}_{\mathrm{trop}}(f)$.

When f is a multihomogeneous real polynomial, we define $V_{\mathrm{trop}}(f)$ and $V^{\geq 0}_{\mathrm{trop}}(f)$ similarly as subsets of a product of tropical projective spaces. We will consider tropical hypersurfaces of polynomials that define the Plücker embedding of a partial flag variety.

Definition 3.2. For integers $0 < r \leq s < n$, the (*single-exchange*) *Plücker relations of type* $(r, s; n)$ are polynomials in variables $\{x_I : I \in \binom{[n]}{r} \cup \binom{[n]}{s}\}$ defined as

$$\mathscr{P}_{r,s;n} = \left\{ \sum_{j\in J\setminus I} \operatorname{sign}(j, I, J) x_{I\cup j} x_{J\setminus j} \;\middle|\; I \in \binom{[n]}{r-1}, J \in \binom{[n]}{s+1} \right\},$$

where $\operatorname{sign}(j, I, J) = (-1)^{|\{k\in J | k<j\}|+|\{i\in I | j<i\}|}$. When $r = s$, the elements of $\mathscr{P}_{r,r;n}$ are called the *Grassmann–Plücker relations* (of type $(r; n)$), and when $r < s$, the elements of $\mathscr{P}_{r,s;n}$ are called the *incidence-Plücker relations* (of type $(r, s; n)$).

As in the introduction, let $\boldsymbol{r} = (r_1 < \cdots < r_k)$ be a sequence of increasing integers in $[n]$. We let $\mathscr{P}_{\boldsymbol{r};n} = \bigcup_{r\leq s,\ r,s\in\boldsymbol{r}} \mathscr{P}_{r,s;n}$, and let $\langle \mathscr{P}_{\boldsymbol{r};n} \rangle$ be the ideal generated by the elements of $\mathscr{P}_{\boldsymbol{r};n}$. It is well-known that for any field $\Bbbk$ the ideal $\langle \mathscr{P}_{\boldsymbol{r};n} \rangle$ set-theoretically carves out the partial flag variety $\mathrm{Fl}_{\boldsymbol{r};n}(\Bbbk)$ embedded in $\prod_{i=1}^k \mathbb{P}(\Bbbk^{\binom{[n]}{r_i}})$ via the standard Plücker embedding [Fulton 1997, Section 9]. Similarly, the Plücker relations define the tropical analogues of partial flag varieties as follows.

Definition 3.3. The *tropicalization* $\mathrm{TrFl}_{\boldsymbol{r};n}$ of $\mathrm{Fl}_{\boldsymbol{r};n}$, *nonnegative tropicalization* $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n}$ of $\mathrm{Fl}_{\boldsymbol{r};n}$, *flag Dressian* $\mathrm{FlDr}_{\boldsymbol{r};n}$, and *nonnegative flag Dressian* $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ are subsets of $\prod_{i=1}^k \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$ defined as

$$\mathrm{TrFl}_{\boldsymbol{r};n} = \bigcap_{f\in\langle \mathscr{P}_{\boldsymbol{r};n}\rangle} V_{\mathrm{trop}}(f) \quad \text{and} \quad \mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n} = \bigcap_{f\in\langle \mathscr{P}_{\boldsymbol{r};n}\rangle} V^{\geq 0}_{\mathrm{trop}}(f),$$

$$\mathrm{FlDr}_{\boldsymbol{r};n} = \bigcap_{f\in \mathscr{P}_{\boldsymbol{r};n}} V_{\mathrm{trop}}(f) \quad \text{and} \quad \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n} = \bigcap_{f\in \mathscr{P}_{\boldsymbol{r};n}} V^{\geq 0}_{\mathrm{trop}}(f).$$

When $k = 1$, i.e., when $\boldsymbol{r}$ consists of one integer d, one obtains the (nonnegative) tropicalization of the Grassmannian $\mathrm{TrGr}_{d;n}^{(\geq 0)}$ and the (nonnegative) Dressian $\mathrm{Dr}_{d;n}^{(\geq 0)}$ studied in [Speyer and Sturmfels 2004; Speyer and Williams 2005; 2021; Arkani-Hamed et al. 2021b]. Like Fl_n, we write only n in the subscript when $\boldsymbol{r} = (1, 2, \ldots, n)$.

Remark 3.4. In [Joswig et al. 2023, Section 6], the authors define the "positive flag Dressian" to consist of the elements $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k) \in \mathrm{FlDr}_{\boldsymbol{r};n}$ whose constituents μ_i are each in the strictly positive Dressian. In our language, this is equal to considering the points of

$$\bigcap_{f \in \bigcup_{i=1}^k \mathscr{P}_{r_i,r_i;n}} V_{\mathrm{trop}}^{\geq 0}(f) \cap \bigcap_{f \in \bigcup_{r_i < r_j} \mathscr{P}_{r_i,r_j;n}} V_{\mathrm{trop}}(f)$$

that have no ∞ coordinates. In a similar vein, we could consider defining the "nonnegative flag Dressian" to be the elements of the flag Dressian whose constituents are in the nonnegative Dressian. This gives a strictly larger set than our definition of the nonnegative flag Dressian, and has the shortcoming that the equivalence of (a) and (b) in Theorem A would no longer hold; see Example 4.4.

We record a useful equivalent description of the (nonnegative) tropicalization of a partial flag variety using Puiseux series. Recall the notion of the tropical semifield from Definition 1.3.

Definition 3.5. Let $\mathcal{C} = \mathbb{C}\{\{t\}\}$ be the field of Puiseux series with coefficients in $\mathbb{C}$, with the usual valuation map $\mathrm{val} : \mathcal{C} \to \mathbb{T}$. Concretely, for $f \neq 0$, $\mathrm{val}(f)$ is the exponent of the initial term of f, and $\mathrm{val}(0) = \infty$. Let

$$\mathcal{C}_{>0} = \{f \in \mathcal{C} \setminus \{0\} : \text{the initial coefficient of } f \text{ is real and positive}\} \quad \text{and} \quad \mathcal{C}_{\geq 0} = \mathcal{C}_{>0} \cup \{0\}.$$

For a point $p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C}) \subseteq \prod_{i=1}^k \mathbb{P}(\mathcal{C}^{\binom{[n]}{r_i}})$, applying the valuation $\mathrm{val} : \mathcal{C} \to \mathbb{T}$ coordinate-wise to the Plücker coordinates gives a point $\mathrm{val}(p) \in \prod_{i=1}^k \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$. Noting that $\mathrm{val}(\mathcal{C}) = \mathbb{Q} \cup \{\infty\} \subset \mathbb{T}$, we say that a point in $\prod_{i=1}^k \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$ has *rational coordinates* if it is a point in $\prod_{i=1}^k \mathbb{P}((\mathbb{Q} \cup \{\infty\})^{\binom{[n]}{r_i}})$. Let $\mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C}_{\geq 0})$ be the subset of $\mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C})$ consisting of points with all coordinates in $\mathcal{C}_{\geq 0}$, i.e., the points $p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C}) \subseteq \prod_{i=1}^k \mathbb{P}(\mathcal{C}^{\binom{[n]}{r_i}})$ that have a representative in $\prod_{i=1}^k \mathcal{C}_{\geq 0}^{\binom{[n]}{r_i}}$.

Proposition 3.6. *The set* $\{\mathrm{val}(p) : p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C})\}$ *equals the set of points in* $\mathrm{TrFl}_{\boldsymbol{r};n}$ *with rational coordinates. Likewise, the set* $\{\mathrm{val}(p) : p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C}_{\geq 0})\}$ *equals the set of points in* $\mathrm{TrFl}_{\boldsymbol{r};n}^{\geq 0}$ *with rational coordinates. Moreover, we have*

$$\mathrm{TrFl}_{\boldsymbol{r};n} = \textit{the closure of } \{\mathrm{val}(p) : p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C})\} \textit{ in } \prod_{i=1}^k \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}}) \quad \textit{and}$$

$$\mathrm{TrFl}_{\boldsymbol{r};n}^{\geq 0} = \textit{the closure of } \{\mathrm{val}(p) : p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C}_{\geq 0})\} \textit{ in } \prod_{i=1}^k \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}}).$$

Proof. The first equality is known as the (extended) *fundamental theorem of tropical geometry* [Maclagan and Sturmfels 2015, Theorems 3.2.3 and 6.2.15]. The second equality is the analogue for nonnegative tropicalizations, established in [Speyer and Williams 2005, Proposition 2.2]. $\square$

Remark 3.7. The need to restrict to rational coordinates and the need to take the closure in Proposition 3.6 can be removed if we let $\mathcal{C}$ be the Maltsev–Neumann ring $\mathbb{C}((\mathbb{R}))$ (see [Poonen 1993, Section 3]) which satisfies $\mathrm{val}(\mathcal{C}) = \mathbb{T}$; see also [Markwig 2010].

Let us also record an equivalent description of the (nonnegative) flag Dressian when $\boldsymbol{r}$ is a sequence of consecutive integers. We need the following definition. As is customary in matroid theory, we write Sij for the union $S \cup \{i, j\}$ of subsets S and $\{i, j\}$ of $[n]$.

Definition 3.8. The set $\mathscr{P}^{(3)}_{r,r;n}$ of *three-term Grassmann–Plücker relations* (of type $(r; n)$) is the subset of $\mathscr{P}_{r,r;n}$ consisting of polynomials of the form

$$x_{Sij}x_{Sk\ell} - x_{Sik}x_{Sj\ell} + x_{Si\ell}x_{Sjk}$$

for a subset $S \subseteq [n]$ of cardinality $r-2$ and a subset $\{i < j < k < \ell\} \subseteq [n]$ disjoint from S. Similarly, the set $\mathscr{P}^{(3)}_{r,r+1;n}$ of *three-term incidence–Plücker relations* (of type $(r, r+1)$) is the subset of $\mathscr{P}_{r,r;n}$ consisting of polynomials of the form

$$x_{Si}x_{Sjk} - x_{Sj}x_{Sik} + x_{Sk}x_{Sij}$$

for a subset $S \subseteq [n]$ of cardinality $r-1$ and a subset $\{i < j < k\} \subseteq [n]$ disjoint from S.

Let $\mathscr{P}^{(3)}_{\boldsymbol{r};n}$ be the union of the three-term Grassmann–Plücker and three-term incidence-Plücker relations, which we refer to as the *three-term Plücker relations*.

Proposition 3.9. *Suppose $\boldsymbol{r} = (r_1 < \cdots < r_k)$ consists of consecutive integers. Then a point $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k) \in \prod_{i=1}^{k} \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$ is in the (nonnegative) flag Dressian if and only if its support $\underline{\boldsymbol{\mu}} = (\underline{\mu}_1, \dots, \underline{\mu}_k)$ is a flag matroid and $\boldsymbol{\mu}$ satisfies the (nonnegative-)tropical three-term Plücker relations. More explicitly, we have*

$$\mathrm{FlDr}_{\boldsymbol{r};n} = \left\{ \boldsymbol{\mu} \in \prod_{i=1}^{k} \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}}) \;\middle|\; \underline{\boldsymbol{\mu}} \text{ is a flag matroid and } \boldsymbol{\mu} \in \bigcap_{f \in \mathscr{P}^{(3)}_{\boldsymbol{r};n}} V_{\mathrm{trop}}(f) \right\}, \text{ and}$$

$$\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n} = \left\{ \boldsymbol{\mu} \in \prod_{i=1}^{k} \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}}) \;\middle|\; \underline{\boldsymbol{\mu}} \text{ is a flag matroid and } \boldsymbol{\mu} \in \bigcap_{f \in \mathscr{P}^{(3)}_{\boldsymbol{r};n}} V^{\geq 0}_{\mathrm{trop}}(f) \right\}.$$

Proof. We will use the language and results from the study of matroids over hyperfields. See [Baker and Bowler 2019] for hyperfields and relation to matroid theory, and see [Gunn 2019, Section 2.3] for a description of the signed tropical hyperfield $\mathbb{TR}$, for which we note the following fact: The underlying set of $\mathbb{TR}$ is $(\mathbb{R} \times \{+, -\}) \cup \{\infty\}$, so given $c \in \mathbb{T}$, one can identify it with the element $(c, +) \in \mathbb{R} \times \{+, -\}$ of $\mathbb{TR}$ if $c < \infty$ and ∞ otherwise.

In the language of hyperfields, for a homogeneous polynomial f in m variables and a hyperfield $\mathbb{F}$, one has the notion of the "hypersurface of f over $\mathbb{F}$," which is a subset $V_{\mathbb{F}}(f)$ of $\mathbb{P}(\mathbb{F}^m)$. When $\mathbb{F}$ is the tropical hyperfield $\mathbb{T}$, this coincides with $V_{\text{trop}}(f)$ in Definition 3.1. When $\mathbb{F}$ is the signed tropical hyperfield $\mathbb{TR}$, a point $w \in \mathbb{T}^m$, when considered as a point of $\mathbb{TR}^m$, is in $V_{\mathbb{TR}}(f)$ if and only if it is in $V_{\text{trop}}^{\geq 0}(f)$. Thus, in the language of flag matroids over hyperfields [Jarra and Lorscheid 2024], the flag Dressian is the partial flag variety $\text{Fl}_{\boldsymbol{r};n}(\mathbb{T})$ over $\mathbb{T}$, and the nonnegative flag Dressian is the subset of the partial flag variety $\text{Fl}_{\boldsymbol{r};n}(\mathbb{TR})$ over $\mathbb{TR}$ consisting of points that come from $\mathbb{T}$.

Now, both the tropical hyperfield and the signed tropical hyperfield are perfect hyperfields because they are doubly distributive [Baker and Bowler 2019, Corollary 3.45]. Our proposition then follows from [Jarra and Lorscheid 2024, Theorem 2.16 and Corollary 2.24], which together state the following: When $\boldsymbol{r}$ consists of consecutive integers, for a perfect hyperfield $\mathbb{F}$, a point $p \in \prod_{i=1}^{k} \mathbb{P}(\mathbb{F}^{\binom{[n]}{r_i}})$ is in the partial flag variety $\text{Fl}_{\boldsymbol{r};n}(\mathbb{F})$ over $\mathbb{F}$ if and only if the support of p is a flag matroid and p satisfies the three-term Plücker relations over $\mathbb{F}$. □

For completeness, we include the proof of the following fact.

Lemma 3.10. *The signed tropical hyperfield $\mathbb{TR}$ is doubly distributive. That is, for any $x, y, z, w \in \mathbb{TR}$, one has an equality of sets $(x \boxplus y) \cdot (z \boxplus w) = xz \boxplus xw \boxplus yz \boxplus yw$.*

Proof. If any one of the four x, y, z, w is ∞, then the desired equality is the usual distributivity of the signed tropical hyperfield. Thus, we now assume that all four elements are in $\mathbb{R} \times \{+, -\}$, and write $x = (x_{\mathbb{R}}, x_{\mathbb{S}}) \in \mathbb{R} \times \{+, -\}$ and similarly for y, z, w. If $x_{\mathbb{R}} > y_{\mathbb{R}}$, then $xz_{\mathbb{R}} > yz_{\mathbb{R}}$ and $xw_{\mathbb{R}} > yw_{\mathbb{R}}$, so the equality follows again from the usual distributivity. So we now assume that all four elements have the same value in $\mathbb{R}$, and the equality then follows from the fact that the signed hyperfield $\mathbb{S}$ is doubly distributive. □

Remark 3.11. Even when $\boldsymbol{r}$ does not consist of consecutive integers, [Jarra and Lorscheid 2024, Theorem 2.16] implies that the flag Dressian and the nonnegative flag Dressian are carved out by fewer polynomials than $\mathscr{P}_{\boldsymbol{r};n}$ in the following way: Denoting by

$$\mathscr{P}_{\boldsymbol{r};n}^{adj} = \bigcup_{i=1}^{k} \mathscr{P}_{r_i,r_i;n} \cup \bigcup_{i=1}^{k-1} \mathscr{P}_{r_i,r_{i+1};n},$$

one has

$$\text{FlDr}_{\boldsymbol{r};n} = \bigcap_{f \in \mathscr{P}_{\boldsymbol{r};n}^{adj}} V_{\text{trop}}(f) \quad \text{and} \quad \text{FlDr}_{\boldsymbol{r};n}^{\geq 0} = \bigcap_{f \in \mathscr{P}_{\boldsymbol{r};n}^{adj}} V_{\text{trop}}^{\geq 0}(f).$$

This generalizes the fact that a sequence of matroids $(M_1, \dots, M_k)$ is a flag matroid if and only if (M_i, M_{i+1}) is a flag matroid for all $i = 1, \dots, k-1$ [Borovik et al. 2003, Theorems 1.7.1 and 1.11.1].

The following corollary of Proposition 3.9 is often useful in computation. It states that the nonnegative tropical flag Dressian is in some sense "convex" inside the tropical flag Dressian.

Corollary 3.12. *Suppose that $\boldsymbol{r} = (r_1 < \cdots < r_k)$ consists of consecutive integers, and suppose we have points $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_\ell \in \prod_i^k \mathbb{T}^{\binom{[n]}{r_i}}$ that are in $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$. Then, if a nonnegative linear combination $c_1\boldsymbol{\mu}_1 + \cdots + c_\ell\boldsymbol{\mu}_\ell$ is in $\mathrm{FlDr}_{\boldsymbol{r};n}$, it is in $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$.*

Proof. We make the following general observation: Suppose $f = c_\alpha x^\alpha - c_\beta x^\beta + c_\gamma x^\gamma$ is a three-term polynomial in $\mathbb{R}[x_1, \ldots, x_m]$ with $c_\alpha, c_\beta, c_\gamma$ positive. Then an element $u \in \mathbb{T}^m$ satisfies the positive-tropical relation of f if and only if $\beta \bullet u = \min\{\alpha \bullet u, \gamma \bullet u\}$. Hence, if $u_1, \ldots, u_\ell \in \mathbb{T}^m$ each satisfy this relation, then a nonnegative linear combination of them can satisfy the tropical relation of f only if the term at β achieves the minimum, that is, only if the positive-tropical relation is satisfied. The corollary now follows from this general observation and Proposition 3.9. □

3.2. *Equivalence of (a) and (b) in Theorem A.* Let $\boldsymbol{r}$ be a sequence of consecutive integers $(a, \ldots, b)$ for some $1 \leq a \leq b \leq n$. We will show that $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n} = \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$. The inclusion $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n} \subseteq \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ is immediate from Definition 3.3. We will deduce $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n} \supseteq \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ by utilizing the two known cases of the equality $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n} = \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ — when $\boldsymbol{r} = (r)$ and when $\boldsymbol{r} = (1, 2, \ldots, n)$.

We start by recalling that tropicalization behaves well on subtraction-free rational maps.

Definition 3.13. Let $f = \sum_{a \in \mathcal{A}} c_a x^a \in \mathbb{R}[x_1, \ldots, x_m]$ be a real polynomial, where $\mathcal{A}$ is a finite subset of $\mathbb{Z}^m_{\geq 0}$ and $0 \neq c_a \in \mathbb{R}$. We define the *tropicalization* $\mathrm{Trop}(f) : \mathbb{R}^m \to \mathbb{R}$ to be the piecewise-linear map $w \mapsto \min_{a \in \mathcal{A}}(a \bullet w)$, where as before, $a \bullet w = a_1 w_1 + \cdots + a_m w_m$.

Note that $\mathrm{Trop}(f_1 f_2) = \mathrm{Trop}(f_1) + \mathrm{Trop}(f_2)$. Moreover, if f_1 and f_2 are two polynomials with positive coefficients, and $a_1, a_2 \in \mathbb{R}_{>0}$, then $\mathrm{Trop}(a_1 f_1 + a_2 f_2) = \min(\mathrm{Trop}(f_1), \mathrm{Trop}(f_2))$. These facts imply the following simple lemma, which appears as [Rietsch and Williams 2019, Lemma 11.5]; see [Speyer and Williams 2005, Proposition 2.5] and [Pachter and Sturmfels 2004] for closely related statements.

Lemma 3.14. *Let $f = (f_1, \ldots, f_n) : \mathcal{C}^m \to \mathcal{C}^n$ be a rational map defined by polynomials $f_1, \ldots, f_n$ with positive coefficients (or more generally by subtraction-free rational expressions). Let $(x_1, \ldots, x_m) \in (\mathcal{C}_{\geq 0})^m$, such that $f(x_1, \ldots, x_m) = (y_1, \ldots, y_n)$. Then*

$$(\mathrm{Trop}(f))(\mathrm{val}(x_1), \ldots, \mathrm{val}(x_m)) = (\mathrm{val}(y_1), \ldots, \mathrm{val}(y_n)).$$

The next result states that we can extend points in the nonnegative Dressian to points in the nonnegative two-step flag Dressian.

Proposition 3.15. *Given $\mu_d \in \mathrm{Dr}^{\geq 0}_{d;n}$ with rational coordinates, there exists $\mu_{d+1} \in \mathrm{Dr}^{\geq 0}_{d+1;n}$ such that $(\mu_d, \mu_{d+1}) \in \mathrm{FlDr}^{\geq 0}_{d,d+1;n}$. Similarly, there exists $\mu_{d-1} \in \mathrm{Dr}^{\geq 0}_{d-1;n}$ such that $(\mu_{d-1}, \mu_d) \in \mathrm{FlDr}^{\geq 0}_{d-1,d;n}$.*

The proof of Proposition 3.15 requires the following refined results about Rietsch's cell decomposition of the nonnegative flag variety.

Theorem 3.16. *The nonnegative flag variety has a cell decomposition into positive Richardsons*

$$\mathrm{Fl}_n(\mathcal{C}_{\geq 0}) = \bigsqcup_{v \leq w} \mathcal{R}_{v,w}(\mathcal{C}_{>0})$$

where each cell $\mathcal{R}_{v,w}(\mathcal{C}_{>0})$ can be parametrized using a map

$$\phi_{v,w} : (\mathcal{C}_{>0})^{\ell(w)-\ell(v)} \to \mathcal{R}_{v,w}(\mathcal{C}_{>0}).$$

Moreover, this parametrization can be expressed as an embedding into projective space (e.g., using the flag minors) using polynomials in the parameters with positive coefficients.

Proof. The first statement comes from [Marsh and Rietsch 2004, Theorem 11.3]; Marsh and Rietsch were working over $\mathbb{R}$ and $\mathbb{R}_{>0}$ but the same proof holds over Puiseux series. The statement that the parametrization can be expressed as an embedding into projective space using positive polynomials comes from [Rietsch and Williams 2008, Proposition 5.1]. □

Corollary 3.17. *Each m-dimensional positroid cell $S_{\mathcal{B}}(\mathcal{C}_{>0})$ in the nonnegative Grassmannian $\mathrm{Gr}_{d,n}(\mathcal{C}_{\geq 0})$ is the projection $\pi_d(\mathcal{R}_{v,w}(\mathcal{C}_{>0}))$ of some positive Richardson of dimension $m = \ell(w) - \ell(v)$ in $\mathrm{Fl}_n(\mathcal{C}_{\geq 0})$, so we get a subtraction-free rational map*

$$\pi_d \circ \phi_{v,w} : (\mathcal{C}_{>0})^m \to \mathcal{R}_{v,w}(\mathcal{C}_{>0}) \to S_{\mathcal{B}}(\mathcal{C}_{>0}).$$

Proof. That fact that each positroid cell is the projection of a positive Richardson was discussed in Section 2.1. The result now follows from Theorem 3.16. □

Proof of Proposition 3.15. Using [Arkani-Hamed et al. 2021b, Theorem 9.2], the fact that $\mu_d \in \mathrm{Dr}^{\geq 0}_{d;n}$ with rational coordinates implies that $\mu_d = \mathrm{val}(\{\Delta_I(V_d)\})$ for some subspace $V_d \in \mathrm{Gr}_{d,n}(\mathcal{C}_{\geq 0})$, and hence V_d lies in some positroid cell $S_{\mathcal{B}}(\mathcal{C}_{>0})$ over Puiseux series.

By Corollary 3.17, V_d is the projection of a point $(V_1, \dots, V_n)$ of $\mathrm{Fl}_n(\mathcal{C}_{\geq 0})$, which in turn is the image of a point $(x_1, \dots, x_m) \in (\mathcal{C}_{>0})^m$, and the Plücker coordinates $\Delta_I(V_j)$ of each V_j are expressed as positive polynomials $\Delta_I(x_1, \dots, x_m)$ in the parameters $x_1, \dots, x_m$.

In particular, we have subtraction-free maps

$$\pi_d \circ \phi_{v,w} : (\mathcal{C}_{>0})^m \to \mathrm{Fl}_n(\mathcal{C}_{\geq 0}) \to \mathrm{Gr}_{d,n}(\mathcal{C}_{\geq 0})$$

taking

$$(x_1, \dots, x_m) \mapsto \{\Delta_I(x_1, \dots, x_m) \mid I \subset [n]\} \mapsto \left\{\Delta_I(x_1, \dots, x_m) \;\middle|\; I \in \binom{[n]}{d}\right\}.$$

The fact that the maps $\phi_{v,w}$ and π_d are subtraction-free implies by Lemma 3.14 that we can tropicalize them, obtaining maps

$$\mathrm{Trop}(\pi_d \circ \phi_{v,w}) : \mathbb{R}^m \to \mathrm{TrFl}_n^{\geq 0} \to \mathrm{TrGr}_{d,n}^{\geq 0}$$

taking

$$(\mathrm{val}(x_1), \dots, \mathrm{val}(x_m)) \mapsto \{\mathrm{val}(\Delta_I(x_1, \dots, x_m)) \mid I \subset [n]\} \mapsto \left\{\mathrm{val}(\Delta_I(x_1, \dots, x_m)) \mid I \in \binom{[n]}{d}\right\}.$$

We now let $\mu_{d+1} = \{\mathrm{val}(\Delta_I(V_{d+1})) \mid I \in \binom{[n]}{d+1}\}$ and $\mu_{d-1} = \{\mathrm{val}(\Delta_I(V_{d-1})) \mid I \in \binom{[n]}{d-1}\}$. By construction we have that all the three-term (incidence) Plücker relations hold for (μ_d, μ_{d+1}), and similarly for (μ_{d-1}, μ_d). Therefore $(\mu_d, \mu_{d+1}) \in \mathrm{FlDr}^{\geq 0}_{d,d+1;n}$ and $(\mu_{d-1}, \mu_d) \in \mathrm{FlDr}^{\geq 0}_{d-1,d;n}$. □

The following consequence of Proposition 3.15 is very useful.

Corollary 3.18. *Let $a' \leq a \leq b \leq b'$ be positive integers, and let $\boldsymbol{r} = (a, a+1, \dots, b)$ and $\boldsymbol{r}' = (a', a'+1, \dots, b')$ be sequences of consecutive integers. Then any point $(\mu_a, \dots, \mu_b) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ with rational coordinates can be extended to a point $(\mu_{a'}, \mu_{a'+1}, \dots, \mu_a, \dots, \mu_b, \dots, \mu_{b'}) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r}';n}$.*

Proof. We start with $\boldsymbol{\mu} = (\mu_a, \mu_{a+1}, \dots, \mu_b) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$. We take μ_b and repeatedly use Proposition 3.15 to construct μ_{b+1}, then $\mu_{b+2}, \dots, \mu_{b'}$. Similarly we take μ_a and use Proposition 3.15 to construct $\mu_{a-1}, \mu_{a-2}, \dots, \mu_{a'}$. Now by construction $(\mu_{a'}, \mu_{a'+1}, \dots, \mu_a, \dots, \mu_b, \dots, \mu_{b'})$ satisfies:

- $\mu_i \in \mathrm{Dr}^{\geq 0}_{i;n}$ for $i = a', a'+1, \dots, b'$.
- All three-term incidence-Plücker relations hold (because the three-term incidence-Plücker relations occur only in consecutive ranks).

Therefore $(\mu_{a'}, \mu_{a'+1}, \dots, \mu_{b'}) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r}';n}$ by Proposition 3.9. □

Theorem 3.19. *Let $\boldsymbol{r} = (a, a+1, \dots, b)$ be a sequence of consecutive integers, and let $\boldsymbol{\mu} \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ with rational coordinates. Then $\boldsymbol{\mu} \in \mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n}$.*

Proof. We start with $\boldsymbol{\mu} = (\mu_a, \mu_{a+1}, \dots, \mu_b) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ and use Corollary 3.18 to construct $(\mu_1, \dots, \mu_n) \in \mathrm{FlDr}^{\geq 0}_n$. Now [Boretsky 2022, Theorem 5.21$^{\mathrm{trop}}$] states that $\mathrm{FlDr}^{\geq 0}_n = \mathrm{TrFl}^{\geq 0}_n$. Hence, we have $(\mu_1, \dots, \mu_n) \in \mathrm{TrFl}^{\geq 0}_n$, so $(\mu_a, \mu_{a+1}, \dots, \mu_b) \in \mathrm{TrFl}^{\geq 0}_{\boldsymbol{r},n}$. □

Proof of (a) $\Longleftrightarrow$ (b) in Theorem A. We only need show that (b)$\Longrightarrow$(a), i.e., that $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n} \supseteq \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$, since the other direction is trivial. But this follows from Theorem 3.19 because the points in $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ with rational coordinates are dense in $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$, and $\mathrm{TrFl}^{\geq 0}_{\boldsymbol{r};n}$ is closed. □

Remark 3.20. Note that our method of proof crucially used the fact that $\boldsymbol{r}$ is a sequence of consecutive integers: we used Proposition 3.15 to fill in the ranks from b through n and from a down to 1. But if say we were considering $\boldsymbol{r} = \{a, b\}$ with $b - a > 1$ and $\boldsymbol{\mu} = (\mu_a, \mu_b)$, we could not guarantee using Proposition 3.15 that we could construct $\mu_{b-1}, \mu_{b-2}, \dots, \mu_{a+1}$ in a way that is consistent with μ_a.

Remark 3.21. Recall from Theorem 2.3 that if $\boldsymbol{r}$ is a sequence of consecutive integers, the two notions of the positive/nonnegative part of the flag variety (see Definition 2.2) coincide. The method used to prove the equivalence of (a) and (b) in Theorem A can be applied in a nontropical context to prove Theorem 2.3 in an alternate way. We start by noting that the result holds when $\boldsymbol{r} = (a)$, which is to say, for the nonnegative Grassmannian [Talaska and Williams 2013, Corollary 1.2] and also when $\boldsymbol{r} = (1, 2, \dots, n)$, which is to say, for the nonnegative complete flag variety [Boretsky 2022, Theorem 5.21]. To prove the result for $\boldsymbol{r} = (a, a+1, \dots, b)$, we start with a flag $V_\bullet = (V_a, \dots, V_b)$ in ranks $\boldsymbol{r}$ whose Plücker coordinates are all nonnegative, so that $V_\bullet$ is Plücker nonnegative. As in Proposition 3.15, we can use the $\boldsymbol{r} = (a)$ case to argue that the flag can be extended to lower ranks in such a way that all the Plücker coordinates are nonnegative. Dually, we can extend to higher ranks from the $\boldsymbol{r} = (b)$ case. This yields a complete flag $(V_1, \ldots, V_n)$ with all nonnegative Plücker coordinates. We can then apply the result in the complete flag case to conclude that $(V_1, \ldots, V_n)$ lies in $\mathrm{Fl}_n^{\geq 0}$. Thus, $V_\bullet$ is a projection of the nonnegative complete flag $(V_1, \ldots, V_n)$ and itself lies in $\mathrm{Fl}_{\boldsymbol{r};n}^{\geq 0}$, which is to say, $V_\bullet$ is Lusztig nonnegative.

The *strictly positive tropicalization* of a partial flag variety $\mathrm{TrFl}_{\boldsymbol{r};n}^{>0}$ is the subset of $\mathrm{TrFl}_{\boldsymbol{r};n}^{\geq 0}$ consisting of points whose coordinates are never ∞. Define similarly the *strictly positive flag Dressian* $\mathrm{FlDr}_{\boldsymbol{r};n}^{>0}$. The weaker version of Theorem 3.19 stating that $\mathrm{TrFl}_n^{>0} = \mathrm{FlDr}_n^{>0}$ was established in [Joswig et al. 2023, Lemma 19] as follows. One starts by noting that if $\mu \in \mathrm{Dr}_{r+m;n+m}^{\geq 0}$, then the sequence of minors $(\mu_r, \ldots, \mu_{r+m})$ where $\mu_{r+i} = \mu \backslash \{n+1, \ldots, n+i\}/\{n+i+1, \ldots, n+m\}$ is a point in $\mathrm{FlDr}_{r,\ldots,r+m;n}^{\geq 0}$. Then, the crucial step is a construction in discrete convex analysis [Murota and Shioura 2018, Proposition 2] that shows that every element of $\mathrm{FlDr}_n^{>0}$ arises from an element of $\mathrm{Dr}_{n;2n}^{>0}$ in this way. One then appeals to $\mathrm{Gr}_{r;n}^{>0} = \mathrm{Dr}_{r;n}^{>0}$ established in [Speyer and Williams 2021].

Example 3.22 shows that the above argument does not work if one replaces "strictly positive" with "nonnegative." In particular, the crucial step fails: that is, not every element of $\mathrm{FlDr}_n^{\geq 0}$ arises from an element of $\mathrm{Dr}_{n;2n}^{\geq 0}$ in such a way.

Example 3.22. Let (M_1, M_2, M_3) be matroids on $[3]$ whose sets of bases are $(\{1, 3\}, \{13\}, \{123\})$. The matrix

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

shows that it is a flag positroid. However, we claim that there is no positroid M of rank 3 on $[6]$ such that $M_1 = M \backslash 4/56$, $M_2 = M \backslash 45/6$, and $M_3 = M \backslash 456$. Since all three cases involve deletion by 4, if we replace $M \backslash 4$ by M', and decrease each of 5, 6 by 1, then we are claiming that there is no positroid M' of rank 3 on $[5]$ such that

$$M_1 = M'/45, \quad M_2 = M' \backslash 4/5, \quad \text{and} \quad M_3 = M' \backslash 45. \tag{4}$$

From $M_1 = M'/45$ and $M_2 = M' \backslash 4/5$, we have that $M'/5$ has bases $\{14, 34, 13\}$, and similarly, we have $M' \backslash 4$ has bases $\{135, 123\}$. Hence, the set of bases of M' contains $\{123, 135, 145, 345\}$, and does not contain $\{125, 235, 245\}$. By considering the Plücker relation

$$p_{134}p_{235} = p_{123}p_{345} + p_{135}p_{234},$$

we see that no positroid satisfies these properties.

4. Positively oriented flag matroids

In this section we explain the relationship between the nonnegative flag Dressian and positively oriented flag matroids, and we apply our previous results to flag matroids. In particular, we prove Corollary 1.5, which says that every positively oriented flag matroid of consecutive ranks is realizable. We also prove

Corollary 4.8, which says that a positively oriented flag matroid of consecutive ranks $a, \dots, b$ can be extended to ranks $a', \dots, b'$ (for $a' \le a \le a \le b$).

4.1. ***Oriented matroids and flag matroids.*** We give here a brief review of oriented matroids in terms of Plücker relations. Let $\mathbb{S} = \{-1, 0, 1\}$ be the *hyperfield of signs*. For a polynomial $f = \sum_{a \in \mathcal{A}} c_a x^a \in \mathbb{R}[x_1, \dots, x_m]$, we say that an element $\chi \in \mathbb{S}^m$ is *in the null set* of f if the set $\{\mathrm{sign}(c_a)\chi^a\}_{a \in \mathcal{A}}$ is either $\{0\}$ or contains $\{-1, 1\}$.

Definition 4.1. An *oriented matroid* of rank r on $[n]$ is a point $\chi \in \mathbb{S}^{\binom{[n]}{r}}$, called a *chirotope*, such that χ is in the null set of f for every $f \in \mathscr{P}_{r,r;n}$. Similarly, an *oriented flag matroid* of ranks $\boldsymbol{r}$ is a point $\boldsymbol{\chi} = (\chi_1, \dots, \chi_k) \in \prod_{i=1}^{k} \mathbb{S}^{\binom{[n]}{r_i}}$ such that $\boldsymbol{\chi}$ is in the null set of f for every $f \in \mathscr{P}_{\boldsymbol{r};n}$.

While these definitions may seem different from those in the standard reference [Björner et al. 1999] on oriented matroids, Definition 4.1 is equivalent to [Björner et al. 1999, Definition 3.5.3] by [Baker and Bowler 2019, Example 3.33]. The definition of oriented flag matroid here is equivalent to the definition of a sequence of oriented matroid quotients [Björner et al. 1999, Definition 7.7.2] by [Jarra and Lorscheid 2024, Example above Theorem D].

Definition 4.2. A *positively oriented matroid* is an oriented matroid χ such that χ only takes values 0 or 1. Similarly, we define a *positively oriented flag matroid* to be an oriented flag matroid $\boldsymbol{\chi}$ such that $\boldsymbol{\chi}$ only takes values 0 or 1.

A positroid M defines a positively oriented matroid $\chi = \chi_M$ where χ takes value 1 on its bases and 0 otherwise. da Silva [1987] conjectured that every positively oriented matroid arises in this way; this conjecture was subsequently proved in [Ardila et al. 2017] and then [Speyer and Williams 2021].

Theorem 4.3 [Ardila et al. 2017]. *Every positively oriented matroid χ is realizable, i.e., χ has the form χ_M for some positroid M.*

By Theorem 4.3, each positively oriented flag matroid is a sequence of positroids which is also an oriented flag matroid.

In this section we will prove Corollary 1.5, which generalizes Theorem 4.3, and says that every positively oriented flag matroid $(\chi_1, \dots, \chi_k)$ of consecutive ranks $r_1 < \cdots < r_k$ can be realized by a flag positroid. But before we prove it, let us give an example that shows that imposing the oriented flag matroid condition is stronger than imposing that we have a realizable flag matroid whose consistent matroids are positroids.

Example 4.4. We give an example of a realizable flag matroid that has positroids as its constituent matroids but is not a flag positroid. This example also appeared in [Joswig et al. 2023, Example 5] and [Bloch and Karp 2023, Example 6]. Let (M, M') be matroids of ranks 1 and 2 on [3] whose sets of bases are $\{1, 3\}$ and $\{12, 13, 23\}$, respectively. Both are positroids. We can realize (M, M') as a flag matroid using the matrix

$$\begin{bmatrix} a & 0 & b \\ c & d & e \end{bmatrix},$$

where the nonvanishing minors $a, b, ad, -bd, ae-bc$ are nonzero. In order to realize (M, M') as a flag positroid, we need to choose real numbers a, b, c, d, e such that all these minors are strictly positive. However, $a > 0$ and $ad > 0$ implies $d > 0$, while $b > 0$ and $-bd > 0$ implies $d < 0$.

This example is consistent with Corollary 1.5 because (M, M'), when considered as a sequence of positively oriented matroids, is not an oriented flag matroid.

4.2. ***From the nonnegative flag Dressian to positively oriented flag matroids.*** We start with the following simple observation. While the proof is very simple, we label it a "theorem" to emphasize its importance.

Theorem 4.5. *The set of positively oriented flag matroids of ranks* $\boldsymbol{r}$ *can be identified with the set of points of the nonnegative flag Dressian* $\mathrm{FlDr}_{\boldsymbol{r};n}^{\geq 0}$ *whose coordinates are all either* 0 *or* ∞*.*

Proof. Given a point $\chi = (\chi_1, \dots, \chi_m) \in \{0, 1\}^m \subset \mathbb{S}^m$,[2] we define $t(\chi) = (t_1, \dots, t_m) \in \mathbb{T}^m$ by setting $t_i = 0$ if $\chi_i = 1$ and $t_i = \infty$ if $\chi_i = 0$. Then, we observe that χ is in the null set of a polynomial $f \in \mathbb{R}[x_1, \dots, x_m]$ if and only if the image of $t(\chi)$ in $\mathbb{P}(\mathbb{T}^m)$ is a point in $V_{\mathrm{trop}}^{\geq 0}(f)$. Therefore, each positively oriented flag matroid $\boldsymbol{\chi}$ can be identified with the element $t(\boldsymbol{\chi})$ in the nonnegative flag Dressian $\mathrm{FlDr}_{\boldsymbol{r};n}^{\geq 0}$. □

We now prove that every positively oriented flag matroid $\boldsymbol{\chi} = (\chi_1, \dots, \chi_k)$ of consecutive ranks $r_1 < \cdots < r_k$ is realizable.

Proof of Corollary 1.5. By the lemma, we may identify a positively oriented flag matroid $\boldsymbol{\chi}$ as an element $t(\boldsymbol{\chi})$ of the nonnegative flag Dressian. Because the ranks $\boldsymbol{r}$ are consecutive integers, the equivalence (a)$\Longleftrightarrow$(b) of Theorem A implies that $t(\boldsymbol{\chi})$ is thus a point in $\mathrm{TrFl}_{\boldsymbol{r};n}^{\geq 0}$. Because $t(\boldsymbol{\chi})$ has rational coordinates (all non-∞ coordinates are 0), Proposition 3.6 implies that $t(\boldsymbol{\chi}) = \mathrm{val}(p)$ for some $p \in \prod_{i=1}^{k} \mathbb{P}(\mathcal{C}_{\geq 0}^{\binom{[n]}{r_i}})$. Setting the parameter t in each Puisseux series of p to 0 now gives the realization of $\boldsymbol{\chi}$ as a flag positroid. □

As in Question 1.6, we do not know whether the corollary holds when $\boldsymbol{r}$ does not consist of consecutive integers. The following example shows that one cannot reduce to the consecutive ranks case.

Example 4.6. We give an example of a flag positroid (M, M') on $[4]$ of ranks $(1, 3)$ such that there is no flag positroid (M, M_2, M') with rank of M_2 equal to 2. Let the sets of bases of M and M' be $\{1, 2, 3, 4\}$ and $\{123, 234\}$, respectively. The matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

for example shows that (M, M') is a flag positroid. However, this flag positroid cannot be extended to a flag positroid with consecutive ranks. To see this, note that any realization of (M, M') as a flag positroid,

[2]Note that $(\chi_1, \dots, \chi_m)$ is not a sequence of chirotopes in this proof, instead each $\chi_i \in \mathbb{S}$.

after row-reducing by the first row, is of the form

$$\begin{bmatrix} 1 & a & b & c \\ 0 & x & y & 0 \\ 0 & z & w & 0 \end{bmatrix}$$

where $a, b, c > 0$ and $xw - yz > 0$. The minors of the matrix formed by the first two rows include $x, y, -cx, -cy$, which cannot be all nonnegative since $c > 0$ and not both of x and y are zero.

Remark 4.7. Let us sketch an alternate proof of Corollary 1.5 that relies only on the weaker version of (a)$\Longleftrightarrow$(b) in Theorem A that the strictly positive parts agree, i.e., that $\mathrm{TrFl}^{>0}_{\boldsymbol{r};n} = \mathrm{FlDr}^{>0}_{\boldsymbol{r};n}$. For a matroid M of rank d, define $\rho_M \in \mathbb{R}^{\binom{[n]}{d}}$ by $\rho_M(S) = d - \mathrm{rk}_M(S)$ for $S \in \binom{[n]}{d}$, where rk_M is the rank function of M. If M is a positively oriented matroid, then ρ_M is a point in the positive Dressian $\mathrm{Dr}^{>0}_{d,n}$ [Speyer and Williams 2021, proof of Theorem 5.1]. One can use this to show that if $\boldsymbol{M} = (M_1, \dots, M_k)$ is a positively oriented flag matroid of consecutive ranks $\boldsymbol{r}$, then the sequence $\boldsymbol{\rho} = (\rho_{M_1}, \dots, \rho_{M_k})$ is a point in $\mathrm{FlDr}^{>0}_{\boldsymbol{r};n}$. Since $\mathrm{TrFl}^{>0}_{\boldsymbol{r};n} = \mathrm{FlDr}^{>0}_{\boldsymbol{r};n}$ and $\boldsymbol{\rho}$ has rational coordinates, Proposition 3.6 implies that there is a point $p \in \mathrm{Fl}_{\boldsymbol{r};n}(\mathcal{C}_{\geq 0})$ with $\mathrm{val}(p) = \boldsymbol{\rho}$. Consider the coordinate $p(S) \in \mathcal{C}$ of p at a subset $S \in \binom{[n]}{r_i}$. By construction, the initial term of $p(S)$ is ct^q for some positive real c and a nonnegative integer q, where q is zero exactly when S is a basis of M_i. Thus, setting the parameter t to 0 in the Puiseux series of p gives a realization of $\boldsymbol{M}$ as a flag positroid.

We now use Theorem 4.5 to give a matroidal analogue of Corollary 3.18.

Corollary 4.8. *Let $a' \leq a \leq b \leq b'$ be positive integers, and let $(M_a, M_{a+1}, \dots, M_b)$ be a positively oriented flag matroid on $[n]$ of consecutive ranks $a, a+1, \dots, b$, that is, a sequence of positroids $M_a, \dots, M_b$ which is also an oriented flag matroid. Then we can extend it to a positively oriented flag matroid $(M_{a'}, M_{a'+1}, \dots, M_a, \dots, M_b, \dots, M_{b'})$ of consecutive ranks $a', a'+1, \dots, b'$.*

Proof. As in Theorem 4.5, we view the positively oriented flag matroid $(M_a, \dots, M_b)$ as a point of the nonnegative flag Dressian $(\mu_a, \dots, \mu_b) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ whose coordinates are all either 0 or ∞. The desired statement *almost* follows from Proposition 3.15: we just need to check that we can extend $(\mu_a, \dots, \mu_b)$ in a way which preserves the fact that coordinates are all either 0 or ∞. This is true, and we prove it by following the proof of Proposition 3.15 and replacing all instances of the positive Puiseux series $\mathcal{C}_{>0}$ by the positive Puiseux series with *constant coefficients*, that is, by $\mathbb{R}_{>0}$. Alternatively, we can use our result that $(M_a, \dots, M_b)$ is realizable by a flag positroid, and then argue as in Remark 3.21. □

5. Subdivisions of flag matroid polytopes

5.1. ***Flag Dressian and flag matroidal subdivisions.*** Consider a point $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k) \in \prod_{i=1}^{k} \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$ such that its support $\underline{\boldsymbol{\mu}}$ is a flag matroid. By construction, the vertices of the flag matroid polytope $P(\underline{\boldsymbol{\mu}})$ have the form $\boldsymbol{e}_{B_1} + \cdots + \boldsymbol{e}_{B_k}$ where B_i is a basis of the matroid $\underline{\mu}_i$ for each $i = 1, \dots, k$.

Definition 5.1. We define $\mathcal{D}_{\boldsymbol{\mu}}$ to be the coherent subdivision of $P(\underline{\boldsymbol{\mu}})$ induced by assigning each vertex $\boldsymbol{e}_{B_1}+\cdots+\boldsymbol{e}_{B_k}$ of $P(\underline{\boldsymbol{\mu}})$ the weight $\mu_1(B_1)+\cdots+\mu_k(B_k)$. That is, the faces of $\mathcal{D}_{\boldsymbol{\mu}}$ correspond to the faces of the lower convex hull of the set of points

$$\{(\boldsymbol{e}_{B_1}+\cdots+\boldsymbol{e}_{B_k},\, \mu_1(B_1)+\cdots+\mu_k(B_k)) \in \mathbb{R}^n \times \mathbb{R} : \boldsymbol{e}_{B_1}+\cdots+\boldsymbol{e}_{B_k} \text{ a vertex of } P(\underline{\boldsymbol{\mu}})\}.$$

The points of the flag Dressians are exactly the ones for which the subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ consists of flag matroid polytopes.

Theorem 5.2 [Brandt et al. 2021, Theorem A(a) and (c)]. *A point $\boldsymbol{\mu} \in \prod_{i=1}^{k} \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$ is in the flag Dressian* $\mathrm{FlDr}_{\boldsymbol{r};n}$ *if and only if the all faces of the subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ are flag matroid polytopes.*

When $\boldsymbol{r}$ consists of consecutive integers $(a, a+1, \dots, b)$, the nonnegative analogue of this theorem is the equivalence of (b) and (c) in Theorem A, which states that a point $\boldsymbol{\mu} \in \prod_{i=a}^{b} \mathbb{P}(\mathbb{T}^{\binom{[n]}{i}})$ is in the nonnegative flag Dressian $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ if and only if all faces of the subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ are flag positroid polytopes. A different nonnegative analogue of Theorem 5.2 that holds for $\boldsymbol{r}$ not necessarily consecutive, but loses the flag positroid property, can be found in Remark 5.6.

5.2. ***The proof of (b)⟹(c)⟹(d)⟹(e) in Theorem A.*** We start by recording two observations. The first is a well-known consequence of the greedy algorithm for matroids; see for instance [Ardila and Klivans 2006, Proposition 4.3]. For a matroid M on $[n]$ and a vector $\boldsymbol{v} \in \mathbb{R}^n$, let $\mathrm{face}(P(M), \boldsymbol{v})$ be the face of the matroid polytope $P(M)$ that maximizes the standard pairing with $\boldsymbol{v}$.

Proposition 5.3. *Let M be a matroid on $[n]$ and let $\mathscr{S} = (\varnothing \subsetneq S_1 \subsetneq \cdots \subsetneq S_\ell \subsetneq [n])$ be a chain of nonempty proper subsets of $[n]$. For a vector $\boldsymbol{v}_{\mathscr{S}}$ in the relative interior of the cone $\mathbb{R}_{\geq 0}\{\boldsymbol{e}_{S_1}, \dots, \boldsymbol{e}_{S_\ell}\}$, we have*

$$\mathrm{face}(P(M), \boldsymbol{v}_{\mathscr{S}}) = P(M^{\mathscr{S}}),$$

where $M^{\mathscr{S}} = M|S_1 \oplus M|S_2/S_1 \oplus M|S_3/S_2 \oplus \cdots \oplus M/S_\ell$ is the direct sum of minors of M.

For $\boldsymbol{M} = (M_1, \dots, M_k)$ a flag matroid, since $P(\boldsymbol{M})$ is the Minkowski sum $P(M_1)+\cdots+P(M_k)$, we likewise have that $\mathrm{face}(P(\boldsymbol{M}), \boldsymbol{v}_{\mathscr{S}}) = P(\boldsymbol{M}^{\mathscr{S}}) = P(M_1^{\mathscr{S}})+\cdots+P(M_k^{\mathscr{S}})$, *where $\boldsymbol{M}^{\mathscr{S}} = (M_1^{\mathscr{S}}, \dots, M_k^{\mathscr{S}})$. In particular, the face of a flag matroid polytope is a flag matroid polytope.*

The second observation concerns the following operations that we will show preserve the nonnegative flag Dressian. Recall that for $w \in \mathbb{T}^{\binom{[n]}{r}}$, its support $\underline{w}$ is $\{S \in \binom{[n]}{r} : w_S \neq \infty\}$:

- We consider a point $w \in \mathbb{T}^{\binom{[n]}{r}}$ as a set of weights on the vertices $\{\boldsymbol{e}_S : S \in \underline{w}\}$ of $P(\underline{w}) \subset \mathbb{R}^n$. Given an affine-linear function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and an element $w \in \mathbb{T}^{\binom{[n]}{r}}$, we define

$$\varphi w \in \mathbb{T}^{\binom{[n]}{r}} \quad \text{by} \quad (\varphi w)(S) = \varphi(\boldsymbol{e}_S) + w(S) \text{ for } S \in \binom{[n]}{r}.$$

- For a point $w \in \mathbb{T}^{\binom{[n]}{r}}$, denote by $w^{\mathrm{in}} \in \mathbb{T}^{\binom{[n]}{r}}$ its *initial part*, i.e.,

$$w^{\mathrm{in}}(S) = \begin{cases} 0 & \text{if } w(S) = \min\{w(S') : S' \in \binom{[n]}{r}\}, \\ \infty & \text{otherwise.} \end{cases}$$

Proposition 5.4. *Let $\boldsymbol{r} = (r_1, \dots, r_k)$ be a sequence of increasing integers in $[n]$. Suppose $\boldsymbol{w} = (w_1, \dots, w_k) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$. Then, the following hold*:

(1) *The support $\underline{\boldsymbol{w}}$ is a positively oriented flag matroid. In particular, it is a flag positroid when $\boldsymbol{r} = (r_1, \dots, r_k)$ consists of consecutive integers.*

(2) *We have $\varphi \boldsymbol{w} = (\varphi w_1, \dots, \varphi w_k) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ for any affine-linear functional φ on $\mathbb{R}^n$.*

(3) *We have $\boldsymbol{w}^{\mathrm{in}} = (w_1^{\mathrm{in}}, \dots, w_k^{\mathrm{in}}) \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$.*

Proof. We may consider $\underline{\boldsymbol{w}}$ as an element $\prod_{i=1}^{k} \mathbb{P}(\mathbb{T}^{\binom{[n]}{r_i}})$ by assigning the value 0 to a subset S if it is in the support of $\boldsymbol{w}$ and ∞ otherwise. Then, we have $\underline{\boldsymbol{w}} \in \mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ because the terms in each of the tropical Plücker relations that achieve the minimum when evaluated at $\boldsymbol{w}$ continue to do so when evaluated at $\underline{\boldsymbol{w}}$. The statement (1) follows from Theorem 4.5 and Corollary 1.5

The support is unchanged by φ, so $\underline{\varphi \boldsymbol{w}}$ is a flag matroid. The statement (2) now follows because for each of the positive-tropical Plücker relations, the operation φ preserves the terms at which the minimum is achieved.

The support $\underline{\boldsymbol{w}^{\mathrm{in}}}$ is a flag matroid by Theorem 5.2 and because $P(\underline{\boldsymbol{w}^{\mathrm{in}}})$ is a face in the subdivision $\mathcal{D}_{\boldsymbol{w}}$ of $P(\underline{\boldsymbol{w}})$. The statement (3) now follows because for each of the positive-tropical Plücker relations, the operation $^{\mathrm{in}}$ either preserves the terms at which the minimum is achieved or changes all the terms involved to ∞. □

Remark 5.5. While it's not needed here, we note that Proposition 5.4 is the "positive" analogue of the following statement, which is proved similarly: If $\boldsymbol{w} \in \mathrm{FlDr}_{\boldsymbol{r};n}$, then (1) $\underline{\boldsymbol{w}}$ is a flag matroid, (2) $\varphi \boldsymbol{w} \in \mathrm{FlDr}_{\boldsymbol{r};n}$, and (3) $\boldsymbol{w}^{\mathrm{in}} \in \mathrm{FlDr}_{\boldsymbol{r};n}$; see also [Brandt et al. 2021, Corollary 4.3.2] for related statements.

Proof of (b) $\implies$ (c). Every face in the coherent subdivision is the initial one after an affine-linear transformation. Hence, the implication follows from Proposition 5.4. □

Remark 5.6. One may modify the statement (c) to the following:

(c') Every face in the coherent subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ of $P(\underline{\boldsymbol{\mu}})$ is the flag matroid polytope of a positively oriented flag matroid.

Similar argument as above shows that (b)$\implies$(c') even when $\boldsymbol{r}$ doesn't consist of consecutive integers. One can also verify the converse (c')$\implies$(b) in this more general case as follows:

Suppose for contradiction (c') but not (b) for some $\boldsymbol{\mu}$. Then Theorem 5.2 implies that $\boldsymbol{\mu}$ is in the flag Dressian, and thus the failure of (b) implies that there is a Plücker relation where the minimum occurs at least twice but at the terms whose coefficients have the same sign. Proposition 5.4 implies that, replacing $\boldsymbol{\mu}$ by $\varphi\boldsymbol{\mu}$ for some φ if necessary, we may conclude that the same is true for that Plücker relation evaluated at $\boldsymbol{\mu}^{\mathrm{in}}$. But then $\boldsymbol{\mu}^{\mathrm{in}}$, which arise as a face in the subdivision, is not a positively oriented flag matroid by Theorem 4.5, contradicting (c').

There is no equivalence of (c') and (e) since three-term incidence relations exist only for consecutive ranks.

The implication (c)$\Longrightarrow$(d) is immediate.

Proof of (d) $\Longrightarrow$ (e). First, (d) implies that every edge of the subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ of $P(\underline{\boldsymbol{\mu}})$ is a flag matroid polytope, i.e., it is parallel to $\boldsymbol{e}_i - \boldsymbol{e}_j$ for some $i \neq j \in [n]$. Hence the edges of $P(\underline{\boldsymbol{\mu}})$ have the same property, so $\underline{\boldsymbol{\mu}}$ is a flag matroid.

We start with the case $a = b$, where $\boldsymbol{\mu}$ is just (μ). We need check the validity of the three-term positive-tropical Grassmann–Plücker relations, say for an arbitrary choice of $S \subset [n]$ and $\{i < j < k < \ell\} \subseteq [n] \setminus S$. Let $\mathscr{S}$ be a maximal chain $S_1 \subsetneq \cdots \subsetneq S_m$ of subsets of $[n]$ with the property that $S_a = S$ and $S_{a+1} = S \cup \{ijk\ell\}$ for some $a \in [m]$. Then, Proposition 5.3 implies that for a vector $\boldsymbol{v}_{\mathscr{S}}$ in the relative interior of the cone $\mathbb{R}_{\geq 0}\{\boldsymbol{e}_{S_1}, \ldots, \boldsymbol{e}_{S_m}\}$, we have

$$\operatorname{face}(P(\underline{\mu}), \boldsymbol{v}_{\mathscr{S}}) = P(\underline{\mu}^{\mathscr{S}}) \simeq P(\underline{\mu}|S \cup ijk\ell/S).$$

For the second identification, we have used that

- the matroid polytope of a direct sum of matroids is the product of the matroid polytopes;
- with the exception of $(S_a, S_{a+1}) = (S, S \cup ijk\ell)$, all other minors of the matroid $\underline{\mu}$ corresponding to (S_b, S_{b+1}) in the chain have their polytopes being a point because $|S_{b+1} \setminus S_b| = 1$.

Let $\widehat{r}$ be the rank of the matroid minor $\underline{\mu}|S \cup ijk\ell/S$. For a basis $\widehat{B}$ of $\underline{\mu}|S \cup ijk\ell/S$, let B be the basis of $\underline{\mu}$ such that the vertex $\boldsymbol{e}_B$ of $P(\underline{\mu})$ corresponds to the vertex $\boldsymbol{e}_{\widehat{B}}$ of $P(\underline{\mu}|S \cup ijk\ell/S)$ under the identification above. Identifying $[4] = \{1 < 2 < 3 < 4\}$ with $\{i < j < k < \ell\}$, we may thus consider "restricting" μ to the face $P(\underline{\mu}|S \cup ijk\ell/S)$ to obtain an element $\widehat{\mu} = \mu|S \cup ijk\ell/S \in \mathrm{Dr}_{\widehat{r};4}$ defined by

$$\widehat{\mu}(\widetilde{B}) = \begin{cases} \mu(B) & \text{if } \widehat{B} \text{ a basis of } \underline{\mu}|S \cup ijk\ell/S, \\ \infty & \text{otherwise,} \end{cases} \qquad \text{for } \widehat{B} \in \binom{[4]}{\widehat{r}}.$$

It is straightforward to check that for points in a Dressian on four elements, the three-term positive-tropical Grassmann–Plücker relations are satisfied if and only if every 2-dimensional faces in the corresponding subdivision are positroid polytopes. Since the faces of the subdivision $\mathcal{D}_{\widehat{\mu}}$ of $P(\underline{\mu}|S \cup ijk\ell/S)$ are a subset of the faces of the subdivision $\mathcal{D}_{\mu}$, we have that μ satisfies the three-term tropical-positive Grassmann–Plücker relation involving $ijk\ell$ and S.

Let us now treat the case $a < b$, whose proof is similar. We check the validity of the three-term positive-tropical incidence-Plücker relations, say for an arbitrary choice of $S \subset [n]$ and $\{i < j < k\} \subseteq [n] \setminus S$. Let $\mathscr{S}$ be a maximal chain $S_1 \subsetneq \cdots \subsetneq S_m$ of subsets of $[n]$ with the property that $S_a = S$ and $S_{a+1} = S \cup ijk$ for some $a \in [m]$. Then, Proposition 5.3 implies that for a vector $\boldsymbol{v}_{\mathscr{S}}$ in the relative interior of the cone $\mathbb{R}_{\geq 0}\{\boldsymbol{e}_{S_1}, \ldots, \boldsymbol{e}_{S_m}\}$, we have

$$\operatorname{face}(P(\underline{\boldsymbol{\mu}}), \boldsymbol{v}_{\mathscr{S}}) = P(\underline{\boldsymbol{\mu}}^{\mathscr{S}}) \simeq P(\underline{\boldsymbol{\mu}}|S \cup ijk/S).$$

For the second identification, we have used that

- the matroid polytope of a direct sum of matroids is the product of the matroid polytopes;
- with the exception of $(S_a, S_{a+1}) = (S, S \cup ijk)$, all other minors of the constituent matroids of $\underline{\boldsymbol{\mu}}$ corresponding to (S_b, S_{b+1}) in the chain have their polytopes being a point because $|S_{b+1} \setminus S_b| = 1$.

Note that the polytope $P(\underline{\boldsymbol{\mu}}|S \cup ijk/S)$ is at most 2-dimensional since it is a flag matroid polytope on 3 elements. Similarly to the $a = b$ case, we may "restrict" $\boldsymbol{\mu}$ to the face $P(\boldsymbol{\mu}|S \cup ijk/S)$ to obtain an element $\widehat{\boldsymbol{\mu}} = \mu|S \cup ijk/S \in \mathrm{FlDr}_3$, it is straightforward to verify that the unique three-term positive-tropical incidence relation involving S and ijk is satisfied if and only if the subdivision $\mathcal{D}_{\widehat{\boldsymbol{\mu}}}$ consists only of flag positroid polytopes. Since the faces of the subdivision $\mathcal{D}_{\widehat{\boldsymbol{\mu}}}$ are a subset of the faces of the subdivision $\mathcal{D}_{\boldsymbol{\mu}}$, we have that $\boldsymbol{\mu}$ satisfies the three-term incidence relation involving S and $\{i, j, k\}$. □

6. Three-term incidence relations

6.1. *The proof of (e)⟹(b) in Theorem A.* In the case that $a = b$ in Theorem A, the implication (e)⟹(b) is the content of Proposition 3.9.

To prove the implication when $a < b$, we will show the following key theorem.

Theorem 6.1. *Suppose $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathbb{P}(\mathbb{T}^{\binom{[n]}{r}}) \times \mathbb{P}(\mathbb{T}^{\binom{[n]}{r+1}})$ satisfies every three-term positive-tropical incidence relation. If the support $\underline{\boldsymbol{\mu}}$ is a flag matroid, then $\boldsymbol{\mu} \in \mathrm{FlDr}^{\geq 0}_{r,r+1;n}$.*

Proof of (e) ⟹ (b). Since $\boldsymbol{r}$ consists of consecutive integers, Theorem 6.1 implies that if $\underline{\boldsymbol{\mu}}$ is a flag matroid and $\boldsymbol{\mu}$ satisfies every three-term positive-tropical incidence relation, then $\boldsymbol{\mu}$ also satisfies every three-term positive-tropical Grassmann–Plücker relation. Hence $\boldsymbol{\mu}$ is an element of $\mathrm{FlDr}^{\geq 0}_{\boldsymbol{r};n}$ by Proposition 3.9. □

The proof of Theorem 6.1 relies on the following technical lemma.

Lemma 6.2. *Suppose $w \in \mathbb{T}^{\binom{[5]}{2}}$ satisfies all three-term positive-tropical Grassmann–Plücker relations involving the element 5. Suppose moreover that $w_{i5} < \infty$ for some $i = 1, 2, 3, 4$. Then $w \in \mathrm{Dr}^{\geq 0}_{2;5}$, i.e., w also satisfies the three-term positive-tropical Grassmann–Plücker relation not involving 5.*

Proof. The idea of the proof of Lemma 6.2 is that in the usual Grassmannian $\mathrm{Gr}_{2,5}$, if we can invert certain Plücker coordinates, then we can write the three-term Grassmann–Plücker relation not involving 5 as a linear combination of three of the other three-term Grassmann–Plücker relations. In particular, we have the following identity, which is easy to verify.

Lemma 6.3. *If $p_{25} \neq 0$ (respectively, $p_{35} \neq 0$) then $p_{13}p_{24} - p_{12}p_{34} - p_{14}p_{23}$ can be written in the following ways.*

$$\begin{aligned}
&p_{13}p_{24} - p_{12}p_{34} - p_{14}p_{23} \\
&\quad = (p_{13}p_{25} - p_{12}p_{35} - p_{15}p_{23})\frac{p_{24}}{p_{25}} - (p_{14}p_{25} - p_{12}p_{45} - p_{15}p_{24})\frac{p_{23}}{p_{25}} + (p_{24}p_{35} - p_{23}p_{45} - p_{25}p_{34})\frac{p_{12}}{p_{25}} \\
&\quad = (p_{13}p_{25} - p_{12}p_{35} - p_{15}p_{23})\frac{p_{34}}{p_{35}} - (p_{14}p_{35} - p_{13}p_{45} - p_{15}p_{34})\frac{p_{23}}{p_{35}} + (p_{24}p_{35} - p_{23}p_{45} - p_{25}p_{34})\frac{p_{13}}{p_{35}}.
\end{aligned}$$

We next note that we can interpret the first (respectively, second) expression in Lemma 6.3 tropically as long as $w_{25} < \infty$ (respectively, $w_{35} < \infty$).

Case 1: $w_{25} < \infty$. Then we can make sense of the terms on the right hand side of the first expression of Lemma 6.3 tropically. Since the three-term positive tropical Plücker relations involving 5 hold, and

$w_{25} < \infty$, we have

$$w_{13}+w_{25}+w_{24}-w_{25}=\min(w_{12}+w_{35}+w_{24}-w_{25},\, w_{15}+w_{23}+w_{24}-w_{25}),$$
$$w_{14}+w_{25}+w_{23}-w_{25}=\min(w_{12}+w_{45}+w_{23}-w_{25},\, w_{15}+w_{24}+w_{23}-w_{25}),$$
$$w_{24}+w_{35}+w_{12}-w_{25}=\min(w_{23}+w_{45}+w_{12}-w_{25},\, w_{25}+w_{34}+w_{12}-w_{25}).$$

We now simplify these expressions and underline terms that agree, obtaining

$$w_{13}+w_{24}=\min(\underline{w_{12}+w_{35}+w_{24}-w_{25}},\, \underline{\underline{w_{15}+w_{23}+w_{24}-w_{25}}}), \tag{5}$$
$$w_{14}+w_{23}=\min(\utilde{w_{12}+w_{45}+w_{23}-w_{25}},\, \underline{\underline{w_{15}+w_{24}+w_{23}-w_{25}}}), \tag{6}$$
$$\underline{w_{24}+w_{35}+w_{12}-w_{25}}=\min(\utilde{w_{23}+w_{45}+w_{12}-w_{25}},\, w_{34}+w_{12}). \tag{7}$$

There are now eight cases to consider, based on whether the minimum is achieved by the first or second term in each of (5), (6), (7). All cases are straightforward. If the minimum is achieved by the first term in (5) and the second term in (7), then we find that $w_{13}+w_{24}=w_{12}+w_{34}\le w_{14}+w_{23}$. In the other six cases, we find that $w_{13}+w_{24}=w_{14}+w_{23}\le w_{12}+w_{34}$. Therefore the positive tropical Plücker relation involving $1,2,3,4$ is satisfied.

Case 2: $w_{35}<\infty$. The argument for Case 2 is the same as for Case 1, except we use the tropicalization of the second identity in Lemma 6.3.

Case 3: $w_{25}=w_{35}=\infty$. In this case, since 5 is not a loop, either $w_{15}<\infty$ or $w_{45}<\infty$. Suppose that $w_{15}<\infty$. Then the positive tropical Plücker relations

- $w_{13}+w_{25}=\min(w_{12}+w_{35},\, w_{15}+w_{23})$,
- $w_{14}+w_{25}=\min(w_{12}+w_{45},\, w_{15}+w_{24})$,
- $w_{14}+w_{35}=\min(w_{13}+w_{45},\, w_{15}+w_{34})$,

imply that $w_{23}=w_{24}=w_{34}=\infty$, and hence the positive tropical Plücker relation involving $1,2,3,4$ is satisfied. The case where $w_{45}<\infty$ is similar. $\square$

For $w\in\mathbb{T}^{\binom{[n]}{r}}$, define its dual $w^\perp\in\mathbb{T}^{\binom{[n]}{n-r}}$ by $w^\perp(I)=w([n]\setminus I)$. It is straightforward to verify that w is an element of $\mathrm{Dr}_{r;n}$ (resp. $\mathrm{Dr}^{\ge 0}_{r;n}$) if and only if $w^\perp$ is an element of $\mathrm{Dr}_{n-r;n}$ (resp. $\mathrm{Dr}^{\ge 0}_{n-r;n}$). This matroid duality gives the following dual formulation of Lemma 6.2.

Corollary 6.4. *Suppose $w\in\mathbb{T}^{\binom{[5]}{3}}$ satisfies all three-term positive-tropical Grassmann–Plücker relations that contain a variable indexed by $S\in\binom{[5]}{3}$ with $5\notin S$. If $\underline{w}$ is a matroid such that 5 is not a coloop, then $w\in\mathrm{Dr}^{\ge 0}_{3;5}$, i.e., w also satisfies the three-term positive-tropical Grassmann–Plücker relation whose every variable contains 5 in its indexing subset.*

We are now ready to prove Theorem 6.1. We expect that the proof of Theorem 6.1 here adapts well to give an analogous statement for arbitrary perfect hyperfields.

Proof of Theorem 6.1. Given such $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathbb{P}(\mathbb{T}^{\binom{[n]}{r}}) \times \mathbb{P}(\mathbb{T}^{\binom{[n]}{r+1}})$, define $\widetilde{\boldsymbol{\mu}} \in \mathbb{P}(\mathbb{T}^{\binom{[n+1]}{r+1}})$ by

$$\widetilde{\boldsymbol{\mu}}(S) = \begin{cases} \mu_1(S \setminus (n+1)) & \text{if } (n+1) \in S, \\ \mu_2(S) & \text{otherwise.} \end{cases}$$

Because $\underline{\boldsymbol{\mu}}$ is a flag matroid, we have that $\underline{\widetilde{\boldsymbol{\mu}}}$ is a matroid, with the element $(n+1)$ that is neither a loop nor a coloop. We observe that $\widetilde{\boldsymbol{\mu}} \in \mathrm{Dr}^{\geq 0}_{r+1;n+1}$ if and only if $\boldsymbol{\mu} \in \mathrm{FlDr}^{\geq 0}_{r,r+1;n}$ because the validity of the three-term positive-tropical Grassmann–Plücker relations for $\widetilde{\boldsymbol{\mu}}$ is equivalent to the validity of both the three-term positive-tropical incidence relations and the three-term positive-tropical Grassmann–Plücker relations for $\boldsymbol{\mu}$.

We need to check that $\widetilde{\boldsymbol{\mu}}$ satisfies every three-term positive-tropical Grassmann–Plücker relation of type $(r+1; n+1)$. Consider the three-term relation associated to the subset $S \subseteq [n+1]$ of cardinality $r-1$ and $\{i < j < k < \ell\} \subseteq [n+1]$ disjoint from S. We have three cases:

- $\ell = n+1$. In this case, erasing the index $n+1$ in the expression for the corresponding three-term Grassmann–Plücker relation yields a three-term incidence relation of type $(r, r+1; n)$, which is satisfied by our assumption on $\boldsymbol{\mu}$.
- $(n+1) \in S$. In this case, considering the minor $\widetilde{\boldsymbol{\mu}}|S \cup ijk\ell/(S \setminus (n+1))$ and then applying Corollary 6.4 implies that the three-term Grassmann–Plücker relation is satisfied.
- $(n+1) \notin S \cup ijk\ell$. In this case, considering the minor $\widetilde{\boldsymbol{\mu}}|S \cup ijk\ell(n+1)/S$ and then applying Lemma 6.2 implies that the three-term Grassmann–Plücker relation is satisfied.

In every case the three-term positive-tropical Grassmann–Plücker relation is satisfied, as desired. □

7. Projections of positive Richardsons to positroids

One recurrent theme in our paper has been the utility of projecting a complete flag positroid (equivalently, a positive Richardson) to a positroid (or a positroid cell). This has come up in Rietsch's cell decomposition of a nonnegative (partial) flag variety, in our proofs in Section 3.2, and in the expression of a Bruhat interval polytope as a Minkowski sum of positroid polytopes in Remark 2.7. Positive Richardsons can be indexed by pairs (u, v) of permutations with $u \leq v$. Meanwhile, by work of Postnikov [2007], positroid cells of $\mathrm{Gr}^{\geq 0}_{d,n}$ can be indexed by *Grassmann necklaces*. In this section we will give several concrete combinatorial recipes for constructing the positroids obtained by projecting a (complete) flag positroid. We will also discuss the problem of determining when a collection of positroids can be identified with a (complete) flag positroid.

7.1. *Indexing sets for cells of* $\mathbf{Gr}^{\geq 0}_{d,n}$. As discussed in Section 2.1, there are two equivalent ways of thinking about the positroid cell decomposition of $\mathrm{Gr}^{\geq 0}_{d,n}$:

$$\mathrm{Gr}^{\geq 0}_{d,n} = \bigsqcup S^{>0}_{\mathcal{B}} = \bigsqcup_{u,v} \pi(\mathcal{R}^{>0}_{u,v}).$$

In the union on the right, π is the projection from Fl_n to $\mathrm{Gr}_{d,n}$, and u, v range over all permutations $u \leq v$ in S_n, such that v is a minimal-length coset representative of W/W_d, and $W_d = \langle s_1, \ldots, s_{d-1}, \hat{s}_d, s_{d+1}, \ldots, s_{n-1}\rangle$. We write W^d for the set of minimal-length coset representatives of W/W_d. Recall that a *descent* of a permutation z is a position j such that $z(j) > z(j+1)$. We have that W^d is the subset of permutations in S_n which have at most one descent, and if it exists, that descent must be in position d.

Even if $v \notin W^d$, the projection of $\mathcal{R}_{u,v}^{>0}$ to $\mathrm{Gr}_{d,n}^{\geq 0}$ is still a positroid, which we will characterize below. We start by defining *Grassmann necklaces* [Postnikov 2007].

Definition 7.1. Let $\mathcal{I} = (I_1, \ldots, I_n)$ be a sequence of subsets of $\binom{[n]}{d}$. We say $\mathcal{I}$ is a *Grassmann necklace* of *type* (d, n) if the following holds:

- If $i \in I_i$, then $I_{i+1} = (I_i \setminus i) \cup j$ for some $j \in [n]$.
- If $i \notin I_i$, then $I_{i+1} = I_i$.

In order to define the bijection between these Grassmann necklaces and positroids, we need to define the i-*Gale order* on $\binom{[n]}{d}$.

Definition 7.2. We write $<_i$ for the following *shifted linear order* on $[n]$.

$$i <_i i+1 <_i \cdots <_i n <_i 1 <_i \cdots <_i i-1.$$

We also define the i-*Gale order* on d-element subsets by setting

$$\{a_1 <_i \cdots <_i a_d\} \leq_i \{b_1 <_i \cdots <_i b_d\}$$

if and only if $a_\ell \leq_i b_\ell$ for all $1 \leq \ell \leq d$.

Given a positroid M, we define a sequence $\mathcal{I}_M = (I_1, \ldots, I_n)$ of subsets of $[n]$ by letting I_i be the minimal basis of M in the i-Gale order. The following result is from [Postnikov 2007, Theorem 17.1].

Proposition 7.3. *For any positroid M, $\mathcal{I}_M$ is a Grassmann necklace. The map $M \mapsto \mathcal{I}_M$ gives a bijection between positroids of rank d on $[n]$ and Grassmann necklaces of type (d, n).*

7.2. *Projecting positive Richardsons to positroids.* In this section we will give several descriptions of the constituent positroids appearing in a complete flag positroid (that is, a flag matroid represented by a positive Richardson). We start by reviewing a cryptomorphic definition of flag matroid, based on [Borovik et al. 2003, Sections 1.7–1.11].

A *flag* $F = F_1 \subset F_2 \subset \cdots \subset F_k$ on $[n]$ is an increasing sequence of finite subsets of $[n]$. A *flag matroid* is a collection $\mathcal{F}$ of flags satisfying the *maximality property*. Recall that e_S denotes the 01 indicator vector in $\mathbb{R}^n$ associated to a subset $S \subset [n]$. For a flag $F = F_1 \subset F_2 \subset \cdots \subset F_k$ we let $e_F = e_{F_1} + \cdots + e_{F_k}$. In this language, the flag matroid polytope of $\mathcal{F}$ is $P_{\mathcal{F}} = \mathrm{Conv}\{e_F \mid F \in \mathcal{F}\}$, whose vertices are precisely the points e_F for $F \in \mathcal{F}$.

In the complete flag case, each point e_F is a permutation vector $(z(1), \ldots, z(n))$ for some $z \in S_n$. Note that we can read off $z := z(F)$ from F by setting $z(i) = j$, where j is the unique element of $F_i \setminus F_{i-1}$.

Given $u \leq v$ in Bruhat order, we define the *Bruhat interval flag matroid* $\mathcal{F}_{u,v}$ to be the complete flag matroid whose flags are precisely

$$\{z([1]) \subset z([2]) \subset \cdots \subset z([n])\} \text{ for } u \leq z \leq v,$$

where $[i]$ denotes $\{1, 2, \ldots, i\}$ and $z([i])$ denotes $\{z(1), \ldots, z(i)\}$. Then by the above discussion, the (twisted) Bruhat interval polytope

$$\tilde{P}_{u,v} = \text{Conv}\{(n+1-z^{-1}(1), n+1-z^{-1}(2), \ldots, n+1-z^{-1}(n)) \mid u \leq z \leq v\}$$

is the flag matroid polytope of the Bruhat interval flag matroid $\mathcal{F}_{u,v}$.

This observation leads naturally to the following definition.

Definition 7.4. Consider a complete flag matroid $\mathcal{F}$ on $[n]$, which we identify with a collection $\mathcal{S}$ of permutations on $[n]$. By the maximality property [Borovik et al. 2003, Section 1.7.2] and its relation to the tableau criterion for Bruhat order [Borovik et al. 2003, Theorem 5.17.3], $\mathcal{S}$ contains a unique permutation u (respectively, v) which is minimal (respectively, maximal) in Bruhat order among all elements of $\mathcal{S}$. We say that $\mathcal{F}_{u,v}$ is the *Bruhat interval envelope* of $\mathcal{F}$.

It follows from Definition 7.4 that the Bruhat interval envelope of a complete flag matroid $\mathcal{F}$ contains $\mathcal{F}$; however, in general this inclusion is strict. It is an equality precisely when $\mathcal{F}$ is a Bruhat interval flag matroid.

Recall that if $F = (F_1, \ldots, F_n)$ and $G = (G_1, \ldots, G_n)$ are flags, we say that F is *less than or equal to G in the $\leq_j$ Gale order* (and write $F \leq_j G$) if and only if $F_i \leq_j G_i$ for all $1 \leq i \leq n$. (We also talk about the "usual" Gale order with respect to the total order $1 < 2 < \cdots < n$.) The *Maximality Property* for flag matroids implies that for any flag matroid $\mathcal{F}$, there is always a unique element which is maximal (and a unique element which is minimal) with respect to $\leq_j$.

We now give a Grassmann necklace characterization of the positroid constituents of a complete flag positroid, which follows from the previous discussion plus Proposition 7.3.

Proposition 7.5. *Consider a complete flag positroid $\boldsymbol{M} = (M_1, \ldots, M_n)$ on $[n]$, that is, the flag positroid associated to any point of $\mathcal{R}_{u,v}^{>0}$, for some $u \leq v$. For each $1 \leq j \leq n$, let $z^{(j)}$ be the Gale-minimal permutation with respect to $\leq_j$ in the interval $[u, v]$. Then the Grassmann necklace of the positroid M_j is $(z^{(1)}([j]), z^{(2)}([j]), \ldots, z^{(n)}([j]))$.*

Example 7.6. Consider the flag positroid associated to a point of $\mathcal{R}_{u,v}^{>0}$, where $u = (1, 2, 4, 3)$ and $v = (4, 2, 1, 3)$ (which we abbreviate as 1243 and 4213). The interval $[u, v]$ consists of

$$[u, v] = \{1243, 1423, 2143, 2413, 4123, 4213\}.$$

We now use Proposition 7.5, and find that the Gale-minimal permutations of $[u, v]$ with respect to $\leq_1$, $\leq_2$, $\leq_3$, $\leq_4$ are 1243, 2413, 4123, 4123. Therefore the Grassmann necklaces for the constituent positroids M_1, M_2, M_3 and M_4 are (1, 2, 4, 4), (12, 24, 14, 14), (124, 124, 124, 124), and (1234, 1234, 1234, 1234).

Alternatively, we can read off the flags in the flag positroid from the permutations in $[u, v]$, obtaining the flags

$$\{1 \subset 12 \subset 124, 1 \subset 14 \subset 124, 2 \subset 12 \subset 124, 2 \subset 24 \subset 124, 4 \subset 14 \subset 124, 4 \subset 24 \subset 124\}.$$

(Note that for brevity, we have omitted the subset 1234 from the end of each flag above.) We can now read off the bases of M_1, M_2, M_3, M_4 from the flags, obtaining $\{1, 2, 4\}$, $\{12, 14, 24\}$, $\{124\}$, and $\{1234\}$. We can then directly calculate the Grassmann necklaces from these sets of bases, getting the same answer as above.

If we compute the Minkowski sum of the positroids M_1, M_2, M_3, M_4 above, we obtain the twisted Bruhat interval polytope $\tilde{P}_{1243,4213} = P_{2314,4312}$, whose vertices are

$$\{(4, 3, 1, 2), (4, 2, 1, 3), (3, 4, 1, 2), (3, 2, 1, 4), (2, 4, 1, 3), (2, 3, 1, 4)\},$$

as noted in Remark 2.7.

The following result gives an alternative description of the constituent positroids of a complete flag matroid, this time in terms of bases.

Lemma 7.7 [Kodama and Williams 2015, Lemma 3.11; Billey and Weaver 2022, Theorem 1.4]. *Consider a complete flag positroid, that is, a flag matroid represented by a point of a positive Richardson $\mathcal{R}_{u,v}^{>0}$, where $u, v \in S_n$ and $u \le v$ in Bruhat order. Choose $1 \le d \le n$. Let π denote the projection from Fl_n to $\mathrm{Gr}_{d,n}$. Then the bases of the rank d positroid represented by $\pi(\mathcal{R}_{u,v}^{>0})$ are $\{z([d]) \mid u \le z \le v\}$.*

Finally, we remark that [Bloch and Karp 2023, Remark 5.24] gives yet another description of the constituent positroids of a complete flag positroid, this time in terms of pairs of permutations.

7.3. *Characterizing when two adjacent-rank positroids form an oriented matroid quotient.* We have discussed how to compute the projection of a complete flag positroid to a positroid. Moreover, it is well-known that every positroid is the projection of a complete flag positroid. In this section we will give a criterion for determining when two positroids M_i and M_{i+1} on $[n]$ of ranks i and $i+1$ can be obtained as the projection of a complete flag positroid (see Theorem 7.14).

We recall the definition of oriented matroid quotient in the setting at hand.

Definition 7.8. We say that two positroids M_i and M_{i+1} on $[n]$ of ranks i and $i+1$ *form an oriented matroid quotient* if (M_i, M_{i+1}) is an oriented flag matroid.

The following statement is a direct consequence of Corollary 4.8.

Proposition 7.9. *Let M_i and M_{i+1} be positroids on $[n]$ of ranks i and $i+1$. Then there is a complete flag positroid with M_i and M_{i+1} as constituents if and only if (M_i, M_{i+1}) form an oriented matroid quotient.*

Proposition 7.10. *Suppose that $(M_1, \dots, M_n)$ is a sequence of positroids of ranks $1, 2, \dots, n$ on $[n]$, such that each pair M_i and M_{i+1} forms an oriented matroid quotient. Then $(M_1, \dots, M_n)$ is a complete flag positroid. Moreover, it is realized by a point of the positive Richardson $\mathcal{R}_{u,v}^{>0}$, where we can explicitly construct u and v as follows*:

- *Let* $B_1^{\min}, \dots, B_n^{\min}$ *(respectively,* $B_1^{\max}, \dots, B_n^{\max}$*) be the bases of* $M_1, \dots, M_n$ *which are minimal (maximal) with respect to the usual Gale ordering. Then* $u, v \in S_n$ *are defined by*

$$u(i) = B_i^{\min} \setminus B_{i-1}^{\min} \quad \textit{and} \quad v(i) = B_i^{\max} \setminus B_{i-1}^{\max}.$$

Proof. As in Theorem 4.5, we identify each positroid M_i with the image $t(\chi_i)$ of its chirotope χ_i; we have that $t(\chi_i)$ lies in $\mathrm{Dr}_{i;n}^{\geq 0}$. The fact that each pair M_i, M_{i+1} forms an oriented matroid quotient means that $(t(\chi_1), \dots, t(\chi_n))$ satisfies all three-term incidence-Plücker relations, and hence $(t(\chi_1), \dots, t(\chi_n)) \in \mathrm{FlDr}_n^{\geq 0}$. Since $\mathrm{FlDr}_n^{\geq 0} = \mathrm{TrFl}_n^{\geq 0}$, we have proved that $(M_1, \dots, M_n)$ is a complete flag positroid.

To prove the characterization of u and v, we use Lemma 7.7. In particular, it follows from Lemma 7.7 and the Tableaux Criterion for Bruhat order that the Gale-minimal and Gale-maximal bases of the rank d positroid $\pi(\mathcal{R}_{u,v}^{>0})$ are $u([d])$ and $v([d])$. The result now follows. □

As we've seen in Example 4.4 it is a subtle question to determine whether a pair of positroids M_1 and M_2 of ranks r and $r+1$ form an oriented matroid quotient. One way is to construct an n by $r+1$ matrix such that the minor in rows $1, \dots, r$ and columns I is nonzero if and only if I is a basis of M_1 while the maximal minor in rows $1, \dots, r+1$ and columns J is nonzero if and only if J is a basis of M_2. Another way is to check the three-term relations over the signed tropical hyperfield, as in Proposition 3.9. We do not have an efficient way to do either of these things. Instead, in Theorem 7.14, we will give an algorithmic, combinatorial way to verify whether M_1 and M_2 form an oriented matroid quotient.

Construction 1. Given two positroids M_1 and M_2 on the ground set $[n]$ of ranks r and $r+1$, respectively, which form a positively oriented matroid quotient, we construct a positroid $M := M(M_1, M_2)$ of rank $r+1$ on the ground set $[n+1]$ where $n+1$ is neither a loop nor a coloop. The bases of M are precisely

$$\mathcal{B}(M) = \mathcal{B}(M_2) \cup \{B \cup \{n+1\} \mid B \in \mathcal{B}(M_1)\}.$$

Construction 2. Conversely, given a rank r positroid M on ground set $[n+1]$, where $(n+1)$ is neither a loop nor coloop, we construct two positroids $M_1 := M_1(M)$ and $M_2 := M_2(M)$ which form a positively oriented matroid quotient, as follows. Let $\tilde{A}$ be a matrix realizing M; therefore its Plücker coordinates are nonnegative. We apply row operations to rewrite $\tilde{A}$ in the form

$$A = \left[\begin{array}{ccccc|c} & & & & & 0 \\ & & A' & & & 0 \\ & & & & & 0 \\ \hline * & * & * & \cdots & * & 1 \end{array}\right].$$

Let M_1 denote the matroid on $[n]$ realized by A' and let M_2 denote the matroid on $[n]$ realized by A' together with the row of $*$'s below it. Then M_1 and M_2 are both positroids (since the Plücker coordinates of A' and A are all nonnegative), and they form a positively oriented quotient. Moreover, it is clear that $M_1 = M \setminus (n+1)$ and $M_2 = M/(n+1)$.

The idea of our algorithm is to translate Constructions 1 and 2 into operations on Grassmann necklaces, so that Construction 1 is well-defined even if M_1 and M_2 fail to form a positively oriented quotient. Clearly if we start with positroids M_1 and M_2 forming a positively oriented matroid quotient, then Construction 1 followed by 2 is the identity map. Conversely, if Construction 1 followed by 2 *is* the identity map, then since Construction 2 always outputs a positively oriented matroid quotient, we must have started with positroids forming a positively oriented matroid quotient.

We let $\min_i\{S_1, \dots, S_k\}$ denote the minimum of the sets $S_1, \dots, S_k$ in the $\leq_i$ order.

Proposition 7.11. *Let M_1 and M_2 be positroids of consecutive ranks which form a positively oriented quotient. Let $\mathcal{I}_{M_j} = (I_1^{(j)}, \dots, I_n^{(j)})$ be the Grassmann necklace of M_j for $j = 1, 2$. Define*

$$J_i = \begin{cases} I_1^{(2)} & \text{for } i = 1, \\ \min_i\{I_i^{(1)} \cup \{n+1\}, I_i^{(2)}\} & \text{for } 2 \leq i \leq n, \\ I_1^{(1)} \cup \{n+1\} & \text{for } i = n+1. \end{cases}$$

Then $\mathcal{J} = (J_1, \dots, J_{n+1})$ is the Grassmann necklace of the positroid $M = M(M_1, M_2)$ on $[n+1]$ whose bases are precisely

$$\mathcal{B}(M) = \mathcal{B}(M_2) \cup \{B \cup \{n+1\} \mid B \in \mathcal{B}(M_1)\}.$$

Proof. It suffices to show that each basis of M is i-Gale greater than $J^{(i)}$ for all $i \in [n+1]$. One also need to check that the $J^{(i)}$ are in fact bases of M but this is clear by definition.

Note that the $\leq_i$ minimal flag of a flag matroid consists of the $\leq_i$ minimal bases of each of its constituent matroids [Borovik et al. 2003, Corollary 7.2.1]. Thus, $I_t^{(1)} \subset I_t^{(2)}$ for each $t \in [n]$.

First, let $S \subset [n]$ be a basis of M_2. For $i \in [n]$, we have $S \geq_i I_i^{(2)} \geq_i J_i$. Since neither S nor $I_i^{(2)}$ contain $n+1$, $S \geq_{n+1} I_1^{(2)}$. By our earlier observation, $I_1^{(2)} = I_1^{(1)} \cup \{a\}$ for some $a \in [n]$. Thus, $I_1^{(2)} \geq_{n+1} I_1^{(1)} \cup \{n+1\}$. We conclude that $S \geq_i J_i$ for all $i \in [n+1]$.

Next, consider $S \cup \{n+1\}$ for S a basis of M_1. For $2 \leq i \leq n$, we have $S \cup \{n+1\} \geq_i I_i^{(1)} \cup \{n+1\} \geq_i J_i$. Since neither S nor $I_i^{(1)}$ contain $n+1$, we have $S \geq_1 I_1^{(1)}$ and $S \cup \{n+1\} \geq_{n+1} I_1^{(1)} \cup \{n+1\} = J_{n+1}$. Since $I_1^{(2)} = I_1^{(1)} \cup \{a\}$, we have $I_1^{(1)} \cup \{n+1\} \geq_1 I_1^{(2)} = J_1$. We conclude that $S \cup \{n+1\} \geq_i J_i$ for all $i \in [n+1]$. $\square$

If M_1 and M_2 form a positively oriented quotient, we should obtain them from the positroid $M = M(M_1, M_2)$, constructed as in Proposition 7.11, by deleting and contracting $n+1$. The following result explains how these operations affect Grassmann necklaces.

Proposition 7.12 [Oh 2008, Proposition 7 and Lemma 9]. *Let M be a positroid on $[n+1]$ such that $n+1$ is neither a loop nor a coloop, with Grassmann necklace $(J_i)_{i=1}^{n+1}$. Then the Grassmann necklaces $(K_1^{(1)}, \dots, K_n^{(1)})$ and $(K_1^{(2)}, \dots, K_n^{(2)})$ of $M_1 = M/(n+1)$ and $M_2 = M \setminus (n+1)$, are as follows*:

$$K_i^{(1)} = \begin{cases} J_i \setminus \{n+1\}, & n+1 \in J_i, \\ J_i \setminus \{\max_i(J_i \setminus J_{n+1})\}, & n+1 \notin J_i, \end{cases}$$

$$K_i^{(2)} = \begin{cases} (J_i \setminus \{n+1\}) \cup \{\min_i(J_{n+1} \setminus J_i)\}, & n+1 \in J_i, \\ J_i, & n+1 \notin J_i. \end{cases}$$

Taken together, the last two results yield a recipe for verifying whether two positroids, given in terms of their Grassmann necklaces, form a positively oriented quotient. First apply the construction of Proposition 7.11. If that yields a Grassmann necklace, apply Proposition 7.12 and see if that yields the original Grassmann necklaces. If so, the two Grassmann necklaces form a positively oriented quotient.

Our next goal is to streamline this recipe. Let $\mathcal{I}^{(1)} = (I_1^{(1)}, \dots, I_n^{(1)})$ and $\mathcal{I}^{(2)} = (I_1^{(2)}, \dots, I_n^{(2)})$ be Grassmann necklaces of positroids of ranks r and $r+1$, respectively. Note that a necessary condition for the positroids corresponding to $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ forming a positively oriented quotient is that $I_i^{(1)} \subset I_i^{(2)}$ for all $i \in [n]$. Now, we define a subset S as follows: For each i, if $I_i^{(1)} \cup \{n+1\} <_i I_i^{(2)}$, let $i \in S$. Since $I_i^{(2)} = I_i^{(1)} \cup a$ for some $a \in [n]$, this is as simple as checking whether $a <_i n+1$. If the positroids corresponding to $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ form a positively oriented quotient, applying Proposition 7.11 and then Proposition 7.12 should leave them unchanged. It is straightforward to see that $i \in S$ if and only if $n+1 \in J_i$ in Proposition 7.12. In particular, since $\mathcal{J}$ is a Grassmann necklace, S must either be an interval of the form $[d, n]$, or empty.

Next we claim that, once we verify that S is an interval of the form $[d, n]$ or is empty, then it follows automatically that $\mathcal{J}$, as constructed in Proposition 7.11, is a Grassmann necklace.

Lemma 7.13. *Let $\mathcal{I}^{(1)} = (I_1^{(1)}, \dots, I_n^{(1)})$ and $\mathcal{I}^{(2)} = (I_1^{(2)}, \dots, I_n^{(2)})$ be Grassmann necklaces of types (r, n) and $(r+1, n)$, respectively. Construct $\mathcal{J} = (J_1, \dots, J_{n+1})$ as in Proposition 7.11. Let $S = \{i \in [n] \mid I_i^{(1)} \cup (n+1) <_i I_i^{(2)}\}$. If $S = [d, n]$ for some $d \le n$ or $S = \varnothing$, then $\mathcal{J}$ is a Grassmann necklace.*

Proof. It is clear from the definition that $\mathcal{J}$ satisfies the Grassmann necklace condition for each pair of consecutive sets J_i and J_{i+1} except for when $i = k-1$, $i = n$ and $i = n+1$ (where we label sets cyclically so that $J_{n+2} = J_1$).

If $S \neq \varnothing$, then $J_n = I_n^{(1)} \cup \{n+1\}$. This makes it clear that the Grassmann necklace condition holds for J_n and J_{n+1}. Also, sing the fact that $I_i^{(1)} \subset I_i^{(2)}$ for all i, it is not hard to verify the Grassmann necklace condition for J_{n+1} and J_1.

This leaves us to check the condition for J_{k-1} and J_k. In this case, $J_{k-1} = I_{k-1}^{(2)}$ and $J_k = I_k^{(1)} \cup \{n+1\}$. Our goal is to show that $J_k = (J_{k-1} \setminus \{k-1\}) \cup \{a\}$ for some $a \in [n+1]$. It is immediately obvious that we necessarily have $a = n+1$. Thus, we are left to show that $I_k^{(1)} \cup \{n+1\} = (I_{k-1}^{(2)} \setminus \{k-1\}) \cup \{n+1\}$, or that $I_k^{(1)} = I_{k-1}^{(2)} \setminus \{k-1\}$.

Let a_i be defined by $I_i^{(1)} = (I_{i-1}^{(1)} \setminus \{i-1\}) \cup \{a_i\}$, let b_i be defined by $I_i^{(2)} = (I_{i-1}^{(2)} \setminus \{i-1\}) \cup \{b_i\}$ and let c_i be defined by $I_i^{(2)} = I_i^{(1)} \cup \{c_i\}$. We observe that $I_k^{(1)} = (I_{k-1}^{(1)} \setminus \{k-1\}) \cup \{a_k\} = (I_{k-1}^{(2)} \setminus \{c_{k-1}, k-1\}) \cup \{a_k\}$. Also, $I_k^{(1)} = I_k^{(2)} \setminus \{c_k\} = (I_{k-1}^{(2)} \setminus \{c_k, k-1\}) \cup \{b_k\}$. Comparing these two equalities, we conclude that either $a_k = c_{k-1}$ and $b_k = c_k$, or $c_{k-1} = c_k$ and $a_k = b_k$. The first case is what we want to prove, so let us show by contradiction that the second case cannot occur.

Assume $c_k = c_{k-1}$ and $a_k = b_k$. By assumption, $I_{k-1}^{(1)} \cup \{c_{k-1}\} = I_{k-1}^{(2)} <_{k-1} I_{k-1}^{(1)} \cup \{n+1\}$ and $I_k^{(1)} \cup \{n+1\} <_k I_k^{(2)} = I_k^{(1)} \cup \{c_k\}$. Thus, $c_{k-1} <_{k-1} n+1$ and $c_k >_k n+1$. Since $c_k = c_{k-1}$, this means they are both equal to $k-1$. However, if $c_k = k-1$, then M_2 has $k-1$ as a coloop. it follows that $b_k = k-1$, which means $a_k = k-1$ as well. Thus, in this case, $I_k^{(1)} = I_{k-1}^{(1)} = I_{k-1}^{(2)} \setminus \{k-1\}$, as desired.

Finally, if $A = \varnothing$, we can check that the Grassmann necklace condition holds for J_{n+1} and J_1 as before. The we are just left to verify this condition for J_n and J_{n+1}. We can apply the same logic but with J_{k-1} replaced by $J_n = I_n^{(2)}$ and J_k replaced by $J_{n+1} = I^{(1)} \cup \{n+1\}$. Specifically, we find $I_1^{(1)} = (I_n^{(2)} \setminus \{c_n, n\}) \cup \{a_1\} = I_n^{(2)} \setminus \{c_1, n\}) \cup \{b_1\}$. We then must show that it is impossible for $c_1 = c_n$ and $a_1 = b_1$. However, $I_n^{(1)} \cup \{c_n\} = I_n^{(2)} <_n I_n^{(1)} \cup \{n+1\}$. Moreover, it is always true that $I_1^{(1)} \cup \{n+1\} <_{n+1} I_1^{(1)} \cup \{c_1\} = I_2^{(1)}$. Using $c_1 = c_n$, we then find $c_n <_n (n+1)$ and $c_n >_{n+1} (n+1)$ which means that $c_1 = c_n = n$ and we can conclude as in the previous paragraph. □

Combining Propositions 7.11, 7.12 and Lemma 7.13, we obtain the following:

Theorem 7.14. *Fix positroids M_1 and M_2 on $[n]$ of ranks r and $r+1$, respectively. Let $\mathcal{I} = \mathcal{I}_{M_1} = (I_1, \ldots, I_n)$ and $\mathcal{J} = \mathcal{I}_{M_2} = (J_1, \ldots, J_n)$ be their Grassmann necklaces. We now set $S = \{i \in [n] \mid I_i \cup \{n+1\} \leq_i J_i\}$, where $\leq_i$ denotes the $\leq_i$ Gale order on $[n+1]$. Define $a_i = \max_i(J_i \setminus I_1)$ and $b_i = \min_i(I_1 \setminus I_i)$. Then M_1 and M_2 form a positively oriented quotient if and only if the following conditions hold*:

(1) *For $i \in [n]$, $I_i \subset J_i$.*

(2) *S is an interval of the form $[d, n]$ or $S = \varnothing$.*

(3) *For $i \notin S$, $I_i = J_i \setminus \{a_i\}$.*

(4) *For $i \in S$, $J_i = I_i \cup \{b_i\}$.*

Proof. First, suppose that we have a positively oriented quotient. As explained earlier, the first two conditions always hold for positively oriented quotients. We know that applying the constructions of Propositios 7.11 and 7.12 in sequence should preserve our positively oriented quotient. Observing what conditions this imposes on the constituent Grassmann necklaces yields conditions (3) and (4).

Conversely, if the conditions in the theorem statement hold, then by Lemma 7.13, applying the construction of Proposition 7.11 to $\mathcal{I}$ and $\mathcal{J}$ yields another Grassmann necklace $\mathcal{K}$ on $[n+1]$ such that $n+1$ is neither a loop nor a coloop of the positroid corresponding to $\mathcal{K}$. Then, conditions (3) and (4) guarantee that applying the construction of Proposition 7.12 to $\mathcal{K}$ will recover $\mathcal{I}$ and $\mathcal{J}$. The result of applying Proposition 7.12 to the Grassmann necklace of a positroid M with $n+1$ neither a loop nor a coloop is the pair of Grassmann necklaces corresponding to $M/(n+1)$ and $M \setminus (n+1)$, which form a positively oriented quotient. □

Example 7.15. Let $\mathcal{I} = (123, 235, 356, 456, 561, 613)$ and $\mathcal{J} = (1235, 2356, 3456, 4562, 5612, 6123)$. Then $A = \{4, 5, 6\}$ is an interval with upper endpoint $n = 6$. Note that $a_1 = 5$, $a_2 = 6$ and $a_3 = 6$, while $b_4 = 1$, $b_5 = 2$ and $b_6 = 2$. The positroids with these Grassmann necklaces do not form a positively oriented quotient since it is false that $I_3 = J_3 \setminus \{a_3\}$.

However, if we start with the Grassmann necklaces $\mathcal{I} = (123, 235, 345, 456, 561, 613)$ and $\mathcal{J} = (1235, 2356, 3456, 4562, 5612, 6123)$, then the values of the a_i and b_i are unchanged. It is straightforward to verify that the conditions of Theorem 7.14 hold and so the positroids corresponding to $\mathcal{I}$ and $\mathcal{J}$ do in fact form a positively oriented quotient.

We now have a tool that allows us to recognize flag positroids in consecutive ranks without finding a realization or certifying the incidence relations over the signed hyperfield.

Corollary 7.16. *Suppose* $(M_a, M_{a+1}, \ldots, M_b)$ *is a sequence of positroids of ranks* $a, a+1, \ldots, b$*. Then* $(M_a, M_{a+1}, \ldots, M_b)$ *is a flag positroid if and only if for* $a \leq i < b$*, the pair of positroids* (M_i, M_{i+1}) *satisfy the conditions of Theorem 7.14.*

Proof. By Proposition 7.10, it suffices to check that each such pair forms a positively oriented quotient, which is precisely the content of Theorem 7.14. □

8. Fan structures for and coherent subdivisions from $\mathrm{TrGr}^{>0}_{d;n}$ and $\mathrm{TrFl}^{>0}_{n}$

In this section we make some brief remarks about the various fan structures for $\mathrm{TrFl}^{>0}_{\boldsymbol{r};n}$ and coherent subdivisions from points of $\mathrm{TrFl}^{>0}_{\boldsymbol{r};n}$. Code written for the computations here can be found at https://github.com/chrisweur/PosTropFlagVar. We take a detailed look at the Grassmannian and complete flag variety, in particular the case of $\mathrm{TrFl}^{>0}_{4}$.

8.1. ***Fan structures.*** There are multiple possibly different natural fan structures for $\mathrm{TrFl}^{>0}_{\boldsymbol{r};n}$:

(i) The Plücker fan (induced by the three-term tropical Plücker relations).

(ii) The secondary fan (induced according to the coherent subdivision as in Corollary 8.3).

(iii) The Gröbner fan (induced according to the initial ideal of the ideal $\langle \mathscr{P}_{\boldsymbol{r};n} \rangle$).

(iv) The simultaneous refinement of the fans dual to the Newton polytopes of the Plücker coordinates, when the Plücker coordinates are expressed in terms of a "positive parametrization" of $\mathrm{Fl}^{>0}_{\boldsymbol{r};n}$, such as an $\mathcal{X}$*-cluster chart.*

(v) (If the cluster algebra associated to $\mathrm{Fl}_{\boldsymbol{r};n}$ has finitely many cluster variables) the same fan as above but with (the larger set of) cluster variables replacing Plücker coordinates.

Note that by definition, fan (v) is always a refinement of (iv).

In the case of the positive tropical Grassmannian, the fan structures in (iv) and (v) were studied in [Speyer and Williams 2005, Definition 4.2 and Section 8], where the authors observed that for $\mathrm{Gr}_{2,n}$, fan (iv) (which coincides with (v)) is isomorphic to the *cluster complex* of type A_{n-3};[3] for $\mathrm{Gr}_{3,6}$ and $\mathrm{Gr}_{3,7}$, fan (iv) is isomorphic to a coarsening of the corresponding cluster complex, while fan (v) is isomorphic to the cluster complex (of types D_4 and E_6, respectively). Conjecture 8.1 of [Speyer and Williams 2005] says that fan (v) (associated to the positive tropicalization of a full rank cluster variety of finite type) should be isomorphic to the corresponding cluster complex. This conjecture was essentially resolved in [Jahn et al. 2021; Arkani-Hamed et al. 2021a] by working with F-polynomials.

Theorem 14 of [Olarte et al. 2019] states that the Plücker fan and the secondary fan structures for Dressians coincide, and hence implies that (i) and (ii) coincide because the positive Dressian and the positive tropical Grassmannian are the same [Speyer and Williams 2021]. For $\mathrm{TrGr}_{2,n}$, the results of

[3] See [Fomin and Zelevinsky 2003] for background on the cluster complex.

[Speyer and Sturmfels 2004, Section 4] imply that (i), (ii), and (iii) agree, and combining this with [Speyer and Williams 2005, Section 5] implies that all five fan structures agree for $\mathrm{TrGr}_{2,n}^{>0}$. For $\mathrm{TrGr}_{3,6}^{>0}$, we computed that (iii) and (v) strictly refine (i), but the two fan structures are not comparable.

We can consider the same fan structures in the case of the positive tropical complete flag variety. When $n = 3$, the fan $\mathrm{TrFl}_n^{>0}$ modulo its lineality space is a one-dimensional fan, and all fan structures coincide. For TrFl_n (before taking the positive part), one can find computations of the fan (iii) for $n = 4$ and $n = 5$ in [Bossinger et al. 2017, Section 3], the fan (i) and its relation to (iii) for $n = 4$ in [Brandt et al. 2021, Example 5.2.3], and the fan (ii) and its relation to (iii) for $n = 4$ in [Joswig et al. 2023, Section 5]. Returning to the positive tropicalization, Bossinger [2022, Section 5.1] computed the fan structure (iii) for $\mathrm{TrFl}_4^{>0}$, and found it was dual to the three-dimensional associahedron; in particular, there are 14 maximal cones and the f-vector is $(14, 21, 9, 1)$. Using the positive parametrization of [Boretsky 2022] (a graphical version of the parametrizations of [Marsh and Rietsch 2004]) for $\mathrm{TrFl}_n^{>0}$, we computed the polyhedral complex underlying (iv) for $n = 4$ in Macaulay2 by computing the normal fan of the Minkowski sum of the Newton polytopes of the Plücker coordinates expressed in the chosen parametrization; we obtained the f-vector $(13, 20, 9, 1)$. We also computed (v) after incorporating the additional non-Plücker cluster variable $p_2 p_{134} - p_1 p_{234}$. Combining these, we find that for $n = 4$, (i)=(iv) and (ii)=(iii). We also find that both (ii) and (v) strictly refine (i)=(iv) and are both isomorphic to the normal fan of the three-dimensional associahedron, but are not comparable fan structures.

The fact that the fan structure (v) of $\mathrm{TrFl}_4^{>0}$ is dual to the three-dimensional associahedron is consistent with [Speyer and Williams 2005, Conjecture 8.1] and the fact that Fl_4 has a cluster algebra structure of finite type A_3 [Geiss et al. 2008, Table 1], whose cluster complex is dual to the associahedron.

We now give a graphical way to think about the fan structure on $\mathrm{TrFl}_4^{>0}$, building on the ideas of [Speyer and Williams 2005] and [Brandt et al. 2021, Example 5.2.3].

Example 8.1. A *planar tree* on $[n]$ is an unrooted tree drawn in the plane with n leaves labeled by $1, 2, \dots, n$ (in counterclockwise order). By [Speyer and Williams 2005], $\mathrm{TrGr}_{2;n}^{>0}$ parametrizes metric planar trees, and its cones correspond to the various combinatorial types of planar trees. In particular, if we assign real-valued lengths to the edges of a planar tree, then the negative of the distance between leaf i and j encodes the positive tropical Plücker coordinate w_{ij} of a point in the corresponding cone. In particular, it is easy to see that the negative distances w_{ij} associated to such a planar tree satisfy the positive tropical Plücker relations.

Now as in [Brandt et al. 2021, Example 5.2.3], we note that for a valuated matroid μ whose underlying matroid is the uniform matroid $U_{2,4}$, the tropical linear spaces $\mathrm{trop}(\mu)$ and $\mathrm{trop}(\mu^*)$ associated to μ and its dual μ^* are translates of each other. This allows us to identify points $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ of $\mathrm{TrFl}_4^{>0}$ with planar trees on the vertices $\{1, 2, 3, 4, 5, 5'\}$ such that the vertices $\{1, 2, 3, 4, 5\}$ and separately the vertices $\{1, 2, 3, 4, 5'\}$ appear in counterclockwise order. To see this, note that (using the same idea as Construction 1 from Section 7) we can identify (μ_1, μ_2), with Plücker coordinates $(w_1, \dots, w_4; w_{12}, \dots, w_{34})$, with an element (w_{ab}) of $\mathrm{TrGr}_{2,5}^{>0}$: we simply set $w_{a5} := w_a$ for $1 \le a \le 4$.

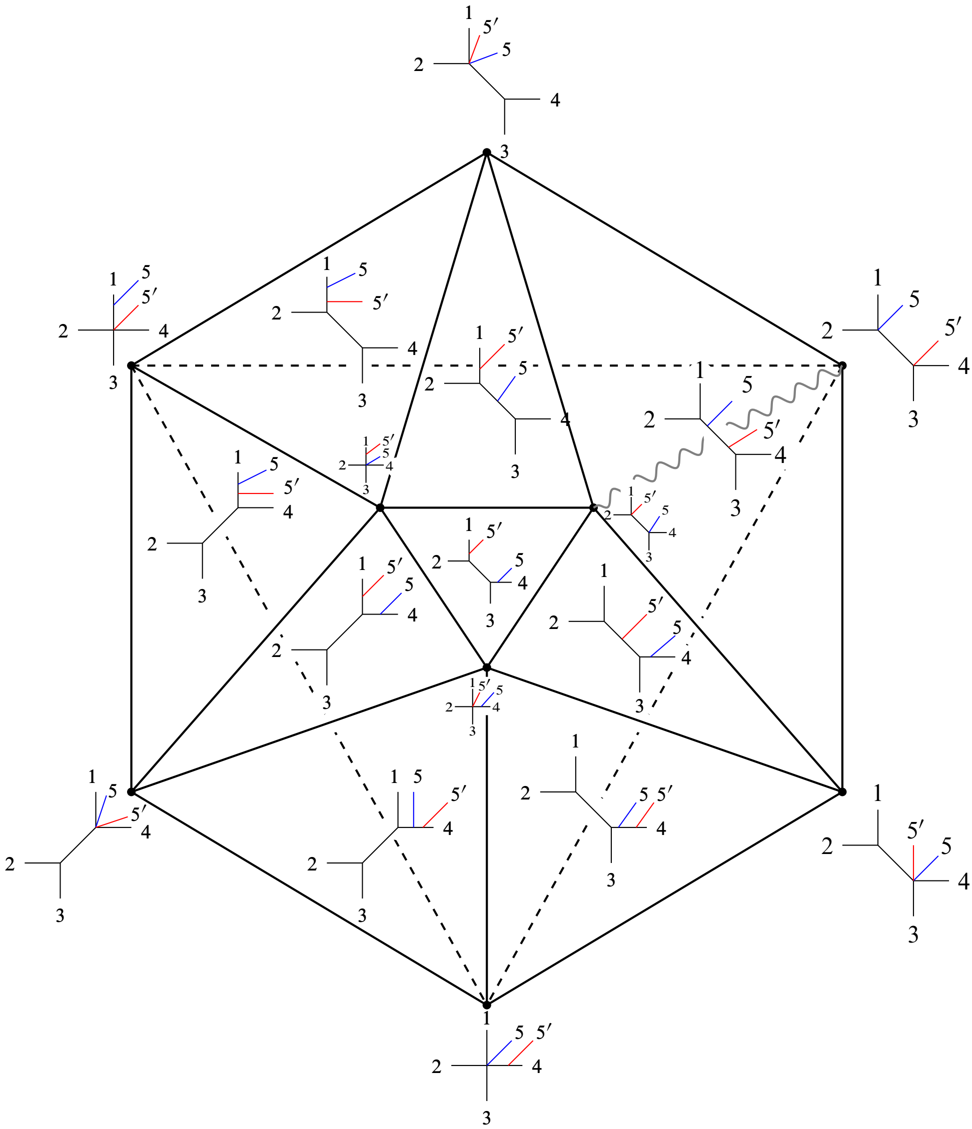

Figure 2. The fan structure (ii)=(iii) of $\mathrm{TrFl}_4^{>0}$.

Similarly, we identify (μ_2, μ_3), where μ_3 has Plücker coordinates $(w_{123}, \ldots, w_{234})$, with an element of $\mathrm{TrGr}_{2,5}^{>0}$: we simply set $w_{d5'} := w_{abc}$, where $\{a, b, c\} := [4] \setminus \{d\}$.

This gives us the Plücker fan structure (i)=(iv) with thirteen maximal cones, as shown in Figure 2. To get the Gröbner fan structure (iii) we subdivide one of the cones into two, along the squiggly line shown in Figure 2. This squiggly line occurs when $\mathrm{dist}(x_1, \mathit{blue}) = \mathrm{dist}(x_2, \mathit{red})$, where x_1 and x_2 are the two black trivalent nodes in the tree on [4]. To obtain the fan structure (v), instead of the squiggly line, the square face is subdivided along the other diagonal.

Using the computation of TrFl_5 in [Bossinger et al. 2017], which can be found on github at https://github.com/Saralamboglia/Toric-Degenerations/blob/master/Flag5.rtf and Corollary 3.12, we further computed that TrFl_5^+ with (iii) has 938 maximal cones (906 of which are simplicial) and that (iv) has 406 maximal cones. According to [Speyer and Williams 2005, Conjecture 8.1], the (v) fan structure for TrFl_5^+ has 672 maximal cones.

8.2. *Coherent subdivisions.* We next discuss coherent subdivisions coming from the positive tropical Grassmannian and positive tropical complete flag variety. When $\mathrm{Fl}_{\boldsymbol{r};n}$ is the Grassmannian $\mathrm{Gr}_{d,n}$ and the support $\underline{\boldsymbol{\mu}}$ is the uniform matroid, Theorem A gives rise to the following corollary (which was first proved in [Lukowski et al. 2023] and [Arkani-Hamed et al. 2021b]).

Corollary 8.2. *Let $\boldsymbol{\mu} = (\mu_d) \in \mathbb{P}(\mathbb{T}^{\binom{[n]}{d}})$, and suppose it has no ∞ coordinates. Then the following statements are equivalent*:

- $\boldsymbol{\mu} \in \mathrm{TrGr}_{d,n}^{>0}$, *that is, $\boldsymbol{\mu}$ lies in the strictly positive tropical Grassmannian.*
- *Every face in the coherent subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ of the hypersimplex $\Delta_{d,n}$ induced by $\boldsymbol{\mu}$ is a positroid polytope.*

The coherent subdivisions above (called *positroidal subdivisions*) were further studied in [Speyer and Williams 2021], where the finest positroidal subdivisions were characterized in terms of series-parallel matroids. Furthermore, all finest positroidal subdivisions of $\Delta_{d,n}$ achieve equality in Speyer's *f-vector theorem*; in particular, they all consist of $\binom{n-2}{d-1}$ facets [Speyer and Williams 2021, Corollary 6.7].

When $\mathrm{Fl}_{\boldsymbol{r};n}$ is the complete flag variety Fl_n, and the support $\underline{\boldsymbol{\mu}}$ is the uniform flag matroid, Theorem A gives rise to the following corollary, which appeared in [Joswig et al. 2023, Theorem 20].

Corollary 8.3. *Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \in \prod_{i=a}^{b} \mathbb{P}(\mathbb{T}^{\binom{[n]}{i}})$, and suppose it has no ∞ coordinates. Then the following statements are equivalent.*

- $\boldsymbol{\mu} \in \mathrm{TrFl}_n^{>0}$, *that is, $\boldsymbol{\mu}$ lies in the strictly positive tropical flag variety.*
- *Every face in the coherent subdivision $\mathcal{D}_{\boldsymbol{\mu}}$ of the permutohedron* Perm_n *induced by $\boldsymbol{\mu}$ is a Bruhat interval polytope.*

In light of the results of [Speyer and Williams 2021], it is natural to ask if one can characterize the finest coherent subdivisions of the permutohedron Perm_n into Bruhat interval polytopes. Furthermore, do they all have the same f-vector?

Explicit computations for TrFl_4 show that the answer to the second question is no. We find that TrFl_4 with the fan structure (iii) (which agrees with (ii) by [Joswig et al. 2023, Section 5]) has 78 maximal cones. We choose a point in the relative interior of each of the 78 cones to use as a height function (thinking of points in TrFl_4 as weights on the vertices of Perm_4 as in (c) of Theorem A), then use Sage to compute the corresponding coherent subdivision of Perm_4. As expected, precisely 14 of the 78 cones induce subdivisions of Perm_4 into Bruhat interval polytopes, see Table 1.

<table>
<tr><th>height function (P_1, P_2, P_3, P_4; P_{12}, P_{13}, P_{14}, P_{23}, P_{24}, P_{34}; P_{123}, P_{124}, P_{134}, P_{234})</th><th>Bruhat interval polytopes in subdivision</th><th>f-vector</th></tr>
<tr><td>$(15, -1, -7, -7; 4, -2, -2, -2, -2, 4; -7, -7, -1, 15)$</td><td>$P_{3214,4321}$, $P_{3124,4231}$, $P_{2314,3421}$, $P_{2134,3241}$, $P_{1324,2431}$, $P_{1234,2341}$</td><td rowspan="12">$(24, 46, 29, 6)$</td></tr>
<tr><td>$(15, 3, -9, -9; 4, -8, -8, -4, -4, 20; -1, -1, -1, 3)$</td><td>$P_{2413,4321}$, $P_{3124,4231}$, $P_{2314,4231}$, $P_{2134,3241}$, $P_{1324,2431}$, $P_{1234,2341}$</td></tr>
<tr><td>$(15, -7, -1, -7; -2, 4, -2, -2, 4, -2; -7, -1, -7, 15)$</td><td>$P_{3142,4321}$, $P_{3124,4312}$, $P_{2143,3421}$, $P_{2134,3412}$, $P_{1243,2431}$, $P_{1234,2413}$</td></tr>
<tr><td>$(-1, -1, -1, 3; 4, -8, -4, -8, -4, 20; 15, 3, -9, -9)$</td><td>$P_{2413,4321}$, $P_{1423,4231}$, $P_{1342,4231}$, $P_{1324,4213}$, $P_{1243,4132}$, $P_{1234,4123}$</td></tr>
<tr><td>$(-7, -7, -1, 15; 4, -2, -2, -2, -2, 4; 15, -1, -7, -7)$</td><td>$P_{1432,4321}$, $P_{1423,4312}$, $P_{1342,4231}$, $P_{1324,4213}$, $P_{1243,4132}$, $P_{1234,4123}$</td></tr>
<tr><td>$(-1, -7, -7, 15; -2, -2, 4, 4, -2, -2; 15, -7, -7, -1)$</td><td>$P_{3142,4321}$, $P_{2143,4312}$, $P_{2134,4213}$, $P_{1342,3421}$, $P_{1243,3412}$, $P_{1234,2413}$</td></tr>
<tr><td>$(-9, -9, 3, 15; 20, -4, -8, -4, -8, 4; 3, -1, -1, -1)$</td><td>$P_{1432,4321}$, $P_{1423,4312}$, $P_{1342,4231}$, $P_{1324,4213}$, $P_{1324,4132}$, $P_{1234,3142}$</td></tr>
<tr><td>$(11, -7, -7, 3; -6, -6, 4, 4, 2, 2; 11, -7, -7, 3)$</td><td>$P_{3142,4321}$, $P_{2143,4312}$, $P_{2134,4213}$, $P_{2143,3421}$, $P_{1243,2431}$, $P_{1234,2413}$</td></tr>
<tr><td>$(3, 3, -3, -3; 20, -10, -10, -10, -10, 20; -3, -3, 3, 3)$</td><td>$P_{2413,4321}$, $P_{3124,4231}$, $P_{2314,4231}$, $P_{1324,2431}$, $P_{1324,3241}$, $P_{1234,3142}$</td></tr>
<tr><td>$(3, -1, -1, -1; 20, -4, -4, -8, -8, 4; -9, -9, 3, 15)$</td><td>$P_{3214,4321}$, $P_{3124,4231}$, $P_{2314,3421}$, $P_{1324,3241}$, $P_{1324,2431}$, $P_{1234,3142}$</td></tr>
<tr><td>$(-3, -3, 3, 3; 20, -10, -10, -10, -10, 20; 3, 3, -3, -3)$</td><td>$P_{2413,4321}$, $P_{1423,4231}$, $P_{1342,4231}$, $P_{1324,4132}$, $P_{1324,4213}$, $P_{1234,3142}$</td></tr>
<tr><td>$(3, -7, -7, 11; 2, 2, 4, 4, -6, -6; 3, -7, -7, 11)$</td><td>$P_{3142,4321}$, $P_{3124,4312}$, $P_{1342,3421}$, $P_{2134,3412}$, $P_{1243,3412}$, $P_{1234,2413}$</td></tr>
<tr><td>$(11, -1, -7, -3; -2, -8, -4, -4, 0, 18; 11, -1, -7, -3)$</td><td>$P_{2413,4321}$, $P_{2143,4231}$, $P_{2134,4213}$, $P_{1243,2431}$, $P_{1234,2413}$</td><td rowspan="2">$(24, 45, 27, 5)$</td></tr>
<tr><td>$(-3, -7, -1, 11; 18, 0, -4, -4, -8, -2; -3, -7, -1, 11)$</td><td>$P_{3142,4321}$, $P_{3124,4312}$, $P_{1342,3421}$ $P_{1324,3412}$, $P_{1234,3142}$</td></tr>
</table>

Table 1. Table documenting the 14 finest coherent subdivisions of Perm_4 into Bruhat interval polytopes. There are two possible f-vectors, each of which can be realized in multiple ways.

Of the 14 coherent subdivisions coming from maximal cones of $\mathrm{TrFl}_4^{\geq 0}$, 12 of them contain 6 facets, while the other 2 contain 5 facets. Table 1 lists the facets and f-vectors of each of these 14 subdivisions. Note that each Bruhat interval polytope $P_{v,w}$ which appears as a facet satisfies $\ell(w) - \ell(v) = 3$. Thus, any

height function $(P_1, P_2, P_3, P_4; P_{12}, P_{13}, P_{14}, P_{23}, P_{24}, P_{34}; P_{123}, P_{124}, P_{134}, P_{234})$	Bruhat interval polytopes in subdivision	f-vector
$(-1,-1,-1,0;-1,-1,0,-1,0,0;0,0,0,0)$	$P_{1243,4321}$, $P_{1234,4213}$	(24, 39, 18, 2)
$(-1,-1,-1,0;0,0,0,0,0,0;0,0,0,0)$	$P_{1342,4321}$, $P_{1234,4312}$	
$(1,0,0,0;0,0,0,0,0,0;0,0,0,0)$	$P_{2134,4321}$, $P_{1234,2431}$	
$(1,0,0,0;0,0,0,1,1,1;0,0,0,0)$	$P_{3124,4321}$, $P_{1234,3421}$	
$(0,0,0,0;-1,-1,-1,-1,-1,0;0,0,0,0)$	$P_{2413,4321}$, $P_{1234,4231}$	(24, 40, 19, 2)
$(0,0,0,0;1,0,0,0,0,0;0,0,0,0)$	$P_{1324,4321}$, $P_{1234,3142}$	
$(-1,-1,0,0;-1,-1,-1,-1,-1,0;0,0,0,0)$	$P_{1423,4321}$, $P_{1342,4231}$, $P_{1324,4213}$, $P_{1234,4132}$	(24, 42, 23, 4)
$(0,-1,-1,0;0,0,1,0,0,0;0,0,0,0)$	$P_{3142,4321}$, $P_{1243,3421}$, $P_{2134,4312}$, $P_{1234,2413}$	
$(1,1,0,0;1,0,0,0,0,0;0,0,0,0)$	$P_{2314,4321}$, $P_{1324,2431}$, $P_{3124,4231}$, $P_{1234,3241}$	

Table 2. Table documenting the 9 coarsest coherent subdivisions of Perm_4 into Bruhat interval polytopes. There are three possible f-vectors, each of which can be realized in multiple ways.

Bruhat interval polytope $P_{v',w}$ properly contained inside $P_{v,w}$ would have the property that $\ell(w')-\ell(v') \leq 2$, and hence $\dim(P_{v',w'}) \leq 2$. Since Perm_4 is 3-dimensional, all 14 of these subdivisions are finest subdivisions.

We note that the 12 finest subdivisions whose f-vector is $(24, 46, 29, 6)$ are subdivisions of the permutohedron into cubes. Subdivisions of the permutohedron into Bruhat interval polytopes which are cubes have been previously studied in [Harada et al. 2019, Sections 5 and 6; Lee et al. 2021; Nadeau and Tewari 2023, Section 6]. In particular, there is a subdivision of Perm_n into $(n-1)!$ Bruhat interval polytopes

$$\{P_{u,v} \mid u=(u_1\dots,u_n) \text{ with } u_n=n, \text{ and } v=(v_1,\dots,v_n) \text{ with } v_i=u_i+1 \text{ modulo } n\}.$$

The first subdivision in Table 1 has this form.

We can further study the f-vectors of subdivisions of $\mathrm{TrFl}_4^{>0}$ which are coarsest (without being trivial), rather than finest. In this case, we observe three different f-vectors, each of which occurs in multiple subdivisions. The detailed results of our explicit computations on coarsest subdivisions can be found in Table 2.

Acknowledgements. Boretsky is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Boretsky a été financé par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [Ref. no. 557353-2021]. Eur is partially supported by the US National

Science Foundation (DMS-2001854). Williams is partially supported by the National Science Foundation (DMS-1854512 and DMS-2152991). We are grateful to Tony Bloch, Michael Joswig, Steven Karp, Georg Loho, Dante Luber, and Jorge Alberto Olarte for sharing their work with us, which partially inspired this project. We also thank Yue Ren, Vasu Tewari, and Jorge Olarte for useful comments. Finally, we are grateful to Lara Bossinger for several invaluable discussions about fan structures.

References

[Ardila and Klivans 2006] F. Ardila and C. J. Klivans, "The Bergman complex of a matroid and phylogenetic trees", *J. Combin. Theory Ser. B* **96**:1 (2006), 38–49. MR Zbl

[Ardila et al. 2010] F. Ardila, A. Fink, and F. Rincón, "Valuations for matroid polytope subdivisions", *Canad. J. Math.* **62**:6 (2010), 1228–1245. MR Zbl

[Ardila et al. 2016] F. Ardila, F. Rincón, and L. Williams, "Positroids and non-crossing partitions", *Trans. Amer. Math. Soc.* **368**:1 (2016), 337–363. MR Zbl

[Ardila et al. 2017] F. Ardila, F. Rincón, and L. Williams, "Positively oriented matroids are realizable", *J. Eur. Math. Soc.* (*JEMS*) **19**:3 (2017), 815–833. MR Zbl

[Arkani-Hamed et al. 2021a] N. Arkani-Hamed, S. He, and T. Lam, "Cluster configuration spaces of finite type", *SIGMA Symmetry Integrability Geom. Methods Appl.* **17** (2021), art. id. 092. MR Zbl

[Arkani-Hamed et al. 2021b] N. Arkani-Hamed, T. Lam, and M. Spradlin, "Positive configuration space", *Comm. Math. Phys.* **384**:2 (2021), 909–954. MR Zbl

[Baker and Bowler 2019] M. Baker and N. Bowler, "Matroids over partial hyperstructures", *Adv. Math.* **343** (2019), 821–863. MR Zbl

[Benedetti et al. 2022] C. Benedetti, A. Chavez, and D. Tamayo Jiménez, "Quotients of uniform positroids", *Electron. J. Combin.* **29**:1 (2022), art. id. 1.13. MR Zbl

[Billey and Weaver 2022] S. C. Billey and J. E. Weaver, "Criteria for smoothness of Positroid varieties via pattern avoidance, Johnson graphs, and spirographs", preprint, 2022. Zbl arXiv 2207.06508

[Björner et al. 1999] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler, *Oriented matroids*, 2nd ed., Encyclopedia of Mathematics and its Applications **46**, Cambridge University Press, 1999. MR Zbl

[Bloch and Karp 2023] A. M. Bloch and S. N. Karp, "On two notions of total positivity for partial flag varieties", *Adv. Math.* **414** (2023), art. id. 108855. MR Zbl

[Boretsky 2022] J. Boretsky, "Totally nonnegative tropical flags and the totally nonnegative flag Dressian", preprint, 2022. Zbl arXiv 2208.09128v2

[Borovik et al. 2003] A. V. Borovik, I. M. Gelfand, and N. White, *Coxeter matroids*, Progress in Mathematics **216**, Birkhäuser, Boston, 2003. MR Zbl

[Bossinger 2021] L. Bossinger, "Birational sequences and the tropical Grassmannian", *J. Algebra* **585** (2021), 784–803. MR Zbl

[Bossinger 2022] L. Bossinger, "Tropical totally positive cluster varieties", preprint, 2022. Zbl arXiv 2208.01723

[Bossinger et al. 2017] L. Bossinger, S. Lamboglia, K. Mincheva, and F. Mohammadi, "Computing toric degenerations of flag varieties", pp. 247–281 in *Combinatorial algebraic geometry*, edited by G. G. Smith and B. Sturmfels, Fields Inst. Commun. **80**, Fields Inst. Res. Math. Sci., Toronto, ON, 2017. MR Zbl

[Brandt et al. 2021] M. Brandt, C. Eur, and L. Zhang, "Tropical flag varieties", *Adv. Math.* **384** (2021), art. id. 107695. MR Zbl

[Cachazo et al. 2019] F. Cachazo, N. Early, A. Guevara, and S. Mizera, "Scattering equations: from projective spaces to tropical Grassmannians", *J. High Energy Phys.* 6 (2019), art. id. 039. MR Zbl

[Early 2022] N. Early, "From weakly separated collections to matroid subdivisions", *Comb. Theory* **2**:2 (2022), art. id. 2. MR Zbl

[Fomin and Zelevinsky 2003] S. Fomin and A. Zelevinsky, "Cluster algebras, II: finite type classification", *Invent. Math.* **154**:1 (2003), 63–121. MR Zbl

[Fulton 1997] W. Fulton, *Young tableaux*, London Mathematical Society Student Texts **35**, Cambridge University Press, 1997. MR Zbl

[Geiss et al. 2008] C. Geiss, B. Leclerc, and J. Schröer, "Preprojective algebras and cluster algebras", pp. 253–283 in *Trends in representation theory of algebras and related topics*, edited by A. Skowroński, Eur. Math. Soc., Zürich, 2008. MR Zbl

[Gelfand and Serganova 1987] I. M. Gelfand and V. V. Serganova, "Combinatorial geometries and the strata of a torus on homogeneous compact manifolds", *Uspekhi Mat. Nauk* **42**:2(254) (1987), 107–134, 287. MR Zbl

[Gelfand et al. 1987] I. M. Gelfand, R. M. Goresky, R. D. MacPherson, and V. V. Serganova, "Combinatorial geometries, convex polyhedra, and Schubert cells", *Adv. in Math.* **63**:3 (1987), 301–316. MR Zbl

[Gunn 2019] T. Gunn, "A Newton polygon rule for formally-real valued fields and multiplicities over the signed tropical hyperfield", preprint, 2019. Zbl arXiv 1911.12274

[Harada et al. 2019] M. Harada, T. Horiguchi, M. Masuda, and S. Park, "The volume polynomial of regular semisimple Hessenberg varieties and the Gelfand–Tsetlin polytope", *Tr. Mat. Inst. Steklova* **305** (2019), 344–373. MR Zbl

[Herrmann et al. 2009] S. Herrmann, A. Jensen, M. Joswig, and B. Sturmfels, "How to draw tropical planes", *Electron. J. Combin.* **16**:2 (2009), art. id. 6. MR Zbl

[Herrmann et al. 2014] S. Herrmann, M. Joswig, and D. E. Speyer, "Dressians, tropical Grassmannians, and their rays", *Forum Math.* **26**:6 (2014), 1853–1881. MR Zbl

[Jahn et al. 2021] D. Jahn, R. Löwe, and C. Stump, "Minkowski decompositions for generalized associahedra of acyclic type", *Algebr. Comb.* **4**:5 (2021), 757–775. MR Zbl

[Jarra and Lorscheid 2024] M. Jarra and O. Lorscheid, "Flag matroids with coefficients", *Adv. Math.* **436** (2024), art. id. 109396. MR Zbl

[Joswig et al. 2023] M. Joswig, G. Loho, D. Luber, and J. A. Olarte, "Generalized permutahedra and positive flag dressians", *Int. Math. Res. Not.* **2023**:19 (2023), 16748–16777. MR Zbl

[Kodama and Williams 2015] Y. Kodama and L. Williams, "The full Kostant–Toda hierarchy on the positive flag variety", *Comm. Math. Phys.* **335**:1 (2015), 247–283. MR Zbl

[Kung 1986] J. P. S. Kung, "Strong maps", pp. 224–253 in *Theory of matroids*, edited by N. White, Encyclopedia Math. Appl. **26**, Cambridge University Press, 1986. MR Zbl

[Le and Fraser 2019] I. Le and C. Fraser, "Tropicalization of positive Grassmannians", *Selecta Math.* (*N.S.*) **25**:5 (2019), art. id. 75. MR Zbl

[Lee et al. 2021] E. Lee, M. Masuda, and S. Park, "Toric Bruhat interval polytopes", *J. Combin. Theory Ser. A* **179** (2021), art. id. 105387. MR Zbl

[Lukowski et al. 2023] T. Lukowski, M. Parisi, and L. K. Williams, "The positive tropical Grassmannian, the hypersimplex, and the $m = 2$ amplituhedron", *Int. Math. Res. Not.* **2023**:19 (2023), 16778–16836. MR Zbl

[Lusztig 1994] G. Lusztig, "Total positivity in reductive groups", pp. 531–568 in *Lie theory and geometry*, edited by J.-L. Brylinski et al., Progr. Math. **123**, Birkhäuser, Boston, 1994. MR Zbl

[Maclagan and Sturmfels 2015] D. Maclagan and B. Sturmfels, *Introduction to tropical geometry*, Graduate Studies in Mathematics **161**, American Mathematical Society, Providence, RI, 2015. MR Zbl

[Markwig 2010] T. Markwig, "A field of generalised Puiseux series for tropical geometry", *Rend. Semin. Mat. Univ. Politec. Torino* **68**:1 (2010), 79–92. MR Zbl

[Marsh and Rietsch 2004] R. J. Marsh and K. Rietsch, "Parametrizations of flag varieties", *Represent. Theory* **8** (2004), 212–242. MR Zbl

[Murota and Shioura 2018] K. Murota and A. Shioura, "On equivalence of $M^\natural$-concavity of a set function and submodularity of its conjugate", *J. Oper. Res. Soc. Japan* **61**:2 (2018), 163–171. MR Zbl

[Nadeau and Tewari 2023] P. Nadeau and V. Tewari, "Remixed Eulerian numbers", *Forum Math. Sigma* **11** (2023), art. id. e65. MR Zbl

[Oh 2008] S. Oh, “Contraction and restriction of positroids in terms of decorated permutations”, preprint, 2008. Zbl arXiv 0804.0882

[Olarte et al. 2019] J. A. Olarte, M. Panizzut, and B. Schröter, “On local Dressians of matroids”, pp. 309–329 in *Algebraic and geometric combinatorics on lattice polytopes*, edited by T. Hibi and A. Tsuchiya, World Sci. Publ., Hackensack, NJ, 2019. MR Zbl

[Oxley 2011] J. Oxley, *Matroid theory*, 2nd ed., Oxford Graduate Texts in Mathematics **21**, Oxford University Press, 2011. MR Zbl

[Pachter and Sturmfels 2004] L. Pachter and B. Sturmfels, “Tropical geometry of statistical models”, *Proc. Natl. Acad. Sci. USA* **101**:46 (2004), 16132–16137. MR Zbl

[Poonen 1993] B. Poonen, “Maximally complete fields”, *Enseign. Math.* (2) **39**:1-2 (1993), 87–106. MR Zbl

[Postnikov 2007] A. Postnikov, “Total positivity, Grassmannians, and networks”, 2007, available at http://math.mit.edu/~apost/papers/tpgrass.pdf. Zbl

[Rietsch 1998] K. C. Rietsch, *Total positivity and real flag varieties*, Ph.D. thesis, Massachusetts Institute of Technology, 1998, available at https://www.proquest.com/docview/304473828. Zbl

[Rietsch 2006] K. Rietsch, “Closure relations for totally nonnegative cells in G/P”, *Math. Res. Lett.* **13**:5-6 (2006), 775–786. MR Zbl

[Rietsch and Williams 2008] K. Rietsch and L. Williams, “The totally nonnegative part of G/P is a CW complex”, *Transform. Groups* **13**:3-4 (2008), 839–853. MR Zbl

[Rietsch and Williams 2019] K. Rietsch and L. Williams, “Newton–Okounkov bodies, cluster duality, and mirror symmetry for Grassmannians”, *Duke Math. J.* **168**:18 (2019), 3437–3527. MR Zbl

[da Silva 1987] I. P. da Silva, *Quelques propriétés des matroides orientés*, Ph.D. thesis, Université Paris VI, 1987.

[Speyer 2008] D. E. Speyer, “Tropical linear spaces”, *SIAM J. Discrete Math.* **22**:4 (2008), 1527–1558. MR Zbl

[Speyer and Sturmfels 2004] D. Speyer and B. Sturmfels, “The tropical Grassmannian”, *Adv. Geom.* **4**:3 (2004), 389–411. MR Zbl

[Speyer and Williams 2005] D. Speyer and L. Williams, “The tropical totally positive Grassmannian”, *J. Algebraic Combin.* **22**:2 (2005), 189–210. MR Zbl

[Speyer and Williams 2021] D. Speyer and L. K. Williams, “The positive Dressian equals the positive tropical Grassmannian”, *Trans. Amer. Math. Soc. Ser. B* **8** (2021), 330–353. MR Zbl

[Talaska and Williams 2013] K. Talaska and L. Williams, “Network parametrizations for the Grassmannian”, *Algebra Number Theory* **7**:9 (2013), 2275–2311. MR Zbl

[Tsukerman and Williams 2015] E. Tsukerman and L. Williams, “Bruhat interval polytopes”, *Adv. Math.* **285** (2015), 766–810. MR Zbl

Communicated by Victor Reiner
Received 2022-09-16 Revised 2023-07-10 Accepted 2023-09-03

jboretsky@math.harvard.edu *Department of Mathematics, Harvard University, Cambridge, MA, United States*

ceur@math.harvard.edu *Department of Mathematics, Harvard University, Cambridge, MA, United States*

williams@math.harvard.edu *Department of Mathematics, Harvard University, Cambridge, MA, United States*

ALGEBRA AND NUMBER THEORY 18:7 (2024)
https://doi.org/10.2140/ant.2024.18.1375

Maximal subgroups of exceptional groups and Quillen's dimension

Kevin I. Piterman

Given a finite group G and a prime p, let $\mathcal{A}_p(G)$ be the poset of nontrivial elementary abelian p-subgroups of G. The group G satisfies the Quillen dimension property at p if $\mathcal{A}_p(G)$ has nonzero homology in the maximal possible degree, which is the p-rank of G minus 1. For example, D. Quillen showed that solvable groups with trivial p-core satisfy this property, and later, M. Aschbacher and S. D. Smith provided a list of all p-extensions of simple groups that may fail this property if p is odd. In particular, a group G with this property satisfies Quillen's conjecture: G has trivial p-core and the poset $\mathcal{A}_p(G)$ is not contractible.

In this article, we focus on the prime $p = 2$ and prove that the 2-extensions of finite simple groups of exceptional Lie type in odd characteristic satisfy the Quillen dimension property, with only finitely many exceptions. We achieve these conclusions by studying maximal subgroups and usually reducing the problem to the same question in small linear groups, where we establish this property via counting arguments. As a corollary, we reduce the list of possible components in a minimal counterexample to Quillen's conjecture at $p = 2$.

1. Introduction

Since the early 70s, there has been a growing interest in the p-subgroup posets and their connections with finite group theory, the classification of the finite simple groups, finite geometries, group cohomology and representation theory. The poset $\mathcal{S}_p(G)$ of nontrivial p-subgroups of a group G was introduced by Kenneth Brown [1975]. In that paper, Brown worked with the Euler characteristic $\chi(G)$ of groups G satisfying certain finiteness conditions and established connections between the p-fractional part of $\chi(G)$ and the p-subgroup structure of G. One of the consequences of his results is the commonly known "Homological Sylow theorem", which states that the Euler characteristic of $\mathcal{S}_p(G)$ is 1 modulo $|G|_p$, the order of a Sylow p-subgroup of G.

Some years later, Daniel Quillen [1978] introduced the poset $\mathcal{A}_p(G)$ of nontrivial elementary abelian p-subgroups of a finite group G and exhibited several applications of the topological properties of these posets. Indeed, the study of elementary abelian p-subgroups goes back to Quillen's earlier work on the Bredon cohomology of G-spaces and his proof of the Atiyah–Swan conjecture, that relates the Krull dimension of a ring to the dimension of $\mathcal{A}_p(G)$ (see [Quillen 1971]).

Piterman is supported by the Alexander von Humboldt Foundation and FWO grant 12K1223N.
MSC2020: 05E18, 20D20, 20D30, 20G41.
Keywords: p-subgroups, exceptional groups of Lie type, Quillen's conjecture.

Quillen [1978] showed that $\mathcal{S}_p(G)$ and $\mathcal{A}_p(G)$ are (G-equivariantly) homotopy equivalent, and provided a new proof of Brown's result. In fact, when G is the set of rational points of a semisimple algebraic group over a finite field of characteristic p, these posets are homotopy equivalent to the building of G and, hence, they have the homotopy type of a wedge of spheres of dimension $l-1$, where l is the rank of the underlying algebraic group. Furthermore, in that case, the homology $\widetilde{H}_*(\mathcal{A}_p(G))$ affords the classical Steinberg module for G.

Quillen also exhibited other connections between intrinsic algebraic properties of G and the topology of these posets. For instance, he showed that $\mathcal{A}_p(G)$ is disconnected if and only if G contains a strongly p-embedded subgroup. Recall that the classification of the groups with this property is indeed one of the many important steps towards the classification of the finite simple groups (see, for example, Section 7.6 of [Gorenstein et al. 1998]).

On the other hand, Quillen proved that if G has a fixed point on $\mathcal{A}_p(G)$ (or, equivalently on $\mathcal{S}_p(G)$), then these posets are contractible. Note that G has a fixed point if and only if its p-core $O_p(G)$ is nontrivial. In view of this and further evidence, Quillen conjectured that the reciprocal to this statement should hold. That is, if $\mathcal{A}_p(G)$ is contractible then there is a fixed point, or, equivalently, $O_p(G) \neq 1$ (see Conjecture 2.9 of [Quillen 1978]). In other words, Quillen's conjecture asserts that $\mathcal{A}_p(G)$ is contractible if and only if $O_p(G) \neq 1$.

A significant part of Quillen's article is devoted to proving the solvable case of this conjecture. In [Quillen 1978] it is shown that for a p-nilpotent group G with abelian Sylow p-subgroups and $O_p(G) = 1$, $\mathcal{A}_p(G)$ is homotopy equivalent to a nontrivial wedge of spheres of the maximal possible dimension, which is $m_p(G) - 1$, the p-rank of G minus 1. Then, if G is any solvable group with $O_p(G) = 1$, G contains a p-nilpotent subgroup $O_{p'}(G)A$, with $A \in \mathcal{A}_p(G)$ of maximal p-rank and $O_p(O_{p'}(G)A) = 1$, and thus $\widetilde{H}_{m_p(G)-1}(\mathcal{A}_p(G)) \neq 0$.

Later, Michael Aschbacher and Stephen D. Smith [1993] formalised this property and gave a name to it: an arbitrary group G with $\widetilde{H}_{m_p(G)-1}(\mathcal{A}_p(G)) \neq 0$ is said to satisfy the *Quillen dimension property at p*, or $(\mathcal{QD})_p$ for short. Therefore, a solvable group G with $O_p(G) = 1$ satisfies $(\mathcal{QD})_p$ and thus Quillen's conjecture. Furthermore, it was shown that p-solvable groups satisfy this property by using Quillen's techniques and, in addition, the CFSG (see [Díaz Ramos 2018; Smith 2011]). These results also suggest that a stronger statement of the conjecture may hold: if $O_p(G) = 1$ then $\widetilde{H}_*(\mathcal{A}_p(G); \mathbb{Q}) \neq 0$. Therefore, from now on, by Quillen's conjecture we will be referring to this stronger version.

It is not hard to see that not every group G with $O_p(G) = 1$ satisfies $(\mathcal{QD})_p$. For example, we mentioned that finite groups of Lie type in characteristic p satisfy the conjecture, but since the Lie rank is usually strictly smaller than the p-rank, they fail $(\mathcal{QD})_p$. This has led to the development of new methods to prove Quillen's conjecture. One of the most notorious advances in the conjecture was achieved by Aschbacher and Smith [1993]. They established Quillen's conjecture for a group G if $p > 5$ and in addition, roughly, all the p-extensions of finite unitary groups $\mathrm{PSU}_n(q)$, with q odd and $p \mid q+1$, satisfy $(\mathcal{QD})_p$ (see Main Theorem of [Aschbacher and Smith 1993] for the precise statement). Here, a p-extension of a group L is a split extension of L by an elementary abelian p-subgroup of $\mathrm{Out}(L)$. In [Aschbacher and Smith 1993] it is not shown that the group G satisfies $(\mathcal{QD})_p$. Instead, they proved that if every p-extension of a fixed component of G satisfies $(\mathcal{QD})_p$, then

$$\widetilde{H}_*(\mathcal{A}_p(G);\mathbb{Q}) \neq 0 \quad \text{if } O_p(G)=1$$

(under suitable inductive hypotheses). This result restricts the possibilities of the components of a minimal counterexample to Quillen's conjecture: every component has a p-extension failing $(\mathcal{QD})_p$. In view of this result and the classification of the finite simple groups, Aschbacher and Smith described for $p \geq 3$, all the possible p-extensions of simple groups which may potentially fail $(\mathcal{QD})_p$. This is the $(\mathcal{QD})$-List, Theorem 3.1, of [Aschbacher and Smith 1993]. Moreover, it is conjectured in [Aschbacher and Smith 1993] that the unitary groups $\mathrm{PSU}_n(q)$ with q odd and $p \mid q+1$ should not appear in this list, and so the extra hypothesis on the unitary groups in the main result of [Aschbacher and Smith 1993] could be omitted. Nevertheless, this problem remains open (see [Piterman and Welker 2022] for recent results in this direction).

In the last few years, there have been further developments in the Quillen conjecture [Piterman 2021; Piterman et al. 2021; Piterman and Smith 2022a; 2022b]. Recently, in [Piterman and Smith 2022b], new tools for the study of the conjecture have been provided. For example, it is shown that the Aschbacher–Smith general approach to the conjecture can be extended to *every* prime p by reducing reliance on results of [Aschbacher and Smith 1993] stated only for odd primes and invoking the classification. In particular, Theorem 1.1 of [Piterman and Smith 2022b] shows that Main Theorem of [Aschbacher and Smith 1993] extends to $p \geq 3$, keeping the additional constraint on the unitary groups. On the other hand, for $p=2$, one important obstruction for this extension is the lack of a $(\mathcal{QD})$-List for this prime. Roughly, Corollary 1.8 of [Piterman and Smith 2022b] concludes that a minimal counterexample to Quillen's conjecture contains a component of Lie type in characteristic $r \neq 3$, and every component of G has a 2-extension failing $(\mathcal{QD})_2$.

In view of these results on Quillen's conjecture, in this article, we focus on showing that the 2-extensions of the finite simple groups of exceptional Lie type in odd characteristic satisfy $(\mathcal{QD})_2$, with a small number of exceptions. This improves the conclusions of [Piterman and Smith 2022b] on Quillen's conjecture for $p=2$, and allows us to conclude then that exceptional groups of Lie type in odd characteristic different from 3 cannot be components of a minimal counterexample to the conjecture (see Corollary 1.2 below).

The main result of this article is the following theorem, whose proof is given through different propositions in Section 5.

Theorem 1.1. *Let L be a finite simple group of exceptional Lie type in odd characteristic. That is, $L = {}^3D_4(q)$, $F_4(q)$, $G_2(q)$, ${}^2G_2(q)'$, $E_6(q)$, ${}^2E_6(q)$, $E_7(q)$ or $E_8(q)$, with q odd. Then every 2-extension of L satisfies the Quillen dimension property at $p=2$, except possibly in the following cases*:

- ${}^3D_4(9)$ *extended with field automorphisms*;
- $F_4(3)$, $F_4(9)$ *extended with field automorphisms*;
- 2-*extensions of* $G_2(3)$, $G_2(9)$ *extended with field automorphisms*;
- ${}^2G_2(3)'$, $E_8(3)$, $E_8(9)$ *extended with field automorphisms.*

Indeed, the extensions of $G_2(3)$, $G_2(9)$ and ${}^2G_2(3)'$ mentioned above do fail $(\mathcal{QD})_2$ by Example 5.3 and Proposition 5.1.

To achieve the conclusions of Theorem 1.1, in most cases we exhibit a maximal subgroup M of a 2-extension LB of L such that $m_2(M) = m_2(LB)$ and M satisfies $(\mathcal{QD})_2$. Since there is an inclusion $\widetilde{H}_{m_2(LB)-1}(\mathcal{A}_2(M)) \hookrightarrow \widetilde{H}_{m_2(LB)-1}(\mathcal{A}_2(LB))$ in the top-degree homology groups, this establishes $(\mathcal{QD})_2$ for LB (see Lemma 3.3). In some cases, the subgroup M arises from suitable parabolic subgroups. More concretely, when it is possible, we pick P to be a maximal parabolic subgroup of L which is stabilised by B and such that $M := PB$ realises the 2-rank of LB. Then we get a 2-nilpotent configuration UA, where U is the unipotent radical of P, A is an elementary abelian 2-subgroup realising the 2-rank of PB, and $O_2(UA) = C_A(U) = 1$ by one of the corollaries of the Borel–Tits theorem. Hence, by Quillen's results on the solvable case, UA satisfies $(\mathcal{QD})_2$, and thus also M and LB.

When the choice of such parabolic P is not possible, we pick one of the maximal rank subgroups of L. Here, the components of the maximal subgroup M are usually smaller exceptional groups, low-dimensional linear group $A_1(q)$ and $A_2(q)$ or unitary groups ${}^2A_2(q)$. Therefore, we first prove that the 2-extensions of simple linear and unitary groups in dimensions 2 and 3 satisfy $(\mathcal{QD})_2$.

Although there is a large literature on maximal subgroups of exceptional groups of Lie type, we will only need the results from [Cohen et al. 1992; Kleidman 1988; Liebeck et al. 1992; Liebeck and Seitz 1990; 2004].

Finally, from Theorem 1.1 and the results of [Piterman and Smith 2022b] for $p = 2$, we can conclude:

Corollary 1.2. *Let G be a minimal counterexample to Quillen's conjecture for $p = 2$. Then G contains a component of Lie type in characteristic $r \neq 3$. Moreover, every such component fails $(\mathcal{QD})_2$ in some 2-extension and belongs to one of the following families*:

$$\mathrm{PSL}_n(2^a)\ \ (n \geq 3), \quad D_n(2^a)\ \ (n \geq 4), \quad E_6(2^a),$$
$$\mathrm{PSL}_n^{\pm}(q)\ \ (n \geq 4), \quad B_n(q)\ \ (n \geq 2), \quad C_n(q)\ \ (n \geq 3), \quad D_n^{\pm}(q)\ \ (n \geq 4),$$

where $q = r^a$ and $r > 3$.

The 2-extensions of $\mathrm{PSL}_2(q)$, $\mathrm{PSL}_3(q)$ and $\mathrm{PSU}_3(q)$ satisfy $(\mathcal{QD})_2$ by Propositions 4.2, 4.5 and 4.6, respectively, with exceptions when $q = 3, 5, 9$. Nevertheless, the results of [Piterman and Smith 2022b] eliminate these possibilities from a minimal counterexample.

Further results on the Quillen dimension property at $p = 2$ for the classical groups could be pursued by combining the methods presented in this article with the results of [Díaz Ramos 2018; Díaz Ramos and Mazza 2022].

The paper is organised as follows. In Section 2 we set the notation and conventions that we will need to work with the finite groups of Lie type. We also provide some useful properties to work out the p-extensions and compute p-ranks. In Section 3 we gather previous results on the Quillen dimension property and related tools that will help us establish this property. Then in Section 4 we establish $(\mathcal{QD})_2$

for some 2-extensions of linear groups and recall the structure of the centralisers of graph automorphisms, following Table 4.5.1 of [Gorenstein et al. 1998]. In Section 5 we prove each case of Theorem 1.1.

All groups considered in this article are finite. We suppress the notation for the homology coefficients, and we assume that they are always taken over $\mathbb{Q}$. The interested reader may note that our results can be extended to homology with coefficients in other rings. Finally, we emphasise that we adopt the language and conventions of [Gorenstein et al. 1998]. This is particularly important when we name the different types of automorphisms of groups of Lie type. Computer calculations were performed with [GAP].

2. Preliminaries

We assume that the reader is familiar with the construction of the finite groups of Lie type as fixed points of Steinberg endomorphisms, and the basic properties concerning root systems of reductive algebraic groups. We will follow the language of [Gorenstein et al. 1998], which also contains the required background on finite groups of Lie type. In this section, we will only recall some notation and names, and state results that will be used later.

We denote by C_n, D_n, Sym_n and Alt_n the cyclic group of order n, the dihedral group of order n, the symmetric group on n points and the alternating group on n points.

If G is a group, then $\mathrm{Aut}(G)$, $\mathrm{Inn}(G)$ and $\mathrm{Out}(G)$ denote the automorphism group, the group of inner automorphisms and the outer automorphism group of G respectively. We denote by $Z(G)$ the centre of G. We usually write $G:H$, or simply GH, for a split extension of G by H. When an extension of G by H may not split, we denote it by $G.H$. By an element g (resp. a subgroup B) of G inducing outer automorphisms on $L \leq G$ we mean that $g \in N_G(L)$ embeds into $\mathrm{Aut}(L) \setminus \mathrm{Inn}(L)$ (resp. $B \leq N_G(L)$ embeds in $\mathrm{Aut}(L)$ with $B \cap \mathrm{Inn}(L) = 1$). Finally, $H \circ_m K$ denotes a central product of H and K by a central cyclic subgroup of order m. That is, $H \circ_m K = (H \times K)/\mathrm{C}_m$, where C_m embeds into both $Z(H)$ and $Z(K)$.

We will usually use the notation n in a group extension to denote a cyclic group of order n, and n^m a direct product of m copies of cyclic groups of order n. A number between brackets $[n]$ in the structure description of an extension means some group of order n.

In this article, we are mainly interested in extensions by elementary abelian groups. Below we recall the definition of p-extension given in the introduction and introduce some useful notation.

Definition 2.1. Let L be a finite group and p a prime number. A p-extension of L is a split extension LB of L by an elementary abelian p-group B inducing outer automorphisms on L.

If $L \leq G$, we denote by $\mathcal{O}_G(L)$ the poset of elements $B \in \mathcal{A}_p(N_G(L))$ such that $B \cap LC_G(L) = 1$ (that is, B induces outer automorphisms on L). We write $\mathcal{O}_2(L)$ for $\mathcal{O}_{\mathrm{Aut}(L)}(L)$ at $p = 2$. We also let $\hat{\mathcal{O}}_G(L) = \mathcal{O}_G(L) \cup \{1\}$ and $\hat{\mathcal{O}}_2(L) = \mathcal{O}_2(L) \cup \{1\}$.

Definition 2.2. For a prime number p, we say that a group G satisfies the *Quillen dimension property at p* if $\mathcal{A}_p(G)$ has nonzero homology in dimension $m_p(G) - 1$, where $m_p(G)$ denotes the p-rank of G:

$$\tilde{H}_{m_p(G)-1}(\mathcal{A}_p(G)) \neq 0. \qquad (\mathcal{QD})_p$$

A remarkable study of the Quillen dimension property for odd primes p was carried out in Theorem 3.1 of [Aschbacher and Smith 1993]. This theorem provides a list with the p-extensions of simple groups that might fail $(\mathcal{QD})_p$, for $p \geq 3$. In particular, this list contains the p-extensions of unitary groups $\mathrm{PSU}_n(q)$ with q odd and $p \mid q+1$. However, Conjecture 4.1 of [Aschbacher and Smith 1993] basically claims that these groups should not belong to this list. In fact, it is shown there that if $n < q(q-1)$ then these p-extensions satisfy $(\mathcal{QD})_p$. Nevertheless, this problem remains open.

The aim of this article is to achieve some progress on a similar list for the prime $p = 2$. Therefore, we will focus on showing that 2-extensions of certain simple groups satisfy $(\mathcal{QD})_2$. To that end, we introduce the following convenient definition.

Definition 2.3. A group L satisfies (E-$(\mathcal{QD})$) if every 2-extension of L satisfies $(\mathcal{QD})_2$:

$$\text{For every } B \in \hat{\mathcal{O}}_2(L),\ LB \text{ satisfies } (\mathcal{QD})_2. \tag{E-$(\mathcal{QD})$}$$

In order to establish $(\mathcal{QD})_p$ for p-extensions, it is crucial to be able to compute p-ranks of extensions. The following result, extracted from Lemma 4.2 in [Piterman et al. 2021], will be a useful tool to compute p-ranks of extensions.

Lemma 2.4 (p-rank of extensions). *Let $G = N.K$ be an extension of finite groups, and let p be a prime number. Then*

$$m_p(G) = \max_{A \in \mathcal{S}} \bigl(m_p(C_N(A)) + m_p(A)\bigr),$$

where $\mathcal{S} = \{A \in \mathcal{A}_p(G) \cup \{1\} : A \cap N = 1\}$. In particular, $m_p(G) \leq m_p(N) + m_p(K)$, and if K has order prime to p then $\mathcal{A}_p(G) = \mathcal{A}_p(N)$ and $m_p(G) = m_p(N)$.

We will implicitly use this result at many points of the proofs. Note that, in order to apply this lemma, we should be able to compute centralisers of elementary abelian 2-subgroups, usually inducing outer automorphisms. We will often proceed as follows: if LB is a 2-extension of L, then take a suitable decomposition $B = B_0 \oplus B_1$, with $|B_1| = 2$. Suppose that we can inductively compute the 2-rank of LB_0. Then, by Lemma 2.4, we have

$$m_2(LB) = \max\bigl\{m_2(LB_0),\ 1 + m_2(C_{LB_0}(t)) : t \in LB \setminus LB_0 \text{ is an involution}\bigr\}. \tag{2-1}$$

Moreover, this computation depends only on the conjugacy classes of the involutions t, and, in most of the cases that we are interested in, such classes are completely classified.

Now we recall, rather informally, the names of the different *types* of automorphisms of a simple group of Lie type K defined over a field of odd characteristic, following Definition 2.5.13 of [Gorenstein et al. 1998]. We refer to [Gorenstein et al. 1998] for the full details. Let $t \in \mathrm{Aut}(K)$ be an involution and $K^* = \mathrm{Inndiag}(K)$. Then we have the following names for t:

(1) inner-diagonal if $t \in K^*$;

(2) inner if $t \in \mathrm{Inn}(K)$;

(3) diagonal if $t \in K^* \setminus \mathrm{Inn}(K)$;

(4) field automorphism if $t \in \mathrm{Aut}(K) \setminus K^*$ is $\mathrm{Aut}(K)$-conjugate to a field automorphism of the ground field and K is not ${}^2A_n(q)$, ${}^2D_n(q)$ or ${}^2E_6(q)$;

(5) graph if $t \in \mathrm{Aut}(K) \setminus K^*$, roughly, is $\mathrm{Aut}(K)$-conjugate to an involution arising as an automorphism of the underlying Dynkin diagram (except for $K = G_2(q)$), or else from a field automorphism in cases ${}^2A_n(q)$, ${}^2D_n(q)$ and ${}^2E_6(q)$; and

(6) graph-field automorphism if it can be expressed as a product gf of a graph involution g and a field automorphism f, or else $K = G_2(q)$ and t arises from an $\mathrm{Aut}(K)$-conjugate of an involution automorphism of the underlying Coxeter diagram.

It follows from Proposition 4.9.1 of [Gorenstein et al. 1998] that the centralisers of field involutions t verify that $m_2(C_K(t)) = m_2(K)$ and $m_2(C_{K^*}(t)) = m_2(K^*)$. By (2-1), we see that $m_2(K\langle t\rangle) = m_2(K)+1$. Below we reproduce a simplified version of this proposition.

Proposition 2.5. *Let $K = {}^d\Sigma(q)$ be a group of Lie type in adjoint version in characteristic r, and let x be a field or graph-field automorphism of prime order p. Set $K_x = O^{r'}(C_K(x))$. Then the following hold:*

(1) *If x is a field automorphism then $K_x \cong {}^d\Sigma(q^{1/p})$.*

(2) *If x is a graph-field automorphism then $d = 1$, $p = 2$ or 3, and $K_x \cong {}^p\Sigma(q^{1/p})$.*

(3) *K_x is adjoint and $C_{\mathrm{Inndiag}(K)}(x) \cong \mathrm{Inndiag}(K_x)$.*

(4) *Field (resp. graph-field) automorphisms are all $\mathrm{Inndiag}(K)$-conjugate, except for graph-fields for $K = D_4(q)$ and $p = 3$.*

The previous proposition does not determine, a priori, the structure of $C_K(x)$, but just of the centraliser taken over the inner-diagonal automorphism group. Since we are interested in computing $m_2(C_K(x))$, it will be crucial for us to decide when a diagonal involution can centralise a field or graph-field automorphism x. We recall below Lemma 12.8 of [Gorenstein et al. 2018, Chapter 17], which provides a partial solution to this problem.

Lemma 2.6. *Let $K \cong \mathrm{PSL}_2(q)$, $\mathrm{P\Omega}_{2n+1}(q)$, $\mathrm{PSp}_{2n}(q)$ or $E_7(q)$, where q is a power of an odd prime r. Let ϕ be a field automorphism of order 2, and let $K_\phi = O^{r'}(C_K(\phi))$. Then*

$$\mathrm{Inndiag}(K_\phi) = C_{\mathrm{Inndiag}(K)}(\phi) = C_{\mathrm{Inn}(K)}(\phi).$$

In particular, ϕ does not commute with diagonal involutions of $\mathrm{Inndiag}(K)$.

We will mainly work with Table 4.5.1 of [Gorenstein et al. 1998] to compute the 2-ranks of extensions by diagonal and graph involutions, mostly for the groups of type $A_m^{\pm}(q)$ and the exceptional groups. In the next paragraph, we briefly and informally describe how to read such a table. See [Gorenstein et al. 1998, pp. 171–182] for a complete and accurate description of Table 4.5.1.

This table records the K^*-conjugacy classes of inner-diagonal and graph involutions t of a finite group of Lie type K in adjoint version, and the structure of their centralisers when taken over $K^* = \mathrm{Inndiag}(K)$.

The centraliser of an involution t is denoted by $C^* = C_{K^*}(t)$. The first column of Table 4.5.1 denotes the family for which the involutions are listed (A_n, B_n, C_n, etc.) The second column indicates the restrictions for these classes to exist, while the third column is a label for the conjugacy class of that involution. For the purposes of this article, we will not need to interpret the fourth column. In the fifth column, it is indicated when such classes are of inner type (denoted by 1), diagonal type (denoted by d) or graph type (several notation like g, g'). The notation $1/d$ indicates that it is inner if the condition inside the parentheses at the right holds, and it is diagonal otherwise. From the sixth column to the end, the structure of the centraliser C^* is described. Roughly, C^* is an extension of a central product of groups of Lie type $L^* = O^{r'}(C^*)$ (column six), whose versions are specified in the column "version" and whose centres can be recovered from the column $Z(L^*)$. An extra part centralising this product can be computed from the column $C_{C^{*\circ}}(L^*)$. Here $C^{*\circ} = L^*T^*$ is the connected-centraliser, where T^* is a certain r'-subgroup arising from a torus T normalised by t and inducing inner-diagonal automorphisms on L^*. From the columns L^*, version, $Z(L^*)$ and $C_{C^{*\circ}}(L^*)$, one can compute the "inner-part" of $C^{*\circ}$. Finally, from the last two columns we can recover the outer automorphisms of L^* arising in $C^{*\circ}$ (in general of diagonal type) and the remaining part of $C^*/C^{*\circ}$, which is often an involution acting on the components of L^* (as field or graph automorphism, or by switching two components) and on the central part $C_{C^*}(L^*)$ (which is usually cyclic and the involution acts by inversion). To recover the action of the last column, the symbols i, $\leftrightarrow$, ϕ, γ, 1 mean, respectively, an action by inversion, a swap of two components, a field automorphism of order 2, a graph automorphism of order 2, and an inner action on a component or trivial action on $C_{C^{*\circ}}(L^*)$.

3. Tools to achieve $(\mathcal{QD})_p$

In this section, we provide tools and collect results that will help us to establish $(\mathcal{QD})_2$ on certain 2-extensions. Many of these tools were introduced and exploited by Aschbacher–Smith to determine the $(\mathcal{QD})$-list in [Aschbacher and Smith 1993].

The following proposition is an easy consequence of the Künneth formula for the join of spaces and the fact that $\mathcal{A}_p(H \times K) \simeq \mathcal{A}_p(H) * \mathcal{A}_p(K)$ (see [Quillen 1978, Proposition 2.6]).

Proposition 3.1. *If p is a prime and H, K satisfy $(\mathcal{QD})_p$, then $H \times K$ satisfies $(\mathcal{QD})_p$.*

The following lemma corresponds to Lemmas 0.11 and 0.12 of [Aschbacher and Smith 1993].

Lemma 3.2. *Let $N \trianglelefteq G$ be such that $N \leq O_{p'}(G)$. Then there is an inclusion*

$$\widetilde{H}_*(\mathcal{A}_p(G/N)) \subseteq \widetilde{H}_*(\mathcal{A}_p(G)).$$

In particular, $m_2(G) = m_2(G/N)$, and if G/N satisfies $(\mathcal{QD})_p$ then so does G.

If $N \leq Z(G)$, then the quotient map induces a poset isomorphism $\mathcal{A}_p(G) \cong \mathcal{A}_p(G/N)$.

The following observation is an easy consequence of the inclusion between the homology groups of top-degree.

Lemma 3.3. *Let $H \leq G$ be such that $m_p(H) = m_p(G)$. If H satisfies $(\mathcal{QD})_p$, then so does G.*

Next, we recall one of the essential results on the Quillen dimension property.

Theorem 3.4 (Quillen). *If G is a solvable group with $O_p(G) = 1$, then G satisfies $(\mathcal{QD})_p$.*

This theorem settles the solvable case of Quillen's conjecture (see [Quillen 1978, Theorem 12.1]). Later, it was extended to the family of p-solvable groups by using the CFSG if p is odd. We refer to Chapter 8 of [Smith 2011] for further details on Quillen's conjecture and the Quillen dimension property.

In view of Theorem 3.4 and the inclusion lemma (Lemma 3.3), it is convenient to look for solvable subgroups of G with maximal p-rank. Some standard solvable subgroups in a group of Lie type L arise by taking extensions of unipotent radicals by elementary abelian subgroups of their normalisers. These extensions lie then inside parabolic subgroups. The following result on parabolic subgroups will help us to achieve (E-$(\mathcal{QD})$) for arbitrary groups of Lie type (see [Aschbacher and Smith 1993, Step v on p. 506]).

Lemma 3.5. *Let L be a simple group of Lie type, and p a prime not dividing the characteristic of L. Suppose that LB is a p-extension of L and that there exists a B-invariant proper parabolic subgroup $P \leq L$ such that $m_p(LB) = m_p(PB)$. Then LB satisfies $(\mathcal{QD})_p$.*

Proof. Let $R := O_r(P)$, where r is the characteristic of the ground field. Then, as a consequence of the Borel–Tits theorem, $C_{\mathrm{Aut}(L)}(R) \leq R$ (see Corollary 3.1.4 of [Gorenstein et al. 1998]). In particular, if $T \leq PB$ realises the p-rank of PB, then T normalises R, and $C_T(R) \leq R \cap T = 1$. This means that T is faithful on R, i.e., $O_p(RT) = 1$, and $m_p(RT) = m_p(PB) = m_p(LB)$. Then RT is a solvable group with trivial p-core, and by Theorem 3.4, RT satisfies $(\mathcal{QD})_p$. By Lemma 3.3, LB satisfies $(\mathcal{QD})_p$. □

Lemma 3.6. *Let L be a simple group of Lie type defined in odd characteristic. Suppose that P is a proper parabolic subgroup of L containing a Sylow 2-subgroup of L (that is, $|L : P|$ is odd). Then L and the extension of L by a field automorphism of order 2 satisfy $(\mathcal{QD})_2$.*

Proof. Let L and P be as in the hypotheses of the lemma. Since P has odd index in L, it contains a Sylow 2-subgroup of L. Therefore, $m_2(P) = m_2(L)$ and by Lemma 3.5, L satisfies $(\mathcal{QD})_2$.

Next, let $B \in \hat{\mathcal{O}}_2(L)$ be cyclic inducing field automorphisms. By passing through algebraic groups and root systems, it can be shown that B normalises some conjugate of P, which we may assume is P itself. Thus, after conjugation, we suppose that $B \leq N_{\mathrm{Aut}(L)}(P)$. Note that a Sylow 2-subgroup of PB is a Sylow 2-subgroup of LB, so $m_2(PB) = m_2(LB)$. By Lemma 3.5, LB satisfies $(\mathcal{QD})_2$. □

We close this section with a few more results on low p-ranks. The following lemma follows from the p-rank 2 case of Quillen's conjecture. See [Quillen 1978, Proposition 2.10].

Lemma 3.7. *If $\mathcal{A}_p(G)$ is connected, $m_p(G) = 2$ and $O_p(G) = 1$, then G satisfies $(\mathcal{QD})_p$.*

It will be convenient to recall the classification of groups with a strongly 2-embedded subgroup, that is, those groups with disconnected 2-subgroup poset. See [Gorenstein et al. 1998, Theorem 7.6.1] and [Quillen 1978, Sec. 5].

Theorem 3.8. *Let $p=2$ and G be a finite group. Then $\mathcal{A}_2(G)$ is disconnected if and only if $O_2(G)=1$ and one of the following holds*:

(1) $m_2(G)=1$;

(2) $\Omega_1(G)/O_{p'}(\Omega_1(G)) \cong \mathrm{PSL}_2(2^n)$, $\mathrm{PSU}_3(2^n)$ *or* $\mathrm{Sz}(2^{2n-1})$ *for some* $n \geq 2$.

In particular, from the isomorphisms among the simple groups, we see that

$$\mathrm{Alt}_5 \cong \mathrm{PSL}_2(5) \cong \mathrm{PSL}_2(2^2), \quad {}^2G_2(3)' \cong \mathrm{PSL}_2(2^3),$$

are included in the list of item (2).

Indeed, sometimes in low dimensions, we will be able to conclude $(\mathcal{QD})_p$ by computing the sign of the Euler characteristic of $\mathcal{A}_p(G)$. Therefore, we will use the following well-known expression of this invariant. We write $\bar{E}$ for the conjugacy class of a subgroup E of G.

Proposition 3.9. *The reduced Euler characteristic of $\mathcal{A}_p(G)$ is*

$$\widetilde{\chi}(\mathcal{A}_p(G)) = \sum_{\bar{E} \in \mathcal{A}_p(G)/G \cup \{1\}} (-1)^{m_p(E)-1} p^{\binom{m_p(E)}{2}} |G : N_G(E)|.$$

Proof. This follows from the results of [Jacobsen and Møller 2012], as we briefly explain below. By [Jacobsen and Møller 2012, Example 2.10], we have

$$\widetilde{\chi}(\mathcal{A}_p(G)) = -\sum_{E \in \mathcal{A}_p(G) \cup \{1\}} \widetilde{\chi}(\mathcal{A}_p(E) \setminus \{E\}).$$

Since $\mathcal{A}_p(E) \setminus \{E\}$ is the poset of proper nonzero subspaces of the vector space E of dimension $m_p(E)$ over the finite field of p elements, we see that

$$\widetilde{\chi}(\mathcal{A}_p(E) \setminus \{E\}) = (-1)^{m_p(E)-2} p^{\binom{m_p(E)}{2}}.$$

Grouping by conjugacy classes yields the formula given in the statement of the proposition. □

Finally, the next lemma will help us to produce nonzero homology by inductively looking into the homology of the Quillen poset of a certain normal subgroup and centralisers of outer elements acting on it. The main reference for this lemma is [Segev and Webb 1994].

Lemma 3.10. *Let G be a finite group and p a prime number. Suppose that $L \trianglelefteq G$ is a normal subgroup such that $\mathcal{O}_G(L)$ consists only of cyclic subgroups. Then we have a long exact sequence*

$$\cdots \to \widetilde{H}_{m+1}(\mathcal{A}_p(G)) \to \bigoplus_{B \in \mathcal{O}_G(L)} \widetilde{H}_m(\mathcal{A}_p(C_L(B))) \xrightarrow{i_*} \widetilde{H}_m(\mathcal{A}_p(L)) \xrightarrow{j_*} \widetilde{H}_m(\mathcal{A}_p(G)) \to \cdots$$

where i_* and j_* are the natural maps induced by the inclusions $\mathcal{A}_p(C_L(B)) \subseteq \mathcal{A}_p(L)$ and $\mathcal{A}_p(L) \subseteq \mathcal{A}_p(G)$, respectively.

In particular, the following hold:

(1) *Let X be the union of the subposets $\mathcal{A}_p(C_L(B))$ for $B \in \mathcal{O}_G(L)$. We have indeed a factorisation*

$$\begin{tikzcd} \bigoplus_{B\in\mathcal{O}_G(L)} \widetilde{H}_m\big(\mathcal{A}_p(C_L(B))\big) \arrow[rr, "i_*"] \arrow[dr, "i'_*"'] & & \widetilde{H}_m(\mathcal{A}_p(L)) \\ & \widetilde{H}_m(X) \arrow[ur, "k_*"'] & \end{tikzcd} \tag{3-1}$$

where also i'_* and k_* are induced by the inclusions $\mathcal{A}_p(C_L(B)) \subseteq X$ and $X \subseteq \mathcal{A}_p(L)$, respectively.

(2) $m_p(G) \leq m_p(L) + 1$.

(3) *If $\widetilde{H}_{m_p(L)-1}(\mathcal{A}_p(C_L(B))) = 0$ for all $B \in \mathcal{O}_G(L)$, then $H_{m_p(L)}(\mathcal{A}_p(G)) = 0$.*

(4) *We have a bound*

$$\begin{aligned} \dim H_{m_p(L)}(\mathcal{A}_p(G)) &\geq \sum_{B\in\mathcal{O}_G(L)} \dim \widetilde{H}_{m_p(L)-1}(\mathcal{A}_p(C_L(B))) - \dim \widetilde{H}_{m_p(L)-1}(X) \\ &\geq \sum_{B\in\mathcal{O}_G(L)} \dim \widetilde{H}_{m_p(L)-1}(\mathcal{A}_p(C_L(B))) - \dim \widetilde{H}_{m_p(L)-1}(\mathcal{A}_p(L)). \end{aligned}$$

(5) *If $m_p(G) = m_p(L) + 1$ and G fails $(\mathcal{QD})_p$, then, for $m = m_p(L) - 1$, we get inclusions*

$$\bigoplus_{B\in\mathcal{O}_G(L)} \widetilde{H}_m\big(\mathcal{A}_p(C_L(B))\big) \hookrightarrow \widetilde{H}_m(X) \hookrightarrow \widetilde{H}_m(\mathcal{A}_p(L)).$$

Proof. The long exact sequence arises from the main result of [Segev and Webb 1994]. Then equation (3-1) in item (1) is an immediate consequence of this sequence. Item (2) holds by Lemma 2.4. Items (3)–(5) follow by looking into the last terms of the long exact sequence, at $m = m_p(L)$. □

4. Some linear groups satisfy $(\mathcal{QD})_2$

In this section, we prove that the linear groups $\mathrm{PSL}_2(q)$ and $\mathrm{PSL}_3(q)$ satisfy (E-$(\mathcal{QD})$) for every q, with a few exceptions for $q = 3, 5, 9$. These cases will serve as basic cases for the exceptional groups, where we will occasionally find linear groups as direct factors in some of their maximal subgroups.

From [Gorenstein et al. 1998, Proposition 4.10.5], we recall the 2-ranks of the small dimensional linear groups:

Proposition 4.1. *If q is a power of an odd prime and $n = 2, 3$, then $\mathrm{PSL}_n^{\pm}(q)$ and $\mathrm{PGL}_n^{\pm}(q)$ have 2-rank 2.*

We begin by studying the linear group of dimension 2.

Proposition 4.2. *Let $L \cong \mathrm{PSL}_2(q)$ with q odd and $q \neq 3$. Then every 2-extension LB of L satisfies $(\mathcal{QD})_2$, with the following exceptions*:

(1) $L \cong \mathrm{PSL}_2(5)$, $B = 1$;

(2) $L \cong \mathrm{PSL}_2(9)$, *$B$ induces field automorphisms of order* 2.

Moreover, every 2-extension of $\mathrm{Inndiag}(L) \cong \mathrm{PGL}_2(q)$ *satisfies* $(\mathcal{QD})_2$, *except in case* (2).

2-extension LB	$C_L(B)$	$m_2(LB)$
$B=1$	L	2
$B=\langle\phi\rangle$	$\mathrm{PGL}_2(q^{1/2})$	3
$B=\langle d\rangle$	$\mathrm{D}_{q+\epsilon}$	2

Table 1. 2-extensions of $\mathrm{PSL}_2(q)$, $q\geq 5$ odd. Here $q\equiv\epsilon \pmod 4$, $\epsilon\in\{1,-1\}$.

Proof. We consider the possible 2-extensions of L. In any case, we know that L is simple and that $\mathrm{Out}(L)=\mathrm{C}_2\times\mathrm{C}_a$, where $\mathrm{C}_2\cong\mathrm{Outdiag}(L)$ and C_a is the group of field automorphisms of $\mathbb{F}_q$. Suppose that ϕ is an order 2-field automorphisms of $\mathbb{F}_q$ (if it exists), and that $d\in\mathrm{Inndiag}(L)\setminus L$ is a diagonal involution. Then the 2-extensions of L are given in Table 1.

This table follows since every involution of $\mathrm{Aut}(\mathrm{PSL}_2(q))\setminus\mathrm{PGL}_2(q)$ is a field automorphism. Recall also that field and diagonal automorphisms of order 2 do not commute by Lemma 2.6. The structure of the centraliser for d follows from the first row of Table 4.5.1 of [Gorenstein et al. 1998]. Finally, observe that $L\langle d\rangle=\mathrm{Inndiag}(L)$ and $m_2(\mathrm{Inndiag}(L)\langle\phi\rangle)=3$ since $m_2(L)=m_2(\mathrm{Inndiag}(L))=2$ by Proposition 4.1.

By computing the Euler characteristic, we prove that each 2-extension of L satisfies $(\mathbb{Q}\mathcal{D})_2$. First, 2-extensions LB and $\mathrm{Inndiag}(L)\langle\phi\rangle$ have connected $\mathcal{A}_2$-poset by Theorem 3.8, except for $L=\mathrm{PSL}_2(5)$, $B=1$. Therefore, by Lemma 3.7, L and $\mathrm{Inndiag}(L)$ satisfy $(\mathbb{Q}\mathcal{D})_2$, except for $L=\mathrm{PSL}_2(5)$. Note that $\mathcal{A}_2(\mathrm{PSL}_2(5))=\mathcal{A}_2(\mathrm{Alt}_5)=\mathcal{A}_2(\mathrm{PSL}_2(4))$ is homotopically discrete with 5 points, and the 2-extension $\mathrm{PGL}_2(5)\cong\mathrm{Sym}_5$ does satisfy $(\mathbb{Q}\mathcal{D})_2$. This yields the conclusions of the statement for the case $q=5$.

Next we show $(\mathbb{Q}\mathcal{D})_2$ for the 2-extensions $L\langle\phi\rangle$ and $\mathrm{Inndiag}(L)\langle\phi\rangle$. Since both have 2-rank 3, by Lemma 3.3 it is enough to show that $L\langle\phi\rangle$ satisfies $(\mathbb{Q}\mathcal{D})_2$. In order to do this, we compute the dimensions of $H_1(\mathcal{A}_2(L))$ and $H_1\big(\mathcal{A}_2(C_L(\phi))\big)$.

Since in this situation, q is a square, $q\neq 5$. Second, if $q=25$, $C_L(\phi)=\mathrm{PGL}_2(5)$. Hence, in any case, the dimension of these degree 1 homology groups can be computed from the reduced Euler characteristic of the underlying $\mathcal{A}_2$-poset. Here we use the formula given in Proposition 3.9. Thus, for $K=L$ or $C_L(\phi)$,

$$\dim H_1(\mathcal{A}_2(K))=-\widetilde{\chi}(\mathcal{A}_2(K))=1-\#\text{ of involutions in }K+2\cdot\#\text{ of 4-subgroups of }K. \tag{4-1}$$

In Table 2 we describe these numbers.

Proof of Table 2. The number of involutions and 4-subgroups of $\mathrm{PSL}_2(q)$ follows from Dickson's classification of the subgroups of $\mathrm{PSL}_2(q)$ (see also Theorem 6.5.1 of [Gorenstein et al. 1998]).

group	number of involutions	number of 4-subgroups
$\mathrm{PSL}_2(q)$	$\frac{1}{2}q(q+\epsilon)$	$\frac{1}{24}q(q^2-1)$
$\mathrm{PGL}_2(q)$	q^2	$\frac{1}{6}q(q^2-1)$

Table 2. Here $q\equiv\epsilon \pmod 4$, $\epsilon\in\{1,-1\}$.

The number of involutions of $\mathrm{PGL}_2(q)$ follows since there is a unique conjugacy class of diagonal involutions d by Table 4.5.1 of [Gorenstein et al. 1998]. Thus, the number of elements in such conjugacy class is equal to $\frac{1}{2}q(q-\epsilon)$, which gives q^2 after adding the number of involutions in $\mathrm{PSL}_2(q)$.

Finally, to compute the number of four-subgroups of $\mathrm{PGL}_2(q)$ we proceed as follows: each four-subgroup of $\mathrm{PGL}_2(q)$ is either contained in $\mathrm{PSL}_2(q)$ or else it contains a unique involution of $\mathrm{PSL}_2(q)$ and 2 diagonal involutions. Therefore, for a given diagonal involution d, there is a one-to-one correspondence between 4-subgroups containing d and involutions in $C_L(d) \cong \mathrm{D}_{q+\epsilon}$. This shows that each diagonal involution is contained in $(q+\epsilon)/2$ 4-subgroups. Since we have $\frac{1}{2}q(q-\epsilon)$ diagonal involutions, the total number of 4-subgroups in $\mathrm{PGL}_2(q)$ containing diagonal involutions is

$$\frac{q(q-\epsilon)}{2}\cdot\frac{(q+\epsilon)}{2}\cdot\frac{1}{2}=\frac{q(q^2-1)}{8}.$$

Thus the total number of 4-subgroups in $\mathrm{PGL}_2(q)$ is

$$\frac{q(q^2-1)}{24}+\frac{q(q^2-1)}{8}=\frac{q(q^2-1)}{6}.$$

This completes the proof of Table 2. □

Indeed, by Table 2, we get concrete values for the dimensions of the degree 1 homology groups of $\mathcal{A}_2(\mathrm{PSL}_2(q))$ and $\mathcal{A}_2(\mathrm{PGL}_2(q))$:

$$\dim H_1\big(\mathcal{A}_2(\mathrm{PSL}_2(q))\big)=-\widetilde{\chi}\big(\mathcal{A}_2(\mathrm{PSL}_2(q))\big)=\tfrac{1}{12}(q-\epsilon)(q^2-(6-\epsilon)q-\epsilon 12), \tag{4-2}$$

$$\dim H_1\big(\mathcal{A}_2(\mathrm{PGL}_2(q))\big)=-\widetilde{\chi}\big(\mathcal{A}_2(\mathrm{PGL}_2(q))\big)=\tfrac{1}{3}(q-3)(q^2-1). \tag{4-3}$$

Now we need to describe the number of field automorphisms in $\mathrm{PSL}_2(q)\langle\phi\rangle$ and in $\mathrm{PGL}_2(q)\langle\phi\rangle$.

Recall that the field automorphisms of $\mathrm{PSL}_2(q)\langle\phi\rangle$ are all $\mathrm{PGL}_2(q)$-conjugate, with centraliser $C_{\mathrm{PGL}_2(q)}(\phi)=C_{\mathrm{PSL}_2(q)}(\phi)$. Thus, the number of field automorphisms of order 2 in $\mathrm{PSL}_2(q)\langle\phi\rangle$ is exactly

$$\frac{|\,\mathrm{PGL}_2(q)|}{|C_{\mathrm{PSL}_2(q)}(q)|}=\frac{q(q^2-1)}{q^{1/2}(q-1)}=q^{1/2}(q+1).$$

This gives $q^{1/2}(q+1)$ involutions in $\mathrm{PSL}_2(q)\langle\phi\rangle\setminus\mathrm{PSL}_2(q)$. Let $L=\mathrm{PSL}_2(q)$, $B=\langle\phi\rangle$. By Lemma 3.10, the values in Table 2 and formula (4-1), we conclude that

$$\begin{aligned}\dim H_2(\mathcal{A}_2(LB)) &\geq q^{1/2}(q+1)\dim H_1\big(\mathcal{A}_2(\mathrm{PGL}_2(q^{1/2}))\big)-\dim H_1\big(\mathcal{A}_2(\mathrm{PSL}_2(q))\big)\\ &=q^{1/2}(q+1)\tfrac{1}{3}(q^{1/2}-3)(q-1)-\tfrac{1}{12}(q-1)(q^2-5q-12)\\ &=\tfrac{1}{4}(q^{1/2}-1)(q-1)(q^{3/2}-3q-4).\end{aligned}$$

Note that $q\equiv 1 \pmod 4$. The above number is positive for all $q\geq 13$, which is our case since q is an even power of an odd prime and $q\neq 9$ by hypothesis. We conclude that $LB=\mathrm{PSL}_2(q)\langle\phi\rangle$ satisfies

$(\mathcal{QD})_2$. Then also $\mathrm{PGL}_2(q)\langle\phi\rangle$ satisfies $(\mathcal{QD})_2$. Moreover,

$$\begin{aligned}\dim H_2\big(\mathcal{A}_2(\mathrm{PGL}_2(q)\langle\phi\rangle)\big) &\geq \dim H_2\big(\mathcal{A}_2(\mathrm{PSL}_2(q)\langle\phi\rangle)\big)\\ &\geq \tfrac{1}{4}(q^{1/2}-1)(q-1)(q^{3/2}-3q-4).\end{aligned} \tag{4-4}$$

We have shown that every possible 2-extension of $\mathrm{PSL}_2(q)$ and $\mathrm{PGL}_2(q)$ satisfies $(\mathcal{QD})_2$, except for the cases described in the statement of the theorem. □

We note that the excluded cases in Proposition 4.2 actually fail $(\mathcal{QD})_2$. Indeed, $\mathrm{PSL}_2(5)$ fails $(\mathcal{QD})_2$ since it has 2-rank 2 and $\mathcal{A}_2(\mathrm{PSL}_2(5)) = \mathcal{A}_2(\mathrm{PSL}_2(4))$ is homotopically discrete. The following example provides the details that show that $\mathrm{PSL}_2(9)\langle\phi\rangle$ and $\mathrm{PGL}_2(9)\langle\phi\rangle$ fail $(\mathcal{QD})_2$, where ϕ is a field automorphism of order 2.

Example 4.3. Let $L = \mathrm{PSL}_2(9)$ and let $A = \mathrm{Aut}(L)$. Then $A/L \cong \mathrm{C}_2 \times \mathrm{C}_2$, so every 2-extension of L is a nontrivial normal subgroup of A. This gives 3 possible 2-extensions of L, but not 4. Let ϕ be a field automorphism of L and d a diagonal automorphism of L, both of order 2. Then the possible 2-extensions of L are

(1) L, with 2-rank 2, satisfies $(\mathcal{QD})_2$ with $H_1(\mathcal{A}_2(L))$ of rank 16;

(2) $L\langle\phi\rangle$, with 2-rank 3, fails $(\mathcal{QD})_2$ since $C_L(\phi) \cong \mathrm{Sym}_4$, which has nontrivial 2-core

$$O_2(C_L(\phi)) \cong \mathrm{C}_2 \times \mathrm{C}_2 \neq 1;$$

(3) $L\langle d\rangle = \mathrm{PGL}_2(9)$, with 2-rank 2, satisfies $(\mathcal{QD})_2$ with $H_1(\mathcal{A}_2(L)\langle d\rangle)$ of rank 160 and $C_L(d) \cong \mathrm{D}_{10}$.

Note that $\mathrm{Aut}(L)$ has 2-rank 3 and does not satisfy $(\mathcal{QD})_2$, and it is not a 2-extension of L since diagonal and field automorphisms do not commute in $\mathrm{Aut}(L)$. Also $\mathrm{PGL}_2(9)\langle\phi\rangle$ fails $(\mathcal{QD})_2$ since $C_{\mathrm{PGL}_2(9)}(\phi) = C_L(\phi)$ has nontrivial 2-core.

There is also a remaining almost simple group N with $L < N < \mathrm{Aut}(L)$, not contained in the previous cases. This is the extension $N = \mathrm{PSL}_2(9).2 \cong \mathrm{Alt}_6\,.2$, and it satisfies that $\mathcal{A}_2(N) = \mathcal{A}_2(L)$. Therefore, although this group N is not a 2-extension of L, it is a "nonsplit 2-extension", and it does satisfy $(\mathcal{QD})_2$.

Finally, these computations show that $\mathcal{A}_2(L) \hookrightarrow \mathcal{A}_2(\mathrm{Aut}(L))$ induces an inclusion in homology, and hence a nonzero map. By the main result of [Piterman and Smith 2022a], $\mathrm{PSL}_2(9)$ is not a component of a minimal counterexample to Quillen's conjecture.

Our next aim is to show that 2-extensions of $\mathrm{PSL}_3(q)$ satisfy $(\mathcal{QD})_2$, with only a few exceptions. We will need the following lemma which records the values of the Euler characteristic of the Quillen poset of some linear groups and the unitary groups in dimension 3.

Lemma 4.4. *For $L = \mathrm{PSL}_n(q)$ and n odd, we have*

$$\widetilde{\chi}(\mathcal{A}_2(L)) = \widetilde{\chi}\big(\mathcal{A}_2(\mathrm{PGL}_n(q))\big) = \frac{(-1)^n}{n}\prod_{i=1}^{n-1}(q^i-1)\,f_n(q),$$

where $f_n(q)$ denotes a polynomial as described in [Welker 1995]. For instance, $f_3(q)=q^3+3q^2+3q+3$. Moreover, since $\mathcal{A}_2(L)$ is Cohen–Macaulay of dimension $n-2$, the above Euler characteristic computes the dimension of $H_{n-2}(\mathcal{A}_2(L))$.

If $L=\mathrm{PSU}_3(q)$, then

$$\widetilde{\chi}(\mathcal{A}_2(L))=\widetilde{\chi}\big(\mathcal{A}_2(\mathrm{PGU}_3(q))\big)=-\tfrac{1}{3}(q^6-2q^5-q^4+2q^3-3q^2+3).$$

Proof. The value of the Euler characteristic for $\mathrm{PGL}_n(q)$ follows from Proposition 4.1 and Theorem 4.4 of [Welker 1995] (note that there is a typo in the formula of Theorem 4.4, and the product over i should be up to $r-1$). Also, since n is odd, by Proposition 7.5 of [Piterman and Welker 2022], $\mathcal{A}_2(\mathrm{PSL}_n(q))=\mathcal{A}_2(\mathrm{PGL}_n(q))=\mathcal{A}_2(\mathrm{GL}_n(q))_{>Z}$ where Z is the cyclic subgroup of order 2 of $Z(\mathrm{GL}_n(q))$. By [Quillen 1978] (see also [Welker 1995]), $\mathcal{A}_2(\mathrm{PSL}_n(q))$ is Cohen–Macaulay of dimension $n-2$.

The formula for $\mathrm{PGU}_3(q)$ follows from Example 7.6 of [Piterman and Welker 2022]. □

Next, we show that the 2-extensions of $\mathrm{PSU}_3(q)$ satisfy $(\mathbb{QD})_2$, except for $q=3$. These cases will be important during our analysis for $\mathrm{PSL}_3(q)$, especially when working with 2-extensions by graph-field automorphisms.

Proposition 4.5. *Let $L=\mathrm{PSU}_3(q)$ with q odd. Then L satisfies (E-$(\mathbb{QD})$) if $q\neq 3$. Moreover, let ϕ be a graph automorphism of order 2 of L. Then we have*

$$\begin{aligned}\dim H_2\big(\mathcal{A}_2(\mathrm{PGU}_3(q)\langle\phi\rangle)\big)&\geq \dim H_2\big(\mathcal{A}_2(\mathrm{PSU}_3(q)\langle\phi\rangle)\big)\\ &\geq \tfrac{1}{3}(q^2-1)(q+1)\left(\frac{q^2(q^2-q+1)}{(3,q+1)}(q-3)-(q^3-3q^2+3q-3)\right),\end{aligned}$$

which is a positive polynomial for $q>3$. Finally, for $q=3$, $\mathrm{PSU}_3(3)$ satisfies $(\mathbb{QD})_2$ but $\mathrm{PSU}_3(3)\langle\phi\rangle$ fails $(\mathbb{QD})_2$.

Proof. We have that $\mathcal{A}_2(L)$ is connected by Theorem 3.8, and $m_2(L)=2$ by Proposition 4.1. Thus L satisfies $(\mathbb{QD})_2$ by Lemma 3.7. Moreover, by Lemma 4.4,

$$\dim H_1(\mathcal{A}_2(L))=-\widetilde{\chi}(\mathcal{A}_2(L))=\tfrac{1}{3}(q^6-2q^5-q^4+2q^3-3q^2+3). \tag{4-5}$$

Next, the only possible nontrivial 2-extension of L is by a graph automorphism ϕ of order 2 (which indeed arises from the field automorphism $x\mapsto x^q$). Let $L_1=L\langle\phi\rangle$ be such extension. By Table 4.5.1 of [Gorenstein et al. 1998],

$$C_{\mathrm{PGU}_3(q)}(\phi)\cong \mathrm{Inndiag}(\Omega_3(q))=\mathrm{PGL}_2(q).$$

This implies that $C_L(\phi)=\mathrm{PGL}_2(q)$. Moreover, there is a unique $\mathrm{PGU}_3(q)$-conjugacy class of graph automorphisms, and such elements act by inversion on $\mathrm{Outdiag}(L)=(3,q+1)$. Thus the conjugacy class of ϕ in $\mathrm{Out}(L)$ has size $(3,q+1)$, and this gives rise to exactly $(3,q+1)$ extensions $L\langle\phi'\rangle\leq\mathrm{Aut}(L)$ of L by a conjugate ϕ' of ϕ, and these extensions are $\mathrm{Aut}(L)$-conjugate. We conclude then that the number of graph automorphisms contained in L_1 is

$$n_g := \frac{|\operatorname{PGU}_3(q)|}{|\operatorname{PGL}_2(q)|(3,q+1)} = \frac{q^2(q^3+1)}{(3,q+1)}.$$

Finally, by Lemma 3.10, we conclude that

$$\begin{aligned}\dim H_2\big(\mathscr{A}_2(\operatorname{PGU}_3(q)\langle\phi\rangle)\big) &\geq \dim H_2\big(\mathscr{A}_2(\operatorname{PSU}_3(q)\langle\phi\rangle)\big)\\ &\geq n_g \dim H_1\big(\mathscr{A}_2(\operatorname{PGL}_2(q))\big) - \dim H_1\big(\mathscr{A}_2(\operatorname{PSU}_3(q))\big)\\ &= \frac{q^2(q^3+1)}{(3,q+1)}\tfrac{1}{3}(q-3)(q^2-1) - \tfrac{1}{3}(q^6-2q^5-q^4+2q^3-3q^2+3)\\ &= \tfrac{1}{3}(q^2-1)(q+1)\bigg(\frac{q^2(q^2-q+1)}{(3,q+1)}(q-3)-(q^3-3q^2+3q-3)\bigg).\end{aligned}$$

This polynomial is positive for all $q > 3$. Therefore, L_1 satisfies $(\mathbb{Q}\mathscr{D})_2$ if $q \neq 3$.

When $q = 3$, $C_L(\phi) = \operatorname{PGL}_2(3)$ has nontrivial 2-core, so $H_1\big(\mathscr{A}_2(C_L(\phi))\big) = 0$, and by Lemma 3.10(3), $H_2(\mathscr{A}_2(L_1)) = 0$. □

Now we have the necessary background to prove that $\operatorname{PSL}_3(q)$ satisfies (E-$(\mathbb{Q}\mathscr{D})$), except for a small number of cases.

Proposition 4.6. *Let $L = \operatorname{PSL}_n(q)$ with n, q odd. The following assertions hold*:

(1) *L, and L extended by a field involution, satisfy* $(\mathbb{Q}\mathscr{D})_2$.

(2) *If $n = 3$, then every 2-extension of L satisfies $(\mathbb{Q}\mathscr{D})_2$, with the following exceptions that fail* $(\mathbb{Q}\mathscr{D})_2$:
   - *$L = \operatorname{PSL}_3(3)$ extended by a graph automorphism, and*
   - *$L = \operatorname{PSL}_3(9)$ extended by a group generated by a field involution and a graph automorphism.*

Proof. Let $L = \operatorname{PSL}_n(q)$, with n odd, and consider the stabiliser P of a 1-dimensional subspace of the underlying module $V = \mathbb{F}_q^n$. Then P is a parabolic subgroup with structure $P \cong [q^{n-1}]L_P$, where L_P, a Levi complement for P, has structure $\operatorname{SL}_{n-1}(q) \circ_{(n,q-1)} \mathrm{C}_{q-1}$. Thus $|L_P| = |\operatorname{GL}_{n-1}(q)|/(n, q-1)$ and the index of P in L is

$$|L:P| = \frac{q^{n(n-1)/2}\prod_{i=2}^{n}(q^i-1)}{q^{n-1}\cdot q^{(n-1)(n-2)/2}\prod_{i=1}^{n-1}(q^i-1)} = \frac{q^n-1}{q-1} = q^{n-1}+q^{n-2}+\cdots+q+1.$$

Since n is odd, the index of P in $\operatorname{PSL}_n(q)$ is odd. By Lemma 3.6, $L = \operatorname{PSL}_n(q)$ and L extended by a field involution satisfy $(\mathbb{Q}\mathscr{D})_2$. This proves item (1).

Before moving to the case $n = 3$, we list all the possible 2-extensions of L. Denote by ϕ, γ and δ a field automorphism of order 2, a graph automorphism and a graph-field automorphism of L, respectively, such that $[\phi, \gamma] = 1$ and $\delta = \phi\gamma$. Let also $L^* = \operatorname{PGL}_n(q)$. Then the 2-extensions of L are

(i) L;

(ii) $L\langle\phi\rangle$, with $C_{L^*}(\phi) \cong \operatorname{PGL}_n(q^{1/2})$ by Proposition 2.5;

(iii) $L\langle\gamma\rangle$, with $C_L(\gamma) \cong \operatorname{Inndiag}(\Omega_n(q))$ by Table 4.5.1 of [Gorenstein et al. 1998];

(iv) $L\langle\delta\rangle$, with $C_{L^*}(\delta) \cong \mathrm{PGU}_n(q^{1/2})$ by Proposition 2.5;

(v) $L\langle\phi,\gamma\rangle$, with $C_L(\phi,\gamma) \cong \mathrm{Inndiag}(\Omega_n(q^{1/2}))$ by (iii) and Proposition 2.5.

Now suppose that $n = 3$, that is $L = \mathrm{PSL}_3(q)$. We know that the extensions of cases (i) and (ii) above satisfy $(\mathbb{QD})_2$ by the parabolic argument. So it remains to show that the 2-extensions by graph, graph-field and both graph and field automorphisms, satisfy $(\mathbb{QD})_2$. To that end, we compute the dimensions of the top-degree homology groups, similar to what we did for $\mathrm{PSL}_2(q)$ in the proof of Proposition 4.2.

First, recall that we have the following number of involutions of each type. Let $B = \langle\phi,\gamma\rangle$.

$$\begin{aligned}
n_f &:= \#\text{ field involutions in } L\langle\phi\rangle = \#\text{ field involutions in } LB\\
&= \frac{|\mathrm{PGL}_3(q)|}{|\mathrm{PGL}_3(q^{1/2})|(3, q^{1/2}+1)},\\
n_g &:= \#\text{ graph involutions in } L\langle\gamma\rangle = \#\text{ graph involutions in } LB\\
&= \frac{|\mathrm{PGL}_3(q)|}{|\mathrm{PGL}_2(q)|(3, q-1)},\\
n_{gf} &:= \#\text{ graph-field involutions in } L\langle\delta\rangle = \#\text{ graph-field involutions in } LB\\
&= \frac{|\mathrm{PGL}_3(q)|}{|\mathrm{PGU}_3(q^{1/2})|(3, q^{1/2}-1)}.
\end{aligned}$$

To compute these numbers, we have used the structure of the centraliser in each case, the fact that there is a unique L^*-conjugacy class for each type of involution, and the structure of $\mathrm{Out}(L) = (3, q-1) : \langle\phi,\gamma\rangle$ (see Theorem 2.5.12 of [Gorenstein et al. 1998]).

Let t be a field, graph or graph-field involution of L, and let $L_1 = L\langle t\rangle$. Then the number n_t of involutions in $L_1 \setminus L$ is n_f, n_g or n_{gf}, accordingly to the type of t. Note also that $m_2(L_1) = m_2(L) + 1 = 3$.

By Lemma 3.10,

$$\dim H_2(\mathcal{A}_2(L_1)) \ge n_t \cdot \dim H_1\big(\mathcal{A}_2(C_L(t))\big) - \dim H_1(\mathcal{A}_2(L)). \tag{4-6}$$

We compute $d(t) := \dim H_1\big(\mathcal{A}_2(C_L(t))\big)$ in each case, by using Lemma 4.4 and (4-3). Note that $\Omega_1(C_L(\phi)) = \mathrm{PSL}_3(q^{1/2})$ by item (ii) above. Also $C_L(\gamma) = \mathrm{PGL}_2(q)$ by the classical isomorphism $\mathrm{Inndiag}(\Omega_3(q)) \cong \mathrm{PGL}_2(q)$. By Lemma 4.4, we have

$$\begin{aligned}
d(\phi) &= \dim H_1\big(\mathcal{A}_2(\mathrm{PSL}_3(q^{1/2}))\big) = \tfrac{1}{3}(q^{1/2}-1)(q-1)(q^{3/2}+3q+3q^{1/2}+3),\\
d(\gamma) &= \dim H_1\big(\mathcal{A}_2(\mathrm{PGL}_2(q))\big) = \tfrac{1}{3}(q-3)(q^2-1),\\
d(\delta) &= \dim H_1\big(\mathcal{A}_2(\mathrm{PGU}_3(q^{1/2}))\big) = \tfrac{1}{3}(q^3-2q^{5/2}-q^2+2q^{3/2}-3q+3).
\end{aligned}$$

Let $d := \dim H_1(\mathcal{A}_2(L))$. Since $\mathcal{A}_2(L)$ is connected and $m_2(L) = 2$, by Lemma 4.4 we have

$$d = -\tilde{\chi}(\mathcal{A}_2(L)) = \tfrac{1}{3}(q-1)(q^2-1)(q^3+3q^2+3q+3).$$

Now it is routine to verify that $n_t d(t) > d$ if $t = \gamma$ or $t = \delta$, if and only if $(t, q) \ne (\gamma, 3)$. Indeed, for $q = 3$, $C_L(\gamma) = \mathrm{PGL}_2(3) \cong \mathrm{Sym}_4$ has nontrivial 2-core, so $d(\gamma) = 0$ and in consequence, $H_2(L\langle\gamma\rangle) = 0$.

This shows that $L\langle\gamma\rangle$ fails $(\mathcal{QD})_2$ if $q=3$. Therefore, a 2-extension of L by a field, graph or graph-field involution satisfies $(\mathcal{QD})_2$ if and only if $q\neq 3$ when L is extended by a graph involution.

It remains to show that $LB = L\langle\phi,\gamma\rangle$ verifies $(\mathcal{QD})_2$. For this case, we take $L_f = L\langle\phi\rangle$, $L_2 = LB$ and consider the long exact sequence of Lemma 3.10 at $m=2$ there (since $m_2(L_2)=4$). That is, we need to show that $H_3(\mathcal{A}_2(L_2))\neq 0$.

Note that the set of involutions $t\in L_2\setminus L_1$ is exactly the set of all graph and graph-field automorphisms of the extension $L_2=LB$. Let $d_g := \dim H_2\big(\mathcal{A}_2(\mathrm{PGL}_2(q)\langle\phi\rangle)\big)$, $d_{gf} := \dim H_2\big(\mathcal{A}_2(\mathrm{PGU}_3(q^{1/2})\langle\phi\rangle)\big)$ and $d_f := \dim H_2(\mathcal{A}_2(L_f))$. Therefore, by Lemma 3.10,

$$\dim H_3(\mathcal{A}_2(L_2)) \geq n_g d_g + n_{gf} d_{gf} - d_f. \tag{4-7}$$

We show that the right-hand side of this equation is positive if $q\neq 9$ by providing proper bounds of the dimensions d_g, d_{gf} and d_f.

By (4-4),

$$d_g = \dim H_2\big(\mathcal{A}_2(\mathrm{PGL}_2(q)\langle\phi\rangle)\big) \geq \tfrac{1}{4}(q^{1/2}-1)(q-1)(q^{3/2}-3q-4). \tag{4-8}$$

Next, by Proposition 4.5,

$$d_{gf} \geq \tfrac{1}{3}(q-1)(q^{1/2}+1)\left(\frac{q(q-q^{1/2}+1)}{(3,q^{1/2}+1)}(q^{1/2}-3) - (q^{3/2}-3q+3q^{1/2}-3)\right), \tag{4-9}$$

which is positive for all $q>9$.

Finally, we need to bound d_f from above. Indeed, by Lemma 3.10 at $m=2$, we have

$$\begin{aligned} d_f &= \dim H_2(\mathcal{A}_2(L_f)) = \dim H_2\big(\mathcal{A}_2(\mathrm{PSL}_3(q)\langle\phi\rangle)\big) \\ &\leq n_f \dim H_1\big(\mathcal{A}_2(\mathrm{PSL}_3(q^{1/2}))\big) \\ &= \frac{q^{3/2}(q+1)(q^{3/2}+1)}{(3,q^{1/2}+1)}\tfrac{1}{3}(q^{1/2}-1)(q-1)(q^{3/2}+3q+3q^{1/2}+3). \end{aligned}$$

Now we check with the given bounds that $n_g d_g + n_{gf} d_{gf} - d_f$ is positive if and only if $q>9$. In fact, if $q=9$, similar arguments show $H_3(\mathcal{A}_2(LB))=0$ since $d_g=0$ by Example 4.3 and $d_{gf}=0$ by Proposition 4.5.

We conclude that every 2-extension of $\mathrm{PSL}_3(q)$ satisfies $(\mathcal{QD})_2$, except for $\mathrm{PSL}_3(3)$ extended by a graph automorphism and for $\mathrm{PSL}_3(9)$ extended by field and graph automorphisms, which actually fail $(\mathcal{QD})_2$. □

5. The Quillen dimension property on exceptional groups of Lie type

We use the results of the preceding sections to show that, with only finitely many exceptions, the 2-extensions of the finite simple groups of exceptional Lie type satisfy $(\mathcal{QD})_2$. For that purpose, it will be convenient to recall first which 2-extensions can arise in each case. Table 3 records the 2-ranks of the exceptional groups of Lie type in adjoint version and the structure of the outer automorphism group. The

group	2-rank	Outdiag	Out / Outdiag
${}^3D_4(q)$	3	1	3Φ
$G_2(q)$	3	1	$\Phi\Gamma$, where $\lvert\Phi\Gamma:\Gamma\rvert = 2$ if $q = 3^a$, and $\Gamma = 1$ otherwise
${}^2G_2(q)$	3	1	Φ (odd order)
$F_4(q)$	5	1	Φ
$E_6(q)$	6	$(3, q-1)$	$\Phi \times \Gamma$, $\Gamma \cong \mathrm{C}_2$
${}^2E_6(q)$	6	$(3, q+1)$	2Φ
$E_7(q)$	8	2	Φ
$E_8(q)$	9	1	Φ

Table 3. Out / Outdiag is cyclic unless specified; $\Phi = \mathrm{Aut}(\mathbb{F}_q) \cong \mathrm{C}_a$, where $q = r^a$, r is an odd prime, and the usual conventions for the twisted types hold. Also, Γ is a set of graph automorphisms.

2-ranks were extracted from [Cohen and Seitz 1987; Gorenstein et al. 1998, Proposition 4.10.5]. From Table 3, we will compute the possible 2-extensions in each case.

Recall that we follow the terminology of [Gorenstein et al. 1998]. In particular, by a group of Lie type K we mean the finite group $O^{r'}(C_\sigma(\overline{K}))$, where $\overline{K}$ is a simple $\overline{\mathbb{F}}_r$-algebraic group and σ a Steinberg endomorphism of $\overline{K}$. Also, K is the adjoint version if $Z(K) = 1$. Unless we specify the version, we will always work with the adjoint versions of the exceptional groups of Lie type.

Cases $G_2(q)$ and ${}^2G_2(q)$. We start by proving that the Ree groups ${}^2G_2(q)$ satisfy $(\mathcal{QD})_2$ if and only if $q \neq 3$. Note that, by Table 3 for example, ${}^2G_2(q)$ has no nontrivial 2-extension.

Proposition 5.1. *Let L be the Ree group ${}^2G_2(q)$, where q is a power of 3 by an odd positive integer. Then the following hold*:

(1) *L has no nontrivial 2-extensions.*

(2) *A Sylow 2-subgroup of L is an elementary abelian group of order 8, so $m_2(L) = 3$.*

(3) *2-subgroups of equal order of L are conjugate.*

(4) *L satisfies $(\mathcal{QD})_2$ if and only if $q \neq 3$. Moreover, if $q > 3$ then*

$$\dim H_2(\mathcal{A}_2(L)) \geq \tilde{\chi}(\mathcal{A}_2(L)) = \tfrac{1}{21}(q^2-1)(q^5 - 8q^4 + 15q^3 + 21) > 0. \tag{5-1}$$

(5) *For $q = 3$, $\mathcal{A}_2(L) = \mathcal{A}_2(\mathrm{PSL}_2(8))$ is homotopy equivalent to a discrete space of 8 points.*

Proof. Items (1)–(3) are well-known facts about the Ree groups and can be found in [Ward 1966].

If $L = {}^2G_2(3)$, then $L' = \mathrm{PSL}_2(8)$ has index 3 in L, and $\mathcal{A}_2(L) \cong \mathcal{A}_2(\mathrm{PSL}_2(8))$ is homotopy equivalent to a discrete space with 8 points. Since $m_2(L) = 3$, we conclude that L fails $(\mathcal{QD})_2$ for $q = 3$. This proves item (5) and the "only if" part of item (4).

Now suppose that $q \neq 3$ and $L = {}^2G_2(q)$. Since $\mathcal{A}_2(L)$ has dimension 2 by item (2), we show that its second homology group is nonzero. To that end, it is enough to see that its Euler characteristic is positive since $\mathscr{A}_2(L)$ is connected for $q \neq 3$ by Theorem 3.8. Indeed,

$$\widetilde{\chi}(\mathscr{A}_2(L)) = \dim H_2(\mathscr{A}_2(L)) - \dim H_1(\mathscr{A}_2(L)) \leq \dim H_2(\mathscr{A}_2(L)).$$

We invoke Theorem C of [Kleidman 1988] to describe the normalisers of 2-subgroups: the centraliser of an involution is $2 \times \mathrm{PSL}_2(q)$, the normaliser of a four-subgroup is $(2^2 \times \mathrm{D}_{(q+1)/2}) : 3$, and the normaliser of a Sylow 2-subgroup is $2^3 : 7 : 3$. From this information, items (2), (3) and Proposition 3.9, we can compute the Euler characteristic of $\mathscr{A}_2(L)$:

$$\begin{aligned}\widetilde{\chi}(\mathscr{A}_2(L)) &= -1 + \frac{|L|}{2|\,\mathrm{PSL}_2(q)|} - 2\frac{|L|}{6(q+1)} + 8\frac{|L|}{168} \\ &= -1 + q^3(q^3+1)(q-1)\Big(\frac{1}{q(q^2-1)} - \frac{1}{3(q+1)} + \frac{1}{21}\Big) \\ &= \tfrac{1}{21}(q^2-1)(q^5 - 8q^4 + 15q^3 + 21).\end{aligned}$$

Since the polynomial $q^5 - 8q^4 + 15q^3 + 21 = q^4(q-8) + 15q^3 + 21$ is positive for every prime power $q \neq 4$, we conclude that $H_2(\mathscr{A}_2(L)) \neq 0$. In consequence, L satisfies $(\mathbb{Q}\mathscr{D})_2$ if $q \neq 3$. This completes the proof of item (4), and hence of this proposition. □

For the case $G_2(q)$, we refer the reader to the classification of maximal subgroups of $G_2(q)$ by P. Kleidman [1988]. We will follow the terminology of that article.

Proposition 5.2. *Let $L = G_2(q)$, with q odd. Then every 2-extension of L satisfies $(\mathbb{Q}\mathscr{D})_2$, except possibly for the 2-extensions of $G_2(3)$ and the 2-extension of $G_2(9)$ by a field involution.*

Proof. Let $L = G_2(q)$. We prove first that $G_2(q)$ and its extension by a field automorphism of order 2 satisfy $(\mathbb{Q}\mathscr{D})_2$, by exhibiting a maximal subgroup of the same rank that satisfies $(\mathbb{Q}\mathscr{D})_2$.

By Theorem A in [Kleidman 1988], $G_2(q)$ contains a subgroup $K_+ = \mathrm{SL}_3(q):2$. Let $L_+ = F^*(K_+) \cong \mathrm{SL}_3(q)$ and $Z = Z(L_+)$. Then $L_0 := L_+/Z = \mathrm{PSL}_3(q)$ and $H_0 := K_+/Z = L_0\langle\gamma\rangle$, where γ induces a graph automorphism on L_0 (see Proposition 2.2 and its proof in [Kleidman 1988]). By Proposition 4.6, L_0 satisfies $(\mathbb{Q}\mathscr{D})_2$ if $q \neq 3$, so H_0 satisfies $(\mathbb{Q}\mathscr{D})_2$.

On the other hand, $m_2(L) = 3$ by Table 3, and also $m_2(L_0) = 3$ by the proof of Proposition 4.6. Recall from Lemma 3.2 that

$$\widetilde{H}_*(\mathscr{A}_2(H_0)) = \widetilde{H}_*(\mathscr{A}_2(K_+/Z)) \subseteq \widetilde{H}_*(\mathscr{A}_2(K_+)).$$

In particular, we get the following inclusions between the top-degree homology groups

$$\widetilde{H}_2(\mathscr{A}_2(H_0)) \subseteq \widetilde{H}_2(\mathscr{A}_2(K_+)) \subseteq \widetilde{H}_2(\mathscr{A}_2(L)),$$

which show that L satisfies $(\mathbb{Q}\mathscr{D})_2$ if $q \neq 3$.

Next, a nontrivial 2-extension of $L = G_2(q)$ can only be given by field automorphisms of order 2 if q is not a power of 3. Moreover, by the construction of the subgroup K_+ given in [Kleidman 1988], field automorphisms of $G_2(q)$ induce field automorphisms on (a suitable conjugate of) K_+, and hence on the quotient H_0. Thus, for $B \in \mathscr{O}_2(L)$ inducing field automorphisms, we may take

K_+ fixed by B, and then $K_+B \cong \mathrm{SL}_3(q) : (2 \times B)$ after a suitable choice of conjugates (recall that $\mathrm{Out}(\mathrm{SL}_3(q)) = (3, q-1) : (\mathrm{Aut}(\mathbb{F}_q) \times \Gamma)$, where $\Gamma = 2$ is a group of graph automorphisms). Similar as before, we have a split extension $K_+B/Z = L_0B'$, where $B' = \langle \gamma \rangle \times B \in \mathcal{O}(L_0)$. By Proposition 4.6, L_0B' satisfies $(\mathbb{Q}\mathcal{D})_2$ if $q \neq 9$. Analogously to the previous case, $m_2(L_0B') = 4 = m_2(L) = m_2(K_+B)$, and we get an inclusion in the degree 3 homology groups, showing that K_+B and LB satisfy $(\mathbb{Q}\mathcal{D})_2$. Therefore, an extension of L by a field automorphism of order 2 satisfies $(\mathbb{Q}\mathcal{D})_2$ if $q \neq 9$.

It remains to analyse the case $q = 3^a$. By Table 4.5.1 of [Gorenstein et al. 1998] (see also Theorem 2.5.12 of [Gorenstein et al. 1998]), only field or graph-field automorphisms can arise in $\mathrm{Aut}(L)$. We have shown above that the extension of L by a field automorphism of order 2 satisfies $(\mathbb{Q}\mathcal{D})_2$ if $q \neq 9$. Thus we need to prove that if t is a graph-field automorphism of L, then $L\langle t\rangle$ satisfies $(\mathbb{Q}\mathcal{D})_2$. In that case, $q = 3^{2a+1}$ and by Proposition 2.5, $C_L(t) = {}^2G_2(q)$, which has 2-rank 3. Therefore $m_2(L\langle t\rangle) = 4$. However, by Theorem B of [Kleidman 1988], every maximal subgroup of $L\langle t\rangle$ containing t is either 2-local or has 2-rank at most 3. This shows that we cannot proceed as before via maximal subgroups. In view of this, we will proceed by using the long exact sequence of Lemma 3.10.

We have subgroups $M_0 := C_L(t) = {}^2G_2(q)$, $M_1 := G_2(3)\langle t\rangle \leq L\langle t\rangle$ and $M_2 := {}^2G_2(3)$ such that $M_2 \leq M_1 \cap M_0$. Fix A a Sylow 2-subgroup of M_2. By Proposition 5.1(2) and [Kleidman 1988, Lemma 2.4], A is also a Sylow 2-subgroup of M_0 and it is self-centralising in L, i.e., $C_L(A) = A$. A direct computation also shows that $N_{M_1'}(A) = A.\mathrm{PSL}_3(2)$, which immediately implies $N_L(A) = A.\mathrm{PSL}_3(2)$.

Now, suppose by the way of contradiction that $L\langle t\rangle$ fails $(\mathbb{Q}\mathcal{D})_2$, that is, the homology group $H_3(\mathcal{A}_2(L\langle t\rangle))$ vanishes. Recall that $C_L(t) = {}^2G_2(q)$ and there is a unique L-conjugacy class of involutions $t' \in L\langle t\rangle - L$ by Proposition 2.5(4). Let $X = \bigcup_{C_L(t)x \in L/C_L(t)} \mathcal{A}_2(C_L(t^x))$. By Lemma 3.10, we get inclusions

$$\bigoplus_{L/C_L(t)} H_2\big(\mathcal{A}_2(C_L(t))\big) \hookrightarrow H_2(X) \hookrightarrow H_2(\mathcal{A}_2(L)). \tag{5-2}$$

Set

$$d := \dim H_2(X), \quad d' := \dim \bigoplus_{L/C_L(t)} H_2\big(\mathcal{A}_2(C_L(t))\big) = |L : C_L(t)| \dim H_2\big(\mathcal{A}_2({}^2G_2(q))\big).$$

Equation (5-2) shows that $d' \leq d$. However, we will prove that $d < d'$, arriving then at a contradiction.

On one hand, we have that X is a union of $\mathcal{A}_2$-posets. Therefore, below each point, we have a wedge of spheres of maximal possible dimension. This means that the homology of X can be obtained from the chain complex that in degree i is freely generated by the spheres below each point of X of height i. In particular, for $i = 2$, the points of height 2 correspond to the conjugates of A, the fixed Sylow 2-subgroup of $M_0 = C_L(t)$ and M_2. Thus,

$$d = \dim H_2(X) \leq |L : N_L(A)| \cdot \#(\text{spheres below } A) = \frac{q^6(q^6-1)(q^2-1)}{168}.$$

On the other hand, by Proposition 5.1(4),

$$d' \geq |L : C_L(t)| \cdot \widetilde{\chi}\big(\mathcal{A}_2({}^2G_2(q))\big) = q^3(q^3-1)(q+1)\tfrac{1}{21}(q^2-1)(q^5-8q^4+15q^3+21).$$

Finally, from these bounds, we prove that $d' > d$ if $q \geq 7$. We can bound

$$\begin{aligned} d'-d &\geq q^3(q^3-1)(q+1)\tfrac{1}{21}(q^2-1)(q^5-8q^4+15q^3+21) - \frac{q^6(q^6-1)(q^2-1)}{168} \\ &= q^3(q^3-1)(q+1)\tfrac{1}{168}(q^2-1)\big(8(q^5-8q^4+15q^3+21)-q^3(q^2-q+1)\big) \\ &> 8(q^5-8q^4+15q^3+21)-q^3(q^2-q+1) \\ &= 7(q^4(q-9)+17q^3+24). \end{aligned}$$

The latter polynomial is clearly positive for $q \geq 9$, and also for $q = 7$ by direct computation. Since $q \geq 7$ by hypothesis, we conclude that $d' > d$. This gives a contradiction to equation (5-2), and thus shows that $H_3(\mathcal{A}_2(L\langle t\rangle)) \neq 0$; that is, $L\langle t\rangle$ satisfies $(\mathcal{QD})_2$. This finishes the proof of the proposition. □

Example 5.3. Let $L = G_2(3)$. We show that $\mathcal{A}_2(L)$ is homotopy equivalent to a wedge of spheres of dimension 1. In particular, since $m_2(L) = 3$, L fails $(\mathcal{QD})_2$. Moreover, by Lemma 3.10, also the unique nontrivial 2-extension of L (by a graph-field automorphism) fails $(\mathcal{QD})_2$.

We construct a subposet of $\mathcal{A}_2(L)$ of dimension 1 and homotopy equivalent to $\mathcal{A}_2(L)$. First, take the subposet $\mathfrak{i}(\mathcal{A}_2(L)) = \{A \in \mathcal{A}_2(L) : A = \Omega_1\big(Z\big(\Omega_1(C_L(A))\big)\big)\}$, which is homotopy equivalent to $\mathcal{A}_2(L)$ (see [Piterman 2019, Remark 4.5]). Next, there are two conjugacy classes of elementary abelian 2-subgroups of order 8, and both are contained in $\mathfrak{i}(\mathcal{A}_2(L))$. For one of these classes, say represented by A, the normaliser $N_L(A)$ has order 192. Then it can be shown that $\mathfrak{i}(\mathcal{A}_2(L))_{<A}$ is contractible. Therefore, if we remove the L-conjugates of A from $\mathfrak{i}(\mathcal{A}_2(L))$ we get a subposet $\mathfrak{si}(\mathcal{A}_2(L))$ homotopy equivalent to $\mathfrak{i}(\mathcal{A}_2(L))$. Now, there is a unique conjugacy class of four-subgroups in this new subposet $\mathfrak{si}(\mathcal{A}_2(L))$, and each such subgroup is contained in a unique element of order 8 of $\mathfrak{si}(\mathcal{A}_2(L))$. Again, we can remove all the four-subgroups from $\mathfrak{si}(\mathcal{A}_2(L))$ and obtain a new subposet Y homotopy equivalent to $\mathcal{A}_2(L)$. Since Y consists only of elements of order 2 and 8, we conclude that Y has dimension 1. Finally, an extra computation shows that $\tilde{\chi}(\mathcal{A}_2(L)) = -11584$. Therefore $\mathcal{A}_2(L)$ is homotopy equivalent to a wedge of 11584 spheres of dimension 1. In particular, L fails $(\mathcal{QD})_2$.

This also shows that $L = G_2(9)$ extended by a field automorphism of order 2 fails $(\mathcal{QD})_2$: if ϕ is a field involution, then $C_L(\phi) = G_2(3)$, and thus $H_2\big(\mathcal{A}_2(C_L(\phi))\big) = 0$ by the previous computation. Then by Lemma 3.10, we conclude that $H_3(\mathcal{A}_2(L)) = 0$.

Cases $^3D_4$ and $F_4(q)$.

Proposition 5.4. *The group $L = {}^3D_4(q)$ satisfies (E-$(\mathcal{QD})$) if $q \neq 9$ is odd. Also ${}^3D_4(9)$ satisfies $(\mathcal{QD})_2$.*

Proof. Recall that $m_2(L) = 3$ by Table 3. Then a graph automorphism of order 3 of ${}^3D_4(q)$ centralises a subgroup $K = G_2(q)$. Also, if ϕ denotes a field automorphism of order 2 of L, then, after choosing a suitable conjugate, we may assume that ϕ induces a field automorphism on K. By Proposition 5.2 and its proof, $m_2(K) = 3 = m_2(L)$, $m_2(K\langle\phi\rangle) = 4 = m_2(L\langle\phi\rangle)$, and both K and $K\langle\phi\rangle$ satisfy $(\mathcal{QD})_2$ for $q \neq 3, 9$ respectively. Also note that $G_2(9)$ satisfies $(\mathcal{QD})_2$. By Lemma 3.3, L and $L\langle\phi\rangle$ satisfy $(\mathcal{QD})_2$ if $q \neq 3, 9$, respectively. Since these are the only possible 2-extensions of L by Table 3, this concluded with the proof of our proposition for $q \neq 3$.

If $q = 3$, a computation of the Euler characteristic of $\mathcal{A}_2(L)$ in GAP with the Posets package [Fernández et al. 2019] shows that $\widetilde{\chi}(\mathcal{A}_2(L)) = 882634225472$. Since $\mathcal{A}_2(L)$ is connected by Theorem 3.8, we see that $H_2(\mathcal{A}_2(L)) \neq 0$, that is, L satisfies $(\mathbb{QD})_2$. □

Proposition 5.5. *If $L = F_4(q)$, with $q \neq 3, 9$ odd, then L satisfies (E-$(\mathbb{QD})$). Also $F_4(9)$ satisfies $(\mathbb{QD})_2$.*

Proof. Suppose that $q \neq 3, 9$ is an odd prime power. Then L contains a subgroup $H := \mathrm{PGL}_2(q) \times G_2(q)$ (see the main result of [Liebeck and Seitz 2004]). Note that H satisfies $(\mathbb{QD})_2$ by Propositions 3.1, 4.2 and 5.2. Since both L and H have 2-rank 5 by Table 3, we conclude that L satisfies $(\mathbb{QD})_2$.

Let $B \in \mathcal{O}_2(L)$, so B is generated by a field automorphism of order 2. Thus it acts by field automorphisms in a direct product subgroup isomorphic to H, which we may assume without loss of generality that it is our H. Then $\widetilde{H} = \mathrm{PGL}_2(q)B \times G_2(q^{1/2})$, which is a subgroup of HB, satisfies $(\mathbb{QD})_2$ by Propositions 3.1, 4.2 and 5.2. Since $m_2(\widetilde{H}) = 6 = m_2(LB)$, we conclude that LB also satisfies $(\mathbb{QD})_2$.

We have shown that every possible 2-extension of L satisfies $(\mathbb{QD})_2$, so L satisfies (E-$(\mathbb{QD})$).

If $q = 9$, then $\mathrm{PGL}_2(9) \times G_2(9)$ satisfies $(\mathbb{QD})_2$ by Propositions 3.1, 4.2 and 5.2. Therefore, $F_4(9)$ satisfies $(\mathbb{QD})_2$. □

Cases $E_6(q)$ and ${}^2E_6(q)$.

Proposition 5.6. *Let $L = E_6^{\epsilon}(q)$ (any version), $\epsilon \in \{\pm 1\}$, and q odd. Then L satisfies (E-$(\mathbb{QD})$).*

Proof. Let $L = E_6^{\epsilon}(q)$ in adjoint version (i.e., simple), where $\epsilon \in \{\pm 1\}$. For a 2-extension LB of the adjoint version L, we see that $m_2(LB) = m_2(L_u \widetilde{B})$, where L_u is the universal version of $E_6^{\epsilon}(q)$ and $\widetilde{B}$, isomorphic to B, is just a lift of the action of B on L_u (this is possible since $Z(L_u) = (3, q - \epsilon)$ is odd). Thus $LB = L_u \widetilde{B}/Z(L_u)$, and by Lemma 3.2, $\widetilde{H}_*(\mathcal{A}_2(LB)) \subseteq \widetilde{H}_*(\mathcal{A}_2(L_u B))$. Therefore, if L satisfies (E-$(\mathbb{QD})$), then so does the universal version of $E_6^{\epsilon}(q)$.

We will show that there exists a parabolic subgroup P of L such that for any 2-extension LB, a suitable conjugate of P is normalised by B (so we can suppose it is P itself), and $m_2(PB) = m_2(LB)$.

This parabolic subgroup P arises from the A_5 subdiagram in E_6, so $P = U\,\mathrm{GL}_6^{\epsilon}(q)/Z(L_u)$, where $\mathrm{GL}_6^{\epsilon}(q)/Z(L_u)$ denotes the Levi complement. Then $m_2(P) = 6$, which realises the 2-rank of L. Furthermore, a graph, graph-field or field automorphism of order 2 of L (the last two only for $\epsilon = 1$) stabilises this subdiagram (and hence P), inducing a graph (resp. graph-field or field) automorphism on $\mathrm{GL}_6^{\epsilon}(q)/Z(L_u)$. Denote by t such automorphism. Then $m_2(L\langle t\rangle) \leq m_2(L) + 1 = 7$. We claim that

$$m_2(P\langle t\rangle) = m_2(\mathrm{GL}_6^{\epsilon}(q)\langle t\rangle) = 7 = m_2(L\langle t\rangle). \tag{5-3}$$

Note that $m_2(P\langle t\rangle) = m_2(\mathrm{GL}_6^{\epsilon}(q)\langle t\rangle)$, for the lifted action of t on $\mathrm{GL}_6^{\epsilon}(q)$. Then it is clear that (5-3) holds if t induces a field automorphism (so $\epsilon = 1$), since the stabiliser of t in $\mathrm{GL}_6(q)$ is $\mathrm{GL}_6(q^{1/2})$. Similarly, if t is a graph-field automorphism then $\epsilon = 1$ and $C_{\mathrm{GL}_6(q)}(t) = \mathrm{GU}_6(q^{1/2})$, which has 2-rank 6. Then, in these two situations, $m_2(P\langle t\rangle) = 7$.

Now assume that t is a graph involution. For $\epsilon = 1$, t acts on $\mathrm{GL}_6(q)$, so $\mathrm{GL}_6(q)\langle t\rangle$ contains a graph automorphism g inducing the map $x \mapsto (x')^{-1}$, where x' denotes the transpose of x. Therefore,

$C_{\mathrm{GL}_6(q)}(g) = \{x \in \mathrm{GL}_6(q) : xx' = 1\} = \mathrm{GO}_6(q)$ is the subgroup of orthogonal matrices, which has 2-rank 6. This implies that $m_2(\mathrm{GL}_6(q)\langle t\rangle) = 6$. If $\epsilon = -1$, t is a graph involution acting on $\mathrm{GU}_6(q)$, so up to conjugation t is indeed the map $x \mapsto x^q$. Therefore, $C_{\mathrm{GU}_6(q)}(t) = \mathrm{GO}_6(q)$, so $m_2(\mathrm{GU}_6(q)\langle t\rangle) = 6$. In any case, we see that $m_2(P\langle t\rangle) = 7$.

Finally, suppose that we have $B = \langle \phi, \gamma\rangle$, where ϕ is a field automorphism of order 2 and γ a graph automorphism of order 2 of $L = E_6(q)$. We can suppose that B stabilises P (and thus its unipotent radical), and its Levi complement $\mathrm{GL}_6(q)/Z(L_u)$. Thus, γ induces a graph automorphism on the stabiliser of ϕ in $\mathrm{GL}_6(q)$, which is isomorphic to $\mathrm{GL}_6(q^{1/2})$. As we saw above, $m_2(\mathrm{GL}_6(q^{1/2})\langle\gamma\rangle) = 7$. Therefore, $m_2(\mathrm{GL}_6(q)B) = 8$. Since $m_2(B) = 2$ and $m_2(E_6(q)) = 6$, we conclude that $m_2(E_6(q)B) = 8$, so the 2-rank is realised in PB.

To conclude, note that a 2-extension of L is one of:

(1) L, of 2-rank $6 = m_2(P)$,

(2) $L\langle\gamma\rangle$ of 2-rank 7, with γ a graph automorphism of order 2, which also stabilises P and $m_2(P\langle\gamma\rangle) = 7$,

(3) $L\langle\phi\rangle$ of 2-rank 7, with ϕ a field automorphism of order 2 ($\epsilon = 1$) that stabilises P, and so $m_2(P\langle\phi\rangle) = 7$,

(4) $L\langle\gamma\phi\rangle$ of 2-rank 7, with $\gamma\phi$ a graph-field automorphism of order 2 ($\epsilon = 1$), which stabilises P, and thus $m_2(P\langle\gamma\phi\rangle) = 7$,

(5) $L\langle\gamma, \phi\rangle$ of 2-rank 8, with ϕ a field automorphism of order 2 ($\epsilon = 1$) commuting with γ a graph automorphism of order 2, and $\langle\gamma, \phi\rangle$ also stabilises P with $m_2(P\langle\gamma, \phi\rangle) = 8$.

From this, we conclude that any 2-extension of the simple group $E_6^\epsilon(q)$ satisfies $(\mathbb{Q}\mathfrak{D})_2$. By the remark at the beginning of the proof, we conclude that any version of $E_6^\epsilon(q)$ satisfies (E-($\mathbb{Q}\mathfrak{D}$)). □

Case $E_7(q)$.

Proposition 5.7. *Let $L = E_7(q)$ with q odd. Then L satisfies (E-($\mathbb{Q}\mathfrak{D}$)).*

Proof. Let $L = E_7(q)$. By Table 3, if ϕ denotes a field automorphism of order 2 of L, the 2-extensions of L are

$$L, \quad \mathrm{Inndiag}(L), \quad L\langle\phi\rangle.$$

Note that $\mathrm{Inndiag}(L)\langle\phi\rangle$ is not a 2-extension since field and diagonal automorphisms of order 2 do not commute in view of Lemma 2.6.

Next, we study the 2-ranks of these extensions, so we need to understand the centralisers of the outer involutions. From Table 3, $m_2(L) = 8$. We claim that $m_2(\mathrm{Inndiag}(L)) = 8 = m_2(L)$. Indeed, consider $K = E_7(q^2)$ in adjoint version. Then $m_2(K) = 8$. Let ϕ' be a field automorphism of order 2 for K. Then, by Proposition 2.5 and Lemma 2.6,

$$K \geq C_K(\phi') = C_{\mathrm{Inndiag}(K)}(\phi') = \mathrm{Inndiag}(E_7(q)) \cong \mathrm{Inndiag}(L).$$

From this we see that $m_2(\mathrm{Inndiag}(L)) = 8 = m_2(L)$. In particular, $\mathrm{Inndiag}(L)$ satisfies $(\mathbb{QD})_2$ if L does. Moreover, this also proves that if ϕ is a field automorphism of order 2 for L then

$$m_2(\mathrm{Inndiag}(L)\langle\phi\rangle) = 9 = m_2(L\langle\phi\rangle).$$

From these observations, we conclude that, in order to establish (E-$(\mathbb{QD})$) for $E_7(q)$, it is enough to show that $E_7(q)$ and $E_7(q)\langle\phi\rangle$ satisfy $(\mathbb{QD})_2$.

To this end, we exhibit a maximal parabolic subgroup of $E_7(q)$ of 2-rank 8. We see that D_6 is a subdiagram of E_7, so we have a maximal parabolic subgroup in $E_7(q)$ of the form

$$P = U(D_6(q).(q-1)).$$

Here U denotes the unipotent radical of P, and the subgroup $H = D_6(q)$ is a quotient of $\mathrm{Spin}_{12}^+(q)$ by a central subgroup of order 2. Indeed, $H = \mathrm{HSpin}_{12}^+(q)$ and it lies in the centraliser of the involution that generates the centre of a Sylow 2-subgroup T of L (see the t_1 involution of the $E_7(q)$ entry in Table 4.5.1 of [Gorenstein et al. 1998]). From this, we show that the Levi complement $L_P = D_6(q).(q-1)$ of P has 2-rank 8. Let t be the involution in the centre of L_P. Then $C_L(t) = (\mathrm{SL}_2(q) \circ_2 \mathrm{HSpin}_{12}^+(q)).2$ by Table 4.5.1 of [Gorenstein et al. 1998]. Since $t \in Z(T)$, $T \leq C_L(t)$. Also, $\mathrm{SL}_2(q)$ has a unique involution, so the 2-rank of T is realised in a subgroup of the extension $M := \mathrm{HSpin}_{12}^+(q).2$. Here, the 2 at the end comes from diagonal automorphisms of the half-spin group, as in the Levi complement above. Therefore, if we identify M as a subgroup of L_P, we conclude that $m_2(L_P) = m_2(M) = m_2(E_7(q))$.

Moreover, after suitable choices of conjugates, a field automorphism ϕ of order 2 must normalise P and act as a field automorphism on our M. Since $C_M(\phi)$ contains a subgroup isomorphic to $\mathrm{HSpin}_{12}^+(q^{1/2}).2$, we see that $P\langle\phi\rangle$ has 2-rank 9, which is the 2-rank of the 2-extension $E_7(q)\langle\phi\rangle$.

By Lemma 3.5, L and $L\langle\phi\rangle$ satisfy $(\mathbb{QD})_2$. Finally, by the previous discussion, we conclude that L satisfies (E-$(\mathbb{QD})$). □

Case $E_8(q)$.

Proposition 5.8. *The simple group $E_8(q)$, $q \neq 3, 9$ odd, satisfies (E-$(\mathbb{QD})$). Also $E_8(9)$ satisfies $(\mathbb{QD})_2$.*

Proof. Let $L = E_8(q)$. By Table 5.1 of [Liebeck et al. 1992], L contains a maximal subgroup

$$H \cong (3, q-1).(\mathrm{PSL}_3(q) \times E_6(q)).(3, q-1).2.$$

Note that

$$F^*(H) = (3, q-1).(\mathrm{PSL}_3(q) \times E_6(q)), \quad \text{and} \quad H_+ := H/Z(F^*(H)) = (\mathrm{PSL}_3(q) \times E_6(q)).(3, q-1).2,$$

where $(3, q-1)$ induces diagonal automorphism on each component of H_+, and the 2 induces a graph involution, also acting on both components. In particular, by taking the centraliser of a graph involution on the $\mathrm{PSL}_3(q)$ component, we see that H_0 contains a subgroup K_0 isomorphic to

$$\mathrm{PGL}_2(q) \times \mathrm{Inndiag}(E_6(q))\langle\gamma\rangle,$$

where γ is a graph involution of $E_6(q)$ centralising $\mathrm{PGL}_2(q)$. Now, recall that

$$m_2(L)=9 \quad \text{and} \quad m_2(\mathrm{PGL}_2(q))=2.$$

Since $m_2(E_6(q)\langle\gamma\rangle)=7$ by item (2) of the proof of Proposition 5.6, we see that

$$m_2(K_0)=m_2(\mathrm{PGL}_2(q))+m_2(E_6(q)\langle\gamma\rangle)=2+7=9=m_2(L).$$

Therefore K_0 realises the 2-rank of L.

By Table 3, $E_8(q)$ extended by a field automorphism of order 2, say ϕ, is the unique nontrivial 2-extension. From the construction of the maximal subgroup H and K_0 (see [Liebeck et al. 1992]), we can pick a suitable L-conjugate of ϕ (and we suppose it is the same ϕ) such that it normalises H and, after passing to the quotient, normalises K_0 and induces a field automorphism on both factors of K_0. In particular, we have a subgroup K_1 of $K_0\langle\phi\rangle$ of the form

$$\mathrm{PGL}_2(q^{1/2})\times \mathrm{Inndiag}(E_6(q))\langle\gamma',\phi\rangle,$$

where we have chosen $\gamma'\in\mathrm{Inndiag}(E_6(q))\langle\gamma\rangle$ to be a graph automorphism commuting with ϕ, and $\mathrm{PGL}_2(q^{1/2})=C_{\mathrm{PGL}_2(q)}(\phi)$. Therefore, by item (5) in the proof of Proposition 5.6,

$$m_2(K_1)=2+m_2(E_6(q)\langle\gamma',\phi\rangle)=2+8=10.$$

Since $m_2(L\langle\phi\rangle)\leq m_2(L)+1=10$, we conclude that $m_2(K_1)=m_2(L\langle\phi\rangle)$.

Finally, note that K_0 and K_1 satisfy $(\mathcal{QD})_2$ if $q\neq 3,9$ respectively, by Propositions 4.2, 5.6 and 3.1. Hence, by Lemmas 3.2 and 3.3, L and $L\langle t\rangle$ satisfy $(\mathcal{QD})_2$ if $q\neq 3,9$, respectively.

Therefore every 2-extension of $E_8(q)$ satisfies $(\mathcal{QD})_2$, with the exceptions given in the statement. This concludes the proof of the proposition. □

Acknowledgements

The author thanks Stephen D. Smith for many helpful discussions concerning the algebraic properties of groups of Lie type and Volkmar Welker for his suggestions on a preliminary version of the article. He is also very grateful to Gunter Malle and the anonymous referee for the many comments that helped to improve the presentation of the article.

References

[Aschbacher and Smith 1993] M. Aschbacher and S. D. Smith, "On Quillen's conjecture for the p-groups complex", *Ann. of Math.* (2) **137**:3 (1993), 473–529. MR Zbl

[Brown 1975] K. S. Brown, "Euler characteristics of groups: the p-fractional part", *Invent. Math.* **29**:1 (1975), 1–5. MR Zbl

[Cohen and Seitz 1987] A. M. Cohen and G. M. Seitz, "The r-rank of the groups of exceptional Lie type", *Nederl. Akad. Wetensch. Indag. Math.* **90**:3 (1987), 251–259. MR Zbl

[Cohen et al. 1992] A. M. Cohen, M. W. Liebeck, J. Saxl, and G. M. Seitz, "The local maximal subgroups of exceptional groups of Lie type, finite and algebraic", *Proc. London Math. Soc.* (3) **64**:1 (1992), 21–48. MR Zbl

[Díaz Ramos 2018] A. Díaz Ramos, "On Quillen's conjecture for p-solvable groups", *J. Algebra* **513** (2018), 246–264. MR Zbl

[Díaz Ramos and Mazza 2022] A. Díaz Ramos and N. Mazza, "A geometric approach to Quillen's conjecture", *J. Group Theory* **25**:1 (2022), 91–112. MR Zbl

[Fernández et al. 2019] X. Fernández, K. I. Piterman, and I. Sadofschi Costa, "Posets — Finite posets and finite topological spaces", 2019, available at https://github.com/isadofschi/posets. (GAP package).

[GAP] *GAP — Groups, Algorithms, and Programming, Version 4.12.1*, available at https://www.gap-system.org.

[Gorenstein et al. 1998] D. Gorenstein, R. Lyons, and R. Solomon, *The classification of the finite simple groups, 3: Almost simple K-groups*, Mathematical Surveys and Monographs **40.3**, American Mathematical Society, Providence, RI, 1998. MR Zbl

[Gorenstein et al. 2018] D. Gorenstein, R. Lyons, and R. Solomon, *The classification of the finite simple groups, 8: The generic case, completed*, Mathematical Surveys and Monographs **40.8**, American Mathematical Society, Providence, RI, 2018. MR Zbl

[Jacobsen and Møller 2012] M. W. Jacobsen and J. M. Møller, "Euler characteristics and Möbius algebras of p-subgroup categories", *J. Pure Appl. Algebra* **216**:12 (2012), 2665–2696. MR Zbl

[Kleidman 1988] P. B. Kleidman, "The maximal subgroups of the Chevalley groups $G_2(q)$ with q odd, the Ree groups ${}^2G_2(q)$, and their automorphism groups", *J. Algebra* **117**:1 (1988), 30–71. MR Zbl

[Liebeck and Seitz 1990] M. W. Liebeck and G. M. Seitz, "Maximal subgroups of exceptional groups of Lie type, finite and algebraic", *Geom. Dedicata* **35**:1-3 (1990), 353–387. MR Zbl

[Liebeck and Seitz 2004] M. W. Liebeck and G. M. Seitz, *The maximal subgroups of positive dimension in exceptional algebraic groups*, Mem. Amer. Math. Soc. **802**, 2004. MR Zbl

[Liebeck et al. 1992] M. W. Liebeck, J. Saxl, and G. M. Seitz, "Subgroups of maximal rank in finite exceptional groups of Lie type", *Proc. London Math. Soc.* (3) **65**:2 (1992), 297–325. MR Zbl

[Piterman 2019] K. I. Piterman, "A stronger reformulation of Webb's conjecture in terms of finite topological spaces", *J. Algebra* **527** (2019), 280–305. MR Zbl

[Piterman 2021] K. I. Piterman, "An approach to Quillen's conjecture via centralisers of simple groups", *Forum Math. Sigma* **9** (2021), art. id. e48. MR Zbl

[Piterman and Smith 2022a] K. I. Piterman and S. D. Smith, "Eliminating components in Quillen's conjecture", *J. Algebra* **607** (2022), 681–732. MR Zbl

[Piterman and Smith 2022b] K. I. Piterman and S. D. Smith, "Some results on Quillen's conjecture via equivalent-poset techniques", preprint, 2022. arXiv 2204.13055

[Piterman and Welker 2022] K. I. Piterman and V. Welker, "Homotopy properties of the complex of frames of a unitary space", preprint, 2022. arXiv 2208.12626

[Piterman et al. 2021] K. I. Piterman, I. Sadofschi Costa, and A. Viruel, "Acyclic 2-dimensional complexes and Quillen's conjecture", *Publ. Mat.* **65**:1 (2021), 129–140. MR Zbl

[Quillen 1971] D. Quillen, "The spectrum of an equivariant cohomology ring, I, II", *Ann. of Math.* (2) **94** (1971), 549–572, 573–602. MR Zbl

[Quillen 1978] D. Quillen, "Homotopy properties of the poset of nontrivial p-subgroups of a group", *Adv. in Math.* **28**:2 (1978), 101–128. MR Zbl

[Segev and Webb 1994] Y. Segev and P. Webb, "Extensions of G-posets and Quillen's complex", *J. Austral. Math. Soc. Ser. A* **57**:1 (1994), 60–75. MR Zbl

[Smith 2011] S. D. Smith, *Subgroup complexes*, Mathematical Surveys and Monographs **179**, American Mathematical Society, Providence, RI, 2011. MR Zbl

[Ward 1966] H. N. Ward, "On Ree's series of simple groups", *Trans. Amer. Math. Soc.* **121** (1966), 62–89. MR Zbl

[Welker 1995] V. Welker, "Direct sum decompositions of matroids and exponential structures", *J. Combin. Theory Ser. B* **63**:2 (1995), 222–244. MR Zbl

Communicated by Pham Huu Tiep
Received 2023-01-16 Revised 2023-05-19 Accepted 2023-09-03

piterman@mathematik.uni-marburg.de *Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Marburg, 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.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

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@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 18 No. 7 2024