RIBET'S CONVERSE TO HERBRAND'S THEOREM

CARL ERICKSON

Contents

Date: May 2008.

1. Introduction

In 1976 K. Ribet [20] proved a refinement of Kummer's criterion for the regularity of an odd prime p. Kummer's criterion relates the condition that p is regular, i.e. that the ideal class group of the cyclotomic field $\mathbb{Q}(\mu_p)$ has no p-torsion, to the p-divisibility of important analytic quantities called Bernoulli numbers. J. Herbrand [12] in 1932 proved one direction of a more precise version of Kummer's criterion. Namely, Herbrand's theorem states that if one decomposes the p-part of the class group in terms of the action of $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$ on it, a certain Bernoulli number is divisible by p if a corresponding part of the decomposition is non-trivial. Herbrand's theorem uses Stickelberger's theorem and classical algebraic number theory. Ribet proved the converse using techniques in arithmetic geometry: given a certain Bernoulli number divisible by p, he produces a Galois representation that cuts out an unramified p-extension of $\mathbb{Q}(\mu_p)$ associated via class field theory to such a non-trivial part of the class group. This essay is primarily intended as a description of Ribet's techniques, focusing especially on his congruences between modular forms and on one of Ribet's major tools, the Eichler-Shimura relation. However, we will still attempt to "connect all the dots." Let us begin by giving background to make precise what I have said above and to motivate Ribet's work.

1.1. Background on Cyclotomic Fields.

In 1851, E. Kummer proved Fermat's Last Theorem for a large number of odd prime exponents p. These primes, called *regular* primes, are those odd primes p such that p does not divide the ideal class number of $\mathbb{Q}(\mu_p)$, where μ_p is a primitive pth root of unity. For if p is regular, one can argue by contradiction as follows.

Let $x, y, z \in \mathbb{Z}$ be a non-trivial, pairwise coprime counterexample to Fermat's last theorem, i.e.

$$x^p + y^p = z^p,$$

and furthermore assume $p \nmid xy$. Factoring these integers in $\mathbb{Z}[\mu_p]$ yields an equality of principal integral ideals,

$$\prod_{i=0}^{p-1}(x + \mu_p^i y) = (z)^p, \tag{1.1}$$

One may showing that the factors on the left of (1.1) are relatively prime ideals; consequently, each factor is a pth power of some ideal $\mathfrak{a}_i$. Then, on the critical hypothesis that p is regular, each $\mathfrak{a}_i$ is a principal ideal. Arguing in terms of the generators of these ideals then provides a basic proof of the rest of Fermat's last theorem (see for example [27], Chs. 1,9).

This connection with Fermat's last theorem is one of the many reasons that cyclotomic fields figure prominently in algebraic number theory. A few more reasons include

- The ring of integers of a cyclotomic field is well understood, in contrast to most high-degree number fields. The maximal order of the nth cyclotomic field $\mathbb{Q}(\mu_n)$ is $\mathbb{Z}[\mu_n]$, whereas the ring of integers of an arbitrary high-degree number field may be virtually impossible for a computer to determine. Consequently, prime decomposition behavior in cyclotomic fields is well understood.
- The nth cyclotomic field is an abelian Galois extensions of $\mathbb{Q}$ with galois group $(\mathbb{Z}/n\mathbb{Z})^\times$. Since by the Kronecker-Weber theorem all abelian extensions of $\mathbb{Q}$ lie in

some cyclotomic extension, class field theory over $\mathbb{Q}$ may be reduced to cyclotomic fields. See for example [14].

- They and their generalizations, CM-fields, were among the first families of number fields of arbitrarily high degree about which we can say anything very interesting about their class groups and regulators (see Appendix A.1).

It is the last type of fact in the above list that we would like to know, and Kummer proved, in order to say more about Fermat's last theorem. Namely, he gave "Kummer's criterion" for when a prime p is regular.

Theorem 1.1 (Kummer; [27], Thm. 5.34)**.** *An odd prime p is irregular if and only if there exists an even integer $2 \leq k \leq p-3$ such that p divides the numerator of the kth Bernoulli number B_k, given by the Taylor series*

$$\frac{t}{e^t - 1} = \sum_{n \geq 0} \frac{B_n}{n!} t^n. \tag{1.2}$$

Remark 1.2. Proving Kummer's criterion in its original setting is very interesting but would take us too far afield. Thus Appendix A.1 proves one direction of Kummer's criterion. Also, this appendix provides lemmas on relations between Bernoulli numbers and the class number of $\mathbb{Q}(\mu_p)$, Ribet's use of which we describe in §3.

Certainly Kummer's criterion is a good computational tool for finding regular primes; mathematicians have exploited it heavily. For instance, the first few Bernoulli numbers are

$$B_2 = \frac{1}{6}, B_4 = \frac{-1}{30}, B_6 = \frac{1}{42}, B_8 = \frac{-1}{30}, B_{10} = \frac{5}{66}, B_{12} = \frac{-691}{2730}, \tag{1.3}$$

allowing us to verify that 691 is an irregular prime and that $3, 5, 7, 11, 13$ are regular. However, the key theoretical importance of Kummer's criterion is that, as I noticed professors often comment, *Bernoulli numbers are analytic objects*. That is, generalized Bernouli numbers appear as special values of L-functions, and the Bernoulli numbers are associated with the simplest L-function:

$$\zeta(1-n) = \frac{-B_n}{n}, \quad n = 1, 2, 3, \ldots \tag{1.4}$$

where $\zeta(s)$ is the Riemann zeta function. We may therefore restate Kummer's criterion in this

Corollary 1.3. *An odd prime p is irregular if and only if there exists an even integer $2 \leq k \leq p-3$ such that p divides the numerator of $\zeta(1-k)$.*

A great deal of advances in number theory have to do with linking special values of L-functions to arithmetic problems. Ribet's converse to Herbrand's theorem is one such result.

1.2. **The Converse to Herbrand's Theorem.** Herbrand [12] proved a refinement of Kummer's criterion, showing that the p-divisibility of a specific Bernoulli number could only occur if a corresponding character occurs in the action of $\mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$ on the p-part of the class group of $\mathbb{Q}(\mu_p)$. While Herbrand used classical algebraic number theoretic tools, Ribet in 1976 proved the converse by applying newly discovered techniques in arithmetic geometry. Ribet's proof is the subject of this essay, and we will investigate

the constitution of his proof in detail. First we will get clear on precisely what Herbrand and Ribet proved.

Let us fix the following notation, following Ribet's paper [20]. Let A be the class group of $\mathbb{Q}(\mu_p)$ and let C be the $\mathbb{F}_p$-vector space A/A^p where $\mathbb{F}_p$ is the finite field with p elements. The $\mathbb{F}_p$-vector space structure is induced by the structure of A as a $\mathbb{Z}$-module. Note that

$$\dim_{\mathbb{F}_p} A/A^p = p\text{-rank of } A.$$

Clearly the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts on C through its abelian quotient $\Delta = \mathrm{Gal}(\mathbb{Q}(\mu_p)/\mathbb{Q})$. Later (§5.1), we will show that whenever a representation $\rho : G \longrightarrow GL_2(\bar{\mathbb{F}}_p)$ has finite image, it is semisimple if and only if the order of its image is prime to p. Hence because $|\Delta| = p-1$, the Galois representation C is semisimple. Therefore it is a direct sum of powers of the standard Galois character

$$\chi : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \Delta \xrightarrow{\sim} \mathbb{F}_p^\times \hookrightarrow \bar{\mathbb{F}}_p^\times, \tag{1.5}$$

given by the relation

$$\sigma(\mu_p) = \mu_p^{\chi(\sigma)}, \quad \text{for all } c \in C.$$

Thus the direct sum of powers of χ that composes C may be canonically written

$$C = \bigoplus_{i \pmod{p-1}} C(\chi^i), \tag{1.6}$$

where $C(\chi^i)$ is the χ^i-isotypical component of C as a Δ-module. That is, $C(\chi^i)$ is the subspace of c in C such that $\sigma(c) = \chi^i(\sigma) \cdot c$. To observe that this decomposition is canonical, note, for example, that $C(\chi)$ is the unique subspace of elements $c \in C$ such that Δ acts on the $\mathbb{F}_p$-modules $C(\chi)$ and $\langle \mu_p \rangle$ in the same way.

Ribet proved the following

Theorem 1.4 ([20], Theorem 1.1)**.** *Let k be an even integer, $2 \leq k \leq p-3$. Then p divides the numerator of B_k if an only if $C(\chi^{1-k}) \neq 0$.*

Ribet completed the proof of this equivalence by showing that if p divides B_k, then $C(\chi^{1-k}) \neq 0$. Classically, Herbrand had proved the converse by refining Stickelberger's theorem (see [27], §§6.2-6.3). Together, the results of Herbrand and Ribet describe the action of Δ on the p-part of the class group of $\mathbb{Q}(\mu_p)$ in terms of analytic quantities.

Of course, it is not a complete description. For example, the quantity of $2 \leq k \leq p-3$ such that $p \mid B_k$ is a lower bound on the p-rank of C, but as long as p is irregular there is no a priori upper bound.[1] Also, note that this theorem applies only to even integers k. The question of whether $C(\chi^{1-k}) \neq 0$ for an odd k is the same as the question of the truth of Vandiver's conjecture that the class number $\mathbb{Q}(\mu_p)^+$ is prime to p (see [27] for information on Vandiver's conjecture). In all known examples, Vandiver's conjecture holds, and actually, Ribet's theorem is a consequence of the truth of Vandiver's conjecture [15].

There are other ways in which Ribet's result could superficially appear to be uninteresting. For example, it followed from the main conjecture of Iwasawa theory proved by Mazur and Wiles [19] in 1984 [15], and also in an even more elementary fashion from techniques in Euler systems developed by V. Kolyvagin [16] and F. Thaine [26] (in [27],

[1]The Iwasawa main conjecture proved by B. Mazur and A. Wiles [19] implies that $C(\chi^i)$ is one-dimensional. Such topics will be discussed further in the conclusion.

§15.2). Yet as C. Khare [15] notes, "The proof of Ribet is still valuable as it *explicitly* constructs abelian unramified extensions of exponent p of $\mathbb{Q}(\mu_p)$ with controlled behavior." Furthermore, Ribet's strategy was expanded upon by Wiles successively throughout the 1980s, culminating in the Iwasawa conjecture for totally real fields [30].

Let's take a look at Ribet's overall strategy.

1.3. **Ribet's Proof of the Converse.** From the summary above we know that Ribet's proof depends on the geometric construction of a certain Galois representation. Here we summarize more deeply, describing Ribet's strategy and the way it will be presented in this essay.

The first step is to understand how the construction of a certain special representation imply the converse to Herbrand's theorem. I consider the proof rather interesting, because I had encountered Galois representations in the number theoretic atmosphere around me, but did not understand how they are canonical enough to be number theoretically applicable. In §2 we will add flesh to Ribet's treatment of deducing the converse to Herbrand from the existence of the representation, which, naturally, is brief because it depends on basic facts about Galois representation and class field theory.

Once we know that the existence of such a special representation (Theorem 2.3 below) implies the converse to Herbrand's theorem, it remains to construct the representation. We will accomplish this in three steps: constructing in §3 a cusp eigenform congruent modulo a prime over p to the Eisenstein series whose constant cofficient is B_k; associating to it in §4 its abelian variety A_f and the accompanying p-adic Galois representations, and finally in §5 showing via the Eichler-Shimura relation's connection between the modular form and the representation that the reduction of this representation is of the correct form. We will focus on two parts of the proof. Firstly, we will follow Ribet directly to give a detailed explanation of how to construct such a cusp form. But our second point of emphasis is one that Ribet merely quotes: it is the Eichler-Shimura relation (Theorem 4.19), which studies the modular curves associated to the cusp form and shows that the action of Hecke operators on the cusp form is equivalent to the action of Frobenius on the reduction of the modular curve modulo p.

At this point it is not possible to give much more motivating detail, as these details must be built up. However, I believe it is possible to give a few key motivating statements that, while broad and imprecise, are extremely helpful for understanding what is going on.

Let us suppose that $p \mid B_k$. We then construct in §3 a cusp eigenform that is congruent modluo p to the Eisenstein series G_k of weight k on $\mathrm{SL}_2(\mathbb{Z})$. Note well that we have already used the fact that $p \mid B_k$, since the constant coefficient of G_k is B_k (up to a p-unit) and cusp forms have no constant coefficient. Now from Shimura's construction of the abelian variety associated to the eigenform, we will get by the end of §4 a Galois representation

$$\rho : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(K_{\mathfrak{p}})$$

where $K_{\mathfrak{p}}$ is a finite extension of the p-adic numbers $\mathbb{Q}_p$ and $V_{\mathfrak{p}}$ is a two- dimensional $K_{\mathfrak{p}}$-representation called the Tate module of the abelian variety. This representation will not be diagonalizable, and in fact we will prove that it is irreducible. However, the Eichler-Shimura relation gives us a connection between the coefficients of the modular form and this Galois action. Therefore, just as our cusp eigenform "looks like" a cusp form

modulo p, so will the reduction modulo p of the representation "look like" a representation coming from an Eisenstein series. The representation associated to Eisenstein series ${G_k}^2$ is $1 \oplus \chi^{k-1}$ modulo p ([11], Thm. 9.6.6). From this fact we can show that the reduction of ρ modulo p is reducible and of the form

$$\begin{pmatrix} 1 & * \\ 0 & \chi^{k-1} \end{pmatrix} \tag{1.7}$$

This reduction is obtained by choosing a Galois-stable lattice $T \subset V_{\mathfrak{p}}$ and considering the action of Galois on $T/\mathfrak{p}T$. Ribet's clever work on reducible reductions of simple $K_{\mathfrak{p}}$-representations, which will be discussed in §5.2, then implies that there exists a T such that the representation in Equation (1.7) is not semi-simple, i.e. not diagonalizable, and as mentioned above it follows that the image of $\bar{\rho}$ has order divisible by p. As this image is the Galois group of some normal extension of $\mathbb{Q}$ (see Definition 2.4), it turns out that it is precisely these elements of order p that correspond to p-extensions of $\mathbb{Q}(\mu_p)$. The most difficult part of Ribet's work is to show that this part of the representation (ergo the p-extensions) are unramified at p. The geometric details of Ribet's approach are beyond the scope of this essay.

1.4. **Preliminaries.** I have intended to write this essay toward an audience of number theory students in Part III while still keeping it to a reasonable length. Thus I presume comfort with notions from algebraic geometry such as Riemann-Roch, Picard groups, etc., the basic concepts of representation theory, Riemann surfaces and spaces of differentials on them, algebraic number theory including infinite Galois theory and L-functions, and elliptic curves at the level of the Part III course. Preliminaries on modular forms and other especially critical topics are given here according to their level of importance, but hastily. I try to indicate sources for both the original major advances and sources that help the student like me get a grip on them. Likewise, I have tried, mostly in §1.3 above, to give a good deal of the "philosophy" involved in how I came to understand the material.

1.5. **Acknowledgments and Sources.** This document was originally written as an essay fulfilling the requirements of one exam for my Cambridge Part III course in 2007-2008. I'm very appreciative to my essay supervisor Dr. Tobias Berger for the interesting topic and helpful meetings.

Of course, none of the core material is original to me. Proofs are drawn from the sources cited. The reader will note that I am especially dependent upon Ribet's paper [20], *A Modular Construction of Unramified p-extensions of* $\mathbb{Q}(\mu_p)$, and on F. Diamond and J. Shurman's [11] *A First Course in Modular Forms*. The other sources I consulted first hand were Washington's book on cyclotomic fields [27], Shimura's book on Automorphic Forms [23], Khare's notes on Ribet's proof [15], and to a lesser extent [17], [4], [10], [18], and [24]. Other sources cited are the original works cited in the sources I consulted, or advanced articles that I used to write the conclusion (§6) or refer to as tangents.

[2]Since in this essay we only know about representations assocaited to modular forms of weight 2, we should clarify that this representation comes from from the Eisenstein series $G_{2,\varepsilon}$ of weight 2 which is congruent modulo p to G_k. Deligne's [6] work is needed to associate a representation to G_k.

2. Extensions of $\mathbb{Q}(\mu_p)$ and Representations

Ribet's"main theorem," our Theorem 1.4, describes the action of Galois on the class group A of $\mathbb{Q}(\mu_p)$. However, his efforts are dedicated to constructing a special Galois representation. Therefore we will begin as Ribet does, addressing why the existence of this representation implies the main theorem. There are two steps: first, we will use class field theory to write down the theorem in terms of extensions of $\mathbb{Q}$. Then we will show that this form of the theorem follows from the existence of the representation.

Let us without delay state the following theorem, which, as we will prove, is equivalent to the main theorem, Theorem 1.4.

Theorem 2.1 ([20], Theorem 1.2)**.** *Suppose $p \mid B_k$. Then there exists a Galois extension $E/\mathbb{Q}$ containing $\mathbb{Q}(\mu_p)$ such that*

(1) *The extension $E/\mathbb{Q}(\mu_p)$ is unramified.*

(2) *The group $\mathrm{Gal}(E/\mathbb{Q}(\mu_p))$ is a non-trivial abelian group of type $(p, \ldots, p)$, i.e. killed by p.*

(3) *If $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$ and $\tau \in \mathrm{Gal}(E/\mathbb{Q}(\mu_p))$ then $\sigma\tau\sigma^{-1} = \chi(\sigma)^{1-k} \cdot \tau$.*

Of course, this theorem will be proved later. For now, we show that it is equivalent to the main theorem.

Proposition 2.2. *Theorem 2.1 is equivalent to the main theorem, Theorem 1.4.*

Proof. This is an exercise in class field theory. See G. Janusz's book [14] for a well presented version of classical class filed theory including the facts we now require.

The equivalence $C \neq 0$ if and only if parts (1) and (2) of Theorem 2.3 clearly follows from the definition of the Hilbert class field and the fact that the Artin map is an isomorphism. Thus it remains to show that part (3) is equivalent to the $C(\chi^{1-k})$ part of C being nontrivial. We will accomplish this via the "functorality of the Artin symbol," which when applied to the present case states that

$$\sigma \left[\frac{E/\mathbb{Q}(\mu_p)}{\mathfrak{a}}\right] \sigma^{-1} = \left[\frac{E/\mathbb{Q}(\mu_p)}{\sigma\mathfrak{a}}\right] \tag{2.1}$$

where $\mathfrak{a}$ is a fractional ideal of $\mathbb{Q}(\mu_p)$, σ is an element of $\mathrm{Gal}(E/\mathbb{Q})$ (though it is clear from the right hand side of the equality that its action depends only on which coset modulo Δ it belongs to), and $\left[\frac{E/\mathbb{Q}(\mu_p)}{\cdot}\right]$ is the Artin symbol for the unramified abelian extension $E/\mathbb{Q}(\mu_p)$.

Choose some $\tau \in \mathrm{Gal}(E/\mathbb{Q}(\mu_p))$ and let H be the Hilbert class field of $\mathbb{Q}(\mu_p)$. The Artin symbol for $E/\mathbb{Q}(\mu_p)$ is a quotient of the symbol for $H/\mathbb{Q}(\mu_p)$. Hence, just as the former symbol is surjective by Takagi's existence theorem, so is the latter. Therefore there exists a fractional ideal $\mathfrak{a}$ of $\mathbb{Q}(\mu_p)$ that the Artin symbol for $E/\mathbb{Q}(\mu_p)$ sends $\mathfrak{a}$ to τ, i.e. $\tau = \left[\frac{E/\mathbb{Q}(\mu_p)}{\mathfrak{a}}\right]$. Assuming that there is some k for which the relation $\sigma\tau\sigma^{-1} = \chi^{1-k} \cdot \tau$ holds, then by the functorality relation (2.1) we have

$$\left[\frac{E/\mathbb{Q}(\mu_p)}{\sigma\mathfrak{a}}\right] = \sigma\tau\sigma^{-1} = \chi(\sigma)^{1-k} \cdot \tau = \chi(\sigma)^{k-1} \cdot \left[\frac{E/\mathbb{Q}(\mu_p)}{\mathfrak{a}}\right] = \left[\frac{E/\mathbb{Q}(\mu_p)}{\chi^{1-k}(\sigma) \cdot \mathfrak{a}}\right]. \tag{2.2}$$

The equality of the leftmost and rightmost terms implies that $\chi(\sigma)^{1-k}\mathfrak{a}$ is in the same ideal class as $\sigma\mathfrak{a}$ modulo the kernel of the Artin symbol for $E/\mathbb{Q}(\mu_p)$. Since this kernel is A^p (recall $C = A/A^p$), we know that if there is some $\tau \in \mathrm{Gal}(E/\mathbb{Q}\mu_p)$ such that $\sigma\tau\sigma^{-1} = \chi^{1-k}(\sigma)\cdot\tau$ then $C(\chi^{1-k}) \neq 0$. One may easily check that the converse follows from formula (2.2) as well, completing the proof. □

Theorem 2.1, which restates the main theorem in terms of Galois extensions, follows from the existence of the Galois representation that Theorem 2.3 below claims to exist. I can say that understanding why Theorem 2.3 implies Theorem 2.1 was valuable as an exercise because it is an example of how a the existence of a certain Galois representation and certain number fields are connected. This idea is useful, for example, in showing that some Galois representations cannot exist, e.g. [25], which was a first step towards Serre's conjecture.

We should note that Theorem 2.1 and Theorem 2.3 are not a priori equivalent; the existence of certain number field extensions does not, as far as I know, imply the existence of a *specific*, much less *modular*, representation that cuts them out. However, we do know that these theorems are equivalent because of Herbrand's theorem.

This theorem is the true "main theorem" of Ribet's paper. It establishes the existence of a certain Galois representation that cuts out exactly the kind of number fields we need to prove Theorem 2.1, and will take the rest of our efforts to prove.

Theorem 2.3 ([20], Theorem 1.3)**.** *Suppose $p \mid B_k$. Then there exists a finite field $\mathbb{F} \supseteq \mathbb{F}_p$ and a continuous representation*

$$\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}) \tag{2.3}$$

such that

(A) *$\bar{\rho}$ is unramified at all primes $\ell \neq p$.*

(B) *The representation $\bar{\rho}$ is reducible (over $\mathbb{F}$) in such a way that $\bar{\rho}$ is isomorphic to a representation of the form*

$$\begin{pmatrix} 1 & * \\ 0 & \chi^{k-1} \end{pmatrix}.$$

That is, $\bar{\rho}$ is an extension of the 1-dimensional representation with character χ^{k-1} by the trivial 1-dimensional representation.

(C) *The image of $\bar{\rho}$ has order divisible by p. In other words, $\bar{\rho}$ is not diagonalizable.*

(D) *Let D_p be a decomposition group for p is $G_{\mathbb{Q}}$. Then $\bar{\rho}(D_p)$ has order prime to p, i.e. $\bar{\rho}\,|_{D_p}$ is diagonalizable.*

To complete our preliminaries and begin working on proving Theorem 2.3, we now prove that it implies Theorem 2.1. Actually, as Ribet notes, Theorem 2.3 implies Theorem 2.1 with $\mathbb{Q}(\mu_p)$ replaced by $\mathbb{Q}(\mu_p^{1-k})$, which has degree $(p-1)/(p-1,k-1)$ over $\mathbb{Q}$. Of course this version of Theorem 2.1 implies the desired one.

Let us record a few useful definitions.

Definition 2.4. A field $K \subset \bar{\mathbb{Q}}$ is *cut out* by a Galois represetation ρ of $G_{\mathbb{Q}}$ provided that $\ker\rho$ is the unique subgroup of $G_{\mathbb{Q}}$ fixing K. Note that $\mathrm{Gal}(K/\mathbb{Q}) \cong G_{\mathbb{Q}}/\ker\rho$.

Definition 2.5. Call a Galois represetation ρ of $G_{\mathbb{Q}}$ *unramified at* $\wp$, a prime in the number field K, provided that for any inertia group $I_{\mathfrak{p}}$ of a maximal ideal $\mathfrak{p} \subset \bar{\mathbb{Z}}$ over $\wp$, $I_{\mathfrak{p}} \subset \ker \rho$.

Observe that a Galois representation ρ is unramified at a rational prime p, then p does not ramify in the (Galois) extension of $\mathbb{Q}$ cut out by ρ. This is the simple yet important fact that makes it important to construct a representation with highly controlled ramification.

Proposition 2.6. *Theorem 2.3 implies Theorem 2.1*

Proof. The image of $\bar{\rho}$ is finite, so it is isomorphic to the Galois group of a finite extension $E/\mathbb{Q}$. Therefore, write $\bar{\rho}$ for the injection $\bar{\rho} : \mathrm{Gal}(E/\mathbb{Q}) \hookrightarrow GL_2(\mathbb{F})$. Recalling the definition of χ in Equation (1.5) and noting especially that it factors through Δ, we note that part (B) implies that $\mathbb{Q}(\mu_p^{1-k}) \subset E$.

Now we claim that $E/\mathbb{Q}(\mu_p^{1-k})$ is Galois and $\mathrm{Gal}(E/\mathbb{Q}(\mu_p^{1-k}))$ is of type $(p, p, \ldots, p)$, i.e. it is elementary abelian. The extension $\mathrm{Gal}(E/\mathbb{Q}(\mu_p^{1-k}))$ is Galois because if $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$ fixes $\mathbb{Q}(\mu_p^{1-k})$, then

$$\bar{\rho}(\sigma) = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}, \tag{2.4}$$

and matrices of this form are clearly a normal subgroup of $\bar{\rho}(\mathrm{Gal}(E/\mathbb{Q}))$. The extension has type $(p, \ldots, p)$ because matrices of the form in Equation (2.4) must have order dividing p. Part (C) says that there exist matrices of order p, therefore as the quotient $\mathrm{Gal}(\mathbb{Q}(\mu_p^{1-k}/\mathbb{Q})$ of $\mathrm{Gal}(E/\mathbb{Q})$ has order prime to p, the extension $E/\mathbb{Q}(\mu_p^{1-k})$ is nontrivial. This establishes part (2) of Theorem 2.1.

Ramification properties have yet to be addressed. Because by part (A) $\bar{\rho}$ is unramified at all primes $\ell \neq p$, we need only address the prime p. Of course, the extension $\mathbb{Q}(\mu_p^{1-k})$ is totally ramified at p. It remains to show that $E/\mathbb{Q}(\mu_p^{1-k})$ is unramified at (the unique prime over) p. This is the case because of part (D): the decomposition group D_p has order prime to p, but the ramification index of p in E divides the order of D_p. Therefore the prime over p in $\mathbb{Q}(\mu_p^{1-k})$ does not ramify in E, completing our proof of part (1).

Finally we prove part (3), that $\sigma\tau\sigma^{-1} = \chi(\sigma)^{1-k} \cdot \tau$ when $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$ and $\tau \in \mathrm{Gal}(E/\mathbb{Q}(\mu_p)$. This follows from representing σ and τ in matrix form via $\bar{\rho}$, i.e.

$$\bar{\rho}(\sigma) = \begin{pmatrix} 1 & a_\sigma \\ 0 & \chi(\sigma)^{k-1} \end{pmatrix} \quad \text{and} \quad \bar{\rho}(\tau) = \begin{pmatrix} 1 & a_\tau \\ 0 & 1 \end{pmatrix}. \tag{2.5}$$

The representatives for σ and τ are as Equation (2.5) prescribes because χ factors through Δ and τ fixes μ_p, ergo χ kills τ. Then we simply conjugate as in the statement of part (3) above to find

$$\begin{aligned} \bar{\rho}(\sigma)\bar{\rho}(\tau)\bar{\rho}(\sigma)^{-1} &= \begin{pmatrix} 1 & a_\sigma \\ 0 & \chi(\sigma)^{k-1} \end{pmatrix} \begin{pmatrix} 1 & a_\tau \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -a_\sigma \cdot \chi(\sigma)^{1-k} \\ 0 & \chi(\sigma)^{1-k} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \chi(\sigma)^{1-k} a_\tau \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & a_\tau \\ 0 & 1 \end{pmatrix}^{\chi(\sigma)^{1-k}} = \chi(\sigma)^{1-k} \cdot \bar{\rho}(\tau) \end{aligned}$$

where the final "$\cdot$" means that $\chi(\sigma)^{1-k}$, an element of $\mathbb{F}_p^\times$, acts naturally on the $\mathbb{F}_p$-module $\mathrm{Gal}(E/\mathbb{Q}(\mu_p^{1-k}))$.

Thus we have verified part (1), and Proposition 2.6 is proved. □

3. Congruences between Modular Forms

In this section we will construct on the assumption that $p \mid B_k$ a certain weight 2 cusp eigenform f that is congruent to the Eisenstein series (3.1) modulo p. This will allow us in §§4 and 5 to use the Eichler-Shimura relation to produce a representation as in Theorem 2.3. Keep in mind the comments in §1.3, that what we will create is a cusp eigenform that looks a lot like an Eisenstein series modulo p and therefore will have a similar representation modulo p.

To fix notation, a whirlwind tour of modular forms is in order. Definitions for the reader not familiar with modular forms may be found in the Appendix §A.3.

3.1. **Modular Forms and Eisenstein Series.** *Modular forms* are holomorphic functions on the upper half plane $\mathcal{H}$ that are

(1) invariant of some weight $k \in \mathbb{Z}^+$ under precomposition with the fractional linear actions of a congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$, namely, with respect to the *weight-k operator* (also known as the slash operator) $[\gamma]_k$ for all $\gamma \in \Gamma$; and

(2) can be extended continuously to $\mathcal{H}^* = \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q})$.

The modular forms of weight k on Γ constitute a complex vector space denoted $M_k(\Gamma)$. If the holomorphy restrictions on f are loosened to meromorphy, then f is called an *automorphic form*. We will exclusively work with modular forms on the usual congruence subgroups $\Gamma_0(N)$ and $\Gamma_1(N)$ (Definition A.3.1). Such modular forms have a unique Fourier expansion

$$f : \mathcal{H} \to \mathbb{C}, \quad z \to \sum_{n \geq 0} a_n(f) q^n$$

where $q = e^{2\pi i z}$ and $a_n(f)$ represents the nth Fourier coefficient of f.

The naturality of modular forms and the important subspace of cusp forms is best explained through their role as differentials of hyperbolic Riemann sufaces. The orbits of Γ in $\mathcal{H}$, denoted $Y(\Gamma)$, and its compactification, $X(\Gamma) = \Gamma \backslash \mathcal{H}^*$, are called *modular curves*. They are Riemann surfaces, and the compactification is accomplished by adjoining the *cusps* of Γ (Definition A.3.4).

The space $M_k(\Gamma)$ of modular forms is naturally isomorphic to the vector space of holomorphic differentials on $Y(\Gamma)$, loosely via the map $f \mapsto f(dz)^{k/2}$. However, because dz has simple poles at the cusps, not all of $M_k(\Gamma)$ maps to holomorphic differentials. This motivates the naturality and importance of the subspace of *cusp forms* $S_k(\Gamma)$, defined by requiring that a modular form vanish at the cusps of Γ. Note that because there is always a cusp "at infinity" (i.e. at $z = i\infty$, or equivalently $q = 0$), a cusp form f must have no constant coefficient, i.e. $a_0(f) = 0$. In this paper, we will focus primarily on the weight $k = 2$ and $\Gamma = \Gamma_1(p)$ for an odd prime p. It is important to know as in Example A.3.5 that $\Gamma_0(p)$ has two cusps. See Appendix A.3 or Ch. 2 of [23] for further details.

Eisenstein series are the archetypal examples of modular forms. Let k be an even integer, $k \geq 4$. The Bernoulli number B_k is (up to p-unit) the constant term of the Eisenstein series $G_k \in M_k(\mathrm{SL}_2(\mathbb{Z}))$,

$$G_k(z) = -\frac{B_k}{2k} + \sum_{n\geq 1}\sum_{d|n} d^{k-1}q^n, \tag{3.1}$$

which is constructed in Appendix A.2 to show its naturality. Note that by Equation (1.4), we can replace $-B_k/2k$ in formula (3.1) with $\zeta(1-k)/2$. That is, the constant coefficient of this Eisenstein series is an L-value special. This phenomenon will generalize to other Eisenstein series below. The Eisenstein series in Equation (3.1) is the starting point for our congruence, for when $p \mid B_k$, this Eisenstein series "looks like" a cusp form, which must have no constant coefficient, modulo p.

The other Eisenstein series that we require are Eisenstein series of weight 1 and 2 on $\Gamma_1(p)$). This prompts us to take a brief interlude to introduce the concept of a "type" of a modular form on $M_k(\Gamma_1(N))$, which will be heavily utilized. The basic idea is that since $\Gamma_1(N) \subset \Gamma_0(N)$, a modular form on $\Gamma_1(N)$ is not necessarily modular on $\Gamma_0(N)$, but this is nearly the case.

Definition 3.1. Let f be a modular form of weight k on $\Gamma_1(N)$. Such f is said to have *level* N. Then (by [11], §5.2) there exists a unique Dirichlet character ε to the modulus N such that

$$f[\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)]_k = \varepsilon(d) \cdot f \quad \text{for all } \gamma \in \Gamma_0(N).$$

The character ε is called the *type* of f on $\Gamma_0(N)$. We write $f \in M_k(N, \varepsilon)$.

Remark 3.2. In fact, this definition gives one of the two types of Hecke operators on level N, the *diamond operator* $\langle d\rangle$ for $(d, N) = 1$. It is defined as $\langle d\rangle f = f[\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)]_k$ where $\gamma \in \Gamma_0(N)$ and $f \in M_k(\Gamma_1(N))$. Note that the action is not trivial since $\Gamma_0(N)$ properly contains $\Gamma_1(N)$ for $N > 2$. See the comments around formula (A.3.3) for further details.

There are two Eisenstein series (up to scalar multiple) on $\Gamma_1(p)$ for each non-trivial even type ε, and one Eisenstein series when ε is the trivial character (by dimension formulas, [11], Theorem 3.5.1; see also [23]). While sums such as that in (3.1) do not converge for $k = 1$ or $k = 2$, the weights that we require, there are similar constructions (see [11], §§4.6 and 4.8). Here we will simply write down the Eisenstein series for non-trivial types.

Definition 3.3. Let ε be a non-trivial even type as above. Then the two Eisenstein series in $M_2(p, \varepsilon)$ are

$$G_{2,\varepsilon} = L(-1, \varepsilon)/2 + \sum_{n\geq 1}\sum_{d|n} \varepsilon(d)dq^n, \tag{3.2}$$

and the semi-cusp form

$$s_{2,\varepsilon} = \sum_{n\geq 1}\sum_{d|n} \varepsilon(n/d)dq^n \tag{3.3}$$

The *semi-cusp form* $s_{2,\varepsilon}$ is so called because it vanishes at infinity (which is clear from its lack of a constant coefficient) but does not vanish at the other cusp of $\Gamma_0(p)$ (the cusps are recorded in Equation (A.3.1)).

Now, the weight 1 forms, which only exist for ε odd just as those with even weight exist only for ε even.

Definition 3.4. Let ε be an odd type on $(\mathbb{Z}/p\mathbb{Z})^\times$. Then the *Eisenstein series of weight 1 and type ε on* $\Gamma_0(p)$ is

$$G_{1,\varepsilon} = L(0,\varepsilon)/2 + \sum_{n\geq 1}\sum_{d|n}\varepsilon(d)q^n. \tag{3.4}$$

Note how in both weights 1 and 2 an L-function value (see L-function definition, Eq. A.1.1) appears as the constant term in the place of a Riemann zeta function value in (3.1). These values are in fact generalized Bernoulli numbers, which are dealt with in the Appendix and defined in Definition A.1.22.

3.2. **The Construction.** Now we may focus on the modular forms relevant to our construction. First, fix notation. Let $\wp$ be a prime in $\mathbb{Q}(\mu_{p-1})$ dividing p, noting that p splits completely in $\mathbb{Q}(\mu_{p-1})$. We will need to discuss congruences modulo p in terms of this prime because our modular forms are of non-trivial type on $\Gamma_0(p)$, and therefore even the Eisenstein series' coefficients generate $\mathbb{Q}(\mu_{p-1})$ over $\mathbb{Q}$ (see Definition 3.1) Also, in analogy to the distinguished Galois character χ, permanently fix ω as the unique type on $\Gamma_0(p)$ such that

$$\omega(d) \equiv d \pmod{\wp} \quad \text{for all } d \in \mathbb{Z}. \tag{3.5}$$

Remark 3.5 ([27], p. 57)**.** In fact, any such ω also satisfies $\omega(d) \equiv d \pmod{p}$ for all d in $\mathbb{Z}$, even though p is not a principal ideal. It is known as the Teichmüller character, and is discussed further in Appendix A.1.

The main remaining prerequisite for the proofs of this construction is the theory of Hecke operators. I have decided to expound on these operators in §3, since there has been a good deal of background in this section already and one needs to know relatively little, and so I have omitted as much detail as possible from the following list. However, all of these definitions will be fleshed out in §3 or Appendix A.3.

Fact 3.6. The following are facts about Hecke actions on modular forms.

(1) For any k, there exist for every $n \geq 1$ a Hecke operator T_n on $M_k(\Gamma_1(N))$ that restricts to $S_k(\Gamma_1(N))$ and preserve type spaces $M_k(N,\varepsilon)$.

(2) There is always a basis of simultaneous eigenvectors of for the T_n with $(n,N)=1$, since these operators commute. Such a simultaneous eigenvector is called an *eigenform.*

(3) If f is an eigenform in $S_2(\Gamma_1(p))$ for all $T_n, (n,N)=1$, then it is an eigenform for *all* T_n.

(4) If f is an eigenform with respect to T_n, then the eigenvalue is $a_n(f)/a_1(f)$. Since the T_ℓ for ℓ prime generate the Hecke operators, an eigenvector with respect to these Hecke operators has that the eigenvalue $\lambda(n)$ of T_n is equal to $a_n(f)/a_1(f)$ for all n.

(5) The Eisenstein series that we have written down are Hecke eigenforms.

Here is the construction we wish to prove.

Theorem 3.7 ([20], Theorem 3.7)**.** *Suppose that $p \mid B_k$. Then there exists a cusp form $f = \sum_{n\geq 1} a_n q^n$ of weight 2 and type ω^{k-2} which is a normalized ($a_1 = 1$) eigenform for all Hecke operators and which satisfies*

$$a_\ell \equiv 1 + \ell^{k-1} \equiv 1 + \omega^{k-2}(\ell)\ell \pmod{\mathfrak{p}} \tag{3.6}$$

for all primes $\ell \neq p$, where $\mathfrak{p}$ is a certain prime ideal over p in the field K generated by the coefficients of f, which does not depend on ℓ.

Remark 3.8. Ribet makes a few enlightening comments, which we now repeat here, as to why he chose to go the route of working with weight 2 forms. Deligne [6] associated to modular forms of arbitrary weight a representation via Galois cohomology, and Serre suggested to Ribet that a congruence such as the one he proves might exist for a cusp eigenform that, like G_k, has weight k. However, at the time of Ribet's work there was only enough known about these representations to prove parts (A), (B), and (C) of Theorem 2.3, and not part (D). Then Ribet uses another idea of Serre, that representations coming from such a cusp eigenform "ought to be visible" modulo p on the Jacobian variety $J_1(p)$ associated to $\Gamma_1(p)$. Since forms of weight 2 on $\Gamma_1(p)$ are differentials on this variety, it is natural to look at them. As Khare [15] notes, this is an example of a principle discovered after the time of Ribet's paper, that "modulo p everything is weight 2."

3.3. **Proof.** Now let us proceed to work toward Theorem 3.7. Among our major tools are the Eisenstein series, which are useful because they are eigenforms whose coefficients we know, and the Deligne-Serre lemma (Lemma 3.14), which will allow us to produce a true eigenform out of a formal q-expansion that we only know to be an eigenform modulo $\wp$. For these first few lemmas, drop the assumption that $p \mid B_k$.

The following proposition is critical to our progress in two ways. It provides the basic congruence between Eisenstein series between G_k and an cusp eigenform when $p \mid B_k$, and also allows us to construct a special series in Proposition 3.10 using already known facts (Lemma 3.12) about how many k satisfy $p \mid B_k$.

Proposition 3.9 ([20], Lemma 3.1)**.** *Let k be even, $2 \leq k \leq p-3$. Then the modular forms $G_{2,\omega^{k-2}}$ and $G_{1,\omega^{k-1}}$ have $\wp$-integral q-expansions in $\mathbb{Q}(\mu_{p-1})$ which are congruent modulo $\wp$ to the q-expansion*

$$G_k(z) = -B_k/2k + \sum_{n\geq 1} \sum_{d|n} d^{k-1} q^n. \tag{3.7}$$

Proof. The assertion is clear except for the constant coefficients, because the nth coefficient ($n \geq 1$) of $G_{2,\omega^{k-2}}$ (resp. $G_{1,\omega^{k-1}}$) is $\omega^{k-2}(d)d$ (resp. $\omega^{k-1}(d)$), which is plainly congruent modulo $\wp$ to that of Equation (3.7) by the special property of ω (Equation (3.5)).

Therefore only the constant coefficients are of concern. However, the congruence between $-B_k/2k$ and the constant coefficients for $G_{2,\omega^{k-2}}$ follows directly from Proposition A.1.26 in the Appendix, and the congruence for $G_{1,\omega^{k-1}}$ follows from Fact A.1.24 and just a bit of computation. □

With the basic congruence of Proposition 3.9 in place, we use it to produce a modular form of any non-trivial even type that does *not* look like a cusp form modulo $\wp$. This form will be used to produce such a cusp form when $p \mid B_k$.

Proposition 3.10. *Let k be as above. Then there exists a modular form g of weight 2 and type ω^{k-2} whose q-expansion coefficients are $\wp$-integers in $\mathbb{Q}(\mu_{p-1})$ and whose constant term is 1.*

Before proving Proposition 3.10, we need these lemmata.

Lemma 3.11. *Let t be the number of even integers n, $2 \leq n \leq p-3$, such that p divides B_n. Then $p^t \mid h_p^-$, the negative part (Defn. A.1.21) of the class number h_p of $\mathbb{Q}(\mu_p)$.*

Proof. This is Proposition A.1.27 in the Appendices. □

Lemma 3.12. *The negative part h_p^- of the class number of $\mathbb{Q}(\mu_p)$ is bounded by*

$$h_p^- < p^{(p+3)/4} 2^{-(p-1)/4}.$$

Proof. In [17], Thm. 7.1, and the discussion afterwards, we find that $\pm D_p = p^{(p-3)/2} h_p^-$ where D_p is the determinant of a dimension $(p-1)/2$ matrix with each entry an integer from 1 to $p-1$. The absolute value of the determinant is bounded by the product of the Euclidean lengths of the row vectors (Hadamard's inequality), from which we derive the desired inequality. □

Now we can produce a "unit series" of sorts, which will allow us to cancel constant terms to produce semi cusp forms later.

Proof. (Proposition 3.10) It suffices to construct a g whose constant term is a $\wp$-unit, since it may be multiplied by another unit to get the desired form. We know from Proposition 3.9 that the Eisenstein series $G_{2,\omega^{k-2}}$ will suffice unless $p \mid B_k$. In this case, consider the set of pairs of even integers

$$(n, m), 2 \leq n, m \leq p-3, \text{ such that } n + m \equiv k \pmod{p-1}.$$

Then the product $G_{1,\omega^{k-1}} G_{1,\omega^{k-1}}$ is a modular form of weight 2 and type ω^{k-2} whose q-expansion coefficients are $\wp$-integers. Furthermore, its constant term is a $\wp$-unit unless $p \mid B_n B_m$. Therefore, our proposition is true unless for every such pair (n, m), p divides one of the two Bernoulli numbers B_n, B_m. Since there are $(p-1)/2$ Bernoulli numbers in question, we need only show that p divides less than $(p-1)/4$ of them to complete the proof.

By Lemma 3.11, if t is the quantity of even integers n, $2 \leq n \leq p-3$, such that p divides B_n, then $p^t \mid h_p^-$. Yet we know from Lemma 3.12 that $h_p^- < p^{(p+3)/4} 2^{-(p-1)/4}$. We are therefore done because $h_p^- = 1$ for $p \leq 19$,[3] and $p \leq 2^{(p-1)/4}$ for $p > 19$, implying that $h_p^- < p^{(p-1)/4}$ as desired. □

Having assembled the necessary tools to make the congruence, return to the usual notation. Fix an integer k as above, i.e. even from 2 to $p-3$, and assume that $p \mid B_k$. Fix also $\varepsilon = \omega^{k-2}$.

[3] In fact these are the only cyclotomic fields $\mathbb{Q}(\mu_p)$ with p prime and unique factorization. See [27], Ch. 11.

Proposition 3.13 ([20], Proposition 3.4)**.** *There exists a semi cusp form $f = \sum_{n\geq 1} a_n q^n$ such that the a_n are $\wp$-integers in $\mathbb{Q}(\mu_{p-1})$ and such that*

$$f \equiv G_k \equiv G_{2,\varepsilon} \pmod{\wp}$$

as q-expansions.

Proof. Let c be the constant term of $G_{2,\varepsilon}$. Then $f = G_{2,\varepsilon} - c \cdot g$ has constant term 0 in its q-expansion since the constant coefficient of g is 1. Therefore f is a semi cusp form, since it vanishes at the cusp ∞. The congruence $G_{2,\varepsilon} \equiv G_k \pmod{\wp}$ proved in Proposition 3.9 implies that their respective constant coefficients $-B_k/2k$ and c are congruent as well. Since $p \mid B_k$, we then have that $\wp \mid c$. Thus $f \equiv G_{2,\varepsilon} \pmod{\wp}$, completing the proof. □

The Deligne-Serre lifting lemma, which we now quote, is a very useful tool that makes modulo p congruences on modular forms worthwhile. Note, however, that it is stated completely module-theoretically.

Lemma 3.14 (Deligne-Serre lifting lemma; [9], Lemme 6.11)**.** *Let M be a free module of finite type over a discrete valuation ring R; write $\mathfrak{m}$ for the maximal ideal of R, k the residue field, and K the field of fractions. Let $\mathcal{T}$ be a pairwise commutative set of endomorphisms of M. Let $f \in M/\mathfrak{m}M$ be a nonzero common eigenvector of all $T \in \mathcal{T}$, and let $a_T \in k$ be the corresponding eigenvalues. Then, there exists a discrete valuation ring R' containing R with maximal ideal $\mathfrak{m}'$ such that $\mathfrak{m}' \cap R = \mathfrak{m}$, fraction field K' such that $[K' : K] < \infty$, and a nonzero element f' of*

$$M' = R' \otimes_R M,$$

which is an eigenvector of all $T \in \mathcal{T}$ with corresponding eigenvalues a'_T such that $a'_T \equiv a_T \pmod{\mathfrak{m}'}$.

The Deligne-Serre lemma will lift our Eisenstein series to a cusp form, if we can verify that the Eisenstein series are eigenforms modulo p. In fact, even more is true: as stated above, they are eigenforms before reduction. The following lemma records this fact.

Lemma 3.15. *The Eisenstein series $G_{2,\varepsilon}$ and $s_{2,\varepsilon}$ are Hecke eigenforms for all Hecke operators on $M_2(p, \varepsilon)$.*

Proof. It is elementary to verify that these forms are eigenforms for Hecke operators T_n with $(n, p) = 1$, but for weight 2 forms the situation is not so hard for $(n, p) \neq 1$. As recorded above, Proposition 5.2.3 of [11] implies the lemma. □

The following proposition constructs f that will turn out to have the properties desired of the construction. However, it leaves the fact that f is an eigenform with respect to T_p, which will be proved afterwards to complete Theorem 3.7.

Proposition 3.16 ([20], Prop. 3.5)**.** *Assume that $p \mid B_k$. There exists a non-zero cusp form f' of type ε which is an eigenform for all Hecke operators T_n with $(n, p) = 1$ and which has the property that for each prime $\ell \neq p$ the eigenvalue $\lambda(\ell)$ of T_ℓ acting on f' satisfies*

$$\lambda(\ell) \equiv 1 + \ell^{k-1} \equiv 1 + \varepsilon(\ell)\ell \pmod{\mathcal{M}},$$

where $\mathcal{M}$ is a certain prime (independent of ℓ) lying over $\wp$ in the field $\mathbb{Q}(\mu_{p-1}, \lambda(n))$ generated by the eigenvalues over $\mathbb{Q}(\mu_{p-1})$.

Proof. On application of the Deligne-Serre Lifting Lemma, Lemma 3.14, every part of the proposition will be complete except the claim that f' is a cusp form and not merely a semi cusp form. Thus we begin by applying the Deligne-Serre lemma.

Let R be the localization at $\wp$ of the ring of integers $\mathcal{O}_{\mathbb{Q}(\mu_{p-1})}$ of $\mathbb{Q}(\mu_{p-1})$. Let $\mathcal{T}$ be the set of Hecke operators $\{T_n : (n,p) = 1\}$ on $M_k(p,\varepsilon)$, which, as Fact 3.6 notes, commute pairwise as required. Since these operators commute, Lemma 3.15 implies that the following decomposition respects the action of operators in $\mathcal{T}$ (cf. [11], §5.11).

$$M_2(p,\varepsilon) = S_2(p,\varepsilon) \oplus \langle G_{2,\varepsilon} \rangle \oplus \langle s_{2,\varepsilon} \rangle$$

Therefore set

$$M = \big(S_2(p,\varepsilon) \oplus \langle s_{2,\varepsilon} \rangle\big) \cap R[[q]],$$

i.e. M is the space of semi cusp forms of weight k and type ε on $\Gamma_0(p)$ with q-expansion coefficients in R. This is a free module of finite rank over R. Replace f from Proposition 3.13 with its reduction modulo $\wp$, and that proposition subsequently implies that f is equal to the reduction of $G_{2,\varepsilon}$ modulo $\wp$, hence by Lemma 3.15 is an eigenform with T_n-eigenvalues $\sum_{d|n} \varepsilon(d)d$ for $(d,n) = 1$. Using the terminology of the Deligne-Serre lemma, a_{T_n} is this same eigenvalue, and more precisely,

$$a_{T_\ell} \equiv 1 + \varepsilon(\ell)\ell \pmod{\wp} \text{ for } \ell \neq p \text{ prime.}$$

Applying the Deligne-Serre lemma to these M, R, $\mathcal{T}$, f, and a_{T_n} precisely as it is recorded in Lemma 3.14, we find that the resulting f' has exactly the properties required, except that it is a semi-cusp eigenform with respect to $\mathcal{T}$ and not necessarily a cusp eigenform. In particular, $\mathcal{M}$ is a prime over $\wp$ in a finite extension K' of $\mathbb{Q}(\mu_{p-1})$ and f' has coefficients with non-negative valuation with respect to every prime in K' over $\wp$. Its Hecke eigenvalues are $a'_T = \lambda(n)$ for $(n,p) = 1$.

It remains to verify that f' is a cusp form. The key is that the space of semi cusp forms in $M_k(p,\varepsilon)$ is the direct sum of the cusp forms and the semi cusp *eigenspace* $\langle s_{2,\varepsilon} \rangle$. Hence it suffices to show that f' cannot be $s_{2,\varepsilon}$. This is readily verified: the eigenvalue of T_ℓ acting on $s_{2,\varepsilon}$ is $\varepsilon(\ell)+\ell$, which cannot be congruent modulo $\wp$ to the corresponding f'-eigenvalue $1 + \varepsilon(\ell)$ unless ε is trivial, which we have ruled out. □

Though f' has only been proven to be an eigenform with respect to T_n with $(n,p) = 1$, facts about Hecke operators listed in Remark 3.6 imply that f' in an eigenform with respect to T_p as well. Hence we complete the proof of the construction, Theorem 3.7.

Proof. (Theorem 3.7) Let f' be the eigenform for Hecke operators $T_n, (n,p) = 1$ given by Proposition 3.16 above. Since all eigenforms in $S_2(\Gamma_1(N))$ are newforms, then the remarks above (and the fuller explanation found in Definition A.3.16 and Fact A.3.17 below) imply that f' is an eigenform with respect to all Hecke operators T_n. As remarked above, for every eigenform g of a single T_n, the eigenvalue is $a_n(g)/a_1(g)$. Therefore there is a scalar multiple of f', call it f, that has Fourier expansion

$$f = \sum_{n \geq 1} \lambda(n) q^n.$$

where $\lambda(n)$ is the eigenvalue of f with respect to T_n and $\lambda(\ell) \equiv 1 + \varepsilon(\ell)\ell$ for all primes $\ell \neq p$, which is exactly what Theorem 3.7 demands. □

4. The Eichler-Shimura Relation

The Eichler-Shimura relation appears as a citation in Ribet's work, but is a critical tool. In this section our goal is to understand the Eichler-Shimura relation, even though we will not be able to furnish all of the scheme-theoretic details.

The Eichler-Shimura relation states, loosely, that for most primes ℓ the Hecke action of T_ℓ on the modular curve of $\Gamma_1(N)$ is congruent modulo ℓ to a Frobenius action. The Eichler-Shimura relation will be useful in the following way: Factoring the Jacobian of the modular curve according to a basis of normalized cusp eigenforms on $\Gamma_1(p)$ (Theorem 4.12) so that the action of T_ℓ on that factor is as the ℓth coefficient of our special modular form f. Restricting the Eichler-Shimura relation to the factor associated to a certain eigenform f matches up the coefficients of f with a Galois representation that is an extension of the Frobenius action.

Almost everything in the previous two sentences needs to be developed, and their usefulness is mainly to state concisely what we will accomplish. In §4.1, the objects that Hecke operators act on - modular forms, modular curves, divisors, Jacobians - will be developed. The action of Hecke will be described on these objects according to need in §4.2. Then §4.3 will explain what we mean by an action on a curve being "congruent modulo ℓ" to another action. This requires defining reductions of curves and morphisms. We will then be prepared to prove the Eichler-Shimura relation in §4.4. Finally, in §4.5 we will construct the $\wp$-adic Tate module, which is the representation we have been looking for. Throughout, the presentation draws heavily on chapters 5 to 9 of [11].

4.1. **Hecke Objects.**

For lack of a better term, "Hecke objects" are the sets that Hecke operators act on. In this section we will build them up without reference to Hecke's action on them. The actions will be developed as needed in the next section. The main objects that we build up here are moduli spaces of elliptic curves and Jacobians of modular curves.

Modular forms are the best known Hecke object, but in fact no more details are needed in addition to those given in §3.

The moduli spaces of elliptic curves are actually modular curves. Recall the definitions of the modular curves $Y(\Gamma)$ and $X(\Gamma)$ from Definition A.3.8 if necessary. We will work entirely with the modular curves $Y_1(N) = Y(\Gamma_1(N))$ and its compactification $X_1(N) = X(\Gamma_1(N))$. This is natural after knowing of the most basic example, i.e. for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H} = Y_1(1) = Y(1)$.

Example 4.1. Each complex elliptic curve is uniquely holomorphically isomorphic to a complex torus $\mathbb{C}/\Lambda$ where Λ is a lattice in $\mathbb{C}$, and that for each Λ there is a unique $\mathrm{SL}_2(\mathbb{Z})z \in Y(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$, $z \in \mathbb{C}$, such that the lattice $\Lambda_z := [1, z]$ is homothetic to Λ. Therefore,

$$\Gamma(1)z \mapsto E_z = \mathbb{C}/[1, z]$$

is a bijection between the modular curve $Y(1)$ and the set of complex elliptic curves up to isomorphism. Thus $Y(1)$ is a moduli space.

It is reasonable to expect that if $1 < [\Gamma(1) : \Gamma] < \infty$, then $Y(\Gamma)$ is a moduli space of elliptic curves with some extra data. These will be called "enhanced elliptic curves (for Γ)." Since it is only the case $\Gamma = \Gamma_1(p)$ that concerns us, we will define the moduli space only for $\Gamma_1(N)$.

Definition 4.2. An *enhanced elliptic curve for* $\Gamma_1(N)$ is a pair (E, Q) where E is a complex elliptic curve and Q is a point of E of order precisely N. We say that two such pairs $(E, Q), (E', Q')$ are equivalent if there exists an isomorphism $E \xrightarrow{\sim} E'$ such that $Q \mapsto Q'$. We let $W_1(N)$ denote the set of enhanced elliptic curves for $\Gamma_1(N)$ modulo this equivalence relation.

This proposition shows that these enhanced elliptic curves modulo equivalence in $W_1(N)$ carry exactly the extra data that will naturally match up, similarly to Example 4.1, with $Y_1(N)$.

Proposition 4.3 ([11], Theorem 1.5.1(b))**.** *The moduli space for* $\Gamma_1(N)$ *is*

$$W_1(N) = \{[E_z, 1/N + \Lambda_z] : z \in \mathcal{H}\}.$$

Two points $[E_z, 1/N+\Lambda_z]$ *and* $[E_{z'}, 1/N+\Lambda_{z'}]$ *are equal if and only if* $\Gamma_1(N)z = \Gamma_1(N)z'$. *Thus there is a bijection*

$$\psi_1 : W_1(N) \xrightarrow{\sim} Y_1(N)$$

We shall write out half of the proof since it provides a good background example for the upcoming computation of Hecke actions on these curves.

Proof. Choose any point $[E, Q]$ of $W_1(N)$. Choose $z' \in \mathcal{H}$ such that $E \xrightarrow{\sim} E_{z'} = \mathbb{C}/\Lambda_{z'}$. Thus $Q = (cz' + d)/N + \Lambda_{z'}$ for some $c, d \in \mathbb{Z}$. Since the order of Q is precisely N, it is clear that $(N, (c, d)) = 1$. Therefore there exist $a, b, k \in \mathbb{Z}$ such that $ad - bc - kN = 1$ and the matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ reduced modulo N is in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Adding multiples of N to the entries of γ doesn't affect Q, so we may therefore assume that $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. Let $z = \gamma \cdot z'$. Then because the action of γ scales the lattice $\Lambda_{z''}$ by $(cz'' + d)^{-1}$, we have that $(cz' + d)\Lambda_z = \Lambda_{z'}$. Thus we may complete the first part of the proof by verifying that

$$(cz' + d)\left(\frac{1}{N} + \Lambda_z\right) = \frac{cz' + d}{N} + \Lambda_{z'} = Q.$$

This shows that $[E, Q] = [\mathbb{C}/\Lambda_z, 1/n + \Lambda_z]$ for some $z \in \mathcal{H}$ as desired.

We leave the second part of the proof as an exercise. □

The formulation above leaves us ready to compute the action of Hecke operators on $Y_1(N)$ as a moduli space.

The other main Hecke objects we discuss are the Jacobians and Picard groups of the modular curve $X_1(N)$. Recall that the Jacobian of a Riemann surface is essentially integration of differentials modulo homology, as follows.

Given a compact Riemann surface X of genus $g > 0$, recall that $\Omega^1(X)$ is the g-dimensional $\mathbb{C}$-vector space of degree 1 holomorphic differentials on X. We expect that the dual space $\Omega^1(X)^\wedge = \mathrm{Hom}_{\mathbb{C}}(\Omega^1(X), \mathbb{C})$ be given by path integration. However, integration is path dependent, with paths that are not homotopy equivalent generating different integrals. The homotopy group $\pi_1(X)$ gives paths up to homotopy equavalence, but since path *integration* does not depend upon the order of paths, the obstruction to removing path dependence from integration is the abelianization of $\pi_1(X)$, namely the homology group $\mathrm{H}_1(X, \mathbb{Z})$. If we think of A_i and B_i as the two inequivalent path integrations around each of the g handles, then they form a $\mathbb{Z}$-basis for $\mathrm{H}_1(X, \mathbb{Z})$. One may verify that $\Omega^1(X)^\wedge = \mathrm{H}_1(X, \mathbb{Z}) \otimes \mathbb{R}$. The Jacobian is the quotient of $\Omega^1(X)^\wedge$ by $\mathrm{H}_1(X, \mathbb{Z})$, that is "integration modulo homology."

Definition 4.4. Let X be a compact Riemann surface of genus g. Then the *Jacobian* of X is

$$\mathrm{Jac}(X) = \Omega^1(X)^\wedge / \mathrm{H}_1(X, \mathbb{Z}),$$

which is isomorphic to the g-dimensional complex torus $\mathbb{C}^g/\Lambda_g$.

We also record the Abel-Jacobi theorem, which will provide an important algebraic perspective on the Jacobian necessary to reduce the Jacobian, a complex object, to $\mathbb{Q}$ and subsequently to finite fields. Let the Abel-Jacobi map $\Psi : X \to \mathrm{Jac}(X)$ be

$$\Psi(x) = \int_{x_0}^{x} \omega$$

where $\omega \in \Omega^1(X)$ and $x_0 \in X$ is a fixed base point. Extend this linearly to $\mathrm{Div}(X)$ and observe that the dependence on x_0 vanishes on the degree-0 divisors $\mathrm{Div}^0(X)$. Thus we have a canonical map $\Psi : \mathrm{Div}^0(X) \to \mathrm{Jac}(X)$. The Abel-Jacobi theorem draws an isomorphism between the Picard group of X and $\mathrm{Jac}(X)$.

Theorem 4.5 ([5], Thm. 1.5)**.** *The map $\Psi : \mathrm{Div}^0(X) \to \mathrm{Jac}(X)$ defined above has a kernel consisting precisely of the group of principal divisors on X. Hence Ψ induces an isomorphism from $\mathrm{Pic}^0(X)$ to $\mathrm{Jac}(X)$.*

Let $\mathrm{J}_1(N)$ denote the the Jacobian of $X_1(N)$. Proposition A.3.9 stated that

$$\psi : S_2(\Gamma_1(N)) \xrightarrow{\sim} \Omega^1(X_1(N)),$$

so the dual spaces may also be identified via

$$S_2(\Gamma_1(N))^\wedge = \psi^\wedge(\Omega^1(X_1(N))^\wedge). \tag{4.1}$$

Sending the homology into $S_2(\Gamma_1(N))$ via the same map, the Jacobian of the modular curve $X_1(N)$ may be taken to be $\mathrm{J}_1(N) = S_2(\Gamma_1(N))^\wedge / \mathrm{H}_1(X_1(N), \mathbb{Z})$. With this identity in place, the action of Hecke on the Jacobian can by simply defined as precomposition by the action on modular forms, if the action preserves homology.

With the moduli space perspective on the modular curves and the concept of the Jacobian prepared, Hecke actions will be defined on them.

4.2. **Hecke Actions.** Here we make a brisk description of the Hecke actions that required for the purposes of this paper. A fuller explanation may be found in Appendix A.3.

Hecke actions in all of their guises come from a double coset of $\mathrm{GL}_2^+(\mathbb{Q})$. In general, double cosets $\Gamma_1\gamma\Gamma_2$ send modular forms, modular curves, etc. with respect to the congruence subgroup Γ_1 to corresponding objects defined with respect to another congruence subgroup Γ_2. The orbits of the action of Γ_1 on the double coset, i.e. $\Gamma_1\backslash\Gamma_1\gamma\Gamma_2$, are finite in number. Therefore, since the domain for the operators are invariant with respect to Γ_1, the operator is defined by a finite number of matrices in $\mathrm{GL}_2^+(\mathbb{Q})$. It is in terms of these representative matrices that we will discuss Hecke operators.

Our Hecke operators will be double coset operators with $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$, so they map the corresponding modular forms of level N, modular curves, etc., to themselves. An example of a Hecke operator already mentioned in this paper appeared in Remark 3.2. For $\gamma \in \Gamma_0(N)$ the double coset $\Gamma_1(N)\gamma\Gamma_1(N)$ consists of one orbit of the left action of $\Gamma_1(N)$ and is represented by γ. Since the action depends only on the bottom right entry d of γ, this Hecke operator is written as the diamond operators $\langle d \rangle$. The action on modular forms, for example, was given in Remark 3.2.

The other type of Hecke operator will play a much larger role. It is this type of operator that will be called a Hecke operator in contrast to the diamond operator. They are written T_n for all positive integers n; however, the double coset definition is complicated for general n, so we will only discuss its matrix decomposition for $n = p$ a prime. This will end up being no problem, since, as has been referred to in the previous sections, the Hecke operators of prime index suitably generate the others.

The Hecke operator T_p is the double coset of $\Gamma_1(N)$ defined by $\gamma = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. The $\Gamma_1(N)$-orbits in the double coset $\Gamma_1(N)\gamma\Gamma_1(N)$ are represented by matrices according to the following

Proposition 4.6. *Let p be a prime and N be a positive integer. If $p \nmid N$, then a system of representatives $B(p,N)$ of $\Gamma_1(N)\backslash\Gamma_1(N)\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)\Gamma_1(N)$ is given by*

$$B(p,N) = \left\{ \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} \right\}_{j=0}^{p-1} \cup \left\{ \begin{pmatrix} m & n \\ N & p \end{pmatrix} \cdot \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \right\} \text{ where } mp - nN = 1.$$

Let the matrices in the left set be denoted $\beta_j = \left(\begin{smallmatrix} 1 & j \\ 0 & p \end{smallmatrix}\right)$, and the right factor β_∞.

If $p \mid N$, then $B(p,N)$ is given by

$$B(p,N) = \{\beta_j\}_{j=0}^{p-1}.$$

Proof. See [11], Prop. 5.2.1. □

With the matrix representatives of the double coset for $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ defined, the Hecke operator T_p may be defined in various context in terms of the action of these matrices. For example, the action on modular forms $M_k(\Gamma_1(N))$ is given by the sum of the weight-k operators for the $B(p,N)$ (see Proposition A.3.12). Later we will see that this action on modular forms induces an action on the Jacobian $\mathrm{J}_1(N)$ via precomposition.

The other Hecke objects, modular curves and their moduli spaces, have a simpler action since there is no weight factor. A point $\Gamma_1(N)z \in X_1(N)$ where $z \in \mathcal{H}^*$ is sent to a divisor via

$$T_p : \Gamma_1(N)z \mapsto \sum_{\beta \in B(p,N)} \Gamma_1(N)(\beta \cdot z)$$

where the action of β is the usual fractional linear transformation. This map extends $\mathbb{Z}$-linearly to degree-0 divisors $\mathrm{Div}^0(X_1(N))$ and descends to the Picard group $\mathrm{Pic}^0(X_1(N))$ ([11], §§6.2-6.3). Of course, the action of Hecke operators on the moduli space will be equivalent. The moduli space formulation of the Hecke action will be the principal way that we calculate the reduction of the Hecke action to nonzero characteristic, so here we we calculate the action of T_p and $\langle d \rangle$ on $W_1(N)$ explcitly.

Proposition 4.7. *The Hecke action $T_p : \mathrm{Div}(W_1(N)) \to \mathrm{Div}(W_1(N))$ on the moduli space $W_1(N)$ the $\mathbb{Z}$-linear extension of*

$$W_1(N) \ni [E,Q] \mapsto \sum_C [E/C, Q + C]$$

where the sum is taken over all order p subgroups $C \subset E$ such that $C \cap \langle Q \rangle = \{0_E\}$. Likewise, the operator $\langle d \rangle$ behaves as $\langle d \rangle : [E,Q] \mapsto [E, dQ]$.

Proof. By Propostion 4.3, it suffices to prove this proposition for $[E, Q] = [\Lambda_z, 1/N + \Lambda_z]$ for arbitrary $z \in \mathcal{H}$. As $\beta_j z = (z + j)/p$ for $0 \leq j < p$ and $\beta_\infty z = 1/p$, the associated elliptic curve to each of these points is E/C_j where $C_j = \langle (z + j)/p \rangle + \Lambda_z$ and $C_\infty = \langle 1/p \rangle + \Lambda_z$. Each of these subgroups satisfy the condition laid out in the proposition above, that they are order p subgroups that only trivially intersect $\langle Q \rangle = \langle 1/N \rangle$ unless $p \mid N$, in which case C_∞ intersects $\langle Q \rangle$, However, it is precisely when $p \mid N$ that $\beta_\infty \notin B(p, N)$, so on verifying that the C_j are all possible subgroups of order p, the proof for T_p is complete.

The part on $\langle d \rangle$ is a quick exercise. □

The Hecke action on the moduli space $W_1(N)$ complete, we move on to the action on Jacobian.

Recall the formulation of the Jacobian in Equation (4.1), i.e.

$$\mathrm{J}_1(N) = S_2(\Gamma_1(N))^\wedge / \mathrm{H}_1(X_1(N), \mathbb{Z}).$$

A Hecke operator acts on the dual space $S_2(\Gamma_1(N))^\wedge$ by precomposition, that is, for $\varphi \in S_2(\Gamma_1(N))^\wedge$,

$$(T(\varphi))(f) = \varphi(T \cdot f) \quad \text{for all } f \in S_2(\Gamma_1(N)).$$

Thus the Hecke algebra would act on $\mathrm{J}_1(N)$ by composition on the right if it preserves homology. In fact this is the case.

Proposition 4.8 ([11], Prop. 6.3.2)**.** *The Hecke operators $T = T_p$ and $T = \langle d \rangle$ act by precomposition on the Jacobian associated to $\Gamma_1(N)$,*

$$T : \mathrm{J}_1(N) \longrightarrow \mathrm{J}_1(N), \quad [\varphi] \mapsto [\varphi \circ T] \text{ for } \varphi \in S_2(\Gamma_1(N))^\wedge,$$

where $[\varphi]$ represents the equivalence class of $\varphi \in S_2(\Gamma_1(N))^\wedge$ modulo $\mathrm{H}_1(X_1(N), \mathbb{Z})$.

Proof. Omitted. See Remark. □

Remark 4.9. Verifying Proposition 4.8 involves factoring the Hecke operators through two intermediate modular curves. Unfortunately this perspective on Hecke actions, while it is tractable at the level of this essay, becomes most important in geometric reasoning involved in proving the Eichler-Shimura relation that is beyond the scope of this essay. As it is lengthy as well, I have chosen to omit it. The proof may be found in [11].

From this point forward, we must quote liberally from the theory of newforms, discussed in the Appendix. The main facts needed may be found in Definition A.3.16 and Fact A.3.17. Also, we will require the Hecke algebra $\mathbb{T}$, defined in Definition A.3.15.

Proposition 4.8 is significant because it shows that the Hecke algebra consists of automorphisms of a free finitely generated $\mathbb{Z}$-module, namely the homology $\mathrm{H}_1(X_1(N), \mathbb{Z})$. This must be the case, otherwise the Hecke action would not descend from $S_2(\Gamma_1(N))^\wedge$ to $\mathrm{J}_1(N)$. The consequences are the following.

Proposition 4.10. *These facts follow from the fact that $\mathbb{T}$ consist of automorphisms of a free finitely generated $\mathbb{Z}$-module.*

(1) *The Hecke algebra $\mathbb{T}$ is itself a finitely generated $\mathbb{Z}$-algebra.*

(2) *Each T_n satsifies a monic polynomial equation with integer coefficients, so its eigenvalues, and in turn the coefficients of a normalized Hecke eigenform, are algebraic integers.*

(3) *Let* $f(z) = \sum_{n\geq 1} a_n(f)q^n$ *be a normalized eigenform. Then the image* $\mathbb{Z}[a_n(f)]$ *of the homomorphism*

$$\lambda_f : \mathbb{T} \to \mathbb{C}, \quad Tf = \lambda_f(T)f$$

is a finitely generated $\mathbb{Z}$*-module, and therefore lies in a number field, denoted* K_f*. If we set* $I_f = \ker(\lambda_f)$*, then* $\mathbb{T}/I_f \xrightarrow{\sim} \mathbb{Z}[a_n(f)]$*.*

.

Proof. Clear, though perhaps only on reading Appendix A.3. □

Now we are able to construct the main geometric object from which we will derive our Galois representation, the abelian variety A_f associated to a newform f. At the very beginning of this section, I commented that the Eichler-Shimura relation, while it applies to the entire modular curve $X_1(N)$, would be used to construct a Galois representation after restricting it to a certain factor of the Jacobian of the modular curve corresponding to our special cusp eigenform f from §2. The following definition of an abelian variety is this certain factor.

Definition 4.11. Let f be a newform of level N. The abelian variety associated to f is the quotient variety

$$A_f = \mathrm{J}_1(N)/I_f\mathrm{J}_1(N),$$

where I_f has been defined in item (3) above.

For $\Phi = [\varphi] + I_f\mathrm{J}_1(N) \in A_f$, it follows from strong multiplicity one ([11], p. 198) that for a eigenform $g \in S_k(\Gamma_1(N))$, we have $\Phi(g) = 0$ if and only if $I_f \cdot g \neq 0$. Recalling from Fact A.3.17 that a Galois conjugate of a newform is again a newform, it is then easy to check that f^σ is killed by I_f, and we may once more argue by multiplicity one that these are the only such eigenforms. Therefore if we set $V_f \subset S_2(\Gamma_1(N))$ to be the $\mathbb{C}$-span of the Galois orbit of f, we know that the action of the abelian variety on cusp forms factors through the $V_f^\wedge$ modulo homology's restriction to $V_f^\wedge$. This reasoning is made rigorous in [11], Proposition 6.6.4.

Thus we may conclude that following diagram,

$$\begin{array}{ccc} \mathrm{J}_1(N) & \xrightarrow{T_p} & \mathrm{J}_1(N) \\ \downarrow & & \downarrow \\ A_f & \xrightarrow{a_p(\cdot)^*} & A_f \end{array}, \tag{4.2}$$

commutes, in the sense that the restriction of T_p to $\varphi \in A_f$ is given by linearly extending the action on elements g of an eigenbasis given by

$$(a_p(g)^*\varphi)(g) = \begin{cases} a_p(g)\varphi(g) & \text{if } g = f^\sigma \\ 0 & \text{otherwise.} \end{cases}$$

to all of $S_2(\Gamma_1(N))$ by linearity.

Because of this restriction result, it is natural to expect (on the theory of newforms) that the Jacobian factors into abelian varieties. We quote the decomposition.

Theorem 4.12 ([11], Thm. 6.6.6)**.** *The Jacobian assocaited to $\Gamma_1(N)$ is isogenous to a direct sum of Abelian varieties associated to equivalence classes of newforms,*

$$\mathrm{J}_1(N) \to \bigoplus_f A_f^{m_f}.$$

Here the sum is taken over a set of representatives $f \in S_2(\Gamma_1(M_f))$ at levels M_f dividing N, and each m_f is the number of divisors of N/M_f.

This decomposition will be useful again when we take the Eichler-Shimura relation, which describes the Hecke action on the entire Jacobian of $\Gamma_1(p)$ reduced modulo ℓ in terms of a Frobenius action, and restrict this to A_f to get our Galois representation.

4.3. **Reduction of Algebraic Curves.** Here we will discuss reduction in two senses. First of all, we will discuss how the results from the previous section, which were phrased in terms of Riemann surfaces and $\mathbb{C}$-vector spaces, apply in a very similar form to the same algebraic curves defined over $\mathbb{Q}$. Then, we will discuss the geometry of the reduction of these varieties modulo a prime.

The Riemann existence theorem says that the analytic structure of functions and differentials on Riemann surfaces comes from a corresponding algebraic structure. That is, for example, our Riemann surface $X_1(N)$ is a complex algebraic curve, and therefore has a function field of transcendence degree 1 over $\mathbb{C}$ given by the meromorphic functions on $X_1(N)$. In the case $N = 1$, it is well known that the function field $\mathbb{C}(X(1))$ of $X(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}^*$ is the field of rational functions $\mathbb{C}(j)$, where $j(z)$ is the elliptic modular function. Viewing $X(1)$ as a moduli space of "non-enhanced" complex elliptic curves $W_1(1)$ as in Example 4.1, the point $E_z = \mathbb{C}/\Lambda_z$ in $W_1(1)$ is related to the function field $\mathbb{C}(X(1))$ in that E_z has j-invariant $j(z)$.

The case that concerns us, the function field $\mathbb{C}(X_1(N))$, is less pervasively known, but can still be easily complex analytically verified (see [11], §7.5) to be

$$\mathbb{C}(X_1(N)) = \mathbb{C}(j, f_1), \text{ where } f_1(z) = \frac{g_2(z)}{g_3(z)}\wp_z(1/N), \tag{4.3}$$

and where g_2, g_3 are certain multiples of our familiar Eisenstein series G_4, G_6 respectively. Also, $\wp_z$ is the Weierstrass $\wp$-function for the lattice $[1, z]$ ([11], Prop. 7.5.1).

Using the moduli space interpretation of E we can reprove formula (4.3) in a more geometric fashion. The additional effort will be worthwhile since this technique gives us a model for $X_1(N)$ over $\mathbb{Q}$, so we move forward thusly.

Consider the *universal elliptic curve*

$$E_j : y^2 + xy = x^3 - \frac{36}{j - 1728}x - \frac{1}{j - 1728},$$

an elliptic curve over $\mathbb{Q}(j)$, with *formal* j-invariant j. The notation E_j is not to be confused with E_z. In fact $E_z \xrightarrow{\sim} E_{j(z)}$ via the Weierstrass $\wp$-function. Then the element $1/N + \Lambda_z$ of E_z maps via $\wp$ to

$$Q_z(1/N) = \left(\frac{g_2(z)}{g_3(z)}\wp_z(1/N), \left(\frac{g_2(z)}{g_3(z)}\right)^{3/2}\wp_z'(1/N)\right),$$

which is a sensible place for the function field generators in Equation (4.3) come from given the description of the moduli space $W_1(N) \cong Y_1(N)$ in Proposition 4.3. This leads

us to think that a function field for $X_1(N)$ over $\mathbb{Q}$ could potentially be $K_1 = \mathbb{Q}(j, f_1)$ as well. To show that $\mathbb{Q}$ is algebraically closed in K_1, we will calculate the action of $G = \mathrm{Gal}(\overline{\mathbb{Q}(j)}/\mathbb{Q}(j))$ on the N-torsion $E_j[N]$, which we know from the theory of elliptic curves is congruent to $(\mathbb{Z}/N\mathbb{Z})^2$ ([24], Cor. 6.4). Take $Q = Q_z(1/N), P = Q_z(z/N)$ as a basis for $E_j[N]$. Then we have a G-representation $\rho : G \to \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ defined by the relation

$$\begin{bmatrix} P^\sigma \\ Q^\sigma \end{bmatrix} = \rho(\sigma) \begin{bmatrix} P \\ Q \end{bmatrix}, \quad \sigma \in G.$$

By the results in [11], §§7.5-7.6, this representation is surjective and cuts out

$$H = \mathrm{Gal}(\mathbb{Q}(j, E_j[N], \mu_N)/\mathbb{Q}(j)) \xrightarrow{\sim} \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}).$$

This is to be expected because the Weil pairing $e_N : E_j[N] \times E_j[N] \to \langle \mu_N \rangle$ ([24], §8.1) is Galois equivariant and $e_N(P^\sigma, Q^\sigma) = e_N(P, Q)^{\det \rho}$.

Despite the appearance of μ_N, we are hoping that the field $K_1 \subset \mathbb{Q}(k, E_j[N], \mu_N)$ intersects trivially with $\mathbb{Q}(\mu_N)$. This is the case if the image under ρ of the subgroup of H fixing K_1 surjects via "det" onto $(\mathbb{Z}/N\mathbb{Z})^\times$. This image is readily calculable because fixing f_1 means fixing Q, therefore these are the matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}.$$

These matrices' determinants take on all possible values, so we have verified that $\mathbb{Q}(j, f_1)$ is the function field of a nonsingular projective algebraic curve over $\mathbb{Q}$. It remains to show only that this model is the same one as over $\mathbb{C}$. This follows upon the fact that f_1 has the same degree minimal polynomial over both $\mathbb{Q}(j)$ and $\mathbb{C}(j)$ ([11], Exer. 7.7.2). Thus from now on, when we write $X_1(N)$ we mean a complex algebraic curve with a model over $\mathbb{Q}$.

At this point we have finished one part of the reduction: from analytic objects over $\mathbb{C}$ to algebraic objects over $\mathbb{Q}$.

Remark 4.13. While the Hecke actions on the moduli space have not been brought along down to $\mathbb{Q}$, it turns out that Hecke operators are defined over $\mathbb{Q}$. The Jacobian has not been brought along, but by Weil's theory [28] in its algebraic geometric Picard group form it remains valid as we reduce. Therefore we will use the Picard group from now on. See Chapter 7 of [11] for details.

The next step is to reduce our curves over $\mathbb{Q}$ to curves over the finite field $\mathbb{F}_p$. Requiring scheme theory, these topics are beyond the scope of this essay, though [11], §8.5 overviews the situation from a classical point of view. However, they are quite important and we will try to give indications here of what must be done. The first main notion is that of *good reduction*. Good reduction of some object X will be written $\widetilde{X}$. Work such as [22] extends to algebraic curves and Jacobians the most basic notion of good reduction, that of elliptic curves. In order to be able to discuss good reduction effectively in the elliptic curve case, a canonical model is required. This is the global minimal Weierstrass equation. The analogy in the case of algebraic curves (resp. Jacobians) are canonical proper models (resp. Neron or "minimal" models) over Spec $\mathbb{Z}$. Here is one example, drawn from [10].

Example 4.14 ([10], Ex. 8.0.1)**.** Consider the scheme $\mathcal{Y} = \mathrm{Spec}\ (\mathbb{Z}[j])$ over Spec $\mathbb{Z}$ and the isomorphism $\phi : Y_1(1) \to \mathcal{Y}(\mathbb{C})$ which sends $\mathrm{SL}_2(\mathbb{Z})z$ to the element of $\mathcal{Y}(\mathbb{C})$ defined by $j \to j(z)$. The pair $(\mathcal{Y}, \phi)$ is a model for $Y_1(1)$ over $\mathbb{Z}$.

For our purposes, however, we may draw on the intuition from basic algebraic geometry.

Definition 4.15. Let $\mathbb{Z}_{(p)}$ be the localization of $\mathbb{Z}$ at p. Following [11] §8.5, say that a nonsingular affine curve C defined by polynomials $\varphi_1, \ldots, \varphi_m \in R = \mathbb{Z}_{(p)}[x_1, \ldots, x_n]$ has *good reduction at* p if they generate a prime ideal in R and the reductions $\widetilde{\varphi}_1, \ldots, \widetilde{\varphi}_m \in \mathbb{F}_p[x_1, \ldots, x_n]$ defines a nonsingular affine algebraic curve $\widetilde{C}$ over $\mathbb{F}_p$. Call $\widetilde{C}$ the reduction of C at p. Similarly extend this definition to projective curves by homogenizing at looking at affine pieces.

Before moving on and focusing on good reduction, we should note that studying the geometry in the case of bad primes is relevant to Ribet's proof. For it is the Deligne-Rappoport results [7] on the fibres at bad primes are a critical tool used by Ribet to control the ramification of the representation at the prime p.

For the computations in the next section that derive the Eichler-Shimura relation, it is the moduli space perspective on the Hecke operators that is most intrumental. Thus information on the reduction of this moduli space from the original setting is needed. Let $W_1(N)$ now represent the space of elliptic curves over $\bar{\mathbb{Q}}$, which is parameterized by the $\bar{\mathbb{Q}}$-points of $X_1(N)_{/\mathbb{Q}}$. To reduce the moduli space modulo p where $p \nmid N$, chose a maximal ideal $\mathfrak{p}$ of $\bar{\mathbb{Z}}$ over p and restrict to those elliptic curves with good reduction modulo $\mathfrak{p}$. Write

$$W_1(N)' = \{[E, Q] \in W_1(N) : E \text{ has good reduction at } \mathfrak{p}\}$$

Likewise write $\widetilde{W_1}(N)$ for the moduli space of elliptic curves over $\bar{\mathbb{F}}_p$. The reduction map is therefore

$$W_1(N)' \to \widetilde{W_1}(N), \quad [E_j, Q] \mapsto [\widetilde{E}_j, \widetilde{Q}].$$

Remark 4.16. Note how the reduction of this moduli space depends on p not dividing N: for simplicity say $N = p$. Then there are $p^2 - 1$ points of order p in $E[p]$, but at most $p - 1$ such points in $\widetilde{E}$. The supersingular case makes it especially obvious that there are some elements of $[E, Q] \in W_1(p)$ that do not reduce well to $\widetilde{W_1}(p)$. This fact is made geometrically rigorous for the modular curve in Igusa's theorem [13] that $X_1(N)$ has good reduction at p for $p \nmid N$.

Finally, we record a few facts about the moduli space reduction that we have just written out apply more generally to nonsinglular projective algebraic curves over $\mathbb{Q}$ with good reduction at p. See [11], §8.5 for further details.

Fact 4.17. The natural reduction map $C \to \widetilde{C}$ is surjective, and induces a surjective map on degree-0 divisors. Principal divisors are sent to principal divisors, ergo this map descends to a surjective map $\mathrm{Pic}^0(C) \to \mathrm{Pic}^0(\widetilde{C})$.

Fact 4.18. Morphisms between curves $h : C \to \widetilde{C'}$ of positive genus reduce naturally to a morphism $\tilde{h} : \widetilde{C} \to \widetilde{C'}$ such that

$$\begin{array}{ccc} C & \xrightarrow{h} & C' \\ \downarrow & & \downarrow \\ \widetilde{C} & \xrightarrow{\tilde{h}} & \widetilde{C'} \end{array}$$

commutes.

4.4. **The Eichler-Shimura Relation.** There are three basic ingredients that go into the Eichler-Shimura relation. First, there is the Hecke action on the moduli space $Y_1(N)$ given in Proposition 4.7. Igusa's theorem [13] (see also [11], Thm. 8.6.1) states that $Y_1(N)$ has good reduction over p for all $p \nmid N$. It remains only to reduce the Hecke action modulo $p \nmid N$ and find that it is given by the action of Frobenius as follows.

Let σ_p be the Frobenius action $x \to x^p$ on the coordinates of an algebraic curve over $\mathbb{F}_p$.

Theorem 4.19 (Eichler-Shimura Relation, [11], Theorem 8.7.2)**.** *Let $p \nmid N$. The following diagram commutes:*

$$\begin{array}{ccc} \mathrm{Pic}^0(X_1(N)) & \xrightarrow{T_p} & \mathrm{Pic}^0(X_1(N)) \\ \downarrow & & \downarrow \\ \mathrm{Pic}^0(\widetilde{X}_1(N)) & \xrightarrow{\sigma_{p,*}+\widetilde{\langle p\rangle}_*\sigma_p^*} & \mathrm{Pic}^0(\widetilde{X}_1(N)) \end{array}$$

where the starred maps are the pushforwards and pullbacks.

Remark 4.20. Following the treatment in [11], there will be a major gap in rigor in our proof of the Eichler-Shimura relation. Recall from Remark 4.9 that understanding the Hecke action T_p on $X_1(N)$ involves factoring through two intermediate modular curves. These modular curves have bad reduction at p, putting their study beyond the scope of this essay. On the other hand, there is no such impediment for the operators $\langle d\rangle$ where $(d, N) = 1$, so their reduction, used in the statement of Theorem 4.19, follows directly from Fact 4.18. However, as Diamond and Shurman comment, it suffices to assume that a commutative diagram

$$\begin{array}{ccc} \mathrm{Pic}^0(X_1(N)) & \xrightarrow{T_p} & \mathrm{Pic}^0(X_1(N)) \\ \downarrow & & \downarrow \\ \mathrm{Pic}^0(\widetilde{X}_1(N)) & \xrightarrow{\widetilde{T}_p} & \mathrm{Pic}^0(\widetilde{X}_1(N)) \end{array}$$

exists, and then compute $\widetilde{T}_p$ ([11], p. 349). This is what our proof will consist of.

Most of the work (and the most interesting work, in my opinion) is in this lemma, which computes the reductions modulo p of the factors $[E/C, C + Q]$ for E with which the Hecke action on the moduli space was expressed in Proposition 4.7. Note that this lemma assumes that E has ordinary reduction at p.

Lemma 4.21 ([11], Lem. 8.7.1)**.** *Let E be an elliptic curve over $\bar{\mathbb{Q}}$ with good ordinary reduction at $\mathfrak{p}$ maximal in $\bar{\mathbb{Q}}$ over p and let $Q \in E$ be a point of order precisely N, $p \nmid N$. Let C_0 be the order p kernel of the reduction map $E[p] \to \widetilde{E}[p]$. For any order p subgroup C of E,*

$$\widetilde{[E/C, C+Q]} = \begin{cases} [\widetilde{E}^{\sigma_p}, \widetilde{Q}^{\sigma_p}] & \text{if } C = C_0 \\ [\widetilde{E}^{\sigma_p^{-1}}, [p]\widetilde{Q}^{\sigma_p^{-1}}] & C \neq C_0 \end{cases}$$

Proof. Suppose $C = C_0$. Let $E' = E/C$ and let $Q' = Q + C = \varphi(Q)$, where $\varphi : E \to E'$ is the quotient isogeny. Let ψ be the dual isogeny of φ. The proof may be reduced to showing that $\tilde{\psi}$ is seperable. It will suffice to prove this case because one of $\tilde{\psi}$ and $\tilde{\varphi}$ must be a Frobenius endomorphism as they have degree p, and in the ordinary reduction case,

this endomorphism is inseparable while its dual is seperable ([24], Thm. 3.1). We may then write $\tilde{\varphi} = i \circ \sigma_p$ where $i : \widetilde{E}^{\sigma_p} \to \widetilde{E'}$ is an isomorphism and $i(\widetilde{Q}^{\sigma_p}) = \widetilde{(Q')}$. Hence

$$[\widetilde{E'}, \widetilde{Q'}] = [\widetilde{E}^{\sigma_p}, \widetilde{Q}^{\sigma_p}]$$

as desired.

Now prove that $\tilde{\psi}$ is separable. Consider the commutative diagram

$$\begin{array}{ccc} E'[p] & \xrightarrow{\psi} & E[p] \\ \downarrow & & \downarrow \\ \widetilde{E'}[p] & \xrightarrow{\tilde{\psi}} & \widetilde{E}[p] \end{array}$$

Since E has ordinary reduction, so does its isogenous image E', so the bottom two groups have order p. Because $\varphi \circ \psi(E'[p]) = [p]E'[p] = 0$ and both φ and ψ have degree p, it follows that $\psi(E'[p]) = \ker \varphi = C$. By assumption, $\psi(E'[p]) = C_0$ which is the kernel of the reduction given by the right side downward arrow, thus the diagram sends $E'[p]$ to $\{0\} \subset \widetilde{E}[p]$. However, the left side downward arrow is surjective, so we conclude that $\ker \tilde{\psi} = \widetilde{E'}[p]$. Then since its degree equals the order of its kernel, $\tilde{\psi}$ is a separable isogeny.

On the other hand suppose that $C \neq C_0$. Use the same notation as above, so that φ is the quotient isogeny and ψ is its dual. In analogy to $C = \ker \varphi$ and C_0, let $C' = \ker \psi$ and let C'_0 be the kernel of the reduction of $E'[p]$, an order p subgroup of $E'[p]$. We claim that $C' = C'_0$. Since $C_0 \neq C = \ker \varphi$, the subgroup $\varphi(C_0)$ has order p. Similar to before, $\varphi(C_0)$ must be contained in $\ker \psi$ since $\psi \circ \varphi = [p] = 0 \mid_{E[p]}$. Thus $\varphi(C_0) = C'$. But since C_0 is the kernel of the left arrow and C'_0 is the kernel of the right, it must be the case that $\varphi(C_0) \leq C'_0$, hence $\varphi(C_0) = C'_0$ since both have order p. This means that $C' = C'_0$.

This puts us back in the first case ($C = C_0$) with φ replaced by ψ. Applying those arguments to ψ, E', and Q' (note that $\psi(Q) = [p]Q$) means that $\tilde{\psi} = i \circ \sigma_p$ where σ_p is the Frobenius endomorphism on $\widetilde{E'}$ and $i : \widetilde{E'}^{\sigma_p} \to \widetilde{E}$ is an isomorphism such that $\widetilde{Q'}^{\sigma_p} = [p]\widetilde{Q}$. Apply σ_p^{-1} to i (coefficientwise) so that $i^{\sigma_p^{-1}} : \widetilde{E'} \to \widetilde{E}^{\sigma_p^{-1}}$ sends $\widetilde{Q'} \mapsto [p]\widetilde{Q}^{\sigma_p^{-1}}$. Thus we have an equivalence of enhanced elliptic curves

$$[\widetilde{E'}, \widetilde{Q'}] = [\widetilde{E}^{\sigma_p^{-1}}, [p]\widetilde{Q}^{\sigma_p^{-1}}],$$

completing the lemma. □

As the above lemma applies only to E that are ordinary at p, we quote the results for when E is supersingular at p.

Lemma 4.22 ([11], Exer. 8.7.1)**.** *Let E be an elliptic curve over $\bar{\mathbb{Q}}$ with supersingular reduction at $\mathfrak{p}$ a maximal ideal in $\bar{\mathbb{Z}}$ over p and let $Q \in E$ be a point of order precisely N. Then for any order p subgroup C of E,*

$$[\widetilde{E/C}, \widetilde{Q + C}] = [\widetilde{E}^{\sigma_p}, \widetilde{Q}^{\sigma_p}] = [\widetilde{E}^{\sigma_p^{-1}}, [p]\widetilde{Q}^{\sigma_p^{-1}}].$$

Note that by this lemma, it does no harm to our considerations of the action of T_p on $\widetilde{W}_1(N)$ to assume that every elliptic curve in $\widetilde{W}_1(N)$ has ordinary reduction at p.

At this point the Eichler-Shimura relation can be readily imagined.

Proof. (Sketch) Since we are assuming that $p \nmid N$, by Proposition the action of T_p on $W_1(N)$ is

$$T_p : [E, Q] \mapsto \sum_{[E:C]=p} [E/C, Q + C],$$

which by Lemma 4.21 reduces modulo p on E with good reduction at p to

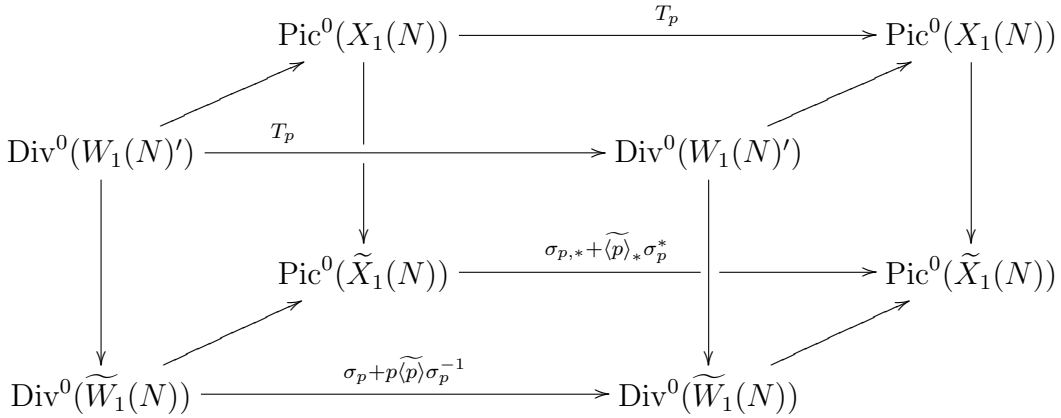
$$\begin{aligned} T_p([\widetilde{E}, \widetilde{Q}]) = \sum_C [\widetilde{E/C}, \widetilde{Q + C}] &= [\widetilde{E}^{\sigma_p}, \widetilde{Q}^{\sigma_p}] + p([\widetilde{E}^{\sigma_p^{-1}}, [p]\widetilde{Q}^{\sigma_p^{-1}}]) \\ &= (\sigma_p + p\widetilde{\langle p \rangle}\sigma_p^{-1})[\widetilde{E}, \widetilde{Q}], \end{aligned}$$

as there are $p + 1$ subgroups of E of order p, one being C_0. Note that the reduction of $\langle d \rangle$ is easy to compute (see Proposition 4.7).

The correlation proved between the T_p action on $W_1(N)'$ and $\widetilde{W_1}(N)$ extends $\mathbb{Z}$-linearly to divisors, that is,

$$\begin{array}{ccc} \mathrm{Div}^0(W_1(N)') & \xrightarrow{T_p} & \mathrm{Div}^0(W_1(N)') \\ \downarrow & & \downarrow \\ \mathrm{Div}^0(\widetilde{W_1}(N)) & \xrightarrow{\sigma_p + p\widetilde{\langle p \rangle}\sigma_p^{-1}} & \mathrm{Div}^0(\widetilde{W_1}(N)) \end{array}$$

commutes. This is the front square in the diagram,



where the back square is the one that we wish to show commutes. The side squares commute by Igusa's work [13], and the top part of the square commutes if one takes an algebraic perspective on the Hecke action Proposition 4.8 and reduces it to $\mathbb{Q}$ as per Remark 4.13 and then reduces modulo p. The bottom square is verified to be commutative in [11], Exer. 8.7.2. Therefore the Eichler-Shimura relation is complete. □

Remark 4.23. The action $\sigma_p + p\widetilde{\langle p \rangle}\sigma_p^{-1}$ on divisors becomes $\sigma_{p,*} + \widetilde{\langle p \rangle}\sigma_p^*$ is a natural extension of the fact that $[p] = \sigma_{p,*}\sigma_p^*$ on elliptic curves (which are their own Jacobian).

We conclude by noting that the rigorous proof of the Eichler-Shimura relation may be found in Shimura's book [23], §7.4. It draws heavily on his theory of canonical models, found in [23], §6.7.

4.5. **The Resulting Galois Representation.** Let us return to the notation of the previous sections: instead of working with general $\Gamma_1(N)$ and the assocaited modular curves, etc., consider an odd prime p such that $p \mid B_k$ for appropriate k, and let f be the cusp eigenform in $S_2(\Gamma_1(p))$ constructed in §3. We wish to apply the Eichler-Shimura

relation and the factorization of the Jacobian (Theorem 4.12) to construct an ℓ-adic representation of $G_{\mathbb{Q}}$ closely associated to f. Consequently, a change in notation is in order. The prime p replaces the integer N as the level, and q takes the place of p, i.e. $q \neq p$. We proceed following [11], §9.5.

Clearly the ℓ^n torsion of $\mathrm{J}_1(p)$ is of rank $2g$ where g is the genus of $X_1(p)$, since $\mathrm{J}_1(p) \cong \mathbb{C}^g/\Lambda_g$. These torsion points are algebraic over $\mathbb{Q}$ since we take Pic^0 instead of J and consider $\mathrm{Pic}^0(X_1(p))$, which has a good model over $\mathbb{Q}$, and moreover since we may take this model to be a Neron model with good reduction for $q \nmid p$, we have an isomorphism $\mathrm{Pic}^0(X_1(p))[\ell^n] \xrightarrow{\sim} \mathrm{Pic}^0(\widetilde{X}_1(p))[\ell^n]$ where the reduction is modulo q and $\ell \nmid qN$.[4] Galois will act on these torsion groups, but the comprehensive way to capture the action is to take its inverse limit.

Definition 4.24. The *ℓ-adic Tate module of $X_1(p)$* is

$$V_\ell(\mathrm{Pic}^0(X_1(p))) = \varprojlim_n \{\mathrm{Pic}^0(X_1(p))[\ell^n]\} \otimes \mathbb{Q}.$$

The Tate module $V_\ell(\mathrm{Pic}^0(X_1(p)))$ is a free $\mathbb{Q}_\ell$-vector space of dimension $2g$.

As $\mathrm{Pic}^0(X_1(p))$ is defined over $\mathbb{Q}$, $G_{\mathbb{Q}}$ acts on $\mathrm{Pic}^0(X_1(p))(\bar{\mathbb{Q}})$; since it has the structure of an abelian variety over $\mathbb{Q}$, the action induces an automorphism on $\mathrm{Pic}^0(X_1(p))[\ell^n]$, and is compatible with the inverse limit defining the Tate module, i.e.

$$\begin{array}{ccc} & G_{\mathbb{Q}} & \\ \swarrow & & \searrow \\ \mathrm{Aut}(\mathrm{Pic}^0(X_1(p))[\ell^n]) & \longleftarrow & \mathrm{Aut}(\mathrm{Pic}^0(X_1(p))[\ell^{n+1}]) \end{array}$$

This compatibility gives us a continuous representation

$$\rho_{X_1(p),\ell} : G_{\mathbb{Q}} \to \mathrm{GL}_{2g}(\mathbb{Z}_\ell) \subset \mathrm{GL}_{2g}(\mathbb{Q}_\ell).$$

As the action of Hecke operators on $\mathrm{Pic}^0(X_1(p))$ is defined over $\mathbb{Q}$ (see Remark 4.13), the Hecke action on the Jacobian commutes with the Galois action. The upcoming theorem shows that they are intertwined, extending, in a sense, the Eichler-Shimura relation from finite fields to the global situation. First, a few facts should be collected.

Definition 4.25. Let $\mathfrak{q}$ be a maximal ideal in $\bar{\mathbb{Z}}$ over a rational prime q. An *absolute Frobenius element of q* is any element in the decomposition group $D_{\mathfrak{q}}$ of $\mathfrak{q}$ that acts as the Frobenius automorphism on $\bar{\mathbb{Z}}/\mathfrak{q} = \bar{\mathbb{F}}_q$. Sometimes, by abuse of notation, we write Frob_q to denote an arbitrary absolute Frobenius element for some such $\mathfrak{q}$ over q.

Clearly these Frobenius elements will be a useful tool in "extending" the Eichler-Shimura relation to a global context.

Fact 4.26. It follows from the Cebotarev density theorem (see for example [14]) that if

$$F = \{\mathrm{Frob}_{\mathfrak{q}}\}_{\mathfrak{q}|q} \subset G_{\mathbb{Q}} \tag{4.4}$$

is a set of absolute Frobenius elements, one for each rational prime q, then this set is dense in $G_{\mathbb{Q}}$. Thus knowing the behavior of a continuous representation on such a set prescribes the representation.

[4] This situation exemplifies how abelian varieties extend the theory of elliptic curves.

Theorem 4.27 ([11], Thm. 9.5.1)**.** *The Galois representation $\rho_{X_1(p),\ell}$ is unramified at every prime $q \nmid \ell p$. For any such q let $\mathfrak{q} \subset \bar{\mathbb{Z}}$ be any maximal ideal over q. Then $\rho_{X_1(p),\ell}(\mathrm{Frob}_p)$ satisfies the polynomial equation*

$$x^2 - T_p x \langle p \rangle p = 0.$$

Proof. Choose $\mathfrak{q}$ in $\bar{\mathbb{Z}}$ over $q \nmid \ell p$. There is a commutative diagram

$$\begin{array}{ccc} D_{\mathfrak{q}} & \longrightarrow & \mathrm{Aut}(\mathrm{Pic}^0(X_1(p))[\ell^n]) \\ \downarrow & & \downarrow \\ G_{\mathbb{F}_q} & \longrightarrow & \mathrm{Aut}(\mathrm{Pic}^0(\widetilde{X}_1(p))[\ell^n]). \end{array}$$

Since the right side arrow is an isomorphism as was mentioned above, and the inertia group $I_{\mathfrak{q}}$ is the kernel of the left side map, the representation is unramified (see Definition 2.5).

To prove the second part of the theorem, the Eichler-Shimura relation restricts to ℓ^n-torsion so that

$$\begin{array}{ccc} \mathrm{Pic}^0(X_1(p))[\ell^n] & \xrightarrow{T_q} & \mathrm{Pic}^0(X_1(p))[\ell^n] \\ \downarrow & & \downarrow \\ \mathrm{Pic}^0(\widetilde{X}_1(p))[\ell^n] & \xrightarrow{\sigma_{q,*}+\widetilde{\langle q \rangle}_* \sigma_q^*} & \mathrm{Pic}^0(\widetilde{X}_1(p))[\ell^n]. \end{array}$$

Replacing the top arrow with $\mathrm{Frob}_{\mathfrak{q}} + \langle q \rangle q \mathrm{Frob}_{\mathfrak{q}}^{-1}$ also commutes (see Definition 4.25). Since the vertical arrows are isomorphisms,

$$T_q = \mathrm{Frob}_{\mathfrak{q}} + \langle q \rangle q \mathrm{Frob}_{\mathfrak{q}}^{-1} \text{ on } \mathrm{Pic}^0(X_1(p))[\ell^n].$$

Since this holds for all n, so the equality extends to the Tate module V_ℓ. The minimal polynomial for the action of $\mathrm{Frob}_{\mathfrak{q}}$ follows from the equality. □

All that is left to secure the relation that we desire is to apply (the algebraic reduction of) Theorem 4.12 to restrict this relation to the factor A_f of $\mathrm{J}(p)$, the abelian variety associated to f. This follows from the following

Lemma 4.28 ([11], Lem. 9.5.2)**.** *The restriction map $\mathrm{Pic}^0(X_1(p))[\ell^n] \to A_f[\ell^n]$ is surjective and its kernel is stable under $G_{\mathbb{Q}}$.*

Proof. Choose $y \in A_f[\ell^n]$. Then on writing $y = x + I_f \mathrm{Pic}^0(X_1(p))$ for some $x \in \mathrm{Pic}^0(X_1(p))$, it must be the case that $\ell^n x \in I_f J_1(p)$. It is easily verifed that ℓ^n-multiplcation is surjective on $I_f J_1(p)$, thus $\ell^n x = \ell^n x'$ for some $x' \in I_f \mathrm{Pic}^0(X_1(p))$. Therefore $x - x'$ is in $\mathrm{Pic}^0(X_1(p))[\ell^n]$ and maps to y, showing that the restriction map is surjective as desired.

The second part of the lemma follows from the fact that Hecke and Galois actions on $\mathrm{Pic}^0(X_1(p))$ commute (see [11]). □

Recall the notation on abelian varieties from §4.1, namely $K_f = \mathbb{Z}[\{a_n(f)] \otimes \mathbb{Q}$, $\deg K_f = d$. Then the action of $G_{\mathbb{Q}}$ on $\mathrm{Pic}^0(X_1(p))[\ell^n]$ induces an action on $A_f[\ell^n]$, subsequently on the Tate module $V_\ell(A_f)$. Thus $V_\ell(A_f)$ is a representation, written commonly as

$$\rho_{A_f,\ell} : G_{\mathbb{Q}} \to \mathrm{GL}(V_\ell(A_f)) = GL_{2d}(\mathbb{Q}_\ell), \tag{4.5}$$

and which inherits properties such as continuity and restricted ramification from $\rho_{X_1(p),\ell}$. Since by formula (4.2) T_q acts as $a_q(f)$ on A_f, and $\langle q \rangle$ acts as $\varepsilon(q)$ where ε is the type of f, the equation

$$x^2 - a_q(f)x + \varepsilon(q)q = 0$$

is satisfied by $\mathrm{Frob}_{\mathfrak{q}}$. After showing that $V_\ell(A_f)$ is in fact free of rank 2 over $K_f \otimes \mathbb{Q}$ ([11], Lemma 9.5.3), we may conclude with this final theorem, the construction of the representation associated to f, which we state in greater generality (replacing p with N).

Theorem 4.29. *Let $f \in S_2(N, \varepsilon)$ be a newform with number field K_f. Let ℓ be a prime. For each prime λ of K_f lying over ℓ there is a 2-dimensional Galois representation*

$$\rho_{f,\lambda} : G_{\mathbb{Q}} \to \mathrm{GL}(V_\ell(A_f)) = \mathrm{GL}_2(K_{f,\lambda}),$$

where $K_{f,\lambda}$ is the λ-adic completion of K_f. This representation is unramified at every prime $q \nmid \ell N$. For any such q let $\mathfrak{q} \subset \bar{\mathbb{Z}}$ be any maximal ideal lying over q. Then $\rho_{f,\lambda}(\mathrm{Frob}_{\mathfrak{q}})$ satisfies the polynomial equation

$$x^2 - a_q(f)x + \varepsilon(q)q = 0.$$

Finally we have produced a Galois representation associated to a cusp eigenform in $S_2(\Gamma_1(p))$. The theorem above shows that the representation may be studied via the coefficients of the modular form. The next section takes facts about this representation and the form of the coefficients of the special cusp eigenform produced in §3 to produce a reduction that cuts out exactly the Galois extension that we want, proving most of the main theorem.

5. Properties of the Representation

By the work in §4 culminating in Theorem 4.29, any appropriate cusp eigenform yields a Galois representation such that the coefficients of the eigenform correspond describe the action of absolute Frobenius elements in $G_{\mathbb{Q}}$. Note well, however, that this is a different kind of representation from the one that we are trying to produce to prove Theorem 2.3. Theorem 4.29 is a representation over a local field, whereas we are looking for a representation over a finite field. The representation sought will be a reduction of the p-adic representation. This section will take the representation from the cusp eigenform constructed in Theorem 3.7 and describe an appropriate choice of reduction which fulfills parts (A), (B), and (C) of Theorem 2.3. Our goal is the following theorem.

Theorem 5.1. *Assume that $p \mid B_k$. Let f be the newform of weight 2, level p, and type $\varepsilon = \omega^{k-2}$ constructed in Theorem 3.7. Let K be the completion of the coefficient field of f at $\mathfrak{p}$, the specific prime over p given in Theorem 3.7, and $\mathcal{O}, \mathbb{F}$ be its ring of integers and residue field. Let $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(K)$ be the representation associated to f by Theorem 4.29. Then there exists a reduction*

$$\bar{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$$

of ρ such that

(1) *$\bar{\rho}$ is unramified at all primes $\ell \neq p$;*

(2) *The representation $\bar{\rho}$ is reducible over $\mathbb{F}$ such that it is isomorphic to a representation of the form*

$$\begin{pmatrix} 1 & * \\ 0 & \chi^{k-1} \end{pmatrix};$$

(3) *$\bar{\rho}$ is not semisimple, or equivalently (Lemma 5.3, its image has order divisible by p.*

That is, there exists a representation $\bar{\rho}$ satsifying all of Theorem 2.3 except part (D).

Note that since ρ is unramified at all primes $\ell \neq p$ by its construction in Theorem 4.29, even the reader who does not know the definition of a "reduction" of ρ would expect that $\bar{\rho}$ satisfies property (1) automatically. Indeed this is the case. It remains only to verify parts (2) and (3).

5.1. **Reductions of p-adic Representations.** As first steps toward proving Theorem 5.1, we will define the reduction of a p-adic representation, and then discuss basic properties of reduced representations. For example, Lemma 5.3, the equivalence of semisimplicity and the absence of elements of order p in the image of a reduced representation modulo p, substantiates the equivalence claimed in part (3) of Theorem 5.1.

This this section (§5.1), we will use the same notation as in Theorem 5.1, but work with general objects of those type (i.e. K is any finite extension of $\mathbb{Q}_p$, etc.) Also fix $\mathcal{O}$ as the integer ring of K with uniformizer π.

Given a d-dimensional p-adic Galois representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}(V) = \mathrm{GL}_2(K)$,[5] it is reasonable to be curious why such a representation has a natural reduction modulo p since not all elements of $\mathrm{GL}_d(\mathbb{Q}_p)$ are p-integral. Yet the fact that the representations constructed above have integral trace and determinant suggest that such a reduction is possible at least in that case. But in fact, this is always the case. Some lattice T of V is always left stable because $G_{\mathbb{Q}}$ is compact and acts continuously on V (Proposition A.4.2).

WIth this lattice T in hand, $G_{\mathbb{Q}}$ acts on $T/\pi T$, which is a vector space of dimension two over $\mathbb{F}$. This action is the reduction of ρ.

Definition 5.2. Let ρ be a Galois representation on V that fixes lattice T as above. Then the induced map

$$\bar{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}(T/\pi T) = \mathrm{GL}_2(\mathbb{F})$$

is the *reduction* of ρ attached to T.

Recall that a semi-simplification of a representation is the direct sum of its Jordan-Hölder factors. By the Brauer-Nesbitt Theorem ([4], Thm. 30.16), the semi-simplfication of $\bar{\rho}$ does not depend on the choice of lattice T. Thus $\bar{\rho}$ is unique (up to equivalence of course) if any reduction of ρ is simple.

However, the opposite case, when some $\bar{\rho}$ is reducible, is the case that Ribet wants to deal with (cf. Theorem 5.1). In this case, the Brauer-Nesbitt theorem implies that there

[5]The discussion is restricted to dimension 2, but these beginning comments on the existence of reductions applies to arbitrary dimension.

are two characters $\varphi_1, \varphi_2 : G_{\mathbb{Q}} \to \mathbb{F}^\times$ such that the semisimplification of $\bar{\rho}$ is $\bar{\rho} = \varphi_1 \oplus \varphi_2$ for any reduction $\bar{\rho}$ of ρ. Hence, $\bar{\rho}$ may be written in one of the two forms

$$\begin{pmatrix} \varphi_1 & * \\ 0 & \varphi_2 \end{pmatrix}, \quad \begin{pmatrix} \varphi_1 & 0 \\ * & \varphi_2 \end{pmatrix},$$

depending on which character φ_i gives the action of $\bar{\rho}$ on its fixed subspace. We call these forms upper triangular and lower triangular respectively.

While the following lemma is not critical to producing an appropriate reduction, it is a basic fact about representations over finite fields that will be crucial to later results. Namely, by producing a reduced representation that is reducible but not semisimple, the elements of order p that then exist are those that correspond to the unramified p-extensions of $\mathbb{Q}(\mu_p)$.

Lemma 5.3. *Let $\bar{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be a representation on a finite field $\mathbb{F}$. Then $\bar{\rho}$ is semisimple if and only if its image has order prime to the characteristic p of $\mathbb{F}$.*

Proof. ([17], pp. 182–183) Choose some element $\alpha \in Im(\bar{\rho})$. Its Jordan normal form in $\bar{\mathbb{F}}$ is then one of

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}, \quad a, d \in \mathbb{F}^\times.$$

For $n \geq 1$ the nth power of these matrices are

$$\begin{pmatrix} a^n & na^{n-1} \\ 0 & a^n \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a^n & 0 \\ 0 & d^n \end{pmatrix}$$

respectively. Plainly, p divides the order of the left matrix and does not divide the order of the right side matrix. As the left matrix does not act semisimply whereas the right matrix does, the lemma has been verified. □

5.2. **Ribet's Lemma on Reducible Reductions.** With these preliminaries in place, we may move on to prove a critical lemma from Ribet's paper, that results in a reduced representation that is reducible but not semi-simple so that, for example, Lemma 5.3 applies. While Ribet's idea in [20] of producing a certain representation to deduce algebraic number theoretic properties were used heavily in further developments, T. Berger commented to me that of theorems in [20] it is this representation theoretic lemma that mathematicians have built upon most. Note that it implies that either character φ_i may act on the fixed subspace, depending on the choice of lattice.

Proposition 5.4 ([20], Prop. 2.1)**.** *Suppose that the K-representation ρ is simple but that its reductions are reducible. As above let φ_1, φ_2 be the characters associated to the reductions of ρ. Then $G_{\mathbb{Q}}$ leaves stable some lattice $T \subset V$ for which the associated reduction is of the form $\begin{pmatrix} \varphi_1 & * \\ 0 & \varphi_2 \end{pmatrix}$ but is not semi-simple.*

Proof. Following Ribet, to begin we set out two preliminary facts.

Choosing a $G_{\mathbb{Q}}$-stable lattice T of V and a $\mathcal{O}$-basis for this lattice allows ρ to be viewed as a map $G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O})$. A matrix $M \in \mathrm{GL}_2(K)$ such that $M\rho(G_{\mathbb{Q}})M^{-1} \subseteq \mathrm{GL}_2(\mathcal{O})$ defines another $G_{\mathbb{Q}}$-stable lattice MT with basis the image under M of the basis for T. From this lattice we get a new reduction

$$G_{\mathbb{Q}} \to M\rho(G)M^{-1} \hookrightarrow \mathrm{GL}_2(\mathcal{O}) \to \mathrm{GL}_2(\mathbb{F}). \tag{5.1}$$

Secondly, the proof will use heavily the identity

$$P\begin{pmatrix} a & \pi b \\ c & d \end{pmatrix} P^{-1} = \begin{pmatrix} a & b \\ \pi c & d \end{pmatrix}, \tag{5.2}$$

where $P = \begin{pmatrix} 1 & 0 \\ 0 & \pi \end{pmatrix}$.

Now we begin the proof proper. Choose a lattice T in V. If the reduction of ρ associated to T has lower triangular form $\begin{pmatrix} \varphi_1 & 0 \\ * & \varphi_2 \end{pmatrix}$, then the top right entry of every matrix in $\rho(G_{\mathbb{Q}})$ has positive valuation. Hence by Equation (5.2) we have $P\rho(G_{\mathbb{Q}})P^{-1} \subset \mathrm{GL}_2(\mathcal{O})$ and the new reduction as in Equation (5.1) is of the form $\begin{pmatrix} \varphi_1 & * \\ 0 & \varphi_2 \end{pmatrix}$. Therefore we may assume that the reduction is of the form desired.

To complete the proof, we assume that all reductions $\bar{\rho}$ of ρ of upper triangular form are semisimple, and prove that ρ is then reducible. This will prove the proposition by contradiction.

Set M_0 as the 2×2 identity matrix. Inductively, we will define a converging set of matrices

$$M_i = \begin{pmatrix} 1 & t_i \\ 0 & 1 \end{pmatrix}$$

such that $M_i\rho(G)M_i^{-1}$ consists of elements of $\mathrm{GL}_2(\mathcal{O})$ whose lower-left entries are divisible by π and whose upper right entries are divisible by π^i. This will imply that ρ is reducible because the matrix $M = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$ with $t = \lim t_i$ will then be such that $M\rho(G)M^{-1}$ consists of matrices whose upper right entries are 0.

Now, the induction step. Assume that $M_i\rho(G)M_i^{-1}$ consists of matrices of the form

$$\begin{pmatrix} a & \pi^i b \\ \pi c & d \end{pmatrix}$$

where $a, b, c, d \in \mathcal{O}$. By the conjugation formula (5.2), the matrices in $P^i M_i \rho(G) M_i^{-1} P^{-i}$ are of the form $\begin{pmatrix} a & b \\ \pi^{i+1}c & d \end{pmatrix}$, thereby describing a reduction modulo π which is in upper triangular form. By assumption, such a representation is semisimple; therefore there exists $u \in \mathcal{O}$ such that $U = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}$ diagonalizes the (mod π) representation. Therefore the matrices in

$$UP^i M_i \rho(G) M_i^{-1} P^{-i} U^{-1} \quad \text{have the form} \quad \begin{pmatrix} a & \pi b \\ \pi^{i+1}c & d \end{pmatrix}$$

as conjugation by U does not modify the bottom left entry. Conjugating by P^{-i} allows us to conclude that since

$$(P^{-i}UP^iM_i)\rho(G)(P^{-i}UP^iM_i)^{-1}$$

consists of integral matrices whose bottom left entries are divisible by π and whose upper right corner entries are divisible by π^{i+1}, setting

$$M_{i+1} = P^{-i}UP^iM_i = \begin{pmatrix} 1 & t_i + \pi^i u \\ 0 & 1 \end{pmatrix}$$

completes the induction. This form of M_{i+1} makes it plain to see that $\{t_i\}$ converges. $\square$

5.3. **Constructing the Desired Reduction.** All of the tools are in place to complete the proof of (all but one part of) Ribet's [20] main theorem (Theorem 2.3), constructing a reduced representation with special properties. We have already verified that the representation ρ associated to the special cusp eigenform f is unramified at all primes except p, so it remains to show that there exists a lattice in the representation such that the associated reduction has the form

$$\begin{pmatrix} 1 & * \\ 0 & \chi^{k-1} \end{pmatrix}$$

and is not semisimple. Ribet's lemma on reducible reductions of simple representations (Proposition 5.4) reduces the task to showing that ρ is irreducible and then finding a lattice such that the associated reduction is reducible. Then all that is left is to show that the reduction on such a lattice has semisimplification of the form $1 \oplus \chi^{k-1}$, where χ was defined in fomula (1.5).

We will prove that ρ is irreducible first, after recalling notation.

Working on the assumption such that $p \mid B_k$, k even, $2 \leq k \leq p-3$, recall the ensemble of notation from Theorem 5.1, namely, the cusp eigenform $f \in S_2(p, \varepsilon = \omega^{k-2}$, the continuous representation

$$\rho = \rho_{f,\mathfrak{p}} : G_{\mathbb{Q}} \to \mathrm{GL}(V_{\mathfrak{p}}(A_f)) = \mathrm{GL}_2(K) \tag{5.3}$$

from the Tate module $V = V_\ell(A_f)$, and so forth. Recall that $f \equiv G_k \pmod{\mathfrak{p}}$ by construction, so that

$$a_\ell(f) \equiv 1 + \ell^{k-1} \pmod{\mathfrak{p}}, \tag{5.4}$$

as recorded in Theorem 3.7. Thus by Theorem 4.29, the key consequence of the Eichler-Shimura relation, we know that an absolute Frobenius element Frob_ℓ over ℓ acts on V with trace and determinant

$$\mathrm{Tr}(\mathrm{Frob}_\ell) = a_\ell, \qquad \det(\mathrm{Frob}_\ell) = \ell \cdot \varepsilon(\ell).$$

This concludes the facts required.

Recall from Fact 4.26 that any system of absolute Frobenius elements F (defined in Equation (4.4)) is dense in $G_{\mathbb{Q}}$. Therefore since ρ and thus its determinant are continuous, the determinant may be uniquely extended from $F \subset G_{\mathbb{Q}}$ to a continuous homomorphism $G_{\mathbb{Q}} \to K_{\mathfrak{p}}$. The unique character extending $\mathrm{Frob}_\ell \mapsto \ell$ is the standard cyclotomic character

$$\chi_* : G_{\mathbb{Q}} \to \mathbb{Z}_p^\times \subset K^\times,$$

which is defined (naturally extending χ, see Equation (1.5)) by the relation

$$\sigma(\mu_{p^n}) = \mu_{p^n}^{\chi(\sigma)} \text{ for all } \sigma \in G_{\mathbb{Q}}, n \in \mathbb{Z}^+.$$

Note that χ_* cuts out the field $\mathbb{Q}(\mu_{p^\infty}) = \bigcup_n \mathbb{Q}(\mu_{p^n})$. Likewise, view ε as a character of $G_{\mathbb{Q}}$ via

$$\varepsilon : \sigma \mapsto \varepsilon(\chi_*(\sigma)). \tag{5.5}$$

Now we can prove

Proposition 5.5 ([20], Prop. 4.1)**.** *The $K_{\mathfrak{p}}$ representation ρ is irreducible.*

Proof. Suppose the proposition is false. Then the semisimplification of ρ (the unique semisimple representation with the same Jordan-Hölder factors) is abelian, hence the direct sum of two characters $\rho_1, \rho_2 : G_{\mathbb{Q}} \to K_{\mathfrak{p}}^{\times}$. Each ρ_i is "locally algebraic" in Serre's terminology [21] because it is an abelian representation from an abelian variety. Consequently, [21], Prop. III.1.2 implies that each ρ_i may be written as an integral power $\chi_*^{n_i}$ of χ_* on an open subgroup of an inertia group for p in $G_{\mathbb{Q}}$. This implies that $\rho_i = \chi^i \varepsilon_i$, where ε_i is a character of finite order ramified only at p. Regarding the Galois characters ε_i and χ as Dirichlet characters (taking both χ and ε_i through the "reverse" of formula (5.5)), we have for $\ell \neq p$ the relations

$$\ell^{n_1+n_2}\varepsilon_1(\ell)\varepsilon_2(\ell) = \ell \cdot \varepsilon(\ell)$$
$$a_\ell = \varepsilon_1(\ell)\ell^{n_1} + \varepsilon_2(\ell) \cdot \ell^{n_2}$$

because of formula (5.4). From the first relation we observe that $n_1 + n_2 = 1$, so that one of the n_i, say n_1, is at least 1, and $n_2 \leq 0$. Therefore, by the second relation, $|a_\ell| \geq \ell - 1$ for all $\ell \neq p$. Since by the list of Bernoulli numbers in Equation (1.3) we may take $\ell \geq 7$, this is a contradiction to the Riemann hypothesis of the Weil conjectures (theorems of Deligne [8]) that $|a_\ell| \leq 2\sqrt{\ell}$. □

Having proved that ρ is irreducible, it remains only to find a lattice such that the reduction is of the correct form (Equation (5.3)). In fact, any $G_{\mathbb{Q}}$-invariant lattice suffices!

Proposition 5.6 ([20], Prop. 4.2)**.** *There exists an $\mathcal{O}$-lattice $T \subset V$ invariant by $G_{\mathbb{Q}}$ for which the action of $G_{\mathbb{Q}}$ on $T/\pi T$ may be described matrically as*

$$\begin{pmatrix} 1 & * \\ 0 & \chi^{k-1} \end{pmatrix}$$

and is furthermore semisimple.

Proof. As a preliminary, note that χ_* reduces modulo π to χ, that is,

$$G_{\mathbb{Q}} \xrightarrow{\chi_*} \mathbb{Z}_p^{\times} \longrightarrow \mathbb{F}_p^{\times} \hookrightarrow \mathbb{F}^{\times}, \qquad G_{\mathbb{Q}} \xrightarrow{\chi} \mathbb{F}^{\times}$$

By Proposition 5.5, ρ is irreducible. Hence Ribet's lemma on representations, Proposition 5.4, implies that irreducible representations with reducible reductions are not semisimple. Thus if we can find a reducible reduction $\bar{\rho}$ of ρ, the last part of the proposition is complete. The Brauer-Nesbitt theorem states that the reduction of a representation has a well defined semisimplification, hence it suffices to find a lattice T such that the associated reduction has semisimplification $1 \oplus \chi^{k-1}$. As there exist reducible representations with this semisimplification, the reducibility of $\bar{\rho}$ follows from finding such a lattice T. Actually, any $G_{\mathbb{Q}}$ stable lattice will suffice, so chose a lattice and write it as T.

The Eichler-Shimura relation implies that an absolute Frobenius element for $\ell \neq p$ acts on $T/\pi T$ with trace $a_\ell \pmod{\pi}$ and discriminant $\ell \cdot \varepsilon(\ell) \pmod{\pi}$. Because of the congruence between f and G_k (Theorem 3.7) we know these numbers to be congruent to $\ell^{k-1} + 1$ and $\ell^{k-1} \pmod{\pi}$. By the Chebotarev density theorem (Fact 4.26) and the fact that $\ell^{k-1} \equiv \chi^{k-1} \pmod{\pi}$ where the trace and determinant of the action of $G_{\mathbb{Q}}$ on $T/\pi T$ are $1+\chi^{k-1}$ and χ^{k-1} respectively. Hence, every $\sigma \in G_{\mathbb{Q}}$, has the same characteristic roots as a representation of the form $1 \oplus \chi^{k-1}$, so it follows by the Brauer-Nesbitt theorem that these two representations have the same semisimplification, i.e. the reduced representation associated to T has semisimplification $1 \oplus \chi^{k-1}$ as desired. □

With the above proof complete, the representation that we must construct to prove Theorem 2.3 has been assembled up to property (D) of Theorem 2.3. Ribet's proof is beyond the scope of this essay.

6. Conclusion

Studying Ribet's paper [20] has been a very enlightening exercise in a rather literal sense. Becoming more familiar with how Galois representations, modular forms, and classical algebraic number theory come together has helped me understand the context for other mathematics that I hear about from day to day. Perhaps this is because the converse to Herbrand's theorem is in the center of a forceful historical stream of research. This seems to be the case to me, heuristically, because having looked at Wiles' papers through the 1980s, they each seem to be building on each other but starting in many ways with Ribet's paper. For example, Wiles in [31], the first paper to follow Ribet's, directly extends Ribet's results. Letting $C(\chi^i)$, i odd with $2 < i \leq p-3$, be the component of the entire p-Sylow subgroup of the class group A of $\mathbb{Q}(\mu_p)$ that is isotypical for χ^i, he proves that

Theorem 6.1 ([31], Thm. 1.1)**.** *If $C(\chi^i)$ is cyclic, then its order is precisely p^m where m is the p-adic valuation of $B_{1,\omega^{-1}}$.*

The assumption that $C(\chi^i)$ is cyclic was completed with the proof of the main conjecture of Iwasawa theory [19] a few years later. One question to ask is whether B_{1,ω^i} is a better quantity to look at than the usual Bernoulli numbers B_k. Does B_k have the same p-adic valuation as its paired B_{1,ω^i}? If not, then the B_{1,ω^i} seem to be the right quantities to look at.

The geometric reasoning that Ribet uses to prove property (D) appears to be a next step to take, while at the same Wiles comments in one of his papers in the 1980s that he will attempt to minimize the role of geometry, presumably in favor of number theoretic arguments. I'm interested to find out more about what results in the Ribet and Wiles vein since then have been discovered with geometry and without.

In preparing this essay I spent a good deal of time with Diamond and Shurman's book [11]. While it's very impressive what is covered in the book, I came to be even more impressed with Shimura's book [23] although I did not have the time to delve into it. I agree with F. Calegari's review [3] of [11] that "More recent works such as [11] contrast and complement [23] more than replace it." Shimura's book appears to be the source to go to for the constructions that we sketched in §4.

A. Appendices

A.1. **Proving One Direction of Kummer's Criterion.** Let p be an odd prime. Kummer's criterion states a remarkable connection between values of the Riemann zeta function, which are analytic quantities, and the p-divisibility of the ideal class number $h = h_p$ of the cyclotomic field $\mathbb{Q}(\mu_p)$.

Recall the following statements from the Introduction.

Theorem A.1.1 (Kummer; [27], Thm. 5.34)**.** *Let $\zeta(s)$ be the Riemann zeta function. Then p is irregular if and only if p divides the numerator of at least one of $\zeta(-1), \zeta(-3), \ldots, \zeta(4-p)$.*

If we define the Bernoulli numbers by

$$\frac{t}{e^t - 1} = \sum_{n=0}^{\infty} B_n \frac{t^n}{n!},$$

then in fact $\zeta(-n+1) = -B_n/n$ for $n = 1, 2, \ldots$ and it is not hard to see that B_n is zero for all odd n except $B_1 = -1/2$. Thus we can restate Kummer's criterion as

Corollary A.1.2. *A prime p is irregular if and only if it divides the numerator of at least one of the Bernoulli numbers $B_2, B_4, \ldots, B_{p-3}$.*

One more topic that should be mentioned before going on is p-adic zeta and L-functions. The beginning of this concept was Kummer's "Kummer congruences" (which will be useful later)

Theorem A.1.3 ([27], Cor. 5.14)**.** *For all positive even $n \equiv m \not\equiv 0 \pmod{p-1}$,*

$$\frac{B_n}{n} \equiv \frac{B_m}{m} \pmod{p}$$

is an equivalence of p-integral quantities.

In terms of zeta values, this implies that $\zeta(1-n) \equiv \zeta(1-m) \pmod{p}$ for such m, n. This is the first step toward showing that this ζ may be extended to a continuous function on $\mathbb{Z}_p$. In the modern perspective, Kummer congruences are viewed as a property of p-adic L-functions. As the conclusion (§6) discussed, these functions are strongly connected with extensions of Ribet's work [20] that this paper has described.

A similarly useful theorem, which also goes to explain why $m, n \equiv 0 \pmod{p-1}$ is excluded from the Kummer congruence, is the von Staudt-Clausen theorem

Theorem A.1.4 (von Staudt-Clausen; see [27], Thm. 5.10)**.** *Let n be an even positive integer. Then the fractional part of the Bernoulli number B_n is given by*

$$B_n \equiv -\sum_{(p-1)|n} \frac{1}{p} \pmod{\mathbb{Z}}.$$

The most basic implications are the following.

Corollary A.1.5. *The Bernoulli number B_n is p-integral unless $(p-1) \mid n$. If $(p-1) \mid n$ then pB_n is p-integral.*

Let us now set out to overview the proof of Kummer's criterion in the direction of $p \mid B_k \implies p \mid h$.

The key will be to relate arithmetic data of $K = \mathbb{Q}(\mu_p)$ to its maximal real subfield $K^+ = \mathbb{Q}(\mu_p)^+$, in the following progression:

(1) Study the relation between the integral units in K and in K^+, producing a relation between the regulators R_K and R_{K^+}. In fact, we will find that $R_K/R_{K^+} = 2^{(p-3)/2}$.

(2) Write the Dedekind zeta functions $\zeta_K(s), \zeta_{K^+}(s)$ for K and K^+ in terms of Dirichlet characters and their L-functions.

(3) Apply the analytic class number formula to the quotient $\zeta_K(s)/\zeta_{K^+}(s)$.

(4) Use the the conductor-discriminant formula and the functional equation for the L-functions to get cancellation in all factors of the equation except h/h^+ (where h^+ is the class number of K^+ and $\prod_\chi L(0,\bar{\chi})$ for odd χ.

(5) Show that the class group of K^+ injects into that of K via the natural inclusion of ideals, so that h/h^+ is an integer and has arithmetic meaning (it's called the negative part of the class number, h^-).

(6) Write these L-values as generalized Bernoulli numbers: in the same way that $\zeta(0) = -B_1$, get $L(0,\bar{\chi}) = -B_{1,\bar{\chi}}$. Namely,

$$h^- = 2p \prod_{\text{odd } \chi \in X_K} \left(-\frac{1}{2}B_{1,\chi}\right) = 2p \prod_{k=2 \text{ even}}^{p-1} \left(-\frac{1}{2}B_{1,\omega^{k-1}}\right)$$

where ω is a distinguished character called the Teichmuller character.

(7) Use this formula for generalized Bernoulli numbers,

$$B_{1,\chi} = \frac{1}{p}\sum_{a=1}^{p} \chi(a)a,$$

and the special property of the Teichmuller character show that

$$B_{1,\omega^{p-2}} \equiv \frac{-1}{p} \mod \mathbb{Z}_p,$$

so

$$h^- \equiv \prod_{k=2,\text{ even}}^{p-3} \left(-\frac{1}{2}B_{1,\omega^{k-1}}\right) \pmod{p}$$

(8) Apply the Kummer congruence to get $B_{1,\omega^{k-1}} \equiv B_k/k \pmod{p}$ (all quantities being p-integral).

The above sketches the proof that $p \mid B_k$ for $k = 2, 4, \ldots, p-3$ implies that $p \mid h$, which is one direction of Kummer's criterion. To show the other direction, one proves that $p \mid h^+ \implies p \mid h^-$ and then applies the same congruence. This is accomplished by dealing with the even characters and showing that p-divisibility of their Bernoulli numbers is related to those of odd character Bernoulli numbers. Unfortunately we will forgo this here.

In order to motivate our calculation of quantities related to the units of $K = \mathbb{Q}(\mu_p)$, we begin with the analytic class number formula, which connects the residue at $s = 1$ of the Dedekind zeta function of a number field with arithmetic invariants.

Definition A.1.6. Let F be a number field. The *Dedekind zeta function* of F is

$$\zeta_F(s) = \prod_{\wp} (1 - (N\wp)^{-s})^{-1},$$

where N is the absolute norm and the product is over the primes of K.

Theorem A.1.7 (Analytic Class Number Formula; see e.g. [14])**.** *Let F be a number field. The Dedekind zeta function $\zeta_F(s)$ has a simple pole at $s = 1$ with residue*

$$\frac{2^{r_1}(2\pi)^{r_2} h_F R_F}{w_F \sqrt{|d(F)|}},$$

where r_1 (resp. r_2) is the number of real (resp. conjugate paired complex) embeddings $F \hookrightarrow \mathbb{C}$, h_F is the ideal class number of F, R_F is the regulator of F, w_F is the number of roots of unity in F, and $d(F)$ is the field discriminant of F.

Since K is an abelian number field, its Dedekind zeta function may be factored into L-functions of the associated Dirichlet characters. We quote the following result the theory of Dirichlet characters, which in this setting are the most basic Galois representations. This theory has been implicit in dealing with the reduced representations needed to prove Ribet's theorem.

Proposition A.1.8 ([27], Thm. 4.3)**.** *Let $F/\mathbb{Q}$ be a Galois extension contained in $\mathbb{Q}(\mu_n)$ for some $n \in \mathbb{Z}^+$. Identify $G_n = \mathrm{Gal}(\mathbb{Q}(\mu_n)/\mathbb{Q})$ with $(\mathbb{Z}/n\mathbb{Z})^\times$ canonically (cf. χ in Eq. (1.5)) and let $X_F \leq G_n$ be the subgroup of Dirichlet characters whose kernel contains the subgroup of G_n fixing F, i.e. such that they cut out F. Then*

$$\zeta_F(s) = \prod_{\chi \in X_F} L(s, \chi)$$

where the L-function $L(s, \chi)$ is

$$L(s, \chi) = \sum_{n=0}^{\infty} \frac{\chi(n)}{n^s} = \prod_q (1 - \frac{\chi(q)}{q^s})^{-1}. \tag{A.1.1}$$

These Dirichlet characters are a more general version of the characters χ and χ_* encountered in this paper, though χ_* is the extension of the Dirichlet characters described here to $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$.

Applying the above facts to the number fields K and K^+ in particular, we calculate the L-values appearing in the ratio of their analytic class number formulas.

Corollary A.1.9. *Let X_K be the group of Dirichlet characters associated to K by Proposition A.1.8. The analytic class number formula for K and K^+ and the decomposition of their associated Dedekind zeta functions into L-functions of Dirichlet characters implies that*

$$\prod_{1 \neq \chi \in X_K} L(1, \chi) = \frac{2^{r_1}(2\pi)^{r_2} h_K R_K}{w_K \sqrt{|d(K)|}},$$

$$\prod_{1 \neq \chi \text{ even} \in X_K} L(1, \chi) = \frac{2^{r_1}(2\pi)^{r_2} h_{K^+} R_{K^+}}{w_{K^+} \sqrt{|d(K^+)|}}.$$

Proof. The first formula follows from the fact that if $\chi = \mathbf{1}_p$ is the trivial character, then $L(s, \chi)$ has a simple pole at $s = 1$ with residue 1. The remaining factors are real numbers which are not equal to zero by the proof of Dirichlet's theorem on primes in arithmetic progressions (see e.g. [27], Cor.4.4). The first formula follows.

For the second formula, note that because an odd character of $\mathrm{Gal}(K/\mathbb{Q})$ corresponds to complex conjugation and K^+ is the maximal totally real subfield of K (or alternatively because $[K : K^+] = 2$ uniquely), the subgroup of Dirichlet characters of $(\mathbb{Z}/n\mathbb{Z})^\times$ associated to K^+ is the group of even characters, i.e. those that send -1 to 1. □

Let $h^+ = h_{K^+}$ be the class number of the maximal real subfield $K^+ = \mathbb{Q}(\mu_p + \mu_p^{-1})$ of K. Recall that we want to prove one direction of Kummer's criterion by relating arithmetic data of K to that of K^+. In fact, what we will do is divide out their Dedekind zeta

functions and calculate all of of the ratios, leaving only that a ratio of class numbers h/h^+ is equal to L-values of odd characters. Subsequently, these L-values will be equated with Bernoulli numbers.

To begin progress toward proving Kummer's criterion, getting a handle on integral units in K relative to K^+ is critical, because this will allow us to compute several factors in their class number formulas.

Remark A.1.10. By abuse of terminology we will often call the integral units of a number field F simply "the units of F."

Proposition A.1.11. *These are facts about K.*

(1) *K is a totally complex field, that is there are $r_1 = 0$ real embeddings of K into $\mathbb{C}$, and $r_2 = (p-1)/2$ conjugate pairs of complex embeddings.*

(2) *The maximal (totally) real subfield of K is $K^+ = \mathbb{Q}(\mu_p + \mu_p^{-1})$, and its ring of integers is $\mathcal{O}_{K^+} = \mathbb{Z}[\mu_p + \mu_p^{-1}]$. We have $[K : K^+] = 2$.*

(3) *K and K^+ have the same unit rank, ergo $\mathcal{O}_{K^+}^\times \hookrightarrow \mathcal{O}_K^\times$ has finite index.*

Remark A.1.12. Facts like this hold for a more general class of fields called CM-fields.

Proof. To show (1): Note that every pth root of unity not equal to 1 is primitive, so the embeddings $K \hookrightarrow \mathbb{C}$ are given by $\mu_p \mapsto \mu_p^a$ for $a = 1, 2, \ldots, p-1$. Clearly each of these is not a real embedding. Thus they are complex embeddings, and as $\deg(K/\mathbb{Q}) = r_1 + 2r_2$, the result follows.

To show (2): Clearly $\mu_p + \mu_p^{-1}$ is real, so K^+ is a totally real field. It has index 2 in K because μ satisfies the irreducible polynomial

$$X^2 - (\mu + \mu^{-1})X + 1 = 0.$$

To show (3): By Dirichlet's unit theorem and (1), the unit rank of K is $r_1 + r_2 - 1 = (p-1)/2 - 1$. Since K^+ is totally real, its unit rank is its degree minus 1, which is also $(p-1)/2 - 1$. □

Thus we have calculated the terms r_1, r_2 for both K and K^+ in their analytic class number formulas. Part (3) Proposition A.1.11 is a first step toward making the calculations about regulators and number of roots of unity. As K^+ is totally real, its only roots of unity are ± 1. It is not hard to show that $\pm\mu_p^n$ are the roots of unity in K, so the ratio w_K/w_{K^+} in the analytic class number formula is p.

The regulator is generally a difficult term to calculate, and accordingly, we will not be able to do this. However, the following proposition will allow the ratio of regulators of K and K^+ to be written down.

Proposition A.1.13. *For any unit ε of $\mathbb{Z}[\mu_p]$, there exists a unit $\varepsilon_1 \in \mathcal{O}_{K^+}^\times$ and an integer r such that $\varepsilon = \mu_p^r \cdot \varepsilon_1$. Thus the index of the units of $\mathcal{O}_{K^+}$ in $\mathcal{O}_K$ is p.*

Proof. Choose ε as above and set $\alpha = \varepsilon/\bar{\varepsilon}$. Clearly α is an algebraic integer with absolute value 1; also, all of its conjugates have absolute value 1, since they commute with conjugation.

Claim. An algebraic integer α whose Galois conjugates all have absolute value 1 must be a root of unity.

Proof. Say that the degree of α is d. Then each of its powers have degree no more than d. Let $f(x)$ be the minimal polynomial for a power of α. Then the ith coefficient of f is bounded by the binomial coefficient $\binom{i}{d}$ since all conjugates of α are bounded by 1. Therefore there are only finitely many such polynomials, ergo finitely many powers of α. □

The only roots of unity in K are $\pm\mu_p^a$, so $\varepsilon/\bar{\varepsilon} = \pm\mu_p^a$ for some a. We will now show that $\pm = +$.

Assume that $\pm = -$. Since ε is an integer, recall that $(p) = (\mu - 1)^{p-1}$ and write

$$\begin{aligned}\varepsilon &= b_0 + b_1\mu_p + \cdots + b_{p-2}\mu_p^{p-2} \\ &\equiv b_0 + b_1 + \cdots + b_{p-2} \pmod{\mu_p - 1}.\end{aligned}$$

Since $\bar{\varepsilon} = b_0 + b_1\mu_p^i + \cdots$, the same congruence is true for $\bar{\varepsilon}$. Therefore,

$$\varepsilon = -\mu_p^a\bar{\varepsilon} \equiv -\varepsilon \pmod{\mu_p - 1},$$

and $2\varepsilon \equiv 0 \pmod{\mu - 1}$. But this is impossible because $(\mu_p - 1)$ is relatively prime to 2 and ε is a unit.

Thus we conclude that $\varepsilon/\bar{\varepsilon} = \mu_p^a$. Letting $2r \equiv a \pmod{p}$ and $\varepsilon_1 = \mu_p^{-r}\varepsilon$, we get $\varepsilon = \mu_p^r\varepsilon_1$ and $\bar{\varepsilon}_1 = \varepsilon_1$, completing the proof. □

Now we are already able to calculate the ratio of regulators of K and K^+. Recall this

Definition A.1.14. The *regulator* of a number field F is

$$R_F = \left|\det(\delta_i \log|\sigma_i(\varepsilon_j)|)_{1\le i,j\le r}\right|,$$

where $\{\varepsilon_j\}$ is a set of generators for the units of $\mathcal{O}_F$ modulo roots of unity and the σ_i, and $r = r_1 + r_2 - 1$ of the $r + 1$ embeddings $F \hookrightarrow \mathbb{C}$ (up to conjugate pair) are chosen to be σ_i. (The choice of which one is omitted does not matter) The δ_i factor is 1 for a real embedding σ_i and 2 for a representative σ_i of a pair of complex conjugate embeddings.

The ratio of regulators now follows immediately from the proposition.

Corollary A.1.15. *The ratio of the regulator R_K of K to the regulator R_{K^+} of K^+ is*

$$\frac{R_K}{R_{K^+}} = 2^{(p-3)/2}. \tag{A.1.2}$$

Proof. Since the units of K^+ have index p in those of K, and p is also the index of the roots of unity of K^+ in K, the inclusion $\mathcal{O}_{K^+}^\times \hookrightarrow \mathcal{O}_K^\times$ sends a set of generators of $\mathcal{O}_{K^+}^\times$ modulo roots of unity to an analogous set in $\mathcal{O}_K^\times$. Therefore the only difference in the calculation of their regulators is the δ_i factors. Since K is totally complex and K^+ is totally real, the calculations in Proposition A.1.11 completes the proof. □

At this point all of the ratios of data appearing in the analytic class number formulas for K and K^+ have been determined except h/h^+ and the ratio of the field discriminants. While these discriminants are easily calculable, they will fall out in the calculation because of the following two facts.

Fact A.1.16 (Conductor-Discriminant Formula; [27], Thm. 3.11)**.** Let F be a number field associated to the group X_F of Dirichlet characters. Then the discriminant of F is given by

$$d(F) = (-1)^{r_2} \prod_{\chi \in X_F} f_\chi,$$

where f_χ is the conductor of the character, i.e. the minimal modulus for which χ is a dirichlet character.

Corollary A.1.17. *Let K, K^+ be as usual. Then $|d(K)| = |d(K^+)|^2 = p^{p-1}$.*

Proof. Every non-trivial character in X_K has the same modulus ($= p$) and there are $p-1$ such characters, while there are $(p-1)/2$ non-trivial characters in the subgroup X_{K^+}. □

Fact A.1.18 ([27], Cor. 4.6)**.** Let $\tau(\chi)$ be the Gauss sum associated to the Dirichlet character χ, and let X_F be the Dirichlet characters associated to the Galois extension $F/\mathbb{Q}$. Then

$$\prod_{\chi \in X_F} \tau(\chi) = \begin{cases} \sqrt{|d(K)|} & \text{if } K \text{ is totally real} \\ i^{\deg(K/\mathbb{Q})/2}\sqrt{|d(K)|} & \text{if } K \text{ is complex.} \end{cases}$$

Corollary A.1.19. *Let K be as usual and let X' be the subset of odd characters of X_K. Then*

$$\prod_{\chi \in X'} \tau(\chi) = i^{(p-1)/2} p^{(p-1)/4}$$

Proof. Immediate. □

Finally, we may calculate the ratio of the class number formulas for K and K^+ written in terms of L-functions in Corollary A.1.9. The quotient relates the product of L-functions of odd Dirichlet characters to the modulus p to the ratios of arithmetic data calculated above.

$$\begin{aligned} \prod_{\chi \text{ odd}} L(1,\chi) &= \left(\frac{2^{r_1(K)}(2\pi)^{r_2(K)}}{2^{r_1(K^+)}(2\pi)^{r_2(K^+)}}\right)\left(\frac{R_K}{R_{K^+}}\right)\left(\frac{w_{K^+}}{w_K}\right)\left(\frac{\sqrt{|d(K^+)|}}{\sqrt{|d(K)|}}\right)\left(\frac{h}{h^+}\right) \\ &= \left(\frac{(2\pi)^{(p-1)/2}}{2^{(p-1)/2}}\right) \cdot 2^{(p-3)/2} \cdot p^{-1} \cdot p^{-(p-1)/4} \cdot \left(\frac{h}{h^+}\right) \end{aligned}$$

Now we apply the functional equation of L-functions for odd Dirichlet characters and observe that

$$L(1,\chi) = \frac{\tau(\chi)\pi}{i f_\chi} L(0,\bar{\chi}) \tag{A.1.3}$$

where $\tau(\chi)$ and f_χ were defined in the Facts above and $\bar{\chi}$ denotes the complex conjugate of χ. We want to know what Equation (A.1.3) looks like as a product over odd characters of X_K. By the Corollaries above,

$$\prod_{\chi \text{ odd}} \frac{\tau(\chi)\pi}{i f_\chi} L(0,\bar{\chi}) = \left(\frac{\pi}{ip}\right)^{(p-1)/2} \prod_{\chi \text{ odd}} \tau(\chi) \prod_{\chi \text{ odd}} L(0,\bar{\chi}) = \pi^{(p-1)/2} p^{-(p-1)/4} \prod_{\chi \text{ odd}} L(0,\chi).$$

Substituting this expression in for $\prod_{\chi \text{ odd}} L(1,\chi)$, we see that the factors $\pi^{(p-1)/2}$ and $p^{-(p-1)/4}$ cancel[6] to yield the following equality.

[6] And though we have glossed over it, the choices of square root of p needed to define $p^{(p-1)/4}$ for $p \equiv 3 \pmod 4$ are the same.

Proposition A.1.20.

$$\frac{h}{h^+} = \left(\frac{p}{2^{(p-3)/2}}\right) \cdot \prod_{odd\ \chi \in X_K} L(0,\chi).$$

It is not a priori the case that h/h^+ is an integer, but this is the case. Since K/K^+ is ramified at p (and at ∞), it follows by class field theory that the class number of h^+ divides h. Additionally, one may verify that the ideal class group of K^+ injects into that of K naturally, under inclusion of ideals. Therefore the quotient not only is an integer, but has arithmetic meaning. We draw the following definition.

Definition A.1.21. The *negative part* h^- *of the class number* h *of* K is the quotient such that $h^-h^+ = h$.

We aim to show that $p \mid B_k$ implies that $p \mid h^-$. The next step toward this goal is to match up the L-values above with Bernoulli numbers.

Recall that if $\zeta(s)$ is the Riemann zeta function, $\zeta(0) = -B_1 = 1/2$, and a similarly $\zeta(1-n) = -B_n$ for every positive integer n. In just the same way, one may define generalized Bernoulli numbers for Dirichlet characters, and get a similar relation with the L-function associated to that character.

Definition A.1.22 ([27], p. 31)**.**

$$\sum_{n=0}^{f_\chi} \frac{\chi(a)te^{at}}{e^{f_\chi t}-1} = \sum_{n=0}^{\infty} B_{n,\chi}\frac{t^n}{n!}.$$

We are concerned with these numbers when $n = 1$, because (by [27], Thm. 4.2)

$$L(0,\chi) = -B_{1,\chi}.$$

Therefore we may rewrite Proposition A.1.20 as

Corollary A.1.23.

$$h^- = 2p \prod_{X_K \ni \chi\ odd} \left(-\frac{1}{2}B_{1,\chi}\right),$$

Now we have shown that a factor of h may be written as a product of generalized Bernoulli numbers. All that remains to prove Kummer's criterion is to connect the generalized Bernoulli numbers with the usual ones that we first introduced.

The following formula is the starting point for drawing this connection.

Fact A.1.24 ([27], p. 32)**.** As long as χ is not a trivial Dirichlet character, we have that

$$B_{1,\chi} = \frac{1}{f_\chi}\sum_{a=1}^{f} \chi(a)a,$$

recalling that for any Dirichlet character χ of K, $f_\chi = p$.

At this point it is useful to introduce the Teichmüller character, which was used heavily throughout this paper.

Definition A.1.25. Now choose $\omega : (\mathbb{Z}/p\mathbb{Z})^\times \longrightarrow \mu_{p-1} \subset \mathbb{Q}(\mu_{p-1}) \hookrightarrow \mathbb{C}$ to be the generator of X_K such that

$$\omega(a) \equiv a \pmod{p}.$$

Note well that p is not prime in $\mathbb{Q}(\mu_{p-1})$ (see Remark 3.5).

Here it will be used to canonically associate the set of $B_{1,\chi}$ for odd χ with the B_k for k even, $2 \leq k \leq p-1$.[7] The correct choice is $B_{1,\omega^{k-1}}$, which we now demonstrate.

Proposition A.1.26. *Choose an even integer k, $2 \leq k \leq p-1$. The set of characters $B_{1,\omega^{k-1}}$ for these values of k is the entire set of odd characters in X_K, and*

$$B_{1,\omega^{k-1}} \equiv \begin{cases} \frac{B_k}{k} \pmod{p} & \text{if } k \neq p-1 \\ \frac{-1}{p} \pmod{p\mathcal{O}_{\mathbb{Q}(\mu_{p-1})}} & \text{if } k = p-1. \end{cases}$$

Proof. Choose such a k. Use the easily verified congruence $\omega(n) \equiv n^p \pmod{p^2}$ and Fact A.1.24 with $f_{\omega^{k-1}} = p$ to get that

$$pB_{1,\omega^{k-1}} \equiv \sum_{n=1}^{p-1} n^{1+p(k-1)} \pmod{p^2}.$$

On the other hand, we have ([2], p. 385)

$$pB_t \equiv \sum_{n=1}^{p-1} n^t \pmod{p^2},$$

which is a congruence of p-integral quantities by the von-Staudt–Clausen Theorem (Theorem A.1.4. Hence

$$pB_{1,\omega^{k-1}} \equiv pB_{1+p(k-1)} \pmod{p^2}$$

Say that $k \neq p-1$. Then Since $1 + p(k-1) \not\equiv 0 \pmod{p-1}$, the Kummer congruence (Theorem A.1.3 implies directly that

$$pB_{1,\omega^{k-1}} \equiv p\frac{B_k}{k} \pmod{p^2}.$$

for even k, which is what we desired. If on the other hand $k = p-1$, then

$$pB_{1,\omega^{k-1}} \equiv pB_{1+p(k-1)} \equiv -1 \pmod{p^2}$$

by the von Staudt-Clausen theorem, completing the proof. □

Now we can prove something a bit stronger than the basic statement of the forward direction of Kummer's criterion.

Proposition A.1.27. *The negative part of the class number, h^-, is divisible by p if and only if some Bernoulli number B_k for even k, $2 \leq k \leq p-3$, is divisible by p. Furthermore, if p divides t distinct such Bernoulli numbers B_k, then $p^t \mid h^-$.*

[7] We have restricted our attention to $2 \leq k \leq p-3$ in the rest of the material on Ribet's converse. While it is already clear why this is the case from the von Staudt-Clausen theorem, the following proposition makes it clear that while $k \equiv 0 \pmod{p-1}$ is excluded, it does play an important role.

Proof. By Corollary A.1.23,

$$h^- = 2p \prod_{\text{odd } \chi \in X_K} \left(-\frac{1}{2}B_{1,\chi}\right) = 2p \prod_{k=2 \text{ even}}^{p-1} \left(-\frac{1}{2}B_{1,\omega^{k-1}}\right)$$

since ω is odd (as $\omega(-1) \equiv -1 \pmod{p}$).

The $k = p-1$ term is exceptional; Proposition A.1.26 states that $pB_{\omega^{-1}} \equiv -1 \pmod{p^2}$. Therefore $(2p)(-\frac{1}{2}B_{1,\omega^{p-2}}) \equiv 1 \pmod{p}$, and we end up with

$$h^- \equiv \prod_{k=2,\ \text{even}}^{p-3} \left(-\frac{1}{2}B_{1,\omega^{k-1}}\right) \pmod{p}$$

which by our recent calculation can be written

$$h^- \equiv \prod_{k=2,\ \text{even}}^{p-3} \left(-\frac{1}{2}\frac{B_k}{k}\right) = \prod_{k=2,\ \text{even}}^{p-3} \left(-\frac{1}{2}\zeta(1-k)\right). \tag{A.1.4}$$

As all quantities B_k/k are p-integral, the first part of the proposition is complete.

To prove the second part, simply note that as $B_{1,\omega^{-1}}$ always has p-adic valuation[8] -1, Corollary A.1.23 implies that the p-adic valuation of h^- is the sum of the p-adic valuations of the B_k for $2 \le k \le p-3$, k even. This is even stronger that what the proposition required us to prove. □

This completes one direction of Kummer's criterion, as we record here.

Corollary A.1.28. *If an odd prime p divides the numerator of B_k for some even k, $2 \le k \le p-3$, then p is irregular.*

Remark A.1.29. The converse statement, which would complete Kummer's criterion, follows upon showing that if $p \mid h+$, then $p \mid h^-$ as well. This involves character computations that are best presented with p-adic L-functions ([27], Cor. 8.17).

A.2. **Eisenstein Series on** $\mathrm{SL}_2(\mathbb{Z})$**.** The Eisenstein series G_k used throughout the document is scalar multiple of the following somewhat more natural Eisenstein series,

$$G_k'(z) = \sum_{(c,d)}{}' \frac{1}{(cz+d)^k}, \quad z \in \mathcal{H}, \tag{A.2.1}$$

where the ' indicates that the 0-vector is skipped. We may readily rearrange this sum to get

$$G_k'(z) = \sum_{d \neq 0} d^{-k} + 2\sum_{c=1}^{\infty} \left(\sum_{d \in \mathbb{Z}} (cz+d)^{-k}\right). \tag{A.2.2}$$

It is a nice exercise to observe that this is a modular form of weight k on $\mathrm{SL}_2(\mathbb{Z})$.

Using the identity ([11], p. 5)

$$\sum_{d \in \mathbb{Z}} (z+d)^{-k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{m=1}^{\infty} m^{k-1} q^m$$

[8] Appropriately extended from $\mathbb{Q}$ to $\mathbb{Q}(\mu_{p-1})$.

and noticing the Riemann zeta function in formula (A.2.2) we find that

$$\begin{aligned} G_k'(z) &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!}\sum_{c=1}^{\infty}\sum_{m=1}^{\infty} m^{k-1}q^{cm} \\ &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n, \end{aligned} \tag{A.2.3}$$

where σ_{k-1} is the $(k-1)$-power divisor function

$$\sigma_{k-1}(n) = \sum_{0<m|n} m^{k-1}.$$

Finally, recalling the identity

$$\zeta(k) = -\frac{(2\pi i)^k}{2k!}B_k,$$

which is equivalent to Equation (1.4) via the functional equation for the zeta function, we find that a scalar multiple of G_k' is the familiar G_k appearing in 3.1.

A.3. **Background on Modular Forms.** Though it will require us to introduce some new information and notation quickly, it will be helpful to make a precise statement of what we will construct before we begin to go about it. We begin with a hasty list of relevant definitions having to do with modular forms. Modular forms are complex analytic functions on the upper half plane

$$\mathcal{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$$

that can be extended continuously to $\mathcal{H}$ plus its cusps

$$\mathcal{H}^* = \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q})$$

and that obey certain transformation properties under the fractional linear transformation action certain subgroups of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{C}$. In the course of the following definitions, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be an element of $\mathrm{SL}_2(\mathbb{Z})$. We follow [11], §1.2 in this presentation.

Definition A.3.1. A congruence subgroup of the modular group $\mathrm{SL}_2(\mathbb{Z})$ is a subgroup that contains the principal congruence subgroup

$$\Gamma(N) = \{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv I_2 \pmod{N}\}$$

for some positive integer N. Some of the standard modular subgroups other than the principle congruence subgroups $\Gamma(N)$ itself are

$$\Gamma_1(N) = \left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N}\right\}$$

and

$$\Gamma_0(N) = \left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N}\right\},$$

i.e. those elements such that $c \equiv 0 \pmod{N}$.

Now we can take a first step toward defining a modular form.

Definition A.3.2. Let k be an integer. Say that a meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k over a congruence subgroup* Γ provided that

$$f(\gamma(z)) = (cz + d)^k f(z)$$

for all $\gamma \in \Gamma$ and $z \in \mathcal{H}$.

In order to simplify notation, we introduce the following operator that allows us to write down a concise definition of weak modularity.

Definition A.3.3. The *weight-k operator* $[\cdot]_k : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{End}(\{f : \mathcal{H} \to \mathbb{C}\})$ for an integer k is an action of $\mathrm{SL}_2(\mathbb{Z})$ on functions $f : \mathcal{H} \to \mathbb{C}$, written on the right as

$$(f[\gamma]_k)(z) = (cz + d)^{-k} f(\gamma \cdot z)$$

Thus a meromorphic function f on $\mathcal{H}$ is weakly modular of weight k provided that $f[\gamma]_k = f$ for all $\gamma \in \Gamma$.

Say for the moment that $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Then since $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, a weakly modular function of weight k on $\mathrm{SL}_2(\mathbb{Z})$ is periodic with period 1. Therefore a modular form has a Fourier development (conditional on convergence properties), which we will call a "q-expansion" because we take $q = e^{2\pi i z}$ and then commonly write

$$f(z) = \sum_{n \geq n_0} a_n q^n.$$

More generally, if f is weakly modular on the congruence subgroup $\Gamma \supseteq \Gamma(N)$, then f has such a Fourier development with q replaced by $q_N = e^{2\pi i z/N}$.

Returning to the case that f is weakly modular on $\mathrm{SL}_2(\mathbb{Z})$ and considering f as a function of q, it is then a function on the punctured unit disc $\{z \in \mathbb{C} : 0 < |z| < 1\}$. We say that f is "holomorphic (resp. meromorphic) at infinity" if it can be extended holomorphically (resp. meromorphically) to $q = 0$, the terminology coming from the fact that $q \to 0$ as $\Im(z) \to +\infty$. We require this condition because then not only is $f(dz)^{k/2}$ a meromorphic differential on the Riemann surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$, but it can also be extended to the compact Riemann surface $\mathrm{SL}_2(\mathbb{Z})\backslash(\mathcal{H} \cup \{\infty\})$, a "modular curve" obtained by adjoining a point at infinity.

In fact, this point at infinity is the simplest instance of idea of a cusp is much more general and can apply to all congruence subgroups. The following definitions make rigorous the above comments on $\mathrm{SL}_2(\mathbb{Z})$ and extend the ideas to congruence subgroups.

Definition A.3.4. Let Γ be a congruence subgroup. A *cusp* of Γ is a Γ-equivalence class of $\mathbb{Q} \cup \{\infty\}$.

The following examples are useful to make this definition concrete, and are also the primary examples needed in this paper.

Example A.3.5. When $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, there is only one cusp, since any rational number r/s where $(r, s) = 1$ or $r = 0$ is sent to infinity by $\gamma = \left(\begin{smallmatrix} a & b \\ -s & r \end{smallmatrix}\right)$ where $a, b \in \mathbb{Z}$ are chosen so that $ar + bs = 1$. However, when $\Gamma = \Gamma_0(p)$ where p is prime, then there are two cusps, the class containing ∞ and the class containing 0. These classes are

$$\left\{\frac{r}{s} \in \mathbb{Q} : p \mid s \text{ and } r \neq 0\right\} \cup \{\infty\} \quad \text{and} \quad \left\{\frac{r}{s} \in \mathbb{Q} : p \nmid s\right\}, \tag{A.3.1}$$

respectively.

The condition on weakly modular functions on Γ analogous to the "holomorphic at infinity" condition on $\mathrm{SL}_2(\mathbb{Z})$ is holomorphicity at cusps. This definition is given by sending a cusp of Γ to infinity with $\mathrm{SL}_2(\mathbb{Z})$.

Definition A.3.6. Let f be weakly modular of weight k on Γ. Then f is *holomorphic the cusps of* Γ provided that $f[\gamma]_k$ is holomorphic at infinity for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

Now we have all of the ingredients to define a modular form.

Definition A.3.7. A *modular form* (resp. *automorphic form*) of weight k on a congruence subgroup Γ, $\Gamma(N) \subseteq \Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$, is a function $f : \mathcal{H} \to \mathbb{C}$ such that

(1) f is a weakly modular function of weight k,

(2) f is holomorphic (resp. meromorphic), and

(3) f is holomorphic (resp. meromorphic) at all cusps of Γ.

Moreover, if f is a modular form and $f[\gamma]_k$ vanishes at infinity for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we say that f vanishes at the cusps of Γ and call f a *cusp form*. We denote the $\mathbb{C}$-vector space of such modular/cusp/automorphic forms as

$$M_k(\Gamma), \quad S_k(\Gamma), \quad A_k(\Gamma), \quad \text{respectively.}$$

The best way I know to understand the naturality of these requirements on a modular form is to consider their relationship with modular curves (see [11], Ch. 2).

Definition A.3.8. Let Γ be a congruence subgroup. The *modular curve* $Y(\Gamma)$ is the quotient space of orbits of the action of Γ on $\mathcal{H}$,

$$Y(\Gamma) = \Gamma \backslash \mathcal{H}.$$

Similarly, the *modular curve* $X(\Gamma)$ is

$$X(\Gamma) = \Gamma \backslash \mathcal{H}^*.$$

We write $Y(N) = Y(\Gamma(N)), Y_0(N) = Y(\Gamma_0(N)), Y_1(N) = Y(\Gamma_1(N))$, and similarly for $X(\Gamma)$.

For a given congruence subroup Γ the modular curves $Y(\Gamma)$ and $X(\Gamma)$ are Riemann surfaces, and $X(\Gamma)$ is compact Riemann surface (see [23], §1.5).

Similar to what we noted in the case that $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ above, automorphic forms of weight k on Γ correspond to degree $k/2$ meromorphic differentials on $X(\Gamma)$. Since we are most concerned with the situation when $k = 2$ and our differentials are holomorphic and this situation is simpler than general k, so we quote this more

Proposition A.3.9 ([23], Cor. 2.17)**.** *Let Γ be a congruence subgroup. Then $S_2(\Gamma)$ is isomorphic to the $\mathbb{C}$-vector space $\Omega^1(X)$ of all degree 1 holomorphic differentials under the map $f \mapsto f(dz)$. It follows from the Riemann-Roch theorem that the dimension of these two spaces is equal to the genus of $X(\Gamma)$.*

While one might expect that the space of modular forms $M_2(\Gamma)$ would satisfy this proposition instead of $S_2(\Gamma)$, the key is that dz itself has poles at the cusps that are cancelled by the zeros of cusp forms ([23], §2.4). This is why we narrow our focus to cusp forms in §4.

Having defined modular forms in the previous appendix, we now go on to discuss the action of Hecke operators on them. The discussion of Hecke actions was minimized in the main part of the text as they would distract too much from the main thrusts of the background to Ribet's proof. However, Hecke theory, mostly in the theory of newforms, are used liberally in the main parts of this essay. Therefore we record here the details of Hecke action for reference.

Hecke operators are an example of double coset operators, which, naturally, are based on double cosets.

Definition A.3.10. Let Γ_1 and Γ_2 be congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. A *double coset* in $\mathrm{GL}_2^+(\mathbb{Q})$ is a set

$$\Gamma_1\alpha\Gamma_2 = \{\gamma_1\alpha\gamma_2 : \gamma_i \in \Gamma i\}. \tag{A.3.2}$$

Such double cosets act on modular forms and modular curves, both of which we defined in §2. The double coset A.3.2 sends modular forms on Γ_1 to modular forms on Γ_2. The orbit space $\Gamma_1\backslash\Gamma_1\alpha\Gamma_2$ is finite, therefore we may define the action to be

Definition A.3.11 ([11], Def. 5.1.3)**.** For congruence subgroups Γ_1 and Γ_2 of $\mathrm{SL}_2(\mathbb{Z})$ and $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, the weight-$k$ $\Gamma_1\alpha\Gamma_2$ operator takes functions $f \in M_k(\Gamma_1)$ to

$$f[\Gamma_1\alpha\Gamma_2]_k = \sum_j f[\beta_j]_k$$

where $\{\beta_j\}$ are orbit representatives, i.e. $\Gamma_1\alpha\Gamma_2 = \bigcup_j \Gamma_1\beta_j$ is a disjoint union.

One should verify that these Hecke operators are well-defined despite the choice of $\{\beta_j\}$, and send cusp forms to cusp forms.

The Hecke operators for the purposes of this exposition are double cosets with $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$, and so are automorphisms of $M_k(\Gamma_1(N))$. The first type of Hecke operator is strongly connected to our notion of type from the last section. Choose any $\gamma \in \Gamma_0(N)$, recalling that $\Gamma_0(N) \supset \Gamma_1(N)$ and conventionally $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then on noting that $\Gamma_1(N) \lhd \Gamma_0(N)$ we have

$$f[\Gamma_1(N)\gamma\Gamma_1(N)]_k = f[\gamma]_k.$$

As f is $\Gamma_1(N)$-invariant and the coset of γ in $\Gamma_1(N)/\Gamma_0(N)$ is determined by d, we can write this double coset operator as the *diamond operator*

$$\langle d\rangle\, f = f[\alpha]_k \quad \text{for any } \alpha = \begin{pmatrix} a & b \\ c & \delta \end{pmatrix} \text{ where } \delta \equiv d \pmod N. \tag{A.3.3}$$

To see the connection with the type of a modular form on $\Gamma_1(N)$, note that

$$M_k(N, \varepsilon) = \{f \in M_k(\Gamma_1(N)) : \langle d\rangle\, f = \varepsilon(d)f \text{ for all } d \in (\mathbb{Z}/N\mathbb{Z})^\times\}.$$

Our other Hecke operator, denoted T_p where p is prime, is given by $\gamma = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. The $\Gamma_1(N)$-orbits in the double coset $\Gamma_1(N)\gamma\Gamma_1(N)$ are represented by the matrices $\beta_j = \left(\begin{smallmatrix} 1 & j \\ 0 & p \end{smallmatrix}\right)$ and, if $p \nmid N$, β_∞. Call this set of representatives $B(p, N)$. They appear in the following description of the action of T_p on $f \in M_k(\Gamma_1(N))$.

Proposition A.3.12 ([11], Prop. 5.2.1)**.** *With notation as above, the operator $T_p = [\Gamma_1(N)\gamma\Gamma_1(N)]_k$ on $M_k(\Gamma_1(N))$ is given by*

$$T_p f = \begin{cases} \sum_{j=0}^{p-1} f[\left(\begin{smallmatrix} 1 & j \\ 0 & p \end{smallmatrix}\right)]_k & \text{if } p \mid N \\ \sum_{j=0}^{p-1} f[\left(\begin{smallmatrix} 1 & j \\ 0 & p \end{smallmatrix}\right)]_k + f[\left(\begin{smallmatrix} m & n \\ N & p \end{smallmatrix}\right)\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)]_k & \text{if } p \nmid N, \text{ where } mp - nN = 1. \end{cases}$$

Also, we can verify from Proposition A.3.12 that the effect of T_p on Fourier series is as follows. Let $a_n(f)$ denote the nth coefficient of a modular form f.

Proposition A.3.13 ([11], Proposition 5.2.2)**.** *Let $f \in M_k(\Gamma_1(N))$ and $\varepsilon : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ such that $f \in M_k(\Gamma_0(N), \varepsilon)$. Since $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_1(N)$, f has a Fourier expansion*

$$f(z) = \sum_{n=0}^{\infty} a_n(f) q^n.$$

Then $T_p f \in M_k(\Gamma_0(N), \varepsilon)$ and its Fourier expansion is

$$(T_p f)(z) = \sum_{n=0}^{\infty} \left(a_{np}(F) + \varepsilon(p) p^{k-1} a_{n/p}(f) \right) q^n,$$

where $a_{n/p}(f) := 0$ when $n/p \notin \mathbb{Z}$.

From this action on Fourier coefficients comes the correlation between Hecke eigenvalues and the coefficients. We note in this corollary that the phenomenon is not restricted to Hecke operators of prime index.

Corollary A.3.14. *If a modular form f is an eigenvector of the Hecke operator T_n for, then $a_n(f)/a_1(f)$ is the eigenvalue of T_n.*

We used this fact many times in §3.

Hecke operators T_n can be defined for any $n \in \mathbb{Z}^+$ in terms of the T_p; for example $T_m T_n = T_{mn}$ for $(m, n) = 1$. However, it is the prime index operators that will be most useful to us since they are most simple but also generate the T_n for all n over $\mathbb{Z}$. The Hecke operators T_n and $\langle d \rangle$ where $(pn, N) = 1$ are are pairwise commutative, preserve cusp forms, and are normal (i.e. commute with their adjoint) with respect to the Petersson inner product on $M_k(\Gamma_1(N))$.[9] Together, all of these Hecke operators form the algebra

$$\mathbb{T}^0 = \mathbb{Z}[\{\langle n \rangle, T_n : (n, N) = 1\}]$$

which is a subalgebra of the full Hecke algebra.

Definition A.3.15. The *Hecke algebra* is the $\mathbb{Z}$-algebra

$$\mathbb{T} = \mathbb{Z}[\{T_n, \langle n \rangle : n \in \mathbb{Z}^+\}],$$

where the weight and level of the Hecke operators is left implicit.

All of the material mentioned can be studied further in [11], §§5.2-5.4.

Because of these properties we have listed, the spectral theorem of linear algebra implies that there exists an orthogonal basis of simultaneous eigenforms for the Hecke subalgebra $\mathbb{T}^0$. However, we would like to find eigenforms with respect to all Hecke operators, because we want to be able to write down a cusp form that is unique up to a constant if we are given a system of Hecke eigenvalues. In this case, if f has $a_n(f) = 1$, we say that f is a *normalized* eigenform, and consequently the nth coefficient $a_n(f)$ is the eigenvalue $\lambda(n)$ of T_n. This extension from an eigenform for $\mathbb{T}^0$ to one for $\mathbb{T}$ is possible if we study *newforms*, a theory due to Atkin-Lehner [1], which we now overview.

[9] At least one of the two factors in the inner product must be a cusp form in order to define the Petersson inner product, but this still suffices.

Definition A.3.16. The space $S_k(\Gamma_1(N))$ is the direct sum the *old* subspace $S_k(\Gamma_1(N))^{\text{old}}$ and the *new* subspace $\langle S_k(\Gamma_1(N))^{\text{new}} \rangle$. Clearly when $M \mid N$, then $S_k(\Gamma_1(M)) \subset S_k(\Gamma_1(N))$, so some modular forms in $S_k(\Gamma_1(N))$ may be inherited. And in fact whenever $aM \mid N$ then $g(az) \in S_k(\Gamma_1(N))$ is inhertied from $g \in S_k(\Gamma_1(M))$. These inherited forms compose $S_k(\Gamma_1(N))^{\text{old}}$, and its orthogonal complement under the Petersson inner product is the new subspace. A *newform* is a member of a set $S_k(\Gamma_1(N))^{\text{new}}$, which is an orthogonal normalized basis of eigenforms for the new subspace.

We cite the following information from the theory of newforms.

Fact A.3.17. The Hecke algebra $\mathbb{T}$ preserves the decomposition into new and old subspaces. While there are eigenforms in $S_k(\Gamma_N)$ with respect to $\mathbb{T}^0$, if such an eigenform is in the new subspace then it is an eigenform for $\mathbb{T}$ as well. This is not necessarily the case in the old subspace. Finally, any Galois conjugate of a newform is again a newform.

Proof. See [5], Theorem 1.22 for a fuller statement, or [17] for the first few facts. The result on Galois conjugation can be found in [11], Theorem 6.5.4. □

As in §2, we will continue to deal only with $S_2(\Gamma_1(p))$, which has no oldforms since $M_2(\mathrm{SL}_2(\mathbb{Z}))$ is trivial. We used this fact in the proof of Proposition 3.16, namely, we constructed an eigenform with respect to $\mathbb{T}^0$ and then claimed that since it was a newform it must be an eigenform for $\mathbb{T}$ and therefore have a completely prescribed Fourier expansion when normalized.

A.4. **Galois Representations.** In this appendix we collect a few useful definitions and facts that are a useful reference for Galois representation. Actually, the collection as a whole is too short and not particularly useful, but I hope to add to it.

Definition A.4.1. Let d be a positive integer. A *d-dimensional p-adic Galois representation* is a d-dimensional topological vector space V over K, where K is a finite extension of $\mathbb{Q}_p$, that is also a $G_{\mathbb{Q}}$-module such that the action

$$V \times G_{\mathbb{Q}} \to V, \qquad (v, \sigma) \mapsto v^{\sigma}$$

is continuous. If V' is another such representation and there is a continuous $G_{\mathbb{Q}}$ module isomorphism of K-vector spaces $V \xrightarrow{\sim} V'$ then V and V' are said to be equivalent.

Note that we have the usual ambiguity of term "representation" to mean both the map into the space of automorphisms of a vector space and the vector space itself.

The fact that any representation is similar to an integral one is an interesting application of the p-adic topology.

Proposition A.4.2 ([11], Prop. 9.3.5)**.** *Let $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_d(K)$ be a Galois representation. Then ρ is similar to a Galois representation $\rho' : G_{\mathbb{Q}} \to \mathrm{GL}_d(\mathcal{O}_K)$.*

Proof. Let $V = K^d$ and $\Lambda = \mathcal{O}_K^d$. Since Λ is compact in V and $G_{\mathbb{Q}}$ is compact as well, so is the image $\Lambda' = \rho(G_{\mathbb{Q}})\Lambda$. Therefore the image lies in $\lambda^{-r}\Lambda$ for some $r \in \mathbb{Z}^+$. The image is finitely generated, it contains Λ so its rank is at least d, it is free since $\mathcal{O}_K$ is a principal ideal domain, and so its rank is precisely d. It is preserved by the action of $G_{\mathbb{Q}}$. Thus any $\mathcal{O}_K$ basis of Λ' gives the desired ρ'. □

References

1. A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160. MR MR0268123 (42 #3022)
2. A. I. Borevich and I. R. Shafarevich, *Number theory*, Translated from the Russian by Newcomb Greenleaf. Pure and Applied Mathematics, Vol. 20, Academic Press, New York, 1966. MR MR0195803 (33 #4001)
3. Frank Calegari, *Book Review: "A First Couse in Modular Forms," by F. Diamond and J. Shurman*, Bull. AMS **43** (2006), 415–421.
4. Charles W. Curtis and Irving Reiner, *Representation theory of finite groups and associative algebras*, Pure and Applied Mathematics, Vol. XI, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1962. MR MR0144979 (26 #2519)
5. Henri Darmon, Fred Diamond, and Richard Taylor, *Fermat's last theorem*, Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), Int. Press, Cambridge, MA, 1997, pp. 2–140. MR MR1605752 (99d:11067b)
6. P. Deligne, *Formes modulaires et représentations ℓ-adiques*, Sém. Bourbaki no. 355, 1968/69, Springer, Berlin, 1971, pp. 139–172. Lecture Notes in Math., Vol. 179.
7. P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1973, pp. 143–316. Lecture Notes in Math., Vol. 349. MR MR0337993 (49 #2762)
8. Pierre Deligne, *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math. (1974), no. 43, 273–307. MR MR0340258 (49 #5013)
9. Pierre Deligne and Jean-Pierre Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. (4) **7** (1974), 507–530 (1975). MR MR0379379 (52 #284)
10. Fred Diamond and John Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem (Toronto, ON, 1993–1994), CMS Conf. Proc., vol. 17, Amer. Math. Soc., Providence, RI, 1995, pp. 39–133. MR MR1357209 (97g:11044)
11. Fred Diamond and Jerry Shurman, *A first course in modular forms*, Graduate Texts in Mathematics, vol. 228, Springer-Verlag, New York, 2005. MR MR2112196 (2006f:11045)
12. Jacques Herbrand, *Sur les classes des corps circulaires*, J. Math. Pures et Appliquées (9) **11** (1932), 417–441.
13. Jun-ichi Igusa, *Kroneckerian model of fields of elliptic modular functions*, Amer. J. Math. **81** (1959), 561–577. MR MR0108498 (21 #7214)
14. Gerald J. Janusz, *Algebraic number fields*, second ed., Graduate Studies in Mathematics, vol. 7, American Mathematical Society, Providence, RI, 1996. MR MR1362545 (96j:11137)
15. C. Khare, *Notes on Ribet's converse to Herbrand*, Unpublished notes.
16. V. A. Kolyvagin, *Euler systems*, The Grothendieck Festschrift, Vol. II, Progr. Math., vol. 87, Birkhäuser Boston, Boston, MA, 1990, pp. 435–483. MR MR1106906 (92g:11109)
17. Serge Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1976, Grundlehren der mathematischen Wissenschaften, No. 222. MR MR0429740 (55 #2751)
18. ———, *Cyclotomic fields I and II*, second ed., Graduate Texts in Mathematics, vol. 121, Springer-Verlag, New York, 1990, With an appendix by Karl Rubin. MR MR1029028 (91c:11001)
19. B. Mazur and A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Invent. Math. **76** (1984), no. 2, 179–330. MR MR742853 (85m:11069)
20. Kenneth A. Ribet, *A modular construction of unramified p-extensions of* $\mathbb{Q}(\mu_p)$, Invent. Math. **34** (1976), no. 3, 151–162. MR MR0419403 (54 #7424)
21. Jean-Pierre Serre, *Abelian l-adic representations and elliptic curves*, McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam, 1968. MR MR0263823 (41 #8422)
22. Jean-Pierre Serre and John Tate, *Good reduction of abelian varieties*, Ann. of Math. (2) **88** (1968), 492–517. MR MR0236190 (38 #4488)
23. Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan, vol. 11, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kano Memorial Lectures, 1. MR MR1291394 (95e:11048)
24. Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 106, Springer-Verlag, New York, 1992, Corrected reprint of the 1986 original. MR MR1329092 (95m:11054)

25. John Tate, *The non-existence of certain Galois extensions of* $\mathbf{Q}$ *unramified outside* 2, Arithmetic geometry (Tempe, AZ, 1993), Contemp. Math., vol. 174, Amer. Math. Soc., Providence, RI, 1994, pp. 153–156. MR MR1299740 (95i:11132)
26. Francisco Thaine, *On the ideal class groups of real abelian number fields*, Ann. of Math. (2) **128** (1988), no. 1, 1–18. MR MR951505 (89m:11099)
27. Lawrence C. Washington, *Introduction to cyclotomic fields*, second ed., Graduate Texts in Mathematics, vol. 83, Springer-Verlag, New York, 1997. MR MR1421575 (97h:11130)
28. André Weil, *Variétés abéliennes et courbes algébriques*, Actualités Sci. Ind., no. 1064 = Publ. Inst. Math. Univ. Strasbourg 8 (1946), Hermann & Cie., Paris, 1948. MR MR0029522 (10,621d)
29. A. Wiles, *On ordinary λ-adic representations associated to modular forms*, Invent. Math. **94** (1988), no. 3, 529–573. MR MR969243 (89j:11051)
30. ———, *The Iwasawa conjecture for totally real fields*, Ann. of Math. (2) **131** (1990), no. 3, 493–540. MR MR1053488 (91i:11163)
31. Andrew Wiles, *Modular curves and the class group of* $\mathbf{Q}(\zeta_p)$, Invent. Math. **58** (1980), no. 1, 1–35. MR MR570872 (82j:12009)