Übung 3

Tutorübung zu Grundlagen: Rechnernetze und Verteilte Systeme
(Gruppen MI-T7 / DO-T5 SS 2015)

Michael Schwarz

Technische Universität München
Fakultät für Informatik

06.05.2015 / 07.05.2015

Wir lernen zwei Schranken für die maximale Übertragungsrate / minimal notwendige Bandbreite kennen

Kanalkapazität

Wir haben bereits gesehen, dass

- ein Kanal wie ein Tiefpass wirkt und
- zusätzliches Rauschen die Übertragung stört.

Wegen der für Kanäle typischen Tiefpasscharakteristik kann man von einer Kanalbandbreite B sprechen:

- Niedrige Frequenzen passieren ungehindert (Tiefpass)
- Hohe Frequenzen werden gedämpft
- Ab einer bestimmten Frequenz ist die Dämpfung so stark, dass die betreffenden Signalanteile vernachlässigt werden können

Vereinfacht nehmen wir eine scharfe Grenze für B an:

- Frequenzanteile $|f| < B$ passieren
- Frequenzanteile $|f| \geq B$ werden gesperrt

Wie hoch ist die erzielbare Datenrate auf einem Kanal mit Bandbreite B?

Hierfür benötigen wir einen Zusammenhang zwischen

- der Kanalbandbreite B,
- der Anzahl M unterscheidbarer Signalstufen und
- dem Verhältnis zwischen der Leistung des Nutzsignals und des Rauschens.

Rauschfreier, M-ärer Kanal

Angenommen es können nicht nur zwei sondern $M = 2^N$ unterscheidbare Symbole übertragen werden. Wie ändert sich die erzielbare Datenrate?

Wir erinnern uns an Quantisierung und Entropie:

- Mit einer Wortbreite von N bit lassen sich $M = 2^N$ diskrete Signalstufen darstellen.
- Emittiert eine Quelle alle Zeichen (Signalstufen) mit der gleichen Wahrscheinlichkeit, so ist die Entropie (und damit die mittlere Information) der Quelle maximal.

Folglich erhalten wir für die Übertragungsrate

- über einen Kanal der Bandbreite B
- die maximale Rate $R_{\text{max}} = 2B \log_2(M)$ bit.

Hartleys Gesetz

Auf einem Kanal der Bandbreite B mit M unterscheidbaren Signalstufen ist die Kanalkapazität durch

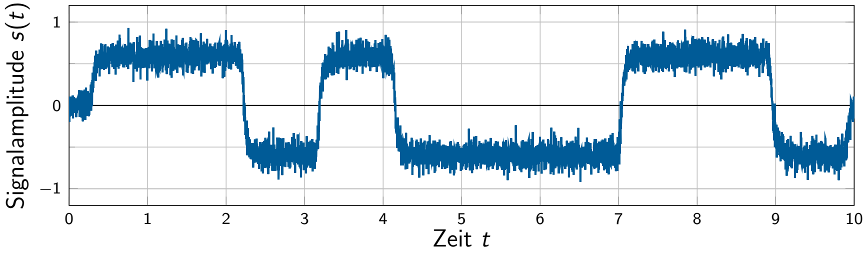
$$C_H = 2B \log_2(M) \text{ bit}$$

nach oben begrenzt.

Interessant: Wenn wir beliebig viele Signalstufen voneinander unterscheiden könnten, wäre die erzielbare Datenrate unbegrenzt! Wo ist das Problem?

Rauschen

- Rauschen macht es schwer, Signalstufen auseinanderzuhalten
- Je feiner die Signalstufen gewählt werden, desto schwieriger wird dies



Maß für die Stärke des Rauschens:

$$\mathrm{SNR} = \frac{\text{Signalleistung}}{\text{Rauschleistung}} = \frac{P_S}{P_N}$$

Das Signal to Noise Ratio (SNR) wird in der Einheit dB angegeben:

$$\mathrm{SNR}\,\mathrm{dB} = 10 \cdot \log_{10}(\mathrm{SNR})$$

Beispiel: $P_S = 1\,\mathrm{mW}$, $P_N = 0.5\,\mathrm{mW}$

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{1}{0.5}\right)\,\mathrm{dB} \approx 3.0\,\mathrm{dB}$$

Signalleistung

Definition (Leistung eines Signals)

Der Erwartungswert des Quadrats der Signalamplitude entspricht der Signalleistung. Die Varianz (Streuung) der Signalamplitude entspricht der Signalleistung ohne deren Gleichanteil und stellt die informationstragende Leistung eines Signals dar.

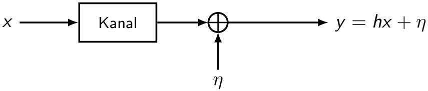
Die Signalleistung eines erwartungswertfreien[3] Signals $y(t)$ beträgt

$$\begin{aligned}\mathrm{Var}[y] &= \int_{-\infty}^{\infty} y^2 p(y)\ \mathrm{d}y \\ &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} y^2(t)\ \mathrm{d}y = P_y.\end{aligned}$$

[3] Als „erwartungsfrei“ bezeichnet man im üblichen Sprachgebrauch eine Zufallsvariable, deren Erwartungswert null ist. Dies kann zu Missverständnissen führen, da es auch Zufallsvariablen gibt, deren Erwartungswert nicht existiert.

Rauschbehafteter, *M*-ärer Kanal

- Keine Quantisierungsfehler (da wir nur an der Rausch-Abhängigkeit interessiert sind)
- Das (erwartungswertfreie) Sendesignal x werde durch den Kanal um den Faktor $h < 1$ gedämpft und von (erwartungswertfreiem) unabhängigem additivem Rauschen η überlagert



Für die Signalleistung[4] P_y am Kanalausgang erhalten wir

$$P_y = \mathrm{Var}[y] = \mathrm{Var}[hx + \eta] = h^2\mathrm{Var}[x] + \mathrm{Var}[\eta].$$

Für den Spezialfall, dass $\eta \sim \mathcal{N}(0,\sigma_\eta^2)$ (Gaussian Channel), lässt sich zeigen, dass ein ebenfalls normalverteiltes $x \sim \mathcal{N}(0,\sigma_x^2)$ die Datenrate maximiert. Wir erhalten in diesem Fall

$$P_y = h^2\sigma_x^2 + \sigma_\eta^2.$$

Setzt man nun P_y ins Verhältnis zur Rauschleistung $P_N = \sigma_\eta^2$, so erhalten wir

$$\frac{h^2\sigma_x^2 + \sigma_\eta^2}{\sigma_\eta^2} = 1 + \frac{h^2\sigma_x^2}{\sigma_\eta^2} = 1 + \frac{P_S}{P_N} = 1 + \mathrm{SNR}.$$

P_S ist die Leistung des Nutzsignals, die beim Empfänger ankommt.

[4] erwartungswertfreie ergodische Signale: $\mathrm{Var}[y] = \int_{-\infty}^{\infty} y^2 p(y)\,dy = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} y^2(t)\,dt = P_y$.

Shannon-Hartley-Theorem

Auf einem Kanal der Bandbreite B mit additiven weißen Rauschen mit Rauschleistung P_N und Signalleistung P_S beträgt die obere Schranke für die erreichbare Datenrate

$$C_S = B \log_2 \left(1 + \frac{P_S}{P_N} \right) \text{ bit.}$$

Herleitung des Theorems: siehe Shannons Veröffentlichung Communication in the Presence of Noise von 1949. **Vergleich mit Hartleys Gesetz (aufschlussreich!):**

$$C_H = 2B \log_2(M) = 2B \log_2 \left(\frac{b-a}{\Delta} \right).$$

- Die Intervallgrenzen a,b beziehen sich hier auf das unquantisierte Signal
- Mit $\alpha = a + \Delta/2$ und $\beta = b - \Delta/2$ als minimale bzw. maximale quantisierte Signalamplitude erhalten wir

$$C_H = 2B \log_2 \left(\frac{\beta - \alpha + \Delta}{\Delta} \right) = B \log_2 \left(\left(1 + \frac{\beta - \alpha}{\Delta} \right)^2 \right). \tag{3}$$

Wird (3) ausmultipliziert, kommt aber etwas anderes raus!

- C_S **berücksichtigt nur additives Rauschen des Kanals, aber keine Quantisierungsfehler.**
- C_H **berücksichtigt nur die Signalstufen und damit die Quantisierung („Quantisierungsrauschen“), aber keine Kanaleinflüsse.**
- Der fehlende gemischte Term, wenn man (3) ausmultipliziert und mit C_S vergleicht, liegt in der Unabhängigkeitsannahme des Nutzsignals und des Rauschens begründet ($\mathrm{E}[x\eta] = \mathrm{E}[x]\mathrm{E}[\eta]$).
  Der Quantisierungsfehler ist natürlich nicht unabhängig vom Eingangssignal – aus diesem Grund lässt sich (3) nicht ohne Näherung in die selbe Form wie C_S bringen.

Zusammenfassung

Die Kanalkapazität C ist durch zwei Faktoren beschränkt:

- **Die Anzahl M der unterscheidbaren Symbole**
  Selbst ein rauschfreier Kanal hilft nichts, wenn wir nur zwei Symbole nutzen (können).
- **Signal-to-Noise Ratio (SNR)**
  Ist das Signal-zu-Rausch-Verhältnis SNR zu gering, muss ggf. der Abstand Δ zwischen den Signalstufen erhöht und damit die Anzahl unterscheidbarer Symbole verringert werden, um eine zuverlässige Unterschei-
  dung gewährleisten zu können.

Für die tatsächliche Kanalkapazität C gilt also folgende obere Schranke:

$$C < \min\{C_H, C_S\} = \min\left\{2B\log_2(M), B\log_2(1+\mathrm{SNR})\right\} \text{ bit.}$$

Anmerkungen:

- Das ist nur ein Modell – mit stark vereinfachenden Annahmen.
- Wie man einen Kanalcode mit genau der richtigen Menge Redundanz konstruieren kann, so dass C maximiert wird, ist ein offenes Problem der Informationstheorie. ($\leftarrow$ Challenge!)
- Wir sprechen hier von Datenraten im informationstheoretischen Sinn, d. h. die zu übertragenden Daten liegen redundanzfrei vor. Dies ist in praktischen Systemen nie gewährleistet:
  - Nutzdaten werden vor dem Senden nicht zwangsläufig (und niemals optimal) komprimiert
  - Zusätzlich zu den Nutzdaten werden Kontrollinformationen (Header) benötigt ($\rightarrow$ später)

$\Rightarrow$ Die tatsächlich erzielbare Netto-Datenrate liegt unterhalb der informationstheoretischen Schranke.

Übergang Basisbandsignal => Bandpasssignal

Modulation [3]

Bislang haben wir nur Basisbandsignale betrachtet:

- Zeitlich verschobene Grundimpulse werden gewichtet
- Zeitlich begrenzte Grundimpulse (wir haben nur solche kennengelernt) besitzen ein unendlich ausgedehntes Spektrum
- Sofern der Übertragungskanal exklusiv für die Basisbandübertragung zur Verfügung steht, ist das zunächst kein Problem

Was ist, wenn der Kanal von mehreren Übertragungen zeitgleich verwendet wird?

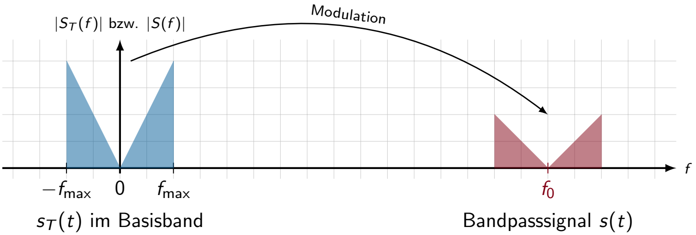
- Das Basisbandsignal (bzw. dessen Grundimpulse) wird tiefpass-gefiltert, was eine Begrenzung des Spektrums (und damit einer leichten Verfälschung des Zeitsignals) entspricht
- Anschließend kann das gefilterte Basisbandsignal auf ein Trägersignal moduliert werden
- Dies entspricht einer Verschiebung des Spektrums (Multiplikation im Zeitbereich ist bedingt eine Verschiebung im Frequenzbereich)
- Teilen sich mehrere Übertragungen auf diese Art einen Kanal, so sprechen wir von Frequency Division Multiplex (FDM)

Prinzipieller Ablauf digitaler Modulationsverfahren

- Die Grundimpulse $g(t)$ werden mittels Tiefpassfilterung auf eine maximale Frequenz $f_{\max}$ beschränkt. Die so gefilterten Impulse bezeichnen wir als $g_T(t)$.
- Das Modulationssignal $s_T(t)$ wird wie im Basisband durch eine gewichtete Überlagerung zeitlich verschobener Grundimpulse erzeugt.
- Das Modulationssignal wird auf ein Trägersignal der Frequenz f_0 aufmoduliert:

$$s(t) = s_T(t) \cdot \sin(2\pi f_0 t) = \left(\sum_{n=0}^{\infty} d_n \cdot g_T(t - nT) \right) \sin(2\pi f_0 t).$$

Schematischer Ablauf im Frequenzbereich:

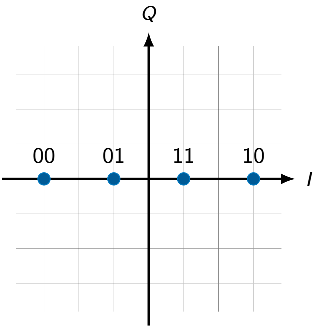


4-ASK (Amplitude Shift Keying)

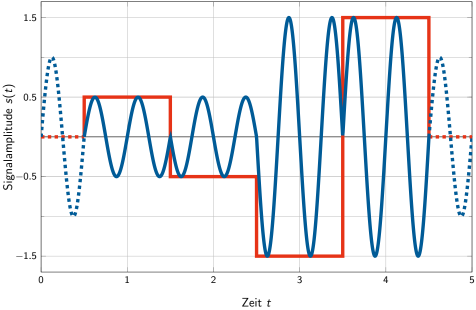
- Es werden 4 Signalstufen unterschieden $\Rightarrow$ 2 bit/Symbol
- Es wird nur die Amplitude des Trägersignals moduliert

Beispiel: Signalraum $S = \left\{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\right\}$

- Je zwei aufeinanderfolgende Bits des Datenstroms werden auf ein Symbol $d \in S$ abgebildet, z. B. $00 \mapsto -\frac{3}{2}$, $01 \mapsto -\frac{1}{2}$, ...
- Die Symbolsequenz d_n verändert die Amplitude eines Grundimpulses (z. B. Rechteckimpuls)
- Das so entstehende Basisbandsignal wird mit einem Trägersignal multipliziert (Modulation)



(a) Signalraumzuordnung

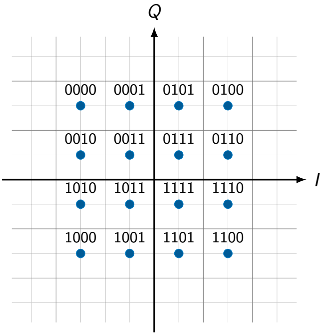


(b) Sendesignal $s(t)$ (blau), Modulationssignal $s_T(t)$ (rot)

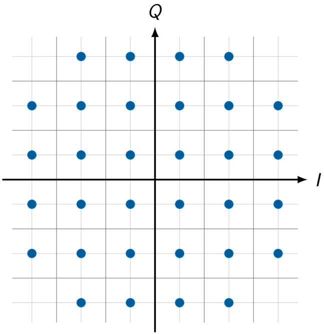
Quadratur-Amplituden-Modulation (QAM)

- Man kann kosinus- und sinus-förmige Trägersignale mischen
- Trennung durch Orthogonalität von Sinus und Kosinus möglich
- Der Kosinus wird als Inphase-Anteil, der Sinus als Quadratur-Anteil bezeichnet
- Die Datenrate lässt sich auf diese Weise verdoppeln

$$s(t) = \left(\sum_{n=0}^{\infty} d_{In} \cdot g_T(t - nT) \right) \cos(2\pi f_0 t) - \left(\sum_{n=0}^{\infty} d_{Qn} \cdot g_T(t - nT) \right) \sin(2\pi f_0 t)$$



(c) 16-QAM



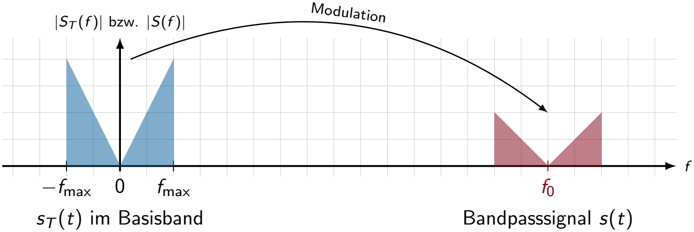
(d) 32-QAM

- QAM verdoppelt die Datenrate nochmals?
- Sensationell, wir haben Shannon widerlegt!

- QAM verdoppelt die Datenrate nochmals?
- Sensationell, wir haben Shannon widerlegt!

Natürlich nicht: [4] Durch die Frequenzverschiebung belegt das Bandpasssignal die doppelte Bandbreite im Vergleich zum Basisbandsignal. Es entsteht ein

- oberes Seitenband, welches den nicht-negativen Frequenzanteilen im Basisband entspricht, sowie ein
- unteres Seitenband, welches den nicht-positiven Frequenzanteilen im Basisband entspricht.



- Durch die Modulation wurde also die benötigte Bandbreite verdoppelt
- Dieser „verlorene Freiheitsgrad“ kann durch die Mischung von Sinus- und Kosinus-Trägern wieder kompensiert werden.

Die obere Schranke für die erzielbare Datenrate gilt natürlich weiterhin.

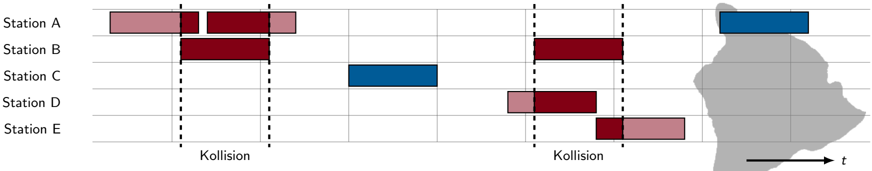
ALOHA und Slotted ALOHA

Random Access (ALOHA)

- Entwickelt an der Universität von Hawaii (1971), c.f. Prof. Abramson
- Ursprünglich für kabellose Datenübertragungen
- Ziel: Verbindung von Oahu mit den anderen hawaiianischen Inseln

Funktionsweise

- Jede Station sendet an eine zentrale Station (vgl. „Basisstation“ in modernen WLANs), sobald Daten vorliegen
- Senden zwei Stationen gleichzeitig, kommt es zu Kollisionen
- Erfolgreich übertragene Nachrichten werden vom Empfänger auf anderer Frequenz quittiert („out-of-band“ Bestätigungsverfahren auf Link-Layer, keine Kollisionen zwischen Nachrichten und Bestätigungen)



Das Kanalmodell ist vergleichsweise einfach. Es existieren math. Beschreibungen für den sog. ALOHA Random Access Channel.

Erreichbarer Durchsatz mit ALOHA

Vereinfachende Annahmen:

- Mittlere bis beliebig große Anzahl an Knoten ($N > 15$)
- Gleiche, unabhängige und geringe Sendewahrscheinlichkeit auf allen Knoten
- Nachrichten konstanter Größe (Sendedauer T)

Modellierung:

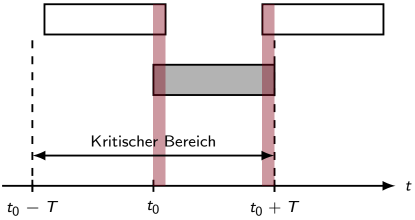
- Ob ein bestimmter Knoten i innerhalb des Zeitintervalls $[t, t+T)$ zu senden beginnt oder nicht entspricht einem Bernoulli-Experiment mit Erfolgs- bzw. Sendewahrscheinlichkeit p_i
- Da die Sendewahrscheinlichkeit für alle Knoten gleich ist, gilt $p_i = p \quad \forall i = 1, \dots, N$
- Da wir N Knoten haben, die jeweils unabhängig voneinander zu senden beginnen, wird dasselbe Bernoulli-Experiment N-mal wiederholt
- Das ist nichts anderes als eine Binomialverteilung, welche die Anzahl der Erfolge einer Serie gleichartiger und unabhängiger Versuche beschreibt
- Für sinnvoll großes N kann die Binomialverteilung durch eine Poisson-Verteilung[2] approximiert werden (s. Übung)
- Die mittlere erwartete Anzahl von Nachrichten pro Intervall ist gegeben als $Np = \lambda$

Das Ereignis X_t, dass im Intervall $[t, t+T)$ genau k Knoten senden, ist poisson-verteilt:

$$\Pr[X_t = k] = \frac{\lambda^k e^{-\lambda}}{k!}.$$

[2] Verteilung der seltenen Ereignisse

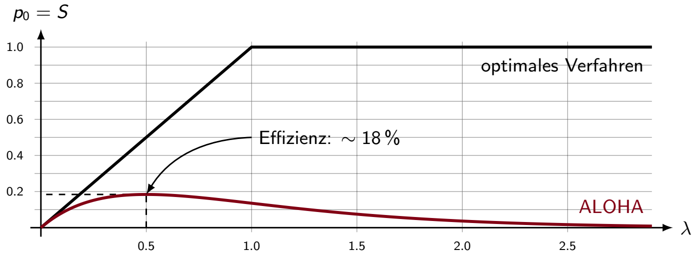
- Eine beliebige Station sende nun zum Zeitpunkt t_0 eine Nachricht
- Eine Kollision tritt genau dann auf, wenn eine andere Station im Intervall $(t_0 - T, t_0 + T]$ versucht, ebenfalls zu übertragen
- Die Übertragung ist also erfolgreich, wenn innerhalb des Intervalls $[t_0, t_0 + T]$ genau eine Übertragung stattfindet und im Intervall $(t_0 - T, t_0)$ keine Übertragung begonnen hat.



- Mit der Dichtefunktion $\Pr[X_t = k] = \frac{\lambda^k e^{-\lambda}}{k!}$ erhalten wir
- die Wahrscheinlichkeit p_0 für eine erfolgreiche Übertragung:

$$p_0 = \Pr[X_{t_0 - T} = 0] \cdot \Pr[X_{t_0} = 1] = e^{-\lambda} \cdot \lambda e^{-\lambda} = \lambda e^{-2\lambda}$$

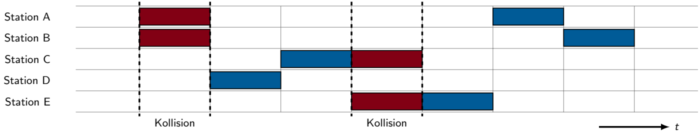
Die Erfolgswahrscheinlichkeit p_0 kann gegen die Senderate λ aufgetragen werden:



- Wir wissen, dass innerhalb eines beliebigen Intervalls $[t, t + T)$ höchstens eine Übertragung erfolgreich sein kann
- Dementsprechend entspricht die Anzahl S der erfolgreichen Nachrichten pro Intervall gleichzeitig der Wahrscheinlichkeit für eine erfolgreiche Übertragung
- Bei einem optimalen Verfahren würde die Anzahl erfolgreicher Nachrichten S linear mit der Senderate ansteigen, bis die maximale Anzahl von Nachrichten pro Zeitintervall erreicht ist (hier ist das genau eine Nachricht pro Intervall)
- Steigt die Senderate weiter, würde dies ein optimales Verfahren nicht beeinträchtigen

Variante: Slotted ALOHA

Stationen dürfen nicht mehr zu beliebigen Zeitpunkten mit einer Übertragung beginnen, sondern nur noch zu den Zeitpunkten $t = nT$, $n = 0, 1, \ldots$



Kritischer Bereich ist nur noch T anstelle von $2T$ $\Rightarrow$ $S = \lambda \cdot e^{-\lambda}$.