Pacific Journal of Mathematics

Volume 332 No. 2 October 2024

PACIFIC JOURNAL OF MATHEMATICS

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

msp.org/pjm

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

The subscription price for 2024 is US $645/year for the electronic version, and $875/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 1945-5844 electronic, 0030-8730 printed) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.195

HOMOTOPY VERSUS ISOTOPY: 2-SPHERES IN 5-MANIFOLDS

DANICA KOSANOVIĆ, ROB SCHNEIDERMAN AND PETER TEICHNER

We give a complete obstruction for two homotopic embeddings of a 2-sphere into a 5-manifold to be isotopic. The results are new even though the methods are classical, the main tool being the elimination of double points via a level preserving Whitney move in codimension 3. Moreover, we discuss how this recovers a particular case of a result of Dax on metastable homotopy groups of embedding spaces. It follows that "homotopy implies isotopy" for 2-spheres in simply connected 5-manifolds and for 2-spheres admitting algebraic dual 3-spheres.

1. Introduction and results

A curious consequence of our generalizations [Schneiderman and Teichner 2022; Kosanović and Teichner 2024a] of the 4-dimensional light bulb theorems of David Gabai [2020; 2021] is that homotopic 2-spheres $R, R' : S^2 \hookrightarrow M$, embedded in a 4-manifold M with a common dual sphere, are smoothly isotopic in M if and only if they are isotopic in the 5-manifold $M \times \mathbb{R}$, see [Schneiderman and Teichner 2022, Corollary 1.5]. The complete isotopy obstruction in [Schneiderman and Teichner 2022, Theorem 1.1] is given by the Freedman–Quinn invariant

$$\mathrm{fq}(R, R') := [\mu_3(H)] \in \frac{\mathbb{F}_2 T_M}{\mu_3(\pi_3 M)},$$

where $\mu_3(H)$ is Wall's self-intersection invariant of the track

$$H : S^2 \times [0, 1] \looparrowright M \times \mathbb{R} \times [0, 1]$$

of a generic homotopy between R and R' in $M \times \mathbb{R}$. Moreover, $\mathbb{F}_2 T_M$ is the $\mathbb{F}_2$-vector space with basis $T_M := \{g \in \pi_1 M \mid g^2 = 1 \neq g\}$, the set of involutions in $\pi_1 M$. It turns out that μ_3 also gives a homomorphism $\mu_3 : \pi_3 M \to \mathbb{F}_2 T_M$, whose cokernel eliminates the choice of homotopy in the definition of fq. Michael Freedman and Frank Quinn [1990, Chapter 10] introduced this invariant, while studying topological concordance classes of embedded 2-spheres in 4-manifolds.

MSC2020: 57K45, 57R40, 58D10.

Keywords: 2-spheres in 5-manifolds, homotopy implies isotopy, level preserving Whitney trick, metastable homotopy groups, Freedman–Quinn invariant, Dax invariant.

This isotopy classification also follows from [Kosanović and Teichner 2024a, Theorem 1.1] via a more powerful invariant due originally to Jean-Pierre Dax [1972] which detects relative isotopy classes of neatly embedded 2-disks having a common dual in ∂M. Dax extends the parametrized double-point elimination method of André Haefliger [1961a; 1961b], which is in turn an extension of the Whitney trick [Whitney 1944]. Haefliger's results were used by Lawrence Larmore [1978, Theorem 6.0.1] to show a special case of Dax's result, see (1.14) below.

In the current paper we consider the "homotopy versus isotopy" question for 2-spheres in general 5-manifolds and show that there is again a self-intersection invariant of a homotopies which detects isotopy classes and takes values in a quotient of the group ring of the ambient fundamental group. The dimensions under consideration here are right at the transition between low- and high-dimensional topology, with successful Whitney moves generally available in the presence of vanishing algebraic obstructions.

With this in mind, our exposition will be aimed at describing this transition from the point of view of the low-dimensional topologist, rather than starting by presenting results in full generality. In particular, we will:

(1) Explain how the isotopy classification can be described by self-intersection invariants of homotopies, using a level-preserving Whitney trick.

(2) Explain how Dax's work recovers the same result, from the perspective of space level techniques and homotopy groups of embedding spaces.

Both approaches can be generalized to describe isotopy classifications for compact n-manifolds embedded in $(2n+1)$-manifolds, as will be described in upcoming work [Kosanović et al. $\geq$ 2024] which recasts Dax [1972] in full generality in this language; see Theorem 4.1 below.

We next give a quick outline of the main results, and refer to the rest of the introduction for details. For any fixed embedded sphere $U : S^2 \hookrightarrow N^5$ in a 5-manifold N, we will define the set $\mathbb{A}_{U_*}$ as a certain quotient of the group ring $\mathbb{Z}\pi_1 N$, see Definition 1.10. The image in $\mathbb{A}_{U_*}$ of the self-intersections $\mu_3(H)$ of a generic track $H : S^2 \times I \looparrowright N \times I$ of a homotopy between U and $R : S^2 \hookrightarrow N$ will by design only depend on U and R, not on H. Here the notation U is meant to suggest that one can think of this fixed 2-sphere as an "unknot", although we emphasize that there are no restrictions on its homotopy class. Denoting this invariant by $\mathrm{fq}_{U_*}(R) := [\mu_3(H)] \in \mathbb{A}_{U_*}$, a basic statement of our main result is the following corollary of Theorem 1.11:

Corollary 1.1. *Homotopic spheres U and R are isotopic if and only if the formula $\mathrm{fq}_{U_*}(R) = 0 \in \mathbb{A}_{U_*}$ holds. Moreover, any element in $\mathbb{A}_{U_*}$ is realized as $\mathrm{fq}_{U_*}(R)$ for an embedded sphere R.*

We note that in this 5-dimensional setting, the result does not require any dual spheres (unlike in four dimensions), see Corollary 1.13. If $\pi_1 N$ is trivial, then $\mathbb{A}_{U_*} = \{0\}$ and we get:

Corollary 1.2. *Homotopy implies isotopy for* 2*-spheres in simply connected* 5*-manifolds.*

There is a straightforward proof of Corollary 1.2, using cusp-cancellation (in dimension 6) and the theorem of John Hudson [1970] that in codimension > 2 concordance implies isotopy. We will give a self-contained proof of the general classification result, by providing a level-preserving version of the Whitney move in codimension > 2 (Proposition 2.5).

If $\pi_1 N$ is not trivial then our classification result is similar to the 4-dimensional setting with common duals, namely fq_{U_*} gives the unique obstruction for homotopic embeddings to be isotopic. Our main work will be in spelling out the precise range of this obstruction and showing that all values in $\mathbb{A}_{U_*}$ are realized. A new issue that arises in the current setting is the distinction between based and free homotopies, whereas the assumption of common duals in the 4-dimensional setting essentially allowed for consideration of only based homotopies (see [Gabai 2020, Theorem 6.1; Schneiderman and Teichner 2022, Lemma 2.1]). We will occasionally emphasize this issue by applying the adjective “free” to the terms “homotopy” and “isotopy”, even though by traditional definitions it would suffice to just omit the adjective “based”.

1.1. ***2-knots in 5-manifolds.*** We now turn to precise formulations of our main results, working in the smooth oriented category throughout. Fixing a basepoint in such a 5-manifold N, and a basepoint in S^2, we have the following commutative diagram which will guide the discussion of our invariants:

$$\begin{array}{ccc} \pi_0 \operatorname{Emb}_*(S^2, N) & \xrightarrow{p_*} & \pi_0 \operatorname{Map}_*(S^2, N) \\ \downarrow{\scriptstyle \bmod \pi_1 N} & & \downarrow{\scriptstyle \bmod \pi_1 N} \\ \pi_0 \operatorname{Emb}(S^2, N) & \xrightarrow{p} & \pi_0 \operatorname{Map}(S^2, N) \end{array}$$

Here

$$\pi_0 \operatorname{Emb}_*(S^2, N) := \{\text{based embeddings } S^2 \hookrightarrow N\}/\text{based isotopy}$$

is the embedded version of $\pi_0 \operatorname{Map}_*(S^2, N) = \pi_2 N$, and

$$\pi_0 \operatorname{Emb}(S^2, N) := \{\text{embeddings } S^2 \hookrightarrow N\}/\text{free isotopy}$$

is the embedded version of $\pi_0 \operatorname{Map}(S^2, N) = [S^2, N]$. Both horizontal arrows forget the fact that we have embeddings, and the vertical arrows divide out the $\pi_1 N$-actions (using embedded tubes along closed paths at the basepoint on the left-hand side). Both p and p_* are surjective (that is, any (based) map is homotopic to a (based) embedding) by general position: the dimension of the double point set is $5-6=-1$ (so generically this set is empty), since codimensions of generic intersections add.

Along with using the label $*$ for based objects, our notational convention is to use a bracket to denote the homotopy class of an embedded object, which is otherwise considered up to isotopy. For example, the upper horizontal map p_* sends $R_* \in \pi_0 \operatorname{Emb}_*(S^2, N)$ to $[R_*] \in \pi_0 \operatorname{Map}_*(S^2, N)$, and the left vertical map sends R_* to $R \in \pi_0 \operatorname{Emb}(S^2, N)$. We are ultimately interested in the fibers of p, but it turns out to be convenient to first understand the fibers of p_*.

1.2. ***The based isotopy invariant.*** As recalled in (2.2) below, the quotient

$$\mathbb{A} := \frac{\mathbb{Z}\pi_1 N}{\langle g + g^{-1}, 1\rangle}$$

of the integral fundamental group ring $\mathbb{Z}\pi_1 N \cong \mathbb{Z}\pi_1(N \times I)$ is the usual target for the self-intersection invariant

(1.3) $\mu_3 : \{\text{simply connected 3-manifolds immersed in the 6-manifold } N \times I\} \to \mathbb{A},$

which is invariant under homotopy rel boundary.

Let us fix a based embedding $U_* : S^2 \hookrightarrow N$, and define a homomorphism of abelian groups $\phi_{[U_*]} : \pi_3 N \to \mathbb{A}$ by

$$\phi_{[U_*]}(A) := \mu_3(A) + [\lambda_N(A, [U_*])], \tag{1.4}$$

where μ_3 denotes the self-intersection invariant on $\pi_3(N \times I) \cong \pi_3 N$, and λ_N is the intersection pairing between $\pi_3 N$ and $\pi_2 N$ taking values in $\mathbb{Z}\pi_1 N$.

Definition 1.5.
$$\mathbb{A}_{[U_*]} := \frac{\mathbb{A}}{\phi_{[U_*]}(\pi_3 N)}.$$

Note that the abelian group $\mathbb{A}_{[U_*]}$ only depends on the based homotopy class $[U_*] \in \pi_2 N$. We now consider the fiber of p_* over $[U_*] \in \pi_2 N$:

$$p_*^{-1}([U_*]) = \{R_* : S^2 \hookrightarrow N \mid R_* \text{ is based homotopic to } U_*\}/(\text{based isotopy}).$$

Definition 1.6. For $R_* \in p_*^{-1}[U_*]$, let $H_* : S^2 \times I \looparrowright N \times I$ be a generic track of a based homotopy from U_* to R_*, and define

$$\operatorname{fq}_{[U_*]}(R_*) := [\mu_3(H_*)] \in \mathbb{A}_{[U_*]}.$$

By the following theorem, $\operatorname{fq}_{[U_*]}(R_*)$ does not depend on the choice of homotopy, and vanishes if and only if R_* and U_* are based isotopic.

Theorem 1.7. *The map* $\mathrm{fq}_{[U_*]} : p_*^{-1}[U_*] \to \mathbb{A}_{[U_*]}$ *is a bijection, whose inverse is given by a geometric action on* U_*.

The action of $g \in \pi_1 N$ on $U_* \in \pi_0 \operatorname{Emb}_*(S^2, N)$ is by a "finger move" along g, which in this setting is an ambient connected sum of U_* with its meridian sphere m_{U_*} along a tube following a loop representing g. Elements in the group ring act by multiple finger moves, which turn out to involve signs and preserve the relations in the quotient $\mathbb{A}$ of the group ring (see Section 2.3). The proof of Theorem 1.7 shows the following.

Corollary 1.8. *The abelian group* $\mathbb{A}$ *acts on* $\pi_0 \operatorname{Emb}_*(S^2, N)$ *compatibly with the* $\pi_1 N$*-actions, preserving* p_* *and acting transitively on its fibers, with the stabilizer of* U_* *equal to* $\phi_{[U_*]}(\pi_3 N)$.

1.3. ***The free isotopy invariant.*** Now consider *free* homotopy versus isotopy, i.e., the set

$$p^{-1}[U] := \{R : S^2 \hookrightarrow N \mid R \text{ is freely homotopic to } U\}/(\text{free isotopy})$$

for $U : S^2 \hookrightarrow N$ a fixed embedding in $\operatorname{Emb}(S^2, N) \subset \operatorname{Map}(S^2, N)$. Choose $U(e) \in N$ as the basepoint for N, where e denotes the basepoint for S^2.

To define the target of an invariant that characterizes $p^{-1}[U]$ we will define an affine action on $\mathbb{A}_{[U_*]}$ (the range of the bijection in the based setting of Theorem 1.7) by the group

$$\operatorname{Stab}[U_*] := \{s \in \pi_1 N : s \cdot [U_*] = [U_*]\},$$

that is, the stabilizer subgroup of $[U_*] \in \pi_2 N$ of the usual action of $\pi_1 N$ on $\pi_2 N$.

Recall that an *affine transformation* T of an abelian group A is given by an endomorphism ℓ and a translation a_0 of A, i.e., $T(a) = a_0 + \ell(a)$, where $a_0 = T(0)$. An *affine action* of a group on A is a homomorphism to the group of affine transformations of A. In our case, the linear action of $s \in \operatorname{Stab}[U_*]$ will be $a \mapsto sas^{-1}$, whereas the translational part will be given by U_s, both of which we explain next.

Firstly, we claim that the linear action $(s, a) \mapsto sas^{-1}$ of $\operatorname{Stab}[U_*]$ on $\mathbb{Z}\pi_1 N$ descends to $\mathbb{A}_{[U_*]}$: for $A \in \pi_3 N$ we have

$$\mu_3(g \cdot A) = g \mu_3(A)\, g^{-1} \quad \text{and} \quad \lambda_N(g \cdot A, [U_*]) = g \lambda_N(A, [U_*]),$$

so if $g \cdot [U_*] = [U_*]$ then the last expression also equals $g\lambda_N(A, [U_*])g^{-1}$, implying

$$g\phi_{[U_*]}(A)g^{-1} = \phi_{[U_*]}(g \cdot A) \in \phi_{[U_*]}(\pi_3 N).$$

Secondly, for $s \in \operatorname{Stab}[U_*]$ there is a generic track $J_s : S^2 \times I \looparrowright N \times I$ of a free self-homotopy of U_* such that the projection of $J_s(e, -)$ to N represents s. It turns out that

$$U_s := [\mu_3(J_s)] \in \mathbb{A}_{[U_*]}$$

only depends on s and the isotopy class U_* (and not on J_s, see Lemma 3.1). It is easy to show that under concatenation, for $s, r \in \mathrm{Stab}[U_*]$ this behaves as

$$U_{s\cdot r} = U_s + sU_r\, s^{-1}.$$

This implies that the formula

$$\text{(aff)} \qquad {}^s a := U_s + sas^{-1}$$

satisfies

$$\text{(1.9)} \quad {}^{sr}a = U_{sr} + srar^{-1}s^{-1} = U_s + sU_r\, s^{-1} + srar^{-1}s^{-1} = {}^s(U_r + rar^{-1}) = {}^s({}^r a).$$

In other words, the composition in the group $\mathrm{Stab}[U_*]$ turns into the composition of affine transformations, so $s \mapsto {}^s a$ defines an affine action of $s \in \mathrm{Stab}[U_*]$ on $a \in \mathbb{A}_{[U_*]}$. This action and the following definition will be examined and clarified below in Section 3.2.

Definition 1.10. Denote by $\mathbb{A}_{U_*}$ the quotient set of this affine action by $\mathrm{Stab}[U_*]$ on $\mathbb{A}_{[U_*]}$.

Note that the definition of $\mathbb{A}_{U_*}$ depends on the based isotopy class U_*, a fixed basing of U. However, the following result shows that it characterizes the set of *free* isotopy classes of embeddings homotopic to U.

Theorem 1.11. *There is a bijection*

$$\mathrm{fq}_{U_*} : p^{-1}[U] \to \mathbb{A}_{U_*}, \quad R \mapsto [\mu_3(H)],$$

where H is a generic track of any free homotopy from U to R.

Here both the computation of $\mu_3(H)$ and the definition of $\mathbb{A}_{U_*}$ use the same basing U_*, and the inverse of the bijection is defined using the same geometric action as in Corollary 1.8.

Curiously, in the based setting of Theorem 1.7, both the target $\mathbb{A}_{[U_*]}$ and the set $p_*^{-1}[U_*]$ only depend on the based homotopy class $[U_*] \in \pi_2 N$, whereas in the free setting of Theorem 1.11, the target $\mathbb{A}_{U_*}$ depends on the *based isotopy* class $U_* \in \pi_0 \,\mathrm{Emb}_*(S^2, N)$, while the set $p^{-1}[U]$ depends on the *free homotopy* class $[U] \in \pi_0 \,\mathrm{Map}(S^2, N)$ of the embedding U.

Example. If U_* is the trivial 2-sphere then $U_s = 0$ for all $s \in \mathrm{Stab}[U_*] = \pi_1 N$ because any self-homotopy J_s can be chosen to be a self-isotopy. As a consequence:

Corollary 1.12 (null-homotopic isotopy classes). *Null-homotopic free isotopy classes of 2-spheres in N are in bijection with $\mathbb{A}_{U_*} = \mathbb{A}_{[U_*]}/\pi_1 N$, the quotient of the abelian group*

$$\mathbb{A}_{[U_*]} = \frac{\mathbb{Z}[\pi_1 N]}{\langle 1, g + g^{-1}, \mu_3(A) : g \in \pi_1 N, A \in \pi_3 N \rangle},$$

by the conjugation action of $\pi_1 N$.

Note that although the free self-isotopies of U have vanishing self-intersection invariants, they still contribute to the indeterminacy of the invariant fq_{U_*} by arbitrarily conjugating double-point group elements in the computation of $\mu_3(H)$.

Example. If $\lambda_N(G, [U_*]) = 1$ for some $G \in \pi_3 N$ such that $\mu_3(G) = 0 \in \mathbb{A}$, then $\phi_{[U_*]}(\pi_3 N) = \mathbb{A}$, since for any $\sum g_i \in \mathbb{Z}[\pi_1 N]$ we have

$$\phi_{[U_*]}\big(\textstyle\sum g_i \cdot G\big) = 0 + \big[\lambda_N\big(\sum g_i \cdot G, [U_*]\big)\big] = \big[\sum g_i \cdot \lambda_N(G, [U_*])\big] = \big[\sum g_i\big].$$

It follows that in this case $\mathbb{A}_{[U_*]}$ contains a single element, and hence so does $\mathbb{A}_{U_*}$.

Corollary 1.13 (5-dimensional light bulb theorem). ***Homotopy implies isotopy*** *for spheres $U_* : S^2 \hookrightarrow N^5$ admitting an* ***algebraic dual*** *$G \in \pi_3 N$ as above.*

See Section 3.3 for more examples.

1.4. ***Isotopy classification via mapping spaces.*** In Section 4 we present a slightly different perspective to the problem of isotopy classification. Namely, the fibers of the map $p : \pi_0 \operatorname{Emb}(S^2, N) \to \pi_0 \operatorname{Map}(S^2, N)$ can also be determined using the homotopy exact sequence associated to the inclusion of mapping spaces $\operatorname{Emb}(S^2, N) \subset \operatorname{Map}(S^2, N)$:

$$\begin{array}{ccccc} \pi_1(\operatorname{Map}(S^2, N), U) & & & & \\ \big\downarrow j & & & & \\ \pi_1(\operatorname{Map}(S^2, N), \operatorname{Emb}(S^2, N), U) & \longrightarrow & \pi_0 \operatorname{Emb}(S^2, N) & \xrightarrow{p} & \pi_0 \operatorname{Map}(S^2, N). \end{array}$$

Here we picked an embedding $U : S^2 \hookrightarrow N$ as a basepoint, and the leftmost absolute π_1 is a group that acts on the relative π_1 (which is just a set) such that the quotient set is isomorphic to the fiber $p^{-1}[U]$ of p over $[U] \in \pi_0 \operatorname{Map}(S^2, N)$.

The relative π_1 is the first nonvanishing relative homotopy group, and that is exactly what was computed by Jean-Pierre Dax [1972]. He translated this (and also other relative homotopy groups in the "metastable range") to certain bordism groups. Computations of this bordism group (which is 0-dimensional in the first nonvanishing case) were carried out in [Kosanović and Teichner 2024b, Theorem 4.14], for the cases $(\operatorname{Imm}_\partial(V, X), \operatorname{Emb}_\partial(V, X))$ when embeddings have nonempty boundary condition, and V is 1-connected. In [Kosanović et al. $\geq$ 2024] we extend this to closed and disconnected manifolds, and compare to maps instead of immersions. Specializing [Kosanović et al. $\geq$ 2024] to $V = S^2$ and $d = 5$ leads to Theorem 4.1: there is a bijection

$$\mathsf{Dax} : \pi_1(\operatorname{Map}(S^2, N), \operatorname{Emb}(S^2, N), U) \to \mathbb{A}, \quad h \mapsto \mu_3(H), \tag{1.14}$$

where H is a homotopy from U to an embedding that represents h, and $\mu_3(H) \in \mathbb{A}$ is the self-intersection invariant of a generic track H of the homotopy, as in (1.3).

In order to compute the set $p^{-1}[U]$ from this viewpoint it remains to understand the action of the absolute π_1 on the relative π_1 in the sequence displayed above. To the best of our current knowledge this does not seem to have been done previously, and guided by our classification using self-intersection invariants of homotopy tracks we proceed as follows: Firstly, we use the fibration sequence

$$\mathrm{Map}_*(S^2, N) \to \mathrm{Map}(S^2, N) \to N$$

to obtain exactness in the column of the following diagram, where the first map is the inclusion i and the second map evaluates at the basepoint $e \in S^2$:

$$\begin{array}{ccc}
\pi_3 N \cong \pi_1(\mathrm{Map}_*(S^2, N), U_*) & & \\
\downarrow i & & \\
\pi_1(\mathrm{Map}(S^2, N), U) & \xrightarrow{\ j\ } & \pi_1(\mathrm{Map}(S^2, N), \mathrm{Emb}(S^2, N), U) \xrightarrow[\cong]{\mathsf{Dax}} \mathbb{A} \\
\downarrow ev_e & & \\
\mathrm{Stab}[U_*] & &
\end{array}$$

By definition the composite $\mathsf{Dax} \circ j \circ i$ sends $\beta \in \pi_3 N$ to $\mathsf{Dax}(A)$ where A is a loop in $\mathrm{Map}_*(S^2, N)$, based at U. We will see that this precisely agrees with $\phi_{[U_*]}$ from (1.4).

Moreover, we will see that the induced action of $s \in \mathrm{Stab}[U_*]$ on the quotient of $\mathbb{A}$ by $\mathsf{Dax} \circ j \circ i(\pi_3 N)$ sends $a = \mathsf{Dax}(H)$ to

$$\mathsf{Dax}(J_s) + sas^{-1}, \tag{1.15}$$

where J_s is a free self-homotopy of U_* such that $ev_e(J_s) = s$, i.e., the projection of $J_s(e, -)$ to N represents s. This action is precisely (aff), so we recover

$$p^{-1}[U] \cong \mathbb{A}_{U_*}$$

as in Theorem 1.11, except that instead of fq_{U_*} this map is now naturally called Dax. This will be stated as Theorem 4.7.

Remark 1.16. Using the analogous fibration sequence

$$\mathrm{Emb}_*(S^2, N) \to \mathrm{Emb}(S^2, N) \to N,$$

we show in [Kosanović et al. ≥ 2024] that there is an isomorphism

$$i^{\mathrm{rel}}: \pi_1(\mathrm{Map}_*(S^2, N), \mathrm{Emb}_*(S^2, N), U_*) \xrightarrow{\ \cong\ } \pi_1(\mathrm{Map}(S^2, N), \mathrm{Emb}(S^2, N), U).$$

In Theorem 4.6 we will show that $\mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_* = \phi_{[U_*]}$ precisely gives the indeterminacy for the based setting: $p_*^{-1}[U_*] \cong \mathbb{A}/\phi_{[U_*]}(\pi_3 N)$, as in Theorem 1.7. Moreover, this shows that $\mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_* = \mathsf{Dax} \circ j \circ i$. Therefore, similarly to our first approach, in this approach we see: a linear action for the based setting (just the quotient by the image of $\phi_{[U_*]}$), and an affine action for the free setting (the further quotient by the action (aff)).

2. Intersection invariants and homotopies

2.1. ***3-manifolds in 6-manifolds.*** Recall that for a smooth oriented 6-manifold P^6, the intersection and self-intersection invariants give maps

$$\lambda_3 : \pi_3 P \times \pi_3 P \to \mathbb{Z}\pi_1 P \quad \text{and} \quad \mu_3 : \pi_3 P \to \mathbb{Z}\pi_1 P/\langle g + g^{-1}, 1\rangle.$$

To compute λ_3 geometrically, start by representing the two homotopy classes by transverse based maps $S^3 \to P$, and then count each intersection point p with a sign ϵ_p determined by orientations and a group element $g_p \in \pi_1 P$ represented by a sheet-changing loop through p. Here a *based map* is equipped with a *whisker*, which is an arc running between a basepoint on the image of the map and the basepoint of the ambient manifold P. Note that by general position a map of a manifold of codimension > 1 is ambient isotopic to a map whose basepoint is equal to the basepoint of the ambient manifold.

Similarly, for μ_3 one represents the homotopy class by a generic map $A : S^3 \looparrowright P$ and counts self-intersections, again with signs and group elements. In this dimension, switching the ordering of sheets at a double point p changes ε_p to $-\varepsilon_p$, and changes g_p to g_p^{-1}, explaining the relation $g + g^{-1} = 0$ in the range of μ_3. The relation $1 = 0$ makes $\mu_3(A)$ only depend on the homotopy class of A, since a cusp homotopy introduces a double point with arbitrary sign and trivial group element. Changing the whisker on A changes $\mu_3(A)$ by a conjugation, with the corresponding group element represented by the difference of the whiskers. The argument for homotopy invariance of μ_3 arises from considering the double-point arcs and circles of the track of a generic homotopy $S^3 \times I \looparrowright P^6 \times I$ of A.

Using the involution $\bar{g} := g^{-1}$ on $\mathbb{Z}\pi_1 P$, the "quadratic form" (λ_3, μ_3) satisfies

$$\begin{aligned} \mu_3(A + B) &= \mu_3(A) + \mu_3(B) + [\lambda_3(A, B)], \\ \lambda_3(A, A) &= \mu_3(A) - \overline{\mu_3(A)} \in \mathbb{Z}\pi_1 P, \end{aligned} \tag{2.1}$$

where the second formula has no content for the coefficient at the trivial element in $\pi_1 P$: Since λ_3 is skew-hermitian, it vanishes on the left-hand side, whereas it is automatically zero on the right-hand side that is defined by picking a representative of $\mu_3(A) \in \mathbb{Z}\pi_1 P$ and then applying the involution to that specific choice. We will be interested in the case that $P = N \times I$ is the product of a 5-manifold N with an interval I, and we denote the target of μ_3 by

$$\mathbb{A} := \frac{\mathbb{Z}\pi_1 N}{\langle g + g^{-1}, 1\rangle} \cong \frac{\mathbb{Z}\pi_1 P}{\langle g + g^{-1}, 1\rangle}.$$

2.2. ***The self-intersection invariant for homotopies of 2-spheres in 5-manifolds.*** The above descriptions of λ_3 and μ_3 can also be applied to properly immersed simply connected 3-manifolds with boundary in a 6-manifold. In this setting the invariants are computed just as above, by summing signed double point group elements, and are invariant under homotopies that restrict to isotopies on the boundary. For a smooth oriented 5-manifold N, and any homotopy $H : S^2 \times I \to N$ between embedded spheres, we define the self-intersection invariant

$$\mu_3(H) \in \mathbb{A} \tag{2.2}$$

to be the self-intersection invariant of a generic track $S^2 \times I \looparrowright N \times I$ for H. We will sometimes use the same letter H to denote either the homotopy or its track when the context is clear.

The "time" parameter (the I-factor) of a homotopy will generally be the unit interval $I = [0, 1]$, although it will be frequently suppressed from notation or reparametrized without mention.

For the purpose of computing the self-intersection invariant $\mu_3(H)$, the whisker on the track of H will be taken at the "start" $H(S^2 \times 0) \subset N \times 0 \subset N \times I$ of the homotopy unless explicitly stated otherwise. So for a homotopy H from $U_* : S^2 \hookrightarrow N$ to an embedding, the whisker for U_* will generally be used to compute $\mu_3(H)$.

Note that choosing a whisker on the track of a homotopy to provide a "basing" for the purposes of computing an intersection invariant is different than saying that the homotopy is a "based homotopy", which is "a homotopy through based maps".

2.3. ***Geometric action of*** $\mathbb{A}$. For $g \in \pi_1 N$ and $U_* : S^2 \hookrightarrow N$, we define $g \cdot U_* : S^2 \hookrightarrow N$ as follows, see Figure 2.3 for several examples. Firstly, note that the normal bundle of U_* is 3-dimensional, so its meridian m_F is a 2-sphere; we choose it over a point near the basepoint $z := U_*(e)$ and orient it according to the orientations of S^2 and N. We then define $g \cdot U_*$ as an ambient connected sum of U_* with m_F along a tube following an arc representing g, where the arc starts and ends near z and has interior disjoint from U_*. This is well defined up to isotopy since removing a neighborhood of U_* does not change the fundamental group and "homotopy implies isotopy" for arcs in this dimension.

Similarly, $(-g) \cdot U_*$ is defined to be the connected sum of U_* with the oppositely oriented meridian sphere $-m_F$. Linear combinations $\sum_i n_i g_i$ act by multiple connected sums along g_i into copies of m_F for $n_i > 0$ respectively $-m_F$ for $n_i < 0$. It is not hard to check that the relations $g + g^{-1} = 0 = 1$ carry over to isotopies of these connected sums, see Figure 2.3. Therefore, we have an action of $\mathbb{A}$ on the set $\pi_0 \operatorname{Emb}_*(S^2, N)$.

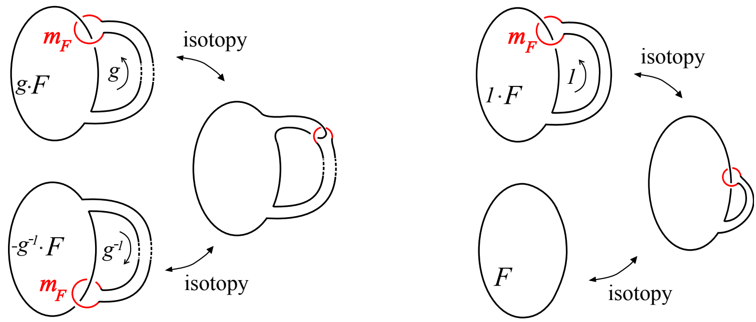

Figure 2.3. The relations $g = -g^{-1}$ and $1 = 0$ realized by isotopies.

Since each meridian sphere m_F bounds a normal 3-ball that intersects U_* exactly once, we get the following result.

Lemma 2.4. *For any $a \in \mathbb{A}$, shrinking the meridian spheres along their 3-balls gives a based homotopy H_a from $a \cdot U_*$ to U_* with $\mu_3(H_a) = a$.* □

2.4. *The level-preserving Whitney move.*

Proposition 2.5. *The track $H : S^2 \times I \to N^5 \times I$ of a homotopy between two embeddings is homotopic (rel boundary) to the track of an isotopy if and only if its self-intersection invariant $\mu_3(H) \in \mathbb{A}$ vanishes.*

It will follow from the proof that if the original homotopy is a based homotopy, then the resulting isotopy can be taken to be based. In fact, the construction given in the proof can be taken to be supported away from any I-family of whiskers.

We remark that since the classical Whitney move works for immersed 3-manifolds in 6-manifolds [Milnor 1965, Theorem 6.6], the vanishing of $\mu_3(H) \in \mathbb{A}$ immediately implies that the track $S^2 \times I \looparrowright N \times I$ is homotopic (rel boundary) to a concordance, so Proposition 2.5 would then follow from Hudson's theorem that concordance implies isotopy in codimensions ≥ 3 [Hudson 1970]. Rather than invoking Hudson's result, our proof of Proposition 2.5 will show that one can arrange for the Whitney moves to preserve I-levels in order to directly achieve an isotopy rather than just a concordance.

Proof of Proposition 2.5. The "only if" direction is clear since μ_3 is invariant under homotopy and vanishes on embeddings.

To prove the "if" direction, we first introduce some streamlined notation that will only be used in the proof of Proposition 2.5, including the ancillary Lemma 2.6.

Notation. For any subset $\sigma \subset I$, denote by $H_\sigma := H|_{S^2\times\sigma}$ the restriction to $S^2 \times \sigma$ of the track $H : S^2 \times I \to N \times I$. By the standard abuse of the notation, $H_\sigma := H(S^2 \times \sigma)$ is also the image of this map, and is contained in the subset $N_\sigma := N \times \sigma \subset N \times I$.

Lemma 2.6. *For H as in Proposition 2.5 with $\mu_3(H) = 0 \in \mathbb{A}$, it may be arranged by a homotopy rel ∂ that there exist finitely many distinct points $c_i \in I$ such that the transverse self-intersections of H occur in pairs $\{p_i, q_i\} \subset H_{c_i}$ with $g_{p_i} = g_{q_i}$ and $\epsilon_{p_i} = -\epsilon_{q_i}$ for each i.*

Assuming Lemma 2.6 (which will be proved just below), Proposition 2.5 follows by a standard application of Whitney moves to eliminate each of the self-intersection pairs of Lemma 2.6 in a way that yields the track of an isotopy. We describe details here for completeness, with the key observation being that each Whitney disk can be chosen to be contained in a level.

Dropping the subscript i from the notation, let p and q be a pair of self-intersections of H_c as in Lemma 2.6. Since H_c is a map of a 2-sphere $S^2 \times c$, the self-intersections p and q are not transverse for H_c, but there exists some small $\delta > 0$ such that $\{p, q\} = H_{[c-\delta,c+\delta]} \pitchfork H_{[c-\delta,c+\delta]}$, the transverse self-intersections of the immersed 3-manifold $H_{[c-\delta,c+\delta]}$ in the 6-manifold $N_{[c-\delta,c+\delta]}$.

Since p and q have the same group elements $g_p = g_q$ and opposite signs $\epsilon_p = -\epsilon_q$, there exists a *Whitney disk* $W \subset N_{[c-\delta,c+\delta]}$ pairing p and q. By general position we may assume that W is embedded in the 5-dimensional slice $N_c \subset N_{[c-\delta,c+\delta]}$ with interior disjoint from H_c. The Whitney disk boundary $\partial W = \alpha \cup \beta$ is the union of embedded arcs α and β contained in H_c, with $\alpha \cap \beta = \{p, q\}$. Let $\bar{\alpha}$ and $\bar{\beta}$ be slightly longer arcs in H_c containing α and β, respectively, such that $\bar{\alpha}$ and $\bar{\beta}$ extend just beyond p and q.

Let $A, B \subset H_{[c-\delta,c+\delta]}$ denote regular 3-ball neighborhoods of $\bar{\alpha}$ and $\bar{\beta}$ in $H_{[c-\delta,c+\delta]}$. Each of A and B is "the image of a local sheet of a 2-sphere H_t moving in time", with A_t and B_t each embedded in H_t for $t \in [c-\delta, c+\delta]$, and $A \cap B = \{p, q\}$. It follows that $A : D^2 \times I \hookrightarrow N \times I$ is the track of an isotopy A_t, and similarly for B.

The Whitney move that eliminates p and q will be described using a particular choice of coordinates for an open neighborhood $V \subset N_{[c-\delta,c+\delta]}$ containing W.

By [Milnor 1965, Lemma 6.7], the neighborhood V may be chosen to be diffeomorphic to $\overline{W} \times \mathbb{R}^2_A \times \mathbb{R}^2_B$, where:

- $\overline{W} \subset N_c$ is a smooth 2-disk formed from W by attaching a half-open collar to ∂W.
- $V \cap A = \bar{\alpha} \times \mathbb{R}^2_A \times (0, 0)$ and $V \cap B = \bar{\beta} \times (0, 0) \times \mathbb{R}^2_B$.

Let $\bar{\alpha}(s)$ be a smooth isotopy of the arc $\bar{\alpha}$ in $\overline{W}$, for $0 \le s \le 1$, such that $\bar{\alpha}(0) = \bar{\alpha}$, and $\bar{\alpha}(1)$ passes just above $\beta \subset \bar{\beta}$ as in Figure 2.7. In particular, $\bar{\alpha}(s)$ is supported near $W \subset \overline{W}$ for all $0 \le s \le 1$, and $\bar{\alpha}(1)$ is disjoint from B.

Let $\rho : \mathbb{R}^2 \to [0, 1]$ be a smooth bump function $(u, v) \mapsto \rho(u, v)$ such that

- $\rho(u, v) = 1$ if $\sqrt{u^2 + v^2} \le 1$,
- $\rho(u, v) = 0$ if $\sqrt{u^2 + v^2} \ge 2$.

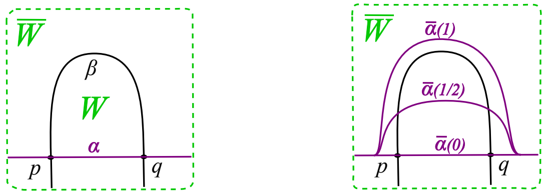

Figure 2.7. A smooth isotopy of the arc $\bar\alpha$ in $\overline{W}$.

Now we use $\bar\alpha(s)$ and ρ to define the Whitney move as the result of an isotopy $A(s),\ 0 \le s \le 1$ of A which fixes B, where (a_1, a_2) runs through $\mathbb{R}^2_A$:

$$A(s) = \bar\alpha(s\rho(a_1, a_2)) \times (a_1, a_2) \times (0, 0) \subset \overline{W} \times \mathbb{R}^2_A \times \mathbb{R}^2_B,$$

Then $A(0) = A$, and the result of the W-Whitney move on A is $A' := A(1)$, so that $A' \cap B = \varnothing$.

Note that $A(s) = A$ near ∂V, and hence we can extend $A(s)$ to be the identity outside V. This defines a homotopy $H(s)$ of $H = H(0)$ such that $H' := H(1)$ satisfies $H' \pitchfork H' = (H \pitchfork H) - \{p, q\}$.

By construction $A(s)$ only moves points of A along the $\overline{W}$-factor, which is orthogonal to the I-factor of $N \times I$ since $\overline{W} \subset N_c$. This means that for each s, $A(s)$ consists of the track of an isotopy $A_t(s)$, and similarly for $B(s)$.

Performing Whitney moves on all the self-intersection pairs of Lemma 2.6 yields the track of an isotopy as desired. □

Proof of Lemma 2.6. The condition $\mu_3 H = 0 \in \mathbb{A}$ means that the (finite) set of transverse self-intersections of the generic track $H : S^2 \times I \looparrowright N \times I$ can be decomposed into finitely many pairs $\{p_i, q_i\}$ with $g_{p_i} = g_{q_i}$ and $\epsilon_{p_i} = -\epsilon_{q_i}$ (for appropriately chosen sheets, and after perhaps performing a single cusp homotopy on H).

Suppose, for some i we have $p_i \in H_{t_1} \cap H_{t_1} \subset H \pitchfork H$ and $q_i \in H_{t_2} \cap H_{t_2} \subset H \pitchfork H$, with $t_1 < t_2$. We will describe how to change H by an isotopy rel boundary which "moves" q_i to $q_i' \in H_{t_1} \cap H_{t_1}$ and is supported away from all other self-intersections of H. The construction will show more generally that self-intersections of H can be arranged to occur at any chosen times, while preserving signs and group elements.

Special case. Consider the special case that H has just a single pair $\{p, q\} = H \pitchfork H$ of self-intersections with $p \in H_{t_1} \cap H_{t_1}$ and $q \in H_{t_2} \cap H_{t_2}$, and with $0 < t_1 < t_2 < 1$ in $I = [0, 1]$. Suppose that (x, t_2) and (y, t_2) are the two distinct preimages in $S^2 \times t_2 \subset S^2 \times I$ of $q = H_{t_2}((x, t_2)) = H_{t_2}((y, t_2))$. Define vertical arcs $a := x \times [t_1, t_2]$ and $b := y \times [t_1, t_2]$ in the domain $S^2 \times I$ (see Figure 2.8, left).

For $0 \le s \le 1$ we will define smooth isotopies $\psi_s : S^2 \times I \to S^2 \times I$ supported near $a \cup b$ and projecting to the identity on S^2 such that the smooth family of homotopies $H \circ \psi_s$ satisfies $H = H \circ \psi_0$ and $(H \circ \psi_1) \pitchfork (H \circ \psi_1) = \{p, q'\} \subset (H \circ \psi_1)_{t_1}$.

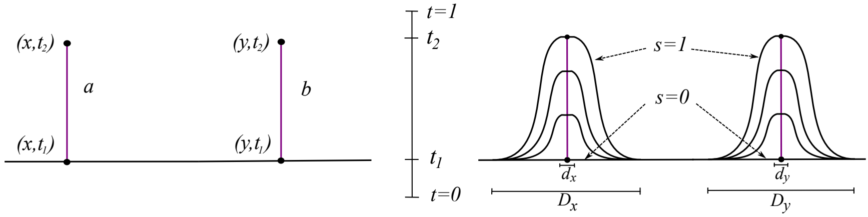

Figure 2.8. Schematic pictures in the domain $S^2 \times I$, with the I-factor running vertically from bottom to top. Left: the vertical arcs a and b, and a horizontal subarc of $S^2 \times t_1$. Right: images of the subarc of $S^2 \times t_1$ containing the nested disks $d_x \subset D_x$ and $d_y \subset D_y$ under the isotopy ψ_s for $s = 0,\ s = 1$, and for two other intermediate values of s.

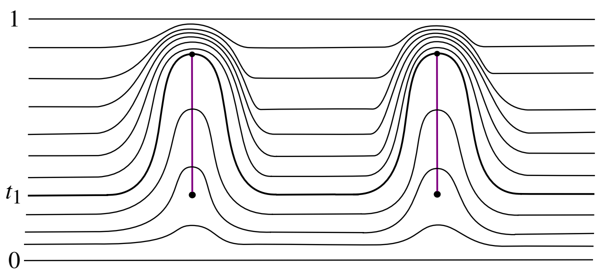

Figure 2.9. A schematic picture of the images under ψ_1 of some horizontal slices of $S^2 \times I$.

First we define ψ_s on $S^2 \times t_1$ as the sum of two local bump functions of height $s(t_2 - t_1)$ in the positive I-direction centered at (x, t_1) and (y, t_1) (see Figure 2.8, right). More specifically, let $x \in d_x \subset D_x \subset S^2$ and $y \in d_y \subset D_y \subset S^2$ be small concentric pairs of nested disks around x and y, respectively. Then

$$\psi_s(z, t_1) = \begin{cases} (z, t_1) & \text{if } z \notin D_x \cup D_y, \\ (z, t_1 + s(t_2 - t_1)) & \text{if } z \in d_x \cup d_y, \\ (z, t_1 + \operatorname{sig}(z)\, s(t_2 - t_1)) & \text{if } z \in (D_x \setminus \operatorname{int}(d_x)) \cup (D_y \setminus \operatorname{int}(d_y)). \end{cases}$$

Here the sigmoid function $\operatorname{sig}(z)$ smoothly interpolates between $\operatorname{sig}(z) = 0$ for $z \in \partial D_x \cup \partial D_y$ and $\operatorname{sig}(z) = 1$ for $z \in \partial d_x \cup \partial d_y$.

Now extend ψ_s to all of $S^2 \times I$ by tapering the bump functions down to zero as t moves away from t_1, so that $\psi_s(z, 0) = (z, 0)$ and $\psi_s(z, 1) = (z, 1)$ for all s. See Figure 2.9 for an illustration of the extended $\psi_1 : S^2 \times I \to S^2 \times I$.

Next we check that $H' := H \circ \psi_1$ has the desired properties. Observe that since ψ_s restricts to the identity map on the complement of $(D_x \cup D_y) \times I$, we have $H' = H$ when restricted to $(S^2 \setminus (D_x \cup D_y)) \times I$. In particular, $H'((S^2 \setminus (D_x \cup D_y)) \times I)$ has only the single transverse self-intersection point $p \in H \pitchfork H$ which occurs at t_1.

Now consider the restriction of H' to $(D_x \cup D_y) \times I$. We have that by construction, $H'((D_x \cup D_y) \times I)$ has just the single transverse self-intersection such that $q' = H'(d_x, t_1) \cap H'(d_y, t_1)$, which is the image of the projection to N_{t_1} of $q \in H \pitchfork H$.

Now it remains to check that there are no transverse intersections between $H'((D_x \cup D_y) \times I)$ and $H'((S^2 \setminus (D_x \cup D_y)) \times I)$. For each $t \in I$, suppose that $\text{proj}_t : N \times I \to N_t$ is the projection map. By general position, for each $t \in I$ the image $\text{proj}_t \circ H(a \cup b) \subset N_t$ is an embedded arc γ_t. Since the image $H'((D_x \cup D_y) \times I)$ in the 6-manifold $N \times I$ can be arranged to be contained in an arbitrarily small neighborhood of the 2-dimensional union $\cup_{t \in I} \gamma_t$, it follows by general position that $H'((D_x \cup D_y) \times I)$ has no transverse intersections with the 3-dimensional $H'((S^2 \setminus (D_x \cup D_y)) \times I)$.

General case. Since by general position any number of self-intersections can be assumed to have preimages projecting to distinct points in S^2, the above construction moving $q \in H_{t_2} \cap H_{t_2}$ to $q' \in H'_{t_1} \cap H'_{t_1}$ for $t_1 < t_2$ can be carried out iteratively (or even simultaneously) for any chosen subset of self-intersections while fixing the complementary subset. $\square$

2.5. ***Based self-homotopies.*** Recall from Section 1.2 of the introduction that, for a fixed based embedding $U_* : S^2 \hookrightarrow N$, we denote by $\mathbb{A}_{[U_*]}$ the quotient of $\mathbb{A}$ by the image of the indeterminacy homomorphism $\phi_{[U_*]} : \pi_3 N \to \mathbb{A}$ defined by $A \mapsto \mu_3(A) + [\lambda_N(A, [U_*])]$.

The following lemma will be used to show that our invariants are well defined.

Lemma 2.10. *If $J_* : S^2 \times I \looparrowright N \times I$ is a generic track of a based self-homotopy of U_*, then*

$$\mu_3(J_*) = 0 \in \mathbb{A}_{[U_*]}.$$

Proof. Since J_* is a based self-homotopy, it agrees with $U_* \times I$ on the 2-skeleton $S^2 \times \{0, 1\} \cup e \times I$ of $S^2 \times I$, with $e \in S^2$ the basepoint. So they only differ on the 3-cell, where $U_* \times I$ is represented by $B_U := U_*(D^2) \times I$ (here D^2 is the complement in S^2 of a small disk around e) and J_* is represented by a generic 3-ball $B_J : D^3 \looparrowright (N \times I) \smallsetminus \nu(U_*(e) \times I)$. By construction, the boundaries of these two 3-balls are parallel copies of an embedded 2-sphere in the boundary of a small neighborhood of $U_* \times \{0, 1\} \cup (U_*(e) \times I)$.

Gluing B_U and B_J together along a small embedded cylinder $S^2 \times I$ between their boundaries yields a map of a 3-sphere $A := B_J \cup (-B_U) : S^3 \to N \times I$. To prove the lemma we will show that $\mu_3(J_*) = \phi_{[U_*]}(A)$.

First note that on one hand, all contributions to $\mu_3(J_*)$ come from the self-intersections of the immersed 3-ball B_J. On the other hand, contributions to $\mu_3(A)$ come from the self-intersections of B_J and the intersections between B_J and the embedded 3-ball $-B_U$. The latter intersections are precisely counted by $-\lambda_3(J_*, U_* \times I)$, see (2.1). Therefore,

$$\mu_3(J_*) - \mu_3(A) = \lambda_3(J_*, U_* \times I),$$

and since $\lambda_N(A, U_*) = \lambda_3(A, U_* \times I) = \lambda_3(J_*, U_* \times I)$, we obtain

$$\mu_3(J_*) = \mu_3(A) + \lambda_N(A, U_*) = \phi_{[U_*]}(A). \qquad \square$$

2.6. ***From homotopy to isotopy by adding 3-spheres.*** The following lemma will be used to show that our invariants are injective.

Lemma 2.11. *Suppose $H : S^2 \times I \looparrowright N \times I$ is a generic track of a homotopy between embeddings $R : S^2 \hookrightarrow N$ and $R' : S^2 \hookrightarrow N$ such that the homotopy restricts to an embedding $U : S^2 \hookrightarrow N$ for some point $t_0 \in I$. Then R is isotopic to R' if $\mu_3(H) = 0 \in \mathbb{A}_{[U_*]}$. Moreover, if H is a based homotopy, then the resulting isotopy may be taken to be based.*

Proof. Since $\mu_3(H) = 0 \in \mathbb{A}_{[U_*]}$, there exists $A \in \pi_3 N$ such that $\mu_3(H) = \phi_{[U_*]}(A)$. By using a small ambient isotopy we may assume that H restricts to a product $U \times [t_0 - \epsilon, t_0 + \epsilon]$ on a small interval around t_0. Represent A by a generic regular homotopy $f_t : S^2 \times [t_0 - \epsilon, t_0 + \epsilon] \to N$ from a local trivial sphere $f_{t_0 - \epsilon} = f_{t_0+\epsilon}$ in N to itself via the isomorphism $\pi_3 N \cong \pi_1(\mathrm{Map}_*(S^2, N))$. Taking a smooth family of ambient connected sums of f_t with $U \times t \subset U \times [t_0 - \epsilon, t_0 + \epsilon]$ yields a self-homotopy J^A of U. We can assume the guiding paths for these connected sums have interiors disjoint from every f_t and $U \times t$, so that $\mu_3(J^A) = \phi_{[U_*]}(A)$. Reversing the t-parameter of J^A yields the track $-J^A$ of a self-homotopy of U with $\mu_3(-J^A) = -\phi_{[U_*]}(A)$.

Now, deleting $H \times [t_0 - \epsilon, t_0 + \epsilon]$ from $H \times I$ and gluing in $-J^A$ yields a based homotopy H^0 between R and R' with $\mu_3(H^0) = 0 \in \mathbb{A}$. It follows from Proposition 2.5 that H^0 is homotopic rel boundary to an isotopy R between R'. If H is a based homotopy, then this resulting isotopy inherits the extended whiskers from H. $\square$

3. Homotopy versus isotopy

In Section 3.1 we recall the statement of Theorem 1.7 describing the based setting, and give a proof. In Section 3.2 we clarify and prove Theorem 1.11, describing the free setting.

Our convention is to write concatenations of homotopies as unions from left to right, with a minus sign indicating that the orientation of the I-factor has been reversed. Recall (Section 2.2) that for the purposes of computing the self-intersection invariant $\mu_3(H)$ of a homotopy H the whisker on the track of H will be assumed to be taken at the "start" $H(S^2 \times 0) \subset N \times 0 \subset N \times I$ unless otherwise explicitly specified.

3.1. ***The based setting.*** Theorem 1.7 states that for a fixed based embedding $U_* : S^2 \hookrightarrow N^5$ the map

$$p_*^{-1}[U_*] \to \mathbb{A}_{[U_*]}, \quad R_* \mapsto \mathrm{fq}_{[U_*]}(R_*) := [\mu_3(H_*)]$$

for H_* any based homotopy from U_* to R_*, is a bijection.

Here we have $p_*^{-1}[U_*]$ the set of based isotopy classes of embedded spheres $R_* : S^2 \hookrightarrow N^5$ that are based homotopic to U_*. Moreover, the group $\mathbb{A}_{[U_*]}$ is the quotient of $\mathbb{A} := \mathbb{Z}\pi_1 N/\langle g + g^{-1}, 1\rangle$ by the image of the indeterminacy homomorphism $\phi_{[U_*]} : \pi_3 N \to \mathbb{A}$ defined by

$$A \mapsto \mu_3(A) + [\lambda_N(A, [U_*])].$$

$\mathrm{fq}_{[U_*]}$ *is well defined.* It suffices to show that $\mathrm{fq}_{[U_*]}(R_*) \in \mathbb{A}_{[U_*]}$ is independent of the choice of H_*.

Taking the union along R_* of any two based homotopies H_*, H'_* from U_* to R_* gives a based self-homotopy $J_* = H_* \cup_{R_*} -H'_*$ of U_* such that

$$\mu_3(J_*) = \mu_3(H_* \cup_{R_*} -H'_*) = \mu_3(H_*) - \mu_3(H'_*),$$

where we are using that μ_3 is additive under concatenations of based homotopies and changes sign under reversing the orientation of the time parameter. Since $\mu_3(J_*)$ lies in the image of $\phi_{[U_*]}$ by Lemma 2.10, we have $[\mu_3(H_*)] = [\mu_3(H'_*)] \in \mathbb{A}_{[U_*]}$.

$\mathrm{fq}_{[U_*]}$ *is injective.* If $\mathrm{fq}_{[U_*]}(R_*) = \mathrm{fq}_{[U_*]}(R'_*)$, then there exist based homotopies H_* and H'_* from U_* to R_* and R'_*, such that $\mu_3(H_*) = \mu_3(H'_*) \in \mathbb{A}_{[U_*]}$. Taking the union of these homotopies along U_* gives a based homotopy $H''_* := H_* \cup_{U_*} -H'_*$ from R_* to R'_* with $\mu_3(H''_*) = \mu_3(H_*) - \mu_3(H'_*) = 0 \in \mathbb{A}_{[U_*]}$. It follows from Lemma 2.11 that R_* is based isotopic to R'_*.

$\mathrm{fq}_{[U_*]}$ *is surjective.* Surjectivity follows directly from Lemma 2.4.

3.2. ***The free setting.*** This section clarifies the target of the invariants in the free setting, and proves Theorem 1.11, which we recall here for the reader's convenience: For U_* a fixed basing of an embedding $U : S^2 \hookrightarrow N^5$, the map

$$p^{-1}[U] \to \mathbb{A}_{U_*}, \quad R \mapsto \mathrm{fq}_{U_*}(R) := [\mu_3(H)],$$

where H is any free homotopy from U to R, is a bijection.

Here $p^{-1}[U]$ is the set of isotopy classes of embedded spheres $R : S^2 \hookrightarrow N^5$ that are freely homotopic to U. Moreover, the group $\mathbb{A}_{U_*}$ is the quotient set of the based target $\mathbb{A}_{[U_*]}$ of Theorem 1.7 by the affine action of $\mathrm{Stab}[U_*] < \pi_1 N$ given by ${}^s a = U_s + sas^{-1}$ for all $a \in \mathbb{A}_{[U_*]}$ and $s \in \mathrm{Stab}[U_*]$, with the definition of $U_s \in \mathbb{A}_{[U_*]}$ given just after Lemma 3.1 in the next subsection.

The affine action of $\mathrm{Stab}[U_*]$ *on* $\mathbb{A}_{[U_*]}$. For each $s \in \mathrm{Stab}[U_*]$ there is a track

$$J_s : S^2 \times I \to N \times I$$

of a free self-homotopy of U_* such that the projection of $J_s(e, -)$ represents s, where the basepoint $e \in S^2$ is the preimage of the basepoint of U_*. For instance, such a J_s can be taken to be a (level-preserving) subset of the track of any based homotopy between $s \cdot U_*$ and U_*. We say that J_s *represents* $s \in \mathrm{Stab}[U_*]$, and call s the *core* of J_s, frequently using the subscript notation to indicate this representation.

Note the following three properties of core elements:

(1) Any self-homotopy whose core is the trivial element of $\pi_1 N$ is homotopic rel boundary to a based self-homotopy.

(2) Concatenating self-homotopies multiplies the core elements: $J_{sr} = J_s \cup J_r$.

(3) Reversing a self-homotopy inverts its core: $-J_s = J_{s^{-1}}$.

It follows from these three properties that given two free self-homotopies J_s and J'_s of U_* representing the same element $s \in \mathrm{Stab}[U_*]$, we can form a based self-homotopy $J_1 = J_{ss^{-1}} := J_s \cup -J'_s$ of U_* representing the trivial element $1 \in \mathrm{Stab}[U_*]$, with $\mu_3(J_1) = \mu_3(J_s) - \mu_3(J'_s)$. Together with Lemma 2.10 we immediately get:

Lemma 3.1. *If J_s and J'_s are two free self-homotopies of U_* representing the same element $s \in \mathrm{Stab}[U_*]$, then $\mu_3(J_s) - \mu_3(J'_s) = 0 \in \mathbb{A}_{[U_*]}$.* □

As a result of Lemma 3.1, the element

$$U_s := [\mu_3(J_s)] \in \mathbb{A}_{[U_*]}$$

is well defined, and hence so is the affine action ${}^s a := U_s + sas^{-1}$ of $\mathrm{Stab}[U_*]$ on $\mathbb{A}_{[U_*]}$. This clarifies Definition 1.10 of the target of the free isotopy invariant $\mathrm{fq}_{U_*} \in \mathbb{A}_{U_*}$ as the quotient set of the based isotopy target $\mathrm{fq}_{[U_*]} \in \mathbb{A}_{[U_*]}$ under this action.

The following lemma illustrates how the affine action describes the effect of free self-homotopies on the self-intersection invariant.

Lemma 3.2. *If H is a homotopy from U_* to an embedding R, and J_s is a free self-homotopy of U_* representing $s \in \mathrm{Stab}[U_*]$, then the free homotopy $J_s \cup H$ from U to R satisfies*

$$\mu_3(J_s \cup H) = \mu_3(J_s) + s\mu_3(H)\, s^{-1} \in \mathbb{A}_{[U_*]}.$$

Proof. It is clear that each double point of the track of $J_s \cup H$ is either a double point of J_s or H. By our convention, the computation of $\mu_3(J_s \cup H)$ uses the whisker for U_* at the start of J_s. Thus, double point loops of H get conjugated by representatives of s while traversing J_s, so all the double-point group elements of $\mu_3(H)$ get conjugated by s. □

fq_{U_*} *is well defined.* It suffices to show that $\mathrm{fq}_{U_*}(R) = [\mu_3(H)] \in \mathbb{A}_{U_*}$ is independent of the choice of H. For H and H' two choices of free homotopies from U_* to R, the concatenation $J_s := H \cup -H'$ is a self-homotopy of U_* representing some $s \in \mathrm{Stab}[U_*]$, and by Lemma 3.2 we have

$$\mu_3(J_s) = \mu_3(H) - s\mu_3(H')\, s^{-1} \in \mathbb{A}.$$

So

$$\mu_3(H) = \mu_3(J_s) + s\mu_3(H')\, s^{-1} = {}^s(\mu_3(H')) \in \mathbb{A}_{[U_*]},$$

which implies $[\mu_3(H)] = [\mu_3(H')] \in \mathbb{A}_{U_*}$, and hence $\mathrm{fq}_{U_*}(R)$ is well defined.

fq_{U_*} *is injective.* If $\mathrm{fq}_{U_*}(R) = \mathrm{fq}_{U_*}(R')$, then by the definition of the target $\mathbb{A}_{U_*}$ there exist homotopies H and H' from U_* to R and R', respectively, such that

$$\mu_3(H') = U_s + s\mu_3(H)\, s^{-1} \in \mathbb{A}_{[U_*]}$$

for some $s \in \mathrm{Stab}[U_*]$.

Consider the homotopy $H'' := -H' \cup J_s \cup H$ from R' to R, where J_s is any self-homotopy of U_* representing s. Using the whisker on U_* in $-H' \cap J_s \subset H''$ we have the following computation in $\mathbb{A}_{[U_*]}$:

$$\begin{aligned}\mu_3(H'') &= \mu_3(-H') + \mu_3(J_s) + s\mu_3(H)\, s^{-1}\\ &= -\mu_3(H') + U_s + s\mu_3(H)\, s^{-1}\\ &= -(U_s + s\mu_3(H)\, s^{-1}) + U_s + s\mu_3(H)\, s^{-1}\\ &= 0.\end{aligned}$$

It follows from Lemma 2.11 that R is isotopic to R'.

fq_{U_*} *is surjective.* Surjectivity follows directly from Lemma 2.4.

3.3. *Examples.* Recall that Corollary 1.12 states that free isotopy classes of null-homotopic 2-spheres in N are in bijection with $\mathbb{A}_{U_*} = \mathbb{A}_{[U_*]}/\pi_1 N$, where the action is by conjugation.

Here we examine some examples of free isotopy classes of essential 2-spheres:

Example. Consider $U_* = S^2 \times \{p\} \subset N = S^2 \times M^3$.

Then $U_s = 0$ for all $s \in \mathrm{Stab}[U_*] = \pi_1 N$, because any self-homotopy J_s can be chosen to be a self-isotopy which moves p around a loop representing s while fixing the S^2-factor.

So the affine action has trivial translations (see Section 1.3) and free isotopy classes of spheres homotopic to U_* are in bijection with $\mathbb{A}_{U_*} = \mathbb{A}_{[U_*]}/\pi_1 M$ (with conjugation action).

Example. Consider again $U_* = S^2 \times \{p\} \subset N = S^2 \times M^3$.

Assume $[g,h] \neq 1 \in \pi_1 M \cong \pi_1 N$. If U^g_* is the result of doing a g-finger move on U_*, then $U^g_s = g - sgs^{-1}$ for each $s \in \mathrm{Stab}[U^g_*] = \pi_1 N$. Here $U^g_s = \mu_3(J_s)$ where J_s is a self-homotopy of U^g_* that undoes the g-finger move, then moves p around a loop representing s while fixing S^2, and then redoes the g-finger move. In particular,

$$U^g_h = g - hgh^{-1} \neq 0 \in \mathbb{A}_{[U^g_*]} = \mathbb{A}_{[U_*]}$$

if $\pi_3 M = 0$. (To see that $\mathbb{A}_{[U^g_*]} = \mathbb{A}_{[U_*]}$ if $\pi_3 M = 0$, observe that any representative of a generator of $\pi_3(S^2 \times M) \cong \pi_3 S^2$ is homotopic to a generic immersion contained in the product of S^2 with a 3-ball, and hence has only double points with trivial group element, implying that μ_3 vanishes.)

So in this case the affine action ${}^h a = (g - hgh^{-1}) + hah^{-1}$ defining $\mathbb{A}_{U^g_*}$ as a quotient of $\mathbb{A}_{[U^g_*]} = \mathbb{A}_{[U_*]}$ has nontrivial translations, illustrating how the target of the free isotopy invariant depends in general on the isotopy class of the fixed embedding, and not just on its homotopy class.

This suggests the following questions: When does a homotopy class of 2-spheres in N contain an isotopy class such that the corresponding affine action has trivial translations? Are stabilizers of elements of $\pi_2 N$ always represented by some embedded $S^2 \times S^1 \subset N \times S^1$?

4. A space level approach following Dax

In this section we reprove our two main results, Theorems 1.7 and 1.11, using a space level approach given in [Dax 1972] and [Kosanović et al. ≥ 2024].

4.1. *The relative homotopy group.* Following [Kosanović and Teichner 2024b] and [Dax 1972], we compute in [Kosanović et al. ≥ 2024] the relative homotopy group $\pi_{d-2\ell}(\mathrm{Map}(V,X), \mathrm{Emb}(V,X), U)$ for any ℓ-manifold V and d-manifold X, and a fixed embedding $U : V \hookrightarrow X$ taken as the basepoint. In our case of interest, $V = S^2$ and $X = N$ of dimension $d = 5$, the relevant result is as follows.

Theorem 4.1 [Kosanović et al. ≥ 2024]. *Let N be an oriented connected 5-manifold and $U_* : S^2 \hookrightarrow N$ a smooth based embedding. Then there are bijections*

$$\begin{array}{c} \mathsf{Dax} : \pi_1(\mathrm{Map}_*(S^2,N), \mathrm{Emb}_*(S^2,N), U_*) \\ \cong \Big\downarrow i^{\mathrm{rel}} \\ \pi_1(\mathrm{Map}(S^2,N), \mathrm{Emb}(S^2,N), U) \\ \cong \Big\downarrow \\ \mathbb{A} \end{array}$$

given on a class $[H]$ as the sum over double points of the associated group elements of the track of $H : I \to \mathrm{Map}(S^2,N)$, defined by $I \times S^2 \to I \times N$, $(t,v) \mapsto (t, H_t(v))$.

In other words, $\mathsf{Dax}([H]) = \mu_3(H)$ *is precisely the self-intersection invariant from* (2.2).

Remark 4.2. Using Lemma 2.4 one can define an explicit inverse of Dax. This is completely analogous to the realization map $\mathfrak{r}$ in [Kosanović and Teichner 2024b; Kosanović et al. ≥ 2024].

Our main square from Section 1.1 extends to a commutative diagram:

$$\begin{array}{ccc}
\pi_1(\mathrm{Map}_*(S^2,N),U_*) & \xrightarrow{\ i\ } & \pi_1(\mathrm{Map}(S^2,N),U) \\
\downarrow{\scriptstyle j_*} & & \downarrow{\scriptstyle j} \\
\pi_1(\mathrm{Map}_*,\mathrm{Emb}_*,U_*) & \xrightarrow{\ i^{\mathrm{rel}}\ } & \pi_1(\mathrm{Map},\mathrm{Emb},U) \\
\downarrow & & \downarrow \\
\pi_0\,\mathrm{Emb}_*(S^2,N) & \twoheadrightarrow & \pi_0\,\mathrm{Emb}(S^2,N) \\
\downarrow{\scriptstyle p_*} & & \downarrow{\scriptstyle p} \\
\pi_0\,\mathrm{Map}_*(S^2,N) & \twoheadrightarrow & \pi_0\,\mathrm{Map}(S^2,N)
\end{array} \tag{4.3}$$

The left column is the final part of the long exact sequence of the pair in the based case,

$$(\mathrm{Map}_*,\mathrm{Emb}_*) := (\mathrm{Map}_*(S^2,N),\mathrm{Emb}_*(S^2,N)),$$

whereas the right column is from the long exact sequence of the pair in the corresponding free case.

We use the following standard facts about homotopy groups of mapping spaces, see [Kosanović et al. ≥ 2024] for details.

Lemma 4.4. *There are isomorphisms*

$$\pi_k(\mathrm{Map}_*(S^2,N),U_*) \to \pi_{k+2}(N)$$

for $k \geq 0$, *and a bijection*

$$\pi_0\,\mathrm{Map}(S^2,N) \cong \pi_2 N/\{\alpha - g\alpha\}$$

for the usual action of $g \in \pi_1 N$ *on* $\alpha \in \pi_2 N$. *For any* $\beta \in \mathrm{Map}_*(S^2,N)$ *there is an exact sequence*

$$\pi_3 N \cong \pi_1(\mathrm{Map}_*(S^2,N),\beta) \xrightarrow{\ i\ } \pi_1(\mathrm{Map}(S^2,N),\beta) \xrightarrow{\ ev_e\ } \mathrm{Stab}\,\beta,$$

where

$$\mathrm{Stab}\,\beta := \{g \in \pi_1 N : g\beta = \beta \in \pi_2 N\},$$

and ev_e *is induced by the map* $\mathrm{Map}(S^2,N) \to N$ *given by* $f \mapsto f(e)$.

Combining (4.3) and Lemma 4.4 with Theorem 4.1, and denoting by $[U_*]$ the class of U_* in $\pi_3 N \cong \pi_1 \operatorname{Map}_*(S^2, N)$, we have the commutative diagram:

$$
\begin{array}{ccccc}
\pi_3 N & \xrightarrow{\quad i \quad} & \pi_1(\operatorname{Map}(S^2, N), U) & \xrightarrow{ev_e} \!\!\!\!\! \rightarrow & \operatorname{Stab}[U_*] \\
\big\downarrow{\scriptstyle j_*} & & \big\downarrow{\scriptstyle j} & & \\
\pi_1(\operatorname{Map}_*, \operatorname{Emb}_*, U_*) & \xrightarrow[\cong]{i^{\mathrm{rel}}} & \pi_1(\operatorname{Map}, \operatorname{Emb}, U) & \xrightarrow[\cong]{\mathsf{Dax}} & \mathbb{A} \\
\big\downarrow & & \big\downarrow & & \\
\pi_0 \operatorname{Emb}_*(S^2, N) & \longrightarrow\!\!\!\!\!\rightarrow & \pi_0 \operatorname{Emb}(S^2, N) & & \\
\big\downarrow\!\!\!\!\downarrow{\scriptstyle p_*} & & \big\downarrow\!\!\!\!\downarrow{\scriptstyle p} & & \\
\pi_2 N & \longrightarrow\!\!\!\!\!\rightarrow & \pi_0 \operatorname{Map}(S^2, N) & &
\end{array}
\tag{4.5}
$$

which will imply desired results, as explained next.

4.2. ***The proofs.*** The following recovers Theorem 1.7.

Theorem 4.6. *There is a short exact sequence of sets*

$$\mathbb{A}/(\mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_*(\pi_3 N)) \rightarrowtail \pi_0 \operatorname{Emb}_*(S^2, N) \xrightarrow{p_*} \!\!\!\!\! \rightarrow \pi_0 \operatorname{Map}_*(S^2, N) \cong \pi_2 N$$

and $\mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_* = \phi_{[U_*]}$ *from* (1.4) *of* Section 1.2.

Proof. From diagram (4.5) we have $\ker(p_*) = \operatorname{coker}(j_*) \cong \operatorname{coker}(\mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_*)$, so it only remains to identify the last homomorphism. And indeed, for a class $A \in \pi_3 N$ the element $j_*(A) : I \to \operatorname{Map}_*$ is a self-homotopy of U that represents A and $\mathsf{Dax}(j_*(A)) = \mu_3(j_*(A))$ by definition. Now, arguing as in the proof of Lemma 2.10 we see that the track of $j_*(A)$ has $\mu_3(j_*(A)) = \mu_3(A) + \lambda_N(A, U_*)$, therefore $\mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_*(A) = \phi_{[U_*]}(A)$ as desired. □

Similarly, the following recovers Theorem 1.11.

Theorem 4.7. *There is a short exact sequence of sets*

$$\big(\mathbb{A}/(\phi_{[U_*]}(\pi_3 N))\big)_{s \mapsto {}^s a} \rightarrowtail \pi_0 \operatorname{Emb}(S^2, N) \xrightarrow{p} \!\!\!\!\! \rightarrow \pi_0 \operatorname{Map}(S^2, N),$$

where on the left we take the quotient by the action $s \mapsto {}^s a$ *of* $\operatorname{Stab}[U_*]$ *from* (aff) *of* Section 1.3.

Proof. From diagram (4.5) we have $\ker(p) = \operatorname{coker}(j) \cong \operatorname{coker}(\mathsf{Dax} \circ j)$. Using the leftmost column we can compute $\operatorname{coker}(\mathsf{Dax} \circ j)$ in two steps:

(1) First take the cokernel of $\mathsf{Dax} \circ j \circ i$.

(2) Then mod out the induced action of $\operatorname{Stab}[U_*]$, using any section of ev_e.

Note that the action in (2) is well defined, and the set of coinvariants is independent of the section, since in $\mathrm{coker}(j \circ i)$ we have modded out $\ker(ev_e)$.

For (1), we simply note that $\mathsf{Dax} \circ j \circ i = \mathsf{Dax} \circ i^{\mathrm{rel}} \circ j_*$ by the commutativity of the leftmost square in (4.5), and this is equal to $\phi_{[U_*]}$ by Theorem 4.6.

For (2), to compute the action, we pick any section; by definition, this sends $s \in \mathrm{Stab}[U_*]$ to any

$$J_s \in \pi_1(\mathrm{Map}(S^2, N), U),$$

which we view as a free self-homotopy of U_*, for which $ev_e(J_s) = J_s(-, e)$ represents s.

Then $s \in \mathrm{Stab}[U_*]$ acts by sending $a = \mathsf{Dax}(H)$ to $\mathsf{Dax}(J_s \cup H)$. Since we have $\mathsf{Dax}(J_s \cup H) = \mathsf{Dax}(J_s) + s\mathsf{Dax}(H)\, s^{-1}$ by Lemma 3.2 (where μ_3 notation was used in place of Dax), we see that the action of s on a is given as claimed by

$$\mathsf{Dax}(J_s) + sas^{-1} = \mu_3(J_s) + sas^{-1} = {}^s a. \qquad \square$$

Acknowledgements

Kosanović was partially supported by the Science Fund of the Republic of Serbia, grant 7749891, Graphical Languages, GWORDS. Schneiderman was supported by a Simons Foundation *Collaboration Grant for Mathematicians*. All the authors thank the Max Planck Institute for Mathematics in Bonn, as well as SwissMAP Research Station (SRS) in Les Diablerets, for support during this project. We thank the referee for valuable comments.

References

[Dax 1972] J.-P. Dax, "Étude homotopique des espaces de plongements", *Ann. Sci. École Norm. Sup.* (4) **5** (1972), 303–377. MR Zbl

[Freedman and Quinn 1990] M. H. Freedman and F. Quinn, *Topology of 4-manifolds*, Princeton Mathematical Series **39**, Princeton University Press, Princeton, NJ, 1990. MR Zbl

[Gabai 2020] D. Gabai, "The 4-dimensional light bulb theorem", *J. Amer. Math. Soc.* **33**:3 (2020), 609–652. MR Zbl

[Gabai 2021] D. Gabai, "Self-referential discs and the light bulb lemma", *Comment. Math. Helv.* **96**:3 (2021), 483–513. MR Zbl

[Haefliger 1961a] A. Haefliger, "Differentiable imbeddings", *Bull. Amer. Math. Soc.* **67** (1961), 109–112. MR Zbl

[Haefliger 1961b] A. Haefliger, "Plongements différentiables de variétés dans variétés", *Comment. Math. Helv.* **36** (1961), 47–82. MR Zbl

[Hudson 1970] J. F. P. Hudson, "Concordance, isotopy, and diffeotopy", *Ann. of Math.* (2) **91** (1970), 425–448. MR Zbl

[Kosanović and Teichner 2024a] D. Kosanović and P. Teichner, "A new approach to light bulb tricks: disks in 4-manifolds", *Duke Math. J.* **173**:4 (2024), 673–721. MR Zbl

[Kosanović and Teichner 2024b] D. Kosanović and P. Teichner, "A space level light bulb theorem in all dimensions", *Comment. Math. Helv.* **99**:4 (2024), 799–860.

[Kosanović et al. $\geq$ 2024] D. Kosanović, R. Schneiderman, and P. Teichner, "Metastable homotopy groups of embedding spaces revisited", in preparation.

[Larmore 1978] L. L. Larmore, "Isotopy groups", *Trans. Amer. Math. Soc.* **239** (1978), 67–97. MR Zbl

[Milnor 1965] J. Milnor, *Lectures on the h-cobordism theorem*, Princeton University Press, Princeton, NJ, 1965. MR Zbl

[Schneiderman and Teichner 2022] R. Schneiderman and P. Teichner, "Homotopy versus isotopy: spheres with duals in 4-manifolds", *Duke Math. J.* **171**:2 (2022), 273–325. MR Zbl

[Whitney 1944] H. Whitney, "The self-intersections of a smooth n-manifold in $2n$-space", *Ann. of Math.* (2) **45** (1944), 220–246. MR Zbl

Received April 11, 2024. Revised July 9, 2024.

DANICA KOSANOVIĆ
DEPARTMENT OF MATHEMATICS
ETH ZÜRICH
ZÜRICH
SWITZERLAND
danica.kosanovic@math.ethz.ch

ROB SCHNEIDERMAN
DEPARTMENT OF MATHEMATICS
LEHMAN COLLEGE
CITY UNIVERSITY OF NEW YORK
BRONX, NY
UNITED STATES
robert.schneiderman@lehman.cuny.edu

PETER TEICHNER
MAX-PLANCK-INSTITUT FÜR MATHEMATIK
BONN
GERMANY
teichner@mac.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.219

A NEW CONVERGENCE THEOREM FOR MEAN CURVATURE FLOW OF HYPERSURFACES IN QUATERNIONIC PROJECTIVE SPACES

SHIYANG LI, HONGWEI XU AND ENTAO ZHAO

We investigate the smooth convergence of the mean curvature flow of hypersurfaces in the quaternionic projective spaces. We prove that if the initial hypersurface satisfies a new nonlinear curvature pinching condition, then the mean curvature flow converges smoothly to a round point in finite time. Our result improves a smooth convergence theorem due to Pipoli and Sinestrari (2017).

1. Introduction

There are many famous geometric evolution equations, such as the Ricci flow, the mean curvature flow and others. Huisken [9] studied the mean curvature flow from the perspective of partial differential equations, and he proved that convex hypersurface in the Euclidean space converges to a round point along the flow. Afterwards, Huisken [10; 11] obtained convergence results for mean curvature flow of convex hypersurfaces in Riemannian manifolds and pinched hypersurfaces in spheres. Following the argument of Huisken [9], Andrews and Baker [1] proved a convergence theorem for the mean curvature flow of closed submanifolds satisfying a suitable pinching condition in the Euclidean space. Later, Baker [2], Liu et al. [20] proved sharp convergence theorems for the mean curvature flow in the spheres and the hyperbolic spaces, respectively. Liu, Xu and Zhao [19] studied the mean curvature flow of arbitrary codimensional submanifolds in the Riemannian manifold and proved a smooth convergence theorem. Lei and Xu [15] verified an optimal convergence theorem for the mean curvature flow of submanifolds in hyperbolic spaces, which implies the first optimal differentiable sphere theorem for submanifolds with positive Ricci curvature. It should be remarked that an optimal topological sphere theorem for complete submanifolds with positive Ricci curvature in a space form with nonnegative curvature has been proved previously by Shiohama and Xu [26]. Lei and Xu [15] also proved sharp convergence theorems for the mean curvature

MSC2020: primary 53E10; secondary 53C40.
Keywords: mean curvature flow, convergence theorem, curvature pinching, real hypersurfaces, quaternionic projective spaces.

flow of submanifolds in the sphere [13; 14], which also improve the convergence theorems due to Huisken [10] and Baker [2]. See [16; 18; 21] for recent progress in the smooth convergence theory for mean curvature flow of arbitrary codimensions. As consequences of these smooth convergence theorems, the submanifolds satisfying the initial curvature conditions are diffeomorphic to the standard sphere. We remark that some of these differentiable sphere theorems are also proved by using the Ricci flow, which has proven to be a very useful tool in understanding the topology of Riemannian manifolds, see [3; 4; 6; 7; 8; 22; 23; 24; 27; 28].

Pipoli and Sinestrari [25] obtained a convergence theorem for mean curvature flow of small codimension in the complex projective spaces. Later, Lei and Xu [17] investigated the smooth convergence of mean curvature flow of arbitrary codimensional submanifolds in the complex projective spaces, which improved and extended the convergence theorem due to Pipoli and Sinestrari [25]. In this paper, we investigate the mean curvature flow in the quaternionic projective spaces. We mainly consider the codimension-one case.

Let M be an n-dimensional closed manifold, and let $F : M^n \times [0, T) \to N^{n+1}$ be a one-parameter family of smooth hypersurfaces immersed in a Riemannian manifold (N, h). We say that $M_t = F_t(M)$ is a solution to the mean curvature flow if F_t satisfies

$$\begin{cases} \frac{\partial}{\partial t} F = -H\nu, \\ F(\,\cdot\,, 0) = F_0(\,\cdot\,), \end{cases} \tag{1-1}$$

where $F_t(\,\cdot\,) = F(\,\cdot\,, t)$, H and ν are the mean curvature of M and the unit outward normal vector of M respectively, such that $\vec{H} = -H\nu$ is the mean curvature vector of M.

Pipoli and Sinestrari [25] obtained a convergence theorem for the mean curvature flow of hypersurfaces in the quaternionic projective spaces, and the proof is the same as their convergence theorem for mean curvature flow of hypersurfaces in the complex projective spaces.

Theorem 1.1 [25]. *Let M^n $(n \geqslant 11)$ be a closed real hypersurface in quaternionic projective space $\mathbb{QP}^{(n+1)/4}(4)$, and M_t be the mean curvature flow starting from M. Assume that M satisfies the following pinching condition*:

$$|h|^2 < \frac{1}{n-1} H^2 + 2.$$

Then the flow has a smooth solution on the maximal time interval $[0, T)$ with $T < \infty$. Moreover, the pinching condition is preserved and M_t collapses to a round point as $t \to T$.

We note that here and in the remaining part of the paper, $n = 4m - 1$ for $m \geqslant 2$. The aim of the present paper is to prove the following refinement of Theorem 1.1.

Theorem 1.2. *Let M^n be an $n(\geqslant 7)$-dimensional closed real hypersurface in quaternionic projective space $\mathbb{QP}^{(n+1)/4}(4)$, and M_t be the mean curvature flow starting from M. Assume that M satisfies the following pinching condition*:

$$|h|^2 < \varphi(H^2).$$

Then the flow has a smooth solution on the maximal time interval $[0, T)$ with $T < \infty$. Moreover, the pinching condition is preserved and M_t collapses to a round point as $t \to T$.

In Theorem 1.2, $\varphi(H^2)$ is given by

$$\varphi(H^2) = 2 + a_n + \left(b_n + \frac{1}{n-1}\right) H^2 - \sqrt{b_n^2 H^4 + 2a_n b_n H^2}, \tag{1-2}$$

where

$$a_n = \sqrt{8(n-5)(n-1)\, b_n}, \quad b_n = \min\left\{\frac{n-5}{8(n-1)}, \frac{2n-5}{(n+2)(n-1)}\right\}.$$

Remark 1.3. By a computation, we have $\varphi(x) > \frac{x}{n-1} + 2$ for $x \geqslant 0$. So, Theorem 1.2 is an improvement of Theorem 1.1. Furthermore, we have $\varphi(x) \geqslant 4\sqrt{n-1} - 6$ for $7 \leqslant n \leqslant 17$, and $\varphi(x) > 2 + \frac{8\sqrt{2}}{5}\sqrt{n-5}$ for $n \geqslant 18$.

It is well known that $\mathbb{QP}^1$ is just the round sphere. By [11; 14], the similar smooth convergence theorem holds for mean curvature flow of closed hypersurfaces in $\mathbb{QP}^1$.

By Theorem 1.2, we have:

Corollary 1.4. *Let M^n be an $n(\geqslant 7)$-dimensional closed real hypersurface in quaternionic projective space $\mathbb{QP}^{(n+1)/4}(4)$. If $|h|^2 < \varphi(H^2)$, then M is diffeomorphic to the standard sphere.*

The rest of the paper is organized as follows. In Section 2, we introduce some notations, formulas and basic equations in submanifold theory, and prove a gradient inequality involving the second fundamental form and the mean curvature for hypersurfaces in the quaternionic projective spaces. We also recall some evolution equations along the mean curvature flow in this section. In Section 3, we show that the pinching condition $|h|^2 < \varphi(H^2)$ is preserved along the mean curvature flow. We also derive an evolution inequality of

$$f_\sigma = \frac{|\mathring{h}|^2}{(\varphi - H^2/n)^{1-\sigma}}.$$

A pinching estimate for the traceless second fundamental form is obtained in Section 4. We give an estimate of the gradient of the mean curvature in Section 5, which is used to compare the mean curvature at different points. In Section 6, we show that the hypersurface shrinks to a single round point in finite time.

2. Notations and formulas

Let $\mathbb{QP}^m$ be the m-dimensional quaternionic projective space with the Fubini–Study metric g_{FS}. Let J_{k_0}, $k_0 = 1, 2, 3$ be complex structures on $\mathbb{QP}^m$. We denote by $\overline{\nabla}$ the Levi–Civita connection of $(\mathbb{QP}^m, g_{\mathrm{FS}})$. Since the Fubini–Study metric is a Kähler metric, we have $\overline{\nabla} J_{k_0} = 0$ for $k_0 = 1, 2, 3$. The curvature tensor $\bar{R}$ of $\mathbb{QP}^m$ can be written as

$$\begin{aligned}
&\bar{R}(X, Y, Z, W)\\
&= \langle X, Z\rangle\langle Y, W\rangle - \langle X, W\rangle\langle Y, Z\rangle\\
&\qquad + \sum_{k_0=1}^{3}\big(\langle X, J_{k_0}Z\rangle\langle Y, J_{k_0}W\rangle - \langle X, J_{k_0}W\rangle\langle Y, J_{k_0}Z\rangle + 2\langle X, J_{k_0}Y\rangle\langle Z, J_{k_0}W\rangle\big)
\end{aligned}$$

and J_{k_0}, $k_0 = 1, 2, 3$ satisfies

$$J_{k_0}^2 = -\,\mathrm{Id}, \quad J_1 J_2 = -J_2 J_1 = J_3, \quad J_1 J_3 = -J_3 J_1 = -J_2, \quad J_2 J_3 = -J_2 J_3 = J_1.$$

Let (M^n, g) be an n-dimensional Riemannian submanifold in $(\mathbb{QP}^m, g_{\mathrm{FS}})$. Let q be its codimension, i.e., $n + q = 4m$. At a point $p \in M$, let $T_p M$ and $N_p M$ be the tangent space and normal space, respectively. For a vector in $T_p M \oplus N_p M$, we denote by $(\,\cdot\,)^T$ and $(\,\cdot\,)^N$ its projections onto $T_p M$ and $N_p M$, respectively. We use the symbols ∇ and $\nabla^\perp$ to represent the connections of tangent bundle TM and normal bundle NM. Denote by $\Gamma(E)$ the space of smooth sections of a vector bundle E. For $X, Y \in \Gamma(TM)$, $\xi \in \Gamma(NM)$, the connections ∇ and $\nabla^\perp$ are given by $\nabla_X Y = (\overline{\nabla}_X Y)^T$ and $\nabla^\perp_X \xi = (\overline{\nabla}_X \xi)^N$. The second fundamental form of M is defined as $h(X, Y) = (\overline{\nabla}_X Y)^N$.

We mainly consider the codimension-one case. Throughout this paper, we shall make the following convention on indices:

$$1 \leqslant A, B, C, \dots \leqslant n+1, \quad 1 \leqslant i, j, k, \dots \leqslant n.$$

We choose a local orthonormal frame $\{e_i\}$ for the tangent bundle and let $\nu = e_{n+1}$ be the unit normal vector field. Denote by $\{\omega^i\}$ the dual frame of $\{e_i\}$. Let h and H denote the second fundamental form and the mean curvature given by

$$h = \sum_{i,j} h_{ij}\,\omega^i \otimes \omega^j \quad \text{and} \quad H = \sum_i h_{ii}.$$

Let $\mathring{h} = h - \frac{1}{n} H g$ be the traceless second fundamental form. We have the relations

$$|\mathring{h}|^2 = |h|^2 - \frac{1}{n} H^2, \quad |\nabla \mathring{h}|^2 = |\nabla h|^2 - \frac{1}{n}|\nabla H|^2.$$

Setting $J^{(k_0)}_{AB} = \langle e_A, J_{k_0} e_B\rangle$ for $k_0 = 1, 2, 3$, we have

$$J^{(k_0)}_{AB} = -J^{(k_0)}_{BA}, \quad \sum_B J^{(k_0)}_{AB} J^{(k_0)}_{BC} = J^{(k_0)}_{AC}, \quad \sum_B J^{(1)}_{AB} J^{(2)}_{BC} = J^{(3)}_{AC}.$$

Similarly, we have

$$\sum_B J^{(1)}_{AB}\,J^{(3)}_{BC} = -J^{(2)}_{AC}, \quad \sum_B J^{(2)}_{AB}\,J^{(3)}_{BC} = J^{(1)}_{AC}.$$

Also, $J^{(k_0)}_{AA} = 0$ for any A and k_0.

Let h_{ijk} and H_i denote the components of ∇h and ∇H, the covariant differentiations of h and H, respectively. We have the following sharp gradient inequality (see Remark 2.2).

Lemma 2.1. *For a hypersurface in* $\mathbb{QP}^{(n+1)/4}$, *we have*

$$|\nabla h|^2 \geqslant \frac{3}{n+2}|\nabla H|^2 + 6(n-3).$$

Proof. Set $S = \sum S_{ijk}\,\omega^i \otimes \omega^j \otimes \omega^k$, where $S_{ijk} = \frac{1}{3}(h_{ijk} + h_{jki} + h_{kij})$. Then S_{ijk} is totally symmetric for i, j, k. Using the same technique as in Lemma 2.2 in [9], we have

$$|S|^2 \geqslant \frac{3}{n+2}\sum_i\Big(\sum_k S_{kki}\Big)^2.$$

By the Codazzi equation, we have

$$\begin{aligned}\sum_k S_{kki} &= \tfrac{1}{3}\sum_k (h_{ikk} + h_{kki} + h_{kik})\\ &= \tfrac{1}{3}\sum_k (h_{kki} + 2h_{kik})\\ &= \tfrac{1}{3}\sum_k (h_{kki} + 2h_{kki} - 2\bar{R}_{n+1kik}) = H_i - \tfrac{2}{3}\sum_k \bar{R}_{n+1kik}.\end{aligned}$$

As

$$\bar{R}_{n+1kik} = \sum_{k_0=1}^{3}(J^{(k_0)}_{n+1i}\,J^{(k_0)}_{kk} - J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki} + 2J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ik}),$$

one has

$$\begin{aligned}-\tfrac{2}{3}\sum_k \bar{R}_{n+1kik} &= -\tfrac{2}{3}\sum_k\sum_{k_0=1}^{3}(J^{(k_0)}_{n+1i}\,J^{(k_0)}_{kk} - J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki} + 2J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ik})\\ &= -\tfrac{2}{3}\sum_k\sum_{k_0=1}^{3}(-J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki} - 2J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki})\\ &= 2\sum_{k_0=1}^{3}\sum_k (J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki}).\end{aligned}$$

Then we get

$$\sum_k S_{kki} = H_i + 2\sum_{k_0=1}^{3}\sum_k (J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki}).$$

This implies

$$\Big(\sum_k S_{kki}\Big)^2 = (H_i)^2 + 4\sum_{k_0=1}^{3}\sum_k H_i(J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki}) + 4\Big[\sum_{k_0=1}^{3}\sum_k (J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ki})\Big]^2.$$

Since

$$4\sum_{k_0=1}^{3}\Big[\sum_{i,k} H_i\, J^{(k_0)}_{n+1k}\, J^{(k_0)}_{ki}\Big]+4\sum_i\Big[\sum_{k_0=1}^{3}\sum_k J^{(k_0)}_{n+1k}\, J^{(k_0)}_{ki}\Big]^2$$
$$=4\sum_{k_0=1}^{3}\Big[\sum_i H_i\sum_A J^{(k_0)}_{n+1A}\, J^{(k_0)}_{Ai}\Big]+4\sum_i\Big[\sum_{k_0=1}^{3}\sum_A J^{(k_0)}_{n+1A}\, J^{(k_0)}_{Ai}\Big]^2$$
$$=4\sum_{k_0=1}^{3}\Big[\sum_i H_i\,\delta_{n+1i}\Big]+4\sum_i\Big[\sum_{k_0=1}^{3}\delta_{n+1i}\Big]^2=0,$$

one has

$$|S|^2\geqslant\frac{3}{n+2}|\nabla H|^2. \tag{2-1}$$

On the other hand, by the Codazzi equation, we have

$$\begin{aligned}|S|^2=\sum(S_{ijk})^2&=\tfrac{1}{9}\sum(h_{ikk}+h_{kki}+h_{kik})^2\\&=\tfrac{1}{3}\sum(h_{ijk})^2+\tfrac{2}{3}\sum h_{ijk}\,h_{ikj}\\&=\tfrac{1}{3}\sum(h_{ijk})^2+\tfrac{2}{3}\sum h_{ijk}(h_{ijk}+\bar R_{n+1ijk})\\&=\sum(h_{ijk})^2+\tfrac{2}{3}\sum\bar R_{n+1jki}\,\bar R_{n+1ijk}\\&=|\nabla h|^2+\tfrac{2}{3}\sum\bar R_{n+1jki}\,\bar R_{n+1ijk}.\end{aligned}$$

Since

$$\bar R_{ABCD}=\delta_{AC}\,\delta_{BD}-\delta_{AD}\,\delta_{BC}+\sum_{k_0=1}^{3}(J^{(k_0)}_{AC}J^{(k_0)}_{BD}-J^{(k_0)}_{AD}J^{(k_0)}_{BC}+2J^{(k_0)}_{AB}J^{(k_0)}_{CD}),$$

one has

$$\begin{aligned}\sum\bar R_{n+1jki}\,\bar R_{n+1ijk}=\sum_{i,j,k}&\Big[\sum_{k_0=1}^{3}(J^{(k_0)}_{n+1k}\,J^{(k_0)}_{ji}-J^{(k_0)}_{n+1i}\,J^{(k_0)}_{jk}+2J^{(k_0)}_{n+1j}\,J^{(k_0)}_{ki})\Big]\\&\times\Big[\sum_{l_0=1}^{3}(J^{(l_0)}_{n+1j}\,J^{(l_0)}_{ik}-J^{(l_0)}_{n+1k}\,J^{(l_0)}_{ij}+2J^{(l_0)}_{n+1i}\,J^{(l_0)}_{jk})\Big].\end{aligned}$$

For each k_0, according to the special property of matrix $(J^{(k_0)}_{AB})$, by direct computation we have

$$\sum_k(J^{(k_0)}_{n+1k})^2=-\sum_k J^{(k_0)}_{n+1k}\,J^{(k_0)}_{kn+1}=-\sum_A J^{(k_0)}_{n+1A}\,J^{(k_0)}_{An+1}=\delta_{(n+1)(n+1)}=1,$$

and

$$\sum_k J^{(k_0)}_{n+1k}\,J^{(l_0)}_{Bk}=-\sum_k J^{(k_0)}_{n+1k}\,J^{(l_0)}_{kB}=-\sum_A J^{(k_0)}_{n+1A}\,J^{(l_0)}_{AB}=\pm J^{(j_0)}_{n+1B},$$

where $\pm$ depends on j_0, k_0, l_0. By some computations, we obtain

$$\sum\bar R_{n+1jki}\,\bar R_{n+1ijk}=-9(n-3).$$

Hence

$$|S|^2=|\nabla h|^2+\tfrac{2}{3}(-9(n-3))=|\nabla h|^2-6(n-3).$$

Combining this with (2-1) implies

$$|\nabla h|^2 \geqslant \frac{3}{n+2}|\nabla H|^2 + 6(n-3). \qquad \square$$

Remark 2.2. For hypersurface M^{4m-1} in $\mathbb{QP}^m$, one has

$$|\nabla h|^2 \geqslant \frac{3}{4m+1}|\nabla H|^2 + 24(m-1).$$

In particular, one has $|\nabla h|^2 \geqslant 24(m-1)$, which has been proved previously by Dong [5]. Dong also proved that a real hypersurface satisfying $|\nabla h|^2 = 24(m-1)$ is one of the generalized equators $M^Q_{p,q}$. See, e.g., [5; 12] for the detailed construction of generalized equators. From this we see that our gradient inequality is sharp.

Let $F : M \times [0, T) \to \mathbb{QP}^{(n+1)/4}$ be a mean curvature flow of hypersurface in the quaternionic projective space $\mathbb{QP}^{(n+1)/4}$. Set $M_t = F_t(M)$, where $F_t(\cdot) = F(\cdot, t)$. Following [1; 25], we have the evolution equations.

Lemma 2.3. *For mean curvature flow $F : M \times [0, T) \to \mathbb{QP}^{(n+1)/4}$, we have*

$$\frac{\partial}{\partial t}|h|^2 = \Delta|h|^2 - 2|\nabla h|^2 - 2n|h|^2 + 2|h|^4 + 18|h|^2 + 4H^2 + 12S_1,$$
$$\frac{\partial}{\partial t}H^2 = \Delta H^2 - 2|\nabla H|^2 + 2H^2(|h|^2 + n + 9),$$

where

$$S_1 = \sum_{k_0=1}^{3} \sum_{i,j,k,l} (\mathring{h}_{ij}\, \mathring{h}_{kl}\, J^{(k_0)}_{il}\, J^{(k_0)}_{jk} - \mathring{h}_{ik}\, \mathring{h}_{jk}\, J^{(k_0)}_{il}\, J^{(k_0)}_{jl}).$$

To do computations involving $(J^{(k_0)}_{AB})$ for $k_0 = 1, 2, 3$, the following well-known property of skew-symmetric matrix will be important.

Proposition 2.4. *Let A be a real skew-symmetric matrix. Then there exists an orthogonal matrix C, such that $C^{-1}AC$ takes the following form*:

$$\begin{pmatrix} 0 & \lambda_1 & & & & & & \\ -\lambda_1 & 0 & & & & & & \\ & & 0 & \lambda_3 & & & & \\ & & -\lambda_3 & 0 & & & & \\ & & & & 0 & \lambda_5 & & \\ & & & & -\lambda_5 & 0 & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \end{pmatrix}. \tag{2-2}$$

We use a notation

$$\tilde{i} = \begin{cases} i+1, & i \text{ is odd}, \\ i-1, & i \text{ is even}. \end{cases}$$

If a matrix (a_{ij}) takes the form as (2-2), then $a_{ij} = 0$, for all $j \neq \tilde{i}$.

3. Preservation of curvature pinching

For each fixed $k_0 \in \{1, 2, 3\}$, we choose a local orthonormal frame $\{e_i\}$ such that the matrix $(J_{ij}^{(k_0)})$ takes the form of (2-2). In fact, let $\{\epsilon_1, \dots, \epsilon_n, \epsilon_{n+1}\}$ be a local orthonormal frame on $\mathbb{QP}^{(n+1)/4}$ such that $\epsilon_1, \dots, \epsilon_n$ are tangent to M and ϵ_{n+1} is normal to M. Let $\tilde{J}_{AB}^{(k_0)} = \langle \epsilon_A, J^{(k_0)} \epsilon_B \rangle$. Since $(\tilde{J}_{ij})_{n\times n}$ is antisymmetric and n is odd, there is an orthonormal matrix $C = (c_{ij})_{n\times n}$, where c_{ij}'s are local functions, such that

$$(c_{ij}^{-1} \tilde{J}_{jk}^{(k_0)} c_{kl})_{n\times n} = \begin{pmatrix} 0 & \lambda_1 & & \\ -\lambda_1 & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}.$$

Here $(c_{ij}^{-1})_{n\times n} = (c_{ij})_{n\times n}^{-1}$. Set $e_i = \sum_{j=1}^n c_{ij}^{-1} \epsilon_j,\ e_{n+1} = \epsilon_{n+1}$. Then

$$\begin{aligned} J_{ij}^{(k_0)} &= \langle e_i, J^{(k_0)} e_j \rangle \\ &= \Big\langle \sum_k c_{ik}^{-1} \epsilon_k, J^{(k_0)} \Big(\sum_l c_{jl}^{-1} \epsilon_l \Big) \Big\rangle = \sum_{k,l} c_{ik}^{-1} \tilde{J}_{kl}^{(k_0)} c_{jl}^{-1} = \sum_{k,l} c_{ik}^{-1} \tilde{J}_{kl}^{(k_0)} c_{lj}. \end{aligned}$$

This implies

$$(J_{ij}^{(k_0)})_{n\times n} = \begin{pmatrix} 0 & \lambda_1 & & \\ -\lambda_1 & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}.$$

Thus we have

$$\begin{aligned} \sum_{i,j,k,l} (\mathring{h}_{ij} \mathring{h}_{kl} J_{il}^{(k_0)} J_{jk}^{(k_0)} - \mathring{h}_{ik} \mathring{h}_{kj} J_{il}^{(k_0)} J_{jl}^{(k_0)}) &= \sum_{i,k} \big(-\mathring{h}_{i\tilde{k}} \mathring{h}_{k\tilde{i}} J_{i\tilde{i}}^{(k_0)} J_{k\tilde{k}}^{(k_0)} - (\mathring{h}_{\tilde{i}k} J_{i\tilde{i}}^{(k_0)})^2 \big) \\ &= -\tfrac{1}{2} \sum_{i,k} (\mathring{h}_{\tilde{i}k} J_{i\tilde{i}}^{(k_0)} + \mathring{h}_{i\tilde{k}} J_{k\tilde{k}}^{(k_0)})^2 \\ &\leqslant 0. \end{aligned}$$

Therefore,

$$S_1 = \sum_{k_0=1}^{3} \Big[\sum_{i,j,k,l} (\mathring{h}_{ij} \mathring{h}_{kl} J_{il}^{(k_0)} J_{jk}^{(k_0)} - \mathring{h}_{ik} \mathring{h}_{jk} J_{il}^{(k_0)} J_{jl}^{(k_0)}) \Big] \leqslant 0.$$

So we get from Lemma 2.3 that

$$\frac{\partial}{\partial t} |\mathring{h}|^2 \leqslant \Delta |\mathring{h}|^2 - 2|\nabla \mathring{h}|^2 + 2|\mathring{h}|^2 (|h|^2 - n + 9). \tag{3-1}$$

For a real number $\varepsilon \in (0, 1)$, by the definition of φ, we define $\varphi_\varepsilon : [0, +\infty) \to \mathbb{R}$ by

$$\varphi_\varepsilon(x) = d_\varepsilon + c_\varepsilon x - \sqrt{b^2 x^2 + 2abx + e}, \tag{3-2}$$

where $a = \sqrt{8(n-5)(n-1)\, b}$, $b = \min\big\{ \frac{n-5}{8(n-1)}, \frac{2n-5}{(n+2)(n-1)} \big\}$, $c_\varepsilon = b + \frac{1}{n-1+\varepsilon}$, $d_\varepsilon = 2 - 2\varepsilon + a$, $e = \sqrt{\varepsilon}$. We define $\varphi = \varphi_0$.

Lemma 3.1. *The function φ has the following property.*

(i) $\frac{x}{n-1}+2<\varphi(x)<\frac{x}{n-1}+n$,

(ii) $\varphi(x)>4\sqrt{n-1}-6$ *if* $n=7, 11, 15$ *and* $\varphi(x)>2+\frac{8\sqrt{2}}{5}\sqrt{n-5}$ *if* $n=4m-1$, $m\geqslant 5$.

Proof. By direct computations, we get

$$\varphi'(x)=c_0-\frac{bx+a}{\sqrt{x^2+2ax/b}},\qquad \varphi''(x)=\frac{a^2}{b(x^2+2ax/b)^{3/2}}.$$

Since $\left(\varphi(x)-\frac{x}{n-1}\right)''=\varphi(x)''>0$ and $\lim_{x\to\infty}\varphi'(x)=\frac{1}{n-1}$, we have $\varphi'(x)<\frac{1}{n-1}$. Hence we get

$$2=\lim_{x\to\infty}\left(\varphi(x)-\frac{x}{n-1}\right)<\varphi(x)-\frac{x}{n-1}\leqslant\varphi(0)=2+a<n.$$

We figure out that

$$\min_{x\geqslant 0}\varphi(x)=\varphi\left(\frac{ac_0}{b\sqrt{c_0^2-b^2}}-\frac{a}{b}\right)=d_0-\frac{ac_0}{b}+\frac{a}{b}\sqrt{c_0^2-b^2}.$$

If $n=7, 11, 15$, we have $\min_{x\geqslant 0}\varphi(x)=4\sqrt{n-1}-6$. If $n=4m-1,\ m\geqslant 5$, then we have

$$\min_{x\geqslant 0}\varphi(x)=2+\sqrt{\frac{8(n-5)}{2n-5}}(\sqrt{5n-8}-\sqrt{n+2})>2+\frac{8\sqrt{2}}{5}\sqrt{n-5}.\qquad\square$$

Let $\mathring{\varphi}_\varepsilon=\varphi_\varepsilon-\frac{1}{n}x$. We will prove the following lemma.

Lemma 3.2. *For sufficiently small ε, the function $\mathring{\varphi}_\varepsilon$ satisfies*

(i) $\mathring{\varphi}_\varepsilon'+2x\mathring{\varphi}_\varepsilon''<\frac{2(n-1)}{n(n+2)}$,

(ii) $\mathring{\varphi}_\varepsilon(x)(\varphi_\varepsilon(x)-n+9)-x\mathring{\varphi}_\varepsilon'(x)(\varphi_\varepsilon(x)+n+9)<6(n-3)$,

(iii) $\mathring{\varphi}_\varepsilon(x)-x\mathring{\varphi}_\varepsilon'(x)>1$.

Proof. By direct computations, we have

$$\mathring{\varphi}_\varepsilon'=c_\varepsilon-\frac{1}{n}-\frac{b^2x+ab}{\sqrt{b^2x^2+2abx+e}},$$
$$\mathring{\varphi}_\varepsilon''=\frac{(b^2x+ab)^2-b^2(b^2x^2+2abx+e)}{(b^2x^2+2abx+e)^{3/2}},$$
$$\mathring{\varphi}_\varepsilon'''=-\frac{3b^3(a^2-e)(bx+a)}{(b^2x^2+2abx+e)^{5/2}}.$$

Then we have

$$\mathring{\varphi}_\varepsilon' + 2x\mathring{\varphi}_\varepsilon'' = c_\varepsilon - \frac{1}{n} - \frac{b^3x^2(bx+3a)+eb(3bx+a)}{(b^2x^2+2abx+e)^{3/2}}$$
$$< b + \frac{1}{n-1+\varepsilon} - \frac{1}{n} < \frac{2(n-1)}{n(n+2)},$$

as $b = \min\left\{\frac{n-5}{8(n-1)}, \frac{2n-5}{(n+2)(n-1)}\right\}$, so we get the inequality (i).

Setting

$$f(x) = \mathring{\varphi}_\varepsilon(\varphi_\varepsilon - n + 9) - x\mathring{\varphi}_\varepsilon'(\varphi_\varepsilon + n + 9).$$

Then

$$\begin{aligned} f(x) &= d_\varepsilon(d_\varepsilon - n + 9) + e_\varepsilon + (2 + ab + c_\varepsilon(d_\varepsilon - 2n))\,x \\ &\quad - (b^2x^2 + 2abx + e)^{-1/2} \\ &\qquad \times \left[b((d_\varepsilon - 2n)\,b + ac)\,x^2 + (3(d_\varepsilon - n + 3)\,ab + ec)\,x + e(2d_\varepsilon - n + 9)\right]. \end{aligned}$$

Then for ε small enough we get

$$\begin{aligned} \lim_{x\to+\infty} f(x) &= \frac{a^2c_\varepsilon}{b} + d_\varepsilon(d_\varepsilon - n + 9) + a(n - 2d_\varepsilon - 9) + e\left(1 - \frac{c_\varepsilon}{b}\right) \\ &= 6(n-3) + \frac{2\varepsilon(n^2 - (18-3\varepsilon)\,n + 33 - 15\varepsilon + 2\varepsilon^2)}{n-1+\varepsilon} + e\left(1 - \frac{c_\varepsilon}{b}\right) \\ &= 6(n-3) + \frac{2\varepsilon(n^2 - (18-3\varepsilon)\,n + 33 - 15\varepsilon + 2\varepsilon^2)}{n-1+\varepsilon} - \frac{\sqrt{\varepsilon}}{(n-1+\varepsilon)\,b} \\ &< 6(n-3), \end{aligned}$$

and

$$\begin{aligned} &f'(x) \\ &= 2 + ab + c_\varepsilon(d_\varepsilon - 2n) \\ &\quad - \frac{1}{(b^2x^2+2abx+e)^{3/2}}\Big[b^3((d_\varepsilon - 2n)\,b + ac_\varepsilon)\,x^3 + 3ab^2((d_\varepsilon - 2n)\,b + ac_\varepsilon)\,x^2 \\ &\qquad + [3a^2b^2(d_\varepsilon - n + 3) - 3eb^2(n+3) + 3abc_\varepsilon\, e]\,x \\ &\qquad + eab(d_\varepsilon - 2n) + e^2c_\varepsilon\Big]. \end{aligned}$$

Then we have

$$\lim_{x\to+\infty} f'(x) = 2 + ab + c_\varepsilon(d_\varepsilon - 2n) - (ac_\varepsilon + b(d_\varepsilon - 2n)) = 0$$

and

$$\begin{aligned} f''(x) &= \frac{3b^2(a^2-e)}{(b^2x^2+2abx+e)^{5/2}} \\ &\qquad \times \left[b(b(d_\varepsilon + 6) - ac_\varepsilon)\,x^2 + (ab(d_\varepsilon - n + 3) - ec_\varepsilon)\,x - e(n+3)\right]. \end{aligned}$$

For $b = \min\left\{\frac{n-5}{8(n-1)}, \frac{2n-5}{(n+2)(n-1)}\right\}$, we obtain

$$b(d_\varepsilon + 6) - ac_\varepsilon < 0 \quad \text{and} \quad ab(d_\varepsilon - n + 3) - ec_\varepsilon < 0.$$

So $f''(x) < 0$. Then we have $f'(x) > 0$. From this we deduce that

$$f(x) < \lim_{x\to+\infty} f(x) < 6(n-3).$$

Thus, inequality (ii) is proved.

We have

$$\mathring{\varphi}_\varepsilon - x\mathring{\varphi}'_\varepsilon = d_\varepsilon - \frac{abx+e}{\sqrt{b^2x^2+2abx+e}} > d_\varepsilon - \frac{abx}{\sqrt{b^2x^2}} - \frac{e}{\sqrt{e}} = 2-2\varepsilon-\sqrt[4]{\varepsilon}.$$

This implies inequality (iii). $\square$

Suppose that M_0 is an $n(\geqslant 7)$-dimensional closed hypersurface in $\mathbb{QP}^{(n+1)/4}$ satisfying $|h|^2 < \varphi(H^2)$. Let

$$F : M^n \times [0, T) \to \mathbb{QP}^{(n+1)/4}$$

be a mean curvature flow with initial value M_0. We will show that the pinching condition is preserved along the flow. For convenience, we denote $\mathring{\varphi}_\varepsilon(H^2)$, $\mathring{\varphi}'_\varepsilon(H^2)$, $\mathring{\varphi}''_\varepsilon(H^2)$ by $\mathring{\varphi}_\varepsilon$, $\mathring{\varphi}'_\varepsilon$, $\mathring{\varphi}''_\varepsilon$, respectively.

Theorem 3.3. *If the initial value M_0 satisfies $|h|^2 < \varphi(H^2)$, then there exists a small positive number ε, such that for all $t \in [0, T)$, we have $|h|^2 < \varphi(H^2) - \varepsilon H^2 - \varepsilon$.*

Proof. Since M_0 is closed, there exists a small positive number ε_1, such that M_0 satisfies $|\mathring{h}|^2 < \mathring{\varphi}_{\varepsilon_1}$.

From Lemma 3.2(i), we have

$$\left(\frac{\partial}{\partial t} - \Delta\right)\mathring{\varphi}_{\varepsilon_1} = -2(\mathring{\varphi}'_{\varepsilon_1} + 2H^2\cdot\mathring{\varphi}''_{\varepsilon_1})|\nabla H|^2 + 2H^2\cdot\mathring{\varphi}'_{\varepsilon_1}(\varphi_{\varepsilon_1}+n+9)$$
$$\geqslant -\frac{4(n-1)}{n(n+2)}|\nabla H|^2 + 2H^2\cdot\mathring{\varphi}'_{\varepsilon_1}(\varphi_{\varepsilon_1}+n+9). \tag{3-3}$$

Let $U = |\mathring{h}|^2 - \mathring{\varphi}_{\varepsilon_1}$. We get

$$\tfrac{1}{2}\left(\frac{\partial}{\partial t} - \Delta\right)U \leqslant -|\nabla\mathring{h}|^2 + \frac{2(n-1)}{n(n+2)}|\nabla H|^2 + |\mathring{h}|^2(|h|^2-n+9) - H^2\cdot\mathring{\varphi}'_{\varepsilon_1}(|h|^2+n+9).$$

By Lemma 2.1, we have

$$-|\nabla\mathring{h}|^2 + \frac{2(n-1)}{n(n+2)}|\nabla H|^2 < -6(n-3).$$

At the point where $U = 0$, we get

$$\tfrac{1}{2}\left(\frac{\partial}{\partial t} - \Delta\right)U \leqslant -6(n-3) + \mathring{\varphi}_{\varepsilon_1}(\varphi_{\varepsilon_1}-n+9) - H^2\cdot\mathring{\varphi}'_{\varepsilon_1}(\varphi_{\varepsilon_1}+n+9) < 0.$$

Applying the maximum principle, we obtain $U < 0$ for all $t \in [0, T)$. Choose a suitable small positive number ε, we complete the proof of Theorem 3.3. □

Let

$$f_\sigma = \frac{|\mathring{h}|^2}{(\mathring{\varphi})^{1-\sigma}},$$

where $\sigma \in (0, \varepsilon^2)$ is a positive constant. The following lemma is very useful for deriving the pinching estimate for $|\mathring{h}|^2$.

Lemma 3.4. *If M_0 satisfies $|h|^2 < \varphi(H^2)$, then there exists a small positive number ε, such that the following inequality holds along the mean curvature flow*:

$$\frac{\partial}{\partial t} f_\sigma \leqslant \Delta f_\sigma + \frac{2}{\mathring{\varphi}}|\nabla f_\sigma|\,|\nabla\mathring{\varphi}| - \frac{2\varepsilon f_\sigma}{n|\mathring{h}|^2}|\nabla\mathring{h}|^2 + 2\sigma|h|^2 f_\sigma - \frac{\varepsilon}{n} f_\sigma.$$

Proof. By a straightforward computation, we have

$$\left(\frac{\partial}{\partial t} - \Delta\right) f_\sigma = f_\sigma\left[\frac{1}{|\mathring{h}|^2}\left(\frac{\partial}{\partial t} - \Delta\right)|\mathring{h}|^2 - \frac{1-\sigma}{\mathring{\varphi}}\left(\frac{\partial}{\partial t} - \Delta\right)\mathring{\varphi}\right] + 2(1-\sigma)\frac{\langle\nabla f_\sigma, \nabla\mathring{\varphi}\rangle}{\mathring{\varphi}} - \sigma(1-\sigma) f_\sigma \frac{|\nabla\mathring{\varphi}|^2}{|\mathring{\varphi}|^2}.$$

Using (3-1) and (3-3), we have

$$\begin{aligned}(3\text{-}4)\qquad \left(\frac{\partial}{\partial t} - \Delta\right) f_\sigma \leqslant{}& 2 f_\sigma\left[-\frac{|\nabla\mathring{h}|^2}{|\mathring{h}|^2} + \frac{2(n-1)}{n(n+2)}\frac{|\nabla H|^2}{\mathring{\varphi}}\right] \\ &+ 2 f_\sigma\left[|h|^2 + 9 - n - (1-\sigma)\frac{H^2\cdot\mathring{\varphi}'}{\mathring{\varphi}}(|h|^2 + n + 9)\right] \\ &+ \frac{2}{\mathring{\varphi}}|\nabla f_\sigma|\,|\nabla\mathring{\varphi}|.\end{aligned}$$

From Lemma 3.2 and Theorem 3.3, we have

$$\begin{aligned}-\frac{|\nabla\mathring{h}|^2}{|\mathring{h}|^2} + \frac{2(n-1)}{n(n+2)}\frac{|\nabla H|^2}{\mathring{\varphi}} &\leqslant -\frac{|\nabla\mathring{h}|^2}{|\mathring{h}|^2} + \frac{|\nabla\mathring{h}|^2 - 6(n-3)}{\mathring{\varphi}} \\ &\leqslant \frac{|\mathring{h}|^2 - \mathring{\varphi}}{|\mathring{h}|^2\mathring{\varphi}}|\nabla\mathring{h}|^2 - \frac{6(n-3)}{\mathring{\varphi}} \\ &\leqslant -\varepsilon\frac{|H|^2 + 1}{|\mathring{h}|^2\mathring{\varphi}}|\nabla\mathring{h}|^2 - \frac{6(n-3)}{\mathring{\varphi}} \\ &\leqslant -\frac{\varepsilon}{n|\mathring{h}|^2}|\nabla\mathring{h}|^2 - \frac{6(n-3)}{\mathring{\varphi}}.\end{aligned}$$

From (ii) and (iii) of Lemma 3.2, we have

$$
\begin{aligned}
&|h|^2+9-n-(1-\sigma)\frac{H^2\cdot\mathring{\varphi}'}{\mathring{\varphi}}(|h|^2+n+9)\\
&\quad=\frac{1-\sigma}{\mathring{\varphi}}[(\mathring{\varphi}-H^2\cdot\mathring{\varphi}')|h|^2-H^2\cdot\mathring{\varphi}'(n+9)]-n+9+\sigma|h|^2\\
&\quad\leqslant\frac{1-\sigma}{\mathring{\varphi}}[(\mathring{\varphi}-H^2\cdot\mathring{\varphi}')(\varphi-\varepsilon H^2-\varepsilon)-H^2\cdot\mathring{\varphi}'(n+9)]-n+9+\sigma|h|^2\\
&\quad=\frac{1-\sigma}{\mathring{\varphi}}[(\mathring{\varphi}-H^2\cdot\mathring{\varphi}')\varphi-H^2\cdot\mathring{\varphi}'(n+9)]-n+9+\sigma|h|^2\\
&\qquad\qquad-\frac{(1-\sigma)\,\varepsilon}{\mathring{\varphi}}(\mathring{\varphi}-H^2\cdot\mathring{\varphi}')(H^2+1)\\
&\quad\leqslant(1-\sigma)\Big[n-9+\frac{6(n-3)}{\mathring{\varphi}}\Big]-n+9+\sigma|h|^2-\frac{(1-\sigma)\,\varepsilon}{\mathring{\varphi}}(H^2+1)\\
&\quad\leqslant\sigma|h|^2+\frac{6(n-3)}{\mathring{\varphi}}-\frac{\varepsilon}{2n}.
\end{aligned}
$$

Inserting these two estimates into (3-4) will complete the proof. □

4. An estimate for traceless second fundamental form

Suppose that the initial value M_0 satisfies the condition in Theorem 1.2. For convenience, we put $W=\mathring{\varphi}$. By the conclusion of the previous section, there exists a sufficiently small positive number ε, such that for all $t\in[0,T)$, the following pinching condition holds:

$$
|\mathring{h}|^2<W-\varepsilon H^2. \tag{4-1}
$$

From this inequality and the definition of W, we have $W<\frac{H^2}{n(n-1)}+n$.

We consider the auxiliary function

$$
f_\sigma=\frac{|\mathring{h}|^2}{W^{1-\sigma}}.
$$

In this section, we will show that f_σ decays exponentially.

Lemma 4.1. *There exist positive numbers ε and C_1 depending only on M_0, such that*

$$
\frac{\partial}{\partial t}f_\sigma\leqslant\Delta f_\sigma+\frac{2C_1}{|\mathring{h}|}|\nabla f_\sigma|\,|\nabla\mathring{h}|-\frac{\varepsilon f_\sigma}{n|\mathring{h}|^2}|\nabla\mathring{h}|^2+2\sigma|h|^2f_\sigma-\frac{\varepsilon}{n}f_\sigma. \tag{4-2}
$$

Proof. According to Lemma 3.4, we have the following inequality with some suitable small $\varepsilon>0$:

$$
\frac{\partial}{\partial t}f_\sigma\leqslant\Delta f_\sigma+\frac{2}{W}|\nabla f_\sigma|\,|\nabla W|-\frac{2\varepsilon f_\sigma}{n|\mathring{h}|^2}|\nabla\mathring{h}|^2+2\sigma|h|^2f_\sigma-\frac{\varepsilon}{n}f_\sigma.
$$

By the definition of W, there exists a constant B_1, such that $|\nabla W| < B_1|\nabla H^2|$ and $|H| < B_1\sqrt{W}$. Let C_1 be a constant such that $2B_1^2|\nabla H| \leqslant C_1|\nabla\mathring{h}|$. From Lemma 2.1, we have

$$\frac{|\nabla W|}{W} \leqslant \frac{2B_1|H|\,|\nabla H|}{\sqrt{W}|\mathring{h}|} \leqslant \frac{2B_1^2|\nabla H|}{|\mathring{h}|} \leqslant \frac{C_1|\nabla\mathring{h}|}{|\mathring{h}|}. \tag{4-3}$$ □

We need the following estimate for the Laplacian of $|\mathring{h}|^2$.

Lemma 4.2. $\Delta|\mathring{h}|^2 \geqslant 2\langle\mathring{h}, \nabla^2 H\rangle + 2|\mathring{h}|^2(\varepsilon|h|^2 - 2n^2) - 18|\mathring{h}|\,|H|$.

Proof. We have

$$\Delta|\mathring{h}|^2 = 2|\nabla\mathring{h}|^2 + 2\mathring{h}\cdot\Delta\mathring{h} = 2|\nabla\mathring{h}|^2 + 2\sum_{i,j}\mathring{h}_{ij}\cdot\Delta h_{ij}$$

and

$$\begin{aligned}\sum_{i,j}\mathring{h}_{ij}\cdot\Delta h_{ij} &= \langle\mathring{h}, \nabla^2 H\rangle + \sum_{i,p,j} H h_{ip}\, h_{pj}\, h_{ij} - |h|^4 \\ &\quad + 3H\sum_{i,j}\sum_{k_0=1}^{3} J_{in+1}^{(k_0)}\, J_{jn+1}^{(k_0)}\,\mathring{h}_{ij} - (n+9)\,|\mathring{h}|^2 + 2n|\mathring{h}|^2 - 6S_1 \\ &\geqslant \langle\mathring{h}, \nabla^2 H\rangle + \sum_{i,p,j} H h_{ip} h_{pj} h_{ij} - |h|^4 + (n-9)\,|\mathring{h}|^2 - 9|\mathring{h}|\,|H|.\end{aligned}$$

It follows from the proof of the Lemma 4.2 in [17], we choose a local orthonormal frame such that

$$H = |H|\, e_{n+1} \quad \text{and} \quad \mathring{h} = \operatorname{diag}\{\mathring{\lambda}_1, \dots, \mathring{\lambda}_n\}.$$

So we have

$$\begin{aligned}\sum_{i,p,j} H h_{ip}\, h_{pj}\, h_{ij} - |h|^4 \\ &= H\sum_i \mathring{\lambda}_i^3 + \frac{1}{n}H^2|\mathring{h}|^2 - |\mathring{h}|^4 \\ &\geqslant -|H|\frac{n-2}{\sqrt{n(n-1)}}|\mathring{h}|^3 + \frac{1}{n}H^2|\mathring{h}|^2 - |\mathring{h}|^4 \\ &= |\mathring{h}|^2\left(\frac{1}{n}H^2 - |\mathring{h}|^2 - \frac{n-2}{\sqrt{n(n-1)}}|\mathring{h}|\,|H|\right) \\ &\geqslant |\mathring{h}|^2\left[\frac{1}{n}H^2 - \left(\frac{H^2}{n(n-1)} + n - \varepsilon H^2\right) - (n-2)\left(\frac{H^2}{n(n-1)} + n\right)\right] \\ &= |\mathring{h}|^2(\varepsilon H^2 - n(n-1)) \\ &> |\mathring{h}|^2(\varepsilon|h|^2 - n^2),\end{aligned}$$

where we have used $|\mathring{h}|^2 < W - \varepsilon H^2$ and $W < \frac{H^2}{n(n-1)} + n$. □

From (4-3) and Lemma 4.2, we have

$$\begin{aligned}\Delta f_\sigma &= f_\sigma\left(\frac{\Delta|\mathring{h}|^2}{|\mathring{h}|^2}-(1-\sigma)\frac{\Delta W}{W}\right)-2(1-\sigma)\frac{\langle\nabla f_\sigma,\nabla W\rangle}{W}+\sigma(1-\sigma)\,f_\sigma\frac{|\nabla W|^2}{W^2}\\ &\geqslant f_\sigma\frac{\Delta|\mathring{h}|^2}{|\mathring{h}|^2}-(1-\sigma)\,f_\sigma\frac{\Delta W}{W}-\frac{2C_1|\nabla f_\sigma|\,|\nabla\mathring{h}|}{|\mathring{h}|}\\ &\geqslant \frac{2\langle\mathring{h},\nabla^2H\rangle}{\mathring{W}^{1-\sigma}}+2f_\sigma(\varepsilon|h|^2-2n^2)-(1-\sigma)\frac{f_\sigma\Delta W}{W}-\frac{2C_1|\nabla f_\sigma||\nabla\mathring{h}|}{|\mathring{h}|}-\frac{18f_\sigma|H|}{|\mathring{h}|}.\end{aligned}$$

Multiplying both sides of the above inequality by f_σ^{p-1}, we get

$$\begin{aligned}2\varepsilon f_\sigma^p|h|^2 \leqslant f_\sigma^{p-1}\Delta f_\sigma+(1-\sigma)\frac{f_\sigma^p\Delta W}{W}-\frac{2f_\sigma^{p-1}\langle\mathring{h},\nabla^2H\rangle}{W^{1-\sigma}}\\ +\frac{2C_1f_\sigma^{p-1}|\nabla f_\sigma|\,|\nabla\mathring{h}|}{|\mathring{h}|}+4n^2f_\sigma^p+\frac{18f_\sigma|H|}{|\mathring{h}|}.\end{aligned}\tag{4-4}$$

Then integrate both sides of (4-4) over M_t. By the divergence theorem, we get

$$\int_{M_t}f_\sigma^{p-1}\Delta f_\sigma\,d\mu_t=-(p-1)\int_{M_t}f_\sigma^{p-2}|\nabla f_\sigma|^2\,d\mu_t.\tag{4-5}$$

From (4-4), we have

$$\begin{aligned}\int_{M_t}\frac{f_\sigma^p}{W}\Delta W\,d\mu_t&=-\int_{M_t}\left\langle\nabla\left(\frac{f_\sigma^p}{W}\right),\nabla W\right\rangle d\mu_t\\ &=\int_{M_t}\left(-\frac{pf_\sigma^{p-1}}{W}\langle\nabla f_\sigma,\nabla W\rangle+\frac{f_\sigma^p}{W^2}|\nabla W|^2\right)d\mu_t\\ &\leqslant\int_{M_t}\left(\frac{C_1pf_\sigma^{p-1}}{|\mathring{h}|}|\nabla f_\sigma|\,|\nabla\mathring{h}|+\frac{C_1^2f_\sigma^p}{|\mathring{h}|^2}|\nabla\mathring{h}|^2\right)d\mu_t.\end{aligned}\tag{4-6}$$

We also have

$$\begin{aligned}&-\int_{M_t}\frac{f_\sigma^{p-1}\langle\mathring{h},\nabla^2H\rangle}{W^{1-\sigma}}d\mu_t\\ &=\int_{M_t}\nabla_i\left(\frac{f_\sigma^{p-1}}{W^{1-\sigma}}\mathring{h}_{ij}\right)\nabla_jH\,d\mu_t\\ &=\int_{M_t}\left[\frac{(p-1)\,f_\sigma^{p-2}}{W^{1-\sigma}}\mathring{h}_{ij}\nabla_if_\sigma-\frac{(1-\sigma)\,f_\sigma^{p-1}}{W^{2-\sigma}}\mathring{h}_{ij}\nabla_iW+\frac{f_\sigma^{p-1}}{W^{1-\sigma}}\nabla_i\mathring{h}_{ij}\right]\nabla_jH\,d\mu_t\\ &\leqslant\int_{M_t}\left[\frac{(p-1)\,f_\sigma^{p-1}}{|\mathring{h}|}|\nabla f_\sigma|+\frac{f_\sigma^{p-1}}{W^{2-\sigma}}|\mathring{h}|\,|\nabla W|+\frac{f_\sigma^{p-1}}{W^{1-\sigma}}n|\nabla\mathring{h}|\right]|\nabla H|\,d\mu_t\end{aligned}\tag{4-7}$$

$$\leqslant \int_{M_t}\left[\frac{(p-1)\, f_\sigma^{p-1}}{|\mathring{h}|}|\nabla f_\sigma| + \frac{C_1 f_\sigma^{p-1}}{W^{1-\sigma}}|\nabla \mathring{h}| + \frac{f_\sigma^{p-1}}{W^{1-\sigma}} n|\nabla \mathring{h}|\right] n|\nabla \mathring{h}|\, d\mu_t$$
$$\leqslant \int_{M_t}\left[\frac{n(p-1)\, f_\sigma^{p-1}}{|\mathring{h}|}|\nabla f_\sigma|\,|\nabla \mathring{h}| + \frac{(C_1 n + n^2) f_\sigma^{p}}{|\mathring{h}|^2}|\nabla \mathring{h}|^2\right] d\mu_t.$$

Putting (4-4)–(4-7) together, we get

$$\int_{M_t}|h|^2 f_\sigma^p\, d\mu_t \leqslant C_2 \int_{M_t}\left[\frac{p f_\sigma^{p-1}}{|\mathring{h}|}|\nabla f_\sigma|\,|\nabla \mathring{h}| + \frac{f_\sigma^p}{|\mathring{h}|^2}|\nabla \mathring{h}|^2 + f_\sigma^p + \frac{f_\sigma^p |H|}{|\mathring{h}|}\right] d\mu_t,$$

where C_2 is a positive constant depending only on M_0.

Combining Lemma 4.2 and (4-2), we get

$$\begin{aligned}
\text{(4-8)}\quad & \frac{\partial}{\partial t}\int_{M_t} f_\sigma^p\, d\mu_t \\
&= p\int_{M_t} f_\sigma^{p-1}\frac{\partial}{\partial t} f_\sigma\, d\mu_t - \int_{M_t} f_\sigma^p H^2\, d\mu_t \\
&\leqslant p\int_{M_t} f_\sigma^{p-2}\Bigg[-(p-1)|\nabla f_\sigma|^2 + (2C_1 + 2\sigma C_2 p)\frac{f_\sigma}{|\mathring{h}|}|\nabla f_\sigma|\,|\nabla \mathring{h}| \\
&\qquad\qquad - \left(\frac{\varepsilon}{2n} - 2\sigma C_2\right)\frac{f_\sigma^2}{|\mathring{h}|^2}|\nabla \mathring{h}|^2\Bigg] d\mu_t \\
&\qquad - p\int_{M_t} f_\sigma^p\left(\frac{\varepsilon}{n} - 2\sigma C_2 + \frac{6(n-3)\,\varepsilon}{2n|\mathring{h}|^2} - \frac{2\sigma C_2 |H|}{|\mathring{h}|} + \frac{|H|^2}{p}\right) d\mu_t.
\end{aligned}$$

Now we will show that the L^p-form of f_σ decays exponentially.

Lemma 4.3. *There exist positive constants C_3, p_0, σ_0 depending only on M_0, such that for all $p \geqslant p_0$ and $\sigma \leqslant \sigma_0/\sqrt{p}$, we have*

$$\left(\int_{M_t} f_\sigma^p\, d\mu_t\right)^{1/p} < C_3\, e^{-\varepsilon t}.$$

Proof. The expression in the square bracket of the right side of (4-8) is a quadratic polynomial. With p_0 large enough and σ_0 small enough, its discriminant satisfies

$$(2C_1 + 2\sigma C_2 p)^2 - 4(p-1)\left(\frac{\varepsilon}{2n} - 2\sigma C_2\right) < 0 \quad \text{and} \quad \frac{12\varepsilon}{7} \geqslant p\sigma^2 C_2^2.$$

We have

$$\begin{aligned}
\frac{\varepsilon}{n} - 2\sigma C_2 + \frac{6(n-3)\,\varepsilon}{2n|\mathring{h}|^2} - \frac{2\sigma C_2 |H|}{|\mathring{h}|} + \frac{|H|^2}{p} &\geqslant \frac{\varepsilon}{n} - 2\sigma C_2 + \frac{12\varepsilon}{7|\mathring{h}|^2} - \frac{p\sigma^2 C_2^2}{|\mathring{h}|^2} \\
&\geqslant \frac{\varepsilon}{n} - 2\sigma C_2 > \frac{\varepsilon}{2n}.
\end{aligned}$$

Here we have used the inequality $\frac{\varepsilon}{2n} - 2\sigma C_2 > 0$, which is implied by the choices of p_0 and σ_0. Then we get

$$\frac{d}{dt}\int_{M_t} f_\sigma^p \, d\mu_t \leqslant -\frac{p\varepsilon}{2n}\int_{M_t} f_\sigma^p \, d\mu_t.$$

So we get $\int_{M_t} f_\sigma^p \, d\mu_t \leqslant e^{-p\varepsilon/2n}\int_{M_0} f_\sigma^p \, d\mu_0$, which completes the proof. □

Let $g_\sigma = f_\sigma e^{\varepsilon t/2}$. By the Sobolev inequality on submanifolds and a Stampacchia iteration procedure, we obtain that g_σ is uniformly bounded for all t (see [9] or [14] for the details). Then we obtain the following theorem.

Theorem 4.4. *There exist positive constants ε, σ and C_0 depending only on M_0, such that for all $t \in [0, T)$, we have*

$$|\mathring{h}|^2 \leqslant C_0(H^2+1)^{1-\sigma} e^{-\varepsilon t/2}.$$

5. A gradient estimate

We derive an estimate for $|\nabla H|^2$ along the mean curvature flow. Firstly, the same as Proposition 4.3 in [25], we have:

Lemma 5.1. *There exists a positive constants $C_4 > 1$ depending only on n, such that*

$$\frac{\partial}{\partial t}|\nabla H|^2 \leqslant \Delta|\nabla H|^2 + C_4(H^2+1)\,|\nabla h|^2.$$

Secondly, we need the following estimates.

Lemma 5.2. *Along the mean curvature flow, we have*

(i) $\frac{\partial}{\partial t}H^4 \geqslant \Delta H^4 - 12nH^2|\nabla h|^2 + \frac{4}{n}H^6$,

(ii) $\frac{\partial}{\partial t}|\mathring{h}|^2 \leqslant \Delta|\mathring{h}|^2 - \frac{1}{3}|\nabla h|^2 + C_5|\mathring{h}|^2(H^2+1)$,

(iii) $\frac{\partial}{\partial t}(H^2|\mathring{h}|^2) \leqslant \Delta(H^2|\mathring{h}|^2) - \frac{1}{6}H^2|\nabla h|^2 + C_6(H^2+1)^2|\mathring{h}|^2 + C_7|\nabla h|^2$,

where C_5, C_6, C_7 are sufficiently large constants.

Proof. (i) From Lemma 2.3, we derive that

$$\frac{\partial}{\partial t}H^4 = \Delta H^4 - 12H^2|\nabla H|^2 + 4H^4(|h|^2+n+9).$$

From Lemma 2.1, we have $12H^2|\nabla H|^2 \leqslant 12nH^2|\nabla h|^2$. Obviously, inequality (i) holds.

(ii) We have

$$\frac{\partial}{\partial t}|\mathring{h}|^2 = \Delta|\mathring{h}|^2 - 2|\nabla\mathring{h}|^2 + 2|h|^2\,|\mathring{h}|^2 + 18|\mathring{h}|^2 - 2n|\mathring{h}|^2 + 12S_1.$$

From Lemma 2.1, we get $|\nabla\mathring{h}|^2 \geqslant \frac{1}{6}|\nabla h|^2$. Choose a large constant C_5, we obtain inequality (ii).

(iii) It follows the evolution equation that

$$\frac{\partial}{\partial t}(H^2|\mathring{h}|^2) = \Delta(H^2|\mathring{h}|^2) - 2\langle\nabla H^2, \nabla|\mathring{h}|^2\rangle + 4|\mathring{h}|^2H^2|h|^2 - 2|\mathring{h}|^2|\nabla H|^2 \\ - 2H^2|\nabla\mathring{h}|^2 + 36H^2|\mathring{h}|^2 + 12S_1H^2.$$

From Lemma 2.1, we get $-2H^2|\nabla\mathring{h}|^2 \leqslant -\frac{1}{3}H^2|\nabla h|^2$. From the preserved pinching condition $|\mathring{h}|^2 < W$, we have

$$4|\mathring{h}|^2H^2|h|^2 + 36H^2|\mathring{h}|^2 \leqslant C_6(H^2+1)^2|\mathring{h}|^2.$$

Using Theorem 4.4, we have

$$-2\langle\nabla H^2, \nabla|\mathring{h}|^2\rangle \leqslant 8|H|\,|\nabla H|\,|\mathring{h}|\,|\nabla h| \leqslant 8n\sqrt{C_0}|H|(H^2+1)^{(1-\sigma)/2}|\nabla h|^2.$$

By Young's inequality, there exists a positive constant C_7, such that

$$-2\langle\nabla H^2, \nabla|\mathring{h}|^2\rangle \leqslant \big(C_7 + \tfrac{1}{6}H^2\big)|\nabla h|^2. \qquad \square$$

Now we prove a gradient estimate for mean curvature.

Theorem 5.3. *For any $\eta \in (0, \sqrt{\varepsilon}/4\pi n)$, there exists a number $\Psi(\eta)$ depending only on η and M_0, such that*

$$|\nabla H|^2 < [(\eta H)^4 + \Psi^2(\eta)]\, e^{-\varepsilon t/4}.$$

Proof. Define a scalar function

$$f = (|\nabla H|^2 + B_1|\mathring{h}|^2 + B_2H^2|\mathring{h}|^2)\, e^{\varepsilon t/4} - (\eta H)^4,$$

where B_1, B_2 are two positive constants.

From Lemmas 5.1 and 5.2, we obtain

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} - \Delta\right) f \\
&= \frac{\varepsilon}{4}(|\nabla H|^2 + B_1|\mathring{h}|^2 + B_2H^2|\mathring{h}|^2)\, e^{\varepsilon t/4} \\
&\qquad + e^{\varepsilon t/4}\left(\frac{\partial}{\partial t} - \Delta\right)(|\nabla H|^2 + B_1|\mathring{h}|^2 + B_2H^2|\mathring{h}|^2) - \eta^4\left(\frac{\partial}{\partial t} - \Delta\right)H^4 \\
&\leqslant \frac{\varepsilon}{4}(|\nabla H|^2 + B_1|\mathring{h}|^2 + B_2H^2|\mathring{h}|^2)\, e^{\varepsilon t/4} \\
&\qquad + e^{\varepsilon t/4}\big\{\big(C_4(H^2+1)|\nabla h|^2\big) + B_1\big(-\tfrac{1}{3}|\nabla h|^2 + C_5(H^2+1)|\mathring{h}|^2\big) \\
&\qquad\qquad + B_2\big(-\tfrac{1}{6}H^2|\nabla h|^2 + C_7|\nabla h|^2 + C_6(H^2+1)^2|\mathring{h}|^2\big)\big\} \\
&\qquad - \eta^4\left(-12nH^2|\nabla h|^2 + \frac{4}{n}H^6\right) \\
&= H^2|\nabla h|^2\left[e^{\varepsilon t/4}\left(C_4 - \frac{B_2}{6}\right) + 24n\eta^4\right] + e^{\varepsilon t/4}\left[|\nabla h|^2\left(C_4 - \frac{B_1}{3} + C_7B_2\right) + \frac{\varepsilon}{4}|\nabla H|^2\right] \\
&\qquad + e^{\varepsilon t/4}|\mathring{h}|^2\left[B_1C_5(H^2+1) + B_2C_6(H^2+1)^2 + \frac{\varepsilon}{4}(B_1 + B_2H^2)\right] - \frac{4\eta^4}{n}|H|^6.
\end{aligned}$$

Choose constants B_1 and B_2, such that $C_4 - \frac{B_2}{6} < -1$ and $C_4 - \frac{B_1}{3} + C_7 B_2 < -1$. Then applying Theorem 4.4, we get

$$\left(\frac{\partial}{\partial t} - \Delta\right) f \leqslant e^{-\varepsilon t/4}\left[B_3(H^2+1)^2(H^2+1)^{1-\sigma} - \frac{4\eta^4}{n} H^6\right]. \tag{5-1}$$

Consider the expression in the bracket of (5-1). Since the coefficient of H^6 is negative, it has a upper bound $\Psi_2(\eta)$. Then we have $\left(\frac{\partial}{\partial t} - \Delta\right) f \leqslant e^{-\varepsilon t/4}\,\Psi_2(\eta)$. It follows from the maximum principle that f is bounded. This completes the proof of Theorem 5.3. □

6. Convergence

In order to estimate the diameter of M_t, we need the well-known Myers's theorem:

Theorem 6.1 (Myers's theorem). *Let Γ be a geodesic of length at least $\pi/\sqrt{K}$ on M. If the Ricci curvature satisfies $\operatorname{Ric}(X) \geqslant (n-1)K$ for each unit vector $X \in T_x M$, at any point $x \in \Gamma$, then Γ has conjugate points.*

Now we show that under the assumption of Theorem 1.2, the mean curvature flow converges to a round point.

Theorem 6.2. *If M_0 satisfies $|\mathring{h}|^2 < \mathring{\varphi}$, then $T < \infty$ and M_t converges to a round point as $t \to T$.*

Proof. Assume $T = \infty$. Let $|H|_{\min}(t) = \min_{M_t} |H|$, $|H|_{\max}(t) = \max_{M_t} |H|$.

We claim that $H^2 \cdot e^{\varepsilon t/8}$ is uniformly bounded on $[0, \infty)$. Suppose not, then there is a time τ such that $|H|^2_{\max}(\tau) \cdot e^{\varepsilon\tau/8} > \Psi/\eta^2$. By Theorem 5.3, for every small positive number η, there exists a positive number Ψ, such that $|\nabla H| < [(\eta H)^2 + \Psi]e^{-\varepsilon t/8}$. Then we have $|\nabla H| < 2\eta^2 |H|^2_{\max}$ on M_τ.

From Lemma 4.1 in [27], the sectional curvature K of M satisfies

$$K \geqslant \tfrac{1}{2}\left(2 + \frac{1}{n-1} H^2 - |h|^2\right). \tag{6-1}$$

By Theorem 4.4, we obtain

$$K \geqslant \tfrac{1}{2}\left(2 + \frac{1}{n(n-1)} H^2 - C_0 (H^2+1)^{1-\sigma} e^{-\varepsilon t/2}\right).$$

Hence, we can pick τ large enough such that $K \geqslant (1/2n^2) H^2$ on M_τ.

Let x be a point on M_τ where $|H|$ achieves its maximum. Consider all the geodesics of length at most $(4\eta |H|_{\max})^{-1}$ starting from x. As $|\nabla H^2| < 4\eta^2 |H|^3_{\max}$, we have

$$H^2 \geqslant |H|^2_{\max} - 4\eta^2 |H|^3_{\max} \cdot (4\eta |H|_{\max})^{-1} = (1-\eta)|H|^2_{\max}$$

along such geodesics. Since $|\nabla H| < 2\eta^2|H|^2_{\max}$ and $K \geqslant (1/2n^2)\,H^2$ on M_τ, one has

$$K \geqslant \frac{1}{2n^2}(1-\eta)\,|H|^2_{\max}$$

along such geodesics provided

$$\eta \in \left(0,\, \min\left\{\frac{1}{32\pi n},\, \frac{\sqrt{\varepsilon}}{4\pi n}\right\}\right).$$

By Myers's theorem, these geodesics can reach any point of M_τ. This implies

$$H^2 \geqslant (1-\eta)\,|H|^2_{\max} \quad \text{on } M_\tau.$$

Combining this inequality with $|H|^2_{\max}(\tau)\cdot e^{\varepsilon\tau/8} > \Psi/\eta^2$ and Theorem 5.3, we get

$$|\nabla H|^2 < (\eta H)^4 + \frac{\eta^4}{(1-\eta)^2}|H|^4_{\min}(\tau), \quad t \geqslant \tau.$$

From the evolution equation of H^2, we have that for $t \geqslant \tau$,

$$\left(\frac{\partial}{\partial t} - \Delta\right)H^2 \geqslant -2|\nabla H|^2 + \frac{2}{n}H^4 \geqslant \frac{1}{n}H^4 - \frac{1}{2n}|H|^4_{\min}(\tau) \tag{6-2}$$

for $\eta > 0$ sufficiently small. By the maximum principle, we get $H^2 \geqslant |H|^2_{\min}(\tau)$ for $t \geqslant \tau$. Then (6-2) yields

$$\left(\frac{\partial}{\partial t} - \Delta\right)H^2 \geqslant \frac{1}{2n}H^4, \quad t \geqslant \tau.$$

By the maximum principle, H^2 blows up in finite time. This contradicts the infinity of T. Therefore, we obtain $H^2 \leqslant Ce^{-\varepsilon t/8}$ for $t \in [0, \infty)$ for a uniform positive constant C. By Theorem 4.4, $|h|^2 = |\mathring{h}|^2 + \frac{1}{n}|H|^2 \leqslant Ce^{-\varepsilon t/8}$ for $t \in [0, \infty)$, which implies that M_t converges to a closed totally geodesic hypersurface M_∞ as $t \to \infty$. However, there is no closed totally geodesic hypersurface in $\mathbb{QP}^{(n+1)/4}$, see, e.g., Corollary 7.2 in [25]. Therefore, we get a contradiction, and hence $T < \infty$.

So T is finite, and $\max_{M_t}|h|^2$ blows up as $t \to T$. From the preserved pinching condition, $|H|_{\max}(t)$ also blows up as $t \to T$. By Theorem 5.3, for any $\eta \in (0, \sqrt{\varepsilon}/4n\pi)$, there exists a positive number $\Psi = \Psi(\eta) > 1$, such that

$$|\nabla H| < (\eta H)^2 + \Psi \quad \text{for } t \in [0, T).$$

Since $|H|_{\max}(t)$ blows up as $t \to T$, there exists a time τ_1 depending on η, such that

$$|H|^2_{\max} \geqslant \max\left\{\frac{2\Psi}{\eta^2},\, \frac{8n}{\varepsilon}\right\} \quad \text{on } M_{\tau_1},$$

where $\varepsilon > 0$ is as in (4-1). Then we get $|\nabla H| \leqslant 2\eta^2 |H|^2_{\max}$ on M_{τ_1}. Similarly as above, we obtain

$$|H|^2_{\min} \geqslant (1-\eta)|H|^2_{\max} \quad \text{on } M_{\tau_1}. \tag{6-3}$$

By (4-1) and (6-1), one has $K \geqslant \frac{1}{2}(\varepsilon H^2 - n)$ for all $t \in [0, T)$. Hence for small $\eta > 0$, we have the following estimate at $t = \tau_1$:

$$K \geqslant \tfrac{1}{4}\varepsilon H^2 + \tfrac{1}{4}[\varepsilon(1-\eta)|H|^2_{\max} - 2n] \geqslant \tfrac{1}{4}\varepsilon H^2 + \tfrac{1}{8}(\varepsilon |H|^2_{\max} - 4n) \geqslant \tfrac{1}{16}\varepsilon |H|^2_{\max}.$$

This implies $\operatorname{diam}(M_{\tau_1}) \leqslant 4\pi/(\sqrt{\varepsilon}|H|_{\max})$.

Furthermore, by Theorem 5.3 and (6-3) one has that for $t \geqslant \tau_1$,

$$|\nabla H|^2 < 2(\eta H)^4 + 2\Psi^2 \leqslant 2(\eta H)^4 + \tfrac{1}{4}\eta^4 |H|^2_{\max}(\tau_1) \leqslant 2(\eta H)^4 + \tfrac{1}{2}\eta^4 |H|^2_{\min}(\tau_1).$$

Hence for $t \geqslant \tau_1$, we have

$$\left(\frac{\partial}{\partial t} - \Delta\right) H^2 \geqslant -2|\nabla H|^2 + \frac{2}{n} H^4 \geqslant \frac{1}{n} H^4 - \frac{1}{2n} |H|^4_{\min}(\tau_1) \tag{6-4}$$

provided that $\eta > 0$ is sufficiently small. By the maximum principle, we get $H^2 \geqslant |H|^2_{\min}(\tau_1)$ for $t \geqslant \tau_1$. Then (6-4) yields

$$\left(\frac{\partial}{\partial t} - \Delta\right) H^2 \geqslant \frac{1}{2n} H^4, \quad t \geqslant \tau_1.$$

By the maximum principle, $|H|^2_{\min}(t)$ is increasing on $[\tau_1, T)$. So

$$|H|^2_{\max}(t) \geqslant |H|^2_{\min}(t) \geqslant |H|^2_{\min}(\tau_1) \geqslant \tfrac{1}{2}|H|^2_{\max}(\tau_1) \geqslant \max\left\{\frac{\Psi}{\eta^2}, \frac{4n}{\varepsilon}\right\}$$

for all $t \geqslant \tau_1$ and for every $\eta > 0$ sufficiently small. Hence $|\nabla H| \leqslant 2\eta^2 |H|^2_{\max}$ for all $t \geqslant \tau_1$. By a similar argument, we get $|H|^2_{\min} \geqslant (1-\eta)|H|^2_{\max}$ for all η sufficiently small and all $t \geqslant \tau_1$. This implies $|H|_{\min}/|H|_{\max} \to 1$ as $t \to T$.

Since for $t \geqslant \tau_1$,

$$K \geqslant \tfrac{1}{4}\varepsilon H^2 + \tfrac{1}{8}(\varepsilon |H|^2_{\max} - 4n) \geqslant \tfrac{1}{16}\varepsilon |H|^2_{\max},$$

we have

$$\operatorname{diam}(M_t) \leqslant \frac{4\pi}{\sqrt{\varepsilon}|H|_{\max}(t)}$$

for all $t \geqslant \tau_1$. So $\operatorname{diam}(M_t) \to 0$, and by a similar argument as in [10], M_t shrinks to a single point as $t \to T$.

Now we dilate the metric of the ambient space such that the hypersurface maintains its volume along the flow. Using the same method as in [19], we can prove that the sequence of time-slices of rescaled flow corresponding to any sequence of times that tends to infinity has a subsequence that converges to a round sphere. This proves that the limit point of the mean curvature flow is round. □

Acknowledgements

The authors would like to thank the referees for valuable comments and suggestions. The research was supported by the National Natural Science Foundation of China, grants 12071424, 12171423, 12471051.

References

[1] B. Andrews and C. Baker, "Mean curvature flow of pinched submanifolds to spheres", *J. Differential Geom.* **85**:3 (2010), 357–395. MR Zbl

[2] C. Baker, "The mean curvature flow of submanifolds of high codimension", preprint, 2011. arXiv

[3] S. Brendle, "A general convergence result for the Ricci flow in higher dimensions", *Duke Math. J.* **145**:3 (2008), 585–601. MR Zbl

[4] S. Brendle and R. Schoen, "Manifolds with 1/4-pinched curvature are space forms", *J. Amer. Math. Soc.* **22**:1 (2009), 287–307. MR Zbl

[5] Y. X. Dong, "Real hypersurfaces of a quaternionic projective space", *Chinese Ann. Math. Ser. A* **14**:1 (1993), 75–80. MR Zbl

[6] J.-R. Gu and H.-W. Xu, "The sphere theorems for manifolds with positive scalar curvature", *J. Differential Geom.* **92**:3 (2012), 507–545. MR Zbl

[7] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geometry* **17**:2 (1982), 255–306. MR Zbl

[8] R. S. Hamilton, "The formation of singularities in the Ricci flow", pp. 7–136 in *Surveys in differential geometry*, *II* (Cambridge, MA, 1993), Int. Press, Cambridge, MA, 1995. MR Zbl

[9] G. Huisken, "Flow by mean curvature of convex surfaces into spheres", *J. Differential Geom.* **20**:1 (1984), 237–266. MR Zbl

[10] G. Huisken, "Contracting convex hypersurfaces in Riemannian manifolds by their mean curvature", *Invent. Math.* **84**:3 (1986), 463–480. MR Zbl

[11] G. Huisken, "Deforming hypersurfaces of the sphere by their mean curvature", *Math. Z.* **195**:2 (1987), 205–219. MR Zbl

[12] H. B. Lawson, Jr., "Rigidity theorems in rank-1 symmetric spaces", *J. Differential Geometry* **4** (1970), 349–357. MR Zbl

[13] L. Lei and H. Xu, "Mean curvature flow of arbitrary codimension in spheres and sharp differentiable sphere theorem", preprint, 2015. arXiv

[14] L. Lei and H. Xu, "A new version of Huisken's convergence theorem for mean curvature flow in spheres", preprint, 2015. arXiv

[15] L. Lei and H. Xu, "An optimal convergence theorem for mean curvature flow of arbitrary codimension in hyperbolic spaces", preprint, 2015. arXiv

[16] L. Lei and H.-W. Xu, "New developments in mean curvature flow of arbitrary codimension inspired by Yau rigidity theory", pp. 327–348 in *Proceedings of the Seventh International Congress of Chinese Mathematicians*, *I*, Adv. Lect. Math. (ALM) **43**, Int. Press, Somerville, MA, 2019. MR Zbl

[17] L. Lei and H. Xu, "Mean curvature flow of arbitrary codimension in complex projective spaces", *Chinese Ann. Math. Ser. B* **44**:6 (2023), 857–892. MR Zbl

[18] K. Liu, H. Xu, and E. Zhao, "Deforming submanifolds of arbitrary codimension in a sphere", preprint, 2012. arXiv

[19] K. Liu, H. Xu, and E. Zhao, "Mean curvature flow of higher codimension in Riemannian manifolds", preprint, 2012. arXiv

[20] K. Liu, H. Xu, F. Ye, and E. Zhao, "Mean curvature flow of higher codimension in hyperbolic spaces", *Comm. Anal. Geom.* **21**:3 (2013), 651–669. MR Zbl

[21] K. Liu, H. Xu, F. Ye, and E. Zhao, "The extension and convergence of mean curvature flow in higher codimension", *Trans. Amer. Math. Soc.* **370**:3 (2018), 2231–2262. MR Zbl

[22] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. arXiv

[23] G. Perelman, "Finite extinction time for the solutions to the Ricci flow on certain three-manifolds", preprint, 2003. arXiv

[24] G. Perelman, "Ricci flow with surgery on three-manifolds", preprint, 2003. arXiv

[25] G. Pipoli and C. Sinestrari, "Mean curvature flow of pinched submanifolds of $\mathbb{CP}^n$", *Comm. Anal. Geom.* **25**:4 (2017), 799–846. MR Zbl

[26] K. Shiohama and H. Xu, "The topological sphere theorem for complete submanifolds", *Compositio Math.* **107**:2 (1997), 221–232. MR Zbl

[27] H.-W. Xu and J.-R. Gu, "An optimal differentiable sphere theorem for complete manifolds", *Math. Res. Lett.* **17**:6 (2010), 1111–1124. MR Zbl

[28] H.-W. Xu and E.-T. Zhao, "Topological and differentiable sphere theorems for complete submanifolds", *Comm. Anal. Geom.* **17**:3 (2009), 565–585. MR Zbl

Received March 18, 2024. Revised July 16, 2024.

SHIYANG LI
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
lisycms@zju.edu.cn

HONGWEI XU
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
xuhw@zju.edu.cn

ENTAO ZHAO
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
zhaoet@zju.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.243

HECKE EIGENVALUES AND FOURIER–JACOBI COEFFICIENTS OF SIEGEL CUSP FORMS OF DEGREE 2

MURUGESAN MANICKAM, KARAM DEO SHANKHADHAR AND VASUDEVAN SRIVATSA

The nonvanishing of the first Fourier–Jacobi coefficient of a Siegel eigenform F gives us that the vanishing of its m-th Fourier–Jacobi coefficient $F \mid \rho_m$ implies the vanishing of its m-th eigenvalue $\lambda_F(m)$. Conversely, we prove that for any odd, squarefree m if $\lambda_F(m)$ is zero then $F \mid \rho_m$ vanishes. While investigating this converse question and its important consequences, we generalize certain existing results of Kohnen and Skoruppa (1989) for index 1 Jacobi cusp forms to any arbitrary index, which are also of independent interest.

1. Introduction

In [6], Kohnen and Skoruppa introduced a novel Dirichlet series attached to any two Siegel cusp forms of degree 2 involving their Fourier–Jacobi coefficients. More importantly they could connect the Dirichlet series attached to a Siegel eigenform and any Siegel cusp form in the Maass space to the spinor zeta function of the Siegel eigenform. In particular, this connection gives us that the image of the m-th Fourier–Jacobi coefficient under certain adjoint operator is same as the m-th eigenvalue times the first Fourier–Jacobi coefficient of the Siegel eigenform (see (1)). Formally this could be viewed as an analogue of the relation between Fourier coefficients and eigenvalues of the Hecke eigenforms in the degree 1 case. Therefore it is natural to explore the relation between Fourier–Jacobi coefficients and eigenvalues further. In this paper, we take up this problem and investigate it in detail.

To state our results precisely, let us first introduce some notation. Throughout this article, k stands for an even integer and $k \geqslant 4$. Let $S_k(\Gamma_2)$ be the space of Siegel cusp forms of weight k for the symplectic group $\Gamma_2 := \mathrm{Sp}_4(\mathbb{Z})$. Let $J_{k,m}^{\mathrm{cusp}}$ denote the space of Jacobi cusp forms of weight k and index m for the group $\mathrm{SL}_2(\mathbb{Z}) \ltimes (\mathbb{Z} \times \mathbb{Z})$. For any $l \geqslant 1$, let $V_{m,l} : J_{k,m}^{\mathrm{cusp}} \to J_{k,ml}^{\mathrm{cusp}}$ be the linear operator defined by [3, page 41, (2)]

MSC2020: 11F11, 11F37, 11F46, 11F50.
Keywords: Siegel modular forms, Hecke operators, Hecke eigenvalues, Fourier–Jacobi coefficients, Jacobi forms, modular forms of integral and half-integral weight.

and let $T_{J,l} : J_{k,m}^{\text{cusp}} \to J_{k,m}^{\text{cusp}}$ be the l-th Hecke operator on Jacobi forms defined by [3, page 41, (3)]. For $m = 1$, Kohnen and Skoruppa [6, page 549, Proposition (i)] calculated the Fourier coefficients of the adjoint operator $V_{1,l}^*$ of the operator $V_{1,l}$. For any $l \geqslant 1$, let $T_{s,l}$ denote the l-th Hecke operator on $S_k(\Gamma_2)$ and let ρ_l denote the l-th projection of any element in $S_k(\Gamma_2)$ to its l-th Fourier–Jacobi coefficient in $J_{k,l}^{\text{cusp}}$. By using a result of Kohnen and Skoruppa [6, page 541, Theorem 2], one gets the following interesting identity [7, Lemma 2.1]. For any $F \in S_k(\Gamma_2)$ and $l \geqslant 1$, we have

(1) $$F \,|\, T_{s,l} | \rho_1 = F \,|\, \rho_\ell \,|\, V_{1,l}^*.$$

Note that the above identity gives the first Fourier–Jacobi coefficient of the image of the l-th Hecke operator $T_{s,l}$. Manickam [7] used this identity crucially to establish the nonvanishing of the first Fourier–Jacobi coefficient of any Siegel eigenform in $S_k(\Gamma_2)$. By using this nonvanishing result, the identity (1) gives us the following result. For any Siegel eigenform $F \in S_k(\Gamma_2)$ and any $l \geqslant 1$, we have

(2) $$F \,|\, \rho_l = 0 \implies \lambda_F(l) = 0,$$

where $F \,|\, T_{s,l} = \lambda_F(l)\, F$. In this article, we investigate the converse of (2) and its interesting consequences through certain important generalizations.

We first calculate the Fourier coefficients of the adjoint operator $V_{m,l}^*$, which generalizes the above mentioned result of Kohnen and Skoruppa [6, page 549, Proposition (i)] to any index $m \geqslant 1$. Our approach is quite different from the one taken in the literature.

Proposition 1.1. *Let $\phi \in J_{k,ml}^{\text{cusp}}$ be a Jacobi cusp form with the Fourier series expansion*

$$\phi(\tau, z) = \sum_{\substack{n,r\in\mathbb{Z} \\ r^2<4lmn}} c_\phi(n, r)\, q^n \xi^r, \quad q = e^{2\pi i\tau},\ \xi = e^{2\pi i z}.$$

Then we have

$$\phi \,|\, V_{m,l}^*(\tau, z) = \sum_{\substack{n,r\in\mathbb{Z} \\ r^2<4mn}} c_{\phi | V_{m,l}^*}(n, r)\, q^n \xi^r,$$

where

$$c_{\phi | V_{m,l}^*}(n, r) := \sum_{d\,|\,l} d^{k-2} \sum_{\substack{s=0 \\ d\,|\,(ms^2+rs+n)}}^{d-1} c_\phi\left(\frac{(ms^2+rs+n)\,l}{d^2}, \frac{(r+2ms)\,l}{d}\right).$$

Let $J_{k,m}^{\text{cusp, new}}$ denote the space of Jacobi cusp newforms of weight k and index m, considered and studied extensively in [11, page 138]. As a consequence of Proposition 1.1 we derive the following identity of the operators on $J_{k,m}^{\text{cusp, new}}$ which generalizes the result of Kohnen and Skoruppa [6, page 549, Proposition (ii)] in the index 1 case to any arbitrary index $m \geqslant 1$. We follow the steps of the proof sketched in the index 1 case with appropriate modifications. For the sake of completion and for the benefit of the readers we provide the proof in Section 3 highlighting the main steps.

Proposition 1.2. *Let $\phi \in J_{k,m}^{\text{cusp, new}}$ and l be any positive integer coprime to m. Then*

$$\phi \,|\, V_{m,l}\, V_{m,l}^{*} = \phi \Big| \sum_{d|l} d^{k-2}\, \psi(d)\, T_{J,(l/d)},$$

where $\psi(d) = d \prod_{p|d}\big(1+\frac{1}{p}\big)$.

Now, we generalize the identity (1) to get the m-th Fourier–Jacobi coefficient of the image of Siegel cusp forms under the Hecke operator T_{s,p^δ}, where p is a prime and δ is a positive integer.

Theorem 1.3. *Let $F \in S_k(\Gamma_2)$ and p be any prime. Then for any two positive integers δ and m with $p \nmid m$, we have*

$$F \,|\, T_{s,p^\delta} \,|\, \rho_m = F \,|\, \rho_{mp^\delta} | V_{m,p^\delta}^{*}. \tag{3}$$

Also, we have the following two identities:

$$F \,|\, T_{s,p} \,|\, \rho_p = F \,|\, \rho_{p^2} \,|\, V_{p,p}^{*} + p^{k-2}\, F \,|\, \rho_1 \,|\, V_{1,p} \tag{4}$$

and

$$F \,|\, T_{s,p^2} \,|\, \rho_p = F \,|\, \rho_{p^3} | V_{p,p^2}^{*} + p^{k-2} F \,|\, \rho_p \,|\, T_{J,p} + p^{2k-4} F \,|\, \rho_p. \tag{5}$$

Note that the algebra of the Hecke operators acting on the space $S_k(\Gamma_2)$ is generated by $T_{s,p}$ and T_{s,p^2}, where p varies over primes. Using the fact that the operator $V_{1,p} : J_{k,1}^{\text{cusp}} \to J_{k,p}^{\text{cusp}}$ is injective together with the identity (4), we have:

Corollary 1.4. *Let p be any prime. For any Siegel eigenform $F \in S_k(\Gamma_2)$ at least one of the Fourier–Jacobi coefficients $F \,|\, \rho_p$ and $F \,|\, \rho_{p^2}$ is nonzero.*

For any Siegel eigenform $F \in S_k(\Gamma_2)$, we have $F \,|\, \rho_{p^2} \,|\, V_{1,p^2}^{*} = \lambda_F(p^2)\, F \,|\, \rho_1$ from (1). On the other hand, by applying $V_{1,p}^{*}$ on both sides of the identity (4) and then by using (1) together with Proposition 1.2, we get:

Corollary 1.5. *Let $F \in S_k(\Gamma_2)$ and p be any prime such that $F \,|\, T_{s,p} = \lambda_F(p)\, F$. Then we have $F \,|\, \rho_{p^2} \,|\, V_{p,p}^{*} V_{1,p}^{*} = (\lambda_F^2(p) - p^{2k-3} - p^{2k-4})\, F \,|\, \rho_1 - p^{k-2}\, F \,|\, \rho_1 \,|\, T_{J,p}$.*

Our next result shows that any nonzero Fourier–Jacobi coefficient of odd, squarefree index of a Siegel cusp form cannot be a newform. In particular, we prove the following theorem.

Theorem 1.6. *Let $F \in S_k(\Gamma_2)$ and $m \geqslant 3$ be any odd, squarefree integer. If $F \,|\, \rho_m \in J_{k,m}^{cusp,new}$ then $F \,|\, \rho_m = 0$.*

Our next result shows that the converse of (2) is also true for l odd, squarefree. More precisely, we prove:

Theorem 1.7. *Let $F \in S_k(\Gamma_2)$ be a Siegel eigenform with n-th eigenvalue $\lambda_F(n)$. Then for any odd, squarefree positive integer m, we have*

$$\lambda_F(m) = 0 \iff F \mid \rho_m = 0.$$

Remark 1.8. The reverse direction $\Leftarrow$ of Theorem 1.7 follows from [7] (see (2)) and that only direction $\Rightarrow$ is proved here. To establish $\Longrightarrow$ part for a given odd, squarefree positive integer m we require the Siegel cusp form to be eigenvector only for the Hecke operators $T_{s,l}$ with $l \mid m$. Also, we only need Proposition 1.1, the identity (3) of Theorem 1.3 and Theorem 1.6, not any other result stated above.

By using the multiplicative property of the eigenvalues of a Siegel eigenform together with Theorem 1.7, we get:

Corollary 1.9. *Let $F \in S_k(\Gamma_2)$ be a Siegel eigenform. Then for any odd prime p, we have*

$$F \mid \rho_p = 0 \implies F \mid \rho_m = 0$$

for any odd, squarefree positive integers m with $p \mid m$.

If m is any positive integer such that $\lambda_F(m) \neq 0$ then (2) implies the existence of infinitely many symmetric, half-integral, positive definite matrices T such that the quadratic form T represents m and $a_F(T) \neq 0$. Conversely, we establish the following two corollaries of Theorem 1.7 assuring the nonvanishing of certain eigenvalues.

Corollary 1.10. *Let $F \in S_k(\Gamma_2)$ be a Siegel eigenform with n-th eigenvalue $\lambda_F(n)$ and T be a symmetric, half-integral, positive definite matrix such that the T-th Fourier coefficient $a_F(T) \neq 0$. If m is any odd, squarefree, positive integer represented by the quadratic form T then $\lambda_F(m) \neq 0$.*

Corollary 1.11. *Let $F \in S_k(\Gamma_2)$ be a Siegel eigenform with n-th eigenvalue $\lambda_F(n)$. Then there exists a positive integer $1 \leqslant n \leqslant \frac{k}{2} - 2$ such that for any odd, squarefree, positive integer m of the form $x^2 + ny^2$ we have $\lambda_F(m) \neq 0$.*

Remark (concluding remark). One may ask more generally about the nonvanishing of the m-th eigenvalue $\lambda_F(m)$ of a Siegel eigenform F if its m-th Fourier–Jacobi coefficient $F \mid \rho_m$ is nonzero. In this paper, we answer it affirmatively for any odd, squarefree m but could not address this question for arbitrary m. However, the intermediate results obtained by us while addressing the question highlight the importance of the theory of Jacobi forms and provide better understanding of certain Hecke-type operators on Jacobi forms.

The question of nonvanishing of Fourier–Jacobi coefficients of Siegel cusp forms of arbitrary degree and eigenvalues of Siegel eigenforms of degree 2 is also considered in [2]. However, the results obtained there are of different nature and do not address the question asked here in this paper.

2. Prerequisites

We refer to [1], [3] and [10] for definitions and basic properties of Jacobi and Siegel modular forms. In this section we fix notation and recall certain results.

Jacobi forms. Let G^J be the group of triplets $[M, X, \xi]$, $M \in \mathrm{SL}_2(\mathbb{R})$, $X \in \mathbb{R}^2$, $\xi \in \mathbb{C}$ with $|\xi| = 1$, via the multiplication

$$[M, X, \xi][M', X', \xi'] = [MM', XM' + X', \xi\xi' e^{2\pi i \det\left(\begin{smallmatrix} XM' \\ X' \end{smallmatrix}\right)}].$$

The group G^J acts on the set of functions $\{\phi : \mathbb{H} \times \mathbb{C} \to \mathbb{C}\}$ as

$$\phi|_{k,m}\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}, (\lambda, \mu), \xi\right](\tau, z) = \xi^m (c\tau + d)^{-k} e^{2\pi i m\left(-\frac{c(z+\lambda\tau+\mu)^2}{c\tau+d} + \lambda^2\tau + 2\lambda z + \lambda\mu\right)} \phi\left(\frac{a\tau + b}{c\tau + d}, \frac{z + \lambda\tau + \mu}{c\tau + d}\right).$$

We consider the action of the discrete subgroup $\mathrm{SL}_2(\mathbb{Z}) \ltimes (\mathbb{Z} \times \mathbb{Z})$ of G^J on the set of functions on $\mathbb{H} \times \mathbb{C}$ by fixing $\xi = 1$. Let $J_{k,m}$ (resp. $J_{k,m}^{\mathrm{cusp}}$) denote the space of Jacobi forms (resp. Jacobi cusp forms) of weight k and index m for the group $\mathrm{SL}_2(\mathbb{Z}) \ltimes (\mathbb{Z} \times \mathbb{Z})$. For any $l \geqslant 1$, let U_l, V_l and T_l be the operators acting on $J_{k,m}$ defined and studied systematically in [3, Section 4]. We are denoting them respectively by $U_{m,l}$, $V_{m,l}$ and $T_{J,l}$ throughout the paper to avoid certain potential confusions. The operator $T_{J,l} : J_{k,m} \to J_{k,m}$ is called the l-th Hecke operator on Jacobi forms.

Any $\phi(\tau, z) \in J_{k,m}^{\mathrm{cusp}}$ with Fourier series expansion

$$\phi(\tau, z) = \sum_{n, r \in \mathbb{Z},\, r^2 < 4mn} c_\phi(n, r)\, q^n \xi^r, \qquad q = e^{2\pi i\tau},\ \xi = e^{2\pi i z},$$

admits the following theta decomposition [3, pages 58–59]:

$$\phi(\tau, z) = \sum_{\mu=0}^{2m-1} h_\mu(\tau)\, \theta_{m,\mu}(\tau, z), \tag{6}$$

where

$$h_\mu(\tau) := \sum_{\substack{N \geqslant 1 \\ N \equiv -\mu^2 (\mathrm{mod}\, 4m)}} c_\phi\left(\frac{N + \mu^2}{4m}, \mu\right) q^{N/4m}, \qquad \theta_{m,\mu}(\tau, z) := \sum_{\substack{r \in \mathbb{Z} \\ r \equiv \mu (\mathrm{mod}\, 2m)}} q^{r^2/4m} \xi^r.$$

By using the transformation law of the Jacobi form ϕ and the Jacobi theta functions $\theta_{m,\mu}$ with respect to the inversion $(\tau, z) \to \left(-\frac{1}{\tau}, \frac{z}{\tau}\right)$, we get

$$h_\mu\left(\frac{-1}{\tau}\right) = \frac{\tau^k}{\sqrt{2m\tau/i}} \sum_{\nu=0}^{2m-1} e^{\pi i \mu\nu/m}\, h_\nu(\tau). \tag{7}$$

Let $J_{k,m}^{\text{cusp, new}}$ be the space of Jacobi cusp newforms considered in [11, page 138] giving the direct sum decomposition

$$J_{k,m}^{\text{cusp}} = J_{k,m}^{\text{cusp, new}} \overset{\perp}{\bigoplus} \Bigg(\bigoplus_{\substack{l \geqslant 1,\, d \geqslant 1 \\ ld^2 \mid m,\, ld^2 > 1}} J_{k,(m/ld^2)}^{\text{cusp, new}} \mid U_{m/ld^2,\,d}\, V_{m/l,\,l} \Bigg). \tag{8}$$

Note that the first direct sum in the above decomposition is orthogonal. If m is squarefree, then for any divisor $l > 1$ of m there is only one copy of the newforms space of index m/l given by $J_{k,(m/l)}^{\text{cusp, new}} \mid V_{(m/l),l}$ in the oldforms direct sum decomposition. By using the Shimura correspondence and the Atkin–Lehner theory for modular forms on the congruence subgroups $\Gamma_0(N)$, we get that for squarefree index m all the direct sums in the above decomposition (8) are orthogonal with respect to the Petersson inner product. For a detailed proof of this fact we refer to [5, Lemma 4]. In [8, Section 5.1], the space of Jacobi cusp newforms has been defined differently but in [7, page 406] it is observed that this newforms space is same as the one considered earlier in [11]. To prove Theorem 1.6, we use an important property of newforms [8, Corollary 5.3] saying that the (n,r)-th Fourier coefficient $c_\phi(n,r)$ of a Jacobi cusp form $\phi \in J_{k,m}^{\text{cusp, new}}$ depends only on the discriminant $r^2 - 4mn$ and not on $r \pmod{2m}$.

Siegel modular forms. The real symplectic unimodular group of degree 2 is defined by

$$\mathrm{Sp}_4(\mathbb{R}) = \{M \in \mathrm{GL}_4(\mathbb{R}) : MJ\,{}^tM = J\},$$

where $J = \left(\begin{smallmatrix} 0_2 & I_2 \\ -I_2 & 0_2 \end{smallmatrix}\right)$, ${}^tM$ denotes the transpose matrix of the matrix M, 0_2 is the 2×2 zero matrix and I_2 is the 2×2 identity matrix. Let $\Gamma_2 := \mathrm{Sp}_4(\mathbb{Z})$ be the subgroup of $\mathrm{Sp}_4(\mathbb{R})$ consisting of matrices with integer entries. Let

$$\mathbb{H}_2 := \{Z \in M_2(\mathbb{C}) : Z = {}^tZ,\ \mathrm{Im}(Z) > 0\}$$

be the Siegel upper half-space of degree 2. We denote the space of Siegel modular forms (resp. cusp forms) on $\mathbb{H}_2$ of weight k for the group Γ_2 by $M_k(\Gamma_2)$ (resp. $S_k(\Gamma_2)$). There is an algebra of Hecke operators acting on the space $M_k(\Gamma_2)$ which preserves $S_k(\Gamma_2)$. For any $l \geqslant 1$, let $T_{s,l}$ denote the l-th Hecke operator on $S_k(\Gamma_2)$. An element in $S_k(\Gamma_2)$ is called a *Siegel eigenform* if it is a common eigenvector of all the Hecke operators $T_{s,l},\ l \geqslant 1$. Note that the space $S_k(\Gamma_2)$ is a Hilbert space under the Petersson inner product.

Any $F \in S_k(\Gamma_2)$ has the Fourier series expansion of the form

$$F(Z) = \sum_T a_F(T)\, e^{2\pi i\, \mathrm{trace}(TZ)},$$

where the sum varies over the set of symmetric, half-integral, positive definite 2×2 matrices. Writing $Z = \left(\begin{smallmatrix} \tau & z \\ z & \tau' \end{smallmatrix}\right) \in \mathbb{H}_2$, where τ, τ' are in the complex upper half-plane $\mathbb{H}$ and $z \in \mathbb{C}$, we get the following Fourier–Jacobi decomposition [3, Theorem 6.1]:

$$F(Z) = F(\tau, z, \tau') = \sum_{m \geqslant 1} \phi_m(\tau, z)\, e^{2\pi i m \tau'},$$

where

$$\phi_m(\tau, z) := \sum_{\substack{n, r \in \mathbb{Z} \\ r^2 < 4nm}} A_F \begin{pmatrix} n & r/2 \\ r/2 & m \end{pmatrix} q^n \xi^r$$

belongs to the space $J_{k,m}^{\mathrm{cusp}}$ and is called the m-th Fourier–Jacobi coefficient of F.

3. Proof of Propositions 1.1 and 1.2

Proof of Proposition 1.1. Let l, m be any two positive integers. Let $\Gamma := \mathrm{SL}_2(\mathbb{Z})$. On the space $J_{k,m}^{\mathrm{cusp}}$, the index changing Hecke operator $V_{m,l}$ is defined by

$$\begin{aligned} \phi \,|\, V_{m,l}(\tau, z) &:= l^{k-1} \sum_{\substack{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma \backslash M_2(\mathbb{Z}) \\ ad - bc = l}} (c\tau + d)^{-k} e^{ml\left(\frac{-cz^2}{c\tau + d}\right)} \phi\left(\frac{a\tau + b}{c\tau + d}, \frac{lz}{c\tau + d}\right) \\ &= l^{(k/2)-1} \sum_{\substack{\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in M_2(\mathbb{Z}) \\ ad = l,\, b \pmod d}} \phi_{\sqrt{l}} \Big|_{k,ml} \left[\frac{1}{\sqrt{l}} \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}, (0, 0), 1\right](\tau, z), \end{aligned}$$

where $\phi_{\sqrt{l}}(\tau, z) := \phi(\tau, \sqrt{l}\, z)$. To prove our claim, first we calculate the image of Jacobi Poincaré series $P_{k,m;n,r}$, $n, r \in \mathbb{Z}$ with $r^2 - 4mn < 0$, under the operator $V_{m,l}$. By using the definition of Jacobi Poincaré series, we have

$$\begin{aligned} &P_{k,m;n,r} \,|\, V_{m,l} \\ &= \sum_{\substack{\left(\begin{smallmatrix} \alpha & \beta \\ 0 & \delta \end{smallmatrix}\right) \in M_2(\mathbb{Z}) \\ \alpha\delta = l,\, \beta (\mathrm{mod}\, \delta)}} l^{(k/2)-1} \Bigg(\sum_{\substack{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_\infty \backslash \Gamma \\ \lambda \in \mathbb{Z}}} e(n\tau + rz) \Big|_{k,m} \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}, (\lambda a, \lambda b), 1\right] \Bigg)_{\sqrt{l}} \Bigg|_{k,ml} \\ &\qquad\qquad \left[\frac{1}{\sqrt{l}} \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}, (0, 0), 1\right]. \end{aligned}$$

Using the definition of $\phi_{\sqrt{l}}(\tau, z) = \phi(\tau, \sqrt{l}\, z)$ and then adjusting the stroke operators in the inner sum, we obtain

$$\begin{aligned} P_{k,m;n,r} \,|\, V_{m,l} = \sum_{\substack{\left(\begin{smallmatrix} \alpha & \beta \\ 0 & \delta \end{smallmatrix}\right) \in M_2(\mathbb{Z}) \\ \alpha\delta = l,\, \beta\, (\mathrm{mod}\, \delta)}} l^{(k/2)-1} \sum_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_\infty \backslash \Gamma,\, \lambda \in \mathbb{Z}} e(n\tau + r\sqrt{l} z) \Big|_{k,ml} \\ \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \left(\frac{\lambda a}{\sqrt{l}}, \frac{\lambda b}{\sqrt{l}}\right), 1\right] \left[\frac{1}{\sqrt{l}} \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}, (0, 0), 1\right]. \end{aligned}$$

In the Jacobi group $\mathrm{SL}_2(\mathbb{R}) \ltimes (\mathbb{R}^2 \times \mathbb{S}^1)$, where $\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$, we have

$$\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \left(\frac{\lambda a}{\sqrt{l}}, \frac{\lambda b}{\sqrt{l}}\right), 1\right]\left[\frac{1}{\sqrt{l}}\begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}, (0,0), 1\right] = \left[\frac{1}{\sqrt{l}}\begin{pmatrix} \alpha' & \beta' \\ 0 & \delta' \end{pmatrix}, \left(\frac{\lambda\alpha'}{l}, \frac{\lambda\beta'}{l}\right), 1\right]\left[\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}, (0,0), 1\right]$$

for some $\left(\begin{smallmatrix} \alpha' & \beta' \\ 0 & \delta' \end{smallmatrix}\right) \in M_2(\mathbb{Z})$ with $\alpha'\delta' = l$ and $\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \Gamma$ with the following crucial property. These matrices vary over a complete set of representatives of the indexing sets in the above summation as the matrices $\left(\begin{smallmatrix} \alpha & \beta \\ 0 & \delta \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ vary over the same, respectively. Using this and interchanging the order of the summations, we get

$$P_{k,m;n,r} \,|\, V_{m,l} = \sum_{\substack{\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \Gamma_\infty \backslash \Gamma \\ \lambda \in \mathbb{Z}}} l^{(k/2)-1} \sum_{\substack{\left(\begin{smallmatrix} \alpha' & \beta' \\ 0 & \delta' \end{smallmatrix}\right) \in M_2(\mathbb{Z}) \\ \alpha'\delta' = l,\, \beta' \ (\mathrm{mod}\ \delta)'}} e(n\tau + r\sqrt{l}z)\Big|_{k,ml} \left[\frac{1}{\sqrt{l}}\begin{pmatrix} \alpha' & \beta' \\ 0 & \delta' \end{pmatrix}, \left(\frac{\lambda\alpha'}{l}, \frac{\lambda\beta'}{l}\right), 1\right]\left[\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}, (0,0), 1\right].$$

For any $\lambda \in \mathbb{Z}$, we write $\lambda = \lambda'\delta' + s$ with $s \pmod{\delta}'$. Then λ' varies over $\mathbb{Z}$ and s varies over a complete residue system mod δ'. Therefore, we have

$$P_{k,m;n,r} \,|\, V_{m,l} = \sum_{\substack{\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \Gamma_\infty \backslash \Gamma \\ \lambda' \in \mathbb{Z}}} l^{(k/2)-1} \sum_{\substack{\delta' \,|\, l,\, \beta' \ (\mathrm{mod}\ \delta)' \\ s \ (\mathrm{mod}\ \delta)'}} e(n\tau + r\sqrt{l}z)\Big|_{k,ml} \left[\frac{1}{\sqrt{l}}\begin{pmatrix} l/\delta' & \beta' \\ 0 & \delta' \end{pmatrix}, \left(\frac{s}{\delta'}, \frac{(s+\lambda'\delta')\beta'}{l}\right), 1\right]\left[\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}, (\lambda' a', \lambda' b'), 1\right].$$

Let us first simplify the inner sum. We set

$$\begin{aligned} I_{k,m;l} &:= \sum_{\substack{\delta | l, \beta (\mathrm{mod}\ \delta) \\ s (\mathrm{mod}\ \delta)}} e(n\tau + r\sqrt{\ell}z)\Big|_{k,ml} \left[\frac{1}{\sqrt{l}}\begin{pmatrix} l/\delta & \beta \\ 0 & \delta \end{pmatrix}, \left(\frac{s}{\delta}, \frac{(s+\lambda'\delta)\beta}{l}\right), 1\right] \\ &= l^{k/2} \sum_{\substack{\delta | l \\ s (\mathrm{mod}\ \delta)}} \delta^{-k} e\left(\left(\frac{l}{\delta^2}(ms^2 + rs + n)\,\tau\right) + \left(\frac{l}{\delta}(r + 2sm)\,z\right)\right) \sum_{\beta (\mathrm{mod}\ \delta)} e\left(\frac{\beta}{\delta}(ms^2 + rs + n)\right) \\ &= l^{k/2} \sum_{\substack{\delta \,|\, l, s (\mathrm{mod}\ \delta) \\ \delta \,|\, (ms^2 + rs + n)}} \delta^{-k+1} e\left(\left(\frac{l}{\delta^2}(ms^2 + rs + n)\,\tau\right) + \left(\frac{l}{\delta}(r + 2sm)\,z\right)\right) \end{aligned}$$

Therefore, we have

$$\begin{aligned}
&P_{k,m;n,r} \mid V_{m,l} \\
&= \sum_{\delta \mid l} \left(\frac{l}{\delta}\right)^{k-1} \sum_{\substack{s \,(\mathrm{mod}\ \delta) \\ \delta \mid (ms^2+rs+n)}} \sum_{\substack{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_\infty \backslash \Gamma \\ \lambda \in \mathbb{Z}}} \\
&\qquad e\left(\left(\frac{l}{\delta^2}(ms^2+rs+n)\,\tau\right) + \left(\frac{l}{\delta}(r+2sm)\,z\right)\right)\bigg|_{k,ml} \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}, (\lambda a, \lambda b), 1\right] \\
&= \sum_{\delta \mid l} \left(\frac{l}{\delta}\right)^{k-1} \sum_{\substack{s \,(\mathrm{mod}\ \delta) \\ \delta \mid (ms^2+rs+n)}} P_{k,ml;\, \frac{(ms^2+rs+n)\,l}{\delta^2},\, \frac{(r+2sm)\,l}{\delta}}.
\end{aligned}$$

Next, we have

$$\begin{aligned}
&c_{\phi \mid V^*_{m,l}}(n,r) \\
&\quad = \frac{(2\pi(4mn-r^2))^{k-3/2}}{2^{k-5/2}\Gamma(k-3/2)\, m^{k-2}} \langle \phi \mid V^*_{m,l},\, P_{k,m;n,r} \rangle \\
&\quad = \frac{(2\pi(4mn-r^2))^{k-3/2}}{2^{k-5/2}\Gamma(k-3/2)\, m^{k-2}} \left\langle \phi,\, \sum_{\delta \mid l} \left(\frac{l}{\delta}\right)^{k-1} \sum_{\substack{s \,(\mathrm{mod}\ \delta) \\ \delta \mid (ms^2+rs+n)}} P_{k,ml;\, \frac{(ms^2+rs+n)\,l}{\delta^2},\, \frac{(r+2sm)\,l}{\delta}} \right\rangle \\
&\quad = \sum_{d \mid l} d^{k-2} \sum_{\substack{s \,(\mathrm{mod}\ d) \\ d \mid (ms^2+rs+n)}} c_\phi\left(\frac{(ms^2+rs+n)\,l}{d^2},\, \frac{(r+2sm)\,l}{d}\right). \qquad \square
\end{aligned}$$

Proof of Proposition 1.2. For all the facts used in this proof about the operators $T_{J,l}$, $V_{m,l}$ and the space $J_{k,m}^{\text{cusp, new}}$, we refer to [3; 11]. Since l and m are coprime, the right-hand side operator

$$\mathbb{T}_{J,l} := \sum_{d \mid l} d^{k-2}\, \psi(d)\, T_{J,(l/d)}$$

is multiplicative. Moreover, the operator $V_{m,l}$ is multiplicative and the Hecke operator $T_{J,n}$ commutes with the operator $V_{m,l}$ if $\gcd(n, lm) = 1$. Therefore it is enough to establish the identity for prime powers, that is, $l = p^\alpha$, where p is a prime and α is any positive integer. Since the space $J_{k,m}^{\text{cusp, new}}$ has a basis of simultaneous eigenfunctions of all the Hecke operators $T_{J,n}$ with $\gcd(n,m) = 1$, it is enough to check the identity for such eigenforms. Let $\varphi \in J_{k,m}^{\text{cusp, new}}$ be any such eigenform. The Hecke operators $T_{J,n}$ with $\gcd(n,m) = 1$ are hermitian and commute with $T_{J,l'}$ and $V_{m,t}$ for $\gcd(nl', m) = 1$ and $\gcd(n, mt) = 1$. Therefore the Jacobi forms $\varphi \mid V_{m,l}\, V^*_{m,l}$ and $\varphi \mid \mathbb{T}_l$ are again simultaneous eigenfunctions of all the Hecke operators $T_{J,n}$ for $\gcd(n, lm) = 1$ with eigenvalues same as of φ. By using multiplicity one result, we get that the Jacobi forms $\varphi \,|\, V_{m,l}\, V_{m,l}^{*}$ and $\varphi \,|\, \mathbb{T}_l$ both are constant multiples of φ. To show that both are same we compare their (n, r)-th Fourier coefficients with the condition that $r^2 - 4mn$ is a fundamental discriminant and prove that they are equal. We have

$$c_{\varphi \,|\, V_{m,l}\, V_{m,l}^{*}}(n, r) = \sum_{t \,|\, l} t^{k-1} \Biggl(\sum_{d \,|\, t} \frac{1}{d} \sum_{\substack{s=0 \\ d \,|\, ms^2+rs+n}}^{d-1} 1 \Biggr) c_{\varphi}\Bigl(n\frac{l^2}{t^2}, r\frac{l}{t} \Bigr). \tag{9}$$

The cardinality of the set

$$\{s \pmod d : ms^2 + rs + n \equiv 0 \pmod d\}$$

is same as the cardinality of the set $\{x \pmod{2d} : x^2 \equiv (r^2 - 4mn) \pmod{4d}\}$. Let us denote this cardinality by $N_d(r^2 - 4mn)$. By using [3, page 50, (16)], we have

$$c_{\varphi}\Bigl(n\frac{l^2}{t^2}, r\frac{l}{t} \Bigr) = \sum_{\delta \,|\, (l/t)} \mu(\delta)\, \chi_D(\delta)\, \delta^{k-2}\, c_{\varphi \,|\, T_{J,(l/\delta t)}}(n, r),$$

where $D = r^2 - 4mn$ and χ_D denotes the Dirichlet character $\bigl(\frac{D}{\cdot}\bigr)$. By using the above observations in (9), we see that it is sufficient to prove the following formal identity of the operators:

$$\sum_{t \,|\, l} t^{k-1} \sum_{d \,|\, t} \frac{N_d(D)}{d} \sum_{\delta \,|\, (l/t)} \mu(\delta)\, \chi_D(\delta)\, \delta^{k-2}\, T_{J,\,(l/\delta t)} = \sum_{t \,|\, l} t^{k-2}\, \psi(t)\, T_{J,\,(l/t)}. \tag{10}$$

Since D is a fundamental discriminant, by using [3, page 21, (6)] we get that

$$N_p(D) = (1 + \chi_D(p)) \quad \text{and} \quad N_{p^a}(D) = N_p(D)$$

for any prime p, positive integer a. By using these facts we get that the coefficients of T_{J,p^a}, $1 \leqslant a \leqslant \alpha$, in both sides of (10) are equal. □

4. Proof of Theorem 1.3

We prove the identities by equating the Fourier coefficients on both sides. First let us write down the Fourier coefficients of $F \,|\, T_{s,p^\delta}$, where p is a prime and δ is a positive integer [9, Corollaries 2.2, 2.4 and 2.5]. For any positive integer l and any finite sequence of integers $\{a_1, a_2, \ldots, a_n\}$, we use the notation $\delta_l(a_1, a_2, \ldots, a_n)$ defined to be 1 if $l \,|\, \gcd(a_1, a_2, \ldots, a_n)$ and 0 otherwise. Let

$$F(Z) = \sum_{T = \left(\begin{smallmatrix} n & r/2 \\ r/2 & m \end{smallmatrix}\right) > 0} A_F(n, r, m)\, e^{2\pi i \,\mathrm{trace}(TZ)},$$

where $T > 0$ indicates that T is positive definite. Then we have

$$A_{F\,|\,T_{s,p}}(n,r,m) = A_F(pn, pr, pm) + p^{k-2}\Bigg(\sum_{\substack{\alpha=0 \\ p|(n+r\alpha+m\alpha^2)}}^{p-1} A_F\Big(\frac{n+r\alpha+m\alpha^2}{p}, r+2m\alpha, mp\Big) + \delta_p(m)\, A_F\Big(np, r, \frac{m}{p}\Big)\Bigg) + p^{2k-3}\delta_p(n,r,m)\, A_F\Big(\frac{n}{p}, \frac{r}{p}, \frac{m}{p}\Big). \tag{11}$$

Also, we have

$$\begin{aligned} &A_{F\,|\,T_{s,p^2}}(n,r,m) \\ &= A_F(p^2n, p^2r, p^2m) + p^{2k-3}\delta_p(n,r,m)\, A_F(n,r,m) + p^{4k-6}\delta_{p^2}(n,r,m) A_F\Big(\frac{n}{p^2}, \frac{r}{p^2}, \frac{m}{p^2}\Big) \\ &\quad + p^{k-2}\Bigg(\sum_{\substack{\alpha=0 \\ p|(n+r\alpha+m\alpha^2)}}^{p-1} A_F\Big(p\Big(\frac{n+r\alpha+m\alpha^2}{p}, r+2m\alpha, mp\Big)\Big) + \delta_p(m)\, A_F\Big(p\Big(np, r, \frac{m}{p}\Big)\Big)\Bigg) \\ &\quad + p^{2k-4}\Bigg(\sum_{\substack{\alpha=0 \\ p^2\,|\,(n+r\alpha+m\alpha^2)}}^{p^2-1} A_F\Big(\frac{n+r\alpha+m\alpha^2}{p^2}, r+2m\alpha, mp^2\Big) + \sum_{\substack{\beta=0 \\ p^2\,|\,(rp\beta+m)}}^{p-1} A_F\Big(np^2, r+2np\beta, n\beta^2+\frac{rp\beta+m}{p^2}\Big)\Bigg) \\ &\quad + p^{3k-5}\Bigg(\sum_{\substack{\alpha=0 \\ p^2\,|\,(n+r\alpha+m\alpha^2),\, p\,|\,r,\, p\,|\,m}}^{p-1} A_F\Big(\frac{n+r\alpha+m\alpha^2}{p^2}, \frac{r+2m\alpha}{p}, m\Big) + \delta_p(n,r)\,\delta_{p^2}(m)\, A_F\Big(n, \frac{r}{2p}, \frac{m}{p^2}\Big)\Bigg). \end{aligned} \tag{12}$$

If $p \nmid m$ then we have

$$\begin{aligned} &A_{F\,|\,T_{s,p^\delta}}(n,r,m) \\ &= A_F(p^\delta(n,r,m)) + \sum_{\beta=1}^{\delta} p^{(k-2)\beta}\Bigg(\sum_{\substack{\alpha=0 \\ p^\beta\,|(n+r\alpha+m\alpha^2)}}^{p^\beta-1} A_F\big(p^{\delta-\beta}((n+r\alpha+m\alpha^2)p^{-\beta}, r+2m\alpha, mp^\beta)\big)\Bigg). \end{aligned} \tag{13}$$

First we compare the coefficients of both sides of (3). Let (n, r) be any pair of integers with $r^2 < 4mn$. By using (13), we have

$$\begin{aligned}(14)\quad & c_{F\,|\,T_{s,p^\delta}\,|\,\rho_m}(n,r)\\ &= A_{F\,|\,T_{s,p^\delta}}(n,r,m)\\ &= A_F(p^\delta(n,r,m))\\ &\quad+\sum_{\beta=1}^{\delta} p^{(k-2)\beta}\Bigg(\sum_{\substack{\alpha=0\\ p^\beta\,|(n+r\alpha+m\alpha^2)}}^{p^\beta-1} A_F\big((n+r\alpha+m\alpha^2)p^{\delta-2\beta},(r+2m\alpha)p^{\delta-\beta},mp^\delta\big)\Bigg).\end{aligned}$$

By using Proposition 1.1, we get

$$\begin{aligned}(15)\quad & c_{F\,|\,\rho_{mp^\delta}\,|\,V^*_{m,p^\delta}}(n,r)\\ &= c_{F\,|\,\rho_{mp^\delta}}(p^\delta n,\, p^\delta r)\\ &\quad+\sum_{\beta=1}^{\delta} p^{(k-2)\beta}\Bigg(\sum_{\substack{s=0\\ p^\beta\,|(n+rs+ms^2)}}^{p^\beta-1} c_{F\,|\,\rho_{mp^\delta}}\big((n+rs+ms^2)p^{\delta-2\beta},\,(r+2ms)p^{\delta-\beta}\big)\Bigg).\end{aligned}$$

Now by comparing (14) and (15) we get that $F\,|\,T_{s,p^\delta}\,|\,\rho_m = F\,|\,\rho_{mp^\delta}\,|\,V^*_{m,p^\delta}$.

Next we compare the coefficients of both sides of (4) and then of (5). By using (11), we have

$$\begin{aligned}(16)\quad & c_{F\,|\,T_{s,p}\,|\,\rho_p}(n,r)\\ &= A_{F|T_{s,p}}(n,r,p)\\ &= A_F(pn,\,pr,\,p^2)\\ &\qquad + p^{k-2}\Bigg(\sum_{\substack{\alpha=0\\ p|(n+r\alpha)}}^{p-1} A_F\bigg(\frac{n+r\alpha+p\alpha^2}{p},\,r+2p\alpha,\,p^2\bigg)+A_F(np,r,1)\Bigg)\\ &\qquad + p^{2k-3}\,\delta_p(n,r)\,A_F\bigg(\frac{n}{p},\frac{r}{p},1\bigg).\end{aligned}$$

By using Proposition 1.1, we have

$$(17)\quad c_{F\,|\,\rho_{p^2}|V^*_{p,p}}(n,r)=\sum_{d\,|\,p} d^{k-2}\sum_{\substack{s=0\\ d\,|\,(ps^2+rs+n)}}^{d-1} c_{F\,|\,\rho_{p^2}}\bigg(\frac{(ps^2+rs+n)p}{d^2},\,(r+2ps)\frac{p}{d}\bigg).$$

By using [3, Theorem 4.2, 7], we have

$$(18)\qquad c_{F\,|\,\rho_1|V_{1,p}}(n,r)=\sum_{d\,|\,(n,r,p)} d^{k-1} c_{F\,|\,\rho_1}\bigg(\frac{np}{d^2},\frac{r}{d}\bigg).$$

Comparing (16), (17) and (18), we get that

$$F\,|\,T_{s,p}\,|\,\rho_p = F\,|\,\rho_{p^2}\,|\,V^*_{p,p} + p^{k-2}\,F\,|\,\rho_1\,|\,V_{1,p}.$$

By using (12), we have

(19)
$$
\begin{aligned}
&c_{F\,|\,T_{s,p^2}\,|\,\rho_p}(n,r)\\
&= A_{F\,|\,T_{s,p^2}}(n,r,p)\\
&= A_F(p^2n,\,p^2r,\,p^3) + p^{2k-3}\,\delta_p(n,r)\,A_F(n,r,p)\\
&\quad + p^{k-2}\Bigg(\sum_{\substack{\alpha=0\\ p\,|\,(n+r\alpha)}}^{p-1} A_F\big(n+r\alpha+p\alpha^2,\,(r+2p\alpha)p,\,p^3\big) + A_F(np^2,\,rp,\,p)\Bigg)\\
&\quad + p^{2k-4}\Bigg(\sum_{\substack{\alpha=0\\ p^2\,|\,(n+r\alpha+p\alpha^2)}}^{p^2-1} A_F\Big(\frac{n+r\alpha+p\alpha^2}{p^2},\,r+2p\alpha,\,p^3\Big)\\
&\qquad\qquad + \sum_{\substack{\beta=0\\ p\,|\,(r\beta+1)}}^{p-1} A_F\Big(np^2,\,r+2np\beta,\,n\beta^2+\frac{r\beta+1}{p}\Big)\Bigg)\\
&\quad + p^{3k-5}\Bigg(\sum_{\substack{\alpha=0\\ p^2\,|\,(n+r\alpha+p\alpha^2),\,p\,|\,r}}^{p-1} A_F\Big(\frac{n+r\alpha+p\alpha^2}{p^2},\,\frac{r+2p\alpha}{p},\,p\Big)\Bigg).
\end{aligned}
$$

By using Proposition 1.1, we have

(20)
$$
c_{F\,|\,\rho_{p^3}\,|\,V^*_{p,p^2}}(n,r) = \sum_{d\,|\,p^2} d^{k-2} \sum_{\substack{s=0\\ d\,|\,(ps^2+rs+n)}}^{d-1} c_{F\,|\,\rho_{p^3}}\Big(\frac{(ps^2+rs+n)p^2}{d^2},\,(r+2ps)\frac{p^2}{d}\Big).
$$

By using [3, page 56, (24)], we have

(21)
$$
c_{F\,|\,\rho_p\,|\,T_p}(n,r) =
\begin{cases}
c_{F\,|\,\rho_p}(p^2n,\,pr) & \text{if } p\nmid r,\\
c_{F\,|\,\rho_p}(p^2n,\,pr) - p^{k-2}c_{F\,|\,\rho_p}(n,r) & \text{if } p\,|\,r,\ p\nmid n,\\
c_{F\,|\,\rho_p}(p^2n,\,pr) + p^{k-2}(p-1)\,c_{F\,|\,\rho_p}(n,r) & \\
\quad + p^{2k-3}\sum_{\alpha=0,\ p^2\,|\,(n+r\alpha+p\alpha^2)}^{p-1} c_{F\,|\,\rho_p}\big(\frac{n+r\alpha+p\alpha^2}{p^2},\,\frac{r+2p\alpha}{p}\big) & \text{if } p\,|\,r,\ p\,|\,n.
\end{cases}
$$

Suppose $p\nmid r$. Then there exists unique $\beta \in \{0, 1, \dots p-1\}$ such that $p\,|\,r\beta+1$. Suppose that $r\beta+1 = lp$ for some $l \in \mathbb{Z}$. Then we have

(22)
$$
\begin{pmatrix} -p & r\\ -\beta & l\end{pmatrix}\begin{pmatrix} n & r/2\\ r/2 & m\end{pmatrix}\begin{pmatrix} -p & -\beta\\ r & l\end{pmatrix} = \begin{pmatrix} np^2 & (r+2np\beta)/2\\ (r+2np\beta)/2 & n\beta^2+(r\beta+1)/p\end{pmatrix}.
$$

By comparing the three equations (19), (20), (21) and also using (22), we get that

$$
F\,|\,T_{s,p^2}\,|\,\rho_p = F\,|\,\rho_{p^3}\,|\,V^*_{p,p^2} + p^{k-2}F\,|\,\rho_p\,|\,T_{J,p} + p^{2k-4}F\,|\,\rho_p.
$$

5. Proof of Theorem 1.6

Suppose $F\,|\,\rho_m = \varphi_m \in J_{k,m}^{\text{cusp, new}}$. By using the Eichler–Zagier isomorphism Z_m given in [8, Theorem 5.4], we get that $\varphi_m\,|\,Z_m$ is in the space

$$S_{k-1/2}^{+,m,\text{new}}(4m) := \left\{ f \in S_{k-1/2}^{+,\text{ new}}(4m) : a_f(n) = 0 \text{ unless } (-1)^{k-1} n \equiv \square \pmod{4m} \right\},$$

where $S_{k-1/2}^{+,\text{ new}}(4m)$ is the subspace of newforms inside the Kohnen's plus space $S_{k-1/2}^{+}(4m)$ studied in [4]. Moreover, we have

$$a_{\varphi_m\,|\,Z_m}(|D|) = c_{\varphi_m}\left(\frac{r^2-D}{4m}, r\right)$$

for any $0 > D$, $r \in \mathbb{Z}$ with $D \equiv r^2 \pmod{4m}$. Let p be an odd prime dividing m and let U_{p^2} be the level dividing operator $\sum_{n\geqslant 1} a(n)\, q^n \mapsto \sum_{n\geqslant 1} a(p^2 n)\, q^n$. Since $\varphi_m\,|\,Z_m \in S_{k-1/2}^{+,\text{ new}}(4m)$, by using [4, Theorem 1] we get that

$$\varphi_m\,|\,Z_m\,|\,(U_{p^2} + p^{k-2} w_p) = 0,$$

where w_p is the involution operator $w_{p,k-1/2}^{m}$ defined in [4, Section 2, page 39]. From [8, Lemma 5.9] we know that w_p acts as the identity operator on $S_{k-1/2}^{+,m}(4m)$. Therefore we have $\varphi_m\,|\,Z_m\,|\,(U_{p^2} + p^{k-2}) = 0$. Hence for any $0 > D$, $r \in \mathbb{Z}$, $D \equiv r^2 \pmod{4m}$, we have

$$c_{\varphi_m}\left(p^2 \frac{r^2-D}{4m}, pr\right) + p^{k-2} c_{\varphi_m}\left(\frac{r^2-D}{4m}, r\right) = 0.$$

For any $n \geqslant 1$, $r \in \mathbb{Z}$ with $r^2 < 4mn$, by taking $D = r^2 - 4mn$ we have

$$c_{\varphi_m}(p^2 n, pr) + p^{k-2} c_{\varphi_m}(n, r) = 0. \tag{23}$$

Suppose $F\,|\,T_{s,p} = G \in S_k(\Gamma_2)$. By using (11), we write down $\left(np, r, \frac{m}{p}\right)$-th coefficients of $F\,|\,T_{s,p}$ to get

$$p^{k-2} \sum_{\nu \pmod p,\, p\,|\,\left(r\nu + \nu^2 \frac{m}{p}\right)} A_F\left(n + \frac{r\nu + \nu^2(m/p)}{p}, r + 2\nu\frac{m}{p}, m\right) + A_F(p^2 n, pr, m) = A_G\left(np, r, \frac{m}{p}\right).$$

Suppose $p \nmid r$. Then there are exactly two choices for $\nu \pmod p$ in the left-hand side sum namely $\nu = 0$ and $\nu = -r(\overline{m/p})$, where $\overline{m/p}$ denotes the inverse of m/p modulo p. Assume that $(m/p)(\overline{m/p}) = 1 + lp$ for some $l \in \mathbb{Z}$. Then we have

$$p^{k-2}\left(c_{\varphi_m}(n, r) + c_{\varphi_m}(n + r^2 l(\overline{m/p}), -r - 2rlp)\right) + c_{\varphi_m}(p^2 n, pr) = c_{G\,|\,\rho_{m/p}}(np, r).$$

Since $(r + 2rlp)^2 - 4m(n + r^2 l(\overline{m/p})) = r^2 - 4mn$ and φ_m is in the space $J_{k,m}^{\text{cusp, new}}$, by using [8, Corollary 5.3] we get that $c_{\varphi_m}(n + r^2 l(\overline{m/p}), -r - 2rlp) = c_{\varphi_m}(n, r)$.

Hence we have

$$2p^{k-2}c_{\varphi_m}(n,r)+c_{\varphi_m}(p^2n,\,pr)=c_{G\,|\,\rho_{m/p}}(np,r). \tag{24}$$

From (23) and (24), we get that $c_{\varphi_m}(n,r)=c_{G\,|\,\rho_{m/p}}(np,r)$ for any $n\geqslant 1,\ r\in\mathbb{Z}$ with $p\nmid r,\ r^2<4mn$. But by using [3, Theorem 4.2, (7)] we get that

$$c_{G\,|\,\rho_{m/p}(np,r)}=c_{G\,|\,\rho_{m/p}\,|\,V_{m/p,p}}(n,r).$$

But $\phi_m\in J_{k,m}^{\text{cusp, new}}$, therefore by using [11, Lemma 3.1] we get that $\varphi_m=F\,|\,\rho_m=0$.

6. Proof of Theorem 1.7 and its corollaries

Proof of Theorem 1.7. We prove the theorem by induction on the number of prime factors of m.

Let p be any odd prime such that $\lambda_F(p)=0$. By using the decomposition given by (8), we have the following orthogonal decomposition $F\,|\,\rho_p=\phi_1\,|\,V_{1,p}+\phi_p$, where $\phi_1\in J_{k,1}^{\text{cusp, new}}$ and $\phi_p\in J_{k,p}^{\text{cusp, new}}$. Note that $J_{k,1}^{\text{cusp, new}}=J_{k,1}^{\text{cusp}}$. By using the identity (1) and the fact that $\lambda_F(p)=0$ we get that $F\,|\,\rho_p\,|\,V_{1,p}^*=0$. Then we have

$$\langle\phi_1\,|\,V_{1,p},\phi_1\,|\,V_{1,p}\rangle=\langle F\,|\,\rho_p,\phi_1\,|\,V_{1,p}\rangle=\langle F\,|\,\rho_p\,|\,V_{1,p}^*,\phi_1\rangle=0.$$

Therefore we have $F\,|\,\rho_p=\phi_p\in J_{k,p}^{\text{cusp, new}}$. By applying Theorem 1.6, we get that $F\,|\,\rho_p=0$.

Let m be any odd, squarefree, positive integer which is a multiple of at least 2 primes. Then again by using the decomposition (8), we have

$$F\,|\,\rho_m\in J_{k,m}^{\text{cusp}}=\bigoplus_{l\,|\,m,\,l\neq m}J_{k,l}^{\text{cusp, new}}\,|\,V_{l,(m/l)}\oplus J_{k,m}^{\text{cusp, new}}.$$

Note that all the direct sums in the above decomposition are orthogonal. We write

$$F\,|\,\rho_m=\sum_{l\,|\,m,\,l\neq m}\varphi_l\,|\,V_{l,(m/l)}+\varphi_m,$$

where $\varphi_l\in J_{k,l}^{\text{cusp, new}}$ and $\varphi_m\in J_{k,m}^{\text{cusp, new}}$. Suppose $\lambda_F(m)=0$. First, by using the identity (1) we deduce that $\varphi_1\,|\,V_{1,m}=0$. Next, by using the multiplicative property of $\lambda_F(m)$ we get that $\lambda_F(p)=0$ for some odd prime $p\,|\,m$. For any $l\,|\,m,\ l\neq m$ with $p\nmid l$, by using the fact that $V_{l,(m/l)}=V_{l,(m/lp)}V_{(m/p),\,p}$ we have

$$\begin{aligned}\langle\varphi_l\,|\,V_{l,(m/l)},\varphi_l\,|\,V_{l,(m/l)}\rangle&=\langle F\,|\,\rho_m,\varphi_l\,|\,V_{l,(m/l)}\rangle\\&=\langle F\,|\,\rho_m\,|\,V_{l,(m/l)}^*,\varphi_l\rangle=\langle F\,|\,\rho_m\,|\,V_{(m/p),\,p}^*V_{l,(m/lp)}^*,\varphi_l\rangle.\end{aligned}$$

By using the identity (3) for $\delta=1$, we have $F\,|\,\rho_m\,|\,V_{(m/p),\,p}^*=\lambda_F(p)\,F\,|\,\rho_{m/p}=0$. On the other hand, for any $l\,|\,m,\ l\neq m$ with $p\,|\,l$, let p' be any odd prime dividing m/l.

Again, by using the fact that $V_{l,(m/l)} = V_{l,(m/lp')}\,V_{(m/p'),\,p'}$ and the identity (3), we have

$$\begin{aligned}\langle \varphi_l \,|\, V_{l,(m/l)}, \varphi_l \,|\, V_{l,(m/l)}\rangle &= \langle F \,|\, \rho_m \,|\, V^*_{l,(m/l)}, \varphi_l\rangle \\ &= \langle F \,|\, \rho_m \,|\, V^*_{(m/p'),\,p'}\, V^*_{l,(m/lp')}, \varphi_l\rangle \\ &= \lambda_F(p')\langle F \,|\, \rho_{m/p'} \,|\, V^*_{l,(m/lp')}, \varphi_l\rangle.\end{aligned}$$

Since $\lambda_F(m/p') = \lambda_F(p)\,\lambda_F(m/pp') = 0$ and m/p' has fewer prime factors than m, by using the induction hypothesis we get that $F \,|\, \rho_{m/p'} = 0$. Hence we get that $F \,|\, \rho_m = \varphi_m \in J^{\text{cusp, new}}_{k,m}$. Now, by using Theorem 1.6 we get that $F \,|\, \rho_m = 0$. □

Proof of Corollary 1.10. Let $T = \begin{pmatrix} a & b/2 \\ b/2 & c\end{pmatrix}$ and $m = ax_0^2 + bx_0y_0 + cy_0^2$ for some $x_0, y_0 \in \mathbb{Z}$. Since m is squarefree, we have $\gcd(x_0, y_0) = 1$. Let $A = \begin{pmatrix} x_1 & x_0 \\ y_1 & y_0\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ and $S = {}^tATA$, where ${}^tA$ denotes the transpose of A. Then the lower right entry of S would be m. We have $a_F(S) = a_F({}^tATA) = a_F(T) \neq 0$ and hence $F \,|\, \rho_m \neq 0$. Using Theorem 1.7, we get the corollary. □

Proof of Corollary 1.11. Since F is a Siegel eigenform, we have the nonvanishing of the first Fourier–Jacobi coefficient of F [7], that is, $F \,|\, \rho_1 \neq 0$. Since $F \,|\, \rho_1 \in J^{\text{cusp}}_{k,1}$, by using (6) we have the following theta decomposition $F \,|\, \rho_1 = h_0\,\theta_{1,0} + h_1\theta_{1,1}$. Since $F \,|\, \rho_1 \neq 0$, by using (7) we get that $h_0 \neq 0$. Since $h_0 \in S_{k-1/2}(4)$ and $\dim S_{k-1/2}(4) = k/2 - 2$ for k even, there exists an n_0 with $1 \leqslant n_0 \leqslant k/2 - 2$ such that the n_0-th Fourier coefficient $a_{h_0}(n_0)$ of h_0 is nonzero. Then we have

$$a_{h_0}(n_0) = a_F\begin{pmatrix} n_0 & 0 \\ 0 & 1\end{pmatrix} \neq 0.$$

Now by using Corollary 1.10, we conclude the proof. □

Acknowledgement

The authors would like to thank the referee for bringing the reference [5] to our attention and for other useful comments which improved the presentation of the article.

References

[1] A. N. Andrianov, "Euler products that correspond to Siegel's modular forms of genus 2", *Uspehi Mat. Nauk* **29**:3(177) (1974), 43–110. In Russian; translated in *Russian Math. Surveys* **29**:3 (1974), 45–116. MR Zbl

[2] S. Böcherer, J. H. Bruinier, and W. Kohnen, "Non-vanishing of scalar products of Fourier–Jacobi coefficients of Siegel cusp forms", *Math. Ann.* **313**:1 (1999), 1–13. MR Zbl

[3] M. Eichler and D. Zagier, *The theory of Jacobi forms*, Progress in Mathematics **55**, Birkhäuser, Boston, MA, 1985. MR Zbl

[4] W. Kohnen, "Newforms of half-integral weight", *J. Reine Angew. Math.* **333** (1982), 32–72. MR Zbl

[5] W. Kohnen, "On the growth of Fourier coefficients of Siegel cusp forms of degree two", *Res. Math. Sci.* **9**:3 (2022), art. id. 50. MR Zbl

[6] W. Kohnen and N.-P. Skoruppa, "A certain Dirichlet series attached to Siegel modular forms of degree two", *Invent. Math.* **95**:3 (1989), 541–558. MR Zbl

[7] M. Manickam, "On the first Fourier–Jacobi coefficient of Siegel modular forms of degree two", *J. Number Theory* **219** (2021), 404–411. MR Zbl

[8] M. Manickam and B. Ramakrishnan, "On Shimura, Shintani and Eichler–Zagier correspondences", *Trans. Amer. Math. Soc.* **352**:6 (2000), 2601–2617. MR Zbl

[9] D. McCarthy, "Multiplicative relations for Fourier coefficients of degree 2 Siegel eigenforms", *J. Number Theory* **170** (2017), 263–281. MR Zbl

[10] A. Pitale, *Siegel modular forms: a classical and representation-theoretic approach*, Lecture Notes in Mathematics **2240**, Springer, Cham, 2019. MR Zbl

[11] N.-P. Skoruppa and D. Zagier, "Jacobi forms and a certain space of modular forms", *Invent. Math.* **94**:1 (1988), 113–146. MR Zbl

Received June 13, 2024. Revised October 3, 2024.

MURUGESAN MANICKAM
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH BHOPAL
BHOPAL
INDIA
manickam@iiserb.ac.in

KARAM DEO SHANKHADHAR
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH BHOPAL
BHOPAL
INDIA
karamdeo@iiserb.ac.in

VASUDEVAN SRIVATSA
INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR
MALAYSIA
vatsa588@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.261

CONTINUOUS SOBOLEV FUNCTIONS WITH SINGULARITY ON ARBITRARY REAL-ANALYTIC SETS

YIFEI PAN AND YUAN ZHANG

Near every point of a real-analytic set in $\mathbb{R}^n$, we make use of Hironaka's resolution of singularities theorem to construct a family of continuous functions in $W^{1,1}_{\text{loc}}$ such that their weak derivatives have (removable) singularities precisely on that set.

1. Introduction

Given a domain U in $\mathbb{R}^n$, $n \geq 1$, denote by $W^{k,p}_{\text{loc}}(U)$ the Sobolev space consisting of functions on U whose k-th order weak derivatives exist and belong to $L^p_{\text{loc}}(U)$, $k \in \mathbb{Z}^+$, $p \geq 1$. We investigate a Sobolev property for the reciprocals of logarithms of the modulus of real-analytic functions near their zero sets. Namely, given a real-analytic nonconstant function f on U, consider

$$v := \frac{1}{\ln|f|} \quad \text{on } U. \tag{1-1}$$

As we are solely interested in the Sobolev behavior of v near $f = 0$, and additional singularities would be introduced near $|f| = 1$, we further assume, say, $|f| < \frac{1}{2}$ on U. Consequently v is continuous on U. Letting $f^{-1}(0)$ be the zero set of f in U, we have $v|_{f^{-1}(0)} = 0$, and v is differentiable on $U \setminus f^{-1}(0)$. Note that $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) \geq 1$ in general.

According to a classical result of Stein [1993, pp. 71], $\ln|f| \in \text{BMO}$ for any polynomial f. On the other hand, Shi and Zhang [2022] showed that for a real-analytic f on U, if $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) \geq 2$, then $\ln|f| \in W^{1,1}_{\text{loc}}(U)$. It is important to note that this codimension assumption is essential and cannot be dropped. In comparison to this result, although v in (1-1) exhibits slightly greater regularity than $\ln|f|$, our first main theorem shows that v belongs to $W^{1,1}_{\text{loc}}(U)$ regardless of the codimension of $f^{-1}(0)$.

MSC2020: primary 32C07; secondary 14E15, 46E35.
Keywords: logarithms, real-analyticity, Sobolev spaces, resolution of singularities.

Theorem 1.1. *Let U be a domain in $\mathbb{R}^n$, $n \geq 1$. Let f be a real-analytic nonconstant function on U and $|f| < \frac{1}{2}$ on U. Then*:

(1) $\frac{1}{\ln|f|} \in W^{1,1}_{\mathrm{loc}}(U)$.

(2) *If* $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) = 1$, *then* $\frac{1}{\ln|f|} \notin W^{1,p}_{\mathrm{loc}}(U)$ *for any* $p > 1$.

The main idea of the proof is to use the coarea formula to transform the integrals under consideration into new ones along level sets of the function f. The L^1-integrability and the L^p-nonintegrability for $p > 1$ that we seek are thus consequences of certain quantitative properties of the level sets of f, which can be conveniently established by utilizing the powerful Hironaka's resolution of singularities theorem and the Łojasiewicz gradient inequality. A novelty of Theorem 1.1 is to provide ample $W^{1,1}_{\mathrm{loc}}$ functions. For instance, $\frac{1}{\ln|P(x)|} \in W^{1,1}_{\mathrm{loc}}$ for any polynomial P near its zeros. It is also interesting to point out that Theorem 1.1 indicates that Sobolev spaces in general do not satisfy an openness property, in the sense that there exists a class of functions in $W^{k,p}_{\mathrm{loc}}(U)$ for some $p \geq 1$ but not in $W^{k,q}_{\mathrm{loc}}(U)$ for any $q > p$.

Unfortunately our method cannot be applied directly in the smooth category, due to the absence of a Hironaka-type resolution property for smooth functions. It is natural to wonder if there is an easy way to verify the optimal Sobolev property of v, say, for any finitely vanishing smooth function f. For instance, consider the function $f(x, y) := y^2 - \sin(e^{1/x}\pi)e^{-1/x^2}$, which is smooth near $0 \subset \mathbb{R}^2$ and vanishes to second order at 0. It turns out, with a straightforward computation, that $\frac{1}{\ln|f|} \in W^{1,1}$ near 0.

As a consequence of Theorem 1.1, the weak derivative ∇v exists on U. Specifically, this implies that the singularity set $f^{-1}(0)$ of ∇v in the classical sense is actually a removable singularity in the weak sense. In other words, Theorem 1.1 allows us to construct, for any given real-analytic set, a continuous function in $W^{1,1}_{\mathrm{loc}}$ such that its weak derivative has a removable singularity precisely on that set.

Corollary 1.2. *Let A be a real-analytic set in $\mathbb{R}^n$. For every $p \in A$, there exists an open neighborhood V of p and a continuous function $u \in W^{1,1}_{\mathrm{loc}}(V)$, such that the set of removable singularities of ∇u is $A \cap V$.*

Finally, we study the Sobolev property of v in the special case when f is a holomorphic function on $U \subset \mathbb{C}^n$. Note that in this case $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) = 2$ unless $f \neq 0$ on U.

Theorem 1.3. *Let U be a domain in $\mathbb{C}^n$. Let f be a holomorphic nonconstant function on U and $|f| < \frac{1}{2}$ on U. Then*:

(1) $\frac{1}{\ln|f|} \in W^{1,2}_{\text{loc}}(U)$.

(2) *If* $f^{-1}(0) \neq \varnothing$, *then* $\frac{1}{\ln|f|} \notin W^{1,p}_{\text{loc}}(U)$ *for any* $p > 2$.

Corollary 1.4. *Let A be a complex analytic set in $\mathbb{C}^n$. For every $p \in A$, there exists an open neighborhood V of p and a continuous function $u \in W^{1,2}_{\text{loc}}(V)$, such that the set of removable singularities of ∇u is $A \cap V$.*

In view of Theorems 1.1 and 1.3, it seems to have suggested a correlation between the codimension of the level sets and the Sobolev integrability index. Thus, one may ask whether $v \in W^{1,d}_{\text{loc}}(U)$ if $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) = d$ for some $0 \le d \le n$. Unfortunately we do not have an answer to this question in general.

2. Proof of Theorem 1.1

Recall that the coarea formula states that, given $\phi \in L^1(U)$ and a real-valued Lipschitz function f on U,

$$\int_U \phi(x)\,|\nabla f(x)|\,dV_x = \int_{-\infty}^{\infty} \int_{f^{-1}(t)} \phi(x)\,dS_x\,dt. \tag{2-1}$$

Here given $t \in \mathbb{R}$, S_x is the $(n-1)$-dimensional Hausdorff measure of the level set $f^{-1}(t)$ of f defined by

$$f^{-1}(t) := \{x \in U : f(x) = t\}.$$

Towards the proof of the main theorems, we shall fix the real-analytic (or holomorphic) function f and use the following notation: two quantities A and B are said to satisfy $A \lesssim B$ if $A \le CB$ for some constant $C > 0$ which depends only on the f under consideration. We say $A \gtrsim B$ if and only if $B \lesssim A$, and $A \approx B$ if and only if $A \lesssim B$ and $B \lesssim A$ at the same time.

Given a set $A \subset \mathbb{R}^n$, denote by $m(A)$ the Hausdorff measure of A at its Hausdorff dimension. We first utilize Hironaka's resolution of singularities theorem to show the Hausdorff measure of level sets of real-analytic functions is bounded (from above). This will be essential in proving a Harvey–Polking type removable singularity lemma for the weak derivatives of v.

Theorem 2.1 [Atiyah 1970]. *Let f be a real-analytic nonconstant function defined near a neighborhood of $0 \in \mathbb{R}^n$. Then there exists an open set $U \subset \mathbb{R}^n$ near 0, a real-analytic manifold $\widetilde{U}$ of dimension n and a proper real-analytic map $\phi : \widetilde{U} \to U$ such that*:

(1) *The function* $\phi : \widetilde{U} \setminus \widetilde{f^{-1}(0)} \to U \setminus f^{-1}(0)$ *is an isomorphism*, *where* $\widetilde{f^{-1}(0)} := \{p \in \widetilde{U} : \phi(p) \in f^{-1}(0)\}$.

(2) *For each* $p \in \widetilde{U}$, *there exist local real-analytic coordinates* $(y_1, \dots, y_n)$ *centered at* p, *such that near* p *one has*

$$f \circ \phi(y) = u(y) \cdot \prod_{i=1}^{n} y_i^{k_i},$$

where u *is real-analytic and* $u \neq 0$, $k_i \in \mathbb{Z}^+ \cup \{0\}$.

Lemma 2.2. *Let* f *be a real-analytic nonconstant function on* U. *Then*

$$m(f^{-1}(t)) \lesssim 1 \quad \textit{for all } |t| \ll 1.$$

Proof. Without loss of generality, assume $0 \in U$ and $f(0) = 0$. Under the setup of Hironaka's resolution Theorem 2.1, for every $p \in \widetilde{f^{-1}(0)}$, let $(\widetilde{V}, \psi)$ be a coordinate chart near p in $\widetilde{U}$ such that, for $y \in \psi(\widetilde{V}) \subset \mathbb{R}^n$,

$$f \circ \Phi(y) := f \circ \phi \circ \psi^{-1}(y) = u(y) \cdot \prod_{i=1}^{n} y_i^{k_i}.$$

By properness of ϕ, $V := \phi(\widetilde{V})$ is an open subset of U near $\phi(p)$. Since ϕ is smooth on $\widetilde{U}$, by shrinking U if necessary, $\Phi : \psi(\widetilde{V}) \to V$ is smooth up to the boundary of $\psi(\widetilde{V})$. By change of coordinates formula,

$$\begin{aligned} m(f^{-1}(t) \cap V) &= \int_{\{f(x)=t\} \cap \phi(\widetilde{V})} dS_x \\ &= \int_{\{f \circ \Phi(y)=t\} \cap \psi(\widetilde{V})} \Phi^* \, dS_x \lesssim \int_{\{f \circ \Phi(y)=t\} \cap \psi(\widetilde{V})} dS_y. \end{aligned}$$

Thus, in view of this and the fact that $u \neq 0$ on $\widetilde{U}$, the proof boils down to showing that the $(n-1)$-dimensional Hausdorff measure satisfies

$$m(A^n(t)) \lesssim 1 \quad \text{for all } 0 < t \ll 1, \tag{2-2}$$

where

$$A^n(t) = \left\{ y \in \mathbb{R}^n : \prod_{i=1}^{n} y_i^{k_i} = t, 0 < y_i < 1, i = 1, \dots, n \right\}. \tag{2-3}$$

Here the constant multiple for "$\lesssim$" in (2-2) is only dependent on k_i, $i = 1, \dots, n$. Clearly, one only needs to prove the case when all $k_i > 0$. Let $k := \sum_{i=1}^{n} k_i$.

We shall employ the mathematical induction on the dimension n to prove (2-2) for all level sets in the form of (2-3). The $n = 1$ case is trivial. Assume the $n = l$ case holds. Namely, for every level set $A^l(t)$ in $\mathbb{R}^l$ defined by (2-3), $m(A^l(t)) \lesssim 1$

for $0 < t \ll 1$. When the dimension n equals $l+1$, one first has

$$A^{l+1}(t) \subset \bigcup_{j=1}^{l+1} A_j^{l+1}(t),$$

where, for each $j = 1, \dots, l+1$,

$$A_j^{l+1}(t) := \left\{ y \in \mathbb{R}^{l+1} : t^{1/k} \le y_j < 1, 0 < y_i < 1 \text{ if } i \neq j, \text{ and } \prod_{\substack{1 \le i \le l+1 \\ i \neq j}} y_i^{k_i} = t y_j^{-k_j} \right\}.$$

Since $A_j^{l+1}(t)$ is a finite union of smooth hypersurfaces in $\mathbb{R}^{l+1}$ away from a set of dimension $l-1$, by Fubini's theorem, the l-dimensional Hausdorff measure satisfies

$$m(A_j^{l+1}(t)) = \int_{t^{1/k}}^{1} \int_{\prod_{1 \le i \le l+1, i \neq j} y_i^{k_i} = t y_j^{-k_j}, 0 < y_i < 1, i \neq j} dS_{\hat{y}_j}\, dy_j,$$

and thus

$$m(A^{l+1}(t)) \le \sum_{j=1}^{l+1} \int_{t^{1/k}}^{1} \int_{\prod_{1 \le i \le l+1, i \neq j} y_i^{k_i} = t y_j^{-k_j}, 0 < y_i < 1, i \neq j} dS_{\hat{y}_j}\, dy_j. \tag{2-4}$$

Further denote $\hat{y}_j := (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_{l+1}) \in \mathbb{R}^l$,

$$t' := t y_j^{-k_j},$$

and

$$A_j^l(t') := \left\{ \hat{y}_j \in \mathbb{R}^l : 0 < y_i < 1, i \neq j, \text{ and } \prod_{\substack{1 \le i \le l+1^l \\ i \neq j}} y_i^{k_i} = t' \right\}.$$

Noting that $t' < t^{1-k_j/k}$ when $y_j > t^{1/k}$, we obtain from (2-4)

$$m(A^{l+1}(t)) \le (1 - t^{1/k}) \sum_{j=1}^{l+1} \sup_{0 < t' < t^{1-k_j/k}} m(A_j^l(t')).$$

On the other hand, since $k_j < k$, one has $t^{1-k_j/k} \ll 1$ when $t \ll 1$. By the induction assumption and the fact that $A_j^l(t')$ is in $\mathbb{R}^l$,

$$\sup_{0 < t' < t^{1-k_j/k}} m(A_j^l(t')) \lesssim 1 \quad \text{for all } 0 < t \ll 1.$$

This finally gives

$$m(A^{l+1}(t)) \lesssim 1 \quad \text{for all } 0 < t \ll 1.$$

The lemma is proved. □

Lemma 2.3. *Given a real-analytic nonconstant function f on U with $|f| < \frac{1}{2}$ on U, let v be defined in (1-1), and*

$$g := \frac{\nabla f}{f \cdot (\ln|f|)^2} \quad \text{on } U. \tag{2-5}$$

Then $g \in L^1_{\text{loc}}(U)$. One has

$$\nabla v = g \quad \text{on } U$$

in the sense of distributions.

Proof. First, we show that $g \in L^1_{\text{loc}}(U)$. Since f is real-analytic on U, shrinking U if necessary, one can assume f to be (globally) Lipschitz on U. Making use of the coarea formula (2-1), one gets

$$\begin{aligned}\int_U |g(x)|\, dV_x &= \int_U \frac{|\nabla f(x)|}{|f(x)|(\ln|f(x)|)^2}\, dV_x \\ &\leq \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{f^{-1}(t)} \frac{1}{|f(x)|(\ln|f(x)|)^2}\, dS_x\, dt = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{m(f^{-1}(t))}{|t|(\ln|t|)^2}\, dt.\end{aligned}$$

Lemma 2.2 further allows us to infer

$$\int_U |g(x)|\, dV_x \lesssim \int_0^{\frac{1}{2}} \frac{1}{t(\ln t)^2}\, dt = \int_{\ln 2}^{\infty} \frac{1}{s^2}\, ds < \infty.$$

Next, we show that, given any testing function $\eta \in C_c^\infty(U)$,

$$-\int_U v\nabla\eta = \int_U \eta g. \tag{2-6}$$

Since v is differentiable away from $f^{-1}(0)$, a direct computation gives

$$\nabla v = g \quad \text{on } U \setminus f^{-1}(0). \tag{2-7}$$

In particular, (2-6) is trivially true if $K := f^{-1}(0) \cap \operatorname{supp} \eta = \varnothing$.

If $K \neq \varnothing$, given $\epsilon > 0$ let

$$K_\epsilon := \{x \in U : \operatorname{dist}(x, K) \leq \epsilon\},$$

where $\operatorname{dist}(x, K)$ is the distance function from x to the set K. Let $\rho_\epsilon \in C^\infty(U)$ be such that $\rho_\epsilon = 0$ in K_ϵ, $\rho_\epsilon = 1$ in $U \setminus K_{3\epsilon}$ and $|\nabla \rho_\epsilon| \lesssim \frac{1}{\epsilon}$ on U. See, for instance, [Hörmander 2003, Theorem 1.2.1-2]. Then $\rho_\epsilon \eta \in C_c^\infty(U \setminus f^{-1}(0))$. Using (2-7) we immediately have

$$-\int_U v\nabla(\rho_\epsilon \eta) = \int_U \rho_\epsilon \eta g,$$

or equivalently,

$$-\int_U v\eta\nabla\rho_\epsilon - \int_U v\rho_\epsilon\nabla\eta = \int_U \rho_\epsilon\eta g. \tag{2-8}$$

We shall prove

$$\lim_{\epsilon\to 0}\int_U v\eta\nabla\rho_\epsilon = 0. \tag{2-9}$$

If so, then passing $\epsilon \to 0$ in (2-8), we obtain the desired equality (2-6) as a consequence of Lebesgue's dominated convergence theorem.

To prove (2-9), first by the assumption on ρ_ϵ,

$$\left|\int_U v\eta\nabla\rho_\epsilon\right| = \left|\int_{K_{3\epsilon}\setminus K_\epsilon} v\eta\nabla\rho_\epsilon\right| \lesssim \frac{C}{\epsilon}\int_{K_{3\epsilon}\setminus K_\epsilon}|v| \tag{2-10}$$

for some constant C dependent only on η. Since f is Lipschitz on U, for any $x_0 \in f^{-1}(0)$, $|f(x)| = |f(x) - f(x_0)| \lesssim |x - x_0|$. In particular,

$$|f(x)| \lesssim \operatorname{dist}(x, f^{-1}(0)).$$

Thus for all $x \in K_{3\epsilon}\setminus K_\epsilon$ (equivalently, $\epsilon < \operatorname{dist}(x, f^{-1}(0)) < 3\epsilon$), one has

$$|v(x)| = \frac{1}{\big|\ln|f(x)|\big|} \lesssim \frac{1}{\big|\ln\operatorname{dist}(x, f^{-1}(0))\big|} \approx \frac{1}{|\ln\epsilon|}$$

for all ϵ small enough. Hence by (2-10)

$$\left|\int_U v\eta\nabla\rho_\epsilon\right| \lesssim \frac{Cm(K_{3\epsilon})}{\epsilon\,|\ln\epsilon|}. \tag{2-11}$$

On the other hand, according to a nontrivial result of Loeser [1986, Theorem 1.1] and its consequent remarks,

$$m(K_{3\epsilon}) \lesssim \epsilon^{\operatorname{codim}_{\mathbb{R}} f^{-1}(0)} \lesssim \epsilon.$$

Here the last inequality has used the fact that $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) \geq 1$ due to the real-analyticity of f. The equality (2-9) follows by combining the above with (2-11). □

Proof of Theorem 1.1. Since $|f| < \frac{1}{2}$, we have $\big|\ln|f|\big| > \ln 2$ and so $|v| < \frac{1}{\ln 2} \in L^\infty(U)$. Part (1) follows from this and Lemma 2.3. For part (2), we only need to show that the function g defined in (2-5) does not belong to L^p_{loc} for any $p > 1$ near any neighborhood of $f^{-1}(0)$.

First, according to the Łojasiewicz inequality, by shrinking U if necessary, there exists some constant $\beta \in (0, 1)$ such that

$$|\nabla f(x)| \gtrsim |f(x)|^\beta, \quad x \in U. \tag{2-12}$$

As a consequence of this,

$$\begin{aligned}\int_U |\nabla v(x)|^p \, dV_x &= \int_U \frac{|\nabla f(x)|}{|\nabla f(x)|^{-(p-1)} |f(x)|^p \big|\ln|f(x)|\big|^{2p}} \, dV_x \\ &\gtrsim \int_U \frac{|\nabla f(x)|}{|f(x)|^{p-(p-1)\beta} \big|\ln|f(x)|\big|^{2p}} \, dV_x.\end{aligned}$$

Utilizing the coarea formula, we have, for some $\epsilon_0 > 0$,

$$\begin{aligned}\int_U |\nabla v(x)|^p \, dV_x &\gtrsim \int_{-\epsilon_0}^{\epsilon_0} \int_{f^{-1}(t)} \frac{1}{|f(x)|^{p-(p-1)\beta} \big|\ln|f(x)|\big|^{2p}} \, dS_x \, dt \\ &= \int_{-\epsilon_0}^{\epsilon_0} \frac{m(f^{-1}(t))}{|t|^{p-(p-1)\beta} \big|\ln|t|\big|^{2p}} \, dt.\end{aligned}$$

Since $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) = 1$, there is some $x_0 \in f^{-1}(0) \cap U$, such that $|\nabla f(x_0)| \neq 0$. Let V be a neighborhood of x_0 in U such that $|\nabla f| \gtrsim 1$ on V. Then for all t small enough, $m(f^{-1}(t) \cap V) \gtrsim 1$. Consequently, $m(f^{-1}(t)) \gtrsim 1$ for $0 < t \ll 1$. Thus

$$\int_U |\nabla v(x)|^p \, dV_x \gtrsim \int_0^{\epsilon_0} \frac{1}{t^{p-(p-1)\beta} |\ln t|^{2p}} \, dt.$$

Note that $p-(p-1)\beta > 1$ necessarily when $p > 1$ and $\beta < 1$. Hence the last term is unbounded. The proof is complete. □

Proof of Corollary 1.2. Since A is real-analytic, there exists an open neighborhood $V \subset \mathbb{R}^n$ of p and a real-analytic function f on V such that $A \cap V = \{x \in V : f(x) = 0\}$. Then $u = \frac{1}{\ln|f|}$ is the desired function satisfying the assumptions. □

For functions (such as $\ln|f|$) with singularities, its composition with another logarithm typically exhibits reduced singularities. The following theorem shows that composing extra logarithms does not improve Sobolev regularity in general.

Theorem 2.4. *Let U be a domain in $\mathbb{R}^n$, $n \geq 1$. Let f be a real-analytic nonconstant function on U and $|f| < \frac{1}{10}$ on U. Then*:

(1) $\frac{1}{\ln\ln|f|} \in W^{1,1}_{\text{loc}}(U)$.

(2) *If* $\operatorname{codim}_{\mathbb{R}} f^{-1}(0) = 1$, *then* $\frac{1}{\ln\ln|f|} \notin W^{1,p}_{\text{loc}}(U)$ *for any* $p > 1$.

Proof. Applying a similar approach as in the proof of Lemma 2.3, we first have

$$\nabla\left(\frac{1}{\ln\big|\ln|f|\big|}\right) = \frac{\nabla f}{f \cdot \ln|f| \cdot \big(\ln\big|\ln|f|\big|\big)^2} \quad \text{on } U$$

in the sense of distributions. Making use of the coarea formula and Lemma 2.2,

$$\begin{aligned}\int_U \left|\nabla\left(\frac{1}{\ln\left|\ln|f|\right|}\right)\right| &\le \int_{-\frac{1}{10}}^{\frac{1}{10}} \int_{f^{-1}(t)} \frac{1}{|f(x)|\,\left|\ln|f(x)|\right|\left(\ln\left|\ln|f(x)|\right|\right)^2}\, dS_x\, dt \\ &\lesssim \int_0^{\frac{1}{10}} \frac{1}{t|\ln t|(\ln|\ln t|)^2}\, dt \\ &= \int_{\ln 10}^{\infty} \frac{1}{t(|\ln t|)^2}\, dt = \int_{\ln\ln 10}^{\infty} \frac{1}{t^2}\, dt \lesssim 1.\end{aligned}$$

In the case when $p > 1$, there exists some $0 < \beta < 1$ by (2-12), and some small $\epsilon_0 > 0$ such that

$$\begin{aligned}\int_U \left|\nabla\left(\frac{1}{\ln\left|\ln|f|\right|}\right)\right|^p &\gtrsim \int_U \frac{|\nabla f(x)|}{|f(x)|^{p-(p-1)\beta}\left|\ln|f(x)|\right|^p\left(\ln\left|\ln|f(x)|\right|\right)^{2p}}\, dV_x \\ &\gtrsim \int_0^{\epsilon_0} \frac{1}{t^{p-(p-1)\beta}|\ln t|\,(\ln|\ln t|)^2}\, dt.\end{aligned}$$

Since $p-(p-1)\beta > 1$, the last term is divergent. This completes the proof of the theorem. □

3. Proof of Theorem 1.3

To prove Theorem 1.3 for holomorphic functions, we shall need the following well-known complex version Hironaka's resolution of singularities theorem. See, for instance, [Smith 2016].

Theorem 3.1. *Let f be a holomorphic function defined near a neighborhood of $0 \in \mathbb{C}^n$. Then there exists an open set $U \subset \mathbb{C}^n$ near* 0, *a complex manifold $\widetilde{U}$ of dimension n and a proper holomorphic map $\phi : \widetilde{U} \to U$ such that*:

(1) *The function $\phi : \widetilde{U} \setminus \widetilde{f^{-1}(0)} \to U \setminus f^{-1}(0)$ is a biholomorphism, where $\widetilde{f^{-1}(0)} := \{p \in \widetilde{U} : \phi(p) \in f^{-1}(0)\}$.*

(2) *For each $p \in \widetilde{U}$, there exist local holomorphic coordinates $(w_1, \dots, w_n)$ centered at p, such that near p one has*

$$f \circ \phi(w) = u(w) \cdot \prod_{i=1}^{n} w_i^{k_i},$$

where u is holomorphic and $u \neq 0$, $k_i \in \mathbb{Z}^+ \cup \{0\}$.

Proof of Theorem 1.3. (1) Since $\bar{\partial} f = 0$, and according to Lemma 2.3,

$$\partial v = \frac{\partial f}{2f \cdot (\ln|f|)^2} \in L^1_{\mathrm{loc}}(U)$$

in the sense of distributions, we only need to show that

$$\frac{\partial f}{f\cdot(\ln|f|)^2}\in L^2_{\text{loc}}(U).$$

On the other hand, making use of Hironaka's resolution of singularities Theorem 3.1 for holomorphic functions, for every $p\in\widetilde{f^{-1}(0)}$, let $(\widetilde{V},\psi)$ be a coordinate chart near p in $\widetilde{U}$ such that, for $w\in\psi(\widetilde{V})\subset\{w\in\mathbb{C}^n:|w_j|<\frac{1}{2}\}$,

$$\tilde{f}(w):=f\circ\phi\circ\psi^{-1}(w)=u(w)\cdot\prod_{i=1}^n w_i^{k_i},$$

where $u\neq 0$ on $\psi(\widetilde{V})$ and $k_i\in\mathbb{Z}^+\cup\{0\}$. Let $V:=\phi(\widetilde{V})$, $\Phi:=\phi\circ\psi^{-1}$, and Jac_Φ be the complex Jacobian of the holomorphic map Φ. Note that the inverse matrix $(\text{Jac}_\Phi)^{-1}$ is smooth on $\psi(\widetilde{V}\setminus\widetilde{f^{-1}(0)})$, and

$$|(\text{Jac}_\Phi)^{-1}(w)\cdot\det(\text{Jac}_\Phi)(w)|\lesssim 1\quad\text{for all }\ w\in\psi(\widetilde{V}\setminus\widetilde{f^{-1}(0)}).$$

By change of variables formula,

$$\begin{aligned}\int_V\frac{|\partial_z f(z)|^2}{|f(z)|^2(\ln|f(z)|)^4}\,dV_z&\\ &=\int_{\Phi^{-1}(V\setminus f^{-1}(0))}\Phi^*\Big(\frac{|\partial_z f(z)|^2}{|f(z)|^2(\ln|f(z)|)^4}\,dV_z\Big)\\ &\lesssim\int_{\psi(\widetilde{V}\setminus\widetilde{f^{-1}(0)})}\frac{|\partial_w\tilde{f}(w)|^2|(\text{Jac}_\Phi)^{-1}(w)|^2}{|\tilde{f}(w)|^2(\ln|\tilde{f}(w)|)^4}\big|\det(\text{Jac}_\Phi(w))\big|^2\,dV_w\\ &\lesssim\int_{\psi(\widetilde{V})}\frac{|\partial_w\tilde{f}(w)|^2}{|\tilde{f}(w)|^2(\ln|\tilde{f}(w)|)^4}\,dV_w.\end{aligned}$$

Thus, the proof boils down to showing that, for $j=1,\dots,n$,

$$\int_{\psi(\widetilde{V})}\frac{|\partial_{w_j}\tilde{f}(w)|^2}{|\tilde{f}(w)|^2(\ln|\tilde{f}(w)|)^4}\,dV_w\lesssim 1. \tag{3-1}$$

For simplicity, let $j=1$ in (3-1). If $k_1=0$, then

$$\partial_{w_1}\tilde{f}(w)=\partial_{w_1}u(w)\cdot\prod_{i=1}^n w_i^{k_i}.$$

Since $\frac{1}{(\ln|\tilde{f}(w)|)^4}\lesssim 1$ and $u\neq 0$, when w is near 0,

$$\frac{|\partial_{w_1}\tilde{f}(w)|^2}{|\tilde{f}(w)|^2(\ln|\tilde{f}(w)|)^4}=\frac{|\partial_{w_1}u(w)|^2}{|u(w)|^2(\ln|\tilde{f}(w)|)^4}\lesssim 1.$$

So (3-1) holds. If $k_1 > 0$, then

$$\partial_{w_1} \tilde{f}(w) = \partial_{w_1} u(w) \cdot \prod_{i=1}^{n} w_i^{k_i} + k_1 u(w) \cdot w_1^{k_1-1} \cdot \prod_{i=2}^{n} w_i^{k_i}.$$

Hence

$$\begin{aligned} \frac{|\partial_{w_1} \tilde{f}(w)|^2}{|\tilde{f}(w)|^2 (\ln |\tilde{f}(w)|)^4} &\lesssim \frac{|\partial_{w_1} u(w)|^2}{|u(w)|^2 (\ln |\tilde{f}(w)|)^4} + \frac{k_1^2}{|w_1|^2 (\ln |\tilde{f}(w)|)^4} \\ &\lesssim 1 + \frac{1}{|w_1|^2 (\ln |\tilde{f}(w)|)^4}. \end{aligned}$$

Note that when w is close to 0,

$$\big|\ln |\tilde{f}(w)|\big| = \left| \ln |u(w)| + \sum_{i=1}^{n} k_i \ln |w_i| \right| \gtrsim -\ln |w_1|. \tag{3-2}$$

This leads to

$$\begin{aligned} \int_{\psi(\tilde{V})} \frac{|\partial_{w_1} \tilde{f}(w)|^2}{|\tilde{f}(w)|^2 (\ln |\tilde{f}(w)|)^4} \, dV_w &\lesssim 1 + \int_{\psi(\tilde{V})} \frac{1}{|w_1|^2 \big|\ln |w_1|\big|^4} \, dV_w \\ &\lesssim 1 + \int_0^{\frac{1}{2}} \frac{1}{s (\ln s)^4} \, ds \lesssim 1. \end{aligned}$$

Equation (3-1) and thus part (1) are proved.

(2) Let U_1 be an open subset of U such that $f^{-1}(0) \cap U_1$ is regular. Then there exists a holomorphic coordinate change on U_1 such that under the new coordinates $(w_1, \dots, w_n)$, one has $w_n = f(z)$. As a consequence of this,

$$\begin{aligned} \int_U \left| \frac{\partial_z f}{f \cdot (\ln |f|)^2} \right|^p dV_z &\geq \int_{U_1} \left| \frac{\partial_z f}{f \cdot (\ln |f|)^2} \right|^p dV_z \\ &\approx \int_{U_1} \frac{1}{|w_n|^p \big|\ln |w_n|\big|^{2p}} \, dV_w \gtrsim \int_0^{\epsilon_0} \frac{1}{s^{p-1} |\ln s|^{2p}} \, ds \end{aligned}$$

for some $\epsilon_0 > 0$. Since $p > 2$, the last term is unbounded. This proves part (2). □

Proof of Corollary 1.4. The proof is similar to that of Corollary 1.2, with Theorem 1.1 substituted by Theorem 1.3, and is omitted. □

An application of Theorem 1.3 is to provide ample data to the $\bar{\partial}$ problem in complex analysis, in particular, within the framework of Hörmander's classical L^2 theory for $\bar{\partial}$-closed forms with L^2_{loc} coefficients. Normally, generating smooth data is straightforward. In the following, we construct data with singularity on complex analytic varieties, where Hörmander's theory can still be applied.

Example 1. Let Ω be a pseudoconvex domain in $\mathbb{C}^n$. Let f be a nonconstant holomorphic function on Ω such that $f^{-1}(0) \neq \varnothing$. Choose a monotone increasing function $\chi \in C^\infty([0, \infty))$ such that $\chi(t) = t$ if $0 \le t \le \frac{1}{4}$, and $\chi(t) = \frac{1}{3}$ if $t \ge 1$. Then $g = \frac{1}{\ln \chi(|f|)} \in W^{1,2}_{\text{loc}}(\Omega)$ by Theorem 1.3. Furthermore, $u := \bar\partial g$ is a $\bar\partial$-closed $(0, 1)$ form with L^2_{loc} coefficients with singularities precisely on $f^{-1}(0)$.

References

[Atiyah 1970] M. F. Atiyah, "Resolution of singularities and division of distributions", *Comm. Pure Appl. Math.* **23** (1970), 145–150. MR Zbl

[Hörmander 2003] L. Hörmander, *The analysis of linear partial differential operators, I: Distribution theory and Fourier analysis*, Springer, 2003. MR Zbl

[Loeser 1986] F. Loeser, "Volume de tubes autour de singularités", *Duke Math. J.* **53**:2 (1986), 443–455. MR Zbl

[Shi and Zhang 2022] Z. Shi and R. Zhang, "Sobolev differentiability properties of logarithmic modulus of real analytic functions", preprint, 2022. arXiv 2205.02159

[Smith 2016] K. Smith, "Jyväskylä Summer school: resolution of singularities", lecture notes, University of Michigan, 2016, available at https://sites.lsa.umich.edu/kesmith/wp-content/uploads/sites/1309/2024/06/JyvaskylaIntro.pdf.

[Stein 1993] E. M. Stein, *Harmonic analysis*: *real-variable methods*, *orthogonality*, *and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR Zbl

Received June 6, 2024. Revised October 18, 2024.

YIFEI PAN
DEPARTMENT OF MATHEMATICAL SCIENCES
PURDUE UNIVERSITY FORT WAYNE
FORT WAYNE, IN
UNITED STATES
pan@pfw.edu

YUAN ZHANG
DEPARTMENT OF MATHEMATICAL SCIENCES
PURDUE UNIVERSITY FORT WAYNE
FORT WAYNE, IN
UNITED STATES
zhangyu@pfw.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.273

GRADING OF AFFINIZED WEYL SEMIGROUPS OF KAC–MOODY TYPE

PAUL PHILIPPE

For any Kac–Moody root datum $\mathcal{D}$, D. Muthiah and D. Orr have defined a partial order on the semidirect product W^a_+ of the integral Tits cone with the vectorial Weyl group of $\mathcal{D}$, and a compatible length function. We classify covers for this order and show that this length function defines a $\mathbb{Z}$-grading of W^a_+, generalizing the case of affine ADE root systems and giving a positive answer to a conjecture of Muthiah and Orr.

Introduction 273
1. Preliminaries 278
1.1. Definitions and notation 278
1.2. Geometric interpretation 284
1.3. Preliminary results 287
2. Restriction to constant dominance classes 291
2.1. Relative length on W 293
2.2. Relation with the affinized Bruhat length 295
2.3. Vectorial covers with nonconstant coweight 296
3. Properly affine covers 300
3.1. A few properties of properly affine covers 300
3.2. Another expression for the affinized length difference 302
3.3. Almost-dominance in properly affine covers 307
3.4. Properly affine covers and quantum roots 311
3.5. Conclusion 317
Acknowledgements 319
References 319

Introduction

Motivation.

Reductive groups over p-adic fields. Let $\boldsymbol{G}$ be a split reductive group scheme with the data of a Borel subgroup $\boldsymbol{B}$ containing a maximal torus $\boldsymbol{T}$. Let $W = N_{\boldsymbol{G}}(\boldsymbol{T})/\boldsymbol{T}$

MSC2020: primary 20F55, 20G44; secondary 20C08, 22E67.
Keywords: Kac–Moody groups, root systems, Coxeter groups, affine Weyl groups, affinized Bruhat order.

be its vectorial Weyl group and Y be its coweight lattice: $Y = \mathrm{Hom}(\mathbb{G}_m, \boldsymbol{T})$. The action of W on $\boldsymbol{T}$ induces an action of W on Y and allows us to form the semidirect product $W^a = Y \rtimes W$. This group, called the extended affinized Weyl group of $\boldsymbol{G}$, appears naturally in the geometry and the representation theory of $\boldsymbol{G}$ over discretely valued fields. A foundational work in this regard was done by N. Iwahori and H. Matsumoto [1965], when they exhibited a Bruhat decomposition of $\boldsymbol{G}(\mathbb{Q}_p)$ indexed by W^a.

Let $\mathcal{K}$ be a nonarchimedean local field with ring of integers $\mathcal{O}_{\mathcal{K}} \subset \mathcal{K}$, uniformizer $\pi \in \mathcal{O}_{\mathcal{K}}$ and residue field $\Bbbk_{\mathcal{K}} = \mathcal{O}_{\mathcal{K}}/\pi$. Let $G = \boldsymbol{G}(\mathcal{K})$, let $K = \boldsymbol{G}(\mathcal{O}_{\mathcal{K}})$ be its integral points and let I be its Iwahori subgroup, defined as

$$I = \{g \in K \mid g \in \boldsymbol{B}(\Bbbk_{\mathcal{K}}) \mod \pi\}.$$

The extended affinized Weyl group can be understood as $N_G(\boldsymbol{T}(\mathcal{K}))/\boldsymbol{T}(\mathcal{O}_{\mathcal{K}})$, so it admits a lift in G. Then, G admits a decomposition in I-double cosets indexed by W^a, the Iwahori–Matsumoto–Bruhat decomposition:

$$G = \bigsqcup_{\pi^\lambda w \in W^a} I\pi^\lambda w I. \tag{0.1}$$

The group W^a is a finite extension of a Coxeter group and thus admits a Bruhat order which arises from the geometry of the homogeneous space G/I: for any $\pi^\lambda w \in W^a$, $I\pi^\lambda w I$ is a subvariety of pure dimension $\ell(\pi^\lambda w)$ in G/I, and its closure admits a disjoint decomposition in I orbits:

$$\overline{I\pi^\lambda w I} = \bigsqcup_{\pi^\mu v \leq \pi^\lambda w} I\pi^\mu v I, \tag{0.2}$$

which extends the Iwahori–Matsumoto decomposition. The connection between the geometry of G/I and the combinatorial structure of W^a is deeper. In particular, R-Kazhdan–Lusztig polynomials introduced by Kazhdan and Lusztig [1980], defined as the number of points of certain intersections in G/I, are also given by a recursive formula based on the Bruhat order and the Bruhat length of W^a.

These polynomials appear in many topics around reductive groups over local fields, we aim to develop analogous polynomials when $\boldsymbol{G}$ is replaced by a general Kac–Moody group.

Extension to Kac–Moody groups. Replace $\boldsymbol{G}$ by a general split Kac–Moody group. Kac–Moody group functors are entirely defined by the underlying Kac–Moody root datum $\mathcal{D}$, as defined in [Rémy 2002, §2], and reductive groups correspond to root data of finite type. Then the Iwahori–Matsumoto decomposition no longer holds on $G = \boldsymbol{G}(\mathcal{K})$. However there is a partial Iwahori–Matsumoto decomposition: there exists a subsemigroup G^+ of G such that

$$G^+ = \bigsqcup_{\pi^\lambda w \in W^a_+} I\pi^\lambda w I. \tag{0.3}$$

The indexing set for this decomposition W^a_+ is a subsemigroup of $W^a = Y \rtimes W$, and it appears naturally in other related contexts, for example, when trying to construct an Iwahori–Hecke algebra for G [Braverman et al. 2016; Bardy-Panse et al. 2016]. Let us briefly explain how W^a_+ is defined.

Let Φ be the real root system of the root datum $\mathcal{D}$. It is an infinite set (unless $\mathcal{D}$ is reductive) of linear forms on Y coming with a subset of positive roots $\Phi_+ \subset \Phi$ such that $\Phi = \Phi_+ \sqcup -\Phi_+$. Let $Y^{++} = \{\lambda \in Y \mid \forall \alpha \in \Phi_+,\ \alpha(\lambda) \geq 0\}$ and $Y^+ = W \cdot Y^{++}$. Then W^a_+ is defined as $Y^+ \rtimes W$. In the reductive case, Y^+ coincides with Y and thus $W^a_+ = W^a$. However, W^a can no longer be conceived as a finite extension of a Coxeter system, hence there is a priori no Bruhat order on W^a_+, let alone on W^a. A well-behaved topology on G^+/I would allow us to define an order on W^a_+ through the analog of decomposition (0.2), but G^+/I does not seem to have a natural variety, nor even an ind-variety structure.

An order and two lengths on W^a_+. In Appendix B2 of their article on the construction of an Iwahori–Hecke algebra for G an affine Kac–Moody group over a p-adic field [Braverman et al. 2016], A. Braverman, D. Kazhdan and M. Patnaik propose the definition of a preorder on W^a_+ which would replace the Bruhat order of W^a and they conjecture that it is a partial order. In [2018], D. Muthiah extends the definition of this preorder to any Kac–Moody group G, defines a $\mathbb{Z}\oplus\varepsilon\mathbb{Z}$-valued length compatible with this preorder and hence shows that it is an order. In [2019], D. Muthiah and D. Orr then show that this length can be evaluated at $\varepsilon = 1$ to obtain a $\mathbb{Z}$-valued length strictly compatible with the order on W^a_+.

In order to build a Kazhdan–Lusztig theory of p-adic Kac–Moody groups, we want to understand how close this order is to the Bruhat order of an affine Coxeter group, which properties still hold and which do not. The definition of a $\mathbb{Z}$-length is already a significant step, but many important properties, which are known to hold for Bruhat orders, remain unknown in this context. Several were proved only for Kac–Moody root systems of affine simply laced type using the specific structure of an affinized Weyl group of W in this context.

Choice of vocabulary. The order on W^a_+ is often mentioned in the literature as "the double affine Bruhat order" and the associated length as "the double affine Bruhat length" because it is most studied in the case of G a Kac–Moody group of affine type (in which case W is an affine Weyl group). We refer to it as "the affinized

Bruhat order" and "the affinized Bruhat length", denoted by ℓ^a, because we do not suppose that W is an affine Weyl group. Note that, if W is finite, then the affinized Bruhat length and order on W^a_+ are just the ones induced by its Coxeter group structure.

Main result. Our main result is a positive answer to [Muthiah and Orr 2019, Conjecture 1.5] in full generality.

For any partial order $\leq$ on a set X, we say that $\boldsymbol{y}$ covers $\boldsymbol{x}$ if $\boldsymbol{x} \neq \boldsymbol{y}$ and $\{z \in X \mid \boldsymbol{x} \leq z \leq \boldsymbol{y}\} = \{\boldsymbol{x}, \boldsymbol{y}\}$. A grading of X is a length function ℓ on X strictly compatible with $\leq$ and such that $\boldsymbol{y}$ covers $\boldsymbol{x}$ if and only if $\boldsymbol{x} \leq \boldsymbol{y}$ and $\ell(\boldsymbol{y}) - \ell(\boldsymbol{x}) = 1$. Gradings thus give an easy classification of covers and more generally of saturated chains in X. The Bruhat length for a Coxeter group equipped with the Bruhat order is the prototypical example of a grading.

Muthiah and Orr [2019] prove that if Φ is of affine ADE type, the affinized Bruhat length gives a $\mathbb{Z}$-grading of W^a_+ for the affinized Bruhat order and conjecture this to be true in general. Our main result is a positive answer to this conjecture:

Theorem A. *Let $\mathcal{D}$ be any Kac–Moody root datum. Then the affinized length ℓ^a on W^a_+ defines a $\mathbb{Z}$-grading of W^a_+ strictly compatible with the affinized Bruhat order. Otherwise said, let $\boldsymbol{x}, \boldsymbol{y} \in W^a_+$ be such that $\boldsymbol{x} \leq \boldsymbol{y}$. Then*

$$\boldsymbol{y} \text{ covers } \boldsymbol{x} \text{ if and only if } \ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x}) = 1. \tag{0.4}$$

Along the way, we obtain several geometric properties of covers for the affinized Bruhat order which we expect to be insightful even if the root datum is reductive (so W is finite and W^a_+ is an affine Weyl group) as they only rely on the Coxeter structure of W. In particular, we obtain in Proposition 3.20 a classification of covers which generalize results obtained using quantum Bruhat graphs, in the reductive setting by T. Lam and M. Shimozono [2010, Proposition 4.4] and F. Schremmer [2024, Proposition 4.5], and in the affine simply laced setting by A. Welch [2022, Theorem 2].

Further directions. In an upcoming joint work with A. Hébert, we prove that any element of W^a_+ admits a finite number of covers for the affinized Bruhat order. We use this finiteness in the context of masures to define R-Kazhdan–Lusztig polynomials, following Muthiah's strategy exposed in [2019] and the work on twin masures of N. Bardy-Panse, A. Hébert and G. Rousseau [Bardy-Panse et al. 2022]. Our understanding of covers is useful to compute these R-polynomials and we intend to use R-polynomials to define P-Kazhdan–Lusztig polynomials.

Another interesting (but quite long reach) question is the following: W^a_+ appears as the affinization of W, which may be taken as an affinized version of a finite

Coxeter group. Can we iterate the affinization process, e.g., to obtain a valid theory for reductive or Kac–Moody groups on valued fields of higher dimensions?

Lastly, little is known on the preorder defined on the whole semidirect product W^a; it could be insightful to study it and to connect it to the failure of the full Iwahori–Matsumoto decomposition of G.

Organization of the paper.

Proof strategy. The global strategy is to construct a nontrivial chain from $\boldsymbol{x}$ to $\boldsymbol{y}$ every time $\boldsymbol{y} \geq \boldsymbol{x}$ satisfies $\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x}) > 1$. Let proj^{Y^+} denote the projection $W^a_+ = Y^+ \rtimes W \to Y^+$. We distinguish two cases which depend on the form of $\boldsymbol{x}$ and $\boldsymbol{y}$: The first case is when $\mathrm{proj}^{Y^+}(\boldsymbol{y})$ lies in the orbit of $\mathrm{proj}^{Y^+}(\boldsymbol{x})$, we call such covers the vectorial covers. The other case is when $\mathrm{proj}^{Y^+}(\boldsymbol{y}) \notin W \cdot \mathrm{proj}^{Y^+}(\boldsymbol{x})$, we call such covers the properly affine covers.

For vectorial covers we show that the affinized Bruhat order on the set $\{\boldsymbol{z} \in W^a_+ \mid \boldsymbol{x} \leq \boldsymbol{z} \leq \boldsymbol{y}\}$ is, in some sense, a lift of several Bruhat-like orders on W. We are then able to construct chains between $\boldsymbol{x}$ and $\boldsymbol{y}$ from chains in W, and we deduce a classification of vectorial covers. The characterization of properly affine covers is, at first glance, more involved. Through a careful study of the relation between the vectorial chamber containing $\mathrm{proj}^{Y^+}(\boldsymbol{x})$ and the vectorial chamber containing $\mathrm{proj}^{Y^+}(\boldsymbol{y})$, we show that the length difference $\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x})$ can be rewritten in a more workable form, making clear the conditions for which it is equal to one. Then the difficulty is to build, explicitly, a nontrivial chain every time one of these conditions is not satisfied.

Organization. Section 1 consists of preliminaries. In Section 1.1 we formally define everything we mentioned in this introduction. In particular we give the definition of the affinized Bruhat order and the two affinized Bruhat lengths as they are given in [Muthiah and Orr 2019]. To be more flexible, we chose to define the affinized Bruhat preorder on the whole affinized Weyl group $W^a = Y \rtimes W$, on which it may not be an order.

We show, amongst other preliminary results, that we indeed recover the affinized Bruhat order on W^a_+ from this preorder in Section 1.3.

We also give, in Section 1.2, a geometric interpretation of W^a_+ and its affinized Bruhat order, which is to be compared with the interpretation of the Bruhat order in the Coxeter complex of a Coxeter group. Even though it is not clearly mentioned in the rest of the paper, this geometric interpretation was very useful to construct chains and understand W^a_+.

In Section 2, we prove Theorem A for vectorial covers. We define relative versions of the Bruhat order on W in Section 2.1 and we connect these relative

Bruhat orders to the affinized Bruhat length in Section 2.2. This is enough to prove Theorem A when $\mathrm{proj}^{Y^+}(\boldsymbol{y}) = \mathrm{proj}^{Y^+}(\boldsymbol{x})$ (see Theorem 2.13). Using finer results on parabolic quotients in Section 2.3, we extend it to vectorial covers such that $\mathrm{proj}^{Y^+}(\boldsymbol{y}) \in W \cdot \mathrm{proj}^{Y^+}(\boldsymbol{x}) \setminus \{\mathrm{proj}^{Y^+}(\boldsymbol{x})\}$ (see Theorem 2.18).

In Section 3, we deal with properly affine covers. We first show in Section 3.1 that these covers are of a very specific form. Namely, if $\boldsymbol{x} = \pi^{v(\lambda)}w$ with $v, w \in W$ and $\lambda \in Y^{++}$, then $\boldsymbol{y}$ needs to be of the form $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ or $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$ for some $\beta \in \Phi_+$.

The strategy is then to get enough necessary conditions on v, w, λ, β for $\boldsymbol{y}$ to cover $\boldsymbol{x}$, in order to obtain a simplified expression for $\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x})$. Proposition 3.3 gives a first result in this direction. In Section 3.2 we fully exploit this strategy to obtain (3.14) for the length difference.

Finally, in Sections 3.3 and 3.4, we construct various chains from $\boldsymbol{x}$ to $\boldsymbol{y}$ to prove that the quantities appearing in (3.14) need to be minimal when $\boldsymbol{y}$ covers $\boldsymbol{x}$, which allows us to conclude the argument in Section 3.5.

1. Preliminaries

1.1. ***Definitions and notation.*** Let $\mathcal{D} = (A, X, Y, (\alpha_i)_{i\in I}, (\alpha_i^\vee)_{i\in I})$ be a Kac–Moody root datum as defined in [Rémy 2002, §8]. It is a quintuplet such that:

(1) I is a finite indexing set and $A = (a_{ij})_{(i,j)\in I\times I}$ is a generalized Cartan matrix.

(2) X and Y are two dual free $\mathbb{Z}$-modules of finite rank, and we write $\langle \cdot, \cdot \rangle$ for the duality bracket.

(3) $(\alpha_i)_{i\in I}$ (resp. $(\alpha_i^\vee)_{i\in I}$) is a family of linearly independent elements of X (resp. Y): the simple roots (resp. simple coroots).

(4) For all $(i, j) \in I^2$ we have $\langle \alpha_i^\vee, \alpha_j \rangle = a_{ij}$.

1.1.1. *Vectorial Weyl group.* For every $i \in I$ set $s_i \in \mathrm{Aut}_{\mathbb{Z}}(X) : x \mapsto x - \langle \alpha_i^\vee, x\rangle \alpha_i$. The generated group $W = \langle s_i \mid i \in I\rangle$ is the *vectorial Weyl group* of the Kac–Moody root datum.

The duality bracket $\langle \cdot, \cdot \rangle$ induces a contragredient action of W on Y, explicitly $s_i(y) = y - \langle y, \alpha_i\rangle \alpha_i^\vee$. The bracket is then W-invariant.

The vectorial Weyl group W is a Coxeter group with set of simple reflections $S = \{s_i \mid i \in I\}$; in particular it has a Bruhat order $<$ and a length function ℓ compatible with the Bruhat order. We refer to [Björner and Brenti 2005] for general definitions and properties of Coxeter groups. A *reflection* in a Coxeter group is any element conjugated to a simple reflection.

1.1.2. *Real roots.* Let $\Phi = W \cdot \{\alpha_i \mid i \in I\}$ be the set of real roots of $\mathcal{D}$. It is a root system in the classical sense, but possibly infinite. In particular let $\Phi_+ = \bigoplus_{i\in I} \mathbb{N}\alpha_i \cap \Phi$ be the set of positive real roots. Then $\Phi = \Phi_+ \sqcup -\Phi_+$, and we write $\Phi_- = -\Phi_+$ for the set of negative roots.

The set $\Phi^\vee = W \cdot \{\alpha_i^\vee \mid i \in I\}$ is the set of *coroots*, and its subset $\Phi_+^\vee = \bigoplus_{i\in I} \mathbb{N}\alpha_i^\vee \cap \Phi^\vee$ is the set of *positive coroots*.

To each root β corresponds a unique coroot $\beta^\vee$: if $\beta = w(\alpha_i)$ then $\beta^\vee = w(\alpha_i^\vee)$. This map $\beta \mapsto \beta^\vee$ is well defined, bijective between Φ and $\Phi^\vee$ and sends positive roots to positive coroots. Note that $\langle \beta^\vee, \beta\rangle = 2$ for all $\beta \in \Phi$.

To each root β we associate a reflection $s_\beta \in W$: if $\beta = w(\pm\alpha_i)$ then $s_\beta := ws_iw^{-1}$. Explicitly it is the map $X \to X$ defined by $s_\beta(x) = x - \langle \beta^\vee, x\rangle\beta$. For any $\beta \in \Phi$ we have $s_\beta = s_{-\beta}$ and the map $\beta \mapsto s_\beta$ forms a bijection between the set of positive roots and the set $\{ws_iw^{-1} \mid (w, i) \in W \times I\}$ of reflections of W.

1.1.3. *Inversion sets.* For any $w \in W$, let

$$\mathrm{Inv}(w) = \Phi_+ \cap w^{-1}(\Phi_-) = \{\alpha \in \Phi_+ \mid w(\alpha) \in \Phi_-\}.$$

These sets are strongly connected to the Bruhat order, as by [Kumar 2002, 1.3.13], for all $\alpha \in \Phi_+$

$$\alpha \in \mathrm{Inv}(w) \iff ws_\alpha < w \iff s_\alpha w^{-1} < w^{-1}. \tag{1.1}$$

They are related to the Bruhat length: $\ell(w) = |\mathrm{Inv}(w)|$ [Kumar 2002, 1.3.14].

1.1.4. *Fundamental chamber and Tits cone.* We define the (closed) integral fundamental chamber by $Y^{++} = \{\lambda \in Y \mid \langle \lambda, \alpha_i\rangle \geq 0\ \forall i \in I\}$. If $\lambda \in Y^{++}$, we say that it is a *dominant coweight*. Then, the integral Tits cone is $Y^+ := \bigcup_{w\in W} w(Y^{++})$. It is a convex cone of Y; in particular it is a semigroup for the group operation of Y, and it is equal to Y if and only if W is finite, if and only if Φ is finite, if and only if A is of finite type (see [Kumar 2002, 1.4.2]).

The integral fundamental chamber Y^{++} is a fundamental domain for the action of W on Y^+, and for any $\lambda \in Y^+$ we define λ^{++} to be the unique element of Y^{++} in its W-orbit.

There is a height function on Y^+, defined as follows:

Definition 1.1. Let $(\Lambda_i)_{i\in I}$ be a set of fundamental weights, that is to say $\langle \alpha_i^\vee, \Lambda_i\rangle = \delta_{ij}$ for any $i, j \in I$. We fix it once and for all. Let $\rho = \sum_{i\in I} \Lambda_i$. Then for any $\lambda \in Y$ define the *height* of λ as

$$\mathrm{ht}(\lambda) = \langle \lambda, \rho\rangle. \tag{1.2}$$

The height depends on the choice of fundamental weights, but its restriction to $Q^\vee = \bigoplus_{i\in I} \mathbb{Z}\alpha_i^\vee$ do not:

$$\mathrm{ht}\Big(\sum_{i\in I} n_i\alpha_i^\vee\Big) = \sum_{i\in I} n_i.$$

Remark 1.2. The height function takes integral values on $Q^\vee$, but not necessarily on Y. In general, one can choose the fundamental weights such that $\mathrm{ht}(Y) \subset \frac{1}{N_{\mathrm{ht}}}\mathbb{Z}$ for some $N_{\mathrm{ht}} \in \mathbb{Z}_{>0}$. As noted by D. Muthiah and A. Puskás [2024, Remark 2.13], if $\mathcal{D}$ is of finite or affine type then the fundamental weights may be chosen such that $N_{\mathrm{ht}} \in \{1, 2\}$, but for more general Kac–Moody root systems the optimal choice for N_{ht} may be arbitrarily large.

1.1.5. *Parabolic subgroups, minimal coset representatives.* For $\lambda \in Y^+$, let Φ_λ denote the set $\{\alpha \in \Phi \mid \langle\lambda, \alpha\rangle = 0\}$ and $W_\lambda = \mathrm{Stab}_W(\lambda)$. We say that λ is *regular* if $\Phi_\lambda = 0$, or equivalently if $W_\lambda = 1_W$. More generally we say that λ is *spherical* if W_λ is finite.

Let $v \in W$ be such that $\lambda = v\lambda^{++}$. Then $W_\lambda v = vW_{\lambda^{++}}$ and, since λ^{++} is dominant, $W_{\lambda^{++}}$ is a standard parabolic subgroup, that is, a group of the form $W_J = \langle s \mid s \in J\rangle$ where $J \subset S$ is a set of simple reflections. More precisely, $J = \{s \in S \mid s(\lambda^{++}) = \lambda^{++}\}$.

By standard Coxeter group theory (see, for instance, [Björner and Brenti 2005, Section 2.2]), for any $u \in W$, the left coset $uW_{\lambda^{++}} = uW_J$ has a unique representative of minimal length which we denote by u^J, and one has a decomposition $u = u^J u_J$ with $u_J \in W_J$ such that

$$\ell(u) = \ell(u^J) + \ell(u_J). \tag{1.3}$$

Notation 1.3. (1) For any $J \subset S$, we denote by W^J the set of minimal length representatives for W_J-cosets in W:

$$w \in W^J \iff \forall \widetilde{w} \in W_J,\ \ell(w\widetilde{w}) > \ell(w) \iff \forall s \in J,\ \ell(ws) > \ell(w). \tag{1.4}$$

If $\lambda \in Y^{++}$ is such that $W_\lambda = W_J$, then we may use W^λ as an alternative notation for W^J.

(2) For any $\lambda \in Y^+$ (not necessarily dominant), we denote by v^λ the minimal length element in W which satisfies $\lambda = v^\lambda\lambda^{++}$:

$$v^\lambda = \min\{v \in W \mid \lambda = v\lambda^{++}\}. \tag{1.5}$$

In other words, for any $u \in W$ such that $\lambda = u\lambda^{++}$, we have $v^\lambda = u^J$, where J is the set of simple reflections such that $W_J = W_{\lambda^{++}}$.

1.1.6. *Affinized Weyl semigroup.* The action of W on Y allows us to form the semidirect product $Y \rtimes W$, which we denote by W^a. We denote its elements by $\pi^\lambda w$ with $\lambda \in Y$, $w \in W$.

By definition, $Y^+ \subset Y$ is stable by the action of W on Y; therefore we can form $W^a_+ = Y^+ \rtimes W$ which is a subsemigroup of W^a. This semigroup is called the affinized Weyl semigroup. Muthiah and Orr [2019] define a Bruhat order and an associated length function on W^a_+ which we aim to study in this article.

Denote by $\mathrm{proj}^{Y^+} : W^a_+ \to Y^+$ the canonical projection, which sends $\pi^\lambda w$ onto λ. Denote by $\mathrm{proj}^{Y^{++}} : W^a_+ \to Y^{++}$ the projection to Y^{++}: $\mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = (\mathrm{proj}^{Y^+}(\boldsymbol{x}))^{++}$. Let us call $\mathrm{proj}^{Y^+}(\boldsymbol{x})$ the *coweight* of $\boldsymbol{x}$, and $\mathrm{proj}^{Y^{++}}(\boldsymbol{x})$ its *dominance class*.

1.1.7. *Affinized roots.* Let $\Phi^a = \Phi \times \mathbb{Z}$ be *the set of affinized roots* and denote by $\beta + n\pi$ the affinized root (β, n). The affinized root $\beta + n\pi$ is said to be positive if $n > 0$ or ($n = 0$ and $\beta \in \Phi_+$) and we write Φ^a_+ for the set of positive affinized roots. We have $\Phi^a = \Phi^a_+ \sqcup -\Phi^a_+$.

The semidirect product W^a acts on Φ^a by

$$\pi^\lambda w(\beta + n\pi) = w(\beta) + \big(n + \langle \lambda, w(\beta)\rangle\big)\pi. \tag{1.6}$$

For any $n \in \mathbb{Z}$, its sign is denoted $\mathrm{sgn}(n) \in \{-1, +1\}$, with the convention that $\mathrm{sgn}(0) = +1$. Note that $|n| = \mathrm{sgn}(n)n$. We also define the sign of an affinized root: $\mathrm{sgn}(\beta + n\pi) \in \{-1, +1\}$ and $\mathrm{sgn}(\beta + n\pi) = +1$ if and only if $\beta + n\pi \in \Phi^a_+$.

For $n \in \mathbb{Z}$ and $\beta \in \Phi_+$, set

$$\beta[n] = \mathrm{sgn}(n)\beta + |n|\pi \in \Phi^a_+, \tag{1.7}$$

$$s_{\beta[n]} = \pi^{n\beta^\vee} s_\beta. \tag{1.8}$$

We also define $\beta[n] \in \Phi^a_+$ for $\beta \in \Phi_-$ by $\beta[n] = (-\beta)[-n]$, and $s_{\beta[n]} = s_{-\beta[-n]} = \pi^{n\beta^\vee} s_\beta$. The affinized root $\beta[n]$ is therefore the positive affinized root within the pair $\{\beta + n\pi, -(\beta + n\pi)\}$. Note that $s_{\beta[0]}$ is the vectorial reflection s_β.

1.1.8. *Bruhat order on W^a_+.* Recall Braverman, Kazhdan and Patnaik's definition of the Bruhat order $<$ introduced in [Braverman et al. 2016, Section B.2]: Let $\boldsymbol{x} \in W^a_+$ and let $\beta[n] \in \Phi^a_+$ be such that $\boldsymbol{x} s_{\beta[n]} \in W^a_+$. Then,

$$\boldsymbol{x} < \boldsymbol{x} s_{\beta[n]} \iff \mathrm{sgn}(\beta + n\pi) = \mathrm{sgn}(\boldsymbol{x}(\beta + n\pi)) \iff \boldsymbol{x}(\beta[n]) \in \Phi^a_+. \tag{1.9}$$

Explicitly, if $\boldsymbol{x} = \pi^\lambda w \in W^a_+$, the right-hand side condition can be written as

$$\mathrm{sgn}(n)\big(n + \langle \lambda, w(\beta)\rangle\big) > 0 \qquad \text{or} \qquad n = -\langle \lambda, w(\beta)\rangle \quad \text{and} \quad \mathrm{sgn}(n) w(\beta) > 0.$$

Then we extend this relation by transitivity, which makes it a preorder on W^a_+. Originally, Braverman, Kazhdan and Patnaik defined it only for affine vectorial

Weyl groups, but the definition extends to any vectorial Weyl group and Muthiah [2018] showed that it is an order on W^a_+ in general.

1.1.9. *Extension to W^a.* As the whole semidirect product W^a acts on Φ^a_+, (1.9) makes sense for any $\boldsymbol{x} \in W^a$, and we define $<$ on W^a as the closure by transitivity of the relation defined through (1.9) for $\boldsymbol{x} \in W^a$. We show in the next section that if $\boldsymbol{x} < \boldsymbol{y}$ and $\boldsymbol{y} \in W^a_+$, then $\boldsymbol{x} \in W^a_+$. This ensures that the restriction of the W^a-preorder to W^a_+ coincides with Braverman, Kazhdan, Patnaik's order on W^a_+. However $<$ may not be an order on W^a.

1.1.10. *Bruhat order through a right action.* We consider multiplication by reflections on the left. To switch between the right and left actions note that

$$s_{\beta[n]}\pi^\lambda w = \pi^{s_\beta\lambda+n\beta^\vee} s_\beta w = \pi^{\lambda+(n-\langle\lambda,\beta\rangle)\beta^\vee} w s_{w^{-1}(\beta)} = \pi^\lambda w s_{w^{-1}(\beta)[n-\langle\lambda,\beta\rangle]}. \tag{1.10}$$

In particular,

$$s_{\beta[0]}\pi^\lambda w = s_\beta \pi^\lambda w = \pi^{s_\beta\lambda} s_\beta w \quad\text{and}\quad s_{\beta[\langle\lambda,\beta\rangle]}\pi^\lambda w = \pi^\lambda s_\beta w = \pi^\lambda w s_{w^{-1}(\beta)}. \tag{1.11}$$

Using (1.10), the affinized Bruhat order can be recovered using a right action of W^a on Φ^a_+.

Proposition 1.4. *Let $\pi^\lambda w \in W^a$ and $(\beta, n) \in (\Phi \times \mathbb{Z}) \setminus (\Phi_- \times \{0\})$. Then*

$$s_{\beta[n]}\pi^\lambda w > \pi^\lambda w \iff \operatorname{sgn}(n) w^{-1}(\beta) + (|n| - \operatorname{sgn}(n)\langle\lambda, \beta\rangle)\pi \in \Phi^a_+. \tag{1.12}$$

Remark 1.5. The root appearing in the right-hand side of (1.12) is the affinized root $(\pi^\lambda w)^{-1}(\beta[n])$.

Proof. Let $\pi^\lambda w \in W^a$ and $\beta + n\pi \in \Phi^a$. Then by (1.9) and (1.10),

$$s_{\beta[n]}\pi^\lambda w > \pi^\lambda w \iff \operatorname{sgn}(\beta + n\pi) = \operatorname{sgn}\big(w^{-1}(\beta) + (n - \langle\lambda, \beta\rangle)\pi\big).$$

If $(\beta, n) \notin \Phi_- \times \{0\}$, then $\beta[n] = \operatorname{sgn}(n)(\beta + n\pi)$ so this is equivalent to

$$\operatorname{sgn}(n)\big(w^{-1}(\beta) + (n - \langle\lambda, \beta\rangle)\pi\big) \in \Phi^a_+,$$

which is (1.12). □

Note that (1.12) is no longer correct if $\beta \in \Phi_-$ and $n = 0$, in which case it needs to be applied to $(-\beta)[0]$. Applying reflections on the left is better suited for the geometric interpretation we will give in Section 1.2.

1.1.11. *Terminology on partially ordered sets.* For $p \le q \in \mathbb{Z}$, we denote by $[\![p, q]\!]$ the set $\{r \in \mathbb{Z} \mid p \le r \le q\}$. If $p > q$, then $[\![p, q]\!]$ is another notation for $[\![q, p]\!]$. We also write $]\!]p, q[\![$ for $[\![p, q]\!] \setminus \{p, q\}$.

Let $(\mathcal{P}, \le)$ be a partially ordered set. For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}$, we say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are comparable if $\boldsymbol{x} \le \boldsymbol{y}$ or $\boldsymbol{y} \le \boldsymbol{x}$. We say that $\boldsymbol{y}$ *covers* $\boldsymbol{x}$, written as $\boldsymbol{x} \lhd \boldsymbol{y}$, if $\boldsymbol{x} \ne \boldsymbol{y}$ and

$\{z \mid x \leq z \leq y\} = \{x, y\}$. If $\mathcal{P} = W_a^+$, covers $x \lhd y$ such that $\mathrm{proj}^{Y^{++}}(y) = \mathrm{proj}^{Y^{++}}(x)$ are called *vectorial covers* and covers which are not vectorial covers are called *properly affine covers*.

A *chain* from x to y is a finite sequence $(x_0, \dots, x_n)$ such that $x_0 = x$, $x_n = y$ and $x_k \leq x_{k+1}$ for all $k \in [\![0, n-1]\!]$. If $\mathcal{P} = W_+^a$ (resp. if $\mathcal{P}$ is a Coxeter group), we add the condition that $x_{k+1}x_k^{-1}$ is an affinized reflection (resp. a reflection). A chain is *saturated* if $x_k \lhd x_{k+1}$ for all $k \in [\![0, n-1]\!]$. We say that a subset $\mathcal{C}$ of $\mathcal{P}$ is *convex* if, for all $x, y \in \mathcal{C}$ and $z \in \mathcal{P}$,

$$x \leq z \leq y \implies z \in \mathcal{C}. \tag{1.13}$$

Equivalently, $\mathcal{C}$ is convex if and only if any chain from one element of $\mathcal{C}$ to another is contained in $\mathcal{C}$.

Let $\ell : \mathcal{P} \to A$ be a function with values in a totally ordered set $(A, \leq_A)$. We say that it is *order-preserving* if, for all $x, y \in \mathcal{P}$,

$$x \leq y \implies \ell(x) \leq_A \ell(y). \tag{1.14}$$

We say that ℓ is a *strictly compatible* (A-valued) length function if

$$x < y \iff x, y \text{ are comparable and } \ell(x) <_A \ell(y). \tag{1.15}$$

We say that a strictly compatible $\mathbb{R}$-valued length function ℓ defines a *$\mathbb{Z}$-grading* of $\mathcal{P}$ if

$$x \lhd y \iff x \leq y \text{ and } \ell(y) = \ell(x) + 1. \tag{1.16}$$

For instance, the Bruhat length on a Coxeter group W is strictly compatible with the Bruhat order, and defines a $\mathbb{N}$-valued grading of W. Muthiah and Orr associated length functions strictly compatible with the Bruhat order on W_+^a, generalizing the classical Bruhat length on Coxeter groups. We now formally introduce these lengths.

1.1.12. *Length functions on W_+^a.*

Definition 1.6. The *affinized length function* is the map $W_+^a \to \mathbb{R} \oplus \varepsilon\mathbb{Z}$ defined by

$$\ell_\varepsilon^a(\pi^\lambda w) = 2\,\mathrm{ht}(\lambda^{++}) + \varepsilon\big(\big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle \lambda, \alpha\rangle \geq 0\}\big| - \big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle \lambda, \alpha\rangle < 0\}\big|\big).$$

The *affinized length with real values* is the affinized length function on which we set $\varepsilon = 1$:

$$\ell^a(\pi^\lambda w) = 2\,\mathrm{ht}(\lambda^{++}) + \big(\big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle \lambda, \alpha\rangle \geq 0\}\big| - \big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle \lambda, \alpha\rangle < 0\}\big|\big).$$

Theorem 1.7 [Muthiah 2018, Theorem 4.24; Muthiah and Orr 2019, Theorem 3.6]. *The affinized length function and the affinized length function with real values are* *strictly compatible with the affinized Bruhat order on W^a_+. In other words, for any $\boldsymbol{x} \in W^a_+$ and $\beta[n] \in \Phi^a_+$,*

$$\boldsymbol{x} s_{\beta[n]} > \boldsymbol{x} \iff \ell^a_\varepsilon(\boldsymbol{x} s_{\beta[n]}) > \ell^a_\varepsilon(\boldsymbol{x}) \iff \ell^a(\boldsymbol{x} s_{\beta[n]}) > \ell^a(\boldsymbol{x}). \tag{1.17}$$

In particular the affinized Bruhat order is a partial order.

Remark 1.8. The affinized length functions depend on the choice made for the height function. However since $\mathrm{proj}^{Y^{++}}(s_{\beta[n]}\boldsymbol{x}) \in \mathrm{proj}^{Y^{++}}(\boldsymbol{x}) + Q^\vee$ for any $\boldsymbol{x} \in W^a_+$ and $\beta[n] \in \Phi^a_+$ such that $s_{\beta[n]}\boldsymbol{x} \in W^a_+$ (this is a consequence of Corollary 1.12 below and of (1.10)), by Remark 1.2 the length difference between two comparable elements do not depend on the choice of height function. By the same remark if $\mathcal{D}$ is of finite or affine type, then the height function may be chosen such that ℓ^a takes integral values; therefore it was first introduced by Muthiah and Orr as "the affinized length with integral values". In general type, ℓ^a may take nonintegral values for every choice of height function but this could be artificially fixed: one could also define a strictly compatible length with integral values $\ell^a_{\mathbb{Z}}$ by setting $\ell^a_{\mathbb{Z}}(\boldsymbol{x}) = \lfloor \ell^a(\boldsymbol{x}) \rfloor$ (where $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$ denotes any $\mathbb{Z}$-equivariant function).

In what follows we will mostly use ℓ^a and rarely mention ℓ^a_ε. We now refer to ℓ^a as the affinized Bruhat length.

1.2. ***Geometric interpretation.*** We introduced everything in a very algebraic way, but there is a strong geometric intuition behind root systems, vectorial Weyl groups and the vectorial Bruhat order, developed, for instance, in the context of buildings in [Ronan 1989]. There is also a geometrical interpretation of the Bruhat order on W^a_+ which we develop in this paragraph; it takes place in the standard apartment of the masure associated to a Kac–Moody group with underlying Kac–Moody datum $\mathcal{D}$.

Let $V = Y \otimes_{\mathbb{Z}} \mathbb{R}$. The lattice X embeds in its dual $V^\vee$ and the vectorial Weyl group W acts naturally on it. Inside V we have the (closed) fundamental chamber $C^v_f = \{v \in V \mid \langle v, \alpha_i \rangle \geq 0\}$ and the Tits cone $\mathcal{T} = W \cdot C^v_f$. A *vectorial chamber* is a set of the form $w \cdot C^v_f$ for $w \in W$. Since the interior of C^v_f has trivial stabilizer in W, the set of chambers is in natural bijection with W by $w \mapsto C^v_w := w \cdot C^v_f$.

To each root $\beta \in \Phi_+$ let $M_\beta = \{x \in V \mid \langle x, \beta \rangle = 0\}$; it is a hyperplane of V and, if $\beta = w(\alpha_i)$ with α_i a simple root, then $C^v_w \cap C^v_{ws_i} \subset M_\beta \cap \mathcal{T}$. The intersection $C^v_w \cap C^v_{ws_i}$ is called the panel of type s_i of w.

We can put a structure of simplicial complex on $\mathcal{T}$, for which chambers are the cells of maximal rank and panels are the cells of maximal rank within nonchambers. This simplicial complex is a realization of the *Coxeter complex* of (W, S). Each wall splits the Tits cone in two parts, and separate the set of vectorial chambers in two: say

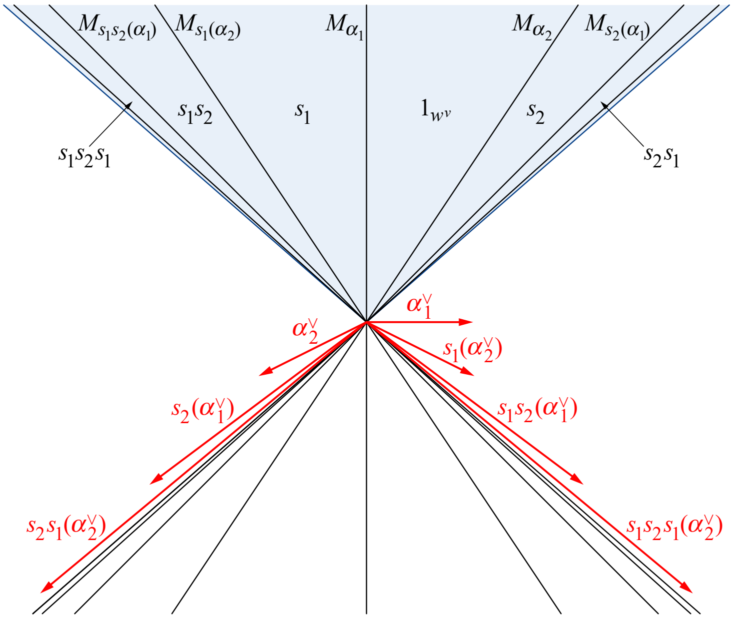

Figure 1. The Tits cone for a root system of Cartan Matrix $\left(\begin{smallmatrix} 2 & -3 \\ -2 & 2 \end{smallmatrix}\right)$.

that C^v_w is on the positive side of M_β if $w^{-1}(\beta) > 0$. In particular since β is a positive root, the positive side is always the one which contains the dominant chamber.

Then the vectorial Bruhat order can be interpreted by $s_\beta w > w$ if and only if, when we split $\mathcal{T}$ along M_β, the chambers C^v_w and C^v_f are in the same connected component of $\mathcal{T}$, that is to say C^v_w is on the positive side of M_β.

The inversion set of w^{-1}, $\mathrm{Inv}(w^{-1})$, can be interpreted as the set of walls separating the chamber $C^v_w = w \cdot C^v_f$ from the fundamental chamber C^v_f.

In Figure 1 we represent the Tits cone and its structure for a root system of rank 2 with Cartan matrix $\left(\begin{smallmatrix} 2 & -3 \\ -2 & 2 \end{smallmatrix}\right)$, which is of indefinite type. The Tits cone is colored in blue, and the vectorial chamber C^v_w is labeled by w. It is an approximation since W is infinite.

Let us now turn to the interpretation of the W^a_+-Bruhat order. Let $\mathbb{A}$ be a real affine space with direction V, we call $\mathbb{A}$ the (standard) affine apartment associated to $\mathcal{D}$. The tangent space of $\mathbb{A}$ is canonically isomorphic to $T\mathbb{A} = \mathbb{A} \times V$, with, for any $x \in \mathbb{A}$, $T_x\mathbb{A} = \{x\} \times V$.

The semigroup W^a_+ has an affine action on $\mathbb{A}$, given by $\pi^\lambda w(x) = -\lambda + w(x)$, which induces an action on $T\mathbb{A}$ given by $\pi^\lambda w((x, v)) = (-\lambda + w(x), w(v))$. To any positive affinized root $\beta[n] \in \Phi^a_+$ corresponds an affine hyperplane

$$M_{\beta[n]} = \{x \in \mathbb{A} \mid \langle x, \beta\rangle + n = 0\}, \tag{1.18}$$

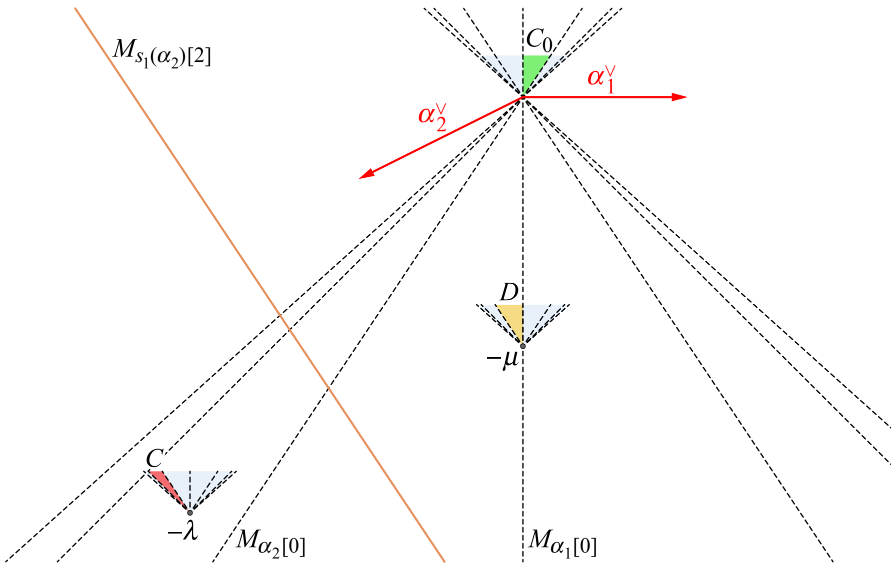

Figure 2. The affine apartment for a root system of Cartan Matrix $\begin{pmatrix} 2 & -3 \\ -2 & 2 \end{pmatrix}$.

the *affine wall* associated to the affinized root $\beta[n]$. For any $x \in M_{\beta[n]}$ we have $T_x M_{\beta[n]} = \{x\} \times M_\beta \subset T_x\mathbb{A}$.

For any $\pi^\lambda w \in W_+^a$ let

$$C_{\pi^\lambda w} = \{-\lambda\} \times C_w^v \subset T_{-\lambda}\mathbb{A} \subset T\mathbb{A}, \tag{1.19}$$

we call it the *alcove of type* $\pi^\lambda w$. Mirroring the classical situation, $C_0 = \{0\} \times C_f^v$ is a fundamental domain for the action of W_+^a on $Y^+ \times \mathcal{T} \subset T\mathbb{A}$ and W_+^a acts on $\{C_{\boldsymbol{x}} \mid \boldsymbol{x} \in W_+^a\}$ simply transitively. Affine walls separate naturally the set of alcoves in two and we call the side containing C_0 the positive side.

Then the W_+^a-Bruhat order can be interpreted geometrically:

$$s_{\beta[n]}\pi^\lambda w > \pi^\lambda w \iff C_{\pi^\lambda w} \text{ is on the positive side of } M_{\beta[n]}. \tag{1.20}$$

We give an illustration of the affine apartment in Figure 2.

In Figure 2 we represent the affine apartment for the same root datum as in Figure 1. The blue polygons represent the local Tits cones at three different points: the origin, $-\lambda \in -Y^+$ and $-\mu$, which is the image of $-\lambda$ by the reflection along the wall $M_{s_1(\alpha_2)[2]}$ (represented in yellow).

We have highlighted three alcoves: In green the alcove C_0; in red the alcove $C = C_{\pi^\lambda s_1 s_2}$ and in yellow $D = C_{\pi^\mu s_1}$ which is the image of C by $s_{s_1(\alpha_2)[2]}$. We see that D is on the same side of $M_{s_1(\alpha_2)[2]}$ as the fundamental alcove C_0; thus $\pi^\lambda s_1 s_2 = s_{s_1(\alpha_2)[2]}(\pi^\mu s_1) > \pi^\mu s_1$.

Note that $-\lambda$ lies in the negative vectorial chamber $-s_2C^v_f$, that is to say that $s_2\lambda$ is dominant. Therefore $\pi^\lambda s_2$ is the minimal length element of $\pi^\lambda W$. We will make this more explicit in Section 2.2.

1.2.1. *Notation for segments.* For any two elements $x, y \in V = Y \otimes_{\mathbb{Z}} \mathbb{R}$, we define

$$[x, y] = \{tx + (1-t)y \mid 0 \le t \le 1\} \quad \text{and} \quad]x, y[= \{tx + (1-t)y \mid 0 < t < 1\}.$$

Note in particular that, if $x \in Y$ and $y = x + n\beta^\vee$ for $n \in \mathbb{Z}$ and $\beta \in \Phi$, then for any $m \in [\![0, n]\!]$ we have $x + m\beta^\vee \in [x, y] \cap Y$.

1.3. *Preliminary results.* Since the affinized Bruhat order is generated on W^a_+ by the relations $s_{\beta[n]}\boldsymbol{x} > \boldsymbol{x} \iff \ell^a(s_{\beta[n]}\boldsymbol{x}) > \ell^a(\boldsymbol{x})$ for affinized roots $\beta[n] \in \Phi^a_+$, covers are always of this form. In the rest of the paper, we always apply affinized reflections on the left.

Lemma 1.9. *Let $\pi^\lambda w \in W^a$ and $\beta[n] \in \Phi^a_+$. Write $\pi^\mu w'$ for $s_{\beta[n]}\pi^\lambda w$ and suppose that $(\pi^\lambda w)^{-1}(\beta[n]) \in \Phi^a_+$. Then $\lambda \in [\mu, s_\beta\mu]$. In particular*

$$\mu \in Y^+ \implies \lambda \in Y^+. \tag{1.21}$$

Proof. Explicitly,

$$\pi^\mu w' = \pi^{n\beta^\vee} s_\beta . \pi^\lambda w = \pi^{s_\beta\lambda + n\beta^\vee} s_\beta w.$$

Thus

$$\mu = s_\beta\lambda + n\beta^\vee = \lambda + (n - \langle\lambda, \beta\rangle)\beta^\vee \quad \text{and} \quad s_\beta\mu = \lambda - n\beta^\vee.$$

Moreover, since $(\pi^\lambda w)^{-1}(\beta[n]) \in \Phi^a_+$, by (1.12),

$$|n| - \operatorname{sgn}(n)\langle\lambda, \beta\rangle = \operatorname{sgn}(n)(n - \langle\lambda, \beta\rangle) \ge 0.$$

Therefore, unless $n - \langle\lambda, \beta\rangle = 0$, n and $n - \langle\lambda, \beta\rangle$ have same sign, and thus $\lambda = s_\beta\mu + n\beta^\vee = \mu - (n - \langle\lambda, \beta\rangle)\beta^\vee$ lies in $[s_\beta\mu, \mu]$. If $n - \langle\lambda, \beta\rangle = 0$ then $\mu = \lambda$ and the result remains true.

The Tits cone $\mathcal{T}$ is convex [Kumar 2002, Proposition 1.4.2c)] and W-stable, so if $\mu \in \mathcal{T}$, then $[\mu, s_\beta\mu]$ is contained in $\mathcal{T}$ for any $\beta \in \Phi$. Therefore in the situation above, if $\mu \in Y^+ = \mathcal{T} \cap Y$, then $\lambda \in [\mu, s_\beta\mu] \cap Y \subset \mathcal{T} \cap Y = Y^+$, and thus $\mu \in Y^+ \implies \lambda \in Y^+$. □

We directly obtain from Lemma 1.9 the following result.

Proposition 1.10. *The affinized Bruhat order defined on W^a_+ coincides with the restriction of the preorder defined through (1.9) on the whole semidirect product W^a.*

1.3.1. *Properties of the height function.* We give here a few elementary results on the height function, which will be useful in our study of the affinized Bruhat length. They are also used in [Muthiah and Orr 2019, Section 3].

Proposition 1.11. *For any $w \in W$,*

$$\rho - w^{-1}(\rho) = \sum_{\gamma \in \mathrm{Inv}(w)} \gamma. \tag{1.22}$$

Proof. This is [Kumar 2002, 1.3.22, Corollary 3], we prove it by induction on the length of w.

(1) If w is a simple reflection s_α then $\mathrm{Inv}(s_\alpha) = \{\alpha\}$ and $\rho - s_\alpha(\rho) = \langle \alpha^\vee, \rho\rangle \alpha = \alpha$ since $\langle \alpha^\vee, \rho\rangle = 1$ by definition of ρ.

(2) Suppose the result is true for elements of length n, and suppose that $\ell(w) = n+1$. Then write $w = w_1 s_\alpha$ for α a simple root and w_1 an element of length n. Then

$$\rho - w(\rho) = \rho - w_1(\rho) + w_1(\rho - s_\alpha(\rho)) = \sum_{\gamma \in \mathrm{Inv}(w_1^{-1})} \gamma + w_1(\alpha)$$

and since $\mathrm{Inv}(w^{-1}) = \mathrm{Inv}(w_1^{-1}) \sqcup \{w_1(\alpha)\}$ we get the result for w. □

Corollary 1.12. *For any positive root $\beta \in \Phi_+$ we have*

$$2\,\mathrm{ht}(\beta^\vee) = \sum_{\gamma \in \mathrm{Inv}(s_\beta)} \langle \beta^\vee, \gamma\rangle. \tag{1.23}$$

All the terms in the sum are positive.

Proof. Let $\beta \in \Phi_+$ be a positive root. Note that $-s_\beta(\beta^\vee) = \beta^\vee$ and thus $\langle \beta^\vee, \rho\rangle = \langle -s_\beta(\beta^\vee), \rho\rangle = \langle \beta^\vee, -s_\beta(\rho)\rangle$. Therefore by Proposition 1.11,

$$2\,\mathrm{ht}(\beta^\vee) = 2\langle \beta^\vee, \rho\rangle = \langle \beta^\vee, \rho - s_\beta(\rho)\rangle = \sum_{\gamma \in \mathrm{Inv}(s_\beta)} \langle \beta^\vee, \gamma\rangle.$$

Also for any $\gamma \in \mathrm{Inv}(s_\beta)$, by definition $\gamma \in \Phi_+$ and $s_\beta(\gamma) = \gamma - \langle \beta^\vee, \gamma\rangle \beta^\vee \in \Phi_-$ so, since β is a positive root, the coefficient $\langle \beta^\vee, \gamma\rangle$ is necessarily positive. □

Corollary 1.13. *Let $\mu \in Y^+$ and $u \in W$ be such that $\mu = u(\mu^{++})$. Then*

$$\mathrm{ht}(\mu^{++}) = \mathrm{ht}(\mu) - \sum_{\tau \in \mathrm{Inv}(u^{-1})} \langle \mu, \tau\rangle. \tag{1.24}$$

The terms in this sum are nonpositive integers and

$$\mathrm{ht}(\mu) \leq \mathrm{ht}(\mu^{++}). \tag{1.25}$$

The inequality is strict unless μ is dominant.

Proof. By definition $\mathrm{ht}(\mu^{++}) = \langle u^{-1}(\mu), \rho\rangle = \langle \mu, u(\rho)\rangle$, and, by Proposition 1.11,

$$\mathrm{ht}(\mu^{++}) = \langle \mu, u(\rho)\rangle = \Big\langle \mu, \rho - \sum_{\tau\in \mathrm{Inv}(u^{-1})} \tau \Big\rangle = \mathrm{ht}(\mu) - \sum_{\tau\in\mathrm{Inv}(u^{-1})} \langle \mu, \tau\rangle.$$

Moreover, for any $\tau \in \Phi$, we have $\langle \mu, \tau\rangle = \langle \mu^{++}, u^{-1}(\tau)\rangle$, so $\tau \in \mathrm{Inv}(u^{-1}) \Longrightarrow \langle \mu, \tau\rangle \le 0$ and the terms of the above sum are all nonpositive; we deduce (1.25). If μ is not dominant, then there exists $\tau \in \Phi_+$ such that $\langle \mu, \tau\rangle < 0$, and thus

$$\mathrm{ht}(\mu) < \mathrm{ht}(s_\tau \mu) \le \mathrm{ht}(\mu^{++}). \qquad \square$$

Amongst other things, Corollary 1.13 directly implies the following result, which was first indicated to the author by Hébert and Muthiah.

Lemma 1.14. *Let $\lambda \in Y^+$ and $\beta \in \Phi_+$ such that $s_\beta\lambda \ne \lambda$. Suppose that $\mu \in]\lambda, s_\beta\lambda[$. Then* $\mathrm{ht}(\mu^{++}) < \mathrm{ht}(\lambda^{++})$.

Proof. Note that we do not suppose $\mu \in Y$. The height function is extended to $V = Y \otimes_{\mathbb{Z}} \mathbb{R}$ linearly. Let $t \in]0, 1[$ be such that $\mu = t\lambda + (1-t)s_\beta\lambda$, and let $v \in W$ be such that $\mu^{++} = v\mu$. Then $\mathrm{ht}(\mu^{++}) = \mathrm{ht}(v\mu) = t\,\mathrm{ht}(v\lambda) + (1-t)\,\mathrm{ht}(vs_\beta\lambda)$. By Corollary 1.13, $\mathrm{ht}(v\lambda) \le \mathrm{ht}(\lambda^{++})$ and $\mathrm{ht}(vs_\beta\lambda) \le \mathrm{ht}(\lambda^{++})$ and, since $s_\beta\lambda \ne \lambda$, at least one of the two inequality is strict. We deduce $\mathrm{ht}(\mu^{++}) < \mathrm{ht}(\lambda^{++})$. $\square$

Proposition 1.15. *Let $\boldsymbol{x} \in W^a_+$ and $\beta[n] \in \Phi^a_+$ such that $s_{\beta[n]}\boldsymbol{x} \in W^a_+$. Then*

$$\mathrm{proj}^{Y^{++}}(s_{\beta[n]}\boldsymbol{x}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{x}) \iff n \in \big\{0, \langle \mathrm{proj}^{Y^+}(\boldsymbol{x}), \beta\rangle\big\}. \tag{1.26}$$

Proof. To simplify notation, let $\lambda \in Y^+$ denote $\mathrm{proj}^{Y^+}(\boldsymbol{x})$. If $n \in \{0, \langle\lambda, \beta\rangle\}$ then by (1.11), $\mathrm{proj}^{Y^+}(s_{\beta[n]}\boldsymbol{x}) \in \{s_\beta(\lambda), \lambda\}$ and therefore it has same dominance class.

Conversely, if $n \in]\!]0, \langle\lambda, \beta\rangle[\![$ then

$$\mathrm{proj}^{Y^+}(s_{\beta[n]}\boldsymbol{x}) = s_\beta(\lambda) + n\beta^\vee \in]\lambda, s_\beta(\lambda)[,$$

and if $n \notin [\![0, \langle\lambda, \beta\rangle]\!]$ then

$$\lambda \in]s_\beta(\lambda) + n\beta^\vee, \lambda - n\beta^\vee[= \big]\mathrm{proj}^{Y^+}(s_{\beta[n]}\boldsymbol{x}), s_\beta(\mathrm{proj}^{Y^+}(s_{\beta[n]}\boldsymbol{x}))\big[.$$

Either way by Lemma 1.14, $\mathrm{ht}(\mathrm{proj}^{Y^{++}}(s_{\beta[n]}\boldsymbol{x})) \ne \mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{x}))$ and in particular $\mathrm{proj}^{Y^{++}}(s_{\beta[n]}\boldsymbol{x}) \ne \mathrm{proj}^{Y^{++}}(\boldsymbol{x})$. $\square$

Remark 1.16. If $n = \langle\lambda, \beta\rangle$, then by (1.11), $s_{\beta[n]}\pi^\lambda w = \pi^\lambda w s_{w^{-1}(\beta)}$. Therefore Proposition 1.15 indicates that, if $\boldsymbol{y} = s_{\beta[n]}\boldsymbol{x}$, then $\mathrm{proj}^{Y^{++}}(\boldsymbol{y}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{x})$ if and only if $\boldsymbol{y}$ is obtained from $\boldsymbol{x}$ by applying a vectorial reflection either on the left-hand side (if $n = 0$) or on the right-hand side (if $n = \langle\lambda, \beta\rangle$). This justifies the terminology for vectorial covers and properly affine covers.

Proposition 1.17. *Let $\boldsymbol{x}, \boldsymbol{y} \in W_+^a$ and suppose that $\boldsymbol{x} \le \boldsymbol{y}$. Then*

$$\mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{x})) \le \mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{y})), \tag{1.27}$$

with equality if and only if $\mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{y})$.

In particular the function $\mathrm{ht} \circ \mathrm{proj}^{Y^{++}} : W_+^a \to \mathbb{R}$ *is order-preserving.*

Proof. It is enough to prove it for cover relations, if $\boldsymbol{y} = s_{\beta[n]}\boldsymbol{x}$ for some $\beta[n] \in \Phi_+^a$. In that case, by Lemma 1.9 we have $\mathrm{proj}^{Y^+}(\boldsymbol{x}) \in \left[\mathrm{proj}^{Y^+}(\boldsymbol{y}), s_\beta(\mathrm{proj}^{Y^+}(\boldsymbol{y}))\right]$. If $\mathrm{proj}^{Y^+}(\boldsymbol{x}) \in \{\mathrm{proj}^{Y^+}(\boldsymbol{y}), s_\beta \,\mathrm{proj}^{Y^+}(\boldsymbol{y})\}$ then they have the same dominance class: $\mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{y})$ and we obtain the equality case.

Otherwise, $\mathrm{proj}^{Y^+}(\boldsymbol{x}) \in \left]\mathrm{proj}^{Y^+}(\boldsymbol{y}), s_\beta(\mathrm{proj}^{Y^+}(\boldsymbol{y}))\right[$, necessarily $s_\beta(\mathrm{proj}^{Y^+}(\boldsymbol{y})) \ne \mathrm{proj}^{Y^+}(\boldsymbol{y})$ and by Lemma 1.14 we deduce $\mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{x})) < \mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{y}))$. □

Corollary 1.18. *For any $\lambda^{++} \in Y^{++}$, the set $\{\boldsymbol{x} \in W_+^a \mid \mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = \lambda^{++}\}$ is convex for the affinized Bruhat order.*

Proof. By Proposition 1.17 the function $W_+^a \to \mathbb{R} : \boldsymbol{x} \mapsto \mathrm{ht} \circ \mathrm{proj}^{Y^{++}}$ is compatible with the affinized Bruhat order. Suppose that $\boldsymbol{x}, \boldsymbol{y} \in W_+^a$ satisfy $\mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{y})$ and $\boldsymbol{x} \le \boldsymbol{y}$. Let $\boldsymbol{z} \in W_+^a$ be such that $\boldsymbol{x} \le \boldsymbol{z} \le \boldsymbol{y}$. Then by Proposition 1.17, $\mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{x})) \le \mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{z})) \le \mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{y})) = \mathrm{ht}(\mathrm{proj}^{Y^{++}}(\boldsymbol{x}))$. By the equality case in Proposition 1.17, we deduce $\mathrm{proj}^{Y^{++}}(\boldsymbol{z}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{x})$. □

Remark 1.19. Note that, for $\lambda \in Y^+$, the set $\{\boldsymbol{x} \in W_+^a \mid \mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = \lambda^{++}\}$ is the double W-orbit of π^λ:

$$\{\boldsymbol{x} \in W_+^a \mid \mathrm{proj}^{Y^{++}}(\boldsymbol{x}) = \lambda^{++}\} = W\pi^\lambda W. \tag{1.28}$$

We show in Section 2 that the right W-orbits $\pi^\lambda W$ are also convex for the affinized Bruhat order.

We end this section with several metric properties of Coxeter groups, the results stated are proved in the context of Coxeter complexes and buildings in [Ronan 1989].

1.3.2. *Metric properties of Coxeter groups.* On any Coxeter group (W_0, S_0) we define a map $d : W_0 \times W_0 \to W_0$ by $d(v, w) = v^{-1}w$, called the *vectorial distance* of W_0. It is W_0-invariant: $d(uv, uw) = d(v, w)$ for any $u, v, w \in W_0$. We also define $d^{\mathbb{N}} = \ell \circ d$ where ℓ is the Bruhat length on (W_0, S_0) (note that ℓ and $d^{\mathbb{N}}$ depend on the set of simple reflections S_0, but the vectorial distance does not). These maps have properties analogous to the standard distance axioms, which justify the name (see [Ronan 1989, Chapter 3, §1]).

An unfolded gallery (resp. a gallery) in W_0 from w to v is a sequence $w = w_0, \dots, w_n = v$ such that $d^{\mathbb{N}}(w_i, w_{i+1}) = 1$ (resp. $d^{\mathbb{N}}(w_i, w_{i+1}) \in \{0, 1\}$) for all

$i \in [\![0, n-1]\!]$. A gallery is said to be minimal if its length n is equal to $d^{\mathbb{N}}(w_1, w_n)$, and a minimal gallery is necessarily unfolded. We refer to [Ronan 1989, Chapter 2] for properties of minimal galleries, but note that if $(w_0, \dots, w_n)$ is a minimal gallery then $d^{\mathbb{N}}(w_0, w_i) = i$ and thus $(w_0, \dots, w_i)$ is a minimal gallery from w_0 to w_i. Since the distance is W_0 invariant, $(vw_0, \dots, vw_n)$ is also a minimal gallery for any $v \in W_0$. The next lemma is a reformulation of [Ronan 1989, Proposition 2.8].

Lemma 1.20. *Let (W_0, S_0) be a Coxeter system and let $v_1, v_2, w \in W_0$ be such that v_2 is not on a minimal gallery from v_1 to w. Then there is a reflection $r \in W_0$ such that $d(v_1, rw) > d(v_1, w)$ and $d(v_2, rw) < d(v_2, w)$.*

Proof. If v_2 is not on a minimal gallery from v_1 to w, by [Ronan 1989, Proposition 2.8] there is a root α — seen as a half-apartment: $\alpha = \{u \in W_0 \mid \ell(u) < \ell(s_\alpha u)\}$ — such that $v_1, w \in \alpha$ and $v_2 \notin \alpha$. Then consider the folding along α, defined by

$$\forall u \in W_0,\ \rho_\alpha(u) = \begin{cases} s_\alpha u & \text{if } u \notin \alpha, \\ u & \text{otherwise.} \end{cases}$$

It reduces the vectorial distance (see [Ronan 1989, §2]); hence

$$\begin{aligned} d(v_1, w) &= d(\rho_\alpha(v_1), \rho_\alpha(s_\alpha w)) < d(v_1, s_\alpha w), \\ d(v_2, s_\alpha w) &= d(s_\alpha v_2, w) = d(\rho_\alpha(v_2), \rho_\alpha(w)) < d(v_2, w). \end{aligned}$$ □

Recall that for $J \subset S$, W_J is the subgroup generated by the set of simple reflections J. The Coxeter system (W_J, J) is an example of Coxeter system for which we will use Lemma 1.20. For any $w \in W$, the coset wW_J is convex, in the sense that, if $w_1, w_2 \in wW_J$, then any minimal gallery from w_1 to w_2 lies in wW_J (see [Ronan 1989, Lemma 2.10]).

Definition 1.21. For any $J \subset S$ and $v, w \in W$, the *projection of w on vW_J* is the unique element of vW_J which reaches $\min_{\tilde{v} \in vW_J} d^{\mathbb{N}}(w, \tilde{v})$. It is denoted by $\text{proj}_{vW_J}(w)$. Any minimal gallery from v to an element of wW_J goes through $\text{proj}_{vW_J}(w)$ (see [Ronan 1989, Theorem 2.10]).

2. Restriction to constant dominance classes

We study the affinized Bruhat order restricted to a dominance class, that is to say, for a given $\lambda^{++} \in Y^{++}$, we study the restriction of the affinized Bruhat order to the subset $(\text{proj}^{Y^{++}})^{-1}(\lambda^{++}) = W\pi^{\lambda^{++}}W$. By Corollary 1.18 these are convex subsets for the affinized Bruhat order. We start by showing that, for any $\lambda \in Y^+$, the subset $\pi^\lambda W = (\text{proj}^{Y^+})^{-1}(\lambda)$ of $(\text{proj}^{Y^{++}})^{-1}(\lambda^{++})$ is also convex for the affinized Bruhat order.

Lemma 2.1. *Let $\lambda \in Y^+$, recall Notation 1.3 for v^λ. Then* $\mathrm{Inv}((v^\lambda)^{-1}) \cap \Phi_\lambda = \varnothing$. *In particular for any $\beta \in \Phi_+$,*

$$\mathrm{ht}(\lambda) < \mathrm{ht}(s_\beta \lambda) \iff \langle \lambda, \beta \rangle < 0 \iff s_\beta v^\lambda < v^\lambda. \tag{2.1}$$

Proof. Let $\lambda \in Y^+$ and $\alpha \in \mathrm{Inv}((v^\lambda)^{-1}) \cap \Phi_\lambda$. Then since $\alpha \in \Phi_\lambda$, s_α fixes λ, that is, $s_\alpha \in W_\lambda$. Moreover $(v^\lambda)^{-1}(\alpha) < 0$ so $s_\alpha v^\lambda < v^\lambda$, which contradicts the minimality of v^λ (note that, as $W_\lambda v^\lambda = v^\lambda W_{\lambda^{++}}$, v^λ is also the minimal representative for the right coset $W_\lambda v^\lambda$). Hence $\mathrm{Inv}((v^\lambda)^{-1}) \cap \Phi_\lambda = \varnothing$, and therefore any $\beta \in \mathrm{Inv}((v^\lambda)^{-1})$ satisfies $\langle \lambda, \beta \rangle \neq 0$.

For $\beta \in \Phi_+$, $\langle \lambda, \beta \rangle = \langle \lambda^{++}, (v^\lambda)^{-1}(\beta) \rangle$. Since λ^{++} is dominant, if this is negative then $\beta \in \mathrm{Inv}((v^\lambda)^{-1})$, and since $\mathrm{Inv}((v^\lambda)^{-1})$ and Φ_λ are disjoint, the converse is also true. Since $\beta \in \mathrm{Inv}((v^\lambda)^{-1}) \iff s_\beta v^\lambda < v^\lambda$ we deduce the second equivalence in (2.1). Moreover $\mathrm{ht}(s_\beta \lambda) = \mathrm{ht}(\lambda) - \langle \lambda, \beta \rangle\, \mathrm{ht}(\beta^\vee)$ by linearity of the height function, and since $\mathrm{ht}(\beta^\vee) > 0$, the first equivalence in (2.1) is clear. □

Remark 2.2. The fact that $\mathrm{Inv}((v^\lambda)^{-1}) \cap \Phi_\lambda = \varnothing$ is visible geometrically in the Coxeter complex of W, in which Φ_λ is the set of walls containing λ and $\mathrm{Inv}(v^{-1})$ is the set of walls separating C^v_f and C^v_v. The chamber $C^v_{v^\lambda}$ is the closest chamber from the fundamental chamber amongst the chambers containing λ in their closure, in other words, $v^\lambda = \mathrm{proj}_{W_\lambda}(1_W)$.

Proposition 2.3. *Suppose that $\pi^\lambda w \in W^a_+$ and $r \in W$ is a reflection such that $r\lambda \neq \lambda$. Then*

$$\pi^{r\lambda} r w > \pi^\lambda w \iff r v^\lambda < v^\lambda. \tag{2.2}$$

For any $\lambda^{++} \in Y^{++}$, the restriction of the function $\mathrm{ht} \circ \mathrm{proj}^{Y^+}$ *to* $(\mathrm{proj}^{Y^{++}})^{-1}(\lambda^{++})$ *is order-preserving.*

Proof. Suppose that $r \in W$ is a reflection which does not fix λ. By definition there exists a positive root $\beta \in \Phi_+$ such that $r = s_\beta$ and, since r does not fix λ, $\langle \lambda, \beta \rangle \neq 0$. Note that $\pi^{r\lambda} r w = s_{\beta[0]} \pi^\lambda w$ so, using (1.12), we have

$$\pi^{r\lambda} r w > \pi^\lambda w \iff -\langle \lambda, \beta \rangle > 0 \iff \langle \lambda, \beta \rangle < 0.$$

By Lemma 2.1 this is equivalent to $r v^\lambda < v^\lambda$, and to $\mathrm{ht}(\lambda) < \mathrm{ht}(r\lambda)$. This is enough to obtain (2.2). Moreover by convexity of $(\mathrm{proj}^{Y^{++}})^{-1}(\lambda^{++})$ (see Corollary 1.18) and by Proposition 1.15 it also implies that $\mathrm{ht} \circ \mathrm{proj}^{Y^+} : (\mathrm{proj}^{Y^{++}})^{-1}(\lambda^{++}) \to \mathbb{R}$ is order-preserving. □

Note that the function $\mathrm{ht} \circ \mathrm{proj}^{Y^+}$ is not order-preserving on the whole semigroup W^a_+. For example suppose that $\lambda \in Y^{++}$ and $\beta \in \Phi_+$ are such that $\lambda + \beta^\vee$ is also dominant. Then we can check that $\pi^{s_\beta(\lambda)} < \pi^{s_\beta(\lambda)-\beta^\vee} s_\beta$ whereas $\mathrm{ht}(s_\beta(\lambda)-\beta^\vee) < \mathrm{ht}(s_\beta(\lambda))$.

Proposition 2.3 implies the convexity of left W-cosets:

Corollary 2.4. *Let $\lambda \in Y^+$. Then the set $\pi^\lambda W = \{\boldsymbol{x} \in W^a_+ \mid \mathrm{proj}^{Y^+}(\boldsymbol{x}) = \lambda\}$ is convex for the affinized Bruhat order.*

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \pi^\lambda W$ such that $\boldsymbol{x} < \boldsymbol{y}$ and let $\boldsymbol{x} = \boldsymbol{x}_0 < \boldsymbol{x}_1 < \cdots < \boldsymbol{x}_n = \boldsymbol{y}$ be a chain from $\boldsymbol{x}$ to $\boldsymbol{y}$; in particular for all $k \in [\![0, n-1]\!]$, let $\beta_k[n_k] \in \Phi^a_+$ be such that $\boldsymbol{x}_{k+1} = s_{\beta_k[n_k]}\boldsymbol{x}_k$. For $k \in [\![0, n]\!]$, write $\boldsymbol{x}_k = \pi^{\lambda_k} w_k$ with $\lambda_k \in Y^+$, $w_k \in W$. By convexity of $(\mathrm{proj}^{Y^{++}})^{-1}(\lambda^{++})$, $\mathrm{proj}^{Y^{++}}$ is constant along the chain, therefore by Proposition 1.15, for all $k \in [\![0, n-1]\!]$ we have $\lambda_{k+1} \in \{\lambda_k, s_{\beta_k}(\lambda_k)\}$. From Proposition 2.3 we deduce that $v^{\lambda_{k+1}} \leq v^{\lambda_k}$. Since $\lambda_0 = \lambda_n = \lambda$, $v^{\lambda_0} = v^{\lambda_n} = v^\lambda$, and thus $v^{\lambda_k} = v^\lambda$, so $\lambda_k = \lambda$ for all $k \in [\![0, n]\!]$. Hence $\pi^\lambda W$ is convex. □

2.1. *Relative length on W.* We define a relative length and a relative Bruhat order on W, which naturally arises in the study of the affinized length ℓ^a on W^a_+. This connection was already observed by Muthiah and Orr [2018].

Definition 2.5. For any $v, w \in W$ let

$$\ell_v(w) = |\mathrm{Inv}(w^{-1}) \setminus \mathrm{Inv}(v^{-1})| - |\mathrm{Inv}(w^{-1}) \cap \mathrm{Inv}(v^{-1})|. \tag{2.3}$$

This is a signed version of the Bruhat length, in particular $\ell_1 = \ell$.

We associate an order to ℓ_v by setting, for any element $w \in W$ and any reflection $r \in W$, $w <_v wr$ if and only if $\ell_v(w) < \ell_v(wr)$, and then let $<_v$ be the order generated by these relations. It is strictly compatible with ℓ_v. In particular $<_1$ is the classical Bruhat order.

As does the Bruhat length, the lengths ℓ_v have a geometric interpretation in the Coxeter complex associated to (W, S). For M a wall of the Coxeter complex and $w \in W$, let $\varepsilon_w(M) = -1$ if M separates C^v_f and C^v_w, and $\varepsilon_w(M) = +1$ otherwise. Then

$$\ell_v(w) = \sum_{M \in \varepsilon_w^{-1}(-1)} \varepsilon_{v^{-1}}(M). \tag{2.4}$$

We will use this relative length to give an alternative definition of the affinized length. Let us first give an explicit formula for ℓ_v depending only on the classical length $\ell = \ell_1$.

Lemma 2.6. *If $sv > v$ with $v \in W$ and s a simple reflection then for any $w \in W$, $\ell_{sv}(w) = \ell_v(sw) - 1$.*

Proof. For any $w \in W$, the map $\gamma \mapsto s\gamma$ defines a bijection:

$$\mathrm{Inv}(w^{-1}) \setminus \{\alpha_s\} \cong \mathrm{Inv}(w^{-1}s) \setminus \{\alpha_s\}.$$

Moreover because $sv > v$, $\alpha_s \in \mathrm{Inv}(v^{-1}s)$ and $\alpha_s \notin \mathrm{Inv}(v^{-1})$.

Therefore

$$|\mathrm{Inv}(w^{-1}) \cap \mathrm{Inv}(v^{-1}s) \setminus \{\alpha_s\}| = |\mathrm{Inv}(w^{-1}s) \cap \mathrm{Inv}(v^{-1})|$$

and

$$|\mathrm{Inv}(w^{-1}) \setminus \mathrm{Inv}(v^{-1}s)| = \left|\mathrm{Inv}(w^{-1}s) \setminus (\mathrm{Inv}(v^{-1}) \cup \{\alpha_s\})\right|.$$

(1) If $\alpha_s \in \mathrm{Inv}(w^{-1})$ then $\alpha_s \notin \mathrm{Inv}(w^{-1}s)$ and

$$\begin{aligned}\ell_{sv}(w) &= |\mathrm{Inv}(w^{-1}s) \setminus \mathrm{Inv}(v^{-1})| - \big(|\mathrm{Inv}(w^{-1}s) \cap \mathrm{Inv}(v^{-1})| + 1\big)\\ &= \ell_v(sw) - 1.\end{aligned}$$

(2) If $\alpha_s \notin \mathrm{Inv}(w^{-1})$ then $\alpha_s \in \mathrm{Inv}(w^{-1}s)$ and

$$\begin{aligned}\ell_{sv}(w) &= \big(|\mathrm{Inv}(w^{-1}s) \setminus \mathrm{Inv}(v^{-1})| - 1\big) - |\mathrm{Inv}(w^{-1}s) \cap \mathrm{Inv}(v^{-1})|\\ &= \ell_v(sw) - 1.\end{aligned}$$

□

Proposition 2.7. *For all $v, w \in W$ the relative length $\ell_v(w)$ is given by*

$$\ell_v(w) = \ell(v^{-1}w) - \ell(v). \tag{2.5}$$

Proof. Since $\ell = \ell_1$, we take a reduced expression for v and apply Lemma 2.6 recursively to get the result. □

Corollary 2.8. *For any $v \in W$, the relative length ℓ_v is a grading of $(W, <_v)$.*

Proof. Let $v, w, w' \in W$. By Proposition 2.7, $\ell_v(w') - \ell_v(w) = \ell(v^{-1}w') - \ell(v^{-1}w)$ and w' covers w for $<_v$ if and only if $v^{-1}w'$ covers $v^{-1}w$ for the (standard) Bruhat order. Since the Bruhat length is a grading of $(W, <_1)$ (see [Björner and Brenti 2005, Theorem 2.2.6]), $v^{-1}w'$ covers $v^{-1}w$ if and only if $\ell(v^{-1}w') - \ell(v^{-1}w) = 1$ and $v^{-1}w' = v^{-1}wr$ for some reflection $r \in W$. Hence w' covers w for $<_v$ if and only if $\ell_v(w') - \ell_v(w) = 1$ and $w' = wr$ for some reflection $r \in W$: ℓ_v is a grading of $(W, <_v)$. □

The order $<_v$ also has a geometric interpretation which will be important later on; it is given by the following corollary.

Corollary 2.9. *For any root $\alpha \in \Phi$ and elements $w, v \in W$, we have that $w <_v s_\alpha w$ if and only if, in the Coxeter complex of W, C_w^v and C_v^v are on the same side of the wall M_α.*

Proof. We have $\ell_v(s_\alpha w)-\ell_v(w)=\ell(v^{-1}s_\alpha w)-\ell(v^{-1}w)=d^{\mathbb{N}}(v, s_\alpha w)-d^{\mathbb{N}}(v, w)$. By the definition of the Coxeter complex this is positive if and only if C^v_v and C^v_w are on the same side of the wall M_α. □

Therefore $<_v$ can be interpreted as a shift of the classical Bruhat order, corresponding geometrically to taking C^v_v as the fundamental chamber in the Coxeter complex.

2.2. *Relation with the affinized Bruhat length.* We relate the affinized Bruhat order and the relative order defined in Section 2.1. We start with an alternative expression for the affinized Bruhat length.

Proposition 2.10. *For any coweight $\lambda = v\lambda^{++} \in Y^+$, for any $w \in W$,*

$$\Big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle\lambda, \alpha\rangle \geq 0\}\Big| - \Big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle\lambda, \alpha\rangle < 0\}\Big| = \ell_{v^\lambda}(w). \tag{2.6}$$

Therefore

$$\ell^a_\varepsilon(\pi^\lambda w) = 2\,\mathrm{ht}(\lambda^{++}) + \varepsilon\ell_{v^\lambda}(w), \tag{2.7}$$

$$\ell^a(\pi^\lambda w) = 2\,\mathrm{ht}(\lambda^{++}) + \ell_{v^\lambda}(w). \tag{2.8}$$

Proof. For $\lambda \in Y^+$ and $v \in W$ such that $\lambda = v\lambda^{++}$, $\alpha \in \Phi_+$ satisfies $\langle\lambda, \alpha\rangle \geq 0$ if and only if $\alpha \in \Phi_\lambda \cup (\Phi_+ \setminus \mathrm{Inv}(v^{-1}))$. Hence,

$$\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle\lambda, \alpha\rangle \geq 0\} = (\mathrm{Inv}(w^{-1}) \setminus \mathrm{Inv}(v^{-1})) \bigsqcup (\mathrm{Inv}(w^{-1}) \cap \mathrm{Inv}(v^{-1}) \cap \Phi_\lambda)$$

and

$$\mathrm{Inv}(w^{-1}) \cap \mathrm{Inv}(v^{-1}) = \{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle\lambda, \alpha\rangle < 0\} \bigsqcup (\mathrm{Inv}(w^{-1}) \cap \mathrm{Inv}(v^{-1}) \cap \Phi_\lambda).$$

Therefore

$$\Big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle\lambda, \alpha\rangle \geq 0\}\Big| - \Big|\{\alpha \in \mathrm{Inv}(w^{-1}) \mid \langle\lambda, \alpha\rangle < 0\}\Big| \\ = \ell_v(w) + 2\,|\mathrm{Inv}(w^{-1}) \cap \mathrm{Inv}(v^{-1}) \cap \Phi_\lambda|$$

By Lemma 2.1 we deduce (2.6). □

Remark 2.11. By combining Corollary 1.13 with Proposition 2.10, we obtain the formulas already given by Muthiah and Orr [2019, Proposition 3.10].

Corollary 2.12. *Let $\lambda \in Y^+$ and $w \in W$. Suppose that $\pi^\mu w'' = s_{\beta[n]}\pi^\lambda w$ for some affinized root $\beta[n] \in \Phi^a$ such that $\mu^{++} = \lambda^{++}$. Then*

$$\pi^\mu w'' > \pi^\lambda w \iff \ell_{v^\mu}(w'') > \ell_{v^\lambda}(w). \tag{2.9}$$

For any $\lambda \in Y^+$ and $w, w'' \in W$,

$$\pi^\lambda w < \pi^\lambda w'' \iff w <_{v^\lambda} w''. \tag{2.10}$$

In particular, $\pi^\lambda v^\lambda$ is the minimal element of $\pi^\lambda W$.

Proof. Equivalence (2.9) is a direct consequence of (2.8) and strict compatibility of the affinized Bruhat length and the affinized Bruhat order. It implies by iteration that a chain for the relative order $<_{v^\lambda}$ from w to w'' lifts to a chain for the affinized Bruhat order from $\pi^\lambda w$ to $\pi^\lambda w''$. Conversely, by Corollary 2.4 proj^{Y^+} is constant along any chain from $\pi^\lambda w$ to $\pi^\lambda w''$, and therefore the projection on W of a chain from $\pi^\lambda w$ to $\pi^\lambda w''$ induces a chain for the relative Bruhat order $<_{v^\lambda}$ from w to w''. □

We deduce a partial version of Theorem A, for vectorial covers with constant coweight.

Theorem 2.13. *Let $\boldsymbol{x}, \boldsymbol{y} \in W^a_+$ be such that $\mathrm{proj}^{Y^+}(\boldsymbol{x}) = \mathrm{proj}^{Y^+}(\boldsymbol{y})$ and $\boldsymbol{x} \le \boldsymbol{y}$. Then*

$$\boldsymbol{x} \lhd \boldsymbol{y} \iff \ell^a(\boldsymbol{y}) = \ell^a(\boldsymbol{x}) + 1. \tag{2.11}$$

More precisely, if $\boldsymbol{x} = \pi^\lambda w$ then $\boldsymbol{y} = \pi^\lambda r w$ for some reflection $r \in W$ such that rw covers w for the relative Bruhat order $<_{v^\lambda}$.

Proof. By (2.10), $\pi^\lambda w'$ covers $\pi^\lambda w$ if and only if w' covers w for the relative Bruhat order $<_{v^\lambda}$. By Corollary 2.8, this is equivalent to $\ell_{v^\lambda}(w') = \ell_{v^\lambda}(w) + 1$. Therefore by (2.8) we deduce that $\boldsymbol{x} \lhd \boldsymbol{y} \implies \ell^a(\boldsymbol{y}) = \ell^a(\boldsymbol{x}) + 1$. The converse is immediate by strict compatibility of the affinized Bruhat length (Theorem 1.7). □

2.3. ***Vectorial covers with nonconstant coweight.*** Here, we prove Theorem A for vectorial covers with nonconstant coweight.

Beforehand, we need a few results on parabolic decomposition. The first lemma is an adaptation of a standard result on minimal coset representatives (see [Björner and Brenti 2005, Theorem 2.5.5]), and the second is proved by P-E. Chaput, L. Fresse and T. Gobet in [Chaput et al. 2021].

Lemma 2.14. *Let J be a subset of S, and recall Notation 1.3 for W^J. Let v be an element of W^J and u be any element of W such that $u < v$. Then, there is, for the Bruhat order, a saturated chain*

$$u = u_0 \lhd u_1 \lhd \cdots \lhd u_N = v \tag{2.12}$$

such that, for any $i \in [\![1, N]\!]$, $u_{i-1}^{-1} u_i$ does not belong to W_J.

Proof. If v covers u, it is clear since $u < v$ is a saturated chain, and as v is a minimal coset representative, $u^{-1}v \notin W_J$. By induction it thus suffices, for a general pair (u, v), to construct $u_1 \in W$ such that u_1 covers u, $u^{-1}u_1 \notin W_J$ and $u_1 < v$; the rest of the chain is obtained by induction. Let $s_1 \dots s_n$ be a reduced expression of v. Since $u < v$, there exists a reduced expression of u obtained from $s_1 \dots s_n$ by deleting letters $s_{i_1}, \dots, s_{i_N}$. Choose one such that i_N is minimal. Then let $t \in W$ be the reflection defined by $t = s_n \dots s_{i_N+1} s_{i_N} s_{i_N+1} \dots s_n$. We show that $u_1 = ut$ satisfies the desired properties.

(1) By construction, an expression of ut is obtained from $s_1 \dots s_n$ by deleting the $N-1$ letters $s_{i_1}, \dots, s_{i_{N-1}}$. Therefore $ut < v$.

(2) Since an expression of vt is obtained from $s_1 \dots s_n$ by deleting s_{i_N}, $vt < v$, and since v is the minimal coset representative of vW_J, t does not belong to W_J.

(3) It remains to show that ut covers u. By the first point, we have that $\ell(ut) \leq \ell(u) + 1$, so it suffices to show that $ut \not< u$. Suppose by contradiction that $ut < u$. Then, by the strong exchange property, an expression of ut is obtained from u by deleting one letter s_p of the reduced expression $s_1 \dots \check{s}_{i_1} \dots \check{s}_{i_N} \dots s_n$ (where $\check{s}_i$ denotes a letter s_i taken away from the expression $s_1 \dots s_n$).

(a) Suppose that $p > i_N$. Then t can also be written as $s_n \dots s_{p+1} s_p s_{p+1} \dots s_n$, and $v = (vt)t = s_1 \dots \check{s}_{i_N} \dots \check{s}_p \dots s_n$, which contradicts the hypothesis that $s_1 \dots s_n$ is reduced.

(b) Suppose now that there is $d \leq N-1$ such that $i_d < p < i_{d+1}$ (with the convention that $i_0 = -1$). Then $t = s_n \dots \check{s}_{i_N} \dots \check{s}_{i_{d+1}} \dots s_p \dots \check{s}_{i_{d+1}} \dots \check{s}_{i_N} \dots s_n$, and $u = (ut)t$ can be written from $s_1 \dots s_n$ by deleting the terms of indices $i_1, \dots, i_{N-1}$ and $p < i_N$, but not i_N. This contradicts the minimality of i_N. □

Definition 2.15. For $v, w \in W$, we write

$$v \leq_R w \iff \ell(w) = \ell(v) + \ell(wv^{-1}). \tag{2.13}$$

Remark 2.16. The relation $\leq_R$ is called the weak Bruhat order and it is related to minimal galleries: $v \leq_R w$ if and only if there is a minimal gallery from 1 to w^{-1} going through v^{-1}.

Recall that for $J \subset S$ and $x \in W$, (x^J, x_J) denotes the unique pair of $W^J \times W_J$ such that $x = x^J.x_J$.

Lemma 2.17 [Chaput et al. 2021, Lemma 8.11]. *Let $J \subset S$ be a subset of simple reflections. Let u be an element of W and t be a reflection of $W \setminus W_J$ such that ut covers u. Then $(ut)_J \leq_R u_J$. In other words $((ut)_J)^{-1}$ lies on a minimal gallery from 1 to $(u_J)^{-1}$.*

Theorem 2.18. *Let $\boldsymbol{x}, \boldsymbol{y} \in W^a_+$ be such that $\mathrm{proj}^{Y^{++}}(\boldsymbol{y}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{x})$ and $\boldsymbol{x} \leq \boldsymbol{y}$. Then*

$$\boldsymbol{x} \lhd \boldsymbol{y} \iff \ell^a(\boldsymbol{y}) = \ell^a(\boldsymbol{x}) + 1. \tag{2.14}$$

More precisely, write $\boldsymbol{x} = \pi^\lambda w$. *Let* J *be the set of simple reflections stabilizing* λ^{++} *and let* $v \in W^J$ *be such that* $\lambda = v\lambda^{++}$ (*so* $v = v^\lambda$ *with Notation 1.3*). *Then, if* $\boldsymbol{x} \lhd \boldsymbol{y}$ *and* $\mathrm{proj}^{Y^+}(\boldsymbol{y}) \neq \mathrm{proj}^{Y^+}(\boldsymbol{x})$, *there exists a unique reflection* $r \in W$ *such that*:

(1) *The reflection* r *does not stabilize* λ *and* $\boldsymbol{y} = \pi^{r\lambda} rw$.

(2) *For the Bruhat order on* W, v *covers* rv.

(3) *Set* $u = rv$, *so* $u_J \in W_J$ *denotes the element* $(rv)_J$ *and* $rvu_J^{-1} \in W^J$. *Then* vu_J^{-1} *is on a minimal gallery from* v *to* w *in* W.

Proof. If $\mathrm{proj}^{Y^+}(\boldsymbol{y}) = \mathrm{proj}^{Y^+}(\boldsymbol{x})$ then (2.14) is given by Theorem 2.13. Moreover if $\ell^a(\boldsymbol{y}) = \ell^a(\boldsymbol{x}) + 1$ and $\boldsymbol{y} \geq \boldsymbol{x}$ then, by strict compatibility of ℓ^a, $\boldsymbol{y}$ covers $\boldsymbol{x}$. We are thus reduced to prove that, if $\boldsymbol{y}$ covers $\boldsymbol{x}$ with $\mathrm{proj}^{Y^+}(\boldsymbol{y}) \neq \mathrm{proj}^{Y^+}(\boldsymbol{x})$ and $\mathrm{proj}^{Y^{++}}(\boldsymbol{y}) = \mathrm{proj}^{Y^{++}}(\boldsymbol{x})$, then $\ell^a(\boldsymbol{y}) = \ell^a(\boldsymbol{x}) + 1$.

Write $\boldsymbol{x} = \pi^\lambda w$ and $v = v^\lambda$. By definition of the affinized Bruhat order, if $\boldsymbol{x} \lhd \boldsymbol{y}$ then $\boldsymbol{y}$ is of the form $s_{\beta[n]}\boldsymbol{x}$ for some $\beta[n] \in \Phi_+^a$.

Let $\boldsymbol{y} = s_{\beta[n]}\boldsymbol{x} \in W_+^a$ with $\mathrm{proj}^{Y^{++}}(\boldsymbol{y}) = \lambda^{++}$ and $\mathrm{proj}^{Y^+}(\boldsymbol{y}) \neq \lambda$, in particular $n \neq \langle \lambda, \beta \rangle$. By Proposition 1.15, $n = 0$ so $\boldsymbol{y} = \pi^{r\lambda} rw$ for the reflection $r = s_\beta$ which does not stabilize λ. Let us write $u = rv$, and note that $u^J = v^{r\lambda}$. By (2.8),

$$\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x}) = \ell_{u^J}(rw) - \ell_v(w). \tag{2.15}$$

By definition, $rv = u^J u_J$ with, by (1.3), $\ell(rv) = \ell(u_J) + \ell(u^J)$. We compute

$$\begin{aligned} \ell_{u^J}(rw) - \ell_v(w) &= \ell((rv(u_J)^{-1})^{-1} rw) - \ell(v^{-1}w) + \ell(v) - \ell(u^J) \\ &= \ell(u_J v^{-1} w) - \ell(v^{-1}w) + \ell(v) + \ell(u_J) - \ell(u) \\ &= (\ell(v) - \ell(u)) + \big(\ell(u_J) - (d^{\mathbb{N}}(v, w) - d^{\mathbb{N}}(vu_J^{-1}, w))\big). \end{aligned} \tag{2.16}$$

From (2.15) and (2.16), we deduce

$$\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x}) = (\ell(v) - \ell(u)) + \big(\ell(u_J) - (d^{\mathbb{N}}(v, w) - d^{\mathbb{N}}(vu_J^{-1}, w))\big). \tag{2.17}$$

In (2.17), by the triangle inequality and since $d^{\mathbb{N}}(v, vu_J^{-1}) = \ell(u_J)$, the second term $\ell(u_J) - (d^{\mathbb{N}}(v, w) - d^{\mathbb{N}}(vu_J^{-1}, w))$ is nonnegative, and it is equal to 0 if and only if $d^{\mathbb{N}}(v, w) = d^{\mathbb{N}}(v, vu_J^{-1}) + d^{\mathbb{N}}(vu_J^{-1}, w)$, so if and only if vu_J^{-1} is on a minimal gallery from v to w.

Recall Definition 1.21 of $\mathrm{proj}_{vW_J}(w)$. Since vu_J^{-1} lies in vW_J, a minimal gallery from vu_J^{-1} to w goes through $\mathrm{proj}_{vW_J}(w)$. Thus $\ell(u_J) - (d^{\mathbb{N}}(v, w) - d^{\mathbb{N}}(vu_J^{-1}, w))$ is equal to zero if and only if u_J^{-1} is on a minimal gallery from 1 to $v^{-1}\mathrm{proj}_{vW_J}(w)$ in W_J.

Let us first suppose that u_J^{-1} is not on a minimal gallery from 1 to $v^{-1}\mathrm{proj}_{vW_J}(w)$. We want to deduce that $\boldsymbol{y}$ does not cover $\boldsymbol{x}$. We thus want to produce a nontrivial chain from $\pi^\lambda w$ to $\pi^{r\lambda} rw$. By Lemma 1.20, there is a reflection $t \in W_J$ such that

$$d^{W_J}(1, tv^{-1}\mathrm{proj}_{vW_J}(w)) > d^{W_J}(1, v^{-1}\mathrm{proj}_{vW_J}(w)),$$
$$d^{W_J}(u_J^{-1}, tv^{-1}\mathrm{proj}_{vW_J}(w)) < d^{W_J}(u_J^{-1}, v^{-1}\mathrm{proj}_{vW_J}(w)).$$

In W, this implies $d(vt, w) > d(v, w)$ and $d(vtu_J^{-1}, w) < d(vu_J^{-1}, w)$.

Let $\widetilde{w} = vtv^{-1}w$. We compute

$$\ell_v(\widetilde{w}) - \ell_v(w) = d^{\mathbb{N}}(vt, w) - d^{\mathbb{N}}(v, w) > 0,$$
$$\ell_{u^J}(rw) - \ell_{u^J}(r\widetilde{w}) = d^{\mathbb{N}}(vu_J^{-1}, w) - d^{\mathbb{N}}(vtu_J^{-1}, w) > 0,$$
$$\ell_{u^J}(r\widetilde{w}) - \ell_v(\widetilde{w}) = \ell(v) - \ell(u) + \ell(u_J) - (d^{\mathbb{N}}(v, \widetilde{w}) - d^{\mathbb{N}}(vu, \widetilde{w})) > 0.$$

Hence by Proposition 2.3 and Corollary 2.12,

$$\pi^\lambda w < \pi^\lambda \widetilde{w} < \pi^{r\lambda} r\widetilde{w} < \pi^{r\lambda} rw. \tag{2.18}$$

Suppose now that vu_J^{-1} is on a minimal gallery from v to w. Then by (2.17), $\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x}) = \ell(v) - \ell(rv)$. Suppose that $\ell(v) - \ell(rv) = N > 1$. Let

$$rv = u_0 \lhd u_1 \lhd \cdots \lhd u_N = v \tag{2.19}$$

be a saturated chain in W from rv to v given by Lemma 2.14 and, for $i \in [\![1, N]\!]$, let $\beta_i \in \Phi_+$ be such that $u_i = s_{\beta_i} u_{i-1}$, so $u_i = s_{\beta_i} \ldots s_{\beta_1} u \geq rv$. Note in particular that

$$\ell(u_i) = \ell(rv) + i = \ell(v) - N + i. \tag{2.20}$$

Let us show that it induces a chain for the affinized Bruhat order

$$\pi^\lambda w = s_{\beta_N[0]} \ldots s_{\beta_1[0]} \pi^{r\lambda} rw < s_{\beta_{N-1}[0]} \ldots s_{\beta_1[0]} \pi^{r\lambda} rw < \cdots < \pi^{r\lambda} rw. \tag{2.21}$$

Since $s_{\beta_i[0]} \ldots s_{\beta_1[0]} \pi^{r\lambda} rw = \pi^{u_i \lambda^{++}} s_{\beta_i} \ldots s_{\beta_1} rw$, by (2.9) it is enough to verify

$$\forall i \in [\![0, n]\!],\ \ell_{u_i^J}(s_{\beta_i} \ldots s_{\beta_1} rw) = \ell_v(w) + N - i. \tag{2.22}$$

We compute

$$\begin{aligned} \ell_{u_i^J}(s_{\beta_i} \ldots s_{\beta_1} rw) &= \ell\big((u_i (u_i)_J^{-1})^{-1} s_{\beta_i} \ldots s_{\beta_1} rw\big) - \ell(u_i^J) \\ &= \ell((u_i)_J v^{-1} w) - \ell(u_i^J). \end{aligned} \tag{2.23}$$

Since the saturated chain $u_0 < u_1 < \cdots < u_N$ is obtained from Lemma 2.14, u_i covers u_{i-1} such that the reflection $u_{i-1}^{-1} u_i$ does not belong to W_J, so by Lemma 2.17, $(u_i)_J \leq_R (u_{i-1})_J$ and by iteration we have $(u_i)_J \leq_R (u_0)_J = u_J$. Otherwise said, $(u_i)_J^{-1}$ is on a minimal gallery from 1 to u_J^{-1}. Therefore $v(u_i)_J^{-1}$ is on a minimal gallery from v to vu_J^{-1}, and hence on a minimal gallery from v to w. We deduce

$$\ell((u_i)_J v^{-1} w) = \ell(v^{-1} w) - \ell((u_i)_J). \tag{2.24}$$

Combining (2.23) and (2.24) we obtain

(2.25) $$\ell_{u_i^J}(s_{\beta_i}\dots s_{\beta_1} r w) = \ell(v^{-1}w) - \ell((u_i)_J) - \ell(u_i^J).$$

Moreover,

(2.26) $$\ell((u_i)_J) + \ell(u_i^J) = \ell(u_i) = \ell(v) - (N - i)$$

by (1.3) and (2.20). Combining (2.25) and (2.26), we deduce (2.22). □

3. Properly affine covers

3.1. ***A few properties of properly affine covers.*** We now turn to the case of covers $\pi^\lambda w < \pi^\mu w'$ in W_+^a with $\mu^{++} \neq \lambda^{++}$. Such covers are called properly affine covers.

By (1.12), if $\pi^\mu s_\beta w = s_{\beta[n]}\pi^\lambda w > \pi^\lambda w$ with $\beta[n] \in \Phi_+^a$, then $n \in \mathbb{Z} \setminus]\!]0, \langle\lambda, \beta\rangle[\![$. Conversely, if $n \in \mathbb{Z} \setminus [\![0, \langle\lambda, \beta\rangle]\!]$ then $s_{\beta[n]}\pi^\lambda w > \pi^\lambda w$, however $s_{\beta[n]}\pi^\lambda w$ may not be in W_+^a as $\lambda + \mathbb{Z}\beta^\vee \not\subset Y^+$. The limit cases $n \in \{0, \langle\lambda, \beta\rangle\}$ correspond to $\lambda^{++} = \mu^{++}$ dealt with in the previous section.

We first show that properly affine covers occur only for minimal n, in the following sense.

Proposition 3.1. *Let $\lambda \in Y^+$ and $w \in W$, and let $\beta \in \Phi$ and $n \in \mathbb{Z}$. Let us define $\sigma = \mathrm{sgn}(\langle\lambda, \beta\rangle) \in \{1, -1\}$. If $\pi^\mu w' = s_{\beta[n]}\pi^\lambda w \triangleright \pi^\lambda w$ is a cover with $\lambda^{++} \neq \mu^{++}$, then $n \in \{-\sigma, \langle\lambda, \beta\rangle + \sigma\}$.*

Proof. For any $\nu \in Y^+$ if we identify the Coxeter complex of W with the positive Tits cone $\mathcal{T} \subset \mathbb{A}$, $C^v_{v^\nu}$ is the closest vectorial chamber, from the fundamental chamber, containing ν in its closure. All the elements of $\lambda + \sigma\mathbb{Z}_{>0}\beta^\vee$ are on the same side of M_β; hence by Corollary 2.9, for any two such $\nu, \nu' \in \lambda + \sigma\mathbb{Z}_{>0}\beta^\vee$ and any $w \in W$,

(3.1) $$w <_{v^\nu} s_\beta w \iff w <_{v^{\nu'}} s_\beta w.$$

Suppose first that $n \in \langle\lambda, \beta\rangle + \sigma\mathbb{Z}_{>1}$ and let $\mu = \lambda + (n - \langle\lambda, \beta\rangle)\beta^\vee$. Then:

(1) If $w <_{v^\mu} s_\beta w$, we have the chain

(3.2) $$\pi^\lambda w < s_{\beta[\langle\lambda,\beta\rangle+\sigma]}\pi^\lambda w = \pi^{\lambda+\sigma\beta^\vee} s_\beta w < \pi^\mu w < \pi^\mu s_\beta w.$$

The second inequality comes from $\pi^\mu w = s_{\beta[n+\sigma]}\pi^{\lambda+\sigma\beta^\vee} s_\beta w$ and (1.12), and the third comes from (2.10).

(2) Else $s_\beta w <_{v^\mu} w$, so by (3.1), $s_\beta w <_{v^{\lambda+\sigma\beta^\vee}} w$ and we have the chain

(3.3) $$\pi^\lambda w < s_{\beta[\langle\lambda,\beta\rangle+\sigma]}\pi^\lambda w = \pi^{\lambda+\sigma\beta^\vee} s_\beta w < \pi^{\lambda+\sigma\beta^\vee} w < \pi^\mu s_\beta w.$$

Here the second inequality comes from (2.10). The third comes from $\pi^\mu s_\beta w = s_{\beta[n+\sigma]}\pi^{\lambda+\sigma\beta^\vee} w$ and (1.12).

Either way, for $n \in \langle\lambda, \beta\rangle + \sigma\mathbb{Z}_{>1}$, $s_{\beta[n]}\pi^\lambda w$ does not cover $\pi^\lambda w$. For $n \in -\sigma\mathbb{Z}_{>1}$ the argument is similar, because all the elements of $s_\beta\lambda - \sigma\mathbb{Z}_{>0}\beta^\vee$ are on the same side of M_β, in particular on the side of $\mu = s_\beta\lambda + n\beta^\vee$.

(1) If $w <_{v^\mu} s_\beta w$, we have a chain

$$\pi^\lambda w < s_{\beta[-\sigma]}\pi^\lambda w = \pi^{s_\beta(\lambda+\sigma\beta^\vee)} s_\beta w < \pi^\mu w < \pi^\mu s_\beta w. \tag{3.4}$$

(2) Else $s_\beta w <_{v^\mu} w$ so $s_\beta w <_{v^{s_\beta(\lambda+\sigma\beta^\vee)}} w$ and we have a chain

$$\pi^\lambda w < s_{\beta[-\sigma]}\pi^\lambda w = \pi^{s_\beta(\lambda+\sigma\beta^\vee)} s_\beta w < \pi^{s_\beta(\lambda+\sigma\beta^\vee)} w < \pi^\mu s_\beta w. \tag{3.5}$$

So the only possible covers (with varying coweights) are for $n \in \{-\sigma, \langle\lambda, \beta\rangle + \sigma\}$. ☐

Remark 3.2. To follow up on Remark 1.16, by (1.10), we have $s_{\beta[\sigma+\langle\lambda,\beta\rangle]}\pi^\lambda w = \pi^\lambda w s_{w^{-1}(\beta)[\sigma]}$, where $\sigma = \operatorname{sgn}(\langle\lambda, \beta\rangle)$. Therefore Proposition 3.1 tells us that, if $\boldsymbol{y}$ covers $\boldsymbol{x}$ in W^a_+, then $\boldsymbol{y}$ is obtained from $\boldsymbol{x}$ applying an affinized reflection $s_{\tilde{\beta}[n]}$ either on the left (for $s_{\beta[0]}$ and $s_{\beta[-\sigma]}$) or on the right (for $s_{\beta[\langle\lambda,\beta\rangle]}$ and $s_{\beta[\langle\lambda,\beta\rangle+\sigma]}$), with $n \in \{-1, 0, 1\}$.

This is still far from a sufficient condition and many cases of potential covers can still be eliminated. We give another necessary condition for $\pi^\mu s_\beta w = s_{\beta[n]}\pi^\lambda w > \pi^\lambda w$ to be a cover; this is a generalization of the chains produced in the proof of Theorem 2.18:

Proposition 3.3. *Let $\pi^\mu s_\beta w = s_{\beta[n]}\pi^\lambda w > \pi^\lambda w$ with $\mu^{++} \neq \lambda^{++}$. Suppose that $s_\beta v^\mu$ is not on a minimal gallery from w to v^λ. Then $\pi^\mu s_\beta w > \pi^\lambda w$ is not a cover.*

Proof. We express the difference of ε-length using (2.7):

$$\ell^a_\varepsilon(\pi^\mu s_\beta w) - \ell^a_\varepsilon(\pi^\lambda w) = 2\operatorname{ht}(\mu^{++} - \lambda^{++}) + \varepsilon(\ell_{v^\mu}(s_\beta w) - \ell_{v^\lambda}(w)). \tag{3.6}$$

If there exists a reflection $r \in W$ such that $\ell_{v^\lambda}(rw) > \ell_{v^\lambda}(w)$ and $\ell_{v^\mu}(s_\beta rw) < \ell_{v^\mu}(s_\beta w)$ then using (2.7) to compute the length ℓ^a_ε, we have a chain

$$\pi^\lambda w < \pi^\lambda rw < \pi^\mu s_\beta rw < \pi^\mu s_\beta w. \tag{3.7}$$

Since $\ell_v(rw) - \ell_v(w) = \ell(v^{-1}rw) - \ell(v^{-1}w)$ for $v, r, w \in W$, Lemma 1.20 guarantees the existence of r, which proves the proposition. ☐

In Figure 3 below, we give an example of a chain constructed this way in the A_1-affine case, with Cartan matrix $\left(\begin{smallmatrix} 2 & -2 \\ -2 & 2 \end{smallmatrix}\right)$.

In this example, α and β are the simple roots of an A_1-affinized root system, and we have chosen λ,w and $\beta[n]$ such that $v^\lambda = s_\alpha$, $v^\mu = s_\alpha s_\beta$ and $w = s_\beta$. $\pi^\lambda w$ corresponds to the alcove C_1 in light blue, and its image $\pi^\mu s_\beta w$ by $s_{\beta[6]}$ corresponds to C_4. Since $r = s_\beta s_\alpha s_\beta$ satisfies $d(v^\lambda, rw) = s_\alpha s_\beta s_\alpha > d(v^\lambda, w) = s_\alpha s_\beta$, and

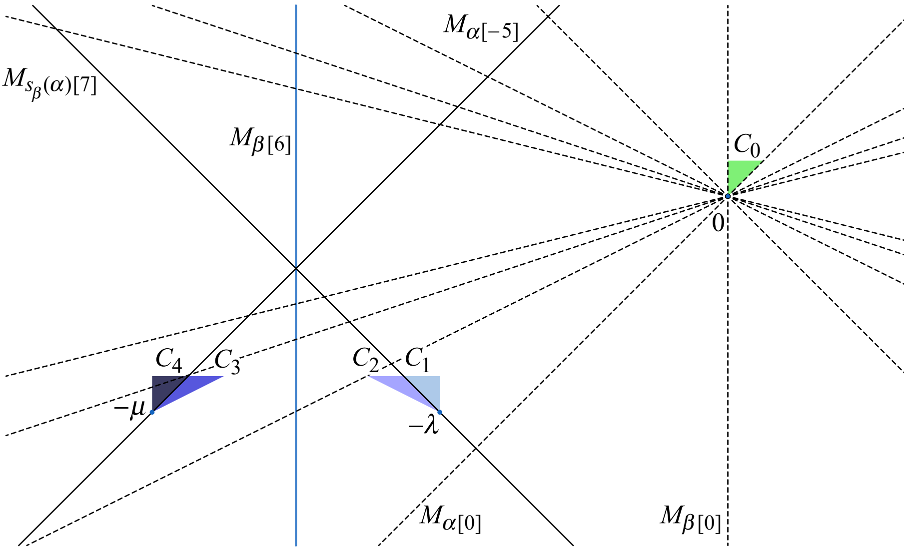

Figure 3. Example of a chain constructed as in Proposition 3.3.

$d(s_\beta v^\mu, rw) = s_\beta < s_\beta s_\alpha = d(s_\beta v^\mu, w)$, there is a chain $\pi^\lambda w < \pi^\lambda rw < \pi^\mu s_\beta rw < \pi^\mu s_\beta w$ which corresponds to the sequence of alcoves (C_1, C_2, C_3, C_4) in Figure 3.

Remark 3.4. Let $v_0, w_0 \in W$, $\mu_0 \in Y$ and $\alpha_0 \in \Phi$. To produce chains, note that (1.12) applied with the affinized reflection $s_{v_0(\alpha_0)[m+\langle\mu_0,\alpha_0\rangle]}$ to $\pi^{v_0(\mu_0)}w_0$ gives

$$\forall m \in \mathbb{Z} \setminus [\![-\langle\mu_0, \alpha_0\rangle, 0]\!],\ \pi^{v_0(\mu_0+m\alpha_0^\vee)} s_{v_0(\alpha_0)} w_0 > \pi^{v_0(\mu_0)} w_0. \tag{3.8}$$

Applying the affinized reflection $s_{v_0(\alpha_0)[-m]}$ to $\pi^{v_0(\mu_0)}w_0$ instead we obtain

$$\forall m \in \mathbb{Z} \setminus [\![-\langle\mu_0, \alpha_0\rangle, 0]\!],\ \pi^{v_0 s_{\alpha_0}(\mu_0+m\alpha_0^\vee)} s_{v_0(\alpha_0)} w_0 > \pi^{v_0(\mu_0)} w_0. \tag{3.9}$$

For $m \in]\!]-\langle\mu_0, \alpha_0\rangle, 0[\![$ the inequalities are reversed. The cases $m \in \{-\langle\mu_0, \alpha_0\rangle, 0\}$ need to be treated more carefully since they depend on the sign of the root $v_0(\alpha_0)$ (because (1.12) holds for the affinized reflection $s_{v_0(\alpha_0)[0]}$ only if $v_0(\alpha_0) \in \Phi_+$), on the sign of $\langle\mu_0, \alpha_0\rangle$ and on the vectorial element w_0.

3.2. ***Another expression for the affinized length difference.*** Outside of the case of vectorial covers dealt with in Theorems 2.13 and 2.18, if we write $\boldsymbol{x} = \pi^{v\lambda} w$ with $\lambda \in Y^{++}$, $v, w \in W$ with v of minimal length in vW_λ, by Proposition 3.1 the only covers are of the form $\boldsymbol{y} \in \{\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w, \pi^{v s_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w\}$ for some $\beta \in \Phi_+$, so the rest of this paper is dedicated to covers of this sort.

Notation 3.5. From now on, unless stated otherwise, we use the following notation:

(1) $\lambda \in Y^{++}$ is a dominant coweight.

(2) $\beta \in \Phi_+$ is a positive root.

(3) $v \in W^\lambda$ is the minimal representative of a W_λ-coset.

(4) $w \in W$ is any element of W.

(5) $\boldsymbol{x} = \pi^{v(\lambda)} w$ and $\boldsymbol{y} \in \{\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w,\ \pi^{v s_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w\}$ are elements of W^+.

The choice to denote by λ a dominant coweight is made in order to avoid the heavier notation λ^{++}. Recall from Notation 1.3 that W^λ is the set of minimal coset representatives of W/W_λ, where W_λ is the standard parabolic subgroup $\{w \in W \mid w(\lambda) = \lambda\}$.

In this subsection, we give another expression for $\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x})$.

The next two lemmas give information on the vectorial chamber of $v(\lambda + \beta^\vee)$.

Lemma 3.6. *Let $\lambda \in Y^{++}$ be a dominant coweight and let $\beta \in \Phi_+$ be a positive root such that $\lambda + \beta^\vee \in Y^+$. Let $u \in W$ be such that $\lambda + \beta^\vee$ belongs to the vectorial chamber C^v_u, that is to say $u^{-1}(\lambda + \beta^\vee) \in Y^{++}$. Then*

$$\ell(s_\beta u) = \ell(s_\beta) + \ell(u). \tag{3.10}$$

Proof. Let $s_{\tau_1} \dots s_{\tau_n}$ be a reduced expression of u, so that $\ell(u) = n$ and

$$\mathrm{Inv}(u^{-1}) = \{\tau_1, s_{\tau_1}(\tau_2), \dots, s_{\tau_1} \dots s_{\tau_{n-1}}(\tau_n)\}.$$

We show that $s_{\tau_{k+1}} \dots s_{\tau_1} s_\beta > s_{\tau_k} \dots s_{\tau_1} s_\beta$ for all $k \in [\![0, n-1]\!]$.

For any $\alpha \in \mathrm{Inv}(u^{-1})$ we have $\langle \lambda + \beta^\vee, \alpha \rangle \le 0$ (because $\lambda + \beta^\vee \in C^v_u$). Since λ is dominant this implies $\langle \beta^\vee, \alpha \rangle \le 0$.

Let $k \in [\![0, n-1]\!]$. Then $\langle \beta^\vee, s_{\tau_1} \dots s_{\tau_k}(\tau_{k+1}) \rangle \le 0$ so

$$s_\beta(s_{\tau_1} \dots s_{\tau_k}(\tau_{k+1})) = s_{\tau_1} \dots s_{\tau_k}(\tau_{k+1}) - \langle \beta^\vee, s_{\tau_1} \dots s_{\tau_k}(\tau_{k+1}) \rangle \beta$$

is a positive root as a sum of positive roots. Thus $s_{\tau_{k+1}} \dots s_{\tau_1} s_\beta > s_{\tau_k} \dots s_{\tau_1} s_\beta$ for any $k \in [\![0, n-1]\!]$ and therefore $\ell(s_\beta u) = \ell(u^{-1} s_\beta) = n + \ell(s_\beta) = \ell(s_\beta) + \ell(u)$. □

Lemma 3.7. *Let $\lambda \in Y^{++}$ be a dominant coweight and let $\beta \in \Phi_+$ be a positive root such that $\lambda + \beta^\vee \in Y^+$. Let $v \in W^\lambda$, $w \in W$ and let u denote the element $v^{\lambda+\beta^\vee}$.*

Then, if $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ (resp. $\pi^{v s_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$) covers $\boldsymbol{x} = \pi^{v(\lambda)} w$,

$$\ell(vu) = \ell(v) + \ell(u) \quad (\textit{resp. } \ell(v s_\beta u) = \ell(v) + \ell(s_\beta u) \textit{ and } \ell(v s_\beta) = \ell(v) + \ell(s_\beta)).$$

Proof. To simplify the notation, write W_J for $W_{(\lambda+\beta^\vee)^{++}}$. Note that, with the notation of Definition 1.21, $vu = \mathrm{proj}_{vuW_J}(v)$ since u is the element of minimal length in uW_J.

Suppose by contradiction that $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ covers $\boldsymbol{x}$ with $\ell(vu) < \ell(v) + \ell(u)$. Then $d^{\mathbb{N}}(1, vu) = \ell(vu) < d^{\mathbb{N}}(1, v) + d^{\mathbb{N}}(v, vu) = \ell(v) + \ell(u)$, so v is not on a minimal gallery from 1 to vu. Therefore by Lemma 1.20, there is a reflection $r \in W$ such that $d(1, rvu) > d(1, vu)$ and $d(1, rv) < d(1, v)$, that is to say $rv < v$ and $rvu > vu$.

By minimality of u, r is not in $vuW_J(vu)^{-1}$: otherwise $rvu \in vuW_J$ satisfies $d(v, rvu) = d(rv, vu) < d(v, vu)$, because foldings reduce the vectorial distance and v, vu are on different sides of the wall M_r associated to r.

Since vu is the projection of v on vuW_J which is convex (see [Ronan 1989, Lemma 2.10]), and since the wall M_r separates v and vu, any element of vuW_J is on the same side of the wall M_r as vu, so $rvu\tilde{u} > vu\tilde{u}$ for any $\tilde{u} \in W_J$. In particular, let $\tilde{u} \in W_J$ be such that $rvu\tilde{u}$ is the minimal coset representative of $rvuW_J$. Then by Proposition 2.3, since $rvu\tilde{u} > vu\tilde{u}$, we have

$$\begin{aligned}\pi^{rv(\lambda+\beta^\vee)} r s_{v(\beta)} w &= \pi^{rvu\tilde{u}((\lambda+\beta^\vee)^{++})} r s_{v(\beta)} w \\ &< \pi^{vu\tilde{u}((\lambda+\beta^\vee)^{++})} s_{v(\beta)} w = \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w.\end{aligned} \tag{3.11}$$

Therefore by Proposition 2.3 for the left- and right-hand side inequalities and (3.8) applied with $(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda, \beta, rv, rw, 1)$ for the middle one, we have a chain

$$\begin{aligned}\pi^{v(\lambda)} w < \pi^{rv(\lambda)} rw &< \pi^{rv(\lambda+\beta^\vee)} s_{rv(\beta)} rw \\ &= \pi^{rv(\lambda+\beta^\vee)} r s_{v(\beta)} w < \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w.\end{aligned} \tag{3.12}$$

Therefore if $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ covers $\boldsymbol{x}$ then $\ell(vu) = \ell(v) + \ell(u)$.

Now assume by contradiction that $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$ covers $\pi^\lambda w$ with $\ell(vs_\beta u) < \ell(v) + \ell(s_\beta u)$. Then, similarly there is a reflection $r \in W$ such that $rv < v$ and $rvs_\beta u\tilde{u} > vs_\beta u\tilde{u}$. By Proposition 2.3 for the left- and right-hand side inequalities and (3.9) applied with $(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda, \beta, rv, rw, 1)$ for the middle one, we have a chain

$$\begin{aligned}\pi^{v(\lambda)} w < \pi^{rv(\lambda)} rw &< \pi^{rvs_\beta(\lambda+\beta^\vee)} s_{rv(\beta)} rw \\ &= \pi^{rvs_\beta(\lambda+\beta^\vee)} r s_{v(\beta)} w < \pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w.\end{aligned} \tag{3.13}$$

We deduce that if $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$ covers $\boldsymbol{x}$ then

$$\ell(vs_\beta u) = \ell(v) + \ell(s_\beta u).$$

By Lemma 3.6 this is $\ell(v) + \ell(s_\beta) + \ell(u)$, by the triangle inequality we deduce that

$$\ell(v) + \ell(s_\beta) \geq \ell(vs_\beta) \geq \ell(vs_\beta u) - \ell(u) = \ell(v) + \ell(s_\beta)$$

and we obtain the second equality in this case. ☐

Proposition 3.8. *Let $\lambda \in Y^{++}$, $v \in W^{\lambda}$, $w \in W$. Let $\beta \in \Phi_+$ be a positive root such that $\lambda + \beta^\vee \in Y^+$ and let u denote $v^{\lambda+\beta^\vee} \in W^{(\lambda+\beta^\vee)^{++}}$.*

Suppose that $\boldsymbol{y} \in \{\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w, \pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w\}$ covers $\boldsymbol{x} = \pi^{v(\lambda)} w$. Then

$$\ell^a(\boldsymbol{y}) - \ell^a(\boldsymbol{x}) = (2\,\mathrm{ht}(\beta^\vee) - \ell(s_\beta)) - 2\Big(\ell(u) + \sum_{\tau \in \mathrm{Inv}(u^{-1})} \langle \lambda + \beta^\vee, \tau \rangle\Big). \tag{3.14}$$

Proof. Let W_J denote the standard parabolic subgroup $W_{(\lambda+\beta^\vee)^{++}}$. Recall that $u = v^{\lambda+\beta^\vee}$ is the minimal element of W such that $u((\lambda+\beta^\vee)^{++}) = \lambda + \beta^\vee$, so it is the minimal representative of the coset uW_J. By Proposition 2.10 we have

$$\begin{aligned} &\ell^a(\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w) - \ell^a(\pi^{v(\lambda)} w) \\ &\qquad = 2\,\mathrm{ht}((\lambda+\beta^\vee)^{++}) - 2\,\mathrm{ht}(\lambda) + \ell_{v^{v(\lambda+\beta^\vee)}}(s_{v(\beta)} w) - \ell_v(w), \end{aligned} \tag{3.15}$$

$$\begin{aligned} &\ell^a(\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w) - \ell^a(\pi^{v(\lambda)} w) \\ &\qquad = 2\,\mathrm{ht}((\lambda+\beta^\vee)^{++}) - 2\,\mathrm{ht}(\lambda) + \ell_{v^{vs_\beta(\lambda+\beta^\vee)}}(s_{v(\beta)} w) - \ell_v(w). \end{aligned} \tag{3.16}$$

We unwrap these formulas with the help of previous results.

(1) In the case $\boldsymbol{y} = \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$, let $\tilde{u} \in W_J$ be such that $vu\tilde{u} = (vu)^J = v^{v(\lambda+\beta^\vee)}$. The term $\ell_{v^{v(\lambda+\beta^\vee)}}(s_{v(\beta)} w) - \ell_v(w)$ rewrites as

$$\ell((u\tilde{u})^{-1} s_\beta v^{-1} w) - \ell(vu\tilde{u}) - \ell(v^{-1}w) + \ell(v).$$

Since $\boldsymbol{y} > \boldsymbol{x}$ is a covering, by Proposition 3.3, $vs_\beta u\tilde{u} = s_{v(\beta)}(vu)^J$ is on a minimal gallery from v to w, so $\ell(v^{-1}w) = \ell((vs_\beta u\tilde{u})^{-1} w) + \ell(s_\beta u\tilde{u})$. Moreover by Lemma 3.6, $\ell(s_\beta u\tilde{u}) = \ell(s_\beta) + \ell(u\tilde{u})$ and, by Lemma 3.7, $\ell(vu) = \ell(v) + \ell(u)$. Finally, by (1.3), since $u = u^J = v^{\lambda+\beta^\vee}$ and $vu\tilde{u} = (vu)^J = v^{v(\lambda+\beta^\vee)}$, we have $\ell(u\tilde{u}) = \ell(u) + \ell(\tilde{u})$ and $\ell(vu) = \ell(vu\tilde{u}) + \ell(\tilde{u})$. Thus

$$\begin{aligned} \ell_{v^{v(\lambda+\beta^\vee)}}(s_{v(\beta)} w) - \ell_v(w) &= \ell((u\tilde{u})^{-1} s_\beta v^{-1} w) - \ell(v^{-1}w) - \ell(vu\tilde{u}) + \ell(v) \\ &= -\ell(s_\beta u\tilde{u}) - \ell(vu) + \ell(\tilde{u}) + \ell(v) \\ &= -\ell(s_\beta) - \ell(u\tilde{u}) - \ell(u) + \ell(\tilde{u}) \\ &= -\ell(s_\beta) - 2\ell(u). \end{aligned} \tag{3.17}$$

(2) In the second case, let $\tilde{u} \in W_J$ be such that $vs_\beta u\tilde{u} = (vs_\beta u)^J = v^{vs_\beta(\lambda+\beta^\vee)}$. Then $\ell_{v^{vs_\beta(\lambda+\beta^\vee)}}(s_{v(\beta)} w) - \ell_v(w)$ rewrites as $\ell((u\tilde{u})^{-1} v^{-1} w) - \ell(vs_\beta u\tilde{u}) - \ell(v^{-1}w) + \ell(v)$. By Proposition 3.3, $\ell((u\tilde{u})^{-1} v^{-1} w) = \ell(v^{-1}w) - \ell(u\tilde{u})$. By (1.3),

$$\ell(u\tilde{u}) = \ell(u) + \ell(\tilde{u}) \quad \text{and} \quad \ell(vs_\beta u\tilde{u}) = \ell(vs_\beta u) - \ell(\tilde{u}).$$

By Lemmas 3.7 and 3.6,

$$\ell(vs_\beta u) = \ell(v) + \ell(s_\beta u) = \ell(v) + \ell(s_\beta) + \ell(u).$$

Thus, in this case,

$$\begin{aligned}(3.18)\quad \ell_{v^{vs_\beta(\lambda+\beta^\vee)}}(s_{v(\beta)}w)-\ell_v(w) &= \ell((u\tilde{u})^{-1}v^{-1}w)-\ell(vs_\beta u\tilde{u})-\ell(v^{-1}w)+\ell(v)\\ &= \ell(v^{-1}w)-\ell(u\tilde{u})-(\ell(vs_\beta u)-\ell(\tilde{u}))-\ell(v^{-1}w)+\ell(v)\\ &= -\ell(s_\beta)-2\ell(u).\end{aligned}$$

(3) By Lemma 1.13 we have

$$\begin{aligned}(3.19)\qquad 2\,\mathrm{ht}((\lambda+\beta^\vee)^{++}) &= 2\Big(\mathrm{ht}(\lambda+\beta^\vee)-\sum_{\tau\in\mathrm{Inv}(u^{-1})}\langle\lambda+\beta^\vee,\tau\rangle\Big)\\ &= 2\Big(\mathrm{ht}(\lambda)+\mathrm{ht}(\beta^\vee)-\sum_{\tau\in\mathrm{Inv}(u^{-1})}\langle\lambda+\beta^\vee,\tau\rangle\Big).\end{aligned}$$

By plugging (3.17), (3.19) into (3.15), and (3.18), (3.19) into (3.16) we obtain, either way,

$$\begin{aligned}\ell^a(\boldsymbol{y})-\ell^a(\boldsymbol{x}) &= 2\,\mathrm{ht}(\lambda)+2\,\mathrm{ht}(\beta^\vee)-2\sum_{\tau\in\mathrm{Inv}(u^{-1})}\langle\lambda+\beta^\vee,\tau\rangle-2\,\mathrm{ht}(\lambda)-\ell(s_\beta)-2\ell(u)\\ &= (2\,\mathrm{ht}(\beta^\vee)-\ell(s_\beta))-2\Big(\ell(u)+\sum_{\tau\in\mathrm{Inv}(u^{-1})}\langle\lambda+\beta^\vee,\tau\rangle\Big).\end{aligned}$$ □

Using Corollary 1.12, it is easy to see that $2\,\mathrm{ht}(\beta^\vee)-\ell(s_\beta)$ is always positive and that, on the contrary, $\ell(u)+\sum_{\tau\in\mathrm{Inv}(u^{-1})}\langle\lambda+\beta^\vee,\tau\rangle$ is always nonpositive. Therefore, the length difference is equal to 1 if and only if in the right-hand side of (3.14), the first term is equal to 1 and the second term cancels out. This motivates the following definitions.

Definition 3.9. A coweight $\mu\in Y^+$ is *almost dominant* if and only if

$$(3.20)\qquad \forall\tau\in\Phi_+,\ \langle\mu,\tau\rangle\geq -1.$$

A root $\beta\in\Phi_+$ is *a quantum root* if and only if

$$(3.21)\qquad \ell(s_\beta)=2\,\mathrm{ht}(\beta^\vee)-1.$$

The notion of quantum roots comes from the definition of quantum Bruhat graphs, (see [Lenart et al. 2015, §4.1]). With Notation 3.5, in Section 3.3 we prove that if $\boldsymbol{y}$ covers $\boldsymbol{x}$ then $\lambda+\beta^\vee$ is almost dominant and we prove in Section 3.4 that β needs to be a quantum root.

Remark 3.10. If $\lambda+\beta^\vee$ is dominant, then the second term in the right-hand side of (3.14) immediately cancels out, since in this case $u=1_W$. In the reductive case, Φ is finite and therefore if λ is far enough in the fundamental chamber (meaning that $\langle\lambda, \alpha_i\rangle$ is large for all $i \in I$, we say that λ is superregular), then $\lambda+\beta^\vee$ is always dominant. Accordingly, covers of $\pi^{v(\lambda)}w$ for λ superregular are easier to classify (see [Lam and Shimozono 2010, Proposition 4.4; Milićević 2021, Proposition 4.4]).

3.3. ***Almost-dominance in properly affine covers.*** We prove that the second term of the right-hand side of (3.14) need to be zero when $\boldsymbol{y}$ covers $\boldsymbol{x}$ (with Notation 3.5), through the following proposition:

Proposition 3.11. *Let $\lambda \in Y^{++}$, $v \in W^\lambda$ and $w \in W$. Let $\beta \in \Phi_+$ be a positive root such that $\lambda+\beta^\vee \in Y^+$ and suppose that $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ or $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$ covers $\pi^{v(\lambda)}w$. Then $\lambda+\beta^\vee$ is almost dominant, that is to say*

$$\forall\tau \in \Phi_+,\ \langle\lambda+\beta^\vee, \tau\rangle \geq -1. \tag{3.22}$$

It is deduced from the following two technical lemmas; we give their proofs after the proof of Proposition 3.11.

Lemma 3.12. *Let $\lambda \in Y^{++}$, $v \in W^\lambda$, $w \in W$, $\beta \in \Phi_+$. Suppose that there exists a pair $(\tau, n) \in \Phi_+ \times \mathbb{Z}$ such that*

(i) $n > 0$,

(ii) $\langle\lambda+n\tau^\vee, \beta\rangle \geq -1$,

(iii) $n < -\langle\lambda+\beta^\vee, \tau\rangle$.

Then, $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$ do not cover $\pi^{v(\lambda)}w$.

Lemma 3.13. *Let $\lambda \in Y^{++}$ and $\beta \in \Phi_+$ be such that $\lambda+\beta^\vee$ lies in Y^+. Let $\tau \in \Phi_+$ be such that $\langle\lambda+\beta^\vee, \tau\rangle \leq -2$ and suppose that $\langle\tau^\vee, \beta\rangle \leq -2$. Then*

$$\langle\lambda+\beta^\vee, s_\tau(\beta)\rangle \geq -1.$$

Proof of Proposition 3.11. We prove the contrapositive: Let $\tau \in \Phi_+$ be a positive root such that $\langle\lambda+\beta^\vee, \tau\rangle \leq -2$. We will produce nontrivial chains from $\pi^{v(\lambda)}w$ to $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$. In particular since λ is dominant, $\langle\beta^\vee, \tau\rangle \leq -2$.

The numbers $\langle\tau^\vee, \beta\rangle$ and $\langle\beta^\vee, \tau\rangle$ have the same sign [Bardy 1996, Lemma 1.1.10], and therefore we have that $\langle\tau^\vee, \beta\rangle \leq -1$.

Suppose first that $\langle\tau^\vee, \beta\rangle \leq -2$. Then $\big(\tau, -(\langle\lambda+\beta^\vee, \tau\rangle+1)\big)$ is a pair which satisfy the conditions of Lemma 3.12:

(i) This is true since $\langle\lambda+\beta^\vee, \tau\rangle \leq -2$, and $-(\langle\lambda+\beta^\vee, \tau\rangle+1) \geq 1 > 0$.

(ii) By Lemma 3.13, $\langle\lambda+\beta^\vee, s_\tau(\beta)\rangle \geq -1$; thus

$$\begin{aligned}\big\langle\lambda-(\langle\lambda+\beta^\vee,\tau\rangle+1)\tau^\vee,\beta\big\rangle &= \langle s_\tau(\lambda+\beta^\vee)-\beta^\vee-\tau^\vee,\beta\rangle\\ &= \langle\lambda+\beta^\vee, s_\tau(\beta)\rangle-2-\langle\tau^\vee,\beta\rangle\\ &\geq \langle\lambda+\beta^\vee, s_\tau(\beta)\rangle \geq -1.\end{aligned}$$

(iii) Clearly $-(\langle\beta^\vee,\tau\rangle+1) < -\langle\beta^\vee,\tau\rangle$.

Suppose now that $\langle\tau^\vee,\beta\rangle=-1$. We show that $(\tau,1)$ is a pair satisfying the conditions of Lemma 3.12:

(i) The first point is trivially verified.

(ii) Since $\langle\tau^\vee,\beta\rangle=-1$ and λ is dominant, $\langle\lambda+\tau^\vee,\beta\rangle\geq -1$.

(iii) Since $\langle\lambda+\beta^\vee,\tau\rangle\leq -2$ we obtain $1<-\langle\lambda+\beta^\vee,\tau\rangle$.

Hence, either way, if such a $\tau\in\Phi_+$ exists, then by Lemma 3.12 $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$ do not cover $\pi^{v(\lambda)}w$. □

Proof of Lemma 3.12. We use conditions (i), (ii), (iii) in the statement to produce chains from $\pi^{v(\lambda)}w$ to $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$.

Suppose first that (ii) is strict. Then we show that we have the chains

$$\pi^{v(\lambda)}w<\pi^{vs_\tau(\lambda+n\tau^\vee)}s_{v(\tau)}w<\pi^{vs_\tau(\lambda+\beta^\vee+n\tau^\vee)}s_{v(\tau)}s_{v(\beta)}w<\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w, \tag{3.23}$$

$$\pi^{v(\lambda)}w<\pi^{v(\lambda+n\tau^\vee)}s_{v(\tau)}w<\pi^{vs_\beta(\lambda+\beta^\vee+n\tau)}s_{v(\beta)}s_{v(\tau)}w<\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w. \tag{3.24}$$

(a) By (i), since λ is dominant and τ is a positive root, applying (3.8) with $(\mu_0,\alpha_0,v_0,w_0,m)=(\lambda,\tau,v,w,n)$, we have

$$\pi^{v(\lambda)}w<\pi^{v(\lambda+n\tau^\vee)}s_{v(\tau)}w. \tag{3.25}$$

Using (3.9) with the same parameters gives

$$\pi^{v(\lambda)}w<\pi^{vs_\tau(\lambda+n\tau^\vee)}s_{v(\tau)}w. \tag{3.26}$$

(b) Since $\langle\tau^\vee,\tau\rangle=2$, (iii) is equivalent to $-n<-\langle\lambda+\beta^\vee+n\tau^\vee,\tau\rangle$, so, using (3.9) for $(\mu_0,\alpha_0,v_0,w_0,m)=(\lambda+\beta^\vee+n\tau^\vee,\tau,vs_\tau,s_{v(\tau)}s_{v(\beta)}w,-n)$, we get

$$\pi^{vs_\tau(\lambda+\beta^\vee+n\tau^\vee)}s_{v(\tau)}s_{v(\beta)}w<\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w. \tag{3.27}$$

Using (3.8) for $(\mu_0,\alpha_0,v_0,w_0,m)=(\lambda+\beta^\vee+n\tau^\vee,\tau,vs_\beta,s_{v(\beta)}s_{v(\tau)}w,-n)$,

$$\pi^{vs_\beta(\lambda+\beta^\vee+n\tau)}s_{v(\beta)}s_{v(\tau)}w<\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w. \tag{3.28}$$

We now split the argument in two cases, according to whether (ii) is strict.

(1) Suppose that (ii) is strict, so $\langle \lambda + n\tau^\vee, \beta\rangle \geq 0$. Then by (3.8) applied with $(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda + n\tau^\vee, \beta, vs_\tau, s_{v(\tau)}w, 1)$, we get

$$\pi^{vs_\tau(\lambda+n\tau^\vee)} s_{v(\tau)} w < \pi^{vs_\tau(\lambda+\beta^\vee+n\tau^\vee)} s_{v(\tau)} s_{v(\beta)} w. \tag{3.29}$$

Moreover by (3.9) applied with $(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda + n\tau^\vee, \beta, v, s_{v(\tau)}w, 1)$,

$$\pi^{v(\lambda+n\tau^\vee)} s_{v(\tau)} w < \pi^{vs_\beta(\lambda+\beta^\vee+n\tau^\vee)} s_{v(\beta)} s_{v(\tau)} w. \tag{3.30}$$

Thus, if (ii) is strict, combining (3.25), (3.30) and (3.28) we obtain the chain (3.24). Moreover combining (3.26), (3.29) and (3.27) we obtain the chain (3.23). This proves Lemma 3.12 in this case.

(2) Suppose now that $\langle \lambda+n\tau^\vee, \beta\rangle = -1$. Then note that $\lambda+n\tau^\vee+\beta^\vee = s_\beta(\lambda+n\tau^\vee)$, and (3.8), (3.9) cannot be used for the middle inequalities of chains (3.24) and (3.23) anymore.

(a) Case of $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$.

(i) If $vs_\tau(\beta) \in \Phi_+$ then we can apply (1.12) to the element $\pi^{vs_\tau(\lambda+n\tau^\vee)} s_{v(\tau)} w$ and the positive affinized root $vs_\tau(\beta)[0]$, and since $\langle \lambda + n\tau^\vee, \beta\rangle = -1 < 0$, we still have

$$\pi^{vs_\tau(\lambda+n\tau^\vee)} s_{v(\tau)} w < s_{vs_\tau(\beta)[0]} \pi^{vs_\tau(\lambda+n\tau^\vee)} s_{v(\tau)} w = \pi^{vs_\tau(\lambda+\beta^\vee+n\tau^\vee)} s_{v(\tau)} s_{v(\beta)} w \tag{3.31}$$

and the chain (3.23) still holds by (3.26), (3.31) and (3.27).

(ii) If $vs_\tau(\beta) \in \Phi_-$ note that, since $\langle \lambda + n\tau^\vee, \beta\rangle < 0$, $\langle \tau^\vee, \beta\rangle < 0$, so $s_\tau(\beta)$ is a positive root. Therefore $vs_\tau(\beta) \in \Phi_-$ is equivalent to $s_\tau(\beta) \in \mathrm{Inv}(v)$. Since $v \in W^\lambda$, by minimality of v we have $\langle \lambda, s_\tau(\beta)\rangle \neq 0$. Then, by Proposition 2.3,

$$\pi^{v(\lambda)} w < \pi^{vs_{s_\tau(\beta)}(\lambda)} s_{vs_\tau(\beta)} w \tag{3.32}$$

and by (3.9) applied with $(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda, \tau, vs_{s_\tau(\beta)}, s_{vs_\tau(\beta)}w, 1)$ we get

$$\pi^{vs_{s_\tau(\beta)}(\lambda)} s_{vs_\tau(\beta)} w < \pi^{vs_\tau s_\beta(\lambda+n\tau^\vee)} s_{v(\tau)} s_{v(\beta)} w = \pi^{vs_\tau(\lambda+\beta^\vee+n\tau^\vee)} s_{v(\tau)} s_{v(\beta)} w. \tag{3.33}$$

For the vectorial elements, we used the fact that $s_{vs_\tau(\beta)} = s_{v(\tau)} s_{v(\beta)} s_{v(\tau)} = vs_\tau s_\beta s_\tau v^{-1}$ and $s_{vs_{s_\tau(\beta)}(\tau)} = vs_\tau s_\beta s_\tau s_\beta s_\tau v^{-1}$, and hence $s_{vs_{s_\tau(\beta)}(\tau)} s_{vs_\tau(\beta)} = vs_\tau s_\beta v^{-1} = s_{v(\tau)} s_{v(\beta)}$.

Combining (3.32), (3.33) and (3.27) we obtain the chain

$$\pi^{v(\lambda)} w < \pi^{vs_\tau s_\beta s_\tau(\lambda)} s_{v(\tau)} s_{v(\beta)} s_{v(\tau)} w < \pi^{v(s_\tau(\lambda+\beta^\vee+n\tau^\vee))} s_{v(\tau)} s_{v(\beta)} w < \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$$

which replaces the chain (3.23).

(b) Case of $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$.

(i) If $w^{-1}vs_\tau(\beta)\in\Phi_-$, by (1.12) applied to $\pi^{v(\lambda+n\tau^\vee)}s_{v(\tau)}w$ and the affinized root $v(\beta)[\langle\lambda+n\tau^\vee,\beta\rangle]$, since $\langle\lambda+n\tau^\vee,\beta\rangle<0$,

$$\pi^{v(\lambda+n\tau^\vee)}s_{v(\tau)}w<\pi^{v(\lambda+n\tau^\vee)}s_{v(\beta)}s_{v(\tau)}w=\pi^{vs_\beta(\lambda+\beta^\vee+n\tau^\vee)}s_{v(\beta)}s_{v(\tau)}w. \tag{3.34}$$

Therefore the chain (3.24) still holds by (3.25), (3.34) and (3.28).

(ii) If $w^{-1}vs_\tau(\beta)\in\Phi_+$, then using (1.12) with $\pi^{v(\lambda)}w$ and the affinized root $vs_\tau(\beta)\big[\langle\lambda,s_\tau(\beta)\rangle\big]$ (which is always possible because if $\langle\lambda,s_\tau(\beta)\rangle=0$ then by minimality of v, $vs_\tau(\beta)\in\Phi_+$), we obtain

$$\pi^{v(\lambda)}w<\pi^{v(\lambda)}s_{v(\tau)}s_{v(\beta)}s_{v(\tau)}w. \tag{3.35}$$

Moreover, by (3.8) applied with $(\mu_0,\alpha_0,v_0,w_0,m)=(\lambda,\tau,v,s_{vs_\tau(\beta)}w,n)$, we get

$$\pi^{v(\lambda)}s_{v(\tau)}s_{v(\beta)}s_{v(\tau)}w<\pi^{v(\lambda+n\tau^\vee)}s_{v(\beta)}s_{v(\tau)}w. \tag{3.36}$$

Hence combining (3.35), (3.36) and (3.28) we obtain a chain

$$\pi^{v(\lambda)}w<\pi^{v(\lambda)}s_{v(\tau)}s_{v(\beta)}s_{v(\tau)}w<\pi^{vs_\beta(\lambda+\beta^\vee+n\tau^\vee)}s_{v(\beta)}s_{v(\tau)}w<\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w.$$

Therefore, in all cases, if such a pair (τ,n) exists, then $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$ do not cover $\pi^{v(\lambda)}w$. □

Proof of Lemma 3.13. The proof relies on the assumption that $\lambda+\beta^\vee$ lies in the Tits cone, which is equivalent to saying that there is only a finite number of positive roots α such that $\langle\lambda+\beta^\vee,\alpha\rangle<0$.

We will show that $\langle\lambda+\beta^\vee,(s_\tau s_\beta)^n(\tau)\rangle\geq 0$ for n large enough implies

$$\langle\lambda+\beta^\vee,s_\tau(\beta)\rangle\geq -1, \tag{3.37}$$

which implies the lemma. To shorten the computation, let us write $a=-\langle\beta^\vee,\tau\rangle$ and $a^\vee=-\langle\tau^\vee,\beta\rangle$. So the assumptions $\langle\lambda+\beta^\vee,\tau\rangle\leq -2$ and $\langle\tau^\vee,\beta\rangle\leq -2$ imply that $a\geq 2+\langle\lambda,\tau\rangle$ and $a^\vee\geq 2$. In the basis (β,τ) of $\mathbb{R}\beta\oplus\mathbb{R}\tau$, the matrix of $s_\tau s_\beta$ is $M=\left(\begin{smallmatrix}-1 & a\\ -a^\vee & aa^\vee-1\end{smallmatrix}\right)$. We have $\chi_M=X^2+(2-aa^\vee)X+1$; thus, since $aa^\vee\geq 4$, $M^2=(aa^\vee-2)M-I_2$. Write $M^n=\mu_nM+\nu_nI_2$ for $n\in\mathbb{N}$. Then an easy computation shows that $\nu_n=-\mu_{n-1}$ and $\mu_{n+1}=(aa^\vee-2)\mu_n-\mu_{n-1}$. In particular since $aa^\vee-2\geq 2$ and $\mu_0=0<\mu_1$, by iteration $(\mu_n)_{n\in\mathbb{N}}$ is strictly increasing.

Let $x=\langle\lambda,\beta\rangle\geq 0$ and $y=\langle\lambda,\tau\rangle\in[\![0,a-2]\!]$. Then

$$\begin{aligned}\langle\lambda+\beta^\vee,(s_\tau s_\beta)^n(\tau)\rangle&=\big\langle\lambda+\beta^\vee,a\mu_n\beta+((aa^\vee-1)\mu_n-\mu_{n-1})\tau\big\rangle\\&=(x+2)\mu_na+((aa^\vee-1)\mu_n-\mu_{n-1})(y-a).\end{aligned} \tag{3.38}$$

Since $\lambda+\beta^\vee$ lies in the Tits cone, $\langle \lambda+\beta^\vee, (s_\tau s_\beta)^n(\tau)\rangle$ is nonnegative for n large enough. Moreover, $\mu_{n-1}<\mu_n$ for all $n\in\mathbb{N}$ and $a-y>0$. Therefore we deduce from (3.38) that, for n large enough,

$$(x+2)\mu_n a \geq (a-y)((aa^\vee-1)\mu_n-\mu_{n-1}) > (a-y)\mu_n(aa^\vee-2).$$

Hence

$$(x+2) > (a-y)\Big(a^\vee-\frac{2}{a}\Big) = aa^\vee - a^\vee y - 2 + 2\frac{y}{a}.$$

Therefore $\langle \lambda+\beta^\vee, s_\tau(\beta)\rangle = x+2+a^\vee y - aa^\vee > -2+2\frac{y}{a}$ and, since it is an integer, we deduce $\langle \lambda+\beta^\vee, s_\tau(\beta)\rangle \geq -1 \geq 1-a^\vee$, which proves the result. □

Corollary 3.14. *Let $\lambda\in Y^{++}$, $v\in W^\lambda$, $w\in W$. Let $\beta\in\Phi_+$ be a positive root such that $\lambda+\beta^\vee\in Y^+$. Suppose that $\boldsymbol{y}\in\{\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w, \pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w\}$ covers $\boldsymbol{x}=\pi^{v(\lambda)}w$.*

Then

$$\ell^a(\boldsymbol{y})-\ell^a(\boldsymbol{x}) = 2\,\mathrm{ht}(\beta^\vee)-\ell(s_\beta). \tag{3.39}$$

Proof. Let $u=v^{\lambda+\beta^\vee}\in W$. Then for any $\tau\in\mathrm{Inv}(u^{-1})$, by Lemma 2.1, $\langle\lambda+\beta^\vee,\tau\rangle$ is negative. By Proposition 3.11, $\langle\lambda+\beta^\vee,\tau\rangle=-1$ for any such τ. Therefore

$$\sum_{\tau\in\mathrm{Inv}(u^{-1})}\langle\lambda+\beta^\vee,\tau\rangle = -|\mathrm{Inv}(u^{-1})| = -\ell(u). \tag{3.40}$$

We then directly obtain (3.39) from (3.14) and (3.40). □

3.4. ***Properly affine covers and quantum roots.*** We now prove in Proposition 3.19 that, with Notation 3.5, if β is not a quantum root, then $\boldsymbol{y}$ does not cover $\boldsymbol{x}$. This is enough, together with Corollary 3.14, to conclude that $\ell^a(\boldsymbol{y})-\ell^a(\boldsymbol{x})=1$. There is a subtlety if the root β lies in a subsystem of Φ of type G_2; we suppose that this is not the case in Lemmas 3.16 and 3.17, and we deal with the G_2 case in Lemma 3.18. Let us first give another characterization of quantum roots.

Lemma 3.15. *A root $\beta\in\Phi_+$ is a quantum root if and only if $\langle\beta^\vee,\gamma\rangle=1$ for all $\gamma\in\mathrm{Inv}(s_\beta)\setminus\{\beta\}$.*

Proof. Recall that a quantum root is a root $\beta\in\Phi_+$ such that $2\,\mathrm{ht}(\beta^\vee)=\ell(s_\beta)+1$. By Corollary 1.12, this is equivalent to

$$\sum_{\gamma\in\mathrm{Inv}(s_\beta)}\langle\beta^\vee,\gamma\rangle = \ell(s_\beta)+1. \tag{3.41}$$

For any $\gamma\in\mathrm{Inv}(s_\beta)$, γ is a positive root and $s_\beta(\gamma)=\gamma-\langle\beta^\vee,\gamma\rangle\beta$ is a negative root, and therefore $\langle\beta^\vee,\gamma\rangle\geq 1$. Moreover, $\langle\beta^\vee,\beta\rangle=2$ and $|\mathrm{Inv}(s_\beta)|=\ell(s_\beta)$. Therefore (3.41) is satisfied if and only if $\langle\beta^\vee,\gamma\rangle$ is exactly one for all $\gamma\in\mathrm{Inv}(s_\beta)\setminus\{\beta\}$. □

Lemma 3.16. *Let $\lambda \in Y^{++}$, $v \in W^{\lambda}$, $w \in W$ and $\beta \in \Phi_+$. Let $\gamma \in \mathrm{Inv}(s_\beta) \setminus \{\beta\}$ be such that $\langle \beta^\vee, \gamma \rangle \geq 2$ and suppose that $\beta \notin \mathrm{Inv}(s_\gamma)$. Then $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ and $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ do not cover $\pi^{vs_\beta(\lambda)} w$.*

Proof. By the contrapositive of Proposition 3.11, we can suppose that $\langle \lambda + \beta^\vee, \tau \rangle \geq -1$ for any $\tau \in \Phi_+$. Let γ be as in the statement and write $\alpha = s_\gamma(\beta) \in \Phi_+$. We will construct nontrivial chains in the same fashion as in the proof of Lemma 3.12. Beforehand, we show by computation that $\langle \lambda + \gamma^\vee, \alpha \rangle \geq -1$. If $\langle \gamma^\vee, \beta \rangle = 1 = -\langle \gamma^\vee, \alpha \rangle$ this is clear since λ is dominant. If $\langle \gamma^\vee, \beta \rangle \geq 2$,

$$\begin{aligned} \langle \lambda + \gamma^\vee, \alpha \rangle &= \big\langle \lambda + \beta^\vee - \alpha^\vee + (1 - \langle \beta^\vee, \gamma \rangle)\gamma^\vee, \alpha \big\rangle \\ &= \langle \lambda + \beta^\vee, \alpha \rangle + (1 - \langle \beta^\vee, \gamma \rangle)\langle \gamma^\vee, \alpha \rangle - 2 \\ &= \langle \lambda + \beta^\vee, \alpha \rangle + (\langle \beta^\vee, \gamma \rangle - 1)\langle \gamma^\vee, \beta \rangle - 2. \end{aligned}$$

Since $\langle \beta^\vee, \gamma \rangle \geq 2$ and $\langle \gamma^\vee, \beta \rangle \geq 2$, $(\langle \beta^\vee, \gamma \rangle - 1)\langle \gamma^\vee, \beta \rangle \geq 2$, and by assumption $\langle \lambda + \beta^\vee, \alpha \rangle \geq -1$. Thus, $\langle \lambda + \gamma^\vee, \alpha \rangle \geq -1$ either way.

We construct chains which are slight modifications of the ones constructed in the proof of Lemma 3.12. We prove that, except in a few particular cases, we have the chains

$$\pi^{v(\lambda)} w < \pi^{v(\lambda+\gamma^\vee)} s_{v(\gamma)} w < \pi^{v(\lambda+\gamma^\vee+\alpha^\vee)} s_{v(\alpha)} s_{v(\gamma)} w < \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w, \tag{3.42}$$

$$\pi^{v(\lambda)} w < \pi^{vs_\gamma(\lambda+\gamma^\vee)} s_{v(\gamma)} w < \pi^{vs_\gamma s_\alpha(\lambda+\gamma^\vee+\alpha^\vee)} s_{v(\gamma)} s_{v(\alpha)} w < \pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w. \tag{3.43}$$

Indeed:

(1) The coweight λ is dominant and $\gamma \in \Phi_+$, so $\langle \lambda, \gamma \rangle \geq 0$ and (3.8) (resp. (3.9)) applied with $(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda, \gamma, v, w, 1)$ proves the leftmost inequality in the chain (3.42) (resp. (3.43)).

(2) Note that $\lambda + \beta^\vee = (\lambda + \gamma^\vee + \alpha^\vee) + (\langle \beta^\vee, \gamma \rangle - 1)\gamma^\vee$. Moreover

$$0 < \langle \beta^\vee, \gamma \rangle - 1 \tag{3.44}$$

and

$$-\langle \lambda + \gamma^\vee + \alpha^\vee, \gamma \rangle = \langle \beta^\vee, \gamma \rangle - \langle \lambda, \gamma \rangle - 2 < \langle \beta^\vee, \gamma \rangle - 1. \tag{3.45}$$

Therefore by applying (3.8) (resp. (3.9)) to

$$(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda + \gamma^\vee + \alpha^\vee, \gamma, v, s_{v(\alpha)} s_{v(\gamma)} w, \langle \beta^\vee, \gamma \rangle - 1)$$
$$(\text{resp. } (\mu_0, \alpha_0, v_0, w_0, m) = \lambda + \gamma^\vee + \alpha^\vee, \gamma, vs_\gamma s_\alpha, s_{v(\gamma)} s_{v(\alpha)} w, \langle \beta^\vee, \gamma \rangle - 1))$$

we obtain the rightmost inequality in the chain (3.42) (resp. (3.43)).

Let us now split cases in order to either prove the second inequality in chains (3.42) and (3.43) or provide alternative chains.

(1) Suppose first that $\langle \lambda + \gamma^\vee, \alpha\rangle \geq 0$. Then (3.8) (resp. (3.9)) applied with

$$(\mu_0, \alpha_0, v_0, w_0, m) = (\lambda + \gamma^\vee, \alpha, v, s_{v(\gamma)}w, 1) \tag{3.46}$$

$$\text{(resp. } (\mu_0, \alpha_0, v_0, w_0, m) = (\lambda + \gamma^\vee, \alpha, vs_\gamma, s_{v(\gamma)}w, 1)) \tag{3.47}$$

prove the middle inequality in the chain (3.42) (resp. (3.43)).

(2) Suppose that $\langle \lambda+\gamma^\vee, \alpha\rangle = -1$. Then $\lambda+\gamma^\vee+\alpha^\vee = s_\alpha(\lambda+\gamma^\vee)$ and the above chains do not always hold. We focus here on the case of $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$.

(a) If $v(\alpha) \in \Phi_+$, since $\langle \lambda + \gamma^\vee, \alpha\rangle < 0$, the inequality $\pi^{vs_\alpha(\lambda+\gamma^\vee)}s_{v(\alpha)}s_{v(\gamma)}w > \pi^{v(\lambda+\gamma^\vee)}s_{v(\gamma)}w$ still holds, by (1.12) applied with $s_{v(\alpha)[0]}$. Therefore the chain (3.42) still holds.

(b) If $v(\alpha) \in \Phi_-$, then $vs_\alpha < v$, and we have a chain

$$\begin{aligned}\pi^{v(\lambda)}w < \pi^{vs_\alpha(\lambda)}s_{v(\alpha)}w &< \pi^{vs_\alpha(\lambda+\gamma^\vee)}s_{v(\alpha)}s_{v(\gamma)}w \\ &= \pi^{v(\lambda+\gamma^\vee+\alpha^\vee)}s_{v(\alpha)}s_{v(\gamma)}w < \pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w.\end{aligned} \tag{3.48}$$

The reflection used for the first inequality is $s_{-v(\alpha)[0]}$, and it holds by (1.12) because $\langle v(\lambda), -v(\alpha)\rangle = -\langle \lambda, \alpha\rangle < 0$. Note that this is nonzero because v is the minimal representative of vW_λ and thus $vs_\alpha < v$ implies $s_\alpha \notin W_\lambda$ so $\langle \lambda, \alpha\rangle \neq 0$. By (3.44) and (3.45) we can use (3.8) with

$$(\mu_0, \alpha_0, v_0, w_0, m) = \begin{cases}(\lambda, \gamma, vs_\alpha, s_{v(\alpha)}w, 1) \\ (\lambda + \alpha^\vee + \gamma^\vee, \gamma, v, s_{v(\alpha)}s_{v(\gamma)}w, \langle \beta^\vee, \gamma\rangle - 1)\end{cases} \tag{3.49}$$

in order to obtain the second and third inequalities of chain (3.48), respectively.

(3) We suppose that $\langle \lambda+\gamma^\vee, \alpha\rangle = -1$. We deal with the case of $\pi^{vs_{v(\beta)}(\lambda+\beta^\vee)}s_{v(\beta)}w$. Then

$$\begin{aligned}\pi^{vs_\gamma s_\alpha(\lambda+\gamma^\vee+\alpha^\vee)}s_{v(\gamma)}s_{v(\alpha)}w &= \pi^{vs_\gamma(\lambda+\gamma^\vee)}s_{vs_\gamma(\alpha)}s_{v(\gamma)}w \\ &= s_{vs_\gamma(\alpha)[\langle\lambda+\gamma^\vee,\alpha\rangle]}\pi^{vs_\gamma(\lambda+\gamma^\vee)}s_{v(\gamma)}w.\end{aligned} \tag{3.50}$$

Moreover $(s_{v(\gamma)}w)^{-1}(vs_\gamma(\alpha)) = w^{-1}v(\alpha)$. Thus, since $\langle \lambda + \gamma^\vee, \alpha\rangle < 0$:

(a) If $w^{-1}v(\alpha) \in \Phi_-$, by (1.12), $\pi^{vs_\gamma(\lambda+\gamma^\vee)}s_{v(\gamma)}w < \pi^{vs_\gamma s_\alpha(\lambda+\gamma^\vee+\alpha^\vee)}s_{v(\gamma)}s_{v(\alpha)}w$ and the chain (3.43) still holds.

(b) If $w^{-1}v(\alpha) \in \Phi_+$, then, since $\langle \lambda, \alpha\rangle = \langle \gamma^\vee, \beta\rangle - 1 > 0$, by (1.12), $\pi^{v(\lambda)}w < s_{v(\alpha)[\langle\lambda,\alpha\rangle]}\pi^{v(\lambda)}w = \pi^{v(\lambda)}s_{v(\alpha)}w$. Then, by (3.9) with $(\mu_0, \alpha_0, v_0, w_0, m) =$

$(\lambda, \gamma, v, s_{v(\alpha)}w, 1)$, we have

$$\pi^{v(\lambda)} s_{v(\gamma)} s_{v(\beta)} s_{v(\gamma)} w < \pi^{vs_\gamma(\lambda+\gamma^\vee)} s_{v(\beta)} s_{v(\gamma)} w = \pi^{vs_\gamma s_\alpha(\lambda+\gamma^\vee+\alpha^\vee)} s_{v(\gamma)} s_{v(\alpha)} w$$

and we have a chain

$$\pi^{v(\lambda)} w < \pi^{v(\lambda)} s_{v(\alpha)} w < \pi^{vs_\gamma s_\alpha(\lambda+\gamma^\vee+\alpha^\vee)} s_{v(\gamma)} s_{v(\alpha)} w < \pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w. \qquad \square$$

Lemma 3.17. *Let $\beta \in \Phi_+$ be a positive root and suppose that there exists $\gamma \in \operatorname{Inv}(s_\beta) \setminus \{\beta\}$ such that $\langle \beta^\vee, \gamma \rangle \geq 2$ and $\langle \beta^\vee, \gamma \rangle \langle \gamma^\vee, \beta \rangle \neq 3$. Then γ can be chosen such that $\beta \notin \operatorname{Inv}(s_\gamma)$.*

Proof. Note that, by [Bardy 1996, Lemma 1.1.10], for any $\beta, \gamma \in \Phi$, $\langle \beta^\vee, \gamma \rangle$ and $\langle \gamma^\vee, \beta \rangle$ have the same sign, so if $\langle \beta^\vee, \gamma \rangle \geq 2$ and $\langle \beta^\vee, \gamma \rangle \langle \gamma^\vee, \beta \rangle \neq 3$, either $\langle \beta^\vee, \gamma \rangle \langle \gamma^\vee \rangle \geq 4$, either $\langle \beta^\vee, \gamma \rangle = 2$ and $\langle \gamma^\vee, \beta \rangle = 1$. We treat separately these cases:

(1) Let us first suppose that there exists $\gamma \in \operatorname{Inv}(s_\beta)$ such that $\langle \beta^\vee, \gamma \rangle = 2$ and $\langle \gamma^\vee, \beta \rangle = 1$. Suppose that $\beta \in \operatorname{Inv}(s_\gamma)$, so $s_\gamma(\beta) = \beta - \gamma < 0$, and $s_\beta(\gamma) = \gamma - 2\beta < 0$. Then we show that $\beta \notin \operatorname{Inv}(s_{\tilde{\gamma}})$ for $\tilde{\gamma} = -s_\beta(\gamma)$:

$$s_{\tilde{\gamma}}(\beta) = s_\beta s_\gamma s_\beta(\beta) = -s_\beta(\beta - \gamma) = \gamma - \beta = -s_\gamma(\beta) > 0.$$

Moreover $s_\beta(\tilde{\gamma}) = -\gamma < 0$ and $\langle \beta^\vee, \tilde{\gamma} \rangle = \langle \beta^\vee, \gamma \rangle = 2$; therefore, γ can be chosen such that $\beta \notin \operatorname{Inv}(s_\gamma)$.

(2) Let us now suppose that there exists $\gamma \in \operatorname{Inv}(s_\beta)$ such that $\langle \beta^\vee, \gamma \rangle \geq 2$ and $\langle \beta^\vee, \gamma \rangle \langle \gamma^\vee, \beta \rangle \geq 4$. Write $\beta = v_\beta(\beta_0) = s_{\alpha_1} \ldots s_{\alpha_n}(\beta_0)$ where the α_i and β_0 are simple roots, and suppose that n is of minimal length amongst possible expressions of β. Therefore $s_{\alpha_1} \ldots s_{\alpha_n} s_{\beta_0} s_{\alpha_n} \ldots s_{\alpha_1}$ is a reduced expression of s_β and

$$\operatorname{Inv}(s_\beta) = \{s_{\alpha_1} \ldots s_{\alpha_{p-1}}(\alpha_p) \mid p \leq n\} \sqcup \{\beta\} \sqcup \{s_{\alpha_1} \ldots s_{\alpha_n} s_{\beta_0} s_{\alpha_n} \ldots s_{\alpha_{n+1-p}}(\alpha_{n-p}) \mid p \leq n\}.$$

Let k be the smallest such that $\gamma_k = s_{\alpha_1} \ldots s_{\alpha_{k-1}}(\alpha_k)$ satisfies $\langle \beta^\vee, \gamma_k \rangle \geq 2$ and $\langle \beta^\vee, \gamma_k \rangle \langle \gamma_k^\vee, \beta \rangle \geq 4$.

The expression $s_{\alpha_1} \ldots s_{\alpha_{k-1}} s_{\alpha_k} s_{\alpha_{k-1}} \ldots s_{\alpha_1}$ is an expression of s_{γ_k}; thus

$$\operatorname{Inv}(s_{\gamma_k}) \subset \{s_{\alpha_1} \ldots s_{\alpha_{p-1}}(\alpha_p) \mid p \leq k-1\} \sqcup \{\gamma_k\} \sqcup \{s_{\alpha_1} \ldots s_{\alpha_k} s_{\alpha_{k-1}} \ldots s_{\alpha_{k+1-p}}(\alpha_{k-p}) \mid p \leq k-1\}.$$

Suppose by contradiction that $\beta \in \operatorname{Inv}(s_{\gamma_k})$. Since v_β is of minimal length, β is not in the first set; thus there is $p \in [\![1, k-1]\!]$ such that $\beta = s_{\alpha_1} \ldots s_{\alpha_k} s_{\alpha_{k-1}} \ldots s_{\alpha_{k+1-p}}(\alpha_{k-p})$.

We show that $\gamma_{k-p} = s_{\alpha_1} \dots s_{\alpha_{k-p-1}}(\alpha_{k-p}) \in \mathrm{Inv}(s_\beta)$ satisfies $\langle \beta^\vee, \gamma_{k-p}\rangle \geq 2$, which contradicts the minimality of k. Note that $\beta = -s_{\gamma_k}(\gamma_{k-p})$. We compute

$$\begin{aligned}\langle \beta^\vee, \gamma_{k-p}\rangle &= \langle -s_{\gamma_k}(\gamma_{k-p}^\vee), \gamma_{k-p}\rangle \\ &= -(2 - \langle \gamma_{k-p}^\vee, \gamma_k\rangle\langle \gamma_k^\vee, \gamma_{k-p}\rangle) \\ &= \langle \beta^\vee, \gamma_k\rangle\langle \gamma_k^\vee, \beta\rangle - 2.\end{aligned}$$

So since $\langle \beta^\vee, \gamma_k\rangle\langle \gamma_k^\vee, \beta\rangle \geq 4$, we get $\langle \beta^\vee, \gamma_{k-p}\rangle \geq 2$, and with a similar computation, we find that $\langle \gamma_{k-p}^\vee, \beta\rangle = \langle \beta^\vee, \gamma_{k-p}\rangle \geq 2$ as well, so $\langle \beta^\vee, \gamma_{k-p}\rangle\langle \gamma_{k-p}^\vee, \beta\rangle \geq 4$. This contradicts the minimality of k and thus $\beta \notin \mathrm{Inv}(s_{\gamma_k})$. □

Lemma 3.18. *Let $\lambda \in Y^{++}$, $v \in W^\lambda$ and $w \in W$. Let $\beta \in \Phi_+$ and let $\gamma \in \mathrm{Inv}(s_\beta)$ be such that $\beta \in \mathrm{Inv}(s_\gamma)$ and $\langle \beta^\vee, \gamma\rangle = 3$, $\langle \gamma^\vee, \beta\rangle = 1$.*

Then $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$ do not cover $\pi^{v(\lambda)} w$.

Proof. We show that, with the assumptions of the statement, β and γ appear as positive roots of a root subsystem of Φ isomorphic to G_2, and we use this system to construct chains replacing the ones in the proof of Lemma 3.16.

First, note that $-s_\gamma(\beta)$ lies in $\mathrm{Inv}(s_\beta)$ (so $s_\beta s_\gamma(\beta)$ is positive). Indeed, we can write, as in the proof of Lemma 3.17, $\beta = s_{\alpha_1} \dots s_{\alpha_n}(\beta_0)$ for a minimal n, and $\gamma = s_{\alpha_1} \dots s_{\alpha_{k-1}}(\alpha_k)$ for some $k \leq n$. Then, since $\beta \in \mathrm{Inv}(s_\gamma)$, β is also of the form $s_{\alpha_1} \dots s_{\alpha_k} s_{\alpha_{k-1}} \dots s_{\alpha_{k-p+1}}(\alpha_{k-p})$ for some $p \leq k-1$, and thus

$$-s_\gamma(\beta) = s_{\alpha_1} \dots s_{\alpha_{k-p-1}}(\alpha_{k-p}) \in \mathrm{Inv}(s_\beta).$$

Therefore we have the following positive roots, and their associated coroots (the notation will become clear afterwards):

(1) $\theta_1 := -s_\gamma(\beta) = \gamma - \beta \in \Phi_+$, with associated coroot $\theta_1^\vee = -s_\gamma(\beta^\vee) = 3\gamma^\vee - \beta^\vee$.
(2) $\tilde\beta := -s_\beta(\gamma) = 3\beta - \gamma \in \Phi_+$, with associated coroot $\tilde\beta^\vee = -s_\beta(\gamma^\vee) = \beta^\vee - \gamma^\vee$.
(3) $\tilde\gamma := s_\beta s_\gamma(\beta) = 2\beta - \gamma \in \Phi_+$, with associated coroot $\tilde\gamma^\vee = s_\beta s_\gamma(\beta^\vee) = 2\beta^\vee - 3\gamma^\vee$.

Let us also define $\theta_2 = s_{\theta_1}(\gamma) = 3\beta - 2\gamma$, with associated coroot $\theta_2^\vee = \beta^\vee - 2\gamma^\vee$. Then one can check that $\{\theta_1, \theta_2\}$ form the positive simple roots of a G_2 root system (in the sense that $\langle \theta_1^\vee, \theta_2\rangle = -3$ and $\langle \theta_2^\vee, \theta_1\rangle = -1$), such that $\gamma = s_{\theta_1}(\theta_2)$, $\beta = s_{\theta_1} s_{\theta_2}(\theta_1)$, $\tilde\gamma = s_{\theta_2}(\theta_1)$ and $\tilde\beta = s_{\theta_2} s_{\theta_1}(\theta_2)$. However, θ_2 may not be a positive root in Φ, and we thus need to distinguish these two cases.

Let us first suppose that θ_2 lies in Φ_+. Notice that

$$\theta_1^\vee + \tilde\beta^\vee + \theta_2^\vee = (3\gamma^\vee - \beta^\vee) + (\beta^\vee - \gamma^\vee) + (\beta^\vee - 2\gamma^\vee) = \beta^\vee, \tag{3.51}$$

and

$$s_{\theta_1}s_{\tilde{\beta}}s_{\theta_2} = s_{\theta_1}(s_{\theta_2}s_{\theta_1}s_{\theta_2}s_{\theta_1}s_{\theta_2})s_{\theta_2} = s_{\theta_1}s_{\theta_2}s_{\theta_1}s_{\theta_2}s_{\theta_1} = s_{\theta_2}s_{\tilde{\beta}}s_{\theta_1} = s_{\beta}. \tag{3.52}$$

Moreover, we have

$$\langle \theta_2^\vee, \tilde{\beta}\rangle = \langle \beta^\vee - 2\gamma^\vee, 3\beta - \gamma\rangle = 1 > 0, \tag{3.53}$$

$$\langle \theta_2^\vee + \tilde{\beta}^\vee, \theta_1\rangle = \langle 2\beta^\vee - 3\gamma^\vee, \gamma - \beta\rangle = -1. \tag{3.54}$$

(1) Suppose first that $\langle \lambda, \theta_1\rangle > 0$. Since λ is dominant and by (3.53), (3.54),

$$-\langle \lambda + \theta_2^\vee, \tilde{\beta}\rangle < 0 \quad \text{and} \quad -\langle \lambda + \theta_2^\vee + \tilde{\beta}^\vee, \theta_1\rangle \leq 0.$$

Using (3.8) (resp. (3.9)) with $(\mu_0, \alpha_0, m) = (\lambda, \theta_2, 1)$ for the first inequality, $(\mu_0, \alpha_0, m) = (\lambda + \theta_2^\vee, \tilde{\beta}, 1)$ for the second and $(\mu_0, \alpha_0, m) = (\lambda + \theta_2^\vee + \tilde{\beta}^\vee, \theta_1, 1)$ for the third (recall (3.51), (3.52)), we obtain the chain (3.55) (resp. (3.56))

$$\pi^{v(\lambda)}w < \pi^{v(\lambda+\theta_2^\vee)}s_{v(\theta_2)}w < \pi^{v(\lambda+\theta_2^\vee+\tilde{\beta}^\vee)}s_{v(\tilde{\beta})}s_{v(\theta_2)}w < \pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w, \tag{3.55}$$

$$\begin{aligned}\pi^{v(\lambda)}w &< \pi^{vs_{\theta_2}(\lambda+\theta_2^\vee)}s_{v(\theta_2)}w \\ &< \pi^{vs_{\theta_2}s_{\tilde{\beta}}(\lambda+\theta_2^\vee+\tilde{\beta}^\vee)}s_{v(\theta_2)}s_{v(\tilde{\beta})}w < \pi^{vs_{\beta}(\lambda+\beta^\vee)}s_{v(\beta)}w.\end{aligned} \tag{3.56}$$

(2) If $\langle \lambda, \theta_1\rangle = 0$, then $-\langle \lambda + \theta_2^\vee + \tilde{\beta}^\vee, \theta_1\rangle = 1$ so the last inequality in the chains (3.55) and (3.56) do not always hold, we have the following case distinction, which we already encountered in Lemmas 3.12 and 3.16:

(a) If $v(\theta_1) \in \Phi_+$, the chain (3.55) still holds, else $vs_{\theta_1} < v$, $\lambda+\beta^\vee = s_{\theta_1}(\lambda+\theta_2^\vee+\tilde{\beta})$ and we instead have the chain

$$\pi^{v(\lambda)}w < \pi^{vs_{\theta_1}(\lambda)}s_{v(\theta_1)}w < \pi^{vs_{\theta_1}(\lambda+\theta_2^\vee)}s_{v(\theta_1)}s_{v(\theta_2)}w < \pi^{vs_{\theta_1}(\lambda+\theta_2^\vee+\tilde{\beta}^\vee)}s_{v(\theta_1)}s_{v(\tilde{\beta})}s_{v(\theta_2)}w,$$

where the last term is actually equal to $\pi^{v(\lambda+\beta^\vee)}s_{v(\beta)}w$.

(b) If $w^{-1}v(\theta_1) \in \Phi_-$, then since $\langle \lambda + \theta_2^\vee + \tilde{\beta}^\vee, \theta_1\rangle < 0$, by (1.12) applied with the affinized root $vs_{\theta_2}s_{\tilde{\beta}}(\theta_1)[\langle \lambda + \theta_2^\vee + \tilde{\beta}^\vee, \theta_1\rangle]$, the third inequality of chain (3.56) still holds, and thus the whole chain remains correct. Otherwise if $w^{-1}v(\theta_1) \in \Phi_+$ we instead have the chain

$$\pi^{v(\lambda)}w < \pi^{v(\lambda)}s_{v(\theta_1)}w < \pi^{vs_{\theta_2}(\lambda+\theta_2^\vee)}s_{v(\theta_2)}s_{v(\theta_1)}w < \pi^{vs_{\theta_2}s_{\tilde{\beta}}(\lambda+\theta_2^\vee+\tilde{\beta}^\vee)}s_{v(\theta_2)}s_{v(\tilde{\beta})}s_{v(\theta_1)}w,$$

where the last term is actually equal to $\pi^{vs_\beta(\lambda+\beta^\vee)}s_{v(\beta)}w$ since $\lambda + \theta_2^\vee + \tilde{\beta}^\vee = s_{\theta_1}(\lambda + \beta^\vee)$.

We now turn to the case of $\theta_2 \in \Phi_-$. Notice that $\beta^\vee = -\theta_2^\vee + \tilde{\gamma}^\vee + \gamma^\vee$ and $s_\beta = s_\gamma s_{\tilde{\gamma}} s_{\theta_2} = s_{\theta_2} s_{\tilde{\gamma}} s_\gamma$. Moreover, $\langle -\theta_2^\vee, \tilde{\gamma}\rangle = \langle 2\gamma^\vee - \beta^\vee, 2\beta - \gamma\rangle = -1$ and

$\langle -\theta_2^\vee + \tilde{\gamma}^\vee, \gamma\rangle = \langle \beta^\vee - \gamma^\vee, \gamma\rangle = 1$. Therefore, since λ is dominant and $-\theta_2$ is a positive root:

(1) If $\langle \lambda, \tilde{\gamma}\rangle > 0$, then using (3.8) (resp. (3.9)) with $(\mu_0, \alpha_0, m) = (\lambda, -\theta_2, 1)$ for the first inequality, $(\lambda - \theta_2^\vee, \tilde{\gamma}, 1)$ for the second and $(\lambda - \theta_2^\vee + \tilde{\gamma}^\vee, \gamma, 1)$ for the third, we obtain the chain (3.57) (resp. (3.58))

$$\pi^{v(\lambda)} w < \pi^{v(\lambda-\theta_2^\vee)} s_{v(\theta_2)} w < \pi^{v(\lambda-\theta_2^\vee+\tilde{\gamma}^\vee)} s_{v(\tilde{\gamma})} s_{v(\theta_2)} w < \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w \tag{3.57}$$

$$\begin{aligned}\pi^{v(\lambda)} w &< \pi^{vs_{\theta_2}(\lambda-\theta_2^\vee)} s_{v(\theta_2)} w \\ &< \pi^{vs_{\theta_2}s_{\tilde{\gamma}}(\lambda-\theta_2^\vee+\tilde{\gamma}^\vee)} s_{v(\theta_2)} s_{v(\tilde{\gamma})} w < \pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w.\end{aligned} \tag{3.58}$$

(2) Suppose now that $\langle \lambda, \tilde{\gamma}\rangle = 0$, so $\lambda - \theta_2^\vee + \tilde{\gamma} = s_{\tilde{\gamma}}(\lambda - \theta_2^\vee)$. Then:

(a) If $v(\tilde{\gamma}) \in \Phi_+$, the chain (3.57) still holds. Else, $vs_{\tilde{\gamma}} < v$ and we instead have the chain

$$\pi^{v(\lambda)} w < \pi^{vs_{\tilde{\gamma}}(\lambda)} s_{v(\tilde{\gamma})} w < \pi^{vs_{\tilde{\gamma}}(\lambda-\theta_2^\vee)} s_{v(\tilde{\gamma})} s_{v(\theta_2)} w < \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w, \tag{3.59}$$

where the first inequality comes from Proposition 2.3 and the two others from (3.8).

(b) If $w^{-1}v(\tilde{\gamma}) \in \Phi_-$, then the chain (3.58) still holds. Else $w^{-1}v(\tilde{\gamma}) \in \Phi_+$ and we instead have the chain

$$\pi^{v(\lambda)} w < \pi^{v(\lambda)} s_{v(\tilde{\gamma})} w < \pi^{vs_{\theta_2}(\lambda-\theta_2^\vee)} s_{v(\theta_2)} s_{v(\tilde{\gamma})} w < \pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w, \tag{3.60}$$

where the first inequality is deduced from (1.12) used with the affinized root $v(\tilde{\gamma})[\langle \lambda, \tilde{\gamma}\rangle]$, and the two others from (3.9) as for the chain (3.58). □

Proposition 3.19. *Let $\lambda \in Y^{++}$, $v \in W^\lambda$ and $w \in W$. Let $\beta \in \Phi_+$ and suppose that $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ or $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$ cover $\pi^{v(\lambda)} w$. Then β is a quantum root.*

Proof. We prove the contrapositive. Suppose that β is not a quantum root. By Lemma 3.15, there is $\gamma \in \mathrm{Inv}(s_\beta) \setminus \{\beta\}$ such that $\langle \beta^\vee, \gamma\rangle \geq 2$. If $\beta \notin \mathrm{Inv}(s_\gamma)$ we apply Lemma 3.16. We can also apply it in case $\langle \beta^\vee, \gamma\rangle\langle \gamma^\vee, \beta\rangle \neq 3$ by Lemma 3.17. Finally if $\beta \in \mathrm{Inv}(s_\gamma)$ and $\langle \beta^\vee, \gamma\rangle\langle \gamma^\vee, \beta\rangle = 3$ we apply Lemma 3.18. Therefore if β is not a quantum root then $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ and $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$ do not cover $\pi^{v(\lambda)} w$. □

3.5. *Conclusion.* We now have everything to prove Theorem A:

Theorem A. *Suppose that $\boldsymbol{y}, \boldsymbol{x} \in W_+^a$ are such that $\boldsymbol{x} \leq \boldsymbol{y}$. Then*

$$\boldsymbol{x} \lhd \boldsymbol{y} \iff \ell^a(\boldsymbol{y}) = \ell^a(\boldsymbol{x}) + 1. \tag{3.61}$$

Proof. If $x \leq y$ with $\ell^a(y) = \ell^a(x) + 1$, then by strict compatibility of ℓ^a (Theorem 1.7), y covers x. Conversely suppose that y covers x. If y and x have same dominance class then Theorem 2.18 implies that $\ell^a(y) = \ell^a(x) + 1$. Else, if $\mathrm{proj}^{Y^+}(y) \notin W \cdot \mathrm{proj}^{Y^+}(x)$, by Proposition 3.1, y is of the form $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ or $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$, for $x = \pi^{v(\lambda)} w$ with $\lambda \in Y^{++}$, $v \in W^\lambda$, $w \in W$ and $\beta \in \Phi_+$. Then, by Corollary 3.14, we have $\ell^a(y) - \ell^a(x) = 2\,\mathrm{ht}(\beta^\vee) - \ell(s_\beta)$. Moreover, by Proposition 3.19, β is a quantum root and therefore, in this case as well,

$$\ell^a(y) - \ell^a(x) = 1. \qquad \square$$

Along the way, we have obtained a classification of covers, which we summarize in Proposition 3.20. This is to be compared with [Schremmer 2024, Proposition 4.5].

Proposition 3.20. *Let $x = \pi^{v(\lambda)} w \in W^a_+$ with $\lambda \in Y^{++}$, $v \in W^\lambda$ and $w \in W$. Let $J \subset S$ be the set of simple reflections such that $W_\lambda = W_J$, and recall Notation 1.3 and Definition 3.9. Then covers of x are the elements of the following form*:

(1) $\pi^{v(\lambda)} s_{v(\beta)} w = x s_{w^{-1}v(\beta)[0]}$ *for $\beta \in \Phi$ such that $\ell(s_\beta v^{-1} w) = \ell(v^{-1} w) + 1$.*

(2) $\pi^{vs_\beta(\lambda)} s_{v(\beta)} w = s_{v(\beta)[0]} x$ *for $\beta \in \Phi_+$ such that*:

- (a) $\langle \lambda, \beta \rangle \neq 0$.
- (b) $\ell(vs_\beta) = \ell(v) - 1$.
- (c) *If u denotes vs_β and u_J the maximal W_J-suffix of u, then vu_J^{-1} is on a minimal gallery from v to w.*

(3) $\pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w = s_{v(\beta)[\langle \lambda, \beta \rangle + 1]} x = x s_{w^{-1}v(\beta)[1]}$ *for $\beta \in \Phi_+$ such that*:

- (a) *β is a quantum root.*
- (b) *$\lambda + \beta^\vee$ is an almost dominant coweight.*
- (c) *For $u = v^{\lambda+\beta^\vee}$, v is on a minimal gallery from 1 to vu, that is to say $\ell(vu) = \ell(v) + \ell(u)$.*
- (d) *For $\tilde{v} = v^{v(\lambda+\beta^\vee)}$, $s_{v(\beta)}\tilde{v}$ is on a minimal gallery from v to w.*

(4) $\pi^{vs_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w = s_{v(\beta)[-1]} x$ *for $\beta \in \Phi_+$ such that*:

- (a) *β is a quantum root.*
- (b) *$\lambda + \beta^\vee$ is an almost dominant coweight.*
- (c) *For $u = v^{\lambda+\beta^\vee}$, v is on a minimal gallery from 1 to $vs_\beta u$.*
- (d) *For $\tilde{v} = v^{vs_\beta(\lambda+\beta^\vee)}$, $s_{v(\beta)}\tilde{v}$ is on a minimal gallery from v to w.*

In particular, suppose that $\lambda \in Y^{++}$ is regular and is such that $\lambda + \beta^\vee$ is also regular for any quantum root $\beta \in \Phi_+$. We then say that λ is superregular. Proposition 3.20 can be simplified for superregular coweights. This is to be compared with [Lam and Shimozono 2010, Proposition 4.4] and [Welch 2022, Theorem 2].

Proposition 3.21. *Let $\boldsymbol{x} = \pi^{v(\lambda)} w \in W_+^a$ with $\lambda \in Y^{++}$ a superregular coweight and $v, w \in W$. Then covers of $\boldsymbol{x}$ are the elements of the following form*:

(1) $\boldsymbol{x} s_{\beta[0]} = \pi^{v(\lambda)} w s_\beta$ *for $\beta \in \Phi_+$ such that $\ell(v^{-1} w s_\beta) = \ell(v^{-1} w) + 1$.*

(2) $s_{\beta[0]} \boldsymbol{x} = \pi^{s_\beta v(\lambda)} s_\beta w$ *for $\beta \in \Phi_+$ such that $\ell(s_\beta v) = \ell(v) - 1$.*

(3) $\boldsymbol{x} s_{w^{-1} v(\beta)[1]} = \pi^{v(\lambda+\beta^\vee)} s_{v(\beta)} w$ *for $\beta \in \Phi_+$ a quantum root such that $\ell(v^{-1} w) = \ell(s_\beta) + \ell(s_\beta v^{-1} w)$ (otherwise said $s_\beta v^{-1} w \leq_R v^{-1} w$).*

(4) $s_{v(\beta)[-1]} \boldsymbol{x} = \pi^{v s_\beta(\lambda+\beta^\vee)} s_{v(\beta)} w$ *for $\beta \in \Phi_+$ a quantum root such that $\ell(v s_\beta) = \ell(v) + \ell(s_\beta)$ (otherwise said $s_\beta \leq_R v s_\beta$).*

For Kac–Moody root systems, the existence of superregular coweights is not clear a priori. However in an upcoming joint work with Hébert we prove that any Kac–Moody root system admits a finite number of quantum roots, which ensures the existence of superregular coweights. We also use this finiteness to deduce that any element of W_a^+ admits a finite number of covers; in particular intervals in W_a^+ are finite.

Acknowledgements

I am especially grateful to Auguste Hébert for his help in understanding the affinized Bruhat order and his feedback throughout the conception of this paper, as well as to Stéphane Gaussent for his supervision and his careful rereading. I would also like to thank the anonymous referee for help and corrections.

References

[Bardy 1996] N. Bardy, *Systèms de racines infinis*, Mémoire Nouvelle Série **65**, Société Mathématique de France, Paris, 1996. MR Zbl

[Bardy-Panse et al. 2016] N. Bardy-Panse, S. Gaussent, and G. Rousseau, "Iwahori–Hecke algebras for Kac–Moody groups over local fields", *Pacific J. Math.* **285**:1 (2016), 1–61. MR Zbl

[Bardy-Panse et al. 2022] N. Bardy-Panse, A. Hébert, and G. Rousseau, "Twin masures associated with Kac–Moody groups over Laurent polynomials", preprint, 2022. arXiv 2210.07603

[Björner and Brenti 2005] A. Björner and F. Brenti, *Combinatorics of Coxeter groups*, Graduate Texts in Mathematics **231**, Springer, 2005. MR Zbl

[Braverman et al. 2016] A. Braverman, D. Kazhdan, and M. M. Patnaik, "Iwahori–Hecke algebras for p-adic loop groups", *Invent. Math.* **204**:2 (2016), 347–442. MR Zbl

[Chaput et al. 2021] P.-E. Chaput, L. Fresse, and T. Gobet, "Parametrization, structure and Bruhat order of certain spherical quotients", *Represent. Theory* **25** (2021), 935–974. MR Zbl

[Iwahori and Matsumoto 1965] N. Iwahori and H. Matsumoto, "On some Bruhat decomposition and the structure of the Hecke rings of $\mathfrak{p}$-adic Chevalley groups", *Inst. Hautes Études Sci. Publ. Math.* 25 (1965), 5–48. MR Zbl

[Kazhdan and Lusztig 1980] D. Kazhdan and G. Lusztig, "Schubert varieties and Poincaré duality", pp. 185–203 in *Geometry of the Laplace operator* (Honolulu, HI, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, RI, 1980. MR Zbl

[Kumar 2002] S. Kumar, *Kac–Moody groups, their flag varieties and representation theory*, Progress in Mathematics **204**, Birkhäuser, Boston, 2002. MR Zbl

[Lam and Shimozono 2010] T. Lam and M. Shimozono, "Quantum cohomology of G/P and homology of affine Grassmannian", *Acta Math.* **204**:1 (2010), 49–90. MR Zbl

[Lenart et al. 2015] C. Lenart, S. Naito, D. Sagaki, A. Schilling, and M. Shimozono, "A uniform model for Kirillov–Reshetikhin crystals, I: Lifting the parabolic quantum Bruhat graph", *Int. Math. Res. Not.* **2015**:7 (2015), 1848–1901. MR Zbl

[Milićević 2021] E. Milićević, "Maximal Newton points and the quantum Bruhat graph", *Michigan Math. J.* **70**:3 (2021), 451–502. MR Zbl

[Muthiah 2018] D. Muthiah, "On Iwahori–Hecke algebras for p-adic loop groups: double coset basis and Bruhat order", *Amer. J. Math.* **140**:1 (2018), 221–244. MR Zbl

[Muthiah 2019] D. Muthiah, "Double-affine Kazhdan–Lusztig polynomials via masures", preprint, 2019. arXiv 1910.13694

[Muthiah and Orr 2018] D. Muthiah and D. Orr, "Walk algebras, distinguished subexpressions, and point counting in Kac–Moody flag varieties", pp. 187–203 in *Representations of Lie algebras, quantum groups and related topics*, Contemp. Math. **713**, Amer. Math. Soc., Providence, RI, 2018. MR Zbl

[Muthiah and Orr 2019] D. Muthiah and D. Orr, "On the double-affine Bruhat order: the $\varepsilon = 1$ conjecture and classification of covers in ADE type", *Algebr. Comb.* **2**:2 (2019), 197–216. MR Zbl

[Muthiah and Puskás 2024] D. Muthiah and A. Puskás, "Pursuing Coxeter theory for Kac–Moody affine Hecke algebras", preprint, 2024. arXiv 2406.14447

[Rémy 2002] B. Rémy, *Groupes de Kac–Moody déployés et presque déployés*, Astérisque **277**, Société Mathématique de France, Paris, 2002. MR Zbl

[Ronan 1989] M. Ronan, *Lectures on buildings*, Perspectives in Mathematics **7**, Academic Press, Boston, 1989. MR Zbl

[Schremmer 2024] F. Schremmer, "Affine Bruhat order and Demazure products", *Forum Math. Sigma* **12** (2024), art. id. 53. MR Zbl

[Welch 2022] A. Welch, "Classification of cocovers in the double affine Bruhat order", *Electron. J. Combin.* **29**:4 (2022), art. id. 4.7. MR Zbl

Received June 12, 2023. Revised October 4, 2024.

PAUL PHILIPPE
INSTITUT CAMILLE JORDAN
UNIVERSITÉ JEAN MONNET
SAINT-ETIENNE
FRANCE
paul.philippe@univ-st-etienne.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.321

CM POINTS ON SHIMURA CURVES VIA QM-EQUIVARIANT ISOGENY VOLCANOES

FREDERICK SAIA

We study complex multiplication points on the Shimura curves $X_0^D(N)_{/\mathbb{Q}}$ and $X_1^D(N)_{/\mathbb{Q}}$, parametrizing abelian surfaces with quaternionic multiplication and extra level structure. A description of the locus of points with CM by a specified order is obtained for general level, via an isogeny-volcano approach in analogy to work of Clark and Saia in the $D=1$ case of modular curves. This allows for a count of all points with CM by a specified order on such a curve, and a determination of all primitive residue fields and primitive degrees of such points on $X_0^D(N)_{/\mathbb{Q}}$. We leverage computations of least degrees towards the existence of sporadic CM points on $X_0^D(N)_{/\mathbb{Q}}$.

1. Introduction 321
2. Background 325
3. QM-equivariant isogenies 336
4. QM-equivariant isogeny volcanoes 338
5. The action of Galois on $\mathcal{G}^D_{K,\ell,f_0}$ 344
6. CM points on $X_0^D(\ell^a)_{/\mathbb{Q}}$ 353
7. Algebraic results on residue fields of CM points on $X^D(1)_{/\mathbb{Q}}$ 362
8. CM points on $X_0^D(N)_{/\mathbb{Q}}$ 365
9. CM points on $X_1^D(N)_{/\mathbb{Q}}$ 369
10. Sporadic CM points on Shimura curves 371
Acknowledgments 381
References 382

1. Introduction

The restriction of the study of torsion of elliptic curves over number fields to the case of complex multiplication (CM) has seen considerable recent progress. In particular, Clark [2022] and Clark and Saia [2022], continuing a program of research in this area from the perspective of CM points on modular curves (see, e.g., [Clark et al. 2013; 2022; Bourdon and Clark 2020]), approach the study of the CM locus on the modular

MSC2020: 11G10, 11G15, 11G18.
Keywords: Shimura curve, quaternionic multiplication, abelian surface, complex multiplication, isogeny.

curves $X_0(M, N)_{/\mathbb{Q}}$ and $X_1(M, N)_{/\mathbb{Q}}$ via a study of CM components of isogeny graphs of elliptic curves over $\overline{\mathbb{Q}}$. For K an imaginary quadratic field and $\Delta = f^2\Delta_K$ the discriminant of the order $\mathfrak{o}(f)$ of conductor f in K, let $j_\Delta \in X(1)_{/\mathbb{Q}}$ denote the closed point corresponding to elliptic curves with CM by the order of discriminant Δ. The work of [Clark 2022; Clark and Saia 2022] results, for instance, in a description of all points in the fiber of the natural map $X_0(M, N)_{/\mathbb{Q}} \to X(1)_{/\mathbb{Q}}$ over j_Δ. This description provides the list of residue fields of Δ-CM points on the first curve, along with a count of closed points in this fiber with each specified residue field.

We study the Shimura curves $X_0^D(N)_{/\mathbb{Q}}$ and $X_1^D(N)_{/\mathbb{Q}}$ parametrizing abelian surfaces with quaternionic multiplication (QM) by the indefinite quaternion algebra B over $\mathbb{Q}$ of discriminant D, along with certain specified level structure. Our main result allows for a similar description of the CM loci on these curves.

In particular, we show that if $x \in X_0^D(N)_{/\mathbb{Q}}$ has CM by the order $\mathfrak{o}(f)$ of conductor f in the imaginary quadratic field K, then the residue field $\mathbb{Q}(x)$ is either a ring class field $K(f')$ for some f' with $f \mid f' \mid Nf$, or is isomorphic to an index 2 subfield of such a field $K(f')$. The ramification index of x with respect to the natural map from $X_0^D(N)$ to $X_0^D(1)$ is always 1 when the CM order has discriminant $f^2\Delta_K = \Delta < -4$. In general, this index is at most 3. The paper culminates in a determination of the residue fields and ramification indices of all CM points on $X_0^D(N)$, and putting together the casework based on the quaternion discriminant, level and CM order gives a result of the following form.

Theorem 1.1. *There exists an algorithm which, given as input*

- *an indefinite quaternion discriminant D over $\mathbb{Q}$,*
- *a positive integer N coprime to D and*
- *an imaginary quadratic discriminant $\Delta = f^2\Delta_K$,*

returns as output the complete list of tuples ($\mathtt{is\_fixed}$, f', e, c), *consisting of*

- *a boolean* $\mathtt{is\_fixed}$,
- *a positive integer f' (necessarily with $f \mid f'$),*
- *an integer $e \in \{1, 2, 3\}$ and*
- *a positive integer c*

such that there exist exactly c closed $\mathfrak{o}$-CM points x on $X_0^D(N)_{/\mathbb{Q}}$ with the properties

- *the residue field of x over K is $K(x) \cong K(f')$, the ring class field of conductor f' associated to K,*
- *$\mathbb{Q}(x) \cong K(f')$ if* $\mathtt{is\_fixed}$ *is False,*
- *$[K(f') : \mathbb{Q}(x)] = 2$ if* $\mathtt{is\_fixed}$ *is True and*
- *x has ramification index e with respect to the natural map to $X_0^D(1)_{/\mathbb{Q}}$.*

This algorithm, outlined in Algorithm 8.2, has been implemented, and is publicly available at [Saia 2024], as is code for all other computations described. All computations were performed in Magma [Bosma et al. 1997].

The outline towards developing this algorithm is as follows: in Section 2 we provide relevant background and prior results on CM points on the Shimura curves of interest. This includes results on concrete decompositions of QM abelian surfaces with CM as products of CM elliptic curves. The main result here is Theorem 2.13. In Sections 3 and 4, we then consider QM-equivariant isogenies and the QM-equivariant ℓ-isogeny graph $\mathcal{G}_\ell^D$. We prove in Theorem 4.5 that a CM component of this graph for a prime ℓ and quaternion discriminant D has the structure of an ℓ-volcano for CM discriminant $\Delta < -4$. We handle the slight deviation from the structure of an ℓ-volcano in the $\Delta \in \{-3, -4\}$ case in Proposition 5.3.

We study the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on such components in Section 5, allowing for an enumeration of closed-point equivalence classes of paths in these graphs and hence a description the CM locus on a prime-power level Shimura curve $X_0^D(\ell^a)_{/\mathbb{Q}}$ as provided in Section 6. The algebraic results of Section 7 then feed into a description of the CM locus on $X_0^D(N)_{/\mathbb{Q}}$ for general level N coprime to D provided in Section 8, which provides the algorithm mentioned in Theorem 1.1.

The ability to transition to information about the $\mathfrak{o}$-CM locus on $X_1^D(N)_{/\mathbb{Q}}$ is explained in Section 9, in which we prove the following result. While this does not determine the list of residue fields of CM points on $X_1^D(N)$ in the vein of Theorem 1.1, it allows us to count all CM points on $X_1^D(N)$ of specified degree and list their corresponding CM orders. Otherwise put, this is enough data to determine, for a fixed discriminant D and degree d, all levels N such that there exists a QM abelian surface (A, ι) and a torsion point $P \in A(\overline{\mathbb{Q}})$ of order N such that the induced point $[A, \iota, P] \in X_1^D(N)$ has residue field of degree d.

Theorem 1.2. *Suppose that $x \in X_0^D(N)_{/\mathbb{Q}}$ is a point with CM by the imaginary quadratic order of discriminant Δ. Let $\pi_1 : X_1^D(N)_{/\mathbb{Q}} \to X_0^D(N)_{/\mathbb{Q}}$ and π_0 : $X_0^D(N)_{/\mathbb{Q}}$ denote the natural morphisms. Then*:

(1) *The scheme-theoretic fiber of π_1 over x consists of a single closed point.*

(2) *The map π_1 is unramified over x if any of the following hold*:

- $\Delta < -4$,
- *x is ramified with respect to π_0 or*
- $N \leq 3$.

(3) *If $N \geq 4$ and x is unramified with respect to π_0, then, in the $\Delta \in \{-3, -4\}$ case,*

$$e_{\pi_1}(x) = \begin{cases} 2 & \text{if } \Delta = -4, \\ 3 & \text{if } \Delta = -3, \end{cases} \qquad \text{and} \qquad f_{\pi_1}(x) = \begin{cases} \phi(N)/4 & \text{if } \Delta = -4, \\ \phi(N)/6 & \text{if } \Delta = -3, \end{cases}$$

are the ramification index and residue degree of x, respectively, with respect to π_1.

We define a *primitive residue field* (respectively, a *primitive degree*) *of an* $\mathfrak{o}$*-CM point on* $X_0^D(N)_{/\mathbb{Q}}$ to be one that does not properly contain (respectively, does not properly divide) that of another $\mathfrak{o}$-CM point on the same curve. Our work allows for a determination of all primitive residue fields and primitive degrees of $\mathfrak{o}$-CM points on $X_0^D(N)_{/\mathbb{Q}}$, as discussed in Section 8.4. An abridged version of our main result on primitive residue fields and degrees is as follows, with Theorem 8.3 providing the complete result:

Theorem 1.3. *Suppose that K splits B, let f be a positive integer, and let N be a positive integer relatively prime to D with prime-power factorization $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$. One of the following occurs:*

(1) *There is a unique primitive residue field L of $\mathfrak{o}(f)$-CM points on $X_0^D(N)_{/\mathbb{Q}}$, with L an index 2, totally complex subfield of a ring class field $K(Hf)$ for some $H \mid N$.*

(2) *There are exactly 2 primitive residue fields of such points, with one of the same form as L in part (1) and the other being a ring class field of the form $K(Cf)$ with $C < H$ and $C \mid N$.*

Knowledge of all primitive degrees provides the ability to compute the *least* degree $d_{\mathfrak{o},\mathrm{CM}}(X_0^D(N))$ of an $\mathfrak{o}$-CM point on $X_0^D(N)_{/\mathbb{Q}}$ for any imaginary quadratic order $\mathfrak{o}$. In Section 10, we discuss minimizing over orders $\mathfrak{o}$ to compute the least degree $d_{\mathrm{CM}}(X_0^D(N))$ of a CM point on $X_0^D(N)_{/\mathbb{Q}}$, and Proposition 10.1 allows one to transition from this to computations of least degrees of CM points on $X_1^D(N)_{/\mathbb{Q}}$.

A closed point x on a curve $X_{/\mathbb{Q}}$ is said to be *sporadic* if there are finitely many points y on $X_{/\mathbb{Q}}$ with $\deg(y) \leq \deg(x)$. We apply our least degree computations towards the existence of sporadic CM points on $X_0^D(N)_{/\mathbb{Q}}$ with the following end result (see Theorem 10.9).

Theorem 1.4. *Let $\mathcal{F}$ be the set of all 393 pairs (D, N) appearing in Table 1 or Table 2. If $(D, N) \notin \mathcal{F}$ consists of a quaternion discriminant $D > 1$ over $\mathbb{Q}$ and a positive integer N which is relatively prime to D, then $X_0^D(N)_{/\mathbb{Q}}$ has a sporadic CM point. If (D, N) is such a pair with*

$$(D, N) \notin \mathcal{F} \cup \{(91, 5)\},$$

then $X_1^D(N)_{/\mathbb{Q}}$ has a sporadic CM point.

The appearance of the pair (91, 5) in this result comes down to the fact that while $X_0^{91}(5)_{/\mathbb{Q}}$ has a sporadic CM point of degree 2, the curve $X_1^{91}(5)_{/\mathbb{Q}}$ has 4 as the least degree of a CM point. See Theorem 10.9(4) for details.

Our work determining residue fields of CM points on $X_0^D(N)_{/\mathbb{Q}}$ can be viewed as a generalization of prior work on the Diophantine arithmetic of Shimura curves via an alternate approach (specifically work of Jordan [1981] and González and

Rotger [2006] — see Theorem 2.8). Of course, our results are aimed towards better understanding the torsion of low-dimensional abelian varieties over number fields, via restriction to a case with extra structure. On this point, the question of which number fields admit abelian surfaces with certain specified rational torsion subgroups is closely related to our results, just as in the classical modular curve case. A result of Jordan (see Theorem 2.6) clarifies this relationship.

Unlike the modular curves $X_0(N)_{/\mathbb{Q}}$, the curves $X_0^D(N)_{/\mathbb{Q}}$ for $D > 1$ have no cusps. For this reason, understanding the CM points on Shimura curves may be of even greater interest, as they provide the most accessible examples of low-degree points and could afford techniques (see, e.g., [Bayer and Travesa 2007]) for computing models in the absence of techniques involving expansions around cusps.

Additionally, while our approach is in analogy to that of [Clark 2022] and [Clark and Saia 2022] in the modular curve case, there are interesting deviations arising in this work due to technical differences in the $D > 1$ case. Namely, while the field of moduli $\mathbb{Q}(x)$ of any CM point $x \in X(1)_{/\mathbb{Q}}$ has a real embedding, a result of Shimura [1975, Theorem 0] states that $X^D(1)_{/\mathbb{Q}}$ has no real points for $D > 1$. This fact also opens the door for the potential of Hasse principle violations by Shimura curves, which has been a subject of significant study (see, e.g., [Clark 2009; Clark and Stankewicz 2018; Rotger et al. 2005; Siksek and Skorobogatov 2003]). If one aims to study the Hasse principle for Shimura curves over some fixed number field (respectively, over a fixed degree), then studying the CM points rational over that field (respectively, over number fields of that degree) seems to be a natural initial point of investigation, and so our results may be of interest in that direction.

2. Background

2.1. ***Shimura curves.*** The main source here is the foundational work of Shimura [1967], while for the background material on quaternion algebras and quaternion orders we recommend the classic [Vignéras 1980] as well as the modern treatment in [Voight 2021]. Throughout, we let $B/\mathbb{Q}$ denote the indefinite quaternion algebra of discriminant D over $\mathbb{Q}$. We denote by Ψ an isomorphism

$$\Psi : B \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow{\sim} M_2(\mathbb{R}).$$

As B is indefinite, the discriminant D is the product of an even number of distinct rational primes, namely those at which B is ramified. We will let $\mathcal{O}$ denote a maximal order in B, which is unique up to conjugation. We will also fix, following [Voight 2021, §43.1], an element $\mu \in \mathcal{O}$, satisfying $\mu^2 = -D$, which induces the involution

$$\alpha \mapsto \alpha^* := \mu^{-1}\alpha\mu$$

on $\mathcal{O}$. We refer to μ as a *principal polarization* on $\mathcal{O}$.

We start by defining the moduli spaces we are considering and discussing the moduli interpretations of those families of particular interest to us in this study. Let $\mathcal{O}^1$ denote the units of reduced norm 1 in $\mathcal{O}$, which we realize as embedded in $\mathrm{SL}_2(\mathbb{R})$ via Ψ. The subgroup $\Gamma^D(1) := \Psi(\mathcal{O}^1) \subset \mathrm{SL}_2(\mathbb{R})$ is discrete, and it is cocompact if and only if $D > 1$. Via the action of this subgroup on the upper-half plane $\mathbb{H}$ we define over $\mathbb{C}$ the Shimura curve

$$X^D(1) := \Gamma^D(1)\backslash\mathbb{H}.$$

For $D = 1$ we have $B \cong M_2(\mathbb{Q})$, which recovers the familiar modular curve setting. We are interested in the $D > 1$ case, and so moving forward we make this assumption on D. This implies that $X^D(1)$ is a compact Riemann surface. For any $z \in \mathbb{H}$, we get a rank-4 lattice Λ_z via the action of $\mathcal{O}$ on $(z, 1) \in \mathbb{C}^2$ via the embedding Ψ above:

$$\Lambda_z := \mathcal{O} \cdot \left(\begin{smallmatrix} z \\ 1 \end{smallmatrix}\right) \subseteq \mathbb{C}^2.$$

From this we obtain a complex torus

$$A_z := \mathbb{C}^2/\Lambda_z$$

of dimension 2, which comes equipped with an $\mathcal{O}$-action $\iota_z : \mathcal{O} \hookrightarrow \mathrm{End}(A_z)$. We require some rigidification data, namely a Riemann form, in order to recognize A_z as an abelian surface. It turns out that we always obtain such data in this setting [Voight 2021, Lemma 43.6.23]; there is a *unique* principal polarization $\lambda_{z,\mu}$ on A_z such that the Rosati involution on $\mathrm{End}^0(A) := \mathrm{End}(A) \otimes \mathbb{Q}$ agrees with the involution induced by the polarization μ on $\Psi(\mathcal{O})$.

Definition 2.1. An $(\mathcal{O}, \mu)$-*QM abelian surface* over F is a triple (A, ι, λ) consisting of an abelian surface A over F, an embedding $\iota : \mathcal{O} \hookrightarrow \mathrm{End}(A)$ which we will refer to as the *quaternionic multiplication* (*QM*) *structure*, and a polarization λ on A such that the following diagram is commutative:

$$\begin{array}{ccc} B & \xrightarrow{\ \iota\ } & \mathrm{End}^0(A) \\ \downarrow{\scriptstyle *} & & \downarrow{\scriptstyle \dagger} \\ B & \xrightarrow{\ \iota\ } & \mathrm{End}^0(A) \end{array}$$

where $\dagger$ denotes the Rosati involution corresponding to λ. An *isomorphism of QM abelian surfaces* (A, ι, λ) and (A', ι', λ') is an isomorphism $f : A \to A'$ of abelian surfaces such that $f \circ \iota = \iota' \circ f$ and such that $f^*\lambda' = \lambda$.

With this definition, we have [Voight 2021, Main Theorem 43.6.14] that $X^D(1)$ is the coarse moduli space of $(\mathcal{O}, \mu)$-QM abelian surfaces over $\mathbb{C}$, with the association $z \mapsto [(A_z, \iota_z, \lambda_{z,\mu})]$.

Remark 2.2. For an abelian variety A over a field F, by $\mathrm{End}(A)$ we mean the ring of endomorphisms defined over F. For an extension $F \subseteq L$, we will write A_L for the base change of A to L and $\mathrm{End}(A_L)$ for the ring of endomorphisms rational over L.

More generally, if $\Gamma \leq \Gamma^D(1) \subseteq \mathrm{SL}_2(\mathbb{R})$ is an arithmetic Fuchsian group, we can consider the curve $\Gamma\backslash\mathbb{H}$, and for $\Gamma' \leq \Gamma$ there is a corresponding covering of curves $\Gamma'\backslash\mathbb{H} \to \Gamma\backslash\mathbb{H}$. Our focus will be on the families of Shimura curves $X_0^D(N)$ and $X_1^D(N)$, for N a positive integer with $\gcd(D, N) = 1$, with $X^D(1) = X_0^D(1) = X_1^D(1)$ being a special case of each.

With setup following the careful exposition of [Buzzard 1997, §1], let

$$R := \varprojlim_{\gcd(m,D)=1} \mathbb{Z}/m\mathbb{Z}$$

and fix an isomorphism $\kappa : B \otimes_{\mathbb{Z}} R \to M_2(R)$. This map κ induces, for m relatively prime to D, a map

$$\mathcal{O} \otimes \widehat{\mathbb{Z}} \to M_2(\mathbb{Z}_m).$$

We get from here a map

$$u_m : \mathcal{O}^1 \to \mathrm{GL}_2(\mathbb{Z}_m).$$

The curve $X_0^D(N)$ can then be described as the Shimura curve corresponding to the compact, open subgroup

$$\Gamma_0^D(N) := \Psi\Big(u_N^{-1}\Big(\big\{\big(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\big) \in \mathrm{GL}_2(\mathbb{Z}_N) \mid c \equiv 0 \pmod{N}\big\}\Big)\Big) \leq \Gamma^D(1).$$

That is, $X_0^D(N)(\mathbb{C}) = \Gamma_0^D(N)\backslash\mathbb{H}$. Equivalently, fixing a level N Eichler order $\mathcal{O}_N$ in B, the curve $X_0(N)$ can be described, in the manner mentioned above, as that associated to the arithmetic group of units of reduced norm 1 in $\mathcal{O}_N$. The Shimura curve $X_1^D(N)$ corresponds to the compact, open subgroup

$$\Gamma_1^D(N) := \Psi\Big(u_N^{-1}\Big(\big\{\big(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\big) \in \mathrm{GL}_2(\mathbb{Z}_N) \mid c \equiv 0 \text{ and } d \equiv 1 \pmod{N}\big\}\Big)\Big) \leq \Gamma^D(1).$$

It follows from a celebrated result of Shimura [1967, Main Theorem I] that the curve $X_0^D(N)$ has a canonical model $X_0^D(N)_{/\mathbb{Q}}$, i.e., such that

$$X_0^D(N)_{/\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{C} \cong X_0^D(N),$$

and similarly for the curve $X_1^D(N)$.

Because we are assuming that N is relatively prime to D, the notion of "level N structure" is group-theoretically just as in the modular curve case. In particular, the natural modular map $X_1^D(N)_{/\mathbb{Q}} \to X_0^D(N)_{/\mathbb{Q}}$ is a $(\mathbb{Z}/N\mathbb{Z})^\times/\{\pm 1\}$-cover. Hence, it is an isomorphism for $N \leq 2$ and it has degree $\phi(N)/2$ for $N \geq 3$, where ϕ denotes the Euler totient function. We now recall moduli interpretations for these families of Shimura curves as provided in, for example, [Buzzard 1997, §3].

Definition 2.3. Suppose that (A, ι, λ) and (A', ι', λ') are $(\mathcal{O}, \mu)$-QM abelian surfaces over F. We will call an isogeny $\varphi : A \to A'$ of the underlying abelian surfaces a *QM-cyclic N-isogeny* if $\varphi^*(\lambda') = \lambda$ and both of the following hold:

- The isogeny φ is QM-equivariant. That is, for all $\alpha \in \mathcal{O}$ we have
$$\iota'(\alpha) \circ \varphi = \varphi \circ \iota(\alpha).$$
- The kernel $\ker(\varphi)$ is a cyclic $\mathcal{O}$-module with
$$\ker(\varphi) \cong \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}.$$

For example, a QM-cyclic 1-isogeny is the same as an isomorphism of QM abelian surfaces.

Proposition 2.4. *The Shimura curve $X_0^D(N)_{/\mathbb{Q}}$ is isomorphic to the coarse moduli scheme associated to any of the following moduli problems*:

(1) *Tuples (A, ι, λ, Q), where (A, ι, λ) is an $(\mathcal{O}, \mu)$-QM abelian surface and $Q \le A[N]$ is an order N^2 subgroup of the N-torsion subgroup of A which is also a cyclic $\mathcal{O}$-module.*

(2) *QM-cyclic N-isogenies $\varphi : (A, \iota, \lambda) \to (A, \iota', \lambda')$ of $(\mathcal{O}, \mu)$-QM abelian surfaces.*

The curve $X_1^D(N)_{/\mathbb{Q}}$ has the following moduli interpretation: *triples (A, ι, λ, P), where (A, ι, λ) is a QM abelian surface and $P \in A[N]$ is a point of order N.*

These interpretations hold for any choice of principal polarization μ of $\mathcal{O}$. That is, if μ and μ' are two such polarizations then they both induce the same coarse moduli scheme $X_0^D(N)_{/\mathbb{Q}}$ up to isomorphism (as discussed, for example, in [Rotger 2004, §6]). Of course, the exact moduli interpretation *does* depend on μ, and we refer to [Rotger 2004, Proposition 4.3] for more on how the corresponding spaces fit into the moduli space of principally polarized abelian surfaces. Because a principal polarization λ on a pair (A, ι) is canonically determined from a fixed μ, moving forward we will suppress polarizations and refer simply to QM abelian surfaces (A, ι). By the same point, the condition on the polarizations in the definition of a QM-cyclic N-isogeny is redundant; it follows from the QM-equivariant condition.

Letting $\mathcal{O}_N$ denote an Eichler order of level N in B, the curve $X_0^D(N)_{/\mathbb{Q}}$ has the equivalent interpretation of parametrizing pairs (A, ι) where $A/\mathbb{C}$ is a QM abelian surface and $\iota : \mathcal{O}_N \hookrightarrow \mathrm{End}(A)$. (We just stated that we would no longer remark on polarizations, but we note that the polarization corresponding to such an ι will not be principally polarized, but $(1, N)$-polarized in general.) That said, interpretations (1) and (2) in Proposition 2.4 will be the primary ones used in our study — see Remark 2.9 for related comments. Thus, we will mainly speak of QM by maximal quaternion orders, and it will benefit us to spell out the connection between interpretations (1) and (2) here. Let (A, ι) be a QM abelian surface. The N-torsion of A is acted on by $\iota(\mathcal{O})$, and the corresponding representation factors through $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}/N\mathbb{Z} \cong M_2(\mathbb{Z}/N\mathbb{Z})$. The resulting map must then be equivalent to

$$M_2(\mathbb{Z}/N\mathbb{Z}) \to \mathrm{End}(A[N]) \cong M_4(\mathbb{Z}/N\mathbb{Z}), \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{pmatrix}.$$

This can be viewed as a case of Morita equivalence, but it is worth being explicit here: let e_1 and e_2 denote the standard idempotents in $M_2(\mathbb{Z}/N\mathbb{Z})$,

$$e_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad e_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We then have $A[N] = e_1 \cdot A[N] \oplus e_2 \cdot A[N]$, and $M_2(\mathbb{Z}/N\mathbb{Z})$ acts on this direct sum in precisely the way noted by the above map.

Any proper, nontrivial, $\mathcal{O}$-stable subgroup $Q \leq A[N]$ must then have order N^2 (this justifies our definition of QM-cyclic isogenies, along with the equivalence of the moduli interpretations presented above). Further, such a subgroup Q is determined by a cyclic order N subgroup of $A[N]$: we have $Q = e_1(Q) \oplus e_2(Q)$ where each summand is cyclic of order N, and conversely $Q = \mathcal{O} \cdot e_i(Q)$ for $i = 1, 2$.

For our applications in Section 10, the genera of our Shimura curves of interest will be of use. Let ψ denote the Dedekind psi function. The derivations are standard — for example, the formula for $X_0^D(N)$ can be found in [Voight 2021, Theorem 39.4.20]:

Proposition 2.5. *We have*

$$g(X_0^D(N)) = 1 + \frac{\phi(D)\psi(N)}{12} - \frac{\epsilon_1(D, N)}{4} - \frac{\epsilon_3(D, N)}{3},$$

where

$$\epsilon_1(D, N) = \begin{cases} \prod_{p \mid D} \left(1 - \left(\frac{-4}{p}\right)\right) \prod_{p \mid N} \left(1 + \left(\frac{-4}{p}\right)\right) & \text{if } 4 \nmid N, \\ 0 & \text{if } 4 \mid N, \end{cases}$$

$$\epsilon_3(D, N) = \begin{cases} \prod_{p \mid D} \left(1 - \left(\frac{-3}{p}\right)\right) \prod_{p \mid N} \left(1 + \left(\frac{-3}{p}\right)\right) & \text{if } 9 \nmid N, \\ 0 & \text{if } 9 \mid N, \end{cases}$$

are, respectively, the numbers of elliptic $\mathbb{Z}[\sqrt{-1}]$*-CM and elliptic* $\mathbb{Z}[\frac{1+\sqrt{-3}}{2}]$*-CM points on* $X_0^D(N)$*. For* $N \leq 2$ *we have*

$$X_1^D(N) \cong X_0^D(N),$$

and for $N \geq 3$ *we have*

$$g(X_1^D(N)) = 1 + \frac{\phi(N)\phi(D)\psi(N)}{24}.$$

2.2. ***CM points.*** Let (A, ι) be a QM abelian surface over a number field F, such that

$$\mathrm{End}^0(A) \cong B.$$

If A is nonsimple, such that $A \sim E_1 \times E_2$ is geometrically isogenous (i.e., isogenous over $\overline{\mathbb{Q}}$) to a product of elliptic curves, then it must be the case that E_1 and E_2 are isogenous elliptic curves with complex multiplication (CM). In this case, $A \sim E^2$ where E is a CM elliptic curve, say with corresponding imaginary quadratic CM field K. Here it is forced that K splits the quaternion algebra B:

$$B \otimes_{\mathbb{Q}} K \cong M_2(K).$$

In this case in which A is nonsimple, we refer to (A, ι) as a *QM abelian surface with CM* and we call the induced point $[(A, \iota)] \in X^D(1)_{/\mathbb{Q}}(F)$ a *CM point*. We call a point x on $X_0^D(D)_{/\mathbb{Q}}$ or $X_1^D(N)_{/\mathbb{Q}}$ a CM point if it lies over a CM point on $X^D(1)_{/\mathbb{Q}}$.

Generalizing our definition for isogenies, we call an endomorphism $\alpha \in \mathrm{End}(A)$ *QM-equivariant* if $\alpha \circ \iota(\gamma) = \iota(\gamma) \circ \alpha$ for all $\gamma \in \mathcal{O}$. If (A, ι) has K-CM, then the ring $\mathrm{End}_{\mathrm{QM}}(A)$ of QM-equivariant endomorphisms of A is an imaginary quadratic order in K. This means that we have some $f \in \mathbb{Z}^+$ such that

$$\mathrm{End}_{\mathrm{QM}}(A) \cong \mathfrak{o}(f),$$

where $\mathfrak{o}(f)$ denotes the unique order of conductor f in K. In other words, $\mathfrak{o}(f)$ is the unique imaginary quadratic order of discriminant $f^2\Delta_K$, where Δ_K denotes the discriminant of K, i.e., that of the maximal order $\mathfrak{o}_K = \mathfrak{o}(1)$. We will call this f the *central conductor* of (A, ι). We will refer to $[(A, \iota)] \in X^D(1)$, or to any point in the fiber over $[A, \iota]$ under some covering of Shimura curves $X \to X^D(1)$, as an $\mathfrak{o}(f)$*-CM point* when we wish to make the CM order clear. Note that the QM on A is by definition defined over F, so if A is isogenous to E^2 over an extension L/F then E necessarily has its CM defined over L.

2.3. ***The field of moduli of a QM-cyclic isogeny.***

2.3.1. *The field of moduli.* The *field of moduli* of a QM abelian surface (A, ι) defined over $\overline{\mathbb{Q}}$ is the fixed field of those automorphisms $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $(A,\iota)^\sigma := (A^\sigma, \iota^\sigma)$ is isomorphic to (A,ι) over $\overline{\mathbb{Q}}$. The conjugate abelian surface A^σ is defined as the fiber product $A \otimes_{\operatorname{Spec}\overline{\mathbb{Q}}} \operatorname{Spec}\overline{\mathbb{Q}}$ over σ:

$$\begin{array}{ccc} A^\sigma & \longrightarrow & A \\ \downarrow & & \downarrow \\ \operatorname{Spec}\overline{\mathbb{Q}} & \xrightarrow{\sigma} & \operatorname{Spec}\overline{\mathbb{Q}} \end{array}$$

and ι^σ is defined via the action of σ on endomorphisms of A. (We are suppressing polarizations at this point, but recall this is justified as there is a unique principal polarization on A^σ compatible with ι^σ.) Equivalently, the field of moduli of (A,ι) is the residue field $\mathbb{Q}(x)$ of the corresponding point $x = [(A,\iota)]$ on $X^D(1)_{/\mathbb{Q}}$.

More generally, for a QM-cyclic isogeny $\varphi : (A,\iota) \to (A',\iota')$ defined over $\overline{\mathbb{Q}}$, the *field of moduli* of φ is the fixed field of the group

$$H(\varphi) := \left\{ \sigma \in \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \;\middle|\; \begin{array}{ccc} (A^\sigma,\iota^\sigma) & \xrightarrow{\varphi^\sigma} & ((A')^\sigma,(\iota')^\sigma) \\ \downarrow & & \downarrow \\ (A,\iota) & \xrightarrow{\varphi} & (A',\iota') \end{array} \text{ commutes, and the vertical maps are isomorphisms} \right\}.$$

For clarity: the vertical maps above are those induced by σ and membership of σ in $H(\varphi)$ means that both (A,ι) and (A',ι') are isomorphic to their conjugates by σ. In other words, the field of moduli of φ is the minimal field over which φ is isomorphic to all of its $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-conjugates. Equivalently, it is the residue field of the corresponding point $[\varphi]$ on $X_0^D(N)_{/\mathbb{Q}}$ (which follows from the much more general theory of [Shimura 1966, Theorem 5.1], as exposited more specifically towards our case in [Shimura 1967, p. 60]).

We call a field F a *field of definition* for a QM-cyclic isogeny φ as above, or say that φ is defined or rational over F, if φ and both (A,ι) and (A',ι') can be given by equations defined over F. We then have a model φ' over F so that $\varphi' \otimes_F \overline{\mathbb{Q}} = \varphi$. It follows that if $x \in X_0^D(N)_{/\mathbb{Q}}$ is induced by φ, then any field of definition for φ contains the field of moduli $\mathbb{Q}(x)$.

It is not generally the case that fields of moduli are fields of definition for (polarized) abelian varieties of dimension bigger than 1, and this is a source of difficulty and interest in the study of their arithmetic. For instance, Shimura [1972] proved that the generic principally polarized even-dimension abelian variety *does not* have a model defined over its field of moduli. Particular towards our interests here, a QM abelian surface (or, more generally, a QM-cyclic isogeny) need not have a model over its field of moduli. However, we have the following result of Jordan [1981, Theorem 2.1.3]:

Theorem 2.6 (Jordan). *Suppose that $(A, \iota)/\overline{\mathbb{Q}}$ is a QM abelian surface with QM by B and with* $\mathrm{Aut}_{\mathrm{QM}}(A) = \{\pm 1\}$ *(equivalently, (A, ι) does not have CM by $\Delta \in \{-3, -4\}$). Let $x = [(A, \iota)] \in X_0^D(1)_{/\mathbb{Q}}$ be the corresponding point. Then a field L containing $\mathbb{Q}(x)$ is a field of definition for (A, ι) if and only if L splits B.*

2.3.2. *The field of moduli in the CM case.* Our attention in this study will primarily be aimed at determining fields of moduli, particularly in the presence of CM. We now recall prior work determining the field of moduli of a CM point on $X^D(1)_{/\mathbb{Q}}$.

The answer begins with a fundamental theorem of Shimura [1967, Main Theorem 1]. Fixing an imaginary quadratic field K and a positive integer f, we let $\mathfrak{o}(f)$ denote the order in K of conductor f and $K(f)$ denote the ring class field corresponding to $\mathfrak{o}(f)$.

Theorem 2.7 (Shimura). *Let $x \in X^D(1)_{/\mathbb{Q}}$ be an $\mathfrak{o}(f)$-CM point with residue field $\mathbb{Q}(x)$. Then*

$$K \cdot \mathbb{Q}(x) = K(f)$$

This tells us that in this setting there are two possibilities: either $\mathbb{Q}(x)$ is the ring class field $K(f)$, or it is an index 2 subfield thereof. Jordan [1981, §3] proved when each possibility occurs in the case where x has CM by the maximal order of K (the $f = 1$ case). Work of González and Rotger [2006, §5] allows for a generalization of Jordan's result to arbitrary CM orders.

To state their result, we first set the following notation: for D a quaternion discriminant over $\mathbb{Q}$ and K an imaginary quadratic field splitting the quaternion algebra B of discriminant D over $\mathbb{Q}$, let

$$D(K) := \prod_{\substack{p \mid D \\ \left(\frac{K}{p}\right) = -1}} p.$$

The assumption that K splits B is exactly the assumption that no prime divisor of D splits in K. From this we see that $D(K) = 1$ if and only if all primes dividing D ramify in K, while $D(K) > 1$ exactly when some prime dividing D is inert in K.

Theorem 2.8 (Jordan, González–Rotger). *Let $x \in X^D(1)_{/\mathbb{Q}}$ be an $\mathfrak{o}(f)$-CM point.*

(1) *If $D(K) = 1$, then we have $\mathbb{Q}(x) = K(f)$.*

(2) *Otherwise, $[K(f) : \mathbb{Q}(x)] = 2$. In this case, $\mathbb{Q}(x) \subsetneq K(f)$ is the subfield fixed by*

$$\sigma = \tau \circ \sigma_{\mathfrak{a}} \in \mathrm{Gal}(K(f)/\mathbb{Q}),$$

where τ denotes complex conjugation and $\sigma_{\mathfrak{a}} \in \mathrm{Gal}(K(f)/K)$ is the automorphism associated via the Artin map to a certain fractional ideal $\mathfrak{a}$ of $\mathfrak{o}(f)$ with the property *that*

$$B \cong \left(\frac{\Delta_K, N_{K/\mathbb{Q}}(\mathfrak{a})}{\mathbb{Q}}\right).$$

More specifically, $\mathfrak{a}$ *is such that*

$$\omega_{D(K)}(x^{\sigma_{\mathfrak{a}}}) = \tau(x),$$

where $\omega_{D(K)}$ *denotes the Atkin–Lehner involution on* $X^D(1)_{/\mathbb{Q}}$ *corresponding to* $D(K)$.

Remark 2.9. González and Rotger provide a generalization of Jordan's result to all CM points on $X_0^D(N)_{/\mathbb{Q}}$ for squarefree N. We state their result only for trivial level $N = 1$ in part because it is all we will need, but also because some translation would be needed for the statement of their result as in their work to the conventions of this work. In comparing our work to [González and Rotger 2006], the definition of an $\mathfrak{o}$-CM point on $X_0^D(N)_{/\mathbb{Q}}$ that they work with is different from ours; whereas our definition is that a CM point has $\mathfrak{o}$-CM for an imaginary quadratic order $\mathfrak{o}$ if it lies over an $\mathfrak{o}$-CM point on $X^D(1)_{/\mathbb{Q}}$, their definition is that $x \in X_0^D(N)_{/\mathbb{Q}}$ has $\mathfrak{o}$-CM if it corresponds to a normalized optimal embedding of $\mathfrak{o}$ into an Eichler order of level N in B. The definition used in [González and Rotger 2006] provides a pleasantly uniform result similar to Jordan's $N = 1$ case, with every $\mathfrak{o}(f)$-CM point $x \in X_0^D(N)_{/\mathbb{Q}}$ having field of moduli $\mathbb{Q}(x)$ with $K \cdot \mathbb{Q}(x) \cong K(f)$. It will not be the case in our work, for level $N > 1$, that all $\mathfrak{o}$-CM points have the same residue field. While our set of K-CM points on $X_0^D(N)$ is the same as that as defined in [González and Rotger 2006], the specific orders we attach may not agree.

The convention used by González and Rotger is common in the literature, appearing in the work of Rotger and his collaborators and also in work of Padurariu and Schembri [2023] in which the authors compute rational points on all Atkin–Lehner quotients of geometrically hyperelliptic Shimura curves. The difference in convention one takes is motivated by which moduli problem one chooses for the course moduli scheme $X_0^D(N)$: our choice of working with maximal orders results in having natural modular maps from $X_0^D(N)$ to $X_0^D(1)$ for all N, while working with Eichler orders of level N naturally situates $X_0^D(N)$ as the base Shimura curve. Because we want to work with general level, we work with maximal orders. A main difference between our work and that of [González and Rotger 2006], beyond the generalization from squarefree N to all positive integers N, is that we consider not just the CM points on a fixed curve $X_0^D(N)$ but the *fiber* of the covering $X_0^D(N)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$ over any CM point.

2.4. ***Decompositions of QM abelian surfaces with CM.*** Restricting to the case of a QM abelian surface (A, ι) with CM over $\mathbb{C}$, we have seen that A is isogenous to a square of an elliptic curve with CM. Through a correspondence between QM abelian surfaces with CM and equivalence classes of certain binary quadratic forms, Shioda and Mitani [1974, Theorem 4.1] proved the following strengthening of this fact:

Theorem 2.10 (Shioda–Mitani). *If $(A, \iota)/\mathbb{C}$ is a QM abelian surface with K-CM for an imaginary quadratic field K, then there exist K-CM elliptic curves E_1, E_2 over $\mathbb{C}$ such that*

$$A \cong E_1 \times E_2.$$

The number of distinct decompositions of a given A as above is finite, resulting from finiteness of the class number of any imaginary quadratic order in K. This theorem was generalized to higher-dimensional complex abelian varieties isogenous to a power of a CM elliptic curve independently by Katsura [1975, Theorem] and Lange [1975], and Schoen [1992, Satz 2.4] later provided a simple proof as well. A generalization from $\mathbb{C}$ to an arbitrary field of definition F is a result of Kani [2011, Theorem 2]:

Theorem 2.11 (Kani). *If A/F is an abelian variety which is isogenous to E^n over F, where E/F is an elliptic curve with CM over F, then there exist CM elliptic curves $E_1/F, \ldots, E_n/F$ such that we have an isomorphism*

$$A \cong E_1 \times \cdots E_n$$

over the base field F.

Kani [2011, Theorem 67] says more, which is relevant in the case of QM abelian surfaces with CM: fixing a K-CM elliptic curve E/F with endomorphism ring of conductor f_E, there is a bijection between the set of F-isomorphism classes $[E']$ of elliptic curves E' isogenous to E with CM conductor $f_{E'} \mid f_E$, and the set of F-isomorphism classes of abelian surfaces A/F isogenous to E^2 with corresponding central conductor $f_A = f_E$. Explicitly, this bijection sends an F-isomorphism class $[E']$ to the F-isomorphism class $[E \times E']$.

In order to obtain concrete decompositions of QM abelian surfaces with CM, the remaining task is to identify *which* such products of CM elliptic curves have potential quaternionic multiplication (that is, which can be given QM structures), and to further describe the classes of QM abelian surfaces with CM. The following result provides the number of such classes ([Alsina and Bayer 2004, Theorem 6.13] interprets this count as a certain class number, or equivalently as an embedding number, and [Vignéras 1980, Corollary 5.12] provides a formula for these class numbers which we use in the $N = 1$ case).

Proposition 2.12. *Let K be an imaginary quadratic field splitting B, and let $f \in \mathbb{Z}^+$. Let b denote the number of primes dividing D that are inert in K. The number of geometric $\mathfrak{o}(f)$-CM points on $X^D(1)$ is then $2^b \cdot h(\mathfrak{o}(f))$, where $h(\mathfrak{o}(f))$ denotes the class number of the order $\mathfrak{o}(f)$.*

Ufer [2010] touches on this topic of taking QM structures into account. In particular, he proves the following [Ufer 2010, Theorem 2.7.12]: with the notation of Proposition 2.12, there exists a 2^b-to-1 correspondence

$$\{K\text{-CM points in } X^D(1)(\mathbb{C}\} \to \{K\text{-CM elliptic curves over } \mathbb{C}\}/\cong.$$

Based on the proof therein, it seems that Ufer could have said more, and so we do that here with reference to his argument. As above, let b denote the number of primes dividing D which are inert in K.

Theorem 2.13. *Let $(A, \iota)/\mathbb{C}$ be a QM abelian surface with CM by $\mathfrak{o}(f)$. There is then a unique $\mathfrak{o}(f)$-CM curve $E_A/\mathbb{C}$, up to isomorphism, such that*

$$A \cong \mathbb{C}/\mathfrak{o}(f) \times E_A.$$

There is a 2^b-to-1 correspondence

$$\{\mathfrak{o}(f)\text{-CM points on } X^D(1)\} \to \{\mathfrak{o}(f)\text{-CM elliptic curves over } \mathbb{C}\}/\cong$$

sending a point $[(A, \iota)] \in X^D(1)$ to the class of E_A.

Proof. Part (2) of the proof of [Ufer 2010, Theorem 2.7.12] details the construction of a QM-structure by a maximal order $\mathcal{O}$ in B on $E \times E'$ for E and E' both $\mathfrak{o}(f)$-CM elliptic curves. The product $E \times E'$ with the constructed QM structure then corresponds to a CM point on $X^D(1)$ with central conductor f.

Let E, E' be K-CM elliptic curves. Part (3) of Ufer's proof explains that if the abelian surface $E \times E'$ has potential quaternionic multiplication then in fact it has 2^b nonisomorphic QM structures. Put differently but equivalently to therein: let W be the group generated by the Atkin–Lehner involutions ω_p on $X^D(1)$ for $p \mid D$ inert in K. The group $W \times \mathrm{Pic}(\mathfrak{o}(f))$ then acts simply transitively on the set of $\mathfrak{o}(f)$-CM points on $X^D(1)$. If $[(A, \iota)] \in X^D(1)$ is such a point, then the action of any element $w \in W$ leaves $[A]$ unchanged, providing the claim (this is proved by Jordan [1981] in the $f = 1$ case, and extended to the general case by González and Rotger [2006, Proposition 5.6]). By the count of Proposition 2.12, Theorem 2.10 and the fact that $\mathbb{C}/\mathfrak{o}(f) \times E \cong \mathbb{C}/\mathfrak{o}(f) \times E'$ implies $E \cong E'$, the claimed result follows. □

Corollary 2.14. *Let $(A, \iota)/F$ be a QM abelian surface with CM by $\mathfrak{o} \subseteq \mathfrak{o}_K$. Suppose that we have an F-rational isogeny $A \sim E^2$ to the square of an elliptic curve. Fix E_1/F any elliptic curve with $\mathfrak{o}$-CM. There then exists an $\mathfrak{o}$-CM elliptic curve E_2/F, unique up to isomorphism over F, such that $A \cong E_1 \times E_2$ over F.*

Proof. Let f be the central conductor of A (i.e., such that $\mathfrak{o} = \mathfrak{o}(f)$). By Theorem 2.11 and the discussion of Kani's results following this theorem statement, there exists a CM elliptic curve E_2/F, with endomorphism ring of conductor f_{E_2}

satisfying $f_{E_2} \mid f$, such that $A \cong E_1 \times E_2$ over F. This curve E_2 is unique up to isomorphism over F. Base changing this entire picture to $\mathbb{C}$, we have

$$A_{/\mathbb{C}} \cong E_{1/\mathbb{C}} \times E_{2/\mathbb{C}}.$$

Now because $A_{/\mathbb{C}}$ (and hence $E_{1/\mathbb{C}} \times E_{2/\mathbb{C}}$, by transport of structure through our isomorphism) has QM and $E_{1/\mathbb{C}}$ has CM conductor f, Theorem 2.13 implies that $f_{E_2} = f$ as well. □

3. QM-equivariant isogenies

Our goal in the following section will be to determine the residue field of a CM point on $X_0^D(N)_{/\mathbb{Q}}$ for any N coprime to D, generalizing Theorem 2.8. A main component in accomplishing this is the study of the structure of, and the action of automorphisms on, components of certain isogeny graphs. Paths in these graphs of consideration will be in correspondence with isogenies of QM abelian surfaces which commute with their QM structures.

Here, we prove facts about QM-equivariant isogenies needed in the proceeding section. Much of what we do in both this section and the next is in strong analogy to the case of isogenies of elliptic curves over $\overline{\mathbb{Q}}$ studied in work of Clark [2022] and Clark and Saia [2022]. We provide proofs here for completeness and for clarity of said analogy.

Lemma 3.1. *Let F be a field of characteristic zero, and let (A, ι) be a QM abelian surface over F which does not have CM by an order of discriminant $\Delta \in \{-3, -4\}$. For ℓ a prime number, the number of QM-cyclic ℓ-isogenies with domain (A, ι) which are Gal$(\overline{F}/F)$-stable, up to isomorphism, is either* $0, 1, 2,$ *or* $\ell + 1$.

Proof. Note that ℓ being prime means we are counting isomorphism classes of QM-cyclic ℓ-isogenies. The hypotheses on A are equivalent to $\operatorname{Aut}(A, \iota) = \{\pm 1\}$. In this case, we have a bijective correspondence between isomorphism classes of QM-cyclic ℓ-isogenies and nontrivial, proper cyclic $\mathcal{O}$-submodules of $A[\ell]$. Under this correspondence, the isogenies which are Gal$(\overline{F}/F)$-stable correspond to Gal$(\overline{F}/F)$-stable submodules.

Now we have that $e_1(Q) \leq e_1(A[\ell]) \cong (\mathbb{Z}/\ell\mathbb{Z})^2$ is a cyclic subgroup of order ℓ, and in this way we have a bijective correspondence between the nontrivial proper QM-stable subgroups of $A[\ell]$ and cyclic order ℓ subgroups of $e_1(A[\ell])$. This correspondence preserves the property of being Gal$(\overline{F}/F)$-stable. We have thus reduced to the situation of the elliptic curve case, and may proceed as such: We are counting Gal$(\overline{F}/F)$-stable cyclic order ℓ subgroups of $(\mathbb{Z}/\ell\mathbb{Z})^2$. The total number of cyclic order ℓ subgroups is $\ell + 1$, and if more than 2 such subgroups are fixed then Gal$(\overline{F}/F)$ is forced to act by scalar matrices on $(\mathbb{Z}/\ell\mathbb{Z})^2$. □

3.1. ***Compositions of QM-cyclic isogenies.*** The following result is in analogy with [Clark 2022, Proposition 3.2].

Proposition 3.2. *Suppose that $\varphi = \varphi_2 \circ \varphi_1$ is a QM-cyclic isogeny, where $\varphi_i : (A_i, \iota_i) \to (A_{i+1}, \iota_{i+1})$ is a QM-cyclic isogeny for $i = 1, 2$.*

(1) *We have*

$$\mathbb{Q}(\varphi) \subseteq \mathbb{Q}(\varphi_1) \cdot \mathbb{Q}(\varphi_2).$$

(2) *If (A_2, ι_2) does not have CM by $\Delta \in \{-3, -4\}$, then*

$$\mathbb{Q}(\varphi) = \mathbb{Q}(\varphi_1) \cdot \mathbb{Q}(\varphi_2).$$

Proof. The containment of part (1) is clear. The assumption that (A_2, ι_2) does not have -3 or -4 CM is equivalent to $\mathrm{Aut}((A_2, \iota_2)) = \{\pm 1\}$, and in this case the reverse containment in part (2) follows by the same argument as in [Clark 2022, Proposition 3.2]. □

3.2. ***Reduction to prime power degrees.*** First, let us say something about rationality. Let $\varphi : (A, \iota) \to (A', \iota')$ be a QM-cyclic N-isogeny which is rational over F, where N has prime-power decomposition $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$. Letting $Q = \ker(\varphi)$ be the kernel of this isogeny, we have that φ is isomorphic to the quotient $(A, \iota) \to (A/Q, \iota)$. (The latter pair indeed provides an $\mathcal{O}$-QM abelian surface, as Q is stable under $\iota(\mathcal{O})$ and $\mathcal{O}$ is maximal, though we are abusing notation by referring to the QM-structure on the quotient as ι.) We have a decomposition $Q = C \oplus D$ with each of C and D cyclic of order N, such that $\mathcal{O} \cdot C = \mathcal{O} \cdot D = Q$. This cyclic subgroup C then decomposes as

$$C = \bigoplus_{i=1}^{r} C_i,$$

where $C_i \leq C$ is the unique subgroup of order $\ell_i^{a_i}$. Letting $Q_i = \mathcal{O} \cdot C_i$, each Q_i is QM stable and isomorphic to $(\mathbb{Z}/{\ell_i}^{a_i}\mathbb{Z})^2$.

From the uniqueness of $C_i \leq C$, and hence of the corresponding $\mathcal{O}$-cyclic subgroup $Q_i \leq Q$, we get that each Q_i is F-rational, resulting in F-rational QM-cyclic $\ell_i^{a_i}$-isogenies $\varphi_i : (A, \iota) \to (A/Q_i, \iota)$ for each i. On the other hand, given a collection of F-rational QM-cyclic $\ell_i^{a_i}$-isogenies with kernels Q_i, we get an F-rational QM-cyclic N-isogeny $(A, \iota) \to (A/Q, \iota)$ where $Q = \bigoplus_{i=1}^{r} Q_i$.

As for fields of moduli, more towards our needs for the following section, we have the following:

Proposition 3.3. *Let $N_1, \ldots, N_r \in \mathbb{Z}^+$ be pairwise coprime, let k be a field of characteristic zero, and let $x \in X^D(1)_{/k}$ be a closed point which does not have CM by discriminant $\Delta \in \{-3, -4\}$. For each i, let $\pi_i : X_0^D(N_i)_{/k} \to X^D(1)_{/k}$ be the* *natural map, let* $F_i = \pi_i^{-1}(x)$, *and let* F *be the fiber over* x *of* $\pi : X_0^D(N)_{/k} \to X^D(1)_{/k}$ *where* $N = N_1 \cdots N_r$. *Then*

$$F = F_1 \otimes_{\operatorname{Spec} k(x)} \cdots \otimes_{\operatorname{Spec} k(x)} F_r.$$

Proof. This follows as in the $D = 1$ case of [Clark 2022, Proposition 3.5], using that $X_0^D(N)$ for $D > 1$ is a cover of $X^D(1)$ with the same corresponding subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$ as in the case of $X_0(N) \to X(1)$. □

It follows that if $x \in X_0^D(N)_{/\mathbb{Q}}$ is a point which does not have -3 or -4-CM and $N = \prod_{i=1}^r \ell_i^{a_i}$, with $\pi_i : X_0^D(N)_{/\mathbb{Q}} \to X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$ the natural maps, then

$$\mathbb{Q}(x) = \mathbb{Q}(\pi_1(x)) \cdots \mathbb{Q}(\pi_r(x)).$$

4. QM-equivariant isogeny volcanoes

Fixing a prime ℓ, we describe CM components of ℓ-isogeny graphs of QM abelian surfaces over $\overline{\mathbb{Q}}$. We will use this work to study CM points on the curves $X_0^D(\ell^a)_{/\mathbb{Q}}$ for $a \in \mathbb{Z}^+$ and $D > 1$, in analogy to the $D = 1$ modular curve case of [Clark 2022; Clark and Saia 2022].

This study, like that of [Clark 2022; Clark and Saia 2022], is motivated by the foundational work on isogeny volcanoes over finite fields by Kohel [1996] and by Fouquet [2001] and Fouquet and Morain [2002]. We also recommend, and will refer to, a more recent, expository account of isogeny volcanoes in the finite field setting by Sutherland [2013].

4.1. ***The isogeny graph of QM abelian surfaces.*** Fix a prime number ℓ and an imaginary quadratic field K. In [Clark 2022] and [Clark and Saia 2022], the authors consider the multigraph with vertex set that of j-invariants of K-CM elliptic curves, and with edges corresponding to $\mathbb{C}$-isomorphism classes of cyclic ℓ-isogenies.

Here, we seek an analog for abelian surfaces with QM by a fixed maximal order $\mathcal{O}$ of the indefinite quaternion algebra B of discriminant D over $\mathbb{Q}$, with $\ell \nmid D$. We let $\mathcal{G}_\ell^D$ denote the directed multigraph with

- vertex set consisting of $\mathbb{C}$-isomorphism classes of $\mathcal{O}$-QM abelian surfaces, and
- edges from $v_1 = [(A_1, \iota_1)]$ to $v_2 = [(A_2, \iota_2)]$ corresponding to $\mathbb{C}$-isomorphism classes of QM-cyclic ℓ-isogenies $\varphi : (A_1, \iota_1) \to (A_2, \iota_2)$.

A given vertex v has $\ell + 1$ edges emanating from it, via the correspondence of QM-stable subgroups of $A_1[\ell]$ with cyclic order ℓ subgroups of $e_1(A_1[\ell]) \cong (\mathbb{Z}/\ell\mathbb{Z})^2$ discussed in Lemma 3.1.

Because a QM structure ι determines a unique principal polarization, we have dual edges via dual isogenies as in the elliptic curve case. As long as the source vertex v_1 corresponds to an isomorphism class $[(A, \iota)]$ having only the single nontrivial automorphism $[-1]$, we obtain a bijection between the edges from v_1 to v_2 and those from v_2 to v_1; in this case, outward edges from v_1 are in bijective correspondence with QM-stable subgroups of $A_1[\ell]$ of order ℓ^2. This occurs precisely when $[(A, \iota)]$ does not have CM by discriminant $\Delta = -3$ or $\Delta = -4$.

Our attention will be to vertices in $\mathcal{G}_\ell^D$ corresponding to QM abelian surfaces with CM. For an abelian variety (A, ι) with QM by $\mathcal{O}$ and K-CM, recall from Section 2.2 that the central conductor of (A, ι) is defined to be the positive integer f such that $\mathrm{End}_{\mathrm{QM}}(A) \cong \mathfrak{o}(f) \subseteq \mathfrak{o}_K$.

Lemma 4.1. *Suppose $\varphi : (A, \iota) \to (A', \iota')$ is a QM-cyclic N-isogeny, with (A, ι) a QM abelian surface with K-CM. Then:*

(1) *The QM abelian surface (A', ι') also has K-CM.*

(2) *Let f and f' denote the central conductors of (A, ι) and (A', ι'), respectively. Then f and f' differ by at most a factor of N:*

$$f \mid Nf' \quad \textit{and} \quad f' \mid Nf.$$

Proof. The argument is similar to that of the elliptic curve case. In our context, we need only remember that we care specifically about those endomorphisms commuting with the QM.

Consider the homomorphism

$$F : \mathrm{End}(A, \iota) \to \mathrm{End}(A', \iota'), \qquad \psi \mapsto \varphi \circ \psi \circ \widehat{\varphi}.$$

Because φ is assumed to be QM-equivariant, this restricts to a homomorphism

$$\mathrm{End}_{\mathrm{QM}}(A, \iota) \to \mathrm{End}_{\mathrm{QM}}(A', \iota').$$

As in the argument in the elliptic curves case, the algebras of endomorphisms commuting with the quaternionic multiplication are isomorphic by the multiple $\frac{1}{N}F$ of the map above. That is,

$$K \cong \mathrm{End}_{\mathrm{QM}}(A, \iota) \otimes \mathbb{Q} \cong \mathrm{End}_{\mathrm{QM}}(A', \iota') \otimes \mathbb{Q}.$$

This completes part (a). Moreover, that

$$\frac{1}{N}F : \mathrm{End}_{\mathrm{QM}}(A, \iota) \otimes \mathbb{Q} \to \mathrm{End}_{\mathrm{QM}}(A', \iota')$$

is an isomorphism tells us that

$$N \cdot \mathrm{End}_{\mathrm{QM}}(A, \iota) \subseteq \mathrm{End}_{\mathrm{QM}}(A', \iota'),$$

yielding $f' \mid Nf$. Via the dual argument, we obtain $f \mid Nf'$. □

For an imaginary quadratic field K, we are therefore justified in defining $\mathcal{G}^D_{K,\ell}$ to be the subgraph of $\mathcal{G}^D_\ell$ consisting of vertices corresponding to QM abelian surfaces with K-CM. An edge in $\mathcal{G}^D_{K,\ell}$ corresponds to a class of QM-cyclic ℓ-isogenies $[\varphi : (A, \iota) \to (A', \iota')]$ between QM abelian surfaces with K-CM, and the above lemma tells us that as we move along paths in $\mathcal{G}^D_{K,\ell}$, the central conductors of vertices met have the same prime-to-ℓ part. It follows that $\mathcal{G}^D_{K,\ell}$ has a decomposition

$$\mathcal{G}^D_{K,\ell} = \bigsqcup_{(f_0,\ell)=1} \mathcal{G}^D_{K,\ell,f_0},$$

where $\mathcal{G}^D_{K,\ell,f_0}$ denotes the subgraph of $\mathcal{G}^D_{K,\ell}$ with vertices having corresponding central conductors of the form $f_0\ell^a$ for some $a \in \mathbb{N}$.

Any edge in $\mathcal{G}_{K,\ell,f_0}$ has vertices with corresponding central conductors f and f' satisfying $f/f' \in \{1, \ell, \ell^{-1}\}$. Defining the *level* of a vertex in $\mathcal{G}^D_{K,\ell,f_0}$ having central conductor f to be $\operatorname{ord}_\ell(f)$, we note that a directed edge can do one of three things:

- increase the level by one, in which case we will call the edge *ascending*,
- decrease the level by one, in which case we will call the edge *descending*, or
- leave the level unchanged, in which case we will call the edge *horizontal*.

We will refer to ascending and descending edges collectively as *vertical* edges. For a connected component of $\mathcal{G}^D_{K,\ell,f_0}$, we refer to the subgraph consisting of level 0 vertices and horizontal edges between them as the *surface* of that component. In other words, the vertex set of the surface consists of vertices with corresponding central conductor f_0. This choice of terminology is reflective of the fact that we cannot have an ascending isogeny starting at level 0, and of the fact that horizontal edges can only occur between surface vertices, as the following lemma states.

Lemma 4.2. *Suppose that there is a horizontal edge in $\mathcal{G}^D_{K,\ell,f_0}$ connecting vertices v_1 and v_2. Letting f_i denote the central conductor corresponding to v_i for $i = 1, 2$, we then have $f_1 = f_2 = f_0$. The number of horizontal edges emanating from a given surface vertex in $\mathcal{G}^D_{K,\ell,f_0}$ is $1 + \left(\frac{\Delta_K}{\ell}\right)$, hence is*

- 0 *if ℓ is inert in K,*
- 1 *if ℓ ramified in K, and*
- 2 *if ℓ is split in K.*

Proof. That $f_1 = f_2$ is part of our definition of horizontal edges. What we must prove is that ℓ does not divide $f := f_1 = f_2$.

The given edge corresponds to a QM-cyclic ℓ-isogeny

$$\varphi : (A_1, \iota_1) \to (A_2, \iota_2),$$

where (A_i, ι_i) has central conductor f for $i = 1, 2$. By Theorem 2.13, we have a decomposition of these two QM abelian surfaces resulting in an isomorphic isogeny ψ as below:

$$\begin{array}{ccc} (A_1, \iota_1) & \xrightarrow{\varphi} & (A_2, \iota_2) \\ \downarrow \cong & & \downarrow \cong \\ (E_1 \times E_1', \iota_1) & \xrightarrow{\psi} & (E_2 \times E_2', \iota_2) \end{array}$$

where each E_i and each E_i' is an elliptic curve with K-CM by conductor f for $i = 1, 2$. Restricting ψ to E_1 and to E_1', respectively, yields isogenies of K-CM elliptic curves

$$\begin{aligned} E_1 &\to \psi(E_1) =: E \subseteq E_2 \times E_2', \\ E_1' &\to \psi(E_1') =: E' \subseteq E_2 \times E_2'. \end{aligned} \tag{1}$$

This provides the decomposition

$$E_2 \times E_2' \cong E \times E'.$$

The conductors of the endomorphism rings of E and E', each of which must divide f and have the same coprime to ℓ part as f, must then have least common multiple f. This provides that either E or E' must have CM conductor f.

The conductors of the endomorphism rings of E and E' must each be in the set $\left\{f, \ell f, \frac{1}{\ell} f\right\}$, and must have least common multiple f. This provides that either E or E' must have CM conductor f.

We now consider the corresponding isogeny of K-CM elliptic curves of conductor f from (1). In doing so, [Clark and Saia 2022, Lemma 4.1] tells us that we must have $\ell \nmid f$. There, the result is reached using the correspondence between horizontal ℓ-isogenies of $\mathfrak{o}(f_0)$-CM elliptic curves over $\mathbb{C}$ with proper $\mathfrak{o}(f_0)$-ideals of norm ℓ. This also gives us the count of horizontal isogenies mentioned; we have the count in the elliptic curve case as in [Clark and Saia 2022], and from a horizontal isogeny of elliptic curves as in (1) we generate a QM-cyclic isogeny of our QM abelian surfaces via the QM action. □

Each surface vertex has $1+\left(\frac{\Delta_K}{\ell}\right)$ horizontal edges emanating from it, and therefore has $\ell-\left(\frac{\Delta_K}{\ell}\right)$ descending edges to level 1 vertices. For vertices away from the surface, we have the following:

Lemma 4.3. *If v is a vertex in $\mathcal{G}^D_{K,\ell,f_0}$ at level $L > 0$. then there is one ascending vertex from v to a vertex in level $L-1$, and the remaining ℓ edges from v are descending edges to distinct vertices in level $L+1$.*

Proof. We will use the same type of counting argument one may use in the elliptic curve case, as in [Sutherland 2013, Lemma 6]. The action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $\mathcal{G}^D_{K,\ell,f_0}$

preserves the level of a given vertex, and hence preserves the notions of horizontal, ascending, and descending for edges. As a result, the number of ascending, respectively descending, edges out of v must be the same as for any other vertex at level L by transitivity of this action on vertices at each level.

For $L = 1$, there are

$$\left(\ell - \left(\frac{\Delta_K}{\ell}\right)\right) 2^b h(\mathfrak{o}(f_0)) = 2^b h(\mathfrak{o}(\ell f_0))$$

total descending vertices from surface vertices (where b is as in Proposition 2.12). The equality above states that this is equal to the total number of level 1 vertices, and so the edges must all be to distinct level 1 vertices. For $L > 1$, the result follows inductively using the same counting argument along with the fact that

$$h(\mathfrak{o}(\ell^L f_0)) = \ell \cdot h(\mathfrak{o}(\ell^{L-1} f_0)). \qquad \square$$

4.2. ***QM-equivariant isogeny volcanoes.*** For a prime number ℓ, we define here the notion of an ℓ-volcano. This notion for the most part agrees with that in the existing literature, with the only caveat being that in the original context of isogeny volcanoes over a finite field one has volcanoes of finite depth. In our case, working over an algebraically closed field as in [Clark 2022; Clark and Saia 2022], we adjust the definition to allow for infinite depth volcanoes.

Definition 4.4. Let V be a connected graph with vertices partitioned into levels

$$V = \bigsqcup_{i \geq 0} V_i,$$

such that if $V_d = \varnothing$ for some d, then $V_i = \varnothing$ for all $i \geq d$. If such a d exists, we will refer to the smallest such d as the *depth* of V and to V_d for d the depth as the *floor* of V, and otherwise we will say that the depth of V is infinite.

Fixing a prime number ℓ, the graph V with its partitioning is an *ℓ-volcano* if the following properties hold:

(1) Each vertex not in the floor of V has degree $\ell + 1$, while any floor vertex has degree 1.

(2) The subgraph V_0, which we call *the surface*, is regular of degree 0, 1 or 2.

(3) For $0 < i < d$ (colloquially: "below the surface" and "above the floor"), a vertex in V_i has one *ascending* edge to a vertex in V_{i-1}, and ℓ *descending* edges to distinct vertices in V_{i+1}. This accounts for all edges of V which are not *horizontal*, by which we mean edges which are not between two surface vertices.

The results of the previous section immediately imply the following theorem, declaring that in most cases connected components of the subgraphs $\mathcal{G}^D_{K,\ell,f_0}$ of $\mathcal{G}^D_{K,\ell}$ are isogeny volcanoes. In such a case, we will refer to this graph as a *QM-equivariant* *isogeny volcano*. This justifies our use of terminology regarding edges and vertices in these subgraphs.

Theorem 4.5. *Fix an imaginary quadratic field K, a prime ℓ and a natural number f_0 with $(\ell, f_0) = 1$ and $f_0^2\Delta_K < -4$. Consider the graph $\mathcal{G}^D_{K,\ell,f_0}$ as an undirected graph by identifying edges with their dual edges as described above. Each connected component of this graph has the structure of an ℓ-volcano of infinite depth.*

A *path* in $\mathcal{G}^D_{K,\ell,f_0}$ refers to a finite sequence of directed edges, say $e_1, \ldots, e_r$, such that the terminal vertex of e_i is the initial vertex of e_{i+1} for all $1 \le i \le r-1$. In the $f_0^2\Delta_K < -4$ case, because the edges in $\mathcal{G}^D_{K,\ell,f_0}$ all have canonical inverse edges we are justified in using the following terminology: we call an edge *backtracking* if e_{i+1} is inverse to e_i for some edge e_i in the path. Note that in the case of ℓ ramified in K, a path consisting of two surface edges always is backtracking. If ℓ is split in K, then there is a horizontal cycle at the surface. In this case, concatenation of this cycle with itself any number of times does not result in backtracking.

Our definitions and the results of this section lead us to the following correspondence:

Lemma 4.6. *Suppose that $f_0^2\Delta_K < -4$. We then have a bijective correspondence between the set of geometric isomorphism classes of QM-cyclic ℓ^a-isogenies of QM abelian surfaces with K-CM and central conductor with prime-to-ℓ part f_0, and the set of length a nonbacktracking paths in $\mathcal{G}^D_{K,\ell,f_0}$. This associates to an isogeny its corresponding path in this isogeny graph.*

Proof. This result is in exact analogy to [Clark 2022, Lemma 4.2], and the proof is as therein. □

In Section 6, we will describe the Galois orbits of such paths in order to describe the K-CM locus on $X_0^D(\ell^a)$ via the above correspondence. For this, the following observation will be of use: any nonbacktracking length a path in $\mathcal{G}^D_{K,\ell,f_0}$ for $f_0^2\Delta_K < -4$ can be written as a concatenation of paths P_1, P_2 and P_3, where P_1 is strictly ascending, P_2 is strictly horizontal and hence consists entirely of surface edges, and P_3 is strictly descending, such that the lengths of these paths (which may be 0) sum to a.

4.3. ***The field of moduli of a QM-cyclic ℓ-isogeny.*** A QM-cyclic ℓ-isogeny φ of K-CM abelian surfaces with $\ell \nmid D$ corresponds to an edge e in $\mathcal{G}^D_{K,\ell,f_0}$, say between vertices v and v' in levels L and L', respectively. Assume that the path is nondescending ($L \ge L'$), so either it is horizontal ($L = L'$) or ascending ($L = L' + 1$).

An automorphism fixing e must fix both v and v', and so by Theorems 2.7 and 2.8 we have that either $\mathbb{Q}(\varphi) = K(\ell^L f_0)$, or $[K(\ell^L f_0) : \mathbb{Q}(\varphi)] = 2$. In the latter case, there exists an involution $\sigma \in \mathrm{Gal}(K(\ell^L f_0)/\mathbb{Q})$ fixing v, and we know precisely when this occurs by Theorem 2.8 — that is, when $D(K) = 1$.

Assume that $f_0^2\Delta_K < -4$, such that $\mathcal{G}^D_{K,\ell,f_0}$ has the structure of an ℓ-volcano. (We will deal with the case of $f_0^2\Delta_K \in \{-3,-4\}$ in the remarks leading up to Proposition 5.3.) If e is the unique edge between v and a vertex in level L', then e is fixed by σ if and only if v is. This is the case unless $L = L' = 0$ and ℓ splits in K, in which case there are two edges from v to surface vertices (which are not necessarily unique, or distinct from v). In either of these cases, consider $[(E \times E', \iota)]$, with E having CM by $\mathfrak{o}(f_0)$, a decomposition of our QM abelian surface corresponding to v_1. The two outward edges from v then have corresponding kernels $\iota(\mathcal{O}) \cdot E[\mathfrak{p}]$ and $\iota(\mathcal{O}) \cdot E[\bar{\mathfrak{p}}]$, with $\mathfrak{p}$ a prime ideal in $\mathfrak{o}(f_0)$ of norm ℓ.

We claim that, in this situation, the involution $\sigma \in \mathrm{Gal}(K(f_0)/\mathbb{Q})$ fixing v cannot fix $\mathfrak{p}$, and hence cannot fix our edge e. Indeed, the exact statement of Theorem 2.8 says that $\sigma = \tau\sigma_{\mathfrak{a}}$ for a certain ideal $\mathfrak{a}$ of $\mathfrak{o}(f_0)$, so to fix e it would have to be the case that $\sigma_{\mathfrak{a}}$ acts on e and hence on v by complex conjugation. It follows from [González and Rotger 2006, Lemma 5.10] that this cannot be the case, as $\omega_{D(K)}$ acts nontrivially on v. From this discussion, we reach the following result regarding fields of moduli corresponding to our edges.

Proposition 4.7. *Let φ be a QM-cyclic ℓ-isogeny corresponding to an edge e from v to v' in $\mathcal{G}^D_{K,\ell,f_0}$ as above, with $f_0^2\Delta_K < -4$.*

- *If $D(K) \neq 1$, i.e., if there is a prime $p \mid D$ which is inert in K, then $\mathbb{Q}(\varphi) = K(\ell^L f_0)$.*
- *Suppose that $D(K) = 1$.*
  - *If φ is a QM-cyclic isogeny of QM abelian surfaces with CM by $\mathfrak{o}(f_0)$ and ℓ splits in K, then $\mathbb{Q}(\varphi) = K(f_0)$.*
  - *Otherwise, $[K(\ell^L f_0) : \mathbb{Q}(\varphi)] = 2$, with $\mathbb{Q}(\varphi)$ equal to the field of moduli corresponding to v as described in Theorem 2.8.*

5. The action of Galois on $\mathcal{G}^D_{K,\ell,f_0}$

5.1. ***Action of*** **$\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.** We have an action of $\mathrm{Aut}(\mathbb{C})$ on $\mathcal{G}^D_{K,\ell,f_0}$: an automorphism σ maps a vertex v corresponding to an isomorphism class of QM abelian surfaces $[(A,\iota)]$ to the vertex corresponding to $[(\sigma(A),\sigma(\iota))]$, and edges are mapped to edges via the action on the corresponding isomorphism classes of isogenies. This action factors through $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, and preserves the level of a vertex. It follows that it also preserves the notions of ascending, descending and horizontal for paths.

For a fixed level $L \geq 0$, let $\mathcal{G}^D_{K,\ell,f_0,L}$ denote the portion of $\mathcal{G}^D_{K,\ell,f_0}$ from the surface (level 0) to level L:

$$\mathcal{G}^D_{K,\ell,f_0,L} := \bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}^D_{K,\ell,f_0}.$$

By Theorem 2.7, the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the graph $\mathcal{G}^D_{K,\ell,f_0,L}$ factors through $\mathrm{Gal}(K(\ell^L f_0)/\mathbb{Q})$. If $D(K) \neq 1$, i.e., if there is some prime $p \mid D$ which is inert in K, then Theorem 2.8 says that the action of this group on V_L is free. Otherwise, each vertex v in level L is fixed by some involution σ, and the class of QM abelian surfaces corresponding to v has field of moduli isomorphic to $K(\ell^L f_0)^\sigma$.

We now fix a vertex v in level L in $\mathcal{G}^D_{K,\ell,f_0}$, and suppose that $\sigma \in \mathrm{Gal}(K(\ell^L f_0)/\mathbb{Q})$ is an involution fixing v. (This forces $D(K) = 1$.) In the following two sections, we provide an explicit description of the action of σ on $\mathcal{G}^D_{K,\ell,f_0,L}$ in all cases. First, we note here the number of vertices at each level fixed by σ.

Proposition 5.1. *Let $x \in X^D(1)_{/\mathbb{Q}}$ be an $\mathfrak{o}(\ell^L f_0)$-CM point fixed by an involution $\sigma \in \mathrm{Gal}(K(\ell^L f_0)/\mathbb{Q})$. Let b denote the number of prime divisors of D which are inert in K. For $0 \leq L' \leq L$, the number of vertices of $\mathcal{G}^D_{K,\ell,f_0}$ in level L' fixed by σ is*

$$2^b \cdot \#\mathrm{Pic}(\mathfrak{o}(\ell^{L'} f_0))[2].$$

Proof. By Theorem 2.8, the involution σ is of the form $\sigma = \tau \circ \sigma_0$ for some $\sigma_0 \in \mathrm{Pic}(\mathfrak{o}(\ell^L f_0))$, where τ denotes complex conjugation. The set of vertices of $\mathcal{G}^D_{K,\ell,f_0}$ at level L' has cardinality $2^b \cdot h(\mathfrak{o}(\ell^{L'} f_0))$, consisting of 2^b orbits under the action of $\mathrm{Pic}(\mathfrak{o}(\ell^{L'} f_0))$. Each orbit is a $\mathrm{Pic}(\mathfrak{o}(\ell^{L'} f_0))$-torsor, and σ_0 yields a bijection on each.

As a result, we have that the number of level L' vertices in a given orbit which are fixed by σ is the same as the number of elements of $\mathrm{Pic}(\mathfrak{o}(\ell^{L'} f_0))$ fixed by τ. As shown in [Clark 2022, Proposition 2.6], this count is equal to $\#\mathrm{Pic}(\mathfrak{o}(\ell^{L'} f_0))[2]$, as τ acts on $\mathrm{Pic}(\mathfrak{o}(\ell^{L'} f_0))$ by inverting ideals. □

Regarding this count, by [Cox 2013, Proposition 3.11] we have the following:

Lemma 5.2. *Let r denote the number of distinct odd prime divisors of a fixed imaginary quadratic discriminant Δ. Then $\mathrm{Pic}(\mathfrak{o}_\Delta)[2] \cong (\mathbb{Z}/2\mathbb{Z})^\mu$, where*

$$\mu = \begin{cases} r-1 & \text{if } \Delta \equiv 1 \pmod 4 \text{ or } \Delta \equiv 4 \pmod{16}, \\ r & \text{if } \Delta \equiv 8, 12 \pmod{16} \text{ or } \Delta \equiv 16 \pmod{32}, \\ r+1 & \text{if } \Delta \equiv 0 \pmod{32}. \end{cases}$$

5.2. ***The field of moduli of a QM-cyclic ℓ^a-isogeny.*** Let φ be a QM-cyclic ℓ^a-isogeny of K-CM abelian surfaces inducing a $\Delta = f^2\Delta_K$-CM point on $X_0^D(\ell^a)_{/\mathbb{Q}}$, with $\ell \nmid D$ and $D > 1$. Let P be the length a nonbacktracking path in $\mathcal{G}^D_{K,\ell,f_0}$ corresponding to φ, via Lemma 4.6, for the appropriate $f_0 \in \mathbb{Z}^+$. The ordered edges in P correspond to a decomposition

$$\varphi = \varphi_1 \circ \cdots \circ \varphi_a,$$

where each φ_i is a QM-cyclic ℓ-isogeny. If $\Delta < -4$, then Proposition 3.2 provides

$$\mathbb{Q}(\varphi) = \mathbb{Q}(\varphi_1) \cdots \mathbb{Q}(\varphi_a),$$

and for $f_0^2\Delta_K < -4$ Proposition 4.7 determines $\mathbb{Q}(\varphi_i)$ for each i. Note that if $\mathbb{Q}(\varphi_i)$ is a ring class field for any i, then $\mathbb{Q}(\varphi)$ must contain K

For $f_0^2\Delta_K \in \{-3, -4\}$, it is impossible to have $D(K) = 1$, as Δ_K has only a single prime divisor while D has at least 2. This is of course consistent with, and can be seen from, the general fact that Shimura curves have no real points; the residue field of a -3-CM or -4-CM point on $X^D(1)_{/\mathbb{Q}}$ must be K in this situation. By these observations and the discussion of the Galois action in the previous section, we have the following proposition.

Proposition 5.3. *Let $\varphi : (A, \iota) \to (A', \iota')$ be a QM-cyclic ℓ^a-isogeny. Suppose that (A, ι) has K-CM with central conductor $f_A = \ell^a f_0$ and that (A', ι') has central conductor $f_{A'} = \ell^{a'} f_0$. Let $L = \max\{a, a'\}$. Let P be the path corresponding to φ in $\mathcal{G}^D_{K,\ell,f_0,L}$.*

- *If $D(K) \neq 1$, i.e., if there is a prime $p \mid D$ which is inert in K, then $\mathbb{Q}(\varphi) = K(\ell^L f_0)$.*
- *Suppose that $D(K) = 1$.*
  - *If ℓ splits in K and φ factors through an ℓ-isogeny of QM abelian surfaces with $f_0^2\Delta_K$-CM, then $\mathbb{Q}(\varphi) = K(\ell^L f_0)$.*
  - *Suppose that we are not in the previous case. Let $\sigma \in \mathrm{Gal}(K(\ell^L f_0)/\mathbb{Q})$ be an involution fixing the class of (A, ι) or (A', ι'). If σ fixes the path P, then $\mathbb{Q}(\varphi) = K(\ell^L f_0)^\sigma$. Otherwise, $\mathbb{Q}(\varphi) = K(\ell^L f_0)$.*

We now explicitly analyze the Galois action in all cases as done in [Clark 2022, §5.3] and [Clark and Saia 2022, §4.2] in the $D = 1$ case. Borrowing the notation therein, for a specified K, f_0, and ℓ we let

$$\tau_L := \#\mathrm{Pic}(\mathfrak{o}(\ell^L f_0))[2].$$

By Proposition 5.1, the number of vertices in level L' in $\mathcal{G}^D_{K,\ell,f_0}$ that are fixed by an involution $\sigma \in \mathrm{Pic}(\ell^L f_0)$ of the type we are studying is $2^b \cdot \tau_{L'}$.

5.3. *Explicit description, I: $f_0^2\Delta_K < -4$.* Here, we assume $f_0^2\Delta_K < -4$, such that each component of $\mathcal{G}^D_{K,\ell,f_0}$ has the structure of an ℓ-volcano of infinite depth. This is in exact parallel to [Clark 2022, §5.3], bearing the same structure of results.

Let $0 \le L' \le L$, and let $\sigma \in \mathrm{Pic}(\ell^L f_0)$ be an involution fixing a vertex v in $\mathcal{G}^D_{K,\ell,f_0}$ in level L. In the following lemmas, we describe the action of σ on $\mathcal{G}^D_{K,\ell,f_0,L}$. In each case, we provide example figures (Figures 1–9) of a component of $\mathcal{G}^D_{K,\ell,f_0}$ (up to some finite level). In these graphs, vertices and edges colored purple (gray) are fixed by the action of the designated involution σ, while black edges and vertices are acted on nontrivially by σ. Without loss of generality based on the symmetry of our graph components, we will always take v to be the left-most vertex in level L in our figures.

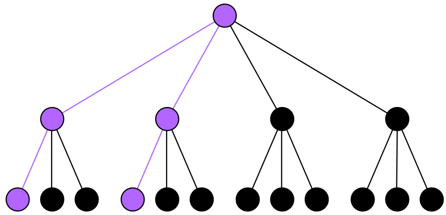

Figure 1. $\ell = 3$ inert in K with $L = 2$.

Lemma 5.4. *Let $\ell > 2$ be a prime which is unramified in K and $f_0 \in \mathbb{Z}^+$ with $f_0^2 \Delta_K < -4$. Let v, L and σ be as above with $L \geq 1$, and consider the action of σ on*

$$\bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}^D_{K,\ell,f_0}.$$

Each surface vertex has two descendants fixed by σ in level 1. For $1 \leq L' < L$, each fixed vertex in level L' has a unique fixed descendent in level $L' + 1$.

Proof. By Lemma 5.2 we have $\tau_1 = 2\tau_0$, while $\tau_{L'} = \tau_{L'+1}$ for $1 \leq L' < L$. The number of edges descending from a given vertex in level $L' \geq 1$ is ℓ, hence is odd, and so we immediately see that each fixed vertex in level L' with $1 \leq L' \leq L$ must have at least one fixed descendant in level $L' + 1$, hence exactly one by our count.

The number of descending edges from a given surface vertex is either $\ell + 1$ or $\ell - 1$ depending on whether ℓ is inert or split in K, hence is even in both cases. With our involution being of the form $\sigma = \tau\sigma_0$, a translated version of the argument of [Clark and Saia 2022, Corollary 5.5] gives that each fixed surface vertex has at least one fixed descendant in level 1. Therefore, each fixed surface vertex must have at least two fixed descendants in level 1 by parity, giving the result. □

Lemma 5.5. *Let $\ell > 2$ be a prime that ramifies in K and $f_0 \in \mathbb{Z}^+$ with $f_0^2 \Delta_K < -4$. Let v, L and σ be as above, and consider the action of σ on*

$$\bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}^D_{K,\ell,f_0}.$$

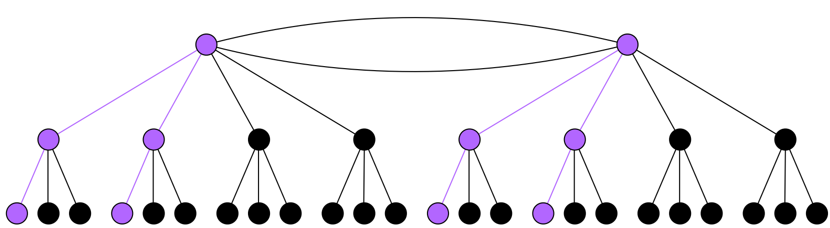

Figure 2. $\ell = 3$ split in K with $L = 2$.

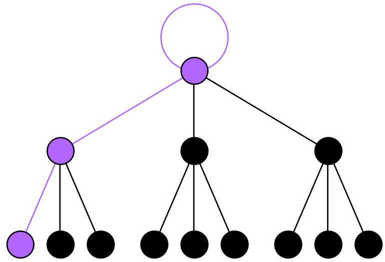

Figure 3. $\ell = 3$ ramified in K with $|V_0| = 1$ and $L = 2$.

Any vertex v' in level L' with $0 \leq L' < L$ which is fixed by σ has exactly one descendant in level $L' + 1$ fixed by σ.

Proof. Each vertex in level L' has ℓ descendants in level $L' + 1$. A descendant of v' must be sent to another descendant of v' by σ, by virtue of v' being fixed by σ. At least one descendant must be fixed by σ by the assumption that ℓ is odd. Lemma 5.2 gives that $\tau_{L'} = \tau_{L'+1}$, and so there must be exactly one fixed descendant of v'. □

Lemma 5.6. *Suppose that $\ell = 2$ is unramified in K and that $f_0^2 \Delta_K \neq -3$. Let v, L and σ be as above with $L \geq 1$, and consider the action of σ on*

$$\bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}^D_{K,2,f_0}.$$

1. *Every surface vertex fixed by σ has a unique fixed descendant in level* 1.
2. *Suppose $L \geq 2$. Each vertex in level* 1 *which is fixed by σ has all of its descendants in levels* 2 *to* $\min(L, 3)$ *fixed by σ.*
3. *Let $3 \leq L' < L$. If v' is a vertex in level L' fixed by σ, then the vertex w in level L' which shares a neighbor in level $L' - 1$ with v' is also fixed by σ, and exactly one of v' and w has its two descendants in level $L' + 1$ fixed by σ.*

Proof. (1) Lemma 5.2 provides $\tau_1 = \tau_0$. If 2 is inert in K, then each fixed surface vertex has three neighbors in level 1, and hence at least one must be fixed. The count

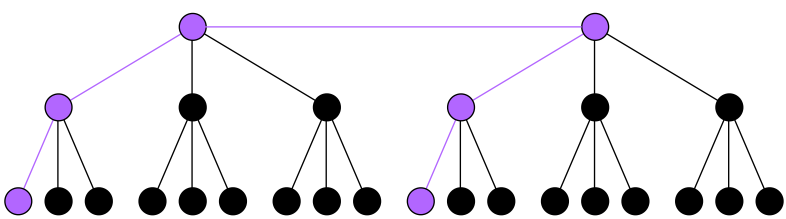

Figure 4. $\ell = 3$ ramified in K with $|V_0| = 2$ and $L = 2$.

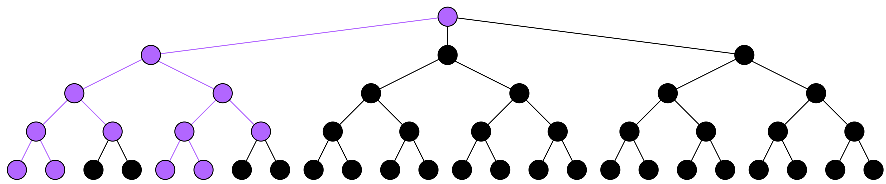

Figure 5. $\ell = 2$ inert with $L = 4$.

then implies exactly one of these neighbors must be fixed. If 2 splits in K, then each fixed surface vertex has exactly one neighbor in level 1 which then must be fixed.

(2) Lemma 5.2 provides $\tau_3 = 2\tau_2$ and $\tau_2 = 2\tau_1$. As each nonsurface vertex has two immediate descendants in the next level, the claim follows.

(3) For $3 \leq L' < L$, we have $\tau_{L'+1} = \tau_L$. Let $v_{L'}$ be a fixed vertex in level L' having a fixed neighbor vertex in level $L'-1$. By a parity argument, there must then be another fixed vertex $w_{L'}$ in level L' with the same neighbor in level $L'-1$ as $v_{L'}$. By the count, it suffices to show that $v_{L'}$ and $w_{L'}$ cannot both have descendants fixed by σ.

Suppose to the contrary that $v_{L'+1}$ and $w_{L'+1}$ are σ-fixed neighbors of $v_{L'}$ and $w_{L'}$, respectively, in level $L'+1$. We find that this cannot be the case as in [Clark 2022, Lemma 5.6c)]; this would imply that we have a QM-cyclic 2^4-isogeny which, upon restriction, would provide a cyclic, real 2^4-isogeny of elliptic curves with CM by $\Delta = 2^{2L+2} f_0^2 \Delta_K$. This in turn implies the existence of a primitive, proper real $\mathfrak{o}(2^{L+1} f_0)$-ideal of index 16, which does not exist. □

In the case of $\ell = 2$ ramifying in K, the discriminant of K must be of the form $\Delta_K = 4m$ for $m \equiv 2$ or $3 \pmod 4$, and so $\Delta_K \equiv 8$ or $12 \pmod{16}$. Hence, the discriminant of the order $\mathfrak{o}(f_0)$ corresponding to the surface of $\mathcal{G}^D_{K,2,f_0}$ will also lie in one of these congruence classes mod 16. Whether these components have a surface loop is answered by the following lemma.

Lemma 5.7. *Consider a component of $\mathcal{G}^D_{K,2,f_0}$ with 2 ramified in K. The surface V_0 of this component consists of a single vertex with a single self-loop if and only if $\Delta_K \in \{-4, -8\}$ and $f_0 = 1$.*

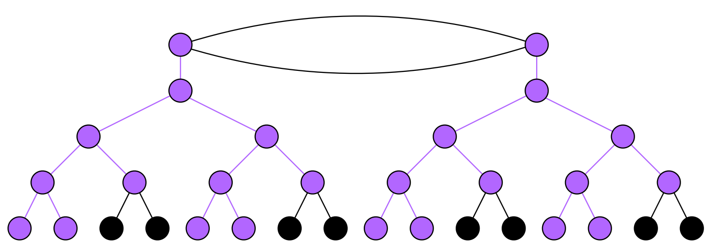

Figure 6. $\ell = 2$ split with $L = 4$.

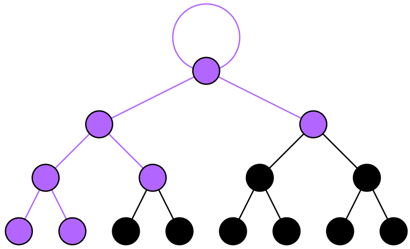

Figure 7. $f_0^2\Delta_K = -8$ and $\ell = 2$ with $L = 3$.

Proof. This proof comes down to a simple argument about ideals of norm 2 in $\mathfrak{o}(f_0)$, as in [Clark 2022, Lemma 5.7] □

The following lemmas therefore cover all possible cases.

Lemma 5.8. *Let $\Delta_K = -8$ and $\ell = 2$, and let v, L and σ be as above with $L \geq 1$. Consider the action of σ on*

$$\bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}_{K,2,1}^{D}.$$

(1) *The two descendants in level 1 of the single surface vertex are fixed by σ.*

(2) *For $1 < L' < L$, there are two vertices in level L' fixed by σ and they have a common neighbor vertex in level $L' - 1$. One of these must have both descendants in level $L'+1$ fixed by σ, while the other has its direct descendants swapped by σ.*

Proof. There is a single vertex on the surface, as the class number of K is 1. Lemma 5.2 tells us that $\tau_1 = 2\tau_0$ in this case, so both descendants of the surface vertex are fixed by σ. For $1 \leq L' < L$, we have

$$\tau_{L'+1} = \tau_{L'} = 2,$$

so one of the fixed vertices in level L' must have both descendants in level $L'+1$ fixed by σ, while the other has its vertices swapped by σ. □

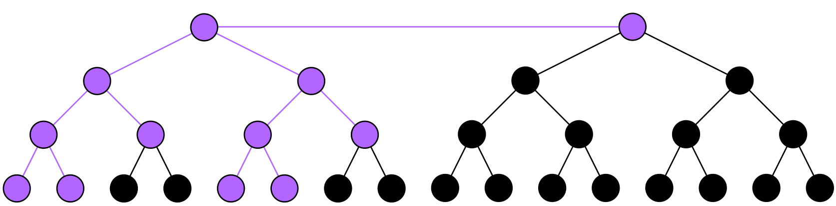

Figure 8. $\Delta_K \neq -4$ with $\ell = 2$, $\text{ord}_2(\Delta_K) = 2$ and $L = 3$.

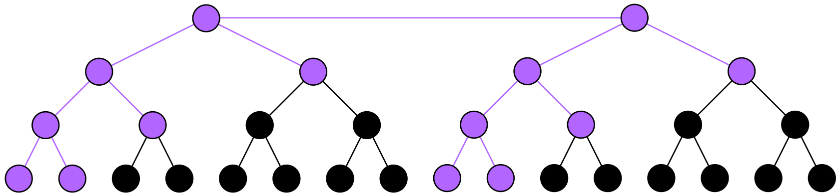

Figure 9. $\Delta_K < -8$ with $\ell = 2$, $\mathrm{ord}_2(\Delta_K) = 3$ and $L = 3$.

Lemma 5.9. *Suppose that* $\Delta_K \equiv 12 \pmod{16}$ *and* $f_0^2\Delta_K \neq -4$ *with* $\ell = 2$. *Let* v, L *and* σ *be as above with* $L \geq 1$. *Consider the action of* σ *on*

$$\bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}_{K,2,f_0}^{D}.$$

(1) *There are two surface vertices, both fixed by* σ. *One surface vertex, which we will denote by* v_0, *has both descendants in level* 1 *fixed by* σ, *while the other has its level* 1 *descendants swapped by* σ.

(2) *If* $L \geq 2$ (*such that the action of* σ *is defined at level* 2), *then each of the four vertices in level* 2 *which descend from* v_0 *are fixed by* σ.

(3) *For* $2 \leq L' < L$ *and for a vertex* v' *in level* L' *fixed by* σ, *let* w *denote the other level* L' *vertex sharing a neighbor vertex in level* $L'-1$ *with* v' (*which must also be fixed by* σ). *Exactly one of* v' *or* w *has both descendants in level* $L'+1$ *fixed by* σ, *while the other vertex has its direct descendants swapped by* σ.

Proof. In this case the surface has two σ-fixed vertices with a single edge between them. We have

$$\tau_1 = \tau_0 \quad \text{and} \quad \tau_2 = 2\tau_1$$

by Lemma 5.2, giving parts (1) and (2). For $2 \leq L' < L$, we have

$$\tau_{L'} = \tau_{L'-1},$$

so half of the σ-fixed vertices in level $L'-1$ must have both descendants in level L' fixed by σ, while the other half have their descendants in level L' swapped by σ. That there must be exactly one pair of fixed vertices in level L' descending from a given fixed vertex in level $L'-2$ follows as in part (3) of Lemma 5.6. □

Lemma 5.10. *Suppose that* $\Delta_K \equiv 8 \pmod{16}$ *with* $\Delta_K < -8$ *and* $\ell = 2$. *Let* v, L *and* σ *be as above with* $L \geq 1$. *Consider the action of* σ *on*

$$\bigsqcup_{i=0}^{L} V_i \subseteq \mathcal{G}_{K,2,f_0}^{D}.$$

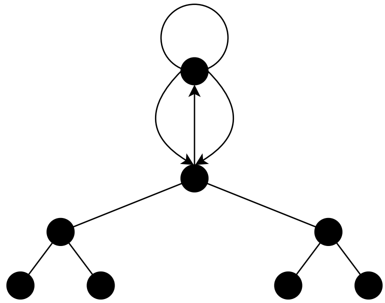

Figure 10. $f_0^2\Delta_K = -4$, $\ell = 2$ up to level 3.

(1) *There are two surface vertices, both fixed by* σ, *and all four vertices in level* 1 *are fixed by* σ.

(2) *For* $1 \leq L' < L$ *and for a vertex* v' *in level* L' *fixed by* σ, *let* w *denote the other level* L' *vertex sharing a neighbor vertex in level* $L' - 1$ *with* v'. *Exactly one of* v' *or* w *has both descendants in level* $L' + 1$ *fixed by* σ, *while the other vertex has its direct descendants swapped by* σ.

Proof. In this case again we have two σ-fixed vertices comprising our surface. Here Lemma 5.2 gives $\tau_1 = 2\tau_0$, providing part (1). For $1 \leq L' < L$, Lemma 5.2 gives $\tau_{L'} = \tau_{L'-1}$. The same argument as in part (3) of Lemma 5.9 then provides part (2). □

5.4. ***Explicit description, II:*** $\boldsymbol{f_0^2\Delta_K \in \{-3, -4\}}$**.** Keeping our notation from the previous section, we now assume $f_0 = 1$ and $\Delta_K \in \{-3, -4\}$. As mentioned earlier, we always have $D(K) \neq 1$ in this case. Therefore, the action of $\mathrm{Gal}(K(\ell^L f)/\mathbb{Q})$ on V_L is free for all $L \geq 0$. This is splendid news for us; while the CM fields $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-3})$ require extra attention at other points in this study, they cause absolutely no difficulties as far as determining the explicit Galois action on $\mathcal{G}^D_{K,\ell,1}$. This is to be compared with the $D = 1$ case of [Clark and Saia 2022, §4], wherein much

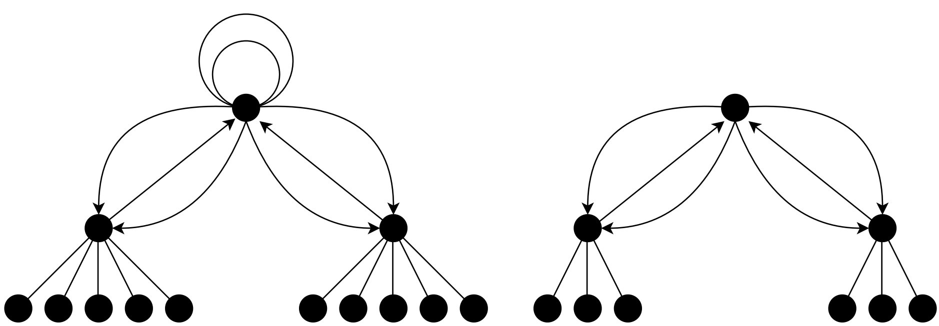

Figure 11. $f_0^2\Delta_K = -4$, ℓ split ($\ell = 5$, left) and inert ($\ell = 3$, right) up to level 2.

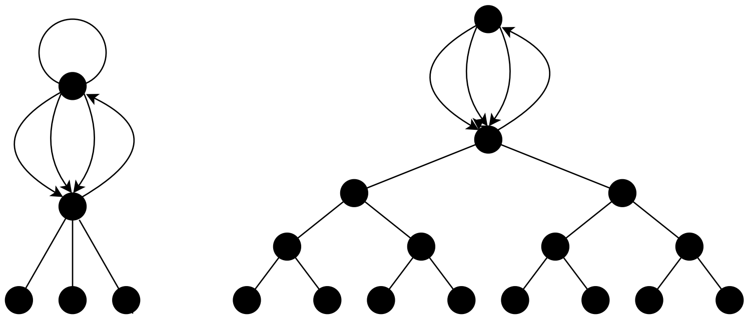

Figure 12. $f_0^2\Delta_K = -3$, $\ell = 3$ up to level 2 (left) and $\ell = 2$ up to level 3 (right).

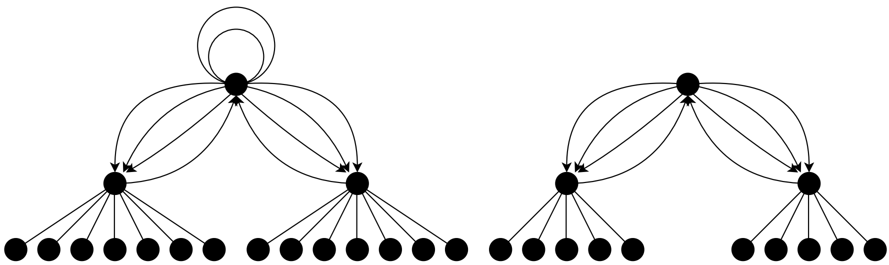

Figure 13. $f_0^2\Delta_K = -3$, ℓ split ($\ell = 7$, left) and inert ($\ell = 5$, right) up to level 2.

care goes into defining and explicitly describing a meaningful action of complex conjugation on CM components of isogeny graphs in these cases.

Still, we provide here example figures (Figures 10–13) of components of $\mathcal{G}_{K,\ell,1}$ (up to finite level L) for each case as reference for the reader for the path-type analysis and enumeration done in Section 6. In these cases, edges from level 0 to 1 have multiplicity as exposited in [Clark and Saia 2022, §3] due to the presence of automorphisms that do not fix kernels of isogenies. We therefore *do not* have a one-to-one identification between edges and "dual" edges in this case, and so as in the referenced study we clearly denote edges with orientation and multiplicity between levels 0 and 1.

6. CM points on $X_0^D(\ell^a)_{/\mathbb{Q}}$

We fix ℓ^a a prime power and $\Delta = f^2\Delta_K = \ell^{2L} f_0^2 \Delta_K$, with $\gcd(f_0, \ell) = 1$, an imaginary quadratic discriminant. We describe the Δ-CM locus on $X_0^D(\ell^a)_{/\mathbb{Q}}$. To this aim, we fully classify all closed-point equivalence classes, by which we mean $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ orbits, of nonbacktracking, length a paths in $\mathcal{G}_{K,\ell,f_0}^D$. We record the number of classes of each type with each possible residue field (up to isomorphism).

In the $f_0^2\Delta_K \in \{-3,-4\}$ cases, the notion of backtracking in $\mathcal{G}^D_{K,\ell,1}$ has subtlety between levels 0 and 1 that is not present in isogeny volcanoes. We address this now: traversing *any* edge from a vertex v in level 0 to a vertex w in level 1 followed by the single edge from w to v corresponds to a composition of dual isogenies, and thus is backtracking. On the other hand, for a given isogeny φ corresponding to the edge e from w to v, there is a single edge from v to w corresponding to its dual $\widehat{\varphi}$. Therefore, traversing e followed by the other edge (respectively, either of the two other edges) from v to w *does not* count as backtracking in the case of $f_0^2\Delta_K = -4$ (respectively, $f_0^2\Delta_K = -3$).

With b denoting the number of prime divisors of D which are inert in K, we have 2^b closed Δ-CM points on $X^D(1)_{/\mathbb{Q}}$, with the fibers over each under the natural map from $X_0^D(\ell^a)_{/\mathbb{Q}}$ to $X^D(1)_{/\mathbb{Q}}$ being isomorphic via Atkin–Lehner involutions. In all cases, we then have

$$\sum_{C(\varphi)} e_\varphi d_\varphi = 2^b \deg(X_0(\ell^a) \to X(1)) = 2^b \psi(\ell^a) = 2^b(\ell^a + \ell^{a-1}),$$

where our sum is over closed-point equivalence classes $C(\varphi)$ of QM-cyclic ℓ^a-isogenies φ with corresponding CM discriminant Δ.

The map $X_0^D(\ell^a)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$ has nontrivial ramification over a closed Δ-CM point if and only if $\Delta \in \{-3,-4\}$. For $\Delta \in \{-3,-4\}$ and path length a, we have that a closed-point equivalence class has ramification, of index 2 or 3 in the respective cases of $\Delta = -4$ and -3, if and only if the paths in the class include a descending edge from level 0 to level 1. This allows for a check on the classifications and counts that we provide.

If $D(K) = 1$, then the path types showing up in our analysis of each $\mathcal{G}^D_{K,\ell,f_0}$ are exactly those appearing in [Clark 2022] and [Clark and Saia 2022]. In this case, each graph $\mathcal{G}^D_{K,\ell,f_0}$ consists of 2^b copies of the analogous graph $\mathcal{G}_{K,\ell,f_0}$ from the $D = 1$ modular curve case. We have shown that the action of relevant involutions on each component is identical to the action of complex conjugation in the $D = 1$ case, up to symmetry of our graphs. In each place where the isomorphism class of a residue field in the referenced $D = 1$ analysis is a rational ring class field, we have in its place here some totally complex, index 2 subfield of a ring class field as described in Theorem 2.8.

If at least one prime dividing D is inert in K, i.e., if $D(K) > 1$, then all of the residue fields of Δ-CM points on $X^D(1)_{/\mathbb{Q}}$, and hence on $X_0^D(\ell^a)_{/\mathbb{Q}}$, are ring class fields. The path types showing up are exactly those in [Clark and Saia 2022], but the counts will in general differ from the case of the previous paragraph. Specifically, a given path type in our analysis in the case of $D(K) = 1$ consists of m classes with corresponding residue field $K(f')$ and n classes with corresponding residue field an index 2 subfield of $K(f')$ for some $f' \in \mathbb{Z}^+$ and $m, n \geq 0$. In the case of $D(K) > 1$, the same path type then consists of $2m+n$ classes, each with corresponding residue field $K(f')$.

Example 6.1. Suppose that $K=\mathbb{Q}(i)$ splits B, and consider the case of $\Delta=-4$ and $\ell^a=2$. We have 2^{b+1} closed-point equivalence classes of QM-cyclic 2-isogenies of QM abelian surfaces with -4-CM. Each corresponding point on $X_0^D(2)_{/\mathbb{Q}}$ has residue degree 1 over its image on $X^D(1)_{/\mathbb{Q}}$, having residue field K. Half of these classes, corresponding to self-loop edges at the surface, have no ramification, while each of the 2^b classes $C(\varphi)$ corresponding to a pair of descending edges to level 1 has $e_\varphi=2$.

A nonbacktracking length a path in $\mathcal{G}_{K,\ell,1}$ starting in level L consists of c ascending edges, followed by h horizontal edges, followed by d descending edges for some $c, h, d \geq 0$ with $c+h+d=a$. We denote this decomposition type of the path with the ordered triple (c, h, d).

6.1. ***Path-type analysis: general case.*** We begin here by considering the portion of the path-type analysis that is independent of ℓ and Δ_K.

I. There are classes consisting of strictly descending paths, i.e., with $(c, h, d)=(0, 0, a)$. If $D(K)\neq 1$, then there are 2^b such classes, each with residue field $K(\ell^a f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(\ell^a f)$.

II. If $a \leq L$, there are classes of strictly ascending paths, that is, with $(c, h, d)=(a, 0, 0)$. If $D(K)\neq 1$, then there are 2^b such classes, each with corresponding residue field $K(f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(f)$.

III. If $L=0$ and $\left(\frac{\Delta_K}{\ell}\right)=0$, then there are classes of paths with $(c, h, d)=(0, 1, a-1)$. If $D(K)\neq 1$, then there are 2^b such classes, each with corresponding residue field $K(\ell^{a-1} f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(\ell^{a-1} f)$.

IV. If $L=0$ and $\left(\frac{\Delta_K}{\ell}\right)=1$, then for each h with $1\leq h\leq a$ there are classes of paths with $(c, h, d)=(0, h, a-h)$ and residue field $K(\ell^{a-h} f)$. There are 2^{b+1} such classes if $D(K)\neq 1$, and there are 2^b such classes otherwise.

X. If $a > L \geq 1$ and $\left(\frac{\Delta_K}{\ell}\right)=1$, then there are classes of paths with $(c, h, d)=(L, a-L, 0)$ and residue field $K(f)$. There are 2^{b+1} such classes if $D(K)\neq 1$, and there are 2^b such classes otherwise.

6.2. ***Path-type analysis: $\ell > 2$.*** Here we assume that ℓ is an odd prime.

V. If $L\geq 2$, then for each c with $1\leq c\leq \min\{a-1, L-1\}$ there are paths which ascend at least one edge but not all the way to the surface, and then immediately descend at least one edge, with $(c, h, d)=(c, 0, a-c)$. Each such class has corresponding residue field $K(\ell^{\max\{a-2c,0\}} f)$. There are $2^b(\ell-1)\ell^{\min\{c,a-c\}-1}$ such paths if $D(K)\neq 1$, and $2^{b-1}(\ell-1)\ell^{\min\{c,a-c\}-1}$ such paths otherwise.

VI. If $a\geq L+1\geq 2$ and $\left(\frac{\Delta_K}{\ell}\right)=-1$, then there are paths which ascend to the surface and then immediately descend at least one edge, with $(c,h,d)=(L,0,a-L)$. If $D(K)\neq 1$, then there are $2^b\ell^{\min\{L,a-L\}}$ classes of such paths with corresponding residue field $K(\ell^{\max\{a-2L,0\}} f)$. Otherwise, there are $2^{b-1}(\ell^{\min\{L,a-L\}}-1)$ classes of such paths with corresponding residue field $K(\ell^{\max\{a-2L,0\}} f)$, and 2^b classes of such paths with corresponding residue field an index 2 subfield of $K(\ell^{\max\{a-2L,0\}} f)$.

VII. If $a\geq L+1\geq 2$ and $\left(\frac{\Delta_K}{\ell}\right)=0$, then there are paths which ascend to the surface and then immediately descend at least one edge, with $(c,h,d)=(L,0,a-L)$. Each such path has corresponding residue field $K(\ell^{\max\{a-2L,0\}} f)$. If $D(K)\neq 1$, then there are $2^b(\ell-1)\ell^{\min\{L,a-L\}-1}$ classes of such paths. Otherwise, there are $2^{b-1}(\ell-1)\ell^{\min\{L,a-L\}-1}$ classes.

VIII. If $a\geq L+1\geq 2$ and $\left(\frac{\Delta_K}{\ell}\right)=0$, then there are paths which ascend to the surface, follow one surface edge, and then possibly descend, with $(c,h,d)=(L,1,a-L-1)$. If $D(K)\neq 1$, then there are $2^b\ell^{\min\{L,a-L-1\}}$ classes of such paths with corresponding residue field $K(\ell^{\max\{a-2L-1,0\}} f)$. Otherwise, there are $2^{b-1}(\ell^{\min\{L,a-L-1\}}-1)$ classes of such paths with corresponding residue field $K(\ell^{\max\{a-2L-1,0\}} f)$, and 2^b classes of such paths with corresponding residue field an index 2 subfield of $K(\ell^{\max\{a-2L-1,0\}} f)$.

IX. If $a\geq L+1\geq 2$ and $\left(\frac{\Delta_K}{\ell}\right)=1$, then there are paths which ascend to the surface and then immediately descend at least one edge, with $(c,h,d)=(L,0,a-L)$. If $D(K)\neq 1$, then there are $2^b(\ell-2)\ell^{\min\{L,a-L\}-1}$ classes of such paths with corresponding residue field $K(\ell^{\max\{a-2L,0\}} f)$. Otherwise, there are $2^{b-1}((\ell-2)\ell^{\min\{L,a-L\}-1}-1)$ classes of such paths with corresponding residue field $K(\ell^{\max\{a-2L,0\}} f)$, and 2^b classes of such paths with corresponding residue field an index 2 subfield of $K(\ell^{\max\{a-2L,0\}} f)$.

XI. If $a\geq L+2\geq 3$ and $\left(\frac{\Delta_K}{\ell}\right)=1$, then for each $1\leq h\leq a-L-1$ there are paths which ascend to the surface, traverse h edges on the surface, and then descend at least one edge, with $(c,h,d)=(L,h,a-L-h)$. Each such path has corresponding residue field $K(\ell^{\max\{a-2L-h,0\}} f)$. If $D(K)\neq 1$, then there are $2^{b+1}(\ell-1)\ell^{\min\{L,a-L-h\}-1}$ classes of such paths. Otherwise, there are $2^b(\ell-1)\ell^{\min\{L,a-L-h\}-1}$ classes.

6.3. ***Path-type analysis:*** $\boldsymbol{\ell=2}$, $\boldsymbol{\left(\frac{\Delta_K}{2}\right)\neq 0}$. Here we assume that $\ell=2$ with Δ_K odd.

V. If $L\geq 2$, we have classes consisting of paths which ascend at least one edge but not all the way to the surface, and then immediately descend at least one edge. We have the following types:

- V_1. If $a\geq 2$, then there are classes with $(c,h,d)=(1,0,a-1)$. If $D(K)\neq 1$, then there are 2^b such classes, each with corresponding residue field $K(2^{a-2} f)$.

Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(2^{a-2}f)$.

V$_2$. If $L \geq a \geq 3$, then there are classes with $(c, h, d) = (a-1, 0, 1)$. If $D(K) \neq 1$, then there are 2^b such classes, each with corresponding residue field $K(2^{a-2}f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(2^{a-2}f)$.

V$_3$. If $a \geq L+1 \geq 4$, then there are paths with $(c, h, d) = (L-1, 0, a-L+1)$. If $D(K) \neq 1$, there are $2^{\min\{a-L+1, L-1\}+b-1}$ classes of such paths with corresponding residue field $K(2^{\max\{a-2L+2,0\}}f)$. Otherwise, there are $2^b(2^{\min\{a-L+1,L-1\}-2} - 1)$ classes of such paths with corresponding residue field $K(2^{\max\{a-2L+2,0\}}f)$, and 2^{b+1} classes of such paths with corresponding residue field an index 2 subfield of $K(2^{\max\{a-2L+2,0\}}f)$.

V$_4$. For each c with $2 \leq c \leq \min\{L-2, a-2\}$, there are paths with $(c, h, d) = (c, 0, a-c)$. Each such path has corresponding residue field $K(2^{\max\{a-2c,0\}}f)$. There are $2^{\min\{c,a-c\}+b-1}$ equivalence classes of such paths if $D(K) \neq 1$. Otherwise, there are $2^{\min\{c,a-c\}+b-2}$ such classes.

VI. If $a \geq L+1 \geq 2$ and $\left(\frac{\Delta_K}{2}\right) = -1$, there are paths that ascend to the surface and then immediately descend at least one edge, with $(c, h, d) = (L, 0, a-L)$. Each such class has corresponding residue field $K(2^{\max\{a-2L,0\}}f)$. If $D(K) \neq 1$, then there are $2^{\min\{L,a-L\}+b}$ classes of such paths. Otherwise, there are $2^{\min\{L,a-L\}-1+b}$ such classes.

XI. If $a \geq L+2 \geq 3$ and $\left(\frac{\Delta_K}{2}\right) = 1$, then for all $1 \leq h \leq a-L-1$ there are paths which ascend to the surface, traverse h horizontal edges, and then descend at least once, with $(c, h, d) = (L, h, a-L-h)$. Each such class has corresponding residue field $K(2^{\max\{a-2L-h,0\}}f)$. If $D(K) \neq 1$, then there are $2^{\min\{L,a-L-h\}+b}$ classes of such paths. Otherwise, there are $2^{\min\{L,a-L-h\}+b-1}$ such classes.

6.4. ***Path-type analysis: $\ell = 2$, $\mathrm{ord}_2(\Delta_K) = 2$.*** Here we assume that $\ell = 2$ with $\mathrm{ord}_2(\Delta_K) = 2$.

V. If $L \geq 2$, we have classes consisting of paths which ascend at least one edge but not all the way to the surface, and then immediately descend at least one edge. We have the following types:

V$_1$. If $a \geq 2$, then there are classes with $(c, h, d) = (1, 0, a-1)$. If $D(K) \neq 1$, then there are 2^b such classes, each with corresponding residue field $K(2^{a-2}f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(2^{a-2}f)$.

V$_2$. If $L \geq a \geq 3$, then there are classes with $(c, h, d) = (a-1, 0, 1)$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(f)$. Otherwise, there are 2^b classes of such paths, each with corresponding residue field an index 2 subfield of $K(f)$.

V$_3$. For each c with $2 \leq c \leq \min\{L-1, a-2\}$, there are paths $(c, h, d) = (c, 0, a-c)$. Each such class has corresponding residue field $K(2^{\max\{a-2c,0\}} f)$. If $D(K) \neq 1$, then there are $2^{\min\{c,a-c\}+b-1}$ classes of such paths. Otherwise, there are $2^{\min\{c,a-c\}+b-2}$ such classes.

VI. If $L \geq 1$, then we have paths which ascend to the surface and then immediately descend at least one edge, with $(c, h, d) = (L, 0, a-L)$. We have the following cases:

VI$_1$. Suppose $L = 1$ and $a \geq 2$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(2^{a-2} f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(2^{a-2} f)$.

VI$_2$. Suppose $a = L+1 \geq 3$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(f)$.

VI$_3$. Suppose $a \geq L+2 \geq 4$. If $D(K) \neq 1$, then there are $2^{\min\{L,a-L\}+b-1}$ classes of such paths, each with corresponding residue field $K(2^{\max\{a-2L,0\}} f)$. Otherwise, there are $2^b(2^{\min\{L,a-L\}-2}-1)$ classes of such paths with corresponding residue field $K(2^{\max\{a-2L,0\}} f)$, and 2^{b+1} classes of such paths with corresponding residue field an index 2 subfield of $K(2^{\max\{a-2L,0\}} f)$.

VIII. If $a \geq L+1 \geq 2$, then we have paths which ascend to the surface, and then traverse the unique surface edge, and then possibly descend, with $(c, h, d) = (L, 1, a-L-1)$. We have the following cases:

VIII$_1$. Suppose $a = L+1$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(f)$.

VIII$_2$. Suppose $a \geq L+2$. Each such path has corresponding residue field $K(2^{\max\{a-2L-1,0\}} f)$. If $D(K) \neq 1$, then there are $2^{\min\{L,a-1-L\}+b}$ classes of such paths. Otherwise, there are $2^{\min\{L,a-1-L\}+b-1}$ such classes.

6.5. *Path-type analysis:* $\ell = 2$, $\mathrm{ord}_2(\Delta_K) = 3$. Here we assume that $\ell = 2$ with $\mathrm{ord}_2(\Delta_K) = 3$. The types of paths occurring here are the same as in the previous section, owing to the fact that the structure of $\mathcal{G}^D_{K,\ell,f_0}$ here is the same as therein. The corresponding residue field counts may differ, though, as the Galois action differs.

V. The analysis of this type is exactly as in Section 6.4.

VI. If $L \geq 1$, then we have paths which ascend to the surface and then immediately descend at least one edge, with $(c, h, d) = (L, 0, a-L)$. We have the following cases:

VI_1. Suppose $L = 1$ and $a \geq 2$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(2^{a-2}f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(2^{a-2}f)$.

VI_2. Suppose $a = L+1 \geq 3$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(f)$.

VI_3. If $a \geq L+2 \geq 4$, then each such class has corresponding residue field $K(2^{\max\{a-2L,0\}}f)$. If $D(K) \neq 1$, then there are $2^{\min\{L,a-L\}+b-1}$ such classes. Otherwise, there are $2^{\min\{L,a-L\}+b-2}$ such classes.

VIII. If $a \geq L+1 \geq 2$, then we have paths which ascend to the surface, and then traverse the unique surface edge, and then possibly descend, with $(c, h, d) = (L, 1, a-L-1)$. We have the following cases:

VIII_1. Suppose $a = L+1$. If $D(K) \neq 1$, then there are 2^b classes of such paths, each with corresponding residue field $K(f)$. Otherwise, there are 2^b such classes, each with corresponding residue field an index 2 subfield of $K(f)$.

VIII_2. Suppose that $a \geq L+2$. If $D(K) \neq -1$, there are $2^{\min\{L,a-1-L\}+b}$ classes of such paths, each with corresponding residue field $K(2^{\max\{a-2L-1,0\}}f)$. Otherwise, there are $2^b(2^{\min\{L,a-1-L\}-1} - 1)$ classes of such paths with corresponding residue field $K(2^{\max\{a-2L-1,0\}}f)$, and 2^{b+1} classes with corresponding residue field an index 2 subfield of $K(2^{\max\{a-2L-1,0\}}f)$.

6.6. ***Primitive residue fields of CM points on $X_0^D(\ell^a)_{/\mathbb{Q}}$.*** Fixing Δ an imaginary quadratic discriminant and $N \in \mathbb{Z}^+$ relatively prime to D, we say that a field F is a *primitive residue field of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$* if

- there is a Δ-CM point $x \in X_0^D(N)_{/\mathbb{Q}}$ with $\mathbb{Q}(x) \cong F$, and
- there does not exists a Δ-CM point $y \in X_0^D(N)_{/\mathbb{Q}}$ with $\mathbb{Q}(y) \cong L$ with $L \subsetneq F$.

The preceding path-type analysis in this section allows us to determine primitive residue fields for prime-power levels $N = \ell^a$. It follows from this analysis that, in all cases, there are at most two primitive residue fields, and that each primitive residue field is either a ring class field or an index 2 subfield of a ring class field.

The cases occurring here are in line with those in [Clark 2022] and [Clark and Saia 2022], though here the primitive residue fields depend on whether $D(K) = 1$. In particular, if some prime dividing D is inert in K, then all residue fields of CM points on $X_0^D(\ell^a)$ are ring class fields, and hence there can only be one primitive residue field. This necessarily happens, for instance, if the class number of K is odd. We provide Case 1.5b with the alternative title of “the dreaded case,” in [Clark

2022], to warn the reader that it will have an important role in later results on primitive residue fields and degrees.

Case 1.1. Suppose $\ell^a = 2$.

Case 1.1a. Suppose $\left(\frac{\Delta}{2}\right) \neq -1$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.1b. Suppose $\left(\frac{\Delta}{2}\right) = -1$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(2f)$. Otherwise, the only primitive residue field is $K(2f)$.

Case 1.2. Suppose $\ell^a > 2$, $L = 0$ and $\left(\frac{\Delta}{\ell}\right) = 1$. If $D(K) = 1$, then the primitive residue fields are $K(f)$ and an index 2 subfield of $K(\ell^a f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.3. Suppose $\ell^a > 2$, $L = 0$ and $\left(\frac{\Delta}{\ell}\right) = -1$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(\ell^a f)$. Otherwise, the only primitive residue field is $K(\ell^a f)$.

Case 1.4. Suppose $\ell^a > 2$, $L = 0$ and $\left(\frac{\Delta}{\ell}\right) = 0$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(\ell^{a-1} f)$. Otherwise, the only primitive residue field is $K(\ell^{a-1} f)$.

Case 1.5. Suppose $\ell > 2$, $L \geq 1$ and $\left(\frac{\Delta_K}{\ell}\right) = 1$.

Case 1.5a. Suppose $a \leq 2L$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.5b *(the dreaded case).* Suppose $a \geq 2L + 1$. If $D(K) = 1$, then the primitive residue fields are $K(f)$ and an index 2 subfield of $K(\ell^{a-2L} f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.6. Suppose $\ell > 2$, $L \geq 1$ and $\left(\frac{\Delta_K}{\ell}\right) = -1$.

Case 1.6a. Suppose $a \leq 2L$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.6b. Suppose $a \geq 2L + 1$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(\ell^{a-2L} f)$. Otherwise, the only primitive residue field is $K(\ell^{a-2L} f)$.

Case 1.7. Suppose $\ell > 2$, $L \geq 1$ and $\left(\frac{\Delta_K}{\ell}\right) = 0$.

Case 1.7a. Suppose $a \leq 2L + 1$. If $D(K) = 1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.7b. Suppose $a \geq 2L+2$. If $D(K)=1$, then the only primitive residue field is an index 2 subfield of $K(\ell^{a-2L-1}f)$. Otherwise, the only primitive residue field is $K(\ell^{a-2L-1}f)$.

Case 1.8. Suppose $\ell = 2$, $a \geq 2$, $L \geq 1$ and $\left(\frac{\Delta_K}{2}\right) = 1$.

Case 1.8a. Suppose $L = 1$. If $D(K) = 1$, then the primitive residue fields are $K(f)$ and an index 2 subfield of $K(2^a f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.8b. Suppose $L \geq 2$ and $a \leq 2L-2$. If $D(K)=1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.8c. Suppose $L \geq 2$ and $a \geq 2L-1$. If $D(K) = 1$, then the primitive residue fields are $K(f)$ and an index 2 subfield of $K(2^{a-2L+2}f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.9. Suppose $\ell = 2$, $a \geq 2$, $L \geq 1$ and $\left(\frac{\Delta_K}{2}\right) = -1$.

Case 1.9a. Suppose $L = 1$. If $D(K) = 1$, then the primitive residue fields are $K(2^{a-2}f)$ and an index 2 subfield of $K(2^a f)$. Otherwise, the only primitive residue field is $K(2^{a-2}f)$.

Case 1.9b. Suppose $L \geq 2$ and $a \leq 2L-2$. If $D(K)=1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.9c. Suppose $L \geq 2$ and $a \geq 2L-1$. If $D(K) \neq 1$, then the primitive residue fields are $K(2^{\max\{a-2L,0\}}f)$ and an index 2 subfield of $K(2^{a-2L+2}f)$. Otherwise, the only primitive residue field is $K(2^{\max\{a-2L,0\}}f)$.

Case 1.10. Suppose $\ell = 2$, $a \geq 2$, $L \geq 1$, $\left(\frac{\Delta_K}{2}\right) = 0$ and $\operatorname{ord}_2(\Delta_K) = 2$.

Case 1.10a. Suppose $a \leq 2L$. If $D(K) \neq 1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.10b. Suppose $a \geq 2L+1$. If $D(K) \neq 1$, then the primitive residue fields are $K(2^{a-2L-1}f)$ and an index 2 subfield of $K(2^{a-2L}f)$. Otherwise, the only primitive residue field is $K(2^{a-2L-1}f)$.

Case 1.11. Suppose $\ell = 2$, $a \geq 2$, $L \geq 1$, $\left(\frac{\Delta_K}{2}\right) = 0$ and $\operatorname{ord}_2(\Delta_K) = 3$.

Case 1.11a. Suppose $a \leq 2L+1$. If $D(K)=1$, then the only primitive residue field is an index 2 subfield of $K(f)$. Otherwise, the only primitive residue field is $K(f)$.

Case 1.11b. Suppose $a \geq 2L+2$. If $D(K)=1$, then the only primitive residue field is an index 2 subfield of $K(2^{a-2L-1}f)$. Otherwise, the only primitive residue field is $K(2^{a-2L-1}f)$.

6.7. ***Primitive degrees of CM points on*** $X_0^D(\ell^a)_{/\mathbb{Q}}$. A positive integer d is a *primitive degree for a Δ-CM point on* $X_0^D(N)_{/\mathbb{Q}}$ if

- there is a Δ-CM point of degree d on $X_0^D(N)_{/\mathbb{Q}}$, and
- there does not exist a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ of degree properly dividing d.

If d is such a degree, then the residue field of a degree d point on $X_0^D(N)_{/\mathbb{Q}}$ is a primitive residue field of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$. For $N = \ell^a$ a prime power, we then have from the previous section that there are at most two primitive degrees.

Although there are several cases that admit two primitive residue fields when $D(K) = 1$, the only case admitting two primitive degrees is Case 1.5b (the dreaded case). In Case 1.5b, our two primitive residue fields are $K(f)$ and an index 2 subfield L of $K(\ell^{a-2L} f)$, with degrees $[K(f) : \mathbb{Q}] = 2h(\mathfrak{o}(f))$ and $[L : \mathbb{Q}] = \ell^{a-2L} h(\mathfrak{o}(f))$, respectively. As ℓ is odd, we indeed have two primitive degrees in this case.

7. Algebraic results on residue fields of CM points on $X^D(1)_{/\mathbb{Q}}$

We develop here algebraic number-theoretic results on fields which arise as residue fields of CM points on $X^D(1)_{/\mathbb{Q}}$ which will feed into our main results. In particular, a determination of composita and tensor products of such fields will be needed in determining information about the CM locus on $X_0^D(N)_{/\mathbb{Q}}$ for general N from information at prime-power levels.

For an imaginary quadratic field K, we let $K(f)$ denote the ring class field corresponding to the imaginary quadratic order $\mathfrak{o}(f)$ of conductor f in K, i.e., that of discriminant $f^2\Delta_K$.

Proposition 7.1. *Let K denote an imaginary quadratic field of discriminant Δ_K.*

(1) *If $\Delta_K \notin \{-3, -4\}$, then for any $f_1, f_2 \in \mathbb{Z}^+$ we have*

$$K(f_1) \cdot K(f_2) = K(\operatorname{lcm}(f_1, f_2)).$$

(2) *Suppose $\Delta_K \in \{-3, -4\}$.*

(a) *For any $f_1, f_2 \in \mathbb{Z}^+$ with $\gcd(f_1, f_2) > 1$, we have*

$$K(f_1) \cdot K(f_2) = K(\operatorname{lcm}(f_1, f_2)).$$

(b) *If the class number of the order of discriminant $f_1^2\Delta_K$ is 1, that is, if $f_1^2\Delta_K \in S = \{-3, -4, -12, -16, -27\}$, then*

$$K(f_1) \cdot K(f_2) = K(f_2).$$

(c) *Suppose we have positive integers $f_1, \ldots, f_r$ which are all pairwise relatively prime and not in the S defined above. Then $K(f_1) \cdots K(f_r) \subsetneq K(f_1 \cdots f_r)$,* *with*

$$[K(f_1\cdots f_r):K(f_1)\cdots K(f_r)]=\begin{cases}2^{r-1} & \text{if } \Delta_K=-4,\\ 3^{r-1} & \text{if } \Delta_K=-3.\end{cases}$$

(3) *In all cases*, $K(f_1)$ *and* $K(f_2)$ *are linearly disjoint over* $K(\gcd(f_1, f_2))$.

Proof. Part (1) is [Clark 2022, Proposition 2.10], while part (2) is [Clark and Saia 2022, Proposition 2.1] and part (3) follows from the combination of these two propositions. □

We now use Proposition 7.1 to get analogs of [Clark 2022, Proposition 2.10] and [Clark and Saia 2022, Proposition 2.2], in which "rational ring class fields" are exchanged for those index 2 subfields of rings class fields which arise as residue fields of CM points on $X^D(1)_{/\mathbb{Q}}$.

Corollary 7.2. *Suppose that* $x_1 \in X_0^D(N_1)_{/\mathbb{Q}}$ *and* $x_2 \in X_0^D(N_2)_{/\mathbb{Q}}$ *are* $\mathfrak{o}(f)$*-CM points, where* $\mathfrak{o}(f)$ *is an imaginary quadratic order in* K. *For* $i = 1, 2$, *let* $f_i \in \mathbb{Z}^+$ *such that*

$$K\cdot\mathbb{Q}(x_i)\cong K(f_i).$$

Let $M = \gcd(N_1, N_2)$ *and* $m = \operatorname{lcm}(N_1, N_2)$, *and suppose that* $x \in X_0^D(M)$ *is a point lying above* x_1 *and* x_2 *which is fixed by an involution* $\sigma \in \operatorname{Gal}(K(M)/K)$. *Let* $\pi : X_0^D(M)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$ *denote the natural map. Then*:

(1) *The fields* $\mathbb{Q}(x_1)$ *and* $\mathbb{Q}(x_2)$ *are linearly disjoint over* $\mathbb{Q}(\pi(x))$.

(2) *We have*

$$\mathbb{Q}(x_1)\otimes_{\mathbb{Q}(\pi(x))}\mathbb{Q}(x_2)\cong\mathbb{Q}(x).$$

(3) *We have*

$$\mathbb{Q}(x_1)\otimes_{\mathbb{Q}(\pi(x))}K(x_2)\cong K(x).$$

Proof. The ring class fields $K(f_1)$ and $K(f_2)$ are linearly disjoint over $K(m)$ by Proposition 7.1. That $\mathbb{Q}(x_1)$ and $\mathbb{Q}(x_2)$ are linearly disjoint over $\mathbb{Q}(\pi(x))$, and that

$$[\mathbb{Q}(x):\mathbb{Q}(x_1)\cdot\mathbb{Q}(x_2)]=[K(x):K(x_1)\cdot K(x_2]=[K(m):K(f_1)\cdot K(f_2)],$$

follow by the same type of arguments as in the analogous case of rational ring class fields in [Clark 2022, Proposition 2.10] and [Clark and Saia 2022, Proposition 2.2], using that $K(x)\cong K(x_1)\cdot K(x_2)\cong K(f_1)\cdot K(f_2)$ via Proposition 3.3 (note that the assumption that x is fixed by σ forces $f^2\Delta_K < -4$, so this proposition applies.).

Part (2) now follows from the preceding remarks, combined with Proposition 7.1. As for part (3), first note that the fact that $\mathbb{Q}(x)$ is fixed by some involution $\sigma \in \operatorname{Gal}(K(M)/K)$ immediately implies that $h(\mathfrak{o}(f)) > 1$ (as $X^D(1)_{/\mathbb{Q}}$ has no real points). We note that the map

$$\mathbb{Q}(x_1)\times K(x_2)\to K(x_1)\cdot K(x_2),\quad (x_1,x_2)\mapsto x_1\cdot x_2,$$

is $\mathbb{Q}(\pi(x))$-bilinear, and the induced map on the tensor product over $\mathbb{Q}(\pi(x))$ must be an isomorphism

$$\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2) \cong K(x_1) \cdot K(x_2)$$

as the two finite $\mathbb{Q}(f)$-algebras here have the same dimension. The result then follows as $K(x_1) \cdot K(x_2) \cong K(x)$. □

Corollary 7.3. *Suppose that $x_1, x_2, \ldots, x_r$ are $\mathfrak{o}(f)$-CM points with $x_i \in X_0^D(N_i)_{/\mathbb{Q}}$ for each $i = 1, \ldots, r$, where $\mathfrak{o}(f)$ is an imaginary quadratic order in K. For each $i = 1, \ldots, r$, let $f_i \in \mathbb{Z}^+$ such that*

$$K \cdot \mathbb{Q}(x_i) \cong K(f_i).$$

Let $M = \gcd(N_1, \ldots, N_r)$ and $m = \operatorname{lcm}(N_1, \ldots, N_r)$. Let $\pi : X_0^D(M)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$ denote the natural map. Let $S = \{-3, -4, -12, -16, -27\}$ be the set of discriminants of imaginary quadratic orders of class number 1 with $\Delta_K \in \{-3, -4\}$.

(1) *Suppose that $r = 2$. If $f_1 \in S$, then we have*

$$K(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2) \cong K(x_2) \times K(x_2).$$

Now assuming $f_1, f_2 \notin S$, if $\Delta_K < -4$ or if $\gcd(f_1, f_2) > 1$ then

$$K(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2) \cong K(M) \times K(M).$$

(2) *Suppose that $\Delta_K \in \{-3, -4\}$, that $f_1, \ldots, f_r \notin S$, and that $f_1, \ldots, f_r$ are all pairwise relatively prime. We then have*

$$\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} \cdots \otimes_{\mathbb{Q}(\pi(x))} \mathbb{Q}(x_r) \cong K(x_1) \otimes_{\mathbb{Q}(\pi(x))} \cdots \otimes_{\mathbb{Q}(\pi(x))} K(x_r) \cong L^r,$$

with L a subfield of $K(M)$ of index 2^{r-1} if $\Delta_K = -4$ and index 3^{r-1} if $\Delta_K = -3$.

Proof. (1) Using part (3) of Corollary 7.2, we have

$$\begin{aligned} K(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2) &\cong (\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x)) \otimes_{\mathbb{Q}(\pi(x))} (\mathbb{Q}(x_2) \otimes_{\mathbb{Q}(\pi(x))} K(x)) \\ &\cong (\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} \mathbb{Q}(x_2)) \otimes_{\mathbb{Q}(\pi(x))} (K(x) \otimes_{\mathbb{Q}(\pi(x))} K(x)) \\ &\cong (\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} \mathbb{Q}(x_2)) \otimes_{\mathbb{Q}(\pi(x))} (K(x) \times K(x)) \\ &\cong (\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2)) \times (\mathbb{Q}(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2)). \end{aligned}$$

The stated result then follows from another use of Corollary 7.2 part (3) if $\mathbb{Q}(x_1)$ properly embeds in the ring class field $K(f_1)$. Otherwise, $\mathbb{Q}(x_i) \cong K(f_i)$ for $i = 1, 2$ and $\mathbb{Q}(\pi(x)) \cong K(f)$. The case of $f_1 \in S$ is then clear, so assume $f_1, f_2 \notin S$ and at least one of $\Delta_K < -4$ or $\gcd(f_1, f_2) > 1$ holds. Then, by Proposition 7.1,

$$\begin{aligned} K(x_1) \otimes_{\mathbb{Q}(\pi(x))} K(x_2) &\cong (K(f_1) \otimes_{K(f)} K(f_2)) \times (K(x_1) \otimes_{K(f)} K(f_2)) \\ &\cong K(M) \times K(M). \end{aligned}$$

(2) This result follows similarly to the above argument using Proposition 7.1 once more. Our assumption that the f_i are relatively prime forces $\mathbb{Q}(x_i)$ to be a ring class field for each i; this assumption gives $K \cdot \mathbb{Q}(\pi(x)) \cong K(1) = K$ as $\Delta_K \in \{-3, -4\}$, and our Shimura curves have no real points so indeed $\mathbb{Q}(\pi(x)) \cong K$. □

8. CM points on $X_0^D(N)_{/\mathbb{Q}}$

We describe the Δ-CM locus on $X_0^D(N)_{/\mathbb{Q}}$ for any $N \in \mathbb{Z}^+$ relatively prime to D and any imaginary quadratic discriminant Δ. For $\Delta < -4$, this description is possible using the foundations we have built thus far, specifically Propositions 3.3 and 5.3, along with the path-type analysis in Section 6. For $\Delta = \Delta_K \in \{-3, -4\}$, however, Proposition 3.3 does not apply.

We first elaborate on the description in the former case, and then provide a result for compiling across prime powers in the case of $\Delta \in \{-3, -4\}$. Following this, we discuss primitive residue fields and degrees of Δ-CM points on $X_0^D(N)_{/\mathbb{Q}}$.

8.1. ***Compiling across prime powers: $\Delta < -4$.*** For a fixed prime ℓ relatively prime to D, let $\Delta = \ell^{2L} f_0^2 \Delta_K$ with $\gcd(f_0, \ell) = 1$ be an imaginary quadratic discriminant. Fixing $a \in \mathbb{Z}^+$, consider the natural map $\pi : X_0^D(\ell^a)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$ and the fiber $\pi^{-1}(x)$ over a Δ-CM point $x \in X_0^D(1)_{/\mathbb{Q}}$. There are 2^b such fibers by Theorem 2.13, and any two are isomorphic via an Atkin–Lehner involution w_p for some prime $p \mid D$ which is inert in K. We then have $\pi^{-1}(x) \cong \operatorname{Spec} A$ with

$$A = \prod_{j=0}^{a} {L_j}^{b_j} \times \prod_{k=0}^{a} K(\ell^k f)^{c_k} \tag{2}$$

for some nonnegative integers b_j, c_k, where L_j is an index 2 subfield of $K(\ell^j f)$ for all $0 \le j \le a$. The explicit values b_j and c_k, based on ℓ^a and Δ, are determined by our path-type analysis in Section 6.

Next assume $\Delta < -4$, let N denote a positive integer relatively prime to D, and consider the fiber $\pi^{-1}(x)$ of the map $\pi : X_0^D(N)_{/\mathbb{Q}} \to X(1)_{/\mathbb{Q}}$ over a Δ-CM point $x \in X^D(1)_{/\mathbb{Q}}$. Let $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$ be the prime-power factorization of N, and for each $1 \le i \le r$ consider the fiber $\pi_i^{-1}(x)$ of $\pi_i : X_0^D(\ell_i^{a_i})_{\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$ over x. We then have

$$\pi_i^{-1}(x) \cong \operatorname{Spec} A_i,$$

with each A_i of the form given in (2). Proposition 3.3 then provides that $\pi^{-1}(x) \cong \operatorname{Spec} A$ with

$$A = A_1 \otimes_{\mathbb{Q}(x)} \cdots \otimes_{\mathbb{Q}(x)} A_r.$$

It follows that A is a direct sum of terms of the form

$$M = M_1 \otimes_{\mathbb{Q}(x)} \cdots \otimes_{\mathbb{Q}(x)} M_r,$$

where for each $1 \le i \le r$ we have that M_i is isomorphic to $K(\ell_i^{j_i} f)$, or a totally complex, index 2 subfield thereof, for some $0 \le j_i \le a_i$.

Let s be the number of indices $1 \le i \le r$ such that K is contained in M_i, i.e., such that $M_i \cong K(\ell_i^{j_i})$ is a ring class field. The results of Section 7 then tell us that

$$M \cong \begin{cases} L & \text{if } s = 0, \\ K(\ell_1^{j_1} \cdots \ell_r^{j_r} f)^{2^{s-1}} & \text{otherwise,} \end{cases}$$

where $L \subsetneq K(\ell_1^{j_1} \cdots \ell_r^{j_r})$ is a totally complex, index 2 subfield in the $s = 0$ case. (Note that $\ell_i^{j_i} \Delta \in \{-12, -16, -27\}$ can only occur, due to the $\Delta < -4$ assumption, if $j_i = 0$, so these possibilities do not require special attention here.)

8.2. ***Compiling across prime powers:*** $\mathbf{\Delta \in \{-3, -4\}}$**.** Here, we determine how to compile residue field information across prime-power level for $\Delta \in \{-3, -4\}$. Our result here should be compared to [Clark and Saia 2022, Proposition 8.2 and Theorem 8.3], wherein more work is required due to the fact that residue fields of -3 and -4-CM points on $X_0(N)_{/\mathbb{Q}}$ do not always contain the CM field K.

Theorem 8.1. *Let $N \in \mathbb{Z}^+$ be coprime to D with prime-power factorization $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$, and suppose $x \in X_0^D(N)_{/\mathbb{Q}}$ is a Δ-CM point with $\Delta \in \{-3, -4\}$. Let $\pi_i : X_0^D(N)_{/\mathbb{Q}} \to X_0^D(\ell_1^{a_1})_{/\mathbb{Q}}$ denote the natural map and let $x_i = \pi_i(x)$ for each $1 \le i \le r$. Let P_i be any path in the closed-point equivalence class of paths in $\mathcal{G}_{K,\ell_i,1}^D$ corresponding to x_i, and let $d_i \ge 0$ be the number of descending edges in P_i (which is independent of the representative path). We then have*

$$\mathbb{Q}(x) \cong K(\ell_1^{d_1} \cdots \ell_r^{d_r}).$$

Proof. Because $\Delta \in \{-3, -4\}$, we know that the residue field of the image of x under the natural map to $X^D(1)_{/\mathbb{Q}}$ is necessarily K. Therefore, $K \subseteq \mathbb{Q}(x_i)$ for each i, and hence $\mathbb{Q}(x_i) \cong K(\ell_i^{d_i})$ for each $1 \le i \le r$.

Let $\varphi : (A, \iota) \to (A', \iota')$ be a QM-cyclic N-isogeny over $\mathbb{Q}(x)$ inducing x (necessarily there is such an isogeny, as $K \subseteq \mathbb{Q}(x)$). Let $Q = \ker(\varphi)$, let $C = e_1(Q)$, and for each $1 \le i \le r$ let $C_i \le C$ be the Sylow ℓ_i subgroup of C. Let $\varphi_i : (A, \iota) \to (A/(\mathcal{O} \cdot C_i), \iota_i)$ be the ℓ_i-primary part of φ, and let f_i denote the central conductor of $(A/(\mathcal{O} \cdot C_i), \iota)$ (where by ι here we really mean the induced QM structure on the quotient). Put

$$\mathcal{I} := \{i \mid d_i > 0\} = \{i \mid \operatorname{ord}_{\ell_i}(f_i) > 0\} \subseteq \{1, \dots, r\}$$

and

$$Q' := \langle \{\mathcal{O} \cdot C_i\}_{i \in \mathcal{I}} \rangle \le Q.$$

Our original isogeny φ then factors as $\varphi = \varphi'' \circ \varphi'$ where $\varphi' : (A, \iota) \to (A/Q', \iota)$. Because a QM-cyclic ℓ_i-isogeny preserves the prime-to-ℓ_i part of the central conductor, the central conductor of $(A/Q', \iota)$ must be divisible by $\ell_1^{d_1} \cdots \ell_r^{d_r}$. We then have

$$K(\ell_1^{d_1} \cdots \ell_r^{d_r}) \subseteq \mathbb{Q}(\varphi') \subseteq \mathbb{Q}(\varphi),$$

and it remains to show the reverse containment. If the central conductor of (A', ι') is also 1, then (up to isomorphism on the target) φ is a QM-equivariant endomorphism of (A, ι), and therefore $\mathbb{Q}(\varphi) \subseteq K$ as desired. Otherwise, the dual isogeny $\varphi^\vee$ induces a Δ'-CM point $x' \in X_0^D(N)_{/\mathbb{Q}}$ with $\Delta' < -4$. We have $\mathbb{Q}(x) \cong \mathbb{Q}(x')$, and the claim then holds via an application of Proposition 3.3 to x'. □

8.3. *The main algorithm.* We have now built up all we need to prove our main result, Theorem 1.1.

Proof of Theorem 1.1. The existence and structure of this algorithm follows from our prior results. We summarize the steps of the algorithm with appropriate references for individual steps here:

Algorithm 8.2 (the $\mathfrak{o}$-CM locus on $X_0^D(N)_{/\mathbb{Q}}$).

Input: an indefinite quaternion discriminant D over $\mathbb{Q}$, a positive integer N coprime to D, an imaginary quadratic discriminant Δ_K and a positive integer f.

Output: the complete list of tuples (`is_fixed`, f', e, c), consisting of a boolean `is_fixed`, a positive integer f', an integer $e \in \{1, 2, 3\}$ and a positive integer c, such that there exist exactly c closed $f^2\Delta_K$-CM points x on $X_0^D(N)_{/\mathbb{Q}}$ with $K(x) \cong K(f')$, with $\mathbb{Q}(x) \cong K(f')$ if `is_fixed` is False and with $[K(f') : \mathbb{Q}(x)] = 2$ otherwise and with ramification index e with respect to the natural map to $X^D(1)_{/\mathbb{Q}}$.

Steps:

- Compute the prime-power factorization $N = \ell_1^{a_1} \cdots \ell_r^{a_r}$ of N.

- For each index $i \in \{1, \ldots, r\}$, compute using the path-type enumeration results of Section 6 information on all $f^2\Delta_K$-CM points on $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$. This information is stored as a list of lists (`is_fixed`$_i$, f_i, e_i, c_i) as in our desired output at general level. (If $D = 1$, this information is originally obtained in the path-type analysis at prime-power level given in [Clark 2022] and [Clark and Saia 2022].)

- For each tuple $(P_1, \ldots, P_r)$, in which each P_i is the information of an $f^2\Delta_K$-CM point on $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$ of the form (`is_fixed`$_i$, f_i, e_i, c_i) as computed in the previous part, compute the information (`is_fixed`, f', e, c) of all $f^2\Delta_K$-CM points on $X_0^D(N)_{/\mathbb{Q}}$ with image a point with information given by P_i under the natural map to $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$ for all $i \in \{1, \ldots, r\}$. This is done as follows:
  - The boolean `is_fixed` is true if and only if the boolean `is_fixed`$_i$ is true for all $i \in \{1, \ldots, r\}$ by Proposition 3.3 and the results of Section 7.

– The CM conductor f' of such a point is equal to the least common multiple of the conductors $f_1, \dots, f_r$ at each prime-power level. This is by Proposition 3.3 and the results of Section 7, as also spelled out at the start of Section 8.1, if $\Delta_K < -4$, and is Theorem 8.1 in the case of $\Delta_K \in \{-3, -4\}$.
– The ramification index e is equal to the maximum among the indices e_i (so in particular is 2 or 3 if and only if $f_0^2 \Delta_K \in \{-3, -4\}$ and at least one of the P_i has $e_i = 2$ or $e_i = 3$).
– If $\Delta < -4$, then the count c is given by the results of Section 8.1. If $\Delta \in \{-3, -4\}$, then this count is given by the results of Section 8.2 if $D > 1$ and is given by the results of [Clark and Saia 2022] if $D = 1$. □

This algorithm has been implemented, and is the function `CM_points_XD0` in the file `shimura_curve_CM_locus.m` in [Saia 2024].

8.4. ***Primitive residue fields of CM points on $X_0^D(N)_{/\mathbb{Q}}$.*** The preceding results imply that the residue field of any Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ is isomorphic to either a ring class field or a totally complex, index 2 subfield of a ring class field as described in Theorem 2.8. As a result, there are at most two primitive residue fields of Δ-CM points on $X_0^D(N)_{/\mathbb{Q}}$. There exists a positive integer C such that an index 2 subfield of $K(Cf)$ is a primitive residue field of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ if and only if for each $1 \leq i \leq r$ there exists a positive integer C_i such that an index 2 subfield of $K(C_i f)$ is a primitive residue field of a Δ-CM point on $X_0^D(\ell_i^{a_i})$.

We begin by investigating the cases in which we do have such a field as a primitive residue field, determining when we have two primitive residue fields and, if so, whether we have two primitive degrees of residue fields. Note that this assumption requires $D(K) = 1$, and hence $\Delta < -4$. Let $H_i = \ell_i^{h_i} \mid \ell_i^{a_i}$ be the unique positive integer such that an index 2 subfield L_i of $K(H_i f)$ is a primitive residue field of a Δ-CM point on $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$ for each $1 \leq i \leq r$. Setting

$$H = H_1 \cdots H_r,$$

we have that a totally complex, index 2 subfield L of $K(Hf)$ is a primitive residue field of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ by the results of Section 8.1.

If L_i is the unique primitive residue field of a Δ-CM point on $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$ for each $1 \leq i \leq r$, then L is the unique primitive residue field for $X_0^D(N)_{/\mathbb{Q}}$. Otherwise, let $C_i = \ell_i^{c_i} \mid \ell_i^{a_i}$ be the smallest positive integer such that there is a Δ-CM point on $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$ with residue field isomorphic to either $K(C_i f)$ or an index 2 subfield thereof for each $1 \leq i \leq r$. Setting

$$C = C_1 \cdots C_r,$$

we then have that $K(Cf)$ is also a primitive residue field for $X_0^D(N)_{/\mathbb{Q}}$.

Now assume that we have two primitive residue fields, $L \subsetneq K(Hf)$ with $[K(Hf) : L] = 2$ and $K(Cf)$, of Δ-CM points on $X_0^D(N)_{/\mathbb{Q}}$. Set

$$d_1 := [L : \mathbb{Q}] \quad \text{and} \quad d_2 := [K(Cf) : \mathbb{Q}].$$

We note $C_i \le H_i$ for each $1 \le i \le r$ by the definitions of these quantities. Further, by assumption we have at least one value of i such that $K(C_i f)$ is a primitive residue field for $X_0^D(\ell_i^{a_i})_{/\mathbb{Q}}$, and thus

$$[K(C_i f) : \mathbb{Q}] \le \tfrac{1}{2}[K(H_i) : \mathbb{Q}] = [L_i : \mathbb{Q}].$$

It follows that $d_2 \le d_1$. Therefore, we have a unique primitive degree of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ if and only if $d_2 \mid d_1$, in which case d_2 is the unique primitive degree. The following result determines when this occurs:

Theorem 8.3. *With the setup and notation as above, let s be the number of indices $1 \le i \le r$ such that $K(H_i f)$ is a primitive residue field of a Δ-CM point on $X_0^D(\ell_i^{a_i})$ (or equivalently, such that $C_i < H_i$).*

(1) *If $s = 0$, then L is the unique primitive residue field of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$, and d_1 is the unique primitive degree.*

(2) *Suppose that $s \ge 1$ and that for some $1 \le i \le r$ with $C_i < H_i$ we are not in Case 1.5b (the dreaded case) with respect to Δ and the prime power $\ell_i^{a_i}$. We then have that L and $K(Cf)$ are the two primitive residue fields of Δ-CM points on $X_0^D(N)_{/\mathbb{Q}}$, while d_2 is the unique primitive degree.*

(3) *Suppose that $s \ge 1$ and that for all $1 \le i \le r$ with $C_i < H_i$ we are in Case 1.5b (the dreaded case) with respect to Δ and the prime power $\ell_i^{a_i}$. We then have that L and $K(Cf)$ are the two primitive residue fields of Δ-CM points on $X_0^D(N)_{/\mathbb{Q}}$, and that d_1 and d_2 are the two primitive degrees of such points.*

Proof. The proof follows exactly as in [Clark 2022, Theorem 9.2]; the main inputs here are the degrees of our residue fields, which are the same for our totally complex, index 2 subfields of ring class fields as they are for the rational ring class fields appearing in the $D = 1$ modular curve study. □

9. CM points on $X_1^D(N)_{/\mathbb{Q}}$

We prove Theorem 1.2, showing that there is a very close relationship between CM points on the Shimura curves $X_0^D(N)_{/\mathbb{Q}}$ and $X_1^D(N)_{/\mathbb{Q}}$. This is a generalization of [Clark and Saia 2022, Theorem 1.2], which was specific to the $D = 1$ case, and allows us to go from our understanding of the Δ-CM locus on $X_0^D(N)_{/\mathbb{Q}}$ based on Section 8 to an understanding of, at the very least, degrees of CM points on $X_1^D(N)_{/\mathbb{Q}}$.

Proof of Theorem 1.2. We first recall some relevant facts about ramification under the natural map $\pi : X_1^D(N)_{/\mathbb{Q}} \to X_0^D(N)_{/\mathbb{Q}}$. All points on $X_1^D(N)_{/\mathbb{Q}}$ not having CM by discriminant $\Delta \in \{-3, -4\}$ are unramified over their image on $X^D(1)_{/\mathbb{Q}}$. For $N \geq 4$, just as in the $D = 1$ case, the curve $X_1^D(N)$ over $\mathbb{C}$ has no elliptic points of periods 2 or 3, from which it follows that all -4 and -3-CM points on $X_1^D(N)_{/\mathbb{Q}}$ are ramified with ramification index 2 or 3, respectively. The curve $X_1^D(2)_{/\mathbb{Q}}$ has a single elliptic point of period 2, unramified with respect to π, lying over each of the 2^b points on $X^D(1)_{/\mathbb{Q}}$ with -4-CM. The curve $X_1^D(3)_{/\mathbb{Q}}$ has a single elliptic point of period 3, unramified with respect to π, lying over each of the 2^b points on $X^D(1)_{/\mathbb{Q}}$ with -3-CM. (One can see these claims regarding elliptic points and ramification from elementary arguments involving congruence subgroups. For example, for $D = 1$ this is [Diamond and Shurman 2005, Exercise 2.3.7].)

First, suppose that $\Delta < -4$. If $N \leq 2$ then the map π is an isomorphism, so assume $N \geq 3$ in which case it is a $(\mathbb{Z}/N\mathbb{Z})^*/\{\pm 1\}$-Galois covering, and hence has degree $\phi(N)/2$. Let f be the conductor of Δ, such that $\Delta = f^2\Delta_K$, and consider a point $\tilde{x} \in \pi^{-1}(x)$. It suffices to show that $[K(\tilde{x}) : K(x)] = \frac{\phi(N)}{2}$, viewing π as a morphism over K.

Take $\varphi : (A, \iota) \to (A', \iota')$ to be a QM-cyclic N-isogeny over $K(x)$ inducing $x \in X_0^D(N)_{/K}$. We know such an isogeny exists over $K(x)$ by Theorem 2.6, because $K(x)$ contains $\mathbb{Q}(x)$ and splits B. By Theorem 2.7 we have $K(x) = K(f)$. We have a well-defined ± 1 Galois representation

$$\overline{\rho_N} : \mathrm{Gal}(\overline{K}/K(f)) \to \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$$

not depending on our choice of representative for x, as $\mathrm{Aut}(A, \iota) = \{\pm 1\}$. Let $Q = \ker(\varphi) \leq A[N]$ and let $P \in Q$ be a choice of generator (of $e_1(Q)$ as an abelian group, or equivalently of Q as an $\mathcal{O}$-module). The action of $\mathrm{Gal}(\overline{K}/K(f))$ on P is then tracked by an isogeny character

$$\lambda : \mathrm{Gal}(\overline{K}/K(f)) \to (\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}.$$

Theorem 2.13 gives that $A_{\mathbb{C}} := A \otimes_{\mathrm{Spec}\, K(x)} \mathrm{Spec}\, \mathbb{C}$ has a decomposition $\psi : A_{\mathbb{C}} \xrightarrow{\sim} \mathbb{C}/\mathfrak{o}(f) \times E_A$, where E_A is a Δ-CM elliptic curve over $\mathbb{C}$. The elliptic curves in this decomposition both have models over $K(f)$, as they both have models over $\mathbb{Q}(f) \cong \mathbb{Q}(j_\Delta)$ where j_Δ is the j-invariant of a Δ-CM elliptic curve. Hence, a $K(f)$-rational model for this product is a twist of A.

It then suffices, as our representation is independent of the choice of $K(f)$-rational model, to consider the case $A = E \times E'$ with E and E' being Δ-CM elliptic curves over $K(f)$. Here, our QM-stable subgroup $Q \leq A[N]$ corresponds to a cyclic subgroup of $E[N]$, and λ is induced by the Galois action on this cyclic subgroup. This ± 1 character λ is surjective by [Bourdon and Clark 2020, Theorem 1.4] (in which the authors state a result of [Stevenhagen 2001] in this form). Therefore, if $\{P, -P\}$ is stable over an extension L of $K(f)$, such that $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$ is in the kernel of λ, we have

$$\tfrac{1}{2}\phi(N) \mid [L : K(f)],$$

and so indeed we have $[K(\tilde{x}) : K(x)] = \phi(N)/2$.

We next tackle case (2)(a), assuming that x is a ramified point of the map $X_0^D(N)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$. In this case, we have that a representative $(A, \iota, Q)_{/K(x)}$ inducing x, where $Q \leq A[N]$ is a QM-cyclic subgroup, is well-defined up to quadratic twist, as all models for (A, ι) are defined over $K(x)$. This is because, working geometrically for a second, a -3 or -4 CM point $x \in X_0^D(N)$ over $\mathbb{C}$ is ramified with respect to the natural map to $X^D(1)$ if and only if it is nonelliptic; it has the trivial stabilizer $\{\pm 1\}$, while its image is an elliptic point of order 3 or 4. The same argument as in the $\Delta < -4$ case above then applies.

We now assume that x is a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ with $\Delta \in \{-3, -4\}$ which is unramified with respect to the map to $X^D(1)_{/\mathbb{Q}}$. If $N = 2$, then π is an isomorphism, so the claim is trivial. If $N = 3$, the fact mentioned above that there is one point lying over each elliptic point on $X^D(1)$ is exactly the inertness claim. For $N \geq 4$, we know that every point in $\pi^{-1}(x)$ is ramified with respect to the map $X_1^D(N)_{/\mathbb{Q}} \to X^D(1)_{/\mathbb{Q}}$, giving the claimed ramification index. The residue degree is therefore *at most* the claimed residue degree in each case.

To provide the lower bound on the residue degree, we modify the argument of the $\Delta < -4$ case slightly in a predictable way. If $\Delta = -4$, then a representative for x is well-defined up to quartic twist. We consider a representative of the form $(E_1 \times E_2, \iota, \mathcal{O} \cdot C)$ where E_1, E_2 are $\mathfrak{o}(f)$-CM elliptic curves and $C \leq E_1[N]$ is a cyclic order N subgroup (again, via the type of argument as in the $\Delta < -4$ case using Theorem 2.13). Let $q_N : \mathfrak{o}_K \to \mathfrak{o}_K / N\mathfrak{o}_K$ denote the quotient map. By tracking the action of Galois on a generator P of C we get a well-defined reduced mod N Galois representation

$$\overline{\rho_N} : \mathrm{Gal}(\mathbb{Q}(x)/\mathbb{Q}) \to (\mathfrak{o}_K / N\mathfrak{o}_K)^\times / q_N(\mathfrak{o}_K^\times),$$

which is surjective (see [Bourdon and Clark 2020, §1.3]). As $\{P, -P, iP, -iP\}$ is stable under the action of $\mathrm{Gal}(\mathbb{Q}(\tilde{x})/\mathbb{Q})$ for $\tilde{x} \in \pi^{-1}(x)$, we must have

$$\tfrac{1}{4}\phi(N) = \#\big(\overline{\rho_N}(\mathrm{Gal}(\mathbb{Q}(\tilde{x})/\mathbb{Q}))\big) \mid [\mathbb{Q}(\tilde{x}) : \mathbb{Q}(x)],$$

giving the result for $\Delta = -4$. For $\Delta = -3$, exchanging "quartic" for "cubic" and μ_4 for μ_3 results in the required divisibility $\frac{1}{6}\phi(N) \mid f_\pi(x)$. □

10. Sporadic CM points on Shimura curves

Fix $D > 1$ an indefinite quaternion discriminant over $\mathbb{Q}$ and $N \in \mathbb{Z}^+$ relatively prime to D. In analogy to prior work on degrees of CM points on certain classical families of modular curves [Clark et al. 2022], we may consider the least degree $d_{\mathrm{CM}}(X)$ of a CM point on a Shimura curve X for the modular Shimura curves $X = X_0^D(N)_{/\mathbb{Q}}$ and $X = X_1^D(N)_{/\mathbb{Q}}$. For an imaginary quadratic order $\mathfrak{o}$, the results of Section 8.4 allow us to compute all primitive residue fields and degrees of $\mathfrak{o}$-CM points on $X_0^D(N)_{/\mathbb{Q}}$, and hence to compute the least degree $\mathrm{d}_{\mathfrak{o},\mathrm{CM}}(X_0^D(N))$ of an $\mathfrak{o}$-CM point on $X_0^D(N)_{/\mathbb{Q}}$. The least degree of an $\mathfrak{o}$-CM point on $X_0^D(N)_{/\mathbb{Q}}$ always satisfies

$$h(\mathfrak{o}) \mid d_{\mathfrak{o},\mathrm{CM}}(X_0^D(N)).$$

Using a complete list of all imaginary quadratic orders $\mathfrak{o}$ of class number up to 100, it then follows that if we have some order $\mathfrak{o}_0$ with

$$d_{\mathfrak{o}_0,\mathrm{CM}}(X_0^D(N)) \leq 100,$$

then we can solve the minimization over orders problem to compute the least degree of a CM point on $X_0^D(N)_{/\mathbb{Q}}$:

$$d_{\mathrm{CM}}(X_0^D(N)) = \min\{d_{\mathfrak{o},\mathrm{CM}}(X_0^D(N)) \mid h(\mathfrak{o}) \leq 100\}.$$

We have implemented an algorithm to compute least degrees over specified orders and, when possible, to compute $d_{\mathfrak{o},\mathrm{CM}}(X_0^D(N))$ exactly as described above. The relevant code, along with all other code used for the computational tasks described in this section, can be found at the repository [Saia 2024]. One may also find there a list of computed exact values of $d_{\mathrm{CM}}(X_0^D(N))$, along with an order minimizing the degree, for all relevant pairs (D, N) with $DN < 10^5$. All computations described in this section are performed using [Bosma et al. 1997].

Theorem 1.2 provides all of the information we need to go from least degrees of CM points on $X_0^D(N)_{/\mathbb{Q}}$ to least degrees of CM points on $X_1^D(N)_{/\mathbb{Q}}$. For ease of the relevant statement, we first generalize some terminology from [Clark et al. 2022]: we will call a pair (D, N) with $N \geq 4$

- *type I* if D splits $\mathbb{Q}(\sqrt{-3})$, we have $\mathrm{ord}_3(N) \leq 1$, and N is not divisible by any prime $\ell \equiv 2 \pmod 3$, and
- *type II* if D splits $\mathbb{Q}(\sqrt{-1})$, we have $\mathrm{ord}_2(N) \leq 1$, and N is not divisible by any prime $\ell \equiv 3 \pmod 4$.

Proposition 10.1. *Let $D > 1$ be a quaternion discriminant over $\mathbb{Q}$ and $N \in \mathbb{Z}^+$ be coprime to D.*

(1) *If (D, N) is type I, then*

$$d_{\mathrm{CM}}(X_1^D(N)) = \tfrac{1}{3}\phi(N).$$

(2) *If (D, N) is not type I and is type II, then*

$$d_{\mathrm{CM}}(X_1^D(N)) = \tfrac{1}{2}\phi(N).$$

(3) *If* (D, N) *is not type I or type II, then*

$$d_{\mathrm{CM}}(X_1^D(N)) = \tfrac{1}{2}\phi(N) \cdot d_{\mathrm{CM}}(X_0^D(N)).$$

Proof. The natural map $X_1^D(N)_{/\mathbb{Q}} \to X_0^D(N)_{/\mathbb{Q}}$ has nontrivial ramification exactly when (D, N) is either type I or type II. In these cases, we have $d_{\mathrm{CM}}(X_0^D(N)) = 2$, which is as small as possible as the $D > 1$ assumption implies these curves have no rational points. The statements then follow immediately from the residue degrees with respect to this map provided by Theorem 1.2. □

For a curve $X_{/\mathbb{Q}}$, let $\delta(X)$ denote the least positive integer d such that X has infinitely many points of degree d. We call a point $x \in X$ *sporadic* if

$$\deg(x) := [\mathbb{Q}(x) : \mathbb{Q}] < \delta(X).$$

That is, x is a sporadic point if there are only finitely many points $y \in X$ with $\deg(y) \leq \deg(x)$. Sporadic points on modular curves have been objects of interest in several recent works, including [Najman 2016; Bourdon and Najman 2021; Clark et al. 2022; Smith 2023; Bourdon et al. 2019; 2024].

In the rest of this section, we apply our least degree computations towards the question of whether the curves $X_0^D(N)_{/\mathbb{Q}}$ and $X_1^D(N)_{/\mathbb{Q}}$ have sporadic CM points.

10.1. ***An explicit upper bound on*** $d_{CM}(X_0^D(N))$**.** In analogy to the Heegner hypothesis of the modular curve case, we make the following definition:

Definition 10.2. Let D be an indefinite quaternion discriminant and N a positive integer relatively prime to D. We will say that an imaginary quadratic discriminant Δ *satisfies the* (D, N)*-Heegner hypothesis* if

(1) for all primes $\ell \mid D$, we have $\left(\frac{\Delta}{\ell}\right) = -1$, and

(2) for all primes $\ell \mid N$, we have $\left(\frac{\Delta}{\ell}\right) = 1$.

If Δ satisfies the (D, N)-Heegner hypothesis, this implies the existence of a Δ-CM point on $X_0^D(N)_{/\mathbb{Q}}$ which is rational over $K(f)$, the ring class field of conductor f where $\Delta = f^2\Delta_K$. This point therefore has degree at most $2 \cdot h(\mathfrak{o}(f))$.

We provide an upper bound on the least degree of a CM point on $X_0^D(N)_{/\mathbb{Q}}$ as follows: Let L be the least positive integer such that

- $\left(\frac{L}{p}\right) = -1$ for all odd primes $p \mid D$,
- $\left(\frac{L}{p}\right) = 1$ for all odd primes $p \mid N$, and
- we have

$$L \equiv \begin{cases} 5 \pmod 8 & \text{if } 2 \mid D, \\ 1 \pmod 8 & \text{otherwise.} \end{cases}$$

Then $0 < L < 8DN$, and so $d_0 = L - 16DN$ is an imaginary quadratic discriminant satisfying the (D, N)-Heegner hypothesis with $-16DN < d_0 < -8DN$. It follows that there exists a fundamental discriminant Δ_K of an imaginary quadratic field K satisfying the (D, N)-Heegner hypothesis with $|\Delta_K| < 16DN$; take K such that d_0 corresponds to an order in K, and hence $d_0 = f^2\Delta_K$ for some positive integer f.

For an imaginary quadratic field K of discriminant $\Delta_K < -4$, we have

$$h_K = h(\mathfrak{o}(\Delta_K)) \leq \frac{e}{2\pi}\sqrt{|d|}\log|d|$$

(see, e.g., [Clark et al. 2013, Appendix]), such that the above provides

$$d_{\mathrm{CM}}(X_0^D(N)) \leq 2 \cdot h_K \leq \frac{4e}{\pi}\sqrt{DN}\log(16DN). \tag{3}$$

10.2. ***Shimura curves with infinitely many points of degree 2.*** If $\delta(X_0^D(N)) = 2$, then as $X_0^D(N)_{/\mathbb{Q}}$ has no real points it certainly does not have a sporadic point. We mention here all pairs (D, N) for which we know $\delta(X_0^D(N)) = 2$ based on the existing literature.

All genus 0 and 1 cases necessarily have $\delta(X_0^D(N)) = 2$, as we have no degree 1 points. Voight [2009] lists all (D, N) for which $X_0^D(N)_{/\mathbb{Q}}$ has genus zero:

$$\{(6, 1), (10, 1), (22, 1)\},$$

and genus one:

$$\{(6, 5), (6, 7), (6, 13), (10, 3), (10, 7), (14, 1), (15, 1), (21, 1), (33, 1), (34, 1), (46, 1)\}.$$

By a result of Abramovich and Harris [1991], a nice curve X defined over $\mathbb{Q}$ of genus at least 2 with $\delta(X) = 2$ is either hyperelliptic over $\mathbb{Q}$, or is bielliptic and emits a degree 2 map to an elliptic curve over $\mathbb{Q}$ with positive rank. The pairs (D, N) for which $X_0^D(N)_{/\mathbb{Q}}$ is hyperelliptic of genus at least 2 were determined by Ogg[1] [1983]:

$$\{(6, 11), (6, 19), (6, 29), (6, 31), (6, 37), (10, 11), (10, 23), (14, 5), (15, 2), (22, 3), (22, 5), (26, 1), (35, 1), (38, 1), (39, 1), (39, 2), (51, 1), (55, 1), (58, 1), (62, 1), (69, 1), (74, 1), (86, 1), (87, 1), (94, 1), (95, 1), (111, 1), (119, 1), (134, 1), (146, 1), (159, 1), (194, 1), (206, 1)\}.$$

As for the bielliptic case, Rotger [2002] has determined all discriminants D such that $X^D(1) = X_0^D(1)$ is bielliptic, and further determined those for which $X^D(1)_{/\mathbb{Q}}$ is

[1] Actually, for the pairs (10, 19) and (14, 5), the referenced work of Ogg says that the corresponding curves are hyperelliptic over $\mathbb{R}$. Ogg does not say whether that is the case over $\mathbb{Q}$, but Guo and Yang [2017] answer negatively for the former pair and positively for the latter.

bielliptic over $\mathbb{Q}$ and maps to a positive rank elliptic curve. All such discriminants D with $g(X_0^D(1)) \geq 2$ and with $X_0^D(1)_{/\mathbb{Q}}$ not hyperelliptic are

$$\{57, 65, 77, 82, 106, 118, 122, 129, 143, 166, 210, 215, 314, 330, 390, 510, 546\}.$$

10.3. ***Sporadic CM points.*** In order to declare the existence of a sporadic CM point on a Shimura curve $X_0^D(N)_{/\mathbb{Q}}$, a main tool for us will be the following result of Frey [1994, Proposition 2] on the least degree $\delta(X)$ over which a nice curve $X_{/F}$ has infinitely many closed points:

Theorem 10.3 [Frey 1994]. *For a nice curve X defined over a number field F,*

$$\tfrac{1}{2}\gamma_F(X) \leq \delta(X) \leq \gamma_F(X),$$

where $\gamma_F(X)$ denotes the F-gonality of X, i.e., is the least degree of a nonconstant F-rational map to the projective line.

It follows from Theorem 10.3 that if

$$d_{\mathrm{CM}}(X_0^D(N)) < \tfrac{1}{2}\gamma_{\mathbb{Q}}(X_0^D(N)), \tag{4}$$

then there exists a sporadic CM point on $X_0^D(N)_{/\mathbb{Q}}$. To complement this, a result of Abramovich provides a lower bound on the gonality of a Shimura curve. Our cases of interest in applying this result are $X_0^D(N) = X_{\Gamma_0^D(N)}$ and $X_1^D(N) = X_{\Gamma_1^D(N)}$ (or, equivalently, $X_0^D(N) = X_{\mathcal{O}_N^1}$, where $\mathcal{O}_N$ is an Eichler order of level N in B, for the former curve).

Theorem 10.4 [Abramovich 1996]. *Let X_Γ be the Shimura curve corresponding to $\Gamma \leq \mathcal{O}^1$ a subgroup of the units of norm 1 in an order $\mathcal{O}$ of B. Then*

$$\tfrac{975}{8192}(g(X_\Gamma) - 1) \leq \gamma_{\mathbb{C}}(X_\Gamma) \leq \gamma_{\mathbb{Q}}(X_\Gamma).$$

Proof. This is a version of [Abramovich 1996, Theorem 1.1], where the constant has been improved using the best known progress due to Kim and Sarnak on Selberg's eigenvalue conjecture [Kim 2003, p. 176]. □

The following result will allow us to transfer information about the existence of sporadic points on $X_0^D(N)_{/\mathbb{Q}}$ to those on $X_1^D(N)_{/\mathbb{Q}}$:

Proposition 10.5. *Let $\pi : X_1^D(N)_{/\mathbb{Q}} \to X_0^D(N)_{/\mathbb{Q}}$ denote the natural modular map. Suppose that $P_0 \in X_0^D(N)_{/\mathbb{Q}}$ satisfies*

$$\deg(P_0) \leq \tfrac{975}{16384}\big(g(X_0^D(N)) - 1\big).$$

Then any $P \in X_1^D(N)_{/\mathbb{Q}}$ with $\pi(P) = P_0$ is sporadic.

Proof. For such a $P \in X_1^D(1)_{/\mathbb{Q}}$, using the notation and results of Proposition 2.5,

$$\begin{aligned}\deg(P) &\leq \deg(P_0)\cdot\deg(\pi)\\ &= \deg(P_0)\cdot \tfrac{1}{2}\phi(N)\\ &\leq \tfrac{975}{16384}\big(\tfrac{1}{12}\phi(D)\psi(N) - \tfrac{1}{4}\epsilon_1(D,N) - \tfrac{1}{3}\epsilon_3(D,N)\big)\cdot\tfrac{1}{2}\phi(N)\\ &\leq \tfrac{975}{16384}\big(\tfrac{1}{24}\phi(N)\phi(D)\psi(N)\big)\\ &= \tfrac{975}{16384}(g(X_1^D(N)-1).\end{aligned}$$

It then follows from Theorem 10.4 that P is sporadic. □

We now obtain a lower bound on the genus of $X_0^D(N)$ that will be amenable to our arguments:

Lemma 10.6. *For $D > 1$ an indefinite quaternion discriminant over $\mathbb{Q}$ and $N \in \mathbb{Z}^+$ relatively prime to D, we have*

$$\begin{aligned}g(X_0^D(N)) - 1 &> \frac{DN}{12}\left(\frac{1}{e^{\gamma}\log\log D + \frac{3}{\log\log D}}\right) - \frac{7\sqrt{DN}}{3}\\ &\geq \frac{DN}{12}\left(\frac{1}{e^{\gamma}\log\log(DN) + \frac{3}{\log\log 6}}\right) - \frac{7\sqrt{DN}}{3}.\end{aligned}$$

Proof. We make use of the trivial bound $\psi(N) \geq N$, and the lower bound

$$\phi(D) > \frac{D}{e^{\gamma}\log\log D + \frac{3}{\log\log D}}.$$

For $M \in \mathbb{Z}^+$, let $\omega(M)$ and $d(M)$ denote, respectively, the number of distinct prime divisors of M and the number of divisors of M. We then have

$$\epsilon_1(D,N), \epsilon_3(D,N) \leq 2^{\omega(DN)} \leq d(DN) \leq d(D)\cdot d(N) \leq 4\sqrt{DN}.$$

Using these bounds along with the fact that $D \geq 6$ and $N \geq 1$, we arrive at the stated inequalities from the genus formula given in Proposition 2.5. □

The combination of this lemma with (3) and (4) guarantees a sporadic CM point on $X_0^D(N)_{/\mathbb{Q}}$ if

$$\frac{4e}{\pi}\sqrt{DN}\log(16DN) \leq \frac{325DN}{65536}\left(\frac{1}{e^{\gamma}\log\log(DN) + \frac{3}{\log\log 6}}\right) - \frac{2275\sqrt{DN}}{16384}.$$

This inequality holds for all pairs (D, N) with $DN \geq 4.27512\cdot 10^{10}$.

Ranging through pairs (D, N) with DN below this bound, we attempt to determine the fundamental imaginary quadratic discriminant Δ_K of smallest absolute value satisfying the (D, N)-Heegner hypothesis. If found, we check whether we have a Δ_K-CM point of degree at most half $\gamma_{\mathbb{Q}}(X_0^D(N))$ via the inequality

$$h_K < \frac{325\phi(D)\psi(N)}{65536} - \frac{2275\sqrt{DN}}{16384}. \tag{5}$$

We confirm that (5) holds, and thus a sporadic CM point on $X_0^D(N)_{/\mathbb{Q}}$ is ensured, for all pairs (D, N) with $DN > 14982$ aside from the 20 pairs comprising the following set $\mathcal{F}_1$:

$$\begin{aligned}\mathcal{F}_1 = \big\{&(101959, 210), (111397, 210), (141427, 210), (154583, 210),\\ &(164749, 210), (165053, 330), (174629, 330), (190619, 210),\\ &(192907, 210), (194051, 210), (199801, 330), (208351, 210),\\ &(218569, 210), (233519, 210), (240097, 210), (272459, 210),\\ &(287419, 210), (296153, 210), (304513, 210), (307241, 210)\big\}.\end{aligned}$$

For each pair $(D, N) \in \mathcal{F}_1$, it is not that (5) does not hold. Rather, there is no imaginary quadratic discriminant of class number at most 100 satisfying the (D, N)-Heegner hypothesis, such that we fail to perform the check using only such discriminants. For each of these pairs, we compute $d_{\mathrm{CM}}(X_0^D(N))$ exactly and find that for each the inequality

$$d_{\mathrm{CM}}(X_0^D(N)) < \frac{325\phi(D)\psi(N)}{32768} - \frac{2275\sqrt{DN}}{8192}$$

holds. By the preceding remarks, this confirms that the curve $X_0^D(N)_{/\mathbb{Q}}$ has a sporadic CM point for all $(D, N) \in \mathcal{F}_1$.

There are exactly 4392 pairs (D, N), each with $DN \leq 14982$, for which (5) does not hold. These are listed in the file `bads_list.m` in [Saia 2024]. For each of these, we perform an exact computation of $d_{\mathrm{CM}}(X_0^D(N))$. By the above, a sporadic CM point on $X_0^D(N)_{/\mathbb{Q}}$ is guaranteed if

$$d_{\mathrm{CM}}(X_0^D(N)) < \tfrac{975}{16384}\big(\tfrac{1}{12}\phi(D)\psi(N) - \tfrac{1}{4}e_1(D, N) - \tfrac{1}{3}e_3(D, N)\big). \tag{6}$$

Lemma 10.7. *There are exactly* 574 *pairs* (D, N) *consisting of a quaternion discriminant* $D > 1$ *over* $\mathbb{Q}$ *and a positive integer* N *coprime to* D *such that* (6) *does not hold. For all such pairs we have* $d_{\mathrm{CM}}(X_0^D(1)) \in \{2, 4, 6\}$, *and the largest value of* D *occurring among such pairs is* $D = 1590$.

Proof. This follows from direct computation. The 574 referenced pairs are listed in the file `fail_dcm_check.m` in [Saia 2024]. □

Lemma 10.8. *Set*

$$\mathcal{D} := \{85, 91, 93, 115, 123, 133, 141, 142, 145, 155, 158, 161, 177, 178, 183, 185, 187, 201, 202, 203, 205, 209, 213, 214, 217, 218, 219, 221, 226, 235, 237, 247, 249, 253, 254, 259, 262, 265, 267, 274, 278, 287, 291, 295, 298, 299, 301, 302, 303, 305, 309, 319, 321, 323, 326, 327, 329, 334, 335, 339, 341, 346, 355, 358, 362, 365, 371, 377, 381, 382, 386, 391, 393, 394, 395, 398, 403, 407, 411, 413, 415, 417, 422, 427, 437, 445, 446, 447, 451, 453, 454, 458, 462, 466, 469, 471, 478, 482, 485, 489, 501, 502, 505, 514, 519, 526, 537, 538, 542, 543, 554, 562, 566, 570, 573, 579, 586, 591, 597, 614, 622, 626, 634, 662, 674, 690, 694, 698, 706, 714, 718, 734, 746, 758, 766, 770, 778, 794, 798, 802, 838, 858, 870, 910, 930, 966, 1110, 1122, 1190, 1218, 1230, 1254, 1290, 1302, 1326, 1410, 1518, 1590\}.$$

and

$$\mathcal{E} := \{(85, 2), (85, 3), (85, 4), (91, 2), (91, 3), (93, 2), (93, 4), (93, 5), (115, 2), (115, 3), (123, 2), (133, 2), (141, 2), (142, 3), (145, 2), (155, 2), (158, 3), (161, 2), (177, 2), (178, 3), (183, 2), (201, 2), (202, 3)\}.$$

For each of the 181 *pairs* (D, N) *with either* $D \in \mathcal{D}$ *and* $N = 1$ *or with* $(D, N) \in \mathcal{E}$, *the curve* $X_0^D(N)_{/\mathbb{Q}}$ *has a sporadic CM point.*

Proof. For each such pair (D, N), we know from Section 10.2 that $X_0^D(1)_{/\mathbb{Q}}$ does not have infinitely many degree 2 points, and hence $X_0^D(N)_{/\mathbb{Q}}$ does not have infinitely many degree 2 points. At the same time, we compute that this curve has a CM point of degree 2, which is therefore necessarily sporadic. □

We are now prepared to end with the main result of this section:

Theorem 10.9. (1) *For each of the* 64 *pairs* (D, N) *in Table 1, the Shimura curve* $X_0^D(N)_{/\mathbb{Q}}$ *has no sporadic points, and we have* $d_{\text{CM}}(X_0^D(N)) = 2$.

(2) *For each of the* 64 *pairs* (D, N) *in Table 1 except for possibly the* 10 *in the set*

$$\{(6, 5), (6, 7), (6, 13), (6, 19), (6, 29), (6, 31), (6, 37), (10, 7), (14, 5), (22, 5)\},$$

the Shimura curve $X_1^D(N)_{/\mathbb{Q}}$ *has no sporadic CM points.*

(3) *There are at most* 329 *pairs* (D, N), *consisting of an indefinite quaternion discriminant* $D > 1$ *over* $\mathbb{Q}$ *and a positive integer* N *coprime to* D, *which do not*

(6, 1)	(6, 5)	(6, 7)	(6, 11)	(6, 13)	(6, 19)	(6, 29)	(6, 31)
(6, 37)	(10, 1)	(10, 3)	(10, 7)	(10, 11)	(10, 23)	(14, 1)	(14, 5)
(15, 1)	(15, 2)	(21, 1)	(22, 1)	(22, 3)	(22, 5)	(26, 1)	(33, 1)
(34, 1)	(35, 1)	(38, 1)	(39, 1)	(39, 2)	(46, 1)	(51, 1)	(55, 1)
(57, 1)	(58, 1)	(62, 1)	(65, 1)	(69, 1)	(74, 1)	(77, 1)	(82, 1)
(86, 1)	(87, 1)	(94, 1)	(95, 1)	(106, 1)	(111, 1)	(118, 1)	(119, 1)
(122, 1)	(129, 1)	(134, 1)	(143, 1)	(146, 1)	(159, 1)	(166, 1)	(194, 1)
(206, 1)	(210, 1)	(215, 1)	(314, 1)	(330, 1)	(390, 1)	(510, 1)	(546, 1)

Table 1. 64 pairs (D, N) with $\gcd(D, N) = 1$ for which $\delta(X_0^D(N)) = 2$, and hence $X_0^D(N)_{/\mathbb{Q}}$ has no sporadic points.

appear among the 64 *listed in Table 1 and for which the Shimura curve $X_0^D(N)_{/\mathbb{Q}}$ does not have a sporadic CM point. These are listed in Table 2.*

(4) *Let (D, N) be a pair consisting of an indefinite quaternion discriminant $D > 1$ over $\mathbb{Q}$ and a positive integer N coprime to D. If (D, N) is not listed in Table 1 or Table 2 and is not equal to* $(91, 5)$, *then the Shimura curve $X_1^D(N)_{/\mathbb{Q}}$ has a sporadic CM point.*

Proof. (1) These Shimura curves $X_0^D(N)_{/\mathbb{Q}}$ are exactly those for which we know that $\delta(X_0^D(N)) = 2$ via Section 10.2. That each such curve has a CM point of degree 2 follows from direct computation.

(2) For each pair in this table, we have

$$\delta(X_1^D(N)) \leq 2 \cdot \deg(X_1^D(N) \to X_0^D(N)) = \max\{2, \phi(N)\}.$$

For each pair in this table other than the 10 listed pairs, we compute that

$$\max\{2, \phi(N)\} \leq d_{\mathrm{CM}}(X_1^D(N)).$$

(3) This is an immediate consequence of the preceding discussion, including Lemmas 10.7 and 10.8.

(4) By Proposition 10.5, we have that $X_1^D(N)_{/\mathbb{Q}}$ has a sporadic CM point for all pairs (D, N) aside from possibly the 574 referred to in Lemma 10.7. Of the 181 pairs listed in Lemma 10.8, we compute that each pair except for $(D, N) = (91, 5)$ satisfies

$$d_{\mathrm{CM}}(X_1^D(N)) = 2 < \delta(X_0^D(1)) \leq \delta(X_1^D(N)),$$

and hence we have a sporadic CM point on $X_1^D(N)_{/\mathbb{Q}}$ for all such pairs. The result then follows from part (2). □

Remark 10.10. For all 329 pairs (D, N) listed in Table 2, we have $d_{\mathrm{CM}}(X_0^D(N)) \in \{2, 4, 6\}$. For all but 56 of these pairs, we have $d_{\mathrm{CM}}(X_0^D(N)) = 2$. For such pairs,

(6, 17)	(6, 23)	(6, 25)	(6, 35)	(6, 41)	(6, 43)	(6, 47)
(6, 49)	(6, 53)	(6, 55)	(6, 59)	(6, 61)	(6, 65)	(6, 67)
(6, 71)	(6, 73)	(6, 77)	(6, 79)	(6, 83)	(6, 85)	(6, 89)
(6, 91)	(6, 95)	(6, 97)	(6, 101)	(6, 103)	(6, 107)	(6, 109)
(6, 113)	(6, 115)	(6, 119)	(6, 121)	(6, 125)	(6, 127)	(6, 131)
(6, 133)	(6, 137)	(6, 139)	(6, 143)	(6, 145)	(6, 149)	(6, 151)
(6, 155)	(6, 157)	(6, 161)	(6, 163)	(6, 167)	(6, 169)	(6, 173)
(6, 179)	(6, 181)	(6, 191)	(6, 193)	(6, 197)	(6, 199)	(6, 203)
(6, 287)	(6, 295)	(6, 319)	(10, 9)	(10, 13)	(10, 17)	(10, 19)
(10, 21)	(10, 27)	(10, 29)	(10, 31)	(10, 33)	(10, 37)	(10, 39)
(10, 41)	(10, 43)	(10, 47)	(10, 49)	(10, 51)	(10, 53)	(10, 57)
(10, 59)	(10, 61)	(10, 63)	(10, 67)	(10, 69)	(10, 71)	(10, 73)
(10, 77)	(10, 79)	(10, 83)	(10, 87)	(10, 89)	(10, 91)	(10, 97)
(10, 103)	(10, 119)	(10, 141)	(10, 161)	(10, 191)	(14, 3)	(14, 9)
(14, 11)	(14, 13)	(14, 15)	(14, 17)	(14, 19)	(14, 23)	(14, 25)
(14, 27)	(14, 29)	(14, 31)	(14, 33)	(14, 37)	(14, 39)	(14, 41)
(14, 43)	(14, 47)	(14, 53)	(14, 59)	(14, 61)	(14, 87)	(14, 95)
(15, 4)	(15, 7)	(15, 8)	(15, 11)	(15, 13)	(15, 14)	(15, 16)
(15, 17)	(15, 19)	(15, 22)	(15, 23)	(15, 26)	(15, 28)	(15, 29)
(15, 31)	(15, 32)	(15, 34)	(15, 37)	(15, 41)	(15, 43)	(15, 47)
(21, 2)	(21, 4)	(21, 5)	(21, 8)	(21, 10)	(21, 11)	(21, 13)
(21, 16)	(21, 17)	(21, 19)	(21, 23)	(21, 25)	(21, 29)	(21, 31)
(21, 38)	(22, 7)	(22, 9)	(22, 13)	(22, 15)	(22, 17)	(22, 19)
(22, 21)	(22, 23)	(22, 25)	(22, 27)	(22, 29)	(22, 31)	(22, 35)
(22, 37)	(22, 51)	(26, 3)	(26, 5)	(26, 7)	(26, 9)	(26, 11)
(26, 15)	(26, 17)	(26, 19)	(26, 21)	(26, 23)	(26, 25)	(26, 29)
(26, 31)	(33, 2)	(33, 4)	(33, 5)	(33, 7)	(33, 8)	(33, 10)
(33, 13)	(33, 16)	(33, 17)	(33, 19)	(34, 3)	(34, 5)	(34, 7)
(34, 9)	(34, 11)	(34, 13)	(34, 15)	(34, 19)	(34, 23)	(34, 29)
(34, 35)	(35, 2)	(35, 3)	(35, 4)	(35, 6)	(35, 8)	(35, 9)
(35, 11)	(35, 12)	(35, 13)	(38, 3)	(38, 5)	(38, 7)	(38, 9)
(38, 11)	(38, 13)	(38, 17)	(38, 21)	(39, 4)	(39, 5)	(39, 7)
(39, 8)	(39, 10)	(39, 11)	(39, 31)	(46, 3)	(46, 5)	(46, 7)
(46, 9)	(46, 11)	(46, 13)	(46, 15)	(46, 17)	(51, 2)	(51, 4)

Table 2. All 329 pairs (D, N) with $D > 1$ for which we remain unsure whether $X_0^D(N)_{/\mathbb{Q}}$ has a sporadic CM point (continued below).

(51, 5)	(51, 7)	(51, 8)	(51, 10)	(51, 11)	(51, 20)	(55, 2)
(55, 3)	(55, 4)	(55, 7)	(55, 8)	(57, 2)	(57, 4)	(57, 5)
(57, 7)	(58, 3)	(58, 5)	(58, 7)	(58, 9)	(58, 11)	(58, 13)
(62, 3)	(62, 5)	(62, 7)	(62, 9)	(62, 11)	(62, 15)	(65, 2)
(65, 3)	(65, 4)	(65, 7)	(69, 2)	(69, 4)	(69, 5)	(69, 7)
(69, 11)	(74, 3)	(74, 5)	(74, 7)	(77, 2)	(77, 3)	(77, 4)
(77, 5)	(77, 6)	(82, 3)	(82, 5)	(82, 7)	(86, 3)	(86, 5)
(86, 7)	(87, 2)	(87, 4)	(87, 5)	(87, 8)	(94, 3)	(94, 5)
(94, 7)	(95, 2)	(95, 3)	(106, 3)	(106, 5)	(106, 7)	(111, 2)
(111, 4)	(118, 3)	(118, 5)	(119, 2)	(119, 3)	(119, 6)	(122, 3)
(122, 5)	(122, 7)	(129, 2)	(129, 7)	(134, 3)	(134, 5)	(134, 9)
(143, 2)	(143, 4)	(146, 3)	(146, 7)	(159, 2)	(166, 3)	(183, 5)
(194, 3)	(215, 2)	(215, 3)	(326, 3)	(327, 2)	(335, 2)	(390, 7)

Table 2. (continued).

it follows that the curve $X_0^D(N)_{/\mathbb{Q}}$ has a sporadic (CM) point if and only if it is not bielliptic with a degree 2 map to an elliptic curve over $\mathbb{Q}$ of positive rank. An extension of the results of [Rotger 2002] mentioned in Section 10.2 to general level N would then allow us to determine whether $X_0^D(N)_{/\mathbb{Q}}$ has a sporadic CM point for all but at most 56 pairs (D, N) with $D > 1$. Such an extension will appear in work of the author and Oana Padurariu [2024].

Acknowledgments

We thank Pete L. Clark for initially suggesting this project, which served as the author's main Ph.D. thesis work. We are very grateful for Clark's excellent advising, and for innumerable helpful conversations. We also thank Dino Lorenzini, Oana Padurariu, Ciaran Schembri and John Voight for helpful comments and encouragement. Much gratitude goes as well to the anonymous referee for a careful reading of, and detailed feedback on, earlier versions of this work. This feedback led to many expository and mathematical improvements, including a strengthening of our results on sporadic points.

We would also like to thank the Simons Foundation for access to the [Bosma et al. 1997] computational algebra system, with which all computations described in this paper were performed, as well as the Georgia Advanced Computing Resource Center at the University of Georgia for access to the high-performance computing cluster, which we used for various computations described in Section 10. Partial support for the author was provided by the Research and Training Group grant DMS-1344994 funded by the National Science Foundation.

References

[Abramovich 1996] D. Abramovich, "A linear lower bound on the gonality of modular curves", *Internat. Math. Res. Notices* 20 (1996), 1005–1011. MR Zbl

[Abramovich and Harris 1991] D. Abramovich and J. Harris, "Abelian varieties and curves in $W_d(C)$", *Compositio Math.* **78**:2 (1991), 227–238. MR Zbl

[Alsina and Bayer 2004] M. Alsina and P. Bayer, *Quaternion orders, quadratic forms, and Shimura curves*, CRM Monograph Series **22**, American Mathematical Society, Providence, RI, 2004. MR Zbl

[Bayer and Travesa 2007] P. Bayer and A. Travesa, "Uniformizing functions for certain Shimura curves, in the case $D = 6$", *Acta Arith.* **126**:4 (2007), 315–339. MR Zbl

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR Zbl

[Bourdon and Clark 2020] A. Bourdon and P. L. Clark, "Torsion points and Galois representations on CM elliptic curves", *Pacific J. Math.* **305**:1 (2020), 43–88. MR Zbl

[Bourdon and Najman 2021] A. Bourdon and F. Najman, "Sporadic points of odd degree on $X_1(N)$ coming from $\mathbb{Q}$-curves", preprint, 2021. arXiv 2107.10909v2

[Bourdon et al. 2019] A. Bourdon, O. Ejder, Y. Liu, F. Odumodu, and B. Viray, "On the level of modular curves that give rise to isolated j-invariants", *Adv. Math.* **357** (2019), art. id. 106824. MR Zbl

[Bourdon et al. 2024] A. Bourdon, D. R. Gill, J. Rouse, and L. D. Watson, "Odd degree isolated points on $X_1(N)$ with rational j-invariant", *Res. Number Theory* **10**:1 (2024), art. id. 5. MR Zbl

[Buzzard 1997] K. Buzzard, "Integral models of certain Shimura curves", *Duke Math. J.* **87**:3 (1997), 591–612. MR Zbl

[Clark 2009] P. L. Clark, "On the Hasse principle for Shimura curves", *Israel J. Math.* **171** (2009), 349–365. MR Zbl

[Clark 2022] P. L. Clark, "CM elliptic curves, I: volcanoes, reality and applications", preprint, 2022. arXiv 2212.13316

[Clark and Saia 2022] P. L. Clark and F. Saia, "CM elliptic curves, II: Volcanoes, reality and applications", preprint, 2022. arXiv 2212.13327

[Clark and Stankewicz 2018] P. L. Clark and J. Stankewicz, "Hasse principle violations for Atkin–Lehner twists of Shimura curves", *Proc. Amer. Math. Soc.* **146**:7 (2018), 2839–2851. MR Zbl

[Clark et al. 2013] P. L. Clark, B. Cook, and J. Stankewicz, "Torsion points on elliptic curves with complex multiplication", *Int. J. Number Theory* **9**:2 (2013), 447–479. MR Zbl

[Clark et al. 2022] P. L. Clark, T. Genao, P. Pollack, and F. Saia, "The least degree of a CM point on a modular curve", *J. Lond. Math. Soc.* (2) **105**:2 (2022), 825–883. MR Zbl

[Cox 2013] D. A. Cox, *Primes of the form $x^2 + ny^2$: Fermat, class field theory, and complex multiplication*, 2nd ed., Wiley, Hoboken, NJ, 2013. MR Zbl

[Diamond and Shurman 2005] F. Diamond and J. Shurman, *A first course in modular forms*, Graduate Texts in Mathematics **228**, Springer, 2005. MR Zbl

[Fouquet 2001] M. Fouquet, *Anneau d'endomorphismes et cardinalité des courbes elliptiques: aspects algorithmiques*, Ph.D. thesis, École Polytechnique, 2001, available at https://theses.fr/2001EPXX0051.

[Fouquet and Morain 2002] M. Fouquet and F. Morain, "Isogeny volcanoes and the SEA algorithm", pp. 276–291 in *Algorithmic number theory* (Sydney, 2002), edited by C. Fieker and D. R. Kohel, Lecture Notes in Comput. Sci. **2369**, Springer, 2002. MR Zbl

[Frey 1994] G. Frey, "Curves with infinitely many points of fixed degree", *Israel J. Math.* **85**:1-3 (1994), 79–83. MR Zbl

[González and Rotger 2006] J. González and V. Rotger, "Non-elliptic Shimura curves of genus one", *J. Math. Soc. Japan* **58**:4 (2006), 927–948. MR Zbl

[Guo and Yang 2017] J.-W. Guo and Y. Yang, "Equations of hyperelliptic Shimura curves", *Compos. Math.* **153**:1 (2017), 1–40. MR Zbl

[Jordan 1981] B. W. Jordan, *On the Diophantine arithmetic of Shimura curves*, Ph.D. thesis, Harvard University, 1981, available at https://www.proquest.com/docview/303012662. MR Zbl

[Kani 2011] E. Kani, "Products of CM elliptic curves", *Collect. Math.* **62**:3 (2011), 297–339. MR Zbl

[Katsura 1975] T. Katsura, "On the structure of singular abelian varieties", *Proc. Japan Acad.* **51**:4 (1975), 224–228. MR Zbl

[Kim 2003] H. H. Kim, "Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2", *J. Amer. Math. Soc.* **16**:1 (2003), 139–183. MR Zbl

[Kohel 1996] D. R. Kohel, *Endomorphism rings of elliptic curves over finite fields*, Ph.D. thesis, University of California, Berkeley, 1996, available at https://www.proquest.com/docview/304241260. MR Zbl

[Lange 1975] H. Lange, "Produkte elliptischer Kurven", *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II* 8 (1975), 95–108. MR Zbl

[Najman 2016] F. Najman, "Torsion of rational elliptic curves over cubic fields and sporadic points on $X_1(n)$", *Math. Res. Lett.* **23**:1 (2016), 245–272. MR Zbl

[Ogg 1983] A. P. Ogg, "Real points on Shimura curves", pp. 277–307 in *Arithmetic and geometry*, *I*, edited by M. Artin and J. Tate, Progr. Math. **35**, Birkhäuser, Boston, 1983. MR Zbl

[Padurariu and Saia 2024] O. Padurariu and F. Saia, "Bielliptic Shimura curves $X_0^D(N)$ with nontrivial level", preprint, 2024. arXiv 2401.08829v2

[Padurariu and Schembri 2023] O. Padurariu and C. Schembri, "Rational points on Atkin–Lehner quotients of geometrically hyperelliptic Shimura curves", *Expo. Math.* **41**:3 (2023), 492–513. MR Zbl

[Rotger 2002] V. Rotger, "On the group of automorphisms of Shimura curves and applications", *Compositio Math.* **132**:2 (2002), 229–241. MR Zbl

[Rotger 2004] V. Rotger, "Shimura curves embedded in Igusa's threefold", pp. 263–276 in *Modular curves and abelian varieties*, edited by J. Cremona et al., Progr. Math. **224**, Birkhäuser, Basel, 2004. MR Zbl

[Rotger et al. 2005] V. Rotger, A. Skorobogatov, and A. Yafaev, "Failure of the Hasse principle for Atkin–Lehner quotients of Shimura curves over $\mathbb{Q}$", *Mosc. Math. J.* **5**:2 (2005), 463–476. MR Zbl

[Saia 2024] F. Saia, "CM-points-Shimura-curves", Github repository, 2024, available at https://github.com/fsaia/CM-Points-Shimura-Curves.

[Schoen 1992] C. Schoen, "Produkte Abelscher Varietäten und Moduln über Ordnungen", *J. Reine Angew. Math.* **429** (1992), 115–123. MR Zbl

[Shimura 1966] G. Shimura, "Moduli and fibre systems of abelian varieties", *Ann. of Math.* (2) **83** (1966), 294–338. MR Zbl

[Shimura 1967] G. Shimura, "Construction of class fields and zeta functions of algebraic curves", *Ann. of Math.* (2) **85** (1967), 58–159. MR Zbl

[Shimura 1972] G. Shimura, "On the field of rationality for an abelian variety", *Nagoya Math. J.* **45** (1972), 167–178. MR Zbl

[Shimura 1975] G. Shimura, "On the real points of an arithmetic quotient of a bounded symmetric domain", *Math. Ann.* **215** (1975), 135–164. MR Zbl

[Shioda and Mitani 1974] T. Shioda and N. Mitani, "Singular abelian surfaces and binary quadratic forms", pp. 259–287 in *Classification of algebraic varieties and compact complex manifolds*, edited by H. Popp, Lecture Notes in Math. **412**, Springer, 1974. MR Zbl

[Siksek and Skorobogatov 2003] S. Siksek and A. Skorobogatov, "On a Shimura curve that is a counterexample to the Hasse principle", *Bull. London Math. Soc.* **35**:3 (2003), 409–414. MR Zbl

[Smith 2023] H. Smith, "Ramification in division fields and sporadic points on modular curves", *Res. Number Theory* **9**:1 (2023), art. id. 17. MR Zbl

[Stevenhagen 2001] P. Stevenhagen, "Hilbert's 12th problem, complex multiplication and Shimura reciprocity", pp. 161–176 in *Class field theory—its centenary and prospect* (Tokyo, 1998), edited by K. Miyake, Adv. Stud. Pure Math. **30**, Math. Soc. Japan, Tokyo, 2001. MR Zbl

[Sutherland 2013] A. V. Sutherland, "Isogeny volcanoes", pp. 507–530 in *ANTS X—Proceedings of the Tenth Algorithmic Number Theory Symposium*, edited by E. W. Howe and K. S. Kedlaya, Open Book Ser. **1**, Math. Sci. Publ., Berkeley, CA, 2013. MR Zbl

[Ufer 2010] D. Ufer, *Shimura–Kurven, endomorphismen und q-parameter*, Ph.D. thesis, Universität Ulm, 2010, available at https://tinyurl.com/ufer-thesis.

[Vignéras 1980] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics **800**, Springer, 1980. MR Zbl

[Voight 2009] J. Voight, "Shimura curves of genus at most two", *Math. Comp.* **78**:266 (2009), 1155–1172. MR Zbl

[Voight 2021] J. Voight, *Quaternion algebras*, Graduate Texts in Mathematics **288**, Springer, 2021. MR Zbl

Received July 24, 2023. Revised September 6, 2024.

FREDERICK SAIA
DEPARTMENT OF MATHEMATICS, STATISTICS, AND COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS CHICAGO
CHICAGO, IL
UNITED STATES
fsaia@uic.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.385

STRATIFICATION OF THE MODULI SPACE OF VECTOR BUNDLES

MONTSERRAT TEIXIDOR I BIGAS

We show that on a generic curve, a bundle obtained by generic successive extensions is stable, so long as the slopes satisfy natural conditions. We compute the dimension of the set of such extensions. We use degeneration methods specializing the curve to a chain of elliptic components. This extends our previous work (1998, 2000).

Take a generic (compact, nonsingular) curve C of genus g defined over the complex numbers. A vector bundle E on C of rank r and degree d is said to be stable (resp. semistable) if for every vector subbundle E_1 of E of rank r_1 and degree d_1, the inequality $\frac{d_1}{r_1} < \frac{d}{r}$ (resp. $\leq$) is satisfied. The moduli space $\mathcal{U}(r, d)(C)$ parametrizes isomorphism classes of stable vector bundles of rank r and degree d on C. It is a nonsingular variety of dimension $r^2(g-1)+1$ which can be compactified by equivalence classes of semistable vector bundles.

Fix $E \in \mathcal{U}(r, d)(C)$ and an integer $r_1 < r$. Define $s_{r_1}(E) = r_1 d - r \max\{\deg E_1\}$ where E_1 moves in the set of subbundles of E of rank r_1. As E is stable, $s_{r_1}(E) > 0$ for all r_1. On the other hand, for a generic E, $s_{r_1}(E)$ is the smallest integer greater than or equal to $r_1(r-r_1)(g-1)$ and congruent with $r_1 d$ modulo r [Lange 1983, Satz, p. 448; Lange and Narasimhan 1983]. Fix then s such that $0 < s \leq r_1(r-r_1)(g-1)$. The (proper) subset of the moduli space of vector bundles given as

$$\mathcal{U}^s(r, d)(C) = \{E \in \mathcal{U}(r, d)(C) \text{ such that } s_{r_1}(E) = s\}$$

generically coincides with the space of stable bundles $E_{r,d}$ that fit in an exact sequence

$$0 \to E_{r_1,d_1} \to E_{r,d} \to \bar{E}_{r-r_1,d-d_1} \to 0.$$

Lange conjectured that a generic choice of the two bundles E_{r_1,d_1}, $\bar{E}_{r-r_1,d-d_1}$ together with a generic choice of the extension would give rise to a stable $E_{r,d}$ and therefore $\mathcal{U}^s(r, d)(C)$ would be nonempty and of the expected dimension. Lange's conjecture was proved in full generality in [Russo and Teixidor 1999].

MSC2020: 14H60.
Keywords: vector bundle, algebraic curve, extensions.

Our goal here is to extend this result to the case of several successive extensions. We show the following:

Theorem 0.1. *Let C be a generic curve of genus g. Fix a rank r and degree d. Choose a collection of integers $r_1 < r_2 < \cdots < r_k = r$ and degrees $d_1, \ldots, d_k = d$ such that*

$$\frac{d_1}{r_1} < \frac{d_2}{r_2} < \cdots < \frac{d_k}{r_k}$$

and

$$\begin{gathered} r_1 d_2 - r_2 d_1 \leq r_1(r_2 - r_1)(g-1), \\ r_2 d_3 - r_3 d_2 \leq r_2(r_3 - r_2)(g-1), \\ \ldots, \\ r_{k-1} d_k - r_k d_{k-1} \leq r_{k-1}(r_k - r_{k-1})(g-1). \end{gathered}$$

Define $\mathcal{U}(r_1, \ldots, r_k; d_1, \ldots, d_k)(C) \subseteq \mathcal{U}(r, d)(C)$ as the set of stable $E_{r,d}$ obtained after a sequence of extensions

$$\begin{gathered} 0 \to E_{r_1,d_1} \to E_{r_2,d_2} \to \bar{E}_{r_2-r_1,d_2-d_1} \to 0, \\ 0 \to E_{r_2,d_2} \to E_{r_3,d_3} \to \bar{E}_{r_3-r_2,d_3-d_2} \to 0, \\ \ldots, \\ 0 \to E_{r_{k-1},d_{k-1}} \to E_{r,d} \to \bar{E}_{r-r_{k-1},d-d_{k-1}} \to 0. \end{gathered}$$

Then, $\mathcal{U}(r_1, \ldots, r_k; d_1, \ldots, d_k)(C)$ is nonempty, irreducible and of codimension (in $\mathcal{U}(r, d)(C)$)

$$\begin{aligned} &(r_1(r_2 - r_1) + r_2(r_3 - r_2) + \cdots + r_k(r_k - r_{k-1}))(g-1) \\ &\qquad - (r_1 d_2 - r_2 d_1 + r_2 d_3 - r_3 d_2 + \cdots + r_{k-1} d_k - r_k d_{k-1}) \end{aligned}$$

Following the ideas in [Teixidor 1998; 2000], we will use a degeneration argument. We first prove the result for a particular reducible nodal curve and then we extend it to the generic curve. The condition on the curve being generic is a by-product of our method of proof. It is unlikely to be necessary. On the other hand, the genericity of the extensions seems essential. This was already the case when we considered a single extension. Then, we were able to give geometric conditions identifying the nonstable extensions (see [Russo and Teixidor 1999]).

We became interested in this question while studying families of rational curves in the moduli space of vector bundles on a fixed curve. The stability of a successive extension is crucial for the construction of families of rational curves in the moduli space (see [Mustopa and Teixidor 2024]).

The result in the case of a single extension has found applications to a variety of other topics, in particular to the study of Brill–Noether theory for vector bundles (see, for instance, [Casalaina-Martin and Teixidor 2011; Hitching et al. 2021; Kopper and Mandal 2023]) and to the existence of Ulrich bundles on ruled surfaces (see [Aprodu et al. 2020]). We expect that this more general result will find similar applications.

1. The problem on the reducible curve

Tensoring with line bundles results in isomorphisms between $\mathcal{U}(r,d)(C)$ and $\mathcal{U}(r,d+tr)(C)$, so there are only, up to isomorphism, at most r nonisomorphic moduli spaces of vector bundles on a curve (in fact, about half of that if you consider also dualization, but this is irrelevant to us now). Without loss of generality, we will assume that $0 \le d < r$.

We will be using bundles on reducible, nodal curves as limits of vector bundles on nonsingular curves. More specifically, we will consider chains of elliptic curves defined as follows: Let $C_1, \dots, C_g$ be g elliptic curves, $P_i \neq Q_i$ arbitrary points on C_i. Glue Q_i to P_{i+1}, $i = 1, \dots, g-1$, to form a nodal curve C of genus g that we call a chain of elliptic curves.

On a reducible curve, stability for a bundle depends on a choice of polarization. A polarization is usually defined as the choice of a line bundle on the variety. For our goal of defining stability of a vector bundle, what matters is the relative degree of the restriction of this line bundle to each component, that is, the numbers

$$a_i = \frac{\deg L_{|C_i}}{\deg L}, \quad i = 1, \dots, g,\ 0 < a_i < 1,\ \sum a_i = 1.$$

Then, a vector bundle E on C of constant rank r is said to be a_i-stable if for every subsheaf F of E

$$\frac{\chi(F)}{\sum a_i \operatorname{rank}(F_{|C_i})} < \frac{\chi(E)}{r}.$$

Fixing a polarization, there is a moduli space of (a_i)-stable vector bundles on the chain of elliptic curves (see [Seshadri 1982]). Stability forces the degree of the restriction of the vector bundle to each component of the curve to vary in certain intervals that depend on the (a_j). The moduli space of (a_j)-stable vector bundles on the chain is reducible, each component $M_{(d_j)}$ corresponding to a choice of allowable degrees d_i on the component C_i. Let us look now at the particular case in which the components of C are elliptic

From results of Atiyah [1957], a vector bundle on an elliptic curve is stable if and only if it is indecomposable (not a direct sum of other bundles) of coprime rank and degree. The indecomposable vector bundles that are not of coprime rank and degree are semistably equivalent to a direct sum of stable (and therefore indecomposable) vector bundles all of the same rank and degree. This is similar to what happens for vector bundle on a rational curve, where each vector bundle is a direct sum of line bundles. On rational curves, vector bundles appearing in families tend to be balanced, that is, most of the time they have degrees that differ in just one unit. An analogous result holds for families of vector bundles on reducible curves whose components are elliptic; we describe these known results below.

Fix a collection (d_j) in the range of degrees allowed by stability and denote by $M_{(d_j)}$ the component of the moduli space of vector bundles on C associated to this choice. Write h_i for the greatest common divisor of d_i, r, $d'_i = \frac{d_i}{h_i}$ and rank $r'_i = \frac{r}{h_i}$. Then, a vector bundle corresponding to a generic point in $M_{(d_j)}$ restricts to C_i to a direct sum of h_i indecomposable bundles of degree d'_i and rank r'_i. That is, a vector bundle corresponding to a generic point in M_{d_i} restricts to C_i to a direct sum of stable vector bundles all of the same rank and degree (see [Teixidor 1991, Theorem; 1995, Theorem 3.2]).

Our goal is to use results from the chain of elliptic curves to deduce similar conditions for nonsingular curves. When dealing with a family of curves in which the general member is nonsingular and the special member is the chain of elliptic curves, we can modify a vector bundle in the family tensoring with a line bundle with support on the special fiber. This action leaves the vector bundle on the general fiber unchanged but moves the degree on the various components of the special fiber by multiples of the rank. This allows us to choose the distribution of degrees among the components up to multiples of r and ignore the actual distribution of degrees among components of the curve imposed by the polarization, focusing instead on the remaining conditions needed for stability.

Proposition 1.1. *Let C be a chain of elliptic curves of genus g. Fix a rank r and degree d, $0 \le d < r$, and a collection of integers $r_1 < r_2 < \cdots < r_k = r$. Choose degrees $d_1, \ldots, d_k = d$ with d_{k-1} the largest degree such that $\frac{d_{k-1}}{r_{k-1}} < \frac{d_k}{r_k}$, d_{k-2} the largest degree such that $\frac{d_{k-2}}{r_{k-2}} < \frac{d_{k-1}}{r_{k-1}}, \ldots, d_1$ the largest degree such that $\frac{d_1}{r_1} < \frac{d_2}{r_2}$. Then, there exists a stable bundle E that contains a chain of subbundles*

$$E_{r_1,d_1} \subseteq E_{r_2,d_2} \subseteq \cdots \subseteq E_{r_k,d_k} = E$$

with E_{r_i,d_i} stable of degree d_i and rank r_i. This E contains at most a finite number of such chains.

Proof. On the moduli space of vector bundles on the chain of elliptic curves, we focus on the component whose restriction have degree $d < r$ on C_1 and degree zero on the remaining components. Write h for the greatest common divisor of d, r. Define d', r' by $d = hd'$, $r = hr'$. As explained above, on the chosen component of the moduli space of vector bundles, the generic vector bundle restricts to a direct sum of h indecomposable bundles of degree d' and rank r' on C_1 and as a direct sum of line bundles of degree zero on $C_2, \dots, C_g$.

More generally and in keeping with the discussion above, for any r_i, d_i, we will say that E_{r_i,d_i} is a generic vector bundle of degree d_i and rank r_i if it is a direct sum of h_i indecomposable vector bundles of coprime rank and degree:

$$E_{r_i,d_i} = \bigoplus_{j=1}^{h_i} F_i^j, \quad h_i = \gcd(r_i, d_i),\ r_i = h_i r_i',\ d_i = h_i d_i',\ \deg F_i = r_i',\ \operatorname{rank} F_i = r_i'.$$

On C_1, choose a generic vector bundle $E^1_{r_1,d_1}$ of degree d_1 and rank r_1, a generic vector bundle $E^1_{r_2,d_2}$ of degree d_2 and rank $r_2, \dots,$ a generic vector bundle $E^1_{r_k,d_k}$ of degree d_k and rank r_k.

The conditions $\frac{d_1}{r_1} < \frac{d_2}{r_2} < \cdots < \frac{d_k}{r_k}$ guarantee (see [Teixidor 2000, Lemma 2.5]) that there exist inclusions

$$E^1_{r_1,d_1} \subseteq E^1_{r_2,d_2} \subseteq \cdots \subseteq E^1_{r_k,d_k}.$$

In fact, as $E^1_{r_i,d_i} = \bigoplus_{j=1}^{h_i} F_i^j$, $E^1_{r,d} = E_{r_k,d_k} = \bigoplus_{j=1}^{h} F_k^j$, $\operatorname{Hom}(E^1_{r_i,d_i}, E^1_{r,d}) = \bigoplus_{j,j'} \operatorname{Hom}(F_i^j, F_k^{j'})$. Then, from

$$\frac{d_i'}{r_i'} = \frac{d_i}{r_i} < \frac{d}{r} = \frac{d'}{r'},$$

the space of morphisms of F_i^j to $F_k^{j'}$ has dimension $r_i'd' - r'd_i'$. Therefore, the space of morphisms of $E^1_{r_i,d_i}$ to $E^1_{r,d}$ has dimension $hh_i(r_i'd' - r'd_i') = r_i d - r d_i \geq 0$, in particular, it is nonempty. We can choose the inclusions from $E^1_{r_i,d_i}$ into E_{r_k,d_k} so that the image does not coincide with any of the finite number of destabilizing subsheaves of E_{r_k,d_k} (it is enough to make sure that none of the morphisms F_i^j to $F_k^{j'}$ is zero).

We now describe a vector bundle on the chain by giving a vector bundle on each component and the gluing at the nodes:

On the curve C^1 take the vector bundle $E^1_{r_k,d_k} = E^1_{r,d}$ we just described. On the curves $C_2, \dots, C_g$, choose a direct sum of r line bundles of degree zero. On each of $C_2, \dots, C_g$, select a first set of r_1 among the r line bundles in the direct sum. Select then a second set of r_2 among the r line bundles containing the initial subset of r_1 already chosen, Select a third set of r_3 line bundles containing the subset of r_2 chosen in the previous step and so on. Form now a bundle on the chain of elliptic curves by gluing the bundles on each component so that when identifying Q_i with P_{i+1}, $i = 2, \dots, g-1$, each of the sets of r_j line bundles $j = 1, \dots, k$ on C_i glues with the set of r_j line bundles on C_{i+1}, $j = 1, \dots, k$ (but the gluings are otherwise generic). At Q_1, glue each set of r_j line subbundles on C_2 with the fiber of the image of the $E^1_{r_j,d_j}$ (but the gluings are otherwise generic). In this way, we obtain bundles of ranks $r_1 < r_2 < \cdots < r_k$ and degrees $d_1, \dots, d_k$ on the whole curve C each contained in the next.

On a reducible nodal curve, gluing vector bundles that are semistable on each of the components and of the degrees allowed by the polarization, one obtains a semistable bundle on the whole curve. Moreover, if one of the bundles we are gluing is stable or if none of the subbundles that contradict stability glue with each other, the whole vector bundle on the reducible curve is stable (see [Teixidor 1991; 1995]).

By construction, the vector bundles on each C_i are semistable. On C_1, the only subbundles of the bundle $E^1_{r_k,d_k} = \bigoplus_{j=1}^h E^j_{r',d'} = \bigoplus_{j=1}^h F^j_k$ that contradict stability are the h subsheaves F^j_k of degree d' and rank r' and their direct sums. Our choice of the inclusions of the subbundles in the bundle on C_1 and the gluings at the nodes guarantee that we have a stable overall bundle.

Note also that our choice of d_i means that $r_i d_{i+1} - r_{i+1}(d_i + 1) \le -1$ or equivalently $r_i d_{i+1} - r_{i+1} d_i \le r_{i+1} - 1$. In the interval $1 \le r_i \le r_{i+1} - 1$, this implies that $r_i d_{i+1} - r_{i+1} d_i \le r_i(r_{i+1} - r_i)$. Therefore, given a subspace of dimension r_i of the fiber of $E^1_{r_{i+1},d_{i+1}}$ at Q_1, there is at most a finite number of subbundles $E^1_{r_i,d_i}$ whose immersion in $E^1_{r_{i+1},d_{i+1}}$ glue with that fixed subspace (see Proposition 2.8 of [Teixidor 2000]). Therefore, the number of chains for a fixed $E_{r,d}$ on the reducible curve is finite. □

2. Extending the result to the nonsingular curve

We start by using the results on the reducible curve to extend it to a generic, nonsingular curve.

Proposition 2.1. *Let C be a generic curve of genus g. Fix a rank r and degree d, $0 \le d < r$, and a collection of integers $r_1 < r_2 < \cdots < r_k = r$. Choose degrees $d_1, \dots, d_k = d$ with d_{k-1} the largest degree such that $\frac{d_{k-1}}{r_{k-1}} < \frac{d_k}{r_k}$, d_{k-2} is the largest degree such that $\frac{d_{k-2}}{r_{k-2}} < \frac{d_{k-1}}{r_{k-1}}, \dots, d_1$ the largest degree such that $\frac{d_1}{r_1} < \frac{d_2}{r_2}$. Then, there exists a stable bundle E that contain a chain of subbundles*

$$E_{r_1,d_1} \subseteq E_{r_2,d_2} \subseteq \cdots \subseteq E_{r_k,d_k} = E$$

with E_{r_i,d_i} stable of degree d_i and rank r_i.

Proof. Take a family of curves where the special fiber is a chain of elliptic curves and the generic curve is nonsingular. Then, the result follows from Proposition 1.1 using the openness of the stability condition. □

We proved stability of the various steps of a chain of extensions under the harder conditions on slopes. This implies the similar result when the slopes are not as close:

Proposition 2.2. *Let C be a generic curve of genus g. Fix a rank r and degree d, $0 \le d < r$, and two collections of integers $r_1 < r_2 < \cdots < r_k = r$, $d_1, \ldots, d_k = d$ such that*

$$\frac{d_1}{r_1} < \frac{d_2}{r_2} < \cdots < \frac{d_k}{r_k}.$$

Then, there exists a stable bundle E that contain a chain of subbundles

$$E_{r_1,d_1} \subseteq E_{r_2,d_2} \subseteq \cdots \subseteq E_{r_k,d_k} = E$$

with E_{r_i,d_i} stable of degree d_i and rank r_i.

Proof. Fix integers r, d, r_1, d_1 with $\frac{d_1}{r_1} < \frac{d}{r}$. The set of vector bundles of rank r and degree d which contain a subbundle of rank r_1 and degree $d_1 - 1$ is contained in the closure of those vector bundles that contain a subbundle of rank r_1 and degree d_1 (see [Russo and Teixidor 1999] Corollary 1.12) Therefore, the result follows from Proposition 2.1. □

Let us now look at dimension and irreducibility:

Proposition 2.3. *Fix integers d_1, d_2, r_1, r_2 such that $\frac{d_1}{r_1} < \frac{d_2}{r_2}$. Let $\mathcal{U}_1$ be an irreducible family of stable vector bundles of rank r_1 and degree d_1. Let $\overline{\mathcal{U}_2}$ be an irreducible family of stable vector bundles of rank $r_2 - r_1$ and degree $d_2 - d_1$. Then, the family of extensions*

$$0 \to E_1 \to E \to \overline{E_2} \to 0,\ \ E_1 \in \mathcal{U}_1,\ \overline{E_2} \in \overline{\mathcal{U}_2}$$

is also irreducible of dimension

$$\dim(\mathcal{U}_1) + \dim(\overline{\mathcal{U}_2}) + r_1(r_2 - r_1)(g-1) + r_1 d_2 - r_2 d_1 - 1.$$

Proof. For fixed E_1, $\overline{E_2}$, the space of extensions as above is parameterized by $H^1(\overline{E_2}^* \otimes E_1)$. We claim that $H^0(\overline{E_2}^* \otimes E_1) = 0$. If this were not the case then there would be a nonzero morphism of $\overline{E_2} \to E_1$. Its image I would be both a quotient of $\overline{E_2}$ and a subsheaf of E_1. The stability of the two bundles implies that

$$\frac{d_1}{r_1} = \mu(E_1) < \mu(I) < \mu(\overline{E_2}) = \frac{d_2 - d_1}{r_2 - r_1}.$$

This contradicts the assumption of our initial choice of ranks and degrees. It follows that $H^0(\overline{E_2}^* \otimes E_1) = 0$ and therefore $H^1(\overline{E_2}^* \otimes E_1)$ has dimension equal to $r_1 d_2 - r_2 d_1 + r_1(r_2 - r_1)(g-1)$, irrespectively of the choice of E_1, $\overline{E_2}$. Then the statement about the dimension follows. □

Proof of Theorem 0.1. Denote by $\mathcal{U}_1$ the space of all vector bundles of degree d_1 and rank r_1, $\overline{\mathcal{U}_2}$ the space of all vector bundles of degree $d_2 - d_1$ and rank $r_2 - r_1, \dots, \overline{\mathcal{U}_k}$ the space of all vector bundles of degree $d_k - d_{k-1}$ and rank $r_k - r_{k-1}$. From Proposition 2.1, the set of bundles E that can be obtained by successive extensions is nonempty. Proposition 2.3 allows us to compute successively the dimensions of the space of extensions, starting with

$$\begin{aligned}\dim(\mathcal{U}_1) &= r_1^2(g-1)+1,\\ \dim(\overline{\mathcal{U}_2}) &= (r_2-r_1)^2(g-1)+1,\\ &\dots,\\ \dim(\overline{\mathcal{U}_k}) &= (r_k-r_{k-1})^2(g-1)+1.\end{aligned}$$

The last claim in Proposition 1.1 ensures that each vector bundle appears only a finite number of times as an extension of the given form. □

Acknowledgements

The author would like to thank the anonymous referees for helping improve the readability of the paper.

References

[Aprodu et al. 2020] M. Aprodu, G. Casnati, L. Costa, R. M. Miró-Roig, and M. Teixidor I Bigas, "Theta divisors and Ulrich bundles on geometrically ruled surfaces", *Ann. Mat. Pura Appl.* (4) **199**:1 (2020), 199–216. MR Zbl

[Atiyah 1957] M. F. Atiyah, "Vector bundles over an elliptic curve", *Proc. London Math. Soc.* (3) **7** (1957), 414–452. MR Zbl

[Casalaina-Martin and Teixidor 2011] S. Casalaina-Martin and M. Teixidor i Bigas, "Singularities of Brill–Noether loci for vector bundles on a curve", *Math. Nachr.* **284**:14-15 (2011), 1846–1871. MR Zbl

[Hitching et al. 2021] G. H. Hitching, M. Hoff, and P. E. Newstead, "Nonemptiness and smoothness of twisted Brill–Noether loci", *Ann. Mat. Pura Appl.* (4) **200**:2 (2021), 685–709. MR Zbl

[Kopper and Mandal 2023] J. Kopper and S. Mandal, "Non-globally generated bundles on curves", *Int. Math. Res. Not.* **2023**:2 (2023), 1094–1118. MR Zbl

[Lange 1983] H. Lange, "Zur Klassifikation von Regelmannigfaltigkeiten", *Math. Ann.* **262**:4 (1983), 447–459. MR Zbl

[Lange and Narasimhan 1983] H. Lange and M. S. Narasimhan, "Maximal subbundles of rank two vector bundles on curves", *Math. Ann.* **266**:1 (1983), 55–72. MR Zbl

[Mustopa and Teixidor 2024] Y. Mustopa and M. Teixidor i Bigas, "Rational curves on moduli spaces of vector bundles", *Internat. J. Math.* (online publication October 2024).

[Russo and Teixidor 1999] B. Russo and M. Teixidor i Bigas, "On a conjecture of Lange", *J. Algebraic Geom.* **8**:3 (1999), 483–496. MR Zbl

[Seshadri 1982] C. S. Seshadri, *Fibrés vectoriels sur les courbes algébriques*, Astérisque **96**, Société Mathématique de France, Paris, 1982. MR Zbl

[Teixidor 1991] M. Teixidor i Bigas, "Moduli spaces of (semi)stable vector bundles on tree-like curves", *Math. Ann.* **290**:2 (1991), 341–348. MR Zbl

[Teixidor 1995] M. Teixidor i Bigas, "Moduli spaces of vector bundles on reducible curves", *Amer. J. Math.* **117**:1 (1995), 125–139. MR Zbl

[Teixidor 1998] M. Teixidor i Bigas, "Stable extensions by line bundles", *J. Reine Angew. Math.* **502** (1998), 163–173. MR Zbl

[Teixidor 2000] M. Teixidor i Bigas, "On Lange's conjecture", *J. Reine Angew. Math.* **528** (2000), 81–99. MR Zbl

Received January 8, 2023. Revised May 21, 2024.

MONTSERRAT TEIXIDOR I BIGAS
DEPARTMENT OF MATHEMATICS
TUFTS UNIVERSITY
MEDFORD, MA
UNITED STATES
mteixido@tufts.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 332, No. 2, 2024
https://doi.org/10.2140/pjm.2024.332.395

CORRECTION TO THE ARTICLE LOCAL MAASS FORMS AND EICHLER–SELBERG RELATIONS FOR NEGATIVE-WEIGHT VECTOR-VALUED MOCK MODULAR FORMS

JOSHUA MALES AND ANDREAS MONO

Volume **322**:2 (2023), 381–406

We correct two errors in our article titled "Local Maaß forms and Eichler–Selberg relations for negative-weight vector-valued mock modular forms".

1. Modifications to the published version

(1) Throughout the paper, we add the assumption that our homogeneous polynomial p inside the Siegel theta function is equal to 1. Otherwise, the Siegel theta function might not split into a positive definite and a negative definite part in general. In particular, one has to add additional assumptions on the isometry ψ as well as on the polynomial p to obtain such a splitting; see [5, Lemma 2.2] and the discussion preceding it. Finding a preimage of Θ_P under the shadow operator ξ might not be guaranteed for nonconstant polynomials p, and our Proposition 4.1 is wrong if $p \neq 1$ since the Laplacian depends on the given polynomial; see [4, Proposition 2.5] for the correct version.

(2) In Theorem 1.2, we need to specialize the signature of the lattice L to $(2, 1)$ instead of $(2, s)$. This is necessary, because the nature of the singularities of the lift is different in higher dimensions; see [1]. In particular, the first condition in our definition of a local Maaß form on page 389 simplifies to the usual scalar-valued modularity condition; see Bringmann, Kane and Viazovska [2, Subsection 2.4] as well. In general, the Siegel theta function is invariant under the discriminant kernel of $\mathrm{O}(L)$ as a function of $Z \in \mathrm{Gr}(L)$; see [3, p. 40]. In the case of signature $(2, 1)$,

The research conducted by the first author for this paper was supported by the Pacific Institute for the Mathematical Sciences (PIMS). The research and findings may not reflect those of the Institute. The second author was supported by the CRC/TRR 191 "Symplectic Structures in Geometry, Algebra and Dynamics", funded by the DFG (project number 281071066).

MSC2020: primary 11F27; secondary 11F37.

Keywords: higher Siegel theta lift, Eichler–Selberg-type relations, local Maaß forms, vector-valued mock modular forms.

we have $\mathrm{Gr}(L) \cong \mathbb{H}$, and choosing a particular lattice of that signature leads to further identifications, which in turn yield the framework of [2]. This is described in Section 5 of our paper.

Acknowledgements

The authors would like to thank Paul Kiefer for pointing out the aforementioned errors and for many helpful conversations. We thank the referee for a helpful comment.

References

[1] C. Alfes, B. Depouilly, P. Kiefer, and M. Schwagenscheidt, "Cycle integrals of meromorphic Hilbert modular forms", preprint, 2024. arXiv 2406.03465

[2] K. Bringmann, B. Kane, and M. Viazovska, "Theta lifts and local Maass forms", *Math. Res. Lett.* **20**:2 (2013), 213–234. MR Zbl

[3] J. H. Bruinier, *Borcherds products on O(2, l) and Chern classes of Heegner divisors*, Lecture Notes in Mathematics **1780**, Springer, 2002. MR Zbl

[4] S. Zemel, "A Gross–Kohnen–Zagier type theorem for higher-codimensional Heegner cycles", *Res. Number Theory* **1** (2015), art. id. 23. MR Zbl

[5] S. Zemel, "Seesaw identities and theta contractions with generalized theta functions, and restrictions of theta lifts", *Ramanujan J.* **63**:3 (2024), 749–771. MR Zbl

Received November 4, 2024. Revised November 13, 2024.

JOSHUA MALES
SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
BRISTOL
UNITED KINGDOM

and

HEILBRONN INSTITUTE FOR MATHEMATICAL RESEARCH
BRISTOL
UNITED KINGDOM

joshua.males@bristol.ac.uk

ANDREAS MONO
DEPARTMENT OF MATHEMATICS
VANDERBILT UNIVERSITY
NASHVILLE, TN
UNITED STATES

andreas.mono@vanderbilt.edu

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

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.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 332 No. 2 October 2024

Homotopy versus isotopy: 2-spheres in 5-manifolds 195
DANICA KOSANOVIĆ, ROB SCHNEIDERMAN and PETER TEICHNER

A new convergence theorem for mean curvature flow of hypersurfaces in quaternionic projective spaces 219
SHIYANG LI, HONGWEI XU and ENTAO ZHAO

Hecke eigenvalues and Fourier–Jacobi coefficients of Siegel cusp forms of degree 2 243
MURUGESAN MANICKAM, KARAM DEO SHANKHADHAR and VASUDEVAN SRIVATSA

Continuous Sobolev functions with singularity on arbitrary real-analytic sets 261
YIFEI PAN and YUAN ZHANG

Grading of affinized Weyl semigroups of Kac–Moody type 273
PAUL PHILIPPE

CM points on Shimura curves via QM-equivariant isogeny volcanoes 321
FREDERICK SAIA

Stratification of the moduli space of vector bundles 385
MONTSERRAT TEIXIDOR I BIGAS

Correction to the article Local Maaß forms and Eichler–Selberg relations for negative-weight vector-valued mock modular forms 395
JOSHUA MALES and ANDREAS MONO