Zeitschrift: Bulletin des Schweizerischen Elektrotechnischen Vereins, des Verbandes Schweizerischer Elektrizitätsunternehmen = Bulletin de l'Association suisse des électriciens, de l'Association des entreprises électriques suisses

Herausgeber: Schweizerischer Elektrotechnischer Verein ; Verband Schweizerischer Elektrizitätsunternehmen

Band: 78 (1987)

Heft: 15

Artikel: Trellis-codierte Modulation

Autor: Ungerböck, G.

DOI: https://doi.org/10.5169/seals-903893

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 12.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Trellis-codierte Modulation

G. Ungerböck

Mit trellis-codierter Modulation lassen sich ohne Veränderung der Datenrate und Bandbreite Codierungsgewinne bis zu 6 dB gegenüber uncodierter Modulation erzielen. Die redundante Codierung wird durch Verwendung erweiterter, mehrwertiger Signalkonstellationen ermöglicht. Anwendungen zeichnen sich heute bei allen bandbegrenzten Übertragungskanälen ab, welche sich bei relativ hohem Signal-Geräuschabstand für die Übertragung mehrwertiger Modulationssignale eignen.

Par rapport à la modulation non codée, celle codée en trellis permet des gains de codage jusqu'à 6 dB, sans modification du taux de données et de la largeur de bande. Le codage redondant est possible avec de constellations polyvalentes de signaux élargies. Son emploi actuel concerne tous les canaux de transmission à bande limitée, qui conviennent à la transmission de signaux de modulation polyvalente et dont l'écart entre signal et bruit est relativement élevé.

Quellenangabe: Die Figuren entsprechen im wesentlichen den in [11; 12] verwendeten Darstellungen.

Adresse des Autors
Dr. *Gottfried Ungerböck*, dipl. El.-Ing., IBM-Forschungslaboratorium Zürich, 8803 Rüschlikon.

1. Einleitung

Bis vor wenigen Jahren hatte der Einsatz von Kanalcodierung zur Verminderung der Fehlerwahrscheinlichkeit digitaler Übertragungsstrecken in der Regel bei gleichbleibender Bandbreite einen Verlust an effektiver Datenrate zur Folge. Die wirkungsvollsten Methoden der Codierung und Decodierung waren darüber hinaus auf die Verwendung binärer Modulation oder Vierphasenmodulation, d.h. die Übertragung mit niedriger spektraler Effizienz (bit $\cdot$ s $^{-1}$Hz^{-1}) beschränkt. Die während der letzten Jahre bekanntgewordene Verfahren der trelliscodierten Modulation (TCM, trellis = Gitterwerk) erlauben nun die Erzielung von Codierungsgewinnen ohne Verminderung der Datenrate und ohne Bandbreitenerweiterung auch bei hohen Werten der spektralen Übertragungseffizienz. TCM-Verfahren gewinnen daher zunehmend an Bedeutung für die mehrwertige digitale Übertragung über bandbegrenzte Kanäle mit relativ hohem Signal-Geräuschabstand.

Bei TCM wird ein durch Codierung bedingter Verlust an Datenrate durch Vergrösserung der Anzahl der zur Modulation zugelassenen diskreten Signalwerte vermieden. Für TCM werden daher grundsätzlich höherwertige Modulationsverfahren benötigt (z.B. 8PSK anstatt 4PSK, wie später gezeigt wird). Bei gleicher Signalleistung und Bandbreite wie bei uncodierter Modulation gelingt es, die mögliche Aufeinanderfolge redundanter Modulationssignale durch Codierung so einzuschränken, dass die codierten Folgen sich im enklidischen Signalraum stärker voneinander unterscheiden als die Signalwertefolgen bei uncodierter Modulation mit gleicher Datenrate. Als Mass für die Unterscheidbarkeit von Signalwertefolgen dient die kleinste Summe der Abstandsquadrate ihrer Einzelsignalwerte (quadrierte euklidische Distanz). Ein dabei erzielter Distanzgewinn führt zu einem Codierungsgewinn, wenn im Empfänger mittels eines Maximum-Likelihood-Sequenz-Decodierers die am wahrscheinlichsten gesendete Signalwertefolge aus der Folge der empfangenen, verrauschten Signalwerte ohne vorangehende Einzelsignalentscheide bestimmt wird (Soft-MLSD).

Bereits mit sehr einfachen TCM-Codes kann ein Codierungsgewinn von 3 dB erzielt werden. Für höhere Komplexität sind TCM-Codes mit Gewinnen bis zu 6 dB bekannt. Diese Gewinne müssen mit einem Mehraufwand vor allem bei der Decodierung bezahlt werden. Der dafür erforderliche Aufwand liegt jedoch heute in vielen Fällen durchaus im Rahmen einer technologisch und wirtschaftlich sinnvollen Realisierung.

Einfache TCM-Codes wurden erstmals 1976 vorgestellt [1]. Einer 1982 erschienenen Arbeit, welche die informationstheoretischen Hintergründe und die Konstruktion optimaler TCM-Codes behandelte [2], folgten bald weitere Veröffentlichungen [3; ...; 6]. Innerhalb kurzer Zeit setzte sich TCM als neues bandbreiteneffizientes Modulationsverfahren für die Datenübertragung über Telefonkanäle durch. Diese Entwicklung wurde durch die Verfügbarkeit schneller digitaler Signalprozessoren (DSP) begünstigt, mit welchen TCM-Codierer und -Decodierer bei den für Telefonkanäle zulässigen Modulationsraten relativ leicht realisiert werden können. Während noch bis vor wenigen Jahren eine Datenrate von 9,6 kbit/s als praktisch obere Grenze für die zuverlässige Übertragung über Telefonkanäle galt, sind heute Modems kommerziell erhältlich, welche mittels TCM (und anderer Verbesserungen im Bereich der Entzerrung, Synchronisation usw.) Datenraten bis 19,2 kbit/s erreichen.

Begriffe

Code: Menge aller Folgen endlicher oder unendlicher Länge von Zeichen oder Signalwerten, welche gemäss einer Codierungsvorschrift erzeugt werden.

Coderate: Das Verhältnis von Informationszeichen zu Codezeichen.

Codierungsgewinn: Durch Codierung erreichte Verbesserung, welche erlaubt, gleiche Fehlerwahrscheinlichkeit bei niedrigerem Signal-Geräuschabstand zu erreichen.

Faltungscode: Code unendlicher Länge. Codefolgen entstehen durch sequentielle Umsetzung von k Folgen von Informationszeichen in n Folgen von Codezeichen. Das Verhältnis k/n (<1) ist die Coderate. Die Umsetzung entspricht in ihrem Charakter der Faltungsoperation in elektrischen Filtern. Die Codestruktur kann durch ein Zustandsdiagramm beschrieben werden. Bei linearen Faltungscodes besteht ein linearer Zusammenhang zwischen erzeugenden und erzeugten Folgen im Sinne der Rechenvorschriften einer endlichen Algebra, z.B. der Modulo-2-Arithmetik.

Binärer Faltungscode: Faltungscode mit binären Zeichenfolgen.

Einflusslänge v (constraint length): Bei Faltungscodierern ohne Rückkopplung für Codes der Rate $1/n$ ist dies die Anzahl der gespeicherten Eingangszeichen, welche zusammen mit dem neuesten Eingangszeichen Einfluss auf die erzeugten Ausgangszeichen haben. Ein binärer Faltungscodierer hat 2^v mögliche Zustände. Die Bezeichnung wird heute für beliebige Faltungscodes zur Angabe der Anzahl der Zustände verwendet.

Euklidische (Signal-)Distanz: Wurzel aus der Summe der Abstandsquadrate zweier Signalwertefolgen.

Freie euklidische Distanz: Kleinste euklidische Distanz zwischen codierten oder uncodierten Folgen von Signalwerten.

Hamming-Distanz: Anzahl der Zeichen, in welchen sich zwei Zeichenfolgen voneinander unterscheiden.

Minimale Hamming-Distanz: Kleinste Anzahl von Zeichen, in welchen sich verschiedene Zeichenfolgen eines Codes voneinander unterscheiden.

MLSD: siehe Kap. 4.2.

Hard-MLSD: Die Bestimmung der am wahrscheinlichsten gesendeten Folge von Informationszeichen aus einer Folge empfangener Codezeichen.

Soft-MLSD: Die direkte Bestimmung der am wahrscheinlichsten gesendeten Folge von Informationszeichen aus einer Folge empfangener Signalwerte ohne vorangehende Einzelzeichenentscheidungen.

Modulationsintervall: Zeitlicher Abstand zwischen zwei nacheinander gesendenten Modulationssignalen.

Modulationsrate: Anzahl der pro Sekunde gesendeten Modulationssignale.

Uncodierte Modulation: Modulation, bei der die Signalwerte der Signalkonstellation in uneingeschränkter Aufeinanderfolge gesendet werden können.

Codierte Modulation: Modulation, bei der die Signalwerte der Signalkonstellation nur in eingeschränkter Aufeinanderfolge gemäss einer Codevorschrift gesendet werden können.

Redundanz: Das Vorhandensein von zusätzlichen Beschreibungen derselben Information.

Signalabstand: siehe euklidische Distanz

Signalkonstellation: Menge der zur Modulation zugelassenen diskreten Signalwerte.

Signalraum: Euklidischer Raum, dessen Dimensionszahl durch die Gesamtzahl der Dimensionen der gesendeten oder empfangenen Signalwerte gegeben ist.

Signalwerte: Werte, mit denen ein Signal moduliert wird bzw. Werte von Signalen zu bestimmten Abtastzeitpunkten.

Spektrale Effizienz: Verhältnis der übertragenen Bitrate zur benötigten Bandbreite.

Viterbi-Algorithmus: von *A.J. Viterbi* 1967 erstmals vorgeschlagener effizienter Algorithmus für Maximum-Likelihood-Sequenz-Decodierung (Hard-MLSD und Soft-MLSD); dieser beruht auf dem schon früher bekannten Optimalitätsprinzip der Dynamischen Programmierung.

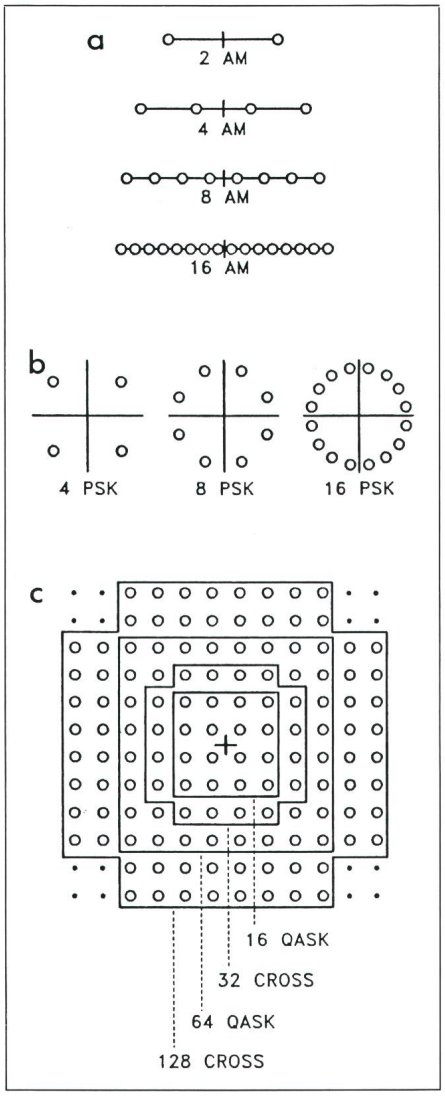
Bereits 1984 wurde die Verwendung eines TCM-Codes mit 8 Codezuständen und einem Codierungsgewinn von 4 dB in der CCITT-Empfehlung V.32 für die Vollduplex-Übertragung über das öffentliche Wahlnetz für die Rate von 9,6 kbit/s festgelegt [7]. Die zukünftige Anwendung von TCM für die Übertragung über andere wichtige bandbegrenzte Nachrichtenkanäle, z.B. im Bereich der Satellitenübertragung [8;...;10], des digitalen Richtfunks oder des Mobilfunks, lässt sich bereits mit grosser Sicherheit vorhersagen.

Der vorliegende Artikel soll als Einführung in das für potentielle Anwender noch ziemlich neue Gebiet der TCM dienen. Eine umfassendere Übersicht findet man in [11; 12].

2. Digitale Modulation/ Demodulation und klassische Kanalcodierung

Zur Übertragung von m bit pro Modulationsintervall T werden in einem konventionellen uncodierten Übertragungssystem diskrete Signalwerte einer 2^m-wertigen Konstellation, wie in Figur 1 für $m = 1$ bis 7 gezeigt, gesendet. Die Signalwerte können in beliebiger Folge auftreten. Im Empfänger werden von den zu den Abtastzeitpunkten empfangenen, verrauschten Signalwerten durch unabhängige, 2^m-wertige Entscheidungen wieder m bit pro Modulationsintervall T gewonnen. Gemäss Nyquistgesetz wird für die verzerrungsfreie Übertragung eindimensionaler Signalwerte mit der Modulationsrate $f_T = 1/T$ Baud – z.B. über einen Basisbandkanal – eine minimale Bandbreite von $B_{min} = f_T/2$ Hz benötigt (positive Frequenzanteile). Für die im folgenden vorzugsweise betrachtete Modulation eines sinusförmigen Trägersignals mit zweidimensionalen Signalwerten (Zweiseitenbandmodulation, Sonderfall: Phasenmodulation) ergibt sich ein minimaler Bandbreitenbedarf von $B_{min} = f_T$ Hz um die gewählte Trägerfrequenz. Daraus folgt bei Verwendung einer 2^m-wertigen Signalkonstellation eine theoretische mögliche spektrale Übertragungseffizienz E (bit · s^{-1} Hz^{-1}) von

$$E = m \cdot f_T / B_{min} = m \qquad (1)$$



Figur 1 Signalkonstellationen

a eindimensionale Amplitudenmodulation (Basisbandübertragung, Einseitenbandmodulation)
b Phasenmodulation (Sonderfall von c)
c zweidimensionale Amplituden- und Phasenmodulation (Zweiseitenbandmodulation, Quadraturmodulation)

Die klassischen Verfahren der Kanalcodierung/-decodierung dienen zur Erkennung bzw. Korrektur von diskreten Übertragungsfehlern konventioneller digitaler Übertragungssysteme. Für die Codierung und Decodierung werden nur die übertragenen diskreten Zeichen in Betracht gezogen. Die meisten der bekannten Faltungs- und Blockcodes sind für binäre Codezeichen (Bits) definiert (Ausnahme: Reed-Solomon-Codes). Mit einem Code der Rate k/n, wobei $k < n$, werden zusätzlich zu je k gesendeten Informationszeichen $n-k$ Prüfzeichen übertragen. Die Anzahl der Codezeichen, in welchen sich zwei Codewörter oder -folgen voneinander unterscheiden, wird als Hamming-Distanz bezeichnet. Solange bei strikter Trennung von Codierung/Decodierung und Modulation/Demodulation einem Decodierer nur die Information der empfangenen Codezeichen zur Verfügung steht, bleibt die Hamming-Distanz das für den Codeentwurf und die Entscheidungen des Decodierers massgebende Kriterium. Eine minimale Hamming-Distanz d_{Hmin} garantiert, dass mindestens $d_{\mathrm{Hmin}}-1$ Fehler erkannt und mindestens $[(d_{\mathrm{Hmin}}-1)/2]$[1] Fehler korrigiert werden können.

Die Verwendung klassischer Methoden der Kanalcodierung führt notwendigerweise zu einer Reduktion der Datenrate im Ausmass der Coderate. Wenn der analoge Kanal eine Bandbreitenerweiterung zulässt, kann der Coderatenverlust durch Erhöhung der Modulationsrate kompensiert werden. Dabei ist allerdings zu berücksichtigen, dass bei vorgegebener Signalleistung die pro Modulationsintervall zur Verfügung stehende Signalenergie abnimmt. Für die hier angestellten Betrachtungen soll jedoch jegliche Bandbreitenerweiterung ausgeschlossen sein. Somit bleibt als weitere Möglichkeit zur Kompensation eines Coderatenverlustes nur die Wahl einer höherwertigen Signalkonstellation, damit bei gleichbleibender Rate der Informationszeichen die Prüfzeichen eines Codes innerhalb derselben Bandbreite mitübertragen werden können. Im Prinzip entspricht dies dem für TCM eingeschlagenen Weg. Die klassischen Methoden der getrennten digitalen Modulation/Demodulation und fehlerkorrigierenden Codierung/Decodierung haben sich jedoch in diesem Fall nicht bewährt.

Zur Illustration dieser Aussage soll hier der Fall von uncodierter Vierphasenmodulation (4PSK, Phase-Shift-Keying) und Achtphasenmodulation (8PSK) mit konventioneller fehlerkorrigierender Kanalcodierung mit einer Coderate von 2/3 betrachtet werden. Beide Systeme übertragen 2 Datenbits pro Modulationsintervall (2 bit · s^{-1} · Hz^{-1}). Es sei der Signal-Geräuschabstand so vorgegeben, dass das uncodierte 4PSK-System eine Fehlerwahrscheinlichkeit von 10^{-5} erreicht. Wegen des geringeren Abstandes zwischen den 8PSK-Signalen beträgt die Fehlerwahrscheinlichkeit des 8PSK-Systems ohne Codierung dann etwa 10^{-2}. Durch klassische Kanalcodierung/-decodierung kann diese erhöhte Fehlerwahrscheinlichkeit mit vertretbarem Aufwand nicht so weit reduziert werden, dass ein nennenswerter Codierungsgewinn gegenüber dem uncodierten 4PSK-System entsteht. Damit beide Systeme ungefähr gleiche Fehlerwahrscheinlichkeit aufweisen, müssen bei dem 8PSK-System bereits Muster von bis zu 3 Fehlern korrigiert werden, z.B. mittels eines binären Faltungscodes der Rate 2/3 mit $d_{\mathrm{Hmin}} = 7$ und einer Einflusslänge v (constraint length) von 6 und durch Verwendung eines Viterbi-Decodierers (Hard-MLSD) mit 64 Zuständen [13].

3. Der Weg zur trellis-codierten Modulation

Für den Misserfolg der fehlerkorrigierenden Codierung in Verbindung mit redundanter, mehrwertiger Modulation lassen sich zwei Gründe angeben. Ein erster Grund ist im Informationsverlust zu sehen, welcher durch die Trennung der Empfängerfunktionen in unabhängige, diskrete Demodulation einzelner Signalwerte und nachfolgende fehlerkorrierende Decodierung entsteht. Dieser Verlust kann durch Sequenz-Decodierung der empfangenen Signalwerte ohne Demodulation einzelner Signalwerte vermieden werden.

Es sei $\{r_n\} = \{a_n + w_n\}$ eine zeitlich unbegrenzte Folge von ein- oder zweidimensionalen Signalwerten, welche im Empfänger (bei Trägermodulation nach Verschiebung des empfangenen Signals ins Basisband) zu den Abtastzeitpunkten empfangen werden. Dabei bezeichnet a_n einen zum n-ten Abtastzeitpunkt verzerrungsfrei empfangenen Signalwert aus der im Sender verwendeten Signalkonstellation und w_n den überlagerten Wert eines als weisses Gaussches Rauschen angenommenen Störsignals. Die Entscheidungsregel eines optimalen Maximum-Likelihood Sequenz-Decodierers (Soft-MLSD) besagt, dass unter allen Folgen $\{a'_n\}$, welche der Sender durch Codierung und Modulation erzeugen kann, die Folge $\{\hat{a}_n\}$ ausgewählt wird, welche den kleinsten euklidischen Abstand, d.h. die kleinste Summe der quadrierten Abstände von der Folge $\{r_n\}$, aufweist. Somit muss gelten

[1] [x] grösster ganzzahliger Wert, der in x enthalten ist.

$$\sum_n |r_n - \hat{a}_n|^2 = \min_{\{a'_n\} \varepsilon C} \left[\sum_n |r_n - a'_n|^7 \right] \quad (2)$$

wobei C die Menge aller möglichen Sendefolgen $\{a_n\}$ darstellt. Wenn sich diese Folgen durch ein Zustandsdiagramm (Trellis-Diagram) beschreiben lassen, d.h. wenn sie nach dem Prinzip eines endlichen Automaten erzeugt werden, kann der Viterbi-Algorithmus [13; 14] zur effizienten Bestimmung der Folge $\{\hat{a}_n\}$ – und damit der gesendeten Datenfolge – verwendet werden (Kap. 4.2). Wie später gezeigt wird, werden dabei gültige Folgen von Signalwerten iterativ in Richtung der positiven Zeitachse ausgedehnt und die Werte der Folge $\{\hat{a}_n\}$ mit einer endlichen Decodierungsverzögerung daraus abgeleitet.

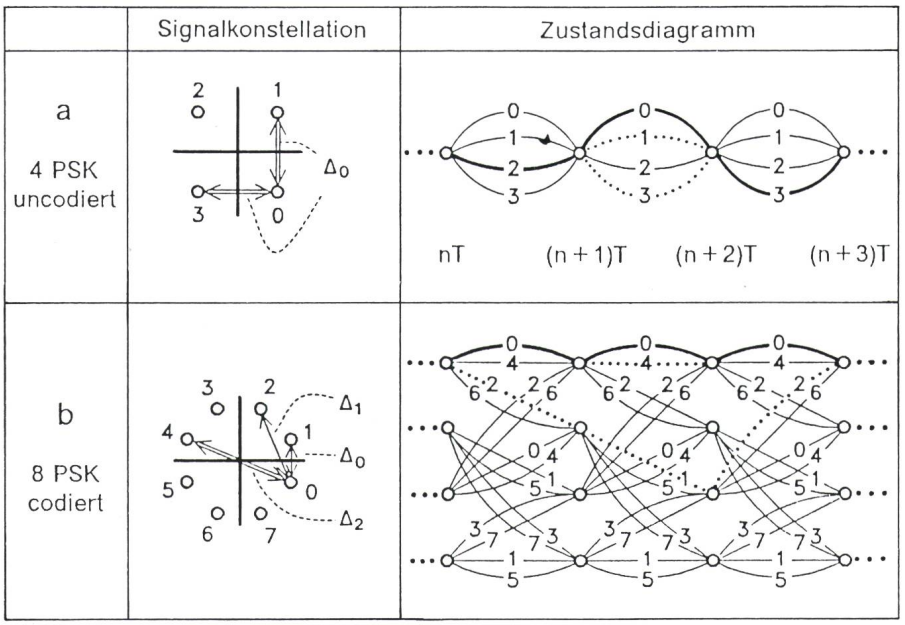
Fehlerereignisse treten bei einem Soft-MLSD-Empfänger in Form einer Vertauschung einer gesendeten Folge $\{a_n\}$ mit einer anderen möglichen Folge $\{a'_n\}$ auf. Die wahrscheinlichsten Fehler ereignen sich zwischen Folgen mit geringstem Abstand. Dieser Abstand wird als freie euklidische Distanz d_{free} bezeichnet:

$$d^2_{\text{free}} = \min_{\{a'_n\} \neq \{a''_n\}} \left[\sum_n |a'_n - a''_n|^2 \right] \quad (3)$$

Folgen mit dem Abstand d_{free} unterscheiden sich nur in wenigen aufeinanderfolgenden Signalwerten und in bestimmten Fällen – ähnlich wie bei uncodierter Modulation – nur in einem Signalwert. Wegen der stationären Eigenschaften der betrachteten Signalcodes kann man sich bei der Bestimmung von d_{free} im allgemeinen auf Fehlerereignisse beschränken, welche zum Zeitpunkt $n = 0$ beginnen.

Hat man sich für einen optimalen Soft-MLSD-Empfänger entschieden, so bedeutet dies noch nicht, dass nun Codierungsgewinne erzielt werden. Die Verwendung von Codes mit optimaler Hamming-Distanz ist nur dann sinnvoll, wenn Codezeichen so in Modulationssignale umgesetzt werden, dass Äquivalenz besteht zwischen Hamming-Distanz und quadrierter euklidischer Distanz. Dies trifft für binäre Codes in Verbindung mit binärer Modulation oder 4PSK-Modulation zu, wobei letztere als binäre Modulation eines Trägersignals in zwei orthogonale Richtungen anzusehen ist. Solche Systeme haben sich seit den späten 60er Jahren für die leistungseffiziente Übertragung mit einer spektralen Effizienz von $\leqslant 2$ bit · $s^{-1}Hz^{-1}$ gut bewährt. Für mehrwertige Signalkonstellationen lässt sich jedoch kein Abbildungsverfahren angeben, welches zu einem proportionalen Zusammenhang zwischen Hamming-Distanz und quadrierter euklidischer Distanz führt. Es ist im allgemeinen sogar unmöglich, einen monotonen Zusammenhang zu finden. Distanzverluste bei der Abbildung bekannter, für Hamming-Distanz optimierter Codes in mehrwertige Signalfolgen standen lange Zeit der Anwendung von Kanalcodierung in Verbindung mit mehrwertiger Modulation im Wege.

Das Problem wurde schliesslich dadurch gelöst, dass es gelang, Signalcodes direkt im Hinblick auf maximale freie euklidische Distanz ohne Bezugnahme auf Hamming-Distanz zu entwerfen. Heuristische Entwurfsversuche führten schon bei geringer Komplexität zu überraschend guten Resultaten. Bei der Suche nach geeigneten Codes spielte auch die Berechnung der Kanalkapazität für mehrwertige Signalkonstellationen unter der Annahme eines optimalen MLSD-Empfängers eine hilfreiche Rolle [2]. Diese Untersuchung zeigte, dass man sich bei der codierten Übertragung von m bit pro Modulationsintervall auf die Verwendung redundanter 2^{m+1}-wertiger Signalkonstellationen beschränken kann. Mit der so gewonnenen Redundanz können theoretisch Codierungsgewinne von 7 bis 8 dB gegenüber uncodierter Modulation erreicht werden.



Figur 2 Signalkonstellation und Zustandsdiagramm

a uncodierte Vierphasenmodulation (4PSK): $\Delta_0 = \sqrt{2} = d_{\text{free}}$

b trellis-codierte Achtphasenmodulation (8PSK) mit vier Codezuständen: $\Delta_0 = 2 \cdot \sin(\pi/8)$, $\Delta_1 = \sqrt{2}$, $\Delta_2 = 2 = d_{\text{free}}$

4. Beispiel: Trellis-codierte 8PSK-Modulation

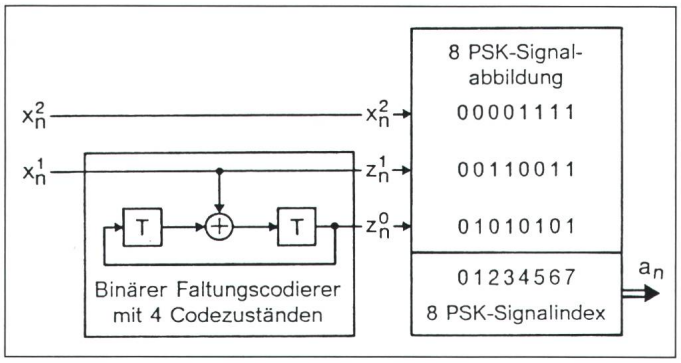
4.1 Trellis-Code mit vier Zuständen

Der im folgenden vorgestellte Code für 8PSK-Modulation wurde 1975 als erster Trellis-Code mit einem nennenswerten Codierungsgewinn gegenüber uncodierter 4PSK-Modulation gefunden. Die Figur 2 zeigt die Signalkonstellationen und Zustandsdiagramme für uncodierte 4PSK-Modulation und trellis-codierte 8PSK-Modulation. Das für uncodierte Modulation gezeigte, unübliche Zustandsdiagramm mit nur einem Zustand (und damit keinem Zustandsspeicher, der die Folge der 4PSK-Signale einschränken könnte) soll zur Verdeutlichung des Unterschiedes zur trellis-codierten Modulation beitragen, bei welcher die jeweils zugelassenen Signalwerte von dem zuletzt erreichten Codezustand abhängen. Jede zusammenhängende Folge von Übergängen in einem Zustandsdiagramm, das in Figur 2 für einen Ausschnitt von drei Modulationsintervallen angegeben ist, ergibt eine erlaubte Folge von Sendesignalen. Da von jedem Zustand vier Übergänge zu einem Folgezustand möglich sind, kommt der Auswahl eines bestimmten Übergangs die Bedeutung des Sendens von 2 Datenbits zu. Die spektrale Effizienz beträgt somit in beiden Fällen (4PSK uncodiert und 8PSK codiert) 2 bit · $s^{-1}Hz^{-1}$.

Im Falle der uncodierten 4PSK-Modulation sind beliebige Folgen der vier Modulationssignale mit den Signalindizes 0, 1, 2, 3 zulässig. Dies wird im Zustandsdiagramm durch vier parallele Übergänge dargestellt. Die freie euklidische Distanz zwischen zwei Folgen ergibt sich hier trivial als die minimale Distanz zwischen einzelnen 4PSK-Signalen, d.h. $d_{free} = \Delta_0$ (4 PSK) $= \sqrt{2}$. Jedes 4PSK-Signal hat zwei Nachbarsignale mit dieser Distanz.

Im Falle der trellis-codierten 8PSK-Modulation treten Übergänge in Paaren von je zwei parallelen Übergängen auf. Den parallelen Übergängen sind 8PSK-Signale mit maximaler Signaldistanz Δ_2 (8 PSK) = 2 zugeordnet. Für jeweils vier von einem Zustand ausgehende bzw. sich in einem Zustand vereinigende Übergänge werden Signale mit minimaler Signaldistanz Δ_1 (8PSK) = Δ_0 (4PSK) = $\sqrt{2}$ verwendet. Der kleinste Abstand zwischen den 8PSK-Signalen beträgt Δ_0 (8PSK) = 2 · sin ($\pi/8$). Alle 8PSK-Signale kommen in dem Zustandsdiagramm mit gleicher Häufigkeit vor. Für diesen 8-PSK-Code lässt sich zeigen, dass $d_{free} = \Delta_2$ (8PSK) = 2. Dieser Abstand ergibt sich zwischen zwei Folgen von 8PSK-Signalen, welche sich nur in einem einzigen parallelen Übergang unterscheiden (Beispiel: Folgen mit Signalindizes .., 0, 0, 0, .. und .., 0, 4, 0,...). Folgen, welche von einem Zustand zu verschiedenen Folgezuständen führen und sich frühestens nach drei Übergängen wieder in einem Zustand vereinigen können, haben wenigstens den quadrierten Abstand Δ_1^2 (8PSK) + Δ_0^2 (8PSK) + Δ_1^2 (8PSK), welcher grösser ist als Δ_2^2 (8PSK) (Beispiel: Folgen mit Signalindizes .., 0, 0, 0, 0, 0, .. und .., 0, 2, 1, 2, 0,..). Folglich gilt $d_{free} = 2$. Dies entspricht einem Distanzgewinn von 3 dB gegenüber uncodierter 4-PSK-Modulation bei gleicher Signalleistung, Bandbreite und Datenrate. Für jede Signalfolge existiert zu jedem Zeitpunkt nur ein Nachbarsignal mit Distanz d_{free}. Die Figur 3 zeigt eine mögliche Senderrealisierung bestehend aus einem binären Faltungscodierer (in Form eines systematischen Codierers mit Rückkopplung [2]) und einem 8PSK-Modulator.

Figur 3
Realisierung eines Codierers und Modulators für trellis-codierte Achtphasenmodulation (8PSK) mit vier Codezuständen



4.2 Maximum-Likelihood-Sequenz-Decodierung

Die Decodierung von Trellis-Codes erfolgt im allgemeinen in zwei Schritten. Im ersten Schritt, Subset-Decoding genannt, wird zu jedem Zeitpunkt n innerhalb der Untermengen (Subsets) von Signalwerten, welche parallelen Zustandsübergängen zugeordnet sind (beim Code von Figur 2b sind dies die Signalpaare mit Indizes 0,4; 1,5; 2,6 und 3,7), der Signalwert mit dem kleinsten Abstand von dem empfangenen Signalwert r_n bestimmt. Diese für Untermengen getroffenen Entscheidungen sind endgültig und entsprechen in ihrem Charakter den unabhängigen Entscheidungen eines diskreten Demodulators bei uncodierter Modulation. Die Subset-Entscheidungen werden zusammen mit den quadrierten Abständen zwischen den ausgewählten Signalwerten und r_n zur weiteren Verwendung zwischengespeichert. In einem weiteren Schritt wird die am wahrscheinlichsten gesendete Signalfolge $\{\hat{a}_n\}$, d.h. die Folge mit der geringsten Summe der Abstandsquadrate von der Signalfolge $\{r_n\}$, mit Hilfe des Viterbi-Algorithmus bestimmt.

Beschreibungen des Viterbi-Algorithmus findet man in allen neueren Standardwerken über Codierung, z.B. in [13]. Die wesentlichen Punkte seien hier kurz zusammengefasst. Es wird davon ausgegangen, dass zum Zeitpunkt n für jeden Codezustand die aus der unendlichen Vergangenheit zu diesem Zustand führende Signalfolge mit minimaler Summe der Abstandsquadrate von r_k, $k = -\infty$ bis $n-1$ bekannt ist. Der Algorithmus verwendet die durch Subset-Decoding gewonnenen Abstandsquadrate von r_n zur Fortsetzung je eines optimalen Pfades zu den Zuständen zum Zeitpunkt $n+1$. Mögliche andere Pfade mit grösserer Abstandssumme werden dabei «vergessen». Die ausgewählten Folgen haben die Eigenschaft, sich in Richtung der Vergangenheit zur Folge $\{\hat{a}_n\}$ zu vereinigen. Daraus ergibt sich die Möglichkeit, zum Zeitpunkt n durch Zurückverfolgen einer Folge bis zu einem Zeitpunkt $n-D$ ein verzögertes Ausgangssignal $\hat{a}_{n-D}$ bzw. die damit übertragene Dateninformation zu bestimmen.

4.3 Fehlerwahrscheinlichkeit

Ein Fehlerereignis wird dann gezählt, wenn die decodierte Folge $\{\hat{a}_n\}$ von der gesendeten Folge $\{a_n\}$ abzuweichen beginnt. Bei uncodierter Modulation und bei TCM-Codes mit parallelen Übergängen kann die Abweichung nach einem Modulationsintervall beendet sein. Wenn die empfangene Signalwertefolge durch additives unkorreliertes Gausssches Rauschen mit der Varianz σ^2 in jeder Signaldimension gestört ist, ergibt sich für die Fehlerwahrscheinlichkeit in guter Näherung bei grossem Signal-Geräuschabstand

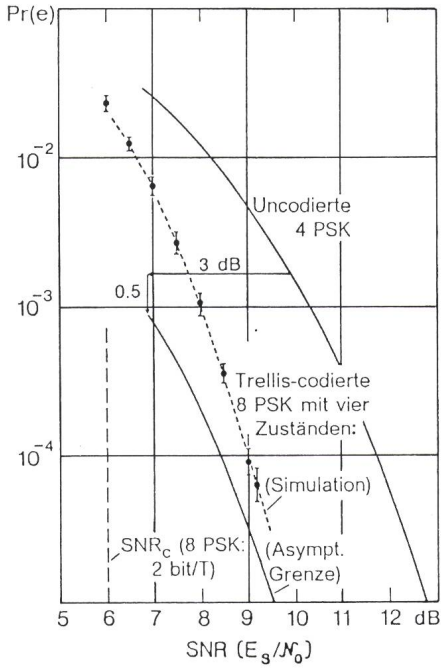
$$Pr(e) \simeq N_{free} \cdot Q(d_{free}/2\sigma) \qquad (4a)$$

wobei $Q(.)$ durch

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-y^2/2)\, dy \qquad (4b)$$

definiert ist. N_{free} bezeichnet die Anzahl der von einem Zustand ausgehenden möglichen Fehlerfolgen mit Distanz d_{free}. In dieser Näherungsformel kommt zum Ausdruck, dass bei hohem Signal-Geräuschabstand die Wahrscheinlichkeit von Fehlerereignissen mit grösserer Distanz als d_{free} vernachlässigbar klein wird (gilt vor allem sehr gut für uncodierte Modulation).

Für uncodierte 4PSK-Modulation ist $d_{free} = \sqrt{2}$ und $N_{free} = 2$. Für trelliscodierte 8PSK-Modulation mit vier Zuständen wurde $d_{free} = 2$ gefunden. Die häufigsten Fehler sind Einzelfehler mit $N_{free} = 1$. Die Figur 4 zeigt $Pr(e)$ als Funktion des Signal-Geräuschabstandes. Die durch Simulation gewonnene Kurve soll zeigen, dass für trelliscodierte Modulation bei niedrigem Signal-Geräuschabstand die Wahrscheinlichkeit des Auftretens von Fehlerereignissen mit grösserer Distanz als d_{free} nicht völlig vernachlässigbar ist. Im wesentlichen bestätigt jedoch



Figur 4 Wahrscheinlichkeit von Fehlerereignissen *Pr(e)* als Funktion des Signal-Geräusch-Abstandes *SNR(E/N₀)* für uncodierte Vierphasenmodulation (4PSK) und trellis-codierte Achtphasenmodulation (8PSK) mit vier Codezuständen

SNR_0 (8PSK: 2 bit/T) = 6 dB bezeichnet den Wert des Signal-Geräusch-Abstandes, bei welchem mit 8PSK-Modulation und Soft-MLSD die Kanalkapazität den Wert von 2 bit pro Modulationsintervall erreicht.

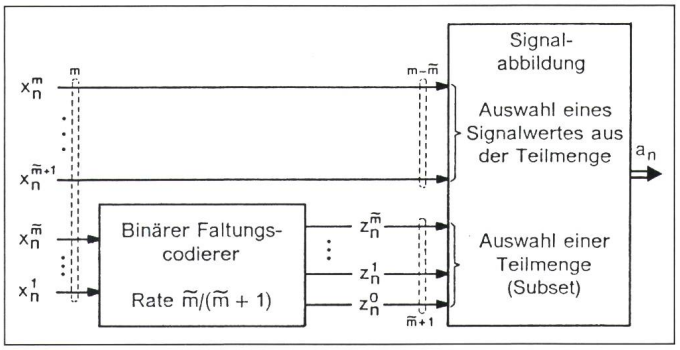
Figur 4 den erwarteten Codierungsgewinn von 3 dB.

Bei TCM-Codes mit mehr als vier Zuständen tritt d_{free} meistens nicht zwischen parallelen Zustandsübergängen auf. Die wahrscheinlichsten Fehlerereignisse erstrecken sich dann über mehrere Modulationsintervalle – z.B. beim CCITT-Code mit acht Zuständen über drei oder vier Modulationsintervalle [11]. Die Werte von N_{free} werden im allgemeinen grösser und von der gesendeten Signalfolge abhängen, so dass es sinnvoll ist, für N_{free} einen Mittelwert anzugeben.

5. Systematischer Entwurf von TCM-Codes

TCM-Codes mit bis zu 16 Zuständen konnten noch in heuristischer Weise gefunden werden. Für den Entwurf von Codes mit höherer Komplexität ist jedoch die Kenntnis allgemeiner Struktureigenschaften unerlässlich. Versuche, heuristisch gefundene TCM-Codes als binäre Faltungscodes mit einer speziellen Abbildung in mehrwertige Modulationssignale zu verstehen, führten zu der in Figur 5 dargestellten Senderstruktur. Für die

Figur 5 Allgemeine Struktur des Codierers und Modulators für trellis-codierte Modulation



trellis-codierte Übertragung vom m Datenbits pro Modulationsintervall werden zunächst $\tilde{m} \leqslant m$ Datenbits $x_n^{\tilde{m}}$, ... x_n^1 durch einen binären Faltungscodierer zu einem $(\tilde{m}+1)$-Codebittupel $z_n = [z_n^{\tilde{m}}, \ldots z_n^0]$ erweitert. Diese Bits wählen eine $2^{m-\tilde{m}}$-wertige Untermenge von Signalen aus einer 2^{m+1}-wertigen Signalkonstellation aus. Die verbleibenden uncodierten Datenbits $x_n^m, \ldots x_n^{\tilde{m}+1}$ bestimmen, welches Signal aus dieser Untermenge gesendet wird.

5.1 Set-Partitioning

Die Figur 6 stellt die wesentlichen Aspekte der Aufteilung einer Signalkonstellation in Untermengen am Beispiel einer 32CROSS-Signalkonstellation dar. Die Aufteilung wird schrittweise so vorgenommen, dass die minimale Distanz Δ_i, $i = 0, 1, 2, \ldots$ zwischen den in den Untermengen enthaltenen Signalen bei jeder Teilung maximal zunimmt. Die Aufteilung wird solange fortgesetzt, bis nach $\tilde{m}+1$ Teilungen die Distanz $\Delta_{\tilde{m}+1}$ in den zuletzt erhaltenen Untermengen den gewünschten Wert von d_{free} erreicht oder übersteigt.

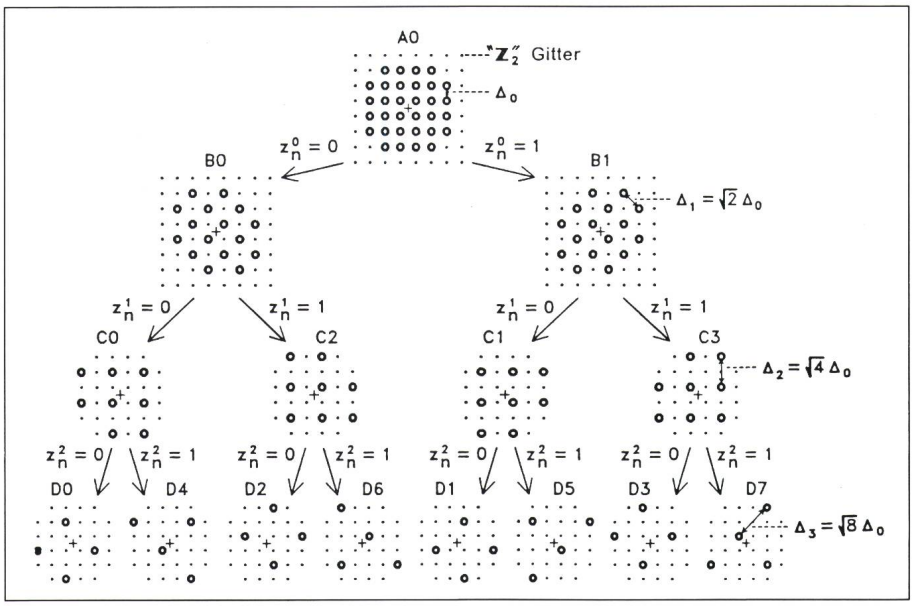
Die Auswahl dieser Untermengen durch die $(\tilde{m}+1)$-Codebittupel z_n erfolgt gemäss Figur 6. Im Zustandsdiagramm des gesuchten TCM-Codes werden die Signalwerte dieser Untermengen parallelen Zustandsübergängen zugeordnet. Damit ergibt sich für die freie euklidische Distanz

$$d_{\text{free}} = \text{Min}\,[\Delta_{\tilde{m}+1}, d_{\text{free}}(\tilde{m})] \qquad (5)$$

wobei $d_{\text{free}}(\tilde{m})$ den kleinsten Abstand zwischen codierten Signalfolgen bezeichnet, welche sich in mehr als einem Zustandsübergang voneinander unterscheiden.

5.2 Faltungscodes für TCM

Die für TCM benötigten binären Faltungscodes der Rate $\tilde{m}/(\tilde{m}+1)$ unterscheiden sich grundlegend von für Hamming-Distanz optimierten Codes. Mit einer vorgegebenen Anzahl von 2^v Codezuständen soll nicht d_{Hmin} maximiert werden, sondern der grösstmögliche Wert von $d_{\text{free}}(\tilde{m})$ erreicht werden. Die Suche nach solchen Codes



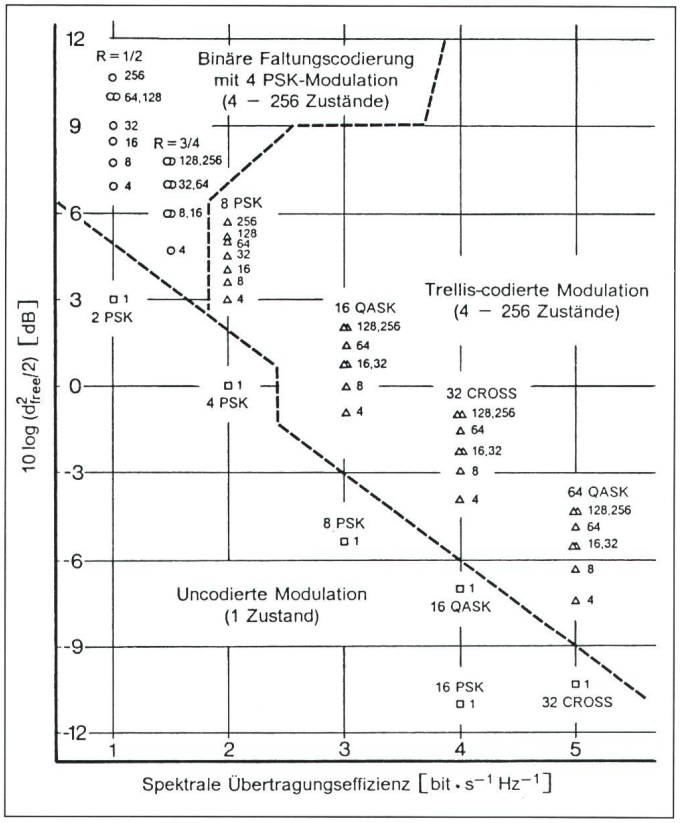
Figur 6 Set-Partitioning, gezeigt am Beispiel einer 32CROSS-Signalkonstellation

wird durch die in [2] abgeleitete, im allgemeinen mit Gleichheit erfüllte Beziehung

$$d^2_{\text{free}}(\tilde{m}) \geq \min_{\{z_n\} \neq \{0\}} \sum_n \Delta^2(z_n) \qquad (6)$$

wesentlich erleichtert. Die quadrierten Distanzwerte $\Delta^2(z_n)$ sind als euklidische Gewichte der $(\tilde{m}+1)$-Codebittupel z_n zu verstehen. Z.B. ist $\Delta(z_n = [....,1]) = \Delta_0$, $\Delta(z_n = [...1,0]) = \Delta_1$, $\Delta(z_n = [..1, 0, 0]) = \Delta_2$ usw., und $\Delta(0) = 0$. Diese Gewichte werden ähnlich wie Hamming-Gewichte bei der Suche nach geeigneten linearen Faltungscodes verwendet. Die Minimierung in der oben gezeigten (Un)Gleichung erfolgt über alle von der Nullfolge verschiedenen binären Codefolgen $\{z_n\}$.

Optimale TCM-Codes für praktisch relevante Signalkonstellationen sind in der Literatur in tabellarischer Form angegeben, z.B. in [12]. Die durch TCM-Codes mit 4 bis 256 Zuständen und zweidimensionalen Signalkonstellationen erzielten Distanzgewinne gegenüber uncodierter Modulation sind in Figur 7 graphisch dargestellt. Es werden auch die entsprechenden Werte für binäre Faltungscodierung mit 4PSK-Modulation gezeigt. Die Figur soll die Bedeutung von TCM als Methode zur Erzielung von Codierungsgewinnen bei Werten der spektralen Übertragungseffizienz von ≥ 2 bit $\cdot$ $s^{-1}Hz^{-1}$ unterstreichen, für welche bisher keine wirkungsvollen Methoden der Kanalcodierung zur Verfügung standen.



Figur 7
Freie euklidische Distanz und spektrale Effizienz für Zweiseitenbandmodulationssysteme: ohne Codierung, mit binärer Faltungscodierung und 4PSK-Modulation (< 2 bit $\cdot$ $s^{-1}Hz^{-1}$) und mit Trellis-Codierung (≤ 2 bit $\cdot$ $s^{-1}Hz^{-1}$)
Die Distanzwerte sind relativ zu d_{free} (4PSK uncodiert) = $\sqrt{2}$ angegeben.

6. Schlussbemerkungen

Mit der trellis-codierten Modulation wurde das Prinzip der Faltungscodierung und Maximum-Likelihood-Sequenzdecodierung auf das Gebiet der mehrwertigen Modulation ausgedehnt. Es wurde dadurch möglich, nennenswerte Codierungsgewinne auch bei hoher spektraler Übertragungseffizienz zu erzielen. Damit konnte eine bedeutsame Lücke auf dem Gebiet der Codierung und Modulation geschlossen werden.

Mit einem Trellis-Code mit 4 Zuständen wird im allgemeinen ein Codierungsgewinn von 3 dB gegenüber uncodierter Modulation erreicht, mit 8 Zuständen kommt ein Gewinn von 4 dB (CCITT-Code) zustande, mit 16 Zuständen ein Gewinn von 4,8 dB usw.

In Modems für die Datenübertragung über Telefonkanäle werden bereits TCM-Codes mit bis zu 64 Codezuständen und mit zwei- bis achtdimensionalen Signalkonstellationen praktisch verwendet. Für Satellitenübertragung und digitalen Richtfunk werden Realisierungen von TCM-Verfahren für Übertragungsraten bis 140 Mbit/s in Betracht gezogen.

Eine ausführlichere Diskussion von Trellis-Codes höherer Komplexität und eine Behandlung der multidimensionalen Trellis-Codierung hätte den Rahmen dieses Artikels gesprengt. Ebenso war es nicht möglich, auf das für manche Anwendungen wichtige Problem der Rotationssymmetrie von Trellis-Codes und die Möglichkeit von phasendifferentieller Codierung einzugehen. Der interessierte Leser sei hier wieder auf [11; 12] hingewiesen, wo auch weitere Literaturhinweise zu finden sind.

Literatur

[1] *G. Ungerböck* and *I. Csajka:* On improving data link performance by increasing the channel alphabet and introducing sequence coding. 1976 International Symposium on Information Theory, Ronneby/Sweden, June 21...24, 1976.

[2] *G. Ungerböck:* Channel coding with multilevel/phase signals. IEEE Trans. IT 28(1982)1, p. 55...67.

[3] *G.D. Forney* a.o.: Efficient modulation for band-limited channels. IEEE J. SAC 2(1984)5, p. 632...647.

[4] *L.-F. Wei:* Rotationally invariant convolutional channel coding with expanded signal space. Part I: 180°. IEEE J. SAC 2(1984)5, p. 659...671.

[5] *L.-F. Wei:* Rotationally invariant convolutional channel coding with expanded signal space. Part II: Nonlinear codes. IEEE J. SAC 2(1984)5, p. 672...686.

[6] *R. Calderbank* and *J.E. Mazo:* A new description of trellis codes. IEEE Trans. IT 30(1984)6, p. 784...791.

[7] A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits. Recommendation V.32. Red book. Volume VIII.1: Data communication over the telephone network. VIIIth CCITT plenary assembly Malaga-Torremolinos 8...19 October 1984.

[8] *G. Ungerböck, J. Hagenauer* and *T. Abdel-Nabi:* Coded 8-PSK experimental modem for the "INTELSAT SCPC system". Proceedings of the 7th International Conference on Digital Satellite Communications, Munich, May 12...16, 1986; p. 299...304..

[9] *R.J.F. Fang:* A coded 8-PSK system for 140-Mbit/s information rate transmission over 80-MHz nonlinear transponders. Proceedings of the 7th International conference on Digital Satellite Communications, Munich, May 12...16, 1986; p. 305...313.

[10] *T. Fujino* a.o.: A 120 Mbit/s 8PSK modem with soft-Viterbi decoding. Proceedings of the 7th International Conference on Digital Satellite Communications, Munich, May 12...16, 1986; p. 315...321.

[11] *G. Ungerböck:* Trellis-coded modulation with redundant signal sets. Part I: Introduction. IEEE Communications Magazine 25(1987)2, p. 5...11.

[12] *G. Ungerböck:* Trellis-coded modulation with redundant signal sets. Part II: State of the art. IEEE Communications Magazine 25(1987)2, p. 12...21.

[13] *G.C. Clark* and *J.B. Cain:* Error-correction coding for digital communications. New York/London, Plenum Press, 1981.

[14] *G.D. Forney:* The Viterbi algorithm. Proc. IEEE 61(1973)3, p. 268...278.

SIEMENS-ALBIS
AKTIENGESELLSCHAFT

Der Schlüssel zur Bürokommunikation:

Mit Hicom ist es möglich, über ein einziges Leitungssystem und über eine einzige Nummer zu telefonieren, teletexen, fernkopieren, videotexen, computern.

Beispiel Videotex-Komforttelefon ALBIS-COMTEL 3210

Eine neue Aera der Kommunikationstechnik ist angebrochen; mit Videotex ist aus dem Telefon ein intelligentes Kommunikations- und Informationssystem geworden.
ALBIS-COMTEL 3210®, das Videotex-Komforttelefon von Siemens-Albis bietet Ihnen eine Fülle von Annehmlichkeiten, von denen Sie bisher kaum zu träumen wagten: Sie haben im Privat- wie im Geschäftsbereich den direkten Draht zu Informationen und Zugriff zu vielen Dienstleistungen – praktisch eine eigene Datei, eine eigene Auskunft, einen eigenen Anschluss an Computer und Datenbanken und vieles mehr.

Gerne informieren wir Sie eingehender über Videotex im allgemeinen und unser ALBIS-COMTEL 3210 im besonderen.

PTT

Für die PTT entwickelt und hergestellt von Siemens-Albis

☐ Senden Sie mir Unterlagen über das Videotex-Komforttelefon ALBIS-COMTEL 3210.

☐ Rufen Sie mich an.

Name: _____ Vorname: _____

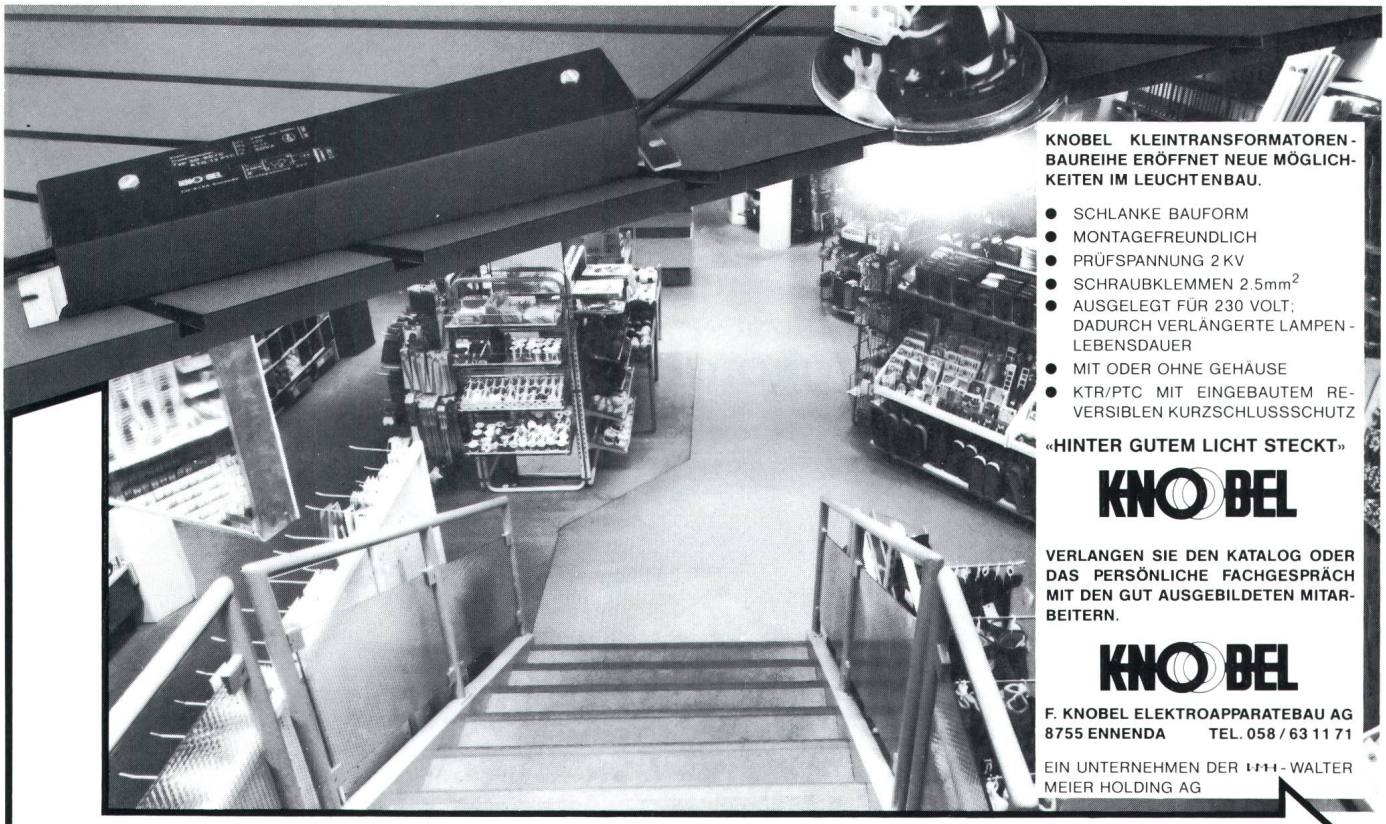
In Firma: _____

_____

Telefon: _____

Nähere Auskünfte und Unterlagen erhalten Sie von Ihrer Fernmeldekreisdirektion, Tel. 113, oder von
Siemens-Albis Aktiengesellschaft
Information 1
8047 Zürich, Freilagerstrasse 40
Telefon 01-495 3111

Computer + Communication von Siemens-Albis



KNOBEL KLEINTRANSFORMATOREN-BAUREIHE ERÖFFNET NEUE MÖGLICHKEITEN IM LEUCHTENBAU.

- SCHLANKE BAUFORM
- MONTAGEFREUNDLICH
- PRÜFSPANNUNG 2 KV
- SCHRAUBKLEMMEN 2.5mm²
- AUSGELEGT FÜR 230 VOLT; DADURCH VERLÄNGERTE LAMPEN-LEBENSDAUER
- MIT Oder OHNE GEHÄUSE
- KTR/PTC MIT EINGEBAUTEM REVERSIBLEN KURZSCHLUSSSCHUTZ

«HINTER GUTEM LICHT STECKT»

KNOBEL

VERLANGEN SIE DEN KATALOG ODER DAS PERSÖNLICHE FACHGESPRÄCH MIT DEN GUT AUSGEBILDETEN MITARBEITERN.

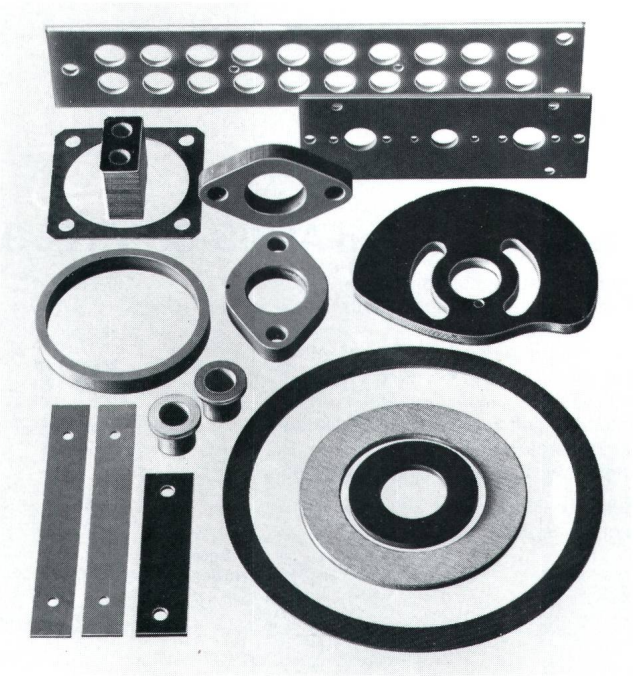
KNOBEL

F. KNOBEL ELEKTROAPPARATEBAU AG
8755 ENNENDA TEL. 058 / 63 11 71

EIN UNTERNEHMEN DER WALTER MEIER HOLDING AG

KNOBEL **TRANSFORMATOREN FÜR NIEDERVOLT-HALOGEN GLÜHLAMPEN**

Wachendorf AG, Technischer Grosshandel, 4002 Basel, Tel. 061-42 90 90



Bearbeitete Teile

aus Hartpapier, Hartgewebe, Fiber, Acetalharz, Polyamid, Glas-Polyester, Glas-Epoxy.

Die von uns geführten Halbzeuge, vor allem die Hart-Thermoplaste und Schichtpressstoffe, können von uns gemäss Ihren Angaben und Zeichnungen spanabhebend, aus Platten, Stäbe oder Rohre in eigener Werkstätte mit umfangreichem Maschinenpark bearbeitet werden.

Ihre Aufträge, seien es Einzelstücke oder Grossserien, erledigen wir kurzfristig.

Verlangen Sie unsere Plattenliste.

Wachendorf