ANNALS OF
MATHEMATICS

Arithmetic quantum unique ergodicity for symplectic linear maps of the multidimensional torus

By Dubi Kelmer

SECOND SERIES, VOL. 171, NO. 2

March, 2010

ANMAAH

Annals of Mathematics, **171** (2010), 815–879

Arithmetic quantum unique ergodicity for symplectic linear maps of the multidimensional torus

By Dubi Kelmer

Abstract

We look at the expectation values for quantized linear symplectic maps on the multidimensional torus and their distribution in the semiclassical limit. We construct super-scars that are stable under the arithmetic symmetries of the system and localize on invariant manifolds. We show that these super-scars exist only when there are isotropic rational subspaces, invariant under the linear map. In the case where there are no such scars, we compute the variance of the fluctuations of the matrix elements for the desymmetrized system and present a conjecture for their limiting distributions.

Introduction

Quantization of discrete chaotic dynamics over a compact phase space has proved to be an effective toy model for understanding phenomena in quantum chaos. The first such model was the quantization of the cat map, a symplectic linear map acting on the 2-dimensional torus [17]. In this paper, we look at the multidimensional analog of this model, the quantization of symplectic linear maps on a multidimensional torus. We generalize some of the results obtained for the two-dimensional case and present some new phenomena occurring in higher dimensions.

Quantum cat map. In an attempt to gain a better understanding of the correspondence between classical and quantum mechanics and, in particular, phenomena in quantum chaos, Hannay and Berry introduced a model for quantum mechanics on the torus [17]. The classical dynamics underlying this model is simply the iteration of a symplectic linear map, $A \in \mathrm{Sp}(2, \mathbb{Z})$, acting on the 2-torus, known colloquially as a cat map. For quantizing the torus, one takes a family of finite dimensional Hilbert spaces of states, $\mathcal{H}_N = L^2(\mathbb{Z}/N\mathbb{Z})$ (where N stands for the inverse of Planck's constant). The quantization of smooth observables $f \in C^\infty(\mathbb{T}^2)$ are operators $\mathrm{Op}_N(f)$ acting on $\mathcal{H}_N$, and the quantization of the classical dynamics is a unitary operator $U_N(A)$, known as the quantum propagator. The connection with the classical system is achieved through an exact form of "Egorov's theorem":

$$U_N(A)^{-1}\mathrm{Op}_N(f)U_N(A) = \mathrm{Op}_N(f \circ A), \quad \forall f \in C^\infty(\mathbb{T}^2).$$

Quantum ergodicity. When the matrix A has no eigenvalues that are roots of unity, the induced classical dynamics is ergodic and mixing. The quantum analog of this, following the correspondence principle, is that the expectation values of an observable $\langle \mathrm{Op}_N(f)\psi, \psi\rangle$ (in an eigenfunction ψ s.t. $U_N(A)\psi = \lambda\psi$) should tend to the phase space average of the observable in the semiclassical limit.

By an analog of Shnirelman's theorem, one can show that indeed almost all of these matrix elements converge to the phase space average [4]. This notion is usually referred to as "Quantum Ergodicity" (QE) and was shown to hold for a large class of ergodic dynamical systems [4], [6], [31], [32]. However, the stronger notion of "Quantum Unique Ergodicity" (QUE), where there are no exceptional subsequences of eigenfunctions, doesn't hold for this model. Indeed, in [10] Faure, Nonnenmacher and De Bièvre managed to construct a subsequence of eigenfunctions, for which the diagonal matrix elements do not converge to the phase space average but concentrate around a periodic orbit. Such exceptional subsequences are also referred to as scars.

Arithmetic quantum unique ergodicity. The existence of scars for the quantum cat map is related to high degeneracies in the spectrum of the quantum propagator. If we denote by $\mathrm{ord}(A, N)$ the smallest integer such that $A^s \equiv I \pmod N$, then the quantum propagator satisfies that $U_N(A)^{\mathrm{ord}(A,N)} = I$, implying spectral degeneracies of order $\frac{N}{\mathrm{ord}(A,N)}$. In particular, since there are infinitely many values of N for which $\mathrm{ord}(A, N)$ is of order $\log(N)$, there could be spectral degeneracies of order $\frac{N}{\log(N)}$. It is precisely for these values of N that the scars in [10] were constructed.

In [23] Kurlberg and Rudnick introduced a group of symmetries of the system, i.e., commuting unitary operators that commute with $U_N(A)$, that remove most of the spectral degeneracies. These operators are called Hecke operators in an analogy to a similar setup on the modular surface [19], [30]. The space $\mathcal{H}_N$ has an orthonormal basis consisting of joint eigenfunctions called "Hecke eigenfunctions." For the desymmetrized system, Kurlberg and Rudnick showed that indeed $\langle \mathrm{Op}_N(f)\psi, \psi\rangle \overset{N\to\infty}{\longrightarrow} \int_{\mathbb{T}^2} f$, for any sequence, $\psi = \psi^{(N)}$, of "Hecke eigenfunctions" [23]. This notion is referred to as arithmetic quantum unique ergodicity, due to the arithmetic nature of these Hecke operators (both here and in the setting on the modular surface).

Higher dimensions. The Hannay-Berry model for the quantum cat map can be naturally generalized for symplectic linear automorphisms of higher-dimensional tori. For quantizing maps on the $2d$-dimensional torus, the Hilbert space of states, $\mathcal{H}_N = L^2(\mathbb{Z}/N\mathbb{Z})^d$, is of dimension N^d (where, again, N stands for the inverse of Planck's constant). The group of quantizable elements is the subgroup $\mathrm{Sp}_\theta(2d,\mathbb{Z})$ defined as

$$\mathrm{Sp}_\theta(2d,\mathbb{Z}) := \left\{ \begin{pmatrix} E & F \\ G & H \end{pmatrix} \in \mathrm{Sp}(2d,\mathbb{Z}) \middle| EF^t, GH^t \text{ are even matrices} \right\}.$$

The quantization of observables $f \in C^\infty(\mathbb{T}^{2d})$ and maps $A \in \mathrm{Sp}_\theta(2d,\mathbb{Z})$ again satisfies "exact Egorov":

$$U_N(A)^{-1}\mathrm{Op}_N(f)U_N(A) = \mathrm{Op}_N(f \circ A), \quad \forall f \in C^\infty(\mathbb{T}^{2d}).$$

Many of the results obtained on the two-dimensional model (i.e. $d = 1$), can be naturally generalized to higher dimensions. Nevertheless, there are still some new and surprising phenomena that occur in higher dimensions.

Results. One new phenomenon that occurs in high dimensions, is the existence of super-scars; that is, joint eigenfunctions of the propagator and all the Hecke operators localized on certain invariant manifolds[1].

Remark 0.1. The scars constructed in [10] (for $d = 1$) are related to the large spectral degeneracies of the propagator. We note that for $d > 1$, there are values of N for which the order $\mathrm{ord}(A, N)$ could grow like N (whenever the characteristic polynomial for A splits modulo N) and possibly even slower (see [29] for some numerical data on the order of A modulo N). Consequently, for these values there are large spectral degeneracies of order N^{d-1}. However, the scarring described here is not related to these degeneracies. In fact, the action of the Hecke operators reduce almost all of the spectral degeneracies (see Proposition 4.4).

Let $A \in \mathrm{Sp}_\theta(2d,\mathbb{Z})$ be a quantizable symplectic map. To any invariant rational isotropic subspace $E_0 \subseteq \mathbb{Q}^{2d}$, we assign a manifold $X_0 \subseteq \mathbb{T}^{2d}$ of dimension $2d - \dim E_0$, invariant under the dynamics.

THEOREM 1. *Let $A \in \mathrm{Sp}_\theta(2d,\mathbb{Z})$ with distinct eigenvalues. Let $E_0 \subseteq \mathbb{Q}^{2d}$ be an invariant subspace that is isotropic with respect to the symplectic form. Then, there is a subsequence of Hecke eigenfunctions $\psi \in \mathcal{H}_{N_i}$, such that the corresponding distributions*

$$f \mapsto \langle \mathrm{Op}_{N_i}(f)\psi, \psi \rangle$$

converge to Lebesgue measure on the manifold X_0.

[1] The name super-scars has been used before in a different context [1].

To illustrate this phenomenon, consider the following simple example (previously presented by Gurevich [14] and by Nonnenmacher [28]). Let $\tilde{A} \in \mathrm{GL}(d, \mathbb{Z})$ and take $A = \begin{pmatrix} \tilde{A}^t & 0 \\ 0 & \tilde{A}^{-1} \end{pmatrix} \in \mathrm{Sp}(2d, \mathbb{Z})$. Then, the space $E_0 = \left\{(\vec{n}_1, 0) \in \mathbb{Q}^{2d}\right\}$ is an invariant isotropic subspace, and the corresponding invariant manifold is $X_0 = \left\{\begin{pmatrix} 0 \\ \vec{p} \end{pmatrix} \in \mathbb{T}^{2d}\right\}$. The action of the quantum propagator corresponding to such a matrix is given by the formula $U_N(A)\psi(\vec{x}) = \psi(\tilde{A}\vec{x})$ (where the action of $\tilde{A} \in \mathrm{GL}(d, \mathbb{Z})$ on $\vec{x} \in (\mathbb{Z}/N\mathbb{Z})^d$ is the obvious one). One can then easily verify that the function $\psi_0(\vec{x}) = \sqrt{N}\delta_0(\vec{x})$ is an eigenfunction of $U_N(A)$. On the other hand, for any $f \in C^\infty(\mathbb{T}^{2d})$ a simple computation gives $\langle \mathrm{Op}(f)\psi_0, \psi_0\rangle = \int_{X_0} f\, dm_{X_0}$; that is the distribution $f \mapsto \langle \mathrm{Op}(f)\psi_0, \psi_0\rangle$ is Lebesgue measure on X_0.

Theorem 1 implies that any matrix $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ that has a rational invariant isotropic subspace is *not* arithmetically QUE. We show that these are the only counter examples.

THEOREM 2. *Let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ be a matrix with distinct eigenvalues. Then, a necessary and sufficient condition for the induced system to be arithmetically* QUE, *is that there are no rational subspaces $E \subseteq \mathbb{Q}^{2d}$ that are invariant under the action of A and are isotropic with respect to the symplectic form.*

Remark 0.2. Note that the existence of a rational invariant isotropic subspace is equivalent to the existence of an isotropic closed connected invariant subgroup of the torus. We can thus reformulate this theorem in these terms, i.e., the condition for arithmetic QUE, is the absence of invariant isotropic sub-tori.

Remark 0.3. It is interesting to note, that the sufficient conditions to insure arithmetic QUE, do not rule out matrices that have roots of unity for eigenvalues. So in a sense, arithmetic QUE can hold also for matrices that are not classically ergodic. This phenomenon already occurs for matrices in $\mathrm{SL}(2, \mathbb{Z})$. For example, $A = \begin{pmatrix} 1 & 1 \\ -2 & -1 \end{pmatrix}$ is not ergodic (because $A^4 = I$). Nevertheless it has two distinct eigenvalues and no rational invariant subspaces, hence arithmetic QUE does hold for this matrix.

For systems that are arithmetically QUE, we can also give a bound on the rate of convergence. For $\vec{n} \in \mathbb{Z}^{2d}$ we denote by $2d_{\vec{n}}$ the dimension of the smallest (symplectic) invariant subspace $E \subseteq \mathbb{Q}^{2d}$ such that $\vec{n} \in E$. For a smooth observable $f \in C^\infty(\mathbb{T}^{2d})$, define $d(f) = \min_{\hat{f}(\vec{n}) \neq 0} d_{\vec{n}}$.

THEOREM 3. *In the case where there are no rational isotropic subspaces, for any smooth $f \in C^\infty(\mathbb{T}^{2d})$ and any normalized Hecke eigenfunction $\psi \in \mathcal{H}_N$, the* *expectation values of* $\mathrm{Op}_N(f)$ *satisfy*:

$$\left|\langle \mathrm{Op}_N(f)\psi, \psi\rangle - \int_{\mathbb{T}^{2d}} f\right| \ll_{f,\varepsilon} N^{-\frac{d(f)}{4}+\varepsilon}.$$

Remark 0.4. The exponent of $\frac{d(f)}{4}$ in this theorem is not optimal. The correct exponent is probably $\frac{d(f)}{2}$, in consistence with the fourth moments (Proposition 3.5) and with the bounds for prime N (Corollary 4.8). For N prime, in the case where there are no invariant rational subspaces, the bound $O(N^{-d/2})$ was independently proved by Gurevich and Hadani [15].

We note that the behavior of the matrix elements of an observable $\mathrm{Op}_N(f)$ is related to the decomposition of N to its prime factors. Consequently, if we restrict ourselves to the case where N is prime, we can obtain much sharper results (e.g., for the bounds on the number of Hecke operators and the dimension of the joint eigenspaces).

We now consider only prime N and restrict to the case where there are no isotropic invariant rational subspaces. In this case the matrix elements of a smooth observable $f \in C^\infty(\mathbb{T}^{2d})$ with respect to a Hecke basis $\{\psi_i\}$ converge to their average $\int_{\mathbb{T}^{2d}} f$ and fluctuate around it. To study these fluctuations, we first give an asymptotic formula for their variance:

$$S_2^{(N)}(f) = \frac{1}{N^d}\sum_i \left|\langle \mathrm{Op}_N(f)\psi_i, \psi_i\rangle - \int f dx\right|^2.$$

Consider the decomposition $\mathbb{Q}^{2d} = \bigoplus E_\theta$ into symplectic irreducible invariant subspaces. To each space, we assign a quadratic form $Q_\theta : \mathbb{Z}^{2d} \to \mathbb{Z}[\lambda_\theta]$, where λ_θ is an eigenvalue of the restriction of A to E_θ, and define the product $Q = \prod Q_\theta$ (see §6 for an explicit construction). For a smooth observable $f \in C^\infty(\mathbb{T}^{2d})$ and an element $\nu \in \prod \mathbb{Z}[\lambda_\theta]$, define modified Fourier coefficients

$$f^\sharp(\nu) = \sum_{Q(\vec{n})=\nu} (-1)^{\vec{n}_1\vec{n}_2}\hat{f}(\vec{n}).$$

Define $d_\nu = \frac{1}{2}\sum_{\nu_\theta\neq 0} \dim E_\theta$ and $d_f = \min_{f^\sharp(\nu)\neq 0} d_\nu$. Note that if $\nu = Q(\vec{n})$, then $d_\nu = d_{\vec{n}}$ as defined in Theorem 3; hence for any smooth f we have $d(f) \leq d_f$. For $f \in \mathbb{C}^\infty(\mathbb{T}^{2d})$ define $V(f) = \sum_{d_\nu = d_f} |f^\sharp(\nu)|^2$.

THEOREM 4. *In the case where there are no rational isotropic subspaces, for a smooth observable* $f \in C^\infty(\mathbb{T}^{2d})$, *as* $N \to \infty$ *through primes, the quantum variance in the Hecke basis satisfies*

$$S_2^{(N)}(f) = \frac{V(f)}{N^{d_f}} + O\left(\frac{1}{N^{d_f+1}}\right).$$

Remark 0.5. We note that when there are symplectic invariant rational subspaces, one can construct observables for which $d_f < d$. We can thus produce a large family of examples (similar to the ones we described in [20]), for which the quantum variance is of a different order of magnitude from the one predicted for generic systems by the Feingold-Peres formula [9], [11].

Remark 0.6. In the case that there are isotropic invariant rational subspaces, the distribution can become degenerate (see Remark 4.6) and there is no definite behavior for the variance.

After establishing the quantum variance, we renormalize to have finite variance $V(f)$ and give a conjecture for the limiting distribution, generalizing the Kurlberg-Rudnick conjecture for the two-dimensional case [24]. To simplify the discussion, we will restrict ourselves to elementary observables of the form $e_{\vec{n}}(\vec{x}) = \exp(2\pi i\vec{n}\cdot\vec{x})$ (see §7 for treatment of any smooth observables).

For an observable $\mathrm{Op}_N(e_{\vec{n}})$, the matrix elements in the Hecke basis can be expressed as a product of certain exponential sums. The sums in the product are of the form:

$$E_q(\nu,\chi) = \frac{1}{|\mathscr{C}|}\sum_{1\neq x\in\mathscr{C}} e_q\left(\nu\kappa\frac{x+1}{x-1}\right)\chi(x)\chi_2(x),$$

where q is some power of N, $\mathscr{C}$ is either the multiplicative group $\mathbb{F}_q^*$ or the group of norm one elements in the quadratic extension $\mathbb{F}_{q^2}/\mathbb{F}_q$, χ is a character of $\mathscr{C}$ and χ_2 is the quadratic character of $\mathscr{C}$, $\nu\in\mathbb{F}_q$ and $\kappa\in\mathbb{F}_{q^2}$ satisfies: $\forall x\in\mathscr{C}$, $\kappa\frac{x+1}{x-1}\in\mathbb{F}_q$.

The Kurlberg-Rudnick conjecture regarding the limit distribution [24], is naturally generalized to a conjecture regarding these exponential sums.

CONJECTURE 5. *For each finite field $\mathbb{F}_q$, fix an element $0\neq\nu\in\mathbb{F}_q$ and consider the set of points on the line defined by the normalized exponential sums $\sqrt{q}E_q(\nu,\chi)$ for all characters $\chi:\mathscr{C}\to\mathbb{C}^*$. Then, as $q\to\infty$, these points become equidistributed on the interval $[-2,2]$ with respect to the Sato-Tate measure. Furthermore, if for each field $\mathbb{F}_q$ we fix a number of distinct elements $\nu_1,\dots,\nu_r\in\mathbb{F}_q$, then the limiting distributions corresponding to $\sqrt{q}E_q(\nu_1,\chi),\dots,\sqrt{q}E_q(\nu_r,\chi)$ are that of r independent random variables.*

We now wish to deduce from this a conjecture regarding the limiting distribution of the matrix elements. However, to do this we need to consider the decomposition of $\mathbb{F}_N^{2d}$ into invariant subspaces under the action of $A \pmod N$ (rather than the decomposition of $\mathbb{Q}^{2d}$ we used for the variance). For $\vec{n}\in\mathbb{Z}^{2d}$, let $E\subset\mathbb{F}_N^{2d}$ be the smallest (symplectic) invariant subspace containing $\vec{n} \pmod N$. Let $E=\bigoplus E_{\bar{\vartheta}}$ be the decomposition of E into irreducible symplectic invariant subspaces and let $2d_{\bar{\vartheta}} = \dim E_{\bar{\vartheta}}$. Then a matrix element for a Hecke eigenfunction $\langle\mathrm{Op}_N(e_{\vec{n}})\psi,\psi\rangle$ can be expressed as the product $\prod_{\bar{\vartheta}} E_{q_{\bar{\vartheta}}}(\nu_{\bar{\vartheta}},\chi_{\bar{\vartheta}})$, where $q_{\bar{\vartheta}} = N^{d_{\bar{\vartheta}}}$, the elements

$\nu_{\bar{\vartheta}}$ are determined by the projections of $\vec{n} \pmod{N}$ to $E_{\bar{\vartheta}}$, and the characters $\chi_{\bar{\vartheta}}$ are determined by the eigenfunction. Consequently, if we denote by $\mathbf{P}_k$ the set of primes for which there are precisely k invariant subspaces $E_{\bar{\vartheta}}$ in the decomposition, we can deduce:

CONJECTURE 6. *As $N \to \infty$ through primes from $\mathbf{P}_k$, the limiting distribution of normalized matrix elements $N^{d_{\vec{n}}/2}\langle \mathrm{Op}_N(e_{\vec{n}})\psi_i, \psi_i\rangle$ is that of a product of k independent random variables, each obeying the semi-circle law.*

It is interesting that while the expression for the variance depends only on the rational properties of A, the limiting distribution already depends specifically on the action of A on $\mathbb{F}_N^{2d}$ and can vary for different values of (prime) N. Moreover, notice that at least one of the sets $\mathbf{P}_k$ is always infinite, so there is a sequence of primes for which there is a limiting distribution. However, there could be other values of k for which the sets $\mathbf{P}_k$ are also infinite, resulting in different limiting distributions (see §7 for some examples).

Outline. This work is composed of three main parts. In the first part (§1), we describe in detail the quantization procedure. In the second part (§§2 and 3), we develop Hecke theory and give the proof of Theorem 3. In the third part (§§4, 5, 6, and 7), we restrict the discussion to the case where Planck's constant is an inverse of a prime number. For these values of Planck's constant, the Hecke operators and eigenfunctions reveal structure closely related to the Weil representation over finite fields. We use this structure to construct scars proving Theorem 1, and to calculate the quantum variance proving Theorem 4. We then generalize the Kurlberg-Rudnick conjecture regarding the limiting distribution of (normalized) matrix elements to deal with higher-dimensional tori.

1. Quantized linear toral automorphisms

The quantization of the cat map on the 2-torus, was originally introduced by Hannay and Berry [17], and is further described in [7], [21], [23]. For higher dimensions, the procedure is mostly analogous and is described in [3], [29]. We take an approach towards the quantization procedure through representation theory, similar to the one taken in [23].

1.1. *Quantization procedure.* We start by giving the outline for the quantization of arbitrary symplectic maps. For a discrete time dynamical system, given by the iteration of a symplectic map A on a phase space X, the quantization procedure can be described as follows: the first step, is constructing a one parameter family of Hilbert spaces $\mathcal{H}_h$, parametrized by Planck's constant. For each space, there is a procedure that assigns to each smooth function $f \in C^\infty(X)$ an operator $\mathrm{Op}_h(f)$ acting on $\mathcal{H}_h$. The connection with the classical system is fulfilled by the requirement that in the limit $h \to 0$, the commutator of the quantization of two observables f, g reproduces the quantization of their Poisson bracket $\{f, g\} = \sum_j (\partial f/\partial p_j)(\partial g/\partial q_j) - (\partial f/\partial q_j)(\partial g/\partial p_j)$:

$$\left\| \frac{1}{i\hbar}[\mathrm{Op}_h(f), \mathrm{Op}_h(g)] - \mathrm{Op}_h(\{f, g\}) \right\| \overset{h\to 0}{\longrightarrow} 0. \tag{1.1}$$

The dynamical part of the quantization is given by discrete time evolution of the algebra of operators. The evolution is through conjugation by a unitary map $U_h(A)$ of $\mathcal{H}_h$ (referred to as the quantum propagator). We require that in the limit $h \to 0$ the classical dynamics is reproduced, in the sense that

$$\left\| U_h(A)^{-1}\mathrm{Op}_h(f)U_h(A) - \mathrm{Op}_h(f \circ A) \right\| \overset{h\to 0}{\longrightarrow} 0. \tag{1.2}$$

In our case, the classical phase space is the multidimensional torus and the classical observables are smooth functions on the torus. For quantizing the torus, the admissible values of Planck's constant are inverses of integers $h = 1/N,\ N \geq 1$. The space of states is $\mathcal{H}_N = L^2((\mathbb{Z}/N\mathbb{Z})^d)$ of dimension N^d with inner product given by $\langle \psi, \phi \rangle = \frac{1}{N^d} \sum_{\vec{x} \pmod N} \psi(\vec{x})\overline{\phi(\vec{x})}$. To each observable $f \in C^\infty(\mathbb{T}^{2d})$, by an analog of Weyl quantization, we assign an operator $\mathrm{Op}_N(f)$ satisfying (1.1). The classical dynamics is given by an iteration of a symplectic linear map $A \in \mathrm{Sp}(2d, \mathbb{Z})$ acting on the torus, so that $\vec{x} = \binom{\vec{p}}{\vec{q}} \in \mathbb{T}^{2d} \mapsto A\vec{x}$ is a symplectic map of the torus. Given an observable $f \in C^\infty(\mathbb{T}^{2d})$, the classical evolution is defined by $f \mapsto f \circ A$. For a certain subset of matrices A, there is a unitary operator $U_N(A)$ acting on $\mathcal{H}_N$ satisfying an exact form of (1.2), i.e.,

$$U_N(A)^{-1}\mathrm{Op}_N(f)U_N(A) = \mathrm{Op}_N(f \circ A). \tag{1.3}$$

We now turn to describe these procedures in more detail.

1.1.1. *Quantizing observables.* In an analogous way to the quantization of observables on $\mathbb{T}^2$ [17], [23], introduce elementary operators $T_N(\vec{n})$ (with $\vec{n} = (\vec{n}_1, \vec{n}_2) \in \mathbb{Z}^{2d}$), acting on $\psi \in \mathcal{H}_N$ via:

$$T_N(\vec{n})\psi(\vec{y}) = e_{2N}(\vec{n}_1 \cdot \vec{n}_2)e_N(\vec{n}_2 \cdot \vec{y})\psi(\vec{y} + \vec{n}_1^t), \tag{1.4}$$

where we use the notation $e_N(x) = e^{\frac{2\pi i x}{N}}$. For notational convenience we also define a twisted version of these operators:

$$\widetilde{T}_N(\vec{n}) := (-1)^{N\vec{n}_1 \cdot \vec{n}_2} T_N(\vec{n}).$$

Remark 1.1. The twisted operators were originally introduced in [14], and make some of the arguments simpler (e.g., the trace formula (1.5)). Moreover, these operators satisfy the intertwining equation (1.6) for all of the symplectic group rather than for the subgroup $\mathrm{Sp}_\theta(2d, \mathbb{Z})$.

The main properties of the twisted elementary operators $\widetilde{T}_N(\vec{n})$ are summarized in the following proposition.

PROPOSITION 1.1. *For the operators $\widetilde{T}_N(\vec{n})$ defined above*:

(1) $\widetilde{T}_N(\vec{n})^* = \widetilde{T}_N(-\vec{n}) = \widetilde{T}_N(\vec{n})^{-1}$ *are unitary operators.*

(2) *The composition of two elementary operators is given by*

$$\widetilde{T}_N(\vec{m})\widetilde{T}_N(\vec{n}) = e_{2N}((1+N^2)\omega(\vec{m},\vec{n}))\widetilde{T}_N(\vec{m}+\vec{n}),$$

implying commutation relation

$$\widetilde{T}_N(\vec{m})\widetilde{T}_N(\vec{n}) = e_N(\omega(\vec{m},\vec{n}))\widetilde{T}_N(\vec{n})\widetilde{T}_N(\vec{m}),$$

where $\omega(\vec{m},\vec{n}) = \vec{m}_1\cdot\vec{n}_2 - \vec{m}_2\cdot\vec{n}_1$ is the symplectic inner product.

(3) *For even N, $\widetilde{T}_N(\vec{n})$ only depends on $\vec{n}$ modulo $2N$, while for odd N it only depends on $\vec{n}$ modulo N.*

The proof is straightforward from (1.4).

For any smooth classical observable $f \in C^\infty(\mathbb{T}^{2d})$ with Fourier expansion $f(\vec{x}) = \sum_{\vec{n}\in\mathbb{Z}^{2d}} \hat{f}(\vec{n})\exp(2\pi i\vec{n}\cdot\vec{x})$, where $\vec{n}\cdot\vec{x} = \vec{n}_1\cdot\vec{p} + \vec{n}_2\cdot\vec{q}$, define its quantization by

$$\mathrm{Op}_N(f) := \sum_{\vec{n}\in\mathbb{Z}^{2d}} \hat{f}(\vec{n})T_N(\vec{n})$$

or alternatively in terms of the twisted operators

$$\mathrm{Op}_N(f) = \sum_{\vec{n}\in\mathbb{Z}^{2d}} \hat{f}(\vec{n})(-1)^{N\vec{n}_1\cdot\vec{n}_2}\widetilde{T}_N(\vec{n}).$$

Using the commutation relation given above and the rapid decay of the Fourier coefficients, relation (1.1) can be verified.

1.1.2. *The Heisenberg group.* The operators $\widetilde{T}_N(\vec{n})$ defined above are connected to a certain representation of a Heisenberg group H_N.

For $N \geq 1$ the corresponding Heisenberg group is taken to be

$$H_N = \left\{(\vec{n},t)|\vec{n}\in(\mathbb{Z}/2N\mathbb{Z})^{2d}\ ,t\in\mathbb{Z}/2N\mathbb{Z}\right\},$$

with a multiplication law given by

$$(\vec{n},t)\cdot(\vec{n}',t') = (\vec{n}+\vec{n}',t+t'+\omega(\vec{n},\vec{n}')).$$

It is easily verified that the center of this group is given by

$$Z(H_N) = \left\{(\vec{n},t)\in H_N|\vec{n}\equiv 0 \pmod{N}\right\}.$$

We construct a unitary representation of H_N on the space $\mathcal{H}_N = L^2((\mathbb{Z}/N\mathbb{Z})^d)$ by setting:

$$\pi(\vec{n},t) = e_{2N}((N^2+1)t)\widetilde{T}_N(\vec{n}).$$

The relations given in Proposition 1.1 insure that this is indeed a representation. Furthermore, the center of H_N acts through the character $\xi(\vec{n},t)=e_{2N}((N^2+1)t)$.

Remark 1.2. This representation can be realized as an induced representation from the one-dimensional representation of the normal subgroup

$$\{(\vec{n},t)|\vec{n}_2=0 \pmod N\}$$

given by $(\vec{n},t)\mapsto e_{2N}((N+1)t)$ for odd N and $(\vec{n},t)\mapsto e_{2N}(t+\vec{n}_1\vec{n}_2)$ for even N.

PROPOSITION 1.2. *Let π be a representation of the Heisenberg group which is given by ξ on the center* (*where ξ is the character defined above*); *then*:

- *The characters of the representation π are supported on the center.*
- *π is irreducible if and only if the dimension of the representation is N^d. In this case, the class of the representation π is determined by the character ξ.*

Proof. See [13, Lemma 1.2]. □

In our case, the dimension $\dim(\pi)=\dim(\mathcal{H}_N)=N^d$, and hence the representation π is irreducible. Furthermore, from the condition on the characters of π, we deduce that the trace of the elementary operators $\tilde{T}_N(\vec{n})$ is given by

$$\mathrm{Tr}(\tilde{T}_N(\vec{n}))=\begin{cases} N^d & \vec{n}\equiv 0 \pmod N \\ 0 & \text{otherwise.}\end{cases} \tag{1.5}$$

In particular, for fixed $\vec{n}\neq 0$ and sufficiently large N, the trace of $\tilde{T}_N(\vec{n})$ vanishes.

COROLLARY 1.3. *For any orthonormal basis for $\mathcal{H}_N$ and any smooth observable $f\in C^\infty(\mathbb{T}^{2d})$, the average of the diagonal matrix elements of* $\mathrm{Op}_N(f)$ *converges to the phase space average as $N\to\infty$.*

1.1.3. *Quantizing maps.* In this section we show how to assign to a symplectic linear map $A\in \mathrm{Sp}_\theta(2d,\mathbb{Z})$ acting on $\mathbb{T}^{2d}$, a unitary operator $U_N(A)$ acting on $L^2((\mathbb{Z}/N\mathbb{Z})^d)$ s.t. for all observables $f\in C^\infty(\mathbb{T}^{2d})$,

$$U_N(A)^{-1}\mathrm{Op}_N(f)U_N(A)=\mathrm{Op}(f\circ A).$$

Any symplectic matrix $A\in \mathrm{Sp}(2d,\mathbb{Z})$ naturally acts on H_N by automorphism via $(\vec{n},t)^A=(\vec{n}A,t)$. Composing the representation π with the action of A thus gives a new representation $\pi^A(\vec{n},t)=\pi(\vec{n}A,t)$ that is again irreducible and acts on the center through the same character $\xi(\vec{n},t)=e_{2N}((1+N^2)t)$.

Therefore by Proposition 1.2, for any $A\in \mathrm{Sp}(2d,\mathbb{Z})$ the representations π,π^A are unitarily equivalent, i.e., there is a unitary intertwining operator $U_N(A)$ satisfying

$$\pi^A(\vec{n},t)=U_N(A)^{-1}\pi(\vec{n},t)U_N(A),\quad \forall(\vec{n},t)\in H_N$$

and, in particular, $\forall \vec{n} \in \mathbb{Z}^{2d}$

$$U_N(A)^{-1}\widetilde{T}_N(\vec{n})U_N(A) = \widetilde{T}_N(\vec{n}A).$$

Assume now that in addition A belongs to the subgroup

$$\mathrm{Sp}_\theta(2d,\mathbb{Z}) = \left\{ \begin{pmatrix} E & F \\ G & H \end{pmatrix} \in \mathrm{Sp}(2d,\mathbb{Z}) \middle| EF^t, GH^t \text{ are even matrices} \right\}.$$

Then $\forall \vec{n} \in \mathbb{Z}^{2d}$, the image $\vec{m} = \vec{n}A$ satisfies $\vec{n}_1 \cdot \vec{n}_2 \equiv \vec{m}_1 \cdot \vec{m}_2 \pmod 2$; hence for all observables $f \in C^\infty(\mathbb{T}^{2d})$,

$$U_N(A)^{-1}\mathrm{Op}_N(f)U_N(A) = \mathrm{Op}(f \circ A).$$

Because the operators $\widetilde{T}_N(\vec{n})$ only depend on $\vec{n}$ modulo $2N$ (respectively modulo N for odd N), the representation π^A also depends only on $A \mod 2N$ (respectively $\pmod N$). We can thus take the intertwining operator $U_N(A)$ to depend only on A modulo $2N$ (respectively N).

Remark 1.3. Note that $U_N(A)$ is defined as an intertwining operator for any $A \in \mathrm{Sp}(2d,\mathbb{Z})$. However, if $A \notin \mathrm{Sp}_\theta(2d,\mathbb{Z})$ then the operator $U_N(A)$ no longer satisfies the Egorov identity. When restricting to the subgroup $\mathrm{Sp}_\theta(2d,\mathbb{Z})$, the definition given here coincides with the standard definition given in [23] (for $d = 1$).

1.2. *Formulas for the quantized cat map.* The irreducibility of π imply (by Schur's lemma) that the map $U_N(A)$ is unique up to multiplication by phase. In other words, if U is a unitary map acting on $\mathcal{H}_N$, satisfying the intertwining equation

$$U\widetilde{T}_N(\vec{n}A) = \widetilde{T}_N(\vec{n})U, \quad \forall \vec{n} \in \mathbb{Z}^{2d}, \tag{1.6}$$

then after multiplying by some phase, $e^{i\alpha}U_N(A) = U$. On the other hand, the contrary is also true; that is, if $U = e^{i\alpha}U_N(A)$, then it obviously satisfies (1.6).

In what follows, we give formulas for operators satisfying (1.6), thus obtaining formulas for the quantized maps.

1.2.1. *Formulas through generators.* The group $\mathrm{Sp}(2d,\mathbb{Z})$ (and hence also $\mathrm{Sp}(2d,\mathbb{Z}/2N\mathbb{Z})$) is generated by the family of matrices

$$\begin{pmatrix} I & F \\ 0 & I \end{pmatrix}, \quad \begin{pmatrix} E^t & 0 \\ 0 & E^{-1} \end{pmatrix}, \quad \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

with $E \in \mathrm{GL}(d,\mathbb{Z})$ and $F \in \mathrm{Mat}(d,\mathbb{Z})$ symmetric [18, Th. 2].

For these matrices the corresponding operators act by the following formulas (up to phase):

$$U_N\begin{pmatrix} I & F\\ 0 & I\end{pmatrix}\psi(\vec{x}) = e_{2N}((1+N^2)\vec{x}\cdot F\vec{x})\psi(\vec{x}). \tag{1.7}$$

$$U_N\begin{pmatrix} E^t & 0\\ 0 & E^{-1}\end{pmatrix}\psi(\vec{x}) = \psi(E\vec{x}). \tag{1.8}$$

$$U_N\begin{pmatrix} 0 & I\\ -I & 0\end{pmatrix}\psi(\vec{x}) = \frac{1}{N^{d/2}}\sum_{\vec{y}\in(\mathbb{Z}/N\mathbb{Z})^d} e_N(\vec{x}\cdot\vec{y})\psi(\vec{y}). \tag{1.9}$$

One can verify directly that these formulas indeed satisfy (1.6). Consequently, the action of any element $U_N(A)$, $A\in\mathrm{Sp}(2d,\mathbb{Z})$ can be obtained by composing the appropriate operators given above for the generators.

1.2.2. *Formulas through averaging.* A different approach to obtain formulas for the operators $U_N(A)$ is through averaging of the representation over the Heisenberg group (similar to the p-adic formula given in [26, p. 37]). With this approach, for any $A\in\mathrm{Sp}(2d,\mathbb{Z})$ satisfying $A\equiv\pm I \pmod 4$, we obtain a formula for the propagator $U_N(A)$ in terms of the elementary operators $\widetilde{T}_N(\vec{n})$. Moreover, if N is odd the formula is valid without the parity condition.

Recall that we defined the operator $U_N(A)$ to be an intertwining operator of the representations π and π^A. It is easily verified that an operator defined by averaging of the form

$$F(\pi,\pi^A) = \sum_{h\in H_N/Z(H_N)} \pi(h)\pi^A(h^{-1})$$

is always an intertwining operator of these representations. Therefore (by Schur's lemma), it will coincide with the original operator after multiplying by some constant (i.e., $F(\pi,\pi^A)=c(A)U_N(A)$). Note that in general this constant might be zero.

PROPOSITION 1.4. *Let $A\in\mathrm{Sp}(2d,\mathbb{Z})$ be a matrix satisfying $A\equiv -I \pmod 4$. Denote by $\ker_N(A-I)$, the kernel of the map $(A-I):(\mathbb{Z}/N\mathbb{Z})^{2d}\to(\mathbb{Z}/N\mathbb{Z})^{2d}$. Then, the intertwining operator $F(\pi,\pi^A)=c(A)U_N(A)$ with*

$$|c(A)|^2 = N^{2d}|\ker_N(A-I)|,$$

and in particular $c(A)\neq 0$.

Proof. Note that we can identify the quotient $H_N/Z(H_N)$ with $(\mathbb{Z}/N\mathbb{Z})^{2d}$, so that

$$F(\pi,\pi^A) = \sum_{(\mathbb{Z}/N\mathbb{Z})^{2d}} \widetilde{T}_N(\vec{n})\widetilde{T}_N(-\vec{n}A). \tag{1.10}$$

Since the operator $U_N(A)$ is unitary, $F(\pi,\pi^A)F(\pi,\pi^A)^* = |c(A)|^2 I$. On the other hand, plugging in (1.10) gives,

$$\begin{aligned}
&F(\pi,\pi^A)F(\pi,\pi^A)^* \\
&\quad=\sum_{\vec n,\vec m}\widetilde T_N(\vec n)\widetilde T_N(-\vec nA)\widetilde T_N(\vec mA)\widetilde T_N(-\vec m)\\
&\quad=\sum_{\vec n,\vec m}e_N(\omega((\vec n-\vec m)A,\vec m))\widetilde T_N(\vec n)\widetilde T_N(-\vec m)\widetilde T_N(-\vec nA)\widetilde T_N(\vec mA)\\
&\quad=\sum_{\vec n,\vec m}e_N(\omega((\vec n-\vec m)A,\vec m)-\omega(\vec n,\vec m))\widetilde T_N(\vec n-\vec m)\widetilde T_N(-(\vec n-\vec m)A).
\end{aligned}$$

Now, change summation variable $\vec k=\vec n-\vec m$ to get

$$\begin{aligned}
F(\pi,\pi^A)F(\pi,\pi^A)^* &= \sum_{\vec k,\vec m}e_N(\omega(\vec k(A-I),\vec m))\widetilde T_N(\vec k)\widetilde T_N(-\vec kA)\\
&=\sum_{\vec k}\widetilde T_N(\vec k)\widetilde T_N(-\vec kA)\sum_{\vec m}e_N(\omega(\vec k(A-I),\vec m)).
\end{aligned}$$

Since the second sum vanishes whenever $\vec k(A-I)\neq 0 \pmod N$, we get that

$$F(\pi,\pi^A)F(\pi,\pi^A)^* = N^{2d}\sum_{\vec k\equiv\vec kA(N)}\widetilde T_N(\vec k)\widetilde T_N(-\vec kA).$$

Finally, when $A\equiv -I \pmod 4$, the condition $\vec k\equiv\vec kA \pmod N$ implies that $\widetilde T_N(\vec k)\widetilde T_N(-\vec kA)=I$, which concludes the proof. □

When $A\equiv -I \pmod 4$, the constant $c(A)$ does not vanish and we can divide by it to get a formula for $U_N(A)$:

$$U_N(A)=\frac{1}{c(A)}F(\pi,\pi^A),\quad (\forall A\equiv -I \pmod 4)). \tag{1.11}$$

When $A\equiv I \pmod 4$ the constant $c(A)$ might be zero. However, in this case $c(-A)\neq 0$ and since $U_N(A)=U_N(-A)U_N(-I)$ we get the formula:

$$U_N(A)=\frac{1}{c(-A)}F(\pi,\pi^{-A})U_N(-I),\quad (\forall A\equiv I \pmod 4)). \tag{1.12}$$

Remark 1.4. When N is odd, the condition $\vec k\equiv\vec kA \pmod N$ implies that $\widetilde T_N(\vec k)\widetilde T_N(-\vec kA)=I$ for any $A\in\mathrm{Sp}(2d,\mathbb Z)$ (without the parity condition). Thus, for odd N we can use both formulas for any symplectic matrix.

From these formulas we get the following corollaries:

COROLLARY 1.5. *Let $A,B\in\mathrm{Sp}(2d,\mathbb Z)$ be matrices that commute modulo N. If $B\equiv\pm I$* (mod 4) (*or if N is odd*), *then the corresponding operators $U_N(A)$,$U_N(B)$ commute as well.*

Proof. If $B \equiv -I \pmod 4$ (or if N is odd), use formula (1.11) for $U_N(B)$ and apply the intertwining (1.6) for the action of $U_N(A)$.

$$U_N(B)U_N(A) = U_N(A)\frac{1}{c(B)} \sum_{\vec{n}\in(\mathbb{Z}/N\mathbb{Z})^{2d}} \widetilde{T}_N(\vec{n}A)\widetilde{T}_N(-\vec{n}BA).$$

Now, change summation variable $\vec{n} \mapsto \vec{n}A$ (using the fact that A and B commute) to get $U_N(B)U_N(A) = U_N(A)U_N(B)$.

Otherwise, use formula (1.12) for $U_N(B)$. As above, the operators $F(\pi, \pi^{-B})$ and $U_N(-I)$ both commute with $U_N(A)$ and hence $U_N(B)$ commutes with $U_N(A)$ as well. □

COROLLARY 1.6. *The trace of* $U_N(A)$ *is given* (*up to phase*) *by*:

- *For* $A \equiv -I \pmod 4$ (*or for odd* N),
$$|\mathrm{Tr}(U_N(A))| = \sqrt{|\ker_N(A-I)|}.$$
- *For* N *even*, *and* $A \equiv I \pmod 4$, *either* $\mathrm{Tr}(U_N(A)) = 0$ *or*
$$|\mathrm{Tr}(U_N(A))| = \sqrt{\frac{|\ker_{2N}(A^2-I)|}{|\ker_N(A+I)|}}.$$

In particular $|\mathrm{Tr}(U_N(A))| \le 2^d \sqrt{|\ker_N(A-I)|}$.

Proof. In the first case, noticing that $\mathrm{Tr}(\widetilde{T}_N(\vec{n})\widetilde{T}_N(-\vec{n}A)) = 0$ when $\vec{n} \neq \vec{n}A \pmod N$, use formula (1.11) and take trace . Now, plug in $|c(A)|$ from Proposition 1.4 to get the result.

Otherwise, use formula (1.12). Using formula (1.11) for $U_N(-I)$ and taking trace we get that $\forall \vec{n} \in \mathbb{Z}^{2d}$,

$$\mathrm{Tr}(\widetilde{T}_N(\vec{n}(A+I))U_N(-I)) = 2^d.$$

Therefore,

$$|\mathrm{Tr}(U_N(A))| = \frac{2^d}{|c(-A)|}\Big|\sum_{\vec{n}(N)} e_{2N}(\omega(\vec{n},\vec{n}A))\Big|.$$

Finally, similar to a Gauss sum, when the sum $\sum e_{2N}(\omega(\vec{n},\vec{n}A))$ does not vanish, its absolute value is given by

$$\Big|\sum_{\vec{n}(N)} e_{2N}(\omega(\vec{n},\vec{n}A))\Big| = \frac{N^d\sqrt{|\ker_{2N}(A^2-I)|}}{2^d}.$$

The bound $|\mathrm{Tr}(U_N(A))| \le 2^d \sqrt{|\ker_N(A-I)|}$ is a consequence of the following observation,

$$|\ker_{2N}(A^2-I)| \le 2^{2d}|\ker_N(A^2-I)| \le 2^{2d}|\ker_N(A-I)||\ker_N(A+I)|. \quad \Box$$

1.3. *Multiplicativity.* The quantum propagators, $U_N(A)$, are unique up to a phase factor and thus define a projective representation of $\mathrm{Sp}(2d, \mathbb{Z}/2N\mathbb{Z})$; that is:

$$U_N(AB) = c(A,B)U_N(A)U_N(B). \tag{1.13}$$

From Corollary 1.5 we infer that: for odd N, if $AB = BA \pmod N$, then $c(A, B) = c(B, A)$ as well. For even N, this holds if $AB = BA \pmod{2N}$ and we restrict to the subgroup of matrices congruent to $\pm I$ modulo 4. This property by itself already allows us to define the Hecke operators (see §2). However, it is more convenient to work with a quantization such that the map $A \mapsto U_N(A)$ forms a representation of the symplectic group. In this section we show that such a quantization indeed exists:

THEOREM 7. *For each $N > 1$, there is a special choice of phases for the propagators, such that the map $A \mapsto U_N(A)$ is a representation of* $\mathrm{Sp}(2d, \mathbb{Z}/N\mathbb{Z})$ *when N is odd. Whereas for even integers, this map is a representation of the subgroup of* $\mathrm{Sp}(2d, \mathbb{Z}/2N\mathbb{Z})$ *composed of all matrices congruent to $\pm I$ modulo* 4.

In order to prove Theorem 7 for all integers, it is sufficient to prove it separately for odd integers and for integers of the form $N = 2^k$ (see [23, §4.1]).

1.3.1. *Odd integers.* When N is an odd integer, we follow a proof of Neuhauser [27]. As we apply this proof for the rings $\mathbb{Z}/N\mathbb{Z}$ (rather than finite fields as done in [27]), we review the proof in some detail:

Let $N \geq 1$ be an odd integer. Note that $-I$ is in the center of $\mathrm{Sp}(2d, \mathbb{Z}/N\mathbb{Z})$, so by Corollary 1.5, $\forall A \in \mathrm{Sp}(2d, \mathbb{Z}/N\mathbb{Z})$,

$$U_N(-I)U_N(A) = U_N(A)U_N(-I).$$

On the other hand, the operator $U_N(-I)$ acts by $U_N(-I)\psi(x) = \psi(-x)$ (formula (1.8)). Hence, the space $\mathcal{H}_N^+ = \{\psi \in \mathcal{H}_N | \psi(-x) = \psi(x)\}$ is an invariant subspace under the action of $\mathrm{Sp}(2d, \mathbb{Z}/N\mathbb{Z})$.

Denote by $U^+(A)$, the restriction of $U_N(A)$ to $\mathcal{H}_N^+$, to get that

$$U^+(AB) = c(A,B)U^+(A)U^+(B). \tag{1.14}$$

By taking determinants of equations (1.13) and (1.14) we get:

$$\det(U_N(AB)) = c(A,B)^{N^d} \det(U_N(A)) \det(U_N(B)),$$
$$\det(U^+(AB)) = c(A,B)^{\frac{N^d+1}{2}} \det(U^+(A)) \det(U^+(B)),$$

(note that the dimension of $\mathcal{H}^+$ is $\frac{N^d+1}{2}$). Define $\kappa(A) = \frac{\det(U_N(A))}{\det(U^+(A))^2}$, then $c(A,B) = \frac{\kappa(A)\kappa(B)}{\kappa(AB)}$ and $A \mapsto \kappa(A)U_N(A)$ is a representation of $\mathrm{Sp}(2d, \mathbb{Z}/N\mathbb{Z})$.

1.3.2. *Dyadic powers.* For integers of the form $N = 2^k$, we take a different approach by induction on the exponent k.

We define a subspace $\mathcal{H}_N^0 \subset \mathcal{H}_N$ of dimension $M^d = (N/2)^d$, invariant under the action of $\mathrm{Sp}_2(2d, 2N)$ (i.e., the matrices congruent to I modulo 2) and under the action of certain elementary operators. We then construct a representation of the Heisenberg group H_M on this space and show that it is equivalent to the original representation on $L^2(\mathbb{Z}/M\mathbb{Z})^d$. We can thus connect the restriction of the quantum propagators to the subspace $\mathcal{H}_N^0$ with the quantum propagators on $\mathcal{H}_M$, for which by induction we already have multiplicativity.

Define the subspace

$$\mathcal{H}_N^0 = \left\{\psi \in \mathcal{H}_N | \psi(\vec{y}) = 0, \quad \forall \vec{y} \neq 0 \pmod 2\right\}$$

and the congruence subgroup

$$\mathrm{Sp}_2(2d, 2N) = \{A \in \mathrm{Sp}(2d, \mathbb{Z}/2N\mathbb{Z}) | A \equiv I \pmod 2\}.$$

LEMMA 1.7. *For $N = 2^k$, $k \geq 2$ and any $A \in \mathrm{Sp}_2(2d, 2N)$, the space $\mathcal{H}_N^0$ is invariant under the action of $U_N(A)$.*

Proof. For any matrix $\begin{pmatrix} E & F \\ G & H \end{pmatrix} \in \mathrm{Sp}_2(2d, 2N)$, we have a Bruhat decomposition:

$$\begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} H^{t^{-1}} & 0 \\ 0 & H \end{pmatrix} \begin{pmatrix} I & H^t F \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ H^{-1}G & I \end{pmatrix}.$$

Consequently, the group $\mathrm{Sp}_2(2d, 2N)$ is generated by the family of matrices

$$u_+(X) = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix}, \quad u_-(Y) = \begin{pmatrix} I & 0 \\ Y & I \end{pmatrix}, \quad s(T) = \begin{pmatrix} T^t & 0 \\ 0 & T^{-1} \end{pmatrix}$$

where $X, Y, T \in \mathrm{Mat}(d, \mathbb{Z}/2N\mathbb{Z})$, $X = X^t$, $Y = Y^t$, $X \equiv Y \equiv 0 \pmod 2$, $T \equiv I \pmod 2$. Therefore, it is sufficient to show that $\mathcal{H}_N^0$ is invariant under the action of the corresponding operators. This can be done directly using the formulas given in Section 1.2.1. □

LEMMA 1.8. *For $N = 2^k$, $k \geq 2$, the space $\mathcal{H}_N^0$ is invariant under the action of $\widetilde{T}_N(\vec{n})$ for all $\vec{n} = (\vec{n}_1, \vec{n}_2)$ such that $\vec{n}_1 \equiv 0 \pmod 2$. Furthermore, if $\vec{n}_1 \equiv 0 \pmod N$ and $\vec{n}_2 \equiv 0 \pmod{N/2}$, then the restriction $\widetilde{T}_N(\vec{n})|_{\mathcal{H}_N^0} = I$.*

Proof. Direct computation using (1.4). □

Define two subgroups of $\mathrm{Sp}_2(2d, 2N)$,

$$S_2(2N) = \left\{ \begin{pmatrix} E & F \\ G & H \end{pmatrix} \in \mathrm{Sp}_2(2d, 2N) \middle| F \equiv 0 \pmod 4 \right\}$$

and

$$\hat{S}_2(2N) = \left\{ \begin{pmatrix} E & F \\ G & H \end{pmatrix} \in \mathrm{Sp}_2(2d, 2N) \middle| G \equiv 0 \pmod 4 \right\}.$$

Let $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$; then the map $A \mapsto -JAJ$ is an obvious isomorphism of these groups (in both directions). Another less trivial isomorphism is given by the map $j : S_2 \to \hat{S}_2$, defined by

$$j \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} E & F/2 \\ 2G & H \end{pmatrix}. \tag{1.15}$$

PROPOSITION 1.9. *For any $N = 2^k$, there is a choice of phases so that for any $A, B \in S_2(2N)$, $U_N(AB) = U_N(A)U_N(B)$. There is another choice such that for any $A, B \in \hat{S}_2(2N)$, $U_N(AB) = U_N(A)U_N(B)$.*

Proof. First note that it suffices to prove multiplicativity for $S_2(2N)$. Because for any $B \in \hat{S}_2(2N)$, there is $\tilde{B} \in S_2(2N)$ such that $B = -J\tilde{B}J$. Therefore, if we have multiplicativity for $S_2(2N)$, we can define for any $B \in \hat{S}_2(2N)$,

$$U_N(B) = U_N(J)^* U_N(\tilde{B}) U_N(J),$$

to get a multiplicativity for $\hat{S}_2(2N)$.

We now show multiplicativity for $S_2(2N)$ by induction on k. For $k = 1$, the group $S_2(4)$ includes only lower triangular matrices for which the formulas given in Section 1.2.1 are multiplicative.

For $k \geq 2$, by Lemma 1.7 the space $\mathcal{H}_N^0$ is invariant under the action of $\mathrm{Sp}_2(2d, \mathbb{Z})$ and hence also under the subgroup $S_2(2N)$. For $A \in S_2(2N)$, denote by $U_N^0(A)$ the restriction of $U_N(A)$ to $\mathcal{H}_N^0$.

Let $M = 2^{k-1} = N/2$ and consider the Heisenberg group H_M defined in Section 1.1.2, together with the representation on $L^2(\mathbb{Z}/M\mathbb{Z})$:

$$\pi(\vec{n}, t) = e_{2M}(t)\widetilde{T}_M(\vec{n}).$$

We now construct another representation on $\mathcal{H}_N^0 \subseteq L^2(\mathbb{Z}/N\mathbb{Z})$:

$$\tilde{\pi}(\vec{n}, t) = e_N(t)\widetilde{T}_N^0((2\vec{n}_1, \vec{n}_2)),$$

where $\widetilde{T}_N^0((2\vec{n}_1, \vec{n}_2))$ is the restriction of $\widetilde{T}_N((2\vec{n}_1, \vec{n}_2))$ to $\mathcal{H}_N^0$ (by Lemma 1.8 this is well defined). From the second part of Lemma 1.8, we see that the action on the center is given by $\tilde{\pi}(Mn, t) = e_N(t)I$. Consequently, by Proposition 1.2 there is a unitary operator $\mathcal{U} : \mathcal{H}_N^0 \to L^2(\mathbb{Z}/M\mathbb{Z})$ such that $\tilde{\pi} = \mathcal{U}^{-1}\pi\mathcal{U}$.

The intertwining equation for $U_N(A)$ imply that the restricted operators satisfy

$$U_N^0(A)^* \tilde{\pi}(n, t) U_N^0(A) = \tilde{\pi}(nj(A), t),$$

where $j : S_2 \to \hat{S}_2$ is the isomorphism defined in (1.15). Thus $\mathcal{U}U_N^0(A)\mathcal{U}^{-1}$ is the intertwining operator between π and $\pi^{j(A)}$ and by the uniqueness of the quantization we get: $\mathcal{U}U_N^0(A)\mathcal{U}^{-1} = \kappa(A)U_M(j(A))$. We can assume by induction that

$A \mapsto U_M(A)$ restricted to $\hat{S}_2(2M)$ is multiplicative. Finally, for $A, B \in S_2(2N)$ we have $U_N(A)U_N(B) = c(A, B)U_N(AB)$; hence the restricted operators satisfy $U_N^0(A)U_N^0(B) = c(A, B)U_N^0(AB)$ as well. Conjugating by $\mathcal{U}$ we get

$$\kappa(A)U_M(j(A))\kappa(B)U_M(j(B)) = c(A, B)\kappa(AB)U_M(j(AB)),$$

implying $c(A, B) = \frac{\kappa(A)\kappa(B)}{\kappa(AB)}$. Therefore, the map $A \mapsto \kappa(A)U_N(A)$ defined on $S_2(N)$ is multiplicative. □

Because the subgroup of matrices congruent to $\pm I$ modulo 4 is a subgroup of $S_2(N)$, this concludes the proof of Theorem 7.

2. **Hecke theory**

In the following section we introduce Hecke theory for the multidimensional torus. For a given symplectic matrix $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ with distinct eigenvalues, we follow the lines of [23] and construct "Hecke operators," a group of commuting operators that commute with the propagator $U_N(A)$. We show that this group of symmetries reduces almost all degeneracies in the spectrum.

Remark 2.1. The requirement that the matrix A has distinct eigenvalues is crucial for our construction. In fact, when there are degenerate eigenvalues, the group of matrices commuting with A modulo N is not necessarily commutative. In such a case, it is not clear how one should define the Hecke group and Hecke operators.

Remark 2.2. In Sections 2.3 and 2.5, in order to simplify the discussion, we will assume there are no rational isotropic subspaces invariant under the action of A. However, we note that the results presented in these sections (i.e., the bound on the number of Hecke operators in Lemma 2.7 and the dimensions of the joint eigenspaces in Proposition 2.8) are still valid without this assumption, and the proofs are analogous.

2.1. *Hecke operators.* In [23] Kurlberg and Rudnick constructed the Hecke operators (for $A \in \mathrm{Sp}(2, \mathbb{Z})$) by identifying integral matrices with elements of the (commutative) integral ring of a certain quadratic extension of the rationals. We follow the same idea, except that for $A \in \mathrm{Sp}(2d, \mathbb{Z})$, the correct ring to work with is the integral ring of a higher extension or rather a product of several such rings.

Let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ with $2d$ distinct eigenvalues. Let $\{\lambda_i\}_{i=1}^{2d}$ be all of its eigenvalues ordered so that $\lambda_{d+i} = \lambda_i^{-1}$. Denote by $\mathfrak{O}_i = \mathbb{Z}[\lambda_i] = \mathbb{Z}[\lambda_i^{-1}]$, $i = 1, \dots, 2d$, and define the ring

$$\mathfrak{O} = \left\{ \beta = (\beta_1, \dots, \beta_{2d}) \in \prod_{i=1}^{2d} \mathfrak{O}_i \,|\, \exists f \in \mathbb{Z}[t],\ f(\lambda_i) = \beta_i \right\}.$$

This ring is naturally isomorphic to the ring $\mathbb{Z}[t]/(P_A)$, where P_A is the characteristic (and minimal) polynomial for A. Thus, there is an embedding $\iota : \mathfrak{D} \hookrightarrow \mathrm{Mat}(2d, \mathbb{Z})$ (contained in the centralizer of A), given by

$$\begin{array}{ccccc} \mathfrak{D} & \to & \mathbb{Z}[t]/(P_A) & \hookrightarrow & \mathrm{Mat}(2d, \mathbb{Z}) \\ \beta & \mapsto & f & \mapsto & f(A). \end{array}$$

LEMMA 2.1. *To any element $\beta = f(\lambda) \in \mathfrak{D}$, define an element $\beta^* \in \prod \mathfrak{D}_i$, such that $\beta_i^* = f(\lambda_i^{-1})$. Then, the map $\beta \mapsto \beta^*$ is an automorphism of $\mathfrak{D}$. Furthermore, to any $\vec{n}, \vec{m} \in \mathbb{Z}^{2d}$ and any $\beta \in \mathfrak{D}$, the symplectic form ω satisfies*:

$$\omega(\vec{n}\iota(\beta), \vec{m}) = \omega(\vec{n}, \vec{m}\iota(\beta^*)).$$

Proof. The map $\beta \mapsto \beta^*$ is obviously injective, and it respects addition and multiplication. Therefore, to show that it is an automorphism it is sufficient to show that for any $\beta \in \mathfrak{D}$, $\beta^* \in \mathfrak{D}$ as well.

Since A is a symplectic map, the polynomial $h(t) = \frac{1-P_A(t)}{t}$ has integer coefficients. Therefore, for all $f \in \mathbb{Z}[t]$ the polynomial $g = f \circ h$ has integer coefficients as well. Notice that this polynomial satisfies $g(\lambda_i) = f(\lambda_i^{-1})$ for all eigenvalues. Hence, if $\beta \in \mathfrak{D}$ such that $\beta = f(\lambda)$, then $\beta^* = g(\lambda) \in \mathfrak{D}$ as well.

The second part is straightforward; indeed, if $\beta = f(\lambda) \in \mathfrak{D}$, then

$$\omega(\vec{n}\iota(\beta), \vec{m}) = \omega(\vec{n} f(A), \vec{m}) = \omega(\vec{n}, \vec{m} f(A^{-1})) = \omega(\vec{n}, \vec{m}\iota(\beta^*)). \qquad \square$$

COROLLARY 2.2. *For any $\beta \in \mathfrak{D}$, the matrix $\iota(\beta)$ is symplectic if and only if $\beta\beta^* = 1$. Furthermore, for any integer $M > 1$, if $\beta\beta^* \equiv 1 \pmod{M\mathfrak{D}}$, then $\iota(\beta)$ is symplectic modulo M.*

Define a "norm map" $\mathcal{N} : \mathfrak{D} \to \mathfrak{D}$ sending $\beta \mapsto \beta\beta^*$. Given an integer $M > 1$, the inclusion $\iota : \mathfrak{D} \hookrightarrow \mathrm{Mat}(2d, \mathbb{Z})$ induces a map $\iota_M : \mathfrak{D}/M\mathfrak{D} \to \mathrm{Mat}(2d, \mathbb{Z}/MZ)$, and the norm map $\mathcal{N}$ induces a well defined map $\mathcal{N}_M : (\mathfrak{D}/M\mathfrak{D})^* \to (\mathfrak{D}/M\mathfrak{D})^*$. The norm map is multiplicative; hence the map $\mathcal{N}_M$ is a group homomorphism and its kernel correspond to symplectic matrices. Consequently,

$$\iota_M(\ker \mathcal{N}_M) \subseteq \mathrm{Sp}(2d, \mathbb{Z}/M\mathbb{Z}),$$

is a commutative subgroup of symplectic matrices that commute with A modulo M. We are now ready to define the Hecke group.

Definition 2.3. Define the Hecke group

$$C_A(N) = \begin{cases} \{\iota_N(\beta) | \beta \in \ker \mathcal{N}_N\} & N \text{ odd} \\ \{\iota_{2N}(\beta) | \beta \in \ker \mathcal{N}_{2N},\ \beta \equiv \pm 1 \pmod 4\} & N \text{ even.} \end{cases}$$

Now take the Hecke operators to be $U_N(B),\ B \in C_A(N)$.

Remark 2.3. Note that if $A \not\equiv \pm I \pmod 4$ and N is even, then $U_N(A)$ is not one of the Hecke operators. Nevertheless, Corollary 1.5 ensures that it still commutes with all of them.

2.2. *Galois orbits and invariant subspaces.* The structure of the Hecke group $C_A(N)$ is closely related to the decomposition of the rational vector space $\mathbb{Q}^{2d} = \bigoplus E_\theta$ into irreducible invariant subspaces under the left action of A. We now make a slight detour and describe this decomposition in terms of Galois orbits of the eigenvalues of A.

Let $\Lambda_{\mathbb{Q}}$ denote the set of eigenvalues of A and $G_{\mathbb{Q}}$ the absolute Galois group. The group $G_{\mathbb{Q}}$ acts on $\Lambda_{\mathbb{Q}}$, and we denote by $\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$ the set of Galois orbits. Since the matrix A is symplectic, if $\lambda \in \Lambda_{\mathbb{Q}}$ is an eigenvalue, then $\lambda^{-1} \in \Lambda_{\mathbb{Q}}$ as well. To each orbit $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$, there is a unique orbit θ^* such that $\lambda \in \theta \Leftrightarrow \lambda^{-1} \in \theta^*$. If $\theta = \theta^*$ we say that the orbit is symmetric and otherwise nonsymmetric. For any orbit θ we define the symplectic orbit $\bar{\theta} = \theta \cup \theta^*$.

PROPOSITION 2.4. *There is a unique decomposition into irreducible left invariant subspaces*: $\mathbb{Q}^{2d} = \bigoplus_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} E_\theta$.

- *To each orbit $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$, there is a corresponding subspace* (*denoted by E_θ*), *such that the eigenvalues of the restriction $A_{|E_\theta}$ are the eigenvalues $\lambda \in \theta$.*
- *For any two orbits θ, θ', unless $\theta' = \theta^*$, E_θ and $E_{\theta'}$ are orthogonal with respect to the symplectic form.*
- *Let $\vec{v}_{\theta^*}$ be a left eigenvector for A with eigenvalue in θ^*. Then*, *the projection of $\vec{n}$ to E_θ with respect to the above decomposition vanishes*, *if and only if $\omega(\vec{n}, \vec{v}_{\theta^*}) = 0$.*

Proof. Appendix A, Lemma A.1, and Corollary A.4. □

Remark 2.4. There is an alternative way to describe this decomposition, using the characteristic polynomial P_A of A. Any invariant irreducible subspace corresponds to an irreducible factor of P_A (which is an integral by Gauss's lemma). The roots of this irreducible factor are then precisely the eigenvalues in the Galois orbit. As a consequence we can deduce that the product of all eigenvalues in one orbit is an integer that divides 1 and can thus be only ± 1 (for a symmetric orbit, by definition, the product is always $+1$).

For each symplectic orbit $\bar{\theta}$ we define the space $E_{\bar{\theta}} = E_\theta + E_{\theta^*}$. Proposition 2.4 then implies that for θ symmetric $E_\theta = E_{\bar{\theta}}$ is a symplectic space (i.e., the restriction of the symplectic form to this subspace is nondegenerate), while for θ nonsymmetric the spaces E_θ, E_{θ^*} are both isotropic (i.e., the restriction of the symplectic form vanishes) and $E_{\bar{\theta}} = E_\theta \oplus E_{\theta^*}$ is again symplectic.

2.3. *Reduction to Galois orbits.* Consider the action of the absolute Galois group $G_{\mathbb{Q}}$, on the set of eigenvalues $\Lambda_{\mathbb{Q}} = \{\lambda_1, \dots, \lambda_{2d}\}$. For each orbit $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$ fix a representative λ_θ (for nonsymmetric orbits we take $\lambda_{\theta^*} = \lambda_\theta^{-1}$). Let $K_\theta = \mathbb{Q}(\lambda_\theta)$ be field extensions and $\mathbb{O}_{K_\theta}$ the corresponding integral rings. For any symmetric orbit, λ_θ and λ_θ^{-1} are Galois conjugates. Consequently, if we denote by $F_\theta = \mathbb{Q}(\lambda_\theta + \lambda_\theta^{-1})$, then K_θ/F_θ are quadratic field extensions.

Note that every element $\beta \in \mathfrak{D}$ is uniquely determined by its components on each orbit $\beta_\theta \in \mathbb{Z}[\lambda_\theta] \subseteq \mathbb{O}_{K_\theta}$ (because if $\lambda_i = \lambda_\theta^\sigma$ for some $\sigma \in G_{\mathbb{Q}}$, then $\beta_i = f(\lambda_i) = f(\lambda_\theta^\sigma) = \beta_\theta^\sigma$). We can thus identify the ring $\mathfrak{D}$ as a subring of $\prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathbb{O}_{K_\theta}$.

LEMMA 2.5. *The norm map $\mathcal{N}$ acts on a component corresponding to a symmetric orbit θ, through the corresponding field extension norm map, $\mathcal{N}_{K_\theta/F_\theta}$, and on a component corresponding to a nonsymmetric orbit θ by $\beta_\theta \mapsto \beta_\theta\beta_{\theta^*}$.*

Proof. Let $\beta \in \mathfrak{D}$, then $\beta_\theta = f(\lambda_\theta)$ for some $f \in \mathbb{Z}[t]$. For any orbit θ,

$$(\mathcal{N}(\beta))_\theta = f(\lambda_\theta) f(\lambda_\theta^{-1}).$$

When the orbit θ is symmetric, this is precisely $\mathcal{N}_{K_\theta/F_\theta}(\beta_\theta))$, and when it is nonsymmetric, it is $\beta_\theta\beta_{\theta^*}$. □

LEMMA 2.6. *There is $s \in \mathbb{N}$, such that*

$$s \prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathbb{O}_{K_\theta} \subseteq \mathfrak{D} \subseteq \prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathbb{O}_{K_\theta}.$$

Proof. The rings $\mathbb{O}_{K_\theta}$ are isomorphic (as $\mathbb{Z}$ modules) to $\mathbb{Z}^{|\theta|}$, so $\prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathbb{O}_{K_\theta} \cong \mathbb{Z}^{2d}$. On the other hand $\mathfrak{D} \cong \mathbb{Z}[t]/P_A \cong \mathbb{Z}^{2d}$ as well (again as $\mathbb{Z}$ modules). The result is now immediate since any subgroup of $\mathbb{Z}^{2d}$ with the same rank satisfies this property. □

We can now estimate the number of Hecke operators.

LEMMA 2.7. *The number of elements in $C_A(N)$ satisfies*

$$N^{d-\varepsilon} \ll_\varepsilon |C_A(N)| \ll_\varepsilon N^{d+\varepsilon}$$

Proof. To simplify the discussion, we will assume that there are no rational isotropic invariant rational subspaces (i.e., all orbits are symmetric). The Hecke group (for N even) is a subgroup of $\iota_{2N}(\ker \mathcal{N}_{2N})$ with index bounded by 2^{d^2}; it is thus sufficient to show that for all N,

$$N^{d-\varepsilon} \ll |\ker \mathcal{N}_N| \ll N^{d+\varepsilon}.$$

For each orbit $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$, the norm map $\mathcal{N}_{K_\theta/F_\theta}$ induces a map on the group of invertible elements

$$\mathcal{N}_{N\mathbb{O}_{F_\theta}} : (\mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta})^* \to (\mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta})^*.$$

Let $\mathcal{C}(N\mathcal{O}_{F_\theta})$ be the kernel of this map. For any $\beta \in \mathfrak{D}$, denote by $\bar{\beta} \in \mathfrak{D}/N\mathfrak{D}$ its class modulo $N\mathfrak{D}$, by β_θ its component in $\mathcal{O}_{K_\theta}$, and by $\bar{\beta}_\theta$ the class of β_θ modulo $N\mathcal{O}_{K_\theta}$. Then the map $\bar{\beta} \mapsto \bar{\beta}_\theta$ is well defined (because if $\beta \in N\mathfrak{D}$, then obviously $\beta_\theta \in N\mathcal{O}_{K_\theta}$), and by Lemma 2.6, the map

$$\mathfrak{D}/N\mathfrak{D} \to \prod_{\theta\in\Lambda} \mathcal{O}_{K_\theta}/N\mathcal{O}_{K_\theta}$$
$$\bar{\beta} \mapsto (\bar{\beta}_\theta)_\theta$$

has kernel and co-kernel of order bounded by $|\mathfrak{D}/s\mathfrak{D}| = s^{2d}$. Furthermore, the restriction of this map to the multiplicative group and to the subgroup of norm one elements also has bounded kernel and co-kernel. Thus, it suffices to show that $\forall\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$

$$N^{d_\theta-\varepsilon} \ll |\mathcal{C}(N\mathcal{O}_{F_\theta})| \ll N^{d_\theta+\varepsilon},$$

where $d_\theta = \frac{|\theta|}{2} = [F_\theta : \mathbb{Q}]$. This is the estimate on the number of norm one elements in the ring $\mathcal{O}_{K_\theta}/N\mathcal{O}_{K_\theta}$ which is proved in Appendix B (Proposition B.3). □

Remark 2.5. If there are invariant rational isotropic subspaces, the proof is analogous. For any symplectic orbit $\bar{\theta} = \theta \cup \theta^*$ corresponding to a nonsymmetric orbit, instead of evaluating the number of elements in $\mathcal{C}(N\mathcal{O}_{K_\theta})$, one needs to evaluate the size of $(\mathcal{O}_{K_\theta}/N\mathcal{O}_{K_\theta})^*$ and show $N^{d_\theta-\varepsilon} \ll |(\mathcal{O}_{K_\theta}/N\mathcal{O}_{K_\theta})^*| \ll N^{d_\theta+\varepsilon}$, where now $d_\theta = \frac{|\bar{\theta}|}{2} = |\theta|$.

2.4. *Additional structure.* So far we have identified a set of commuting integral matrices with the commutative ring $\mathfrak{D}$. We are now going to identify the action of these matrices on $\mathbb{Z}^{2d}$, with the action of $\mathfrak{D}$ on an appropriate ideal $\mathcal{I}$. This identification allows us to think of both the matrices and the lattice points on which they act as elements of the same space $\mathfrak{D}$.

For every orbit $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$, take a left eigenvector $\vec{v}_\theta$ with eigenvalue λ_θ^{-1} and coefficients in $s\mathcal{O}_{K_\theta}$. Therefore $\vec{v} = (\vec{v}_\theta)_\theta$ is a (left) eigenvector with coefficients in $\prod s\mathcal{O}_{K_\theta} \subseteq \mathfrak{D}$, such that $\vec{v}\iota(\beta^*) = \beta\vec{v}$. Define the map $\iota^* : \mathbb{Z}^{2d} \to \mathfrak{D}$, by $\iota^*(\vec{n}) = \omega(\vec{n}, \vec{v})$, and the ideal $\mathrm{Im}(\iota^*) = \mathcal{I} \subseteq \mathfrak{D}$. To see that $\mathcal{I}$ is indeed an ideal, notice that if $\nu = \iota^*(\vec{n}) \in \mathcal{I}$ and $\beta \in \mathfrak{D}$ with $B = \iota(\beta)$, then

$$\beta\nu = \beta\iota^*(\vec{n}) = \beta\omega(\vec{n}, \vec{v}) = \omega(\vec{n}, \vec{v}\iota(\beta^*)) = \omega(\vec{n}\iota(\beta), \vec{v}) = \iota^*(\vec{n}B),$$

so $\beta\nu \in \mathcal{I}$ as well. Furthermore, by the third part of Proposition 2.4, we see that $(\iota^*(\vec{n}))_\theta = 0$ if and only if the projection of $\vec{n}$ to E_θ vanishes. In particular $\iota^*(\vec{n}) = 0$ implies $\vec{n} = 0$ and the map $\iota^* : \mathbb{Z}^{2d} \to \mathcal{I}$ is an isomorphism of $\mathbb{Z}$ modules.

Now, for any integer $M \in \mathbb{N}$, the map ι^* induces a group isomorphism $\iota_M^* : (\mathbb{Z}/M\mathbb{Z})^{2d} \to \mathcal{I}/M\mathcal{I}$. This map is compatible with the map $\iota_M : \mathfrak{D}/M\mathfrak{D} \to$

$\mathrm{Mat}(2d,\mathbb{Z}/M\mathbb{Z})$, in the sense that for any $B=\iota_M(\bar{\beta})$ and $\vec{n}\in(\mathbb{Z}/N\mathbb{Z})^{2d}$, we have $\iota_M^*(\vec{n}B)=\bar{\beta}\iota_M^*(\vec{n})$ in $\mathcal{I}/M\mathcal{I}$.

2.5. *Hecke eigenfunctions.* Since all of the Hecke operators commute with $U_N(A)$, they act on its eigenspaces, and since they commute with each other, there is a basis of joint eigenfunctions of $U_N(A)$ and the Hecke operators. Such a basis is called a Hecke basis. We now show that the Hecke symmetries cancel most of the degeneracies in the spectrum of $U_N(A)$, implying that the Hecke basis is essentially unique.

The action of the Hecke group on the Hilbert space $\mathcal{H}_N$ induces a decomposition into joint eigenspaces

$$\mathcal{H}_N=\bigoplus_{\chi}\mathcal{H}_\chi,$$

where χ runs over the characters of the Hecke group.

PROPOSITION 2.8. *The dimension of any Hecke eigenspace satisfies*

$$\dim\mathcal{H}_\chi\ll_\varepsilon N^\varepsilon.$$

Proof. Again, for simplicity we will assume all orbits are symmetric. The operator

$$\mathcal{P}_\chi=\frac{1}{|C_A(N)|}\sum_{C_A(N)}\chi(B)^{-1}U_N(B))$$

is a projection operator to the eigenspace $\mathcal{H}_\chi$. Consequently, the dimension of $\mathcal{H}_\chi$ is given by its trace, $\dim\mathcal{H}_\chi=\mathrm{Tr}(\mathcal{P}_\chi)$.

By Corollary 1.6, for any $B\in C_A(N)$,

$$|\mathrm{Tr}(U_N(B))|\leq 2^d\sqrt{\ker_N(B-I)}.$$

Note that while for even N the operator $U_N(B)$ depends on B modulo $2N$, this bound only depends on B modulo N. Hence if $B=\iota_N(\beta)\pmod{N}$, then using the identification $\iota_N^*:\mathbb{Z}/N\mathbb{Z}\to\mathcal{I}/N\mathcal{I}$ we can write this bound as

$$|\mathrm{Tr}(U_N(B)|\leq 2^d\sqrt{\#\{\nu\in\mathcal{I}/N\mathcal{I}|\nu(\beta-1)\equiv 0\pmod{N\mathcal{I}}\}}.$$

Since both $\mathfrak{D}$ and the ideal $\mathcal{I}$ are isomorphic (as $\mathbb{Z}$ modules) to $\mathbb{Z}^{2d}$, there is $s'\in\mathbb{N}$ such that $s'\mathfrak{D}\subseteq\mathcal{I}\subseteq\mathfrak{D}$ and we can replace $\mathcal{I}/N\mathcal{I}$ by $\mathfrak{D}/N\mathfrak{D}$ to get

$$|\mathrm{Tr}(U_N(B)|\leq(2s')^d\sqrt{\#\{\nu\in\mathfrak{D}/N\mathfrak{D}|\nu(\beta-1)\equiv 0\pmod{N\mathfrak{D}}\}}.$$

We now replace the sum over $C_A(N)$ with a sum over $\ker(\mathcal{N}_N)$ in the bound $\dim(\mathcal{H}_\chi)\leq\frac{1}{|C_A(N)|}\sum_{C_A(N)}|\mathrm{Tr}(U_N(A))|$ (losing at most a constant factor) to get the bound

$$\dim(\mathcal{H}_\chi)\ll\frac{1}{|C_A(N)|}\sum_{\beta\in\ker\mathcal{N}_N}\sqrt{\#\{\nu\in\mathfrak{D}/N\mathfrak{D}|\nu(\beta-1)\in N\mathfrak{D}\}}.$$

Because the map $\ker(\mathcal{N}_N) \to \prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathcal{C}(N\mathbb{O}_{F_\theta})$ has bounded kernel (as in the proof of Lemma 2.7), after multiplying by some bounded constant we can replace $\mathfrak{D}/N\mathfrak{D}$ with $\prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta}$ and the sum over the group $\ker \mathcal{N}_N$, to a sum over $\prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} \mathcal{C}(N\mathbb{O}_{F_\theta})$ to get

$$\dim(\mathcal{H}_\chi) \ll \frac{1}{|C_A(N)|} \prod_{\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}} S_1(N\mathbb{O}_{K_\theta}),$$

where

$$S_1(N\mathbb{O}_{F_\theta}) = \sum_{\beta \in \mathcal{C}(N\mathbb{O}_{F_\theta})} \sqrt{\#\left\{\nu \in \mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta} | \nu(\beta - 1) = 0\right\}}.$$

It now suffices to show that $\forall \theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}, \quad S_1(N\mathbb{O}_{F_\theta}) \ll_\varepsilon N^{d_\theta + \varepsilon}$ (recall $\frac{1}{C_A(N)} = O_\varepsilon(N^{-d+\varepsilon})$. This is a counting argument on elements in the ring $\mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta}$ that is proved in Appendix B (Proposition B.6). □

Remark 2.6. The proof in the nonsymmetric case is analogous. For any nonsymmetric orbit one needs to bound sums of the form

$$\sum_{(\mathbb{O}_{F_\theta}/N\mathbb{O}_{F_\theta})^*} \sqrt{\#\left\{\nu_1, \nu_2 \in \mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta} | \nu_2(\beta - 1) = \nu_2(\beta^{-1} - 1) = 0\right\}}.$$

This can be done using the same methods.

3. **Arithmetic quantum unique ergodicity**

This section is devoted to proving Theorem 3. We fix a matrix $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ with distinct eigenvalues and no invariant isotropic rational subspaces and show that for any smooth observable $f \in C^\infty(\mathbb{T}^{2d})$, the expectation values for $\mathrm{Op}_N(f)$ in any Hecke eigenfunction ψ, satisfy:

$$|\langle \mathrm{Op}_N(f)\psi, \psi\rangle| \ll_{\varepsilon, f} N^{-\frac{d(f)}{4}+\varepsilon},$$

where $d(f) = \min_{\hat{f}(\vec{n}) \neq 0} d_{\vec{n}}$, and $2d_{\vec{n}}$ is the dimension of the smallest invariant subspace containing $\vec{n}$.

Much of the proof goes along the lines of [23]. The first step is to make a reduction to a theorem regarding elementary observables. Next, we show that it is sufficient to bound the fourth moment of the matrix elements (after restricting the elementary operator to an appropriate subspace). Finally, we use averaging over the Hecke group to transform the moment calculation into a counting problem which is then solved using the connection of the Hecke group with the groups $\mathcal{C}(N\mathbb{O}_{F_\theta}) \subseteq (\mathbb{O}_{K_\theta}/N\mathbb{O}_{K_\theta})^*$.

3.1. *Reduction to elementary observables.* In order to prove Theorem 3, it is sufficient to prove it for elementary observables of the form $\mathrm{Op}_N(e_{\vec{n}}),\ 0 \neq \vec{n} \in \mathbb{Z}^{2d}$; that is, to show that the following theorem holds:

THEOREM 8. *Let* $0 \neq \vec{n} \in \mathbb{Z}^{2d}$ *and let* ψ *be an eigenfunction of all the Hecke operators. Then, the diagonal matrix elements satisfy*

$$|\langle \widetilde{T}_N(\vec{n})\psi, \psi\rangle| \ll_\varepsilon \left\|\vec{n}\right\|^{4d^2} N^{-d_{\vec{n}}/4+\varepsilon}.$$

The proof of Theorem 3 from Theorem 8 is immediate, due to the rapid decay of the Fourier coefficients.

Remark 3.1. The estimate in Theorem 8 is in fact valid also when there are invariant rational isotropic subspaces, as long as $\vec{n}$ is not contained in any of these subspaces.

3.2. *Reduction to a moment calculation.* In order to prove Theorem 8, we estimate the fourth moment of the diagonal matrix elements in a Hecke basis,

$$\sum_\psi |\langle \widetilde{T}_N(\vec{n})\psi, \psi\rangle|^4.$$

However, when summing over all the Hecke eigenfunctions, the fourth moment is of order $N^{d-2d_{\vec{n}}}$ (where $2d_{\vec{n}}$ is the dimension of the smallest (symplectic) invariant subspace containing $\vec{n}$). Thus, when $\vec{n}$ is contained in an invariant subspace of dimension $\leq d$, we cannot use this method directly to bound the size of the individual matrix elements. Instead, we would like to make the sum only over a subset of the Hecke eigenfunctions. For that purpose, for each $\vec{n} \in \mathbb{Z}^{2d}$ and Hecke eigenfunction ψ, we introduce a subspace $\mathcal{H}_{\vec{n},\psi} \subseteq \mathcal{H}_N$, invariant under the action of $\widetilde{T}_N(\vec{n})$ and the Hecke operators. Theorem 8 is then proved by estimating the fourth moment for the restriction of $\widetilde{T}_N(\vec{n})$ to $\mathcal{H}_{\vec{n},\psi}$.

Let $\vec{n} \in \mathbb{Z}^{2d}$. Recall the decomposition into irreducible invariant subspaces $\mathbb{Q}^{2d} = \bigoplus E_\theta$ described in Section 2.2, and let $\Lambda_{\vec{n}}/G_{\mathbb{Q}}$ be the set of orbits $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$ for which the projection of $\vec{n}$ to E_θ vanishes. Denote by $E_{\vec{n}}$ the minimal invariant subspace containing $\vec{n}$. We can decompose $E_{\vec{n}} = \sum_{\theta \notin \Lambda_{\vec{n}}/G} E_\theta$, and, in particular,

$$2d_{\vec{n}} = \dim E_{\vec{n}} = \sum_{\theta \notin \Lambda_{\vec{n}}/G_{\mathbb{Q}}} 2d_\theta,$$

where $2d_\theta = \dim E_\theta = |\theta|$ (recall that E_θ is of even dimension because it is symplectic).

Define the lattice $Z_{\vec{n}} = E_{\vec{n}} \cap \mathbb{Z}^{2d}$. Then by the third part of Proposition 2.4, we have $Z_{\vec{n}} = \left\{\vec{m} \in \mathbb{Z}^{2d} | \iota^*(\vec{m})_\theta = 0, \ \forall \theta \in \Lambda_{\vec{n}}/G_{\mathbb{Q}}\right\}$.

Definition 3.1. For $\psi \in \mathcal{H}_N$ a Hecke eigenfunction, define the subspace $\mathcal{H}_{\vec{n},\psi} \subseteq \mathcal{H}_N$ to be the minimal subspace containing ψ and invariant under the action of all $\widetilde{T}_N(\vec{m})$, $\vec{m} \in Z_{\vec{n}}$.

LEMMA 3.2. *The space* $\mathcal{H}_{\vec{n},\psi}$ *is invariant under the action of the Hecke operators.*

Proof. For $B = \iota_{2N}(\bar{\beta}) \in C_A(N)$ and $\vec{m} \in Z_{\vec{n}}$, let $\vec{m}' = \vec{m}\iota(\beta)$, where $\beta \in \mathfrak{O}$ is a representative of $\bar{\beta}$. Then $\vec{m}' \equiv \vec{m}B \pmod{2N}$ and $\vec{m}' \in Z_{\vec{n}}$ (because $\forall\theta \in \Lambda_{\vec{n}}/G_{\mathbb{Q}}$, $\omega(\vec{m}', \vec{v}_\theta) = \beta_\theta\omega(\vec{m}, \vec{v}_\theta) = 0$). Now, if $\phi' = U_N(B)\phi \in U_N(B)\mathcal{H}_{\vec{n},\psi}$ (for some $\phi \in \mathcal{H}_{\vec{n},\psi}$), then

$$\widetilde{T}_N(\vec{m})\phi' = U_N(B)\widetilde{T}_N(\vec{m}B)\phi = U_N(B)\widetilde{T}_N(\vec{m}')\phi;$$

hence $\widetilde{T}_N(\vec{m})\phi' \in U_N(B)\mathcal{H}_{\vec{n},\psi}$ as well (because $\mathcal{H}_{\vec{n},\psi}$ is invariant under $\widetilde{T}_N(\vec{m}')$). Therefore, the space $U_N(B)\mathcal{H}_{\vec{n},\psi}$ contains ψ and is invariant under the action of $\widetilde{T}_N(\vec{m})$, $\forall\vec{m} \in Z_{\vec{n}}$. Thus, from the minimality condition $\mathcal{H}_{\vec{n},\psi} \subseteq U_N(B)\mathcal{H}_{\vec{n},\psi}$, and since $U_N(B)$ is invertible, we have $U_N(B)\mathcal{H}_{\vec{n},\psi} = \mathcal{H}_{\vec{n},\psi}$. □

When $\Lambda_{\vec{n}} = \varnothing$, then $Z_{\vec{n}} = \mathbb{Z}^{2d}$ and $\mathcal{H}_{\vec{n},\psi} = \mathcal{H}_N$. Otherwise it is a proper subspace and we can give an estimate for its dimension.

PROPOSITION 3.3. *The dimension of the subspace $\mathcal{H}_{\vec{n},\psi}$ satisfies*

$$\dim(\mathcal{H}_{\vec{n},\psi}) \ll_\varepsilon N^{d_{\vec{n}}+\varepsilon},$$

where the implied constant does not depend on $\vec{n}$ or on ψ.

Proof. Consider a subgroup of the Hecke group

$$C_0(N) = \left\{\iota_N(\bar{\beta}) \in C_A(N) | \beta_\theta = 1,\ \forall\theta \notin \Lambda_{\vec{n}}/G_{\mathbb{Q}}\right\},$$

in the sense that there is a representative $\beta \in \mathfrak{O} \subseteq \prod \mathbb{O}_{K_\theta}$ satisfying this condition. Notice that this group acts trivially on $Z_{\vec{n}}$ modulo $2N$ (i.e., $\forall B \in C_0(N)$ and $\forall\vec{m} \in Z_{\vec{n}}$, $\vec{m}B \equiv \vec{m} \pmod{2N}$).

Let $\chi(B)$, $B \in C_A(N)$, be the eigenvalues corresponding to ψ, and consider the subspace

$$\mathcal{H}^0_\chi = \{\phi \in \mathcal{H}_N | U_N(B)\phi = \chi(B)\phi,\ \forall B \in C_0(N)\}.$$

Since $C_0(N)$ acts trivially on $Z_{\vec{n}}$, then $\forall\vec{m} \in Z_{\vec{n}}$ and $B \in C_0(N)$, $U_N(B)\widetilde{T}_N(\vec{m}) = \widetilde{T}_N(\vec{m})U_N(B)$ commute; hence $\mathcal{H}^0_\chi$ is invariant under $\widetilde{T}_N(\vec{m})$, $\forall\vec{m} \in Z_{\vec{n}}$. Obviously $\psi \in \mathcal{H}^0_\chi$; hence, from minimality, $\mathcal{H}_{\vec{n},\psi} \subseteq \mathcal{H}^0_\chi$ and it suffices to bound the dimension of $\mathcal{H}^0_\chi$.

The eigenspace $\mathcal{H}^0_\chi$ decomposes into joint eigenspaces of all the Hecke operators

$$\mathcal{H}^0_\chi = \bigoplus \mathcal{H}_{\chi'},$$

where the sum is only on characters χ' that identify with χ on $C_0(N)$. Note that $\chi'_{|C_0(N)} = \chi_{|C_0(N)}$ implies that they differ by a character of $C_A(N)/C_0(N)$ (and vice-versa). Therefore (by Proposition 2.8) the dimension

$$\dim\mathcal{H}^0_\chi = \sum_{\chi'|_{C_0}=\chi|_{C_0}} \dim H_{\chi'} \ll_\varepsilon N^\varepsilon[C_A(N) : C_0(N)].$$

Following the lines of the proof of Lemma 2.7, one can show $[C_A(N):C_0(N)] \ll_\varepsilon N^{d_{\vec{n}}+\varepsilon}$, concluding the proof. Notice that the implied constants depend only on the set of orbits $\Lambda_{\vec{n}}/G_{\mathbb{Q}}$. But, as there are at most 2^d possibilities for such subsets, we can take the same constant for all the spaces $\mathcal{H}_{\vec{n},\psi}$. □

For $\vec{m} \in Z_{\vec{n}}$, denote by $\tilde{T}^0_N(\vec{m})$ the restriction of $\tilde{T}_N(\vec{m})$ to $\mathcal{H}_{\vec{n},\psi}$. Then, similar to the original operators, the trace of the restricted operators vanishes for sufficiently large N.

LEMMA 3.4. *There is $r \in \mathbb{N}$ (depends only on $Z_{\vec{n}}$) such that for any $\vec{m} \in Z_{\vec{n}}$,*

$$|\mathrm{Tr}(\tilde{T}^0_N(\vec{m}))| \leq \begin{cases} \dim \mathcal{H}_{\vec{n},\psi} & \vec{m} \equiv 0 \pmod{N'} \\ 0 & \textit{otherwise} \end{cases},$$

where $N' = \frac{N}{\gcd(N,r^2)}$.

Proof. Recall that the space $E_{\vec{n}}$ is a symplectic subspace. Let $\{e_i, f_i\}$ be a symplectic basis (i.e., $\omega(e_i, f_j) = \delta_{i,j}$ and $\omega(e_i, e_j) = \omega(f_i, f_j) = 0$), and let $r \in \mathbb{Z}$ such that $re_i, rf_i \in \mathbb{Z}^{2d}$. Fix $\vec{m} \in E_{\vec{n}}$, and consider the decomposition $\vec{m} = \sum_{i=1}^{d_{\vec{n}}}(a_i e_i + b_i f_i)$. Then $ra_i = \omega(\vec{m}, rf_i)$ and $rb_i = -\omega(\vec{m}, re_i)$ are integers.

Notice that for all $i = 1, \dots, d_{\vec{n}}$,

$$\mathrm{Tr}(\tilde{T}^0_N(\vec{m})) = \mathrm{Tr}(\tilde{T}^0_N(-rf_i)\tilde{T}^0_N(\vec{m})\tilde{T}^0_N(rf_i)) = e_N(ra_i)\mathrm{Tr}(\tilde{T}^0_N(\vec{m})),$$

and by a similar argument $\mathrm{Tr}(\tilde{T}^0_N(\vec{m})) = e_N(rb_i)\mathrm{Tr}(TN^0(\vec{m}))$. Consequently, if $ra_i \not\equiv 0 \pmod{N}$ or $rb_i \not\equiv 0 \pmod{N}$, then $\mathrm{Tr}(\tilde{T}^0_N(\vec{m})) = 0$. On the other hand, if $\forall i$, $ra_i \equiv rb_i \equiv 0 \pmod{N}$, then $r^2\vec{m} = \sum_{i=1}^{d_{\vec{n}}}(ra_i re_i + rb_i rf_i) \equiv 0 \pmod{N}$ and $\vec{m} \equiv 0 \pmod{N'}$. □

Remark 3.2. The integer r in the above lemma, depends only on the lattice $Z_{\vec{n}}$, that is determined by the subset $\Lambda_{\vec{n}}/G_{\mathbb{Q}}$. We can thus take r to be the same for all $\vec{n}$ (by taking the lcm for the 2^d possibilities).

The Hecke operators act on the space $\mathcal{H}_{\vec{n},\psi}$ so there is a basis $\{\psi_i\}$ of joint eigenfunctions of all the Hecke operators (we can assume $\psi_1 = \psi$). To prove Theorem 8, we will prove a stronger statement regarding the fourth moment of matrix elements in this basis.

PROPOSITION 3.5. *Let $\{\psi_i\}$ be a basis for $\mathcal{H}_{\vec{n},\psi}$ composed of joint eigenfunctions of all the Hecke operators. Then, the fourth moment satisfies*

$$\sum_i |\langle \tilde{T}_N(\vec{n})\psi_i, \psi_i\rangle|^4 \ll_\varepsilon \left\|\vec{n}\right\|^{16d_{\vec{n}}^2} N^{-d_{\vec{n}}+\varepsilon}.$$

The proof of Theorem 8 from Proposition 3.5 is now immediate. The first element in the sum is obviously bounded by the whole sum, so

$$|\langle \tilde{T}_N(\vec{n})\psi,\psi\rangle|^4 \ll_\varepsilon \left\|\vec{n}\right\|^{16d_{\vec{n}}^2} N^{-d_{\vec{n}}+\varepsilon} \leq \left\|\vec{n}\right\|^{16d^2} N^{-d_{\vec{n}}+\varepsilon}.$$

3.3. *Reduction to a counting problem.* We now reduce Proposition 3.5 into a counting problem which is then solved in the following section.

PROPOSITION 3.6. *Let $\{\psi_i\}$ be a basis for $\mathcal{H}_{\vec{n},\psi}$ composed of joint eigenfunctions of all the Hecke operators. Then, the fourth moment,*

$$\sum_i |\langle \tilde{T}_N(\vec{n})\psi_i,\psi_i\rangle|^4$$

is bounded by $\frac{\dim \mathcal{H}_{\vec{n},\psi}}{|C_A(N)|^4}$ times the number of solutions to

$$\vec{n}(B_1-B_2+B_3-B_4)\equiv 0 \pmod{N'},\ B_i\in C_A(N),$$

where N' is as in Lemma 3.4.

Proof. Define an operator $D=D(\vec{n})$, acting on $\mathcal{H}_{\vec{n},\psi}$ through averaging over the Hecke group:

$$D=\frac{1}{|C_A(N)|}\sum_{B\in C_A(N)} \tilde{T}_N^0(\vec{n}B).$$

Recall that for $B\in C_A(N)$, there is $\vec{m}\in Z_{\vec{n}}$ such that $\vec{n}B\equiv\vec{m} \pmod{2N}$, so this is indeed well defined. The identity $\tilde{T}_N(\vec{n}B)=U_N(B)^*\tilde{T}_N(\vec{n})U_N(B)$ implies $\langle D(\vec{n})\psi_i,\psi_i\rangle=\langle\tilde{T}_N(\vec{n})\psi_i,\psi_i\rangle$. Since for any complex matrix $D=(d_{i,j})$, $\sum_i |d_{i,i}|^4\leq \mathrm{Tr}((DD^*)^2)$, it is sufficient to bound $\mathrm{Tr}((DD^*)^2)$. Now, expand $(DD^*)^2$ as a product of four sums, and take trace (using Lemma 3.4) to get the result. □

By Proposition 2.8 and Lemma 2.7, we know that

$$\frac{\dim \mathcal{H}_{\vec{n},\psi}}{|C_A(N)|^4}\ll_\varepsilon \frac{1}{N^{4d-d_{\vec{n}}-\varepsilon}}.$$

Therefore, in order to prove Proposition 3.5 from Proposition 3.6, it remains to show that the number of solutions to

$$\vec{n}(B_1-B_2+B_3-B_4)\equiv 0 \pmod{N'},\quad B_i\in C_A(N) \tag{3.1}$$

is bounded by $O(\left\|\vec{n}\right\|^{16d_{\vec{n}}^2}N^{4d-2d_{\vec{n}}+\varepsilon})$.

3.4. *Counting solution.* We now bound the number of solutions to (3.1), thus completing the proof of Theorem 3.

PROPOSITION 3.7. *The number of solution to* (3.1) *is bounded by*

$$O_\varepsilon\left(\left\|\vec{n}\right\|^{16d_{\vec{n}}^2}N^{4d-2d_{\vec{n}}+\varepsilon}\right).$$

Proof. Let $\nu = \iota^*(\vec{n}) \in \mathcal{I}$. Then the number of solutions to (3.1) is the same as the number of solutions to

$$\nu(\beta_1 - \beta_2 + \beta_3 - \beta_4) \equiv 0 \pmod{N'\mathcal{I}}, \quad \beta_i \in \ker \mathcal{N}_{2N}. \tag{3.2}$$

In the same way as in the proof of Proposition 2.8, it is sufficient to bound for each θ the number of solutions to

$$\nu_\theta(\beta_1 - \beta_2 + \beta_3 - \beta_4) \equiv 0 \pmod{N'\mathcal{O}_{K_\theta}}, \quad \beta_i \in \mathcal{C}(N\mathcal{O}_{F_\theta}), \tag{3.3}$$

the product of which gives the number of solutions to (3.2) up to some bounded constant.

If $\theta \in \Lambda_{\vec{n}}/G_{\mathbb{Q}}$, then $\nu_\theta = 0$ and the best bound is the trivial bound of $|\mathcal{C}(N\mathcal{O}_{F_\theta})|^4 = O_\varepsilon(N^{4d_\theta+\varepsilon})$. Otherwise, $0 \neq \mathcal{N}_{K_\theta/\mathbb{Q}}(\nu_\theta) \in \mathbb{Z}$ and the number of solutions to (3.3) is bounded by the number of solutions to

$$\beta_1 - \beta_2 + \beta_3 - \beta_4 \equiv 0 \pmod{M\mathcal{O}_{K_\theta}}, \quad \beta_i \in \mathcal{C}(N\mathcal{O}_{F_\theta}) \tag{3.4}$$

where $M = \frac{N'}{\gcd(N', \mathcal{N}_{K_\theta/\mathbb{Q}}(\nu_\theta))}$. The natural map $\mathcal{C}(N\mathcal{O}_{F_\theta}) \to \mathcal{C}(M\mathcal{O}_{F_\theta})$ has kernel of order at most $(\frac{N}{M})^{2d_\theta} \leq (r|\mathcal{N}_{K_\theta/\mathbb{Q}}(\nu_\theta)|)^{2d_\theta} \ll \left\|\vec{n}\right\|^{4d_\theta^2}$; hence the number of solutions to (3.4) is bounded by $\left\|\vec{n}\right\|^{16d_\theta^2}$ times the number of solutions to

$$\beta_1 - \beta_2 + \beta_3 - \beta_4 \equiv 0 \pmod{M\mathcal{O}_{K_\theta}}, \quad \beta_i \in \mathcal{C}(M\mathcal{O}_{F_\theta}). \tag{3.5}$$

Equation (3.5) is invariant under the action of the Galois group $\mathrm{Gal}(K_\theta/F_\theta)$. We thus get a second equation,

$$\beta_1^{-1} - \beta_2^{-1} + \beta_3^{-1} - \beta_4^{-1} \equiv 0 \pmod{M\mathcal{O}_{K_\theta}}, \quad \beta_i \in \mathcal{C}(M\mathcal{O}_{F_\theta}). \tag{3.6}$$

The set of equations ((3.5), (3.6)) is equivalent to the following set of equations (see [23, Lemma 15]):

$$\begin{cases} (\beta_3 - \beta_1)(\beta_3 - \beta_2)(\beta_1 + \beta_2) = 0 & \pmod{M\mathcal{O}_{K_\theta}} \\ \beta_4 = \beta_1 - \beta_2 + \beta_3 = 0 & \pmod{M\mathcal{O}_{K_\theta}}. \end{cases} \tag{3.7}$$

Since β_4 is determined by $\beta_1, \beta_2, \beta_3$, ignoring the second equation only increases the number of solutions. Finally, the number of solutions to the first equation is bounded by $|\mathcal{C}(M\mathcal{O}_{F_\theta})| S_2(M\mathcal{O}_{F_\theta})$ where $S_2(M\mathcal{O}_{F_\theta})$ is the number of solutions to

$$(1 - \beta_1)(1 - \beta_2)(\beta_1 + \beta_2) = 0 \pmod{M\mathcal{O}_{K_\theta}}, \quad \beta_i \in \mathcal{C}(M\mathcal{O}_{F_\theta}), \tag{3.8}$$

that satisfies $S_2(M\mathcal{O}_{F_\theta}) = O(M^{d_\theta+\varepsilon})$ (Proposition B.8).

To conclude, for $\theta \in \Lambda_{\vec{n}}/G_{\mathbb{Q}}$ the number of solutions to (3.3) is bounded by $O(N^{4d_\theta+\varepsilon})$. Otherwise, it is bounded by $O(\left\|\vec{n}\right\|^{16d_\theta^2} N^{2d_\theta+\varepsilon})$. Therefore, since

$\sum_{\theta\notin\Lambda_{\vec{n}}/G_{\mathbb{Q}}} d_\theta = d_{\vec{n}}$, the number of solutions to (3.1) is bounded by

$$O\Big(\prod_{\theta\in\Lambda_{\vec{n}}/G_{\mathbb{Q}}} N^{4d_\theta+\varepsilon} \prod_{\theta\notin\Lambda_{\vec{n}}/G_{\mathbb{Q}}} \|\vec{n}\|^{16d_\theta^2} N^{2d_\theta+\varepsilon}\Big) = O\Big(\|\vec{n}\|^{16d_{\vec{n}}^2} N^{4d-2d_{\vec{n}}+\varepsilon}\Big).$$

□

4. Hecke theory tor prime N

In the following section we restrict the discussion to the case where $N = p$ is a large prime. For this case, the structure of the Hecke group (hence also the behavior of Hecke eigenfunctions and matrix elements) is determined by the decomposition of the vector space $\mathbb{F}_p^{2d}$ (rather than $\mathbb{Q}^{2d}$) into irreducible invariant subspaces. This decomposition can be described using the Frobenius orbits of the eigenvalues of A. Analyzing the action of A on $\mathbb{F}_p^{2d}$ enables us to obtain much sharper results from the ones presented above for composite N.

The main difference between composite and prime N is that instead of integral rings (we used in §2), here we work with finite fields so that the counting arguments become sharp. For example, we can precisely describe the structure of the Hecke group (Lemma 4.2) and obtain sharp bounds for the dimension of the joint Hecke eigenspaces (Proposition 4.12). Compare this to Lemma 2.7 and Proposition 2.8 obtained for composite N.

4.1. *Hecke operators.* Let $A \in \mathrm{Sp}_\theta(2d,\mathbb{Z})$ be a matrix with distinct eigenvalues. Fix a large prime $N = p > \Delta(P_A)$ (the discriminant of the characteristic polynomial). Then we can think of A also as an element of $\mathrm{Sp}(2d,\mathbb{F}_p)$ with distinct eigenvalues. In fact, in order to ensure that $A \pmod p$ has distinct eigenvalues, it is sufficient to assume that $\Delta(P_A) \not\equiv 0 \pmod p$. For $A \in \mathrm{Sp}(2d,\mathbb{F}_p)$ with distinct eigenvalues, the centralizer of A (in the symplectic group) is a commutative subgroup $C_p(A) \subseteq \mathrm{Sp}(2d,\mathbb{F}_p)$, and we can take the Hecke operators to be $U_p(B),\ B \in C_p(A)$. Note that p is odd; hence the operator $U_p(B)$ depends only on B modulo p and this definition of the Hecke operators makes sense.

Remark 4.1. When $N = p$ is a prime ≥ 5, the map $B \mapsto U_p(B)$ (which is a representation of $\mathrm{Sp}(2d,\mathbb{F}_p)$) identifies with the celebrated Weil representation of the symplectic group over the finite field $\mathbb{F}_p$. Consequently, the Hecke operators can be obtained by restricting the Weil representation to a maximal torus. These representations are described at length in [13], and we follow the same lines in our analysis.

4.2. *Reduction to irreducible orbits.* Let Λ_p/G_p denote the Frobenius orbits of the eigenvalues of A modulo p. To each orbit $\vartheta \in \Lambda_p/G_p$ (with representative λ_ϑ) denote by ϑ^* the orbit of λ_ϑ^{-1} and by $\bar{\vartheta} = \vartheta \cup \vartheta^*$ the symplectic orbit. We say that an orbit ϑ is symmetric if $\vartheta = \vartheta^*$ and nonsymmetric otherwise. Denote by $\Lambda_p / \pm G_p$ the set of symplectic orbits and let

$$\mathbb{F}_p^{2d} = \bigoplus_{\Lambda_p/\pm G_p} E_{\bar{\vartheta}}$$

be the orthogonal decomposition into invariant irreducible symplectic subspaces (see Appendix A for more details). For each symplectic orbit $\bar{\vartheta} \in \Lambda_p / \pm G_p$, let $2d_{\bar{\vartheta}} = \dim(E_{\bar{\vartheta}}) = |\bar{\vartheta}|$ denote the dimension of the corresponding subspace.

Remark 4.2. Note that while this decomposition is similar to the decomposition of $\mathbb{Q}^{2d}$ into invariant symplectic subspaces described in Proposition 2.4, they are not the same. The relation between the two decompositions is described in Section 5.2.

To each invariant subspace $E_{\bar{\vartheta}}$, take a symplectic basis. For any $\vec{n} \in \mathbb{F}_p^{2d}$, let $\vec{n}_{\bar{\vartheta}} \in \mathbb{F}_p^{2d_{\bar{\vartheta}}}$ be the projection of $\vec{n}$ to $E_{\bar{\vartheta}}$ in the symplectic basis. Since the decomposition is orthogonal, then for any $\vec{n}, \vec{m} \in \mathbb{F}_p^{2d}$

$$\omega(\vec{m}, \vec{n}) = \sum_{\Lambda_p/\pm G_p} \omega(\vec{m}_{\bar{\vartheta}}, \vec{n}_{\bar{\vartheta}}). \tag{4.1}$$

We thus get an embedding,

$$\prod \mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p) \hookrightarrow \mathrm{Sp}(2d, \mathbb{F}_p), \tag{4.2}$$

through the action of each factor on the corresponding subspace. Denote by $\mathcal{S} \subseteq \mathrm{Sp}(2d, \mathbb{F}_p)$ the image of $\prod \mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p)$. For each $B \in \mathcal{S}$, let $B_{\bar{\vartheta}} \in \mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p)$ denote the restriction of B to $E_{\bar{\vartheta}}$ in the symplectic basis. In order to keep track of dimensions, we denote by $\tilde{T}_p^{(d)}(\cdot), U_p^{(d)}(\cdot)$, the quantized elementary operators and propagators for $\mathbb{T}^{2d}$.

PROPOSITION 4.1. *There is a unitary map*

$$\mathcal{U} : L^2(\mathbb{F}_p^d) \to \bigotimes_{\Lambda_p/\pm G_p} L^2(\mathbb{F}_p^{d_{\bar{\vartheta}}}),$$

such that

(1) *For any* $\vec{n} \in \mathbb{F}_p^{2d}$,

$$\mathcal{U}\tilde{T}_p^{(d)}(\vec{n})\mathcal{U}^{-1} = \bigotimes_{\Lambda_p/\pm G_p} \tilde{T}_p^{(d_{\bar{\vartheta}})}(\vec{n}_{\bar{\vartheta}}).$$

(2) *For any* $B \in \mathcal{S}$,

$$\mathcal{U}U_p^{(d)}(B)\mathcal{U}^{-1} = \bigotimes_{\Lambda_p/\pm G_p} U_p^{(d_{\bar{\vartheta}})}(B_{\bar{\vartheta}}).$$

Proof. Define $T_p^{\otimes}(\vec{n}) = \bigotimes_{\Lambda_p/\pm G_p} \tilde{T}_p^{(d_{\bar{\vartheta}})}(\vec{n}_{\bar{\vartheta}})$. It is easily verified from (4.1) that $T_p^{\otimes}(\vec{n})$ obey the same commutation relation as in Proposition 1.1. Therefore, there is a unitary map $\mathcal{U}$ such that $\mathcal{U}\tilde{T}_p^{(d)}(\vec{n})\mathcal{U}^{-1} = T_p^{\otimes}(\vec{n})$ for all $\vec{n} \in \mathbb{F}_p^{2d}$.

As for the second part, recall that $U_p^{(d_{\bar{\vartheta}})}(B_{\bar{\vartheta}})$ all satisfy the intertwining equation and from the first part $\mathcal{U}\tilde{T}_p^{(d)}(\vec{n})\mathcal{U}^{-1} = T_p^{\otimes}(\vec{n})$. Consequently, if we define $\tilde{U}_p(B) = \mathcal{U}^{-1} \bigotimes_{\Lambda_p/\pm G_p} U_p^{(d_{\bar{\vartheta}})}(B_{\bar{\vartheta}})\mathcal{U}$, then $\tilde{U}_p(B)$ is also an intertwining operator:

$$\tilde{U}_p(B)^{-1}\tilde{T}_p^{(d)}(\vec{n})\tilde{U}_p(B) = \tilde{T}_p^{(d)}(\vec{n}B).$$

Thus, from uniqueness of the quantization, the operators $\tilde{U}_p(B)$ and $U_p^{(d)}(B)$ differ by a character of $\mathcal{S}$ (recall that the quantization is multiplicative). Finally, since $\mathcal{S} \cong \prod \mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p)$ has no nontrivial multiplicative characters,

$$\mathcal{U}^{-1}U_p^{(d)}(B)\mathcal{U} = \bigotimes_{\Lambda_p/\pm G_p} U_p^{(d_{\bar{\vartheta}})}(B_{\bar{\vartheta}}). \qquad \square$$

Notice that any element in $B \in C_p(A)$ leaves the spaces $E_{\bar{\vartheta}}$ invariant; hence $C_p(A) \subseteq \mathcal{S}$. Let $C_p(A_{\bar{\vartheta}}) \subset \mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p)$ be the centralizer of $A_{\bar{\vartheta}}$ in $\mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p)$. Then the embedding (4.2) induces an isomorphism

$$\prod_{\Lambda_p/\pm G_p} C_p(A_{\bar{\vartheta}}) \to C_p(A). \tag{4.3}$$

We can thus recover the quantization of any element in $B \in C_p(A)$ from the tensor product of the quantization of corresponding elements $B_{\bar{\vartheta}} \in C_p(A_{\bar{\vartheta}})$.

We now want to look at the quantization of $A_{\bar{\vartheta}}$ together with its centralizer $C_p(A_{\bar{\vartheta}}) \subset \mathrm{Sp}(2d_{\bar{\vartheta}}, \mathbb{F}_p)$ for one irreducible symplectic orbit $\bar{\vartheta} \in \Lambda_p/\pm G_p$. For the rest of this section, the orbit $\bar{\vartheta}$ will be fixed and for notational convenience, the subscript will be omitted.

4.3. *Irreducible orbit.* Let $A \in \mathrm{Sp}(2d, \mathbb{F}_p)$ be a matrix with $2d$ distinct eigenvalues such that there is only one irreducible symplectic orbit (symmetric or nonsymmetric). We now look at the quantization of A together with its centralizer $C_p(A)$.

Remark 4.3. For a symplectic matrix $A \in \mathrm{Sp}(2d, \mathbb{Z})$, $d \geq 2$, the requirement that $A \pmod p$ has only one irreducible orbit cannot hold for all primes. However, for a two-dimensional matrix $A \in \mathrm{SL}(2, \mathbb{Z})$ this is indeed the case (since for any $A \in \mathrm{SL}(2, \mathbb{F}_p)$ there could be only one orbit). The distinction between symmetric and nonsymmetric orbits in this case correspond to inert and splitting primes respectively (cf. [8], [24]).

4.3.1. *Identification with finite fields.* We now identify the action of the Hecke group $C_p(A)$ on the vector space $\mathbb{F}_p^{2d}$ with the action (of a multiplicative subgroup) of the finite field $\mathbb{F}_{p^{2d}}^*$ on itself by multiplication. Compare this to the identification $\iota_N^* : (\mathbb{Z}/N\mathbb{Z})^{2d} \to \mathcal{I}/N\mathcal{I}$ that we defined in Section 2.4 (note that the identification of the Hecke group here is more precise from the inclusions we used in the proof of Lemma 2.7 to estimate the number of Hecke operators).

Take a pair of eigenvalues λ, λ^{-1} in a field extension of $\mathbb{F}_p$. If we denote by $q = p^d$, then in the symmetric case $\mathbb{F}_p(\lambda) = \mathbb{F}_{q^2}$ and in the nonsymmetric $\mathbb{F}_p(\lambda) = \mathbb{F}_q$. Let $\vec{v}, \vec{v}^*$ be eigenvectors for λ, λ^{-1} respectively. In the symmetric case, where the eigenvalues are Galois conjugates, $\tau(\lambda) = \lambda^{-1}$, we take $\vec{v}^* = \tau(\vec{v})$ to be Galois conjugates as well. By Lemma A.3, in the nonsymmetric case (respectively symmetric), the map

$$(\nu_1, \nu_2) \mapsto \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(\nu_1 \vec{v}) + \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(\nu_2 \vec{v}^*)$$

(respectively $\nu \mapsto \mathrm{Tr}_{\mathbb{F}_{q^2}/\mathbb{F}_p}(\nu \vec{v})$) is an isomorphism from $\mathbb{F}_q \oplus \mathbb{F}_q$ (respectively $\mathbb{F}_{q^2}$) to $\mathbb{F}_p^{2d}$. By Lemma A.5, under this identification,

$$\omega(\vec{n}, \vec{m}) = \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(2\kappa(\mu\nu^* - \nu\mu^*)), \tag{4.4}$$

where $\nu = \omega(\vec{n}, \vec{v}^*), \nu^* = \omega(\vec{n}, \vec{v}),\ \mu = \omega(\vec{m}, \vec{v}^*), \mu^* = \omega(\vec{m}, \vec{v})$, and $\kappa = (2\omega(\vec{v}, \vec{v}^*))^{-1}$.

This identification with finite fields, enables us to identify the centralizer as a subgroup of the multiplicative group $\mathbb{F}_{q^2}^*$ and to identify the orbits of elements in $\mathbb{F}_p^{2d}$ under the action of the centralizer.

LEMMA 4.2. *In the symmetric case, $C_p(A) \cong \ker(\mathcal{N}_{\mathbb{F}_{q^2}/\mathbb{F}_q})$, while in the nonsymmetric case, $C_p(A) \cong \mathbb{F}_q^*$.*

Proof. First for the symmetric case. For any $B \in C_p(A)$ the vectors $\vec{v}, \vec{v}^*$ are eigenvectors with eigenvalues $\beta, \beta^{-1} \in \mathbb{F}_{q^2}$. Therefore, the action of $B \in C_p(A)$ on $\mathbb{F}_{q^2}$ is given by

$$\nu = \omega(\vec{n}, \vec{v}^*) \mapsto \omega(\vec{n}B, \vec{v}^*) = \omega(\vec{n}, \vec{v}^* B^{-1}) = \beta\nu.$$

On the other hand, any element $\beta \in \mathbb{F}_{q^2}$ defines (by multiplication) a linear transformation on $\mathbb{F}_{q^2}$ that commutes with the action of A. Given formula (4.4) for the symplectic form, the condition for the action of $\beta \in \mathbb{F}_{q^2}$ to be symplectic is precisely that $\beta\tau(\beta) = 1$. We can thus identify $C_p(A)$ with the norm one elements in $\mathbb{F}_{q^2}/\mathbb{F}_q$.

For the nonsymmetric case, the action of $C_p(A)$ on $\mathbb{F}_q \oplus \mathbb{F}_q$ is given by $(\nu_1, \nu_2) \mapsto (\beta\nu_1, \beta^{-1}\nu_2)$. Here any element $(\beta_1, \beta_2) \in \mathbb{F}_q \times \mathbb{F}_q$ defines a linear action that commutes with the action of A, and the elements that preserve the symplectic form are precisely the elements (β, β^{-1}). We can thus identify these elements with $\mathbb{F}_q^*$. □

COROLLARY 4.3. *For* $\vec{n} \in \mathbb{F}_p^{2d}$ *define* $\mathfrak{Q}(\vec{n}) = \omega(\vec{n}, \vec{v})\omega(\vec{n}, \vec{v}^*) \in \mathbb{F}_q$. *Let* $\vec{n}, \vec{m} \in \mathbb{F}_p^{2d}$. *If* $\mathfrak{Q}(\vec{n}) = \mathfrak{Q}(\vec{m}) \neq 0$, *then there is* $B \in C_p(A)$ *s.t.* $\vec{n}B = \vec{m}$.

Proof. We use the identification with finite fields. In the symmetric case, let $\nu = \omega(\vec{n}, \vec{v}^*)$ and $\mu = \omega(\vec{m}, \vec{v}^*)$. We thus need to find $\beta \in \ker \mathcal{N}_{\mathbb{F}_{q^2}/\mathbb{F}_q}$, such that $\beta\nu = \mu$. The requirement that $\mathfrak{Q}(\vec{n}) = \mathfrak{Q}(\vec{m}) \neq 0$ implies that $\mathcal{N}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(\mu\nu^{-1}) = 1$ and we can take $\beta = \mu\nu^{-1}$.

In the nonsymmetric case, denote $(\nu, \nu^*) = (\omega(\vec{n}, \vec{v}^*), \omega(\vec{n}, \vec{v}))$ and $(\mu, \mu^*) = (\omega(\vec{m}, \vec{v}^*), \omega(\vec{m}, \vec{v}))$. Then the requirement $\mathfrak{Q}(\vec{n}) = \mathfrak{Q}(\vec{m}) \neq 0$, implies that $\frac{\nu}{\mu} = \frac{\mu^*}{\nu^*}$. Set $\beta = \mu\nu^{-1}$, then $(\beta\nu, \beta^{-1}\nu^*) = (\mu, \mu^*)$. □

Remark 4.4. Notice that the converse is obviously true; that is, if $\vec{m} = \vec{n}B$ for some $B \in C_p(A)$ then $\mathfrak{Q}(\vec{n}) = \mathfrak{Q}(\vec{m})$.

4.3.2. *Hecke eigenspaces.* Consider the quantization of an irreducible element $A \in \mathrm{Sp}(2d, \mathbb{F}_p)$, together with its centralizer $C_p(A)$. To any character χ of $C_p(A)$ let $\mathcal{H}_\chi$ denote the corresponding eigenspace. Both in the symmetric and nonsymmetric cases, the centralizer is a cyclic group of even order. Therefore, there is a unique quadratic character of $C_p(A)$, that we will denote by χ_2.

PROPOSITION 4.4. *For any character* $\chi \neq \chi_2$, $\dim \mathcal{H}_\chi = 1$. *In the symmetric case, the character* χ_2 *does not appear in the decomposition, and in the nonsymmetric case* $\dim \mathcal{H}_{\chi_2} = 2$.

Proof. Consider the projection operator

$$\mathcal{P}_\chi = \frac{1}{|C_p(A)|} \sum_{B \in C_p(A)} \chi^{-1}(B) U_p(B).$$

The dimension of the corresponding eigenspace is then given by its trace

$$\dim(\mathcal{H}_\chi) = \mathrm{Tr}(\mathcal{P}_\chi) = \frac{1}{|C_p(A)|} \sum_{B \in C_p(A)} \chi^{-1}(B) \mathrm{Tr}(U_p(B)). \tag{4.5}$$

From Corollary 1.6 we have that,

$$\left| Tr(U_p(B)) \right| = \sqrt{|\ker(B - I)|}.$$

For any $B \in C_p(A)$, all eigenvalues are Galois conjugates and their inverses; hence, 1 is an eigenvalue of B if and only if $B = I$. Therefore, for all $I \neq B \in C_p(A)$, $\left| Tr(U_p(B)) \right| = 1$ (and obviously $\mathrm{Tr}(U_p(I)) = p^d = q$).

In the symmetric case, $C_p(A)$ is isomorphic to the norm one elements in $\mathbb{F}_{q^2}/\mathbb{F}_q$ and hence of order $q+1$. We can thus bound

$$\dim \mathcal{H}_\chi \leq \frac{2q}{q+1} < 2,$$

but since the dimension is an integer, $\dim(\mathcal{H}_\chi) \leq 1$. Finally, there are $q+1$ characters and $\dim \mathcal{H}_p = q$, so q characters appear with multiplicity one. For now, denote the character that does not appear by $\tilde{\chi}_0$.

In the nonsymmetric case, $|C_p(A)| = q-1$ and the corresponding bound is

$$\dim \mathcal{H}_\chi \leq \frac{2q-2}{q-1} = 2.$$

However, this inequality is actually an equality only if there is no cancellation in the sum (4.5); that is,

$$\forall I \neq B \in C_p(A),\ \chi(B) = \mathrm{Tr}(U_p(B)).$$

Such an equality can hold for at most one character. Thus, for any other character there is a strict inequality and $\dim \mathcal{H}_\chi \leq 1$. Now, from dimension consideration we can deduce that there is a character with multiplicity 2 (denoted again by $\tilde{\chi}_0$) and that all the other characters appear with multiplicity one.

We now show that in both cases $\tilde{\chi}_0$ is the quadratic character. Notice that for a cyclic group of even order the product of all the characters is the quadratic character. Therefore, for any $B \in C_p(A)$ the determinant of $U_p(B)$ is $\chi_2(B)\tilde{\chi}_0(B)^{-1}$ in the symmetric case and $\chi_2(B)\tilde{\chi}_0(B)$ in the nonsymmetric. But since $\mathrm{Sp}(2d, \mathbb{F}_p)$ has no nontrivial characters, then $\forall B \in \mathrm{Sp}(2d, \mathbb{F}_p)$, $\det(U_p(B)) = 1$, and $\tilde{\chi}_0 = \chi_2$. □

Since for $B \in C_p(A) - \{1\}$ the sum over all the characters vanish, the trace of $U_p(B)$ is $-\chi_2(B)$ in the symmetric case and $\chi_2(B)$ in the nonsymmetric. Consequently, we can find the constant in formula (1.11).

COROLLARY 4.5. *For any* $B \in C_p(A)$,

$$U_p(B) = \pm\frac{\chi_2(B)}{q} \sum_{\vec{n} \in \mathbb{F}_p^d} \widetilde{T}_p(\vec{n})\widetilde{T}_p(-\vec{n}B),$$

where the minus sign is for the symmetric case and the plus sign is for the nonsymmetric.

4.3.3. *Explicit formulas and exponential sums.* We now show that the matrix elements of elementary operators can be written explicitly as exponential sums. In [16] Gurevich and Hadani observed that matrix elements of elementary observables could be expressed as $\mathrm{Tr}(\widetilde{T}_p(\vec{n})\mathcal{P}_\chi)$ (where $\mathcal{P}_\chi$ is the projection operator to the corresponding eigenspace). Using this observation, together with the formula for the propagator (Corollary 4.5), we obtain explicit formulas for the matrix elements.

Denote by $e_q(x) = e_p(\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(x))$ the corresponding additive character of $\mathbb{F}_q$. For notational convenience, we will denote by $\mathscr{C} \cong C_p(A)$, the group of norm one elements in $\mathbb{F}_{q^2}/\mathbb{F}_q$ in the symmetric case and the multiplicative group of $\mathbb{F}_q$ in the nonsymmetric.

Definition 4.6. For any character χ of $\mathscr{C}$ and any element $\nu \in \mathbb{F}_q$, define the exponential sum:

$$E_q(\nu, \chi) = \frac{1}{|\mathscr{C}|} \sum_{1 \neq x \in \mathscr{C}} e_q\left(\nu\kappa \frac{x+1}{x-1}\right) \chi\chi_2(x),$$

where $\kappa = (2\omega(\vec{v}, \vec{v}^*))^{-1}$ (note that in the symmetric case, $\kappa\frac{x+1}{x-1} \in \mathbb{F}_q$ so this is well defined).

PROPOSITION 4.7. *Let* $0 \neq \vec{n} \in \mathbb{F}_p^{2d}$ *and* $\tilde{T}_p(\vec{n})$, *the corresponding elementary operator. Let* $\mathscr{Q}(\vec{n}) = \omega(\vec{n}, \vec{v})\omega(\vec{n}, \vec{v}^*) \in \mathbb{F}_q$, *as in Corollary* 4.3. *Let* ψ *be a joint eigenfunction*, *with corresponding character* χ. *Then*, *when* $\chi \neq \chi_2$ *is not the quadratic character* (*relevant only in the nonsymmetric case*),

$$\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle = \pm E_q(\mathscr{Q}(\vec{n}), \chi),$$

where the minus sign is for the symmetric case and plus for the nonsymmetric.

Proof. Since the joint eigenspaces are one-dimensional, an alternative way to write the matrix element is:

$$\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle = \mathrm{Tr}(\tilde{T}_p(\vec{n})\mathscr{P}_\chi),$$

where $\mathscr{P}_\chi = \frac{1}{|C_p(A)|}\sum_{C_p(A)} \chi^{-1}(B)U_p(B)$ is the projection operator to $\mathscr{H}_\chi$ [16]. Plugging in the formula for $U_p(B)$ (Corollary 4.5), gives

$$\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle = \frac{\pm 1}{q|C_p(A)|} \sum_{C_p(A)} \chi^{-1}\chi_2(B) \sum_{\mathbb{F}_p^{2d}} \mathrm{Tr}(\tilde{T}_p(\vec{n})\tilde{T}_p(\vec{m})\tilde{T}_p(-\vec{m}B))$$

(where the minus sign is for the symmetric case). Notice that when $\vec{n} = \vec{m}(B - I)$,

$$\mathrm{Tr}(\tilde{T}_p(\vec{n})\tilde{T}_p(\vec{m})\tilde{T}_p(-\vec{m}B)) = qe_p\left(\frac{p+1}{2}\omega(\vec{n}, \vec{m})\right),$$

and that otherwise the trace vanishes. Therefore, when $B = I$ we get no contribution from the inner sum, and otherwise the only contribution is from $\vec{m} = \vec{n}(B - I)^{-1}$. Consequently,

$$\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle = \frac{\pm 1}{|C_p(A)|} \sum_{C_p(A)\setminus\{I\}} \chi^{-1}\chi_2(B)e_p\left(\frac{p+1}{2}\omega(\vec{n}, \vec{n}(B - I)^{-1})\right).$$

We now use the identification with finite fields described in Section 4.3.1. Replace the sum over the elements in the centralizer with a sum over the elements in $\mathscr{C}$ and for the symplectic form use formula (4.4). Consequently, the formula for the matrix elements now takes the form

$$\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle = \frac{\pm 1}{|\mathscr{C}|}\sum_{\mathscr{C}-\{1\}} e_q\left(\mathscr{Q}(\vec{n})\kappa\frac{1+\beta}{1-\beta}\right)\chi\chi_2(\beta^{-1}).$$

Changing summation variable $x = \beta^{-1}$ concludes the proof. □

We note that in both the symmetric and nonsymmetric cases for $\chi \neq \chi_2$, the Riemann Hypothesis for curves over finite fields implies the bound $|E_q(\nu, \chi)| \leq \frac{2}{\sqrt{q}} + O(\frac{1}{q})$ (see e.g. [25, Ch. 6] or [22]). We can thus deduce:

COROLLARY 4.8. *For any* $0 \neq \vec{n} \in \mathbb{F}_p^{2d}$ *and any* $\psi \in \mathscr{H}_\chi$ *with* $\chi \neq \chi_2$,

$$|\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle| \leq \frac{2}{\sqrt{q}} + O\left(\frac{1}{q}\right).$$

Remark 4.5. Note that for $d = 1$, this gives an alternative proof of the Kurlberg-Rudnick rate conjecture originally proved by Gurevich and Hadani [16].

For the quadratic character (in the nonsymmetric case), $E_q(\mathscr{Q}(\vec{n}), \chi_2)$ is no longer a formula for the corresponding matrix element, but rather for

$$\mathrm{Tr}(\tilde{T}_p(\vec{n})|_{\mathscr{H}_{\chi_2}}) = \langle \tilde{T}_p(\vec{n})\psi_0, \psi_0\rangle + \langle \tilde{T}_p(\vec{n})\psi_1, \psi_1\rangle,$$

where $\{\psi_0, \psi_1\}$ is an orthonormal basis for $\mathscr{H}_{\chi_2}$. Nevertheless, in this case we can find formulas for the eigenfunctions and use them to bound the individual matrix elements.

LEMMA 4.9. *In the nonsymmetric case, there is a normalized eigenfunction* $\psi_0 \in \mathscr{H}_{\chi_2}$, *such that*

$$\langle \tilde{T}_p(\vec{n})\psi_0, \psi_0\rangle = \begin{cases} 0 & \omega(\vec{n}, \vec{v}) \neq 0 \\ 1 & \omega(\vec{n}, \vec{v}) = 0. \end{cases}$$

Furthermore, if $\mathscr{Q}(\vec{n}) \neq 0$ *then for any normalized* $\psi \in \mathscr{H}_{\chi_2}$,

$$|\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle| \leq \frac{2}{\sqrt{q}}.$$

Proof. We use a similar construction to the eigenfunctions constructed by Degli Esposti, Graffi and Isola for two-dimensional cat maps for splitting primes [8].

In the nonsymmetric case, there is a decomposition $\mathbb{F}_p^{2d} = E \oplus E^*$ into two invariant Lagrangian subspaces. Therefore, there is $M \in \mathrm{Sp}(2d, \mathbb{F}_p)$ such that for any $B \in C_p(A)$, $M^{-1}BM = \begin{pmatrix} \tilde{B}^t & 0 \\ 0 & \tilde{B}^{-1} \end{pmatrix}$. Consequently (by formula (1.8)), the functions $\psi_0 = \sqrt{q}U_p(M)\delta_0$ and $\psi_1 = \sqrt{\frac{q}{q-1}}U_p(M)(1-\delta_0)$ are two orthonormal joint eigenfunctions of $U_p(B)$, $B \in C_p(A)$ with the same eigenvalues, and hence a basis for $\mathscr{H}_{\chi_2}$.

Denote by $T_{i,j} = \langle \tilde{T}_p(\vec{n})\psi_i, \psi_j\rangle$. If we denote $\vec{m} = \vec{n}M$, then (by the intertwining equation)

$$T_{0,0} = \langle \tilde{T}_p(\vec{n})\psi_0, \psi_0\rangle = q\langle \tilde{T}_p(\vec{m})\delta_0, \delta_0\rangle.$$

By Lemma A.3, the projection of $\vec{n}$ to the Lagrangian subspace E vanishes (i.e., $\vec{m} = (0, \vec{m}_2)$), if and only if $\omega(\vec{n}, \vec{v}^*) = 0$. Now calculate directly:

$$T_{0,0} = \sum_{\vec{x}} e_p\left(\frac{1}{2}\vec{m}_1 \cdot \vec{m}_2\right) e_p(\vec{m}_2 \cdot \vec{x})\delta_0(\vec{x} + \vec{m}_1)\delta_0(\vec{x}).$$

Therefore, $T_{0,0} = 0$ if $\vec{m}_1 \neq 0$ and $T_{0,0} = 1$ if $\vec{m}_1 = 0$.

When $\mathcal{Q}(\vec{n}) \neq 0$, the projections to both Lagrangian subspaces do not vanish. By a similar computation, one can show that $T_{1,0}$ and $T_{0,1}$ are bounded by $\frac{1}{\sqrt{q-1}}$, and that $T_{1,1}$ is bounded by $\frac{2}{q-1}$. Therefore, since any normalized $\psi \in \mathcal{H}_{\chi_2}$ is of the form $\psi = a_0\psi_0 + a_1\psi_1$, with $|a_0|^2 + |a_1|^2 = 1$, we have

$$|\langle \tilde{T}_p(\vec{n})\psi, \psi\rangle| \leq \sum_{i,j=0}^{1} |a_i a_j T_{i,j}| \leq \frac{2}{\sqrt{q}}. \qquad \square$$

4.3.4. *Moments.* Let $A \in \mathrm{Sp}(2d, \mathbb{F}_p)$ be a matrix with one irreducible symplectic orbit (symmetric or nonsymmetric), and fix $\vec{n} \in \mathbb{F}_p^{2d}$ with $\mathcal{Q}(\vec{n}) \neq 0$. Let $\{\psi_i\}$ be an orthonormal basis of joint eigenfunctions of $C_p(A)$. The different matrix elements $\langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle$ fluctuate around their average

$$\frac{1}{q}\sum_i \langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle = \frac{1}{q}\mathrm{Tr}(\tilde{T}_p(\vec{n})) = 0.$$

Remark 4.6. In the nonsymmetric case, for $0 \neq \vec{n} \in \mathbb{F}_p^{2d}$ such that $\mathcal{Q}(\vec{n}) = 0$, Proposition 4.7 implies that for all characters $\chi \neq \chi_2$ the corresponding matrix elements are identical (and equal $-\frac{1}{p-1}$), so that the fluctuations are trivial.

In [24] Kurlberg and Rudnick gave a conjecture regarding the limiting distribution of these fluctuations (for $d = 1$). Considering that in the formula for the matrix elements (Proposition 4.7), the dimension d only determines the ground field $\mathbb{F}_q = \mathbb{F}_{p^d}$. We can reformulate their conjecture to predict the fluctuations of the corresponding exponential sums (formulated here as Conjecture 5). We now calculate (asymptotically) the second and fourth moments and show agreement with this conjecture.

PROPOSITION 4.10. *Let $\vec{n}, \vec{m} \in \mathbb{F}_p^{2d}$ with $\mathcal{Q}(\vec{n}), \mathcal{Q}(\vec{m}) \neq 0$. Then the mixed second moment satisfies*

$$\frac{1}{q}\sum_i \langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle \overline{\langle \tilde{T}_p(\vec{m})\psi, \psi\rangle} = \begin{cases} \frac{1}{q} + O(\frac{1}{q^2}) & \mathcal{Q}(\vec{n}) = \mathcal{Q}(\vec{m}) \\ O(\frac{1}{q^2}) & \mathcal{Q}(\vec{n}) \neq \mathcal{Q}(\vec{m}). \end{cases}$$

Proof. First, we can replace the sum over eigenfunction to a sum over characters and the matrix element by corresponding exponential sums. By Lemma 4.9, the error that comes from the quadratic character is bounded by $O(\frac{1}{q^2})$ (recall that $\mathcal{Q}(\vec{n}), \mathcal{Q}(\vec{m}) \neq 0$). Now, since the sum over the characters $\chi(x)$ vanish unless $x = 1$,

$$\frac{1}{q}\sum_{\chi} E_q(\mathcal{Q}(\vec{n}), \chi)\overline{E_q(\mathcal{Q}(\vec{m}), \chi)} = \frac{1}{q|\mathscr{C}|}\sum_{x\neq 1} e_q\left((\mathcal{Q}(\vec{n}) - \mathcal{Q}(\vec{m}))\kappa\frac{x+1}{x-1}\right).$$

If $\mathcal{Q}(\vec{n}) = \mathcal{Q}(\vec{m})$ we indeed get $\frac{|\mathscr{C}|-1}{q|\mathscr{C}|} = \frac{1}{q} + O(\frac{1}{q^2})$. Otherwise, note that the map $x \mapsto \frac{x+1}{x-1}$ is injective; hence, the sum is over $q-2$ distinct points in $\mathbb{F}_q$ (or q in the symmetric case) and is therefore bounded by $O(\frac{1}{q^2})$. □

PROPOSITION 4.11. *For $\vec{n} \in \mathbb{F}_p^{2d}$ with $\mathcal{Q}(\vec{n}) \neq 0$, the fourth moment satisfies*

$$\frac{1}{q}\sum_{i} |\langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle|^4 = \frac{2}{q^2} + O\left(\frac{1}{q^{5/2}}\right).$$

Proof. We follow the same lines as in the proof of Proposition 3.5. Consider the averaged operator

$$D = \frac{1}{|\mathscr{C}|}\sum_{B\in C_p(A)} \tilde{T}_p(\vec{n}B).$$

Then, for any eigenfunction ψ_i, the diagonal matrix elements are the same $\langle D\psi_i, \psi_i\rangle = \langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle$, and for any two eigenfunctions ψ_i, ψ_j, corresponding to different characters, the corresponding off diagonal terms vanish $\langle D\psi_i, \psi_j\rangle = 0$. Consequently,

$$\frac{1}{q}\sum_{i} |\langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle|^4 = \frac{1}{q}\mathrm{Tr}((DD^*)^2) + O\left(\frac{1}{q^3}\right),$$

where the error comes from the eigenfunctions corresponding to the quadratic character.

We can calculate $\mathrm{Tr}((DD^*)^2)$ differently by writing it as a product of four sums and then taking its trace, recalling that

$$\mathrm{Tr}(\tilde{T}_p(\vec{n})\tilde{T}_p(\vec{m})) = \begin{cases} q & \vec{n}+\vec{m} = 0 \\ 0 & \vec{n}+\vec{m} \neq 0. \end{cases}$$

Define the set $X = \{B_1, \dots, B_4 \in C_p(A) | \vec{n}(B_1 - B_2 + B_3 - B_4) = 0\}$; then this calculation gives

$$\frac{1}{q}\mathrm{Tr}((DD^*)^2) = \frac{1}{|C_p(A)|^4}\sum_{X} e_p\left(\frac{1}{2}\Big(\omega(\vec{n}B_2, \vec{n}B_1) + \omega(\vec{n}B_4, \vec{n}B_3)\Big)\right).$$

Rewrite this expression using the identification with finite fields. The set X transforms to

$$X=\left\{\beta_1,\dots,\beta_4\in\mathscr{C}\middle|\begin{array}{l}\nu(\beta_1-\beta_2+\beta_3-\beta_4)=0\\ \nu^*(\beta_1^{-1}-\beta_2^{-1}+\beta_3^{-1}-\beta_4^{-1})=0\end{array}\right\}$$

and

$$\frac{1}{q}\mathrm{Tr}((DD^*)^2)=\frac{1}{|\mathscr{C}|^4}\sum_X e_q(\nu\nu^*\kappa(\beta_2\beta_1^{-1}-\beta_2^{-1}\beta_1+\beta_4\beta_3^{-1}-\beta_4^{-1}\beta_3)),$$

where $\nu=\omega(\vec{n},\vec{v}^*)$, $\nu^*=\omega(\vec{n},\vec{v})$ and $\kappa=(2\omega(\vec{v},\vec{v}^*))^{-1}$ as in Section 4.3.1.

Since we assumed $\mathfrak{Q}(\vec{n})=\nu\nu^*\neq 0$, the set X is actually

$$X=\left\{\beta_1,\dots,\beta_4\in\mathscr{C}\middle|\begin{array}{l}\beta_1-\beta_2=\beta_4-\beta_3\\ \beta_1^{-1}-\beta_2^{-1}=\beta_4^{-1}-\beta_3^{-1}\end{array}\right\}.$$

Now make a change of variables:

$$x=\beta_2\beta_1^{-1},\ y=\beta_4\beta_3^{-1},\ z=\beta_3\beta_1^{-1},\ w=\beta_1,$$

or equivalently $\beta_1=w,\beta_2=xw,\beta_3=zw$ and $\beta_4=yzw$. In these variables we get

$$\frac{1}{q}\mathrm{Tr}\left((DD^*)^2\right)=\frac{1}{|\mathscr{C}|^4}\sum_Y e_q\left(\mathfrak{Q}(\vec{n})\kappa(x-x^{-1}+y-y^{-1})\right),$$

where the set

$$Y=\left\{x,y,z,w\in\mathscr{C}\,\middle|\,\begin{array}{l}(1-x)=z(y-1)\\ yz(x-1)=x(1-y)\end{array}\right\}.$$

The set Y can be rewritten as

$$Y=\left\{x,y,z,w\in\mathscr{C}\,\middle|\,\begin{array}{ll}x=y=1 & \text{or}\\ x=z=y^{-1} & \text{or}\\ x=y\text{ and }z=-1 & \end{array}\right\}.$$

Indeed, if $x=1$, then from the second equation $y=1$ and z is arbitrary. Otherwise, replace $y-1=z^{-1}(1-x)$ in the second equation to get $yz^2(x-1)=x(x-1)$, implying that $x=yz^2$. Plug this back to the first equation to get $1-yz^2=z(y-1)$, that is equivalent to $(yz-1)(1+z)=0$. Hence, either $z=-1$ or $yz=1$ which implies (by the first equation) $x=y$, or $x=z$ respectively.

Therefore,

$$\frac{1}{q}\mathrm{Tr}\left((DD^*)^2\right)=\frac{1}{|\mathscr{C}|^3}\left(\sum_{z\in\mathscr{C}}1+\sum_{x\in\mathscr{C}}1+\sum_{x\in\mathscr{C}}e_q\left(2\mathfrak{Q}(n)\kappa\frac{x^2-1}{x}\right)-3\right).$$

Both in the symmetric and nonsymmetric cases, $\mathscr{C}$ is an irreducible algebraic curve of genus 1, defined over the field $\mathbb{F}_q$, and the function $\frac{x^2-1}{x}$ has two simple poles at $0, \infty$. Hence, by [2, Th. 5],

$$\left|\sum_{x\in\mathscr{C}} e_q\left(2\mathfrak{Q}(n)\kappa\frac{x^2-1}{x}\right)\right| \leq 2\sqrt{q}$$

(in fact, in the nonsymmetric case $\mathscr{C} = \mathbb{F}_q^*$ and this is a Kloosterman sum). This estimate implies

$$\frac{1}{q}\mathrm{Tr}\left((DD^*)^2\right) = \frac{2}{q^2} + O\left(\frac{1}{q^{5/2}}\right),$$

concluding the proof. □

4.4. *Formulas for matrix elements.* Let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ be a matrix with distinct eigenvalues, and let $p > \Delta(P_A)$ be a sufficiently large prime. We showed that the quantization of the centralizer $C_p(A)$ is equivalent to a tensor product of the quantizations of $C_p(A_{\bar\vartheta}) \subset \mathrm{Sp}(2d_{\bar\vartheta}, \mathbb{F}_p)$. For each irreducible element, we showed that the joint eigenfunctions are essentially unique and found explicit formulas for the matrix elements. We now describe the Hecke eigenfunctions and corresponding matrix elements in the general case.

Since $C_p(A) \cong \prod_{\bar\vartheta} C_p(A_{\bar\vartheta})$, we can identify any character of $C_p(A)$ as a product $\chi = \prod_{\bar\vartheta} \chi_{\bar\vartheta}$, where $\chi_{\bar\vartheta}$ are characters of $C_p(A_{\bar\vartheta})$. Denote by $\mathscr{H}^{\bar\vartheta}_{\chi_{\bar\vartheta}} \subseteq L^2(\mathbb{F}_p^{d_{\bar\vartheta}})$ the joint eigenspace of all the operators $U_p^{(d_{\bar\vartheta})}(B_{\bar\vartheta})$, $B_{\bar\vartheta} \in C_p(A_{\bar\vartheta})$ (with eigenvalues $\chi_{\bar\vartheta}$). Then, the map $\mathscr{U}$ from Proposition 4.1 maps the eigenspace $\mathscr{H}_\chi$ isomorphically onto the space $\bigotimes \mathscr{H}^{\bar\vartheta}_{\chi_{\bar\vartheta}}$. Furthermore, from Proposition 4.4 we know that these eigenspaces are essentially one-dimensional. We can thus deduce:

PROPOSITION 4.12. *Let $\chi = \prod_{\bar\vartheta} \chi_{\bar\vartheta}$ be a character of $C_p(A)$.*

- *If $\forall\bar\vartheta$, $\chi_{\bar\vartheta}$ is not the quadratic character, then* $\dim \mathscr{H}_\chi = 1$.
- *If $\chi_{\bar\vartheta}$ is the quadratic character for some symmetric orbit $\bar\vartheta$, then* $\dim \mathscr{H}_\chi = 0$.
- *Otherwise,* $\dim \mathscr{H}_\chi = 2^k$, *where k is the number of (nonsymmetric) orbits $\bar\vartheta$ for which $\chi_{\bar\vartheta}$ is the quadratic character.*
- *A basis for this space is given by $\left\{\psi_\chi^\eta | \eta \in (\mathbb{Z}/2\mathbb{Z})^k\right\}$,*

$$\psi_\chi^\eta = \mathscr{U}^{-1}\left(\bigotimes_{\chi_{\bar\vartheta}\neq\chi_2} \psi^{\bar\vartheta}_{\chi_{\bar\vartheta}} \otimes \bigotimes_{\chi_{\bar\vartheta}=\chi_2} \psi^{\bar\vartheta}_{\eta_{\bar\vartheta}}\right),$$

where $\left\{\psi_0^{\bar\vartheta}, \psi_1^{\bar\vartheta}\right\}$ is a basis for $\mathscr{H}^{\bar\vartheta}_{\chi_2}$.

Note that the number of characters for which the quadratic character appears in the decomposition is bounded by $O(p^{d-1})$. Hence, the set $J_p \subseteq \{\psi_1, \dots, \psi_{p^d}\}$ of Hecke eigenfunctions for which the quadratic character does not appear in the decomposition is of density one (i.e., $\lim_{p\to\infty}\frac{\sharp J_p}{p^d}=1$). For these eigenfunctions we can express the matrix elements as a product of exponential sums.

For $\vec{n}\in\mathbb{Z}^{2d}$ and any symplectic Frobenius orbit $\bar{\vartheta}\in\Lambda_p/\pm G_p$, let $\nu_{\bar{\vartheta}}=\mathfrak{Q}_{\vartheta}(\vec{n}_{\bar{\vartheta}})=\omega(\vec{n}_{\bar{\vartheta}},\vec{v}_{\bar{\vartheta}})\omega(\vec{n}_{\bar{\vartheta}},\vec{v}^*_{\bar{\vartheta}})$ as in Proposition 4.7 (where $\vec{v}_{\bar{\vartheta}},\vec{v}^*_{\bar{\vartheta}}$ are eigenvectors of $A_{\bar{\vartheta}}$ and $\vec{n}_{\bar{\vartheta}}$ is the projection of $\vec{n}$ (mod p) to $E_{\bar{\vartheta}}$). Let ψ be a Hecke eigenfunction with corresponding character $\chi=\prod\chi_{\bar{\vartheta}}$. Define

$$E^{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}},\chi_{\bar{\vartheta}})=\begin{cases}-E_{q_{\bar{\vartheta}}}(\nu_{\bar{\vartheta}},\chi_{\bar{\vartheta}}) & \vec{n}_{\bar{\vartheta}}\neq 0,\ \vartheta=\vartheta^*\\ E_{q_{\bar{\vartheta}}}(\nu_{\bar{\vartheta}},\chi_{\bar{\vartheta}}) & \vec{n}_{\bar{\vartheta}}\neq 0,\ \vartheta\neq\vartheta^*\\ 1 & \vec{n}_{\bar{\vartheta}}=0\end{cases},$$

where $E_{q_{\bar{\vartheta}}}(\nu_{\bar{\vartheta}},\chi_{\bar{\vartheta}})$ are the exponential sums which were defined in Definition 4.6 and $q_{\bar{\vartheta}}=p^{d_{\bar{\vartheta}}}$.

If $\chi_{\bar{\vartheta}}\neq\chi_2$ is not the quadratic character for any orbit, then ψ is uniquely determined and $E^{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}},\chi_{\bar{\vartheta}})=\langle T^{(d_{\bar{\vartheta}})}(\vec{n}_{\bar{\vartheta}})\psi^{\bar{\vartheta}}_{\chi_{\bar{\vartheta}}},\psi^{\bar{\vartheta}}_{\chi_{\bar{\vartheta}}}\rangle$. Consequently, the corresponding matrix element is a product of exponential sums,

$$\langle\tilde{T}_p(\vec{n})\psi,\psi\rangle=\prod_{\Lambda_p/\pm G_p}E^{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}},\chi_{\bar{\vartheta}}).\tag{4.6}$$

For characters χ such that the quadratic character appears in the decomposition, the corresponding eigenfunction is no longer unique. However, any $\psi\in\mathcal{H}_\chi$ is of the form $\psi=\sum_\eta a_\eta\psi^\eta_\chi$, where ψ^η_χ are defined in Proposition 4.12 and $\sum|a_\eta|^2=1$. Consequently, the corresponding matrix element is of the form

$$\langle\tilde{T}_p(\vec{n})\psi,\psi\rangle=F(\vec{n},\psi)\prod_{\bar{\vartheta}\notin W_\chi}E^{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}},\chi_{\bar{\vartheta}}),\tag{4.7}$$

where W_χ is the set of nonsymmetric orbits $\bar{\vartheta}$ for which $\chi_{\bar{\vartheta}}$ is the quadratic character and

$$F(\vec{n},\psi)=\sum_{\eta,\eta'}a_\eta a_{\eta'}\prod_{\vartheta\in W_\chi}\langle\tilde{T}_p^{(d_{\bar{\vartheta}})}(\vec{n}_{\bar{\vartheta}})\psi^{\bar{\vartheta}}_{\eta_\vartheta},\psi^{\bar{\vartheta}}_{\eta'_\vartheta}\rangle.$$

5. Super scars

This section is devoted to the proof of Theorem 1. For any rational isotropic subspace $E_0\subset\mathbb{Q}^{2d}$ that is invariant under the action of A (i.e., $\vec{n}\in E_0\Rightarrow\vec{n}A\in E_0$), we assign a corresponding submanifold of the torus $X_0\subseteq\mathbb{T}^{2d}$ of dimension $\dim X_0=2d-\dim E_0$ that is invariant under the induced dynamics (i.e., $\vec{x}=\begin{pmatrix}\vec{p}\\ \vec{q}\end{pmatrix}\in X_0\Rightarrow A\vec{x}\in X_0$). We then construct, for each prime $N=p$, a corresponding Hecke eigenfunction $\psi=\psi^{(p)}$ such that the distribution on the torus given by $f\mapsto\langle\mathrm{Op}_p(f)\psi,\psi\rangle$ weekly converges to Lebesgue measure on X_0.

5.1. *Invariant manifolds.* Let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ be a matrix with distinct eigenvalues. To any invariant isotropic rational subspace $E_0 \subset \mathbb{Q}^{2d}$, define the lattice $Z_0 = E_0 \cap \mathbb{Z}^{2d}$ and assign a closed subgroup of the torus $X_{E_0} \subseteq \mathbb{T}^{2d}$ defined by

$$X_{E_0} = \left\{ \vec{x} \in \mathbb{T}^{2d} | \vec{n} \cdot \vec{x} = 0 \pmod{\mathbb{Z}},\ \forall\, \vec{n} \in Z_0 \right\}.$$

The group $X_{E_0} \cong \mathbb{T}^{2d-d_0}$ is a submanifold with codimension $d_0 = \dim E_0$ and is invariant under the action of A. In general the submanifold X_{E_0} is co-isotropic; nevertheless, when E_0 is a Lagrangian subspace, X_{E_0} is also Lagrangian.

LEMMA 5.1. *Let E_0 be an invariant rational isotropic subspace. Then there is $\vec{x}_0 \in \mathbb{T}^{2d}$ such that*

$$\vec{n} \cdot \vec{x}_0 = \frac{\vec{n}_1 \cdot \vec{n}_2}{2} \pmod{\mathbb{Z}},\ \forall \vec{n} \in Z_0.$$

Proof. It is sufficient to show that there is $\vec{x} \in \mathbb{R}^{2d}$ such that

$$\vec{n} \cdot \vec{x} \equiv \vec{n}_1 \cdot \vec{n}_2 \pmod{2}, \tag{5.1}$$

for all $\vec{n} \in Z_0$ (then $\vec{x}_0$ is the class of $\frac{1}{2}\vec{x}$ modulo $\mathbb{Z}$). Notice, that if (5.1) is satisfied for $\vec{n}, \vec{m} \in Z_0$, then it is also satisfied for $\vec{n} + \vec{m}$. Indeed, for any $\vec{n}, \vec{m} \in Z_0$, because E_0 is isotropic we have $\vec{n}_1 \cdot \vec{m}_2 = \vec{m}_1 \cdot \vec{n}_2$; hence

$$(\vec{n}_1 + \vec{m}_1) \cdot (\vec{n}_2 + \vec{m}_2) \equiv \vec{n}_1 \cdot \vec{n}_2 + \vec{m}_1 \cdot \vec{m}_2 \pmod{2}.$$

Therefore, it is sufficient to check the condition for an integral basis of the lattice Z_0.

Let $\{\vec{n}^{(i)}\}_{i=1}^{d_0}$ be an integral basis. The vectors $\vec{n}^{(i)}$ are linearly independent; hence, the set of equations $\vec{n}^{(i)} \cdot \vec{x} = b_i$ has a solution for any $(b_1, \ldots, b_{d_0}) \in \mathbb{R}^{d_0}$ and in particular for $b_i = \vec{n}_1^{(i)} \cdot \vec{n}_2^{(i)}$. □

We can now define the manifold X_0 to be the coset $X_0 = \vec{x}_0 + X_{E_0}$; that is,

$$X_0 = \left\{ \vec{x} \in \mathbb{T}^{2d} \middle| \vec{n} \cdot \vec{x} = \frac{\vec{n}_1 \cdot \vec{n}_2}{2} \pmod{\mathbb{Z}}, \quad \forall \vec{n} \in Z_0 \right\}.$$

The condition that $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ is quantizable implies that X_0 is still invariant under the induced dynamics.

5.2. *Rational orbits and Frobenius orbits.* For the proof of Theorem 1, we would like to use the properties of the Hecke eigenfunctions and matrix elements described in Section 4. However, since all the results in Section 4 were described in terms of the finite field $\mathbb{F}_p$, we first need to establish the correspondence between invariant rational subspaces for A and invariant subspaces for A modulo p.

Let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ with distinct eigenvalues. Then for any prime $p > \Delta(P_A)$, we can think of A also as an element of $\mathrm{Sp}(2d, \mathbb{F}_p)$ with distinct eigenvalues. Denote by $\Lambda_\mathbb{Q}$ the set of complex eigenvalues of A, and by Λ_p the set of eigenvalues of A (modulo p) in $\bar{\mathbb{F}}_p$ (the algebraic closure of $\mathbb{F}_p$). Let $\mathbb{Q}^{2d} = \bigoplus_{\lambda_\mathbb{Q}/G_\mathbb{Q}} E_\theta$ and $\mathbb{F}_p^{2d} = \bigoplus_{\Lambda_p/G_p} E_\vartheta$ be the decompositions into irreducible invariant subspaces.

For each rational orbit $\theta \in \Lambda_\mathbb{Q}/G_\mathbb{Q}$, denote by $P_\theta = \mathrm{irr}_\mathbb{Q}(\theta)$ the minimal polynomial for some $\lambda_\theta \in \theta$ (this is independent of representative). We say that a Frobenius orbit, $\vartheta \in \Lambda_p/G_p$, lies under θ (denoted by $\vartheta|\theta$) if $\mathrm{irr}_{\mathbb{F}_p}(\vartheta)$ divides P_θ modulo p. We denote by θ^* the orbit of λ_θ^{-1} and note that $\vartheta|\theta \Leftrightarrow \vartheta^*|\theta^*$ in particular, if θ is nonsymmetric (i.e., $\theta \neq \theta^*$), then so is any Frobenius orbit ϑ that lies under θ.

For every rational orbit $\theta \in \Lambda_\mathbb{Q}/G_\mathbb{Q}$, fix an eigenvalue λ_θ. For every Frobenius orbit $\vartheta \in \Lambda_p/G_p$ lying under θ, fix a representative λ_ϑ. For any such choice, there is a corresponding ring homomorphism

$$\pi_{\lambda_\theta,\lambda_\vartheta} : \mathbb{Z}[\lambda_\theta] \to \mathbb{F}_p(\lambda_\vartheta),$$

sending λ_θ to λ_ϑ.

LEMMA 5.2. *Let $\mathfrak{D}_K \subseteq \mathcal{O}_K$ be a subring of the integral ring of a number field $K/\mathbb{Q}$, let $\mathbb{F}_q$ be a finite field of characteristic p, and let $\pi : \mathfrak{D}_K \to \mathbb{F}_q$ be any ring homomorphism. Then, for any $\alpha \in \mathcal{O}_K$ such that $\mathcal{N}_{K/\mathbb{Q}}(\alpha) \neq 0 \pmod p$, the image $\pi(\alpha) \neq 0$ as well.*

Proof. Let $f = \mathrm{irr}_\mathbb{Q}(\alpha)$. Then f is a unit integral polynomial such that $f(\alpha) = 0$. Consequently, if we take $\bar{f} \in \mathbb{F}_q[t]$ (by reduction of f modulo p), then $\bar{f}(\pi(\alpha)) = 0$ as well. On the other hand, we have that $f(0) = \pm\mathcal{N}_{K/\mathbb{Q}}(\alpha) \neq 0 \pmod p$; hence $\bar{f}(0) \neq 0$ and in particular $\pi(\alpha) \neq 0$. □

For any rational orbit $\theta \in \Lambda_\mathbb{Q}/G_\mathbb{Q}$, take eigenvectors $\vec{v}_\theta, \vec{v}_\theta^*$ with coefficients in $\mathbb{Z}[\lambda_\theta]$ and eigenvalues $\lambda_\theta, \lambda_\theta^{-1}$ respectively. For $\vec{n} \in \mathbb{Z}^{2d}$ define

$$N_\theta(\vec{n}) = \mathcal{N}_{\mathbb{Q}(\lambda_\theta)/\mathbb{Q}}(\omega(\vec{n}, \vec{v}_\theta^*)).$$

LEMMA 5.3. *For any element $\vec{n} \in \mathbb{Z}^{2d}$ and any orbit $\theta \in \Lambda_\mathbb{Q}/G_\mathbb{Q}$:*

- *If the projection of $\vec{n}$ to E_θ vanishes, then for any $\vartheta|\theta$ the projection of $\vec{n}$* (mod p) *to E_ϑ also vanishes.*
- *If $p > N_\theta(\vec{n})$ and the projection of $\vec{n}$ to E_θ does not vanish, then for any $\vartheta|\theta$, the projection to E_ϑ does not vanish as well.*

Proof. For any Frobenius orbit $\vartheta|\theta$, let $\vec{v}_\vartheta^* = \pi_{\lambda_\theta,\lambda_\vartheta}(\vec{v}_\theta^*)$. The vectors $\vec{v}_\vartheta^*$ are then eigenvectors with eigenvalues λ_ϑ^{-1}, and

$$\omega(\vec{n}, \vec{v}_\vartheta^*) = \pi_{\lambda_\theta,\lambda_\vartheta}(\omega(\vec{n}, \vec{v}_\theta^*)).$$

By Corollary A.4 the projection of $\vec{n}$ to E_θ vanishes if and only if $\omega(\vec{n}, \vec{v}_\theta^*) = 0$, and the projection of $\vec{n} \pmod{p}$ to E_ϑ vanishes if and only if

$$\omega(\vec{n}, \vec{v}_\vartheta^*) = \pi_{\lambda_\theta, \lambda_\vartheta}(\omega(\vec{n}, \vec{v}_\theta^*)) = 0.$$

The first part is now immediate, and the second part follows from Lemma 5.2. □

5.3. *Construction of eigenfunctions.* Let $Q^{2d} = \bigoplus_{\Lambda_\mathbb{Q}/G_\mathbb{Q}} E_\theta$ be the unique decomposition into irreducible (rational) invariant subspaces. Then any invariant isotropic subspace E_0 is a direct sum

$$E_0 = \bigoplus_{\theta \in \Theta} E_\theta,$$

where $\Theta \subseteq \Lambda_\mathbb{Q}/G_\mathbb{Q}$ is a subset containing nonsymmetric orbits such that $\theta \in \Theta \Rightarrow \theta^* \notin \Theta$.

Fix a large prime $p \geq \Delta(P_A)$, and recall the reduction to irreducible orbits described in Section 4 and the formulas for the eigenfunctions given in Proposition 4.4. We will now construct a Hecke eigenfunction by prescribing the characters $\chi_{\bar{\vartheta}}$ and eigenstates $\psi_{\bar{\vartheta}}$ for each symplectic Frobenius orbit $\bar{\vartheta} \in \Lambda_p / \pm G_p$.

We first determine the characters. For any symmetric orbit $\bar{\vartheta}$ fix an arbitrary character $\chi_{\bar{\theta}} \neq \chi_2$. For any nonsymmetric orbit $\bar{\vartheta}$, there is a unique nonsymmetric rational orbit $\bar{\theta}$ such that $\bar{\vartheta}|\bar{\theta}$. If $\bar{\theta} = \theta \cup \theta^*$ with $\theta, \theta^* \notin \Theta$, then we take $\chi_{\bar{\vartheta}} \neq \chi_2$ to be any character except the quadratic, and otherwise we take $\chi_{\bar{\vartheta}} = \chi_2$ to be the quadratic one.

Now for the eigenfunctions, when $\chi_{\bar{\vartheta}} \neq \chi_2$, the eigenspace $\mathcal{H}^{\bar{\vartheta}}_{\chi_{\bar{\vartheta}}}$ is one-dimensional and $\psi^{\bar{\vartheta}}_{\chi_{\bar{\vartheta}}}$ is determined. Otherwise, there is $\theta \in \Theta$ such that $\bar{\theta} = \theta \cup \theta^*$. Let $\vartheta|\theta$ be the Frobenius orbit under θ and let $\vec{v}_\vartheta$ be an eigenvector for $A_{\bar{\vartheta}}$ with eigenvalue $\lambda_\vartheta \in \vartheta$. We then take $\psi_0^{\bar{\vartheta}} \in \mathcal{H}^{\bar{\vartheta}}_{\chi_2}$ to be the eigenfunction (constructed in Lemma 4.9) satisfying

$$\langle \tilde{T}_p(\vec{n}_{\bar{\vartheta}}) \psi_0^{\bar{\vartheta}}, \psi_0^{\bar{\vartheta}} \rangle = \begin{cases} 1 & \omega(\vec{n}_{\bar{\vartheta}}, \vec{v}_\vartheta) = 0 \\ 0 & \text{otherwise.} \end{cases}$$

To conclude, we take the character $\chi = \prod \chi_{\bar{\vartheta}}$ and eigenfunction

$$\psi = \psi_\chi = \mathcal{U}^{-1} \left(\bigotimes_{\chi_{\bar{\vartheta}} \neq \chi_2} \psi^{\bar{\vartheta}}_{\chi_{\bar{\vartheta}}} \otimes \bigotimes_{\chi_{\bar{\vartheta}} = \chi_2} \psi_0^{\bar{\vartheta}} \right)$$

as in Proposition 4.12.

PROPOSITION 5.4.

$$|\langle \tilde{T}_p(\vec{n}) \psi, \psi \rangle| = \begin{cases} 1 & \vec{n} \in E_0 \\ O(p^{-1/4}) & \vec{n} \notin E_0. \end{cases}$$

Proof. The matrix elements corresponding to ψ are of the form

$$\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle = \prod_{\Lambda_p/\pm G_p} \langle \widetilde{T}_p(\vec{n}_{\bar{\vartheta}})\psi_{\vartheta}, \psi_{\vartheta}\rangle.$$

First for $\vec{n} \in E_0$. For any rational orbit $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$ such that $\theta, \theta^* \notin \Theta$ and any $\vartheta|\theta$, by Lemma 5.3, $\vec{n}_{\bar{\vartheta}} = 0$ and $\langle \widetilde{T}_p(\vec{n}_{\bar{\vartheta}})\psi_{\vartheta}, \psi_{\vartheta}\rangle = 1$. On the other hand, for $\theta \in \Theta$, the projection of $\vec{n}$ to E_{θ^*} vanishes. Since $\vartheta|\theta \Rightarrow \vartheta^*|\theta^*$, again by Lemma 5.3, the projection to E_{ϑ^*} vanishes implying $\omega(\vec{n}_{\bar{\vartheta}}, \vec{v}_{\bar{\vartheta}}) = 0$. Therefore, by construction again $\langle \widetilde{T}_p(\vec{n}_{\bar{\vartheta}})\psi_{\vartheta}, \psi_{\vartheta}\rangle = 1$. This covers all symplectic Frobenius orbits in the product, hence $\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle = 1$.

Now for $\vec{n} \notin E_0$. There is some $\theta \notin \Theta$ such that the projection of $\vec{n}$ to E_{θ} does not vanish. Then, by the second part of Lemma 5.3 (we can assume p is sufficiently large) for any $\vartheta|\theta$, the projection $\vec{n}_{\vartheta} \neq 0$. There are two possibilities: either $\theta^* \in \Theta$ or $\theta^* \notin \Theta$. If $\theta^* \in \Theta$, then $\vec{n}_{\vartheta^*} \neq 0$ implying that $\omega(n_{\bar{\vartheta}}, \vec{v}_{\vartheta}) \neq 0$ so $\langle \widetilde{T}_p(\vec{n}_{\bar{\vartheta}})\psi_{\vartheta}, \psi_{\vartheta}\rangle = 0$ by our construction. Otherwise, the corresponding character is not the quadratic character, and by Corollary 4.8 we have $|\langle \widetilde{T}_p(\vec{n}_{\bar{\vartheta}})\psi_{\vartheta}, \psi_{\vartheta}\rangle| = O(p^{-d_{\bar{\vartheta}}/2})$. Therefore, the whole product satisfies

$$|\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle| \leq O\Bigg(\prod_{\vartheta|\theta} p^{-d_{\bar{\vartheta}}/2}\Bigg) = O(p^{-d_{\theta}/2}). \qquad \square$$

The eigenfunctions constructed above satisfy $\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle = 1$ for all $\vec{n} \in Z_0$. This implies that ψ is also a joint eigenfunction of the operators $\widetilde{T}_p(\vec{n})$ with trivial eigenvalue[2]. This property can be used in order to make an alternative construction of these eigenfunctions. Given the isotropic invariant subspace E_0, the operators $\widetilde{T}_p(\vec{n})$, $\vec{n} \in Z_0$ all commute (because it is isotropic) and one can consider the decomposition into joint eigenspaces. The joint eigenspace corresponding to the trivial eigenvalue is not empty and is invariant under the action of all Hecke operators (because the space E_0 is invariant). Therefore, there is a basis for this space composed of Hecke eigenfunctions each satisfying $\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle = 1$ for any $\vec{n} \in Z_0$. Furthermore, if $\vec{m} \in E_0^*$ (the symplectic complement of E_0), then there is $\vec{n} \in Z_0$ such that $\omega(\vec{n}, \vec{m}) \neq 0$. Consequently, for a sufficiently large p, $\widetilde{T}_p(\vec{m})\psi$ is an eigenfunction of $\widetilde{T}_p(\vec{n})$ with eigenvalue $\neq 1$ and so $\langle \widetilde{T}_p(\vec{m})\psi, \psi\rangle = 0$. If we assume in addition that the space E_0 is a maximal isotropic invariant subspace, then any $\vec{n} \in \mathbb{Z}^{2d}$ is either in $E_0 \cup E_0^*$ or that it does not belong to any invariant isotropic subspace, in which case we have the estimate $\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle = O(p^{-\frac{1}{2}})$. We thus see that any Hecke eigenfunction constructed in this manner, also satisfies Proposition 5.4.

[2] I thank Stéphane Nonnenmacher for pointing that out.

5.4. *Proof of Theorem* 1. We now turn to prove Theorem 1; that is we prove the following proposition.

PROPOSITION 5.5. *As $p \to \infty$ through primes, the distribution on the torus given by*

$$f \mapsto \langle \mathrm{Op}_p(f)\psi, \psi\rangle$$

(*where ψ are the Hecke eigenfunctions constructed above*) *converges to Lebesgue measure on X_0.*

Proof. It is sufficient to show convergence for the test functions $e_{\vec{n}}(\vec{x}) = \exp(2\pi i \vec{n}\cdot\vec{x})$, $\vec{n} \in \mathbb{Z}^{2d}$. For these functions,

$$\int_{\mathbb{T}^{2d}} e_{\vec{n}}(\vec{x}) d\mu_{X_0}(\vec{x}) = \begin{cases} (-1)^{\vec{n}_1\cdot\vec{n}_2} & \vec{n} \in Z_0 \\ 0 & \text{otherwise,} \end{cases}$$

where μ_{X_0} is Lebesgue measure on X_0. For $N = p$, a large (and in particular odd) prime, the corresponding operator $\mathrm{Op}_p(e_{\vec{n}}) = (-1)^{\vec{n}_1\cdot\vec{n}_2}\widetilde{T}_p(\vec{n})$. Therefore, it is sufficient to show that as $p \to \infty$,

$$\langle \widetilde{T}_p(\vec{n})\psi, \psi\rangle \to \begin{cases} 1 & \vec{n} \in Z_0 \\ 0 & \text{otherwise,} \end{cases}$$

and this follows from Proposition 5.4. □

6. Quantum variance

In the following section, we assume that $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ has no invariant isotropic rational subspaces and compute the quantum variance when Planck's constant is the inverse of a large prime $N = p$. First we introduce a quadratic form $Q : \mathbb{Z}^{2d} \to \mathfrak{O}$ that characterizes the Hecke orbits of an element $\vec{n} \in \mathbb{Z}^{2d}$ (in the sense of Proposition 6.1). We then define modified Fourier coefficients, grouping together coefficients belonging to the same Hecke orbits. Finally, we use the structure of the Hecke eigenfunctions described in Section 4 and the relations between the rational orbits and Frobenius orbits described in Section 5.2 to calculate the quantum variance proving Theorem 4.

6.1. *A quadratic form.* Let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ with $2d$ distinct eigenvalues. Recall the notation of Section 2.3. Let $\Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$ denote the orbits of the Galois group $G_{\mathbb{Q}}$ on the set of eigenvalues $\Lambda_{\mathbb{Q}}$. Let $\mathbb{Q}^{2d} = \bigoplus E_\theta$ be the decomposition into irreducible invariant subspaces. Further assume that there are no invariant rational isotropic subspaces, implying that all orbits are symmetric $\theta = \theta^* = \bar{\theta}$ (hence all E_θ are symplectic). Recall the map $\iota^* : \mathbb{Z}^{2d} \to \mathfrak{O}$ (sending $\vec{n} \mapsto \omega(\vec{n}, \vec{v})$) and the norm map $\mathcal{N} : \mathcal{D} \to \mathcal{D}$ (sending $\beta \mapsto \beta\beta^*$) and define the quadratic form

$$\begin{aligned} Q : \mathbb{Z}^{2d} &\to \mathcal{D} \\ \vec{n} &\mapsto \mathcal{N}(\iota^*(\vec{n})). \end{aligned}$$

The projection of $Q(\vec{n})$ to each component is given by

$$Q_\theta(\vec{n}) = \mathcal{N}_{K_\theta/F_\theta}(\omega(\vec{n}, \vec{v}_\theta)),$$

where $\vec{v}_\theta$ is a left eigenvector with eigenvalue λ_θ, and where $K_\theta = \mathbb{Q}(\lambda_\theta)$ and $F_\theta = \mathbb{Q}(\lambda_\theta + \lambda_\theta^{-1})$.

PROPOSITION 6.1. *Let $\vec{n}, \vec{m} \in \mathbb{Z}^{2d}$. Then $Q(\vec{n}) = Q(\vec{m})$ if and only if for all sufficiently large primes, the classes of $\vec{n}$ and $\vec{m}$ modulo p are in the same $C_p(A)$ orbit.*

Proof. We now use the relations between rational orbits and Frobenius orbits described in Section 5.2, to relate Corollary 4.3 to the rational arithmetics. First, assume that $Q(\vec{n}) = Q(\vec{m}) = \nu$. Let $N_0(\nu) = \max_\theta(\mathcal{N}_{F_\theta/\mathbb{Q}}(\nu_\theta))$. We show that for any prime $p > N_0(\nu)$, there is $B \in C_p(A)$ such that $\vec{n}B = \vec{m} \pmod{p}$. It is sufficient to show that for any Frobenius orbit $\bar{\vartheta} \in \Lambda_p/\pm G_p$, there is $B_{\bar{\vartheta}} \in C_p(A_{\bar{\vartheta}})$ such that $\vec{n}_{\bar{\vartheta}} B_{\bar{\vartheta}} = \vec{m}_{\bar{\vartheta}}$. For $\theta \in \Lambda_{\mathbb{Q}}/G_{\mathbb{Q}}$ such that $Q_\theta(\vec{n}) \neq 0$,

$$\mathcal{N}_{F_\theta/\mathbb{Q}}(Q_\theta(\vec{n})) = \mathcal{N}_{F_\theta/\mathbb{Q}}(Q_\theta(\vec{m})) \neq 0 \pmod{p}.$$

Notice that $\vec{v}_{\bar{\vartheta}} = \pi_{\lambda_\theta,\lambda_\vartheta}(\vec{v}_\theta)$ and $\vec{v}^*_{\bar{\vartheta}} = \pi_{\lambda_\theta,\lambda_\vartheta}(\vec{v}^*_\theta)$ are eigenvectors for $A \pmod{p}$ with eigenvalues λ_ϑ and λ_ϑ^{-1} respectively. Consequently, by Lemma 5.2, for any $\vartheta|\theta$,

$$\mathcal{Q}_{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}}) = \omega(\vec{n}_{\bar{\vartheta}}, \vec{v}_{\bar{\vartheta}})\omega(\vec{n}_{\bar{\vartheta}}, \vec{v}^*_{\bar{\vartheta}}) = \pi_{\lambda_\theta,\lambda_\vartheta}(Q_\theta(\vec{n})) \neq 0,$$

and by Corollary 4.3, there is $B_{\bar{\vartheta}} \in C_p(A_{\bar{\vartheta}})$ such that $\vec{n}_{\bar{\vartheta}} B_{\bar{\vartheta}} = \vec{m}_{\bar{\vartheta}}$. On the other hand, if $Q_\theta(\vec{n}) = Q_\theta(\vec{m}) = 0$, then by Lemma 5.3 for any $\vartheta|\theta$, $\vec{n}_\vartheta = \vec{m}_\vartheta = 0$. Since $\theta = \theta^*$ is symmetric, then $\vec{n}_{\vartheta^*} = \vec{m}_{\vartheta^*} = 0$ as well; hence, $\vec{n}_{\bar{\vartheta}} = \vec{m}_{\bar{\vartheta}} = 0$ and we can take any element of $C_p(A_{\bar{\vartheta}})$.

For the other direction, assume $Q(\vec{n}) \neq Q(\vec{m})$. Then there is at least one orbit θ such that $Q_\theta(\vec{n}) \neq Q_\theta(m)$. Consequently, for any prime $p > \mathcal{N}_{F_\theta/\mathbb{Q}}(Q_\theta(\vec{n}) - Q_\theta(m))$ and for any $\vartheta|\theta$, we have that $\mathcal{Q}_{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}}) \neq \mathcal{Q}_{\bar{\vartheta}}(\vec{n}_{\bar{\vartheta}})$. Therefore, $\vec{m}_{\bar{\vartheta}}$ and $\vec{n}_{\bar{\vartheta}}$ are not in the same $\mathcal{C}_p(A_{\bar{\vartheta}})$ orbit implying that $\vec{n}, \vec{m} \pmod{p}$ are not in the same $\mathcal{C}_p(A)$ orbit. □

COROLLARY 6.2. *Let $\vec{n}, \vec{m} \in \mathbb{Z}^{2d}$ such that $Q(\vec{n}) = Q(\vec{m}) = \nu$. For any prime $p > N_0(\nu)$ and any Hecke eigenfunction $\psi \in \mathcal{H}_p$,*

$$\langle \tilde{T}_p(\vec{n})\psi, \psi \rangle = \langle \tilde{T}_p(\vec{m})\psi, \psi \rangle.$$

6.2. *Rewriting of matrix elements.* We now use the form Q to define modified Fourier coefficients and rewrite the matrix elements, incorporating the Hecke symmetries.

Definition 6.3. For $f \in C^\infty(\mathbb{T}^{2d})$ and $\nu \in \mathfrak{D}$, define modified Fourier coefficients,

$$f^\sharp(\nu) = \sum_{Q(\vec{n})=\nu} (-1)^{\vec{n}_1\cdot\vec{n}_2} \hat{f}(\vec{n}).$$

For $\nu \in \mathfrak{D}$ and any Hecke eigenfunction ψ, define

$$V_\nu(\psi) = \langle \tilde{T}_p(\vec{n})\psi, \psi\rangle,$$

where $\vec{n} \in \mathbb{Z}^{2d}$ is any element such that $Q(\vec{n}) = \nu$.

For $\nu \in \mathfrak{D}$ define $N_0(\nu) = \max_\theta(\mathcal{N}_{F_\theta/\mathbb{Q}}(\nu_\theta))$ (as in the proof of Proposition 6.1). For any trigonometric polynomial f, let $N_0(f) = \max_{\hat{f}(\vec{n})\neq 0}(N_0(Q(\vec{n})))$.

PROPOSITION 6.4. *For any trigonometric polynomial f, any prime $p > N_0(f)$ and any Hecke eigenfunction $\psi \in \mathcal{H}_p$:*

$$\langle Op_p(f)\psi, \psi\rangle = \sum_\nu f^\sharp(\nu) V_\nu(\psi).$$

Proof. Apply Corollary 6.2. □

Remark 6.1. Notice that it is possible to have $f \neq 0$ such that all the coefficients $f^\sharp(\nu) = 0$ vanish. For example, fix some $\vec{n} \in \mathbb{Z}^{2d}$ and take $f(\vec{x}) = e_{\vec{n}}(\vec{x}) - e_{\vec{n}A}(\vec{x}) \neq 0$.

6.3. *Proof of Theorem* 4. We now want to prove Theorem 4, that is to show that as $p \to \infty$,

$$S_2^{(p)}(f) = \frac{V(f)}{p^{d_f}} + O\left(\frac{1}{p^{d_f+1}}\right),$$

where $d_f = \min_{f^\sharp(\nu)\neq 0} d_\nu$, $d_\nu = \sum_{\nu_\theta\neq 0} \frac{|\theta|}{2}$, and $V(f) = \sum_{d_\nu = d_f} |f^\sharp(\nu)|^2$.

First, we compute mixed moments of elementary operators

$$S_2^{(p)}(\vec{n},\vec{m}) = \frac{1}{p^d}\sum_i \langle \tilde{T}_p(\vec{n})\psi_i, \psi_i\rangle \overline{\langle \tilde{T}_p(\vec{m})\psi_i, \psi_i\rangle}.$$

LEMMA 6.5. *Let $0 \neq \vec{n}, \vec{m} \in \mathbb{Z}^{2d}$ with $Q(\vec{n}) = \nu$, $Q(\vec{m}) = \mu$ and assume $d_\nu \le d_\mu$. Then, for $p > \max(N_0(\nu), N_0(\mu))$ the mixed second moment satisfies*

$$S_2^{(p)}(\vec{n},\vec{m}) = \begin{cases} \frac{1}{p^{d_\nu}} + O\left(\frac{1}{p^{d_\nu+1}}\right) & \nu = \mu \\ O\left(\frac{1}{p^{d_\nu+1}}\right) & \nu \neq \mu. \end{cases}$$

Proof. First assume that the matrix elements for all Hecke eigenfunctions are in the form of (4.6). Consequently, we can rewrite

$$S_2^{(p)}(\vec{n},\vec{m})=\prod_{\bar{\vartheta}\in\Lambda_p/\pm G_p}\left(\frac{1}{p^{d_{\bar{\vartheta}}}}\sum_{\chi_{\bar{\vartheta}}}E^{(\bar{\vartheta})}(\vec{n}_{\bar{\vartheta}},\chi_{\bar{\vartheta}})E^{(\bar{\vartheta})}(\vec{m}_{\bar{\vartheta}},\chi_{\bar{\vartheta}})\right).$$

Recall that we assumed that there are no nonsymmetric rational orbits; so (by the proof of Proposition 6.1) if $\nu_\theta\neq 0$ then $\forall\vartheta|\theta$, $\mathcal{Q}_{\bar{\vartheta}}(\vec{n}_{\bar{\theta}})\neq 0$ and if $\nu_\theta=0$ then $\forall\vartheta|\theta$, $\vec{n}_{\bar{\theta}}=0$ (similarly for $\vec{m}$ and μ). The result is now immediate from Proposition 4.10.

Now for a general Hecke basis. Any Hecke eigenfunction for which the quadratic character does not appear in the decomposition gives the same contribution to the sum as before (because such an eigenfunction is unique). It is thus sufficient to show that the contribution of all other eigenfunctions is bounded by $O(\frac{1}{p^{d_\nu+1}})$. The number of these eigenfunctions is bounded by $O(p^{d-1})$, so it is sufficient to show that each summand contributes at most $O(\frac{1}{p^{d_\nu}})$ and this is immediate from Corollary 4.8 and Lemma 4.9. □

Theorem 4 now follows from Lemma 6.5 and Proposition 6.4.

Proof. We first prove the case where f is a trigonometric polynomial. Define $N_0(f)=\max_{\hat{f}(\vec{n})\neq 0}\{N_0(Q(\vec{n}))\}$. Then for $p>N_0(f)$ (by Proposition 6.4), we can rewrite

$$\langle \mathrm{Op}_p(f)\psi_i,\psi_i\rangle-\int f\,dx=\sum_{0\neq\nu\in\mathcal{D}}f^\sharp(\nu)V_\nu(\psi_i).$$

Consequently (after changing the order of summation), the quantum variance takes the form

$$S_2^{(p)}(f)=\sum_{0\neq\nu,\mu\in\mathcal{D}}f^\sharp(\nu)\overline{f^\sharp(\mu)}\frac{1}{p^d}\sum_i V_\nu(\psi_i)\overline{V_\mu(\psi_i)}.$$

The second term (by Lemma 6.5) contributes $\frac{1}{p^{d_\nu}}+O(\frac{1}{p^{d_\nu+1}})$ when $\nu=\mu$ and $O(\frac{1}{p^{d_\nu+1}})$ otherwise. Therefore, the leading term is indeed

$$S_2^{(p)}(f)=\frac{1}{p^{d_f}}\sum_{d_\nu=d_f}|f^\sharp(\nu)|^2+O\left(\frac{1}{p^{d_f+1}}\right).$$

Now for any smooth $f\in C^\infty(\mathbb{T}^{2d})$, approximate f by trigonometric polynomials $f_R=\sum_{\|\vec{n}\|\leq R}\hat{f}(\vec{n})e_{\vec{n}}$. Note that since $N_0(Q(\vec{n}))\ll\|\vec{n}\|^{4d^2}$, then $N_0(f_R)\ll R^{4d^2}$. We can thus define $R=R(p)\sim p^{1/4d^2}$, so that $\|\vec{n}\|\leq R$ implies $N_0(Q(\vec{n}))\leq p$. We can take p sufficiently large, so that $d_f=d_{f_R}$, then from the first part

$$S_2^{(p)}(f_R) = \frac{V(f_R)}{p^{d_f}} + O\left(\frac{1}{p^{d_f+1}}\right).$$

On the other hand, we can bound the difference

$$|S_2^{(p)}(f) - S_2^{(p)}(f_R)| \ll_f \sum_{\|n\|>R} \hat{f}(\vec{n}) \ll_{f,\delta} \frac{1}{R^\delta},$$

for any power R^δ. In particular $|S_2^{(p)}(f) - S_2^{(p)}(f_R)| \ll_f \frac{1}{p^{d+1}}$, and in the same way, we also have $|V(f) - V(f_R)| \ll_f \frac{1}{p^{d+1}}$. We thus get that the quantum variance for smooth $f \in C^\infty(\mathbb{T}^{2d})$ satisfies

$$S_2^{(p)}(f) = \frac{V(f)}{p^{d_f}} + O_f\left(\frac{1}{p^{d_f+1}}\right). \qquad \square$$

7. Limiting distributions

As in the previous section, let $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$, with distinct eigenvalues and no invariant rational isotropic subspaces. Given a smooth observable f and a large prime p, consider the normalized matrix elements in the Hecke basis,

$$\mathcal{W}_i(f, p) = p^{d_f/2}\left(\langle \mathrm{Op}_p(f)\psi_i, \psi_i\rangle - \int f dx\right).$$

As $p \to \infty$ these points fluctuate around zero with variance tending to $V(f)$, and we can ask whether they converge to some limiting distribution. Throughout this section, we will assume the validity of the Kurlberg-Rudnick conjecture for the limiting distribution (formulated here as Conjecture 5), and deduce the limiting distributions for $\mathcal{W}_i(f, p)$.

First, for any trigonometric polynomial f, and Hecke eigenfunction ψ_i for which the quadratic character does not appear the decomposition, by formula (4.6) and Proposition 6.4 we have

$$\mathcal{W}_i(f, p) = \sum_{d_\nu = d_f} f^\sharp(\nu) \prod_{\nu_\theta \neq 0} \prod_{\vartheta | \theta} \sqrt{q_{\bar{\vartheta}}} E_{q_{\bar{\vartheta}}}(\nu_{\bar{\vartheta}}, \chi_{\bar{\vartheta}}) + O\left(\frac{1}{\sqrt{p}}\right), \tag{7.1}$$

where $\chi = \prod \chi_{\bar{\vartheta}}$ is the character corresponding to ψ_i, $\nu_{\bar{\vartheta}} = \pi_{\lambda_\theta, \lambda_\vartheta}(\nu_\theta)$ and the error term comes from the elements with $d_\nu > d_f$. By approximating a smooth function f by trigonometric polynomials f_R (as in the proof of Theorem 4) formula (7.1) is also valid for smooth functions. Finally, recall that the subset $J_p \subseteq \{\psi_1, \dots, \psi_{p^d}\}$ of eigenfunctions ψ_i for which formula (7.1) is valid is of density 1. Therefore, Conjecture 5 for the limiting distributions of the exponential sums implies the following limiting distributions for the matrix elements:

CONJECTURE 9. *For any tuple $k = (k_\theta)$, $1 \le k_\theta \le d_\theta$, consider the set of primes $\mathbf{P}_k$ for which under every rational (symmetric) orbit θ, there are precisely k_θ symplectic Frobenius orbits $\bar{\vartheta}|\theta$. Then, as $p \to \infty$ through primes from $\mathbf{P_k}$, there is a limiting distribution for $\mathcal{W}_i(f, p)$, and it is that of the random variable*

$$X_f = \sum_{d_\nu = d_f} f^\sharp(\nu) \prod_{\nu_\theta \neq 0} X^\theta_{\nu_\theta},$$

where the random variables $X^\theta_{\nu_\theta}$ are all independent random variables. Furthermore, each of the variables $X^\theta_{\nu_\theta}$ is a product of k_θ independent random variables with Sato-Tate distribution.

In particular, if we restrict to elementary observables $e_{\vec{n}} = \exp(2\pi i \vec{n} \cdot \vec{x})$, we recover Conjecture 6.

We now give an algorithm for determining which of the sets $\mathbf{P}_k$ are infinite; that is, to determine for a given matrix $A \in \mathrm{Sp}_\theta(2d, \mathbb{Z})$ which limiting distributions can actually occur.

Denote by P_A the characteristic polynomial for A, and assume that P_A is irreducible over the rationals (if it is reducible, one can repeat this process for each irreducible factor). Let λ be a root of P_A and denote by $\tilde{P}_A = \mathrm{irr}_{\mathbb{Q}}(\lambda + \lambda^{-1})$ the minimal polynomial for $\lambda + \lambda^{-1}$. Then $\tilde{P}_A$ is an irreducible integral unit polynomial of degree d. Furthermore, the space $\mathbb{F}_p^{2d} = \bigoplus_{\bar{\vartheta}} E_{\bar{\vartheta}}$ decomposes into k irreducible invariant symplectic subspaces, if and only if $\tilde{P}_A = \prod_{\bar{\vartheta}} \tilde{P}_{\bar{\vartheta}}$ is a product of k irreducible polynomials over $\mathbb{F}_p$ (where $\tilde{P}_{\bar{\vartheta}} = \mathrm{irr}_{\mathbb{F}_p}(\lambda_{\bar{\vartheta}} + \lambda_{\bar{\vartheta}}^{-1})$). Therefore, the set $\mathbf{P}_k$ is precisely the set of primes for which the polynomial $\tilde{P}_A \pmod p$ is a product of k irreducible polynomials. The density of these sets, $\frac{1}{\pi(X)}\#\{p \le X | p \in \mathbf{P}_k\}$, can be calculated by the Chebotarev theorem after calculating the Galois groups for $\tilde{P}_A$. To do this, consider the Galois group as a subgroup of the symmetric group S_d (via its action on the roots of $\tilde{P}_A$). Recall that any element of S_d can be uniquely presented as a product of disjoint cycles. The Chebotarev theorem says that the density of the set $\mathbf{P}_k$ is the relative number of elements in the Galois group that are a product of k cycles. Furthermore, if there are no elements that are a product of k cycles, then $\mathbf{P}_k$ contains at most finitely many primes. For a precise statement and some background on the Chebotarev theorem; see [12, Th. 6.3.1]. We demonstrate this calculation for a few simple examples.

Our first example is a four-dimensional symplectic matrix $A \in \mathrm{Sp}(4, \mathbb{Z})$ for which P_A is irreducible (i.e., no invariant rational subspaces). In this case the polynomial $\tilde{P}_A$ is a quadratic irreducible polynomial. In fact if $P_A(t) = t^4 - at^3 + bt^2 - at + 1$, then $\tilde{P}_A = t^2 - at + b - 2$. Consequently, the condition that $p \in \mathbf{P}_2$ is equivalent to the condition that the quadratic polynomial $t^2 - at + b - 2$ has roots in $\mathbb{F}_p$, which is equivalent to the integer $c = a^2 - 4(b-2)$ being a square modulo p. Therefore, the sets $\mathbf{P}_1, \mathbf{P}_2$ are both unions of arithmetic progressions, and each have density $1/2$.

Our next examples are for matrices $A \in \mathrm{Sp}(6, \mathbb{Z})$ with an irreducible characteristic polynomial (i.e., the polynomial $\tilde{P}_A$ is an irreducible polynomial of degree 3). In this case, we can no longer describe the sets $\mathbf{P}_k$ as arithmetic progressions. However, the classification of the Galois group for degree 3 polynomials is still relatively easy, and we can give the corresponding densities in each case. There are only two possible cases: either the splitting field for $\tilde{P}_A$ is a degree 6 extension, in which case the Galois group is isomorphic to the symmetric group S_3, or that the splitting field is of degree 3 and the Galois group is cyclic of order 3. We will now consider each case separately.

In the symmetric group S_3 there are a total of six elements: two of them ($(1,2,3)$ and $(1,3,2)$) are composed of one cycle, three of them ($(1,2)(3)$, $(1,3)(2)$ and $(2,3)(1)$) are composed of two cycles, and one element (the identity) is composed of three cycles. Consequently, if the Galois group for $\tilde{P}_A$ is S_3 then, by the Chebotarev theorem, the densities of the sets $\mathbf{P}_1, \mathbf{P}_2$ and $\mathbf{P}_3$, are $2/6, 3/6$, and $1/6$ respectively.

The cyclic group has three elements: one is composed of three cycles and two of them ($(1,2,3)$ and $(1,3,2)$) are composed of one cycle. (There are no elements composed of two cycles.) Therefore, when the Galois group for $\tilde{P}_A$ is cyclic, the Chebotarev theorem implies that the density of $\mathbf{P}_1, \mathbf{P}_2$, and $\mathbf{P}_3$ are $2/3, 0$, and $1/3$ respectively. Furthermore, $\mathbf{P}_2$ contains at most finitely many primes and the corresponding limiting distribution is not obtained.

Appendix A. Galois orbits and invariant subspaces

Let E be a $2d$-dimensional vector space defined over a perfect field F (we will consider only the cases where F is a number field or a finite field). Let $\omega : E \times E \to F$ be a symplectic form, and let $A \in \mathrm{Sp}(E, \omega)$ be a symplectic linear map with distinct eigenvalues acting on E from the left. Denote by Λ_F the set of eigenvalues of A (in the algebraic closure of F). Let G_F be the absolute Galois group, and denote by Λ_F / G_F the orbits of the eigenvalues under the action of G_F (in fact, it is sufficient to consider $\mathrm{Gal}(P_A/F)$, the Galois group of the splitting field of the characteristic polynomial P_A).

Since the matrix A is symplectic, if $\lambda \in \Lambda_F$ is an eigenvalue, then $\lambda^{-1} \in \Lambda_F$ as well. To each orbit $\theta \in \Lambda_F / G_F$ there is a unique orbit θ^* such that $\lambda \in \theta \Leftrightarrow \lambda^{-1} \in \theta^*$. If $\theta = \theta^*$ we say that the orbit is symmetric, otherwise we say that the orbit is nonsymmetric.

LEMMA A.1. *There is a unique decomposition into irreducible invariant subspaces*: $E = \bigoplus_{\Lambda_F / G_F} E_\theta$.

- *To each orbit $\theta \in \Lambda_F/G_F$, there is a corresponding subspace (denoted by E_θ), such that the eigenvalues of the restriction $A_{|E_\theta}$ are the eigenvalues $\lambda \in \theta$. In particular* $\dim E_\theta = |\theta|$.
- *For any two orbits θ, θ', unless $\theta' = \theta^*$, E_θ and $E_{\theta'}$ are orthogonal with respect to the symplectic form.*

Proof. Take representatives $\lambda_\theta \in \theta$ with eigenvectors $\vec{v}_\theta$. The space

$$E_\theta = \left\{\mathrm{Tr}_{F(\lambda_\theta)/F}(t\vec{v}_\theta) | t \in F(\lambda_\theta)\right\}$$

is a subspace of E invariant under A, and the eigenvalues of the restriction of A to E_θ are $\lambda \in \theta$. Furthermore, E_θ and $E_{\theta'}$ are orthogonal, unless there is $\sigma \in G_F$ such that $\omega(\sigma(\vec{v}_\theta), \vec{v}_{\theta'}) \neq 0$, and this happens only when $\theta' = \theta^*$. It remains to show that this is the only decomposition. Indeed, if $\tilde{E}$ is an invariant irreducible subspace, then there is an eigenvector $\vec{v}_\theta \in \tilde{E} \otimes \bar{F}$, and since $\tilde{E}$ is defined over F, then all the Galois conjugates $\sigma(\vec{v}_\theta)$ are in this space as well. Therefore, the space $E_\theta \subseteq \tilde{E}$ and since we assumed $\tilde{E}$ is irreducible, then $\tilde{E} = E_\theta$. □

Definition A.2. To each orbit $\theta \in \Lambda_F/G_F$, we define a symplectic orbit $\bar{\theta} = \theta \cup \theta^*$. Correspondingly, to each symplectic orbit, we assign the symplectic subspace $E_{\bar{\theta}} = E_\theta + E_{\theta^*}$. Then for symmetric orbits $E_{\bar{\theta}} = E_\theta$, and for nonsymmetric orbits $E_{\bar{\theta}} = E_\theta \oplus E_{\theta^*}$.

Denote by $\Lambda_F/\pm G_F$ the set of symplectic orbits. Then

$$E = \bigoplus_{\Lambda_F/\pm G_F} E_{\bar{\theta}}$$

is a decomposition to a direct sum of orthogonal symplectic subspaces.

LEMMA A.3. *Let $\lambda_\theta \in \theta$ with corresponding eigenvector $\vec{v}_\theta$ (with coefficients in $F(\lambda_\theta)$). Let $\vec{v}_\theta^*$ be an eigenvector with eigenvalue λ_θ^{-1}. Then the map*

$$\begin{array}{ccc} F(\lambda_\theta) & \to & E_\theta \\ t & \mapsto & \mathrm{Tr}_{F(\lambda_\theta)/F}(t\vec{v}_\theta) \end{array}$$

is a linear isomorphism, with an inverse map given by

$$\begin{array}{ccc} E_\theta & \to & F(\lambda_\theta) \\ \vec{n} & \mapsto & \frac{\omega(\vec{n},\vec{v}_\theta^*)}{\omega(\vec{v}_\theta,\vec{v}_\theta^*)} \end{array}.$$

Proof. The Galois conjugates of $\vec{v}_\theta$ are all eigenvectors with distinct eigenvalues in θ. Therefore, they are linearly independent and the map $t \mapsto \mathrm{Tr}_{F(\lambda)/F}(t\vec{v}_\theta)$ is injective. On the other hand, $F(\lambda_\theta)/F$ is a vector space of dimension $[F(\lambda_\theta) : F] = |\theta|$; hence it is isomorphic to E_θ.

Now let $\vec{n} \in E_\theta$, from the first part there is a decomposition

$$\vec{n} = \mathrm{Tr}_{F(\lambda)/F}(t\vec{v}_\theta) = \sum_{\sigma \in \mathrm{Mor}_F(F(\lambda),\bar{F})} \sigma(t\vec{v}_\theta).$$

Note that for any morphism, $\sigma \in \mathrm{Mor}_F(F(\lambda), \bar{F})$, the symplectic form

$$\omega(\sigma(\vec{v}_\theta), \vec{v}^*_\theta) = \omega(\sigma(\vec{v}_\theta A), \vec{v}^*_\theta A) = \sigma(\lambda)\lambda^{-1}\omega(\sigma(\vec{v}_\theta), \vec{v}^*_\theta).$$

Therefore, for any nontrivial morphism σ we have $\omega(\sigma(\vec{v}_\theta), \vec{v}^*_\theta) = 0$, and indeed $\omega(\vec{n}, \vec{v}^*_\theta) = t\omega(\vec{v}_\theta, \vec{v}^*_\theta)$. □

COROLLARY A.4. *For any element $\vec{n} \in E$, the projection of $\vec{n}$ to E_θ vanishes if and only if $\omega(\vec{n}, \vec{v}^*_\theta) = 0$, where $\vec{v}^*_\theta$ is any eigenvector with eigenvalue in θ^*.*

For each symplectic orbit $\bar{\theta} \in \Lambda_F / \pm G_F$, fix a representative $\lambda_{\bar{\theta}}$ and let $\vec{v}_{\bar{\theta}}, \vec{v}^*_{\bar{\theta}}$ be eigenvectors for $\lambda_{\bar{\theta}}, \lambda^{-1}_{\bar{\theta}}$. In the symmetric case, where $\lambda^{-1}_{\bar{\theta}} = \tau(\lambda_{\bar{\theta}})$ are Galois conjugates, we take $\vec{v}^*_{\bar{\theta}} = \tau(\vec{v}_{\bar{\theta}})$ to be Galois conjugates as well. To each symplectic orbit we also assign a field extension, $F_{\bar{\theta}} = F(\lambda_{\bar{\theta}} + \lambda^{-1}_{\bar{\theta}})$ (note that for $\bar{\theta}$ nonsymmetric $F(\lambda_{\bar{\theta}}) = F_{\bar{\theta}}$ and for $\bar{\theta}$ symmetric $[F(\lambda_{\bar{\theta}}) : F_{\bar{\theta}})] = 2$).

LEMMA A.5. *Let $\vec{n}, \vec{m} \in E$, and denote by $\vec{n}_{\bar{\theta}}, \vec{m}_{\bar{\theta}}$ their projection to $E_{\bar{\theta}}$. Then, the symplectic form*

$$\omega(\vec{n}_{\bar{\theta}}, \vec{m}_{\bar{\theta}}) = \mathrm{Tr}_{F_{\bar{\theta}}/F}(\kappa(\mu\nu^* - \nu\mu^*)),$$

where $\nu = \omega(\vec{n}, \vec{v}^*_{\bar{\theta}}), \nu^* = \omega(\vec{n}, \vec{v}_{\bar{\theta}})$, $\mu = \omega(\vec{m}, \vec{v}^*_{\bar{\theta}}), \mu^* = \omega(\vec{m}, \vec{v}_{\bar{\theta}})$, and $\kappa = \omega(\vec{v}_{\bar{\theta}}, \vec{v}^*_{\bar{\theta}})^{-1}$.

Proof. We prove first in the symmetric case. By Lemma A.3,

$$\vec{n}_{\bar{\theta}} = \mathrm{Tr}_{F(\lambda_{\bar{\theta}})/F}(\kappa\nu\vec{v}_{\bar{\theta}}) = \sum_\sigma \sigma(\kappa\nu\vec{v}_{\bar{\theta}}),$$

where the sum is over $\sigma \in \mathrm{Mor}_F(F(\lambda_{\bar{\theta}}), \bar{F})$. Therefore,

$$\begin{aligned}\omega(\vec{n}_{\bar{\theta}}, \vec{m}_{\bar{\theta}}) &= \omega(\textstyle\sum_\sigma \sigma(\kappa\nu\vec{v}_{\bar{\theta}}), \sum_{\sigma'} \sigma'(\kappa\mu\vec{v}_{\bar{\theta}}))\\ &= \textstyle\sum_{\sigma,\sigma'} \sigma(\kappa\nu)\sigma'(\kappa\mu)\omega(\sigma(\vec{v}_{\bar{\theta}}), \sigma'(\vec{v}_{\bar{\theta}}))\\ &= \mathrm{Tr}_{F(\lambda_{\bar{\theta}})/F}[\textstyle\sum_\sigma \kappa\nu\sigma(\kappa\mu)\omega(\vec{v}_{\bar{\theta}}, \sigma(\vec{v}_{\bar{\theta}}))].\end{aligned}$$

Now notice that $\omega(\vec{v}_{\bar{\theta}}, \sigma(\vec{v}_{\bar{\theta}})) \neq 0 \Leftrightarrow \sigma = \tau$, in which case $\omega(\vec{v}_{\bar{\theta}}, \tau(\vec{v}_{\bar{\theta}})) = \omega(\vec{v}_{\bar{\theta}}, \vec{v}^*_{\bar{\theta}}) = \kappa^{-1} = -\tau(\kappa^{-1})$ and $\tau(\nu) = \nu^*$. Therefore,

$$\omega(\vec{n}_{\bar{\theta}}, \vec{m}_{\bar{\theta}}) = \mathrm{Tr}_{F(\lambda_{\bar{\theta}})/F}(-\kappa\nu\mu^*) = \mathrm{Tr}_{F_{\bar{\theta}}/F}(\kappa(\mu\nu^* - \mu^*\nu)).$$

In the nonsymmetric case,

$$\vec{n}_{\bar{\theta}} = \mathrm{Tr}_{F_{\bar{\theta}}/F}(\kappa\nu\vec{v}_{\bar{\theta}}) + \mathrm{Tr}_{F_{\bar{\theta}}/F}(-\kappa\nu^*\vec{v}^*_{\bar{\theta}}).$$

Here $\omega(\vec{v}_{\bar{\theta}}, \sigma(\vec{v}_{\bar{\theta}})) = 0$ for all automorphisms, and $\omega(\vec{v}_{\bar{\theta}}, \sigma(\vec{v}_{\bar{\theta}})^*) \neq 0$ only if σ is the trivial automorphism. Hence, in this case as well

$$\omega(\vec{n}_{\bar{\theta}}, \vec{m}_{\bar{\theta}}) = \mathrm{Tr}_{F_{\bar{\theta}}/F}(\kappa(\nu^*\mu - \nu\mu^*)). \qquad \square$$

Appendix B. Counting elements in quotient rings

Let F be a number field and K/F a quadratic Galois extension. Denote by $\mathcal{O}_F, \mathcal{O}_K$ the corresponding integral rings. For any ideal $a \subseteq \mathcal{O}_F$, consider the map $\mathcal{N}_a : (\mathcal{O}_K/a\mathcal{O}_K)^* \to (\mathcal{O}_F/a)^*$ induced by the norm map $\mathcal{N}_{K/F}$, and let $\mathcal{C}(a) = \ker(\mathcal{N}_a)$ denote its kernel.

To each ideal $a \subseteq \mathcal{O}_F$ define:

$$S_1(a) = \sum_{\beta\in\mathcal{C}(a)} \sqrt{\#\{\nu \in \mathcal{O}_K/a\mathcal{O}_K | \nu(\beta-1) \equiv 0 \pmod{aO_K}\}},$$
$$S_2(a) = \#\{\beta_1, \beta_2 \in \mathcal{C}(a) | (1-\beta_1)(1-\beta_2)(\beta_1+\beta_2) \equiv 0 \,(\mathrm{mod}\ a\mathcal{O}_K)\}.$$

Eventually we will be interested in estimating these quantities for ideals of the form $N\mathcal{O}_F$ where $N \in \mathbb{N}$ are large integers. By the Chinese reminder theorem, if $a, b \subseteq \mathcal{O}_F$ are co-prime (i.e., $a + b = O_F$), then $\mathcal{O}_F/ab \cong \mathcal{O}_F/a \times \mathcal{O}_F/b$ and $\mathcal{O}_K/ab\mathcal{O}_K \cong \mathcal{O}_K/a\mathcal{O}_K \times \mathcal{O}_K/b\mathcal{O}_K$. Consequently, $\mathcal{C}(ab) \cong \mathcal{C}(a) \times \mathcal{C}(b)$ and the quantities S_1, S_2 are multiplicative (i.e., $S_i(ab) = S_i(a)S_i(b)$). Therefore, it suffices to calculate them for powers of prime ideals.

B.1. *Prime ideals.* In the following proposition we summarize some facts regarding factorization of ideals in extensions of number fields (for proofs and general background on the subject we refer to [5]).

PROPOSITION B.1. *Let K/F be an extension of number fields and $\mathcal{O}_K$, $\mathcal{O}_F$ the corresponding integral rings. Let $P \subseteq \mathcal{O}_F$ be a prime ideal, then the ideal $P\mathcal{O}_K$ decomposes into prime ideals of $\mathcal{O}_K$, $P\mathcal{O}_K = \prod_{i=1}^r \mathcal{P}_i^{e_i}$ where the ideals $\mathcal{P}_i$ are all the ideals lying above P (i.e., $\mathcal{P}_i \cap \mathcal{O}_F = P$). Furthermore*:

(1) *The fields $\mathcal{O}_K/\mathcal{P}_i$ are all finite field extensions of $\mathcal{O}_F/P$. The degree $[\mathcal{O}_K/\mathcal{P}_i : \mathcal{O}_F/P] = f_i$ is called the inertia degree. If the inertia degree $f_i = 1$, then $\forall k \in \mathbb{N}$ the corresponding rings are isomorphic $\mathcal{O}_K/\mathcal{P}_i^k \cong \mathcal{O}_F/P^k$.*

(2) *The exponent e_i is called the ramification index. When not all the ramification indices $e_i = 1$, the ideal P is said to be ramified in $\mathcal{O}_K$. For any number field $F/\mathbb{Q}$, there are only a finite number of ramified ideals (all lying above prime factors of the discriminant).*

(3) *The ramification indices e_i and the inertia degrees f_i satisfy $[K : F] = \sum_{i=1}^r e_i f_i$.*

(4) *If K/F is a Galois extension then all prime ideals of $\mathcal{O}_K$ lying above a prime ideal $P \subseteq \mathcal{O}_F$ are Galois conjugates, the ramification indices and the inertia degrees are fixed $e_i = e$, $f_i = f$, and the former equation takes the form $[K : F] = ref$.*

In particular, in our case $[K : F] = 2$; hence for any fixed prime ideal $P \subseteq \mathcal{O}_F$, there are only three possibilities:

(1) $P\mathcal{O}_K = \mathcal{P}\bar{\mathcal{P}}$ (P splits),

(2) $P\mathcal{O}_K = \mathcal{P}$ (P is inert),

(3) $P\mathcal{O}_K = \mathcal{P}^2$ (P is ramified),

where $x \mapsto \bar{x}$ denotes the nontrivial automorphism of K/F.

In the following proposition we describe the norm map $\mathcal{N}_{P^k}$ in each of these cases.

PROPOSITION B.2. *Let $\mathcal{P} \subseteq \mathcal{O}_K$ and $P = \mathcal{P} \cap \mathcal{O}_F$ be prime ideals.*

(1) *If $P\mathcal{O}_K = \mathcal{P}\bar{\mathcal{P}}$ splits, then $\mathcal{O}_K/P^k\mathcal{O}_K \cong \mathcal{O}_F/P^k \times \mathcal{O}_F/P^k$ as rings. Under this isomorphism, the norm map $\mathcal{N}_{P^k}$ induces the map*

$$\begin{array}{ccc} (\mathcal{O}_F/P^k)^* \times (\mathcal{O}_F/P^k)^* & \to & (\mathcal{O}_F/P^k)^* \\ (x, y) & \mapsto & xy. \end{array}$$

(2) *If $P\mathcal{O}_K = \mathcal{P}$ is inert, then the norm map $\mathcal{N}_{P^k}$ is onto.*

(3) *If $P\mathcal{O}_K = \mathcal{P}^2$ ramifies, then the image of $\mathcal{N}_{P^k}$ is a subgroup of $(\mathcal{O}_F/P^k)^*$ with index 2 if P lies above an odd prime, and index bounded by 2^{d+1} if it lies above 2.*

Proof. We prove for each case separately:

Part 1. When P splits, by the Chinese reminder theorem $\mathcal{O}_K/P^k\mathcal{O}_K \cong \mathcal{O}_K/\mathcal{P}^k \times \mathcal{O}_K/\bar{\mathcal{P}}^k$, and since the inertia degree $f = 1$, we can identify $\mathcal{O}_F/P^k \cong \mathcal{O}_K/\mathcal{P}^k \cong \mathcal{O}_K/\bar{\mathcal{P}}^k$. Under this identification the norm map N_{P^k} sends $(x, y) \in (\mathcal{O}_F/P^k)^* \times (\mathcal{O}_F/P^k)^*$ to $xy \in (\mathcal{O}_F/P^k)^*$.

Part 2. When P is inert we prove by induction on k. For $k = 1$, the inertia degree $[\mathcal{O}_K/\mathcal{P} : \mathcal{O}_F/P] = 2$ and the nontrivial automorphism of K/F induces the nontrivial automorphism of $(\mathcal{O}_K/\mathcal{P})/(\mathcal{O}_F/P)$. Consequently, the norm map $\mathcal{N}_P$ is the field extension norm map that is surjective for finite fields. For $k > 1$ by induction, let $\alpha \in (\mathcal{O}_F/P^k)^*$ and $\alpha_0 \in \mathcal{O}_F$ its representative. By induction $\exists \beta_0 \in \mathcal{O}_K$ such that $\mathcal{N}_{K/F}(\beta_0) \equiv \alpha_0 \pmod{P^{k-1}}$. Denote by $\eta = \mathcal{N}_{K/F}(\beta_0) - \alpha_0 \in P^{k-1}$ and let $x \in \mathcal{O}_K$ be an element such that $\mathrm{Tr}_{K/F}(\bar{\beta}_0 x) = -1 \pmod{P}$ (such an element exists because the trace for extension of finite fields is onto). Now, $\mathcal{N}_{K/F}(\beta_0 + \eta x) - \alpha_0 \in P^k$; hence, for $\beta = [\beta_0 + \eta x] \in \mathcal{O}_K/\mathcal{P}^k$ (the class of $\beta_0 + \eta x$), the norm map $\mathcal{N}_{P^k}(\beta) = \alpha$.

Part 3. When P is ramified and lies above an odd prime again by induction. For $k=1$, P ramified implies $[\mathbb{O}_K/\mathcal{P}:\mathbb{O}_F/P]=1$. Consequently, the nontrivial automorphism of K/F induces the trivial automorphism of $(\mathbb{O}_K/\mathcal{P})/(\mathbb{O}_F/P)$ and the induced map $\mathcal{N}_P$ (after identifying $\mathbb{O}_K/\mathcal{P}\cong\mathbb{O}_F/P$) is the squaring map $x\mapsto x^2$. When the ideal P lies above an odd prime p, the multiplicative group $(\mathbb{O}_F/P\mathbb{O}_F)^*$ is a cyclic group of an even order $(p^{f_P}-1)$ and the image of the map $x\mapsto x^2$ has index 2. For $k>1$ by induction, let $\alpha\in(\mathbb{O}_F/P^k)^*$ and $\alpha_0\in\mathbb{O}_F$ its representative. Then $\exists\beta_0\in\mathbb{O}_K$ such that $\eta=\xi\mathcal{N}_{K/F}(\beta_0)-\alpha_0\in P^{k-1}$, where ξ is a representative of one of the classes of $(\mathbb{O}_F/P^{k-1})^*/\mathrm{Im}(\mathcal{N}_{P^{k-1}})$. The map induced by $\mathrm{Tr}_{K/F}$ on $\mathbb{O}_K/\mathcal{P}\cong\mathbb{O}_F/P$ is simply multiplication by 2 and hence onto. We can thus take $x\in\mathbb{O}_K$ such that $\xi\mathrm{Tr}(\beta_0x)=-1\pmod{P}$. Now $\xi\mathcal{N}_{K/F}(\beta_0+x\eta)-\alpha_0\in P^k$, meaning α is in one of the two classes as well.

When P lies above 2, let h denote the largest integer such that $2\in P^h$. For any $\alpha\in\mathbb{O}_F$ we have that $\alpha^2\equiv1\pmod{P^k}$ implies $\alpha\equiv\pm1\pmod{P^{k-h}}$. Consequently, the kernel of squaring map has order bounded by $2|\mathbb{O}_F/P^h|\le 2|\mathbb{O}_F/2\mathbb{O}_F|\le 2^{d+1}$. □

B.2. *Counting elements.*

PROPOSITION B.3. *The number of norm one elements satisfies*

$$\left(\frac{N}{\log N}\right)^d \ll |\mathcal{C}(N\mathbb{O}_F)| \ll (N\log N)^d.$$

We first compute $|\mathcal{C}(P^k)|$ for $P\subseteq\mathbb{O}_F$, a prime ideal.

LEMMA B.4. *Let $P\in\mathbb{O}_F$ be a prime ideal lying above a rational prime $p\in\mathbb{Z}$. Then, if p is odd,*

$$|\mathcal{C}(P^k)|=|\mathbb{O}_F/P^k|\cdot\begin{cases}(1-\frac{1}{p^{f_P}}) & P \text{ splits}\\ (1+\frac{1}{p^{f_P}}) & P \text{ is inert}\\ \quad 2 & P \text{ is ramified,}\end{cases}$$

where $f_P=[\mathbb{O}_F/P:\mathbb{Z}/p\mathbb{Z}]$ is the inertia degree. If P lies above 2, *we can bound*

$$|\mathcal{C}(P^k)|\le 2^{d+1}|\mathbb{O}_F/P^k|.$$

Proof. We compute $|\mathcal{C}(P^k)|$ in each case separately.

Part 1. When P splits, by Proposition B.2 we can identify the group of norm one elements:

$$\mathcal{C}(P^k)\cong\left\{(x,y)\in(\mathbb{O}_F/P^k)^{*2}|xy=1\pmod{P^k}\right\}\cong(\mathbb{O}_F/P^k)^*.$$

Therefore, $|\mathcal{C}(P^k)|=|(\mathbb{O}_F/P^k)^*|=|\mathbb{O}_F/P^k|(1-\frac{1}{|\mathbb{O}_F/P|})$, and recall that $\mathbb{O}_F/P$ is the finite field with p^{f_P} elements.

Part 2. When P is inert, the map $\mathcal{N}_{P^k} : (\mathcal{O}_K/\mathcal{P}^k)^* \to (\mathcal{O}_F/P^k)^*$ is onto. Therefore,

$$|\mathcal{C}(P^k)| = |\ker(\mathcal{N}_{P^K})| = \frac{|(\mathcal{O}_K/\mathcal{P}^k)^*|}{|(\mathcal{O}_F/P^k)^*|} = \frac{|(\mathcal{O}_K/\mathcal{P}^k)||(1-\frac{1}{p^{f_{\mathcal{P}}}})|}{|(\mathcal{O}_F/P^k)||(1-\frac{1}{p^{f_P}})|}.$$

Now, the inertia degree $[\mathcal{O}_K/\mathcal{P} : \mathcal{O}_F/P] = 2$, which implies $f_{\mathcal{P}} = 2f_P$ and $|\mathcal{O}_F/\mathcal{P}| = |\mathcal{O}_F/P|^2$.

Part 3. For P ramified and odd, the image of $\mathcal{N}_{P^k}$ is of index 2 in $(\mathcal{O}_F/P^k)^*$. Therefore,

$$|\mathcal{C}(P^k)| = 2\frac{|(\mathcal{O}_K/\mathcal{P}^{2k})^*|}{|(\mathcal{O}_F/P^k)^*|} = 2\frac{|(\mathcal{O}_K/\mathcal{P}^{2k})||(1-\frac{1}{p^{f_{\mathcal{P}}}})|}{|(\mathcal{O}_F/P^k)||(1-\frac{1}{p^{f_P}})|}.$$

In this case the inertia degree $[\mathcal{O}_K/\mathcal{P} : \mathcal{O}_F/P] = 1$, so that $f_{\mathcal{P}} = f_P$ and $|\mathcal{O}_K/\mathcal{P}| = |\mathcal{O}_F/P|$. When P lies above 2, the image is of index bounded by 2^{d+1}, which implies the bound on $|\mathcal{C}(P^k)|$. □

We now give the proof of Proposition B.3 for composite N.

Proof. Let $N\mathcal{O}_F = \prod P_i^{k_i}$ be the decomposition to prime ideals. By the Chinese remainder theorem,

$$|\mathcal{C}(N\mathcal{O}_F)| = \prod_{i=1}^{r} |\mathcal{C}(P_i^{k_i})|.$$

Using Lemma B.4 for each component, for all prime ideals P_i, there is a common term of $|\mathcal{O}_F/P_i^{k_i}|$ that contributes precisely

$$\prod |\mathcal{O}_F/P_i^{k_i}| = |\prod(\mathcal{O}_F/P_i^{k_i})| = |\mathcal{O}_F/N\mathcal{O}_F| = N^d.$$

The additional contribution from the inert primes is bounded from below by 1 and from above by

$$\prod_i \left(1+\frac{1}{p^{f_{P_i}}}\right) \le \prod_{p|N}\left(1+\frac{1}{p}\right)^d \ll (\log N)^d$$

(since for every prime $p|N$, there are at most d ideal primes that lie above it). Similarly, the contribution from the split primes is bounded from above by 1 and from below by

$$\prod_i \left(1-\frac{1}{p^{f_{P_i}}}\right) \ge \prod_{p|N}\left(1-\frac{1}{p}\right)^d \gg \left(\frac{1}{\log N}\right)^d.$$

Finally, the contribution from the even and ramified primes is bounded by some constant (recall that there is a bounded number of ramified primes). □

Given a prime ideal $P \subset \mathcal{O}_F$ with ramification index $e \in \{1,2\}$ and any $1 \leq l \leq ek$, consider the congruence subgroup

$$\mathcal{C}^{(l)}(P^k) = \left\{\beta \in \mathcal{C}(P^k) | \beta \equiv 1 \pmod{\mathcal{P}^l}\right\},$$

where $\mathcal{P} \subset \mathcal{O}_K$ is a prime ideal above P (note that it is indeed well defined and does not depend on $\mathcal{P}$).

LEMMA B.5. *If P lies above an odd prime, then*

$$|\mathcal{C}^{(l)}(P^k)| = |\mathcal{O}_F/P|^{k-\lfloor \frac{l}{e} \rfloor}.$$

Otherwise,

$$|\mathcal{O}_F/P|^{k-\lfloor \frac{l}{e} \rfloor} \leq |\mathcal{C}^{(l)}(P^k)| \leq 2^{d+1}|\mathcal{O}_F/P|^{k-\lfloor \frac{l}{e} \rfloor}.$$

Proof. We prove it separately for P split inert or ramified.

Part 1. When P splits, we can identify

$$C(P^k) \cong \left\{(x, x^{-1}) \in (\mathcal{O}_F/P^k)^* \times (\mathcal{O}_F/P^k)^*\right\} \cong (\mathcal{O}_F/P^k)^*.$$

Denote by $(1 + P^l)/(1 + P^k)$ the kernel of the natural projection $(\mathcal{O}_F/P^k)^* \to (\mathcal{O}_F/P^l)^*$. Then, under this identification $C^{(l)}(P^k) \cong (1 + P^l)/(1 + P^k)$, and hence of order

$$|C^{(l)}(P^k)| = |(1 + P^l)/(1 + P^k)| = |\mathcal{O}_F/P|^{k-l}.$$

Part 2. For P inert, we let $\mathcal{N}^{(l)}_{P^k}$ denote the restriction of the norm map to $(1+\mathcal{P}^l)/(1+\mathcal{P}^k)$ (then $\mathcal{C}^{(l)}(P^k) = \ker(\mathcal{N}^{(l)}_{P^k})$). We now show that P odd $\mathcal{N}^{(l)}_{P^k}$ is onto $(1 + P^l)/(1 + P^k)$, whereas if P lies above 2, its image has index bounded by 2^{d+1} (this would conclude the proof for the inert case). First, the image of $\mathcal{N}^{(l)}_{P^k}$ is indeed a subgroup of $(1 + P^l)/(1 + P^k)$ (because if $\beta = 1 \pmod{\mathcal{P}^k}$, then $\mathcal{N}_{K/F}(\beta) = 1 \pmod{P^k}$). Next, note that the image of $\mathcal{N}^{(l)}_{P^k}$ contains all the squares in $(1 + P^l)/(1 + P^k)$. Now, for odd prime, $|(1 + P^l)/(1 + P^k)| = |\mathcal{O}_F/P|^{k-l}$ is a power of p and hence odd. Consequently, the map $x \mapsto x^2$ is an automorphism of $(1 + P^l)/(1 + P^k)$, and $\mathcal{N}^{(l)}_{P^k}$ is onto. When P lies above 2, the map $x \mapsto x^2$ has kernel bounded by $2|\mathcal{O}_F/P|^h$ (as in the proof of Proposition B.2). Consequently, the image of the squaring map (and hence also the image of $\mathcal{N}^{(l)}_{P^k}$) has index bounded by $2|\mathcal{O}_F/P|^h \leq 2^{d+1}$.

Part 3. For P ramified as in the previous case, we can restrict the norm map to the group $(1 + \mathcal{P}^l)/(1 + \mathcal{P}^{2k})$. Here (again by the squaring argument), the restricted map $\mathcal{N}^{(l)}_{P^k}$ is onto $(1+P^{\lceil \frac{l}{2}\rceil})/(1+P^k)$ for P odd and has image of index bounded by 2^{d+1} if $2 \in P$. Consequently, in this case for P odd,

$$|\mathcal{C}^{(l)}(P^k)| = |\mathbb{O}_F/P|^{k-\lfloor \frac{l}{2}\rfloor},$$

while for even prime ideals,

$$|\mathbb{O}_F/P|^{k-\lfloor \frac{l}{2}\rfloor} \le |\mathcal{C}^{(l)}(P^k)| \le 2^{d+1}|\mathbb{O}_F/P|^{k-\lfloor \frac{l}{2}\rfloor}. \qquad \square$$

PROPOSITION B.6. $S_1(N\mathbb{O}_F) \ll_\varepsilon N^{d+\varepsilon}$.

Again we start by computing $S_1(P^k)$ for powers of prime ideals.

LEMMA B.7. *Let $P \in \mathbb{O}_F$ be a prime ideal.*
If P lies above an odd prime, then

$$S_1(P^k) \le |\mathbb{O}_F/P^k| \cdot \begin{cases} (k+1) & P \text{ is inert or splits} \\ (k+1)\sqrt{|\mathbb{O}_F/P|} & P \text{ is ramified.} \end{cases}$$

If P lies above 2, then

$$S_1(P^k) \le 2^{d+2}|\mathbb{O}_F/P^k| \cdot \begin{cases} (k+1) & P \text{ is inert or splits} \\ (k+1)\sqrt{|\mathbb{O}_F/P|} & P \text{ is ramified.} \end{cases}$$

Proof. Let $e \in \{1,2\}$ be the ramification index of P in $\mathbb{O}_K$. The group $\mathcal{C}(P^k)$ decomposes into a disjoint union $\bigcup_{l=0}^{ek} \mathcal{C}^{(l)}(P^k) \setminus \mathcal{C}^{(l+1)}(P^k)$. We can thus rewrite

$$S_1(P^k) = \sum_{l=0}^{ek} \sum_{\mathcal{C}^{(l)}(P^k)\setminus\mathcal{C}^{(l+1)}(P^k)} \sqrt{\#\{\nu \in \mathbb{O}_K/P^k O_K | \nu(\beta-1)=0\}}.$$

For fixed l and any $\beta \in \mathcal{C}^{(l)}(P^k) \setminus \mathcal{C}^{(l+1)}(P^k)$, we have $\beta - 1 \in \mathcal{P}^l \setminus \mathcal{P}^{l+1}$. Therefore, the number of elements $\nu \in \mathbb{O}_K/P^k\mathbb{O}_K$ satisfying $\nu(\beta-1)=0$ is precisely $|\mathbb{O}_F/P|^{2l/e}$, independent of β. We can thus take it out of the sum to get

$$S_1(P^k) = \sum_{l=0}^{ek}(|\mathcal{C}^{(l)}(P^k)| - |\mathcal{C}^{(l+1)}(P^k)|)|\mathbb{O}_F/P|^{l/e}.$$

The result now follows directly from Lemma B.5. $\square$

We now give the proof of Proposition B.6 for composite N.

Proof. Decompose $N\mathbb{O}_F = \prod_{i=1}^r P_i^{k_i}$ into prime ideals. For each prime ideal apply Lemma B.7 to get the bound

$$S_1(N\mathbb{O}_F) = \prod_{i=1}^r S_1(P_i^{k_i}) \ll |\mathbb{O}_F/N\mathbb{O}_F| \prod_{i=1}^r (k_i+1),$$

where the implied constant comes from the contribution of the ramified and even prime ideals. The first term $|\mathcal{O}_F/N\mathcal{O}_F| = N^d$ and the second term can be bounded by $\prod_{i=1}^r (k_i+1) \ll_\varepsilon N^\varepsilon$, completing the proof. □

PROPOSITION B.8.

$$S_2(N\mathcal{O}_F) \ll_\varepsilon N^{d+\varepsilon}.$$

As before, we start by a computation for powers of prime ideals.

LEMMA B.9. *Let $P \in \mathcal{O}_F$ be a prime ideal. If P lies above an odd prime, then*

$$S_2(P^k) \leq |\mathcal{O}_F/P^k| \begin{cases} 6(k+1) & P \text{ is inert or splits} \\ 6(k+1)|\mathcal{O}_F/P| & P \text{ is ramified.} \end{cases}$$

If P is even, then

$$S_2(P^k) \leq 2^{4d} 6(k+1)|\mathcal{O}_F/P^k|.$$

Proof. First note that when P splits, the equation

$$(1-\beta_1)(1-\beta_2)(\beta_1+\beta_2) \equiv 0 \pmod{P^k\mathcal{O}_K},\ \beta_i \in \mathcal{C}(P^k)$$

is invariant under Galois conjugation. Thus, it is equivalent to the equation

$$(1-\beta_1)(1-\beta_2)(\beta_1+\beta_2) \equiv 0 \pmod{\mathcal{P}^k},\ \beta_i \in \mathcal{C}(P^k),$$

where $\mathcal{P}$ is a prime ideal above P. Therefore, in any case, $S_2(P^k)$ is the number of solutions to

$$\text{(B.1)} \qquad (1-\beta_1)(1-\beta_2)(\beta_1+\beta_2) \equiv 0 \pmod{\mathcal{P}^{ek}}, \quad \beta_i \in \mathcal{C}(P^k).$$

When P lies above an odd prime, then $2 \notin \mathcal{P}$ and $\beta_1 \equiv \beta_2 \equiv 1 \pmod{\mathcal{P}} \Rightarrow \beta_1+\beta_2 \equiv 2 \not\equiv 0 \pmod{\mathcal{P}}$. Therefore, the number of solutions to (B.1) is bounded by 3 times the number of solutions to

$$\text{(B.2)} \qquad (1-\beta_1)(1-\beta_2) \equiv 0 \pmod{\mathcal{P}^{ek}}, \quad \beta_i \in \mathcal{C}(P^k).$$

Any solution β_1, β_2 of (B.2) satisfies $\beta_1 \in \mathcal{C}^l(P^k) \setminus \mathcal{C}^{l+1}(P^k)$, $\beta_2 \in \mathcal{C}^{(ek-l)}(P^k)$ for some $0 \leq l \leq ek$. Thus the number of solutions is bounded by

$$S_2(P^k) \leq 3\sum_{l=0}^{ek} (|\mathcal{C}^l(P^k)| - |\mathcal{C}^{l+1}(P^k)|)|\mathcal{C}^{ek-l}(P^k)|,$$

and the result follows from Lemma B.5.

When $2 \in P$, denote by h the largest integer such that $2 \in P^h$ (so that $\mathcal{P}^{eh}|2\mathcal{O}_K$). Now, if $\beta_1 \equiv \beta_2 \equiv 1 \pmod{\mathcal{P}^{eh+1}}$, then $\beta_1+\beta_2 \not\equiv 0 \pmod{\mathcal{P}^{eh+1}}$. Therefore, as in the case of the odd prime, the number of solutions to (B.1) is bounded by 3 times the number of solutions to

$$\text{(B.3)} \qquad (1-\beta_1)(1-\beta_2) \equiv 0 \pmod{\mathcal{P}^{ek-eh}}, \quad \beta_i \in \mathcal{C}(P^k).$$

Now, any such solution satisfies

$$\beta_1 \in \mathcal{C}^{(l)}(P^k) \setminus \mathcal{C}^{(l+1)}(P^k) \quad \text{and} \quad \beta_2 \in \mathcal{C}^{(ek-eh-l)}(P^k)$$

for some $0 \le l \le ek - eh$; hence,

$$S_2(P^k) \le 3 \sum_{l=0}^{ek-eh} (|\mathcal{C}^l(P^k)| - |\mathcal{C}^{(l+1)}(P^k)|)|\mathcal{C}^{ek-eh-l}(P^k)|,$$

and the result follows from Lemma B.5. □

Now for the general case.

Proof. Decompose $N\mathcal{O}_K = \prod_{i=1}^t P_i^{k_i}$ and apply Lemma B.9 for each component

$$S_2(N\mathcal{O}_K) = \prod_{i=1}^r S_2(P_i^{k_i}) \ll \prod_{i=1}^r |\mathcal{O}_F / P_i^{k_i}| 6(k_i + 1) = N^d \prod_{i=1}^r 6(k_i + 1),$$

where the implied constant comes from the even and ramified ideals. The estimate $\prod_{i=1}^r 6(k_i + 1) \ll_\varepsilon N^\varepsilon$ concludes the proof. □

Acknowledgments

I warmly thank my Ph.D. advisor Zeev Rudnick for introducing me to this subject and for his guidance throughout this project. I thank Par Kurlberg for long discussions of his work. I thank Shamgar Gurevich and Ronny Hadani for patiently explaining their work and for their stimulating suggestions. I also thank Stéphane Nonnenmacher for his helpful comments. I thank my friends Lior Rosenzweig and Lior Bary-Soroker for our discussions. This work was supported in part by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities. This work was carried out as part of the author's Ph.D. thesis at Tel Aviv University, under the supervision of Prof. Zeev Rudnick.

References

[1] E. Bogomolny and C. Schmit, Multiplicities of periodic orbit lengths for non-arithmetic models, *J. Phys. A* **37** (2004), 4501–4526. MR 2005f:81073 Zbl 1050.37008

[2] E. Bombieri, On exponential sums in finite fields, *Amer. J. Math.* **88** (1966), 71–105. MR 34 #166 Zbl 0171.41504

[3] F. Bonechi and S. De Bièvre, Controlling strong scarring for quantized ergodic toral automorphisms, *Duke Math. J.* **117** (2003), 571–587. MR 2004i:81079 Zbl 1049.81028

[4] A. Bouzouina and S. De Bièvre, Equipartition of the eigenfunctions of quantized ergodic maps on the torus, *Comm. Math. Phys.* **178** (1996), 83–105. MR 97b:81023 Zbl 0876.58041

[5] H. Cohn, *A Classical Invitation to Algebraic Numbers and Class Fields*, Springer-Verlag, New York, 1978. MR 80c:12001 Zbl 0395.12001

[6] S. DE BIÈVRE and M. DEGLI ESPOSTI, Egorov theorems and equidistribution of eigenfunctions for the quantized sawtooth and baker maps, *Ann. Inst. H. Poincaré Phys. Théor.* **69** (1998), 1–30. MR 99g:58119 Zbl 0922.58074

[7] M. DEGLI ESPOSTI and S. GRAFFI, Mathematical aspects of quantum maps, in *The Mathematical Aspects of Quantum Maps*, *Lecture Notes in Phys.* **618**, Springer-Verlag, New York, 2003, pp. 49–90. MR 2159323 Zbl 1058.81542

[8] M. DEGLI ESPOSTI, S. GRAFFI, and S. ISOLA, Classical limit of the quantized hyperbolic toral automorphisms, *Comm. Math. Phys.* **167** (1995), 471–507. MR 96c:81054 Zbl 0822.58022

[9] B. ECKHARDT, S. FISHMAN, J. KEATING, O. AGAM, J. MAIN, and K. MÜLLER, Approach to ergodicity in quantum wave functions, *Phys. Rev. E* **52** (1995), 5893–5903.

[10] F. FAURE, S. NONNENMACHER, and S. DE BIÈVRE, Scarred eigenstates for quantum cat maps of minimal periods, *Comm. Math. Phys.* **239** (2003), 449–492. MR 2005a:81076 Zbl 1033.81024

[11] M. FEINGOLD and A. PERES, Distribution of matrix elements of chaotic systems, *Phys. Rev. A* **34** (1986), 591–595. MR 87i:81055

[12] M. D. FRIED and M. JARDEN, *Field Arithmetic*, second ed., *Ergeb. Math. Grenzgeb.* **11**, Springer-Verlag, New York, 2005. MR 2005k:12003 Zbl 1055.12003

[13] P. GÉRARDIN, Weil representations associated to finite fields, *J. Algebra* **46** (1977), 54–101. MR 57 #470 Zbl 0359.20008

[14] S. GUREVICH, Weil representation, Deligne sheaf, and proof of the Kurlberg-Rudnick conjecture, Ph.D. thesis, Tel-Aviv University, 2005.

[15] S. GUREVICH and R. HADANI, The higher-dimensional Rudnick-Kurlberg conjecture, 2004, preprint. arXiv math-ph/0409031

[16] ———, Proof of the Kurlberg-Rudnick rate conjecture, in *p-adic Mathematical Physics*, AIP *Conf. Proc.* **826**, Amer. Inst. Phys., Melville, NY, 2006, pp. 74–80. MR 2008f:58033 Zbl 1152.58310

[17] J. H. HANNAY and M. V. BERRY, Quantization of linear maps on a torus-Fresnel diffraction by a periodic grating, *Phys. D* **1** (1980), 267–290. MR 82g:81005

[18] L. K. HUA and I. REINER, On the generators of the symplectic modular group, *Trans. Amer. Math. Soc.* **65** (1949), 415–426. MR 10,684d Zbl 0034.30503

[19] H. IWANIEC, *Spectral Methods of Automorphic Forms*, second ed., *Grad. Studi. Math.* **53**, Amer. Math. Soc., Providence, RI, 2002. MR 2003k:11085 Zbl 1006.11024

[20] D. KELMER, On the quantum variance of matrix elements for the cat map on the 4-dimensional torus, *Int. Math. Res. Not.* (2005), 2223–2236. MR 2007j:81066 Zbl 05004181

[21] S. KNABE, On the quantisation of Arnold's cat, *J. Phys. A* **23** (1990), 2013–2025. MR 91f:81086 Zbl 0715.58014

[22] P. KURLBERG, L. ROSENZWEIG, and Z. RUDNICK, Matrix elements for the quantum cat map: fluctuations in short windows, *Nonlinearity* **20** (2007), 2289–2304.

[23] P. KURLBERG and Z. RUDNICK, Hecke theory and equidistribution for the quantization of linear maps of the torus, *Duke Math. J.* **103** (2000), 47–77. MR 2001f:11065 Zbl 1013.81017

[24] ———, On the distribution of matrix elements for the quantum cat map, *Ann. of Math.* **161** (2005), 489–507. MR 2006h:81091 Zbl 1082.81054

[25] W. C. W. LI, *Number Theory with Applications*, *Series on University Mathematics* **7**, World Scientific Publishing Co., River Edge, NJ, 1996. MR 98b:11001 Zbl 0849.11006

[26] C. MŒGLIN, M. F. VIGNÉRAS, and J. L. WALDSPURGER, *Correspondances de Howe sur un Corps p-Adique*, *Lecture Notes in Math.* **1291**, Springer-Verlag, New York, 1987. MR 91f:11040

[27] M. NEUHAUSER, An explicit construction of the metaplectic representation over a finite field, *J. Lie Theory* **12** (2002), 15–30. MR 2003e:20014 Zbl 1026.22018

[28] S. NONNENMACHER, private communication, 2003.

[29] A. M. F. RIVAS, M. SARACENO, and A. M. OZORIO DE ALMEIDA, Quantization of multidimensional cat maps, *Nonlinearity* **13** (2000), 341–376. MR 2002a:37086 Zbl 0952.37016

[30] Z. RUDNICK and P. SARNAK, The behaviour of eigenstates of arithmetic hyperbolic manifolds, *Comm. Math. Phys.* **161** (1994), 195–213. MR 95m:11052 Zbl 0836.58043

[31] A. I. ŠNIREL'MAN, Ergodic properties of eigenfunctions, *Uspehi Mat. Nauk* **29** (1974), 181–182. MR 53 #6648

[32] S. ZELDITCH, Uniform distribution of eigenfunctions on compact hyperbolic surfaces, *Duke Math. J.* **55** (1987), 919–941. MR 89d:58129 Zbl 0643.58029

(Received January 4, 2006)

E-mail address: kelmerdu@math.uchicago.edu
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO, 5734 UNIVERSITY AVENUE, CHICAGO, IL 60637-1514, UNITED STATES,
http://www.math.uchicago.edu/~kelmerdu/