Robuste Silberscheiben

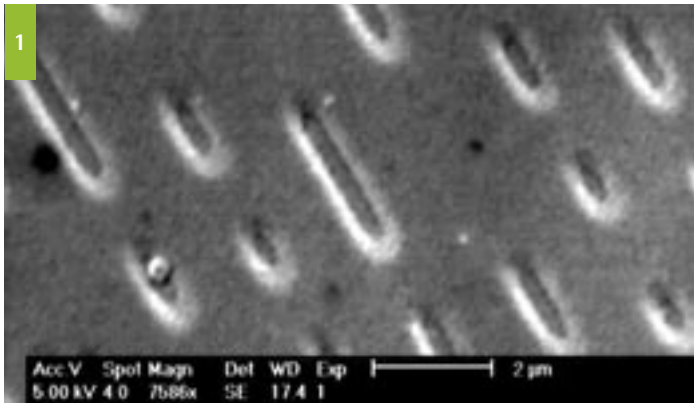
DIE MATHEMATIK AUF DER CD

Heute vor etwa zwanzig Jahren begann der Siegeszug der Compact Disc, vor allem wegen ihrer brillanteren Klangqualität und ihrer größeren Widerstandsfähigkeit gegen Kratzer. Doch was ist der Ursprung dieser Überlegenheit? Es ist die Mathematik, die dem von Philips und Sony entwickelten Compact-Disc-Audiosystem zugrundeliegt. Der reine Hörgenuss wäre ohne Fehler korrigierende Codes undenkbar. Drei Wissenschaftler des Instituts für Algebraische Geometrie geben einen kurzen Einblick in die Mathematik auf der CD.

Von Bits und Pits

Wir müssen zunächst erklären, wie die Speicherung auf der CD technisch vor sich geht. Bei einer Musikaufnahme wird 44100mal pro Sekunde eine Momentaufnahme (ein Muster) des Analogsignals aufgezeichnet. Dabei wird das Analogsignal digitalisiert, d.h. es wird in ein Bitmuster, eine Folge von Bits, übertragen. Ein Bit ist eine 0 oder eine 1. Jedes Muster entspricht dabei einer Folge von 32 Bits, aufgefasst als 4 aufeinander folgende Bytes (ein Byte ist eine Folge von 8 Bits). Diese Bitmuster werden nun auf der CD in einer fünf Kilometer langen spiralförmigen Spur gespeichert, die nur eine Breite von weniger als einem µm hat. Dabei brennt man Vertiefungen auf der CD-Oberfläche ein, die *Pits* genannt werden. Die Teile zwischen zwei aufeinander folgenden Pits heißen *Lands*. Bild 1 zeigt eine Elektronenrastermikroskopaufnahme von Teilen paralleler Spuren. Die Spur wird optisch von einem Laserstrahl in regelmäßigen Zeitabständen abgetastet. Jeder Übergang Land/Pit oder Pit/Land wird als 1 interpretiert. Findet zum Abtastzeitpunkt kein solcher Übergang statt, so wird dies als 0 interpretiert.

Kommen nun Kratzer oder Staub auf die CD, so werden mehrere der Pits abgedeckt und der Laserstrahl liest dadurch fehlerhafte Bitmuster. Trotzdem wird die Musik nicht verfälscht. Woran liegt das? Hier kommen jetzt die Fehler korrigierenden Codes zum Einsatz.



1

Leibniz und das binäre Zahlensystem

Leibniz hat das binäre Zahlsystem unabhängig von anderen (wenn auch nicht als einziger) entwickelt und gezeigt, wie man damit rechnen kann. Wichtig war Leibniz vor allen Dingen der philosophische und theologische Aspekt des binären Zahlensystems, den er in den Jahren 1695-1701 intensiv mit den Herzögen Rudolph August und Anton Ulrich diskutierte. Die im Logo der Leibniz Universität verwendete Handschrift (Bild 2) stammt aus dem Neujahrsbrief vom Januar 1697 an Herzog Rudolf August von Wolfenbüttel. Leibniz interpretierte sein Zahlensystem im Sinne der Schöpfung, in der aus dem Nichts (der Null) und Gottes Wort (der Eins) die ganze Welt entstanden sei. Er fasste dies in dem Satz »omnibus ex nihilo ducendis sufficit unum (um alles aus dem Nichts herzuleiten, genügt Eines)« zusammen.

Auch wenn das Interesse von Leibniz am binären Zahlensystem in erster Linie philosophisch-theologisch begründet war, befasste er sich auch mit möglichen Anwendungen. So konstruierte er eine mechanische Rechenmaschine, die das binäre Zahlsystem benutzt. Ein Modell, das auf den Beschreibungen von Leibniz beruht, wurde auf Grund einer Anregung von L. von Mackensen in den 1970er Jahren gefertigt. Dieses Modell ist heute in der Leibniz Ausstellung der Universität zu besichtigen.

Die grundlegende Idee besteht darin, dass man die Information mit einer gewissen Redundanz abspeichert. Die

einfachste Möglichkeit wäre, die Symbole einfach zu wiederholen. Statt eines Bits 0 speichern wir zum Beispiel die Bitfolge 00000 und statt 1 die Bitfolge 11111 ab. Die Wörter 00000 und 11111 bilden einen so genannten Fehler korrigierenden Code, der Wiederholungscode genannt wird. Die Anzahl der Symbole (hier 5) nennt man Länge des Codes, die Symbolketten 00000 und 11111 nennt man die Wörter des Codes. Wird das Wort 00000 nun (durch Staub auf der CD) irrtümlicherweise als 01001 gelesen, so kann man das ursprüngliche Wort rekonstruieren, wenn man davon ausgeht, dass höchstens zwei Fehler passiert sind. Das liegt daran, dass sich die beiden Codewörter 00000 und 11111 an 5 Stellen unterscheiden.

Dieser Code wäre für die CD aber ein schlechter Code: Wir haben nur zwei Informationen (0 oder 1) und brauchen dafür 5 Stellen zur Fehlerkorrektur. Auch wenn auf der CD sehr viel Platz ist, will man doch ökonomischer mit dem Speicherplatz umgehen.

Hier hilft uns die Mathematik, denn man kann mit den Symbolen 0 und 1 auch rechnen.

Rechnen mit Bitsequenzen

Man kann eine Bitfolge auch als eine Zahl im binären Zahlensystem auffassen. So gilt

$$11111=1\cdot2^4+1\cdot2^3+1\cdot2^2+1\cdot2^1+1\cdot2^0$$

also stellt 11111 die Dezimalzahl 31 dar. Die Zahlen kann man nun zueinander addieren, wie wir es vom Addieren im Dezimalsystem gewohnt sind, z.B.

$$\begin{array}{r} 11010010 \\ 11011101 \\ \hline 110101111 \end{array}$$

Nun wollen wir aber mit Bitsequenzen mit einer festen Anzahl von Stellen, z.B. mit Bytes, rechnen. Dann vergessen wir einfach den Übertrag, der über die vorgegebene Zahl von Stellen hinausgeht. Damit haben wir eine Addition auf den $256=2^8$ Dualzahlen mit 8 Stellen erklärt. Nun wollen wir auch noch eine Multiplikation definieren. Dabei soll das Produkt mit den Dezimalzahlen 0 und 1 wie gewohnt erklärt sein. Das Produkt der anderen Elemente ist komplizierter. Wir wollen dies nur für Bitpaare erklären. Um eine Verwechslung mit den Dezimalzahlen zu vermeiden, führen wir für die Bitpaare die folgende Bezeichnung ein:

00	01	10	11
0	1	a	b

Die Addition und Multiplikation dieser vier Elemente nehmen wir nach den folgenden Tabellen vor:

+	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

·	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

Die Menge, die aus den vier Bitpaaren besteht und auf der diese Addition und Multiplikation definiert ist, nennen wir den Körper $\mathbf{F}_4$.

Das Prinzip des Reed-Solomon-Codes

Wir wollen nun erklären, wie man über einem solchen Körper einen Code konstruieren kann.

Dazu nehmen wir als Beispiel den Körper $\mathbf{F}_4$. Wir stellen uns vor, dass die Information, die wir codieren wollen, durch Paare von Elementen von $\mathbf{F}_4$ gegeben ist. Damit haben wir $16=4^2$ verschiedene Informationen. Die vier Elemente des Körpers $\mathbf{F}_4$ kann man sich geometrisch als vier Punkte auf einem Zahlenstrahl vorstellen. Wir betrachten nun die Funktionen $f(x)=1$ und $g(x)=\frac{1}{x}$, wobei x einer der Punkte 1, a, b ist. Das Paar (a, b) codieren wir dann zum Beispiel dadurch, dass wir die Funktion $a\cdot f+b\cdot g$ auf den Punkten 1, a, b auswerten. Dies ergibt den dreistelligen Ausdruck $10b$. Insgesamt erhält man damit einen Code mit den folgenden Codewörtern:

+	$0\cdot g$	$1\cdot g$	$a\cdot g$	$b\cdot g$
$0\cdot f$	000	1ba	a1b	ba1
$1\cdot f$	111	0ab	b0a	ab0
$a\cdot f$	aaa	b10	0b1	10b
$b\cdot f$	bbb	a01	1a0	01a

Die Idee der Konstruktion eines solchen Codes stammt von den Mathematikern Irving S. Reed und Gustave Solomon. Man nennt diesen Code daher einen *Reed-Solomon-Code (RS-Code)*. Um uns eine Vorstellung von diesem Code zu machen,

Abbildung 1
Elektronenrastermikroskopaufnahme einer CD

Abbildung 2
Das binäre Zahlensystem in der Handschrift von Leibniz
Quelle: Gottfried Wilhelm Leibniz Bibliothek – Niedersächsische Landesbibliothek Hannover, LBR II, I, Vol. 15. Bl. 19v

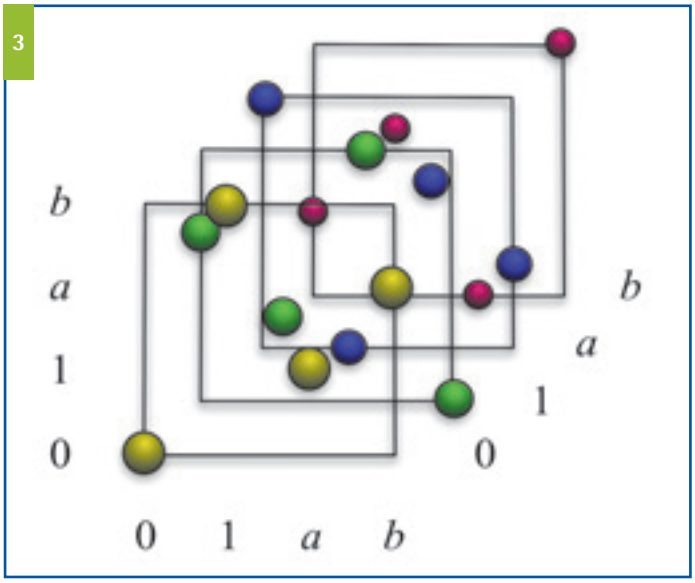


Abbildung 3
Die Codewörter des [3,2,2]-RS-Codes als Punkte in einem 4 x 4 x 4-Raster

Abbildung 4
Illustration der im Text beschriebenen Nullstellenmenge von $x^3=y^2+y$, in der reellen Ebene (4a) und in der als Kreuzhaube dargestellten reellen projektiven Ebene (4b) (vgl. Infokasten 2).

Abbildung 5
Geraden durch einen Punkt

Abbildung 6
Ansichten der reellen projektiven Ebene als Kreuzhaube (6a und 6b) und der reellen Ebene (6c)

fassen wir die Codewörter als Punkte in einem dreidimensionalen 4 x 4 x 4-Raster auf (siehe Bild 3).

Die Wortlänge ist 3. Wir sehen, dass sich je zwei Codewörter an mindestens 2 Stellen unterscheiden und 000 und 0*ab* unterscheiden sich an genau 2 Stellen. Man sagt, der Code hat den *Minimalabstand* 2. Es gibt $16=4^2$ Codewörter. Man sagt deshalb, dass der Code die *Dimension* 2 hat. Man spricht von dem [3,2,2]-RS-Code über $\mathbf{F}_4$. Der Quotient R=2/3 aus Dimension und Länge heißt die *Informationsrate* des Codes. Dieser Code ist besser als der Wiederholungscode, er hat eine wesentlich bessere Informationsrate. Die bessere Informationsrate wird allerdings erkauft durch einen schlechteren Minimalabstand, d.h. die Codewörter sind nicht so unterschiedlich wie beim Wiederholungscode.

Die Hauptaufgabe der Codierungstheorie besteht nun darin, die miteinander konkurrierenden Ziele eines großen Minimalabstandes und einer großen Informationsrate zu verwirklichen. In diesem Sinne sind die RS-Codes sehr gute Codes.

Unser RS-Code ist ein einfaches Modell der Codes, die bei der CD verwendet werden. Dort wird als Körper ein Körper mit 256 Elementen zugrunde gelegt. Das bedeutet, dass einem Element des Körpers gerade ein Byte entspricht. Man verwendet zwei verschiedene RS-Codes über diesem Körper, die noch in bestimmter Weise miteinander verwoben sind. Der Fachausdruck für das Codierungsverfahren ist CIRC *(Cross-Interleaved Reed-Solomon Code)*.

Algebraisch-geometrische Codes

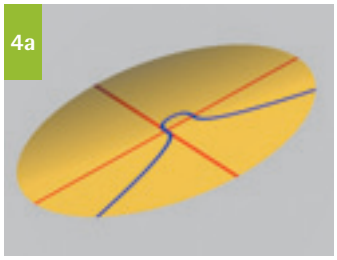
Wie wir bereits erwähnt haben, erhält man RS-Codes durch Auswerten von Funktionen an endlich vielen geeigneten Punkten. Diese wählt man als die Punkte eines geometrischen Objekts über einem endlichen Körper, d.h. über einem Körper mit endlich vielen Elementen. Codes, die man in solcher Weise erhält, sind *algebraisch-geometrische Codes* und stellen ein aktuelles Forschungsgebiet dar. Für den RS-Code werden als geometrische Objekte algebraische Kurven in der Ebene betrachtet, d.h. Nullstellenmengen einer Gleichung in zwei Veränderlichen, wie etwa

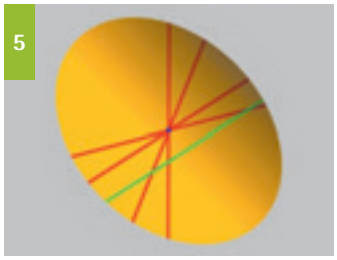
$$x^3=y^2+y,$$

deren Nullstellenmenge in der reellen Ebene mit Koordinaten x und y in Bild 4 dargestellt ist (Genauer gesagt, betrachtet man eigentlich Kurven in der projektiven Ebene; diese wird im Infokasten rechts beschrieben). Bei der Erzeugung von Codes arbeiten wir jedoch über endlichen Körpern und nicht über den reellen Zahlen; über $\mathbf{F}_4$ liefert die obige Gleichung gerade folgende Nullstellen:

(0,0), (1,*a*), (1,*b*), (*a*,*a*), (*a*,*b*), (*b*,*a*), (*b*,*b*).

Durch eine geschickte Wahl des Grundkörpers, der algebraischen Kurve, der Funktionen und der Punkte auf der Kurve, an denen man die Funktionen auswertet, kann man auf diese Weise Codes konstruieren, die sich dadurch auszeichnen, dass sie bei großem Minimalabstand über





eine sehr hohe Informationsrate verfügen. Gleichzeitig lässt die algebraische Konstruktion relativ einfache Decodierungs- bzw. Fehlerkorrekturverfahren zu.

Literatur

- Jack H. van Lint: Die Mathematik der Compact Disc. In: Alles Mathematik. Von Pythagoras zum CD-Player (Hrsg.: M. Aigner, E. Behrends), Vieweg, Braunschweig/Wiesbaden 2002.

Prof. Dr. Wolfgang Ebeling
Jahrgang 1951, ist Professor für Mathematik am Institut für Algebraische Geometrie und Autor eines Buches über Gitter und Codes; Arbeitsgebiete: Algebraische Geometrie, Singularitätentheorie

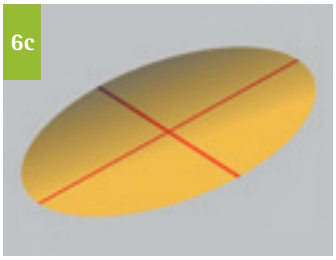
Dr. Anne Frühbis-Krüger
Jahrgang 1970, Habilitation 2005 Universität Kaiserslautern, ist wissenschaftliche Mitarbeiterin am Institut für Algebraische Geometrie; Arbeitsgebiete: Computeralgebra und Singularitätentheorie

Prof. Dr. Klaus Hulek
Jahrgang 1952, ist Professor für Mathematik am Institut für Algebraische Geometrie und Vizepräsident für Forschung der Leibniz Universität Hannover; Arbeitsgebiete: Algebraische und komplexe Geometrie

6a

6b

6c



Kurven in der projektiven Ebene

Bekanntlich schneiden sich zwei Geraden in der Ebene stets in einem Punkt, wenn sie nicht parallel sind. Wie in Bild 5 illustriert kann man daher die Geraden (rot) durch einen Punkt (blau) durch ihren Schnittpunkt mit einer festen Gerade kennzeichnen, die den blauen Punkt nicht trifft (grün) – mit der Ausnahme der zu der grünen Geraden parallelen Geraden. Um hierbei ein einheitliches Verhalten zu erreichen, nehmen nun die Mathematiker einen weiteren Punkt ∞ zu der Geraden hinzu, den Sie mit diesem fehlenden Schnittpunkt identifizieren.

Das neu erhaltene Objekt bezeichnet man als den projektiven Abschluß der (grün gezeichneten) reellen Gerade. Ähnlich kann man auch den projektiven Abschluß der reellen Ebene erhalten; jedoch reicht es dabei nicht, einen einzigen Punkt hinzuzunehmen, man benötigt eine Gerade. Damit ist die projektive Ebene nicht mehr ohne Anstrengung im Dreidimensionalen darstellbar; Bild 6a und Bild 6b zeigen eine mögliche Einbettung der reellen projektiven Ebene, die Kreuzhaube. Darauf sind in rot die Koordinatenachsen der ursprünglichen reellen Ebene (Bild 6c) eingezeichnet. Auch wenn die Darstellung der reellen projektiven Ebene recht kompliziert erscheint, lassen sich viele Eigenschaften geometrischer Objekte im Projektiven wesentlich einfacher und vor allem einheitlicher fassen, weshalb oftmals diese Beschreibung vorzuziehen ist.