

3.8 Kanalkapazität. Die Sätze von McMillan, Feinstein und Shannon.

Dieses Kapitel soll dem mathematisch etwas stärker interessierten Leser den Übergang zum Schrifttum der allgemeinen, mathematischen Informationstheorie ebnen, an welcher in den vorangegangenen Kapiteln einige (in Kapitel 1.3 hervorgehobene) Anpassungen an die Anwendungsbedürfnisse der Kommunikationskybernetik vorgenommen wurden. Diese *Anpassungen* sind selbstverständlich keine Verfälschungen (also nicht von vorneherein *mathematisch* falsch). Sie bestehen vielmehr aus kommunikationskybernetisch benötigten *Unterscheidungen* (vor allem zwischen Information und Unsicherheit), *Verallgemeinerungen* (insbesondere auf die subjektive Information) und Verknüpfungen mit *zeichentheoretischen Begriffen* (insbesondere mit dem Begriff der Superierung). Dies alles fehlt in den klassischen Schriften zur mathematischen Informationstheorie.

Die im Folgenden referierten und (wenngleich nur teilweise und nur für naheliegende Spezialfälle) bewiesenen Theoreme fanden bisher in der Kommunikationskybernetik vergleichsweise wenig Anwendungsmöglichkeiten. Sie bilden aber Kernkenntnisse der allgemeinen Kybernetik auf ihrer fundamentalen, nämlich auf der informationstheoretischen Stufe. Für den Kommunikationskybernetiker stellen sie seine Werkzeuge in den größeren Zusammenhang einer allgemeineren, kohärenten Theorie. Deshalb wird die zeichentheoretische Terminologie (vor allem der Begriff des Superierens, also der Superzeichenbildung) beibehalten, die zu Beginn der sechziger Jahre des letzten Jahrhunderts eigens für kommunikationskybernetische Zwecke entwickelt worden war - und dementsprechend in den vorausgehenden Kapiteln benutzt wurde.

3.8 Kanalkapacito. La teoremoj de McMillan, Feinstein kaj Shannon.

Ĉi tiu ĉapitro ebenigu por la leganto havanta iom pli fortajn matematikajn interesojn la vojon al la literaturo de la ĝenerala, matematika teorio de informacio, kiu estis en la antaŭaj ĉapitroj je kelkaj lokoj (reliefigitaj en la ĉapitro 1.3) adaptita al la aplikadbezonoj de la komunikadkibernetiko. Ĉi tiuj *adapttoj* kompreneble ne estas *falsigoj* (do ne aŭtomate *matematike* malkorektaj). Pli ĝuste ili konsistas el komunikadkibernetike bezonataj *diferencigoj* (precipe inter informacio kaj necerteco), *ĝeneraligoj* (precipe al la subjektiva informacio) kaj kunligoj kun *semiotikaj nocioj* (precipe kun la nocio de supreniĝo). Ĉio ĉi mankas en la klasikaj verkoj de la matematika teorio pri informacio.

La teoremoj ĉisekve sciigitaj kaj (kvankam nur parte kaj nur por sin-sugestaj specialaj kazoj) pruvitaj ĝis nun trovis relative malmultajn aplikadoblojn en la komunikadkibernetiko. Sed ili konstituas kernajn ekkonojn de la ĝenerala kibernetiko sur ties baza, nome sur la informaciteoria ŝtupo. Por la komunikadkibernetikisto ili enkadrigas lian ilaron en la pli ampleksan rilataron de ĝenerala, kohera teorio. Tial estas plue aplikita la semiotika terminologio (precipe la nocio de la supreniĝo, do de la kreado de kunsignoj, alivorte: de „supersignoj“), kiu estis evoluigita, komence de la sesdekaj jaroj de la pasinta jarcento, speciale por komunikadkibernetikaj celoj - kaj estis konforme aplikata en la antaŭaj ĉapitroj.

Vorab seien die Fragen motiviert, formuliert und (noch „ohne Wenn und Aber“, also vorläufig, für den mathematisch uninteressierten Leser jedoch hinreichend genau) beantwortet, um die es sich anschließend handelt. Diese noch verbliebenen Fragen, zu denen in den vorausgegangenen Kapiteln nur Vermutungen aufgestellt und in Spezialfällen bestätigt wurden, betreffen den Zusammenhang zwischen der Information als einem *exakt* definiertem Maß der *Unvorhersehbarkeit* und der Information als einem mindestens *ungefähren* Maß der *Transportschwierigkeit* derselben Nachrichten.

Auf das exakt definierte Unvorhersehbarkeitsmaß wurden in Teil 2 u.a. Maße (1) der Unsicherheit, (2) der Knappheit bzw. Redundanz und (3) der Transinformation aufgebaut.

Mit dem *Unsicherheitsmaß* konnte in Kapitel 3.2 eine *untere Schranke* des Codieraufwands bestimmt werden, die sich in Spezialfällen sogar als *untere Grenze* erwies. Allgemein liegt sie bei zeichenweiser Optimalcodierung um weniger als 1 niedriger als der so minimalisierte Codieraufwand. Das führt zur

Frage 1: Kann dieser Abstand zwischen H/bit und $M(I)/\text{Bit}$ durch immer sparsamere Codierung beliebig klein gemacht werden, so dass Unsicherheit und Codieraufwand tatsächlich quantitativ übereinstimmen?

Die Antwort lautet aufgrund des Satzes von McMillan: Ja. -

Die Bezeichnungen „Knappheit“ und „Redundanz“ suggerieren, es könne nicht nur gleich *viel* Information sondern „*die-selbe* Information“ mit demselben Zeichenrepertoire *knapper* formuliert, also die Zeichenfolge, welche die Nachricht bildet, durch Verzicht auf Weitschweifigkeit *gekürzt* werden. Daher die

Frage 2: Kann eine Nachricht verlustfrei,

Anticipate estu motivataj, vortigataj kaj respondataj (ankoraŭ sen „se“ kaj „sed“, do provizore, tamen por la matematike neinteresita leganto sufiĉe precize) la demandoj, pri kiuj temos pli sube. Ĉi tiuj ankoraŭ restantaj demandoj, pri kiuj la antaŭaj ĉapitroj nur starigis konjektojn konfirmante ilin en specialaj kazoj, koncernas la rilaton inter la informacio kiel *ekzakte* difinita mezuro de la *neatenditeco* kaj la informacio kiel almenaŭ *proksimuma* mezuro de la *transportmalfacileco* de la samaj informoj.

En parto 2 estis starigataj sur la ekzakte difinitan mezuron de la neatenditeco i.a. mezuroj de (1) la necerteco, (2) la koncizeco resp. redundo kaj (3) la transinformacio.

Per la mezuro de la *necerteco* eblis en ĉapitro 3.2 determini *suban baron* de la kodadkostoj, kiu montriĝis en specialaj kazoj eĉ kiel *infimo*. Ĝi ĉiam troviĝas kaze de posigna optimuma kodado je malpli ol 1 sub la tiel minimumigitaj kodadkostoj. Tial la

Demando 1: Ĉu eblas, ĉi tiun distancon inter H/bit kaj $M(I)/\text{Bit}$ laŭplaĉe malgrandigi per pli kaj pli ŝpariga kodado, tiel ke necerteco kaj kodadkostoj fakte laŭkvante koincidas?

La respondo estas pro la teoremo de McMillan: Jes. -

La esprimoj „koncizeco“ kaj „redundo“ sugestas la eblon, per la sama signorepertuaro esprimi *pli dense* ne nur *sammulton* da informacio sed „*la saman* informon“. do *mallongigi* per rezigno pri redundo la signosinsekvon, kiu konstituas la mesaĝon. Tial la

Demando 2: Ĉu mesaĝo transformeb-

also eindeutig umkehrbar, in eine Zeichenfolge desselben Repertoires so verwandelt werden, dass dabei die Länge *auf* den durch die *Knappheit* gemessenen Prozentsatz bzw. *um* den durch die *Redundanz* gemessenen Prozentsatz sinkt?

Aufgrund der Umkehrung des Satzes von McMillan ist auch diese Frage zu bejahen. -

Die Bezeichnung „Trans“-Information suggeriert die Vorstellung eines Hinüberbringens von Information durch einen Kanal. Dieser übermittelt eine Nachricht, sie dabei eventuell verändernd, von einem Ort zu einem anderen (Beispiel: Fernschreiber). oder er speichert sie von einem Zeitpunkt zu einem späteren (Beispiel: Lochkarte), oder er erfüllt beide Aufgaben zugleich (Beispiel: Internet), oder er ordnet (wie eine Schreibmaschine) in anderer Weise einer Folge von Eingabezeichen aus einem Feld Z (im Beispiel: den Tastenanschlägen) eine Folge von Ausgabezeichen aus einem möglicherweise anderen Feld Y zu (im Beispiel: Druckschriftzeichen). Treten dabei keine Störungen auf, dann trägt der Kanal zur Transportschwierigkeit allenfalls wegen seines nur endlich kleinen Zeitbedarfs τ zur Übertragung eines Zeichens bei (und eventuell durch die beschränkte Zeit, während welcher er dem Nutzer zur Verfügung steht, also wegen der endlichen Höchstzahl nutzbarer Zeitquanten), bzw. wegen seines nur endlich kleinen Platzbedarfs σ zur Speicherung eines Zeichens (und eventuell wegen des beschränkten Speicherplatzes, den er zur Verfügung stellt, also wegen der endlichen Höchstzahl nutzbarer Speicherzellen). Aus den Antworten auf die beiden ersten Fragen folgt, dass die Information einer Nachricht in beiden Fällen ein (relatives) Maß für die (durch Zeit- oder Platzbedarf gemessene) Schwierigkeit ist, die Nachricht (durch einen bestimmten Kanal, gekennzeichnet durch eine bestimmte „Kapazität“ C) „hin-

las sen perdo, do senambigue inversig-eble, en signosinsekvon el la sama repertuaro tiel, ke la longeco falas *al* la procentaĵo mezurita kiel *koncizeco* resp. *je* la procentaĵo mezurita kiel *redundo*?

Pro la inversigo de la teoremo de McMillan ankaŭ ĉi tiu demando estas jese respondenda. -

La esprimo „trans“-informacio sugestas la imagon de transenportado de informacio tra kanalo. Ĉi tiu transigas mesaĝon, ĝin eventuale dume ŝanĝante, de loko al alia loko (ekzemplo: teltajpilo), aŭ ĝi storas ĝin de unu tempopunkto al pli malfrua (ekzemplo: trukarto), aŭ ĝi plenumas ambaŭ taskojn samtempe (ekzemplo: interreto), aŭ ĝi alie alordigas (kiel skribmaŝino) al sinsekvo de enigsignoj el kampo Z (en la ekzemplo: al klavfrapoj) sinsekvon de eligsignoj el eble alia kampo Y (en la ekzemplo: presliteroj). Se al tio ne okazas perturboj, tiam la kanalo kontribuas al la transportmalfacileco pro nenio alia ol pro sia nur finie malgranda tempobezono τ por transigi signon (kaj eventuale pro la limigita tempo, dum kiu ĝi estas je dispono al la uzanto, do pro la finia maksimumo de uzeblaj temperoj), resp. pro sia nur finie malgranda spacbezono σ por stori signon (kaj eventuale pro la limigita storspaco disponigata, do pro la finia maksimumo de uzeblaj storĉeloj). El la respondoj al la du unuaj demandoj sekvas, ke la informacio de mesaĝo ambaŭkaze estas (relativa) mezuro por la malfacilo (mezurita per la bezonata tempo aŭ spaco), „transigi“ la mesaĝon (tra certa kanalo, karakterizita per certa „kapacito“ C). Pli precize: La informacio proporcias al la sendad-

überzubringen“. Genauer: Die Information ist proportional zur benötigten Sendezeit ($i = Ct$) oder zum benötigten Prozentanteil am Gesamtspeicherplatz ($i = pC$) - mit der Kanalkapazität C als Proportionalitätsfaktor. Ist aber irgend einer der erwähnten Kanäle so gestört, dass das empfangene Zeichen zwar vom gesendeten noch stochastisch abhängt, jedoch die Unsicherheit über dieses nicht voll beseitigt, dann kann seine Kanalkapazität durch die *höchste* Transinformation gemessen werden, die bei verschiedenen Wahrscheinlichkeitsverteilungen der Sendezeichen erreichbar ist. Das führt zur

Frage 3: Kann eine Nachricht so codiert werden, dass sie mit beliebig geringer Fehlerwahrscheinlichkeit auch durch einen *gestörten* Kanal übertragbar ist, solange die von der Nachrichtenquelle pro Zeiteinheit gelieferte Information unter der Übertragungskapazität bleibt - bzw. wenn die insgesamt zu speichernde Information kleiner ist als die Speicherkapazität?

Die beiden Fundamentalsätze von Shannon (deren Beweise Chintschin auf das Theorem von Feinstein und jenes von McMillan aufbaut) lassen auch auf diese Frage die Antwort Ja zu. -

Wir wenden uns nun zunächst dem Satz von McMillan zu.

Eine Zeichenquelle liefere eine stochastisch unabhängige Folge von Zeichen aus einem Felde Z vom Umfang U . Wir zerlegen diese Folge in Ketten (Segmente) gleicher Länge s . Wir können sie als durch Komplexbildung entstandene Superzeichen auffassen. Es gibt U^s verschiedene solche Ketten. Sind sie genügend lang, dann tritt – mit einem beliebig kleinen Prozentsatz von Ausnahmen (genauer: mit einer beliebig kleinen Ausnahmewahrscheinlichkeit) – ein Zeichen z_k mit einer relativen Häufigkeit in einer bestimmten Kette auf, die etwa gleich

tempo ($i = Ct$) aŭ al la bezonata procentaĵo de la tuta storspaco ($i = pC$) - kun la kanalkapacito C kiel proporci-faktoro. Sed se ajna de la menciitaj kanaloj estas tiel perturbita, ke la ricevita signo ja stokaste dependas de la sendita, sed ne plene nuligas la necertecon pri ĉi tiu, tiam eblas mezuri ĝian kanalkapaciton per la *maksimumo* de la transinformacio, kiu estas atingebla per diversaj probablodistribuoj de la sendsignoj. Tio motivas la starigon de la

Demando 3: Ĉu eblas kodi mesaĝon tiel, ke ĝi estas transigebla ankaŭ tra *perturbita* kanalo kun laŭplaĉe malgranda erarprobablo, se la informacio havigita de la informfonto je tempo unuo malsuperas la transigkapaciton - resp. se la entute storenda informacio malsuperas la storkapaciton?

La du fundamentaj teoremoj de Shannon (kies pruvon Ĥinĉin starigas sur la teoremon de Feinstein kaj de McMillan) permesas jese respondi ankaŭ al ĉi tiu demando. -

Ni nun priokupiĝu unue pri la teoremo de McMillan.

Signofonto havigu stokastike sendependan sinsekvon de signoj el kam-po Z kun amplekso U . Ni disigas ĉi tiun sinsekvon en ĉenojn (segmentojn) kun egala longeco s . Ni povas ilin interpreti kiel kunsignojn ekestajn pro kompleksigo. Ekzistas U^s diversaj tiaj ĉenoj. Se ili estas sufiĉe longaj, tiam aperas – kun laŭplaĉe malgranda procentaĵo da esceptoj (pli precize: kun laŭ plaĉe malgranda esceptprobablo) – signo z_k en certa ĉeno kun relativa ofteco proksimume egala al sia

seiner Wahrscheinlichkeit p_k ist⁸.

Die Zahl seiner Auftritte in der Kette ist dann $N_k = sh_k \approx sp_k$. Die betreffende Kette hat also die Wahrscheinlichkeit

$$(3.8.1) \quad p = \prod_{k=1}^U p_k^{N_k} \approx \left(\prod_{k=1}^U p_k^{p_k} \right)^s$$

Sie enthält daher die in bit gemessene⁹ Information

$$(3.8.2) \quad \text{ld} \frac{1}{p} \approx s \cdot \sum_{k=1}^U p_k \cdot \text{ld} \frac{1}{p_k} = s \cdot H(Z)$$

Durch Auflösung nach p – oder schon direkt aus (3.8.1), wenn man dort p_k durch $2^{\text{ld} p_k} = 2^{-\text{ld} 1/p_k}$ ersetzt – erhält man

$$(3.8.3) \quad p \approx 2^{-s \cdot H}$$

Bis auf einen beliebig kleinen Prozentsatz von Ausnahmen enthalten also alle Ketten der Länge s ungefähr gleichviel Information, nämlich das s -fache der mittleren Information der Einzelzeichen. Sei q die Summe der Wahrscheinlichkeiten aller Ausnahmeketten. Dann haben die „normalen“ Ketten (deren Wahrscheinlichkeit also je etwa 2^{-sH} beträgt) zusammen die Wahrscheinlichkeit $1-q$. Es gibt also etwa $x = (1-q)2^{sH}$ solche Ketten – das sind *weniger* als 2^{sH} .

Wir konstruieren nun einen Bacon-Code für $2^{\lceil sH \rceil}$ Zeichen. Wir können damit alle „normalen“ Ketten codieren – jedes durch $\lceil sH \rceil$ Bit –, und wir behalten mindestens

probablo p_k ⁸.

Tiam ĝia aperofteco en la ĉeno estas $N_k = sh_k \approx sp_k$. La probablo de la koncerna ĉeno do estas

$$(3.8.1) \quad p = \prod_{k=1}^U p_k^{N_k} \approx \left(\prod_{k=1}^U p_k^{p_k} \right)^s$$

Ĝi sekve enhavas la en bit mezuritan⁹ informacion

$$(3.8.2) \quad \text{ld} \frac{1}{p} \approx s \cdot \sum_{k=1}^U p_k \cdot \text{ld} \frac{1}{p_k} = s \cdot H(Z)$$

Per solvo izoliganta p – aŭ jam rekte el (3.8.1), tie substituante p per $2^{\text{ld} p} = 2^{-\text{ld} 1/p}$ – oni ricevas

$$(3.8.3) \quad p \approx 2^{-s \cdot H}$$

Kun laŭplaĉe malgranda procentaĵo da esceptoj do ĉiuj ĉenoj kun longeco s enhavas sammulte da informacio, nome la s -oblon de la aritma informacio de la unupaj signoj. Estu q la sumo de la probabloj de ĉiuj esceptaj ĉenoj. Sekve la sumo de la probabloj de la „normalaj“ ĉenoj (havantaj la probablon po ĉirkaŭ 2^{-sH}) estas $1-q$. Ekzistas do ĉirkaŭ $x = (1-q)2^{sH}$ tiaj ĉenoj – tio estas malpli ol 2^{sH} .

Ni nun konstruas BACON-kodon por $2^{\lceil sH \rceil}$ signoj. Per ili ni povas kodi ĉiujn „normalajn“ ĉenojn – ĉiun per $\lceil sH \rceil$ Bit –, kaj restas al ni almenaŭ *unu* ti-

⁸Die Begründungen diesser Aussage und der nachfolgenden weiteren Rückgriffe auf wahrscheinlichkeitstheoretische Grundlagen ergeben sich aus dem in Kapitel 2.1 Gesagten. Es wird im Folgenden auf Einzelweise auf die dortigen Gleichungen und Ungleichungen verzichtet.

⁹Um die Formeln nicht unnötig zu überladen, verzichten wir (wie es im mathematischen Schrifttum meist geschieht) im Folgenden auf Dimensionsangaben, wo diese sich von selbst verstehen.

⁸ La pravigoj de ĉi tiu aserto kaj de la ĉisekvaj pluaj aplikoj de probablecteoraj bazoj sekvas el la enhavo de ĉapitro 2.1. En la sekva teksto estas rezignite pri unuopaj referencoj al la tiaj egalajoj kaj malegalajoj.

⁹ Por eviti sennecesan komplikigon de la formuloj ni ĉisekve rezignas (kiel plejofte en la matematika literaturo) pri indikoj de evidentaj dimensioj.

eine Binärfolge dieser Länge übrig, welche nicht als Codierung einer Kette benutzt wurde. Nun codieren wir getrennt auch das Repertoire der restlichen $U^s - x$ Ketten durch einen Bacon-Code, also nicht notwendig optimal, und setzen vor ihre Codewörter quasi als Präfix jeweils die nicht für eine „normale“ Kette benutzte Binärfolge. (Der Codebaum wird also für die Ausnahmeketten weiterverzweigt.) Damit erhalten alle Ausnahmeketten Codewörter der Länge

$$\lceil sH \rceil + \lceil \text{ld}(U^s - x) \rceil \leq \lceil sH \rceil + \lceil \text{ld } U^s \rceil = \lceil sH \rceil + \lceil s \cdot \text{ld } U \rceil$$

Bit. Der Codieraufwand (Erwartungswert der Codewortlänge) für alle (normalen und Ausnahme-)Ketten beträgt daher

$$\mathbf{M}(l) / \text{Bit} \leq (1-q) \lceil sH \rceil + q(\lceil sH \rceil + \lceil s \cdot \text{ld } U \rceil) = \lceil sH \rceil + q \lceil s \cdot \text{ld } U \rceil < \\ < sH + 1 + q(s \cdot \text{ld } U + 1) = s(H + q \cdot \text{ld } U) + 1 + q$$

Daraus folgt als Codieraufwand *pro Zeichen*

$$\mathbf{M}(l) / s < H + q \cdot \text{ld } U + (1 + q)/s$$

Bit/Zeichen.

Da q beliebig klein wird, wenn man die Kettenlänge genügend groß wählt, können wir die Ungleichung (3.2.5) nun durch folgende Feststellung ergänzen: Wird die Gleichheit bei einer direkten, optimalen Codierung der *einzelnen Zeichen* nicht erreicht, dann kann man einen *beliebig kleinen* Unterschied zwischen dem zu erwartenden Codieraufwand in Bit/Zeichen und der in bit/Zeichen gemessenen Unsicherheit erreichen, indem man *genügend lange Ketten* codiert. Dies war schon in Kapitel 3.4 vermutet worden. -

Der Beweis wurde hier für den Fall *stochastischer Unabhängigkeit* innerhalb der Zeichenfolge eines Textes skizziert. Dieser einfachste Fall ist kommunikationskybernetisch kaum interessant, da es sich bei ihm nur um Zeichenfolgen ohne

om longa binara sinsekvo, kiu ne estas uzita por kodi iun ĉenon. Nun ni kodas dise ankaŭ la repertuaron de la restaj $U^s - x$ ĉenoj per BACON-kodo, do ne nepre optimale, kaj ni antaŭmetas antaŭ ĉiun unuopan de iliaj kodvortoj kvazaŭ kiel prefikson la binaran sinsekvon ne uzitan por „normala“ ĉeno. (La kodarbo do pludisbranĉiĝas por la esceptaj ĉenoj.) Tiel la kodvortoj de la esceptaj ĉenoj eklongas

Bit. La kodadkostoj (ekspekto de la kodvortlongeco) do estas por ĉiuj (normalaj kaj esceptaj) ĉenoj

El tio sekvas kiel kodadkostoj *je signo*

Bit/signo.

Pro tio, ke q fariĝas laŭplaĉe malgranda, se oni elektas sufiĉe grandan ĉenlongecon, eblas nun kompletigi la malegalaĵon (3.2.5) per la sekva konstato: Se la egaleco ne estas atingita pro rekta, optimuma kodado de la *unuopaj signoj*, tiam eblas atingi *laŭplaĉe malgrandan* diferencon inter la ekspekto de la kodadkostoj en Bit/signo kaj la necerteco mezurita en bit/signo, se oni kodas *sufiĉe longajn ĉenojn*. Tion konjektigis jam la ĉapitro 3.4. -

La pruvo estas ĉi tie skizita por la kazo de *stokasta sendependeco* en la signo-sinsekvo de teksto. Ĉi tiu plej simpla kazo estas apenaŭ interesa por la komunikadkybernetiko, ĉar en ĝi temas nur pri signo-sinsekvoj sen inter-

inneren Zusammenhang handelt. Schon Shannon (1948) bewies die Gültigkeit der Feststellung auch für den allgemeineren Fall bestehender stochastischer Abhängigkeit vom unmittelbaren Vorgängerzeichen. McMillan (1953) bewies schließlich, dass das Gesagte für alle Texte einer stationären ergodischen Quelle¹⁰ gilt. Daher benennt Chintschin (1956) den Satz nach McMillan. -

Jeder Nachrichtenübertragungskanal hat eine bestimmte *Kanalkapazität* (Chintschin formuliert strenger: ergodische Durchlasskapazität). Bei einem Kanal ohne Störung und mit gleicher Übertragungszeit τ für jedes der u verschiedenen, zur Übertragung benutzten Zeichen ist sie

$$(3.8.4) \quad C = (\text{ld } u \text{ bit}) / \tau$$

also berechenbar als Produkt der Zahl der Zeichen, die pro Zeiteinheit (z.B. pro Sekunde) durch den Kanal geleitet werden können, und dem Logarithmus dualis des Repertoireumfangs der Übertragungszeichen. Diesen *Höchstwert* überträgt der Kanal natürlich nur, wenn die Quelle *nicht weniger* Information liefert. Entstammen der Quelle umgekehrt pro Zeiteinheit *mehr* Zeichen, als der Kanal zu übertragen vermag, dann kann bei Vorliegen von ausreichender Redundanz eine Umcodierung so vorgenommen werden, dass eventuell doch eine Übertragung möglich ist. Das zeigt das folgende Beispiel.

Nehmen wir an, eine Quelle sende die 8 Zeichen lange Vokalfolge von Kapitel 1.2 -

naj interrilatoj. Jam Shannon (1948) pruvis la valideco de la konstato ankaŭ por la pli ĝenerala kazo de stokasta dependeco de la senpere antaŭiranta signo. McMillan (1953) fine pruvis, ke la dirito validas por ĉiuj tekstoj de stacionara, ergoda fonto¹⁰. Ĥinĉin (1956) tial nomas la teoremon laŭ McMillan.-

Ĉiu mesaĝtransiga kanalo havas certan *kanalkapaciton* (Ĥinĉin vortigas pli strikte: ergoda trairkapacito). Kaze de neperturbita kanalo, kiu bezonas la saman transigtempon τ por ĉiu de la u diversaj signoj uzitaj por la transigo, ĝi estas

do kalkulebla kiel produto de la nombro de signoj transigeblaj en unu tempounuo (ekz. en unu sekundo) tra la kanalo, kaj la duuma logaritmo de la repertuaramplekso de la transigsignoj. Ĉi tiun *maksimumon* la kanalo kompreneble nur transigas, se la fonto liveras *ne malpli* da informacio. Se inverse en unu tempounuo devenas *pli* da signoj de la fonto, ol povas transigi la kanalo, tiam la ekzisto de sufiĉa redundo ebligas alikodadon tian, ke eventuale tamen la transigo eblas. Tion montras la sekvanta ekzemplo.

Ni supozu, fonto sendus la 8 signojn longan vokalsinsekvon de ĉapitro

¹⁰ Eine Quelle heißt *ergodisch*, wenn die Wahrscheinlichkeit, mit der sie *jetzt* ein bestimmtes Zeichen liefert, von dem Zeichen, das sie an einer bestimmten *früheren* Stelle realisierte, beliebig wenig abhängt, falls sie nur seither genügend viele weitere Zeichen lieferte. - *Stationär* heißt die Quelle, wenn die Wahrscheinlichkeiten, mit denen sie die Zeichen liefert, außer von den Vorgängerzeichen nicht auch noch vom Zeitpunkt der Zeichenlieferung abhängen (wenn also die Quelle einstweilen keinem „Reifungsprozess“ - oder Alterungsprozess - unterlag).

¹⁰ Fonto nomitas *ergoda*, se la probablo, laŭ kiu ĝi *nun* liveras certan signon, laŭplaĉe malmulte dependas de la signo, kiun ĝi liveris je certa *pli frua* loko, se ĝi nur havigis intertempe sufiĉe multajn pluajn signojn. - *Stacionara* nomitas la fonto, se la probabloj, laŭ kiuj ĝi liveras la signojn, dependas krom de la antaŭirintaj signoj ne ankaŭ de la tempopunkto de la signo-havigo (se do intertempe ĉe la fonto ne okazis „maturiĝprocezo“ - aŭ maljuniĝprocezo).

also UOEAEAE – innerhalb von 1 Sekunde, der Kanal könne aber nur 7 Zeichen dieses Repertoires pro Sekunde übertragen, benötige also $\tau = 1/7$ Sekunden pro Zeichen. Da der Vorrat der übertragbaren Zeichen den Umfang $u = 4$ hat, berechnet sich nach (3.8.4) seine Kanalkapazität zu 2 bit pro siebtel Sekunde, also zu $C = 14$ bit/sek. Enthält (je unabhängig vom Vorgängerzeichen) E 1 bit Information, A 2 bit und O und U je 3 bit, dann enthält die gesamte Vokalfolge gerade 14 bit Information, müsste also in *einer* Sekunde übertragbar sein. (H berechnete sich ja zu nur 1,75 bit pro Vokal, also ergab sich eine Redundanz von $12,5\% = 1/8$, um welche sich die Nachricht komprimieren lässt – die Knappheit der Vokalfolge beträgt ja nur $7/8$.) Folgende Umcodierung erlaubt dies.

Bei Verwendung des Optimalcodes von Bild 1.2 wird die Codierung 14 Bit lang. Sie kann in einer Sekunde durch einen Kanal übertragen werden, der *Binärzeichen* überträgt und, wie der betrachtete Kanal, die Kapazität 14 bit/sek hat. Da dieser aber in jeder siebentel Sekunde einen der vier *Vokale* überträgt, codieren wir nach dem Bacon-Code von Bild 1.2c die 14 Bit lange Binärzeichenfolge in eine 7 Vokalzeichen lange Folge um. Das Ergebnis lautet UOEAEOA. Das kann der Kanal innerhalb einer Sekunde übertragen. Am Kanalausgang muss die komprimierte Vokalfolge zunächst nach dem Bacon-Code in die 14 Bit lange Binärfolge entschlüsselt, und diese dann nach dem Huffman-Code als ursprüngliche Folge von 8 Vokalen decodiert werden. -

Sei $H(Z)/\tau$ die „Ergiebigkeit“ (mittlere gelieferte Information) der Quelle (in unserem Beispiel also 1,75 bit pro Siebtelsekunde oder 14 bit/sek), $H(Z|Y)$ die bedingte Unsicherheit über das, was gesendet wurde, falls man die vom Kanal über-

1.2 – do UOEAEAE – dum 1 sekundo, sed la kanalo povus en unu sekundo transigi nur 7 sigojn de ĉi tiu repertuaro, ĝi do bezonus $\tau = 1/7$ sekundojn je signo. Pro tio ke la repertuaro de la traireblaj signoj ampleksas $u = 4$, kalkuliĝas laŭ (3.8.4) kiel kanalkapacito $C = 2$ bit / (1/7) sek = 14 bit/sek. Se (ĉiam sendepende de la antaŭiranta signo) E enhavas 1 bit da informacio, A 2 bit, O kaj U po 3 bit, tiam la tuta vokalsinsekvo enhavas ĝuste 14 bit da informacio, devus do esti transigebla dum unu sekundo. (H ja estas nur 1,75 bit je vokalo, rezultis do la redundo $12,5\% = 1/8$, je kiu eblas kunpremi la mesaĝon – la koncizeco de la vokalsinsekvo ja estas nur $7/8$.) Tion ebligas la jena alikodado.

Aplikinte la optimuman kodon de bildo 1.2 la kodaĵo longas 14 Bit. Ĝi estas transigebla tra kanalo, kiu transigas *binarajn* sigojn kaj havas, kiel la konsiderata kanalo, la kapaciton 14 bit/sek. Sed ĉar ĉi tiu kanalo transigas dum ĉiu sepona sekundo unu el la kvar *vokaloj*, ni alikodas laŭ la BACON-kodo de bildo 1.2c la 14 Bit longan ĉenon da binaraj signoj en 7 vokalsignojn longan sinsekvon. La rezulto estas UOEAEOA. Tion la kanalo povas transigi en unu sekundo. Ĉe la kanalelirejo la densigita vokalsinsekvo devas esti unue malkodita laŭ la BACON-kodo en la 14 Bit longan binarsinsekvon, kaj tiam ĉi tiu laŭ la HUFFMAN-kodo en la originan sinsekvon de 8 vokaloj. -

Estu $H(Z)/\tau$ la „produktiveco“ (aritime havigata informacio) de la fonto (en nia ekzemplo do 1,75 bit je sepona sekundo aŭ 14 bit/sek), $H(Z|Y)$ la kondiĉita malcerteco pri tio, kio estis sendata, kondiĉe ke oni konas la sig-

tragenen Zeichen y_k kennt („das, was herauskommt“ – in unserem Beispiel die tatsächlich übertragenen Vokale), dann kann man mit

$$(3.8.5) \quad C = \max \left(\frac{H(Z)}{\tau} - \frac{H(Z|Y)}{\tau} \right) = \max \left(\frac{T(ZY)}{\tau} \right)$$

die (ergodische) Kapazität des Kanals bezeichnen, wenn das Maximum für alle (ergodischen) Quellen bestimmt wird.

Diese Definition ist so allgemein gefasst, dass das Feld Z der Eingabezeichen des Kanals und das Feld Y seiner Ausgabe nicht nur in der Wahrscheinlichkeitsverteilung sondern schon im Repertoire verschieden sein können. Die Eingabezeichen können ebenso wie die Ausgabezeichen Superzeichen sein, die durch Komplexbildung als Zeichenketten entstanden, wobei die Länge der Ketten ebenso wie die Repertoires der Unterzeichen am Eingang und am Ausgang des Kanals sich unterscheiden können. Nur die Eingabezeit τ für eine solche Kette soll dieselbe sein wie die Zeit, die für die Ausgabe der durch den Kanal zugeordneten Kette benötigt wird.

Falls das empfangene Zeichen stets gleich dem gesandten ist, also $y(t) = z(t)$ für alle Sendezeitpunkte gilt, - oder, allgemeiner: falls $z = f(y)$ gilt, also aus dem empfangenen Zeichen eindeutig auf das gesandte Zeichen geschlossen werden kann - ist die bedingte Unsicherheit $H(Z|Y)$ über das gesandte Zeichen natürlich 0, sobald es vom Kanal übertragen wurde. Da ferner das Maximum von $H(Z)$, wie in Kapitel 2.51 bewiesen wurde, $\text{ld } u$ ist, stimmt in diesem Spezialfall die Definition (3.8.5) mit (3.8.4) überein. Die Transinformation und damit die Kanalkapazität ist im anderen Extremfall 0, nämlich wenn das, was am Kanalausgang ankommt, stochastisch unabhängig ist von dem, was die Quelle sandte, wenn sich also die Unsicherheit über das gesandte

nojn y_k transigitajn de la kanalo („tion, kio eliras“ – en nia ekzemplo la fakte transigitajn vokalojn), tiam eblas difini per

la (ergodan) kapaciton de la kanalo, se estas determinata la maksimumo por ĉiuj (ergodaj) fontoj.

Ĉi tiu difino estas tiom ĝenerale vortigita, ke la kampo Z de la enigitaj signoj de la kanalo kaj la kampo Y de ĝia eligo povas diferenci ne nur en la probablodistribuo sed jam en la repertuaro. La enigitaj same kiel la eligitaj signoj povas esti kunsignoj, kiuj ekstis pro kompleksigo kiel signoĉenoj, tiel ke la longeco de la ĉenoj same kiel la repertuaroj de la subsignoj povas diferenci ĉe la enirejo kaj la elirejo de la kanalo. Nur la enigtempo τ por tia ĉeno estu la sama kiel la tempo bezonata por eligi la ĉenon alordigitan per la kanalo.

Se la ricevita signo ĉiam egalas al la sendita, se do $y(t) = z(t)$ validas por ĉiuj sendadtempopunktoj, - aŭ, ĝenerale: se validas $z = f(y)$, se do eblas senambigue konkludi el la ricevita signo al la sendita - la kondiĉita necerteco $H(Z|Y)$ pri la sendita signo evidente estas 0, ekde kiam ĝi estas transigita de la kanalo. Ĉar krome la maksimumo de $H(Z)$ estas $\text{ld } u$, kiel pruvite en la ĉapitro 2.51, en ĉi tiu speciala kazo la difino (3.8.5) koincidas kun (3.8.4). La transinformacio, sekve la kanalkapacito estas 0 en la alia ekstrema kazo: se tio, kio alvenas ĉe la eliro de la kanalo, estas stokaste sendependa de tio, kion sendis la fonto, se do la necerteco pri la sendita signo ne ŝanĝiĝas pro la

Zeichen durch das empfangene Zeichen nicht ändert. $H(Z|Y)$ ist also bei gegebenem $H(Z)$ ein Maß für die Gestörtheit des Kanals.

Falls Störungen nicht die gesamte Information zerstören, werden sie eine Zeichenkette sehr selten in *irgend eine* andere Zeichenkette verwandeln. Vielmehr wird es zu jeder Zeichenkette v_k eine Menge V_k von Zeichenketten geben, so dass die Wahrscheinlichkeit, dass die abgesandte Zeichenkette v_k in eine Kette aus V_k verwandelt wird, größer ist als $1-p$, wobei p für genügend lange Ketten beliebig klein ist, so dass jedenfalls $p < 1/2$ vorausgesetzt werden kann. Man wählt eine Menge disjunkter Mengen V_k aus und liefert dem Kanal nur die zugehörigen Zeichenketten v_k ein¹¹. Dann kann mit einer beliebig hohen Wahrscheinlichkeit $1-p$ erwartet werden, dass eine gesendete Kette v_k als eine Kette aus der Menge V_k empfangen wird, so dass der Empfänger annimmt, v_k sei gesandt worden. (Die Länge der Ketten kann beliebig groß gewählt werden, und dementsprechend auch die für ihre Eingabe erforderliche Zeit.)

Der Satz von Feinstein (1954) besagt nun, dass die Zahl z der unterscheidbaren Mengen V_k

$$(3.8.6) \quad z = 2^{s \cdot (C \cdot \tau - p)}$$

ist, wobei s die Kettenlänge, C die Kanalkapazität ist.

Durch Vergleich mit dem Satz von McMillan (demnach der Text fast nur die weniger als 2^{sH} Ketten mit je rund sH bit Information enthält) gewinnt schließlich

ricevita signo. Por fiksa $H(Z)$ do $H(Z|Y)$ estas mezuro de la perturbiteco de la kanalo.

Se perturboj ne ruinigas la tutan informon, ili tre malofte transformos signo-ĉenon en *ajnan* alian signo-ĉenon. Supozeble ekzistas por ĉiu signo-ĉeno v_k aro V_k da signo-ĉenoj, tiel ke la probablo, ke la sendita signo-ĉeno v_k estas transformata en ĉenon el V_k , estas pli granda ol $1-p$, kie p por sufiĉe longaj ĉenoj estas laŭplaĉe malgranda; oni do povas ĉiukaze supozi, ke $p < 1/2$. Oni elektas aron da disjunkciaj aroj V_k kaj enigis en la kanalon nur la respektivajn signoĉenojn¹¹. Tiam oni povas ekspekti laŭ laŭplaĉe alta probablo $1-p$, ke sendita ĉeno v_k alvenas kiel ĉeno el la aro V_k , tiel ke la ricevonto supozas, ke estis sendita v_k . (La longeco de la ĉenoj povas esti elektata laŭplaĉe granda, kaj konforme al tio ankaŭ la tempo bezonata por enmeti ilin.)

Nun konstatas la teoremo de Feinstein (1954), ke la nombro z de diferencigeblaj aroj V_k estas

kie s signas la longeco de la ĉenoj, C la kapaciton de la kanalo.

Per komparo kun la teoremo de McMillan (laŭ kiu la teksto preskaŭ nur enhavas la malpli ol 2^{sH} ĉenojn kun po ĉirkaŭ sH bit da informacio)

¹¹ Dies ist eine Verallgemeinerung der in Kapitel 3.7 beschriebenen fehlerkorrigierenden Codierung. In Bild 3.7b sind vier Codewörter (Zeichenketten) v_k markiert, sowie zu jeder davon die Codewortmenge V_k . Die im gegenwärtigen Kapitel vorgenommenen Verallgemeinerungen bestehen darin, dass die Codewörter nicht binär zu sein brauchen, und vor allem darin, dass es nicht mehr *sicher* sein muss, dass ein Codewort in höchstens *einer* Position gestört wird, - dafür reicht jetzt die Wahrscheinlichkeit $1-p > 1/2$ aus.

¹¹ Tio estas ĝeneraligo de la erarkorektiga kodado priskribita en ĉapitro 3.7. En bildo 3.7b estas markitaj kvar kodvortoj (signo-ĉenoj) v_k kaj al ĉiu de ili la kodvortaro V_k . La ĝeneraligoj faritaj en la prezenta ĉapitro konsistas en tio, ke la kodvortoj ne bezonas esti binaraj, kaj precipe en tio, ke ne plu estas *necese*, ke kodvorto estas perturbata en maksimume unu pozicio, - sufiĉas nun, ke tio veras laŭ probablo $1-p > 1/2$.

Chintschin aus (3.8.6) den Beweis für die Fundamentalsätze von Shannon:

1) Falls der Erwartungswert H der Information, die von der (ergodischen) Quelle gesandt wurde, kleiner ist als $C\tau$, kann man stets so codieren, dass mit beliebig hoher Wahrscheinlichkeit am Kanalende erraten werden kann, was die Quelle sandte (z.B. dadurch, das man die hochwahrscheinlichen Ketten des McMillanschen Satzes mit den unterscheidbaren Ketten v_k des Feinsteinschen Satzes codiert, wobei wenigstens eine Feinsteinsche Kette übrig bleibt für die nicht mehr unterschiedenen unwahrscheinlichen Ketten oder für einen Präfix zu deren Codierung).

2) Dabei kann man eine beliebig nahe bei H/τ liegende Übertragungsgeschwindigkeit erzielen, d.h. pro ankommendem Zeichen nahezu H bit von der Quelle stammender Information erhalten.

Eine höhere Übertragungsgeschwindigkeit als C ist durch keine Codierung erreichbar (Zaregradski, 1958). Über den Anstieg der Fehlerwahrscheinlichkeit bei $H > C$, informiert Kolmogoroff (1956).

Übungsaufgabe 3.8(1)

Ein binärer Kanal K^- überträgt während 7 TU (Zeiteinheiten, Zeitelementen, Zeitatomen – z.B. Subjektiven Zeitquanten im Sinne der Informationspsychologie) Kettchen der Länge $l = 7$ Binärzeichen, von denen durch irgendeinen Mechanismus mit Wahrscheinlichkeit $p = 10\%$ genau 1 gestört werden wird (wobei mit gleicher Wahrscheinlichkeit irgend eines der $l = 7$ Zeichen das gestörte sein kann). Mit Wahrscheinlichkeit $1-p$ ist überhaupt kein Element gestört. Man ergänzt den Kanal K^- zum Kanal K^{++} durch Zufügung einer Eingabe- und einer Ausgabevorrichtung. Die Eingabevorrichtung

Ĥinĉin finfine sukcesas per (3.8.6) pruvi la fundamentajn teoremojn de Shannon:

1) Se la ekspekto H de la informacio, kiun sendis la (ergoda) fonto, estas malpli granda ol $C\tau$, oni povas ĉiam tiel kodi, ke laŭ laŭplaĉe alta probablo eblas diveni ĉe la kanalfino, kion la fonto sendis (ekz. per tio, ke oni kodas la altprobablajn ĉenojn de la teoremo de McMillan per la diferencigeblaj ĉenoj v_k de la teoremo de Feinstein, kaj havos almenaŭ unu ĉenon de Feinstein por la ne plu diferencigataj malprobablaj ĉenoj, aŭ por prefikso de ilia kodado).

2) Tiel oni povas atingi transigrapidecon laŭplaĉe proksiman al H/τ , t.e. akiri je alvenanta signo preskaŭ H bit da informacio, kies origino estas la fonto.

Pli alta transigrapideco ol C estas atingebla per nenia kodado (Zaregradski, 1958). Pri la kresko de la erarprobablo kaze de $H > C$ informas Kolmogoroff (1956).

Ekzerctasko 3.8(1)

Binara kanalo K^- transigas dum po 7 TU (tempo-unuoj, temperoj, tempoatomoj – ekzemple subjektivaj tempokvantoj en la informacipsikologia senso) ĉenetojn je $l = 7$ binaraj signoj, de kiuj per ia mekanismo estos laŭ probablo $p = 10\%$ perturbata precize 1 (tiel ke laŭ la sama probablo povas esti perturbita ajna de la $l = 7$ signoj) laŭ probablo $1-p$ neniuj. Oni kompletigas la kanalon K^- al kanalo K^{++} per aldono de enmetilo kaj de elmetilo. La enmetilo kodas eventojn el repertuaro kun la amplekso $U = 16$ laŭ erarkorek-

codiert Ereignisse aus einem Repertoire des Umfangs $U = 16$ gemäß einem fehlerberichtigenden Code mit $l = 7$ Bit langen Kettchen, deren wechselseitiger Hamming-Abstand je wenigstens 3 beträgt. Die Ausgabevorrichtung entschlüsselt in Zeichen, deren Referenda die ursprünglich beobachten und codierten Ereignisse sind.

1. Wieviel Transinformation pro TU liefert der Kanal K^{++} (in Abhängigkeit von der Wahrscheinlichkeitsverteilung der U Ereignisse)? Wie groß ist also die Kanalkapazität des Kanals K^{++} ?

2. Wieviel Transinformation pro TU liefert der binäre Kanal K^- selbst (d.h. ohne Eingabe- und Ausgabevorrichtung), wenn mit gleicher Wahrscheinlichkeit in ihn während $l = 7$ TU irgend eines der $2^l = 128$ $l = 7$ Bit langen Kettchen gerät? Wie groß ist also die Kanalkapazität von K^- mindestens?

3. Deuten Sie die Eingabevorrichtung als Kanal K^+ , der Information weder durch den Raum noch durch die Zeit transportiert, sondern von einem Repertoire in ein anderes, wobei er jeweils $l = 7$ TU an Zeit verbraucht. Wieviel Transinformation pro TU überträgt er? Was kann über seine Kanalkapazität gesagt werden?

4. Analysieren Sie analog die Ausgabevorrichtung als Kanal K^{+} !

tiga kodo per $l = 7$ Bit longaj ĉenoj havantaj unu de la alia po minimume la Hamming-distancon 3. La elmetilo malkodas en signojn, kies referendoj estas la origine observitaj kaj koditaj eventoj.

1. Kiom da transinformacio je TU havigas la kanalo K^{++} (depende de la probablo-distribuo de la U eventoj)? Kiom granda do estas la kanalkapacito de la kanalo K^{++} ?

2. Kiom da transinformacio je TU havigas la binara kanalo K^- mem (t.e. sen la enmetilo kaj sen la elmentilo), se laŭ egala probablo ĝin eniras dum $l = 7$ TU iu el la $2^l = 128$ po $l = 7$ Bit longaj ĉenetoj? Kiom granda do estas la kanalkapacito de K^- minimume?

3. Interpretu la enmetilon kiel kanalon K^+ , kiu transportas informacion nek tra la spaco nek tra la tempo, sed de unu repertuaro en alian, bezonante ĉiam $l = 7$ TU da tempo. Kiom da transinformacio je TU ĝi transigas? Kio estas direbla pri ĝia kanalkapacito?

4. Analizu analoge la elmetilon kiel kanalon K^{+} !