

HITCHIN TYPE MODULI STACKS IN AUTOMORPHIC REPRESENTATION THEORY

ZHIWEI YUN

ABSTRACT. This is an article prepared for the proceedings of the ICM 2018. In the study of automorphic representations over a function field, Hitchin moduli stack and its variants naturally appear and their geometry helps the comparison of trace formulae. We give a survey on applications of this observation to a relative fundamental lemma, the arithmetic fundamental lemma and to the higher Gross-Zagier formula.

1. INTRODUCTION

1.1. Hitchin's original construction. In an influential paper by Hitchin [10], he introduced the famous integrable system, the moduli space of Higgs bundles. Let X be a smooth proper and geometrically connected curve over a field k . Let G be a connected reductive group over k . Let \mathcal{L} be a line bundle over X . An \mathcal{L} -twisted G -Higgs bundle over X is a pair (\mathcal{E}, φ) where \mathcal{E} is a principal G -bundle over X and φ is a global section of the vector bundle $\mathrm{Ad}(\mathcal{E}) \otimes \mathcal{L}$ over X . Here, $\mathrm{Ad}(\mathcal{E})$ is the vector bundle associated to \mathcal{E} and the adjoint representation of G . The moduli stack $\mathcal{M}_{G,\mathcal{L}}$ of \mathcal{L} -twisted Higgs G -bundles over X is called the Hitchin moduli stack. Hitchin defined a map

$$(1.1) \quad f : \mathcal{M}_{G,\mathcal{L}} \rightarrow \mathcal{A}_{G,\mathcal{L}}$$

to some affine space $\mathcal{A}_{G,\mathcal{L}}$ by collecting invariants of φ such as its trace and determinant in the case $G = \mathrm{GL}_n$. The map f is called the *Hitchin fibration*. When $\mathcal{L} = \omega_X$ is the line bundle of 1-forms on X , Hitchin showed that f exhibited the stable part of \mathcal{M}_{G,ω_X} as a completely integrable system. He also gave concrete descriptions of the fibers of f in terms of spectral curves.

1.2. Applications in geometric representation theory. Although discovered in the context of Yang-Mills theory, Hitchin moduli stacks have subsequently played important roles in the development of geometric representation theory.

When $\mathcal{L} = \omega_X$, \mathcal{M}_{G,ω_X} is essentially the total space of the cotangent bundle of the moduli stack Bun_G of G -bundles over X . Therefore the categories of twisted D -modules on Bun_G give quantizations of \mathcal{M}_{G,ω_X} . Beilinson and Drinfeld studied such quantizations and used them to realize part of the geometric Langlands correspondence (namely when the \widehat{G} -connection comes from an oper). This can be viewed as a global analogue of the Beilinson-Bernstein localization theorem. A related construction in positive characteristic was initiated by Bezrukavnikov and Braverman [4] for GL_n and extended to any reductive G by T.-H. Chen and X. Zhu [5]. Hitchin's integrable system also plays a key role in the work of Kapustin and Witten [13] which ties geometric Langlands correspondence to quantum field theory.

Hitchin moduli stacks have also been used to construct representations of the double affine Hecke algebra, giving global analogues of Springer representations. See [24], [25] and [19].

1.3. Applications in automorphic representation theory. B.C. Ngô [16] made the fundamental observation that point-counting on Hitchin fibers is closely related to orbital integrals that appear in the study of automorphic representations. This observation, along with ingenious technical work, allowed Ngô to prove the Lie algebra version of the Fundamental Lemma conjectured by Langlands and Shelstad in the function field case, see [17].

Group versions of the Hitchin moduli stack were introduced by Ngô using Vinberg semigroups. They are directly related to the Arthur-Selberg trace formula, as we will briefly review in Section 2.1. See recent works by [18], [2] and [3] for applications of group versions of the Hitchin moduli stack.

1.4. Contents of this report. This report will focus on further applications of variants of Hitchin moduli stacks to automorphic representation theory.

In Section 2, we explain, in heuristic terms, why Hitchin-type moduli stacks naturally show up in the study of Arthur-Selberg trace formula and more generally, relative trace formulae. A relative trace formula calculates the L^2 -pairing of two distributions on the space of automorphic forms of G given by two subgroups. Such pairings, when restricted to cuspidal automorphic representations, are often related to *special values* of automorphic L -functions. In Section 2.3, we will elaborate on the relative trace formulae introduced by Jacquet and Rallis, for which the fundamental lemma was proved in [23].

In Section 3, we point out a new direction initiated in the works [29], [27] and [28]. Drinfeld introduced the moduli stack of Shtukas as an analogue of Shimura varieties for function fields, which turns out to allow richer variants than Shimura varieties. Cohomology classes of these moduli of Shtukas generalize the notion of automorphic forms. In Section 3.1 we review the basic definitions of Shtukas, and discuss the spectral decomposition for the cohomology of moduli of Shtukas. In Section 3.2, we introduce Heegner-Drinfeld cycles on the moduli of G -Shtukas coming from subgroups H of G . The relative trace in the context of Shtukas is then defined in Section 3.3 as the intersection pairing of two Heegner-Drinfeld cycles. We believe that such pairings, when restricted to the isotypical component of a cuspidal automorphic representation, are often related to *higher derivatives* of automorphic L -functions. We then explain why Hitchin-type moduli stacks continue to play a key role in the Shtuka context, and what new geometric ingredients are needed to study relative trace formulae in this setting.

In Section 3.4-3.5 we survey what has been proven in this new direction. In Section 3.4, we review [27] and [28], in which we obtain formulae relating higher derivatives of automorphic L -functions for PGL_2 and the intersection numbers of Heegner-Drinfeld cycles. Our results generalize the Gross-Zagier formula in the function field case. In Section 3.5 we discuss the analogue of the fundamental lemma in the Shtuka setting. This was originally conjectured by W.Zhang under the name *arithmetic fundamental lemma*. We state an extension of his conjecture for function fields involving higher derivatives of orbital integrals, and sketch our strategy to prove it.

Acknowledgments. I would like to thank my mentors for bringing me to the exciting crossroads of algebraic geometry, representation theory and number theory. The influences of R. MacPherson, P. Deligne, G. Lusztig, B. Gross, B.C. Ngô and R. Bezrukavnikov on my career are especially important. I also learned a lot from my collaborators, colleagues and fellow students, to whom I would like to express my deep gratitude.

2. HITCHIN MODULI STACK AND TRACE FORMULAE

Throughout this article we fix a finite field $k = \mathbb{F}_q$. Let X be a smooth, projective and geometrically connected curve over k of genus g . Let $F = k(X)$ be the function field of X . The places of F can be identified with the set $|X|$ of closed points of X . Let \mathbb{A} denote the ring of adèles of F . For $x \in |X|$, let \mathcal{O}_x be the completed local ring of X at x , and F_x (resp. k_x) be its fraction field (resp. residue field). We also use ϖ_x to denote a uniformizer of \mathcal{O}_x .

In this section we work with the classical notion of automorphic forms for groups over function fields. We shall briefly review the Arthur-Selberg trace formula and the relative trace formulae, and explain why Hitchin type moduli stacks naturally show up in the study of these trace formulae.

2.1. Arthur-Selberg trace formula. The Arthur-Selberg trace formula is an important tool in the theory of automorphic representations. For a detailed introduction to the theory over a number field we recommend Arthur's article [1]. Here we focus on the function field case. The idea that Hitchin moduli stacks give a geometric interpretation of Arthur-Selberg trace formula is due to B.C. Ngô. For more details, we refer to [16] for the Lie algebra version, and [8] for the group version.

We ignore the issue of convergence in the discussion (i.e., we pretend to be working with an anisotropic group G), but we remark that the convergence issue lies at the heart of the theory of Arthur and we are just interpreting the easy part of his theory from a geometric perspective.

2.1.1. The classical setup. Let G be a split connected reductive group over k and we view it as a group scheme over X (hence over F) by base change. Automorphic forms for G are \mathbb{C} -valued smooth functions on the coset space $G(F)\backslash G(\mathbb{A})$. Fix a Haar measure μ_G on $G(\mathbb{A})$. Let \mathcal{A} be the space of automorphic forms for G . For any smooth compactly supported function f on $G(\mathbb{A})$, it acts on \mathcal{A} by right convolution $R(f)$.

The Arthur-Selberg trace formula aims to express the trace of $R(f)$ on \mathcal{A} in two different ways: one as a sum over conjugacy classes of $G(F)$ (the geometric expansion) and the other as a sum over automorphic representations (the spectral expansion). The primitive form of the geometric expansion reads

$$(2.1) \quad \mathrm{Tr}(R(f), \mathcal{A})'' ='' \sum_{\gamma \in G(F)/\sim} J_\gamma(f)$$

where γ runs over $G(F)$ -conjugacy classes in $G(F)$, and $J_\gamma(f)$ is the orbital integral

$$J_\gamma(f) = \mathrm{vol}(G_\gamma(F)\backslash G_\gamma(\mathbb{A}), \mu_{G_\gamma}) \int_{G_\gamma(\mathbb{A})\backslash G(\mathbb{A})} f(g^{-1}\gamma g) \frac{\mu_G}{\mu_{G_\gamma}}(g)$$

where μ_{G_γ} is any Haar measure on the centralizer $G_\gamma(\mathbb{A})$ of γ . We write the equality sign in quotation marks ¹ to indicate that the convergence issue has been ignored in (2.1). We will give a geometric interpretation of the geometric expansion.

Fix a compact open subgroup $K = \prod_{x \in |X|} K_x \subset G(\mathbb{A})$ and assume $\mathrm{vol}(K, \mu_G) = 1$. Let $\mathcal{A}_K = C_c(G(F)\backslash G(\mathbb{A})/K)$ on which the Hecke algebra $C_c(K\backslash G(\mathbb{A})/K)$ acts. For $g \in G(\mathbb{A})$, there is a Hecke correspondence attached to the double coset $KgK \subset G(\mathbb{A})$

$$(2.2) \quad G(F)\backslash G(\mathbb{A})/K \xleftarrow{p_0} G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1}) \xrightarrow{q_0} G(F)\backslash G(\mathbb{A})/K$$

where p_0 is the natural projection and q_0 is induced by right multiplication by g . The action of $f = \mathbf{1}_{KgK}$ on \mathcal{A}_K is given by $\varphi \mapsto q_{0!} p_0^* \varphi$, where $q_{0!}$ means summing over the fibers of q_0 .

¹Later we will use the same notation (equal signs in quotation marks) to indicate heuristic identities.

Upon ignoring convergence issues, the trace of $R(\mathbf{1}_{KgK})$ on \mathcal{A}_K is equal to the cardinality of the restriction of $G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1})$ to the diagonal $G(F)\backslash G(\mathbb{A})/K$ via the maps (p_0, q_0) . In other words, we should form the pullback diagram of groupoids

$$(2.3) \quad \begin{array}{ccc} M_{G,KgK} & \longrightarrow & G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1}) \\ \downarrow & & \downarrow (p_0, q_0) \\ G(F)\backslash G(\mathbb{A})/K & \xrightarrow{\Delta} & G(F)\backslash G(\mathbb{A})/K \times G(F)\backslash G(\mathbb{A})/K \end{array}$$

and we have a heuristic identity

$$(2.4) \quad \mathrm{Tr}(R(\mathbf{1}_{KgK}), \mathcal{A}_K)'' ='' \#M_{G,KgK}.$$

Here $\#\mathfrak{X}$ of a groupoid \mathfrak{X} is a counting of isomorphism classes of objects in \mathfrak{X} weighted by the reciprocals of the sizes of automorphism groups.

2.1.2. Weil's interpretation. Let $K_0 = \prod_{x \in |X|} G(\mathcal{O}_x)$. It was observed by Weil that the double coset groupoid $G(F)\backslash G(\mathbb{A})/K_0$ is naturally isomorphic to the groupoid of G -bundles over X that are trivial at the generic point of X . In fact, starting from $g = (g_x) \in G(\mathbb{A})$, one assigns the G -bundle on X that is glued from the trivial bundles on the generic point $\mathrm{Spec} F$ and the formal disks $\mathrm{Spec} \mathcal{O}_x$ via the ‘‘transition matrices’’ g_x . For a compact open $K \subset K_0$, one can similarly interpret $G(F)\backslash G(\mathbb{A})/K$ as the groupoid of G -bundles with K -level structures. There is an algebraic stack $\mathrm{Bun}_{G,K}$ classifying G -bundles on X with K -level structures, and the above observation can be rephrased as a fully faithful embedding of groupoids

$$(2.5) \quad G(F)\backslash G(\mathbb{A})/K \hookrightarrow \mathrm{Bun}_{G,K}(k).$$

A priori, the groupoid $\mathrm{Bun}_{G,K}(k)$ contains also G -bundles that are not trivial at the generic point, or equivalently G' -bundles for certain inner forms G' of G . Since we assume G is split, the embedding (2.5) is in fact an equivalence.

In the same spirit, we interpret $G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1})$ as the groupoid of triples $(\mathcal{E}, \mathcal{E}', \alpha)$ where $\mathcal{E}, \mathcal{E}'$ are G -bundles with K -level structures on X , and $\alpha : \mathcal{E} \dashrightarrow \mathcal{E}'$ is a *rational* isomorphism between \mathcal{E} and \mathcal{E}' (i.e., an isomorphism of G -bundles over the generic point $\mathrm{Spec} F$) such that the relative position of \mathcal{E} and \mathcal{E}' at each closed point $x \in |X|$ is given by the double coset $K_x g_x K_x$. For example, when $G = \mathrm{GL}_n$, $K_x = \mathrm{GL}_n(\mathcal{O}_x)$ and $g_x = \mathrm{diag}(\varpi_x, 1, \dots, 1)$, then α has relative position $K_x g_x K_x$ at x if and only if α extends to a homomorphism $\alpha_x : \mathcal{E}|_{\mathrm{Spec} \mathcal{O}_x} \rightarrow \mathcal{E}'|_{\mathrm{Spec} \mathcal{O}_x}$, and that $\mathrm{coker}(\alpha_x)$ is one-dimensional over the residue field k_x . There is a moduli stack $\mathrm{Hk}_{G,KgK}$ classifying such triples $(\mathcal{E}, \mathcal{E}', \alpha)$. The above discussion can be rephrased as an equivalence of groupoids

$$(2.6) \quad \mathrm{Hk}_{G,KgK}(k) \cong G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1}).$$

Moreover, $\mathrm{Hk}_{G,KgK}$ is equipped with two maps to $\mathrm{Bun}_{G,K}$ by recording \mathcal{E} and \mathcal{E}' , which allow us to view it as a self-correspondence of $\mathrm{Bun}_{G,K}$

$$(2.7) \quad \mathrm{Bun}_{G,K} \xleftarrow{p} \mathrm{Hk}_{G,KgK} \xrightarrow{q} \mathrm{Bun}_{G,K}$$

Under the equivalences (2.5) and (2.6), the diagram (2.7) becomes the diagram (2.2) after taking k -points.

2.1.3. *Geometric interpretation of the trace.* Continuing further with Weil's observation, we can form the stack-theoretic version of (2.3), and define the stack $\mathcal{M}_{G,KgK}$ by the Cartesian diagram

$$(2.8) \quad \begin{array}{ccc} \mathcal{M}_{G,KgK} & \longrightarrow & \mathrm{Hk}_{G,KgK} \\ \downarrow & & \downarrow (p,q) \\ \mathrm{Bun}_{G,K} & \xrightarrow{\Delta} & \mathrm{Bun}_{G,K} \times \mathrm{Bun}_{G,K} \end{array}$$

so that $\mathcal{M}_{G,KgK}(k) = M_{G,KgK}$.

By the defining Cartesian diagram (2.8), $\mathcal{M}_{G,KgK}$ classifies pairs (\mathcal{E}, φ) where \mathcal{E} is a G -bundle over X with K -level structures, and $\varphi : \mathcal{E} \dashrightarrow \mathcal{E}$ is a rational *automorphism* with relative position given by KgK .

Recall the classical Hitchin moduli stack $\mathcal{M}_{G,\mathcal{L}}$ in Section 1.1. If we write $\mathcal{L} = \mathcal{O}_X(D)$ for some effective divisor D , $\mathcal{M}_{G,\mathcal{L}}$ then classifies pairs (\mathcal{E}, φ) where \mathcal{E} is a G -bundle over X and φ is a rational section of $\mathrm{Ad}(\mathcal{E})$ (an infinitesimal automorphism of \mathcal{E}) with poles controlled by D . Therefore $\mathcal{M}_{G,\mathcal{L}}$ can be viewed as a Lie algebra version of $\mathcal{M}_{G,KgK}$, and $\mathcal{M}_{G,KgK}$ is a group version of $\mathcal{M}_{G,\mathcal{L}}$.

Let \mathfrak{C}_G be the GIT quotient of G by the conjugation action of G ; $\mathfrak{C}_G(F)$ is the set of *stable conjugacy classes* in $G(F)$. There is an affine scheme $\mathcal{B}_{G,KgK}$ classifying rational maps $X \dashrightarrow \mathfrak{C}_G$ with poles controlled by KgK . The Hitchin fibration (1.1) has an analogue

$$(2.9) \quad h_G : \mathcal{M}_{G,KgK} \rightarrow \mathcal{B}_{G,KgK}.$$

Using the map h_G , the counting of $M_{G,KgK}$, hence the trace of $R(\mathbf{1}_{KgK})$, can be decomposed into a sum over certain stable conjugacy classes a

$$(2.10) \quad \mathrm{Tr}(R(\mathbf{1}_{KgK}), \mathcal{A}_K)'' ='' \sum_{a \in \mathcal{B}_{G,KgK}(k)} \#\mathcal{M}_{G,KgK}(a)(k).$$

Here $\mathcal{M}_{G,KgK}(a)$ (a stack over k) is the fiber $h_G^{-1}(a)$. To tie back to the classical story, $\#\mathcal{M}_{G,KgK}(a)(k)$ is in fact a sum of orbital integrals

$$\#\mathcal{M}_{G,KgK}(a)(k) = \sum_{\gamma \in G(F)/\sim, [\gamma]=a} J_\gamma(\mathbf{1}_{KgK})$$

over $G(F)$ -conjugacy classes γ that belong to the stable conjugacy class a .

By the Lefschetz trace formula, we can rewrite (2.10) as

$$(2.11) \quad \mathrm{Tr}(R(\mathbf{1}_{KgK}), \mathcal{A}_K)'' ='' \sum_{a \in \mathcal{B}_{G,KgK}(k)} \mathrm{Tr}(\mathrm{Frob}_a, (\mathbf{R}h_{G!}\mathbb{Q}_\ell)_a).$$

This formula relates the Arthur-Selberg trace to the direct image complex of the Hitchin fibration h_G (called the *Hitchin complex* for G). Although it is still difficult to get a closed formula for each term in (2.11), this geometric point of view can be powerful in comparing traces for two different groups G and H by relating their Hitchin bases and Hitchin complexes.

In the work of Ngô [17], where the Lie algebra version was considered, the Hitchin complex was studied in depth using tools such as perverse sheaves and the decomposition theorem. When H is an endoscopic group of G , Ngô shows that the stable part of the Hitchin complex for H appears as a direct summand of the Hitchin complex for G , from which he deduces the Langlands-Shelstad fundamental lemma for Lie algebras over function fields.

2.2. Relative trace formulae.

2.2.1. *Periods of automorphic forms.* For simplicity we assume G is semisimple. Let $H \subset G$ be a subgroup defined over F , and μ_H a Haar measure on $H(\mathbb{A})$. For a cuspidal automorphic representation π of $G(\mathbb{A})$, the linear functional on π

$$(2.12) \quad \begin{aligned} \mathcal{P}_{H,\pi}^G &: \pi \rightarrow \mathbb{C} \\ \varphi &\mapsto \int_{H(F)\backslash H(\mathbb{A})} \varphi \mu_H \end{aligned}$$

is called the H -period of π . It factors through the space of coinvariants $\pi_{H(\mathbb{A})}$.

One can also consider variants where we integrate φ against an automorphic character χ of $H(\mathbb{A})$. If π has nonzero H -period, it is called H -distinguished. Distinguished representations are used to characterize important classes of automorphic representations such as those coming by functoriality from another group. In case the local coinvariants $(\pi_x)_{H(F_x)}$ are one-dimensional for almost all places x (as in the case for many spherical subgroups of G), one expects the period $\mathcal{P}_{H,\pi}^G$ to be related to special values of L -functions of π .

2.2.2. **Example.** Let $G = \mathrm{PGL}_2$ and $H = A$ be the diagonal torus. Then by Hecke's theory, for a suitably chosen $\varphi \in \pi$, $\int_{H(F)\backslash H(\mathbb{A})} \varphi(t)|t|^{s-1/2} dt$ gives the standard L -function $L(s, \pi)$.

2.2.3. *Relative trace formulae.* Now suppose H_1, H_2 are two subgroups of G . Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ and $\tilde{\pi}$ its contragredient. We get a bilinear form

$$\mathcal{P}_{H_1,\pi}^G \otimes \mathcal{P}_{H_2,\tilde{\pi}}^G : \pi \otimes \tilde{\pi} \rightarrow \mathbb{C}$$

In case the local coinvariants $(\pi_x)_{H_i(F_x)}$ are one-dimensional for all places x and $i = 1, 2$, the $H_1(\mathbb{A}) \times H_2(\mathbb{A})$ invariant bilinear forms on $\pi \times \tilde{\pi}$ are unique up to scalar. Therefore $\mathcal{P}_{H_1,\pi}^G \otimes \mathcal{P}_{H_2,\tilde{\pi}}^G$ is a multiple of the natural pairing on $\pi \times \tilde{\pi}$ given by the Petersson inner product. This multiple is often related to special values of L -functions attached to π . For a systematic treatment of this topic, see the book by Sakellaridis and Venkatesh [20].

An important tool to study the pairing $\mathcal{P}_{H_1,\pi}^G \otimes \mathcal{P}_{H_2,\tilde{\pi}}^G$ is the relative trace formula. We have natural maps of cosets

$$(2.13) \quad H_1(F)\backslash H_1(\mathbb{A}) \xrightarrow{\varphi_1} G(F)\backslash G(\mathbb{A}) \xleftarrow{\varphi_2} H_2(F)\backslash H_2(\mathbb{A})$$

Consider the push-forward of the constant functions on $H_i(F)\backslash H_i(\mathbb{A})$ along φ_i , viewed as distributions on $G(F)\backslash G(\mathbb{A})$. Since we will only give a heuristic discussion of the relative trace formula, we will pretend that the L^2 -pairing of two distributions makes sense. The relative trace of a test function $f \in C_c^\infty(G(\mathbb{A}))$ is the L^2 -pairing

$$(2.14) \quad \mathrm{RTr}_{H_1, H_2}^G(f)'' = \langle \varphi_{1,!} \mathbf{1}_{H_1(F)\backslash H_1(\mathbb{A})}, R(f) \varphi_{2,!} \mathbf{1}_{H_2(F)\backslash H_2(\mathbb{A})} \rangle_{L^2(G(F)\backslash G(\mathbb{A}), \mu_G)}$$

A variant of this construction is to replace the constant function 1 on $H_i(F)\backslash H_i(\mathbb{A})$ by an automorphic quasi-character χ_i of $H_i(\mathbb{A})$.

The relative trace formula is an equality between a spectral expansion of $\mathrm{RTr}_{H_1, H_2}^G(f)$ into quantities related to $\mathcal{P}_{H_1,\pi}^G \otimes \mathcal{P}_{H_2,\tilde{\pi}}^G$ and a geometric expansion into a sum of orbital integrals $J_{H_1, H_2, \gamma}^G(f)$ indexed by double cosets $H_1(F)\gamma H_2(F) \subset G(F)$

$$(2.15) \quad J_{H_1, H_2, \gamma}^G(f) = \mathrm{vol}(H_\gamma(F)\backslash H_\gamma(\mathbb{A}), \mu_{H_\gamma}) \int_{H_\gamma(\mathbb{A})\backslash (H_1 \times H_2)(\mathbb{A})} f(h_1^{-1} \gamma h_2) \frac{\mu_{H_1 \times H_2}}{\mu_{H_\gamma}}(h_1, h_2)$$

where H_γ is the stabilizer of γ under the left-right translation on G by $H_1 \times H_2$.

Let $\mathfrak{C}_{H_1, H_2}^G = \text{Spec } F[G]^{H_1 \times H_2}$. We may call elements in $\mathfrak{C}_{H_1, H_2}^G(F)$ *stable orbits* of G under the action of $H_1 \times H_2$. There is a tautological map

$$\text{inv} : H_1(F) \backslash G(F) / H_2(F) \rightarrow \mathfrak{C}_{H_1, H_2}^G(F).$$

We define

$$(2.16) \quad J_{H_1, H_2}^G(a, f) = \sum_{\gamma \in H_1(F) \backslash G(F) / H_2(F), \text{inv}(\gamma) = a} J_{H_1, H_2, \gamma}^G(f).$$

Now we fix a compact open subgroup $K \subset G(\mathbb{A})$ and let $K_i = K \cap H_i(\mathbb{A})$ for $i = 1, 2$. Choose Haar measures on $H_i(\mathbb{A})$ and $G(\mathbb{A})$ so that K_i and K have volume 1. Consider the test function $f = \mathbf{1}_{KgK}$ as before. Unwinding the definition of the relative trace, we can rewrite (2.14) as

$$(2.17) \quad \text{RTr}_{H_1, H_2}^G(\mathbf{1}_{KgK})'' ='' \# M_{H_1, H_2, KgK}^G$$

where $M_{H_1, H_2, KgK}^G$ is defined by the Cartesian diagram of groupoids

$$\begin{array}{ccc} M_{H_1, H_2, KgK}^G & \longrightarrow & G(F) \backslash G(\mathbb{A}) / (K \cap gKg^{-1}) \\ \downarrow & & \downarrow (p_0, q_0) \\ H_1(F) \backslash H_1(\mathbb{A}) / K_1 \times H_2(F) \backslash H_2(\mathbb{A}) / K_2 & \xrightarrow{(\varphi_1, \varphi_2)} & G(F) \backslash G(\mathbb{A}) / K \times G(F) \backslash G(\mathbb{A}) / K \end{array}$$

2.2.4. Geometric interpretation of the relative trace. Now we give a geometric interpretation of $M_{H_1, H_2, KgK}^G$. We assume H_1 and H_2 are also obtained by base change from split reductive groups over k , which we denote by the same notation. We have maps

$$\text{Bun}_{H_1, K_1} \xrightarrow{\Phi_1} \text{Bun}_{G, K} \xleftarrow{\Phi_2} \text{Bun}_{H_2, K_2}.$$

Taking k -points of the above diagram we recover (2.13) up to modding out by the compact open subgroups K_i and K . We may form the Cartesian diagram of stacks

$$\begin{array}{ccc} \mathcal{M}_{H_1, H_2, KgK}^G & \longrightarrow & \text{Hk}_{G, KgK} \\ \downarrow & & \downarrow (p, q) \\ \text{Bun}_{H_1, K_1} \times \text{Bun}_{H_2, K_2} & \xrightarrow{(\Phi_1, \Phi_2)} & \text{Bun}_{G, K} \times \text{Bun}_{G, K} \end{array}$$

We have $\mathcal{M}_{H_1, H_2, KgK}^G(k) = M_{H_1, H_2, KgK}^G$.

2.2.5. Example. Consider the special case $G = G_1 \times G_1$ (G_1 is a semisimple group over k) and $H_1 = H_2$ is the diagonal copy of G_1 , $K_1 = K_2$, $K = K_1 \times K_2$. Taking $g = (1, g_1)$ for some $g_1 \in G_1(\mathbb{A})$, we get a canonical isomorphism between the stacks $\mathcal{M}_{\Delta(G_1), \Delta(G_1), K(1, g_1)K}^{G_1 \times G_1}$ and $\mathcal{M}_{G_1, K_1 g_1 K_1}$. In this case, the relative trace is the usual trace of $R(\mathbf{1}_{K_1 g_1 K_1})$ on the space of automorphic forms for G_1 .

The moduli stack $\mathcal{M}_{H_1, H_2, KgK}^G$ classifies $(\mathcal{E}_1, \mathcal{E}_2, \alpha)$ where \mathcal{E}_i is an H_i -bundle with K_i -structure over X for $i = 1, 2$; α is a rational isomorphism between the G -bundles induced from \mathcal{E}_1 and \mathcal{E}_2 , with relative position given by KgK .

One can construct a scheme $\mathcal{B}_{H_1, H_2, KgK}^G$ classifying rational maps $X \dashrightarrow \mathfrak{C}_{H_1, H_2}^G$ with poles controlled by KgK , so that $\mathcal{B}_{H_1, H_2, KgK}^G(k) \subset \mathfrak{C}_{H_1, H_2}^G(F)$. For $(\mathcal{E}_1, \mathcal{E}_2, \alpha) \in \mathcal{M}_{H_1, H_2, KgK}^G$, we may restrict α to the generic point of X and take its invariants as a rational map $X \dashrightarrow \mathfrak{C}_{H_1, H_2}^G$. This way we get a map of algebraic stacks

$$(2.18) \quad h_{H_1, H_2}^G : \mathcal{M}_{H_1, H_2, KgK}^G \rightarrow \mathcal{B}_{H_1, H_2, KgK}^G.$$

In the situation of Example 2.2.5, h_{H_1, H_2}^G specializes to the usual Hitchin map (2.9) for G_1 , so we may think of h_{H_1, H_2}^G as an analogue of the Hitchin map for the Hitchin-like moduli $\mathcal{M}_{H_1, H_2, KgK}^G$. Taking k -points of (2.18) we get a map

$$(2.19) \quad M_{H_1, H_2, KgK}^G \rightarrow [H_1 \backslash G / H_2](F) \xrightarrow{\text{inv}} \mathfrak{C}_{H_1, H_2}^G(F)$$

whose fiber over $a \in \mathfrak{C}_{H_1, H_2}^G(F)$ has cardinality equal to $J_{H_1, H_2}^G(a, \mathbf{1}_{KgK})$ defined in (2.16). We may thus decompose the relative trace into a sum of point-counting along the fibers of the map (2.19)

$$(2.20) \quad \begin{aligned} \text{RTr}_{H_1, H_2}^G(\mathbf{1}_{KgK}) & \quad " = " \quad \sum_{a \in \mathfrak{C}_{H_1, H_2}^G(F)} J_{H_1, H_2}^G(a, \mathbf{1}_{KgK}) \\ & = \sum_{a \in \mathcal{B}_{H_1, H_2, KgK}^G(k)} \text{Tr}(\text{Frob}_a, (\mathbf{R}h_{H_1, H_2, !}^G \mathbb{Q}_\ell)_a). \end{aligned}$$

The above formula relates the relative trace to the direct image complex $\mathbf{R}h_{H_1, H_2, !}^G \mathbb{Q}_\ell$. As in the case of the Arthur-Selberg trace formula, we may apply sheaf-theoretic tools to study this direct image complex, especially when it comes to comparing two such complexes.

2.2.6. Example. Consider the case $G = \text{PGL}_2$, and $H_1 = H_2 = A \subset G$ is the diagonal torus. Let $K_1 = K_2 = \prod_x A(\mathcal{O}_x)$ and $K = \prod_x G(\mathcal{O}_x)$. Let $D = \sum_x n_x x$ be an effective divisor. Define the function h_D on $G(\mathbb{A})$ to be the characteristic function of $\text{Mat}_2(\mathbb{O})_D = \{(g_x) | g_x \in \text{Mat}_2(\mathcal{O}_x), v_x(\det g_x) = n_x, \forall x \in |X|\}$. Then $\text{Mat}_2(\mathbb{O})_D$ is a finite union of K -double cosets Kg_iK . We will define a stack $\mathcal{M}_{A, A, D}^G$ which turns out to be the union of the $\mathcal{M}_{A, A, Kg_iK}^G$.

Consider the stack $\widetilde{\mathcal{M}}_{A, A, D}^G$ classifying the data $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2, \varphi)$ where

- $\mathcal{L}_i, \mathcal{L}'_i$ are line bundles over X , for $i = 1, 2$;
- $\varphi : \mathcal{L}_1 \oplus \mathcal{L}_2 \rightarrow \mathcal{L}'_1 \oplus \mathcal{L}'_2$ is an injective map of coherent sheaves such that $\det(\varphi)$, viewed as a section of the line bundle $\mathcal{L}_1^{-1} \otimes \mathcal{L}_2^{-1} \otimes \mathcal{L}'_1 \otimes \mathcal{L}'_2$, has divisor D .

The Picard stack Pic_X acts on $\widetilde{\mathcal{M}}_{A, A, D}^G$ by simultaneously tensoring on \mathcal{L}_i and \mathcal{L}'_i . We define

$$\mathcal{M}_{A, A, D}^G \cong \widetilde{\mathcal{M}}_{A, A, D}^G / \text{Pic}_X.$$

The bi- A -invariant regular functions on G are generated by $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \frac{bc}{ad-bc}$, therefore the space $\mathfrak{C}_{A, A}^G$ is isomorphic to \mathbb{A}^1 . Define the Hitchin base $\mathcal{B}_{A, A, D}^G$ to be the affine space $\text{H}^0(X, \mathcal{O}_X(D))$.

To define the Hitchin map in this case, we write φ above as a matrix $\begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix}$ where φ_{ij} is a section of $\mathcal{L}_j^{-1} \otimes \mathcal{L}'_i$. The determinant $\det(\varphi)$ gives an isomorphism $\mathcal{L}_1^{-1} \otimes \mathcal{L}_2^{-1} \otimes \mathcal{L}'_1 \otimes \mathcal{L}'_2 \cong \mathcal{O}_X(D)$. On the other hand, $\varphi_{12}\varphi_{21}$ gives another section of $\mathcal{L}_1^{-1} \otimes \mathcal{L}_2^{-1} \otimes \mathcal{L}'_1 \otimes \mathcal{L}'_2$. The Hitchin map

$$h_{A, A}^G : \mathcal{M}_{A, A, D}^G \rightarrow \text{H}^0(X, \mathcal{O}_X(D)) = \mathcal{B}_{A, A, D}^G$$

then sends $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2, \varphi)$ to $\varphi_{12}\varphi_{21}$, viewed as a section of $\mathcal{O}_X(D)$ via the identification given by $\det(\varphi)$.

Although $\mathcal{M}_{A, A, D}^G$ is not of finite type, it is the disjoint union of finite type substacks indexed by a subset of $\mathbb{Z}^4 / \Delta(\mathbb{Z})$. Indeed, for $\underline{d} = (d_1, d_2, d'_1, d'_2) \in \mathbb{Z}^4 / \Delta(\mathbb{Z})$ such that $d'_1 + d'_2 = d_1 + d_2 + \deg D$, the substack $\underline{d}\mathcal{M}_{A, A, D}^G$ where $\deg \mathcal{L}_i = d_i$ and $\deg \mathcal{L}'_i = d'_i$ is of finite type. We may write

$\mathrm{RTr}_{A,A}^G(h_D)$ as a formal sum of

$${}^d\mathrm{RTr}_{A,A}^G(h_D) = \#{}^d\mathcal{M}_{A,A,D}^G(k) = \mathrm{Tr}(\mathrm{Frob}, \mathbf{H}_c^*({}^d\mathcal{M}_{A,A,D}^G \otimes \bar{k}, \bar{\mathbb{Q}}_\ell)).$$

2.3. Relative fundamental lemma. In many cases we do not expect to prove closed formulae for relative traces of the form (2.14). Instead, for applications to problems on automorphic representations, it often suffices to establish an identity between relative traces for two different situations (G, H_1, H_2) and (G', H'_1, H'_2) .

2.3.1. General format of RTF comparison. In order to establish such an identity, we need the following structures or results:

- (1) There should be an isomorphism between the spaces of invariants $\mathfrak{C}_{H_1, H_2}^G \cong \mathfrak{C}_{H'_1, H'_2}^{G'}$.
- (2) (fundamental lemma) For almost all $x \in |X|$, and $a_x \in \mathfrak{C}_{H_1, H_2}^G(F_x) \cong \mathfrak{C}_{H'_1, H'_2}^{G'}(F_x)$, we should have an identity of local orbital integrals up to a transfer factor

$$J_{H_1, H_2, x}^G(a_x, \mathbf{1}_{G(\mathcal{O}_x)}) \sim J_{H'_1, H'_2, x}^{G'}(a_x, \mathbf{1}_{G'(\mathcal{O}_x)}).$$

Here $J_{H_1, H_2, x}^G(a_x, f_x)$ is the local analogue of $J_{H_1, H_2}^G(a, f)$ defined in (2.16).

- (3) (smooth matching) For any $x \in |X|$, $a_x \in \mathfrak{C}_{H_1, H_2}^G(F_x) \cong \mathfrak{C}_{H'_1, H'_2}^{G'}(F_x)$, and $f_x \in C_c^\infty(G(F_x))$, there exists $f'_x \in C_c^\infty(G'(F_x))$ such that $J_{H_1, H_2, x}^G(a_x, f_x) = J_{H'_1, H'_2, x}^{G'}(a_x, f'_x)$.

The geometric interpretation (2.20) of the relative trace gives a way to prove the fundamental lemma by comparing the direct image complexes of the Hitchin maps h_{H_1, H_2}^G and $h_{H'_1, H'_2}^{G'}$. Below we discuss one such example.

2.3.2. The relative trace formulae of Jacquet and Rallis. Jacquet and Rallis [12] proposed a relative trace formula approach to the Gan-Gross-Prasad conjecture for unitary groups. They considered two relative trace formulae, one involving general linear groups, and the other involving unitary groups. They formulated both the fundamental lemma and the smooth matching in this context as conjectures. In [23], we used the geometric interpretation sketched in Section 2.2.4 to prove the fundamental lemma conjectured by Jacquet and Rallis, in the case of function fields. In the appendix to [23], J.Gordon used model theory to deduce the mixed characteristic case from the function field case. On the other hand, W. Zhang [31] proved the smooth matching for the Jacquet-Rallis relative trace formula at non-archimedean places. Together with the fundamental lemma proved in [23], W.Zhang deduced the Gan-Gross-Prasad conjecture for unitary groups under some local restrictions.

In the next two examples, we introduce the groups involved in the two trace formulae in [12], and sketch the definition of the moduli stacks relevant to the orbital integrals. Since we proved the fundamental lemma by reducing to its Lie algebra analogue, our moduli stacks will be linearized versions of the Hitchin-type moduli stacks introduced in Section 2.2.4, which are closer to the classical Hitchin moduli stack.

2.3.3. Example. Let F'/F be a separable quadratic extension corresponding to a double cover $\nu : X' \rightarrow X$. Let $\sigma \in \mathrm{Gal}(F'/F)$ be the nontrivial involution. Consider $G = \mathrm{Res}_{F'/F} \mathrm{GL}_n \times \mathrm{Res}_{F'/F} \mathrm{GL}_{n-1}$, $H_1 = \mathrm{Res}_{F'/F} \mathrm{GL}_{n-1}$ and $H_2 = \mathrm{GL}_n \times \mathrm{GL}_{n-1}$ (over F). The embedding $H_1 \rightarrow G$ sends $h \in H_1$ to $\left(\begin{pmatrix} h & \\ & 1 \end{pmatrix}, h \right) \in G$.

The double quotient $H_1 \backslash G / H_2$ can be identified with $\mathrm{GL}_{n-1} \backslash \mathbf{S}_n$, where

$$\mathbf{S}_n = \{g \in \mathrm{Res}_{F'/F} \mathrm{GL}_n \mid \sigma(g) = g^{-1}\}$$

with GL_{n-1} acting by conjugation. The local orbital integral relevant to this relative trace formula is

$$(2.21) \quad J_{x,\gamma}^{\mathrm{GL}}(f) := \int_{\mathrm{GL}_{n-1}(F_x)} f(h^{-1}\gamma h)\eta_x(\det h)dh, \quad \gamma \in \mathrm{S}_n(F_x), f \in C_c^\infty(\mathrm{S}_n(F_x)).$$

Here η_x is the character $F_x^\times \rightarrow \{\pm 1\}$ attached to the quadratic extension F_x'/F_x .

The Lie algebra analogue of $\mathrm{GL}_{n-1}\backslash\mathrm{S}_n$ is $\mathrm{GL}_{n-1}\backslash(\mathfrak{gl}_n \otimes_F F'_-)$ where $F'_- = (F')^{\sigma=-1}$, and GL_{n-1} acts by conjugation. Let V_n be the standard representation of GL_n over F . It is more convenient to identify $\mathrm{GL}_{n-1}\backslash(\mathfrak{gl}_n \otimes_F F'_-)$ with

$$\mathrm{GL}_n \backslash (\mathrm{Hom}_F(V_n, V_n \otimes F'_-) \times (V_n \times V_n^*)^1),$$

where $(V_n \times V_n^*)^1$ consists of $(e, e^*) \in V_n \times V_n^*$ such that $e^*(e) = 1$, and GL_n is acting diagonally on all factors (conjugation on the first factor). The GIT quotient \mathfrak{C} of $\mathrm{Hom}_F(V_n, V_n \otimes F'_-) \times (V_n \times V_n^*)^1$ by GL_n is an affine space of dimension $2n - 1$. For $(\varphi, e, e^*) \in \mathrm{Hom}_F(V_n, V_n \otimes F'_-) \times (V_n \times V_n^*)^1$, we have invariants $a_i(\varphi) \in (F'_-)^{\otimes i}$ that records the i -th coefficient of the characteristic polynomial of φ ($1 \leq i \leq n$), and $b_i = e^*(\varphi^i e) \in (F'_-)^{\otimes i}$ for $1 \leq i \leq n - 1$. The invariants $(a_1, \dots, a_n, b_1, \dots, b_{n-1})$ give coordinates for \mathfrak{C} .

We introduce the following moduli stack \mathcal{M} which serves as a global avatar for the Lie algebra version of the orbital integrals appearing in this relative trace formula. Fix line bundles \mathcal{L} and \mathcal{L}' on X . Let $\mathcal{L}_- = \mathcal{L} \otimes_{\mathcal{O}_X} \mathcal{O}_{X'}^{\sigma=-1}$. The stack \mathcal{M} classifies tuples $(\mathcal{E}, \varphi, s, s^*)$ where \mathcal{E} is a vector bundle of rank n over X , $\varphi : \mathcal{E} \rightarrow \mathcal{E} \otimes \mathcal{L}_-$, $s : \mathcal{L}'^{-1} \rightarrow \mathcal{E}$ and $s^* : \mathcal{E} \rightarrow \mathcal{L}'$ are \mathcal{O}_X -linear maps of coherent sheaves. The ‘‘Hitchin base’’ \mathcal{B} in this situation is the affine space $\prod_{i=1}^n \Gamma(X, \mathcal{L}_-^{\otimes i}) \times \prod_{i=0}^{n-1} \Gamma(X, \mathcal{L}'^{\otimes 2} \otimes \mathcal{L}_-^{\otimes i})$. The Hitchin map $f : \mathcal{M} \rightarrow \mathcal{B}$ sends $(\mathcal{E}, \varphi, s, s^*)$ to the point of \mathcal{B} with coordinates $(a_1(\varphi), \dots, a_n(\varphi), b_0, \dots, b_{n-1})$, where $a_i(\varphi)$ are the coefficients of the characteristic polynomial of φ , and $b_i = s^* \circ \varphi^i \circ s : \mathcal{L}'^{-1} \rightarrow \mathcal{L}_-^{\otimes i} \otimes \mathcal{L}'$.

2.3.4. Example. Let F'/F and $\nu : X' \rightarrow X$ be as in Example 2.3.3. Let W_{n-1} be a Hermitian vector space of dimension $n - 1$ over F' . Let $W_n = W_{n-1} \oplus F'e_n$ with the Hermitian form (\cdot, \cdot) extending that on W_{n-1} and such that $W_{n-1} \perp e_n$, $(e_n, e_n) = 1$. Let U_n and U_{n-1} be the unitary groups over F attached to W_n and W_{n-1} . Consider $G' = U_n \times U_{n-1}$, and the subgroup $H'_1 = H'_2 = U_{n-1}$ diagonally embedded into G' .

The double quotient $H'_1 \backslash G' / H'_2$ can be identified with the quotient $U_{n-1} \backslash U_n$ where U_{n-1} acts by conjugation. For $x \in |X|$, the local orbital integral relevant to this relative trace formula is

$$J_{x,\delta}^{\mathrm{U}}(f) = \int_{U_{n-1}(F_x)} f(h^{-1}\delta h)dh, \quad \delta \in U_n(F_x), f \in C_c^\infty(U_n(F_x)).$$

The Lie algebra analogue of $U_{n-1} \backslash U_n$ is $U_{n-1} \backslash \mathfrak{u}_n$, where \mathfrak{u}_n , the Lie algebra of U_n , consists of skew-self-adjoint endomorphisms of W_n . As in the case of Example 2.3.3, we identify $U_{n-1} \backslash \mathfrak{u}_n$ with $U_n \backslash (\mathfrak{u}_n \times W_n^1)$ where W_n^1 is the set of vectors $e \in W_n$ such that $(e, e) = 1$. The GIT quotient of $\mathfrak{u}_n \times W_n^1$ by U_n can be identified with the space \mathfrak{C} introduced in Example 2.3.3. For $(\psi, e) \in \mathfrak{u}_n \times W_n^1$, its image in \mathfrak{C} is $(a_1(\psi), \dots, a_n(\psi), b_1, \dots, b_{n-1})$ where $a_i(\psi) \in (F'_-)^{\otimes i}$ are the coefficients of the characteristic polynomial of ψ (as an endomorphism of W_n), and $b_i = (\psi^i e, e) \in (F'_-)^{\otimes i}$, since $\sigma(\psi^i e, e) = (e, \psi^i e) = (-1)^i (\psi^i e, e)$.

We introduce a moduli stack \mathcal{N} which serves as a global avatar for the Lie algebra version of the orbital integrals appearing in this relative trace formula. Fix line bundles \mathcal{L} and \mathcal{L}' on X . The stack \mathcal{N} classifies tuples $(\mathcal{F}, h, \psi, t)$ where \mathcal{F} is a vector bundle of rank n on X' , $h : \mathcal{F} \xrightarrow{\sim} \sigma^* \mathcal{F}^\vee$ is a Hermitian form on \mathcal{F} , $\psi : \mathcal{F} \rightarrow \mathcal{F} \otimes \nu^* \mathcal{L}$ is skew-self-adjoint with respect to h and $t : \nu^* \mathcal{L}'^{-1} \rightarrow \mathcal{F}$ is an $\mathcal{O}_{X'}$ -linear map. When ν is unramified, the base \mathcal{B} introduced in Example

2.3.3 still serves as the Hitchin base for \mathcal{N} . The Hitchin map $g : \mathcal{N} \rightarrow \mathcal{B}$ sends $(\mathcal{F}, h, \psi, t)$ to $(a_1(\psi), \dots, a_n(\psi), b_0, \dots, b_{n-1})$, where $a_i(\psi)$ are the coefficients of the characteristic polynomial of ψ , and $b_i = t^\vee \circ h \circ \psi^i \circ t$ descends to an \mathcal{O}_X -linear map $\mathcal{L}'^{-1} \rightarrow \mathcal{L}'^{\otimes i} \otimes \mathcal{L}'$.

2.3.5. **Theorem** ([23]). *Let x be a place of F such that F'/F is unramified over x and the Hermitian space $W_{n,x}$ has a self-dual lattices $\Lambda_{n,x}$. Then for strongly regular semisimple elements $\gamma \in \mathrm{S}_n(F_x)$ and $\delta \in \mathrm{U}_n(F_x)$ with the same invariants in $\mathfrak{C}(F_x)$, we have*

$$J_{x,\gamma}^{\mathrm{GL}}(\mathbf{1}_{\mathrm{S}_n(\mathcal{O}_x)}) = \pm J_{x,\delta}^{\mathrm{U}}(\mathbf{1}_{\mathrm{U}(\Lambda_{n,x})})$$

for some sign depending on the invariants of γ .

The main geometric observation in [23] is that both $f : \mathcal{M} \rightarrow \mathcal{B}$ and $g : \mathcal{N} \rightarrow \mathcal{B}$ are small maps when restricted to a certain open subset of \mathcal{B} . This enables us to prove an isomorphism between the direct images complexes of f and g by checking over the generic point of \mathcal{B} . Such an isomorphism of sheaves, after passing to the Frobenius traces on stalks, implies the identity above, which was the fundamental lemma conjectured by Jacquet and Rallis.

3. HITCHIN MODULI STACK AND SHTUKAS

In this section we consider automorphic objects that arise as cohomology classes of moduli stacks of Shtukas, which are the function-field counterpart of Shimura varieties. These cohomology classes generalize the notion of automorphic forms. The periods and relative traces have their natural analogues in this more general setting. Hitchin-type moduli stacks continue to play an important role in the study of such relative trace formulae. We give a survey of our recent work [27],[28] on higher Waldspurger-Gross-Zagier formulae and [29] on the arithmetic fundamental lemma, which fit into the framework to be discussed in this section.

3.1. **Moduli of Shtukas.** In his seminal paper [6], Drinfeld introduced the moduli of elliptic modules as a function field analogue of modular curves. Later in [7], Drinfeld defined more general geometric object called Shtukas, and used them to prove the Langlands conjecture for GL_2 over function fields. Since then it became clear that the moduli stack of Shtukas should play the role of Shimura varieties for function fields, and its cohomology should realize the Langlands correspondence for global function fields. This idea was realized for GL_n by L.Lafforgue [14] who proved the full Langlands conjecture in this case. For an arbitrary reductive group G , V.Lafforgue [15] proved the automorphic to Galois direction of the Langlands conjecture using moduli stacks of Shtukas.

3.1.1. *The moduli of Shtukas.* The general definition of G -Shtukas was given by Varshavsky [21]. For simplicity of presentation we assume G is split. Again we fix an open subgroup $K \subset K_0$, and let $N \subset |X|$ be the finite set of places where $K_x \neq G(\mathcal{O}_x)$. Choosing a maximal split torus T and a Borel subgroup B containing T , we may therefore talk about dominant coweights of T with respect to B . Let $r \geq 0$ be an integer. Let $\mu = (\mu_1, \dots, \mu_r)$ be a sequence of dominant coweights of T . Recall dominant coweights of T are in bijection with relative positions of two G -bundles over the formal disk with the same generic fiber.

Let $\mathrm{Hk}_{G,K}^\mu$ be the Hecke stack classifying points $x_1, \dots, x_r \in X - N$ together with a diagram of the form

$$\mathcal{E}_0 - \frac{f_1}{\triangleright} \mathcal{E}_1 - \frac{f_2}{\triangleright} \dots - \frac{f_r}{\triangleright} \mathcal{E}_r$$

where \mathcal{E}_i are G -bundles over X with K -level structures, and $f_i : \mathcal{E}_{i-1}|_{X-x_i} \xrightarrow{\sim} \mathcal{E}_i|_{X-x_i}$ is an isomorphism compatible with the level structures whose relative position at x_i is in the closure of that given by μ_i .

A G -Shtuka of type μ with level K is the same data as those classified by $\mathrm{Hk}_{G,K}^\mu$, together with an isomorphism of G -bundles compatible with K -level structures

$$(3.1) \quad \iota : \mathcal{E}_r \xrightarrow{\sim} {}^\tau \mathcal{E}_0.$$

Here, ${}^\tau \mathcal{E}_0$ is the image of \mathcal{E}_0 under the Frobenius morphism $\mathrm{Fr} : \mathrm{Bun}_{G,K} \rightarrow \mathrm{Bun}_{G,K}$. If we are talking about an S -family of G -Shtukas for some k -scheme S , \mathcal{E}_0 is a G -torsor over $X \times S$, then ${}^\tau \mathcal{E}_0 := (\mathrm{id}_X \times \mathrm{Fr}_S)^* \mathcal{E}_0$. There is a moduli stack $\mathrm{Sht}_{G,K}^\mu$ of G -Shtuka of type μ , which fits into a Cartesian diagram

$$(3.2) \quad \begin{array}{ccc} \mathrm{Sht}_{G,K}^\mu & \longrightarrow & \mathrm{Hk}_{G,K}^\mu \\ \downarrow & & \downarrow (p_0, p_r) \\ \mathrm{Bun}_{G,K} & \xrightarrow{(\mathrm{id}, \mathrm{Fr})} & \mathrm{Bun}_{G,K} \times \mathrm{Bun}_{G,K} \end{array}$$

Let us observe the similarity with the definition of the (group version of) Hitchin stack in diagram (2.8): the main difference is that we are replacing the diagonal map of $\mathrm{Bun}_{G,K}$ by the graph of the Frobenius.

Recording only the points x_1, \dots, x_r gives a morphism

$$\pi_{G,K}^\mu : \mathrm{Sht}_{G,K}^\mu \rightarrow (X - N)^r$$

The datum μ is called *admissible* if $\sum_i \mu_i$ lies in the coroot lattice. The existence of an isomorphism (3.1) forces μ to be admissible. Therefore $\mathrm{Sht}_{G,K}^\mu$ is nonempty only when μ is admissible. When $r = 0$, $\mathrm{Sht}_{G,K}^\mu$ is the discrete stack given by the double coset $\mathrm{Bun}_{G,K}(k) = G(F) \backslash G(\mathbb{A}) / K$. For μ admissible, we have

$$d_G(\mu) := \dim \mathrm{Sht}_{G,K}^\mu = \sum_i (\langle 2\rho_G, \mu_i \rangle + 1).$$

3.1.2. Hecke symmetry. Let $g \in G(\mathbb{A})$ and let S be the finite set of $x \in |X| - N$ such that $g_x \notin G(\mathcal{O}_x)$. There is a self-correspondence $\mathrm{Sht}_{G,KgK}^\mu$ (the dependence on g is only through the double coset KgK) of $\mathrm{Sht}_{G,K}^\mu|_{(X-N-S)^r}$ such that both maps to $\mathrm{Sht}_{G,K}^\mu|_{(X-N-S)^r}$ are finite étale. It then induces an endomorphism of the direct image complex $\mathbf{R}\pi_{G,K,!}^\mu \mathrm{IC}(\mathrm{Sht}_{G,K}^\mu)|_{(X-N-S)^r}$. V.Lafforgue [15] used his construction of excursion operators to extend this endomorphism to the whole complex $\mathbf{R}\pi_{G,K,!}^\mu \mathrm{IC}(\mathrm{Sht}_{G,K}^\mu)$ over $(X - N)^r$. If we assign this endomorphism to the function $\mathbf{1}_{KgK}$, it extends by linearity to an action of the Hecke algebra $C_c(K \backslash G(\mathbb{A}) / K)$ on the complex $\mathbf{R}\pi_{G,K,!}^\mu \mathrm{IC}(\mathrm{Sht}_{G,K}^\mu)$, and hence on its geometric stalks and on its cohomology groups.

3.1.3. Intersection cohomology of $\mathrm{Sht}_{G,K}^\mu$. The singularities of the map $\pi_{G,K}^\mu$ are exactly the same as the product of the Schubert varieties $\mathrm{Gr}_{G, \leq \mu_i}$ in the affine Grassmannian Gr_G . It is expected that the complex $\mathbf{R}\pi_{G,K,!}^\mu \mathrm{IC}(\mathrm{Sht}_{G,K}^\mu)$ should realize the global Langlands correspondence for G in a way similar to the Eichler-Shimura correspondence for modular curves. The phenomenon of endoscopy makes stating a precise conjecture quite subtle, but a rough form of the expectation is a $C_c(K \backslash G(\mathbb{A}) / K)$ -equivariant decomposition over $(X - N)^r$

$$(3.3) \quad \mathbf{R}\pi_{G,K,!}^\mu \mathrm{IC}(\mathrm{Sht}_{G,K}^\mu)'' ='' \left(\bigoplus_{\pi \text{ cuspidal}} \pi^K \otimes (\boxtimes_{i=1}^r \rho_\pi^{\mu_i}[1]) \right) \bigoplus (\text{Eisenstein part}).$$

Here π runs over cuspidal automorphic representations of $G(\mathbb{A})$ such that $\pi^K \neq 0$, ρ_π is the \widehat{G} -local system on $X - N$ attached to π by the Langlands correspondence, and $\rho_\pi^{\mu_i}$ is the local system obtained by the composition

$$\pi_1(X - N, *) \xrightarrow{\rho_\pi} \widehat{G}(\overline{\mathbb{Q}}_\ell) \rightarrow \mathrm{GL}(V(\mu_i))$$

where $V(\mu_i)$ is the irreducible representation of the dual group \widehat{G} with highest weight μ_i .

One approach to prove (3.3) is to use trace formulae. One the one hand, consider the action of a Hecke operator composed with a power of Frobenius at some $x \in |X| - N$ acting on the geometric stalk at x of the left side of (3.3), which is $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K,\bar{x}}^\mu)$. The trace of this action can be calculated by the Lefschetz trace formula, and can be expressed as a sum of twisted orbital integrals. On the other hand, the trace of the same operator on the right side of (3.3) can be calculated by the Arthur-Selberg trace formula, and be expressed using orbital integrals. The identity (3.3) would then follow from an identity between the twisted orbital integrals and the usual orbital integrals that appear in both trace formulae, known as the base-change fundamental lemma.

The difficulty in implementing this strategy is that $\mathrm{Sht}_{G,K}^\mu$ is not of finite type, and both the Lefschetz trace and the Arthur-Selberg trace would be divergent. L.Lafforgue [14] treated the case $G = \mathrm{GL}_n$ and $\mu = ((1, 0, \dots, 0), (0, \dots, 0, -1))$ by difficult analysis of the compactification of truncations of $\mathrm{Sht}_{G,K}^\mu$, generalizing the work of Drinfeld on GL_2 .

3.1.4. Cohomological spectral decomposition. We discuss a weaker version of the spectral decomposition (3.3). As mentioned above, $\mathrm{Sht}_{G,K}^\mu$ is not of finite type, so its intersection cohomology is not necessarily finite-dimensional. One can present $\mathrm{Sht}_{G,K}^\mu$ as an increasing union of finite-type open substacks, but these substacks are not preserved by the Hecke correspondences. Despite all that, we expect nice finiteness properties of $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})$ as a Hecke module. More precisely, the spherical Hecke algebra $C_c(K_0^N \backslash G(\mathbb{A}^N) / K_0^N, \overline{\mathbb{Q}}_\ell)$ (superscript N means removing places in N) should act through a quotient $\overline{\mathcal{H}}^N$ (possibly depending on μ) which is a finitely generated algebra over $\overline{\mathbb{Q}}_\ell$, and that $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})$ should be a finitely generated module over $\overline{\mathcal{H}}^N$. Now viewing $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})$ as a coherent sheaf on $\overline{\mathcal{H}}^N$, we should get a canonical decomposition of it in terms of connected components of $\overline{\mathcal{H}}^N$. A coarser decomposition should take the following form

$$(3.4) \quad \mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k}) = \bigoplus_{[P]} \mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})_{[P]}$$

where $[P]$ runs over associated classes of parabolic subgroups of G . The support of $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})_{[P]}$ should be described using the analogous quotient of the Hecke algebra $\overline{\mathcal{H}}_L^N$ for the Levi factor L of P , via the Satake transform from the spherical Hecke algebra for G to the one for L . If G is semisimple, the part $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k}, \overline{\mathbb{Q}}_\ell)_{[G]}$ should be finite-dimensional over $\overline{\mathbb{Q}}_\ell$.

In the simplest nontrivial case $G = \mathrm{PGL}_2$ we have proved the coarse decomposition.

3.1.5. Theorem ([27],[28]). *For $G = \mathrm{PGL}_2$, consider the moduli of Shtukas Sht_G^r without level structures of type $\mu = (\mu_1, \dots, \mu_r)$ where each μ_i is the minuscule coweight. Then there is a decomposition of Hecke modules*

$$\mathrm{H}_c^{2r}(\mathrm{Sht}_G^r \otimes \bar{k}) = (\oplus_\chi \mathrm{H}_c^{2r}(\mathrm{Sht}_G^r \otimes \bar{k})[\chi]) \oplus \mathrm{H}_c^{2r}(\mathrm{Sht}_G^r \otimes \bar{k})_{\mathrm{Eis}}$$

where χ runs over a finite set of characters of the Hecke algebra $C_c(K_0 \backslash G(\mathbb{A}) / K_0)$, and the support of $\mathrm{H}_c^{2r}(\mathrm{Sht}_G^r \otimes \bar{k})_{\mathrm{Eis}}$ is defined by the Eisenstein ideal.

For $i \neq 2r$, $\mathrm{H}_c^i(\mathrm{Sht}_G^r \otimes \bar{k})$ is finite-dimensional.

Similar result holds for a version of Sht_G^r with Iwahori level structures.

We expect the similar techniques to work for general split G and general type μ .

Assume we have an analogue of the above theorem for G . Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ such that $\pi^K \neq 0$, then $C_c(K_0^N \backslash G(\mathbb{A}^N)/K_0^N, \overline{\mathbb{Q}}_\ell)$ acts on π^K by a character χ_π up to semisimplification. Suppose χ_π does not appear in the support of $\mathrm{IH}_c^*(\mathrm{Sht}_G^\mu \otimes \bar{k})_{[P]}$ for any proper parabolic P (in which case we say χ_π is non-Eisenstein), then the generalized eigenspace $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})[\chi_\pi]$ is a finite-dimensional direct summand of $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})$ containing the contribution of π but possibly also companions of π with the same Hecke character χ_π away from N .

3.2. Heegner-Drinfeld cycles and periods. When $r = 0$, the left side of (3.3) is simply the function space $C_c(G(F) \backslash G(\mathbb{A})/K)$ where cuspidal automorphic forms live. In general, we should think of cohomology classes in $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K,\bar{x}}^\mu)$ or $\mathrm{IH}_c^*(\mathrm{Sht}_{G,K}^\mu \otimes \bar{k})$ as generalizations of automorphic forms. We shall use this viewpoint to generalize some constructions in Section 2 from classical automorphic forms to cohomology classes of $\mathrm{Sht}_{G,K}^\mu$.

3.2.1. Heegner-Drinfeld cycles. Let $H \subset G$ be a subgroup defined over k with level group $K_H = K \cap H(\mathbb{A})$. It induces a map $\theta_{\mathrm{Bun}} : \mathrm{Bun}_{H,K_H} \rightarrow \mathrm{Bun}_{G,K}$.

Fix an integer $r \geq 0$. Let $\lambda = (\lambda_1, \dots, \lambda_r)$ be an admissible sequence of dominant coweights of H ; let $\mu = (\mu_1, \dots, \mu_r)$ be an admissible sequence of dominant coweights of G . To relate Sht_H^λ to Sht_G^μ , we need to impose more restrictions on λ and μ . For two coweights μ, μ' of G we write $\mu \leq_G \mu'$ if for some (equivalently all) choices of a Borel $B' \subset G_{\bar{k}}$ and a maximal torus $T' \subset B'$, $\mu'_{B',T'} - \mu_{B',T'}$ is a sum of positive roots. Here $\mu_{B',T'}$ (resp. $\mu'_{B',T'}$) is the unique dominant coweight of T' conjugate to μ (resp. μ').

We assume that $\lambda_i \leq_G \mu_i$ for $0 \leq i \leq r$. In this case there is a natural morphism of Hecke stacks

$$\theta_{\mathrm{Hk}} : \mathrm{Hk}_{H,K_H}^\lambda \rightarrow \mathrm{Hk}_{G,K}^\mu$$

compatible with θ_{Bun} . The Cartesian diagram (3.2) and its counterpart for Sht_H^λ then induce a map over $(X - N)^r$

$$\theta : \mathrm{Sht}_{H,K_H}^\lambda \rightarrow \mathrm{Sht}_{G,K}^\mu.$$

If θ is proper, the image of the fundamental class of $\mathrm{Sht}_{H,K_H}^\lambda$ defines an algebraic cycle $\mathrm{Sht}_{G,K}^\mu$ which we call a *Heegner-Drinfeld cycle*.

3.2.2. Example. Consider the case $G = \mathrm{PGL}_2$, and $H = T$ is a non-split torus of the form $T = (\mathrm{Res}_{F'/F} \mathbb{G}_m)/\mathbb{G}_m$ for some quadratic extension F'/F . Since T is not a constant group scheme over X , our previous discussion does not directly apply, but we can easily define what a T -Shtuka is. The quadratic extension F' is the function field of a smooth projective curve X' with a degree two map $\nu : X' \rightarrow X$. Let $\lambda = (\lambda_1, \dots, \lambda_r) \in \mathbb{Z}^r$ with $\sum_i \lambda_i = 0$, we may consider the moduli of rank one Shtukas $\mathrm{Sht}_{\mathrm{GL}_1, X'}^\lambda$ over X' of type λ . We define Sht_T^λ to be the quotient $\mathrm{Sht}_{\mathrm{GL}_1, X'}^\lambda / \mathrm{Pic}_X(k)$, where the discrete groupoid $\mathrm{Pic}_X(k)$ is acting by pulling back to X' and tensoring with rank one Shtukas. It can be shown that the projection $\mathrm{Sht}_T^\lambda \rightarrow X'^r$ is a finite étale Galois cover with Galois group $\mathrm{Pic}_{X'}(k)/\mathrm{Pic}_X(k)$. In particular, Sht_T^λ is a smooth and proper DM stack over k of dimension r .

Now let $\mu = (\mu_1, \dots, \mu_r)$ be a sequence of dominant coweights of G . Then each μ_i can be identified with an element in $\mathbb{Z}_{\geq 0}$, with the positive coroot corresponding to 1. Admissibility of μ means that $\sum_i \mu_i$ is even. The condition $\lambda_i \leq_G \mu_i$ is saying that $|\lambda_i| \leq \mu_i$ and that $\mu_i - \lambda_i$

is even. When $\lambda_i \leq_G \mu_i$ for all i , the map $\theta : \text{Sht}_T^\lambda \rightarrow \text{Sht}_G^\mu$ simply takes a rank one Shtuka $(\{\mathcal{E}_i\}; \{x'_i\})$ on X' and sends it to the direct image $(\{\nu_* \mathcal{E}_i\}; \{\nu(x'_i)\})$, which is a rank two Shtuka on X .

3.2.3. *Periods.* Fix a Haar measure μ_H on $H(\mathbb{A})$. Under a purely root-theoretic condition on λ and μ , θ^* induces a map

$$\theta^* : \text{IH}_c^{2d_H(\lambda)}(\text{Sht}_{G,K}^\mu \otimes \bar{k}) \rightarrow \text{H}_c^{2d_H(\lambda)}(\text{Sht}_{H,K_H}^\lambda \otimes \bar{k}),$$

and therefore defines a period map

$$\mathcal{P}_{H,\lambda}^{G,\mu} : \text{IH}_c^{2d_H(\lambda)}(\text{Sht}_{G,K}^\mu \otimes \bar{k}) \xrightarrow{\theta^*} \text{H}_c^{2d_H(\lambda)}(\text{Sht}_{H,K_H}^\lambda \otimes \bar{k}) \xrightarrow{\cap[\text{Sht}_{H,K_H}^\lambda] \cdot \text{vol}(K_H, \mu_H)} \overline{\mathbb{Q}}_\ell.$$

The last map above is the cap product with the fundamental class of Sht_H^λ followed by multiplication by $\text{vol}(K_H, \mu_H)$.

Now assume $\text{Sht}_{H,K_H}^\lambda$ has half the dimension of $\text{Sht}_{G,K}^\mu$,

$$(3.5) \quad d_G(\mu) = 2d_H(\lambda).$$

Let π be a cuspidal automorphic representation of $G(\mathbb{A})$. To make sense of periods on π , we assume for the moment that the contribution of π to the intersection cohomology of $\text{Sht}_{G,K}^\mu$ is as predicted in (3.3). Restricting $\mathcal{P}_{H,\lambda}^{G,\mu}$ to the π -part we get

$$\pi^K \otimes (\otimes_{i=1}^r \text{H}_c^1((X-N) \otimes \bar{k}, \rho_\pi^{\mu_i})) \xrightarrow{\theta^*} \text{H}_c^{2d_H(\lambda)}(\text{Sht}_{H,K_H}^\lambda \otimes \bar{k}) \rightarrow \overline{\mathbb{Q}}_\ell.$$

As we expect $\rho_\pi^{\mu_i}$ to be pure, the above map should factor through the pure quotient of $\text{H}_c^1((X-N) \otimes \bar{k}, \rho_\pi^{\mu_i})$, which is $\text{H}_c^1(X \otimes \bar{k}, j_{1*} \rho_\pi^{\mu_i})$ (the cohomology of the middle extension of $\rho_\pi^{\mu_i}$), and which does not change after enlarging N . Now shrinking K and passing to the direct limit, we get

$$\mathcal{P}_{H,\lambda,\pi}^{G,\mu} : \pi \otimes (\otimes_{i=1}^r \text{H}^1(X \otimes \bar{k}, j_{1*} \rho_\pi^{\mu_i})) \rightarrow \overline{\mathbb{Q}}_\ell$$

which is the analogue of the classical period (2.12). Again $\mathcal{P}_{H,\lambda,\pi}^{G,\mu}$ should factor through the coinvariants $\pi_{H(\mathbb{A})} \otimes (\cdots)$.

3.3. Shtuka version of relative trace formula.

3.3.1. *The setup.* Let H_1 and H_2 be reductive subgroups of G over k . Fix an integer $r \geq 0$. Let $\lambda = (\lambda_1, \dots, \lambda_r)$ (resp. κ and μ) be an admissible sequence of dominant coweights of H_1 (resp. H_2 and G). Assume that $\lambda_i \leq_G \mu_i$ and $\kappa_i \leq_G \mu_i$. In this case there are natural morphisms

$$\text{Sht}_{H_1}^\lambda \xrightarrow{\theta_1} \text{Sht}_G^\mu \xleftarrow{\theta_2} \text{Sht}_{H_2}^\kappa$$

Suppose

$$\dim \text{Sht}_{H_1}^\lambda = \dim \text{Sht}_{H_2}^\kappa = \frac{1}{2} \dim \text{Sht}_G^\mu.$$

With the same extra assumptions as in Section 3.2.3, we may define the periods $\mathcal{P}_{H_1,\lambda,\pi}^{G,\mu}$ and $\mathcal{P}_{H_2,\kappa,\tilde{\pi}}^{G,\mu}$ (where $\tilde{\pi}$ is the contragradient of π). We expect the tensor product

$$\mathcal{P}_{H_1,\lambda,\pi}^{G,\mu} \otimes \mathcal{P}_{H_2,\kappa,\tilde{\pi}}^{G,\mu} : \pi \otimes \tilde{\pi} \otimes (\otimes_{i=1}^r (\text{H}^1(X \otimes \bar{k}, j_{1*} \rho_\pi^{\mu_i}) \otimes \text{H}^1(X \otimes \bar{k}, j_{1*} \rho_{\tilde{\pi}}^{\mu_i}))) \rightarrow \overline{\mathbb{Q}}_\ell$$

to factor through the pairing between $\text{H}^1(X \otimes \bar{k}, j_{1*} \rho_\pi^{\mu_i})$ and $\text{H}^1(X \otimes \bar{k}, j_{1*} \rho_{\tilde{\pi}}^{\mu_i})$ given by the cup product (the local systems $\rho_\pi^{\mu_i}$ and $\rho_{\tilde{\pi}}^{\mu_i}$ are dual to each other up to a Tate twist). Assuming this, we get a pairing

$$\mathcal{P}_{H_1,\lambda,\pi}^{G,\mu} \otimes \mathcal{P}_{H_2,\kappa,\tilde{\pi}}^{G,\mu} : \pi_{H_1(\mathbb{A})} \otimes \tilde{\pi}_{H_2(\mathbb{A})} \rightarrow \overline{\mathbb{Q}}_\ell.$$

When $\dim \pi_{H_1(\mathbb{A})} = \dim \tilde{\pi}_{H_2(\mathbb{A})} = 1$, we expect the ratio between the above pairing and the Petersson inner product to be related to *derivatives* of L -functions of π , though I do not know how to formulate a precise conjecture in general.

3.3.2. *The relative trace.* One way to access $\mathcal{P}_{H_1, \lambda, \pi}^{G, \mu} \otimes \mathcal{P}_{H_2, \kappa, \tilde{\pi}}^{G, \mu}$ is to develop a relative trace formula whose spectral expansion gives these periods. Fix a compact open subgroup $K \subset G(\mathbb{A})$ and let $K_i = H_i(\mathbb{A}) \cap K$. Assume θ_i are proper, we have Heegner-Drinfeld cycles

$$\mathcal{Z}_{H_1}^\lambda = \theta_{1*}[\text{Sht}_{H_1, K_1}^\lambda], \quad \mathcal{Z}_{H_2}^\kappa = \theta_{2*}[\text{Sht}_{H_2, K_2}^\kappa]$$

both of half dimension in $\text{Sht}_{G, K}^\mu$. Intuitively we would like to form the intersection number

$$(3.6) \quad I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}(f) = \langle \mathcal{Z}_{H_1}^\lambda, f * \mathcal{Z}_{H_2}^\kappa \rangle_{\text{Sht}_{G, K}^\mu}, \quad f \in C_c(K \backslash G(\mathbb{A})/K)$$

as the ‘‘relative trace’’ of f in this context. Here $f * (-)$ denotes the action of the Hecke algebra on the Chow group of $\text{Sht}_{G, K}^\mu$, defined similarly as in Section 3.1.2. However, there are several technical issues before we can make sense of this intersection number.

- (1) $\text{Sht}_{G, K}^\mu$ may not be smooth so the intersection product of cycles may not be defined.
- (2) Suppose the intersection of $\mathcal{Z}_{H_1}^\lambda$ and $\mathcal{Z}_{H_2}^\kappa$ is defined as a 0-cycle on $\text{Sht}_{G, K}^\mu$, if we want to get a number out of this 0-cycle, we need it to be a proper cycle, i.e., it should lie in the Chow group of cycles with proper support (over k).

The first issue goes away if we assume each μ_i to be a minuscule coweight of G , which guarantees that $\text{Sht}_{G, K}^\mu$ is smooth over $(X - N)^r$. The second issue is more serious and is analogous to the divergence issue for the usual relative trace. In results that we will present later, it won't be an issue because there $\mathcal{Z}_{H_1, K_1}^\lambda$ is itself a proper cycle. In the sequel we will proceed with heuristic arguments as we did in Section 2.2, and ignore these issues.

When λ, κ, μ are all zero, the linear functional $I_{(H_1, 0), (H_2, 0)}^{(G, 0)}$ becomes the relative trace RTr_{H_1, H_2}^G defined in (2.14). Therefore the functional $I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}$ is a generalization of the relative trace.

3.3.3. *Intersection number in terms of Hitchin-like moduli stacks.* In the case of the usual relative trace for $f = \mathbf{1}_{K_0 g K_0}$, we introduced a Hitchin-like moduli stack $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$ whose point-counting is essentially the relative trace of f . We now try to do the same for $I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}$. To simplify notations we assume $K = K_0 = \prod_x G(\mathcal{O}_x)$, hence $K_i = \prod H_i(\mathcal{O}_x)$, and suppress them from the notation for Shtukas.

To calculate the intersection number (3.6), a natural starting point is to form the stack-theoretic intersection of the cycles $\mathcal{Z}_{H_1}^\lambda$ and $f * \mathcal{Z}_{H_2}^\kappa$, i.e., consider the Cartesian diagram

$$(3.7) \quad \begin{array}{ccc} \text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)} & \longrightarrow & \text{Sht}_{G, K_0 g K_0}^\mu \\ \downarrow & & \downarrow \\ \text{Sht}_{H_1}^\lambda \times \text{Sht}_{H_2}^\kappa & \xrightarrow{\theta_1 \times \theta_2} & \text{Sht}_G^\mu \times \text{Sht}_G^\mu \end{array}$$

The expected dimension of $\text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$ is zero. If $\text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$ indeed was zero-dimensional and moreover was proper over k , then $I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}(\mathbf{1}_{K_0 g K_0})$ would be equal to the length of $\text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$. However, neither the zero-dimensionality nor the properness is

true in general. Putting these issues aside, we proceed to rewrite $\mathrm{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$ in Hitchin-like terms.

Recall we have Hecke correspondences $\mathrm{Hk}_G^\mu, \mathrm{Hk}_{H_1}^\lambda$ and $\mathrm{Hk}_{H_2}^\kappa$ for $\mathrm{Bun}_G, \mathrm{Bun}_{H_1}$ and Bun_{H_2} related by the maps $\theta_{1, \mathrm{Hk}}$ and $\theta_{2, \mathrm{Hk}}$. We can also define a Hecke correspondence for $\mathrm{Hk}_{G, K_0 g K_0}$ as the moduli stack classifying $x_1, \dots, x_r \in X$ and a commutative diagram of rational isomorphisms of G -bundles over X

$$(3.8) \quad \begin{array}{ccccccc} \mathcal{E}_0 & \dashrightarrow & \mathcal{E}_1 & \dashrightarrow & \dots & \dashrightarrow & \mathcal{E}_r \\ | & & | & & & & | \\ \downarrow \varphi_0 & & \downarrow \varphi_1 & & & & \downarrow \varphi_r \\ \mathcal{E}'_0 & \dashrightarrow & \mathcal{E}'_1 & \dashrightarrow & \dots & \dashrightarrow & \mathcal{E}'_r \end{array}$$

such that

- (1) The top and bottom rows of the diagram give objects in Hk_G^μ over $(x_1, \dots, x_r) \in X^r$;
- (2) Each column of the diagram gives an object in $\mathrm{Hk}_{G, K_0 g K_0}$, i.e., the relative position of φ_i is given by $K_0 g K_0$ for $0 \leq i \leq r$.

We denote the resulting moduli stack by $\mathrm{Hk}_{\mathrm{Hk}, K_0 g K_0}^\mu$. We have maps $\mathfrak{p}, \mathfrak{q} : \mathrm{Hk}_{\mathrm{Hk}, K_0 g K_0}^\mu \rightarrow \mathrm{Hk}_G^\mu$ by taking the top and the bottom rows; we also have maps $\mathfrak{p}_i : \mathrm{Hk}_{\mathrm{Hk}, K_0 g K_0}^\mu \rightarrow \mathrm{Hk}_{G, K_0 g K_0}$ by taking the i th column. The Hecke correspondence $\mathrm{Sht}_{G, K_0 g K_0}^\mu$ in Section 3.1.2 is defined as the pullback of $(\mathfrak{p}_0, \mathfrak{p}_r) : \mathrm{Hk}_{\mathrm{Hk}, K_0 g K_0}^\mu \rightarrow \mathrm{Hk}_{G, K_0 g K_0} \times \mathrm{Hk}_{G, K_0 g K_0}$ along the graph of the Frobenius morphism for $\mathrm{Hk}_{G, K_0 g K_0}$.

We then define $\mathrm{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$ using the Cartesian diagram

$$\begin{array}{ccc} \mathrm{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa} & \longrightarrow & \mathrm{Hk}_{\mathrm{Hk}, K_0 g K_0}^\mu \\ \downarrow & & \downarrow (\mathfrak{p}, \mathfrak{q}) \\ \mathrm{Hk}_{H_1}^\lambda \times \mathrm{Hk}_{H_2}^\kappa & \xrightarrow{\theta_{1, \mathrm{Hk}} \times \theta_{2, \mathrm{Hk}}} & \mathrm{Hk}_G^\mu \times \mathrm{Hk}_G^\mu \end{array}$$

Now $\mathrm{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$ can be viewed as an r -step Hecke correspondence for $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$. Indeed, $\mathrm{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$ classifies a diagram similar to (3.8), except that \mathcal{E}_i (resp. \mathcal{E}'_i) are now induced from H_1 -bundles \mathcal{F}_i (resp. H_2 -bundles \mathcal{F}'_i), and the top row (resp. bottom row) are induced from an object in $\mathrm{Hk}_{H_1}^\lambda$ (resp. $\mathrm{Hk}_{H_2}^\kappa$). Now each column in such a diagram gives an object in $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$. Recording the i -th column gives a map $m_i : \mathrm{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa} \rightarrow \mathcal{M}_{H_1, H_2, K_0 g K_0}^G$. We claim that there is a Cartesian diagram expressing $\mathrm{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$ as the “moduli of Shtukas for $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$ with modification type (λ, κ) ”:

$$(3.9) \quad \begin{array}{ccc} \mathrm{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)} & \longrightarrow & \mathrm{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa} \\ \downarrow & & \downarrow (m_0, m_r) \\ \mathcal{M}_{H_1, H_2, K_0 g K_0}^G & \xrightarrow{(\mathrm{id}, \mathrm{Fr})} & \mathcal{M}_{H_1, H_2, K_0 g K_0}^G \times \mathcal{M}_{H_1, H_2, K_0 g K_0}^G \end{array}$$

Indeed, this diagram is obtained by unfolding each corner of (3.7) as a fiber product of an r -step Hecke correspondence with the graph of Frobenius, and re-arranging the order of taking fiber products.

Continuing with the heuristics, the intersection number $I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}(\mathbf{1}_{K_0 g K_0})$, which “is” the length of $\text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$, should also be the intersection number of $\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$ with the graph of Frobenius for $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$. In other words, we are changing the order of intersection and leaving the “Shtuka-like” construction to the very last step. It is often true that $\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$ has the same dimension as $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$, and the fundamental class of $\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$ induces an endomorphism of the cohomology of $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$ which we denote by $[\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}]$. The Lefschetz trace formula then gives the following heuristic identity

$$(3.10) \quad I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}(\mathbf{1}_{K_0 g K_0})'' ='' \text{Tr}([\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}] \circ \text{Frob}, \mathbf{H}_c^*(\mathcal{M}_{H_1, H_2, K_0 g K_0}^G \otimes \bar{k}, \mathbb{Q}_\ell)).$$

The equal sign above is in quotation marks for at least two reasons: both sides may diverge; change of the order of intersection needs to be justified.

Let $\mathcal{B}_{H_1, H_2, K_0 g K_0}^G$ be the Hitchin base and h_{H_1, H_2}^G be the Hitchin map as in (2.18). Observe that for various $0 \leq i \leq r$, the compositions $h_{H_1, H_2}^G \circ m_i : \text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa} \rightarrow \mathcal{B}_{H_1, H_2, K_0 g K_0}^G$ are all the same. On the other hand, the Frobenius of $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$ covers the Frobenius of $\mathcal{B}_{H_1, H_2, K_0 g K_0}^G$. Therefore diagram (3.9) induces a map

$$\text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)} \rightarrow \mathcal{B}_{H_1, H_2, K_0 g K_0}^G(k)$$

which simply says that $\text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}$ decomposes into a disjoint union

$$(3.11) \quad \text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)} = \coprod_{a \in \mathcal{B}_{H_1, H_2, K_0 g K_0}^G(k)} \text{Sht}_{(H_1, \lambda), (H_2, \kappa), K_0 g K_0}^{(G, \mu)}(a).$$

The action of $[\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}]$ on the cohomology of $\mathcal{M}_{H_1, H_2, K_0 g K_0}^G$ can be localized to an action on the complex $\mathbf{R}h_{H_1, H_2, !}^G \mathbb{Q}_\ell$ using the formalism of cohomological correspondences. Then we may rewrite (3.10) as

$$(3.12) \quad I_{(H_1, \lambda), (H_2, \kappa)}^{(G, \mu)}(\mathbf{1}_{K_0 g K_0})'' ='' \sum_a \text{Tr}([\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}]_a \circ \text{Frob}_a, (\mathbf{R}h_{H_1, H_2, !}^G \mathbb{Q}_\ell)_a),$$

where a runs over $\mathcal{B}_{H_1, H_2, K_0 g K_0}^G(k)$.

Comparing (3.12) with (2.20), we see the only difference is the insertion of the operator $[\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}]_a$ acting on the stalk $(\mathbf{R}h_{H_1, H_2, !}^G \mathbb{Q}_\ell)_a$.

3.3.4. Example. Suppose $G = \text{PGL}_2$ and $H_1 = H_2 = T$ as in Example 3.2.2. Let $r \geq 0$ be even. We pick $\lambda, \kappa \in \{\pm 1\}^r$ with total sum zero, and form Sht_T^λ and Sht_T^κ . Let $\mu = (\mu_1, \dots, \mu_r)$ consist of minuscule coweights of G . In this situation we have $\dim \text{Sht}_T^\lambda = \dim \text{Sht}_T^\kappa = r = \frac{1}{2} \dim \text{Sht}_G^\mu$. Below we give more explicit descriptions of $\mathcal{M}_{T, T, K_0 g K_0}^G$ and $\text{Hk}_{\mathcal{M}, K_0 g K_0}^{\lambda, \kappa}$. For simplicity, we assume that the double cover $\nu : X' \rightarrow X$ is étale.

As in Example 2.2.6, it is more convenient to work with the test function h_D rather than $\mathbf{1}_{K_0 g K_0}$, where D is an effective divisor on X . We denote the corresponding version of $\mathcal{M}_{T, T, K_0 g K_0}^G$ by $\mathcal{M}_{T, T, D}^G$. To describe $\mathcal{M}_{T, T, D}^G$, we first consider the moduli stack $\widetilde{\mathcal{M}}_{T, T, D}^G$ classifying $(\mathcal{L}, \mathcal{L}', \varphi)$ where \mathcal{L} and \mathcal{L}' are line bundles over X' , and $\varphi : \nu_* \mathcal{L} \rightarrow \nu_* \mathcal{L}'$ is an injective map of coherent sheaves such that $\det(\varphi)$ has divisor D . We then have $\mathcal{M}_{T, T, D}^G = \widetilde{\mathcal{M}}_{T, T, D}^G / \text{Pic}_X$, where Pic_X acts by pulling back to X' and simultaneously tensoring with \mathcal{L} and \mathcal{L}' .

It is more convenient to work with another description of $\mathcal{M}_{T,T,D}^G$. The map $\varphi : \nu_*\mathcal{L} \rightarrow \nu_*\mathcal{L}'$ is equivalent to the data of two maps

$$\alpha : \mathcal{L} \rightarrow \mathcal{L}', \quad \beta : \sigma^*\mathcal{L} \rightarrow \mathcal{L}'$$

where σ is the nontrivial involution of X' over X . The determinant $\det(\varphi) = \text{Nm}(\alpha) - \text{Nm}(\beta)$, as sections of $\text{Nm}(\mathcal{L})^{-1} \otimes \text{Nm}(\mathcal{L}')$. Let $\mathcal{M}_{T,T,D}^{G,\diamond} \subset \mathcal{M}_{T,T,D}^G$ be the open subset where α and β are nonzero. By recording the divisors of α and β , we may alternatively describe $\mathcal{M}_{T,T,D}^{G,\diamond}$ as the moduli of pairs (D_α, D_β) of effective divisors on X' of degree $d = \deg D$, such that there exists a rational function f on X (necessarily unique) satisfying $\text{div}(f) = \nu(D_\alpha) - \nu(D_\beta)$ and $\text{div}(1-f) = D - \nu(D_\beta)$.

The Hecke correspondence $\text{Hk}_{\mathcal{M},D}^{\lambda,\kappa}$ is the composition of r correspondences each of which is either \mathcal{H}_+ or \mathcal{H}_- depending on whether $\lambda_i = \kappa_i$ or not. Over the open subset $\mathcal{M}_{T,T,D}^{G,\diamond}$, \mathcal{H}_+ can be described as follows: it classifies triples of effective divisors $(D_\alpha, D_\beta, D'_\beta)$ on X' such that $(D_\alpha, D_\beta) \in \mathcal{M}_{T,T,D}^{G,\diamond}$, and D'_β is obtained by changing one point of D_β by its image under σ . The two maps $p_+, q_+ : \mathcal{H}_+ \rightarrow \mathcal{M}_{T,T,D}^{G,\diamond}$ send $(D_\alpha, D_\beta, D'_\beta)$ to (D_α, D_β) and (D_α, D'_β) . Similarly, over $\mathcal{M}_{T,T,D}^{G,\diamond}$, \mathcal{H}_- classifies triples of effective divisors $(D_\alpha, D'_\alpha, D_\beta)$ on X' such that D'_α is obtained by changing one point of D_α by its image under σ .

3.4. Application to L -functions. In the work [27], we considered the case $G = \text{PGL}_2$ and the moduli of Shtukas Sht_G^r without level structures, where r stands for the r -tuple $\mu = (\mu_1, \dots, \mu_r)$ consisting of minuscule coweights of G (so r is even). Let $\nu : X' \rightarrow X$ be an unramified double cover. The Heegner-Drinfeld cycle we considered was the one introduced in Example 3.2.2, i.e., Sht_T^λ for $\lambda \in \{\pm 1\}^r$. We consider the lifting of the natural map $\theta : \text{Sht}_T^\lambda \rightarrow \text{Sht}_G^r$

$$\theta' : \text{Sht}_T^\lambda \rightarrow \text{Sht}_G^r := \text{Sht}_G^r \times_{X^r} X'^r.$$

Since Sht_T^λ is proper of dimension r , the Heegner-Drinfeld cycle $\mathcal{Z}_T^\lambda := \theta'_*[\text{Sht}_T^\lambda]$ is an r -dimensional proper cycle in the $2r$ -dimensional Sht_G^r . Therefore \mathcal{Z}_T^λ defines a class $Z_T^\lambda \in \text{H}_c^{2r}(\text{Sht}_G^r \otimes \bar{k}, \mathbb{Q}_\ell)(r)$.

Now let π be an everywhere unramified cuspidal automorphic representation of $G(\mathbb{A})$ with coefficients in $\overline{\mathbb{Q}_\ell}$. By the coarse cohomological spectral decomposition for $\text{H}_c^{2r}(\text{Sht}_G^r \otimes \bar{k}, \mathbb{Q}_\ell)$ (see Theorem 3.1.5), we may project Z_T^λ to the χ_π -isotypical summand, and denote the resulting class by $Z_{T,\pi}^\lambda \in \text{H}_c^{2r}(\text{Sht}_G^r \otimes \bar{k}, \mathbb{Q}_\ell)[\chi_\pi]$.

3.4.1. Theorem ([27]). *We have*

$$\langle Z_{T,\pi}^\lambda, Z_{T,\pi}^\lambda \rangle_{\text{Sht}_G^r} = \frac{q^{2-2g}}{2(\log q)^r} \frac{\mathcal{L}^{(r)}(\pi_{F'}, 1/2)}{L(\pi, \text{Ad}, 1)}$$

where

- $\langle Z_{T,\pi}^\lambda, Z_{T,\pi}^\lambda \rangle_{\text{Sht}_G^r}$ is the self-intersection number of the cycle class $Z_{T,\pi}^\lambda$.
- $\pi_{F'}$ is the base change of π to $F' = k(X')$.
- $\mathcal{L}(\pi_{F'}, s) = q^{4(g-1)(s-1/2)} L(\pi_{F'}, s)$ is the normalized L -function of $\pi_{F'}$ such that $\mathcal{L}(\pi_{F'}, s) = \mathcal{L}(\pi_{F'}, 1-s)$.

In [28], we extended the above theorem to allow the automorphic representation π to have square-free level structures (which means the local representations π_v are either unramified or an unramified twist of the Steinberg representation), and to allow ramifications for the double cover $\nu : X' \rightarrow X$. We consider the moduli stack $\text{Sht}_G^r(\Sigma; \Sigma_\infty)$ where Σ is a finite set of places where

we add Iwahori level structures to the G -Shtukas; $\Sigma_\infty \subset \Sigma$ is a subset of places where we impose supersingular conditions. The admissibility condition forces r to have the same parity as $\#\Sigma_\infty$.

3.4.2. Theorem ([28]). *Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ with square-free level Σ . Assume the double cover $\nu : X' \rightarrow X$ is unramified over Σ . Let $\Sigma_\infty \subset \Sigma$ be the places that are inert in F' . Let $r \in \mathbb{Z}_{\geq 0}$ be of the same parity as $\#\Sigma_\infty$. Then for any $r_1, r_2 \in \mathbb{Z}_{\geq 0}$ such that $r_1 + r_2 = r$, there is an explicit linear combination $Z_T^{r_1, r_2}$ of the cycles $\{Z_T^\lambda; \lambda \in \{\pm 1\}^r\}$ such that*

$$\langle Z_{T, \pi}^{r_1, r_2}, Z_{T, \pi}^{r_1, r_2} \rangle_{\text{Sht}_G^r(\Sigma; \Sigma_\infty)} = \frac{q^{2-2g+\rho/2-N} \mathcal{L}^{(r_1)}(\pi, 1/2) \mathcal{L}^{(r_2)}(\pi \otimes \eta_{F'/F}, 1/2)}{2(-\log q)^r L(\pi, \text{Ad}, 1)}.$$

where

- $N = \deg \Sigma$, and ρ is the degree of the ramification locus of ν .
- $\mathcal{L}(\pi, s) = q^{(2g-2+N/2)(s-1/2)} L(\pi, s)$ is the normalized L -function of π such that $\mathcal{L}(\pi, s) = \mathcal{L}(\pi, 1-s)$.
- $\eta_{F'/F}$ is the character of $F^\times \backslash \mathbb{A}^\times$ corresponding to the quadratic extension F'/F .
- $\mathcal{L}(\pi \otimes \eta_{F'/F}, s) = q^{(2g-2+\rho+N/2)(s-1/2)} L(\pi \otimes \eta_{F'/F}, s)$ is the normalized L -function of $\pi \otimes \eta_{F'/F}$ such that $\mathcal{L}(\pi \otimes \eta_{F'/F}, s) = \mathcal{L}(\pi \otimes \eta_{F'/F}, 1-s)$.

When $r = 0$, the above theorem is a special case of the Waldspurger formula [22], and our proof in this case is very close to the one given by Jacquet [11]. When $r = 1$ and $\#\Sigma_\infty = 1$, the above theorem is an analogue of the Gross-Zagier formula (see [9]) which expresses the first derivative of the base-change L -function of a cuspidal Hecke eigenform in terms of the height of Heegner points on the modular curve. However our proof is very different from the original proof of the Gross-Zagier formula in that we do not need to explicitly compute either side of the formula.

3.4.3. Relation with the B-SD conjecture. Theorem 3.4.2 is applicable to those π coming from semistable elliptic curves E over the function field F . The relation of our result and the Birch-Swinnerton-Dyer conjecture for E can be roughly stated as follows. Take r to be the vanishing order of $L(E_{F'}, s) = L(\pi_{F'}, s - 1/2)$ at $s = 1$. According to the expectation (3.3), $Z_{T, \pi}^\lambda$ is an element in $\pi^K \otimes H^1(X' \otimes \bar{k}, j_{!*} \nu^* \rho_\pi)^{\otimes r}$. The 2-dimensional ℓ -adic Galois representation ρ_π attached to π is the Tate module of E , therefore $L(E_{F'}, s) = \det(1 - q^{-s} \text{Frob} | H^1(X' \otimes \bar{k}, j_{!*} \nu^* \rho_\pi))$. The standard conjecture predicts that the Frobenius acts semisimply on $H^1(X' \otimes \bar{k}, j_{!*} \nu^* \rho_\pi)$, hence the multiplicity of the Frobenius eigenvalue q should be r . We expect $Z_{T, \pi}^\lambda$ to lie in $\pi^K \otimes \wedge^r (H^1(X' \otimes \bar{k}, j_{!*} \nu^* \rho_\pi)^{\text{Fr}=q})$, and giving a basis for this hypothetically 1-dimensional space. However, currently we do not have a way to construct rational points on E from the Heegner-Drinfeld cycle Sht_T^λ .

3.4.4. The method to prove Theorems 3.4.1 is by comparing the Shtuka version of the relative trace $I_{T, T}^G(f) = \langle \mathcal{Z}_T^\lambda, f * \mathcal{Z}_T^\lambda \rangle_{\text{Sht}_G^r}$ as in (3.6) with the usual relative trace of the kind in Example 2.2.6. More precisely, for the triple (G, A, A) considered in Example 2.2.6, we consider the relative trace involving a complex variable s

$$\text{RTr}_{A, (A, \eta)}^G(f, s) = \langle \varphi! | \cdot |_{A(F) \backslash A(\mathbb{A})}^s, \varphi!(\eta | \cdot |^s)_{A(F) \backslash A(\mathbb{A})} \rangle_{L^2(G(F) \backslash G(\mathbb{A}), \mu_G)}.$$

Here $|\cdot| : A(F) \backslash A(\mathbb{A}) = F^\times \backslash \mathbb{A}^\times \rightarrow q^{\mathbb{Z}}$ is the global absolute value function, and $\eta = \eta_{F'/F}$. Let $J_r(f)$ be the r th derivative of $\text{RTr}_{A, (A, \eta)}^G(f, s)$ at $s = 0$. The key to the proof is to establish the following identity of relative traces for all spherical Hecke functions f

$$(3.13) \quad I_{T, T}^G(f) = (\log q)^{-r} J_r(f).$$

To prove this identity, it suffices to consider $f = h_D$ for effective divisors D on X (see Example 2.2.6). The moduli stacks $\mathcal{M}_{A,A,D}^G$ in Example 2.2.6 and $\mathcal{M}_{T,T,D}^G$ in Example 3.3.4 share the same Hitchin base $\mathcal{B}_D = \mathbb{H}^0(X, \mathcal{O}_X(D))$. We may fix a degree d and let D vary over effective divisors of degree d and get Hitchin maps $f_d : \mathcal{M}_{A,A,d}^G \rightarrow \mathcal{B}_d$ and $g_d : \mathcal{M}_{T,T,d}^G \rightarrow \mathcal{B}_d$. Formulae (2.20) and (3.12) suggest that we should try to prove an identity between the direct image complexes of f_d and g_d . The new geometric input here is the action of the Hecke correspondences $[\mathrm{Hk}_{\mathcal{M}_d}^{\lambda,\lambda}]$ on the complex $\mathbf{R}g_{d!}\mathbb{Q}_\ell$, which is the r -th iteration of the action of the correspondence $[\mathcal{H}_+]$ defined in Example 3.3.4. It turns out that the eigenvalues of the action of $[\mathcal{H}_+]$ on $\mathbf{R}g_{d!}\mathbb{Q}_\ell$ match exactly with the factors coming from taking the derivative of the relative trace $\mathrm{RTr}_{A,(A,\eta)}^G(f, s)$, which explains why derivatives of automorphic quantities are indeed geometric.

3.5. Arithmetic fundamental lemma. Generalizing Theorems 3.4.1 and 3.4.2 to higher rank groups would involve intersecting non-proper cycles in an ambient stack which is not of finite type. This is the same issue as the non-convergence of the naive relative trace (2.14), therefore a certain truncation and regularization procedure is needed. There is, however, a local version of such results that can be proved for higher rank groups. One example of such a local version is the Arithmetic Fundamental Lemma formulated by W. Zhang [30] originally for Rapoport-Zink spaces. In [29], we stated a higher derivative extension of W.Zhang's conjecture in the function field case, and sketched a proof. This was the first time higher derivatives of automorphic quantities were related to geometry, and it partially motivated the later work [27].

3.5.1. Local Shtukas. The moduli of Shtukas has a local version. Fix a local function field F_x with ring of integers \mathcal{O}_x . In the diagram (3.2) defining the moduli of Shtukas, we may replace Bun_G by the affine Grassmannian Gr_G , and replace Hk_G^μ by an iterated Hecke correspondence for Gr_G over a formal disk $\Delta_r = \mathrm{Spf}(\mathcal{O}_x \widehat{\otimes} \cdots \widehat{\otimes} \mathcal{O}_x)$ of dimension r . We also have the freedom of changing the Frobenius morphism on Gr_G by the composition $b \circ \mathrm{Fr} : \mathrm{Gr}_G \rightarrow \mathrm{Gr}_G$, where b is an element of the loop group of G giving the datum of a G -isocrystal. The resulting object ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ by forming the Cartesian square as (3.2) is called the moduli space of local Shtukas, and it is a formal scheme over Δ_r . The special fiber of ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ is an iterated version of the affine Deligne-Lusztig variety. The twisted centralizer G_b of b is an inner form of a Levi subgroup of G , and $G_b(F_x)$ acts on ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$.

For a subgroup $H \subset G$ with a sequence of coweights $\lambda = (\lambda_1, \dots, \lambda_r)$ such that $\lambda_i \leq_G \mu_i$ for all i and an H -isocrystal b_H compatible with b , we have a morphism $\theta^{\mathrm{loc}} : {}^b_H\mathrm{Sht}_H^{\lambda,\mathrm{loc}} \rightarrow {}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ over Δ_r . This morphism is a closed embedding because $\mathrm{Gr}_H \rightarrow \mathrm{Gr}_G$ is. We define the local Heegner-Drinfeld cycle ${}^b_H\mathcal{Z}_H^{\lambda,\mathrm{loc}}$ as the image of θ^{loc} .

If we have two local Heegner-Drinfeld cycles ${}^{b_1}\mathcal{Z}_{H_1}^{\lambda,\mathrm{loc}}$ and ${}^{b_2}\mathcal{Z}_{H_2}^{\kappa,\mathrm{loc}}$ in ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ with complementary dimensions, and if μ_i are minuscule and the reduced structure of their intersection is proper over k , we may ask for their intersection number in ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$. More generally, if $\delta \in G_b(F_x)$, we may consider the intersection number

$$I_\delta = \langle {}^{b_1}\mathcal{Z}_{H_1}^{\lambda,\mathrm{loc}}, \delta \cdot {}^{b_2}\mathcal{Z}_{H_2}^{\kappa,\mathrm{loc}} \rangle_{{}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}}$$

using the action of $G_b(F_x)$ on ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$. When $\mu = 0$, this is the same as the local orbital integral $J_{H_1, H_2, \delta}^G(\mathbf{1}_{G(\mathcal{O}_x)})$ (see (2.15)) for the relative trace formula of the triple (G, H_1, H_2) .

3.5.2. Example. Let F'_x/F_x be an unramified quadratic extension, with ring of integers \mathcal{O}'_x and residue field k'_x . Fix a Hermitian vector space $W_{n,x}$ of dimension n over x , and let U_n be the unitary group of $W_{n,x}$. We define the moduli of local Shtukas $\mathrm{Sht}_{U_n}^{r,\mathrm{loc}}$ over $\Delta'_r =$

$\mathrm{Spf}(\mathcal{O}'_x \widehat{\otimes}_{k'_x} \cdots \widehat{\otimes}_{k'_x} \mathcal{O}'_x)$ in the following way. Let Gr_{U_n} be the affine Grassmannian classifying self-dual lattices in $W_{n,x}$. Since U_n is split over F'_x , the base change $\mathrm{Gr}_{U_n} \otimes_{k_x} k'_x$ can be identified with the affine Grassmannian $\mathrm{Gr}_{\mathrm{GL}_n} \otimes_{k_x} k'_x$ classifying \mathcal{O}'_x -lattices in $W_{n,x}$. We have the local Hecke correspondence $\mathrm{Hk}_{U_n}^{\mathrm{loc}}$ over Δ'_1 which, after identifying $\mathrm{Gr}_{U_n} \otimes_{k_x} k'_x$ with $\mathrm{Gr}_{\mathrm{GL}_n} \otimes_{k_x} k'_x$, corresponds to the upper modification of lattices in $W_{n,x}$ of colength one. Let $\mathrm{Hk}_{U_n}^{r,\mathrm{loc}}$ be the r -fold composition of $\mathrm{Hk}_{U_n}^{\mathrm{loc}}$ as a correspondence, so $\mathrm{Hk}_{U_n}^{r,\mathrm{loc}} \rightarrow \Delta'_r$. Then $\mathrm{Sht}_{U_n}^{r,\mathrm{loc}}$ is defined using the Cartesian diagram

$$\begin{array}{ccc} \mathrm{Sht}_{U_n}^{r,\mathrm{loc}} & \longrightarrow & \mathrm{Hk}_{U_n}^{r,\mathrm{loc}} \\ \downarrow & & \downarrow \\ \mathrm{Gr}_{U_n} & \xrightarrow{(\mathrm{id}, \mathrm{Fr})} & \mathrm{Gr}_{U_n} \times \mathrm{Gr}_{U_n} \end{array}$$

The group $U_n(F_x)$ acts on $\mathrm{Sht}_{U_n}^{r,\mathrm{loc}}$. We remark that $\mathrm{Sht}_{U_n}^{r,\mathrm{loc}}$ is formally smooth over Δ'_r of relative dimension $r(n-1)$.

Now suppose we are in the situation of Example 2.3.4 so we have an orthogonal decomposition $W_{n,x} = W_{n-1,x} \oplus F'_x e_n$ and $(e_n, e_n) = 1$. Adding a standard lattice $\mathcal{O}'_x e_n$ gives an embedding $\mathrm{Gr}_{U_{n-1}} \hookrightarrow \mathrm{Gr}_{U_n}$ compatible with the Hecke modifications, hence induces an embedding $\mathrm{Sht}_{U_{n-1}}^{r,\mathrm{loc}} \hookrightarrow \mathrm{Sht}_{U_n}^{r,\mathrm{loc}}$. Consider the diagonal map

$$\Delta : \mathrm{Sht}_{U_{n-1}}^{r,\mathrm{loc}} \rightarrow \mathrm{Sht}_{U_{n-1}}^{r,\mathrm{loc}} \times_{\Delta'_r} \mathrm{Sht}_{U_n}^{r,\mathrm{loc}}.$$

The image of Δ gives an $r(n-1)$ -dimensional cycle $\mathcal{Z}_x^{r,\mathrm{loc}}$ in the $2r(n-1)$ -dimensional ambient space $\mathrm{Sht}_{U_{n-1}}^{r,\mathrm{loc}} \times_{\Delta'_r} \mathrm{Sht}_{U_n}^{r,\mathrm{loc}}$. For strongly regular semisimple $\delta \in U_n(F_x)$ (with respect to the conjugation action by U_{n-1}) we form the intersection number

$$I_{x,\delta}^{\mathrm{U},r} = \langle \mathcal{Z}_x^{r,\mathrm{loc}}, (\mathrm{id} \times \delta) \mathcal{Z}_x^{r,\mathrm{loc}} \rangle_{\mathrm{Sht}_{U_{n-1}}^{r,\mathrm{loc}} \times_{\Delta'_r} \mathrm{Sht}_{U_n}^{r,\mathrm{loc}}}.$$

This makes sense because the support of the intersection of the two cycles $\mathcal{Z}_x^{r,\mathrm{loc}}$ and $(\mathrm{id} \times \delta) \mathcal{Z}_x^{r,\mathrm{loc}}$ is proper. When $r = 0$, we have $I_{x,\delta}^{\mathrm{U},r}$ is either equal to $J_{x,\delta}^{\mathrm{U}}(\mathbf{1}_{U(\Lambda_{n,x})})$ as in Example 2.3.4 if a self-dual lattice $\Lambda_{n,x} \subset W_{n,x}$ exists, or 0 otherwise.

To state the higher arithmetic fundamental lemma, we introduce a variant of the orbital integral (2.21) with a complex variable s

$$J_{x,\gamma}^{\mathrm{GL}}(f, s) = \int_{\mathrm{GL}_{n-1}(F_x)} f(h^{-1}\gamma h) \eta_x(\det h) |\det h|^s dh, \quad \gamma \in \mathrm{S}_n(F_x), f \in C_c^\infty(\mathrm{S}_n(F_x)).$$

For γ strongly regular semisimple, $J_{x,\gamma}^{\mathrm{GL}}(f, s)$ is a Laurent polynomial in q_x^s , where $q_x = \#k_x$. Let

$$J_{x,\gamma}^{\mathrm{GL},r}(f) = \left(\frac{d}{ds} \right)^r \Big|_{s=0} J_{x,\gamma}^{\mathrm{GL}}(f, s).$$

3.5.3. Theorem ([29]). *Let $\gamma \in \mathrm{GL}_n(F_x)$ and $\delta \in U_n(F_x)$ be strongly regular semisimple with the same invariants in the sense of Examples 2.3.3 and 2.3.4. Then*

$$(3.14) \quad I_{x,\delta}^{\mathrm{U},r} = c(\log q_x)^{-r} J_{x,\gamma}^{\mathrm{GL},r}(\mathbf{1}_{\mathrm{S}_n(\mathcal{O}_x)})$$

with an explicit constant c depending on r and the invariants of γ .

The proof consists of the following main steps.

- (1) Prove a global analogue of (3.14). Consider the triple (G, H_1, H_2) as in Example 2.3.4 with coweights μ for G and λ for $H_1 = H_2$ being the first fundamental coweights. Let the triple (G', H'_1, H'_2) be as in Example 2.3.3. Recall that for the global situation, the intersection number $I_{(H_1, \lambda), (H_2, \lambda)}^{(G, \mu)}(\mathbf{1}_{K_0 g K_0})$ is the degree of a certain 0-cycle on $\text{Sht}_{(H_1, \lambda), (H_2, \lambda), K_0 g K_0}^{(G, \mu)}$ introduced in the diagram (3.7). On the other hand, we have a decomposition (3.11) of $\text{Sht}_{(H_1, \lambda), (H_2, \lambda), K_0 g K_0}^{(G, \mu)}$ into a disjoint union of $\text{Sht}_{(H_1, \lambda), (H_2, \lambda), K_0 g K_0}^{(G, \mu)}(a)$ indexed by k -points of the base $\mathcal{B}_{H_1, H_2, K_0 g K_0}^G$. For strongly regular semisimple a , $\text{Sht}_{(H_1, \lambda), (H_2, \lambda), K_0 g K_0}^{(G, \mu)}(a)$ is proper for any g , so we can talk about the degree of the a -component of the zero cycle $\mathcal{Z}_{H_1}^\lambda \cdot (f * \mathcal{Z}_{H_2}^\lambda)$, denoted $\langle \mathcal{Z}_{H_1}^\lambda, f * \mathcal{Z}_{H_2}^\lambda \rangle_a$. The global analogue of (3.14) means proving an identity of the form (3.13), but with both sides replaced by their a -components. One can prove such a global identity by analyzing the direct image complexes of the Hitchin maps h_{H_1, H_2}^G and $h_{H'_1, H'_2}^{G'}$ using sheaf-theoretic methods, as we did in [23] and [27].
- (2) Deduce the arithmetic fundamental lemma from the global identity. The moduli $\text{Sht}_G^{\mu, \text{loc}}$ of local Shtukas for G is related to a formal completion of the global moduli stack Sht_G^μ by a uniformization diagram, analogous to the one relating Rapoport-Zink spaces and Shimura varieties. Using the uniformization, one can express $\langle \mathcal{Z}_{H_1}^\lambda, f * \mathcal{Z}_{H_2}^\lambda \rangle_a$ as a finite sum, where each summand is a product of usual orbital integrals and intersection numbers of the form $I_{x_i, \delta}^{U, r_i}$ (with $\sum r_i = r$). There is a similar product formula for the global orbital integral for (G', H'_1, H'_2) . By choosing a appropriately we may deduce the local identity (3.14) from the global one using the product expansions and the known fundamental lemma (Theorem 2.3.5).

REFERENCES

- [1] Arthur, J. An introduction to the trace formula. In *Harmonic analysis, the trace formula, and Shimura varieties*, 1–263, Clay Math. Proc., 4, Amer. Math. Soc., Providence, RI, 2005.
- [2] Bouthier, A. La fibration de Hitchin-Frenkel-Ngô et son complexe d'intersection. *Ann. Sci. Éc. Norm. Supér.* (4) 50 (2017), no. 1, 85–129.
- [3] Bouthier, A.; Ngô B.C.; Sakellaridis, Y. On the formal arc space of a reductive monoid. *Amer. J. Math.* 138 (2016), no. 1, 81–108.
- [4] Braverman, A.; Bezrukavnikov, R. Geometric Langlands correspondence for D -modules in prime characteristic: the $\text{GL}(n)$ case. *Pure Appl. Math. Q.* 3 (2007), no. 1, Special Issue: In honor of Robert D. MacPherson. Part 3, 153–179.
- [5] Chen, T-H.; Zhu, X. Geometric Langlands in prime characteristic. *Compos. Math.* 153 (2017), no. 2, 395–452.
- [6] Drinfeld, V. G. Elliptic modules. *Mat. Sb. (N.S.)* 94(136) (1974), 594–627, 656.
- [7] Drinfeld, V. G. Moduli varieties of F -sheaves. (Russian) *Funktsional. Anal. i Prilozhen.* 21 (1987), no. 2, 23–41.
- [8] Frenkel, E.; Ngô, B.C. Geometrization of trace formulas. *Bull. Math. Sci.* 1 (2011), no. 1, 129–199.
- [9] Gross, B.; Zagier, D. Heegner points and derivatives of L -series. *Invent. Math.* 84 (1986), no. 2, 225–320.
- [10] Hitchin, N. Stable bundles and integrable systems. *Duke Math. J.* 54 (1987), no. 1, 91–114.
- [11] Jacquet, H. Sur un résultat de Waldspurger. *Ann. Sci. École Norm. Sup.* (4) 19 (1986), no. 2, 185–229.
- [12] Jacquet, H.; Rallis, S. On the Gross-Prasad conjecture for unitary groups. *On certain L -functions*, 205–264, Clay Math. Proc., 13, Amer. Math. Soc., Providence, RI, 2011.
- [13] Kapustin, A.; Witten, E. Electric-magnetic duality and the geometric Langlands program. *Commun. Number Theory Phys.* 1 (2007), no. 1, 1–236.
- [14] Lafforgue, L. Chtoucas de Drinfeld et correspondance de Langlands. *Invent. Math.* 147 (2002), no. 1, 1–241.
- [15] Lafforgue, V. Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale, <http://arxiv.org/abs/1209.5352>, 2013.
- [16] Ngô, B.C. Fibration de Hitchin et endoscopie. *Invent. Math.* 164 (2006), 399–453.
- [17] Ngô, B.C. Le lemme fondamental pour les algèbres de Lie, *Publ. Math. Inst. Hautes Études Sci.* 111 (2010), 1–169.

- [18] Ngô, B.C. On a certain sum of automorphic L -functions. In *Automorphic forms and related geometry: assessing the legacy of I. I. Piatetski-Shapiro*, 337–343, Contemp. Math., 614, Amer. Math. Soc., Providence, RI, 2014.
- [19] Oblomkov, A.; Yun, Z. Geometric representations of graded and rational Cherednik algebras. *Adv. Math.* 292 (2016), 601–706.
- [20] Sakellaridis, Y.; Venkatesh, A. Periods and harmonic analysis on spherical varieties. To appear in *Astérisque*, 296pp.
- [21] Varshavsky, Y. Moduli spaces of principal F -bundles. *Selecta Math. (N.S.)* 10 (2004), no. 1, 131–166.
- [22] Waldspurger, J.-L. Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie. *Compositio Math.* 54 (1985), no. 2, 173–242.
- [23] Yun, Z. The fundamental lemma of Jacquet and Rallis (with an appendix by J.Gordon). *Duke Math. J.* 156 (2011), no. 2, 167–227.
- [24] Yun, Z. Global Springer theory, *Adv. Math.* 228 (2011) 266–328.
- [25] Yun, Z. Langlands duality and global Springer theory. *Compos. Math.* 148 (2012) 835–867.
- [26] Yun, Z. The spherical part of the local and global Springer actions, *Math. Ann.* 359 (3-4) (2014) 557–594.
- [27] Yun, Z.; Zhang, W. Shtukas and the Taylor expansion of L -functions. *Ann. of Math. (2)* 186 (2017), 767–911.
- [28] Yun, Z.; Zhang, W. Shtukas and the Taylor expansion of L -functions, II. Preprint. arXiv:1712.08026.
- [29] Yun, Z. The arithmetic fundamental lemma over function fields. Preprint.
- [30] Zhang, W. On arithmetic fundamental lemmas. *Invent. Math.* 188 (2012), no. 1, 197–252.
- [31] Zhang, W. Fourier transform and the global Gan-Gross-Prasad conjecture for unitary groups. *Ann. of Math. (2)* 180 (2014), no. 3, 971–1049.

DEPARTMENT OF MATHEMATICS, MIT, 77 MASSACHUSETTS AVE, CAMBRIDGE, MA 02139
E-mail address: `zyun@mit.edu`