Bemerkungen zum 1-Bit-Fehler korrigierbaren zyklischen Code

(zu „Grundkurs Codierung“, 3. Auflage 2006, Vieweg Verlag, ISBN 3-528-25399-1,
Unterkapitel 3.7.3, Seiten 111 ff)

update vom 19.11.1997

Frage: Warum müssen die Generatorpolynome g(x) irreduzibel über dem Z_2 sein, damit man einen 1-Bit-Fehler-korrigierbaren Code **maximaler** Länge erhält?

Ausgangslage: Binäre zyklische Codes sind Polynome über dem Körper Z_2, die Codewort-Bits stellen die Polynomkoeffizienten dar. Alle Operationen lassen sich daher gemäß den Gesetzmäßigkeiten und Rechenregeln für solche Polynome durchführen und nutzen. Die Codewörter v(x) werden als Vielfache des Generatorpolynomens g(x) mit dem Informationspolynom u(x) gebildet, entweder als

$$v(x) = u(x) \cdot g(x)$$

oder als

$$v(x) = u(x) \cdot x^{\text{grad } g(x)} + u(x) \cdot x^{\text{grad } g(x)} \text{ MOD } g(x).$$

Beide Formen sind gleichwertig, ergeben aber unterschiedliche Codewörter. Die erstgenannte Form ist einfacher zu berechnen, jedoch kann man die Informationsbits nicht mehr unmittelbar in v(x) erkennen, während diese bei der zweiten Form alle als u(x) den linken Teil des Codewortes bilden (daher systematischer Code).

Generatorpolynome g(x) sind im allgemeinen - d.h. bei t-Bit-Fehler-korrigierbaren Codes - Produkte von t geeigneten irreduziblen Polynomen, genauer solcher irreduzibler Polynome, die unabhängige Nullstellen im gewählten Galoisfeld $GF(2^m)$ haben.

Im Sonderfall des 1-Bit-Fehler-korrigierbaren Codes besteht das Generatorpolynom g(x) nur aus einem einzigen irreduziblen Polynom. Die Codelänge n entspricht der Ordnung des durch g(x) definierten primitiven Elements α und hat mit m = grad g(x) daher den Wert $n = 2^m - 1$. Für reduzible Generatorpolynome gleichen Grades wird diese Länge nicht erreicht. Wie lässt sich *ohne* Kenntnis der Eigenschaften von Galoisfeldern erklären, warum solche Codes eine maximale Länge haben, wenn sie mit irreduziblen Generatorpolynomen gebildet werden? Mit maximaler Länge ist im übrigen diejenige Anzahl von Codewortstellen gemeint, bei der sich ein 1-Bit-Fehler noch *sicher* lokalisieren lässt.

Antwort: Die Fehlerkorrektur erfolgt mit Hilfe des Syndrompolynoms s(x). Dies entsteht durch MOD g(x)-Division des Empfangswortes w(x) als

$$s(x) = w(x) \text{ MOD } g(x).$$

Da w(x) die Summe des Codewortes v(x) und eines eventuellen Fehlers e(x) ist, fällt bei der MOD g(x) - Division der Anteil des Codwortes heraus und es bleibt nur der Anteil des Fehlers übrig:

$$w(x) = v(x) + e(x) = q(x) \cdot g(x) + e(x)$$

$$s(x) = \text{MOD } w(x) = \text{MOD } e(x)$$

Das Syndrompolynom hat also höchstens den Grad m-1 und damit maximal m Koeffizienten. Damit lassen sich höchstens 2^m verschiedene Zustände darstellen. Da der Fall “Fehlerfreiheit” oder e(x) = 0 bereits den Zustand

$$s(x) = \quad 0$$

belegt, bleiben theoretisch noch 2^m - 1 Zustände für die Kennzeichnung von 1-Bit-Fehlerpositionen. Eine größere Codelänge als $n = 2^m - 1$ ist also nicht sinnvoll. Allerdings stellt dies zunächst nur eine notwendige Bedingung dar, um überhaupt 2^m - 1 - Fehlerpositionen festlegen zu können. Hinreichend wird die Bedingung erst, wenn die 2^m - 1 Zustände auch wirklich paarweise verschieden sind.

Genau diese paarweise Verschiedenheit wird aber nur dann erreicht, wenn g(x) irreduzibel ist. Betrachten wir zunächst 2 Beispiele. Das irreduzible Generatorpolynom $g(x) = x^3 + x + 1$ erzeugt für die 7 verschiedenen Fehlerpositionen die folgenden Syndrome s(x) (hier grau hinterlegt):

		Syndrom
$x^0 =$		1
$x^1 =$		x
$x^2 =$		x^2
$x^3 =$	$(x^3 + x + 1) +$	$x + 1$
$x^4 =$	$x \cdot (x^3 + x + 1) +$	$x^2 + x$
$x^5 =$	$(x^2 + x) \cdot (x^3 + x + 1) +$	$x^2 + x + 1$
$x^6 =$	$(x^3 + x + 1) \cdot (x^3 + x + 1) +$	$x^2 + 1$

Man sieht, dass alle 7 Syndrome verschieden sind. Nimmt man als Generatorpolynom das zerlegbare (= reduzible) Polynom

$$g(x) = \quad x^3 + x^2 + x + 1 = \quad (x + 1)^3 ,$$

so ergibt sich

		Syndrom
$x^0 =$		1
$x^1 =$		x
$x^2 =$		x^2
$x^3 =$	$(x^3 + x^2 + x + 1) +$	$x^2 + x + 1$
$x^4 =$	$(x + 1) \cdot (x^3 + x^2 + x + 1) +$	1
$x^5 =$	$(x^2 + x) \cdot (x^3 + x^2 + x + 1) +$	x
$x^6 =$	$(x^3 \quad x^2) \cdot (x^3 + x^2 + x + 1) +$	x^2

Hier sind nur 4 Syndrompolynome verschieden.

Es lässt sich allgemein zeigen, dass – bei zyklischen Hammingcodes zur 1-Bit-Fehlerkorrektur - nur für irreduzible Generatorpolynome die maximale Zahl verschiedener Syndrompolynome entsteht. Betrachten wir dazu irgendein Generatorpolynom g(x) des Grades m (das aber als Koeffizienten bei x^0 immer 1 hat, da es sonst durch x teilbar wäre). Jedes Fehlerpolynom lässt sich dann darstellen als

$$e(x) = \quad q(x) \cdot g(x) + s(x),$$

wobei sich das Syndrompolynom s(x) als Rest der Division e(x)/g(x) bzw. e(x) MOD g(x) ergibt.

Da die Zahl der verschiedenen Syndrompolynome maximal 2^m-1 beträgt, wenn man vom Fall „Fehlerfreiheit“ mit s(x) = 0 absieht, wird man beim systematischen Ausrechnen der Syndrompolynome für verschiedene Fehlerpolynome auf jeden Fall einmal auf ein gleiches treffen, sagen wir s*(x). Bei zwei verschiedenen Fehlern erhält man also für ein bestimmtes Paar i, j mit j > i (i und j kennzeichnen die Fehlerposition im Codewort v, i = 0 bedeutet z. B., dass der Fehler in der ganz rechten Stelle aufgetreten ist)

$$e_1 (x) = \quad x^i = \quad q_1(x) \cdot g(x) + s^*(x)$$

und

$$e_2 (x) = \quad x^j = \quad q_2(x) \cdot g(x) + s^*(x).$$

$q_1(x)$ und $q_2(x)$ sind für $i \neq j$ ungleich, da die höchste Potenz in g(x) in beiden Fällen dieselbe ist (nämlich m) und mit der höchsten Potenz in q(x) die höchste Potenz in e(x) sichergestellt wird

Addition ergibt:

$$x^i + x^j = x^i \cdot (x^{j-i} + 1) = [q_1(x) + q_2(x)] \cdot g(x)$$

Linke und rechte Seite besitzen dieselben Teiler. Das Generatorpolynom g(x) ist kein Teiler von x^i , da es eine 1 enthält, also muß es Teiler von $x^{(j-i)} +1$ sein. Der Fermatsche Satz (siehe z. B. Unterkapitel 2.2, Seite 46) besagt, dass irreduzible Polynome g(x) ein Polynom $x^r +1$ erst dann teilen, wenn

$$r \geq 2^m - 1$$

ist. Für reduzible Polynome gilt immer ein kleinerer Wert. Also müssen irreduzible Polynome gewählt werden, um die maximale „Ausbeute“ an Codewortstellen zu erhalten.

Für einen Grad m = 3 des Generatorpolynoms ist j - i = 7. Syndrompolynome s(x) werden sich erst dann wiederholen, wenn die Fehlerpositionen wenigstens 8 Stellen auseinander stehen. Andersherum ausgedrückt dürfen Fehler bei den x-Potenzen 0, 1, 2, 3, 4, 5 oder 6 stehen, ohne dass zwei gleiche Syndrompolynome erscheinen.

Bei m = 4 sind es entsprechend $2^4 - 1 = 15$ Stellen, bei m = 5 werden es $2^5 - 1 = 31$ Stellen usw. Der zyklische Code für 1-Fehler-Korrektur ist also dem Hammingcode gleichwertig.