Invent math (2012) 188:197–252
DOI 10.1007/s00222-011-0348-1

On arithmetic fundamental lemmas

Wei Zhang

To my friend Lin Chen (1981–2009)

Received: 17 November 2010 / Accepted: 18 July 2011 / Published online: 25 August 2011

Abstract We present a relative trace formula approach to the Gross–Zagier formula and its generalization to higher-dimensional unitary Shimura varieties. As a crucial ingredient, we formulate a conjectural arithmetic fundamental lemma for unitary Rapoport–Zink spaces. We prove the conjecture when the Rapoport–Zink space is associated to a unitary group in two or three variables.

Contents

W. Zhang (✉)
Department of Mathematics, Harvard University, One Oxford Street, Cambridge, MA 02138, USA
e-mail: wzhang@math.harvard.edu

1 Introduction

In 1980s, Gross and Zagier [10] established a formula that relates the Neron–Tate height of Heegner points on modular curves to the central derivative of certain L-functions associated to modular forms. Ever since the publication of [10], it has inspired a large amount of works in number theory (for example, [18, 25, 26, 30, 31], etc.). A conjectural generalization of the Gross–Zagier formula to higher-dimensional Shimura varieties has been proposed, for instance, in recent preprints of Gan–Gross–Prasad [3] and Zhang [32]. In this article, we present a relative trace formula approach to the arithmetic Gan–Gross–Prasad conjecture for unitary groups (cf. Sect. 3.2). As a first step we consider the places where the relevant Shimura variety has good reduction. In particular, we formulate arithmetic fundamental lemmas for unitary groups and verify the cases of unitary groups in two or three variables. Briefly, the arithmetic fundamental lemma is an identity between the derivative of certain orbital integrals and the arithmetic intersection numbers on unitary Rapoport–Zink space. As shown in Theorem 3.11, the arithmetic fundamental lemmas is a crucial ingredient to establish a Gross–Zagier type formula for high dimensional unitary Shimura varieties.

We explain some of the history. At almost the same time of the Gross–Zagier paper [10], Waldspurger proved a parallel formula that relates certain toric periods (instead of "heights") to the central value (instead of "derivative") of the same type of L-functions. In early 1990s, Gross–Prasad formulated a conjectural generalization of Waldspurger's formula to (special) orthogonal groups [8, 9] which are usually called Gross–Prasad conjectures (local and global) in the literature. Recently, Gan–Gross–Prasad [3] formulated a similar conjecture for more classical groups including the unitary groups. Besides recent work of Waldspurger on the local conjecture for orthogonal groups, a lot of work has been done on the global Gross–Prasad conjecture for orthogonal and unitary groups. Among them we have a by-no-means complete list: Waldspurger's work on $SO(2) \times SO(3)$ [24], Ichino's work on $SO(3) \times SO(4)$ ([12] etc.), and Ginzburg–Jiang–Rallis's work on higher rank groups ([4, 5] etc.).

Note that the global Gross–Prasad and Gan–Gross–Prasad conjecture [3] asserts an identity relating certain period integrals for the groups $U(n-1)\times U(n)$ (or the orthogonal counterparts) to the central values of some Rankin–Selberg L-functions. Their arithmetic versions are identities relating heights of certain algebraic cycles on Shimura varieties to the central derivative of some Rankin–Selberg L-functions. Ever since Gross–Zagier's paper [10], a series of work of S. Zhang and recent joint work of Yuan–Zhang–Zhang [25, 26] have mostly established the arithmetic version for $SO(2)\times SO(3)$ and $SO(3)\times SO(4)$. However, the methods employed in these works do not generalize to higher dimensional cases.

Recently Jacquet and Rallis initiated a relative trace formula (for short, RTF) approach to the global Gross–Prasad conjecture for unitary groups $U(n-1)\times U(n)$ [13]. This method is completely different from that of Waldspurger [24] or Ginzburg–Jiang–Rallis [4, 5]. Briefly speaking, the relative trace formula approach of Jacquet–Rallis consists of a comparison of two relative trace formulas, one on the general linear groups and the other on the unitary groups. The spectral side of the former deals with central values of Rankin–Selberg L-functions via its integral representation, while the spectral side of the latter deals with the period integral on unitary groups. Then the global Gross–Prasad conjecture should follow from the comparison of the geometric sides of the two relative trace formulas. This comparison leads Jacquet and Rallis to their conjectural existence of *smooth transfer* and *fundamental lemma* (see Sect. 3.1 and Sect. 2.2 resp.). A proof of the existence of smooth transfer for non-archimedean p-adic fields has been announced recently [29]. But we will not use it here in an essential way. We note that the fundamental lemma of Jacquet–Rallis takes the form

$$O(\gamma,1_{K_S},s)|_{s=0}=\pm O(\delta,1_{K'}) \tag{1.1}$$

where the LHS is an orbital integral on some symmetric space S and the RHS is an orbital integral on some unitary group. For more details see Conjecture 2.4. In the case of positive characteristic, this has been proved recently by Yun [27]. By Gordon [6], Yun's result can be transferred to a p-adic field when p is large enough.

Inspired by Jacquet–Rallis's approach as well as the work [25, 26], we will formulate a relative trace formula approach to the arithmetic Gan–Gross–Prasad conjecture for unitary groups $U(n-1)\times U(n)$. On one hand, we take the first derivative of the relative trace formula (RTF, for short) on the general linear groups. On the other hand we also formulate an "arithmetic relative trace formula" on unitary groups which deals with the Beilinson–Bloch heights of certain algebraic cycles. The comparison of the two RTFs leads to what we will call "*arithmetic fundamental lemma*". Roughly speaking the arithmetic fundamental lemma is an identity (cf. (1.1))

$$\frac{d}{ds}O(\gamma, 1_{K_S}, s)|_{s=0} = \pm O'(\delta, 1_{K'}) \tag{1.2}$$

where the LHS is the derivative of Jacquet–Rallis orbital integral and the RHS is an intersection number on unitary Rapoport–Zink space. More precisely, see Conjecture 2.9. We will prove the arithmetic fundamental lemma for $U(n-1)\times U(n)$ with $n=2,3$ (see Theorems 2.10, 5.5).

This paper is organized as follows. We reverse the order of presentation to separate the local conjecture from the global one. In Sect. 2 we formulate the local conjecture of the arithmetic fundamental lemma for unitary Rapoport–Zink spaces and as an example we verify the case $n=2$. In Sect. 3 we recall the global construction of relative trace formula and height pairing. We consider the places where the Shimura variety has good reduction. The key results are in Sect. 3.3. The rest of the paper is devoted to proving the arithmetic fundamental lemma for $n=3$. In Sect. 4 we calculate the derivatives of orbital integrals. In Sect. 5 we recall some results of Rapoport–Zink, Keating and Kudla–Rapoport and then prove the arithmetic fundamental lemma for $n=3$ (Theorem 5.5).

2 Local conjecture: arithmetic fundamental lemma

2.1 Transfer of orbits

Orbits In [13], Jacquet–Rallis defines the transfer of orbits for Lie algebras. Here we need to work with the group version. Though the definition of the transfer of orbits is the same as in the case of Lie algebras, the proof of the existence does require some work.

For the moment we let F be a field and E be a separable quadratic extension. We fix an embedding of $GL_{n-1}(E)$ into $GL_n(E)$ by $g\mapsto \mathrm{diag}[g,1]$ and we will freely consider an element of $GL_{n-1}(E)$ as an element of $GL_n(E)$.

We consider the action of $GL_{n-1}(E)$ on $GL_n(E)$ by conjugation. First introduced by Rallis and Schiffmann [20, Sect. 6], we say that an element $g=\left(\begin{smallmatrix}A & u\\ v & d\end{smallmatrix}\right)\in GL_n(E)$ is *regular* if the column vectors A^iu for $i=0,1,\dots,n-2$ are linearly independent and so are the row vectors vA^i for $i=0,1,\dots,n-2$. Or equivalently eg^i and $g^i({}^te)$ for $i=0,1,\dots,n-1$ are linearly independent respectively where we denote $e=(0,0,\dots,0,1)$.

Lemma 2.1 *Let g be as above. Then the following are equivalent*

(1) *g is regular.*
(2) $D(A,u,v)=:\det(a_{ij})\neq 0$ *where* $a_{ij}=vA^{i+j}u$.

(3) *The stabilizer of g is trivial and the $GL_{n-1}(\overline{E})$-orbit of g is a Zariski closed subset in $GL_n(\overline{E})$ where $\overline{E}$ is an algebraic closure of E.*

Proof (1) $\Leftrightarrow$ (2) is trivial. For the third one, we refer to [20, Theorem 6.1]. $\square$

Remark 1 By item (3) above, the notion of "regular" here corresponds to the notion of "regular semi-simple" for the usual conjugacy class in a reductive group.

For $g \in GL_n(E)$ of the form $\left(\begin{smallmatrix} A & u \\ v & d \end{smallmatrix}\right)$, we call the $2n-1$ numbers its *invariants*

$$(A_i)_{i=1}^{n-1}, \qquad (B_j)_{j=1}^{n-1}, \qquad d, \tag{2.1}$$

where

$$\det(X - A) = X^{n-1} - A_1 X^{n-1} + \cdots + (-1)^{n-1} A_{n-1}, \qquad B_j = v A^{j-1} u.$$

Then it is easy to see that in a $GL_{n-1}(E)$-orbit (even $GL_{n-1}(\overline{E})$-orbit) these invariants take constant values. And it is easy to see that the converse is also true:

Lemma 2.2 *Two regular elements are $GL_{n-1}(E)$-conjugate if and only if they have the same invariants.*

Proof See [20, Proposition 6.2]. $\square$

Now we introduce some subsets of $GL_n(E)$ and restrict the $GL_{n-1}(E)$ orbits to these subsets. First we let $H \subset GL_n(E)$ be the image of $GL_{n-1}(F)$ via the fixed embedding $GL_{n-1}(E) \hookrightarrow GL_n(E)$. Let $g \mapsto \bar{g}$ be the involution on G given by the Galois conjugate in $Gal(E/F)$ and consider the symmetric space

$$S_n(F) = \{s \in GL_n(E) | s\bar{s} = 1\}.$$

Then we consider the $H(F)$-action on $S_n(F)$ by conjugation by

$$h \circ s := hsh^{-1}, \quad h \in H(F).$$

Denote the space of orbits under the action of $H(F)$ on $S_n(F)$ by

$$\mathbb{O}(S_n) := H(F) \backslash\backslash S_n(F),$$

where the double slash is to indicate the action by conjugation.

Let $Her_n(E)$ denote the space of Hermitian matrices of size $n \times n$. For a non-degenerate $J \in Her_n(E)$ we will denote by $U(J)$ the unitary group

$$U(J) := \{g \in GL_n(E) | gJg^* = J\}, \quad g^* = {}^t\bar{g}.$$

For $J \in Her_{n-1}(E)$, naturally we will consider the $U(J)$ as a subgroup $U(J \oplus 1)$ where $J \oplus 1 = \mathrm{diag}[J, 1] \in Her_n(E)$. We denote the space of orbits by

$$\mathbb{O}(U(J \oplus 1)) := U(J)(F) \backslash\backslash U(J \oplus 1)(F),$$

where the double slash is to indicate the action of $U(J)(F)$ by conjugation. For $J_1, J_2 \in Her_{n-1}(E)$, they are called equivalent if there is an $h \in GL_{n-1}(E)$ such that $J_1 = hJ_2h^*$. If $J_1, J_2 \in Her_{n-1}(E)$ are equivalent, it is clear that there is a natural identification between the set of orbits $\mathbb{O}(U(J_1 \oplus 1))$ and $\mathbb{O}(U(J_2 \oplus 1))$ given explicitly by $g \mapsto hgh^{-1}$.

We consider both S_n and $U(J \oplus 1)$ as subsets of $GL_n(E)$ so that we may speak of regularity for their elements. Since H and $U(J)$ are then subsets of $GL_{n-1}(E)$, we may speak of regularity for H-orbits in S_n and $U(J)$-orbits in $U(J \oplus 1)$. Also we define the invariants of the orbit as we did right before Lemma 2.2.

For $\gamma \in \mathbb{O}(S_n)$ and $\delta \in \mathbb{O}(U(J \oplus 1))$, we say they match each other and denote this relation by $\gamma \leftrightarrow \delta$ if $\delta = h\gamma h^{-1}$ for some $h \in GL_{n-1}(E)$ (or equivalently, they have the same invariants). Or we say γ and δ are *transfers* of each other.

Lemma 2.3 *The transfer of orbits defines a bijection between regular orbits*

$$\coprod_J \mathbb{O}(U(J \oplus 1))_{reg} \simeq \mathbb{O}(S_n)_{reg},$$

where the sum over J runs over all equivalence classes of non-degenerate $J \in Her_{n-1}(E)$.

Proof First we prove that the transfer of any regular $s \in S_n(F)$ exists. Note that s and ${}^ts$ have the same invariants. By Lemma 2.2 we can find an element $g \in GL_{n-1}(E)$ such that

$$gsg^{-1} = {}^ts.$$

By $s\bar{s} = 1$ we have ${}^ts = {}^t\bar{s}^{-1}$ and hence

$$s^*gs = g.$$

And taking the conjugate of the original equation gives us

$$\bar{g}\bar{s}\bar{g}^{-1} = s^* \quad \Leftrightarrow \quad \bar{g}s^{-1}\bar{g}^{-1} = s^*.$$

Thus we have

$$s^*\bar{g}s = \bar{g}.$$

Taking transpose of the original equation gives us

$$s^*({}^t g)s = {}^t g.$$

By the regularity of s, its stabilizer is trivial and hence

$$g = \bar{g} = {}^t g.$$

This proves that $s \in U(g \oplus 1)$.

Second we prove that the transfer of any $\delta \in U(J \oplus 1)$ exists for any J. Let $\delta \in U(\beta)_{reg}$ for $\beta = J \oplus 1 \in Her_n(E)$. For regular δ, by Lemma 2.2, there exists $g \in GL_{n-1}(E)$ such that

$$g\delta g^{-1} = {}^t\delta.$$

By $\delta^*\beta\delta = \beta$ we have

$$\delta^*\beta g^{-1}({}^t g)g = \beta \quad \Leftrightarrow \quad g\beta^{-1}\delta^*\beta g^{-1}({}^t\delta) = 1.$$

Take conjugate to get

$$\bar{\beta}\bar{g}^{-1}\delta^*\bar{g}\bar{\beta}^{-1}({}^t\delta) = 1.$$

Comparing the last two equations, and since the stabilizer is trivial, we have

$$\bar{\beta}\bar{g}^{-1} = g\beta^{-1} \quad \Leftrightarrow \quad g\beta^{-1}\overline{g\beta^{-1}} = 1.$$

So we have $g\beta^{-1} \in S_n(F)$ and by Hilbert Satz–90 there exists $h \in GL_{n-1}(E)$ such that $h\bar{h}^{-1} = g\beta^{-1}$. Substitute back

$$h\bar{h}^{-1}\delta^*\bar{h}h^{-1}({}^t\delta) = 1 \quad \Leftrightarrow \quad (\bar{h}^{-1}\delta^*\bar{h})(h^{-1}({}^t\delta)h) = 1.$$

Setting $s = {}^t h\delta({}^t h^{-1})$, then $s \in S_n(F)$ and it is $GL_{n-1}(E)$-conjugate to δ.

Finally we need to show the uniqueness of transfer. This follows from that fact that the intersection of each regular $GL_{n-1}(E)$-orbit with $S_n(F)$ ($U(J \oplus 1)(F)$, resp.)—if not empty —gives exactly one H-orbit ($U(J)$-orbit, resp.). □

Remark 2 From the proof, we see that if F is a non-archimedean local field, then $\gamma \leftrightarrow \delta \in \mathbb{O}(U(J \oplus 1))_{reg}$ for the isometric class J uniquely determined by

$$\det((e\gamma^{i-j}e^t)) \in \det(J) \cdot NE^\times,$$

where $(e\gamma^{i-j}e^t)$ is the $n \times n$-matrix whose (i, j)-entry is $e\gamma^{i-j}e^t$.

Fundamental lemma of Jacquet–Rallis From now on we assume that F is a local field. When F is non-archimedean, we let $\mathcal{O}_F, \mathcal{O}_E$ be the ring of integers and fix a uniformizer ϖ of F and let q be the cardinality of the residue field k of F.

For $f \in \mathcal{C}_c^\infty(S_n(F))$ and a regular $\gamma \in S_n(F)$, we define the "orbital" integral

$$O(\gamma, f, s) := \int_H f(h\gamma h^{-1})|\det(h)|^{-s}\eta(h)\,dh \tag{2.2}$$

where η is the quadratic character associated to E/F by class field theory and by abuse of notation $\eta(h) := \eta(\det(h))$. The Haar measure on H is normalized such that the volume of its standard maximal compact subgroup $GL_{n-1}(\mathcal{O}_F)$ is one when F is non-archimedean. By the regularity of γ, the H-orbit of γ is a closed subset of S_n. Therefore the restriction of f to the H-orbit of γ is smooth with compact support. It follows that when F is non-archimedean, the orbital integral is actually a finite sum and hence gives a polynomial of q^s and q^{-s}. And when F is archimedean, this is always absolutely convergent.

Similarly we define the orbital integral for $f \in \mathcal{C}_c^\infty(U(J \oplus 1))$ and a regular $\delta \in U(J \oplus 1)(F)$

$$O(\delta, f) := \int_{U(J)} f(h^{-1}\delta h)\,dh. \tag{2.3}$$

Now we assume further that F is non-archimedean with odd residue characteristic p and E is an unramified quadratic extension of F.

The fundamental lemma of Jacquet–Rallis concerns the orbital integrals of some special test functions. More precisely, let $K = GL_n(\mathcal{O}_E)$ be the standard maximal compact subgroup of $GL_n(E)$ and let $K_S = K \cap S_n$ and 1_{K_S} the characteristic function. Since F is non-archimedean, there are precisely two isomorphism classes, denoted by J_0 and J_1, where the discriminant of J_0 (J_1, resp.) in $F^\times/NE^\times$ is the identity (the other element, resp.). Since E/F is unramified we may choose a self-dual lattice in the n-dimensional Hermitian space defined by $J_0 \oplus 1$ in a way that its intersection with the subspace defined by J_0 is also self-dual. We denote its stabilizer by $K' \subseteq U(J_0 \oplus 1)(F)$. Then the compact open K' and $K' \cap U(J_0)(F)$ are hyperspecial subgroups of $U(J_0 \oplus 1)(F)$ and $U(J_0)(F)$, respectively. We now normalize the measure on $U(J_0)(F)$ to give volume one to $K' \cap U(J_0)(F)$.

Conjecture 2.4 (Fundamental lemma of Jacquet–Rallis) *For $\gamma \in \mathbb{O}(S_n)_{reg}$, there is a sign $\omega(\gamma) \in \{\pm 1\}$ such that*

$$O(\gamma, 1_{K_S}, 0) = \begin{cases} \omega(\gamma) O(\delta, 1_{K'}), & \text{if } \gamma \leftrightarrow \delta \in \mathbb{O}(U(J_0 \oplus 1))_{reg}, \\ 0, & \text{if } \gamma \leftrightarrow \delta \in \mathbb{O}(U(J_1 \oplus 1))_{reg}. \end{cases} \tag{2.4}$$

The sign $\omega(\gamma)$ can be given explicitly. But we will not treat this issue in this paper.

When F is of characteristic $p > 0$ and $p > n$, this is now a theorem proved recently by Yun [27] using similar technique of Ngo's proof of endoscopy fundamental lemma. In the appendix to [27] by Gordan [6], it is proved that Yun's result implies the above conjecture when the residue characteristic of F is large enough.

Remark 3 An earlier version of this paper also proved the fundamental lemma for $n = 3$ by brutal computation.

The second half of the fundamental lemma turns out to be very easy.

Lemma 2.5 *Suppose that* $f \in \mathcal{C}_c^\infty(S_n(F))$ *satisfies* $f(g) = f({}^t g)$. *Then for a regular* $\gamma \in S_n(F)$ *whose transfer* δ *lies in* $U(J \oplus 1)(F)$, *we have*

$$O(\gamma, f, 0) = \eta(\det(J))O(\gamma, f, 0).$$

In particular, $O(\gamma, 1_{K_S}, 0) = 0$ *if* $\gamma \leftrightarrow \delta \in \mathbb{O}(U(J_1 \oplus 1))_{reg}$.

Proof By the symmetry of f we have

$$O(\gamma, f, 0) = \int f(({}^t g^{-1})^t \gamma^t g)\eta(g)\, dg = \int f(g^t \gamma g^{-1})\eta(g)\, dg.$$

Note that for a regular γ there exists a unique $J \in Sym_{n-1}(F)$ such that ${}^t\gamma = J\gamma J^{-1}$ and this J is precisely the one such that γ matches an orbit in $U(J \oplus 1)$ (see the proof of Lemma 2.3). Then we have

$$O(\gamma, f, 0) = \int f(gJ\gamma(gJ)^{-1})\eta(g)\, dg = \eta(\det(J))O(\gamma, f, 0). \qquad \square$$

Note that we have the following for any $h \in H(F)$:

$$O(h\gamma h^{-1}, 1_{K_S}, s) = |\det(h)|^s \eta(h) O(\gamma, 1_{K_S}, s).$$

When $O(\gamma, 1_{K_S}, 0) = 0$, taking the first derivative of the above equality yields:

$$O'(h\gamma h^{-1}, 1_{K_S}, 0) = \eta(h) O'(\gamma, 1_{K_S}, 0)$$

where we define

$$O'(\gamma, 1_{K_S}, 0) := \frac{d}{ds} O(\gamma, 1_{K_S}, s)|_{s=0}. \tag{2.5}$$

Then the integral $O'(\gamma, 1_{K_S}, 0)$ up to a sign depends only on the H-orbit of γ if $\gamma \leftrightarrow \delta \in \mathbb{O}(U(J_1 \oplus 1))_{reg}$.

2.2 Arithmetic fundamental lemma

The arithmetic fundamental lemma alluded to the title is a geometric interpretation of the derivative of the orbital integral $O'(\gamma, 1_{K_S}, 0)$ defined above. The counterpart of the orbital integral on the unitary group will be a certain intersection number on a Rapoport–Zink spaces.

Unitary Rapoport–Zink space We recall some basic facts about Rapoport–Zink spaces [21]. We will restrict ourselves to the situation when F is of characteristic zero, namely F is a finite extension of $\mathbb{Q}_p$ with odd p. Let ϖ be a uniformizer of F. Let $\mathbb{F} = \bar{k}$ be the algebraic closure of the residue field k of F. Let $W(\mathbb{F})$ ($W(k)$, resp.) be the Witt ring of $\mathbb{F}$ (k, resp.). Let $W_F = W(\mathbb{F}) \otimes_{W(k)} F$ be the maximal unramified extension of F and W its ring of integers. We consider F as a subfield of W_F. Let E/F be an unramified quadratic extension.Then there are two embeddings of E into W_F lying above $F \hookrightarrow W_F$: ϕ_0 and ϕ_1 equal to ϕ_0 composed with the Galois conjugate. By a *p-divisible $\mathcal{O}_F$-module* over a $\mathcal{O}_F$-scheme S, we will mean a p-divisible group X over S with an action $\iota : \mathcal{O}_F \to End_S(X)$ such that we have an exact sequence of finite flat group schemes for all $i, j \in \mathbb{Z}_{\geq 0}$:

$$0 \longrightarrow X[\varpi^i] \longrightarrow X[\varpi^{i+j}] \xrightarrow{\varpi^i} X[\varpi^j] \longrightarrow 0$$

and the induced action of $\mathcal{O}_F$ on the Lie algebra $Lie(X)$ is given by the structure morphism $\mathcal{O}_F \to \mathcal{O}_S$. Its dimension is the rank of $Lie(X)$ and its height is the unique integer h such that the rank of $X[\varpi^n]$ is q^{nh}.

Let Nilp_W be the category of W-schemes S on which p is locally nilpotent and we will consider W-schemes as $\mathcal{O}_F$-schemes via the fixed embedding. For a scheme S over W we denote by $\bar{S}$ its special fiber $S \times_{Spec\, W} Spec\, \mathbb{F}$. We will define a notion of *unitary p-divisible O_F-module of signature* (r, s) essentially following Vollaard and Wedhorn [23, Sect. 1.2].

Definition 2.6 A unitary p-divisible O_F-module of signature (r, s) over S is a triple (X, ι_X, λ_X) where

- X is a p-divisible O_F-module over S.
- $\iota_X : O_E \to End_S(X)$ is an injective homomorphism extending the $\mathcal{O}_F$-action. And the dual p-divisible O_F-module $X^\vee$ is thus endowed with an action of $\mathcal{O}_E$ by $\iota_{X^\vee}(a) = (\iota(\bar{a}))^\vee$.
- $\lambda_X : X \to X^\vee$ is an $\mathcal{O}_E$-linear p-principal polarization.
- For $a \in \mathcal{O}_E$, the induced action on the Lie algebra of X has characteristic polynomial given by

$$\mathrm{charpol}(\iota_X(a)|Lie(X))(T) = (T - \phi_0(a))^r (T - \phi_1(a))^s \in O_S[T].$$

Remark 4 Note that our terminology is slightly different from that of [23, Sect. 1.2]. When $F = \mathbb{Q}_p$, our *unitary p-divisible O_F-module of signature* (r, s) is the same as *unitary p-divisible group of signature* (r, s) in [23, Sect. 1.2].

Let $\mathbb{E}$ be the unique (up to isogeny) formal p-divisible $\mathcal{O}_F$-module of dimension one and height two over $\mathbb{F}$. We may fix one isomorphism $End(\mathbb{E}) \simeq \mathcal{O}_D$, where D is the unique division quaternion algebra over F and $\mathcal{O}_D$ its ring of integer. We can endow $\mathbb{E}$ with auxiliary structure to obtain a unitary p-divisible $\mathcal{O}_F$-module of signature $(0, 1)$. To fulfill this, we may fix one embedding $\mathcal{O}_E \hookrightarrow \mathcal{O}_D$ (composing with Galois conjugate if necessary) and endow $\mathbb{E}$ with a p-principal polarization (see [16, Remark 2.5] when $F = \mathbb{Q}_p$) compatible the $\mathcal{O}_E$-action. By abuse of notation we will denote by $\mathbb{E}$ this unitary p-divisible $\mathcal{O}_F$-module. And we will denote by $\overline{\mathbb{E}}$ the polarized unitary $\mathcal{O}_F$-divisible module of signature $(1, 0)$ by switching the embedding $\mathcal{O}_E \hookrightarrow \mathcal{O}_D$ to its conjugate.

From $\mathbb{E}$ and $\overline{\mathbb{E}}$ we can construct a unitary p-divisible $\mathcal{O}_F$-module of signature $(1, n-1)$ by setting $\mathbb{X}_n = \overline{\mathbb{E}} \times \mathbb{E}^{n-1}$ together with the auxiliary structure. Let G_n be the group of quasi-isogeny of $\mathbb{X}_n$. Under the convention at hand we may identify G_n as follows. Let $End_{\mathcal{O}_E}(\mathbb{X}_n)$ be the set of $\mathcal{O}_E$-linear endomorphism of $\mathbb{X}_n$. We can identify $End^0_{\mathcal{O}_E}(\mathbb{X}_n) = End_{\mathcal{O}_E}(\mathbb{X}_n) \otimes \mathbb{Q}$ with the subalgebra $M_{n,E}(D)$ consisting of $x \in End^0(\mathbb{X}) = M_n(D)$ which commute with every element of E, where the embedding of E is given by

$$\begin{aligned} E &\to M_n(D) \\ x &\mapsto \begin{pmatrix} \bar{x} & & & \\ & x & & \\ & & \ddots & \\ & & & x \end{pmatrix}. \end{aligned}$$

Under this identification we see that the group G_n of quasi-isogenies of height zero is identified as the set of elements $g \in M_{n,E}(D)$ such that $gg^* = 1_n$ where $*$ is the Rosati involution induced by the polarization. We may assume that the polarization will make $*$ be given as

$$g^* = {}^t\bar{g},$$

where the bar is induced by the involution (still denoted by bar) on the quaternion algebra D such that the reduced norm is given by $x\bar{x}$.

Now we have an isomorphism between G_n and unitary group $U(J_1 \oplus 1)$ defined by the Hermitian matrix $\beta = J_1 \oplus 1$, $J_1 = \mathrm{diag}[-\varpi, 1, \ldots, 1]$

$$U(J_1 \oplus 1) := \{g \in GL_n(E) | g\beta g^* = \beta\}.$$

Indeed we may give the isomorphism explicitly. Note that we have a unique decomposition

$$D = E + \mathrm{j}E, \quad \mathrm{j}^2 = -\varpi \tag{2.6}$$

for a uniformizer j of D normalizing E. First it is easy to see that $x = (a_{ij}) \in M_n(D)$ commutes with E if and only if $a_{1j}, a_{i1} \in \mathrm{j}E$ for all $i, j \in \{2, 3, \dots, n\}$ and all the other $a_{ij} \in E$. Denoting $\mathcal{J} = \mathrm{diag}[\mathrm{j}, 1 \dots, 1] \in GL_{n-1}(D)$, then the isomorphism is explicitly given by

$$\phi : U(J_1 \oplus 1) \simeq G_n = \{g \in M_{n,E}(D) | gg^* = 1_n\}$$
$$h \mapsto \begin{pmatrix} \mathcal{J}^{-1} & \\ & 1 \end{pmatrix} h \begin{pmatrix} \mathcal{J} & \\ & 1 \end{pmatrix}.$$

From now on we will identify G_n with $U(J_1 \oplus 1)$.

Now we denote by the $\mathcal{N}_n$ the *unitary Rapoport–Zink space* associated to $\mathbb{X}_n$. Namely, $\mathcal{N}_n$ represents the moduli functor that associates to a W-scheme S the set of quadruples $(X, \iota_X, \lambda_X, \rho_X)$, where (X, ι_X, λ_X) is a unitary p-divisible O_F-module of signature $(1, n-1)$ over S and ρ_X is a quasi-isogeny of height zero

$$\rho_X : X \times_S \bar{S} \to \mathbb{X}_n \times_{\mathbb{F}} \bar{S}$$

that respects the auxiliary structure, namely, ρ_X is $\mathcal{O}_E$-linear such that $\rho_X^\vee \circ \lambda_{\mathbb{X}} \circ \rho_X$ is locally an $\mathcal{O}_F^\times$-multiple of $\lambda_X \in Hom_{\mathcal{O}_E}(X, X^\vee) \otimes F$. Two such objects $(X, \iota_X, \lambda_X, \rho_X)$ and $(X', \iota_{X'}, \lambda_{X'}, \rho_{X'})$ are isomorphic if there exists an $\mathcal{O}_E$-linear isomorphism $\alpha : X \to X'$ with $(\alpha \times_W \mathbb{F}) \circ \rho_X = \rho_{X'}$ and such that $\alpha^\vee \circ \lambda_{X'} \circ \alpha$ is locally an $\mathcal{O}_F^\times$-multiple of λ_X.

By the results of Rapoport–Zink [21, Corollary 3.40] and Vollaard–Wedhorn [23], $\mathcal{N}_n$ is a formally smooth formal scheme of relative dimension $n-1$ over $Spf(W)$, separated and locally formally of finite type over W.

Arithmetic fundamental lemma Let $\mathcal{E}$ be the canonical lifting of $\mathbb{E}$, namely the universal object over $\mathcal{N}_1$. Then one has a natural embedding $\mathcal{N}_{n-1} \to \mathcal{N}_n$ given by associating to a unitary p-divisible $\mathcal{O}_F$-module of signature $(1, n-2)$ with height 0 quasi-isogeny ρ_0 over S its product with $\mathcal{E} \times_{Spf\, W} S$ (together with an appropriate choice of the auxiliary structure). Let $\mathcal{N} = \mathcal{N}_{n-1} \times_W \mathcal{N}_n$ and let $\Delta_{\mathcal{N}_{n-1}}$ be the diagonal embedding of $\mathcal{N}_{n-1}$. Then the group of quasi-isogenies $G_{n-1} \times G_n$ acts on $\mathcal{N}$ as automorphisms. For $(g_1, g_2) \in G_{n-1} \times G_n$ we will denote by $(g_1, g_2)^* \Delta_{\mathcal{N}_{n-1}}$ the translation of the sub-formal-scheme $\Delta_{\mathcal{N}_{n-1}}$ under (g_1, g_2).

Definition 2.7 For $g = (g_1, g_2) \in G_{n-1} \times G_n$, we define an intersection number

$$O'(g, 1_{K'}) := (\Delta_{\mathcal{N}_{n-1}} \cdot g^* \Delta_{\mathcal{N}_{n-1}})_{\mathcal{N}} \log q \tag{2.7}$$

to be the Euler–Poincaré characteristic $\chi(\mathcal{O}_{\Delta_{\mathcal{N}_{n-1}}} \otimes^{\mathbb{L}} \mathcal{O}_{g^*\Delta_{\mathcal{N}_{n-1}}})$.

Here $\otimes^{\mathbb{L}}$ is the derived tensor product of $\mathcal{O}_{\mathcal{N}}$-modules and for a sheaf of $\mathcal{O}_{\mathcal{N}}$-module $\mathcal{F}$, we define

$$\chi(\mathcal{F}) = \sum_i (-1)^i \,\mathrm{length}_W(R^i\pi_*\mathcal{F})$$

where $\pi : \mathcal{N} \to Spf\ W$ is the structure morphism. For a bounded complex of sheaves $\mathcal{F}^\bullet$ of $\mathcal{O}_{\mathcal{N}}$-modules, we define (cf. [16])

$$\chi(\mathcal{F}^\bullet) = \sum_i (-1)^i \chi(\mathcal{F}^i).$$

We consider $G_{n-1} \simeq U(J_1)$ as a subgroup of $G_n \simeq U(J_1 \oplus 1)$ and then diagonally embed it into $G_{n-1} \times G_n$. Then $\Delta_{\mathcal{N}_{n-1}}$ is invariant under the action of G_{n-1}. Therefore we see that the intersection number $O'(g, 1_{K'})$ depends only on the G_{n-1}-double coset of g. In particular we may assume that $g = (1, \delta)$ for $\delta \in G_n$ and we say g is regular if δ is regular. In this case we will still denote the intersection number by $O'(\delta, 1_{K'})$ by abuse of notation.

Lemma 2.8 *Assume that $F = \mathbb{Q}_p$. If g is regular, then $O'(g, 1_{K'})$ is finite.*

Proof Assume that $g = (1, \delta)$ for $\delta \in G_n$. Let $\mathcal{F}^\bullet = \mathcal{O}_{\Delta_{\mathcal{N}_{n-1}}} \otimes^{\mathbb{L}} \mathcal{O}_{(1,\delta)^*\Delta_{\mathcal{N}_{n-1}}}$. We will use the special cycles of Kudla–Rapoport and for the notation we refer to [16] and [17] (we will also recall some definitions below in Sect. 5.2). We may identify $\mathcal{N}_{n-1}$ with the special divisor $\mathcal{Z}(u)$ on $\mathcal{N}_n$ for a vector u of norm one in $Hom_{\mathcal{O}_E}(\overline{\mathbb{E}}, \mathbb{X}_n)$. Denote by $\mathcal{Y}$ the intersection of $\Delta_{\mathcal{N}_{n-1}}$ and $(1,\delta)^*\Delta_{\mathcal{N}_{n-1}}$. By the projection from $\mathcal{N}$ to the factor $\mathcal{N}_{n-1}$, we may consider $\mathcal{Y}$ as a sub-formal-scheme of $\Delta_{\mathcal{N}_{n-1}}$. Then $\mathcal{Y}$ is contained in the intersection $\mathcal{Z}(u, \delta u, \ldots, \delta^{n-1}u)$ of the special divisors $\mathcal{Z}(\delta^i u)$, $i = 0, 1, \ldots, n-1$. By the regularity of δ, the fundamental matrix T of the n-tuple $\underline{x} := (\delta^i u)$ is non-singular. Note that the special cycle $\mathcal{Z}(u, \delta u, \ldots, \delta^{n-1}u)$ on $\mathcal{N}_n$ depends only on the $\mathcal{O}_E$-span of $u, \delta u, \ldots, \delta^{n-1}u$. In particular, it depends only on the Jordan decomposition (cf. [16]) of the fudamental matrix T. Hence we may choose special homomorphisms $y_1, y_2, \ldots, y_n$ such that we have an identity $\mathcal{Z}(y_1, \ldots, y_n) = \mathcal{Z}(u, \delta u, \ldots, \delta^{n-1}u)$ as special cycles on $\mathcal{N}_n$ and such that all entries of the fundamental matrix T' of $y_1, \ldots, y_n$ are in $\mathbb{Q}$ (indeed T' can be made to be a diagonal matrix). Then there exists a global special cycle $\mathcal{Z}(T')$ in the sense of [17] and with $\mathcal{Z}(y_1, y_2, \ldots, y_n) \subset \mathcal{Z}(T')$. The formal scheme $\mathcal{Z}(T')$ is the completion along the supersingular locus of the intersection of some global divisors, on which p is nilpotent, and which is a closed subset of the supersingular locus, cf. [17], Sect. 2. Hence $\mathcal{Z}(T')$ is a scheme, and hence so is its closed formal subscheme $\mathcal{Z}(y_1, y_2, \dots, y_n)$. This implies that $\mathcal{Y}$ is also a scheme on which p is nilpotent. $\square$

We are ready to state the following:

Conjecture 2.9 (Arithmetic Fundamental Lemma) *If $\gamma \in S_n(F)_{reg}$ matches $\delta \in \mathbb{O}(U(J_1 \oplus 1))_{reg}$, then $O'(g, 1_{K'})$ is finite and there is a sign $\omega'(\gamma) \in \{\pm 1\}$ such that*

$$O'(\gamma, 1_{K_S}, 0) = \omega'(\gamma) O'(\delta, 1_{K'}). \tag{2.8}$$

As shown in next section, this intersection number serves as a part of the local height pairing of some algebraic cycle on some unitary Shimura varieties via the well-known p-adic uniformization by Rapoport–Zink spaces (cf. Theorem 3.9).

Remark 5 It should also be pointed out that our arithmetic fundamental lemma in its form resembles the local unramified (conjectural) theory of the arithmetic Siegel–Weil formula that was first proposed by Kudla in [15] (cf. [16, 17]).

Remark 6 One may also formulate a Lie algebra version of the arithmetic fundamental lemma. We omit this in this paper since the Lie algebra version does not seem simpler than the group version. The author has verified the Lie algebra version for $n \le 3$ using the same technique as this paper.

Note that $\Delta_{\mathcal{N}_{n-1}}$ and $g^*\Delta_{\mathcal{N}_{n-1}}$ intersect properly (i.e., the scheme theoretical intersection is an artinian scheme) for all regular g only when $n \le 3$. This is the reason that in this article we restrict ourselves to this case and we will prove the arithmetic fundamental lemma when $n \le 3$ (cf. Theorem 2.10 and Theorem 5.5). The proof is by explicit computation of both sides of the target identities. When $n = 2$, it is essentially reduced to Gross's theory of canonical lifting [7]. When $n = 3$, we use a result of Kudla–Rapoport [16] on the structure of special cycles on Rapoport–Zink space for $U(2)$, together the work of Keating on endomorphisms of reductions of quasi-canonical lifting (cf. Sect. 5).

2.3 Example: $n = 2$

As an example, we verify Conjecture 2.9 for $n = 2$. The non-archimedean local height computation in the original Gross–Zagier formula [10] requires essentially only this.

We will denote by E^1 the set of norm one elements in E. We first fix representatives of regular orbits. For the symmetric space S_2, we can choose

$$\gamma(a,d) := \begin{pmatrix} a & \\ & 1 \end{pmatrix} \begin{pmatrix} \bar{d} & t_a \\ -(1-d\bar{d})/t_a & d \end{pmatrix},$$
$$a \in E^1,\ t_a \in \mathcal{O}_E^\times,\ a = t_a/\bar{t}_a,\ d \in E^\times. \tag{2.9}$$

Such a $\gamma(a,d)$ matches an element in the group G_2 of quasi-isogenies of $\mathbb{X}_2 = \overline{\mathbb{E}} \otimes \mathbb{E}$ if and only if $v(1-d\bar{d})$ is odd (in particular, $d \in \mathcal{O}_E$). A matching representative in G_2 can be chosen as

$$\delta(a,d) = \begin{pmatrix} a & \\ & 1 \end{pmatrix} \begin{pmatrix} \bar{d} & b \\ -\bar{b} & d \end{pmatrix} \in M_2(D), \tag{2.10}$$

where b is any element in $\mathrm{j}E$ such that

$$d\bar{d} + b\bar{b} = 1.$$

We define a sign $\omega'(\gamma) \in \{\pm 1\}$ for regular $\gamma \in S_2(F)$ as follows: if $\gamma = h \cdot \gamma(a,d) \cdot h^{-1}$ (such h is unique by the regularity of γ), then

$$\omega'(\gamma) := \eta(\det(h)).$$

Theorem 2.10 *The arithmetic fundamental lemma holds when $n = 2$:*

$$O'(\gamma, 1_{K_S}, 0) = \omega'(\gamma) O'(\delta, 1_{K'}). \tag{2.11}$$

Proof It is easy to see that the derivative of orbital integral is given by

$$O'(\gamma(a,d), 1_{K_S}, 0) = \frac{v(1-d\bar{d})+1}{2}.$$

On the other hand, the intersection multiplicity is the maximal integer m such that $\delta \in M_2(\mathcal{O}_D)$ can be lifted to an endomorphism of the reduction of $\overline{\mathcal{E}} \oplus \mathcal{E}$ to $W/\varpi^m W$. Since $d \in \mathcal{O}_E$ already extends to an endomorphism of $\mathcal{E}$ on W, this is the same as to extend $b \in \mathrm{j}E \subset End(\mathbb{E})$ to a homomorphism from $\mathcal{E}$ to $\overline{\mathcal{E}}$ mod ϖ^m. By Gross' theory of canonical lifting, we immediately obtain

$$m = \frac{v_D(b)+1}{2},$$

where v_D is the valuation in D. This is the same as $\frac{v(b\bar{b})+1}{2} = \frac{v(1-d\bar{d})+1}{2}$. $\square$

3 Global motivation

3.1 Relative trace formula and its derivative

A relative trace formula of Jacquet–Rallis Now let F be a number field and let E be a quadratic extension. We consider the F-algebraic group

$$G' = Res_{E/F}\,(GL_{n-1} \times GL_n)$$

and two subgroups: H_1' is the diagonal embedding of $Res_{E/F}GL_{n-1}$ (where GL_{n-1} is embedded into GL_n in the same way as in Sect. 2.1) and $H_2' = GL_{n-1,F} \times GL_{n,F}$ embedded into G' in the obvious way. For an F-algebraic group H, we will denote by Z_H its maximal F-split torus in the center of H. In particular we see that $Z_{G'} = Z_{H_2'}$.

Here is the construction of our relative trace formula. For $f' \in C_c^\infty(G'(\mathbb{A}))$, fixing a Haar measure on $Z_{G'}(\mathbb{A})$ and the counting measure on $Z_{G'}(F)$ we define a kernel function

$$K_{f'}(x,y) = \int_{Z_{G'}(F)\backslash Z_{G'}(\mathbb{A})} \sum_{\gamma\in G'(F)} f'(x^{-1}z\gamma y)\,dz,$$

or equivalently,

$$K_{f'}(x,y) = \sum_{Z_{G'}(F)\backslash G'(F)} \widetilde{f}'(x^{-1}\gamma y)$$

where

$$\widetilde{f}'(h) := \int_{Z_{G'}(\mathbb{A})} f'(zh)\,dz.$$

It is easy to see that the kernel function is a continuous function on $G'(\mathbb{A}) \times G'(\mathbb{A})$.

Let $\eta = \eta_{E/F} : F^\times\backslash\mathbb{A}^\times \to \{\pm 1\}$ be the quadratic character associated to E/F by class field theory. By abuse of notation we will also denote by η the character of $H_2'(\mathbb{A})$ defined by $\eta(h) := \eta(\det(h_1))$ ($\eta(\det(h_2))$, resp.) if $h = (h_1, h_2) \in GL_{n-1}(\mathbb{A}) \times GL_n(\mathbb{A})$ and n odd (even, resp.).

Fix a Haar measure on $H_i'(\mathbb{A})$ ($i = 1, 2$) and $Z_{H_2'}(\mathbb{A})$. We then consider a distribution on $G'(\mathbb{A})$ indexed by a complex s-variable:

$$I(f',s) = \int_{H_1(F)\backslash H_1'(\mathbb{A})} \int_{Z_{H_2'}(\mathbb{A})H_2'(F)\backslash H_2'(\mathbb{A})} K_{f'}(h_1,h_2)\eta(h_2)|h_1|^s\,dh_1\,dh_2,$$

where we denote for simplicity $|h_1| = |\det(h_1)|$. Without the s-variable, this was proposed by Jacquet–Rallis in [13]. Note that in general this integral may diverge. But for our purpose, it suffices to consider test functions f' satisfying a local condition as follows. Firstly we may identify $H_1'\backslash G'$ with $Res_{E/F}GL_n$. Then we consider the morphism between F-varieties:

$$\nu : Res_{E/F}GL_n \to S_n$$
$$g \mapsto g\bar{g}^{-1}$$

where S_n is as in Sect. 2. By Hilbert Satz-90, this defines an isomorphism of two affine varieties $Res_{E/F}GL_n/GL_{n,F} \simeq S_n$ and at the level of F-points: $GL_n(E)/GL_n(F) \simeq S_n(F)$. In this way we may naturally identify the double cosets

$$\mathbb{O}(G'(F)) := H_1'(F)\backslash G'(F)/H_2'(F)$$

with the quotient $\mathbb{O}(S_n(F)) = GL_{n-1}(F)\backslash\backslash S_n(F)$ which was defined in Sect. 2. In particular we will say that $g \in G'(F)$ is regular if its image in $S_n(F)$ is so. The same holds if we replace F by F_v for a place v of F. The set of regular elements is open and dense in $G'(F_v)$. For $f_v' \in \mathcal{C}_c^\infty(G'(F_v))$, we say that f_v' is regularly supported if the support of f_v' is contained in the regular locus.

Moreover the map ν is compatible with the obvious integral structures on the source and target in the following sense.

Lemma 3.1 *Let v be a non-archimedean place of F that is unramified in the quadratic extension E/F. Assume that the residue characteristic is odd. Then the image of $G'(\mathcal{O}_{F_v})$ in $S_n(F_v)$ is $S_n(\mathcal{O}_{F_v})$.*

Proof It is trivial if v is split in E. Now we assume that v is non-split. In the following we suppress v in the notation. It is clear that the image of $G'(\mathcal{O}_F)$ in $S_n(F)$ is contained in $S_n(\mathcal{O}_F)$. We need to prove that any $s \in K_S := S_n(\mathcal{O}_F)$ there exists $g \in K := GL_n(\mathcal{O}_E)$ such that $g\bar{g}^{-1} = s$. Take a pre-image $h \in GL_n(E)$ of s. By the Iwasawa decomposition $h = kan$ for $k \in GL_n(\mathcal{O}_E)$, $a \in A(E)$ and a unipotent

$$n = e^X \in N(E),\ X = \begin{pmatrix} 0 & & b_{ij} \\ & \ddots & \\ & & 0 \end{pmatrix}.$$

Since E/F is unramified we may write $a = a_0a_1$ where $a_0 \in K$ and $a_1 \in A(F)$. Replace n by $a_1na_1^{-1}$ to obtain $h = ka_0na_1$. Since $s \in K_S$, we must have $\nu(n) = n\bar{n}^{-1} \in K_S$. Note that

$$n = e^{(X-\bar{X})/2}e^{(X+\bar{X})/2}.$$

Since $\nu(n) = e^{X-\bar{X}} \in K_S$ and 2 is a unit in $\mathcal{O}_E$, we conclude that $e^{(X-\bar{X})/2} \in K$. Therefore

$$h = ka_0 e^{(X-\bar{X})/2} h_0, \quad h_0 = e^{(X+\bar{X})/2} a_1 \in GL_n(F),\ ka_0 e^{(X-\bar{X})/2} \in K.$$

This completes the proof. □

Lemma 3.2 *Suppose that $f' = \bigotimes_v f'_v$ is decomposable. Assume that at some place v, f'_v is regularly supported. Then as a function on $H'_1(\mathbb{A}) \times H'_2(\mathbb{A})$, $K_{f'}(h_1, h_2)$ is compactly supported modulo $H'_1(F) \times Z_{H'_2}(\mathbb{A}) H'_2(F)$. In particular, the integral $I(f', s)$ converges absolutely.*

Proof Note that $Z_{G'} = Z_{H'_2}$. It is equivalent to show that the following kernel function

$$\sum_{\gamma \in G'(F)} f'(h_1^{-1} \gamma h_2)$$

is compactly supported modulo $H'_1(F) \times H'_2(F)$. This new kernel function can be written as

$$\sum_{\gamma \in H'_1(F) \backslash G'(F) / H'_2(F)} \sum_{H'_1(F) \times H'_2(F)} f'(h_1^{-1} \gamma_1^{-1} \gamma \gamma_2 h_2),$$

where the outer sum is over regular γ by the local condition at v. First we claim that the outer sum is finite. Let Ω be the support of f'. Note that the invariants of $G'(\mathbb{A})$ defines a continuous map from $G'(\mathbb{A})$ to $X(\mathbb{A})$ where X is the categorical quotient of S_n by $GL_{n-1,F}$. So the image of Ω will be a compact set in $X(\mathbb{A})$. On the other hand the image of $h_1^{-1} \gamma_1^{-1} \gamma \gamma_2 h_2$ is in the discrete set $X(F)$. Moreover for a fixed $x \in X(F)$ there is at most one $H'_1(F) \times H'_2(F)$ double coset with given invariants. This shows the outer sum is finite.

It remains to show that for a fixed $\gamma_0 \in G'(F)$, the function on $H'_1(\mathbb{A}) \times H'_2(\mathbb{A})$ defined by $(h_1, h_2) \mapsto f'(h_1^{-1} \gamma_0 h_2)$ has compact support. Consider the continuous map $H'_1(\mathbb{A}) \times H'_2(\mathbb{A}) \to G'(\mathbb{A})$ given by $(h_1, h_2) \mapsto h_1^{-1} \gamma_0 h_2$. When γ_0 is regular, this defines an homeomorphism onto a closed subset of $G'(\mathbb{A})$. This implies the desired compactness. □

From now on we always assume that at some place v, f'_v is regularly supported. This is not only because we want to simplify the trace formula, but also because otherwise in the height pairing below we would have self-intersection.

The last lemma allows us to decompose the distribution into a sum of orbital integrals

$$I(f', s) = \sum_{\gamma \in \mathbb{O}(G'(F))_{reg}} O(\gamma, f', s),$$

where for a regular γ we define

$$O(\gamma, f', s) := \int_{H_1'(\mathbb{A})} \int_{H_2'(\mathbb{A})} f'(h_1^{-1} \gamma h_2) |h_1|^s \eta(h_2)\, dh_1\, dh_2. \tag{3.1}$$

The sum is finite for a fixed f'. Fix a decomposition of the measure on $H_i'(\mathbb{A})$ as a product of local Haar measures. We may define the corresponding local orbital integral in an obvious way and we have an Euler product:

$$O(\gamma, f', s) = \prod_v O(\gamma, f_v', s). \tag{3.2}$$

To discuss the transfer of test functions, we need to introduce a "transfer factor": it is a compatible family of smooth map

$$\Omega_v : G'(F_v)_{reg} \longrightarrow \mathbb{C}^\times$$

for all places v of F on the regular locus of $G'(F_v)$ such that

- If $g \in G'(F)$ is regular, then $\Omega_v(g) = 1$ for almost all v and $\prod_v \Omega_v(g) = 1$.
- For any $h_1 \in H_1'(F_v), h_2 \in H_2'(F_v)$ and $s \in S_n(F_v)$, then $\Omega_v(h_1^{-1} g h_2) = \eta(h_2)\Omega_v(g)$.

See [29] for an explicit construction of a transfer factor.

Derivative In the following, for some test functions f' we describe the vanishing order of $O(\gamma, f', s)$ at $s = 0$ in terms of some local data associated to a regular element γ.

Now we fix a one dimensional Hermitian space Eu with $\langle u, u\rangle = 1$. Firstly we consider a local quadratic extension E/F. Let W be a (non-degenerate) Hermitian space of dimension $n-1$ and let $V = W \oplus Eu$ be the orthogonal direct sum of Hermitian spaces. We will be varying W in all isomorphism classes of Hermitian spaces of dimension $n-1$. We also allow E to be split $E = F \times F$ and in this case a Hermitian space over E is a free E-module with a pairing with values in E which is E-linear (conjugate E-linear, resp.) for the first (second, resp.) variable. Then there exists a unique isomorphism class (under the obvious notion of isomorphism between two hermitian spaces). The unitary group $U(W)$ is naturally embedded into $U(V)$ as the stabilizer of the vector u. Let

$$G = G_W = U(W) \times U(V)$$

and let $H = H_W$ be the diagonal embedding of $U(W)$. Then the double cosets $H(F)\backslash G(F)/H(F)$ can be naturally identified with $U(W)(F)\backslash\backslash U(V)(F)$ where the action is by conjugation. Then we say that $\delta = (\delta_1, \delta_2) \in U(W)(F) \times U(V)(F)$ is regular if $\delta_1^{-1}\delta_2 \in U(V)(F)$ is regular in the sense of regularity defined in Sect. 2. Then for regular $\gamma = (\gamma_1, \gamma_2) \in G'(F)$ and $\delta = (\delta_1, \delta_2) \in G(F)$, we say that they are transfers of each other if the image $\nu(\gamma_1^{-1}\gamma_2) \in S_n(F)$ and $\delta_1^{-1}\delta_2 \in U(V)(F)$ are transfers of each other. And for a regular $\delta \in G(F)$ and $f \in \mathcal{C}_c^\infty(G(F))$, the orbital integral

$$O(\delta, f) = \int_{H(F)} \int_{H(F)} f(h_1^{-1}\delta h_2)\, dh_1\, dh_2$$

converges absolutely for any fixed choice of measure on $H(F)$.

Definition 3.3 Given $f' \in \mathcal{C}_c^\infty(G'(F))$ and a tuple $f_W \in \mathcal{C}_c^\infty(G_W(F))$ for each W, we say that f' and the tuple $(f_W)_W$ are (smooth) transfer of each other if we have

$$O(\delta, f_W) = \Omega(\gamma) O(\gamma, f', 0)$$

whenever $\delta \in G_W(F)$ is a transfer of a regular $\gamma \in G'(F)$.

The following conjecture is essentially due to Jacquet–Rallis in [13] where they propose an infinitesimal version.

Conjecture 3.4 *There is a transfer factor such that smooth transfers always exist*: *for any* $f' \in \mathcal{C}_c^\infty(G'(F))$, *there exists its transfer* $(f_W)_W$; *and for any tuple* $(f_W)_W$, *there exists its transfer* $f' \in \mathcal{C}_c^\infty(G'(F))$.

This is trivial if $E = F \times F$. By Lemma 3.1, the fundamental lemma of Jacquet–Rallis in Sect. 2 asserts that for a non-archimedean unramified extension E/F, $1_{G'(\mathcal{O}_F)}$ is the transfer of the pair $f_{W_1} = 1_K$, $f_{W_2} = 0$ where W_1 is the unique W with a self-dual lattice L and K the stabilizer of L, and W_2 is the other isometric class of Hermitian space.

The conjecture is proved for non-archimedean F [29]. But in the following we will not make use of this conjecture in an essential way.

Definition 3.5 Given $f' \in \mathcal{C}_c^\infty(G'(F))$, we say that f' is pure of type W if there is a transfer $(f_{W'})_{W'}$ of f' such that $f'_W = 0$ unless $W' \simeq W$. In this case, we also say that f' is pure of type W and a transfer of f_W.

For instance, $1_{G'(\mathcal{O}_F)}$ is pure of type W_1 as asserted by the fundamental lemma of Jacquet–Rallis.

Now we return to the global setting. Let E/F be a quadratic extension of number fields. We will again be varying the Hermitian space W over E

and we similarly define groups G_W, H_W. We will also need the *incoherent* Hermitian space in Kudla's terminology [16]. In our setting this will be an adelic Hermitian space $\mathbb{W} = \prod_v \mathbb{W}_v$ with determinant in $F^\times$ which does not come from the base change of any Hermitian space W defined over E. Then $\mathbb{V} = \mathbb{W} \oplus \mathbb{A}_E u$, being a sum of incoherent and coherent spaces, is incoherent, too. We will denote by $\mathbb{G} = \mathbb{G}_{\mathbb{W}}$, $\mathbb{H} = \mathbb{H}_{\mathbb{W}}$ the adelic groups.

Now fix an incoherent Hermitian space $\mathbb{W}$. We say that $f' = \bigotimes_v f'_v \in \mathcal{C}_c^\infty(G'(\mathbb{A}))$ is pure of type $\mathbb{W}$ if for all v, f'_v is pure of type $\mathbb{W}_v$. We denote by $\mathcal{C}_c^\infty(G'(\mathbb{A}))^{\mathbb{W}}$ the subspace generated by f' pure of type $\mathbb{W}$.

Let W be a Hermitian space over E. We define a finite subset of non-split places of F that in some sense measures the "distance" between W and $\mathbb{W}$:

$$\Sigma(W, \mathbb{W}) := \{v : W_v \ncong \mathbb{W}_v\}.$$

As $\mathbb{W}$ is incoherent, $\Sigma(W, \mathbb{W})$ is non-empty.

For a fixed $\mathbb{W}$ and a non-archimedean non-split place v, we define a nearby Hermitian space $W(v)$ to be the unique isometric class of W such that $\Sigma(W, \mathbb{W}) = \{v\}$.

For a regular $\gamma \in G'(F)$, let $\delta \in G_W(F)$ be a transfer of γ. Note that this determines a Hermitian space W over E. We call the isometric class of W the type of γ and we also denote it by $W(\gamma)$. Then we define

$$\Sigma(\gamma, \mathbb{W}) = \Sigma(W(\gamma), \mathbb{W}).$$

This depends only on the double coset of γ.

Proposition 3.6 *Let $f' = \bigotimes_v f'_v \in \mathcal{C}_c^\infty(G'(\mathbb{A}))^{\mathbb{W}}$ be pure of type $\mathbb{W}$.*

(i) *Let $\gamma \in G'(F)$ be a regular element. Then we have*

$$\mathrm{ord}_{s=0} O(\gamma, f', s) \geq |\Sigma(\gamma, \mathbb{W})|.$$

(ii) *Assume that for some place v_0, f'_{v_0} is regularly supported. Then we have*

$$I(f', 0) \equiv 0.$$

And we have a decomposition of its first derivative $I'(f, 0) := \frac{d}{ds} I(f, s)|_{s=0}$:

$$I'(f', 0) = \sum_v I'_v(f', 0) \tag{3.3}$$

where $I'_v(f', 0) = 0$ unless v is non-split in which case it is given by

$$\begin{aligned} &I'_v(f', 0) \\ &\quad = \sum_{W, \Sigma(W, \mathbb{W}) = \{v\}} \ \sum_{\gamma \in \mathbb{O}(G'(F))_{reg}, W(\gamma) = W} O(\gamma, f'^{v}, 0) \cdot O'(\gamma, f'_v, 0), \end{aligned} \tag{3.4}$$

$$O'(\gamma, f'_v, 0) := \frac{d}{ds} O(\gamma, f'_v, s)|_{s=0}.$$

(iii) *If v is non-archimedean non-split, the outer sum in $I'_v(f', 0)$ contains only one term, namely, the nearby Hermitian space $W(v)$.*

Proof If $v \in \Sigma(\gamma, \mathbb{W})$, since f'_v is pure of type $\mathbb{W}_v$, we have

$$O(\gamma, f'_v, 0) = 0.$$

Thus (i) follows from the fact that each local orbital integral $O(\gamma, f'_v, s)$ of s is holomorphic at $s = 0$ and the product $O(\gamma, f', s) = \prod_v O(\gamma, f'_v, s)$ is absolutely convergent.

In particular, $O(\gamma, f', 0) = 0$ for all regular γ. Under the assumption in (ii), we have $I(f', s) = \sum_\gamma O(\gamma, f', s)$ for regular double cosets $\gamma \in \mathbb{O}(G'(F))$ and the sum is finite by Lemma 3.2. Thus $I(f', 0) = 0$. Moreover, the first derivative vanishes $\frac{d}{ds}|_{s=0} O(\gamma, f', s) = 0$ unless the set $\Sigma(\gamma, \mathbb{W})$ contains a single element. We thus arrange the non-zero terms in $I'(f', 0)$ according to the single element in $\Sigma(\gamma, \mathbb{W})$. The rest of the proposition follows easily. □

Remark 7 Even if f' is not necessarily regularly supported, as long as f' is of pure type of an incoherent $\mathbb{W}$, the vanishing of $I(f, 0)$ should still hold. But this will require a discussion of the orbital integrals of non regular orbits. The pattern of the decomposition of the first derivative resembles that of the first derivative of the Siegel–Eisenstein series attached to an incoherent quadratic or Hermitian space ([15] for a quadratic space, and [17, Sect. 9] for a Hermitian space).

3.2 Arithmetic Gan–Gross–Prasad conjecture

In the following we recall the global motivation. We only recall a coarse form of the arithmetic Gan–Gross–Prasad conjecture ([3, Sect. 27], [32], [28]). However, to even state the conjecture we need to assume some sort of "standard conjectures" about height pairings.

Fix an incoherent Hermitian space $\mathbb{W}$ of dimension $n - 1$. Now we impose the following hypothesis:

(i) F is totally real and E is a CM extension of F.
(ii) $\mathbb{W}$ is totally definite, namely the signature of the Hermitian space $\mathbb{W}_v$ is $(n - 1, 0)$ for all $v|\infty$.

Then as before we form another incoherent $\mathbb{V} = \mathbb{W} \oplus \mathbb{A}_E u$ with $(u, u) = 1$. In particular, for all $v|\infty$, $\mathbb{W}_v$, $\mathbb{V}_v$ are positively definite and hence the group $\mathbb{G}(F_v)$ is compact.

Then we have a Shimura variety denoted by $Sh(\mathbb{H})$ associated to $\mathbb{W}$. It is a projective system of varieties indexed by open subgroups $K \subset \mathbb{H}(\mathbb{A}^\infty)$ (small enough) defined over E of dimension $n-2$. It is characterized by the following property. For any fixed embedding $v : E \to \mathbb{C}$, we have a unique (coherent) Hermitian space $W(v)$ such that $W(v)_u \simeq \mathbb{W}_u$ for all $u \neq v$ and $W(v)_v$ has signature $(n-2, 1)$. We call $W(v)$ the nearby Hermitian space at v. We have a Shimura variety $Sh(H_{W(v)})$ (for the Shimura datum, see [3, Sect. 27]) defined over the reflex field (or possibly its quadratic extension when $n = 2$) $v(E)$ via the embedding $v : E \to \mathbb{C}$. Then the property characterizing $Sh(\mathbb{H})$ is that for all $v : E \to \mathbb{C}$:

$$Sh(\mathbb{H}) \times_E v(E) \simeq Sh(H_{W(v)}).$$

Similarly, we have a Shimura variety $Sh(\mathbb{G})$ of dimension $(n-2) + (n-1) = 2n-3$ defined over E. The group $\mathbb{G}(\mathbb{A}^\infty)$ acts on $Sh(\mathbb{G})$ by Hecke correspondences. We extend this action to $\mathbb{G}(\mathbb{A})$ by demanding that $\mathbb{G}(\mathbb{A}_\infty)$ acts trivially. Let R denote this action.

If $F \neq \mathbb{Q}$, both $Sh(\mathbb{H})$ and $Sh(\mathbb{G})$ are projective. If $F = \mathbb{Q}$ by abuse of notation we will still denote by $Sh(\mathbb{H})$ the toroidal compactification if $Sh(\mathbb{H})$ is not projective. In our case the toroidal compactification is unique. Then we have a closed immersion of projective system of varieties:

$$Sh(\mathbb{H}) \hookrightarrow Sh(\mathbb{G}).$$

It is $\mathbb{H}(\mathbb{A})$-equivariant. Let $Ch^*(Sh(\mathbb{G})_K)$ be the Chow groups with complex coefficients for $K \subset \mathbb{G}(\mathbb{A}^\infty)$. Let $Ch^*(Sh(\mathbb{G})) = \varinjlim_K Ch^*(Sh(\mathbb{G})_K)$ ($Ch_*(Sh(\mathbb{G}))$, resp.) be the inductive (projective, resp.) limit with respect to natural pull-back (push-forward, resp.) maps. Then $Sh(\mathbb{H})$ defines an element denoted by $[Sh(\mathbb{H})]$ in $Ch_{n-2}(Sh(\mathbb{G}))$. Then we have a cycle class map $\mathrm{cl} : Ch^*(Sh(\mathbb{G})) \to H^{2*}(Sh(\mathbb{G})) = \varinjlim_K H^{2*}(Sh(\mathbb{G})_K)$ where the cohomology is any fixed Weil cohomology with complex coefficients. Moreover, for $f \in \mathcal{C}_c^\infty(\mathbb{G})$, we denote by $R(f)$ the Hecke correspondence. *Though not necessary, for the sake of simplicity we will assume that:*

$$f_\infty = 1_{\mathbb{G}(\mathbb{A}_\infty)} \tag{3.5}$$

and for each $v|\infty$ we fix a measure on $\mathbb{G}(F_v)$ such that the volume of $\mathbb{G}(F_v)$ is one (*note that $\mathbb{G}(F_v) = U(\mathbb{W}_v) \times U(\mathbb{V}_v)$ is compact by* (i)). Under this simplification, $R(f)$ is simply the Hecke operator associated to $f^\infty \in \mathcal{C}_c^\infty(\mathbb{G}(\mathbb{A}^\infty))$.

Let $\mathcal{A}(\mathbb{G})$ be the set of irreducible admissible representations π of $\mathbb{G}$ that occur in the middle dimensional cohomology $H^{2n-3}(Sh(\mathbb{G}))$. Note that by definition, $\mathbb{G}(\mathbb{A}_\infty)$ (a compact group) acts trivially. Consider the natural surjective map

$$\mathcal{C}_c^\infty(\mathbb{G}) \to \bigoplus_{\pi\in\mathcal{A}(\mathbb{G})} End(\pi) \simeq \bigoplus_{\pi\in\mathcal{A}(\mathbb{G})} \pi\otimes\widetilde{\pi}.$$

It is bi-$\mathbb{G}(\mathbb{A})$ equivariant where $\mathbb{G}(\mathbb{A})\times\mathbb{G}(\mathbb{A})$ acts on LHS by left and right translation. For $\phi\otimes\widetilde{\phi}\in\pi\otimes\widetilde{\pi}$ considered as an element of $\bigoplus_{\pi\in\mathcal{A}(\mathbb{G})}\pi\otimes\widetilde{\pi}$, we choose any lifting $f_{\phi\otimes\widetilde{\phi}}\in\mathcal{C}_c^\infty(\mathbb{G})$. Then it is essentially conjectured in the work of Beilinson and Bloch [1, 2] specialized to our situation:

Conjecture 3.7

(A) *The cycle class map* $\mathrm{cl}: Ch^{n-1}(Sh(\mathbb{G}))\to H^{2n-2}(Sh(\mathbb{G}))$ *splits* $\mathbb{G}$-*equivariantly. We thus have a* $\mathbb{G}$-*equivariant projection* $Ch^{n-1}(Sh(\mathbb{G}))\to Ch^{n-1}(\mathbb{G})_0=\mathrm{Ker}(\mathrm{cl})$. *We denote by* $[Sh(\mathbb{H})]_0$ *the projection of* $[Sh(\mathbb{H})]$ *and we call it a cohomological trivialization of the cycle* $[Sh(\mathbb{H})]$.

(B) *Let* $\pi\in\mathcal{A}(\mathbb{G})$. *Then* $R(f_{\phi\otimes\widetilde{\phi}})[Sh(\mathbb{H})]_0$ *is independent of the choice of the lifting* $f_{\phi\otimes\widetilde{\phi}}$.

Assuming this conjecture above and assuming that the height pairing $\langle\cdot,\cdot\rangle_B$ of Beilinson–Bloch [1, 2] is well-defined, we may define a linear form on $\pi\otimes\widetilde{\pi}$ as follows:

$$\ell_{\mathbb{H}}(\phi\otimes\widetilde{\phi}) := \langle R(f_{\phi\otimes\widetilde{\phi}})[Sh(\mathbb{H})]_0, [Sh(\mathbb{H})]_0\rangle_{BB}.$$

Note that this is understood as follows. Fix a Haar measure on $\mathbb{H}(\mathbb{A})$ given by a product of measures on $\mathbb{H}(F_v)$ for all v. Choose a compact open $K'\subset\mathbb{G}(\mathbb{A}^\infty)$ and let $K=K'\cap\mathbb{H}(\mathbb{A}^\infty)$. Then we define

$$\begin{aligned}&\langle R(f)[Sh(\mathbb{H})]_0, [Sh(\mathbb{H})]_0\rangle_{BB}\\ &\quad := \mathrm{vol}(K)^2\langle R(f_{\phi\otimes\widetilde{\phi}})[Sh(\mathbb{H})_K]_0, [Sh(\mathbb{H})_K]_0\rangle_{BB},\end{aligned}$$

where the intersection takes place on $Sh(\mathbb{G})_{K'}$. Then this definition depends only on f but not on the choice of compact open K' (also cf. [15]).

Note that $Sh(\mathbb{H})$ is invariant under $\mathbb{H}(\mathbb{A})$. Then the functorial property of height pairing shows that

$$\ell_{\mathbb{H}}\in Hom_{\mathbb{H}\times\mathbb{H}}(\pi\otimes\widetilde{\pi},\mathbb{C}).$$

Remark 8 When $n=2$, $Sh(\mathbb{G})$ is a Shimura curve. $Sh(\mathbb{H})$ is a sum of CM points. When the curve needs a compactification, the cohomological trivialization $Sh(\mathbb{H})_0$ can be achieved by subtracting a suitable linear combination of the boundary components. In general we need to use the method of trivializing cohomology in [25]. Then the conjecture above holds by [25] and the Beilinson–Bloch height pairing is the same as Neron–Tate pairing. So all assumptions hold unconditionally.

We are now ready to state the arithmetic Gan–Gross–Prasad conjecture:

Conjecture 3.8 *Assume the hypothesis* (i), (ii) *and Conjecture* 3.7 *above. Let* $\pi \in \mathcal{A}(\mathbb{G})$. *Then the following are equivalent*:

(a) $L'(\pi, R, 1/2) \neq 0$ *where* $L(\pi, R, s)$ *denotes the Rankin–Selberg L-function of the base change of* π *to* $G'(\mathbb{A})$ (*cf.* [3, *Sect.* 27]).
(b) *There exists a* $\mathbb{W}$ *satisfying* (ii) *such that the linear form* $\ell_{\mathbb{H}} \neq 0$.

Here we define the base change of π to be an automorphic representation of $G'(\mathbb{A})$ which is locally a base change of π_v for almost all places v. The base change is expected to exist by a special case of the Langlands functoriality conjecture and is proved under some conditions (for example, [11]).

Remark 9 When $n = 2$, $Sh(\mathbb{G})$ is a Shimura curve and this conjecture can be essentially deduced from the work of Gross–Zagier [10] and Yuan–Zhang–Zhang [25].

3.3 Main global result

Now we assume that $F = \mathbb{Q}$ and $E = \mathbb{Q}[\sqrt{-d}]$ an imaginary quadratic field of discriminant $-d$. Let $\mathbb{W}$ be a definite incoherent Hermitian space with Hermitian form $\langle \cdot, \cdot \rangle$. When necessary we will convert it into a symplectic form by $(x, y) = \mathrm{tr}(\langle x, y\rangle/\sqrt{-d})$.

We assume part (A) of Conjecture 3.7. We assume also that the Beilinson–Bloch height is well-defined. Then we may form a distribution on $\mathbb{G}$:

$$J(f) = \langle R(f)[Sh(\mathbb{H})]_0, [Sh(\mathbb{H})]_0\rangle_{BB}. \tag{3.6}$$

When $R(f)[Sh(\mathbb{H})]_0$ and $[Sh(\mathbb{H})]_0$ as algebraic cycles are disjoint, the height pairing is a sum of local heights:

$$J(f) = \sum_v J_v(f) \tag{3.7}$$

where the $J_v(f)$ is the sum of local heights at all places w of E lying above the place v of F.

In this article, we restrict ourselves to the case of good reduction namely, those non-archimedean places p such that

- p is unramified inert or split in E,
- $\mathbb{W}_p$ has a self-dual lattice and $f_p = 1_{K_p}$.

Note that to the local height $J_p(f)$, there are contributions from various sources: (1) the cycle $[Sh(\mathbb{H})]$ on the non-compactified Shimura variety; (2) the compactification; (3) the difference between the cycle $[Sh(\mathbb{H})]$ and its cohomological trivialization $[Sh(\mathbb{H})]_0$. Among them, (1) is the main contribution and will be denoted by J_p^m ("m" means "main") and the rest will be denoted by J_p^b ("b" means "boundary"). Then we decompose the local height

$$J_p(f) = J_p^m(f) + J_p^b(f). \tag{3.8}$$

In the following, we will give a sufficient condition for f such that $R(f)Sh(\mathbb{H})$ and $Sh(\mathbb{H})$ is disjoint in the non-compactified Shimura variety. We then construct smooth integral models of $Sh(\mathbb{H})$ and $Sh(\mathbb{G})$ over the completion of the maximal unramified extension of $\mathcal{O}_{E,w}$ to calculate $J_p^m(f)$. So in the end, even though the height pairing is conditionally defined, the term $J_v^m(f)$ can be defined and calculated unconditionally. Finally we compare $J_p^m(f)$ with the term $I_p'(f',0)$ if $f' = \bigotimes f_p'$ and f_p' is a transfer of the tuple given by $f_{W_p} = f_p$ when $W_p = \mathbb{W}_p$ and zero otherwise.

Integral models Now assume that p is inert or split and $\mathbb{W}_p$ has a self-dual lattice L. Let $K_{p,0}$ be the stabilizer of L. It is a hyperspecial subgroup of $\mathbb{H}(\mathbb{Q}_p)$. Similarly we have a hyperspecial subgroup $K'_{p,0}$ of $\mathbb{G}(\mathbb{Q}_p)$.

Let $\mathbb{F}$ be a fixed algebraic closure of the finite field $\mathbb{F}_p$ of order p. Let $\mathcal{O}_{E,(p)}$ be the localization of $\mathcal{O}_E$ at p. Let $W = W(\mathbb{F})$ be the Witt ring of $\mathbb{F}$. We fix an embedding $\mathcal{O}_{E,(p)} \hookrightarrow W$. We construct a smooth integral model of $Sh(\mathbb{H})_{K^pK_0^p}$ over W. The observation is that W contains all prime-to-p roots of unity $\widehat{\mathbb{Z}}^p(1)$ so we may fix a trivialization:

$$\widehat{\mathbb{Z}}^p(1) \simeq \widehat{\mathbb{Z}}^p. \tag{3.9}$$

We will fix such an isomorphism. Essentially following Kottwitz [14] we consider a moduli functor $\mathcal{A}_{K^p}^{\mathbb{H}}$ for a compact open subgroup $K^p \subseteq \mathbb{H}(\mathbb{A}^{\infty,p})$. The functor $\mathcal{A}_{K^p}^{\mathbb{H}}$ on the category $\mathcal{S}_W$ of W-schemes sends each $S \in \mathcal{S}_W$ to the set of isomorphism classes of quadruples $(A, \iota_A, \bar{\lambda}_A, \bar{\eta}_A)$ where

- A is an abelian scheme over S of relative dimension equal to $n-1$.
- $\iota_A : \mathcal{O}_{E,(p)} \to End(A) \otimes \mathbb{Z}_{(p)}$ is an injective homomorphism of $\mathbb{Z}_{(p)}$-algebras. We also make the following convention: if $A^\vee$ is the dual abelian scheme of A, we define an action of $\mathcal{O}_{E,(p)}$ by $\iota_{A^\vee} : \mathcal{O}_{E,(p)} \to End(A^\vee) \otimes \mathcal{O}_{(p)}$,

$$\iota_{A^\vee}(e) = \iota_A(\bar{e})^\vee.$$

- $\bar{\lambda}_A \subset Hom(A, A^\vee) \otimes \mathbb{Q}$ is a one dimensional $\mathbb{Q}$-subspace which contains a p-principal $\mathcal{O}_{E,(p)}$-linear polarization.
- $\bar{\eta}_A$ is a K^p-class of isomorphisms of symplectic forms: $H_1(A, \mathbb{A}^{\infty,p}) \simeq \mathbb{W}^{\infty,p}$ (more than isomorphism up to similitudes). Here the Weil pairing on $H_1(A, \mathbb{A}^{\infty,v})$ takes values in $\mathbb{A}^{\infty,v}(1) \simeq \mathbb{A}^{\infty,v}$ by the trivialization.

More precisely, for a connected scheme S over W, choosing a geometric point s of S, we may think of the rational Tate module of A as the Tate module $H_1(A_s, \mathbb{A}^{\infty,p})$ of A_s with a structure of $\pi_1(S,s)$-module. Then a K^p-level structure is a K^p-orbit of $\mathcal{O}_{E,(p)}$-linear isomorphisms of symplectic forms: $H_1(A_s, \mathbb{A}^{\infty,p}) \simeq \mathbb{W}^{\infty,p}$ fixed by $\pi_1(S,s)$. As S is a scheme over W, the group $\pi_1(S,s)$ preserves the symplectic form on $H_1(A_s, \mathbb{A}^{\infty,p})$ valued in $\mathbb{A}^{\infty,p}$ (not just "up to similitudes", cf. [14, p. 390])

such that (A, ι_A) satisfies Kottwitz's determinant condition of signature $(r,s) = (n-2, 1)$ for all $e \in \mathcal{O}_{E,(p)}$:

$$\mathrm{charpol}(e, Lie(A))(T) = (T-e)^r (T-\bar{e})^s \in \mathcal{O}_S[T]$$

where e is considered as a section of $\mathcal{O}_S$ via the structure morphism $S \to Spec(W)$.

Two quadruples $(A, \iota_A, \bar{\lambda}_A, \bar{\eta}_A)$ and $(A', \iota_{A'}, \bar{\lambda}_{A'}, \bar{\eta}_{A'})$ are called isomorphic if there exists an $\mathcal{O}_{E,(v)}$-linear isogeny of prime-to-p degree: $\alpha : A \to A'$ such that $\alpha^*(\bar{\lambda}_{A'}) = \bar{\lambda}_A$ and $\eta_{A'} \circ H_1(\alpha, \mathbb{A}^{\infty,p}) = \eta_A$.

The functor $\mathcal{A}_{K^p}$ is represented by a smooth quasi-projective scheme over W when K^p is small enough, which we always assume from now on. This gives a smooth integral model of $Sh(\mathbb{H})_{K^p K_{p,0}} \times_E W_{\mathbb{Q}}$.

Similarly we have a smooth integral model $\mathcal{A}^{\mathbb{G}}_{K'^p}$ of $Sh(\mathbb{G})_{K'^p K'_{p,0}}$ over W classifying a pair (A, B) with obvious additional structures. This gives a smooth integral model of $Sh(\mathbb{G}_{K'^p K'_{p,0}}) \times_E W_{\mathbb{Q}}$. The action of the Hecke algebra $\mathcal{H}(\mathbb{G}(\mathbb{A}^{\infty,p} /\!/ K^p))$ extend to étale correspondences on $\mathcal{A}_{K'^p}$. Again we extend the action trivially to the archimedean component $\mathbb{G}(\mathbb{A}_\infty)$. For $f = f^p f_p \in \mathcal{H}(\mathbb{G}(\mathbb{A}) /\!/ K)$ with $f_p = 1_{K'_{p,0}}$, we still denote by $R(f)$ the Hecke correspondence on $\mathcal{A}_{K'^p}$.

We now define a morphism:

$$i : \mathcal{A}^{\mathbb{H}}_{K^p} \hookrightarrow \mathcal{A}^{\mathbb{G}}_{K'^p}$$

when $K^p = K'^p \cap \mathbb{H}(\mathbb{A}^{\infty,p})$. To do so, we fix an elliptic curve $\mathcal{E}$ over W with CM by $\mathcal{O}_E$. It is unique up to prime-to-p isogenies. We denote $\iota_0 : \mathcal{O}_E \hookrightarrow End(\mathcal{E})$ such that the action on Lie algebra is the fixed embedding of $\mathcal{O}_E$ into W (i.e., signature $(1,0)$). The canonical polarization is a principle polarization denoted by λ_0. We fix any $\mathcal{O}_{E,(p)}$-linear isomorphism of symplectic forms $H_1(\mathcal{E}, \mathbb{A}^{\infty,p}) \simeq Eu \otimes \mathbb{A}^{\infty,p}$ where Eu denotes the symplectic space form the trivial hermitian space of dimension one.

We define i by sending $(A, \iota_A, \bar{\lambda}_A, \bar{\eta}_A)$ to $(A, A \times \mathcal{E})$ where $A \times \mathcal{E}$ has the additional structure $(\iota_A \times \iota_0, \bar{\lambda}_A \times \lambda_0, \overline{\eta_A \times \eta_0})$.

Now we fix a Haar measure on $\mathbb{H}(F_v)$ for non-archimedean v such that the volume of $K_{v,0}$ is one for almost all places v including $v = p$. And we take the product measure on $\mathbb{H}(\mathbb{A}^\infty)$. When p is inert, we define the main term of the local height at p as

$$J_p^m(f) = \mathrm{vol}(K)^2 (R(f)\mathcal{A}^{\mathbb{H}}_{K^p} \cdot \mathcal{A}^{\mathbb{H}}_{K^p})_{\mathcal{A}^{\mathbb{G}}_{K'^p}} \cdot \log p^2 \tag{3.10}$$

where the intersection number on $\mathcal{A}^{\mathbb{G}}_{K'^p}$ is defined in a similar way as in Sect. 2.2 (cf. [17, Sect. 11]):

$$(R(f)\mathcal{A}^{\mathbb{H}}_{K^p} \cdot \mathcal{A}^{\mathbb{H}}_{K^p})_{\mathcal{A}^{\mathbb{G}}_{K'^p}} := \chi(\mathcal{O}_{R(f)\mathcal{A}^{\mathbb{H}}_{K^p}} \otimes^{\mathbb{L}} \mathcal{O}_{\mathcal{A}^{\mathbb{H}}_{K^p}}).$$

When p is split, there are two places v_1, v_2 of E lying above p where each place corresponds to an embedding of $\mathcal{O}_E$ into the Witt ring W. For each of them we give one integral model described above. By definition $J_p^m(f)$ is a sum of two terms with each given by the intersection number as above replacing p^2 by p.

p: inert Recall that for p non-split, there is a unique nearby Hermitian space W with $\Sigma(\mathbb{W}, W) = \{p\}$. We then have an algebraic group $G_{W(p)}$ over $\mathbb{Q}$ which contains the unitary group $H_{W(p)} \simeq U(W(p))$ as a subgroup (cf. Sect. 3.1). We will consider the orbital integral $O(\delta, f^p)$. We now take the measure on $\mathbb{H}(\mathbb{A})$ as the product of the measure on $\mathbb{H}(\mathbb{A}^\infty)$ and $\mathbb{H}(\mathbb{A}_\infty)$: on $\mathbb{H}(\mathbb{A}^\infty)$ we take the measure which was normalized in the definition of $J_p^m(f)$, and we normalize the measure on $\mathbb{H}(\mathbb{A}_\infty)$ such that $\mathrm{vol}(\mathbb{H}(F_\infty)) = 1$.

Theorem 3.9 *Suppose that p is inert. Suppose that $f = \bigotimes_v f_v$ satisfies*

(1) $f_p = 1_{K'_{p,0}}$ *is the characteristic function of the hyperspecial subgroup* $K'_{p,0}$.
(2) *For at least one place $v \neq p$, the test function f_v is supported in regular orbits.*
(3) f_∞ *is as in* (3.5).

Then we have a decomposition of $J_p^m(f)$ into a sum of rational regular orbits associated to the nearby unitary groups $G_{W(p)}$:

$$J_p^m(f) = 2 \sum_{\delta \in \mathbb{O}(G_{W(p)}(\mathbb{Q}))_{reg}} O(\delta, f^p) \cdot O'(\delta, f_p)$$

where $O'(\delta, f_p)$ is the intersection number on unitary Rapoport–Zink space defined in Sect. 2, (2.7).

Remark 10 The expression is very similar to the counting points formula on Shimura varieties mod p (for instance, cf. [14, p.376]).

Proof Fix K'^p such that f^p is bi-K'^p-invariant. We first claim that the set theoretical intersection $R(f)\mathcal{A}^{\mathbb{H}}_{K^p} \cap \mathcal{A}^{\mathbb{H}}_{K^p}$ only happens in the supersingular locus. It is easy to see that the intersection on generic fiber is empty (this can also be seen from the argument below). To simplify the notation, we will shorten $\mathcal{A}^{\mathbb{H}}_{K^p}$ etc. to $\mathcal{A}^{\mathbb{H}}$ etc. We also shorten the quadruple $(A, \iota_A, \bar{\lambda}_A, \bar{\eta}_A)$ to A though we always keep the rest in mind. By the moduli interpretation, if $(A, B) \in \mathcal{A}^{\mathbb{G}}(\mathbb{F})$ lies in the intersection, we have (with additional structure) for some $(g_1, g_2) \in \mathbb{G}(\mathbb{A}^{\infty,p})$ regular at the place $v \neq p$

$$B = A \times \mathbb{E}, \quad g_1 B = (g_2 A) \times \mathbb{E}$$

where $\mathbb{E}$ is the reduction of $\mathcal{E}$. We thus may assume that $g_2 = 1$ by replacing g_1 by $g_2^{-1} g_1$. And we have

$$g_1(A \times \mathbb{E}) = A \times \mathbb{E} = B. \tag{3.11}$$

Note that this means there is a prime-to-p isogeny $\beta_1 : B \to B$ that sends the level structure $\bar{\eta}_B$ to $g_1 \circ \bar{\eta}_B$. Let $\alpha_1 : \mathbb{E} \to B$ be the composition of the embedding of $\mathbb{E}$ to B and β_1. In general, for $i = 1, 2, \ldots, n$, replacing g_1 by g_1^i in (3.11), we obtain a prime-to-p isogeny β_i and then α_i. Now we define a homomorphism

$$\alpha := (\alpha_1, \ldots, \alpha_n) : \quad \mathbb{E}^n \to B.$$

Consider the level structure at v where f_v is regularly supported, namely the isomorphism

$$\eta_{B,v} : H_1(B, \mathbb{Q}_v) \simeq \mathbb{V}_v.$$

Note that α_i induces an embedding of a one-dimensional E_v-space

$$\alpha_{i*} : H_1(\mathbb{E}, \mathbb{Q}_v) \to \mathbb{V}_v.$$

Let the image of α_{1*} be $E_v u$. Then it is easy to see that α_i has image $E_v g_1^i u$. Note that the condition that g_1 is regular implies that the vectors $g_1^i u$, $i = 1, 2, \ldots, n$ are linearly independent. Therefore, the homomorphism α induces an isomorphism

$$H_1(\mathbb{E}^n, \mathbb{Q}_v) \simeq H_1(B, \mathbb{Q}_v).$$

This implies that α is an isogeny. We have thus proved that if g_1 is regular, then we have an isogeny $\mathbb{E}^n \sim B$. Note that $\mathbb{E}$ is supersingular by our choice. This proves the claim.

Now we let $\mathcal{A}^{\mathbb{H},ss}$ be the supersingular locus, which forms only a single isogeny class [23]. And let $\mathcal{A}^{\mathbb{H},/ss}$ be the formal completion of $\mathcal{A}^{\mathbb{H}}$ along $\mathcal{A}^{\mathbb{H},ss}$. By the work of Rapoport–Zink [21, Theorem 6.30] on the uniformization of the supersingular isogeny class, we have an isomorphism of formal schemes over the Witt ring W

$$H_{W(p)}(\mathbb{Q})\backslash\mathcal{N}_{n-1} \times \mathbb{H}(\mathbb{A}^{\infty,p})/K^p \simeq \mathcal{A}^{\mathbb{H},/ss}.$$

Here the isomorphism is with respect to a fixed supersingular point and the $\mathbb{Q}$-group $H_{W(p)}$ is the group of quasi-isogenies of this fixed supersingular abelian variety (with additional structure). It can be identified with the unitary group $G_{W(p)}$ associated to the nearby Hermitian space $W(p)$. Similar uniformization holds for $\mathcal{A}^{\mathbb{G}}$:

$$G_{W(p)}(\mathbb{Q})\backslash\mathcal{N} \times \mathbb{G}(\mathbb{A}^{\infty,p})/K'^{p} \simeq \mathcal{A}^{\mathbb{G},/ss}$$

where $\mathcal{N} = \mathcal{N}_{n-1} \times_W \mathcal{N}_n$ is the product of unitary Rapoport–Zink spaces as in Sect. 2.2. And as before we will denote by $\Delta_{\mathcal{N}_{n-1}}$ the diagonal embedding of $\mathcal{N}_{n-1}$.

We denote by $[\Delta_{\mathcal{N}_{n-1}}, h]$ a $G_{W(p)}(\mathbb{Q}) \times K'^{p}$-coset for $h \in \mathbb{H}(\mathbb{A}^{\infty,p})$. Thus we may break $\chi(\mathcal{O}_{R(f)\mathcal{A}^{\mathbb{H}}_{K^p}} \otimes^{\mathbb{L}} \mathcal{O}_{\mathcal{A}^{\mathbb{H}}_{K^p}})$ into a sum

$$\begin{aligned}\chi(&\mathcal{O}_{R(f)\mathcal{A}^{\mathbb{H}}_{K^p}} \otimes^{\mathbb{L}} \mathcal{O}_{\mathcal{A}^{\mathbb{H}}_{K^p}})\\ &= \sum_{g,h_1,h_2} f^{\infty,p}(g)\cdot\chi(\mathcal{O}_{[\Delta_{\mathcal{N}_{n-1}},h_1 g]} \otimes^{\mathbb{L}} \mathcal{O}_{[\Delta_{\mathcal{N}_{n-1}},h_2]}),\end{aligned} \tag{3.12}$$

where $g \in \mathbb{G}(\mathbb{A}^{\infty})/K'$ and $h_1, h_2 \in H_{W(p)}(\mathbb{Q})\backslash\mathbb{H}(\mathbb{A}^{\infty,p})/K$. The term for g, h_1, h_2 is zero unless there exist $\delta \in G_{W(p)}(\mathbb{Q})$ such that

$$h_1 g = \delta h_2 \pmod{K'^{p}}$$

in which case the term is equal to (noting that $g = h_2^{-1}\delta h_1 \pmod{K'^{p}}$)

$$f^{\infty,p}(h_2^{-1}\delta h_1)\cdot\chi(\mathcal{O}_{\Delta_{\mathcal{N}_{n-1}}} \otimes^{\mathbb{L}} \mathcal{O}_{\delta^*\Delta_{\mathcal{N}_{n-1}}}).$$

Note that δ is only well-defined in an $H_{W(p)}(\mathbb{Q})$-double coset. We may rewrite the sum (3.12) as

$$\begin{aligned}&\sum_{\delta\in H_{W(p)}(\mathbb{Q})\backslash G_{W(p)}(\mathbb{Q})/H_{W(p)}(\mathbb{Q})}\left(\sum_{h_i\in\mathbb{H}(\mathbb{A}^{\infty,p})/K^p} f^{\infty,p}(h_1^{-1}\delta h_2)\right)\\ &\quad\cdot\chi(\mathcal{O}_{\Delta_{\mathcal{N}_{n-1}}} \otimes^{\mathbb{L}} \mathcal{O}_{\delta^*\Delta_{\mathcal{N}_{n-1}}}),\end{aligned} \tag{3.13}$$

where the sum is actually taken over regular δ by the regularity of the support. Note that in our formulation of the arithmetic fundamental lemma, we have set

$$O'(\delta, f_p) = \chi(\mathcal{O}_{\Delta_{\mathcal{N}_{n-1}}} \otimes^{\mathbb{L}} \mathcal{O}_{\delta^* \Delta_{\mathcal{N}_{n-1}}}) \log p.$$

Therefore returning to the main term (3.10), we obtain

$$\begin{aligned} J_p^m(f) = 2 \cdot \mathrm{vol}(K)^2 \\ \cdot \sum_{\delta \in H_{W(p)}(\mathbb{Q}) \backslash G_{W(p)}(\mathbb{Q}) / H_{W(p)}(\mathbb{Q})} \left(\sum_{h_i \in \mathbb{H}(\mathbb{A}^{\infty,p})/K^p} f^{\infty,p}(h_1^{-1} \delta h_2) \right) \\ \cdot O'(\delta, f_p). \end{aligned}$$

This is equal to

$$J_p^m(f) = 2 \sum_{\delta \in \mathbb{O}(G_{W(p)}(\mathbb{Q}))_{reg}} O(\delta, f^{\infty,p}) O'(\delta, f_p).$$

Under our choice of f_∞ and the measure on $\mathbb{H}(F_\infty)$, we have

$$O(\delta, f_\infty) = 1$$

for any $\delta \in \mathbb{G}(\mathbb{A}_\infty)$. Hence we have

$$J_p^m(f) = 2 \sum_{\delta \in \mathbb{O}(G_{W(p)}(\mathbb{Q}))_{reg}} O(\delta, f^p) O'(\delta, f_p).$$

□

p: *split*

Theorem 3.10 *Suppose that p is split. Suppose that $f = \bigotimes_v f_v$ satisfies* (1), (2), (3) *as in Theorem* 3.9. *Then we have*

$$J_p^m(f) = 0.$$

Proof Now let v_1, v_2 be the two places of E lying above p. We will prove that the local height at v_1 is zero and the argument is the same for v_2. Let $\mathbb{E}$ be the reduction to $\mathbb{F}$ of $\mathcal{E}$. As p is split, $\mathbb{E}$ is an elliptic curve with ordinary reduction and $End^0(\mathbb{E}) \simeq E$. Let $\mathbb{X}_n = \overline{\mathbb{E}} \times \mathbb{E}^{n-1}$ (together with the additional structure) similar to the supersingular case as in Sect. 2. By the same argument as in the proof of the previous theorem, we see that the intersection is supported in the ordinary locus, more precisely in the isogeny class of $\mathbb{X}_n$. We denote this isogeny class by ξ. Then we have a parametrization of geometric points:

$$I(\mathbb{Q}) \backslash \mathcal{M}(\mathbb{F}) \times \mathbb{G}(\mathbb{A}^{\infty,p})/K'^p \simeq \xi$$

for some Rapoport–Zink space $\mathcal{M}$. Here $I(\mathbb{Q})$ is the group of quasi-isogenies of $\mathbb{X}_n$ preserving the additional structure. It consists of the $\mathbb{Q}$-points of some reductive group I. We will show that $I(\mathbb{Q})$ is “small” in some sense.

Suppose that f_v is regularly supported. We consider the embedding of $I(\mathbb{Q}) \subset \mathbb{G}(F_v)$. It is enough to show that as a subset of $\mathbb{G}(F_v)$, any element δ of $I(\mathbb{Q})$ is not regular. Note that we may describe the group $I(\mathbb{Q})$ as follows. We write $I = I_{n-1} \times I_n$ and it is enough to consider $\delta = (1, g)$ for $g \in I_n(\mathbb{Q})$. Since $End^0(\mathbb{E}) \simeq E$, we may identify $End^0(\mathbb{X}_n) = M_n(E)$ (endomorphisms that do not necessarily preserve the additional structure). Then, without loss of generality, we may suppose that the embedding $\iota : E \to End^0(\mathbb{X}_n)$ is given by $x \to \mathrm{diag}[\bar{x}, x, x, \dots, x] \in M_n(E)$. Since any element in $I(\mathbb{Q})$ commutes with ι, an easy calculation shows that $I_n(\mathbb{Q})$ is a subgroup of the Levi subgroup of $GL_n(E)$ consisting elements of the form $\mathrm{diag}[a, D]$ where $a \in GL_1(E)$, $D \in GL_{n-1}(E)$. Now consider the induced action of $g = (a, D)$ on $H_1(\mathbb{X}_n, \mathbb{Q}_v)$ for the place v. We identify $H_1(\mathbb{X}_n, \mathbb{Q}_v)$ with $\mathbb{W}_v \oplus E_v u$ and consider u as a generator of $H_1(\mathbb{E}, \mathbb{Q}_v)$ as an $E \otimes \mathbb{Q}_v$-module. Then for $g = (a, D)$, $g^i u$ will lie in the subspace $H_1(\mathbb{E}^{n-1}, \mathbb{Q}_v)$ of $H_1(\mathbb{X}_n, \mathbb{Q}_v)$. In particular $u, gu, \dots, g^{n-1}u$ cannot span $H_1(\mathbb{X}_n, \mathbb{Q}_v)$. Hence $g = (a, D)$ cannot be regular and we complete the proof. $\square$

Comparison Finally we can compare the local height $J_p^m(f)$ at a good place with the first derivative of the Jacquet–Rallis relative trace formula. Recall that for $f' \in \mathcal{C}_c^\infty(G'(\mathbb{A}))$ with regular supported f'_v for some place v, we have a decomposition (Proposition 3.6):

$$I'(f', 0) = \sum_p I'_p(f', 0).$$

Theorem 3.11 *Let E be an imaginary quadratic field and let $\mathbb{W}$ be an definite incoherent Hermitian space. Let $f = \bigotimes f_v \in \mathcal{C}_c^\infty(\mathbb{G})$ and let $f' = \bigotimes f'_v$ be pure of type $\mathbb{W}$ and a transfer of f. For a fixed prime p, suppose that f' is regularly supported at some place different from p and $f = \bigotimes_v f_v$ satisfies* (1), (2), (3) *in Theorem* 3.9. *Then when p is split, we have*

$$I'_p(f', 0) = J'_p(f).$$

When p is inert, assuming the arithmetic fundamental lemma (*Conjecture* 2.9) *we have*

$$J_p^m(f) = I'_p(f', 0).$$

Proof When p is split, $I'_p(f', 0) = J'_p(f) = 0$ by Proposition 3.6, Theorem 3.10. When p is inert, this is evident by the definition of transfer, Proposition 3.6 and Theorem 3.9 and the arithmetic fundamental lemma. $\square$

4 Derivative of orbital integrals: $n = 3$

In this section we calculate the derivative of orbital integrals when $n = 3$. The computation is similar to that of Jacquet–Rallis in [13] for Lie algebras. But in several places our situation is more difficult and requires some work (e.g., Lemma 4.7).

4.1 Orbits of $S_3(F)$

Let F be a finite extension of $\mathbb{Q}_p$ with odd p. Instead of considering the orbital integral

$$O(\gamma, s) := \int_H 1_{K_S}(h\gamma h^{-1})|\det(h)|^{-s}\eta(h)\,dh \tag{4.1}$$

for $H = GL_2(F)$, we will perform a substitution to obtain a simpler orbital integral.

Consider the following model of $M_2(F)$ (by endowing it with a "complex" structure)

$$D = \left\{ \begin{pmatrix} t_1 & t_2 \\ \bar{t}_2 & t_1 \end{pmatrix} \,\middle|\, t_1, t_2 \in E, t_1\bar{t}_1 - t_2\bar{t}_2 \neq 0 \right\} \simeq M_2(F).$$

And

$$H' =: D^\times \simeq GL_2(F)$$

where the isomorphism is given explicitly as follows. Let

$$w = \begin{pmatrix} -\sqrt{\epsilon} & \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in GL_2(\mathcal{O}_E)$$

where we fix once for all $\epsilon \in \mathcal{O}_F^\times$ such that $E = F[\sqrt{\epsilon}]$. Then $g \in GL_2(F) \mapsto w^{-1}gw \in H'$ defines an isomorphism. Indeed, we have

$$w^{-1}\begin{pmatrix} a & b \\ c & d \end{pmatrix} w = \begin{pmatrix} u & \bar{v} \\ v & \bar{u} \end{pmatrix},$$

$$u = \frac{a + d + (b + c\epsilon)/\sqrt{\epsilon}}{2}, \qquad v = \frac{a - d + (b - c\epsilon)/\sqrt{\epsilon}}{2} \in E.$$

Since E/F is unramified with odd residue characteristic, we obviously have

$$O(\gamma, s) := \int_{H'} 1_{K_S}(h'\gamma'\bar{h}'^{-1})|\det(h')|^{-s}\eta(h')\,dh'. \tag{4.2}$$

Here $\gamma' = w^{-1}\gamma\bar{w}$ is the corresponding H'-orbit of $S_3(F)$ where $h \in H'$ acts on $S_3(F)$ by

$$h \circ \gamma = h\gamma\bar{h}^{-1}.$$

We thus can speak of matching an orbit γ' of $S_3(F)$ with an orbit δ of some unitary group.

We now give an explicit representative for each regular H'-orbit of $S_3(F)$. To compare with orbits of unitary group we will also calculate their invariants defined in Sect. 2.1. Then a regular $\gamma \in S_n(F)$ matches a regular $\delta \in U(J \oplus 1)$ if and only if their invariants are the same.

Proposition 4.1 *A complete set of representatives of regular H'-orbits $\mathbb{O}(S_3(F))_{reg}$ is given by*

$$\gamma(a, b, d) := \begin{pmatrix} a & 0 & 0 \\ b & -\bar{d} & 1 \\ c & 1 - d\bar{d} & d \end{pmatrix}, \quad c = -a\bar{b} + bd$$

such that

$$a \in E^1,\ b, d \in E, \quad (1 - d\bar{d})^2 - c\bar{c} \neq 0.$$

The invariants of $\gamma(a, b, d)$ are (cf. (2.1)):

$$(-b, a\bar{d}), \qquad (-c, bc + a(1 - d\bar{d})), \qquad d.$$

Moreover, $\gamma(a, b, d)$ matches an (unique) orbit of $\mathbb{O}(U(J_0 \oplus 1))_{reg}$ ($\mathbb{O}(U'(J_1 \oplus 1))_{reg}$ resp.) if and only if the valuation

$$v((1 - d\bar{d})^2 - c\bar{c})$$

is even (odd, resp.). Here J_0, J_1 are as in the statement of the fundamental lemma (cf. (2.4)).

Proof For $s = \left(\begin{smallmatrix} A & \mathbf{b} \\ \mathbf{c} & d \end{smallmatrix}\right) \in S_3(F)_{reg}$ and $\mathbf{b} = {}^t(b_1, b_2)$, then we must have $b_1\bar{b}_1 - b_2\bar{b}_2 \neq 0$. Indeed, it is clear that (b_1, b_2) cannot be the zero vector. Suppose that $b_1\bar{b}_1 - b_2\bar{b}_2 = 0$. Then up to an element in $D^\times$, we may assume that $(b_1, b_2) = (1, 1)$. Then $s' = ws\bar{w}^{-1} \in S_3(F)$ is regular of the same form with $b = {}^t(-\sqrt{\epsilon}, 0)$. In particular $\bar{b} = -b$. Since we have $A\bar{b} + \bar{d}b = 0$, the vector b and $Ab = -A\bar{b} = \bar{d}b$ are indeed linearly dependent. Contradiction!

Therefore by the action of elements of $D^\times$ we may assume that $\mathbf{b} = {}^t(0, 1)$:

$$s = \begin{pmatrix} A & e_2 \\ \mathbf{c} & d \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Since $s\bar{s} = 1_3$, we have

$$\begin{cases} A\bar{e}_2 + e_2\bar{d} = 0 & \Rightarrow \quad A = \begin{pmatrix} a & 0 \\ b & -\bar{d} \end{pmatrix} \\ \mathbf{c}\bar{e}_2 + d\bar{d} = 1 & \Rightarrow \quad \mathbf{c} = (c, 1 - d\bar{d}) \\ A\bar{A} + e_2\bar{\mathbf{c}} = 1_2 & \Rightarrow \quad a\bar{a} = 1, c = -a\bar{b} + bd. \end{cases}$$

To match with the orbits of the unitary group, we need to calculate the invariants of $w\gamma(a, b, d)\bar{w}^{-1}$. It is straightforward to show that its invariants are given by

$$(-b, a\bar{d}), \qquad (-c, bc + a(1 - d\bar{d})), \qquad d.$$

Note that

$$c = -a\bar{b} + bd.$$

It is regular if and only if the determinant Δ of the following is non-zero

$$\begin{pmatrix} -c & bc + a(1 - d\bar{d}) \\ bc + a(1 - d\bar{d}) & -b(bc + a(1 - d\bar{d})) + ac\bar{d} \end{pmatrix},$$

namely $-a^2((1 - d\bar{d})^2 - c\bar{c})$ is non-zero. Since E/F is unramified, a regular γ matches an (unique) orbit of $U(J_0 \oplus 1)$ ($U(J_1 \oplus 1)$, resp.) if and only if $v((1 - d\bar{d})^2 - c\bar{c})$ is even (odd, resp.). $\square$

We will only be concerned with the case where $v((1 - d\bar{d})^2 - c\bar{c})$ is odd and from now on we assume so. In particular, $(1 - d\bar{d})$ and c are nonzero with the same valuation. We will denote

$$u = \frac{\bar{c}}{1 - d\bar{d}} \in \mathcal{O}_E^\times. \tag{4.3}$$

Then we have

$$b = -au - \bar{d}\bar{u}. \tag{4.4}$$

4.2 Orbital integral

For $\gamma = \gamma(a, b, d) \in \mathbb{O}(S_3)_{reg}$, we go back to the orbital integral

$$O(\gamma, s) = \int 1_{K_S}(\bar{g}^{-1}\gamma g)\eta(g)|\det(g)|^s \, dg$$

where the integral is taken over

$$g \in H' = \left\{ \begin{pmatrix} t_1 & t_2 \\ \bar{t}_2 & \bar{t}_1 \end{pmatrix} \,\middle|\, t_1, t_2 \in E, t_1\bar{t}_1 - t_2\bar{t}_2 \neq 0 \right\}$$

and the measure is $dg = \kappa \frac{dt_1\, dt_2}{|t_1\bar{t}_1 - t_2\bar{t}_2|_F^2}$ where dt_1, dt_2 are the Haar measure on E such that the volume of $\mathcal{O}_E$ is one, and the constant

$$\kappa = \frac{1}{(1-q^{-1})(1-q^{-2})}.$$

The constant κ makes sure that the volume of the maximal compact subgroup of H' is one under our choice of measures.

It is clear that one necessary condition for the orbital integral to be non-zero is that both b, d are integral, which we assume from now on.

Let $t = \frac{t_1}{t_2}$. The orbital integral can be split into two parts

$$O(\gamma, s) = O_0(\gamma, s) + O_1(\gamma, s) \tag{4.5}$$

where O_0 is the contribution of g with $v(1 - t\bar{t}) > 0$ and O_1 is that of the rest.

Lemma 4.2 *We have*

$$O_1(\gamma, s) = \sum_{j=0}^{v(1-d\bar{d})} q^{2js}. \tag{4.6}$$

Proof It is easy to see

$$O_1(\gamma, s) = \kappa \int |t_1|_E^s \frac{d^\times t_1\, dt}{|1 - t\bar{t}|^2} + \kappa \int |t_1|_E^s \frac{d^\times t_1\, dt}{|1 - t\bar{t}|^2}$$

where the first term corresponds to the set of

$$g = \begin{pmatrix} t_1 & \\ & \bar{t}_1 \end{pmatrix} \begin{pmatrix} 1 & t \\ \bar{t} & 1 \end{pmatrix}$$

such that $v(t) \geq 0$, $v(1 - t\bar{t}) = 0$ and the second term

$$g = \begin{pmatrix} t_1 & \\ & \bar{t}_1 \end{pmatrix} \begin{pmatrix} t & 1 \\ 1 & \bar{t} \end{pmatrix}$$

such that $v(t) > 0$. We can thus forget about the quantity in the denominators.

Note that the volume of $t \in E$ with $v(1-t\bar{t}) = 0$ is $1 - \frac{q+1}{q^2}$ and the volume of $\varpi \mathcal{O}_E$ is q^{-2}. Then we have

$$O_1(\gamma, s) = \left(1 - \frac{q+1}{q^2}\right)\kappa \int |t_1|_E^s \, d^\times t_1 + q^{-2}\kappa \int |t_1|_E^s \, dt_1^\times$$

where both integrals are over all $t_1 \in E^\times$ such that

$$t_1^{-1}, ct_1, (1-d\bar{d})t_1 \in \mathcal{O}_E.$$

Note that $v(c) = v(1-d\bar{d})$. We obtain that

$$O_1(\gamma, s) = (1-q^{-1})(1-q^{-2})\kappa \sum_{j=0}^{v(1-d\bar{d})} q^{2js}.$$

□

We need to calculate O_0. Therefore we now assume that $v(1-t\bar{t}) > 0$ (hence t is a unit).

Lemma 4.3 *We have*

$$O_0(\gamma, s) = (1-q^{-2})\kappa q^{2v(1-d\bar{d})s} \sum_{j,n} (-1)^n q^{(n-2j)s} q^{2n} \alpha(j, n; \gamma), \tag{4.7}$$

where $\alpha(j, n; \gamma)$ is the volume of t subject to the conditions

$$v(1-t\bar{t}) = n, \qquad v((2at-b)^2 - (b^2 - 4a\bar{d})) \geq n, \qquad v(t+u) \geq n-j, \tag{4.8}$$

and the sum is taken over

$$0 \leq j \leq v(1-d\bar{d}), \quad 1 \leq n. \tag{4.9}$$

Proof Since the integral over t_1 is invariant under $\mathcal{O}_E^\times$, we may make the integral over t_1 into a discrete sum over $m = -v(t_1)$. Now we refresh our notation to be

$$g = \varpi^{-m}\begin{pmatrix} 1 & t \\ \bar{t} & 1 \end{pmatrix}, \qquad v(1-t\bar{t}) \geq 1.$$

The integrality conditions imposed by $\bar{g}^{-1}\gamma g \in GL_3(\mathcal{O}_E)$ are listed below. Firstly, we have

$$\bar{g}^{-1}\gamma g = \frac{1}{1-t\bar{t}}\begin{pmatrix} 1 & -\bar{t} \\ -t & 1 \end{pmatrix}\begin{pmatrix} a & 0 \\ b & -\bar{d} \end{pmatrix}\begin{pmatrix} 1 & t \\ \bar{t} & 1 \end{pmatrix}$$

$$= \frac{1}{1-t\bar{t}} \begin{pmatrix} \bar{d}\bar{t}^2 - b\bar{t} + a & -bt\bar{t} + at + \bar{d}\bar{t} \\ b - at - \bar{d}\bar{t} & -at^2 + bt - \bar{d} \end{pmatrix}.$$

Since all a, b, d are integers, it is not hard to see that the four equations are all equivalent. Hence we only need to require any one of the four, say

$$\frac{-at^2 + bt - \bar{d}}{1 - t\bar{t}} \in \mathcal{O}.$$

And the integrality for the other entries yields

$$v(1 - t\bar{t}) \le m, \qquad v((1 - d\bar{d})t + \bar{c}) \ge m, \qquad v(ct + 1 - d\bar{d}) \ge m.$$

Since $v(t) = 0$ we may replace the last one by

$$v(1 - t\bar{t}) + v(1 - d\bar{d}) \ge m.$$

Now we may write the orbital integral

$$O_0(\gamma, s) = (1 - q^{-2})\kappa \sum_{m,n} (-1)^n q^{(-n+2m)s} q^{2n} \alpha(m, n; \gamma), \tag{4.10}$$

where $\alpha(m, n; \gamma)$ is the volume of the set of t satisfying the following conditions

$$\begin{cases} v(1 - t\bar{t}) = n, v((2at - b)^2 - (b^2 - 4a\bar{d})) \ge n, \\ v\left(t + \dfrac{\bar{c}}{1 - d\bar{d}}\right) \ge m - v(1 - d\bar{d}), \\ n + v(1 - d\bar{d}) \ge m, 1 \le n. \end{cases} \tag{4.11}$$

Let $j = v(1 - d\bar{d}) + n - m \in \mathbb{Z}_{\ge 0}$. Then we have $m = v(1 - d\bar{d}) + n - j$ and the proof is complete. $\square$

Before we proceed, we first establish a useful lemma. For $\zeta \in 1 + \varpi\mathcal{O}_F$, we will denote by $\sqrt{\zeta}$ its unique square root lying in $1 + \varpi\mathcal{O}_F$, namely (note that we assume that the residue characteristic $p > 2$):

$$\sqrt{\zeta} = \sum_{i=0}^{\infty} \binom{1/2}{i} (\zeta - 1)^i.$$

Lemma 4.4 *For $\xi \in \mathcal{O}_E^\times$, we denote by $\beta(n, i; \xi)$ the volume of the domain $D(n, i; \xi)$ where $D(n, i; \xi)$ is the set of $x \in E$ satisfying*

$$v(1 - x\bar{x}) = n, \qquad v(x - \xi) \ge i.$$

If $1 \leq i \leq n$, *we have*

$$\beta(n, i; \xi) = \begin{cases} 0, & \textit{if } v(1 - \xi\bar{\xi}) < i \\ q^{-(n+i)}(1 - q^{-1}), & \textit{if } v(1 - \xi\bar{\xi}) \geq i. \end{cases}$$

Proof It is easy to see one necessary condition for the domain to be non-empty is $v(1 - \xi\bar{\xi}) \geq i$. We thus assume so. And since the volume depends only on $\xi\bar{\xi}$ we may change ξ to $\lambda = \sqrt{\xi\bar{\xi}}$. Then apply the following integral formula to the characteristic function of $D(n, i, \lambda)$: for unramified E/F and a function $f \in \mathcal{S}(E)$,

$$\int_E f(x)\,dx = (1 - q^{-1})^{-1} \int_F \int_{\mathcal{O}_E^\times} f\left(x_t \frac{y}{\bar{y}}\right) dy\,dt$$

where x_t is any element in E with norm $t \in F$. □

4.3 The case $v(b^2 - 4a\bar{d})$ is odd

Then the condition $v((2at - b)^2 - (b^2 - 4a\bar{d})) \geq n$ is equivalent with $v((2at - b)^2) \geq n$ and $n \leq v(b^2 - 4a\bar{d})$.

By Lemma 4.3, we rewrite

$$O_0(\gamma, s) = (1 - q^{-2})q^{2v(1-d\bar{d})s}\kappa \sum_{j,n} (-1)^n q^{(n-2j)s} q^{2n} \alpha(j, n) \tag{4.12}$$

where $\alpha(j, n)$ is the volume of the domain

$$\begin{cases} v(1 - t\bar{t}) = n, \qquad v\left(t - \dfrac{b}{2a}\right) \geq \left[\dfrac{n+1}{2}\right], \qquad v(t + u) \geq n - j, \\ 1 \leq n \leq v(b^2 - 4a\bar{d}), \qquad 0 \leq j \leq v(1 - d\bar{d}). \end{cases}$$

It is easy to see that

$$\alpha(j, n) = \begin{cases} \beta(n, [\frac{n+1}{2}]; \frac{b}{2a}) & \text{if } j \geq [\frac{n}{2}] \text{ and } v(au - \bar{d}\bar{u}) \geq n - j, \\ \beta(n, n - j; u) & \text{if } j \leq [\frac{n}{2}] \text{ and } v(au - \bar{d}\bar{u}) \geq [\frac{n+1}{2}]. \end{cases}$$

(Note that $\frac{b}{2a} + u = (au - \bar{d}\bar{u})/2$.)

Now we are ready to prove

Proposition 4.5 *Assume that* $v(b^2 - 4a\bar{d})$ *is odd. Then we have*

$$O'(\gamma, 0) = \sum_{j=0,\, j \equiv v(1-d\bar{d}) (\mathrm{mod}\ 2)}^{v(1-d\bar{d})} \sigma'(\gamma, j) \log q,$$

where

$$\sigma'(\gamma, j)=\begin{cases}2\sum_{i=0}^{[\frac{v(b^2-4a\bar{d})}{2}]} q^i, & \textit{if } 2j>v(b^2-4a\bar{d}),\\ 2\sum_{i=0}^{j} q^i+\frac{1}{2}(v(b^2-4a\bar{d})+1-j)e_j, & \textit{if } 2j<v(b^2-4a\bar{d}).\end{cases}$$

Here we define

$$e_0=1, \qquad e_j=q^j(1+q^{-1}), \quad j\ge 1.$$

Proof For a fixed j, by Lemma 4.4 the contribution to the sum in $O_0(\gamma, s)$

$$\begin{aligned}&\sum_{1\le n<2j}(-1)^n q^{(n-2j)s}q^{2n}(1-q^{-1})q^{-(n+[\frac{n+1}{2}])}\\ &\qquad+\sum_{n\ge 2j}(-1)^n q^{(n-2j)s}q^{2n}(1-q^{-1})q^{-(2n-j)}\\ &\quad=\sum_{1\le n<2j}(-1)^n q^{(n-2j)s}(1-q^{-1})q^{[\frac{n}{2}]}+\sum_{n\ge 2j}(-1)^n q^{(n-2j)s}(1-q^{-1})q^j.\end{aligned}$$

Here the first sum runs over n such that

$$1\le n\le \min\{j+v(au-\bar{d}\bar{u}), 2v(4-b\bar{b}), 2j-1, v(b^2-a\bar{d})\},$$

and the second term runs over n such that

$$2j\le n\le \min\{2v(au-\bar{d}\bar{u}), j+v(1-u\bar{u}), v(b^2-a\bar{d})\}.$$

Combining $O_1(\gamma, s)$ (4.2) with $O_0(\gamma, s)$ (4.12) we have

$$\begin{aligned}O(\gamma, s)=q^{2v(1-d\bar{d})s}\sum_{j=0}^{v(1-d\bar{d})}\Bigg(&\sum_{0\le n<2j}(-1)^n q^{(n-2j)s}q^{[\frac{n}{2}]}\\ &+\sum_{n\ge 2j}(-1)^n q^{(n-2j)s}q^j\Bigg),\end{aligned}\tag{4.13}$$

where in the inner sum the first term runs over

$$n\le \min\{j+v(au-\bar{d}\bar{u}), 2j-1, 2v(4-b\bar{b}), v(b^2-4a\bar{d})\}$$

and the second runs over

$$2j\le n\le \min\{2v(au-\bar{d}\bar{u}), j+v(1-u\bar{u}), v(b^2-4a\bar{d})\}.$$

Note that we must have $v(d)=0$ (otherwise, $b=-au-\bar{d}\bar{u}$ must be a unit and so is $b^2-4a\bar{d}$). This also implies that $v(b)=0$. Since $v(1-u\bar{u})$ is odd, by $b^2-4a\bar{d}=(au-\bar{d}\bar{u})^2-4a\bar{d}(1-u\bar{u})$ we have

$$v(b^2-4a\bar{d})=v(1-u\bar{u})<2v(au-\bar{d}\bar{u}).$$

By $b(4-b\bar{b})=-4au(1-d\bar{d})-\bar{b}(b^2-4a\bar{d})$, we have $v(4-b\bar{b})\geq \min\{v(1-d\bar{d}), v(b^2-a\bar{d})\}$. Then the sum over j runs from 0 to $v(1-d\bar{d})$; and for each $j\leq \frac{v(b^2-a\bar{d})}{2}$ the sum over n runs from 0 to an odd number $2j-1$ and for each $j\leq \frac{v(b^2-a\bar{d})}{2}$ the sum over n runs from 0 to an odd number $v(b^2-a\bar{d})$. Therefore for each j the contribution to the orbital integral is already zero. Hence $O(\gamma,0)=0$ and the derivative of (4.13) can be taken term-wise for each j. Now this gives us

$$O'(\gamma,0)=\sum_{j=0}^{v(1-d\bar{d})}\sigma(\gamma,j)\log q,$$

where

$$\sigma(\gamma,j)=\begin{cases}\sum_{i=0}^{[\frac{v(b^2-4a\bar{d})}{2}]} q^i, & \text{if } 2j>v(b^2-4a\bar{d}),\\ \sum_{i=0}^{j} q^i+([\frac{v(b^2-4a\bar{d})}{2}]-j)q^j, & \text{if } 2j<v(b^2-4a\bar{d}).\end{cases}$$

Then it is easy to check that this can be rewritten as the sum in the statement of the proposition. □

4.4 The case $v(b^2-4a\bar{d})$ is even

If $v(b^2-4a\bar{d})$ is even, then by $b^2-4a\bar{d}=(au-\bar{d}\bar{u})^2-4a\bar{d}(1-u\bar{u})$, $b^2-4a\bar{d}$ must be a square which will be denoted by τ^2. Then the condition $v((av-b)^2-(b^2-4a\bar{d}))\geq n$ in $O_0(\gamma,s)$ is satisfied if and only if one of the following three is satisfied

$$\begin{cases}(I) & v((av-b)^2)\geq n, v(\tau^2)\geq n,\\ (II) & v(\tau^2)<n, \qquad v(av-b-\tau)\geq n-v(\tau),\\ (III) & v(\tau^2)<n, \qquad v(av-b+\tau)\geq n-v(\tau).\end{cases}$$

Accordingly we will denote the integral as a sum of three terms O^I, O^{II} and O^{III}. For O^I, we have the same expression as in the previous case. And O^{III} can be obtained from O^{II} by exchanging τ to $-\tau$. So we only need to calculate O^{II}.

We have

$$O_0^{II}(\gamma, s) = q^{2v(a-d\bar{d})s} \sum_j q^{-2js} \sum_n q^{ns} q^{2n} \alpha(j, n; \gamma) \tag{4.14}$$

where the sum runs over

$$2v(\tau) < n, \quad 0 \le j \le v(1 - d\bar{d}),$$

and $\alpha(j, n; \gamma)$ is the volume of the domain of t such that

$$v(1 - t\bar{t}) = n, \qquad v(2at - b - \tau) \ge n - v(\tau), \qquad v(t + u) \ge n - j.$$

Similarly for O^{III}. Thus we need to determine the valuation of $4 - N(b \pm \tau)$.

Lemma 4.6 *Let*

$$\alpha = 4 - N(b + \tau), \qquad \beta = 4 - N(b - \tau)$$

and

$$\theta = 16 + 16d\bar{d} - 8b\bar{b} + 8\tau\bar{\tau}.$$

Then we have

$$\alpha\beta\theta = 16^2(1 - d\bar{d})^2(1 - u\bar{u}).$$

Proof We have

$$\alpha\beta = 16 - 8b\bar{b} - 8\tau\bar{\tau} + N(b^2 - \tau^2) = 16 + 16d\bar{d} - 8b\bar{b} - 8\tau\bar{\tau}.$$

Denote $\zeta = 4a\bar{d}/b^2$. Then we have

$$16(1 - d\bar{d}) = 16 - (b\bar{b})^2\zeta\bar{\zeta}, \qquad 16c\bar{c} = b\bar{b}(4 - b\bar{b}\zeta)(4 - b\bar{b}\bar{\zeta})$$

and

$$\begin{aligned}\alpha\beta\theta &= (16 - 8b\bar{b} + (b\bar{b})^2\zeta\bar{\zeta})^2 - 64(b\bar{b})^2(1 - \xi)(1 - \bar{\xi}) \\ &= (16^2 - 32(b\bar{b})^2\zeta\bar{\zeta} + (b\bar{b})^4(\zeta\bar{\zeta})^2) \\ &\quad - 16b\bar{b}(16 - 4b\bar{b}(\zeta + \bar{\zeta}) + (b\bar{b})^2\zeta\bar{\zeta}) \\ &= 16^2((1 - d\bar{d})^2 - c\bar{c}) = 16^2(1 - d\bar{d})^2(1 - u\bar{u}). \end{aligned}$$

□

Lemma 4.7 *Assume that $v(1 - d\bar{d}) > v(au - \bar{d}\bar{u}) = v(\tau)$ and $v(b) = 0$. And let τ be a square root of $b^2 - 4a\bar{d}$ such that $v(au - \bar{d}\bar{u} + \tau) = v(1 - u\bar{u}) - v(\tau)$. Then we have*

$$v(4-N(b+\tau)) = v(1-u\bar{u}) + v(1-d\bar{d}) - 2v(\tau),$$
$$v(4-N(b-\tau)) = v(1-d\bar{d}).$$

Proof For $x \in E$ we denote by $\mathrm{Re}\, x$ ($\mathrm{Im} x$, resp.) the real part $(x+\bar{x})/2$ (the imaginary part $(x-\bar{x})/2$, resp.). Note that $b^2 - (au - \bar{d}\bar{u})^2 = 4a\bar{d}u\bar{u}$. We obtain

$$u\bar{u}\,\mathrm{Im}\left(\frac{\tau}{b}\right)^2 = \mathrm{Im}\left(\frac{-4a\bar{d}u\bar{u}}{b^2}\right) = \mathrm{Im}\left(\frac{au-\bar{d}\bar{u}}{b}\right)^2.$$

Note that

$$\frac{au-\bar{d}\bar{u}}{b} + \frac{\bar{a}\bar{u}-du}{\bar{b}} = -\frac{2u\bar{u}(1-d\bar{d})}{b\bar{b}}.$$

Since $v(1-d\bar{d}) > v(au-\bar{d}\bar{u})$, we obtain $v(\mathrm{Im}\frac{au-\bar{d}\bar{u}}{b}) = v(au-\bar{d}\bar{u})$. Together we have

$$\mathrm{Im}\left(\frac{\tau}{b}\right)^2 = v(1-d\bar{d}) + v(\tau).$$

Now it is easy to estimate that (see also below (4.15)) $v(\frac{\tau}{b}+\frac{\bar{\tau}}{\bar{b}}) > v(\tau)$. We must have

$$v\left(\frac{\tau}{b}+\frac{\bar{\tau}}{\bar{b}}\right) = v(1-d\bar{d}), \qquad v\left(\frac{\tau}{b}-\frac{\bar{\tau}}{\bar{b}}\right) = v(\tau).$$

Since $b+\tau = -2au + (au - \bar{d}\bar{u} + \tau)$, we know that $v(4-N(b+\tau)) \geq \min\{4-4u\bar{u}, v(au-\bar{d}\bar{u}+\tau)\} = v(1-u\bar{u}) - v(\tau) > v(\tau)$ (note that $v(1-u\bar{u}) > 2v(\tau)$). Similarly $v(\beta) \geq v(\tau)$. We thus have $v(\alpha\beta) > 2v(\tau)$. Since $v(\alpha\beta-\gamma) = 2v(\tau)$, we must have $v(\gamma) = 2v(\tau)$. This in turn gives us that

$$v(\alpha\beta) = v(1-u\bar{u}) + 2v(1-d\bar{d}) - 2v(\tau).$$

Now we claim that $v(4-N(b+\tau)) > v(1-d\bar{d})$. And the lemma follows from this claim easily. In fact, since $v(\alpha-\beta) = v(\bar{b}\tau + b\bar{\tau}) = v(1-d\bar{d})$, then the claim implies that $v(\beta) = v(1-d\bar{d})$ and the valuation of α follows.

We now prove the claim. It suffices to prove

$$(4-N(b+\tau))(4+N(b-\tau)) = 16(1-d\bar{d})\left(1-\frac{b\bar{\tau}+\bar{b}\tau}{2(1-d\bar{d})}\right)$$

has valuation strictly large than $v(1-d\bar{d})$ (note that the other factor of LHS is a unit).

Note that our choice of τ implies that

$$\tau = -(au - \bar{d}\bar{u})t, \quad t = \sqrt{1 + \frac{4a\bar{d}(u\bar{u}-1)}{(au-\bar{d}\bar{u})^2}}. \tag{4.15}$$

Thus we have

$$\begin{aligned}
1 - \frac{b\bar{\tau} + \bar{b}\tau}{2(1-d\bar{d})} &= 1 + \frac{\bar{b}(au-\bar{d}\bar{u})t + b(\bar{a}\bar{u} - du)t}{2(1-d\bar{d})} - \frac{b(\bar{a}\bar{u}-du)(t-\bar{t})}{2(1-d\bar{d})} \\
&= 1 - u\bar{u}t - \frac{b(\bar{a}\bar{u}-du)(t-\bar{t})}{2(1-d\bar{d})} \\
&= (1-u\bar{u}) + u\bar{u}(1-t) - \frac{b(\bar{a}\bar{u}-du)(t-\bar{t})}{2(1-d\bar{d})}.
\end{aligned}$$

Since $v(t-1) = v(1-u\bar{u}) - 2v(\tau)$, it suffices to estimate $v(t-\bar{t})$.

$$\begin{aligned}
t - \bar{t} &= \frac{2\mathrm{Im}(1 + \frac{4a\bar{d}(u\bar{u}-1)}{(au-\bar{d}\bar{u})^2})}{t+\bar{t}} \\
&= \frac{(1-u\bar{u})(1-d\bar{d})(adu^2 - \bar{a}\bar{d}\bar{u}^2)}{(t+\bar{t})N(au-\bar{d}\bar{u})^2}.
\end{aligned}$$

Note that $2(adu^2 - \bar{a}\bar{d}\bar{u}^2) = b(\bar{a}\bar{u} - du) - \bar{b}(au - \bar{d}\bar{u})$ has valuation at least $v(\tau)$ (indeed precisely $v(\tau)$!). This implies that

$$v(t-\bar{t}) \geq v(1-u\bar{u}) + v(1-d\bar{d}) - 3v(\tau).$$

And in summary we have proved that

$$\begin{aligned}
v(4 - N(b+\tau)) &= v((4 - N(b+\tau))(4 + N(b-\tau)) \\
&\geq v(1-u\bar{u}) + v(1-d\bar{d}) - 2v(\tau).
\end{aligned}$$

This completes the proof. □

Now we are ready to prove

Proposition 4.8 *If $b^2 - 4a\bar{d} = \tau^2$ is a square, then we have*

$$O'(\gamma, 0) = \sum_{j=0}^{v(1-d\bar{d})} \sigma(\gamma, j) \log q,$$

where

$$\sigma(\gamma, j) = \begin{cases} \sum_{i=0}^{j} q^i + (\frac{v(1-u\bar{u})-1}{2} - j)q^j, & j \leq v(\tau), \\ \sum_{i=0}^{v(\tau)-1} q^i + \frac{1}{2}q^{v(\tau)} + (-1)^{j+v(\tau)+1}(\frac{v(1-u\bar{u})}{2} - v(\tau))q^{v(\tau)} \\ \quad j > v(\tau). \end{cases}$$

Or equivalently

$$O'(\gamma, 0) = \sum_{j=0, j \equiv v(1-d\bar{d}) \pmod 2}^{v(1-d\bar{d})} \sigma'(\gamma, j),$$

where for j

$$\sigma'(\gamma, j) = \begin{cases} 2\sum_{i=0}^{j-1} q^i + \frac{1}{2}(v(1-u\bar{u}) + 1 - 2j)e_j, & j \leq v(\tau), \\ 2\sum_{i=0}^{v(\tau)} q^{v(\tau)} - q^{v(\tau)}, & j > v(\tau). \end{cases}$$

(*Recall* $e_0 = 1, e_j = q^j(1+q^{-1}), j \geq 1$.)

Proof We first assume that $v(b) = v(d) = 0$. Then we have

$$v(b^2 - 4a\bar{d}) = 2v(au - \bar{d}\bar{u}) < v(1 - u\bar{u}).$$

Since $v(au - \bar{d}\bar{u}) \leq v(1 - d\bar{d})$, by $c\bar{b} = -a(\bar{b}^2 - 4a\bar{d}) + (b\bar{b} - 4)d$ we see that

$$v(4 - b\bar{b}) \geq \min\{v(c), 2v(\tau)\} \geq v(\tau).$$

We still choose τ such that $v(au - \bar{d}\bar{u} - \tau) = v(\tau)$ and $v(au - \bar{d}\bar{u} + \tau) = v(1 - u\bar{u}) - v(\tau)$.

We will incorporate O_1 into the first part, namely $O^I(\gamma, s) = O_0^I(\gamma, s) + O_1(\gamma, s)$. Then we have

$$O^I(\gamma) = q^{-2v(1-d\bar{d})s} \sum_{j=0}^{v(1-d\bar{d})} q^{2js} \left(\sum_n (-1)^n q^{-ns+[\frac{n}{2}]} + \sum_{n'} (-1)^{n'} q^{n's+j} \right),$$

where the two inner sums run over, respectively:

$$0 \leq n \leq \min\{j + v(\tau), 2j - 1, 2v(\tau)\}, \quad 2j \leq n' \leq 2v(\tau).$$

And we have

$$O^{II}(\gamma, s) = q^{-2v(c)s} \sum_{j=0}^{v(1-d\bar{d})} q^{2js} \sum_n (-1)^n q^{-ns+v(\tau)},$$

where the inner sum runs over:

$$2v(\tau) < n \le \begin{cases} \min\{j + v(1-u\bar{u}), v(\tau) + v(au - \bar{d}\bar{u} + \tau)\}, \\ \quad \text{if } j \le v(\tau) \\ \min\{j + v(au - \bar{d}\bar{u} + \tau), v(\tau) + v(4 - N(b+\tau))\}, \\ \quad \text{if } j > v(\tau). \end{cases}$$

And one can switch τ to $-\tau$ to obtain O^{III}.

For a fixed j we will collect contributions from three sums and we write $O(\gamma, j, s)$. We will prove that in fact for each j the term $O(\gamma, j, 0) = 0$ so that we can take its derivative. We distinguish two cases:

(1) *The case* $j \le v(\tau)$: Then the inner sums in I are taken over $0 \le n \le 2j-1, 2j \le n' \le 2v(\tau)$, the sum in *II* is taken over $2v(\tau) < n \le v(1-u\bar{u})$ and the sum in *III* is void.
(2) *The case* $v(\tau) < j \le v(1-d\bar{d})$: This happens only if we have $v(1-d\bar{d}) > v(\tau)$. Then only the first one of the two inner sums in I is nonzero and it is taken over $0 \le n \le 2v(\tau)$. By Lemma 4.7 the sum in *II* is taken over $2v(\tau)+1 \le n \le j + v(\tau)$ and the sum in *III* is taken over $2v(\tau)+1 \le n \le j + v(1-u\bar{u}) - v(\tau)$.

In either case it is easy to see $O(\gamma, j, 0) = 0$ and straightforward computation yields the expression of $\sigma(\gamma, j)\log q := \frac{d}{ds}O(\gamma, j, s)|_{s=0}$.

We now consider the case $v(b) > 0$. By $b = -au - \bar{d}\bar{u}$, d must be a unit and so is $b^2 - 4a\bar{d} = \tau^2$. Then I contributes zero to O_0 and O^{II} will be a sum of volume of

$$\begin{cases} v(1-t\bar{t}) = n, \qquad v(2at - b - \tau) \ge n, \qquad v(t+u) \ge n - j. \\ 0 < n, \qquad 0 \le j \le v(1-d\bar{d}). \end{cases}$$

And O^{III} is obtained by changing τ to $-\tau$. Since $(au - \bar{d}\bar{u})^2 - \tau^2 = 4a\bar{d}(1-u\bar{u})$, we may assume that $v(au - \bar{d}\bar{u} + \tau) = v(1-u\bar{u}) - v(\tau) = v(1-u\bar{u})$. Note that the conclusion of Lemma 4.7 still holds. Hence the same formula as above case still holds with $v(\tau) = 0$.

We are left with the case $v(d) > 0$. Then $v(b) = 0$, and $v(\tau) = 0$, $v(1-b\bar{b}) > 0$. Choose τ such that $v(b+\tau) = 0$ and $v(b-\tau) = v(d)$. Using Lemma 4.6, we can show that $v(4 - N(b+\tau)) = v(1-u\bar{u})$ and $v(4 - N(b-\tau)) = 0$. Moreover, $v(au - \bar{d}\bar{u} - \tau) = 0$ and $v(au - \bar{d}\bar{u} + \tau) = v(d) + v(1-u\bar{u})$. These data suffice to yield the desired result. □

5 Intersection numbers: $n = 3$

In this section we calculate the intersection numbers and combining results in previous section we prove the arithmetic fundamental lemma when $n = 3$.

5.1 Orbits of $U(J_1 \oplus 1)$

We will take $J_1 = \mathrm{diag}[-\varpi, 1]$. Then we have

$$SU(J_1) = \left\{ \begin{pmatrix} x & \varpi y \\ \bar{y} & \bar{x} \end{pmatrix} \Bigm| x\bar{x} - \varpi y\bar{y} = 1 \right\}.$$

Any element in $U'(3) := U(J_1 \oplus 1)$ matching $\gamma(a, b, d)$ must have the form

$$\delta = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{c} & d \end{pmatrix} \in U(J_1 \oplus 1) = U'(3).$$

Recall that $1 - u\bar{u}$ is odd, we may fix any choice of v such that

$$u\bar{u} - \varpi v\bar{v} = 1. \tag{5.1}$$

Lemma 5.1 (1) *If $v(1 - d\bar{d})$ is even, one $\delta \in U'(3)$ that matches $\gamma(a, b, d) \in \mathbb{O}(S_3)_{reg}$ can be given by δ of the above form with*

$$A = \begin{pmatrix} au & a\varpi v \\ \bar{d}\bar{v} & \bar{d}\bar{u} \end{pmatrix}, \qquad \boldsymbol{b} = {}^t(0, \xi), \qquad \boldsymbol{c} = -\bar{\xi}(\bar{v}, \bar{u})$$

for any choice ξ and v such that $\xi\bar{\xi} = 1 - d\bar{d}$.

(2) *If $v(1 - d\bar{d})$ is odd, such a δ can be chosen as*

$$A = \begin{pmatrix} \bar{d}\bar{u} & a\bar{d}\varpi v \\ \bar{v} & au \end{pmatrix}, \qquad \mathbf{b} = {}^t(\xi, 0), \qquad \boldsymbol{c} = \frac{a\bar{\xi}}{\varpi}(u, \varpi v)$$

where $\xi\bar{\xi} = -\varpi(1 - d\bar{d})$.

Proof The condition $\delta \in U'(3)$ is equivalent with

$$\begin{cases} AJA^* + bb^* = J, \\ \mathbf{c}JA^* + d\mathbf{b}^* = 0 \quad \Leftrightarrow \quad \mathbf{c}^* = -\dfrac{1}{d}A^*J^{-1}\mathbf{b}, \\ \mathbf{c}J\mathbf{c}^* + d\bar{d} = 1 = \mathbf{b}^*J^{-1}\mathbf{b} + d\bar{d}. \end{cases}$$

If $v(1 - d\bar{d})$ is even, up to action of $SU(J)$ we may assume that $\mathbf{b} = {}^t(0, \xi)$. Then we have $\mathrm{diag}[1, \bar{d}]^{-1}A \in U(J)$. So we may further assume that

$$A = \begin{pmatrix} 1 & \\ & \bar{d} \end{pmatrix} \begin{pmatrix} a'u' & a'\varpi v \\ \bar{v} & \bar{u}' \end{pmatrix}, \quad a' \in E^1, u'\bar{u}' - \varpi v\bar{v} = 1.$$

If δ matches $\gamma(a,b,d)$, by comparing their invariants we have

$$a'u' + \bar{d}\bar{u}' = -b, \qquad a'\bar{d} = ad,$$

from which we solve for a', u'

$$a' = a, \qquad u' = u = \frac{\bar{c}}{1 - d\bar{d}}.$$

Then we further get

$$\mathbf{c} = -\bar{\xi}(\bar{v}, \bar{u}).$$

Similarly for the case $v(1 - d\bar{d})$ odd. □

5.2 Some preliminary results

We now recall some results of Gross, Keating and Kudla–Rapoport (cf. [16]).

For an integer $j \geq 0$ let us consider the order of conductor j

$$\mathcal{O}_j = \mathcal{O}_F + \varpi^j \mathcal{O}_E. \tag{5.2}$$

Let $\mathcal{E}_j$ be a quasi-canonical lift of level j of $\mathbb{E}$ which is defined over the totally ramified extension W_j of W of degree $e_j = q^j(1 + q^{-1})$ if $j \geq 1$ and $e_0 = 1$. Then for our fixed choice of a uniformizer j of D (cf. (2.6)), there exists a unique $\mathcal{O}_j$-linear isogeny $\alpha_j : \mathcal{E} = \mathcal{E}_0 \to \mathcal{E}_j$ which induces the endomorphism j^j on the special fiber $\mathbb{E}$. Moreover the set $H_{0,j} \subset \mathcal{O}_D$ of homomorphisms from $\mathcal{E}_0 \otimes \mathbb{F} = \mathbb{E}$ to $\mathcal{E}_j \otimes \mathbb{F} = \mathbb{E}$ that lift to homomorphisms from $\mathcal{E}_0$ to $\mathcal{E}_j$ is exactly $\mathrm{j}^j \mathcal{O}_E$ (cf. [16, Lemma 6.4]).

For each j, a quasi-canonical lifting $\mathcal{E}_j$ defines a divisor on $\mathcal{N}_2$ in the following manner. For $\mathcal{E}_j$ we take the exterior tensor product $\mathcal{E} \otimes \mathcal{O}_E$ so that it admits an obvious $\mathcal{O}_E$-action [16, Sect. 6]. And one can endow it with a p-principal polarization so that it defines a deformation of the unitary divisible module $\mathbb{E}$. Note that the $\mathcal{N}_2$ is indeed the universal deformation space of the unitary p-divisible $\mathcal{O}_F$-module $\mathbb{X}_2$. We thus have a morphism $\mathit{Spf}\, W_j \to \mathcal{N}_2$ which turns out to be a closed immersion [16, Lemma 6.5]. We denote its image by $\mathcal{Z}_j$.

For $y \in \mathit{Hom}_{\mathcal{O}_E}(\mathbb{E}, \mathbb{X}_2)$, let $\mathcal{Z}(y)$ be the special divisor as defined in [16, Sect. 3] (note that our $\mathbb{E}$ is their $\overline{\mathbb{Y}}$). If we write $y = (y_1, y_2)$ where $y_1 \in \mathrm{j}E$, $y_2 \in E$, then we define $v_D(y) = \min\{v_D(y_1), v_D(y_2)\}$. Then one of our key ingredients is the following result:

Proposition 5.2 ([16], Proposition 8.1) *Assume that one of y_i is zero and let $m = v_D(y)$. As a divisor on $\mathcal{N}_2$ we have*

$$\mathcal{Z}(y) = \sum_{j=0,\, j\equiv m \pmod 2}^{m} \mathcal{Z}_j.$$

Remark 11 Strictly speaking, the paper [16] only deals with the case $F = \mathbb{Q}_p$. Their proof relies on an argument of Zink [16, Proposition 8.2]. It is not surprising that this can be generalized to an arbitrary finite extension F of $\mathbb{Q}_p$ as shown by Liu [19, Proposition 6.14].

Finally we also need to recall a result essentially due to Keating on the endomorphism ring of the reduction of a quasi-canonical lifting. For a quasi-canonical lifting $\mathcal{E}_j$ of level j defined over the ring W_j with a uniformizer t, let $\psi \in \mathcal{O}_D = End(\mathcal{E}_j \otimes \mathbb{F})$ be a homomorphism of its special fiber. Let $n_j(\psi)$ be the maximal integer n such that ψ lifts to an endomorphism of $\mathcal{E}_j$ modulo t^n. We let $\ell_j(\psi)$ be the "distance" of ψ to $\mathcal{O}_j$, namely

$$\ell_j(\psi) := \max\{v_D(x+\psi) | x \in \mathcal{O}_j\}.$$

Equivalently $\ell_j(\psi)$ is the positive integer ℓ such that

$$\psi \in (\mathcal{O}_j + \mathrm{j}^{\ell}\mathcal{O}_D) \setminus (\mathcal{O}_j + \mathrm{j}^{1+\ell}\mathcal{O}_D).$$

Note that if $v_D(\psi) > 2j$, then $\ell_j(\psi) \geq 2j$ must be odd.

Proposition 5.3 Keating *For $j \geq 0$, let $\ell = \ell_j(\psi)$ be defined as above. Then we have*

$$n_j(\psi) = \begin{cases} 2\sum_{i=0}^{\ell/2} q^i - q^{\ell/2}, & \text{if } \ell \leq 2j \text{ is even}, \\ 2\sum_{i=0}^{(\ell-1)/2} q^i, & \text{if } \ell \leq 2j \text{ is odd}, \\ 2\sum_{i=0}^{j-1} q^i + \frac{1}{2}(\ell+1-2j)e_j, & \text{if } \ell \geq 2j+1. \end{cases}$$

Recall that $e_j = q^j(1+q^{-1})$ when $j \geq 1$ and $e_0 = 1$.

For a proof, one may see [22]. For $j = 0$ this was calculated by Gross [7] and used to compute local height of Heegner points in the proof of the original Gross–Zagier formula [10].

5.3 Intersection numbers

For simplicity we will denote $X = \mathcal{N}_2$ and $Y = \mathcal{N}_3$ with the diagonal embedding $X \to X \times Y$. Assume that $\delta = \delta(a, b, d) \in U'(3)$ is of the form in Lemma 5.1 and let $\phi(\delta) \in G_n$ be the element under the isomorphism in Sect. 2.2. We now calculate the intersection number $X \cdot \phi(\delta)^* X$ on $X \times Y$.

Let $\mathcal{Z}_{\phi(\delta)}$ be the sub-formal-scheme of $Y \times Y$ which represents the pairs (A, A') of unitary p-divisible $\mathcal{O}_F$-modules over a base $S \in \mathrm{Nilp}_W$ together with a homomorphism $\alpha : A \to A'$ such that $\alpha \times_S \bar{S} = \phi(\delta) \times_{\mathbb{F}} \bar{S}$. By the rigidity of quasi-isogenies, the two projections from $Y \times Y$ to Y induce isomorphisms between $\mathcal{Z}_{\phi(\delta)}$ and Y. It is then tautological that the intersection number $X \cdot \phi(\delta)^* X$ on $X \times Y$ is the same as the intersection number $(\mathcal{Z}_{\phi(\delta)}|_{X\times X}) \cdot \Delta(X)$ on $X \times X$ where $\mathcal{Z}_{\phi(\delta)}|_{X\times X}$ is the restriction of $\mathcal{Z}_{\phi(\delta)}$ to $X \times X$. Note that the underlying reduced scheme X_{red} of X is a single point and X is the universal deformation space of $\mathbb{X}_2$. We will see that the intersection is proper so that the intersection number is the length of the (artinian) subscheme representing unitary p-divisible $\mathcal{O}_F$-module V of signature $(1,1)$ such that in the following diagram δ deforms to a homomorphism "?" (compatible with the auxiliary structure)

$$\begin{tikzcd} V \times \mathcal{E} \arrow[r, "?"] & V \times \mathcal{E} \\ (V \times \mathcal{E}) \otimes \mathbb{F} \arrow[u] \arrow[r, "\phi(\delta)"] & (V \times \mathcal{E}) \otimes \mathbb{F} \arrow[u] \end{tikzcd}$$

The idea is to determine this length in two steps. Recall that $\delta = \left(\begin{smallmatrix} A & \mathbf{b} \\ \mathbf{c} & d \end{smallmatrix}\right)$ as in Lemma 5.1. Hence,

$$\phi(\delta) = \begin{pmatrix} \mathcal{J}^{-1} A \mathcal{J} & \mathcal{J}^{-1}\mathbf{b} \\ \mathbf{c}\mathcal{J} & d \end{pmatrix}.$$

Note that $d \in \mathcal{O}_E$ deforms. We only need to deform $\mathbf{c}\mathcal{J}$, $\mathcal{J}^{-1}\mathbf{b}$ and $\mathcal{J}^{-1} A \mathcal{J}$. In the first step we will find the locus on X such that the vector $\mathcal{J}^{-1}\mathbf{b}$ deforms. This will use the decomposition of the special divisor recalled in the previous subsection. Roughly speaking, the locus where $\mathcal{J}^{-1}\mathbf{b}$ deforms is a union of quasi-canonical liftings of level up to $v(1 - d\bar{d})$. In the second step we determine the locus such that $\mathcal{J}^{-1} A \mathcal{J}$ deforms. This will use the result of Keating as recalled above. When d is a unit, the vector $\mathbf{c}\mathcal{J}$ automatically deforms if both $\mathcal{J}^{-1}\mathbf{b}$ and $\mathcal{J}^{-1} A \mathcal{J}$ deform. If $v(d) > 0$, then $v(1 - d\bar{d}) = 0$. Then the first step is very simple and the theory of canonical liftings will suffice.

Obviously the restriction $\mathcal{Z}_{\phi(\delta)}|_{X\times X}$ is exactly the sub-functor representing a pair (V, V') of unitary p-divisible $\mathcal{O}_F$-modules such that in the following diagram δ deforms to a homomorphism "?" (compatible with the auxiliary structure)

$$\begin{array}{ccc} V\times\mathcal{E} & \xrightarrow{?} & V'\times\mathcal{E} \\ \uparrow & & \uparrow \\ (V\times\mathcal{E})\otimes\mathbb{F} & \xrightarrow{\phi(\delta)} & (V'\times\mathcal{E})\otimes\mathbb{F} \end{array}$$

We now let p_1, p_2 be the two projections of $X\times X$ to X. Then the image $p_{2*}(\mathcal{Z}_{\phi(\delta)}|_{X\times X})$ of $\mathcal{Z}_{\phi(\delta)}|_{X\times X}$ under the projection p_2 is indeed the sub-formal-scheme of X that represents V such that $\mathcal{J}^{-1}\mathbf{b}:\mathcal{E}\otimes\mathbb{F}\to V\otimes\mathbb{F}$ lifts (compatible with the auxiliary structure). Note that this is exactly the special divisor associated to $y=\mathcal{J}^{-1}\mathbf{b}$.

It is easy to see $v_D(y)=v(1-d\bar{d})$. By Proposition 5.2 we have an equality as divisors on X

$$p_{2*}(\mathcal{Z}_{\phi(\delta)}|_{X\times X})=\sum_{j=0,\, j\equiv v(1-d\bar{d})}^{v(1-d\bar{d})}\mathcal{Z}_j.$$

Note that this also proves that the intersection is proper as long as we see that for all such $\mathcal{Z}_j$, the locus where δ lifts is an artinian scheme. But this is obvious now since $\mathcal{Z}_j$ is of dimension one and δ does not lift to the whole $\mathcal{Z}_j$.

Now we may decompose $\mathcal{Z}_{\phi(\delta)}|_{X\times X}$ as a sum $\sum_j\widetilde{\mathcal{Z}}_j$ where $\widetilde{\mathcal{Z}}_j$ denotes the preimage under p_2 of $\mathcal{Z}_j$ in $X\times X$. And hence we have

$$O'(\delta,1_{K'})=\sum_{j=0,\, j\equiv v(1-d\bar{d})}^{v(1-d\bar{d})}\sigma'(\delta,j)\log q \tag{5.3}$$

where

$$\sigma'(\delta,j)=\widetilde{\mathcal{Z}}_j\cdot\Delta(X).$$

Proposition 5.4 *Suppose $\delta\leftrightarrow\gamma$. Then we have for each $0\le j\le v(1-d\bar{d})$ with the same parity as $v(1-d\bar{d})$:*

$$\sigma'(\delta,j)=\sigma'(\gamma,j).$$

Proof We first consider the case $v(d)=0$. Then $\mathbf{c}\mathcal{J}$ automatically deforms if both $\mathcal{J}^{-1}\mathbf{b}$ and $\mathcal{J}^{-1}A\mathcal{J}$ deform. Since $\mathcal{Z}_j\simeq Spf\, W_j$, the number $\sigma'(\delta,j)$ is the maximal integer n such that the endomorphism $\mathcal{J}^{-1}A\mathcal{J}$ lifts to the reduction of universal p-divisible $\mathcal{O}_F$-module over $\mathcal{Z}_j$ modulo t^n.

By the construction of the divisor $\mathcal{Z}_j$, the universal divisible module over $\mathcal{Z}_j$ is in fact the exterior tensor product $\mathcal{E}_j \otimes \mathcal{O}_E$. We thus have an isomorphism

$$\mathcal{E}_j \otimes \mathcal{O}_E \simeq \bar{\mathcal{E}}_j \times \mathcal{E}_j. \tag{5.4}$$

By taking an $\mathcal{O}_F$-basis $\{1, \sqrt{\epsilon}\}$ of $\mathcal{O}_E$, we may identify $\mathcal{E}_j \otimes \mathcal{O}_E$ with $\mathcal{E}_j \times \mathcal{E}_j$. Then the above isomorphism gives an isomorphism $\bar{\mathcal{E}}_j \times \mathcal{E}_j \simeq \mathcal{E}_j \times \mathcal{E}_j$ with matrix

$$w = \begin{pmatrix} \sqrt{\epsilon} & -\sqrt{\epsilon} \\ 1 & 1 \end{pmatrix}.$$

Then the matrix $\mathcal{J}^{-1} A \mathcal{J}$ transfers to $w\mathcal{J}^{-1} A \mathcal{J} w^{-1} \in End(\mathcal{E}_j^2 \otimes \mathbb{F})$. Then the number $\sigma(\delta, j)$ is the maximal integer n such that all entries of $w\mathcal{J}^{-1} A \mathcal{J} w^{-1}$ lift to endomorphism of the reduction of $\mathcal{E}_j$ modulo t^n. By Proposition 5.3 we only need to determine the minimum (denoted by $\ell_j(\delta)$) among the invariants $\ell_j(\psi)$ for all entries of $w\mathcal{J}^{-1} A \mathcal{J} w^{-1}$.

We first consider the case $v(1 - d\bar{d})$ is even. Then we have

$$A = \begin{pmatrix} au & a\varpi v \\ \bar{d}\bar{v} & \bar{d}\bar{u} \end{pmatrix}, \qquad \mathcal{J}^{-1} A \mathcal{J} = \begin{pmatrix} \bar{a}\bar{u} & -\bar{a}\bar{v}\mathfrak{j} \\ \bar{d}\bar{v}\mathfrak{j} & \bar{d}\bar{u} \end{pmatrix}$$

and

$$\begin{aligned} &w\mathcal{J}^{-1} A \mathcal{J} w^{-1} \\ &\quad = \frac{1}{2} \begin{pmatrix} \bar{a}\bar{u} + \bar{d}\bar{u} - (-\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j} & (\bar{a}\bar{u} - \bar{d}\bar{u})\sqrt{\epsilon} + (\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}\sqrt{\epsilon} \\ (\bar{a}\bar{u} - \bar{d}\bar{u})/\sqrt{\epsilon} - (\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}/\sqrt{\epsilon} & \bar{a}\bar{u} + \bar{d}\bar{u} + (-\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j} \end{pmatrix}. \end{aligned}$$

Then we want to determine the minimum of the invariants ℓ_j of all four entries. It is easy to see this minimum is the same as the minimum of the invariants of the following four entries

$$\bar{a}\bar{u} + \bar{d}\bar{u}, \qquad (-\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}, \qquad (\bar{a}\bar{u} - \bar{d}\bar{u})\sqrt{\epsilon}, \qquad (\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}\sqrt{\epsilon}.$$

For $\psi \in E\mathfrak{j}$, we clearly have $\ell_j(\psi) = v_D(\psi)$. And we also have

$$\min\{v_D((-\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}), v_D((\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}\sqrt{\epsilon})\} = v_D(\bar{v}) + 1 = v(1 - u\bar{u}).$$

So we obtain

$$\min\{\ell_j((-\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}), \ell_j((\bar{a}\bar{v} + \bar{d}\bar{v})\mathfrak{j}\sqrt{\epsilon})\} = v(1 - u\bar{u}).$$

For $\psi \in E$, we have

$$\ell_j(\psi) = \ell_j(\psi - \bar{\psi}) = \begin{cases} 2v(\psi - \bar{\psi}) & \text{if } v(\psi - \bar{\psi}) < j \\ \infty & \text{if } v(\psi - \bar{\psi}) \geq j. \end{cases}$$

Then we see that

$$\min\{\ell_j(\bar{a}\bar{u}+\bar{d}\bar{u}),\ell_j((\bar{a}\bar{u}-\bar{d}\bar{u})\sqrt{\epsilon})\}=\begin{cases}2v(au-\bar{d}\bar{u}) & \text{if } v(au-\bar{d}\bar{u})<j\\ \infty & \text{if } v(au-\bar{d}\bar{u})\geq j.\end{cases}$$

In the case $v(1-d\bar{d})$ is odd, this also holds and we omit the computation. In summary we have proved that the minimal ℓ_j-invariants of the four entries is given by

$$\ell_j(\delta)=\begin{cases}v(1-u\bar{u}) & \text{if } v(au-\bar{d}\bar{u})\geq j;\\ \min\{v(1-u\bar{u}),2v(au-\bar{d}\bar{u})\} & \text{if } v(au-\bar{d}\bar{u})<j.\end{cases}$$

Note that $v(1-u\bar{u})$ is odd. Then we have two cases:

(1) $v(1-u\bar{u})<2v(au-\bar{d}\bar{u})$. This is equivalent to the assumption that $v(b^2-4a\bar{d})$ is odd. Then we always have

$$\ell_j(\delta)=v(1-u\bar{u})=v(b^2-4a\bar{d}).$$

By Proposition 5.3 we have

$$\sigma'(\delta,j)=\begin{cases}2\sum_{i=0}^{[\frac{v(b^2-a\bar{d})}{2}]}q^i, & \text{if } 2j>v(b^2-a\bar{d}),\\ 2\sum_{i=0}^{j}q^i+\frac{1}{2}(v(b^2-4a\bar{d})+1-j)e_j, & \\ \quad \text{if } 2j<v(b^2-a\bar{d}).\end{cases}$$

(2) $v(1-u\bar{u})>2v(au-\bar{d}\bar{u})$. This is equivalent to the assumption that $b^2-4a\bar{d}=\tau^2$ is a square. And we have $v(\tau)=v(au-\bar{d}\bar{u})$. Now we have

$$\ell_j(\delta)=\begin{cases}v(1-u\bar{u}) & \text{if } v(au-\bar{d}\bar{u})\geq j;\\ 2v(au-\bar{d}\bar{u}) & \text{if } v(au-\bar{d}\bar{u})<j.\end{cases}$$

By Proposition 5.3 we obtain

$$\sigma'(\delta,j)=\begin{cases}2\sum_{i=0}^{j-1}q^i+\frac{1}{2}(v(1-u\bar{u})+1-2j)e_j, & j\leq v(au-\bar{d}\bar{u});\\ 2\sum_{i=0}^{v(\tau)}q^{v(au-\bar{d}\bar{u})}-q^{v(au-\bar{d}\bar{u})}, & j>v(au-\bar{d}\bar{u}).\end{cases}$$

Now compare the results with $\sigma'(\gamma,j)$ as in Propositions 4.5 and 4.8. This completes the proof when $v(d)=0$.

When $v(d) > 0$, we only have $j = 0$ and $\ell_0(\delta) = v(1 - u\bar{u})$. We need to deform $\mathbf{c}\mathcal{J} = -\bar{\xi}(\bar{v}, \bar{u})\mathcal{J}$ (Lemma 5.1) where $\xi \in \mathcal{O}_E^\times$ is an automorphism. Under the isomorphism matrix w (5.4), we can use the theory of canonical lifting similar to the process above. It is easy to see that in this case we have

$$\sigma'(\delta, 0) = \frac{v(1 - u\bar{u}) + 1}{2}. \qquad \square$$

Finally we define a sign $\omega'(\gamma) \in \{\pm 1\}$ for regular $\gamma \in S_3(F)$ as follows: if $\gamma = h \cdot \gamma(a, b, d) \cdot h^{-1}$ (such h is unique by the regularity of γ), then

$$\omega'(\gamma) := \eta(\det(h)). \tag{5.5}$$

Theorem 5.5 *The arithmetic fundamental lemma for $n = 3$ holds. Namely, for $\delta \leftrightarrow \gamma$ we have*

$$\omega'(\gamma) O'(\gamma, 1_{K_S}, 0) = O'(\delta, 1_{K'}).$$

Proof This follows immediately from (5.3), Proposition 5.4 and the expression from Propositions 4.5 and 4.8:

$$O'(\gamma, 1_{K_S}, 0) = \sum_{j=0,\, j\equiv v(1-d\bar{d})}^{v(1-d\bar{d})} \sigma'(\gamma, j) \log q. \qquad \square$$

Acknowledgements The author thanks Ye Tian for his inspirational lecture on relative trace formula when visiting Morningside Center of Mathematics at Beijing during summer 2008. The author also thanks Professor Benedict Gross, Herve Jacquet, Michael Rapoport and Shouwu Zhang for some useful suggestions. The author thanks the Mathematical Science Center of Tsinghua University for their hospitality where part of this paper was written. The result of this paper was reported in the AIM workshop on "*Relative Trace formula and Periods of Automorphic Forms*" in August 2009, "*Workshop on Cycles and Special Values of L-series*" at CRM at Barcelona in December 2009 and "*Pan Asia Number Theory Conference*" at Kyoto in September 2010. The author would like to thank the organizers of these workshops. The author is partially supported by NSF Grant DMS #1001631.

References

1. Beilinson, A.: Height pairing between algebraic cycles. In: Current Trends in Arithmetical Algebraic Geometry, Arcata, CA, 1985. Contemp. Math., vol. 67, pp. 1–24. Am. Math. Soc., Providence (1987)
2. Bloch, S.: Height pairings for algebraic cycles. J. Pure Appl. Algebra **34**(2–3), 119–145 (1984). Proceedings of the Luminy Conference on Algebraic K-Theory (Luminy, 1983)
3. Gan, W.T., Gross, B.H., Prasad, D.: Symplectic local root numbers, central critical L-values, and restriction problems in the representation theory of classical groups. Preprint http://www.math.harvard.edu/~gross/preprints/

4. Ginzburg, D., Jiang, D., Rallis, S.: On the nonvanishing of the central value of the Rankin–Selberg L-functions. J. Am. Math. Soc. **17**(3), 679–722 (2004)
5. Ginzburg, D., Jiang, D., Rallis, S.: On the nonvanishing of the central value of the Rankin–Selberg L-functions. II. In: Automorphic Representations, L-functions and Applications: Progress and Prospects, pp. 157–191
6. Gordon, J.: Transfer to characteristic zero—appendix to "The fundamental lemma of Jacquet–Rallis in positive characteristics" by Zhiwei Yun. arXiv:1005.0610
7. Gross, B.H.: On canonical and quasicanonical liftings. Invent. Math. **84**(2), 321–326 (1986)
8. Gross, B.H., Prasad, D.: On the decomposition of a representation of SO_n when restricted to SO_{n-1}. Can. J. Math. **44**(5), 974–1002 (1992)
9. Gross, B.H., Prasad, D.: On irreducible representations of $SO_{2n+1} \times SO_{2m}$. Can. J. Math. **46**(5), 930–950 (1994)
10. Gross, B., Zagier, D.: Heegner points and derivatives of L-series. Invent. Math. **84**(2), 225–320 (1986)
11. Harris, M., Labesse, J.-P.: Conditional base change for unitary groups. Asian J. Math. **8**(4), 653–684 (2004)
12. Ichino, A.: Trilinear forms and the central values of triple product L-functions. Duke Math. J. **145**(2), 281–307 (2008)
13. Jacquet, H., Rallis, S.: On the Gross–Prasad conjecture for unitary groups. http://www.math.columbia.edu/~hj/GrossPrasadFinal.pdf
14. Kottwitz, R.E.: Points on some Shimura varieties over finite fields. J. Am. Math. Soc. **5**(2), 373–444 (1992)
15. Kudla, S.: Central derivatives of Eisenstein series and height pairings. Ann. Math. (2) **146**(3), 545–646 (1997)
16. Kudla, S., Rapoport, M.: Special cycles on unitary Shimura varieties I. Unramified local theory. Invent. Math. **184**(3), 629–682 (2011)
17. Kudla, S., Rapoport, M.: Special cycles on unitary Shimura varieties II: global theory. arXiv:0912.3758
18. Kudla, S.S., Rapoport, M., Yang, T.: Modular Forms and Special Cycles on Shimura Curves. Annals of Mathematics Studies, vol. 161. Princeton University Press, Princeton (2006). x+373 pp
19. Y., Liu, Arithmetic theta lifting and central derivatives of L-functions for $U(1, 1)$. Preprint (2009)
20. Rallis, S., Schiffmann, G.: Multiplicity one conjectures. arXiv:0705.2168
21. Rapoport, M., Zink, Th.: Period Spaces for p-Divisible Groups. Annals of Mathematics Studies, vol. 141. Princeton University Press, Princeton (1996)
22. Vollaard, I.: Endomorphisms of quasi-canonical lifts. Astérisque **312**, 105–112 (2007)
23. Vollaard, I., Wedhorn, T.: The supersingular locus of the Shimura variety for $GU(1, n-1)$, II. Invent. Math. **3**, 591–627 (2011)
24. Waldspurger, J.: Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie. Compos. Math. **54**(2), 173–242 (1985)
25. Yuan, X., Zhang, S.-W., Zhang, W.: Gross–Zagier formula on Shimura curves. Ann. Math. Stud. (submitted)
26. Yuan, X., Zhang, S.-W., Zhang, W.: Triple product L-series and Gross–Schoen cycles. Preprint
27. Yun, Z.: The fundamental lemma of Jacquet–Rallis in positive characteristics. Duke Math. J. **156**(2), 167–228 (2011)
28. Zhang, W.: Relative trace formula and arithmetic Gross–Prasad conjecture. Preprint (2009). http://www.math.harvard.edu/~wzhang/math/RTF/AFL.pdf

29. Zhang, W.: Fourier transform and the global Gan–Gross–Prasad conjecture for unitary groups. Preprint (2011)
30. Zhang, S.: Heights of Heegner points on Shimura curves. Ann. Math. (2) **153**(1), 27–147 (2001)
31. Zhang, S.: Gross–Zagier formula for GL_2. Asian J. Math. **5**(2), 183–290 (2001)
32. Zhang, S.: Linear forms, algebraic cycles, and derivatives of L-series. Available at: www.math.columbia.edu/~szhang/papers/Preprints.htm