

Ein statistisches Modell für die Fehler bei binärer Datenübertragung auf Fernsprechanalenen

VON JOACHIM SWOBODA *

Mitteilung aus dem Institut für Nachrichtenvermittlung und Datenverarbeitung der Universität Stuttgart

(A.E.Ü. 23 [1969], Heft 6, 313–322; eingegangen am 21. Dezember 1968)

DK 681.327.8:621.395.3:621.391.82

Die bekannten Modellvorstellungen von GILBERT bzw. von BERGER und MANDELBRÖT beschreiben nährungsweise die büschelartigen Fehler bei binärer Datenübertragung auf Fernsprechanalenen.

Ein neues Modell, das hier vorgeschlagen wird, arbeitet mit drei bzw. vier verschiedenen Störzuständen für den betrachteten Kanal, wobei zu jedem Zustand eine feste Fehlerfallswahrscheinlichkeit gehört. Jeder Störzustand kann dabei nur unmittelbar nach einer falsch übertragenen Stelle in einen anderen Störzustand übergehen. Für das Modell genügen die zufälligen Abstände zwischen den zeitlich benachbarten Fehlern einer Verteilung, die durch Überlagerung von drei bzw. vier geometrischen Verteilungen entsteht.

Die im Modell enthaltene Annahme, daß aufeinanderfolgende Fehlerabstände statistisch unabhängig sind, stellt eine praktikable Näherung dar. Als spezielle Erweiterung des Modells wird auch eine statistische Abhängigkeit für die Fehlerabstände eingeführt.

Die beiden obengenannten sowie die neue Modellvorstellung werden anhand eigener Meßreihen geprüft. Die gemessenen Fehlerabstandsverteilungen werden durch das neue Modell sehr genau beschrieben, wobei auch der auf die hypothetische Verteilung angewandte χ^2 -Test günstig ausfällt.

A Statistical Model for the Errors Occurring with Binary Data Transmission on Telephone Channels

The well-known model concepts of GILBERT and BERGER/MANDELBRÖT, respectively, describe approximately the clustered errors occurring with binary data transmission on telephone channels.

A new model is here proposed; it operates with three and four states of noise, respectively, for the channel under consideration, a fixed error probability being associated with each state. Each noise state can change into another noise state only immediately after an incorrectly transmitted place. The random separations between the time-adjacent errors of a distribution produced by a superposition of three and four geometrical distributions, respectively, suffice for the model.

The assumption contained in the model that successive error separations are statistically independent is a practical approximation. A statistical dependence is also introduced for the error separations as a special extension of the model concept.

The two aforementioned model concepts as well as the new one are verified by reference to measurement series conducted by the author. The measured error separation distributions are described with high accuracy by the new model; the χ^2 -test applied to the hypothetical distribution also gives a favorable result.

1. Einführung

Bei der Planung eines durch entsprechende Codierung gegen Fehler gesicherten Datenübertragungssystems spielt die Störspannung in den verfügbaren Kanälen eine entscheidende Rolle.

Es wäre daher nützlich, über eine vollständige statistische Beschreibung der Übertragungsfehler eines vorgesehenen Kanals zu verfügen. Dabei soll unter einem Übertragungsfehler die Fälschung einer gesendeten Binärstelle 0 in eine empfangene Binärstelle L verstanden werden bzw. umgekehrt. Auf Grund einer vollständigen statistischen Beschreibung der bei ungesicherter Übertragung auftretenden Fehler läßt sich die Auswirkung der Fehler auf ein geplantes Datenübertragungssystem mit Codesicherung entweder berechnen oder durch eine Funktionsnachbildung der (gesicherten) Datenübertragung einschließlich der Übertragungsfehler auf einem Rechner simulieren.

Von den während der Simulation nachgebildeten Übertragungsfehlern wird erwartet, daß sie in allen

statistischen Eigenschaften mit den wirklich auftretenden Fehlern möglichst gut übereinstimmen. Der Algorithmus zur möglichst wirklichkeitstreuen Nachbildung der Übertragungsfehler werde als *statistisches Kanalmodell* bezeichnet.

Ein besonders einfaches Modell, das jedoch keine „Büschelung“ der Fehler beschreiben kann, ist der statistisch unabhängig gestörte Binärkanal (binary symmetric memoryless channel). Bei diesem wird jede Binärstelle mit einer festen Wahrscheinlichkeit p_E gefälscht (E: error), und zwar statistisch unabhängig davon, ob diese Stelle eine 0 oder ein L ist, und ferner statistisch unabhängig von vorangehenden Fehlern. Ein Abstand von genau t Stellen zwischen einer gefälschten Binärstelle und der nächsten gefälschten Binärstelle tritt dann mit der Wahrscheinlichkeit $p(t) = (1 - p_E)^{t-1} p_E$ (1)

auf. Dabei liegen $t - 1$ nicht gefälschte Stellen dazwischen. Diese geometrische Verteilung für die Fehlerabstände beziehungsweise die zugehörige konstante Einfallswahrscheinlichkeit für Fehler werden jedoch auf Fernsprechanalenen, die für Datenübertragung benutzt werden, kaum beobachtet [1].

* Priv.-Doz. Dr. J. SWOBODA, 79 Ulm-Söflingen, Königstraße 88.

Vielmehr treten die Fehler „in Büscheln“, d.h. kurzzeitig gehäuft auf, und dazwischen liegen längere Zeiten fehlerfreier Übertragung. Insbesondere sind zwei Kanalmodelle, die eine Fehlerbüschelung berücksichtigen, bekannt geworden:

Ein *erstes Kanalmodell* wurde von GILBERT [2] eingeführt. Bei diesem Modell befindet sich der Kanal entweder in einem Gut-Zustand G (G: good) oder in einem Schlecht-Zustand B (B: bad). Im Gut-Zustand treten *keine* Übertragungsfehler auf, und im Schlecht-Zustand B fallen Fehler mit konstanter Wahrscheinlichkeit p_{EB} ein. Während des Schlecht-Zustandes stimmt dieses Modell mit dem erstgenannten Binärkanal überein. Der Kanal kann nach jeder übertragenen Stelle zwischen den beiden Zuständen G und B wechseln. Der Übergang von G nach B erfolgt dabei mit einer festen Übergangswahrscheinlichkeit p_{GB} , und von B nach G mit p_{BG} . Durch die drei Parameter p_{EB} , p_{GB} , p_{BG} ist das Kanalmodell von GILBERT vollständig beschrieben. Die Verteilung der Abstände zwischen benachbarten Fehlern — im folgenden als *Fehlerabstandsverteilung* bezeichnet — läßt sich entsprechend Gl. (1) berechnen. Der Zusammenhang zwischen den drei Kanalparametern und der Fehlerabstandsverteilung ist etwas kompliziert und wird von GILBERT [2] wie folgt angegeben:

$$p(t) = a_1 u_1^{t-2} + a_2 u_2^{t-2} \quad \text{für } t \geq 2, \quad (2)$$

$$p(1) = \bar{p}_{BG} p_{EB}, \quad (3)$$

wobei

$$\bar{p}_{EB} = 1 - p_{EB}, \quad p_{GG} = 1 - p_{GB}, \quad p_{BB} = 1 - p_{BG},$$

$$2u_{1,2} = p_{GG} + p_{BB} \bar{p}_{EB} \pm \sqrt{(p_{GG} + p_{BB} \bar{p}_{EB}) + 4\bar{p}_{EB}(p_{BG} - p_{GG})}, \quad (4)$$

$$a_1 = \frac{p_{BB} \bar{p}_{EB} - u_2}{u_1 - u_2} p_{EB} \left(p_{BB} - \frac{u_2 - p_{GG}}{\bar{p}_{EB}} \right),$$

$$a_2 = \frac{p_{BB} \bar{p}_{EB} - u_2}{u_1 - u_2} p_{EB} \left(p_{GB} - p_{BB} \frac{u_1 - p_{BB} \bar{p}_{EB}}{p_{BG}} \right) \cdot \left(p_{BG} + p_{BB} \frac{u_2 - p_{GG}}{p_{GB}} \right).$$

Die Abstandsverteilung ist vom Typ her eine Überlagerung von zwei geometrischen Verteilungen. Näherungsweise kann die eine der beiden überlagerten Verteilungen als die Abstandsverteilung für die kurzen Abstände *innerhalb* der Büschel gedeutet werden und der andere Anteil als die Abstandsverteilung für die längeren Abstände von Büschel zu Büschel.

Das büschelartige Auftreten der Fehler wird durch die Art der Fehlerabstandsverteilung erfaßt. Verglichen mit dem Kanal von durchwegs konstanter Fehlereinfallwahrscheinlichkeit treten im Kanalmodell von GILBERT sehr kurze und sehr lange Fehlerabstände mit größerer Wahrscheinlichkeit und dafür die mittleren Fehlerabstände mit kleinerer Wahrscheinlichkeit auf.

Es muß noch darauf hingewiesen werden, daß bei dem Kanalmodell von GILBERT das Auftreten einer gefälschten Stelle durchaus davon abhängt, ob vorangehende Stellen fehlerfrei sind oder nicht. Dagegen sind die Fehlerabstände in diesem Modell von vorangehenden Fehlerabständen statistisch *nicht* abhängig: Ein Fehlerabstand beginnt mit dem vorangehenden Fehler, wobei der Kanal stets im Zustand B sein muß. Die Wahrscheinlichkeit für das Auftreten des nächsten Fehlers nach t Schritten hängt von diesem Ausgangszustand und von den drei Kanalparametern p_{GB} , p_{BG} und p_{EB} ab. Da aber der Kanal bei Beginn eines betrachteten Fehlerabstandes stets im Zustand B ist, hat der vorangehende Fehlerabstand keinen Einfluß.

Ein *zweites Kanalmodell* wurde von BERGER und MANDELBROT [3] vorgeschlagen. In diesem Modell wird einerseits vorausgesetzt, daß die Fehlerabstände untereinander statistisch unabhängig sind, und andererseits wird angenommen, daß die Fehlerabstände der Pareto-Verteilung genügen:

$$p(\geq t) = t^{-\alpha} \quad \text{für } t = 1, 2, 3, \dots \quad (5)$$

Dabei ist $p(\geq t)$ die Wahrscheinlichkeit, daß ein Fehlerabstand nicht kürzer als die Länge t ist. Diese Abstandsverteilung verursacht in ähnlicher Weise wie die Abstandsverteilung des Modells von GILBERT eine Büschelung der Fehler.

Die Pareto-Verteilung von Gl. (5) stellt sich in doppelt-logarithmischem Maßstab als Gerade mit der Steigung $-\alpha$ dar:

$$\log p(\geq t) = -\alpha \log t. \quad (6)$$

Der Parameterwert α liegt bei der Approximation praktischer Fälle zwischen 0 und 1. In diesem Bereich $0 < \alpha < 1$ existiert für die Pareto-Verteilung unglücklicherweise kein endlicher Erwartungswert. Wie sehr sich auch die Nachrichteningenieure darüber freuen würden, wenn der Abstand von einem Übertragungsfehler bis zum nächsten Fehler „im Mittel unendlich groß“ wäre; um der Wirklichkeit besser zu entsprechen, wird in dem Vorschlag von BERGER und MANDELBROT die Verteilung an einer Stelle $t = \beta$ abgeschnitten. Der Erwartungswert der abgeschnittenen Verteilung kann durch entsprechende Wahl von β der Messung angepaßt werden. Allerdings treten dann gemäß dem Modell keine größeren Fehlerabstände als β auf. Somit kann eine gemessene Verteilung durch die zwei Parameter α und β approximiert werden.

2. Vergleich der zwei Büschelfehler-Modelle mit Messungen

Um die beschriebenen Kanalmodelle in ihrer Brauchbarkeit zu beurteilen, müssen sie mit Fehlermessungen bei der Datenübertragung in der Praxis verglichen werden. Alle einführend vorgestellten Kanalmodelle werden durch die Verteilung der Fehlerabstände in hinreichender Weise beschrieben. Daher liegt es nahe, für die Beurteilung der Modelle gemessene Fehlerabstandsverteilungen heranzuziehen.

Gemessene Fehlerabstandsverteilungen sind nur zum Teil den bekannten Veröffentlichungen [6], [7], [8] zu entnehmen. Jedoch waren Unterlagen über die Fehlerstatistik bei Datenübertragung über das Fernsprechnetz der Deutschen Bundespost nicht in dem Maße verfügbar, wie es für die Beurteilung der Modelle notwendig war [15]. Daher wurden vom Verfasser umfangreiche, von der Deutschen Bundespost durch bereitgestellte Leitungen unterstützte Messungen ausgeführt, über die an anderer Stelle [9] ausführlich berichtet wird. In der vorliegenden Arbeit wird für die vergleichende Betrachtung der Modelle nur eine dieser zahlreichen Messungen benutzt, die als typisch angesehen werden darf.

Bei dieser Messung wurden während 94 Minuten mit einem 4-Phasen-Modem $6,8 \cdot 10^6$ Binärstellen mit einer Geschwindigkeit von 1200 bit/s und mit einem Sendempfangspegel von $-16/-21$ dB übertragen. Der Übertragungskanal war eine über Hebdrehwähler gewählte Fernsprech-Ortsverbindung. Bei der Messung wurden 3138 fehlerhafte Binärstellen registriert, und die Folge von 3138 Fehlerabständen auf Lochstreifen ausgegeben.

2.1. Vergleich des Modells von Gilbert mit der Messung

Die gemessene Fehlerabstandsverteilung wird als Summenhäufigkeit $h(\geq t)$ dargestellt; dies ist die relative Häufigkeit für Abstände von t oder mehr Binärstellen. Für unmittelbar benachbarte fehlerhafte Binärstellen ist als Fehlerabstand $t = 1$ vereinbart.

Die der Summenhäufigkeit $h(\geq t)$ entsprechende Verteilungsfunktion der Abstände $p(\geq t)$ für das Modell von GILBERT lautet gemäß Gl. (2)

$$p(\geq t) = \sum_{\tau=t}^{\infty} (a_1 u_1^{\tau-2} + a_2 u_2^{\tau-2}) \quad \text{für } t = 2, 3, \dots \quad (7)$$

$$= \frac{a_1}{1 - u_1} u_1^{t-2} + \frac{a_2}{1 - u_2} u_2^{t-2} \quad (8)$$

$$= A_1 u_1^{t-2} + A_2 u_2^{t-2}. \quad (9)$$

Wir entnehmen Gl. (9), daß sich die Verteilungsfunktion für die Abstände $\geq t$ als Überlagerung von zwei geometrischen Folgen ergibt.

Für sehr große Abstände t darf unter der Annahme

$$u_1 > u_2 (< 1) \quad (10)$$

der zweite Term in Gl. (9) vernachlässigt werden:

$$p(\geq t) = A_1 u_1^{t-2} \left[1 + \frac{A_2}{A_1} \left(\frac{u_2}{u_1} \right)^{t-2} \right],$$

$$p(\geq t) \approx A_1 u_1^{t-2} \quad \text{für } t \gg 1. \quad (11)$$

In einer einfach-logarithmischen Darstellung für $p(\geq t)$ über t liegen daher die Punkte für $t \gg 1$ auf einer Geraden mit der Steigung $\log u_1$:

$$\log p(\geq t) \approx (t - 2) \log u_1 + \log A_1 \quad \text{für } t \gg 1.$$

Um die Übereinstimmung mit der Messung zu beurteilen, ist in Bild 1 die gemessene Summenhäufig-

keit $h(\geq t)$ im halb-logarithmischen Maßstab dargestellt. Der lineare Maßstab für die Abstände reicht bis $t = 50000$, was bei einer Übertragungsgeschwindigkeit von 1200 bit/s einem zeitlichen Abstand von etwa 40 s entspricht.

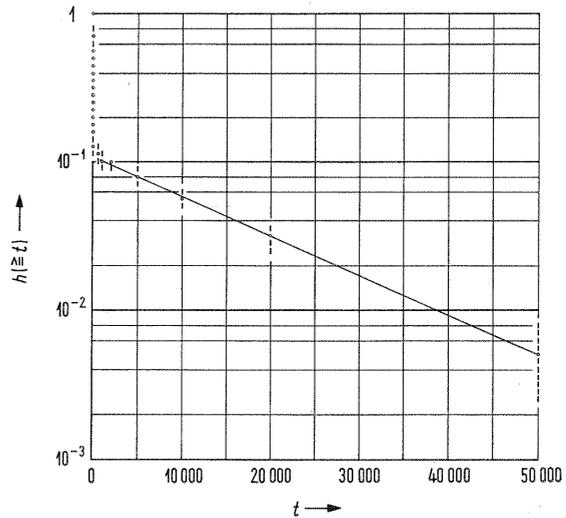


Bild 1. Gemessene Summenhäufigkeit $h(\geq t)$ für Abstände von t oder mehr Schritten zwischen aufeinanderfolgenden Fehlerstellen, dargestellt im halb-logarithmischen Maßstab. Das Bild ist von einem Rechner ausgedruckt und daher auf die Druckpositionen des Zeilendruckers quantisiert. Die senkrechten Intervalle zu jedem Meßpunkt $h(\geq t)$ sind Vertrauensintervalle auf Grund einer endlichen Meßdauer. Die Vertrauensintervalle überdecken mit 95% Sicherheit den „wahren“ Wert. Die Meßpunkte liegen für $t \geq 500$ (500 Stellen entsprechen einer zeitlichen Dauer von 417 ms) in der halb-logarithmischen Darstellung mit guter Näherung auf einer Geraden, d.h., sie genügen einer geometrischen Verteilung.

In Bild 1 liegen die Meßpunkte $h(\geq t)$ für $t \geq 500$ mit guter Näherung auf einer Geraden. Damit widerspricht diese Messung der Annahme nicht, daß die langen Fehlerabstände einer geometrischen Verteilung genügen.

Bei einem Vergleich zwischen dem Modell von GILBERT und der gemessenen Abstandsverteilung ist weiter zu prüfen, ob letztere durch die Überlagerung von gerade zwei geometrischen Folgen darstellbar ist: Durch die „nach Augenmaß“ eingezeichnete Gerade in Bild 1 sind Schätzwerte für die eine geometrische Folge (Gl. (11)) festgelegt worden. Wird von den Meßpunkten $h(\geq t)$ diese erste geometrische Folge abgezogen, dann müßte die zweite geometrische Folge als reduzierte Summenhäufigkeit $h_{\text{red}}(\geq t)$ verbleiben. Die reduzierten Meßpunkte $h_{\text{red}}(\geq t)$ entsprechen der reduzierten Verteilung

$$h_{\text{red}}(\geq t) \triangleq p(\geq t) - A_1 u_1^{t-2} = A_2 u_2^{t-2}. \quad (13)$$

In halb-logarithmischer Darstellung müßte $h_{\text{red}}(\geq t)$ ebenfalls auf einer Geraden liegen.

Bild 2 zeigt die derart reduzierte Summenhäufigkeit $h_{\text{red}}(\geq t)$ im halb-logarithmischen Maßstab. Die lineare Abstandsachse reicht dabei bis $t = 100$

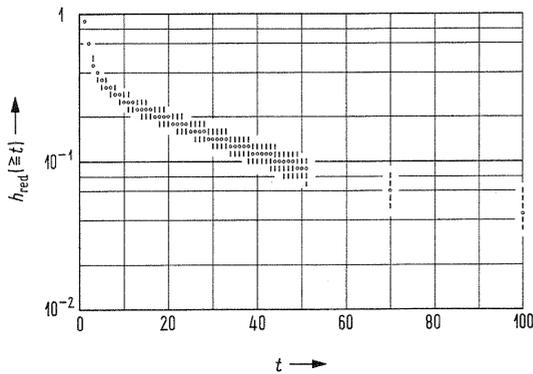


Bild 2. Aus der gemessenen Summenhäufigkeit gewonnene reduzierte Summenhäufigkeit $h_{\text{red}}(\geq t)$, dargestellt im halb-logarithmischen Maßstab. Für einen Kanal nach dem Modell von GILBERT müßten die dargestellten Punkte auf einer Geraden liegen.

Schritte. Die reduzierten Meßpunkte mit ihren Vertrauensintervallen liegen in signifikanter Weise *nicht* auf einer Geraden. Dies zeigt, daß die gemessene Abstandsverteilungsfunktion $h(\geq t)$ durch die Überlagerung von nur zwei geometrischen Folgen nur sehr unbefriedigend approximiert werden kann. Jedoch erscheint es anhand von Bild 2 aussichtsreich, das Modell von GILBERT dahingehend zu erweitern, daß die Abstandsverteilungsfunktion durch die Überlagerung von mehr als zwei geometrischen Folgen dargestellt wird. Dieser Weg wird in dem Vorschlag von Abschnitt 3 besprochen.

Für das Modell von GILBERT jedoch darf gesagt werden, daß im Bereich großer Fehlerabstände die gemessene Verteilung sehr gut approximiert wird. Dann aber ist die Approximation für kleine Fehlerabstände weniger gut.

2.2. Vergleich des Modells von Berger und Mandelbrot mit der Messung

Für den Vergleich des Modells von BERGER und MANDELBROT mit der Messung wird die gemessene Fehlerabstandsverteilung $h(\geq t)$ in Bild 3 im doppelt-logarithmischen Maßstab dargestellt. Die senkrechten Intervalle geben wieder die 95%-Vertrauensintervalle zu den Meßpunkten an.

Die Pareto-Verteilung dieses Modells ist nach Gl. (6) eine Gerade mit der Steigung $-\alpha$. Als Approximation an die Messung ist in Bild 3 eine Gerade mit der Steigung $-\alpha = -0,4$ eingezeichnet. Die Pareto-Verteilung $p(\geq t) = t^{-\alpha}$ hat für $\alpha = 0,4$ einen nicht endlichen Erwartungswert. Daher wird die Verteilung abgeschnitten, z. B. bei $t = \beta = 10^5$, und damit der zweite Parameter β eingeführt. Gemäß diesem Modell müßte dann spätestens nach $10^5 : 1200 \text{ bit/s} \approx 80 \text{ s}$ ein Übertragungsfehler auftreten.

Die Übereinstimmung mit der Messung im Bereich kleiner Abstände bis etwa $t \leq 100$ ist durchaus befriedigend. Jedoch weichen für große Abstände ($100 < t \leq 50000$) die Meßpunkte von der Pareto-Verteilung so stark ab, daß diese Abweichungen nicht als zufällig angesehen werden können. Ab-

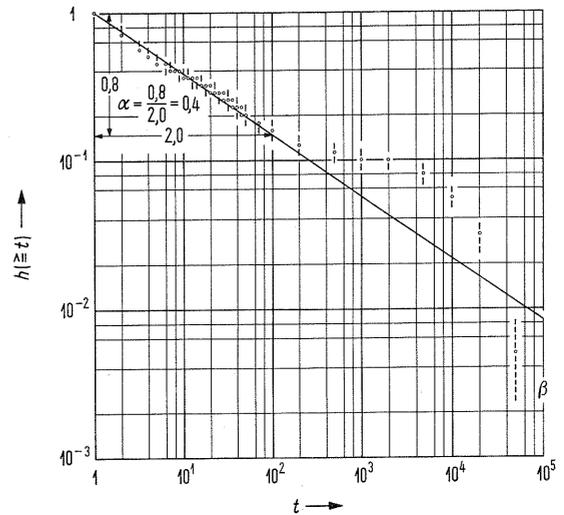


Bild 3. Gemessene Summenhäufigkeit $h(\geq t)$ wie in Bild 1, jedoch hier im doppelt-logarithmischen Maßstab. (Zwei verschiedene Ordinatenwerte zu einem Abszissenwert entstehen aus der Rundung der Werte auf die Druckpositionen.)

Für einen Kanal nach dem Modell von BERGER und MANDELBROT müßten die dargestellten Punkte auf einer Geraden liegen. Für Fehlerabstände von $t > 100$ (was 83 ms entspricht) zeigen sich starke, systematische Abweichungen.

weichungen dieser Art haben sich ebenfalls bei allen hier nicht vorgestellten Messungen gezeigt. Bei den 95%-Vertrauensintervallen müßte man erwarten, daß die theoretische Verteilung nur in 5% der Fälle von diesen Intervallen nicht überdeckt wird.

Damit kann zusammenfassend festgestellt werden, daß die Fehlerabstandsverteilung nur im Bereich kleinerer Abstände durch die Pareto-Verteilung befriedigend beschrieben werden kann.

Bevor im Abschnitt 3 ein neues Kanalmodell vorgestellt wird, das die gemessenen Fehlerabstandsverteilungen besser beschreibt, muß noch untersucht werden, inwieweit die statistische Unabhängigkeit zwischen den Abständen, die den bisher diskutierten Modellen zu eigen ist, durch die Messungen gerechtfertigt ist.

2.3. Gemessene statistische Abhängigkeit zwischen den Fehlerabständen

Die statistische Unabhängigkeit zwischen den Fehlerabständen wird bei dem Kanalmodell nach BERGER und MANDELBROT a priori angenommen und bei dem Modell von GILBERT folgt sie aus den Voraussetzungen.

Um für die Messungen den Grad einer eventuellen statistischen Abhängigkeit zwischen aufeinanderfolgenden Fehlerabständen zu beschreiben, könnte der empirische Kovarianzfaktor [10] herangezogen werden. Dieser beträgt bei der bereits oben eingeführten Messung $k = 3,3\%$. Der kleine Kovarianzfaktor k läßt eine geringe Abhängigkeit vermuten. Da aber eine starke Abhängigkeit nicht unbedingt auf einen großen Kovarianzfaktor führt, wird statt dieses unsicheren Maßes besser der χ^2 -Test [11] auf

die Kontingenztafel benachbarter Abstände angewendet¹.

		j →													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
t ↓	1 - 1	1	+						-	-	-	-			
	2 - 2	2								-					
	3 - 3	3													
	4 - 4	4													
	5 - 6	5													
	7 - 9	6								↘	+				
	10 - 15	7													
	16 - 23	8								+					
	24 - 33	9													
	34 - 49	10													
	50 - 99	11													
	100 - 499	12												+	
	500 - 9999	13												+	+
	10000 - 99999	14													+

Bild 4. Kontingenztafel für Paare aufeinanderfolgender Fehlerabstände. Ein „+“-Zeichen für die Koordinate (i, j) bedeutet, daß nach einem Fehlerabstand der Klasse i ein Abstand aus der Klasse j mit größerer Häufigkeit auftritt, als es bei statistischer Unabhängigkeit zu erwarten wäre. Entsprechendes gilt für das „-“-Zeichen.

Für die betrachtete Messung wurden die Fehlerabstände in 14 Klassen eingeteilt. In Bild 4 sind links die zu den Klassennummern i gehörigen Intervalle der Fehlerabstände t zu entnehmen. Die Matrix mit den Zeichen „-“, „+“ und „leer“ in Bild 4 gibt ein übersichtliches Bild für die Kontingenztafel [14]: Das durch einen Pfeil markierte Zeichen „+“ z. B. bedeutet, daß auf Fehlerabstände, die in die Klasse $i = 6$ fallen, als nächster Fehlerabstand einer aus der Klasse $j = 9$ mit größerer Häufigkeit auftritt, als es bei statistischer Unabhängigkeit erwartet wird. Entsprechend bedeutet ein Zeichen „-“ in der Koordinate (i, j) , daß auf die Abstandsklasse i die Abstandsklasse j zu selten folgt. Die Schranken, bei deren Überschreitung ein „+“ bzw. ein „-“ gesetzt wurde, sind so bemessen, daß bei statistischer Unabhängigkeit im Mittel nur 5% der Koordinaten (i, j) mit „+“- oder „-“-Zeichen in zufälliger Verteilung belegt wären.

Auffällig ist die Konzentration der „+“-Zeichen auf der Hauptdiagonalen, insbesondere rechts unten

¹ Für den χ^2 -Test wird der Merkmalraum (hier die Abstände) in n Klassen eingeteilt. Aus den beobachteten, absoluten Häufigkeiten H_i , die in der Klasse i auftreten, und aus den gemäß der Hypothese erwarteten absoluten Häufigkeiten T_i wird das Abweichungsmaß

$$\chi^2 = \sum_{i=1}^n \frac{(H_i - T_i)^2}{T_i} \quad (14)$$

gebildet. Aus dem Abweichungsmaß χ^2 kann eine sogenannte Signifikanz a berechnet werden. Die Signifikanz gibt die Wahrscheinlichkeit an, mit der man sich täuscht, wenn man die zu prüfende Hypothese *verwirft*. Je kleiner χ^2 ist, um so größer wird a . Signifikanzwerte von einigen Prozent oder mehr werden üblicherweise als so große Täuschungswahrscheinlichkeiten angesehen, daß das Verwerfen der Hypothese zu riskant wird; der χ^2 -Test legt dann der Annahme der Hypothese nichts in den Weg.

in der Kontingenztafel. Letzteres bedeutet, daß bei der Messung auf einen langen Fehlerabstand wieder bevorzugt ein langer Fehlerabstand folgt. Entsprechend zu selten folgt dann auf einen langen Abstand ein kurzer Abstand, und umgekehrt. Auf diese Eigenart wurde bereits von LEWIS und COX [5] erstmals hingewiesen. Ferner fällt auf, daß etwa 15%, und nicht nur 5%, der Koordinatenpunkte mit „+“- oder „-“-Zeichen belegt sind. Dies schlägt sich auch in der Ablehnung der Hypothese der Unabhängigkeit durch den χ^2 -Test nieder: Wenn die Unabhängigkeitshypothese für die Fehlerabstände dieser Messung verworfen wird, dann ist die Irrtumswahrscheinlichkeit α sehr klein, nämlich kleiner als 10/100.

Der χ^2 -Test entscheidet bei einer großen Stichprobe hinsichtlich der Verwerfung einer Hypothese sehr scharf. Deshalb soll unbeschadet dieses χ^2 -Tests ein gegenüber dem Modell von GILBERT und dem Modell von BERGER und MANDELBROT verbessertes Kanalmodell vorgeschlagen werden, das *zunächst* ebenfalls statistisch unabhängige Fehlerabstände postuliert. Wie stark sich der Fehler vernachlässigter Abhängigkeit auswirkt, wird noch im Abschnitt 3.3.3 zu prüfen sein.

3. Vorschlag eines neuen Kanalmodells

3.1. Erste Beschreibungsform des neuen Modells

Dieses Modell stellt in gewisser Hinsicht eine Erweiterung des Modells von GILBERT dar. Nach dem neuen Vorschlag soll die Fehlerstatistik des Kanalmodells für Datenübertragung auf Fernspreitleitungen durch folgende drei Eigenschaften gekennzeichnet sein:

1. Die Wahrscheinlichkeit, daß zwischen einer gefälschten Binärstelle und der nächsten gefälschten Binärstelle $t - 1$ oder mehr ungefälschte Binärstellen liegen, sei

$$p(\geq t) = \sum_{i=1}^g W_i u_i^{t-1} \quad \text{für } t = 1, 2, 3, \dots \quad (15)$$

mit
$$\sum_{i=1}^g W_i = 1.$$

Diese Verteilung ist eine Summe von g geometrischen Verteilungen. Unmittelbar benachbarte Fehler entsprechen dabei dem Fehlerabstand $t = 1$.

2. Die Fälschung einer übertragenen Binärstelle soll statistisch nicht davon abhängen, ob die gesendete Binärstelle eine 0 oder ein L war (Symmetrie).
3. Die Fehlerabstände sollen untereinander statistisch unabhängig sein.

3.2. Zweite Beschreibungsform des neuen Kanalmodells

Dieses Kanalmodell kann auch noch auf eine zweite Weise beschrieben werden. Diese Beschreibungsart könnte als eine physikalische Interpreta-

tion aufgefaßt werden, ohne daß diese Interpretation einen Anspruch auf die Rechtfertigung des Modells erhebt. Ein Modell kann selbstverständlich nur durch Messungen gerechtfertigt werden. Neben der Interpretation ist diese zweite Beschreibungsart jedoch gut geeignet, um Störungen gemäß diesem Kanalmodell bei einer Simulation von Kanal und Datenübertragung auf einem Digitalrechner nachzubilden.

Der Kanal kann sich in g verschiedenen Zuständen

$$Z_1, Z_2, \dots, Z_i, \dots, Z_g$$

befinden. Jedem Zustand entspricht eine verschiedene Störintensität des Kanals, so daß in den Zuständen Z_1, \dots, Z_g die Fehler mit den Wahrscheinlichkeiten

$$p_{E1}, p_{E2}, \dots, p_{Ei}, \dots, p_{Eg}$$

auftreten. Es wird unterstellt, daß ein Zustand Z_i nur zwischen einer *gefälschten* Binärstelle und der unmittelbar darauf folgenden Binärstelle in einen anderen Zustand Z_j übergehen kann. Diese Annahme unterscheidet sich vom Modell von GILBERT, und sie ist wesentlich, weil dadurch eine sehr einfache und geschlossene mathematische Beschreibung ermöglicht wird. (Ein Modell erhält seine Berechtigung durch die Verträglichkeit mit Messungen, und ein einfaches Modell ist einem komplizierten stets vorzuziehen.) Damit ändert der Kanal während eines Fehlerabstandes seinen Zustand nicht.

Der Übergang des Kanalzustandes nach einer gefälschten Binärstelle in einen der Zustände

$$Z_1, \dots, Z_i, \dots, Z_g$$

soll mit den Wahrscheinlichkeiten

$$W_1, W_2, \dots, W_i, \dots, W_g$$

erfolgen, und zwar unabhängig vom bestehenden Zustand. Letztere Annahme bedingt die statistische Unabhängigkeit aufeinanderfolgender Fehlerabstände; dazu wird im Abschnitt 3.4 noch eine Erweiterung vorgeschlagen, welche auch die Annahme einer statistischen Abhängigkeit ermöglicht.

Mit den genannten Voraussetzungen läßt sich die Fehlerabstandsverteilung $p(\geq t)$ leicht angeben: Ein Abstand beginnt und liegt mit der Wahrscheinlichkeit W_i im Zustand Z_i , der die Fehlerwahrscheinlichkeit p_{Ei} hat. Wenn der Abstand größer oder gleich t sein soll, dann müssen mindestens $t-1$ aufeinanderfolgende Stellen ungestört sein. Die Wahrscheinlichkeit, daß mindestens $t-1$ aufeinanderfolgende ungestörte Stellen auftreten und daß diese Folge dem Zustand Z_i angehört, ist

$$W_i(1 - p_{Ei})^{t-1}.$$

Jeder der g Zustände liefert einen solchen Beitrag, und die Wahrscheinlichkeiten der paarweise fremden Ereignisse können addiert werden:

$$p(\geq t) = \sum_{i=1}^g W_i(1 - p_{Ei})^{t-1} \quad (16)$$

$$\text{mit} \quad p(\geq t) = \sum_{i=1}^g W_i = 1. \quad (17)$$

Der Vergleich von Gl. (16) und Gl. (15) zeigt, daß die zweite Beschreibungsform des Kanalmodells mit der ersten Form verträglich ist. Dabei ist die Größe

$$u_i = 1 - p_{Ei} \quad (i = 1, 2, \dots, g) \quad (18)$$

die Wahrscheinlichkeit für eine richtig übertragene Binärstelle im Zustand Z_i .

Das neue Kanalmodell kann wegen der unabhängigen Fehlerabstände und der Fehlerabstandsverteilung als eine Erweiterung des Modells von GILBERT angesprochen werden. Das neue Modell für den Sonderfall $g = 2$ und das Modell von GILBERT sind bis auf die Wahrscheinlichkeit für die Abstände $t = 1$ in ihrem Verhalten identisch. Durch die Einschränkung für den Zustandswechsel nur nach einem Fehlerbit wird aber die mathematische Handhabung des neuen Modells wesentlich vereinfacht.

3.2.1. Weitere Kenngrößen des neuen Kanalmodells

Als weitere Kenngrößen des Kanalmodells sollen noch berechnet werden:

die mittleren Fehlerabstände E_i im Zustand Z_i , die Wahrscheinlichkeit p_{Zi} , daß eine beliebig herausgegriffene Binärstelle vom Kanal im Zustand Z_i übertragen wurde, und

die mittlere, pauschale Stellenfehlerwahrscheinlichkeit p_E .

Der mittlere Fehlerabstand E_i im Zustand Z_i folgt aus der Wahrscheinlichkeit für den Abstand t für Z_i von $(1 - p_{Ei})^{t-1} p_{Ei}$ zu

$$E_i = \sum_{t=1}^{\infty} t(1 - p_{Ei})^{t-1} p_{Ei}. \quad (19)$$

Nach Umformung ergibt sich

$$E_i = \frac{1}{p_{Ei}} \quad (i = 1, 2, \dots, g), \quad (20)$$

also der Kehrwert der Fehlerwahrscheinlichkeit im Zustand Z_i .

Der über alle g Zustände gemittelte pauschale Fehlerabstand E ergibt sich durch Bewertung mit W_i zu

$$E = \sum_{i=1}^g W_i E_i. \quad (21)$$

Die Wahrscheinlichkeit p_{Zi} für den Zustand Z_i ist gleich dem relativen Anteil, den der Zustand Z_i zu der Summe in Gl. (21) beiträgt:

$$p_{Zi} = \frac{W_i E_i}{\sum_{j=1}^g W_j E_j}. \quad (22)$$

Die mittlere, pauschale Stellenfehlerwahrscheinlichkeit p_E ergibt sich entsprechend Gl. (20) als Kehrwert des pauschalen mittleren Fehlerabstandes E :

$$p_E = 1/E. \quad (23)$$

Nach der Vorstellung des Modells und seiner Eigenschaften muß noch geprüft werden, inwieweit dieses die Fehler bei realer Datenübertragung auf Fernsprechkanälen zu beschreiben vermag.

3.3. Prüfung des vorgeschlagenen Kanalmodells durch Messungen

Für die Prüfung des vorgeschlagenen Kanalmodells anhand von Messungen sollen die drei kennzeichnenden Eigenschaften Fehlerabstandsverteilung, Symmetrie und statistische Unabhängigkeit herangezogen werden.

3.3.1. Prüfung der Fehlerabstandsverteilung

Als erstes wird geprüft, ob die gemessene Fehlerabstandsverteilung mit der theoretischen Verteilung des Modells verträglich ist. Als gemessene Abstandsverteilung wird die gleiche Messung herangezogen, mit der im Abschnitt 2 das Modell von GILBERT und das Modell von BERGER und MANDELBROT verglichen wurden. Diese relative Häufigkeitsverteilung $h(\geq t)$ für gemessene Fehlerabstände zeigt Bild 5 in

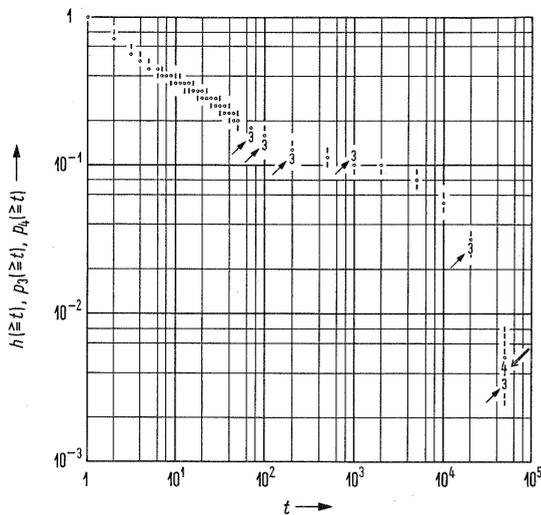


Bild 5. Die Punkte mit den senkrechten Strecken geben die gemessene Summenhäufigkeit $h(\geq t)$ mit den zugehörigen Vertrauensintervallen an. Die Darstellung stimmt mit Bild 3 überein. Darüber hinaus geben die Positionen mit der Ziffer 3 den Verlauf der theoretischen Werte $p_3(\geq t)$ bei der Approximation mit dem Grad $g = 3$ an. Wenn der theoretische Wert $p_3(\geq t)$ auf den Meßwert $h(\geq t)$ fällt (Abweichung kleiner als eine halbe Druckposition), dann ist die Ziffer 3 nicht ausgedruckt. Man sieht, daß nur fünf Punkte um eine Position abweichen (das entspricht 12%), und $p_3(\geq t)$ von $h(\geq t)$ für den Abstand $t = 50000$ um drei Positionen abweicht. Damit stellt $p_3(\geq t)$ bereits eine „gute Approximation“ der Messung dar.

Entsprechend geben Positionen mit der Ziffer 4 den Verlauf der theoretischen Werte $p_4(\geq t)$ bei der Approximation mit dem Grad $g = 4$ an. Es unterscheiden sich $h(\geq t)$ und $p_4(\geq t)$ in so geringem Maße, daß nur für $t = 50000$ eine Abweichung um eine Druckposition auftritt.

doppelt-logarithmischer Darstellung. Um $h(\geq t)$ mittels einer theoretischen Funktion $p(\geq t)$ nach Gl. (15) zu approximieren, muß zuerst festgelegt werden:

Nach welchem Kriterium und nach welchem Verfahren soll approximiert werden?

Mit welchem Grad von g überlagerten geometrischen Verteilungen (in Gl. (15)) soll approximiert werden?

Als Approximationskriterium wurde gewählt²:

$$\sum_t \frac{[h(\geq t) - p(\geq t)]^2}{p(\geq t)[1 - p(\geq t)]} \rightarrow \text{Minimum bezüglich Parameter von } p(\geq t). \quad (24)$$

Dieser Ausdruck liefert eine Maßzahl, die mit der quadratischen Abweichung zwischen Messung und Approximation zusammenhängt, und diese soll durch Wahl der $2g - 1$ freien Parameter $W_1, W_2, W_{g-1}, u_1, u_2, \dots, u_g$ von $p(\geq t)$ zum Minimum gemacht werden. Dafür wurde ein Rechenprogramm [12] mitbenutzt, das nach einem Gradientenverfahren von POWELL und FLETCHER [13] arbeitet.

Die zweite Frage nach dem Grad g der Approximation ist eng verknüpft mit der Beurteilung der Übereinstimmung zwischen gemessener und theoretischer Verteilung. Dem Bild 5 sind der gemessene Verlauf $h(\geq t)$ sowie der theoretische Verlauf $p(\geq t)$ für den Grad $g = 3$ und $g = 4$ zu entnehmen. Das Bild zeigt augenscheinlich eine gute Übereinstimmung zwischen Messung und theoretischen Werten. Die Parameter für die Approximation mit dem Grad $g = 3$ nennt Tabelle I.

Tabelle I.

Zustandsnummer i	mittlerer Fehlerabstand E_i	Anteil der Abstände W_i
1	14400	0,12
2	36	0,32
3	2,2	0,56

Interessant sind die mittleren Fehlerabstände E_i , die zu den drei Störzuständen Z_i gehören: Als kleinster mittlerer Fehlerabstand entspricht $E_3 = 2,2$ den Fehlerabständen innerhalb der Fehlerbüschel. Ein mittlerer Fehlerabstand von $E_3 = 2,0$ würde eine maximale Störung mit der Stellenfehlerwahrscheinlichkeit $p_{E3} = 0,5$ bedeuten. Der mittlere Fehlerabstand $E_2 = 36$ ($\cong 30$ ms) läßt sich als mittlerer Abstand zwischen den Büscheln deuten. Schließlich können die ganz langen Fehlerabstände von $E_1 = 14400$ ($\cong 12$ s) so interpretiert werden, daß die genannten Fehlerbüschel in Büschelgruppen kürzerer Büschel zusammengefaßt werden, die einen mittleren Abstand E_1 von Gruppe zu Gruppe haben.

Mit diesem Eindruck darf man sich jedoch nicht begnügen, vielmehr soll der χ^2 -Test [11] angewendet werden auf folgende Hypothese: „Die Fehlerabstände, die bei einer Messung die Häufigkeitsverteilung $h(\geq t)$ ergaben, gehorchen einer Wahrscheinlichkeitsverteilung $p(\geq t)$ nach Gl. (15), d. h.

² Es wäre naheliegender, zur Schätzung der Parameter für die theoretische Funktion $p(\geq t)$ die „likelihood-Methode“ heranzuziehen [14]. Wird jedoch die Übereinstimmung mit dem χ^2 -Test geprüft, dann führte hier das oben genannte Kriterium auf signifikantere Ergebnisse. Die Benutzung dieses Approximationskriteriums erhält nur durch den Erfolg beim χ^2 -Test seine nachträgliche Rechtfertigung.

einer Überlagerung von g geometrischen Verteilungen.“ Die Approximation an die Meßkurve wurde mit drei verschiedenen Graden für $p(\geq t)$ ausgeführt, nämlich $g = 3, 4$ und 5 .

Dabei lieferte der χ^2 -Test die Signifikanzwerte α von Tabelle II.

Tabelle II.

Grad g	Signif. α
3	< 0,001
4	0,255
5	0,193

Der Tabelle II entnimmt man:

Für den Grad $g = 3$ wird die Verteilungshypothese wegen zu geringer Signifikanz vom χ^2 -Test noch abgelehnt, obwohl die gemessene Fehlerabstandsverteilung durch die Überlagerung von drei geometrischen Verteilungen, wie Bild 5 zeigt, schon recht gut approximiert wird.

Für den Grad $g = 4$ steht bei einer Signifikanz von über 25% der Annahme der Hypothese nichts im Wege. Bei der Stichprobengröße von 3138 Fehlerabständen ist dies eine ausgezeichnete Bestätigung für die angenommene Verteilung.

Für den Grad $g = 5$ fällt gegenüber $g = 4$ die Signifikanz. Dies ist kein Widerspruch. Selbstverständlich kann die Meßkurve durch mehr Freiheitsgrade besser approximiert werden. Jedoch geht in die Berechnung der Signifikanz die Zahl der Freiheitsgrade ein, und bei einem Mehraufwand an Freiheitsgraden wird eine entsprechende Abnahme des Abweichungsmaßes erwartet. Diese stellt sich hier nicht ein und daher wird für $g = 5$ die Signifikanz geringer.

Zusammenfassend darf festgestellt werden: Die Verteilung der Fehlerabstände kann durch die Überlagerung von g geometrischen Verteilungen nach Gl. (15) beschrieben werden. Die Approximation ist besonders gut bei $g = 4$, so daß nicht einmal der χ^2 -Test die Hypothese auf die Abstandsverteilung verwirft. Jedoch kann bereits bei einem Grad von $g = 3$ die Übereinstimmung zwischen Messung und Modell als augenscheinlich gut bezeichnet werden.

3.3.2. Prüfung der Fehlersymmetrie

Als ein zweites Kennzeichen des vorgeschlagenen Modells ist zu prüfen, ob Fehler in gesendeten 0-Stellen mit gleicher Wahrscheinlichkeit auftreten wie in gesendeten L-Stellen. Diese Symmetriehypothese ließe sich ebenfalls mit dem χ^2 -Test nachprüfen. Jedoch wurden die Fehler, die in 0 und L auftreten, in der Auswertung nicht getrennt, sondern getrennte Messungen ausgeführt, bei denen nur 0 oder nur L oder gemischt 0/L gesendet wurde. Bei Symmetrie müßte bei allen drei Messungen eine im Rahmen der statistisch bedingten Schwankungen gleich große Stellenfehlerwahrscheinlichkeit p_E erwartet werden.

In Bild 6 ist die pauschale Fehlerrate p_E für drei Meßserien aufgetragen. Das Vertrauensintervall ei-

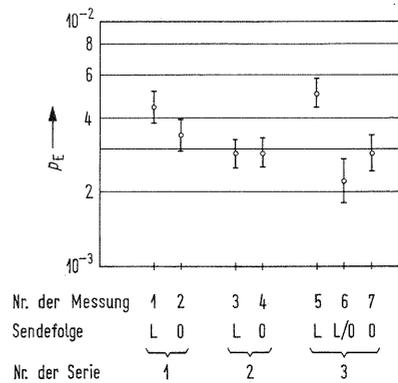


Bild 6. Pauschale Fehlerrate p_E von sieben Messungen, bei denen nur 0, nur L oder L/0 gemischt gesendet wurde. Bei Symmetrie für die Fehler überdecken die Vertrauensintervalle den gemeinsamen „wahren Wert“ einer Serie mit 95% Sicherheit.

ner Messung folgt aus der Stichprobengröße und überdeckt mit 95% Wahrscheinlichkeit den „wahren Wert“. Die Serien 1 und 2 wurden mit einem Phasenmodem und die Serie 3 mit einem Frequenzmodem gemessen. Die Abweichungen zwischen den Messungen einer Serie sind nicht so groß, daß eine Unsymmetrie unterstellt werden muß, sondern erklären sich eher aus unterschiedlichen Bedingungen von Messung zu Messung.

Da die Symmetrie in starkem Maße von der Einstellung der Diskriminatorchwelle im Demodulator des Empfängers abhängt, mögen diese Messungen zur Rechtfertigung der Symmetrieannahme hier genügen.

3.3.3. Zur statistischen Unabhängigkeit zwischen den Fehlerabständen

Bereits im Abschnitt 2.3 wurde festgestellt, daß der χ^2 -Test die Hypothese der statistischen Unabhängigkeit für die Fehlerabstände verwirft. Jedoch bietet der χ^2 -Test kein Maß für die Stärke der Abhängigkeit, da der Test die Hypothese umso sicherer verwirft, je größer nur die Stichprobe ist.

Jedes einfache Abhängigkeitsmaß, etwa durch eine Zahl ausgedrückt, setzt eine einschränkende Bewertung der Abhängigkeit voraus. In diesem Sinn könnte hier folgende Frage interessieren: Mit welcher Wahrscheinlichkeit treten genau i Fehlerstellen in einem Block von n Stellen auf? Oder, was sehr eng mit dieser Frage verknüpft ist: Mit welcher Wahrscheinlichkeit erstrecken sich j aufeinanderfolgende Fehlerabstände gerade insgesamt über t Stellen? Die letztere Frage soll näher untersucht werden.

Es wird die Wahrscheinlichkeitsverteilung für die Summe von j aufeinanderfolgenden Abständen betrachtet: Aus der Messung kann einerseits die Häufigkeitsverteilung $h_j(\geq t)$ von j aufeinanderfolgenden Abständen ermittelt werden. Andererseits kann aus der Häufigkeitsverteilung einfacher Abstände $h_1(\geq t) = h(\geq t)$ jene Häufigkeitsverteilung $h_{uj}(\geq t)$ für j aufeinanderfolgende Abstände berechnet werden, die sich bei statistischer

Unabhängigkeit einstellen würde. Die Frage, ob die Abweichungen zwischen $h_j(\geq t)$ und $h_{uj}(\geq t)$ nur die Auswirkung einer endlich großen Stichprobe sind, oder ob eine systematische Abweichung vorliegt, kann mit dem χ^2 -Test auf die Hypothese gleicher Verteilungen entschieden werden. Falls der χ^2 -Test die Annahme gleicher Verteilungen zuläßt, dann darf für die Berechnung von Abstandssummen mit statistischer Unabhängigkeit gerechnet werden.

Die praktische Berechnung der Häufigkeitsverteilung $h_{uj}(\geq t)$ aus $h(\geq t)$ könnte durch Faltung erfolgen. Jedoch wird der Rechenaufwand unverhältnismäßig groß. Daher wurde die Folge der Abstände in der Messung nachträglich in zufälliger Weise permutiert. Unter Annahme der statistischen Unabhängigkeit hätte sich diese permutierte Folge der Abstände mit gleicher Wahrscheinlichkeit als originale Messung einstellen können. Aus der permutierten Folge kann $h_{uj}(\geq t)$ in gleicher Weise ermittelt werden, wie $h_j(\geq t)$ aus der originalen Folge.

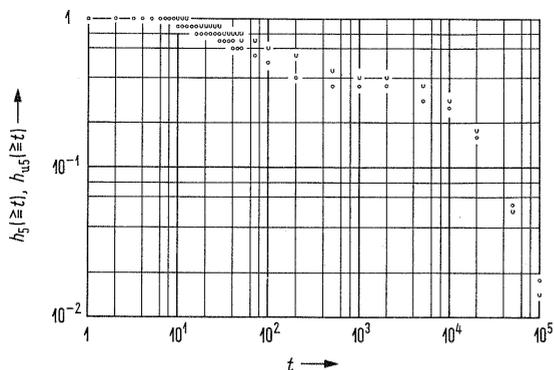


Bild 7. Gemessene Verteilungsfunktion $h_5(\geq t)$ für die Summe von $j = 5$ aufeinanderfolgenden Abständen ($\circ \circ$). Die Verteilungsfunktion $h_{u5}(\geq t)$, die sich bei statistisch unabhängig angenommenen Abständen ergeben würde, ist durch (u u) dargestellt, soweit sie nicht auf die gemessene Abstandsverteilung fällt. Auffällig ist der steilere Abfall von $h_{u5}(\geq t)$.

In Bild 7 sind die Verteilungsfunktion $h_5(\geq t)$ für $j = 5$ aufeinanderfolgende Abstände der originalen Messung sowie die Verteilungsfunktion $h_{u5}(\geq t)$ bei Annahme der statistischen Unabhängigkeit zwischen den Fehlerabständen in doppelt-logarithmischem Maßstab dargestellt. Auffällig ist der steilere Abfall der Verteilungsfunktion $h_{u5}(\geq t)$, was einer kleineren Varianz für die Summe von Abständen bei Annahme der statistischen Unabhängigkeit gegenüber der originalen Messung entspricht. Dies ist eine Auswirkung der positiven Kovarianz, die durch die bevorzugte Folge eines langen Fehlerabstandes auf einen langen Fehlerabstand verursacht wird.

Neben diesem bildlichen Eindruck für die Abweichung soll die Signifikanz a des χ^2 -Tests auf die Verteilungshypothese für einige Werte von j Abständen vorgestellt werden (Tabelle III).

Die Tabelle III zeigt, daß für die Summe von zwei benachbarten Abständen die Annahme der

Unabhängigkeit sogar auf Grund des kritischen χ^2 -Tests nicht verworfen werden muß. (Bei Verwerfung würde man sich mit einer Wahrscheinlichkeit von über 4% täuschen.) Erst bei der Summe von $j \geq 3$ Abständen wird der Einfluß der Abhängigkeit signifikant.

Tabelle III.

Zahl der summierten Abstände j	Signifikanz auf gleiche Verteilungen a
2	0,044
3	< 0,001
5	< 0,001
10	< 0,001

Zusammenfassend kann festgestellt werden, daß die Annahme von statistisch unabhängigen Fehlerabständen als eine erste Näherung an die Wirklichkeit angesehen werden darf. Von dieser Näherungsannahme wird jedoch die ausgezeichnete Approximation der Verteilung einfacher Abstände nicht berührt.

3.4. Erweiterung des vorgeschlagenen Modells

Das bisher vorgestellte Modell für die Fehler bei Datenübertragung befriedigt hinsichtlich der Verteilung der Abstände zwischen den Fehlerstellen der Messung ausgezeichnet. Die Annahme der statistischen Unabhängigkeit zwischen den Fehlerabständen ist eine gute Näherung.

Für besondere Anwendungen könnte der Wunsch entstehen, diese, wenn auch geringe, statistische Abhängigkeit in einem erweiterten Modell zu berücksichtigen, um in dieser Hinsicht die Statistik der Übertragungsfehler noch besser nachzubilden.

In diesem erweiterten Modell bleiben alle bisher angeführten Eigenschaften (siehe Abschnitte 3.1, 3.2 und 3.2.1) mit Ausnahme der statistischen Unabhängigkeit erhalten: Der Kanal befindet sich in einem von g Zuständen Z_i ($i = 1, 2, \dots, g$), wobei im Zustand Z_i die Fehlereinfallwahrscheinlichkeit p_{Ei} ($i = 1, 2, \dots, g$) herrscht. Ein Zustandswechsel kann nur nach einer gefälschten Binärstelle auftreten, so daß während eines Fehlerabstandes der Zustand derselbe bleibt. Die Wahrscheinlichkeit, daß der Zustand nach einem Fehlerbit in den Zustand Z_i übergeht, ist W_i ($i = 1, 2, \dots, g$). Damit behalten die Gl. (15) bis (23) weiterhin ihre Gültigkeit.

Im Gegensatz zu dem einfachen Modell soll nun bei dem erweiterten Modell der Zustand nach einer gefälschten Binärstelle Z_j von dem vorangehenden Zustand Z_i abhängen. Der Übergang des Zustandes von Z_i nach Z_j soll durch die Übergangswahrscheinlichkeiten

$$p_{ij} \quad (i, j = 1, 2, \dots, g) \quad (25)$$

gesteuert werden. p_{ij} ist die Wahrscheinlichkeit für Z_j unter der Bedingung, daß der vorangehende Zustand Z_i war. Für diese Übergangswahrscheinlich-

keiten gilt

$$\sum_{j=1}^g p_{ij} = 1 \quad (i = 1, 2, \dots, g). \quad (26)$$

Für den Sonderfall der Unabhängigkeit sind die Übergangswahrscheinlichkeiten und die Zustandswahrscheinlichkeiten gleich:

$$p_{ij} = W_j \quad (i, j = 1, 2, \dots, g). \quad (27)$$

Durch die Übergangswahrscheinlichkeiten p_{ij} werden die Zustandswahrscheinlichkeiten W_j festgelegt [17], denn es gilt

$$W_j = \sum_{i=1}^g W_i p_{ij} \quad (j = 1, 2, \dots, g). \quad (28)$$

Die zeitliche Folge der einzelnen Zustände bildet in diesem erweiterten Modell eine sogenannte Markoff-Kette. Eine solche Markoff-Kette für Simulationszwecke auf einem Digitalrechner künstlich zu erzeugen, bereitet keine besondere Schwierigkeit [18], [19], [20], wenn die Übergangswahrscheinlichkeiten bekannt sind. Falls der augenblickliche Zustand Z_i ist, dann muß ein Zufallsereignis erzeugt werden mit g möglichen Ausgängen und den zugehörigen Wahrscheinlichkeiten

$$p_{i1}, p_{i2}, \dots, p_{ij}, \dots, p_{ig}.$$

Trifft Ausgang Nummer j ein, so wird Z_j als nächster Zustand angenommen.

Mehr Schwierigkeit bereitet die geeignete Wahl der Übergangswahrscheinlichkeiten p_{ij} . Durch Vorgabe der Fehlerabstandsverteilung $p(\geq t)$ sind nach Gl. (16) die Zustandswahrscheinlichkeiten W_i festgelegt. Damit verbleiben in Gl. (27) für die Wahl der p_{ij} noch viele Freiheitsgrade, weit mehr als notwendig, um Abhängigkeiten der Form „auf einen langen Abstand folgt bevorzugt wieder ein langer Abstand“ auszudrücken.

Um das Modell nicht unnötig zu komplizieren, kann die Zahl der Freiheitsgrade durch freiwillige Beschränkung reduziert werden, z.B. durch Einteilen der Zustände in zwei Klassen: Die eine Klasse enthalte nur den Zustand mit der kleinsten Fehlerwahrscheinlichkeit, was im Mittel auf die längsten Fehlerabstände führt; die andere Klasse enthalte alle anderen Zustände. Der Zustandswechsel innerhalb der gleichen Klasse sei vom vorangehenden Zustand unabhängig. Dagegen trete der Übergang von einer Klasse zur anderen mit kleinerer Wahrscheinlichkeit auf als im Fall der Unabhängigkeit. Auf diese Weise folgt auf den Zustand mit schwächster Störung bevorzugt der gleiche Zustand und damit auch auf einen langen Fehlerabstand bevorzugt wiederum ein langer Fehlerabstand. Bei derart beschränkten Freiheitsgraden bleibt bei vorgegebenen Zustandswahrscheinlichkeiten W_i nur ein Parameter frei, um das Maß der Abhängigkeit auszudrücken.

Die letztgenannte „Zweiklassen-Methode“ für das erweiterte Modell wurde auf dieselbe Messung angewendet [16]. Dabei zeigte sich, daß das Modell durch Hinzunahme einer derartigen Abhängigkeit noch merkbar verbessert werden kann.

4. Abschließende Bemerkung

Das neue statistische Modell zur Beschreibung der Fehler bei Datenübertragung wurde in dieser Arbeit im wesentlichen anhand einer einzigen Messung eingeführt und begründet. Diese eine Messung steht jedoch hier stellvertretend für über 200 ausgeführte Messungen. In einer weiteren Veröffentlichung [9] wird die Zweckmäßigkeit dieses Modells anhand weiterer Messungen noch erhärtet werden. Ferner wird darin eine Statistik über die für das Modell ermittelten Parameter vorgelegt.

Schrifttum

- [1] AULHORN, H., LANGE, H. und MARKO, H., Probleme und Anwendungen der Datenübertragung. Elektron. Rechenanl. **3** [1961], 148–159.
- [2] GILBERT, E. N., Capacity of a burst-noise channel. Bell Syst. tech. J. **39** [1960], 1253–1265.
- [3] BERGER, J. M. und MANDELBROT, B., A new model for error clustering in telephone circuits. IBM J. Res. Develop. **7** [1963], 224–236.
- [4] SUSSMAN, S. M., Analysis of the Pareto-Model for error statistics on telephone circuits. Transact. Inst. Elect. Electron. Engrs. CS-**11** [1963], 213–221.
- [5] LEWIS, P. A. W. und COX, D. R., A statistical analysis of telephone circuit error data. Transact. Inst. Elect. Electron. Engrs. COM-**14** [1966], 382–389.
- [6] ALEXANDER, A. A., GRYB, R. M. und MAST, D. W., Capabilities of the telephone network for data transmission. Bell Syst. tech. J. **39** [1960], 431–476.
- [7] FONTAINE, A. B. und GALLAGER, R. G., Error statistics and coding for binary transmission over telephone circuits. Proc. Inst. Radio Engrs. **49** [1961], 1059–1065.
- [8] TOWNSEND, R. L. und WATTS, R. N., Effectiveness of error control in data communication over the switched network. Bell Syst. tech. J. **43** [1964], 2611–2638.
- [9] SWOBODA, J., Messung und Analyse der Fehler bei Datenübertragung auf Fernsprechanaläen. A.E.Ü. **23** [1969], im Druck.
- [10] GNEDENKO, B. W., Lehrbuch der Wahrscheinlichkeitsrechnung. Akademie-Verlag, Berlin 1957.
- [11] VAN DER WAERDEN, B. L., Mathematische Statistik. Springer-Verlag, Berlin 1965.
- [12] HOHNKER, W., ALGOL-Prozedur MINIMUM. Rechenzentrum der Universität Stuttgart 1965.
- [13] POWELL, M. J. D. und FLETCHER, R., A rapidly convergent descend method for minimization. Computer J. **6** [1963], 163–168.
- [14] UNGER, C., Mathematische Analyse der Fehler bei Datenübertragung. Diplomarbeit am Institut für Nachrichtenvermittlung und Datenverarbeitung der Techn. Hochschule Stuttgart, 1966.
- [15] BOLLINGER, W., Störmodelle eines Fernsprechanals, der für binäre Datenübertragung benutzt wird. Studienarbeit am Institut für Nachrichtenvermittlung und Datenverarbeitung der Techn. Hochschule Stuttgart, 1964.
- [16] KEIL, E., Statistische Modelle für die Fehler bei Datenübertragung. Test der Hypothesen anhand von Messungen. Diplomarbeit am Institut für Nachrichtenvermittlung und Datenverarbeitung der Techn. Hochschule Stuttgart 1967.
- [17] FELLER, W., An introduction to probability theory and its applications. J. Wiley & Sons, New York 1961.
- [18] KRAUS, G., Experimentelle stochastische Prozesse. A.E.Ü. **21** [1967], 19–22.
- [19] KÖRNER, A., LINSBAUER, U., SCHAFFER, B. und WEHRMANN, W., Elektronische Erzeugung stochastischer Stufenprozesse, die durch stationäre Markoffsche Ketten bestimmt sind. A.E.Ü. **21** [1967], 21–30.
- [20] SWOBODA, J., Elektrische Erzeugung von Zufallsprozessen mit vorgebbaren statistischen Eigenschaften. A.E.Ü. **16** [1962], 135–148.
- [21] MERTZ, P., Model of error-burst structure in data transmission. Proc. Nat. Electronics Conf. **16** [1960], 232–240.