ARITHMETIC MODULI OF ELLIPTIC CURVES

BY

NICHOLAS M. KATZ

AND

BARRY MAZUR

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

1985

INTRODUCTION

This book is devoted to giving an account of the arithmetic theory of the moduli spaces of elliptic curves. The main emphasis is on understanding the behavior of these moduli spaces at primes dividing the "level" of the moduli problem being considered. Until recently, this seemed a very difficult problem, because one had no a priori construction of these spaces at the "bad" primes. One defined them as schemes over, say, $\mathbf{Z}[1/N]$ as the solution to some well-posed moduli problem which *only made sense* for elliptic curves over rings in which N was *invertible*, and then one used a process of *normalization* to extend them to schemes over $\mathbf{Z}$, e.g., one took the "proj" of the graded subring of the ring of all modular forms of the type in question consisting of those with *integral* Fourier ("q-expansion") coefficients at the cusps. This procedure produced a scheme over $\mathbf{Z}$, but one had no idea of what moduli interpretation this scheme had, nor a fortiori did one have any idea of the modular interpretation of its reduction modulo p, for p a prime dividing the level.

Historically, the only case where this question was in any way satisfactorily understood was the case of $\Gamma_0(p)$, which made universal sense as the moduli problem "p-isogenies", or "finite flat subgroup-schemes of rank p." This modular interpretation was implicitly known to Kronecker, for whom it appears as the statement that the "modular equation of degree p, reduced mod p, is the curve in the (j_1,j_2)-plane

$$(j_1-(j_2)^p)((j_1)^p-j_2) = 0\ .''$$

One knows the crucial role that the reinterpretation of this Kronecker congruence by Eichler-Shimura as the "congruence relation"

$$T_p = F + V \mod p$$

played in the reduction of the Ramanujan conjecture to Weil's "Riemann Hypothesis for varieties over finite fields."

Eichler-Shimura made use of this relation to prove that for all but finitely many p the Ramanujan conjecture held for any given cusp form of weight two and level N which is a simultaneous eigenfunction of all T_p with p not dividing N. The method was intrinsically incapable of *specifying* the exceptional p, which were *believed* to consist only of primes dividing N.

Partly in order to settle definitively this question of exceptional p for weight two forms, Igusa, in a brilliant series of papers ([Ig 2, 3, 4, 5]), gave a complete and definitive account of the level N moduli scheme over $Z[1/N]$. Except for a difference of mathematical language, and the modular interpretation of the cusps by Deligne-Rapoport, there have been no "improvements" to Igusa's account of what happens over $Z[1/N]$. Although Igusa's papers contain many stimulating speculations about the situation mod p for "bad" p (e.g., the footnote ([Ig 2], p. 472) where he points out that the genus of $\Gamma_0(p)$ is closely related to the number of [super] singular points in characteristic p), there was to be no significant progress in understanding the situation at "bad" p for another decade.

In 1968, Deligne completed the general reduction of Ramanujan's conjecture, for forms of arbitrary weight, and in particular for Δ, to Weil's Riemann Hypothesis for varieties over finite fields. In his article [De 1], he mentions that in fact the $\Gamma_0(p)$ moduli scheme, (with suitable auxiliary prime-to-p rigidification) is actually a *regular* scheme, and in a letter of July 10, 1970 to Parshin he proves this regularity by checking what happens at the supersingular points in characteristic p.

Simultaneously, another theme was developing. Shimura conjectured and Casselman [Cmn 1] proved that for $p = 29, 53, 61, 73, 89, 97$, the Jacobian of (the modular curve for) $\Gamma_1(p)$, modulo the Jacobian of $\Gamma_0(p)$, acquired good reduction over the field $Q(\zeta_p)$. Casselman [Cmn 2]

explained that such theorems of good reduction could be predicted by the "Langlands philosophy", which related, conjecturally, such questions to questions in representation theory which were already well-understood.

The paper of Deligne-Rapoport [De-Ra] in 1972 provided an exhaustive account of what was then known about arithmetic moduli of elliptic curves. It gave a complete account of level N moduli problems over $\mathbf{Z}[1/N]$, including a modular interpretation of the compactified moduli scheme (i.e., including the cusps) as the moduli space of "generalized elliptic curves with auxiliary structure." It also gave a modular interpretation over $\mathbf{Z}$ to the $\Gamma_1(p)$ moduli problem, and with it a proof that the good reduction phenomenon of Shimura-Casselman held for any p. Another innovation was the systematic use of algebraic stacks, as developed by Mumford [Mum 1] and Deligne-Mumford [De-Mu].

The next significant progress came in 1974, with Drinfeld's introduction (in the context of his theory of "elliptic modules") of the notion of a "full level N-structure" on an elliptic curve over an arbitrary scheme, where N need not be invertible, as a pair of points P,Q of order N such that the group-scheme $E[N]$ of points of order N is equal to the sum

$$\sum_{a,b \bmod N} [aP+bQ]$$

as a Cartier divisor inside E. Drinfeld showed that with this definition, the corresponding full level N moduli problem for his "elliptic modules" was regular. It was clear to the experts, although never published, that with Drinfeld's definition applied to usual elliptic curves, one obtained a moduli problem over $\mathbf{Z}$ which was regular, and which, over $\mathbf{Z}[1/N]$, coincided with the usual "full level N" moduli problem. In particular, one now had a *modular* interpretation of its reduction modulo any p, as the moduli space of elliptic curves, together with Drinfeld level N structures, over $\mathbf{F}_p$-algebras. With this modular interpretation, it became a pleasant exercise to calculate explicitly the reduction modulo any prime p.

In a letter to Drinfeld of January 21, 1975, Deligne explained how the Drinfeld idea of using Cartier divisors allowed one to define universally the $\Gamma_1(N)$ problem as well as a "balanced" version of it by saying that a point P in an elliptic curve has "exact order N" if it is killed by N and if the Cartier divisor inside E defined by

$$\sum_{a \bmod N} [aP]$$

is actually a *subgroup*-scheme of E. Deligne also explained that the resulting moduli problem was regular.

In June 1979, the present authors rediscovered Deligne's $\Gamma_1(N)$ idea, and they formulated a Drinfeldian version of $\Gamma_0(N)$ by defining a finite locally free subgroup-scheme G of rank N inside an elliptic curve to be cyclic if locally f.p.p.f. on the base one could find a point P in it which generated it, in the sense that

$$G = \sum_{a \bmod N} [aP]$$

as Cartier divisors inside E. Using this definition of $\Gamma_0(N)$ as the moduli problem of "elliptic curves together with cyclic subgroup-schemes which are finite locally free of rank N," they proved that the $\Gamma_0(N)$ problem was regular and worked out explicitly the reductions mod p of all the "standard" moduli problems.

These calculations of special fibers, together with some intricate (due to wild ramification) calculations of the topological invariants of the special fibers, led to a direct geometric verification of a rather general theorem of good reduction which includes the Shimura-Casselman-Deligne-Rapoport theorem as a special case. For the most part, this good reduction theorem is also a consequence of the above-mentioned Langlands philosophy, which reduces it to a known question in representation theory (cf. [La] and [M W], proof of Prop. 2, §2, Chapter 3).

In writing this book, we have tried simultaneously to be self-contained and to be as general as possible.

In the first chapter, we develop the Drinfeldian notions of level structure through the notion of equality of Cartier divisors in the ambient elliptic curve. With an eye to future applications to the moduli of higher-dimensional abelian varieties, in which the points of order N cease to be a Cartier divisor, we give a reformulation of all the Drinfeldian notions in the context of finite locally free commutative group-schemes, *without* reference to any ambient space. This reformulation, and the questions it raises, may prove to be of some independent interest.

This chapter is followed by a short "Review of Elliptic Curves", in which we recall all the basic facts we will use about elliptic curves. We give either complete proofs or precise references for all of these facts.

In Chapter 3, we apply the general notions developed in the first chapter to the special case of elliptic curves, and we formulate in terms of them the basic moduli problems for elliptic curves.

In Chapter 4, we develop a rudimentary formalism for speaking about these moduli problems, which amounts to working systematically with stacks without ever saying so. We speak rather of "relatively representable moduli problems", a notion which seems admirably suited to our purposes, and which is a throw-back to Mumford's original exposition [Mum 1].

After these preliminary chapters, we turn to the detailed study of the basic moduli problems *as* "open arithmetic surfaces." The basic results on their structure and inter-relations (e.g., which are regular, which are finite flat over which others, which are quotients by finite groups of which others,$\cdots$) are given in Chapters 5, 6 and 7.

The remaining 7 chapters are devoted to the detailed study of these same moduli problems *as* "curves over $\mathrm{Spec}(\mathbf{Z})$."

In Chapter 8, we compactify our moduli problems, relative to $\mathrm{Spec}(\mathbf{Z})$, by adding the cusps. In Chapter 9, we explain how to deal systematically with these moduli problems which are "really" defined over cyclotomic integer rings rather than over $\mathbf{Z}$. In Chapter 10 we give the basic result on the structure of our compactified moduli problems *as* relative curves.

After a brief digression concerning "exotic" isomorphisms of moduli problems in Chapter 11, the remaining three chapters are devoted to the detailed study of the degeneration at bad primes of our moduli problems as relative curves. Chapter 12 gives the theory of the Igusa curves, which are the "basic" p-power level moduli problems in characteristic p. In Chapter 13, we give the detailed structure of the reduction mod p of each of our basic moduli problems as a "disjoint union, with crossings at the supersingular points", of suitable Igusa curves.

In Chapter 14, we apply the specific calculations of the previous chapter to prove a general theorem of good reduction for suitable "pieces" of Jacobians of modular curves.

We would like to thank the IHES for providing the congenial atmosphere in which this book was written. We warmly thank Ofer Gabber, whose innumerable comments and corrections were invaluable to us in preparing the final version of this work. Lauri Hein and Perry Di Verita of Princeton University, and Helen Mann of Princeton University Press prepared the original and final manuscripts respectively. To them and to our editor, Barbara Stump of Princeton University Press, we extend our thanks for their patience in the face of our numerous and unexpected revisions.

NICHOLAS M. KATZ
BARRY MAZUR

TABLE OF CONTENTS

ARITHMETIC MODULI OF ELLIPTIC CURVES

BY

NICHOLAS M. KATZ

AND

BARRY MAZUR

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

1985

Annals of Mathematics Studies

Number 108

The Annals of Mathematics Studies are edited by
William Browder, Robert P. Langlands, John Milnor, and Elias M. Stein
Corresponding editors:
Stefan Hildebrandt, H. Blaine Lawson, Louis Nirenberg, and David Vogan

Clothbound editions of Princeton University Press books are printed on acid-free paper, and binding materials are chosen for strength and durability. Paperbacks, while satisfactory for personal collections, are not usually suitable for library rebinding

ISBN 0-691-08349-5 (cloth)
ISBN 0-691-08352-5 (paper)

Printed in the United States of America
by Princeton University Press, 41 William Street
Princeton, New Jersey

Library of Congress Cataloging in Publication data will
be found on the last printed page of this book

Chapter 1

GENERALITIES ON "A-STRUCTURES" AND "A-GENERATORS"

(1.1) *Review of relative Cartier divisors* (Compare [Mum 2], pp. 61-73.)

(1.1.1) Let S be an arbitrary scheme, and let X be an S-scheme. By an effective Cartier divisor D in X/S we mean a closed subscheme $D \subset X$ such that

$$\begin{cases} D \text{ is flat over } S \\ \text{the ideal sheaf } I(D) \subset \mathcal{O}_X \text{ is an invertible } \mathcal{O}_X\text{-module, i.e.,} \\ \text{it is a locally free } \mathcal{O}_X\text{-module of rank one.} \end{cases}$$

This notion is local on S. When S is affine, say $S = \mathrm{Spec}\,(R)$, it means that we can cover X by affine opens $U_i = \mathrm{Spec}\,(A_i)$, A_i an R-algebra, such that $D \cap U_i$ is defined in U_i by one equation $f_i = 0$, where $f_i \in A_i$ is an element such that

$$\begin{cases} A_i/f_iA_i \text{ is flat over } R \\ f_i \text{ is not a zero-divisor in } A_i\,. \end{cases}$$

The tautological exact sequence on X.

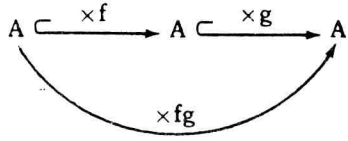
$$0 \to I(D) \to \mathcal{O}_X \to \mathcal{O}_D \to 0\,,$$

becomes on $U_i = \mathrm{Spec}\,(A_i)$ the exact sequence

$$0 \longrightarrow A_i \xrightarrow{\times f_i} A_i \longrightarrow A_i/f_iA_i \longrightarrow 0\,.$$

(1.1.2) Given two effective Cartier divisors D and D' in X/S, their sum $D+D'$ is the effective Cartier divisor in X/S defined locally by the

product of the defining equations of D and D′. Explicitly, if $S = \mathrm{Spec}(R)$ and if on an affine open $\mathrm{Spec}(A)$ of X, D and D′ are defined respectively by equations $f = 0$ and $g = 0$, then $D + D'$ is defined there by $fg = 0$. To check that fg is not a zero-divisor in A, one notes the commutative diagram



To check that A/fgA is flat over R, one notes the short exact sequence

$$0 \longrightarrow A/gA \xrightarrow{\times f} A/fgA \longrightarrow A/fA \longrightarrow 0\,,$$

which exhibits A/fgA as an extension of flat R-modules.

(1.1.3) Given an effective Cartier divisor D in X/S, we may speak of the inverse (as invertible $\mathcal{O}_X$-module) $I^{-1}(D)$ of its ideal sheaf. We have a tautological exact sequence

$$0 \to \mathcal{O}_X \to I^{-1}(D) \to \mathcal{O}_D \underset{\mathcal{O}_X}{\otimes} I^{-1}(D) \to 0\,.$$

The inclusion of $\mathcal{O}_X$ in $I^{-1}(D)$ allows us to view the constant function "1" as a global section of $I^{-1}(D)$, and we may recover D as the scheme of zeroes of this global section of $I^{-1}(D)$.

Conversely, suppose we are given a pair $(\mathcal{L}, \ell)$ consisting of an invertible $\mathcal{O}_X$-module $\mathcal{L}$ on X together with a global section $\ell \in H^0(X, \mathcal{L})$ which sits in a short exact sequence of $\mathcal{O}_X$-modules

$$0 \longrightarrow \mathcal{O}_X \xrightarrow{\times \ell} \mathcal{L} \longrightarrow \mathcal{L}/\mathcal{O}_X \longrightarrow 0$$

with $\mathcal{L}/\mathcal{O}_X$ *flat* over S. Then the scheme of zeroes of the section ℓ of $\mathcal{L}$ is easily seen to be an effective Cartier divisor D in X/S, and there is a unique isomorphism of $(\mathcal{L}, \ell)$ with $(I^{-1}(D), \text{"1"})$.

This construction allows us to interpret effective Cartier divisors in X/S as isomorphism classes of pairs $(\mathcal{L}, \ell)$ as above. From this point of view, the operation "sum of effective Cartier divisors in X/R" is none other than the operation of tensor product:

$$(\mathcal{L}, \ell) + (\mathcal{L}', \ell') = (\mathcal{L} \otimes \mathcal{L}', \ell \otimes \ell') .$$

The zero element for this addition is the pair $(\mathcal{O}_X, 1)$, corresponding to the *empty* Cartier divisor.

(1.1.4) There are two natural situations in which one can define the inverse image of a relative Cartier divisor. First let

$$T \to S$$

be an arbitrary morphism of schemes. Then for any effective Cartier divisor D in X/S, say represented by a pair $(\mathcal{L}, \ell)$, the closed subscheme $D_T \overset{\text{dfn}}{=\!=} D \underset{S}{\times} T$ of $X_T = X \underset{S}{\times} T$ is an effective Cartier divisor in X_T/T, represented by the pair $(\mathcal{L}_T, \ell_T)$ on X_T. To see this, it suffices to treat the case when $S = \mathrm{Spec}(R)$ and $T = \mathrm{Spec}(R')$ are both affine; then the sequence on $X_T = X \otimes R'$

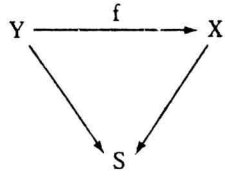
$$\mathcal{O}_X \underset{R}{\otimes} R' \xrightarrow{\ell \otimes 1} \mathcal{L} \otimes R' \longrightarrow \mathcal{L} \otimes R'/\mathcal{O}_X \underset{R}{\otimes} R'$$

is obtained from the short exact sequence on X

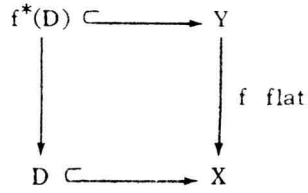
$$0 \longrightarrow \mathcal{O}_X \xrightarrow{\ell} \mathcal{L} \longrightarrow \mathcal{L}/\mathcal{O}_X \longrightarrow 0$$

by applying the functor $\underset{R}{\otimes} R'$. Because $\mathcal{L}/\mathcal{O}_X$ is assumed flat over R, this sequence stays short exact after $\underset{R}{\otimes} R'$, and its last term is R'-flat. Therefore $(\mathcal{L} \underset{R}{\otimes} R', \ell \otimes 1)$ defines an effective Cartier divisor in $X \underset{R}{\otimes} R'/R'$ as required.

Second, let



be a flat morphism of S-schemes. Then any effective Cartier divisor D in X/S gives rise to an effective Cartier divisor $f^*(D)$ in Y/S. Indeed, the cartesian diagram



shows that $f^*(D)$ is flat over D, and hence, D being S-flat, that $f^*(D)$ is flat over S. To see that the ideal sheaf $I(f^*(D))$ is an invertible $\mathcal{O}_Y$-module, we remark that this ideal sheaf is none other than $f^*(I(D))$, as follows from the fact that the short exact sequence on X

$$0 \to I(D) \to \mathcal{O}_X \to \mathcal{O}_D \to 0$$

remains short exact after application of the functor f^*, thanks to the flatness of f.

(1.1.5) We now turn to the question of recognizing which closed subschemes of X are in fact effective Cartier divisors in X/S.

PROPOSITION 1.1.5.1. *Suppose that* S *is locally noetherian, and that* X *is an* S*-scheme of finite type which is flat over* S. *Let* $\mathcal{F}$ *be a coherent sheaf on* X *which is flat over* S. *Then the necessary and sufficient condition that* $\mathcal{F}$ *be flat over* $\mathcal{O}_X$ *is that for every geometric point of* S, *i.e., every morphism* Spec(k) → S *with* k *an algebraically*