Modular Curves and Coding Theory: A Survey

Wen-Ching W. Li

Abstract. In this survey article we explain the role modular curves played in the theory of error-correcting codes. The emphasis is on Elkies’ modularity conjecture which predicts that all asymptotically optimal recursively defined towers of curves over finite fields with square cardinality arise from reductions of modular curves. The known examples support this conjecture. We discuss in detail key properties of the modular towers concerning this conjecture.

1. Introduction

In the past two decades, tremendous progress has been made in the theory of modular forms, including the proof of the Taniyama-Shimura-Weil modularity conjecture, which was a key ingredient in the proof of the Fermat’s Last Theorem, and the Sato-Tate conjecture. At the same time, novel applications of the theory of modular forms have been found. One such example is the explicit construction of Ramanujan graphs by Margulis [Mag88] and, independently, Lubotzky-Phillips-Sarnak [LPS88]. They are optimal expanders which have a wide variety of applications in computer science. These graphs are sparse and highly connected at the same time.

In coding theory, good codes should have high information rate and also high error-correction rate. Like expanders, these are two mutually repelling features. Many good codes are known to be algebraic geometry codes. When the base field \( F_q \) has square cardinality, asymptotically optimal recursively defined towers of curves (or function fields) were explicitly constructed by Garcia-Stichtenoth and company [GS95, GS96a, GS96b, GSR03, GST97]. By properly choosing divisors, these curves give rise to explicitly constructed good codes with arbitrarily long length. Elkies observed that the optimal towers constructed by Garcia and Stichtenoth all arose from reductions of elliptic modular curves, Shimura modular curves, or Drinfeld modular curves (cf. [El97, El00]). Based on these examples, he predicted that this should hold in general, known as the Elkies’ modularity conjecture. (See §6 for details.) The work of Li-Maharaj-Stichtenoth [LMS02] provided numerical evidence to Elkies’ conjecture. This conjecture, if established, shows another instance...
that optimal objects possessing simultaneously repelling features arise from modular forms/modular curves. It confirms once more the general belief that number theory provides a rich source for constructing optimal objects.

The purpose of this survey is to explain the connections between good codes and modular curves. This paper is arranged as follows. Basic concepts in error-correcting codes are reviewed in §2. In §3 we introduce the Goppa codes and algebraic geometry (AG) bound, and also discuss recent improvements on the AG bound. The main ingredient of the AG bound is the quantity $A(q)$ introduced by Ihara [Iha81]. The value of $A(q)$ is known only for square $q$. In §4 we review various upper and lower bounds for $A(q)$. Asymptotically optimal recursively defined towers are the focus of §5. Explicit towers over $\mathbb{F}_q$ with $q$ a square or a cube are recalled. In addition to introducing Elkies’ conjecture, the purpose of §6 is to explain why reductions of elliptic modular curves give rise to asymptotically optimal and recursively defined towers. In the final section, §7, we draw connections with solutions of Picard-Fuchs differential equations attached to modular curves, following the work of Maier [Mar07]. In particular, we show how a recursive defining equation and the splitting set of (the reduction of) an elliptic modular tower can be read off from the functional equation obtained from the normalized analytic solutions of the normalized weak-lifts of the Picard-Fuchs differential equation of the bottom curve along two different projections from the second curve. This viewpoint provides a more transparent and better understanding of the modular towers.

2. Error-correcting codes

Let $q$ be a prime power. A linear code $C$ of length $n$ in $q$ symbols is a subspace of $\mathbb{F}_q^n$. The information rate of $C$ is

$$r(C) = \frac{\dim C}{n}.$$ 

If $C$ is merely a subset of $\mathbb{F}_q^n$, we call it a nonlinear code of length $n$. Its information rate is defined in the same way with $\dim C$ replaced by $\log_q |C|$.

The distance between two elements in $\mathbb{F}_q^n$ is the number of components in which they differ. The minimum distance $d$ of a code $C$ is the shortest distance between two distinct codewords. Thus the balls in $\mathbb{F}_q^n$ centered at the codewords of $C$ with radius $\lfloor (d - 1)/2 \rfloor$ are mutually disjoint. A received word falling in the region covered by these balls is closest to a unique codeword; when the error probability is small, we assume this received word was sent from $c$, and hence we decode it as $c$ and thereby correct the errors occurred during the transmission. Because of this, we say that $C$ corrects at least $\lfloor (d - 1)/2 \rfloor$ errors. The error correction rate of $C$ is defined to be

$$\delta(C) = \frac{d}{n}.$$ 

By definition, both $r$ and $\delta$ lie in the interval $[0, 1]$. A code is considered good if it has high information rate $r$ and at the same time large error-correction rate $\delta$. Notice that $r$ and $\delta$ are mutually conflicting measurements, since this means choosing many vectors far apart from each other in a fixed vector space. To further clarify their relationship, given $\delta \in [0, 1]$, let

$$\alpha_q(\delta) = \sup_C \{ r(C) \mid \delta(C) \geq \delta \}.$$
A good code with error-correction rate $\delta$ should have its information rate $r$ as close to $\alpha_q(\delta)$ as possible.

It follows from definition that $\alpha_q$ is a nonincreasing function in $\delta$. Manin [Man81] showed that $\alpha_q$ is continuous. It is easy to see that $\alpha_q(0) = 1$ and $\alpha_q(1 - \frac{1}{q}) = 0$ (and hence $\alpha_q(\delta) = 0$ for $\delta \in [1 - \frac{1}{q}, 1]$). Unfortunately, no other value of $\alpha_q(\delta)$ is known. It satisfies the linear Plotkin upper bound

$$\alpha_q(\delta) \leq 1 - \frac{q}{q-1} \delta.$$ 

The celebrated Gilbert-Varshamov lower bound says

$$\alpha_q(\delta) \geq 1 - H_q(\delta),$$

where $H_q(x) = \frac{x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x)}{q}$.

This remained the best known lower bound for a long time until Goppa introduced algebraic geometry codes [Gpa81] which gave rise to a better lower bound, called algebraic geometry (AG) bound, for square $q \geq 49$ and $\delta$ in a certain range. The AG bound was improved further in the 21st century. In the next section we briefly recall this development.

3. Goppa codes and algebraic geometry bounds

3.1. Goppa codes and AG bounds. Let $V$ be an irreducible smooth projective curve defined over $\mathbb{F}_q$ with genus $g$. Let $P_1, ..., P_n$ be $\mathbb{F}_q$-rational points on $V$. Choose a divisor $G$ disjoint from the $n$ rational points. One has the Riemann-Roch space

$$\mathcal{L}(G) = \{ f : \text{div}(f) + G \geq 0 \} \cup \{0\}.$$

Goppa introduced a linear algebraic geometry code $C$ of length $n$ over $\mathbb{F}_q$ by evaluating the $\mathbb{F}_q$-rational functions $f$ in $\mathcal{L}(G)$ at the $\mathbb{F}_q$-rational points $P_1, ..., P_n$:

$$C = \{(f(P_1), ..., f(P_n)) : f \in \mathcal{L}(G)\}.$$ 

Since these rational points are away from the poles of $f$, each component $f(P_i)$ is an element in $\mathbb{F}_q$. Moreover, the assignment $f \mapsto (f(P_1), ..., f(P_n))$ is a linear map from $\mathcal{L}(G)$ to $C$. When $\deg G < n$, this map is injective, hence the dimension $k$ of $C$ is equal to $\dim \mathcal{L}(G)$, which is $\geq \deg G - g + 1$ by the Riemann-Roch theorem, and the minimum distance $d$ of $C$ is $\geq n - \deg G$ since a rational function has the same number of zeros as poles in an algebraic closure of $\mathbb{F}_q$. This shows that the information rate $r(C) = \frac{k}{n}$ and the error-correcting rate $\delta(C) = \frac{d}{n}$ satisfy the relation

$$r(C) + \delta(C) \geq 1 + \frac{1}{n} - \frac{1}{g}.$$ 

To describe the best lower bound, denote by $N_q(g)$ the maximum number of rational points among all smooth irreducible genus $g$ curves defined over $\mathbb{F}_q$. Ihara [Iha81] introduced the quantity

$$A(q) = \limsup_{g \to \infty} \frac{N_q(g)}{g}.$$
Taking the limit as $n$ tends to infinity yields the algebraic geometry (AG) bound for the function $\alpha_q$:

\[ \alpha_q(\delta) + \delta \geq 1 - \frac{1}{A(q)}. \]  

(3.2)

3.2. Improved AG bounds. In 2001 Xing [Xn01, Xn03] constructed non-linear subcodes of AG codes by exploiting derivatives of the rational functions at the $F_q$-rational points to obtain improvements of the above AG bound. His work has stimulated a flurry of activities along the same vein. They include the works by Elkies [El01, El03], Niederreiter-Özbudak [NO04, NO07a, NO07b], Stichtenoth-Xing [SX05], Maharaj [Mh07a] and Yang-Qi [YQ09]. The main idea is to properly choose the divisor $G$ to construct an AG code first and then strategically select a subset of this AG code to trade size for increased minimum distance. The detailed statements are often technical. Generally speaking, the improvements are around

\[ \alpha_q(\delta) + \delta \geq 1 - \frac{1}{A(q)} + \log_q(1 + \frac{1}{q^3}) \]

with the best lower bound given in [YQ09] of the form

\[ \alpha_q(\delta) + \delta \geq 1 - \frac{1}{A(q)} + \log_q(1 + \frac{2}{q^3}) + \log_q(1 + \frac{q - 1}{q^6}) \]

for certain values of $q$ and $\delta$ in some interval. The paper [Li07] gives more detail on the construction of nonlinear codes improving AG bound. See also the paper by Niederreiter [Ni] in this volume and the references therein.

4. Upper and lower bounds for $A(q)$

To compare the Gilbert-Varshamov bound with the AG bound, we need an estimate of $A(q)$. The first bound

\[ A(q) \geq \sqrt{q} - 1 \]

was proved in 1981 by Ihara [Iha81] by considering the reduction of Shimura curves, and independently in 1982 by Tsfasman-Vladut-Zink [TVZ82] by considering the reduction of elliptic modular curves. In 1983, Vladut-Drinfeld [VD83] gave the general upper bound

\[ A(q) \leq \sqrt{q} - 1. \]

Combined, they imply $A(q) = \sqrt{q} - 1$ when $q$ is a square. As a consequence, one sees that for square $q \geq 49$ and $\delta$ varying in an interval depending on $q$, the AG bound (3.2) is better than the Gilbert-Varshamov bound (2.1).

The truth of $A(q)$ in unknown for nonsquare $q$. However, some lower bounds were obtained. By considering class field towers analogous to the Golod-Shafarevich tower, Serre [Ser83, Ser85] gave the first lower bound estimate

\[ A(q) > c_1 \log q \]

for a constant $c_1 > 0$ independent of $q$. Then Temkine [Tem01] and Li-Maharaj [LM02] obtained the lower bound

\[ A(q') \geq c_2 r^2 \log q \frac{\log q}{\log r + \log q}, \]
which improves Serre’s bound when the cardinality of the finite field is a high power of a prime. The lower bound of the form

$$A(q^r) \geq cq^{1/2}$$

was established in a sequence of papers during the period 1997-99 by Niederreiter-Xing [NX98, NX99] and in 2002 by Li-Maharaj [LM02]. See [LM02] for a detailed description of the actual lower bound. For cubic powers, the estimates are much tighter. The first was given by Zink [Z85] in 1985 using degenerations of Shimura surfaces:

$$A(p^3) \geq \frac{2(p^2 - 1)}{p + 2}.$$  

The extension of this result with $p$ replaced by prime powers $q$ was done by Bezerra-Garcia-Stichtenoth [BGS05] in 2005 via explicit constructions. Their result also improves the Gilbert-Varshamov bound over $\mathbb{F}_{q^3}$ when $q \geq 7$.

5. Asymptotically optimal recursively defined towers

5.1. Recursively defined towers. A recursively defined tower over $\mathbb{F}_q$ is a strictly increasing tower $\mathcal{T}$ of function fields

$$F_1 \subset F_2 \subset F_3 \cdots$$

satisfying the following conditions:

1. Each $F_i$ is a function field with the same field of constants $\mathbb{F}_q$;
2. Each $F_{i+1}$ is a finite separable extension of $F_i$;
3. The genus $g(F_i) \to \infty$ as $i \to \infty$;
4. $F_1$ is the rational function field $\mathbb{F}(x_1)$, $F_{i+1} = F_i(x_{i+1})$ for $i \geq 1$, and there is a rational function $f(X,Y)$ in variables $X$ and $Y$ with coefficients in $\mathbb{F}_q$ such that $f(x_i, x_{i+1}) = 0$ for $i \geq 1$.

This tower has a geometrical interpretation. Let $V_i$ be the smooth projective curve over $\mathbb{F}_q$ with function field $F_i$. Geometrically, the curves $\{V_i\}$ form a covering

$$\cdots \to V_{i+1} \to V_i \to \cdots \to V_1 = \mathbb{P}_1.$$

Further, $(P_1, ..., P_i) \in V_i$ if and only if $f(P_j, P_{j+1}) = 0$, i.e., $(P_j, P_{j+1}) \in V_2$ for $j = 1, ..., i - 1$. Thus each $V_i$, $i \geq 2$, can be imbedded in $(V_2)^{i-1}$.

5.2. Asymptotically optimal recursively defined towers. For the tower $\mathcal{T}$, define

$$\lambda(\mathcal{T}) = \lim_{i \to \infty} \frac{N(V_i)}{g(V_i)} = \lim_{i \to \infty} \frac{\text{number of places of degree } 1 \text{ in } F_i}{\text{genus of } F_i}.$$  

Here $N(V_i)$ denotes the number of $\mathbb{F}_q$-rational points on $V_i$ and $g(V_i)$ its genus. When $\lambda(\mathcal{T}) = A(q)$, the tower $\mathcal{T}$ is said to be asymptotically optimal.

Since the value of $A(q)$ is known only for square $q$, it is this case that explicit examples of recursively defined asymptotically optimal towers are constructed. We list a few examples below. See the paper by Maharaj [Mh07b] for more details.

1. The first optimal tower was constructed by Garcia-Stichtenoth [GS95] in 1995. The field of constants is $\mathbb{F}_{q^2}$ and the recursive function is

$$f(X,Y) = (XY)^q + YX - X^{q+1}.$$
6 WEN-CHING W. LI

(2) In 1996 Garcia-Stichtenoth [GS96a] gave another optimal tower over $\mathbb{F}_{q^2}$ with

$$f(X, Y) = Y^q + Y - \frac{X^q}{X^{q-1} + 1}.$$ 

It is in fact a subtower of the first tower. Both of these towers are the so-called wild towers, meaning that the successive extensions $F_{i+1}$ over $F_i$ are wildly ramified. Elkies [El00] in 2000 showed that these two towers arose from Drinfeld modular curves by reduction.

(3) Garcia-Stichtenoth [GS96b] published two optimal tame towers in 1996 over $\mathbb{F}_4$ and $\mathbb{F}_9$, respectively. Elkies [El97] showed in 1997 that they came from the elliptic modular curves $\{X_0(3^n)\}$ mod 2 and $\{X_0(2^n)\}$ mod 3, respectively. Using Jacobi identity on theta functions, Solé in 2000 gave a different proof of this fact [So00].

(4) Aided by computer search, Li-Maharaj-Stichtenoth [LMS02] in 2002 exhibited four new optimal towers over $\mathbb{F}_4, \mathbb{F}_9, \mathbb{F}_{25}, \mathbb{F}_{49}$, respectively, whose recursive function $f(X, Y)$ is a polynomial of low degree. These are tame towers, and not subtowers of previously known towers. Elkies [El02] showed that these towers came from elliptic modular curves by reduction.

(5) In 2003, Garcia-Stichtenoth-Rück [GSR03] exhibited an optimal tower over $\mathbb{F}_{p^2}$ with

$$f(X, Y) = Y^2 - \frac{X^2 + 1}{2X}.$$ 

They are from the elliptic modular curves $\{X_0(2^n)\}$ modulo any odd prime $p$.

(6) In 2004 Bezerra-Garcia [BG04] found an optimal subtower of the second tower with the recursive function

$$f(X, Y) = \frac{Y - 1}{Y^q} - \frac{X^q - 1}{X}.$$ 

We also exhibit some recursively defined explicit towers over $\mathbb{F}_{q^3}$ with $\lambda(T) = \frac{2(q^2 - 1)}{q + 2}$.

1. The first such tower was constructed in 2002 by van der Geer and van der Vlugt [GV02] over $\mathbb{F}_8$ with

$$f(X, Y) = Y^2 + Y - X + 1 + 1/X.$$ 

2. In 2005 Bezerra-Garcia-Stichtenoth [BGS05] constructed such a tower over $\mathbb{F}_{q^3}$ with

$$f(X, Y) = \frac{1 - Y}{Y^q} - \frac{X^q + X - 1}{X}.$$ 

The argument was simplified by Bassa-Stichtenoth [BS07].

6. Elkies’ modularity conjecture

6.1. The statement of Elkies’ modularity conjecture. Based on the above examples, Elkies in 2000 proposed the following conjecture [El00]:

**Elkies’ modularity conjecture.** All asymptotically optimal recursively defined towers over $\mathbb{F}_q$ for square $q$ arise from reductions of elliptic modular curves, Shimura modular curves, and Drinfeld modular curves.
More precisely, among these towers, the tame ones over \( \mathbb{F}_p^2 \) are expected to arise from reductions of elliptic modular curves, those over \( \mathbb{F}_q^2 \) for nonprime \( q \) are expected to come from reductions of Shimura curves, and the wild towers should arise from reductions of Drinfeld modular curves. The work [LMS02] provided some numerical evidence to this conjecture for small squares \( q \) and low degree recursive defining polynomials \( f(X,Y) \).

To explain the philosophical reasons behind Elkies’ conjecture, we restrict ourselves to towers of elliptic modular curves \( \mod p \) viewed as curves over \( \mathbb{F}_p^2 \). Recall that for a positive integer \( N \) the modular group
\[
\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d, \in \mathbb{Z}, N \mid ac - bd = 1 \right\}
\]
acts on the Poincaré upper half-plane \( \mathfrak{H} \) by fractional linear transformations. The completion of \( \Gamma_0(N) \backslash \mathfrak{H} \) by adding cusps of \( \Gamma_0(N) \) is a Riemann surface of genus \( g_0(N) \), called the modular curve \( X_0(N) \). It is known that \( X_0(N) \) has a model defined over \( \mathbb{Q} \). Furthermore, for a prime \( p \) coprime to \( N \), the curve \( X_0(N) \) has a good reduction \( \mod p \) with the same genus \( g_0(N) \). We shall first show that for a fixed prime \( p \), the family \( \{X_0(N)\} \mod p \), as \( N \to \infty \) and \( p \) not dividing \( N \), is asymptotically optimal over \( \mathbb{F}_p^2 \). Then we explain why the family \( \{X_0(\ell^n)\}, n \geq 1 \), \( \mod p \) coprime to \( \ell \) is recursively defined. Put together, we see that the family \( \{X_0(\ell^n)\}, n \geq 1 \), \( \mod p \) coprime to \( \ell \) is an asymptotically optimal recursively defined tower.

### 6.2. Asymptotical optimality of elliptic modular towers

We give two ways to see why the family \( \{X_0(N)\} \mod p \) for \( p \) coprime to \( N \) is asymptotically optimal over \( \mathbb{F}_p^2 \). To count the number of rational points over \( \mathbb{F}_p^2 \), the first method is to appeal to the zeta function
\[
Z(X_0(N)/\mathbb{F}_p, t) = \frac{g_0(N)}{\prod_{i=1}^{g_0(N)} (1 - \lambda_i(p)t + pt^2)}.
\]
Here we have used the fact that the roots \( \beta \) and \( \bar{\beta} \) occur in pairs in the numerator of the zeta function \( Z(X_0(N)/\mathbb{F}_p, t) \) to combine them in a quadratic term \( 1 - \lambda_i(p)t + pt^2 \). A simple computation shows that
\[
N(X_0(N)/\mathbb{F}_p) = 1 + p^2 - \sum_{i=1}^{g_0(N)} \lambda_i(p)^2 + 2pg_0(N).
\]
So to compute \( N(X_0(N)/\mathbb{F}_p^2) \) we need to estimate the sum of \( \lambda_i(p)^2 \). For this, we invoke the so-called Eichler-Shimura congruence relation, which interprets \( \lambda_i(p) \) as the eigenvalues of the Hecke operator \( T_p \) on weight 2 cusp forms for \( \Gamma_0(N) \). This then allows us to apply the trace formula for the trace of the Hecke operator \( T_p \) on such cusp forms to conclude
\[
\sum_{i=1}^{g_0(N)} \lambda_i(p)^2 = (p + 1)g_0(N) + o(g_0(N)).
\]
Put together, this shows that, when \( N \) goes to infinity, the main term of \( N(X_0(N)/\mathbb{F}_p^2) \) is \( (p - 1)g_0(N) \) and the error term divided by the genus \( g_0(N) \) goes to zero. Thus the limit of \( N(X_0(N)/\mathbb{F}_p^2)/g_0(N) \) is \( p - 1 \), which is precisely \( A(p^2) \).
The second way is to compute \( \mathbb{F}_{p^2} \)-rational points on \( X_0(N)/\mathbb{F}_{p^2} \) using the moduli interpretation of the curve \( X_0(N) \). There are two kinds of \( \mathbb{F}_{p^2} \)-rational points on this curve: the first kind is the cusps, which are negligible compared to the genus \( g_0(N) \); the second kind corresponds to supersingular elliptic curves over \( \mathbb{F}_{p^2} \), whose cardinality is well-known. This is the approach in [TVZ82].

6.3. Recursively defined elliptic modular towers. In this subsection we consider the family \( \{X_0(\ell^n)\} \) for \( \ell > 1 \) and \( n \geq 1 \). Our purpose is to show that this family can be constructed recursively, as explained by Elkies [El00]. Consequently, modulo primes \( p \) coprime to \( \ell \), we obtain families of asymptotically optimal recursively defined towers over \( \mathbb{F}_{p^2} \).

A point \( z \in Y_0(\ell^n) = X_0(\ell^n) - \{\text{cusps}\} = \Gamma_0(\ell^n) \backslash \mathcal{H} \) represents an equivalence class of the pairs \((E, C_{\ell^n})\) of elliptic curves \( E \) over \( \mathbb{C} \) together with a cyclic subgroup \( C_{\ell^n} \) of order \( \ell^n \). We may also view this pair as an isogeny from \( E \to E/C_{\ell^n} \). Being a cyclic group, \( C_{\ell^n} \) has a unique filtration of cyclic subgroups of the following type
\[
C_{\ell^n} \supset C_{\ell^{n-1}} \supset \cdots \supset C_{\ell}.
\]

In terms of isogenies, this yields a chain
\[
E_0 = E \to E_1 = E/C_{\ell} \to \cdots \to E_n = E/C_{\ell^n}.
\]

Hence this gives rise to an imbedding
\[
\pi_n : Y_0(\ell^n) \to (Y_0(\ell^2))^{n-1}
\]
sending \( E = E_0 \to E_1 \to \cdots \to E_n \) to the point \((E_0 \to E_1 \to E_2, E_1 \to E_2 \to E_3, \ldots, E_{n-2} \to E_{n-1} \to E_n)\).

As a point in \( \mathcal{H} \), \( z \) represents the isogeny from \( E = \mathbb{C}/(\mathbb{Z} + z\mathbb{Z}) \) to \( E/C_{\ell^n} = \mathbb{C}/(\ell^{-n}\mathbb{Z} + z\mathbb{Z}) \cong \mathbb{C}/(\mathbb{Z} + \ell^n z\mathbb{Z}) \). So in terms of points in \( \mathcal{H} \), \( \pi_n \) maps \( z \in Y_0(\ell^n) \) to the point \((z, \ell z, \ldots, \ell^{n-2} z) \in (Y_0(\ell^2))^{n-1}\).

It remains to find the recursive relation \( f(X, Y) \) which characterizes the pair \((z, \ell z)\) for \( z, \ell z \in Y_0(\ell^2) \). Recall the Atkin-Lehner involution \( w_{\ell^n} \) on \( X_0(\ell^n) \) which sends \( z \) to \( \frac{1}{\ell^2} z \), or equivalently, it maps \( E_0 \to E_1 \to \cdots \to E_n \) to the dual isogeny \( E_n \to E_{n-1} \to \cdots \to E_0 \). From the diagram
\[
\begin{align*}
\begin{array}{ccc}
\downarrow \text{proj} & \downarrow \text{proj} & \\
X_0(\ell^2) & \to & X_0(\ell^2) \\
\uparrow w_{\ell^2} & & \uparrow w_{\ell^2} \\
X_0(\ell) & \to & X_0(\ell)
\end{array}
\end{align*}
\]
we get
\[
\text{proj} \circ w_{\ell^2}(z) = -1/\ell^2 z = w_{\ell} \circ \text{proj}(\ell z).
\]
This gives rise to the recursive function \( f(X, Y) \).

When \( X_0(\ell^2) \) has genus zero, parametrize the points on the curve by a Hauptmodul \( x \). The recursive function is a relation between \( x(z) \) and \( x(\ell z) \); the variables added to define the corresponding function fields in the tower are \( x(\ell^i z) \) for \( i \geq 1 \).

Elkies in [El00] computed the following recursive defining equations:
\[
\begin{align*}
f(X, Y) &= (X^2 - 1)((Y + 3)^2/(Y - 1)^2 - 1) - 1 & \text{for } \ell = 2, \\
f(X, Y) &= (X^3 - 1)((Y + 2)^3/(Y - 1)^3 - 1) - 1 & \text{for } \ell = 3, \\
f(X, Y) &= (X^4 - 1)((Y + 1)^4/(Y - 1)^4 - 1) - 1 & \text{for } \ell = 4.
\end{align*}
\]
Note that a tower can be defined by different recursive functions \( f(X, Y) \). For these three towers, recursive functions were also derived from functional equations satisfied by certain hypergeometric functions \( \binom{\alpha, \beta}{1; \cdot} \) obtained from AGM (arithmetic and geometric means) iterations and identities on theta functions. This line of work was initiated by Gauss, and continued by Ramanujan, Borwein-Borwein [B-B91], and Solé [So00].

### 6.4. Splitting set and ramification locus.

Given a tower \( T \), Garcia and Stichtenoth introduced the splitting set of \( T \) to be

\[
S = \{ \text{degree 1 place } v \text{ of } F_1 \mid v \text{ splits completely in all } F_i \}
\]

and the ramification locus of \( T \) to be

\[
R = \{ \text{places } v \text{ of } F_1 \mid v \text{ ramifies in some } F_i \text{ over } F_1 \}.
\]

They showed that if \( T \) is a tame tower, then there is a convenient lower bound for \( \lambda(T) \) in terms of the \( S \) and \( R \):

\[
\lambda(T) \geq \frac{2|S|}{2g(F_1) + \sum_{v \in R} \deg v - 2}.
\]

In particular, when the tower \( T \) is recursively defined by \( f(X, Y) \), choose \( S \) to be the subset of points on the projective line \( \mathbb{P}^1(F_q) \) such that for each \( P \in S \), all roots of \( f(P, Y) = 0 \) are contained in \( S \). Then \( S \) is certainly contained in the splitting set of \( T \).

How does one get the splitting set and ramification locus of the modular tower \( \{ X_0(\ell^n) \} \) mod \( p \)?

One example was given by Garcia-Stichtenoth-Rück in their 2003 paper [GSR03], where they showed that the tower \( \{ X_0(2^n) \} \) over \( F_{p^2} \) for \( p \) odd has

\[
f(X, Y) = Y^2 - \frac{X^2 + 1}{2X}.
\]

As the places contained in its ramification set and splitting set turn out to have degree 1, they are identified with \( F_{p^2} \)-rational points as follows:

\[
R = \{ 0, \infty, \pm 1, \pm \sqrt{-1} \},
\]

and

\[
S = \{ \alpha \in \overline{F}_p : H_p(\alpha^4) = 0 \} \subset F_{p^2}.
\]

Here

\[
H_p(X) = \sum_{0 \leq j \leq (p-1)/2} \binom{(p - 1)/2}{j}^2 X^j
\]

is the Deuring polynomial. As a subtower, the tower \( \{ X_0(4^n) \} \) over \( F_{p^2} \) has the same splitting set and ramification locus. More similar examples can be found in Hasegawa [Ha07].

In the next section, we show how the recursive defining function \( f \) and the polynomial \( H_p \) giving rise to the splitting set above arise from a functional equation satisfied by \( \binom{\frac{1}{2}, \frac{1}{2}}{1; \cdot} \). We shall arrive at this conclusion by considering Picard-Fuchs differential equations attached to modular curves.
7. Connections with Picard-Fuchs differential equations

7.1. Picard-Fuchs differential equations attached to modular curves.

On the modular curve \( X_0(1) = \text{PSL}_2(\mathbb{Z}) \backslash \mathcal{H} \) with Hauptmodul \( J = 1728/j \), choose the well-known Picard-Fuchs differential equation

\[
L_1: D_j^2 + \left[ \frac{1}{J} + \frac{1}{2(J - 1)} \right] D_j + \frac{5/144}{J(J - 1)}.
\]

It has three regular singular points at \( J = 0 \) (corresponding to the cusp \( \infty \)), \( J = 1 \) (corresponding to the elliptic point of order 2), and \( J = \infty \) (corresponding to the elliptic point of order 3). The characteristic exponents at these points are 0, 0, 0, 1/2, 1/12, 5/12, respectively. In a neighborhood of the cusp \( \infty \), the solution space has as a basis one analytic function and one with logarithmic singularity at \( \infty \). The function \( h_1 = 2F_1(1/12, 5/12; 1; J(z)) \) is the unique analytic solution with value 1 at \( \infty \); it is a modular form for \( SL_2(\mathbb{Z}) \) of weight 1 with multiplier.

For a subgroup \( \Gamma \) of \( SL_2(\mathbb{Z}) \) of finite index, the modular curve \( X(\Gamma) \) admits Picard-Fuchs differential operators. These are second order differential equations whose singular points are regular singular. One way to construct them is to "weak-lift" \( L_1 \) along the natural projection from \( X(\Gamma) \) to \( X_0(1) \). Impose some normalization conditions similar to what we have observed on \( L_1 \). In particular, we demand that the regular singular points are the cusps and the elliptic points of \( X(\Gamma) \).

Given \( \ell > 1 \) there are two morphisms from \( X_0(\ell^2) \) to \( X_0(\ell) \) given by \( \phi_\ell = \text{proj} \) and \( \phi_\ell' = w_\ell^{-1} \circ \text{proj} \circ w_\ell \) so that \( \phi_\ell(z) = z \) and \( \phi_\ell'(z) = \ell z \). To proceed, we follow the steps described in Maier [Mar07]:

- Compatibly weak-lift \( L_1 \) to a Picard-Fuchs differential equation \( L_N \) on \( X_0(N) \) along the projection from \( X_0(N) \) to \( X_0(1) \) for \( N = \ell \) and \( \ell^2 \). Let \( h_N \) denote the unique analytic solution in a neighborhood of the cusp \( \infty \) with the constant term 1; it is a weight 1 modular form with multiplier for \( \Gamma_0(N) \) as before.

**Remark 7.1.** In general, if the curve \( X_0(N) \) has \( n \) singular points and genus \( g \), in addition to the characteristic exponents at each singular point, there are still \( n - 3 + 3g \) projective accessory parameters and \( g \) affine accessory parameters that can vary. Consequently, once the characteristic exponents are specified, such a differential equation \( L_N \) is unique when \( n = 3 \) and \( g = 0 \), and nonunique otherwise.

- Find a suitable weak lift of \( L_\ell \) to \( L_\ell' \) on \( X_0(\ell^2) \) along the map \( \phi_\ell' \) such that the two unique normalized analytic solutions \( h_\ell' \) and \( h_\ell'' \) around the cusp \( \infty \) agree.
- Express \( h_\ell' \) and \( h_\ell'' \) using \( \ell \) to get a functional equation of \( h_\ell \), from which the recursive function \( f(X, Y) \) and the splitting set \( S \) can be read off.

For \( \ell = 2, 3, 4, 5, 6, 7 \), the modular curve \( X_0(\ell) \) has genus zero. Choose a Hauptmodul \( x_\ell \) with \( \text{div}(x_\ell) = \infty \ell - 0 \ell \) supported at two cusps. Further, for \( \ell = 2, 3, 4, 5 \), the curve \( X_0(\ell^2) \) also has genus zero. Choose a Hauptmodul \( x_{\ell^2} \) such that \( \text{div}(x_{\ell^2}) = \infty \ell^2 - 0 \ell^2 \). By checking ramifications, one finds \( x_\ell \circ \phi_\ell = R_\ell(x_{\ell^2}) \) for a polynomial \( R_\ell \) of degree \( \ell \), and \( x_{\ell^2} \circ \phi_\ell' = x_{\ell^2} / S_\ell'(x_{\ell^2}) \) for a polynomial \( S_\ell' \) of degree \( \ell - 1 \). Maier [Mar07] showed that for \( \ell = 2, 3, 4, 5 \), the relation \( h_{\ell^2} = h_{\ell''} \) can be expressed using \( h_\ell \) as follows:

\[
h_\ell(R_\ell(x_{\ell^2})) = [S_\ell'(x_{\ell^2}) / S_\ell'(0)]^{-\psi(\ell)/12} h_\ell \left( \frac{x_{\ell^2}}{S_\ell'(x_{\ell^2})} \right),
\]

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.
where $\psi(\ell)$ is the index of $\Gamma_0(\ell)$ mod center in $\text{PSL}_2(\mathbb{Z})$.

For $\ell = 2, 3, 4$, the total number of cusps and elliptic points for $\Gamma_0(\ell)$ is three. So its Picard-Fuchs differential equation $L_\ell$ has three regular singular points, and $h_\ell$ can be expressed in terms of the hypergeometric function $2F_1$. After suitable change of variables, one obtains the following functional equations for $2F_1$ corresponding to $\ell = 2, 3, 4$, respectively:

(a) $2F_1(1/4, 3/4; 1; 1 - (1/1+\ell)^2) = (1 + 3x)^{1/2} 2F_1(1/4, 3/4; 1; x^2)$,

(b) $2F_1(1/3, 2/3; 1; 1 - (1/1+\ell)^3) = (1 + 2x)^{1/2} 2F_1(1/3, 2/3; 1; x^3)$,

(c) $2F_1(1/2, 1/2; 1; 1 - (1/1+\ell)^4) = (1 + x)^2 2F_1(1/2, 1/2; 1; x^4)$.

They agree with the functional equations obtained from AGM. This gives a modular interpretation of the functional equations.

### 7.2. The recursive equation and splitting set.

Now we explain how to read off the recursive equation $f$ and the splitting set $S$ for the elliptic modular tower $\{X_0(\ell^n)\}$ modulo a prime $p$ coprime to $\ell$. We study the case $\ell = 4$ and relate it to the known result in [GSR03] discussed at the end of §6.

Consider hypergeometric functions modulo $p$. In the literature this is often treated by taking truncations to arrive at polynomials of degree $< p$. But then the functional equations no longer hold. Our observation is that a more natural approach is to raise the functions mod $p$ to the power $1 - p$. In so doing, we not only obtain polynomials of degree $< p$, but also retain the functional equations. To illustrate this point, we prove

**Theorem 7.2.** For odd primes $p$ we have

$$2F_1(1/2, 1/2; 1; x)^{1-p} \equiv H_p(x) \pmod{p},$$

where $H_p$ is the Dieuring polynomial (6.1).

**Proof.** The hypergeometric function

$$2F_1(1/2, 1/2; 1; x) = \sum_{n \geq 0} c_n x^n = 1 + \frac{1}{2}x + \frac{3^2}{4^2 2^2} x^2 + \cdots,$$

where

$$c_n = \left(\frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} + \cdots \frac{1}{2} + n - 1)/n!\right)^2 \quad \text{for } n \geq 1,$$

is a formal power series with coefficients in the ring of $p$-adic integers $\mathbb{Z}_p$ which converges to a $p$-adically analytic function on $p\mathbb{Z}_p$. One checks easily that the initial $p$ terms in the series satisfy

$$\sum_{0 \leq n \leq p-1} c_n x^n \equiv H_p(x) \pmod{p\mathbb{Z}_p}.$$

Further, it is not hard to compute $c_n$ modulo $p\mathbb{Z}_p$ for $n \geq p$ to obtain

$$2F_1(1/2, 1/2; 1; x) \equiv H_p(x)H_p(x^p)H_p(x^{p^2}) \cdots \pmod{p\mathbb{Z}_p}$$

$$\equiv H_p(x)H_p(x^p)H_p(x^{p^2}) \cdots \pmod{p\mathbb{Z}_p}.$$

Note that the series $H_p(x)H_p(x^p)H_p(x^{p^2}) \cdots$ converges to an invertible function $h(x) \in \mathbb{Z}_p[[x]]$ with $h(0) = 1$ so that

$$2F_1(1/2, 1/2; 1; x) = h(x)(1 + pg(x))$$
for some function \( g(x) \in \mathbb{Z}_p[[x]] \). Thus
\[
_{2} F_{1}(1/2, 1/2; 1; x)^p = h(x)^p(1 + pg(x))^p
\]
and hence
\[
_{2} F_{1}(1/2, 1/2; 1; x)^{1-p} = \frac{h(x)}{h(x)^p}(1 + pg(x))^{1-p} = H_p(x)(1 + pk(x))
\]
for some \( k(x) \in \mathbb{Z}_p[[x]] \). Modulo \( p \) yields the desired congruence. \( \square \)

Combined with (c) we obtain the functional equation for \( H_p \):
\[
H_p(1 - \left(\frac{1-x}{1+x}\right)^4) \equiv (1 + x)^{2(1-p)}H_p(x^4) \pmod{p}.
\]
So if \( \alpha^4 \) is a root of \( H_p \) and \( \alpha \neq -1 \), then so is \( 1 - \left(\frac{1-\alpha}{1+\alpha}\right)^4 \). Write \( \beta^4 = 1 - \left(\frac{1+\alpha}{1-\alpha}\right)^4 \) so that \( H_p(\beta^4) = 0 \). Repeating this procedure, we obtain a tower \( \{X_0(4^n)\} \pmod{p} \) recursively defined by
\[
f(X,Y) = Y^4 - 1 + \left(\frac{1-X}{1+X}\right)^4,
\]
which records the relation between \( \alpha \) and \( \beta \). This tower has splitting set \( S = \{ \alpha \in \overline{\mathbb{F}}_p : H_p(\alpha^4) = 0 \} \). The smallest field containing \( S \) can be shown to be \( \mathbb{F}_p^2 \).

For \( \ell = 5, 6, 7 \) the Picard-Fuchs differential equation has 4 regular singular points, \( h_\ell \) can be expressed using the local Heun function \( H1 \). Maier [Mar07] also obtained explicit functional equations in this case. Similar functional equations can be derived for \( \{X_0(m\ell^n)\}_{n \geq 1} \), where \( m \) and \( \ell \) are coprime.

The paper by Beelen and Bouw [BB05] explained the tower [GSR03] by Garcia-Stichtenoth-Rück along the same line, by considering the Picard-Fuchs differential equation in characteristic \( p \) directly.

References


This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.


Department of Mathematics, Pennsylvania State University, University Park, PA 16802 U.S.A. and National Center for Theoretical Sciences, Mathematics Division, Third General Building, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C.

E-mail address: wli@math.psu.edu