Ramanujan-Petersson Conjecture

Yasutaka Ihara

submitted 2019/10/24; minor revisions by 2021/09/02

1 Introduction

Hecke theory of modular forms gives rise to various Dirichlet series $L(s)$ enjoying distinguished analytic properties including the functional equation and the Euler product expansion $= \prod_p H^{(p)}(p^{-s})^{-1}$, where p runs over all prime numbers and each $H^{(p)}(u)$ is a polynomial. The famous Ramanujan-Petersson conjecture (*abbrev.* RPC) was on the absolute values of zeros of each *Hecke polynomial* $H^{(p)}(u)$. On the other hand, Hasse's zeta function $Z_V(s)$ of an algebraic variety V, say, over the rational number field $\mathbb{Q}$ is *defined* as the Euler product $\prod_p Z_V^{(p)}(p^{-s})$, where $Z_V^{(p)}(u)$ is the congruence zeta function of the reduction mod p of V (for almost all primes p) which is a rational function. The properties of zeros and poles of $Z_V^{(p)}(u)$ (resp. $Z_V(s)$) are in direct connections with the arithmetic geometry of $V(\mathrm{mod}\ p)$ (resp. V). The most basic Weil conjecture (*abbrev.* WC) was on the absolute values of zeros and poles of $Z_V^{(p)}(u)$ when V mod p is complete and non-singular. The former Hecke L-function is an L-function associtated with an automorphic representation, while the latter Hasse zeta function is that associated with a system of Galois representations arising from ℓ -adic cohomology groups of V. The discovery of a new connection between these two types of Dirichlet series would transform information from one to the other; e.g. WC to RPC.

The RPC was proved finally by P.Deligne in two steps; its reduction to some specific case of WC ([Del1] 1968/69), and proofs of WC itself [Del2I],[Del2II] (1974,80). Now, Sato's contribution (1962) to RPC was crucial in the first step. To be precise, based on Michio Kuga's concrete speculation as to which variety V would be relevant to the RPC, i.e., for which V the zeta function $Z_V(s)$ would be directly connected with the Hecke's $L(s)$, he started trying to find an explicit relation between $L(s)$ and $Z_V(s)$ for the variety V proposed by Kuga, and soon found a pathway connecting these two objects. He reached a point far enough to make one feel convinced that Kuga's speculation is correct and Sato's method is at least basically on the right track.

On this work, he made some oral communications, e.g. a colloquium talk '62 and a seminar talk '63 at the University of Tokyo, with some details in the latter talk but without explanations on how his formal computations can be justified. Then suspicions arose among some leading mathematicians in this

field, partly because of its unfortunate note [note] distributed soon after. And years passed (with some related papers by others published but) without any publication by himself on this subject, until the appearance of the decisive paper [Del1] of Deligne in 1969. Deligne also appeals to Kuga variety but his method of proof is fairly different from that employed by Sato. Sato's method of using arithmetical-geometric interpretation of the trace formula was later further developed and became a principal tool in the study of Shimura varieties, notably by R. Langlands starting with [Lg].

The writer Y.I. of this article was a graduate student in the Master's course until spring '63 and then a research associate (Math. Dept., Univ. Tokyo); he attended Sato's both talks mentioned above. He hopes that the readers will allow him for writing from his own understanding (and non-understanding!) and point of view, and to quote his own relevant paper having the subtitle *to validate M. Sato's identity* ('67) where necessary.

2 Brief history until around 1963

(II-1) The RPC is a conjecture on absolute values of eigenvalues of Hecke operators acting on the space of modular cuspidal newforms, and can be formulated in analytic terms (cf.IV-1), but the development of arithmetic algebraic geometry especially by A.Weil made one aware of similarities and possible connections between RPC and the Weil conjecture WC on absolute values of the Frobenius eigenvalues of complete non-singular algebraic varieties V over finite fields (cf.IV-3). Rather, it should be said that Weil's study related to WC itself had been motivated by this possible connection. For the one-dimensional case, WC is his theorem WT. When $n = \dim V > 1$, it consists of (i) the existence of a good cohomology group H^i for each i $(0 \leq i \leq 2n)$ of characteristic 0; (ii) a conjecture on the complex absolute values of eigenvalues of the Frobenius morphism acting on H^i for each i. Among them (i) was being established by the ℓ-adic etale cohomology theories of A.Grothendieck and M.Artin around 1963-64. The WC before this meant a little weaker conjecture. Thus the basic question was:

Can one find a variety (complete, non-singular) V such that the Weil conjecture for V implies the RPC for modular cuspidal newforms?.

Now we must specify the weight k of the modular form in question, and sometimes also its level N.

(II-2) The case $k = 2$. In this case, RPC was solved by M.Eichler [E1] and G. Shimura [Sh1](for almost all p), J-I. Igusa [Ig](for individual p not dividing N), by reduction to the WT for the modular curve of level N.

(II-3) The case $k > 2$. (i) Inspired by an earlier work of Eichler [E2], Shimura constructed in [Sh2] a $\mathbb{Z}$-lattice D_k of rank $= 2\dim S_k$ in the complex vector space S_k of modular cusp forms of weight k (w.r.t. a given congruence subgroup of

the modular group). The lattice consists of all those forms in S_k having rational integral real parts for all the periods w.r.t. the Eichler integration. It is such that (a) the quotient $A_k = S_k/D_k$ carries a natural structure of a polarized abelian variety over the complex number field $\mathbb{C}$, which in the case of $k = 2$ is nothing but the Jacobian of the associated modular curve, (b) each Hecke operator leaves D_k stable and induces an endomorphism of A_k. This work of Shimura contained a crucial key to the problem. It was, dazzling as it seemed, not the abelian variety A_k itself as the "wanted V". It was the $\mathbb{Z}$-module D_k on which the Hecke operators act, or rather, its base change $\otimes\mathbb{Z}_\ell$ (cf. III-3,4).

(ii) A new candidate, so simple and straightforward, occurred to M. Kuga *on some happy day* (Kuga's usual way of saying) for the above basic question. It is a fiber variety whose base is the modular curve parametrizing elliptic curves with level N structures and whose fiber is the product of $(k-2)$ copies of the parametrized elliptic curve, or its quotient Kummer variety by the diagonal ± 1 . To be precise, *some* smooth compactification of this $(k-1)$-dimensional variety. This family turned out to be the correct candidate. Those who know his once-hidden but decisive contribution well enough started calling it Kuga variety, and I shall follow this naming.

(iii) Kuga and Sato were both staying at IAS, Princeton, in the academic year 1961–62 (Sato's second year, Kuga's first), and Y.I. heard from each of them that Kuga mentioned this idea to Sato on some day in some conversation between them.

(iv) In 1962, around summer or autumn, news reached Japan: "Sato succeeded in reducing RPC to WC". Non-specialists only murmurred "what conjecture to what conjecture?", but for us this was surprising news. On his return to Tokyo (Tokyo University of Education), he was invited to give a talk in the autumn colloquim, and then a seminar talk in winter, both at the University of Tokyo. As a graduate student there, I (Y.I.) attended both. I felt I understood the basic outline. But the proof related to justification of formal calculations was not given and we were not able to fill this ourselves. We tried to make contact with him, but he was even busier as he was going to leave Tokyo soon. On April '63, Sato moved to Osaka Universiy and his main interest in number theory was already on his new conjecture, the "Sato-Tate conjecture".

(v) Then appeared a note[1], to be cited [note], of his seminar talk taken by a young faculty member in "Sugaku-no-Ayumi", an informal periodical distributed among "new generation mathematicians" in Japan. This caused two waves among specialists concerned:

(!) Observation that the variety which Kuga proposed must be the correct one, and the full proof would be obtained along a line not too far from what Sato indicated.

(?) A question about whether it has really been proved. In [note] the points of proofs are too ambiguous; the main technical difficulty encountered is not indicated. Is this, then, the fault of the speaker or the note-taker?

[1]It was without any approval of Sato, and though quoted here just following previous series of quotations, it is not worth translating/reproducing it in this Volume for several reasons.

What I can say are that [note] reflected almost faithfully what was written on the blackboard, that some additional arguments should be necessary for the proof, and that the final active efforts of the note-taker, even to see Sato, ended in vain.

(vi) As far as this subject is concerned, Sato was a passer-by. His main interest by then included his new theory of hyperfunctions and that of zeta functions of prehomogeneous vector spaces, each as the theory-founder. We know that he also possessed strong interest and amazing power of penetration in number theory. But a passer-by has technical handicaps. As regards the technical problem, I understood three years later [**?**] that *had* Sato found some basic old papers of Deuring's (cf. IV-4), either in the library of IAS or by learning from some colleague in number theory, then he would have completed the proof by himself.

(vii) As a flow, the stimulations (!) and (?) led to Deligne's final solutions of RPC. Thus, Sato's unpublished contribution on this subject certainly had provided a crucial step.

Kuga found the correct target,
Sato made the first breakthrough,
This stimulation led others to the final solution.

In fact, this caused stimulations and combinations of different ideas in multiple steps among a few others which led to more than the solution of the RPC; see III.

(II-4) Sato's idea, computations and some unclarified points.

Here, let me give a brief explanation, assuming basic Hecke theory for the case of the level $N = 1$. A more precise account will be left to (V).

[The $H_k^{(p)}(u)$-side] When $L(s) = L_k(s)$ is associated with the space S_k of modular cusp forms of weight k, each local Hecke polynomial $H^{(p)}(u) = H_k^{(p)}(u)$ is in a direct connection with the traces $\mathrm{tr}(T_k(p^m))$ for all $m \geq 1$ of the Hecke operators $T_k(p^m)$ acting on S_k. By Eichler-Selberg trace formula (cf.IV-2), the characteristic part of this trace is a certain *sum over* the points of intersection $\mathbb{T}(p^m).\Delta$. Here, $\mathbb{T}(p^m)$ denotes the associated Hecke correspondence and Δ the diagonal, both being effective divisors on the product $X \times X$ of the base modular curve X with itself, intersecting with each other properly. These intersection points are parametrized by certain classes of imaginary quadratic quantities depending on p^m. Each *summand* depends also on k and is in fact a polynomial of degree $k-2$ of the associated quadratic quantity.

[The $Z_V^{(p)}(u)$-side] The congruence zeta function of the Kuga variety $V = V_k$ for weight k at a prime p is in a similar type of connection with the number $N_k(p^m)$ of $\mathbb{F}_{p^m}$-rational points of $V_k \bmod p$ for all $m \geq 1$, and this number is naturally a *sum over* the $\mathbb{F}_{p^m}$-rational points ξ of $X \bmod p$. And V_k being the Kuga variety, each summand, the number of rational points of the fiber above ξ, is basically the $(k-2)$-th power of the number of rational points of the elliptic curve E_ξ parametrized by ξ, which can be expressed by the p^m-th Frobenius eigenvalues of E_ξ. Thus, each summand is a polynomial of degree $k-2$ of (either

$\pm p^{m/2}$ or) imaginary quadratic Frobenius eigenvalues.

To compare these two, Sato is led to consider the following modifications, for $m > 1$, of $\mathbb{T}(p^m)$ and $T_k(p^m)$:

$$\mathbb{U}(p^m) = \mathbb{T}(p^m) - p\mathbb{T}(p,p)\mathbb{T}(p^{m-2}) \qquad \text{(a Hecke correspondence)} \qquad (1)$$

$$U_k(p^m) = T_k(p^m) - p^{k-1}T_k(p^{m-2}) \qquad \text{(a linear operator on } S_k) \qquad (2)$$

For the notations $\mathbb{T}(n,n')$, cf. e.g. [Sh0]. The operator $U_k(p^m)$ is induced from the correspondence $\mathbb{U}(p^m)$ in the same way as $T_k(p^m)$ is induced from $\mathbb{T}(p^m)$. This simple pair of modifications led to two amazing observations.

(i)! As regards the formula for $\mathrm{tr}(U_k(p^m))$, by cancellations, it appears merely as a *partial sum* in the formula for $\mathrm{tr}(T_k(p^m))$.

(ii)! If Π (resp.Π') denotes the graph of the p-th Frobenius map of $X \pmod p$ (resp.its transpose), then the Kronecker congruence relation $\mathbb{T}(p) \equiv \Pi + \Pi' (\mathrm{mod}\, p)$ leads to such a simple higher degree version as:

$$\mathbb{U}(p^m) \equiv \Pi^m + \Pi'^m \pmod p. \qquad (3)$$

This means in particular that the points of intersection of $\mathbb{U}(p^m) (\mathrm{mod}\ p)$ and $\Delta (\mathrm{mod}\ p)$ are in *two-to-one* correspondence with the $\mathbb{F}_{p^m}$-rational points of $X \pmod p$.

Now, these (i)!(ii)! apparently had inspired Sato with the following observation:

(iii)([note]? [**?**]!) *The summation in the formula for* $\mathrm{tr}(U_k(p^m))$ *is essentially parametrized by the* $\mathbb{F}_{p^m}$*-rational points of* $X (\mathrm{mod}\ p)$.

This can certainly be deduced *if* the divisor $\mathbb{U}(p^m)$ were effective and intersected with Δ properly. But it is not. In fact, take, say $m = 2$. Then the linear expression of the divisor $\mathbb{U}(p^2)$ by irreducible components reads as $\mathbb{T}(1,p^2) - (p-1)\Delta$. (Note that $\mathbb{T}(p,p) = \Delta$ as a divisor.) The sentense in [note] "$\mathrm{tr}(U_k(p^m))$ is a sum over the *fixed points of* $\mathbb{U}(p^m)$" only left careful readers in confusion. Thus, what should be remedied is not the [note] unapproved by Sato, but his intuition based on (i)!(ii)!. This later led Y.I. to a more delicate *pointwise* argument using deeper results of Deuring's to settle the proof of (iii) (cf. III-2, V below).

How to compactify and desingularize the (Kummer) Kuga variety, in order to reduce RPC to WC, seems to me a minor point, but I may have to add that this was totally untouched in his talks in Tokyo mentioned above.

3 Closely related results during 1964-69

(III-1) Kuga-Shimura [K-Sh](1965); cf.also [K1][K2]. In [K-Sh], the authors took up a quaternionic analogue of RPC and of the Kuga variety, and proved in this case that the former can be reduced to WC for the latter for almost all p.

Their argument was based on their congruence relation $\mathbb{T}(p) \equiv \Pi + \Pi'(\mathrm{mod}\ p)$ not only over the base curve but also on the fiber variety, and did not need to use $\mathbb{T}(p^m)$ for $m > 1$, thereby avoiding the difficulty mentioned above. Also in this case, the quaternions being division, the corresponding Kuga variety is already compact. Two basic citations in [K-Sh] of Sato's work are (i) concrete criticisms[2] in the footnote 4 in the Introduction and (ii) an acknowledgment "We are grateful to Mikio Sato for this important observation which inspired us to a great extent."

This track separated itself from Sato's track to use the trace formula, at "the first branch point", but it was *this* paper which inspired Serre the idea to use ℓ-adic method in order to make it applicable to the original elliptic modular case, which was then practised by Deligne (see III-3,4 below). Now, in II-3(vi), I wrote: *had* Sato known Deuring's work, then he would have completed the proof by himself. Here in quite a similar sense, I might add: *had* Kuga and Shimura some familiarity with the ℓ-adic cohomology theory, then they would have succeeded in settling the elliptic modular case too.

(III-2) [Ih](1966-67). In the process of his own study of modular curves in connection with congruence subgroups of $SL_2(\mathbb{Z}[1/p])$, Y.I. observed that Sato's arguments can be justified by using Deuring's basic work [Drg1]. He was at the IAS and at the end of his "virgin talk" in the members' seminar (Jan.,1966), added a few words on this observation. Then some members (A.Borel, A.Weil, R.Langlands, etc.) kindly suggested him strongly to write down an independent paper on this. This appeared as [Ih], with a subtitle "to validate M.Sato's identities". Two ideas are involved.

First is Sato's original idea, that each summand in the formula for $\mathrm{tr}(U_k(p^m))$ has a geometric interpretation related to the number of $\mathbb{F}_{p^m}$-rational points of the reduction mod p of Kuga varieties, and the second is that the points of modular curves[3] over finite fields can be described, in a down-to-earth way, group theoretically. The main theorem expresses, for the case of level $N = 1$, each Hecke polynomial as a product of powers of the congruence zeta functions of $i - 1$ dimensional Kuga varieties for $i = k$ and for smaller even integers $i \geq 2$. It did not reach the point to reduce RPC to WC, as the model was not compactified. This paper did not serve as a basis of [Del1], but instead had the following impact.

Each of the two ideas mentioned above was developed and used in a different language (adelic instead of $(\infty \times p)$-adic), notably by Langlands to the study of zeta functions of higher dimensional Shimura varieties (cf. [Lg] p.499, [Lg2] etc.)

(III-3) Serre [Ser][Ser2](1967/68). First, let us recall a conjecture of Serre [Ser], stated in the case of $N = 1, k = 12$ where the space of cusp forms S_k is

[2]to be precise, contrary to a criticism made there, the case of higher powers of p is in fact treated but not sufficiently, as explained in II-4 above.

[3]and Shimura curves as well, as shown in his later articles

one-dimensional spanned by

$$\Delta(\tau) = q\prod_{m=1}^{\infty}(1-q^m)^{24} = \sum_{n=1}^{\infty}\tau(n)q^n \qquad (q = e^{2\pi i\tau}),$$

and each Hecke operator $T_{12}(n)$ acts as a scalar multiplication of $\tau(n)$. RPC is equivalent to that for each prime p the quadratic polynomial $1-\tau(p)u+p^{11}u^2$ has imaginary roots.Serre conjectured that for any prime number ℓ there exists a continuous representation ρ_ℓ of the absolute Galois group over $\mathbb{Q}$ into $GL_2(\mathbb{Q}_\ell)$ which is unramified outside ℓ such that for any prime number $p \neq \ell$ and a Frobenius element F_p above p one has (the inclusion $\ni$ and) the equality

$$\mathbb{Z}[u] \ni \det(1-\rho_\ell(F_p)u) = 1-\tau(p)u+p^{11}u^2.$$

Thus, if the representation ρ_ℓ exists and is realized in the ℓ-adic cohomology group H^{11} of some complete smooth (Kuga) variety, then RPC can be reduced to WC (cf. IV-3).

Now, Serre had not just conjectured but had an idea how to prove it. His idea was a combination of (i) Kuga-Shimura [K-Sh], together with [E2],[Sh2], and (ii) ℓ-adic theorysed an ℓ-adic Galois module of $\mathbb{Q}_\ell$-rank as small as the desired $2\dim S_k$, to be constructed by using the modules of ℓ-power division points of the fiber elliptic curves and their tensor powers, on which the Hecke operators act via the Shimura isomorphism. In other words, if D_k is the $\mathbb{Z}$-module on which the Hecke operators act, constructed by Shimura described in II-3(i), it is $D_k \otimes \mathbb{Z}_\ell$. I understand that he gave a concrete series of suggestions in his letter initially addressed to J.-L.Verdier (Feb 11, 1967)[Ser2].

(III-4) Deligne (1968/69). Deligne [Del1] reduced RPC to WC. It seems to be basically along Serre's suggestions mentioned above, which means that it inherits those previous works of Kuga, Sato (indirectly), Shimura [Sh2], Kuga-Shimura [K-Sh], and more directly, of Serre [Ser2]. But he overcame various delicate points which, if easy for him, were probably not so for others.

In its Introduction, Deligne cites Sato's contribution as "l'idée fondamentale de Sato-Kuga-Shimura". On the other hand, he calls such previous works as to express Hecke polynomials by congruence zeta functions "first approximation", and sets the target of his own paper on the construction of a (by-Serre-conjectured) ℓ-adic representation of the Galois group as a subquotient of an ℓ-adic cohomology group of the correct dimension of a suitable Kuga variety. Also, the problem related to the difference between cohomology groups of the original Kuga variety and its smooth compactification is solved. An overall strong impression was the change of epoch caused by change of language in algebraic geometry. For the construction of ℓ-adic representations, the Grothendieck cohomology theory was the suitable one.

And a few years later, as mentioned above, Deligne settled the proof of the Weil conjecture itself, in [Del2I](for V: projective),[Del2II] (general).

4 Preparations for precise mathematical descriptions

(IV-1) *The Ramanujan-Petersson conjecture* RPC.

For simplicity, we assume that the level $N = 1$. Let S_k (k positive, even) denote the vector space of all modular cusp forms of weight k on $PSL_2(\mathbb{Z})$. It is equipped with the (positive Hermitian) Petersson metric. The Hecke operators $T_k(n)$ act on S_k as self-adjoint linear operators. They are mutually commutative and hence S_k is spanned by simultaneous eigenforms with *real* eigenvalues $\tau_\nu(n)$ ($1 \leq \nu \leq \dim\ S_k$). The RPC asserts that for each prime p, $|\tau_\nu(p)| \leq 2p^{\frac{k-1}{2}}$, or equivalently, that the roots of the quadratic polynomial $1 - \tau_\nu(p)u + p^{k-1}u^2$ are pairwise complex conjugate for all ν, or equivalently, that the roots of the Hecke polynomial

$$H_k^{(p)}(u) = \det(1 - T_k(p)u + p^{k-1}u^2) \tag{4}$$

are pairwise complex conjugate. The Hecke's Dirichlet series has the Euler product expansion

$$L_k(s) = \det(\sum_{n=1}^{\infty} T_k(n)n^{-s}) = \prod_p H_k^{(p)}(p^{-s})^{-1}. \tag{5}$$

The case $k = 12$, where $\dim S_{12} = 1$, corresponds to the original Ramanujan Conjecture.

(IV-2) *Eichler-Selberg trace formula* [E2][Sel].

$$tr(T_k(n)) = \sum_{\{\rho,\bar{\rho}\}} \sum_{\Theta} (-\frac{h_\Theta}{w_\Theta}) F_k(\rho, \bar{\rho}) + r_k(n), \tag{6}$$

where $\{\rho, \bar{\rho}\}$ runs over all unordered pairs of mutually conjugate irrational imaginary quadratic integers with norm n, and Θ runs over all *orders* of imaginary quadratic fields containing $\rho, \bar{\rho}$. Here, an order of a quadratic field K means a subring of the ring Θ_K of all integers of K of the form $\mathbb{Z} + f.\Theta_K$ with some natural number f called its conductor; w_Θ is the number of roots of unity in Θ; G_Θ is the group of *proper* (i.e., locally principal) Θ-ideal classes (cf. e.g.[Ih]); $h_\Theta = |G_\Theta|$ its cardinality,

$$F_k(X, Y) = (X^{k-1} - Y^{k-1})/(X - Y), \tag{7}$$

and finally, $r_k(n)$, in the case of our interest $k > 2$, is given as

$$r_k(n) = -\sum_{d|n, d \leq \sqrt{n}}{}' d^{k-1} + \delta(\sqrt{n}) \frac{k-1}{12} n^{k/2-1}, \tag{8}$$

where, in the first summand d^{k-1} should be replaced by $(1/2)d^{k-1}$ when $d = \sqrt{n}$, and $\delta(\sqrt{n})$ represents 1 (resp. 0) when $\sqrt{n}$ is rational (resp. irrational).

(IV-3) *The Weil conjecture* [W];[SGA4][SGA5][Del2I][Del2II].
Let V be a complete non-singular algebraic variety over a finite field $\mathbb{F}_q$, and $Z_V(u)$ be its congruence zeta function; namely, if $N_V(q^m)$ for each $m \geq 1$ denotes the number of $\mathbb{F}_{q^m}$-rational points of V, then

$$Z_V(u) = \exp(\sum_{m\geq 1} \frac{N_V(q^m)}{m} u^m) \tag{9}$$

as a formal power series of u. By B. Dwork, this is a rational function. Let $n = \dim V$, and let $H^i(\bar{V}, \mathbb{Q}_\ell)$ $(0 \leq i \leq 2n)$ be the i-th ℓ-adic cohomology group, where $\bar{V} = V \otimes \bar{\mathbb{F}}_q$ and ℓ is any prime number not equal to the characteristic p. The Frobenius element of V over $\mathbb{F}_q$ induces a linear automorphism $F_{V,i}$ of $H^i(\bar{V}, \mathbb{Q}_\ell)$. The WC proved by Deligne asserts that (i)

$$P_{V,i}(u) = \det(1 - uF_{V,i}) \tag{10}$$

is a polynomial *over* $\mathbb{Z}$ independent of ℓ; (ii)

$$Z_V(u) = \prod_{i=0}^{2n} P_{V,i}(u)^{(-1)^{i+1}}, \tag{11}$$

and (iii) if

$$P_{V,i}(u) = \prod_{\nu=1}^{dim(H^i)} (1 - \alpha_{i,\nu} u) \tag{12}$$

denotes the linear decomposition of $P_{V,i}(u)$ over $\mathbb{C}$, then $|\alpha_{i,\nu}| = q^{i/2}$ holds for all i and ν. Note that

$$N_V(q^m) = \sum_{i=0}^{2n} (-1)^i \sum_{\nu} \alpha_{i,\nu}^m. \tag{13}$$

(IV-4) *Elliptic curves over* $\bar{\mathbb{F}}_p$ (Hasse, Deuring [Drg1][Drg2]).

The isomorphism classes of elliptic curves E over $\bar{\mathbb{F}}_p$ are parametrized by their normalized j-invariants $j \in \bar{\mathbb{F}}_p$, $j \leftrightarrow E_j$ [Drg2]. And Deuring's theory [Drg1] gives a complete classification of E in terms of the endomorphism rings $\Theta = \mathrm{End}(E)$. The algebra $\Theta \otimes \mathbb{Q}$ is either an imaginary quadratic field K such that $(\frac{K}{p}) = 1$ (then E (or j) is called *ordinary*), or the definite quaternion algebra $B_{\infty,p}$ with discriminant p over $\mathbb{Q}$ (called *supersingular*).[4] Let (i) O_p^{ord} denote the set of all orders Θ of imaginary quadratic fields K such that $(\frac{K}{p}) = 1$ and that the conductor of Θ is not divisible by p, (ii) O_p^{ss} denote the set of isomorphism classes, or equivalently, the $B_{\infty,p}^\times$-conjugacy classes, of all maximal orders in $B_{\infty,p}$, and (iii) put $O_p = O_p^{ord} \cup O_p^{ss}$.

Then $j \mapsto \mathrm{End}(E_j)$ defines a surjective mapping $\bar{\mathbb{F}}_p \to O_p$ such that for each $\Theta \in O_p$, the preimage is a union of h_Θ/d_Θ distinct conjugacy classes over $\mathbb{F}_p$

[4] "singulär" *vs.* "supersingulär" in [Drg1].

of elements j of degree $d_\Theta = d_j$ over $\mathbb{F}_p$. Here, h_Θ, d_Θ are defined as follows. If $\Theta \in O_p^{ord}$, then $h_\Theta = |G_\Theta|$ as in IV-2 and d_Θ is the order of the element of G_Θ represented by $\mathfrak{p}_\Theta = \mathfrak{p} \cap \Theta$, $\mathfrak{p}$ being a prime factor of p in K. Note that $\mathfrak{p}_\Theta^{d_\Theta} = \pi_\Theta.\Theta$ holds with some $\pi_\Theta \in \Theta$, and that $\pi_\Theta \bar{\pi}_\Theta = p^{d_\Theta}$. When $\Theta \in O_p^{ss}$, we put $h_\Theta = d_\Theta = 1$ when Θ contains an element π_Θ with $\pi_\Theta \bar{\pi}_\Theta = p$, and $h_\Theta = d_\Theta = 2$ otherwise (we may then define $\pi_\Theta = \pm p$).

As before, put $w_\Theta = |\Theta^\times|$. Then $w_\Theta = 2$ (i.e., $\Theta^\times = \{\pm 1\}$) if and only if $j \neq 0, 12^3$, and in this case the pair $\{\pi_\Theta, \bar{\pi}_\Theta\}$ is determined up to the sign, and moreover, the pair of Frobenius eigenvalues $\{\pi_j, \bar{\pi}_j\}$ of any model of E_j over $\mathbb{F}_{p^{d_\Theta}}$ is one of $\pm\{\pi_\Theta, \bar{\pi}_\Theta\}$, with the sign depending on the choice of the model.

One may put all these into the following universal formula. For any $m \geq 1$ and any polynomial $F(X,Y) \in \mathbb{C}[X,Y]$ satisfying $F(X,Y) = F(Y,X) = F(-X,-Y)$, we have, for $p \neq 2, 3$,

$$\sum_{\substack{j \in \bar{\mathbb{F}}_p, \neq 0, 12^3 \\ d_j | m}} F(\pi_j^{m/d_j}, \bar{\pi}_j^{m/d_j}) = \sum_{\substack{\Theta \in O_p, w_\Theta = 2 \\ d_\Theta | m}} h_\Theta F(\pi_\Theta^{m/d_\Theta}, \bar{\pi}_\Theta^{m/d_\Theta}). \tag{14}$$

Two remarks (i) The well-known "Mass formula" reads as

$$\frac{p-1}{12} = \sum_{\Theta \in O_p^{ss}} \frac{2}{w_\Theta} = H_p - \frac{1}{4}\left(1 - (\frac{-1}{p})\right) - \frac{1}{3}\left(1 - (\frac{-3}{p})\right), \tag{15}$$

where

$$H_p = \sum_{\Theta \in O_p^{ss}} d_\Theta = \textit{the class number of } \mathrm{B}_{\infty,p} \tag{16}$$

is the number of supersingular j-invariants. If $\Theta \in O_p^{ss}$ and $w_\Theta = 2$, then $\{\pi_\Theta, \bar{\pi_\Theta}\} = \{\sqrt{-p}, -\sqrt{-p}\}$ for $d_\Theta = 1$, and $= \pm\{p, p\}$ for $d_\Theta = 2$.

(ii) A hint for generalizations (higher levels, Shimura curves). If one uses the corresponding arithmetic groups over $\mathbb{Z}[1/p]$ and their certain type of conjugacy classes, then the distinction between O_p^{ord} and O_p^{ss}, and the description of at least the former, can be understood in a unified and a functorial way [5].

5 Sato's arguments in a justified form

(V-1) We continue II-4, now with formulas. A prime number p will now be fixed and suppressed from the notations. Assume again $N = 1$. The p-part of the Hecke's Dirichlet series associated with the Hecke operators $T_k(n)$ acting on the space S_k of modular cusp forms of even weight k is

$$\det(\sum_{m \geq 0} T_k(p^m) p^{-ms}) = H_k(p^{-s})^{-1}, \tag{17}$$

[5] cf. the Introduction of [Ih] and, just for curious readers, perhaps also some later works of the same author on such arithmetic groups.

where

$$H_k(u) = \det(1 - T_k(p)u + p^{k-1}u^2) = \exp\{-\sum_{m\geq 1} \frac{\mathrm{tr}(U_k(p^m))}{m} u^m\}. \qquad (18)$$

Now from the trace formula (IV-2)(6) for $T_k(p^m)$ one can derive a simple formula for $\mathrm{tr}(U_k(p^m))$. An arrangement as a sum *over* $\Theta \in O_p$, instead of over imaginary quadratic integers, suggested by the work of Deuring (IV-4), makes it simple. Here, we assume $p \neq 2, 3$, which assures us that the unit group $\Theta^\times$, which can be non-commutative in general, is cyclic of orders $2, 4$ or 6.

$$-\mathrm{tr}(U_k(p^m)) = \sum_{\substack{\Theta \in O_p \\ d_\Theta | m}} F_{k,\Theta} + 1 - \delta_k. \qquad (19)$$

Here, $\delta_k = 0$ for $k > 2$, and $= p^m + 1$ for $k = 2$. Each summand $F_{k,\Theta}$ is given as follows.

$$\begin{array}{ll}
\dfrac{h_\Theta}{w_\Theta} \displaystyle\sum_{\zeta \in \Theta^\times} F_k(\zeta \pi_\Theta^{m/d_\Theta},\ \overline{\zeta \pi_\Theta^{m/d_\Theta}}) & \cdots \Theta \in O_p^{ord} \\
F_k((\sqrt{-p})^m, -(\sqrt{-p})^m) & \cdots \Theta \in O_p^{ss},\ m\text{:}\textit{odd}\ (\textit{hence}\ d_\Theta = 1) \\
\dfrac{1}{w_\Theta} \displaystyle\sum_{\zeta \in \Theta^\times} F_k(\zeta p^{m/2}, \bar{\zeta} p^{m/2}) & \cdots \Theta \in O_p^{ss},\ m\text{:}\textit{even},\ d_\Theta = 1 \\
2F_k(p^{m/2}, p^{m/2}) & \cdots \Theta \in O_p^{ss},\ m\text{:}\textit{even},\ d_\Theta = 2.
\end{array}$$

$F_k(X, Y)$ being as in (7)(IV-2). This is a (less down-to-earth) version of Lemma 6 of [Ih][6].

Remarks (i) The collection of all Θ-terms for $\Theta \in O_p^{ss}$, with $d_\Theta = 2$, arises from the remainder term $r_k(n)$ in the trace formula (6) for $T_k(n)$; thus

$$-(r_k(p^m) - p^{k-1} r_k(p^{m-2})) = \frac{p-1}{12} F_k(p^{m/2}, p^{m/2}) + 1. \qquad (20)$$

(ii) For each $\Theta \in O_p^{ord}$, the Θ-term "corresponds to" the *Deuring lifting* to characteristic 0 of an ordinary elliptic curve over $\bar{\mathbb{F}}_p$ *together with* its endomorphism ring. Among the points of the intersection $\mathbb{T}(p^m).\Delta$, what remain in (the ordinary part of) the trace formula for $U_k(p^m)$ are the images of the fixed points on the complex upper half plane of those $\gamma \in M_2(\mathbb{Z})$ with $\det(\gamma) = p^m$ that are not divisible by p in the stronger sense, i.e., not only that $p^{-1}\gamma \notin M_2(\mathbb{Z})$ but also that $p^{-1}\gamma \notin O_K$, the ring of integers of the quadratic field $K = \mathbb{Q}(\gamma)$. When the former $\notin$ is satisfied, the latter $\notin$ is equivalent to that the conductor of the order $K \cap M_2(\mathbb{Z})$ of K is not divisible by p.

(V-2) Now we shall connect the right side of (19) with the p-part of zeta functions of Kuga varieties over the j-line minus $\{0, 12^3\}$. For this purpose

[6] In the formulas (45)(45") following this lemma, $-H - \Delta$ should be read as $-(H - \Delta)$.

we treat the sum over those Θ with $w_\Theta > 2$, to be denoted by $\varepsilon_k(p^m)$, as an exceptional term and read (19) now as:

$$-\mathrm{tr}(U_k(p^m)) = \sum_{\substack{\Theta\in O_p, w_\Theta=2\\ d_\Theta|m}} h_\Theta F_k(\pi_\Theta^{m/d_\Theta},\ \overline{\pi_\Theta^{m/d_\Theta}}) + 1 - \delta_k + \varepsilon_k(p^m). \qquad (21)$$

Combining this with Deuring's (14), we obtain, with a full proof, the key equality:

$$-\mathrm{tr}(U_k(p^m)) = \sum_{\substack{j\in\bar{\mathbb{F}}_p, \neq 0,1 2^3\\ d_j|m}} F_k(\pi_j^{m/d_j}, \bar{\pi}_j^{m/d_j}) + 1 - \delta_k + \varepsilon_k(p^m). \qquad (22)$$

(V-3) Now for the zeta functions of Kuga varieties. Start with any choice of a smooth fiber variety $\mathbb{E}$ over the affine j-line *minus* $\{0, 12^3\}$, defined over $\mathbb{F}_p$, whose fiber above each point j is isomorphic to an elliptic curve E_j over $\mathbb{F}_p(j)$ with the absolute invariant j. For each $k \geq 2$, let $V_k = \mathbb{E}^{k-2}$ (the fiber product), and $V_k' = V_k/\{\pm 1\}$ (the fiberwise Kummer quotient). The latter is not smooth but instead, it is canonical, i.e., independent of the choice of $\mathbb{E}$. Now, the congruence zeta functions of V_k,V_k' over $\mathbb{F}_p$ are

$$Z_{V_k}(u) = \exp\{\sum_{m\geq 1} \frac{N_k(p^m)}{m} u^m\}, \qquad (23)$$

$$Z_{V_k'}(u) = \exp\{\sum_{m\geq 1} \frac{N_k'(p^m)}{m} u^m\}, \qquad (24)$$

where $N_k(p^m)$ resp. $N_k'(p^m)$ denote the number of $\mathbb{F}_{p^m}$-rational points of V_k resp. V_k' rewritten according to the base point parametrization, as

$$N_k(p^m) = \sum_{j\in\mathbb{F}_{p^m}, \neq 0,1 2^3} N_k(\pi_j^{m/d_j}, \bar{\pi_j}^{m/d_j}), \qquad (25)$$

$$N_k'(p^m) = \sum_{j\in\mathbb{F}_{p^m}, \neq 0,1 2^3} N_k'(\pi_j^{m/d_j}, \bar{\pi}_j^{m/d_j}), \qquad (26)$$

where $N_k(X,Y) = ((X-1)(Y-1))^{k-2}$ and $N_k'(X,Y)$ is its *even-degree* part. Note that XY will always be substituted by $(\pi_j\bar{\pi}_j)^{m/d_j} = p^m$.

(V-4) Sato expressed the main term of $Z_{V_k'}(u)$ as the product of powers of $H_\ell(u)$ for even ℓ with $2 \leq \ell \leq k$. By (18)(22)(26), the point is to express the main term of $N_k'(\pi_j^{m/d_j}, \bar{\pi}_j^{m/d_j})$ as a linear combination of $F_\ell(\pi_j^{m/d_j}, \bar{\pi}_j^{m/d_j})$. This is reduced to expressing the polynomial $N_k'(X,Y)$ by $F_\ell(X,Y)$'s:

$$N_k'(X,Y) = a_k.F_k(X,Y) + \sum_{\substack{2\leq\ell<k\\ \ell:even}} B_{k,\ell}(XY)F_\ell(X,Y), \qquad (27)$$

where $a_k = 1$ (resp. 0) for k: even (resp. odd), and

$$B_{k,\ell}(W) = \sum_{0 \le \mu \le k-\ell} b_{k,\ell,\mu} W^\mu \in \mathbb{Z}[W]. \tag{28}$$

Combining these we obtained the proof of Sato's assertion that "the main part" of $Z_{V_k'}(u)$ is given as:

$$H_k(u)^{a_k} \times \prod_{\substack{2 \le \ell < k \\ \ell: even}} \prod_{0 \le \mu \le k-\ell} H_\ell(p^\mu u)^{b_{k,\ell,\mu}}. \tag{29}$$

This contributed to the reduction of RPC to WC (modulo smooth compactifications). What Sato wrote down actually at the end of each of his talks was the corresponding global identity between the zeta and the L-functions, where μ contributes to shifting s to $s-\mu$. Probably, he wished to stress, no less than the possibility of the reduction of RPC to WC, an important consequence on the other direction that the expected analytic properties of the Hasse zeta function of Kuga varieties can be derived from Hecke theory through his argument.

Finally, to see the other direction, i.e., to express each Hecke polynomial of weight k as a product of powers of congruence zeta functions of V_ℓ's, please refer to [Ih]. For this direction, a passage to the Kummer quotient is unnecessary.

(References)

[note] A note from M. Sato's talk "Weil's conjecture and Ramanujan's conjecture", Sugaku-no-Ayumi 10-2 (1963) (in Japanese).

[K1] M. Kuga, "Fibre varieties over a symmetric space whose fibres are abelian varieties", Lect. Notes, Univ. of Chicago, Chicago, Ill. U.S.A.1963/64; Proc. Symp. Pure Math. vol. 9, AMS (1966), 338-346.

[K2] ———, "Hecke's polynomial as a generalized congruence Artin L-function", Proc. Symp. Pure Math. vol. 9, AMS (1966), 333-337.

[K-Sh] M. Kuga and G. Shimura, "On the zeta function of a fibre variety whose fibres are abelian varieties", Ann. of Math., 82 (1965), 478-539.

[Ih] Y. Ihara, "Hecke polynomials as congruence ζ functions in elliptic modular case (To validate M.Sato's identity)", Ann. of Math., 85 (1967), 267-295.

[Ser] J-P. Serre, "Une interprétation des congruences relatives à la fonction τ de Ramanujan", Sém Delange-Pisot-Poitou, 1967/68, no. 14.

[Ser2] J-P. Serre, A letter to J.-L.Verdier (Feb.11,1967), in: Correspondance Serre-Tate Vol II, Appendice, 909-911; eds. P. Colmez, J-P. Serre, Soc.math. de France, 2015

[Del1] P. Deligne, "Formes modulaires et représentations ℓ-adiques", Sém. Bourbaki 21e année 1968/69, no 355.

[Del2I] ———, "La conjecture de Weil I", Publ. math. IHES 43 (1974), 273-307.

[Del2II] ———, "La conjecture de Weil II", Publ. math. IHES 52 (1980), 137-252.

[Drg1] M. Deuring, "Die Typen der Multiplikatorenringe elliptischer Funktionenkörper" Abh. Math. Sem. Hamburg 14(1941),197-272.

[Drg2] M. Deuring, "Invarianten und Normalformen elliptischer Funktionenkörper" Math. Z. 47 (1941), 47-56.

[E1] M. Eichler, "Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion", Arch. Math. 5 (1954), 355-366.

[E2] ———, "Eine Verallgemeinerung der Abelschen Integrale", Math. Z. 67 (1957), 267-298.

[Ig] J-I. Igusa, "Kroneckerian model of fields of elliptic modular functions, Amer. J. Math. 81 (1959), 561-577.

[Lg] R.P. Langlands, "Modular forms and ℓ-adic representations", in: "Modular functions of one variable II" Lecture Notes in Math. 349 (1972), 361-500, Springer.

[Lg2] R.P. Langlands, "Automorphic representations, Shimura varieties, and motives. Ein Märchen", Proc. Symp. Pure Math. 33 (1979), part 2, 205-246.

[Sel] A. Selberg, "Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series", J. Indian Math. Soc. 20 (1956), 47-87.

[SGA4] M. Artin, A. Grothendieck and J-L.Verdier eds. "Théorie des topos et Cohomologie étale des schémas", Sém. de géométrie algébrique du Bois Marie 1963/64 ("SGA 4"), Lecture Notes in Math. 269, 270, 305; Springer (1972)

[SGA5] A. Grothendieck, "Cohomologie ℓ-adique et fonctions L" ("SGA 5"). Lecture Notes in Math. 589, Springer (1977).

[Sh0] G. Shimura, "Introduction to the arithmetic theory of automorphic functions", Publ. Math. Soc. Japan 11, Iwanami Shoten and Princeton Univ. Press (1971).

[Sh1] ——— "Correspondances modulaires et les fonctions ζ de courbes algébriques ", J. Math. Soc. Japan 10 (1958), 1-28.

[Sh2] ——— "Sur les intégrales attachées aux formes automorphes", J. Math. Soc. Japan 11(1959), 291-311.

[W] A. Weil, "Number of solutions of equations in finite fields", Bull. AMS 55 (1949), 497-508.