

DANMARKS TEKNISKE UNIVERSITET



INTRODUKTION TIL DIGITAL KOMMUNIKATION

Noter

34210 Introduktion til digital kommunikation, 2018

Knud J. Larsen

Sektion for Kommunikationsteknologi

DTU

Bygning 343, 2800 Kgs. Lyngby

DTU Fotonik
Department of Photonics Engineering

Indhold

Forord	3
Oversigt over anvendte symboler	5
1 Introduktion - kommunikation og signaler	7
1.1 Analogt/digitalt signal	7
1.2 Deterministisk/stokastisk signal	8
1.3 Periodisk/ikke-periodisk signal	8
1.4 Eksempler på signaler	9
1.5 Kursets opbygning og indhold	10
2 Sampling og kvantisering af analoge signaler	13
2.1 Harmoniske funktioner	13
2.2 Sampling	15
2.3 Kvantisering	18
3 Tid og frekvens	23
3.1 Fourierrækker for periodiske signaler	24
3.2 Fouriertransformation for ikke-periodiske signaler	31
3.3 Fouriertransformation af tidsdiskret signal	36
3.4 Oversigt over Fouriertransformationer	40
3.5 Sampling og aliasering	40
3.6 Energi- og effektspektre	42
4 Elektriske kredsløb og lineære systemer	45
4.1 Lineære systemer	45
4.2 Elektriske kredsløb	47
4.3 Almindelige komponenter	49
4.4 Overføringsfunktion og impulsrespons for elektriske kredsløb	50
4.5 Introduktion til kredsløbsteori	54
4.6 Tidsdiskrete lineære systemer	56
5 Transmission med pulser	59
5.1 Støj i transmissionssystemer	63
5.2 Transmissionssystem som lineært system	74

6	Modulation	81
6.1	Analog amplitudemodulation	82
6.2	Digital modulation	86
6.3	Multidimensionale signaler	96
7	Kanalmodeller	103
7.1	Metalliske transmissionsledninger	104
7.2	Optical transmission lines	120
7.3	Radio channels	125
7.4	References for Chapter 7	131
	Indeks	131

Forord

Disse noter tager udgangspunkt i meget beskedne dele af det matematiske og fysiske pensum på DTU og i gymnasiet/HTX. Med hensyn til DTU forudsættes dele af pensummet i første semester af Matematik 1 kendt. Vi har forsøgt at referere til disse ved referencer af typen [eNote nn]. Til at undersøge signaler og deres egenskaber og til at illustrere begreberne vil vi anvende programsystemet MATLAB, som er egnet til alle former for beregninger på signaler. Dette program blev nødtørftigt introduceret i kurset 34302 og i dette kursus vil der blive demonstreret yderligere funktioner og anvendelser af MATLAB. I Øvelsesbeskrivelsen findes en oversigt over anvendte MATLAB kommandoer og funktioner. MATLAB kommandoer, variable og programmer skrives i disse noter normalt med skrifttypen `courier`.

På grund af de beskedne forudsætningskrav, har vi valgt en meget eksempelorienteret fremgangsmåde uden for mange formler, men helt undværes kan de ikke. Det er altså meningen at man ud fra eksempler og øvelser skal opbygge en mere generel forståelse af emnerne. Senere i studiet på DTU, fx i Matematik 2, Tidskontinuerte signaler og Digital kommunikation, vil der komme en mere teoretisk underbygning af de fremførte påstande og de viste eksempler.

Som et meget vigtigt supplement findes der en separat beskrivelse af øvelser for de enkelte afsnit i noterne.

Dette er tredje udgave af noter for kurset 34210 (og fjortende udgave hvis tidligere kursus 34310 medregnes). Denne udgave er omlagt til et nyt tekstbehandlingssystem, så en del fejl og besynderligheder findes sikkert. Grundlæggende skulle teksten gerne være forbedret ud fra kommentarer og erfaringer fra de sidste år, men der vil utvivlsomt stadig være mange mindre fejl og uklarheder. I løbet af kurset vil underviserne sætte meget pris på at få det at vide.

En stor tak skal lyde til kolleger og tidligere studerende fra kurset for forslag og rettelser. Specielt en meget stor tak til Mathias Boss Jørgensen for et meget omhyggeligt arbejde med omlægningen til L^AT_EX.

Knud J. Larsen
Marts 2018

Oversigt over anvendte symboler

I disse noter anvendes almindelige betegnelser for matematiske funktioner og konstanter, fx π , e , \log og \cos , og almindelige betegnelser for fysiske enheder kendt fra undervisning i fysik, i skole og gymnasium, fx A, V, W, F og Hz, ofte med præfiks som i nedenstående tabel.

Præfiks	Størrelse
T (<i>tera</i>)	10^{12}
G (<i>giga</i>)	10^9
M (<i>mega</i>)	10^6
k (<i>kilo</i>)	10^3
m (<i>milli</i>)	10^{-3}
μ (<i>mikro</i>)	10^{-6}
n (<i>nano</i>)	10^{-9}
p (<i>piko</i>)	10^{-12}

Desuden anvender vi ret mange bogstaver for forskellige størrelser og begreber. Følgende liste er et forsøg på at liste de vigtigste, dvs. bogstaver der anvendes mere globalt i noterne og ikke bare lokalt på en side eller figur. Der gøres også opmærksom på indekset til slut i noterne, hvor begreber (ofte med korte betegnelser, som fx v_f) er listet med sidehenvisninger til definitioner. Disse er udeladt i nærværende liste, så de to lister skal evt. anvendes sammen, hvilket kan forvirre lidt, men forhåbentlig er listen nyttig.

a_k, b_k	symboler (til tiden kT)	f_s	samplingfrekvens
A	amplitude	ϕ	fase
α	dæmpningskoefficient (plus andre mindre anvendelser)	G	konduktans
		g	pulsform
β	fasekonstant	h	impulsrespons
C	kondensator, kapacitans	H	overføringsfunktion
γ	udbredelseskoefficient	i	strøm (funktion af tid)
E	spænding	I	strøm (funktion af frekvens)
E	energi	i, I	ifase-signal
f, F	frekvens	j	imaginær enhed
f_c	bærefrekvens	k, n, m	heltal

L	selvinduktion, induktans	R	bitrate
L	længde af puls	t	tid
P	længde (af ledning)	T	symboltid
λ	bølgelængde	T_S	samplingtid
M	antal værdier (kvantisering, alfabet)	r, v, x, y	signalværdier (funktion af tid)
n	støj ($n(t)$)	V, X, Y	signalværdier (funktion af frekvens)
N	antal værdier (fx i Fouriertransformation)	v, V	spænding (tid, frekvens)
P	effekt	$x(t)$, x_n	samlet impulsrespons
P	periode for funktion	x	stedsangivelse (på ledning)
q, Q	kvadraturfasesignal	Z	impedans
R	modstand, resistans	ω	vinkelfrekvens ($= 2\pi f$)

KAPITEL 1

Introduktion - kommunikation og signaler

I kurset 34311 var emnet kommunikation i en ganske abstrakt form - vi diskuterede ikke meget hvordan kommunikationen rent fysisk foregår. Det fysiske lag i kommunikationen er nødvendigt, da der ellers ingen kommunikation er - hvis man da udelukker tankeoverføring og andre sådanne tvivlsomme fænomener. Det er i det fysiske lag den største kapital er bunden i kommunikationsnet, og det er her muligheder og barrierer for det meste afgøres, fx global rækkevidde for radiosignaler eller begrænsninger i kapacitet. Det er derfor af største vigtighed at forstå de fysiske forhold omkring kommunikation. Det vil hurtigt blive klart at selv denne del af kurset ikke er så fysisk endda - vi baserer os på matematiske modeller af den fysiske verden og opbygger sådanne modeller i teori og viser dem i praksis ved at programmere dem i MATLAB.

Al kommunikation foregår ved udveksling af “signaler”. Signaler er måske et lidt vidt begreb, så vi vil indskrænke os til behandling af signaler der kan repræsenteres ved en fysisk størrelse, fx strøm, spænding, lys eller mere generelt et elektromagnetisk felt. Det udelukker nogle af hverdagens “signaler” som mest er baseret på billeder såsom lyssignaler for vej eller jernbane, vejskilte, tegnsprog eller røgsignaler. Vi vil beskrive signaler som de genereres, transmitteres, modtages og behandles med elektroniske eller optiske komponenter, mest elektroniske. Et signal defineres som:

Et **signal** er en funktion, $v(t)$, af tiden, t , og tidsvariationen bærer på en eller anden form for **information**.

1.1 Analogt/digitalt signal

Umiddelbart vil man ud fra definitionen tænke på at tiden er en kontinuert, reel variabel og funktionen ligeledes. Et sådant signal vil vi kalde et **analogt** signal og vi ser eksempler på sådanne i Figur 1.1 - 1.4. Imidlertid er der en god grund til også at betragte signaler, hvor tiden er **diskret**, dvs. signalet findes kun på visse tidspunkter. Grunden er at moderne realiseringer af signalbehandlingsudstyr er digitale og behandler signalet ét tidspunkt ad gangen. Man kan fx tænke på at gemme signalet i en RAM, hvor hver værdi gemmes på en adresse. Et andet problem er signalets værdier. Digitalt udstyr kræver en **kvantisering** af signalet, dvs. kun visse definerede værdier er mulige, ikke alle de reelle

tal. Et tidsdiskret og kvantiseret signal vil vi kalde et **digitalt** signal. Ofte skal man omsætte et signal mellem analog og digital form, og hvilke problemer det kan give, er emnet for Kapitel 2. Et eksempel på et transmissionssystem hvor sådanne omsætninger foregår ses i Figur 1.5. Her ses også at et digitalt signal, som transmitteres, jo er analogt i sig selv, fx spændingen leveret til den viste antenne. Flere eksempler på dette vil findes senere i kurset, men som en forsmag kan Figur 1.2 - 1.4 studeres. Vi kommer mere ind på fysiske forhold i Kapitel 4 og 7. Dette kursus handler mest om **digital kommunikation** og det skal forstås således at det er informationen der er digital, fx $\{0, 1\}$ (bit), heltal, tegn fra et eller andet alfabet, digitaliseret audio eller digital video. Selvom signalerne på et eller andet tidspunkt er repræsenteret i en analog form, er det karakteristisk for digital kommunikation at signalet sammensættes af et endeligt antal grundformer. Eksempler kan som nævnt ses i Figur 1.2 - 1.4 hvor den digitale information er ± 1 .

1.2 Deterministisk/stokastisk signal

Et karakteristisk punkt i vores definition af et signal er at tidsvariationen bærer på en eller anden form for **information**. Kommunikationsteknologi beskæftiger sig netop med transport og lagring af information, men informationen er i princippet ukendt på forhånd, da der ellers ikke er noget indhold i kommunikationen. Lidt mere teknisk kan man sige at information er en tilfældig proces om hvilken man kun kan udtale sig i termer som sandsynligheden for dette og hint i fremtiden, men aldrig med præcise værdier i al fremtid. Vi kalder et sådant signal **stokastisk** i modsætning til **deterministiske** signaler, hvor man (i princippet) kender alle værdier af signalet i fortid såvel som fremtid. Desværre vil det være for indviklet rigtigt at behandle stokastiske signaler i dette indledende kursus, da alt for mange begreber mangler. Vi indskrænker os derfor til deterministiske signaler, som man også kan lære en del af. Af og til må vi dog anvende lidt stokastisk tankegang for at behandle virkelig kommunikation. Det er fx tilfældet i Kapitel 5.

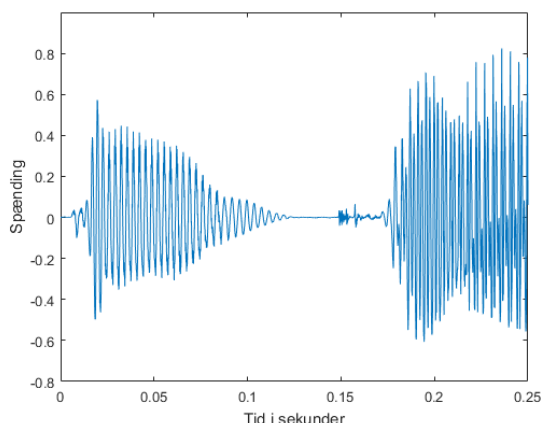
1.3 Periodisk/ikke-periodisk signal

Deterministiske signaler kan yderligere karakteriseres ved deres tidsforløb: De kan være **periodiske**, dvs. signalværdierne gentages efter et stykke tid, perioden P , således at $v(t) = v(t + P)$, og dette går i princippet for sig uendeligt længe, men signalet er altså fuldt beskrevet når man kender perioden og alle signalværdierne i en periode. Periodiske signaler er simple at analysere og anvendes ofte til måleformål. Signaler kan også være **ikke-periodiske**, dvs. det gentager ikke sig selv og signalets værdier skal angives for alle tider. En beskrivelse kunne fx være en matematisk funktion, $v(t) = \text{sinc}(t) = \sin(\pi t)/(\pi t)$, men det kan også være en del mere indviklet som Figur 1.1 - 1.4 nedenfor viser. Stokastiske signaler er selvfølgelig ikke periodiske, men for en del af dem er der en vis lighed med

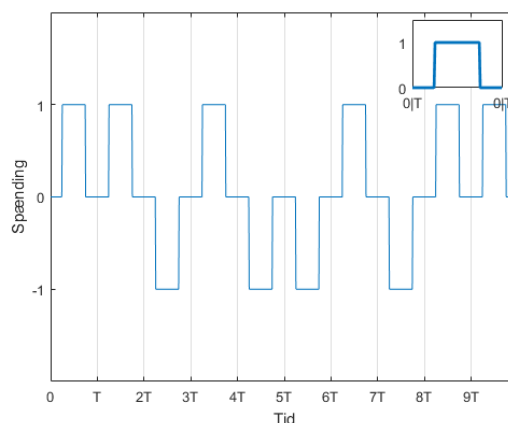
periodiske signaler i og med at egenskaberne er uforandrede over tid (sådanne stokastiske signaler kaldes stationære).

1.4 Eksempler på signaler

Ovenstående definition af signaler kan nok virke noget abstrakt, men eksempler kan ses i følgende fire figurer:



Figur 1.1: Talesignal.

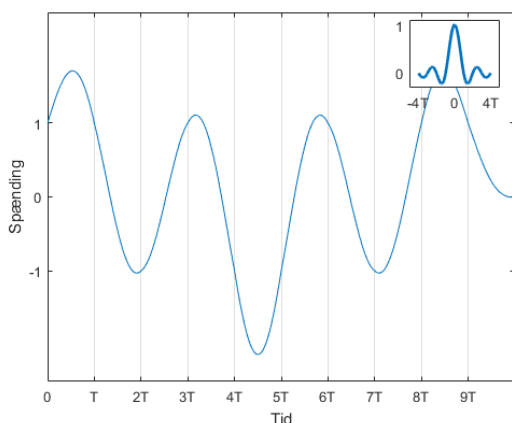


Figur 1.2: Signal med pulser.

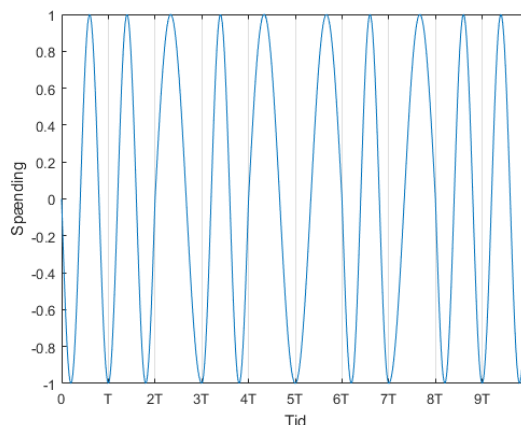
Figur 1.1: Dette viser spændingen fra en mikrofon, når der tales i den, altså et **talesignal**. Den fysiske størrelse, vi interesserer os for i kommunikationsteknologien er altså spændingen, men det skal ikke skjules, at denne spænding er en funktion af en forandring i lufttrykket frembragt af en talende person og mikrofonen fungerer som en omsætning af dette til spænding. Vi vil afholde os fra at sige præcis hvad informationen er i det viste signal, men det er klart ud fra oprindelsen at der er en eller anden form for information. Signalet er selvfølgelig et eksempel taget ud af et mere generelt stokastisk signal, “menneskelig tale”. Det er ikke sikkert at vi ønsker at beskrive det konkrete deterministiske eksempel i detaljer, måske vil vi hellere angive nogle egenskaber ved det generelle signal.

Figur 1.2: Dette kunne vise en spænding der kommer fra et elektrisk kredsløb. Det er karakteristisk at denne spænding kan ses som en sekvens af det samme korte funktionsstykke vist øverst i figuren. Et sådant funktionsstykke til sammensætning af et signal kaldes en **puls**. Alt hvad der sker er at pulsen multipliceres med $+1$ eller -1 og indsættes i sekvensen. Her er det klart at den digitale information er $+1 +1 -1 +1 -1 -1 +1 -1 +1 +1$ som kunne

være en afbildning af en informationssekvens udtrykt med de mere almindelige bit: 1 1 0 1 0 0 1 0 1 1. Igen kunne det være et deterministisk eksempel taget ud af et generelt stokastisk signal, tilfældige ± 1 som styrer den viste puls.



Figur 1.3: Signal med pulser.



Figur 1.4: Moduleret signal.

Figur 1.3: Her er det igen en spænding i et elektrisk kredsløb. Hvad der ikke er så klart er at dette også er en sekvens af pulser. Nu overlapper pulserne bare. Grundpulsen er vist øverst i figuren og signalet er faktisk en repræsentation af den samme digitale information som lige ovenfor. Vi vil senere komme tilbage til hvad der kunne få os til at benytte en sådan pulsform.

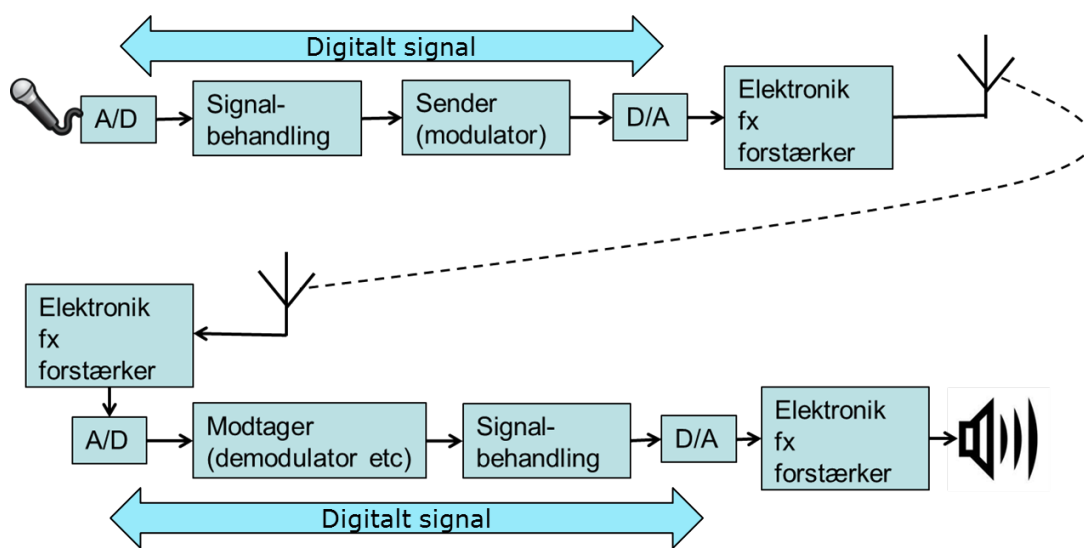
Figur 1.4: Dette er også en repræsentation af et digitalt signal. Man kan godt se at der for hver tidsenhed T sker et eller andet genkendeligt, men det er ikke en simpel pulsform. Derimod er der tidsintervaller hvor signalet svinger fra $+1$ over -1 tilbage til $+1$ og ender på 0 , mens andre starter på 0 og ender med $+1$ efter blot at have besøgt -1 en enkelt gang. Atter andre ligner disse, men er eventuelt forskudt eller vendt i tid. Det synes som om informationen ligger i antallet af svingninger og det er netop tilfældet. Signalet er et eksempel på et **moduleret** signal som vi vil definere nærmere siden. Signalet er også en repræsentation af det samme digitale signal som ovenfor.

1.5 Kursets opbygning og indhold

Hovedformålet er at signaler skal anvendes til at transmittre noget information, men en del af metoderne finder også anvendelse i andre sammenhænge, fx undersøgelse af

signaler i frekvensdomænet som introduceres i Kapitel 3. Vi illustrerer ofte samlede systemer ved hjælp af **blokdiagrammer**, der viser hvordan et signal bevæger sig gennem et system og hvordan det omformes og behandles under denne bevægelse. Et sådant blokdiagram med et eksempel på trådløs kommunikation er vist i Figur 1.5, og systemer for transmission via metalliske ledere ("tråd") eller optiske fibre ser ud på næsten samme måde.

Som det ses, er der en del omsætninger mellem analog og digital form og i Kapitel 2 beskriver vi konvertering af et analogt signal til et digitalt ved sampling og kvantisering. Det vigtigste resultat, samplingsætningen som beviser at man kan genvinde det analoge signal, indføres og udvikles videre i de næste to kapitler. Det viser sig nemlig nødvendigt at kunne beskrive signaler ved frekvenser, og i Kapitel 3 behandler vi frekvensbegrebet og sammenhængen mellem karakteristika udtrykt tidsmæssigt og frekvensmæssigt. Som det ses på figuren er der en del signalbehandling og elektronik og vi introducerer elektriske kredsløb meget kortfattet i Kapitel 4 og beskriver en meget almindelig model for et kommunikationskredsløb, nemlig et lineært system.



Figur 1.5: Eksempel på transmissionssystem med omsætninger mellem analog og digital form, A/D og D/A.

I Kapitel 5 introducerer vi kommunikation med pulsførmede signaler (som i Figur 1.2 og 1.3) og vi beskriver de to væsentligste problemer: Hvordan modtager man på bedst mulige måde støjfyldte signaler og hvordan udformes pulserne bedst muligt (dvs. er Figur 1.2 eller 1.3 det bedste og i hvilken forstand?). I Kapitel 6 beskriver vi modulation af signaler (som i Figur 1.4 og 1.5) og det afsluttende kapitel beskæftiger sig med metalliske ledninger, optiske fibre og radiotransmission, dvs. egenskaber ved de fysiske realiseringer

af forskellige transmissionsmedier.

KAPITEL 2

Sampling og kvantisering af analoge signaler

Som beskrevet i introduktionen vil moderne signalbehandling kræve at naturligt forekommende signaler - som fx et talesignal - kan omsættes på forsvarlig vis til et digitalt signal og tilbage igen. Det er derfor af største vigtighed at forstå denne omsætning og hvilke problemer og udfordringer man møder. Vi er så vant til at lytte og se på signaler der har været digitaliserede at vi måske slet ikke tænker over det som et problem, ja måske endda har forestilling om at det digitaliserede signal er en mere ren form end fx det akustiske signal et orkester frembringer for digital indspilning på en CD.

Omsætningen af det analoge signal foregår ved at måle signalet på visse tidspunkter. Denne proces kaldes **sampling** og almindeligvis er disse tidspunkter ækvidistante, dvs. med en fast afstand som vi i disse noter kalder T_S , **samplingtiden**. Det første problem vi vil behandle, er hvor tæt disse punkter skal anbringes for at gøre det muligt senere at kunne **rekonstruere** det oprindelige analoge signal ud fra de samplede værdier, og hvilken egenskab ved signalet det afhænger af. Et andet problem er signalets værdier. Digitalt udstyr kræver en **kvantisering** af signalet, dvs. kun visse definerede værdier er mulige, ikke alle de reelle tal. Dette problem behandler vi i Afsnit 2.3. I Kapitel 3 vil vi give en metode til at beskrive signalers egenskaber (frekvensanalyse), men foreløbig starter vi med en simpel klasse af signaler og så generaliserer vi resultaterne siden.

2.1 Harmoniske funktioner

En simpel form for signal er en cosinusfunktion:

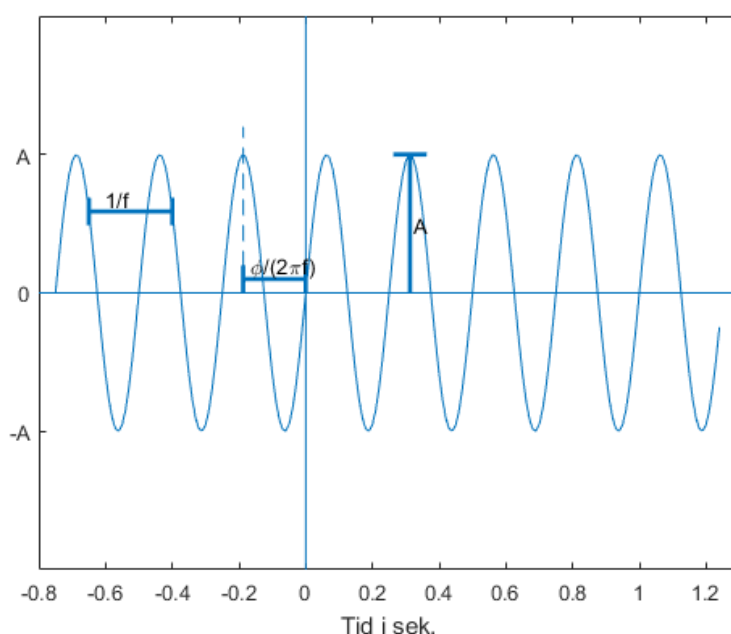
$$v(t) = A \cos(2\pi ft + \phi) \quad (2.1)$$

Denne funktion er illustreret i Figur 2.1. Tiden betegnes som sædvanligt med t . Vi kan se at $v(t)$ er **periodisk**, dvs. at der findes en periode P således at $v(t + P) = v(t)$. Perioden her ses at være $P = 1/f$, hvor f kaldes **frekvensen** og måles i Hertz, Hz (ofte med sædvanlige præfikser som fx MHz = 10^6 Hz). Frekvensen angiver hvor mange perioder der gennemløbes på ét sekund, idet hver gang ft øges med én er der gået én periode da argumentet for cosinusfunktionen er øget med 2π og cosinus er periodisk med denne periode. Frekvenser indgår ofte i argumenter for cosinusfunktioner og lignende

hvor de bliver multiplicerede med 2π som ovenfor. Af skrive- og talemæssige årsager ser man derfor ofte **vinkelfrekvens** defineret som $\omega = 2\pi f$. Det græske bogstav ω (omega) anvendes næsten altid for dette, somme tider dog w , hvis ω ikke lige findes.

$$\omega = 2\pi f, \quad v(t) = A \cos(\omega t + \phi)$$

Da cosinusfunktionen er periodisk vil dens værdi være helt bestemt af hvor argumentet er i forhold til et nærmeste lavere hele antal gange 2π . Denne størrelse kaldes signalets **fase** og er altså normalt et tal mellem 0 og 2π eller mellem $-\pi$ og π . For at signalet ikke nødvendigvis skal have maksimum ved $t = 0$ er størrelsen ϕ (fi) indført som fasen for $t = 0$. Størrelsen A kaldes signalets **amplitude** - somme tider spidsværdi.



Figur 2.1: Harmonisk funktion. De viste parametre forklares i teksten.

En beslægtet funktion er sinus, men da $\sin(x + \pi/2) = \cos(x)$, er dette den samme funktion, blot faseforskuet $\pi/2$. Cosinusfunktionen (og sinus) kaldes en **harmonisk funktion**, da den kan karakteriseres med netop én frekvens, f . Vi skal senere i Kapitel 3 se på hvordan enhver periodisk funktion kan sammensættes af harmoniske funktioner med forskellige frekvenser og hvordan enhver almindelig funktion kan udtrykkes ved harmoniske funktioner med et interval af frekvenser. Harmoniske funktioner er meget vigtige i analysen af signaler, idet det viser sig i Kapitel 4, at der er særligt simple metoder til beregning af kommunikationskredsløbs reaktion på harmoniske funktioner ligesom målinger ofte udføres med harmoniske funktioner. Her vil vi dog først prøve at sample en

harmonisk funktion i næste afsnit.

Ret beset er det en lille smule forkert at kalde (2.1) og andre periodiske funktioner et signal, da tidsvariationen er givet på forhånd og signalet derfor ikke indeholder nogen information. Vi vil alligevel gøre det pga. vigtigheden i analyse af “rigtige” signaler.

2.2 Sampling

Vi kommer nu til det ovenfor omtalte problem: Hvor tæt skal man sample et analogt signal, for at det skal være muligt ud fra sampleværdierne at kunne rekonstruere signalet fuldstændigt dvs. kunne finde enhver værdi af signalet også mellem samplepunkterne. Før vi giver et helt generelt svar på dette vil vi undersøge sampling af harmoniske signaler med frekvensen F . Her viser det sig, at:

Et harmonisk signal $v(t)$ med frekvensen F kan rekonstrueres fuldstændigt ud fra de samplede værdier, $v(kT_S)$:

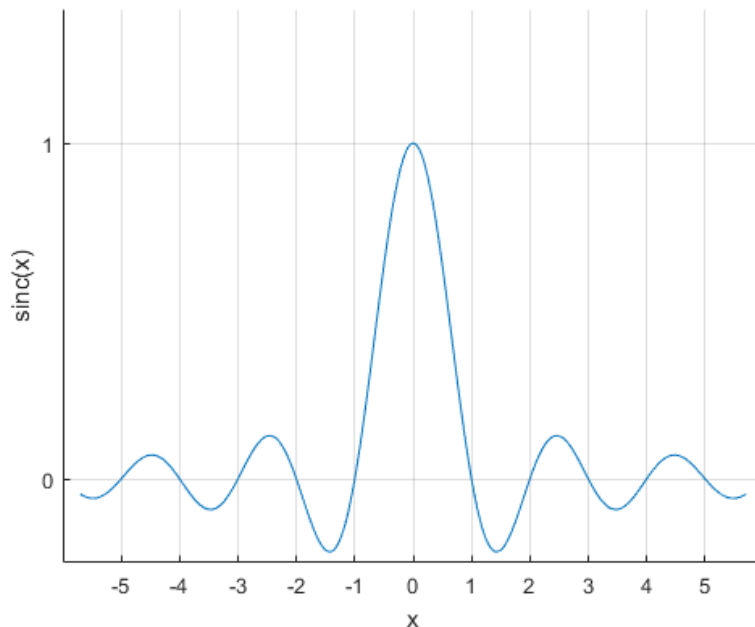
$$v(t) = \sum_{k=-\infty}^{\infty} v(kT_S) \frac{\sin\left(\pi \frac{t-kT_S}{T_S}\right)}{\pi \frac{t-kT_S}{T_S}}$$

hvis **samplingfrekvensen** $f_S = 1/T_S$ vælges $> 2F$.

Det er måske lidt uvant at se en uendelig sum, hvis man ikke tidligere har mødt det i matematik. Selvom der er uendeligt mange led, har summen ovenfor alligevel endelige værdier, $v(t)$, (i matematik siger man at denne uendelige række er konvergent. Dette behandles i kurset Matematik 2, men prøv fx at udregne $\sum 3^{-k}$ hvor k går fra 0 til ∞ . I øvrigt minder det om Taylor’s grænseformel for en funktion, [eNote17]). Vi vil endnu ikke bevise påstanden om den nødvendige samplingfrekvens, men i den efterfølgende øvelse vil vi undersøge dette. Det er ganske overraskende at det skulle være muligt med kun to samples pr. periode. Det er måske ikke så svært at forestille sig dette, hvis de to samplepunkter i en periode rammer tæt ved signalets amplitudeværdier, men man kunne jo også ramme tæt på værdien 0 begge gange (fase $\pi/2$ og $3\pi/2$) og hvordan skulle amplituden A så nogensinde blive bestemt? Løsningen ligger i at man jo har uendeligt mange værdier og når $f_S > 2F$ vil man før eller siden bevæge sig hen på andre faser i signalet. Vi beskriver i øvrigt rekonstruktionen af det analoge signal mere nøje nedenfor når vi har generaliseret til vilkårlige periodiske signaler.

Som det ses, anvender rekonstruktionen funktionen $(\sin \pi x)/(\pi x)$ for at gendanne det tidskontinuerte signal $v(t)$ ud fra de samplede værdier. Denne funktion kaldes ofte **sinc(x)** og vi vil se lidt nærmere på den. Som det vides fra matematik [eNote 17] er funktionen

kontinuert for $x = 0$ med værdien 1. Det er også nemt at overbevise sig om at $\text{sinc}(n) = 0$ for alle heltallige $n \neq 0$ og at $\text{sinc}(x)$ aftager med voksende $|x|$ idet tælleren altid svinger mellem $+1$ og -1 , mens nævneren vokser. Vi viser et udsnit af funktionen på Figur 2.2.



Figur 2.2: Funktionen $\text{sinc}(x)$ - læg mærke til hvor nulgennemgangene er.

Når vi anvender $\sum v(kT_S) \text{sinc}((t - kT_S)/T_S)$ siger vi at der foretages en **interpolation** mellem de samplede værdier for at gendanne signalet mellem sampleværdierne (“inter” betyder mellem). I rekonstruktionsformlen bidrager interpolationsfunktionen for samplingtidspunktet kT_S med værdien 1, men den giver 0 for alle de andre samplingpunkter nT_S for $n \neq k$, idet $(nT_S - kT_S)/T_S = n - k$, et helt tal. Således er det nemt at se at selve signalværdierne $v(kT_S)$ gendannes. At det også virker for t mellem disse er mere opsigtsvækkende og at det kræver uendeligt mange værdier i rekonstruktionen diskuterer vi nedenfor.

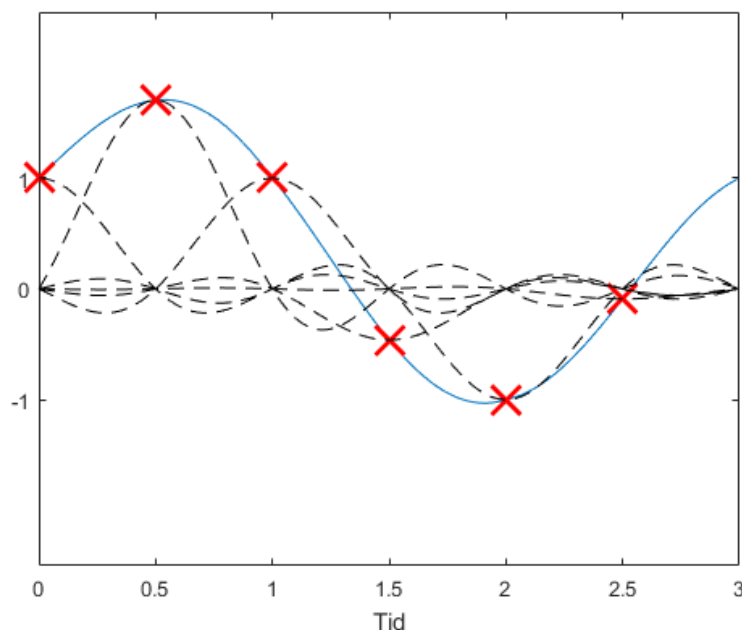
I Kapitel 3 beskriver vi hvordan vilkårlige periodiske signaler kan repræsenteres som en sum af harmoniske funktioner med forskellige frekvenser og hvordan vilkårlige signaler kan beskrives ved harmoniske funktioner med frekvenser i et interval (evt. alle reelle tal). Man kan derfor for et givet signal tale om det eventuelt er **båndbegrænset**, hvilket skal forstås sådan at signalet kan beskrives alene ved frekvenser i et begrænset interval, ikke alle mulige frekvenser. I første omgang er sådanne intervaller fra 0 til en eller anden øverste frekvens F . Om et sådant båndbegrænset signal gælder **samlingsætningen**:

Hvis signalet $v(t)$ er **båndbegrænset** således at højeste frekvens er $F < 1/(2T_S)$ kan signalet rekonstrueres fuldstændigt ud fra de samplede værdier, $v(kT_S)$:

$$v(t) = \sum_{k=-\infty}^{\infty} v(kT_S) \frac{\sin \pi \frac{t-kT_S}{T_S}}{\pi \frac{t-kT_S}{T_S}} = \sum_{k=-\infty}^{\infty} v(kT_S) \text{sinc} \left(\frac{t-kT_S}{T_S} \right) \quad (2.2)$$

Den nødvendige **samplingfrekvens** $f_S = 1/T_S$ skal altså være større end $2F$.

Vi er ikke med det forhåndenværende værktøj i stand til at bevise samplingsætningen. Den tidligere version for harmoniske signaler er klart indeholdt i (2.2), da et harmonisk signal selvfølgelig også er båndbegrænset. Anvendelse af interpolationsfunktionerne $\text{sinc}((t-kT_S)/T_S)$ for rekonstruktion illustreres i Figur 2.3. Bemærk at interpolationsfunktionerne ændres med T_S så de netop går gennem 0 for kT_S , $k \neq 0$.



Figur 2.3: Anvendelse af interpolationsfunktioner for et lille udsnit af et tidskontinuerligt signal. Sampleværdierne er markeret med krydser og den tilhørende interpolationsfunktion er stipt. For hele signalet skal anvendes uendeligt mange samplepunkter og uendeligt lange interpolationsfunktioner.

Når alle disse interpolationsfunktioner lægges oven i hinanden ved summationen, vil samplingtidspunkterne således give præcis den rigtige signal værdi $v(kT_S)$, mens man mellem disse får en interpolation ud fra de samplede værdier og de uendeligt mange bidrag giver netop $v(t)$. Rekonstruktionen skal implementeres på en eller anden måde og her bliver

det et problem at der skal uendeligt mange bidrag til at give værdien $v(t)$, og man i givet fald skal vente uendeligt længe på at få det rekonstruerede signal. Man er derfor nødt til at afkorte interpolationsfunktionen på en eller anden måde. Vi kommer til dette i den tilhørende øvelse, hvor vi simpelthen kun anvender den korrekte interpolationsfunktion i tidsintervallet $-4T_S \leq t \leq 4T_S$ og antager at den er 0 udenfor. Dette giver anledning til problemer med rekonstruktionen, hvis samplingfrekvensen vælges alt for tæt på $2 \cdot (\text{højeste frekvens})$ som øvelsen vil demonstrere. I den almindelige teori for elektriske kredsløb udtrykker man dette lidt anderledes, idet interpolationen realiseres ved hjælp af et såkaldt lavpasfilter (se Kapitel 4), men skal forsinkelsen i dette ikke være uendelig er man nødt til at vælge samplingfrekvensen højere end $2 \cdot (\text{højeste frekvens})$.

Man kunne så stille sig det interessante spørgsmål: Hvad sker der hvis et signal samples med mindre end $2 \cdot (\text{højeste frekvens})$? Det er ikke så nemt at forklare ud fra den formulering af samplingsætningen vi har givet nu, men vi vil vende tilbage til dette problem i Afsnit 3.5. Det kan dog allerede nu siges signalet så ikke kan gendannes: Der genereres nogle andre frekvenser end dem der oprindeligt var i signalet. Disse fremkommer ved at nogle frekvenser i signalet blandes sammen og dette fænomen kaldes **aliasering**. I den efterfølgende øvelse kan dette studeres lidt nærmere. Naturligt forekommende signaler filtreres ofte med et lavpasfilter (se Kapitel 4), et såkaldt **antialiaseringsfilter**, før samplingen så de for høje frekvenser fjernes.

Bemærk i øvrigt at i alle anvendelser af MATLAB er signaler nødvendigvis samplede da de jo skal repræsenteres ved vektorer med et endeligt antal elementer. Vi har ikke betragtet dette som et problem, men blot placeret værdierne tæt nok i tid til at det så pænt ud på plottene. Det er klart at samplingsætningen altid skal tages i betragtning for et givet signal og normalt er den langt mere økonomisk målt i antallet af værdier end vores pænhedskrav. I den efterfølgende øvelse opererer MATLAB med en sådan “oversampling” af signalet for at beskrive interpolationen mellem de samplede værdier. Når signalet samples med en meget kort T_S vil rekonstruktionen praktisk talt bestå af blot at lægge de samplede værdier ved siden af hinanden, og det er jo netop sådan vi tegner kurver i MATLAB.

2.3 Kvantisering

Digitalt udstyr kræver en **kvantisering** af signalet, dvs. kun visse definerede værdier er mulige, ikke alle de reelle tal. Stiltiende har vi mødt dette i vores forskellige MATLAB anvendelser, men det har aldrig været et problem da MATLABs præcision har været noget bedre end eksemplerne har krævet.

I kurset 34311 mødte vi binær repræsentation af et naturligt tal T med n **bit**, b_{n-1} , b_{n-2} ,

..., b_2, b_1, b_0 , hvor hver bit kan være 0 eller 1:

$$T = b_{n-1}2^{n-1} + b_{n-2}2^{n-2} + \dots + b_22^2 + b_12 + b_0$$

Men behandling af signaler kræver lidt flere valgmuligheder:

- Heltal (*integer*) - med og uden fortegn. Fx giver 16 bit tallene 0 - 65535, men de kan også bruges til et fortegn og en størrelse (dvs. mellem ± 32767) eller endnu smartere, det såkaldte 2's komplement, hvor et negativt tal x er repræsenteret ved $2^n - |x|$. Det giver talområdet -32768 - $+32767$, i alt $65536 = 2^{16}$ værdier. Eksempelvis er -1 repræsenteret som n 1'ere. Det smarte ved 2's komplement er at man kan regne direkte på det uden at skulle gøre sig overvejelser om fortegn.
- Fastkommatal (*fixed-point number*). En vis del af de anvendte bits kan afsættes til tallets heldel og en vis del til brøkdelen, eller man kan betragte det som et helt tal, der skal divideres med 2 opløftet til en eller anden potens (kommaets placering). Stadig kun 65536 værdier med 16 bits, men talområdet er ændret.
- Flydendekommatal (*floating-point number*). Her anvendes en del til den såkaldte mantisse der er de betydende cifre i tallet, mens resten anvendes til at repræsentere en eksponent, ganske som vi skriver $2,2251 \cdot 10^{-308}$. I digitale systemer anvendes som regel grundtallet 2 i stedet for 10. MATLAB anvender som standard *IEEE Double-Precision Floating-Point Format*, hvor tal forskellige fra 0 kan have numerisk værdi mellem $\text{realmin} = 2,225 \cdot 10^{-308}$ og $\text{realmax} = 1,7977 \cdot 10^{+308}$. Tallet $\text{eps} = 2,2 \cdot 10^{-16} = 2^{-52}$ angiver mantissens præcision (52 bits), sådan at forstå at $1 - \text{eps}$ er det tal nærmest på 1 som kan skelnes fra 1. I alt anvendes 64 bits, så selv her er der kun et endeligt antal værdier mulige, nemlig 2^{64} .

I signalbehandlingsudstyr må man typisk stille sig tilfreds med en noget mindre præcision end i MATLAB. Fastkommatal anvendes ofte, da de er langt mindre komplekse at realisere, men til gengæld skal man have meget styr over størrelsen af signalerne. For flydendekommatal er operationer besværliggjort af at man ikke bare kan operere på mantissen. Multiplikation er nogenlunde simpel: Mantisserne multipliceres og eksponenterne adderes. Ofte efterfølges det dog af en normalisering hvor mantisserne bringes til at ligge mellem $1/2$ og 1 så der ikke tabes noget præcision ved at mantissebit "falder ud over kanten". Addition kræver at man først justerer tallene så eksponenterne er ens og derefter adderer mantisserne.

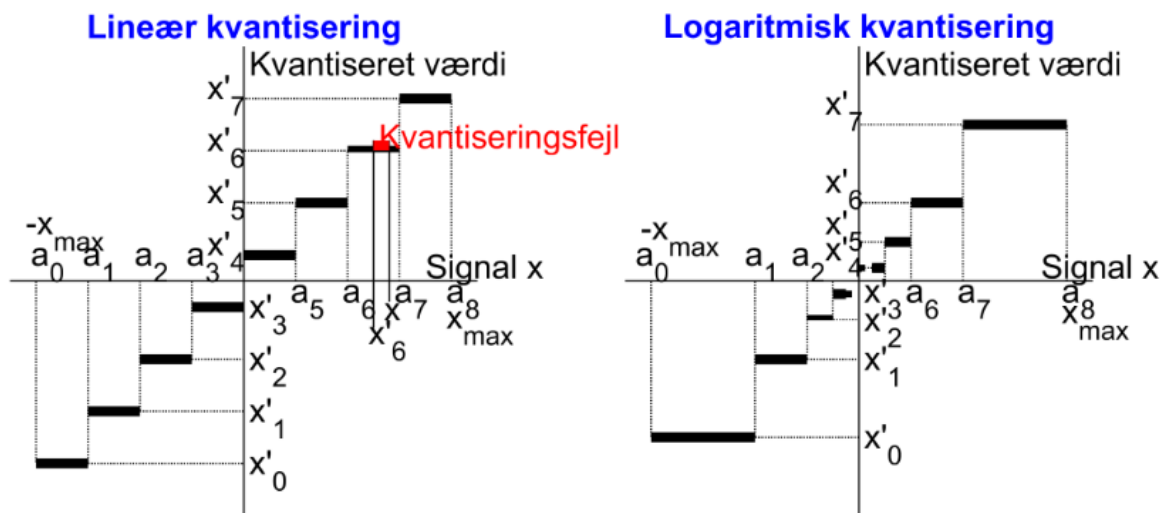
Nu vil vi behandle kvantisering af samplede analoge signaler generelt. Lad os antage at vi kender signalets **udstyringsområde**, dvs. det interval hvori signalets værdier kan forekomme. Nedenfor antager vi dette til at være $[a_0; a_M]$ og ofte ligger endepunkterne a_0 og a_M symmetrisk om 0. Skulle et signal alligevel komme udenfor taler vi om

overstyring. Kvantisering foregår ved at dele udstyringsområdet i M **kvantiseringsintervaller**, $[a_n; a_{n+1}[$:

$$a_0 < a_1 < \dots < a_n < a_{n+1} < \dots < a_M$$

og **nummerere** disse fra 0 til $M - 1$. Dette er - noget anderledes udtrykt - umiddelbart det samme som beskrivelsen af fastkommatal ovenfor, hvor heltallet er nummeret fra 0 til $M - 1$ og længden af de ens intervaller er $udstyringsområde \cdot 2^{-\text{kommaets placering}}$. Når signalet er placeret i et sådant interval, $[a_n; a_{n+1}[$, og repræsenteret ved nummeret, n , er det ikke muligt på nogen måde at genvinde den oprindelige værdi. Man taler om at kvantisering er en **irreversibel** proces, hvor man uundgåeligt mister noget, også kaldet **kvantiseringsfejl**. Når værdien på et senere tidspunkt skal **rekonstrueres** ud fra intervalnummeret må man gøre det så godt som muligt. Som regel vælger man intervallets **midtpunkt** til at repræsentere alle intervallets værdier. Nedenfor bruger vi betegnelsen x'_n for rekonstruktionsværdien for alle x i intervallet $[a_n; a_{n+1}[$. I Figur 2.4 vises kvantiseringsfejlen $x - x'_6$ for et x der falder i intervallet $[a_6; a_7[$.

Hvis de M intervaller vælges ens taler man om **lineær kvantisering** ellers om **ulineær kvantisering**. Begge disse er illustrerede i Figur 2.4. For lineær kvantisering ligger rekonstruktionsværdierne på en ret linie. Lineær kvantisering er derfor ret nemt at implementere.



Figur 2.4: Lineær og ulineær kvantisering (logaritmisk) i 8 niveauer.

Eksempel 2.1 - Lineær kvantisering og signal/støjforhold

Lad udstyringsområdet være $[a_0; a_M] = [-1; +1]$ og $M = 8$ som på figuren. De 8 intervaller bliver så $[-1; -0,75[$, $[-0,75; -0,5[$, $[-0,5; -0,25[$, $[-0,25; 0[$, $[0; 0,25[$, $[0,25; 0,5[$, $[0,5; 0,75[$, $[0,75; 1]$.

$[0, 5; 0, 75[$, $[0, 75; 1[$, som alle har længde 0,25. Den på figuren viste spænding er 0,7 som altså ligger i intervallet $[0, 5; 0, 75[$, interval nr. 6. Intervalnummeret kan udregnes

$$x = 0,7 \quad \Rightarrow \quad n = \left\lfloor \frac{x - (-1)}{0,25} \right\rfloor = 6$$

hvor $\lfloor \cdot \rfloor$ som sædvanligt angiver nedrunding (MATLAB `floor()`). Alle værdier i et givet kvantiseringsinterval rekonstrueres som én værdi, x' , for dette interval - som regel midtpunktet i intervallet. For $x = 0,7$ er det $x'_6 = 0,625$. Det medfører en uoprettelig kvantiseringsfejl, $e = 0,075$, da man generelt ikke kan vide hvor signalet var i intervallet. Kvantiseringsfejlen for et interval af længde Δ (her 0,25) kan variere mellem $\pm\Delta/2$ hvis midtpunktet anvendes. Man kan vise at hvis et stokastisk signal fordeler sig jævnt i intervallet bliver den gennemsnitlige fejl kvadreret $\Delta^2/12$. Når vi anvender kvadratet er det fordi det kan relateres til elektrisk effekt, men mere om det i Kapitel 4. Det er almindelig praksis at udtrykke kvaliteten af et signal ved det såkaldte **signal/støjforhold** $= (\text{signalets værdi})^2 / (\text{støjens værdi})^2$. Hvis vi ser på alle værdier i interval 6 kan vi sige at midtpunktet 0,625 er repræsentativt for signalet så signal/støjforholdet (*SNR* efter engelsk *signal-to-noise ratio*) for dette interval nr. 6 kunne udregnes som

$$SNR_6 = \frac{0,625^2}{0,25^2/12} = 75$$

Tilsvarende får vi $SNR_4 = 3$, $SNR_5 = 27$ og $SNR_7 = 147$. Signal/støjforholdet er altså rigtigt dårligt for svage signaler som dem i interval 4, $[0; 0,25[$, (og 3). \square

Dette eksempel illustrerer at man kunne få en meget bedre kvalitet ved at lade kvantiseringsintervallerne være forskellige (ulinear kvantisering) og afhænge af signalets styrke. Et eksempel på dette er den viste **logaritmiske kvantisering**, hvor kvantiseringsintervallerne vokser med signalet som vist på Figur 2.4. Rekonstruktionsværdierne følger dermed en logaritmefunktion. Signalet kunne fx være et tale- eller musiksignal, hvor kvantiseringsfejls hørbarhed afhænger af signalets størrelse, så man kan tillade sig større fejl for store signalværdier. Fx benytter man sig i telefonsystemer af logaritmisk kvantisering, hvor intervallets størrelse er proportional med signalværdien. Dette muliggør at de mindste intervaller har en nøjagtighed som 12 bits lineær kvantisering, selvom kun 8 bit anvendes (Noterne “Introduktion til netværk og protokoller” om digitalisering). Dette grundsystem kaldes ofte **PCM**. Der samples som beskrevet i noterne med 8000 Hz og informationsstrømmen er således 64000 bit/s. Pga. de nævnte problemer med rekonstruktion ud fra et endeligt antal værdier kan PCM kun overføre signaler præcist op til 3400 Hz - over denne frekvens dæmpes de hurtigt så aliasering undgås (de dæmpes også under 300 Hz af andre tekniske grunde). En musik-CD anvender derimod lineær kvantisering med 16 bits og en samplingfrekvens på 44100 Hz.

Eksempel 2.2 - Logaritmisk kvantisering og signal/støjforhold

Lad udstyringsområdet igen være $[a_0; a_M] = [-1; +1]$ og $M = 8$ som på figuren. De 8 intervaller bliver så $[-1; -0,5[$, $[-0,5; -0,25[$, $[-0,25; -0,125[$, $[-0,125; 0[$, $[0; 0,125[$, $[0,125; 0,25[$, $[0,25; 0,5[$, $[0,5; 1[$, som har længderne 0,125, 0,125 og 0,5. Den på figuren viste spænding er 0,7 som altså ligger i intervallet $[0,5; 1[$, interval nr. 7. Her rekonstruerer vi også som midtpunktet i intervallet. For $x = 0,7$ er det $x'_7 = 0,75$. Det medfører en uoprettelig kvantiseringsfejl, $e = 0,05$ (sjovt nok mindre end for den lineære kvantisering, men det er et tilfælde). Hvis vi ser på alle værdier i interval 7 kan vi sige at midtpunktet 0,75 er repræsentativt for signalet så signal/støjforholdet for dette interval nr. 7 med længde 0,5 kunne udregnes som

$$SNR_7 = \frac{0,75^2}{0,5^2/12} = 27$$

Tilsvarende får vi $SNR_5 = SNR_6 = 27$, altså uafhængigt af signalets styrke, mens $SNR_4 = 3$ stadig er dårligere. Det mindste interval er dog blevet mindre, $[0; 0,125[$, så der er ikke så stor chance for at ramme ind i det. Med lidt større M kan man holde et konstant signal/støjforhold over et ret stort område. For det beskrevne digitale telefon-system er $M = 16$ og 4 bit anvendes til yderligere at inddele de eksponentielt voksende intervaller lineært. \square

Sampling og kvantisering kaldes ofte under ét for **analog/digital-konvertering** forkortet **A/D-konvertering**. Valg af samplingfrekvens og kvantisering har stor betydning for kompleksiteten af udstyr til behandling af signaler, men i dette kursus vil det ikke bekymre os videre frem, da vi anvender MATLABs præcision.

Moderne transmissionssystemer som fx mobiltelefon eller digitalt TV anvender data-kompression for at reducere mængden af producerede data. Grundlæggende udnytter datakompression sammenhængen mellem forskellige samplinger til at anvende et mindre antal bit pr. sample (et eksempel er det såkaldte ADPCM, adaptiv PCM, 4 bit/sample og talekanaler i GSM som reducerer helt ned til omkring 1,7 bit/sample). Dette er et stort emne for senere kurser på DTU. ADPCM anvender foruden sammenhængen også kvantiseringsintervaller der styres af signalets karakter.

KAPITEL 3

Tid og frekvens

Indtil nu har vi fulgt den mest naturlige fremgangsmåde man kan forestille sig i analysen af signaler: Vi har beskrevet dem som en funktion af tiden og afbildet disse tidsfunktioner. Ved behandling af sampling kunne det ses at vi kom i en eller anden form for vanskeligheder med at udtrykke et signals egenskaber: Vi var nødt til at indføre **begrebet frekvens**. Vi gjorde dette for signaler der indlysende havde en frekvens, nemlig harmoniske funktioner, og vi tog derefter en smule forskud på nærværende kapitel ved at hævde at alle signaler kan beskrives ved frekvenser. Det er karakteristisk for al behandling af signaler at man er nødt til at have to betragtningssvinkler. Vi kalder dem domæner, så vi har altså **tidsdomænet** og **frekvensdomænet** og formålet med dette kapitel er at relatere disse til hinanden. Dette er et meget vigtigt emne i behandling og fortolkning af signaler. I vores eksempel om sampling vil man skulle indføre et eller andet begreb i tidsdomænet, fx størst mulige hældning (forandringshastighed) for at kunne beskrive en slags samplingsætning. I frekvensdomænet kan dette gøres langt nemmere og mere præcist. I transmissionssystemer vil der ofte være en båndbegrænsning, der betyder at man kun kan transmittere frekvenser op til en vis frekvens eller frekvenser i et vist interval ("bånd"). Det er fx tilfældet med radiokommunikation hvor der skal være plads til mange transmissioner samtidigt og vi har også i kurset 34311 set at almindelige telefonkanaler er begrænsede til frekvensområdet 300 - 3400 Hz. Det er derfor af yderste vigtighed at kende frekvensindholdet af et signal og eventuelt at kunne påvirke dette. I Kapitel 4 vil vi se at det for de fleste anvendte systemer er langt nemmere at analysere deres opførsel i frekvensdomænet. Dette forhold har gjort at det er nærmest uomgængeligt at forstå relationerne mellem de to domæner og det er grundlag for både teoretiske analyser og praktiske implementeringer. Derefter er det op til brugeren at vælge det domæne hvori problemet lettest løses. Signalet er selvfølgelig altid det samme - det er kun analysen af det der tager forskellige udgangspunkter.

Vi begynder med at studere periodiske signaler, der er lidt mere generelle end de harmoniske funktioner i Kapitel 2 og når dernæst til mere almindelige signaler, specielt de i Kapitel 1 introducerede pulser.

Fra matematik vil den komplekse eksponentialfunktion være kendt [eNote 29] og det kan

hjælpe os til at gøre formlerne lidt enklere. En **kompleks harmonisk funktion** med **frekvensen** f er

$$e^{j2\pi ft} = \cos 2\pi ft + j \sin 2\pi ft = e^{j\omega t} = \cos \omega t + j \sin \omega t \quad (3.1)$$

hvor vi har anvendt j som betegnelse for den imaginære enhed. Dette er almindelig praksis i elektro- og kommunikationsteknologi, vistnok fordi det er nemmere at læse end i (og ofte anvendes i som betegnelse for strøm). Læg mærke til at den komplekse harmoniske funktion har de fra Afsnit 2.1 kendte harmoniske funktioner (cosinus og sinus) med samme frekvens, f , langs akserne. Den komplekse harmoniske funktion kan beskrives som en vektor af længde 1 fra $(0, 0)$ til $(\cos(2\pi ft), \sin(2\pi ft))$ og denne vektor roterer i koordinatsystemet med endepunkt på enhedscirklen når tiden går med en vinkelhastighed svarende til vinkelfrekvensen $\omega = 2\pi f$. Undertiden omtales denne geometriske fortolkning som en viser (engelsk *phasor*), men det tilføjer ikke rigtigt noget i dette kursus.

Fra (3.1) udleder vi Eulers formel (jf. [eNote 29]):

$$\cos(2\pi ft) = \frac{1}{2}e^{j2\pi ft} + \frac{1}{2}e^{-j2\pi ft} \quad (3.2)$$

(der er egentlig også en version for sinus, men den skal vi ikke anvende i dette kursus).

3.1 Fourierrækker for periodiske signaler

Vi vil først beskæftige os med et signal $v(t)$ der er **periodisk med periode P** , dvs. $v(t) = v(t + nP)$ for ethvert heltalligt n . I kurset Matematik 2 (Kapitel 6 og appendiks A) bevises det at et signal der er en periodisk, kvadratisk integrabel funktion kan betragtes som frembragt af en (uendelig) sum af komplekse harmoniske funktioner:

$$v(t) = \sum_{m=-\infty}^{\infty} V_m e^{j2\pi \frac{m}{P} t} \Leftrightarrow V_m = \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt \quad (3.3)$$

Denne fremstilling af signalet kaldes en **Fourierrække** og værdierne V_m der vægter de enkelte harmoniske funktioner kaldes **Fourierkoefficienter**. Frekvensen i den m 'te harmoniske funktion er m/P , hvor $1/P$ kaldes **grundfrekvensen**. Bemærk at vi anvender store bogstaver for Fourierkoefficienterne (**frekvensdomænet**) og små bogstaver for signalet i **tidsdomænet**. Dette vil så vidt muligt blive gjort overalt i dette kursus inklusiv i MATLAB programmer da MATLAB kender forskel på store og små bogstaver og det er godt hele tiden at gøre sig klart hvilket domæne man er i. Vi negligerer i dette kursus forskellige problemer vedrørende om den uendelige række i (3.3) faktisk konvergerer mod $v(t)$ - det er ikke helt tilfældet hvis der er spring (diskontinuiteter) i funktionen og det

behandles i kurset Matematik 2.

Ovenstående sum er udtrykt ved komplekse harmoniske funktioner og der forekommer det lidt besynderlige begreb “negative frekvenser” for $m < 0$. Disse er indført for at lette formuleringen; de er ikke udtryk for nogle nye frekvenser, fx er $-3/P$ og $3/P$ begge udtryk for den samme fysiske frekvens, $3/P$. Det kan ses af Eulers formel, (3.2), at en reel harmonisk funktion, $\cos(2\pi(m/P)t)$, indeholder både den negative frekvens $-m/P$ og den positive frekvens m/P :

$$\cos\left(2\pi\frac{m}{P}t\right) = \frac{1}{2}e^{j2\pi\frac{m}{P}t} + \frac{1}{2}e^{j2\pi\frac{-m}{P}t}$$

Vi skal senere - ved behandling af aliasering og i Kapitel 6 om modulation - se at det er ret smart at tænke de negative frekvenser med også selvom de fysisk er udtryk for samme frekvens.

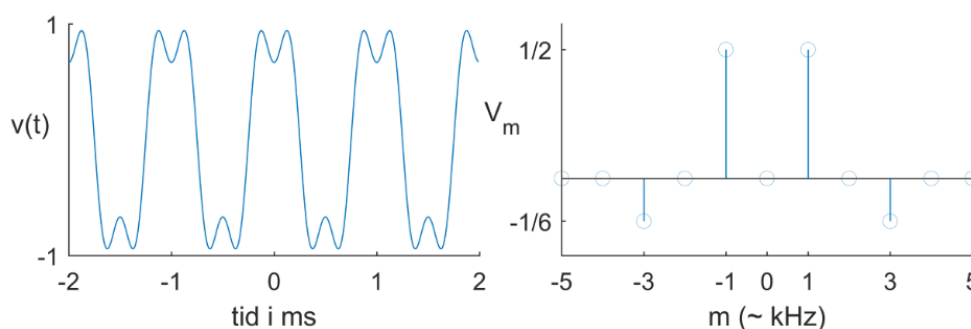
Eksempel 3.1 - Fourierrække for harmonisk funktion

Lad $v(t)$ være sammensat af to reelle harmoniske funktioner

$$v(t) = \cos(2\pi \cdot f_0 t) - \frac{1}{3} \cos(2\pi \cdot 3f_0 t)$$

med $f_0 = 1$ kHz. Perioden for denne $v(t)$ er $P = 1/f_0$ da første led har denne periode og andet led har netop 3 perioder i løbet af P . For at udvikle Fourierrækken anvendes Eulers formel og vi får

$$\begin{aligned} v(t) &= \cos(2\pi \cdot f_0 t) - \frac{1}{3} \cos(2\pi \cdot 3f_0 t) \\ &= \frac{1}{2}e^{j2\pi \cdot f_0 t} + \frac{1}{2}e^{j2\pi \cdot (-f_0)t} - \frac{1}{6}e^{j2\pi \cdot 3f_0 t} - \frac{1}{6}e^{j2\pi \cdot (-3f_0)t} \\ &= \sum_{m=-\infty}^{\infty} V_m e^{j2\pi \cdot m f_0 t} \Rightarrow V_{-3} = V_3 = -\frac{1}{6} \text{ og } V_{-1} = V_1 = \frac{1}{2} \end{aligned}$$



Figur 3.1: Periodisk signal og dets Fourierrække (spektrum)

Værdierne for V_m fås altså ved blot at identificere koefficienter i den uendelige sum og kun $m = \pm 1$ samt $m = \pm 3$ findes. Disse koefficienter fortæller os altså “styrken” (amplituden) af de fysiske frekvenser $f_0 = 1$ kHz og $3f_0 = 3$ kHz. Alle andre Fourierkoefficienter ses at være 0. Figur 3.1 viser $v(t)$ og V_m . Som det ses kan man nemt finde Fourierrækken for summer af sædvanlige harmoniske funktioner, hvor frekvenserne i hvert enkelt led er et multiplum af en grundfrekvens. Hver af de reelle harmoniske funktioner giver anledning til to (komplekse) Fourierkoefficienter og der indgår kun multipla af grundfrekvensen som f_0 i eksemplet. \square

Hvis man går til en lidt mere generel periodisk funktion end i eksemplet, kommer der flere betydende Fourierkoefficienter. Man siger at det periodiske signal har et **spektrum**, $\{V_m\}$, hvor V_m angiver bidraget ved frekvensen m/P , m et helt tal, $-\infty < m < \infty$. Udregning af et signals spektrum kaldes normalt **Fouriertransformation**. Spektret $\{V_m\}$ har i almindelighed komplekse værdier og man kalder modulus $|V_m|$, $-\infty < m < \infty$, for signalets **amplitudespektrum**, mens vinklen $\angle V_m$, $-\infty < m < \infty$, kaldes **fasespektrum**. Spektret er generelt komplekst, men kan være reelt i særlige tilfælde som vi skal se nedenfor.

3.1.1 Egenskaber for spektrum for periodisk funktion

For et reelt signal $v(t)$ kan man nemt vise at V_m og V_{-m} er hinandens kompleks konjugerede:

$$V_{-m} = \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{-j2\pi \frac{-m}{P} t} dt = \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{j2\pi \frac{m}{P} t} dt = V_m^*$$

I et vist tilfælde fås et reelt spektrum, nemlig når $v(t)$ er en lige funktion, dvs. en funktion der er symmetrisk om tiden 0. Her vil $v(t) = v(-t)$ få den imaginære del til at gå ud i integralet:

$$\begin{aligned} V_m &= \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt \\ &= \frac{1}{P} \int_0^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} + v(-t) e^{-j2\pi \frac{m}{P} (-t)} dt = \frac{1}{P} \int_0^{P/2} 2v(t) \cos\left(2\pi \frac{m}{P} t\right) dt \end{aligned}$$

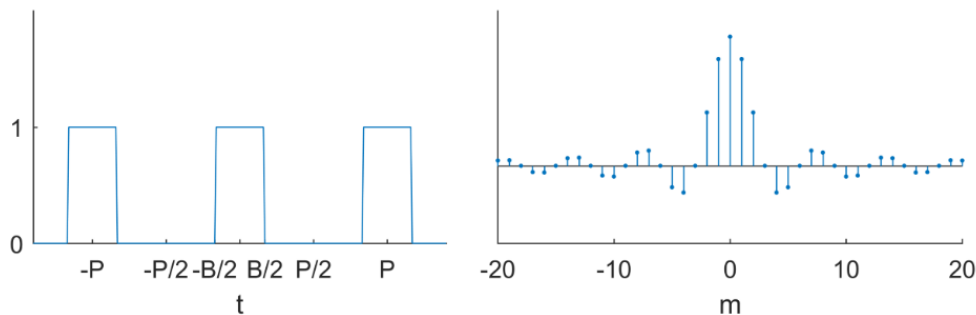
For en sådan lige funktion bliver $V_m = V_{-m}$ og reel. Fourierrækken bliver også meget simplere, da hver frekvens kun bidrager med $2V_m \cos((2\pi m/P)t)$, $m > 0$. Et eksempel følger nedenfor og det fortsættes i den tilknyttede øvelse, som viser hvordan en periodisk funktion opbygges af sin Fourierrække.

Eksempel 3.2 - Fourierrække for firkantsignal

Lad $v(t)$ være et firkantsignal med periode P og lad en periode være **firkantpuls** $p(t)$ med bredden B :

$$p(t) = \begin{cases} 1 & \text{for } |t| \leq \frac{B}{2} \\ 0 & \text{for } \frac{B}{2} < |t| \leq \frac{P}{2} \end{cases}$$

Figur 3.2 viser et eksempel på dette signal med $B = P/3$. Vi har valgt tidsaksen således at $v(t)$ er en lige funktion.



Figur 3.2: Periodisk firkantsignal og dets spektrum.

For vilkårlig B og P får vi Fourierkoefficienterne

$$\begin{aligned} V_m &= \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt = \frac{1}{P} \int_{-B/2}^{B/2} e^{-j2\pi \frac{m}{P} t} dt = \frac{1}{P} \left[\frac{e^{-j2\pi \frac{m}{P} t}}{-j2\pi \frac{m}{P}} \right]_{-B/2}^{B/2} \\ &= \frac{1}{P} \frac{e^{-j2\pi \frac{m}{P} \frac{B}{2}} - e^{-j2\pi \frac{m}{P} (-\frac{B}{2})}}{-j2\pi \frac{m}{P}} = \frac{B \sin \pi m \frac{B}{P}}{P \pi m \frac{B}{P}} = \frac{B}{P} \text{sinc} \left(m \frac{B}{P} \right) \end{aligned}$$

Spektret vises også på Figur 3.2. Man kan se at det er uendeligt bredt i den forstand at der er betydelige dele af spektret helt til uendelig. Værdierne aftager kun med $1/m$ og det er for lidt til at man kan nøjes med en lille del af det for at forklare signalet. Det er altid sådan, at når et signal er begrænset i tid, dvs. lig 0 uden for et eller andet område om $t = 0$, skal der uendeligt mange frekvenser til at opbygge det, men det er ikke altid at de er så betydningsfulde som her. Det er de stejle flanker på signalet der bevirker den langsomme hændelse af spektret; en mere blød puls ville være nemmere at have med at gøre. \square

I udtrykket for Fourierkoefficienterne, (3.3), integreres fra $-P/2$ til $P/2$, men det er uden betydning, blot der **integreres over en periode** kan denne placeres hvor det skal være

tidsmæssigt. Lad os forskyde integrationsgrænserne med tiden a :

$$\begin{aligned}
\frac{1}{P} \int_{-P/2+a}^{P/2+a} v(t) e^{-j2\pi \frac{m}{P} t} dt &= \frac{1}{P} \int_{-P/2+a}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt + \frac{1}{P} \int_{P/2}^{P/2+a} v(t) e^{-j2\pi \frac{m}{P} t} dt \\
&= \frac{1}{P} \int_{-P/2+a}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt + \frac{1}{P} \int_{-P/2}^{-P/2+a} v(\tau + P) e^{-j2\pi \frac{m}{P} (\tau+P)} d\tau \\
&= \frac{1}{P} \int_{-P/2+a}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt + \frac{1}{P} \int_{-P/2}^{-P/2+a} v(\tau) e^{-j2\pi \frac{m}{P} \tau} e^{-j2\pi m} d\tau \\
&= \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt = V_m
\end{aligned}$$

Selvom placeringen af perioden således ingen betydning har, har det derimod stor betydning om funktionen $v(t)$ forskydes i tid i forhold til den tid der anvendes i eksponentialfunktionerne i integralet. Lad $w(t) = v(t - A)$ være **$v(t)$ forskudt tiden A** :

$$\begin{aligned}
W_m &= \frac{1}{P} \int_{-P/2}^{P/2} w(t) e^{-j2\pi \frac{m}{P} t} dt = \frac{1}{P} \int_{-P/2}^{P/2} v(t - A) e^{-j2\pi \frac{m}{P} t} dt \\
&= e^{-j2\pi \frac{m}{P} A} \frac{1}{P} \int_{-P/2}^{P/2} v(t - A) e^{-j2\pi \frac{m}{P} (t-A)} dt = e^{-j2\pi \frac{m}{P} A} \frac{1}{P} \int_{-P/2-A}^{P/2-A} v(\tau) e^{-j2\pi \frac{m}{P} \tau} d\tau \\
&= e^{-j2\pi \frac{m}{P} A} V_m
\end{aligned}$$

Fourierkoefficienterne multipliceres altså med komplekse eksponentialfunktioner hvis en funktion forskydes i tid.

3.1.2 Udregning af Fourierkoefficienter med MATLAB

Integralet i (3.3) giver den principielle sammenhæng og er mest for teoretiske betragtninger, og ovenfor så vi hvordan man nemt kom til resultatet for den harmoniske funktion cosinus.

I praksis vil vi lade MATLAB lave de numeriske beregninger af Fourierkoefficienter. Denne beregning kaldes ofte Fouriertransformation

I en øvelse i kurset 34302 beregnede vi et integral ved at tilnærme det med en sum. Deler vi tidsperioden i N lige store dele hver med længde $T_S = P/N$, vil tiden for del nr. k

være $t = kT_S$, $k = 0, \dots, N - 1$ og vi får da tilnærmelsen

$$\begin{aligned} V_m &= \frac{1}{P} \int_0^P v(t) e^{-j2\pi \frac{m}{P} t} dt \approx \frac{1}{P} \sum_{k=0}^{N-1} v(kT_S) e^{-j2\pi \frac{m}{P} kT_S} \cdot T_S = \frac{1}{P} \sum_{k=0}^{N-1} v(kT_S) e^{-j2\pi \frac{m}{P} k \frac{P}{N}} \cdot \frac{P}{N} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} v(kT_S) e^{-j2\pi \frac{km}{N}} \end{aligned}$$

Da eksponentialfunktionen er periodisk med N får man i tilnærmelsen at spektret for $v(t)$ kun udregnes for nogle frekvenser, $0 \leq m < N - 1$ og $V_{m+N} = V_m$, så det **udregnede spektrum** bliver **periodisk** i modsætning til (3.3). Det betyder at approksimationen kun gælder for $0 \leq m \leq N/2$. Man kunne tænke sig at det virkede for m op til $N - 1$, men pga. det periodiske spektrum giver værdierne fra $N/2$ til $N - 1$ bidrag for de negative frekvenser $-(N/2)/P, \dots, -1/P$ og dem kender man i forvejen når $v(t)$ er et reelt signal, idet V_{-m} som tidligere skrevet er den kompleks konjugerede af V_m . Det ser vi lidt nærmere på i det efterfølgende eksempel.

MATLAB indeholder funktionen `fft(v)` som udregner den viste sum

$$\sum_{k=0}^{N-1} v(kT_S) e^{-j2\pi \frac{km}{N}} \text{ for } m = 0, \dots, N - 1$$

så udtrykt i MATLAB foregår udregningen af V_m ved $V = \text{fft}(v)/N$. Navnet `fft` på MATLABs funktion betyder “Fast Fourier Transform” og hentyder egentlig til at der for visse N findes en meget smart algoritme til udregning af Fouriertransformationen. MATLAB bruger selvfølgelig denne, når N har de rette værdier, men er i øvrigt i stand til at transformere for ethvert N , så navnet burde måske snarere være DFT, Diskret Fourier Transformation. Det er meget vigtigt at bemærke at `fft` i MATLAB kræver at elementet for $t = 0$ står på indeks 1, og at det transformerede resultat har frekvensen 0 i indeks 1 og derefter følger de positive frekvenser op til $N/2$ efterfulgt af de negative (i omvendt orden som beskrevet ovenfor) sluttende med $m = N - 1$. $N/2$ -frekvensen er lidt speciel ved både at have den negative og positive komponent i sig hvis N er lige. Somme tider ønsker man en mere naturlig rækkefølge af frekvensbidragene i vektoren V , fx for et plot, og funktionen `fftshift(V)` omordner V (det er altså ikke en ny transformation) så frekvensbidragene svarer til frekvenserne $-(N/2)/P, \dots, -1/P, 0, 1/P, \dots, (N/2 - 1)/P$ (hvis N er lige).

Hvis vi kender spektret for et periodisk signal, $\{V_m\}$, kan vi anvende (3.3) til at finde signalets værdier i nogle samplepunkter, nemlig for $t = kP/N$:

$$v\left(k \frac{P}{N}\right) = \sum_{m=-\infty}^{\infty} V_m e^{j2\pi \frac{m}{P} \cdot k \frac{P}{N}} \approx \sum_{m=0}^{N-1} V_m e^{j2\pi \frac{km}{N}}$$

hvor vi tilnærmer udtrykket ved at anvende en begrænset del af spektret, fx de N værdier vi har beregnet i MATLAB. MATLAB har en funktion til udførelse af denne **inverse Fouriertransformation** Fouriertransformation, nemlig `ifft(V)` som beregner

$$\frac{1}{N} \sum_{m=0}^{N-1} V_m e^{j2\pi \frac{km}{N}}$$

Denne kan benyttes i ovenstående transformation fra frekvensdomænet (det periodiske spektrum V_m) til tidsdomænet (den periodiske funktion $v(t)$ samplet til tidspunkterne $t = k(P/N) = kT_S$, $k = 0, \dots, N-1$). Man kan se at man skal udregne $v = N * \text{ifft}(V)$ og at man selvfølgelig kun får en tilnærmelse til funktioner $v(t)$ indeholdende de første $N/2$ frekvenser. Bemærk at frekvensbidragene skal være rigtigt placeret, dvs. element 1 i V er frekvensen 0. Vi illustrerer denne tilnærmelse yderligere i den tilknyttede øvelse. Man kan nemt ved en prøve overbevise sig om at $\text{ifft}(\text{fft}(v)) = v$, så MATLABs funktioner er også hinandens inverse ligesom det vi regner ud: $N * \text{ifft}(\text{fft}(v)/N)$. Placeringen af N i disse udregninger kan altså gøres forskelligt; der skal blot indgå en samlet faktor $1/N$ i de to funktioner - vores placering begrundes i Afsnit 3.6, mens MATLAB har sin begrundelse i Afsnit 3.2.3. Sommetider anvendes $1/\sqrt{N}$ i begge funktioner.

Eksempel 3.3 - MATLAB udregning af Fourierkoefficienter

Som i det tidligere eksempel lader vi $v(t)$ være sammensat af to reelle harmoniske funktioner

$$v(t) = \cos(2\pi \cdot f_0 t) - \frac{1}{3} \cos(2\pi \cdot 3f_0 t)$$

med $f_0 = 1$ kHz. Perioden for denne $v(t)$ er $P = 1/f_0 = 1$ ms. Hvis vi sampler en periode $N = 8$ gange fås $T_S = 1/8$ ms, og den transformerede udregnes for frekvenserne $-4f_0, -3f_0, \dots, 3f_0$, hvilket lige netop er nok ud fra vores tidligere viden om at signalet indeholder $\pm f_0$ og $\pm 3f_0$. Programmet er altså

```
P=1/f0;
N=8;
Ts=P/N;
t=(0:N-1)*Ts;
v=cos(2*pi*f0*t)-1/3*cos(2*pi*3*f0*t);
V=real(fft(v)/N)
```

som giver resultatet

```
V =
-0.0000  0.5000  0.0000 -0.1667  0.0000 -0.1667  0.0000  0.5000
```

ganske som tidligere udregnet, da dette er V_m for $m = 0, \dots, 7$, der svarer til frekvenserne

0 til $(N/2 - 1) \cdot f_0$ eller hvis de sidste 4 af disse (korrekt) betragtes som negative: 8 punkter i intervallet $-1/(2T_S) \leq f < 1/(2T_S)$, $-4 \text{ kHz} \leq f < 4 \text{ kHz}$. Læg mærke til at vi har været meget omhyggelige med at starte perioden i $t = 0$ og at have netop 8 punkter (fra 0 til $N - 1 = 7$) i perioden. Hvis samplingstiden T_S ikke er en brøkdel af perioden, bliver den samplede funktion ikke periodisk eller den får en meget længere periode og det bevirker at de udregnede Fourierkoefficienter ikke passer med de rigtige. Da funktionen $v(t)$ er lige omkring $t = 0$, har vi tilladt os at sætte `real` ind i resultatet for at få en lidt mere kompakt udskrift. Uden dette bliver der genereret nogle meget små imaginærdele svarende til afrundingsfejl i MATLAB. Man kunne nok føle sig lidt utryg ved at man kun har regnet spektret ud for ganske få punkter. Hvad hvis der var nogle højere frekvenser? Samme program med $N = 16$ udvider frekvensområdet til $-8f_0 \leq f \leq 7f_0$ (stadig frekvensområdet $-1/(2T_S) \leq f < 1/(2T_S)$, men T_S er nu halvt så stor), Vi får:

```
V =
-0.0000  0.5000  0.0000 -0.1667  0.0000  0.0000  0.0000      0
 0.0000      0  0.0000  0.0000  0.0000 -0.1667  0.0000  0.50000
```

Ingen nye, højere frekvenser. Der kommer blot en masse 0-bidrag (for høje frekvenser) mellem de allerede kendte.

Vi kan derimod være helt sikre på at der ikke er frekvenser ind i mellem de udregnede da vi havde en hel periode P af signalet og der derfor kun opstår frekvenser $m/P = mf_0$. Vi kunne prøve at tage en længere periode bestående af to af de oprindelige perioder:

```
t=[t t];
N=2*N;
v=cos(2*pi*f0*t)-1/3*cos(2*pi*3*f0*t);
V=real(fft(v)/N)
```

som giver

```
V =
-0.0000  0  0.5000  0  0.0000  0 -0.1667  0
 0.0000  0 -0.1667  0  0.0000  0  0.5000  0
```

Afstanden mellem frekvenserne for de udregnede spektrumværdier er nu $1/(2P) = f_0/2 = 1/2 \text{ kHz}$ og der ses at være nogle 0-bidrag ind i mellem de oprindelige frekvensbidrag. \square

3.2 Fouriertransformation for ikke-periodiske signaler

De periodiske signaler er ikke så meget bevendt til kommunikation, da der som tidligere omtalt ikke kan knyttes nogen information til dem. De er derimod anvendelige som må-

lesignaler og lignende, og som det vil ses i Kapitel 4 er det i analysen af systemer vigtigt hvad der foregår ved en bestemt frekvens. For kommunikation har vi brug for pulsaftige signaler, som vi allerede viste i eksemplerne i Kapitel 1. For en sådan puls, $v(t)$, kan man også finde et spektrum, $V(f)$, men det bliver over et kontinuert frekvensområde i modsætning til spektre for de periodiske signaler. Ganske som for periodiske signaler kan tidsfunktionen $v(t)$ betragtes som sammensat af harmoniske funktioner, der hver bidrager med $V(f)e^{j2\pi ft}$. Disse har nu frekvenser, f , i et kontinuert område og derfor er sammenhængen beskrevet ved et integral. Vi får altså:

Et ikke-periodisk signal $v(t)$ kan repræsenteres i frekvensdomænet ved $V(f)$ og sammenhængen mellem $v(t)$ og $V(f)$ beskrives ved **Fouriertransformationen**:

$$v(t) = \int_{-\infty}^{\infty} V(f)e^{j2\pi ft} df \quad \Leftrightarrow \quad V(f) = \int_{-\infty}^{\infty} v(t)e^{-j2\pi ft} dt \quad (3.4)$$

Også her kalder vi $|V(f)|$ for **amplitudespektret** og $\angle V(f)$ for **fasespektret**. Bemærk at vi som tidligere beskrevet anvender store bogstaver for spektret i frekvensdomænet og små bogstaver for signalet i tidsdomænet. Nogle gange har vi dog brug for at kunne betegne selve transformationen og det sker ofte ved \mathcal{F} , fx $V(f) = \mathcal{F}(v(t))$.

3.2.1 Egenskaber for Fouriertransformationen

Hvis vi afgrænser os til at $v(t)$ er et reelt signal fås at $V(-f)$ er kompleks konjugeret af $V(f)$ og en reel, lige funktion $v(t)$ har et reelt spektrum.

En vigtig egenskab er at

$$\mathcal{F}(V(t)) = \int_{-\infty}^{\infty} V(t)e^{-j2\pi ft} dt = \int_{-\infty}^{\infty} V(z)e^{j2\pi(-f)z} dz = v(-f)$$

så funktionerne i tidsdomænet og i frekvensdomænet hører sammen som et par. Det vil vi kommentere meget mere efter det følgende eksempel.

Ligesom for Fourierkoefficienterne er det af stor betydning at "tidsregningen" i funktionen og i eksponentialfunktionen er den samme. Lad $w(t) = v(t - A)$ være en tidsforskydning af $v(t)$:

$$W(f) = \int_{-\infty}^{\infty} w(t)e^{-j2\pi ft} dt = e^{-j2\pi fA} \int_{-\infty}^{\infty} v(t - A)e^{-j2\pi f(t-A)} dt = e^{-j2\pi fA} V(f)$$

Et eksempel på denne sammenhæng vises i en senere Figur 3.9.

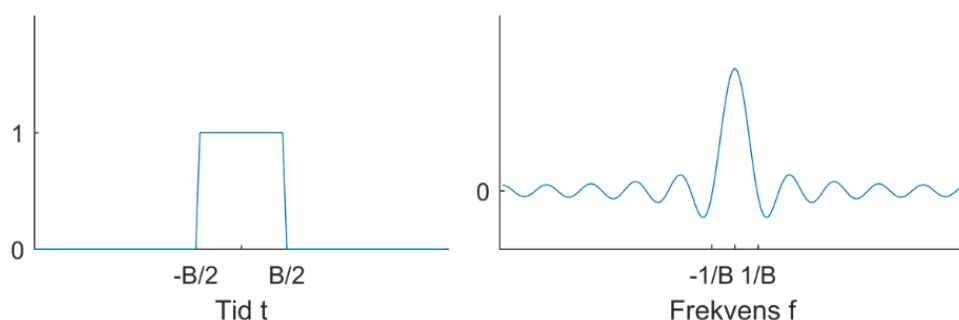
Fouriertransformation behandles også i Matematik 2 (ganske kort i Afsnit 6.6).

Eksempel 3.4 - Fouriertransformation af firkantpuls

Lad $v(t)$ være en firkantpuls med bredden B :

$$v(t) = \begin{cases} 1 & \text{for } |t| \leq \frac{B}{2} \\ 0 & \text{ellers} \end{cases}$$

Figur 3.3 viser et eksempel på en sådan puls. Vi har valgt tidsaksen således at $v(t)$ er en lige funktion.



Figur 3.3: Firkantpuls og dens spektrum.

For vilkårligt B får man

$$\begin{aligned} V(f) &= \int_{-\infty}^{\infty} v(t)e^{-j2\pi ft} dt = \int_{-B/2}^{B/2} e^{-j2\pi ft} dt = \left. \frac{e^{-j2\pi ft}}{-j2\pi f} \right]_{-B/2}^{B/2} \\ &= \frac{e^{-j2\pi f \frac{B}{2}} - e^{-j2\pi f (-\frac{B}{2})}}{-j2\pi f} = B \frac{\sin \pi f B}{\pi f B} = B \text{sinc}(fB) \end{aligned}$$

Spektret vises også på Figur 3.3. Som det kan ses svarer det ganske til spektret for det periodiske firkantsignal i Figur 3.2, men nu er alle frekvenser mulige. Konklusionerne er ganske de samme: Man kan se at spektret er uendeligt bredt i den forstand at der er betydende dele af spektret helt til uendelig. Værdierne aftager kun med $1/f$ og det er for lidt til at man kan nøjes med en lille del af det for at forklare signalet. Det er altid sådan, at når et signal er begrænset i tid, dvs. lig 0 uden for et eller andet område om $t = 0$, skal der uendeligt mange frekvenser til at opbygge det, men det er ikke altid sådan at de er så betydningsfulde som her. Det er de stejle flanker på signalet der bevirker den langsomme hændelsen af spektret; en mere blød puls ville være nemmere at have med at gøre. \square

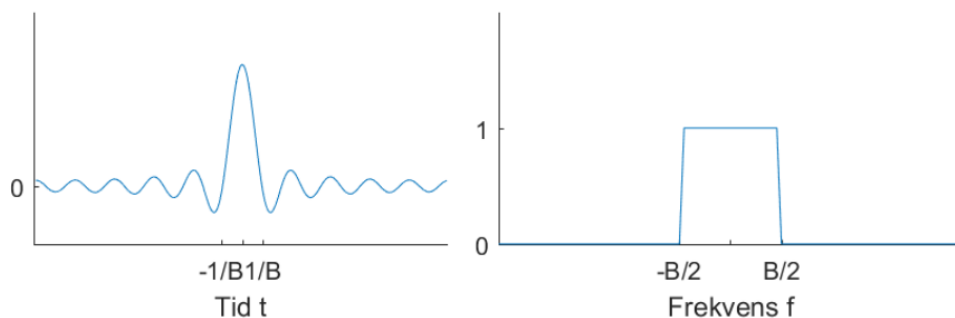
3.2.2 Konsekvenser af begrænsning i tid eller frekvens

Hvordan skaber man så en puls der er begrænset i frekvens, dvs. det vi kalder et **bånd-begrænset** spektrum? Det er nemt at se af den ovenstående egenskab for Fouriertrans-

formationen: $\mathcal{F}(V(t)) = v(-f)$. Vi skal altså fx som puls vælge kurveformen til højre i Figur 3.3, da dens spektrum vil være en firkant som vist til venstre i figuren. Det er illustreret nedenfor og kan også nemt udregnes:

$$\begin{aligned} v(t) &= \int_{-\infty}^{\infty} V(f)e^{j2\pi ft} df = \int_{-B/2}^{B/2} e^{j2\pi ft} df = \left. \frac{e^{j2\pi ft}}{j2\pi t} \right|_{-B/2}^{B/2} \\ &= \frac{e^{j2\pi t \frac{B}{2}} - e^{j2\pi t (-\frac{B}{2})}}{j2\pi t} = B \frac{\sin \pi t B}{\pi t B} = B \text{sinc}(Bt) \end{aligned}$$

Spektret vises også på følgende Figur



Figur 3.4: Tidssignal med båndbegrænset spektrum.

I eksemplerne i Kapitel 1 var det netop forskellen på signalet i Figur 1.2 som med de firkantede pulser er et signal med uendeligt bredt spektrum og det i Figur 1.3 som med de sinc-formede pulser giver et båndbegrænset spektrum. Kort sagt:

Begrænsning i tid giver et uendeligt bredt spektrum, mens begrænsning i frekvens giver et uendelig bredt signal i tid.

Vi har udviklet tidsfunktioner som uendelige summer af komplekse eksponentialfunktioner med en bestemt frekvens (formel (3.3)) og også anvendt sådanne udviklinger for ikke-periodiske funktioner, (3.4). I praktiske anvendelser og i computerbaserede regninger måtte vi begrænse os til et endeligt antal af disse funktioner. Man kunne derfor spørge sig selv om der ikke findes andre funktioner der bedre kan bruges til at repræsentere ikke-periodiske funktioner ved et ret lille antal led i udviklingen. Det gør der faktisk, de såkaldte wavelets. Disse introduceres også i Matematik 2 (Afsnit 6.7)). Frekvensanalysen er dog stadig helt overskyggende i anvendelser.

3.2.3 Udregning af Fouriertransformation med MATLAB

Integralen i Fouriertransformationen i (3.4) er mest for teoretiske overvejelser. Ønsker vi at udregne nogle værdier gør vi som i det foregående afsnit: Anvender MATLAB `fft`-funktionen. **Grundlæggende er `fft` for periodiske signaler og giver som vist i**

det tidligere Afsnit 3.1.2 et periodisk spektrum. Det er derfor nødvendigt at vælge et tidsinterval der er stort nok til at alle signalets væsentlige værdier behandles, dvs. $v(t)$ skal være 0 i kanten af intervallet og udenfor. Kalder vi som i det foregående afsnit det valgte interval P og deler dette i N lige store dele hver med længde $T_S = P/N$, vil tiden for del nr. k være $t = kP/N = kT_S$, $k = -N/2, \dots, N/2 - 1$ og vi får da tilnærmelsen

$$\begin{aligned} V\left(\frac{m}{P}\right) &= \int_{-\infty}^{\infty} v(t)e^{-j2\pi\frac{m}{P}t}dt \approx \int_{-P/2}^{P/2} v(t)e^{-j2\pi\frac{m}{P}t}dt \approx \sum_{k=-N/2}^{N/2-1} v(kT_S)e^{-j2\pi\frac{m}{P}k\frac{P}{N}} \cdot T_S \\ &= T_S \sum_{k=0}^{N/2-1} v(kT_S)e^{-j2\pi\frac{km}{N}} + T_S \sum_{k=-N/2}^{-1} v(kT_S)e^{-j2\pi\frac{km}{N}} \\ &= T_S \sum_{k=0}^{N/2-1} v(kT_S)e^{-j2\pi\frac{km}{N}} + T_S \sum_{k=N/2}^{N-1} v(kT_S)e^{-j2\pi\frac{km}{N}} = T_S \sum_{k=0}^{N-1} v(kT_S)e^{-j2\pi\frac{km}{N}} \end{aligned}$$

Den sidste sum i det næstsidste udtryk fås ved at udnytte den **kunstige periode** $P = NT_S$ vi har indført. Funktionen `fft(v)` udregner den viste sum for $m = 0 \dots N - 1$, så udtrykt i MATLAB foregår udregningen ved $V = T_S * \text{fft}(v)$. Den ekstra faktor giver tidsdimensionen, da det er afstanden mellem samplinger af pulsen. Som for det periodiske tilfælde er værdierne kun en fornuftig approksimation for $0 \leq m \leq N/2$. Værdierne fra $N/2$ til $N - 1$ beskriver nemlig de negative frekvenser $-(N/2)/P, \dots, -1/P$ og dem kender man i forvejen når $v(t)$ er et reelt signal, idet $V(-m/P)$ er den kompleks konjugerede af $V(m/P)$. Det totale frekvensområde kan også udtrykkes som fra $-1/(2T_S)$ til $1/(2T_S)$ og **frekvensopløsningen** er $1/P = 1/(NT_S)$. Det er meget vigtigt at bemærke at `fft` i MATLAB kræver at elementet for $t = 0$ står på indeks 1, og at det transformerede resultat har frekvensen 0 i indeks 1 og derefter følger de positive frekvenser op til $N/2$ efterfulgt af de negative (i omvendt orden) sluttende med $m = N - 1$. $N/2$ -frekvensen er lidt speciel ved både at have den negative og positive komponent i sig. `fftshift`-funktionen kan også anvendes her. Eksempel 3.6 i det følgende afsnit viser et eksempel på dette.

Som skrevet ovenfor skal værdien for $t = 0$ stå på indeks 1 i v . Det kræver muligvis lidt omflytning af elementerne i vektoren v , fx hvis den er defineret med en sekvens af værdier for negative t før de positive kommer. Man kan så flytte den kunstige periode således at værdien for $t = 0$ står først og værdierne for de negative t kommer til sidst i vektoren. Et eksempel på dette kan også ses i Eksempel 3.6 i det følgende afsnit. En anden udvidelse af v kan også komme på tale: Da $v(t)$ skal være 0 i enderne af det relevante tidsinterval, kan man frit forlænge vektoren v ved at tilføje en nulsekvens og derved forøge N . Det kan fx være for at få en bedre frekvensopløsning $1/(NT_S)$ eller for at få et N hvor `fft` er mere beregningseffektiv, dvs. hvor N er en potens af 2. Man skal huske på at hvis v er omorganiseret så værdier for negative t står til sidst, skal nulsekvensen indføres i midten.

Også her kan MATLAB-funktion `ifft(V)` anvendes: $v = \text{ifft}(V)/T_S$, som giver samples af $v(t)$ med afstanden T_S :

$$\begin{aligned} v(kT_S) &= \int_{-\infty}^{\infty} V(f) e^{j2\pi f k T_S} df \approx \int_{-1/(2T_S)}^{1/(2T_S)} V(f) e^{j2\pi f k T_S} df \approx \sum_{m=-N/2}^{N/2-1} V\left(\frac{m}{P}\right) e^{j2\pi \frac{m}{P} k \frac{P}{N}} \cdot \frac{1}{P} \\ &= \frac{1}{P} \sum_{m=0}^{N/2-1} V\left(\frac{m}{P}\right) e^{j2\pi \frac{km}{N}} + \frac{1}{P} \sum_{m=-N/2}^{-1} V\left(\frac{m}{P}\right) e^{j2\pi \frac{km}{N}} \\ &= \frac{1}{P} \sum_{m=0}^{N/2-1} V\left(\frac{m}{P}\right) e^{j2\pi \frac{km}{N}} + \frac{1}{P} \sum_{m=N/2}^{N-1} V\left(\frac{m}{P}\right) e^{j2\pi \frac{km}{N}} = \frac{1}{T_S} \frac{1}{N} \sum_{m=0}^{N-1} V\left(\frac{m}{P}\right) e^{j2\pi \frac{km}{N}} \end{aligned}$$

Vi benyttede i udregningen at det totale frekvensområde i approksimationen er fra $-1/(2T_S)$ til $1/(2T_S)$ med opløsningen $1/P = 1/(NT_S)$ og perioden N/P for $V(m/P)$ samt at MATLABs `ifft` udregner summen med $1/N$. Læg mærke til at MATLAB her anvender samme placering af $1/N$ som vi gør - i modsætning til udtrykkene for de periodiske funktioner. Til gengæld skal man huske T_S her.

3.3 Fouriertransformation af tidsdiskret signal

Samplede signaler og en del kunstigt genererede signaler eksisterer kun på adskilte (diskrete) tidspunkter kT_S . Vi har endnu ikke indført frekvensdomænet for disse, men siden det er den mest moderne måde at realisere systemer på, vil vi gøre det i det følgende. Lad de samplede værdier hedde v_k .

3.3.1 Periodisk tidsdiskret signal

Først vil vi se på et **periodisk** tidsdiskret signal som har $v_k = v_{k+N}$. Periodiske signaler har som omtalt en Fourierrække, og den approksimation til Fourierrækken for det tidskontinuerte periodiske signal som vi regnede ud i Afsnit 3.1.2 kan bruges her:

$$V_m = \frac{1}{N} \sum_{k=0}^{N-1} v_k e^{-j2\pi \frac{km}{N}} \quad (3.5)$$

Fourierrækken bliver selvfølgelig også periodisk med den samme periode, N , (der dog nu regnes i frekvenser). Fouriertransformationen (3.5) kaldes normalt den **diskrete Fouriertransformation**, forkortet **DFT**, og som vi så i de foregående afsnit danner den grundlaget for næsten alt hvad der udregnes numerisk i praksis. Der henvises i øvrigt til Afsnit 3.1.2 for nærmere forklaringer på MATLABs `fft` funktion.

Som for den kontinuerte Fouriertransformation giver et reelt periodisk signal $V_m = V_{-m}^* = V_{N-m}^*$ (det sidste på grund af perioden) og en lige funktion (symmetrisk om $k = 0$, dvs.

$v_k = v_{-k} = v_{N-k}$) giver reel V_m . Som tidligere set ændrer en tidsforkydning af v_k spektret. Lad $w_K = v_{k-A}$:

$$W_m = \frac{1}{N} \sum_{k=0}^{N-1} w_k e^{-j2\pi \frac{km}{N}} = e^{-j2\pi \frac{mA}{N}} \sum_{k=0}^{N-1} v_{k-A} e^{-j2\pi \frac{(k-A)m}{N}} = e^{-j2\pi \frac{mA}{N}} V_m$$

3.3.2 Ikke-periodisk tidsdiskret signal

For et **ikke-periodisk signal** defineres den Fouriertransformerede af den tidsdiskrete sekvens $\{v_k\}$ som

$$V(f) = \sum_{k=-\infty}^{\infty} T_S v_k e^{-j2\pi k f T_S} \Leftrightarrow v_k = \int_{-1/(2T_S)}^{1/(2T_S)} V(f) e^{j2\pi k f T_S} df \quad (3.6)$$

hvor vi må antage at den uendelige sum eksisterer. Dette er blandt emnerne for kurset Matematik 2, så vi vil ikke gå nærmere ind på det her. Umiddelbart er der dog ingen problemer for en sekvens der er begrænset i tid, dvs. $v_k = 0$ uden for et eller andet interval.

Spektret $V(f)$ er som tidligere komplekst og det kan også ses at det er **periodisk** med perioden $1/T_S$, da $V(f + 1/T_S) = V(f)$. Man kan se det som en dualitet: ikke-periodiske, tidsdiskrete signaler giver periodiske, frekvenskontinuerte spektre mens periodiske, tidskontinuerte signaler giver frekvensdiskrete spektre (Fourierrækker).

Som for den kontinuerte Fouriertransformation giver et reelt signal $\{v_k\}$ $V(f) = V(-f)^*$ og en lige funktion (symmetrisk om $k = 0$) giver reel $V(f)$. Igen er det vigtigt at tidsforkydning af v_k ændrer spektrets fase. Lad $w_K = v_{k-A}$:

$$W(f) = \sum_{k=-\infty}^{\infty} T_S w_k e^{-j2\pi k f T_S} = e^{-j2\pi f A T_S} \sum_{k=-\infty}^{\infty} T_S v_{k-A} e^{-j2\pi (k-A) f T_S} = e^{-j2\pi f A T_S} V(f)$$

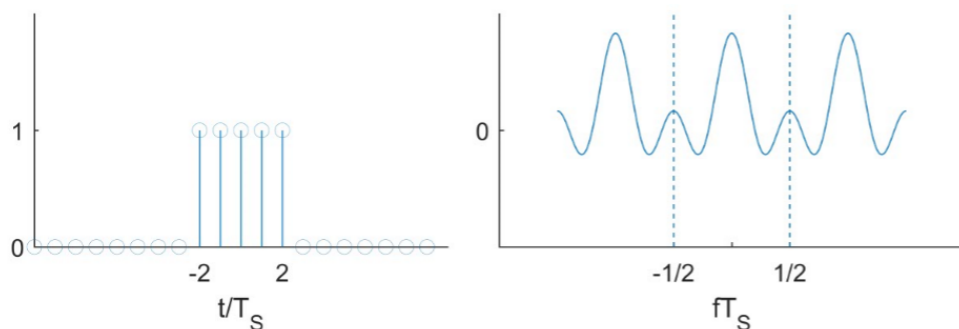
Eksempel 3.5 - Fouriertransformation af firkantpuls i diskret tid

En firkantpuls med bredde $5T_S$ kunne beskrives ved

$$v_k = \begin{cases} 1 & \text{for } |k| \leq 2 \\ 0 & \text{ellers} \end{cases}$$

Da der kun er 5 værdier med $v_k = 1$, er den Fouriertransformerede nem at udregne

$$\begin{aligned} V(f) &= \sum_{k=-\infty}^{\infty} T_S v_k e^{-j2\pi k f T_S} = \sum_{k=-2}^2 T_S e^{-j2\pi k f T_S} \\ &= T_S \left(e^{-j2\pi(-2)fT_S} + e^{-j2\pi(-1)fT_S} + e^{-j2\pi 0 f T_S} + e^{-j2\pi 1 f T_S} + e^{-j2\pi 2 f T_S} \right) \\ &= T_S (1 + 2 \cos 2\pi f T_S + 2 \cos 2\pi f T_S) \end{aligned}$$



Figur 3.5: Firkantpuls (samplet) og dens spektrum.

$V(f)$ bliver reel pga. den lige funktion. Det er også umiddelbart til at se at $V(f)$ er periodisk med perioden $1/T_S$. 3 perioder af spektret er vist i Figur 3.5. Det kan måske være svært at forstå at spektret ikke ligner det i Figur 3.3, men her skal man holde sig for øje at vi har valgt en temmelig grov opløsning for samplingen af firkanten. Da den indeholder 5 punkter er frekvensen $1/B = 1/(5T_S)$ så vi ser kun den midterste del af spektret i Figur 3.3. Med en højere opløsning ville spektret ligne mere i den centrale del, men man skal altid huske at det er periodisk med perioden $1/T_S$ \square .

3.3.3 MATLAB for ikke-periodiske, tidsdiskrete signaler

Ovenstående udregning af spektret var ret simpel, men for mere indviklede signaler må man anvende MATLAB som selvfølgelig er tidsdiskret i sin natur. Situationen er ganske som i Afsnit 3.2.3 hvor vi udregnede spektret for en funktion $v(t)$ der var begrænset til et interval $P = NT_S$ (startende ved $t = 0$ for nemheds skyld). Vi får $T_S = P/N$ og dermed

$$V\left(\frac{m}{NT_S}\right) = \sum_{k=-\infty}^{\infty} T_S v_k e^{-j2\pi \frac{m}{NT_S} T_S} = T_S \sum_{k=0}^{N-1} v_k e^{-j2\pi \frac{km}{N}}$$

Funktionen `fft(v)` udregner den viste sum for $m = 0 \dots N-1$, så udtrykt i MATLAB foregår udregningen ved $V = T_S * \text{fft}(v)$. $V(f)$ er jo en frekvenskontinuert funktion så MATLAB udregner kun nogle samples af spektret. De positive frekvenser findes for $0 \leq m \leq N/2$. Værdierne fra $N/2$ til $N-1$ beskriver nemlig de negative frekvenser $-(N/2)/(NT_S)$, ..., $-1/(NT_S)$ og dem kender man i forvejen når $v(t)$ er et reelt signal, idet $V(-f)$ er den kompleks konjugerede af $V(f)$. Det totale frekvensområde er altså også her fra $-1/(2T_S)$ til $1/(2T_S)$ og frekvensopløsningen er $1/(NT_S)$. Det er som sædvanligt meget vigtigt at bemærke at `fft` i MATLAB kræver at elementet for $t = 0$ står på indeks 1, og at det transformerende resultat har frekvensen 0 i indeks 1 og derefter følger de positive frekvenser op til $N/2$ efterfulgt af de negative (i omvendt orden) sluttende med $m = N-1$. $N/2$ -frekvensen er lidt speciel ved både at have den negative og positive komponent i sig. Transformationen fra frekvensdomænet til tidsdomænet foregår tilsva-

rende ved $v = \text{ifft}(V)/T_s$.

Som i Afsnit 3.2.3 kræves der muligvis lidt omflytning af elementerne i vektoren v ved at udnytte den kunstige periode til at få værdien for $t = 0$ til at stå på indeks 1. Et eksempel på dette kan ses i Eksempel 3.6 nedenfor. Man kan også som beskrevet i Afsnit 3.2.3 forøge N ved at tilføje en nulsekvens. Man skal huske på at hvis v er omorganiseret så værdier for negative t står til sidst, skal nulsekvensen indføres i midten.

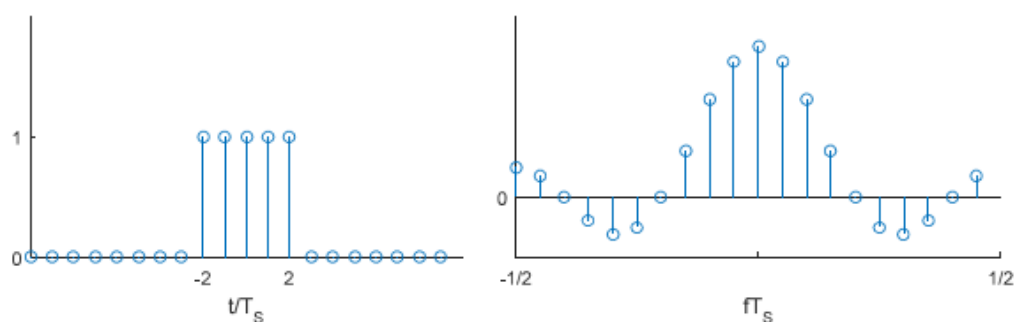
Eksempel 3.6 - Fouriertransformation af firkantpuls i diskret tid beregnet i MATLAB Vi fortsætter det foregående eksempel med en firkantpuls med bredde $5T_s$:

$$v_k = \begin{cases} 1 & \text{for } |k| \leq 2 \\ 0 & \text{ellers} \end{cases}$$

MATLAB skal have tiden 0 på indeks 1, så vi flytter de fem 1-værdier lidt rundt så vi får v_lige nedenfor

```
v_lige=[ones(1,3) zeros(1,15) ones(1,2)];
N=length(v_lige);
V=real(Ts*fft(v_lige));
fm=-N/2:N/2-1;
f=fm*1/(N*Ts);
stem(f,fftshift(V))
```

Det giver de viste 20 værdier af spektret



Figur 3.6: Spektrum for samplet firkantpuls udregnet af MATLAB.

I praktiske anvendelser kender man selvfølgelig en værdi for T_s , fx fås med en samplings-frekvens på 100 Hz, $T_s = 1/100$ s og frekvensområdet går fra -50 Hz til 50 Hz med frekvensopløsning $1/(20T_s) = 5$ Hz. \square

3.4 Oversigt over Fouriertransformationer

De tidligere indførte Fourierrækker og -transformationer samt deres udregninger i MATLAB kan sammenfattes som i nedenstående tabel.

Signaltype	Tid	
	Kontinuert	Diskret
Periodisk (periode $P \sim N$)	Fourierrække $v(t) = \sum_{m=-\infty}^{\infty} V_m e^{j2\pi \frac{m}{P} t}$ $\Leftrightarrow V_m = \frac{1}{P} \int_{-P/2}^{P/2} v(t) e^{-j2\pi \frac{m}{P} t} dt$ MATLAB: <code>fft(v)/N, ifft(V)*N</code>	DFT $v_k = \sum_{m=0}^{N-1} V_m e^{j2\pi \frac{km}{N}}$ $\Leftrightarrow V_m = \frac{1}{N} \sum_{k=0}^{N-1} v_k e^{-j2\pi \frac{km}{N}}$ MATLAB: <code>fft(v)/N, ifft(V)*N</code>
Puls (vilkaarligt signal)	Fouriertransformation $v(t) = \int_{-\infty}^{\infty} V(f) e^{j2\pi ft} df$ $\Leftrightarrow V(f) = \int_{-\infty}^{\infty} v(t) e^{-j2\pi ft} dt$ MATLAB: <code>Ts*fft(v), ifft(V)/Ts</code>	Fouriertransformation $v_k = \int_{-1/(2T_s)}^{1/(2T_s)} V(f) e^{j2\pi f k T_s} df$ $\Leftrightarrow V(f) = T_s \sum_{k=-\infty}^{\infty} v_k e^{-j2\pi f k T_s}$ MATLAB: <code>Ts*fft(v), ifft(V)/Ts</code>

Tabel 3.1: Oversigt over Fouriertransformationer.

3.5 Sampling og aliasering

Som skrevet flere gange ovenfor er spektret af et tidsdiskret signal periodisk. Det gælder altså også, hvis det tidsdiskrete signal er et samlet analogt signal, hvor vi får en sekvens af værdier:

$$\tilde{v}_k = v(kT_s), \quad \text{for } -\infty < k < \infty$$

som har et spektrum med en periode lig samplingfrekvensen $f_s = 1/T_s$:

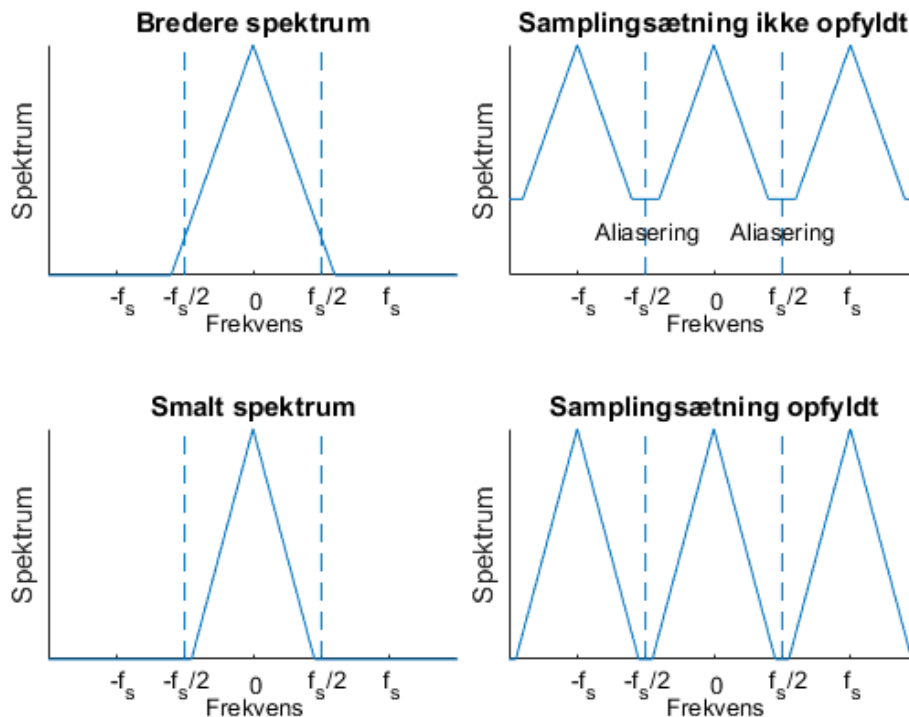
$$\tilde{V}(f) = T_s \sum_{k=-\infty}^{\infty} \tilde{v}_k e^{-j2\pi f k T_s}$$

Man kan vise at dette spektrum kan findes ud fra spektret $V(f)$ for det tidskontinuerte signal $v(t)$ som

$$\tilde{V}(f) = \sum_{m=-\infty}^{\infty} V(f + m f_s) = \sum_{m=-\infty}^{\infty} V\left(f + \frac{m}{T_s}\right)$$

altså en uendelig sum af forskudte kopier af $V(f)$ med afstand f_s . Dette er illustreret på Figur 3.7 i to situationer, nemlig med en for lille samplingsfrekvens (øverst) og med en samplingsfrekvens der opfylder samplingssætningen (nederst).

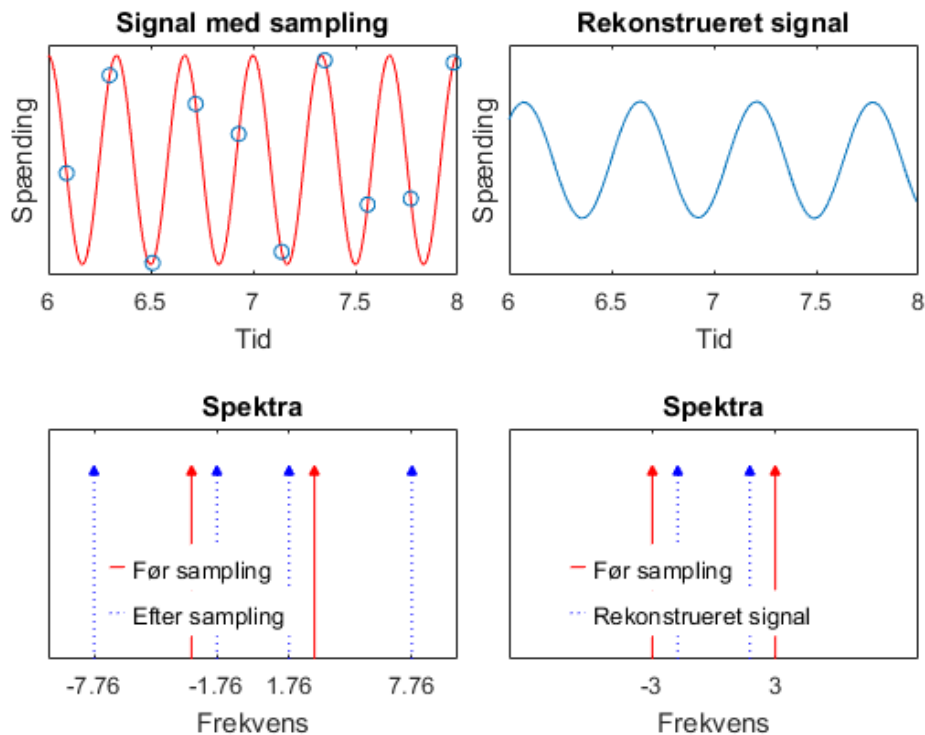
Samplingssætningen i Kapitel 2 kan nu også gives en mere håndfast fortolkning: De funktioner der anvendes til rekonstruktion af det analoge signal har et spektrum begrænset af



Figur 3.7: Spektre af originale og samplede signaler (henholdsvis til venstre og højre) til illustration af samplingssætningen.

frekvensen $1/(2T_S) = f_S/2$, dvs. de tillader frekvenser op til $f_S/2$ at “slippe igennem” og derfor skal samplingsfrekvensen f_S være mindst dobbelt så høj som den højeste frekvens i signalet. Hvad vi helt nøjagtigt mener med at “slippe igennem” vil fremgå af Kapitel 4.

I den øverste del af figuren er samplingsfrekvensen ikke tilstrækkelig høj til at hele signalets spektrum slipper igennem. Her får vi forklaringen på at en for lille samplingsfrekvens frembringer nye frekvenser, **aliasering**. Hvis spektret har frekvenser højere end $f_S/2$ vil det blive overlappet af sig selv forskudt f_S (et “alias” for spektret) og når rekonstruktionen slipper alt med frekvens mindre end $f_S/2$ igennem, vil der også komme noget fra de overlappende aliaser (deraf navnet aliasering) som vist øverst på figuren. Omvendt kan man også se at hvis samplingssætningen er opfyldt, vil rekonstruktionen præcis slippe spektret for signalet igennem og dermed bliver signalet rekonstrueret perfekt. Dette er illustreret nederst på Figur 3.7. Dette kan ses som et bevis på samplingssætningen. Vi illustrerer dette yderligere i Figur 3.8, hvor et 3 Hz sinussignal samples med $f_S = 4,76$ Hz hvor over $2 \cdot 3$ Hz er nødvendigt.. Man ser at det rekonstruerede signal, som er det samplede spektrum inden for frekvensområdet $\pm f_S/2 = \pm 2,38$ Hz, har frekvensen 1,76 Hz.



Figur 3.8: Sampling af 3 Hz signal med samplingsfrekvens 4,76 Hz.

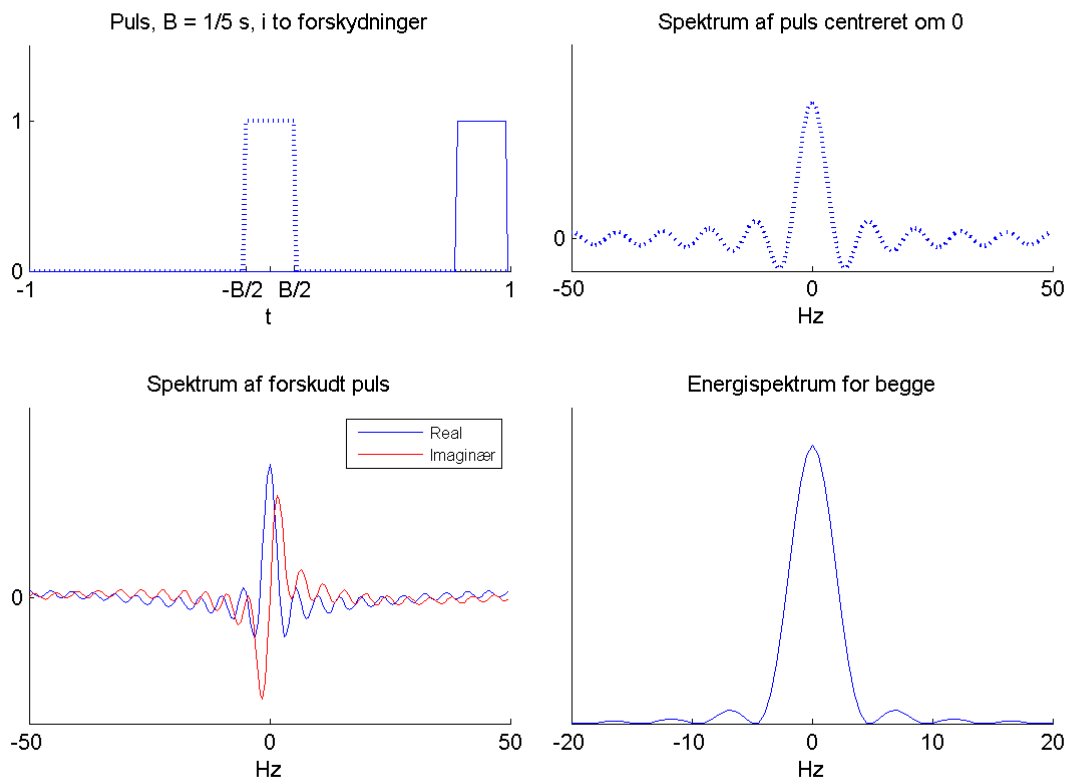
3.6 Energi- og effektspektre

Oftentimes one is interested in a measure for the content of the different frequencies in a signal without worrying about their phases. One can achieve this by simply looking at the numerical values - the amplitude spectrum - but as a rule, the squared values are used, as these have a physical interpretation as energy and effect. We will get closer to the precise relation between energy and effect in Chapter 4, but for signal analysis here it is sufficient to work with the squared numerical values. One of the pitfalls of this is that one does not need to take into account the precise time relations, for example, a signal $[v \ 0 \ 0 \ 0 \ 0]$ will have the same energy spectrum as $[0 \ 0 \ 0 \ 0 \ v]$, where v is any sequence. This illustrates what we see in Figure 3.9. The reason for this is obviously the earlier described relations between the spectrum for a signal and its time shift, here in continuous time, $w(t) = v(t - A)$:

$$S_W(f) = |W(f)|^2 = |e^{-j2\pi f A} V(f)|^2 = |e^{-j2\pi f A}|^2 |V(f)|^2 = S_V(f)$$

Periodiske signaler har et **effektspektrum** og **ikke-periodiske signaler** (såvel tidskontinuerte som tidsdiskrete) har et **energispektrum**. Med betegnelserne fra (3.3) - (3.6) kan de regnes ud som vist i Tabel 3.2.

Effect or energy for a signal can be calculated in the time domain, for example, the effect for a time-discrete



Figur 3.9: Spektre og energispektre for forskudte pulser.

periodisk signal v :

$$P_V = \frac{1}{N} \sum_{k=0}^{N-1} v_k^2$$

og man kan vise at denne effekt netop er summen af effekterne ved de enkelte frekvenser

$$P_V = \sum_{m=0}^{N-1} S_V \left(\frac{m}{NT_S} \right)$$

Dette er begrundelsen for vores placering af $1/N$ i (3.3) og (3.5).

Vi vil ikke komme nærmere ind på **stokastiske signaler**, men disse kan også tillægges et (frekvenskontinuert) effektspektrum og det er faktisk det eneste relevante begreb for sådanne signaler, da man ikke kan tale om disses faser eller fasespektre. Her kommer en anden fidus, nemlig at effekter for uafhængige signaler er bevarede ved addition, dvs. effekten for summen af to uafhængige signaler er blot summen af effekterne uden at man behøver at tage hensyn til tidslige relationer. Effektspektre for tidsdiskrete stokastiske signaler har også aliaser med afstand $1/T_S$.

Signaltype	Tid	
	Kontinuert	Diskret
Periodisk (periode $P = NT_S$)	Effektspektrum	Effektspektrum
	$S_V\left(\frac{m}{P}\right) = V_m ^2 \quad \text{for } -\infty < m < \infty$ for frekvenserne m/P	$S_V\left(\frac{m}{NT_S}\right) = V_m ^2 \quad \text{for } 0 \leq m \leq N-1$ for frekvenserne $m/NT_S = m/P$
Puls (vilkrårligt signal)	Energispektrum	Energispektrum
	$S_V(f) = V(f) ^2 \quad \text{for } -\infty < f < \infty$ for frekvensen f	$S_V(f) = V(f) ^2 \quad \text{for } -\frac{1}{2T_S} \leq f \leq \frac{1}{2T_S}$ for frekvensen f

Tabel 3.2: Oversigt over energi- og effektspektre (betegnelser fra (3.3) - (3.6)).

KAPITEL 4

Elektriske kredsløb og lineære systemer

Når signaler skal frembringes, transmitteres og behandles er det nødvendigt med nogle fysiske komponenter. I mere end 150 år har man benyttet elektriske signaler og behandlet disse ved forskellige elektriske kredsløb. Ved transmission af optiske signaler benyttes en del optiske komponenter, men også her bliver signalerne repræsenterede ved elektriske størrelser et eller flere steder under transmission fra ét punkt til et andet. Det er derfor nødvendigt at etablere et nødtørftigt grundlag for forståelse af elektriske kredsløb. Disse anvendes både som forskellige filtre - vi har omtalt nogle af dem i forbindelse med sampling og rekonstruktion, men der findes mange andre - og som selve transmissionsmediet - tænk på et kabel med metalliske ledere som fx en telefonforbindelse eller tilslutningsskabler til et datanet. Sådanne transmissionsledninger med metalliske ledere behandles senere (Afsnit 7.1) mens vi i dette kapitel ser lidt på den generelle teori og praksis for elektriske kredsløb for tidskontinuerte signaler.

Udover de analoge elektriske kredsløb, der behandler tidskontinuerte signaler, er der også i udstrakt grad behandling af tidsdiskrete signaler i digitale kredsløb som godt nok er elektroniske men kun foretager behandling og lagring til bestemte tidspunkter. Disse kan behandles uden at tage den fysiske realisering i betragtning som det kan ses i Afsnit 4.6.

Fælles for de tidskontinuerte og de tidsdiskrete behandlingskredsløb, vi betragter, er at de er lineære, så vi starter med en kort introduktion til lineære systemer.

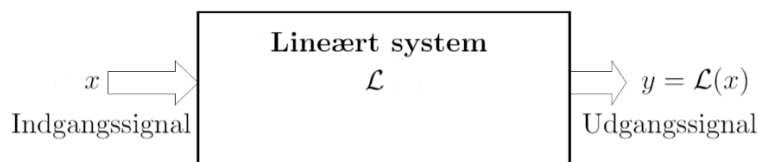
4.1 Lineære systemer

Ordet system er meget anvendt om alt muligt, og der skal da nok være nogle mere eller mindre præcise anvendelser i dette kursus, men netop i signalanalysen er et system ret veldefineret. Vi viser et **system** i Figur 4.1, hvor man kan se et **indgangssignal** $x(t)$ og et **udgangssignal** $y(t) = \mathcal{L}(x)$. Det er meget velovervejet at der ikke står $x(t)$ her, for udgangssignalet på et givet tidspunkt afhænger af hele signalet, ikke blot værdien til tiden t . Hvis udgangssignalet kun afhænger af x 's foregående værdier, dvs. $x(\tau)$ for $\tau \leq t$, kaldes systemet **kausalt**. Fysiske systemer er selvfølgelig kausale. Den vigtigste klasse af

systemer er **lineære systemer**, der har følgende egenskab for to signaler x_1 og x_2 :

$$\begin{aligned} y_1(t) &= \mathcal{L}(x_1) \text{ og } y_2(t) = \mathcal{L}(x_2) \\ \Rightarrow \mathcal{L}(a_1x_1 + a_2x_2) &= \mathcal{L}(a_1x_1) + \mathcal{L}(a_2x_2) = a_1y_1(t) + a_2y_2(t) \end{aligned}$$

hvis a_1 og a_2 er vilkårlige konstanter. For tidsdiskrete signaler findes $x(t)$ kun til diskrete tidspunkter, men ellers er definitionen den samme. I elektroteknikken er signalerne



Figur 4.1: Lineært system med ind- og udgangssignal.

strømme og spændinger og dem ser vi på i de følgende afsnit, men først lidt mere generel teori om lineære systemer.

Et lineært system er **tidsinvariant** hvis det opfylder

$$y(t - \Delta) = \mathcal{L}(x(t - \Delta)) \quad \text{for alle } t \text{ og } \Delta$$

dvs. systemet har altid samme virkning på indgangssignalet uafhængigt af om dette forskydes Δ .

Man kan vise at udgangssignalet for ethvert lineært tidsinvariant system i kontinuert tid kan findes som

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau \quad (4.1)$$

hvor $h(\tau)$ kaldes systemets **impulsrespons**.

Impulsresponsen karakteriserer altså det lineære system fuldstændigt. (Hvis systemet var tidsvarierende skulle h også afhænge af t). Et kausalt system har selvfølgelig impulsrespons 0 for $\tau < 0$, så kun x 's forhistorie kommer til at indgå i $y(t)$. Den matematiske operation mellem $h(t)$ og $x(t)$ kaldes en **foldning** og skrives $y = h * x$, hvor $*$ betyder foldning og ikke multiplikation som fx i MATLAB.

Systemerne behøver ikke at blive ret indviklede før det er ganske svært at finde resultatet af virkningen på et givet indgangssignal i tidsdomænet. Den helt store fidus ved at kunne

betragte signaler både i tids- og frekvensdomænet viser sig ved lineære systemer. I stedet for det væmmelige integral i tidsdomænet får man nemlig følgende sammenhæng:

$$Y(f) = H(f) \cdot X(f),$$

(4.2)

hvor $Y(f)$, $H(f)$ og $X(f)$ er Fouriertransformerede
af $y(t)$, $h(t)$ og $x(t)$

$H(f)$ kaldes systemets **overføringsfunktion** og den er altså den Fouriertransformerede af impulsresponsen $h(t)$. (Igen er $H(f)$ ikke afhængig af tid da systemet er tidsinvariant).

$H(f)$ er en kompleks funktion. $|H(f)|$ angiver ændringen i amplitude for et harmonisk signal med frekvensen f og $\angle H(f)$ giver faseforskydningen. Almindeligvis varierer overføringsfunktionen med frekvensen, så udgangssignalet bliver **forvrænget** (ændret) i forhold til indgangssignalet, da nogle frekvenser overføres bedre end andre. Hvis $H(f) = Ae^{-j2\pi f\tau}$, A og τ konstanter inden for det frekvensområde indgangssignalet optager (dvs. hvor $X(f) \neq 0$), overføres signalet forvrængningsfrit, så $y(t)$ er præcis som $x(t)$ i form. Det kan være skaleret lidt (svarende til A) og forsinket lidt (svarende til τ jf. forskydelses-sammenhæng mellem spektre fra Afsnit 3.2.1).

4.2 Elektriske kredsløb

Elektriske kredsløb anvendes i realiseringen af de fleste kommunikationssystemer så en kort introduktion er nødvendig. Grundlæggende består elektriske kredsløb af en samling komponenter forbundet med ledninger uden at vi i øvrigt vil komme nærmere ind på disses udførelse. Komponenterne beskrives nedenfor i Afsnit 4.3. Det grundlæggende fysiske fænomen er **elektrisk strøm** der består i transport af **elektrisk ladning** som det forhåbentlig er kendt fra skoleundervisningen. Enheden for strøm er A (ampere) der dog er en ret stor størrelse for kommunikationsteknologien, så vi taler som regel om mA eller μA . Strøm er ladningstransport pr. tidsenhed, dvs. enheden for ladning er $As = C$ (coulomb). En tredje vigtig størrelse er **elektrisk spænding** der er evnen til at udføre noget arbejde: Når to punkter har en spændingsforskel på 1 V (volt) betyder det at der for at føre en ladning på 1 C fra det ene punkt til det andet skal udføres et arbejde på 1 J (joule). Spændinger (også kaldet potentialer) ses at være ikke-absolutte størrelser, kun spændingsforskelle er af betydning. Ofte vedtager vi et nulniveau som vi som regel kalder jordpotential og regner spændinger ud fra dette.

Definitionen af spænding betyder at spænding (målt i V) gange ladning (målt i C) er udtryk for arbejde (målt i J) og da arbejde (energi) pr. tidsenhed i fysikken kaldes effekt,

fås at den **elektriske effekt** (P for *power*, måles i W (watt)) er

$$\begin{aligned} P [\text{watt}] &= \frac{\text{arbejde}}{\text{tid}} = \frac{\text{spænding} \cdot \text{ladning}}{\text{tid}} = \text{spænding}(v) \cdot \frac{\text{ladning}}{\text{tid}} \\ &= v [\text{volt}] \cdot \text{strøm}(i [\text{ampere}]) \end{aligned} \quad (4.3)$$

Nedenfor vil det fremgå af spænding og strøm er proportionale i de almindelige komponenter, fx siger Ohms lov at $v = R \cdot i$ for en modstand. Man får da at effekten er

$$P = \frac{v^2}{R}$$

dvs. proportional med spændingen kvadreret. I Afsnit 3.6 definerede vi effektspektrum blot som kvadratet på den Fouriertransformerede af signalet. Vi tog derved ikke hensyn til modstanden R . Det er ret almindeligt at gøre således i signalanalyse, men man skal selvfølgelig tænke sig om hvis man skal udtrykke noget om den fysiske effekt.

Man er ofte ude for at skulle tale om forskellige størrelser eller forhold der er proportionale med effekten og da proportionalitetskonstanterne ikke altid er kendte regner man på spændinger kvadrerede. Et eksempel er forholdet mellem to effekter (P_1 og P_2) frembragt af to spændinger v_1 og v_2 (over samme modstand) som ofte i kommunikationsteknologien udregnes logaritmisk (udtrykkes i de såkaldte **dB** (deciBel)):

$$10 \log \frac{P_1}{P_2} [\text{dB}] = 10 \log \frac{v_1^2}{v_2^2} [\text{dB}] = 20 \log \frac{v_1}{v_2} [\text{dB}] \quad (4.4)$$

Årsagen til anvendelse af logaritmiske udtryk for forhold er at disse ellers kan blive ganske store eller meget små, fx kan et effektforhold på 1:1000 mere overskueligt udtrykkes som -30 dB. Det er klogt kun at tale om effektforhold og glemme spændingsudtrykket i ovenstående formel da det passer bedst til kommunikationssignaler.

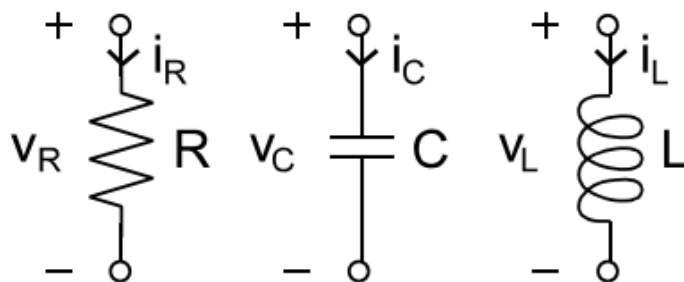
Et andet sted, hvor kvadrerede spændinger kommer ind, er ved tidsvarierende spændinger. Man kan spørge sig om hvilken konstant spænding, V_{eff} , der vil give samme effekt som en tidsvarierende $v(t) = V_p \cdot \cos(2\pi ft)$. Det kan nemt findes ved at se på én periode i spændingen

$$V_{\text{eff}}^2 = \int_0^{1/f} (V_p \cdot \cos 2\pi ft)^2 dt = \frac{V_p^2}{2} \Rightarrow V_{\text{eff}} = \frac{V_p}{\sqrt{2}}$$

Vi kommer mere ind på effektforhold for tidsvarierende spændinger i Afsnit 4.4.2. Vi har afvejet en smule fra den almindelige konvention i disse noter om at variable i frekvensdomænet er med store bogstaver og små bogstaver anvendes i tidsdomænet. V_p har dog en relation til Fourierkoefficienten ved frekvensen f som vi skal komme nærmere ind på i Afsnit 4.4.

4.3 Almindelige komponenter

Elektriske kredsløb indeholder dels passive komponenter som modstande (også kaldet resistor), kondensatorer (også kaldet kapacitor) eller selvinduktioner (også kaldet induktor) forbundet med ledninger og dels aktive komponenter som transistorer eller forstærkere. Alle disse er som beskrevet nedenfor ved normal anvendelse lineære eller tilstræbt lineære, så det er en meget almindelig egenskab. Ved meget store eller meget små signalværdier ophører de aktive komponenter med at være lineære og passive komponenter kan også drives for langt ud - fx kan en modstand gå op i røg. Ved normale signaler kan der for aktive komponenter være mindre afvigelser fra linearitet og det kan være en vigtig kvalitetsparameter for fx en HiFi-forstærker hvor små disse afvigelser kan gøres.



Figur 4.2: De tre almindeligste typer af passive komponenter.

Figur 4.2 viser de tre almindeligste type af passive komponenter. Disse komponenter er alle karakteriserede ved en værdi: Modstand - resistans R målt i Ohm (Ω), kondensator - kapacitans C målt i Farad (F) og selvinduktion - induktans L målt i Henry (H). Værdierne af grundenhederne er ikke umiddelbart passende for almindelige elektriske kredsløb, så oftest anvendes de med præfiks, f.eks. $k\Omega$, $M\Omega$, nF, pF og mH. Disse værdier indgår i den komponentligning der beskriver sammenhængen mellem strøm og spænding for hver type komponent:

- Fra tidligere undervisning kendes nok **Ohms lov**, der siger at spændingen over R er $v_R(t) = R \cdot i_R(t)$.
- For kondensatoren er det lidt mere indviklet: Strømmen $i_C(t) = C \cdot dv_C/dt$, altså strømmen er afhængig af v_C 's differentialkoefficient.
- For selvinduktionen er det lige så indviklet: Spændingen $v_L(t) = L \cdot di_L/dt$, altså spændingen er afhængig af i_L 's differentialkoefficient.

Det ses at alle disse komponentligninger er lineære. I Afsnit 4.4 og 4.5 ser vi på hvordan man kan regne på sammensatte kredsløb med passive komponenter.

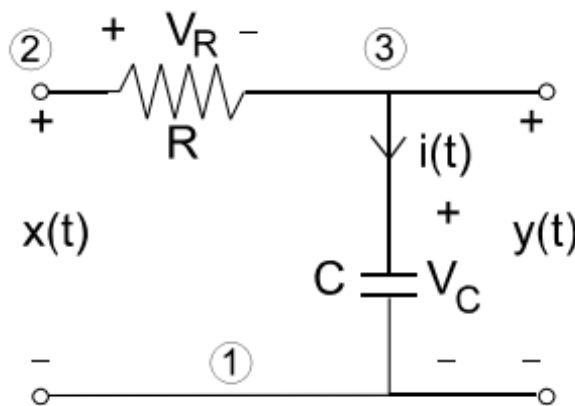
Aktive komponenter er noget mere indviklede, men grundlæggende anvendes de som en eller anden form for forstærkning. Uden at gå for meget i detaljer er den simpleste aktive komponent en **transistor**. Denne har tre tilslutninger, kaldet basis, emitter og kollektor. Forstærkningen fremkommer ved at strømmen gennem kollektor-emitter styres af strømmen til basis og typisk er kollektor-emitterstrømmen mange gange større - transistorens strømforstærkning. Der er ikke helt proportionalitet som for de passive komponenter, men for visse strømintervaller er der en fin proportionalitet. **Forstærkere** laves ved at sammensætte transistorer og passive komponenter, men sådanne sammenkoblinger vil det føre for vidt at komme ind på. Ved sammensætningen kan man fx opnå at udgangsspændingen fra forstærkeren bliver proportional med indgangsspændingen over et stort interval (udstyringsområdet).

4.4 Overføringsfunktion og impulsrespons for elektriske kredsløb

Vi vil nu finde impulsresponsen for et simpelt kredsløb og da det viser sig svært vil vi se hvordan vi kan finde overføringsfunktioner for kredsløb.

Eksempel 4.1 - Impulsrespons for simpelt kredsløb

I Figur 4.3 viser vi en fysisk realisering af et meget simpelt system med en modstand R og en kondensator C . Ind- og udgangssignalerne $x(t)$ og $y(t)$ er spændinger.



Figur 4.3: Simpelt lineært system (Lavpas-filter).

Komponentligningerne giver os $v_R(t) = R \cdot i(t)$ og $i(t) = C \cdot dv_C/dt$. Man kan regne på sammensætningen af de viste komponenter ved at bemærke at $x(t) = v_R(t) + v_C(t) = v_R(t) + y(t)$, idet $v_C(t) = y(t)$ og spændinger kan lægges sammen. For et vilkårligt indgangssignal $x(t)$ får man altså

$$x(t) = v_R(t) + y(t) = R \cdot i(t) + y(t) = RC \frac{dy}{dt} + y(t)$$

Denne inhomogene differentialligning er kendt fra Matematik 1, [eNote 11] og den kan løses med den såkaldte Panserformel:

$$y(t) = e^{-\frac{t}{RC}} \int e^{\frac{z}{RC}} \left(\frac{1}{RC} x(z) \right) dz = e^{-\frac{t}{RC}} \int \frac{1}{RC} e^{\frac{t-\tau}{RC}} x(t-\tau) d\tau = \int \frac{1}{RC} e^{\frac{-\tau}{RC}} x(t-\tau) d\tau$$

som når det sammenlignes med (4.1) viser at systemet har impulsresponsen $h(\tau) = e^{-\tau/RC}/RC$ for $\tau \geq 0$. \square

Når kredsløbet bliver mere indviklet vil opstilling og løsning af differentialligninger blive meget kompliceret, så derfor vil vi i stedet forsøge at finde overføringsfunktionen, $H(f)$, i frekvensdomænet. Som beskrevet i Kapitel 3 giver den Fouriertransformerede $V(f)$ bidraget $V(f)e^{j2\pi ft}$ til $v(t)$. Normalt kalder vi den komplekse størrelse $V(f)$ for den **komplekse repræsentant** for tidsfunktionen $v_f(t) = Re(V(f)e^{j2\pi ft}) = |V(f)| \cos(2\pi ft + \angle V(f))$ som er en harmonisk funktion med amplitude $|V(f)|$ og faseforskydning med vinklen $\angle V(f)$. Som tidligere omtalt kan bidraget $V(f)e^{j2\pi ft}$ beskrives geometrisk som en roterende vektor (engelsk *phasor*), men det tilfører ikke rigtigt noget her. For at finde overføringsfunktionen må vi først se på de enkelte komponenters virkning på et sådant spændingssignal og det tilsvarende strømsignal. Hvis vi refererer til betegnelserne i Figur 4.2 - dog med store bogstaver for strømme og spændinger der nu er Fouriertransformerede - får vi:

- For en resistans gælder $V_R(f)e^{j2\pi ft} = RI_R(f)e^{j2\pi ft}$, så efter bortforkortning af tidsfunktionen får vi $V_R(f) = RI_R(f)$, dvs. Ohms lov gælder også i frekvensdomænet.
- For en kapacitans har vi $I_C(f)e^{j2\pi ft} = C \cdot d(V_C(f)e^{j2\pi ft})/dt = j2\pi fC \cdot V_C(f)e^{j2\pi ft}$ dvs. $I_C(f) = j2\pi fC \cdot V_C(f)$ eller $V_C(f) = (1/(j2\pi fC)) \cdot I_C(f)$. Det er ganske som Ohms lov, men nu med en kompleks konstant foran strømmen.
- For en induktans har vi $V_L(f)e^{j2\pi ft} = L \cdot d(I_L(f)e^{j2\pi ft})/dt = j2\pi fL \cdot I_L(f)e^{j2\pi ft}$ dvs. $V_L(f) = j2\pi fL \cdot I_L(f)$. Det er også ganske som Ohms lov med en kompleks konstant foran strømmen.

Man kan således regne på de komplekse repræsentanter helt uden tidsfunktionen og man får erstattet differentiering med en simpel multiplikation med $j2\pi f$. Proportionalitetskonstanterne i ovenstående ligninger kaldes under ét **impedanser**. Dette er en generalisering af resistansbegrebet fra **Ohms lov**, $R = V/I$. Af ovenstående komponentligninger får vi impedanser for de tre komponenttyper

$$\begin{aligned} \text{Resistans : } Z_R &= \frac{V_R(f)}{I_R(f)} = R \\ \text{Kapacitans : } Z_C &= \frac{V_C(f)}{I_C(f)} = \frac{1}{j2\pi fC} \\ \text{Induktans : } Z_L &= \frac{V_L(f)}{I_L(f)} = j2\pi fL \end{aligned}$$

Eksempel 4.2 - Overføringsfunktion for simpelt kredsløb (lavpasfilter)

Vi kan nu nemt finde overføringsfunktionen for systemet i Figur 4.3:

$$\begin{aligned} X(f) &= V_R(f) + Y(f) = R \cdot I(f) + V_C(f) = R \cdot j2\pi fC \cdot V_C(f) + V_C(f) \\ &= (j2\pi fRC + 1) V_C(f) = (j2\pi fRC + 1) Y(f) \\ \Rightarrow H(f) &= \frac{Y(f)}{X(f)} = \frac{1}{j2\pi fRC + 1} \end{aligned}$$

Af udtrykket for $H(f)$ kan vi nu se hvorfor vi kalder systemet for et **lavpasfilter**: Det tillader lave frekvenser at passere ($H(f) \approx 1$ for $f \ll 1/(2\pi RC)$) mens høje frekvenser spærres ($H(f) \approx 0$ for $f \gg 1/(2\pi RC)$). Som det kan ses af den tilknyttede øvelse, er det viste filter et noget “blødt” lavpasfilter således at forstå at overgangen mellem gennemgang og spærring er meget blød. \square

4.4.1 Lavpasfiltre og sampling

Man kan nemt frembringe bedre lavpasfiltre end i eksemplet, men ikke helt komme til det **ideelle lavpasfilter** som har $H(f) = 1$ for $f \leq B$ og $H(f) = 0$ for $f > B$. Som vi så i Figur 3.3 vil den tilhørende impulsrespons være en sinc-funktion som strækker sig til $\pm\infty$ og aftager meget langsomt. Man kan derfor ikke lave en sådan hvis man ikke vil vente uendeligt længe på udgangssignalet. Dette emne var vi inde på allerede i forbindelse med samplingsætningen i Kapitel 2, hvor rekonstruktionen var en foldning af de samplede værdier med en sådan funktion, dvs. de samplede værdier kan ses som indgangssignal til et lavpasfilter der netop folder med den rigtige funktion.

Med vor nuværende viden kan **samplingsætningen** altså formuleres:

Hvis signalet $v(t)$ er båndbegrænset således at højeste frekvens er $F < 1/(2T_S)$ kan signalet rekonstrueres fuldstændigt ud fra de samplede værdier, $v(kT_S)$, ved at sende disse gennem et ideelt lavpasfilter med afskæringsfrekvens $1/(2T_S)$.

Lavpasfilteret tillader simpelthen signalets frekvenser at slippe igennem og spærrer for højere frekvenser. Disse sidste er et resultat af samplingprocessen som beskrevet i Afsnit 3.5 og de fjernes altså af filteret. Aliasering opstår når disse frekvenser fra det gentagne spektrum også slipper i gennem og det sker når $F > 1/(2T_S)$. Dette blev illustreret i Figur 3.7 og 3.8. Som tidligere omtalt anbringer man ofte et lavpasfilter, antialiaseringsfilter, før samplingen for at begrænse frekvensområdet så meget at aliasering undgås.

4.4.2 Effekt i elektriske kredsløb

Vi vil nu studere effektforhold i tidsvarierende spændinger og tilhørende strømme. Dette har selvfølgelig voldsom interesse i elforsyningen, men også i kommunikationsanvendelser er det vigtigt at se på effekten. Vi betragter først den øjeblikkelige effekt som også er

tidsvarierende. Dernæst ser vi på middeleffekt (gennemsnitseffekt) som vi startede lidt på at diskutere i Afsnit 4.2. For spændingen $v(t) = V \cdot \cos(2\pi ft + \theta)$ med frekvensen f og den tilhørende strøm $i(t) = I \cdot \cos(2\pi ft + \phi)$ bliver den øjeblikkelige effekt

$$\begin{aligned} P(t) &= v(t) \cdot i(t) = V \cdot \cos(2\pi ft + \theta) \cdot I \cdot \cos(2\pi ft + \phi) \\ &= \frac{VI}{2} (\cos(2\pi \cdot 2f \cdot t + \theta + \phi) + \cos(\theta - \phi)) \end{aligned}$$

hvor θ og ϕ er faserne for henholdsvis strøm og spænding og V og I er disses spidsværdier. Denne effekt ses at variere kraftigt hen over en periode og det eneste der rigtigt interesserer os er middeleffekten over en periode, for det er den der fx afleveres som varme i en modstand:

$$\begin{aligned} P &= \int_0^{1/f} v(t) \cdot i(t) dt = \frac{VI}{2} \left(\int_0^{1/f} \cos(2\pi \cdot 2f \cdot t + \theta + \phi) dt + \int_0^{1/f} \cos(\theta - \phi) dt \right) \\ &= \frac{VI}{2} \cdot 0 + \frac{VI}{2} \cdot \cos(\theta - \phi) \end{aligned}$$

Læg mærke til at vi nu skriver effekten uden afhængighed af tiden. Denne effekt kaldes **realeffekten** eller aktiv effekt, da den som omtalt viser den effekt der afsættes i fx en modstand. Udtrykt ved de såkaldte **effektivværdier** af spænding og strøm, V_{eff} og I_{eff} har man altså

$$P = \frac{VI}{2} \cdot \cos(\theta - \phi) = \frac{V}{\sqrt{2}} \cdot \frac{I}{\sqrt{2}} \cdot \cos(\theta - \phi) = V_{\text{eff}} \cdot I_{\text{eff}} \cdot \cos(\theta - \phi)$$

som allerede omtalt i Afsnit 4.2. Faktoren $\cos(\theta - \phi)$ kaldes effektfaktoren og udtrykker at man i reglen får mindre effekt end produktet af effektiv spænding og strøm. For en modstand er $v(t) = Ri(t)$ så her er $\theta - \phi = 0$ så man får det hele afsat i modstanden. Det er normalt V_{eff} man taler om i forbindelse med tidsvarierende spændinger, fx er $V_{\text{eff}} = 230 \text{ V}$ i vore almindelige stikkontakter, så spidsværdien er $V = 230 \cdot \sqrt{2} \text{ V} = 325 \text{ V}$.

Vi vil nu gerne udtrykke effekten ud fra de komplekse repræsentanter $V(f) = |V(f)|e^{j\angle V(f)} = V e^{j\theta}$ og $I(f) = I e^{j\phi}$. Da middelværdier over en periode $(1/f)$ også er nul for disse, får man

$$\begin{aligned} P &= \int_0^{1/f} \text{Re}(|V(f)|e^{j(2\pi ft + \theta)}) \cdot \text{Re}(|I(f)|e^{j(2\pi ft + \phi)}) dt \\ &= \frac{|V(f)I(f)|}{2} \text{Re} \left(\int_0^{1/f} e^{j(2\pi \cdot 2f \cdot t + \theta + \phi)} dt + \int_0^{1/f} e^{j(\theta - \phi)} dt \right) \\ &= \frac{|V(f)I(f)|}{2} \cdot \text{Re}(e^{j(\theta - \phi)}) = \frac{1}{2} \text{Re}(V(f)I^*(f)) \end{aligned}$$

Størrelsen $P_c = 1/2 V(f) I^*(f)$ kaldes den **komplekse effekt**, dens størrelse $|P_c| = 1/2 |V(f)| |I(f)|$ kaldes den **tilsyneladende effekt** eller **skineffekt**. Da man sjældent har styr på de præcise faseforhold i transmissionssystemer, er det ofte skineffekten man betragter og fx udtaler sig om hvor meget den dæmpes (og måler det i dB, se Afsnit 4.2). Imaginærdelen af den komplekse effekt kaldes den **reaktive effekt** (i modsætning til den aktive effekt som er realdelen). Den er ikke af så stor interesse i kommunikationsteknologien, men i elforsyningen er man meget interesseret i denne. Dette skyldes at selv om den reale effekt måske ikke er så stor for en given belastning skal der stadig leveres store strømme hvis der er en stor reaktiv effekt og systemet skal dimensioneres efter dette - man kunne også sige at det var skineffekten man brugte ved dimensionering.

4.5 Introduktion til kredsløbsteori

Kredsløbet i eksempel 4.1 og 4.2 ovenfor var ret simpelt og vi kunne nemt overskue spændinger og strømme. For mere indviklede kredsløb må man gå lidt mere systematisk frem. Et **generelt kredsløb** består af knudepunkter forbundet af komponenter (inklusive påtrykte spændinger). For en systematisk behandling tildeles hvert knudepunkt en spænding, V_i for knudepunkt i , og en del af disse er ubekendte som kan findes ved at løse et lineært ligningssystem som det demonstreres i eksemplet nedenfor. Til at opstille en ligning for hvert knudepunkt haves den såkaldte **Kirchhoffs strømlov**: Summen af alle strømme regnet med fortegn ind mod et knudepunkt er 0. Dette er en ret trivial lov, der bare siger at der ikke forsvinder nogen ladning i kredsløbet (strøm er transport af ladning). Vi demonstrerer dette i eksemplet.

Eksempel 4.3 - Systematisk beregning af overføringsfunktion

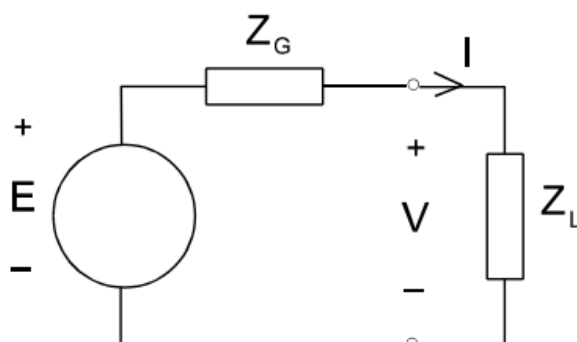
I Figur 4.3 er der 3 knudepunkter forbundet af den påtrykte indgangsspænding $x(t)$, af R og af C . Knudepunkterne er vist på tegningen ved numre i cirkler. Vi brugte uformelt Kirchhoffs strømlov i eksempel 4.1 ved at konstatere at strømmen gennem R og C er ens. Antallet af ubekendte reduceres ved at tildele ét knudepunkt spændingen 0 og indsætte de påtrykte spændinger direkte. For eksemplet kan vi altså sætte $V_1 = 0$ og $V_2 = V_1 + X(f) = X(f)$. Der er så kun én ubekendt tilbage, V_3 og anvendes Kirchhoffs strømlov for dette knudepunkt tillige med de sammenhænge, ligningerne for komponenterne giver, fås

$$\begin{aligned} 0 &= \frac{V_2(f) - V_3(f)}{R} + j2\pi f C \cdot (V_1(f) - V_3(f)) \\ &= \frac{X(f) - V_3(f)}{R} + j2\pi f C \cdot (0 - V_3(f)) \Rightarrow \end{aligned}$$

$$Y(f) = V_3 = \frac{X(f)}{R} \cdot \frac{1}{j2\pi fC + 1/R} = \frac{X(f)}{j2\pi fRC + 1} \Rightarrow$$

$$H(f) = \frac{Y(f)}{X(f)} = \frac{1}{j2\pi fRC + 1}$$

Et andet interessant begreb for fysiske kredsløb er **impedans** for en samling af komponenter, defineret som ovenfor $Z = V(f)/I(f)$, hvor $V(f)$ og $I(f)$ er de komplekse repræsentanter for spænding og strøm gennem samlingen af komponenter. Man kan som forklaret ovenfor regne med impedanser ligesom resistanser, fx er kredsløbet i Figur 4.3 en serieforbindelse af Z_R og Z_C , så den impedans “ X ser ind i”, også kaldet indgangsimpedansen, er $Z_R + Z_C$. Overføringsfunktionen kunne også være fundet simpelt ved at regne på en såkaldt spændingsdeling mellem Z_C og $Z_R + Z_C$. I kurset “Tidskontinuerlige signaler”, vil metoden med analyse i frekvensdomænet blive udviklet til meget mere komplicerede systemer og mange avancerede konklusioner vil blive draget. Bemærk i øvrigt at videregående kurser som regel anvender $\omega = 2\pi f$ som variabel og den komplekse beregningsmetode kaldes ofte “ $j\omega$ -regning”. Formlerne med ω er lidt elegantere. \square



Figur 4.4: En generator med belastning.

Et af de vigtigste resultater i kredsløbsteorien er, at selv om man har et indviklet kredsløb med mange (lineære) komponenter og strøm- og spændingskilder, kan man repræsentere det ved et simpelt kredsløb illustreret i Figur 4.4 bestående af én spændingskilde - også kaldet en generator (vist med en cirkel) - i serie med én impedans (vist som en firkantet boks). Dette kaldes **Thevenin's sætning**. Vi vil ikke komme ind på hvordan man kan finde disse to størrelser (se dog efterfølgende eksempel), blot kalde dem henholdsvis E (spænding, reel værdi) og $Z_G = R_G + jX_G$. Man kan nu stille sig det spørgsmål: Hvordan kan man vælge en belastningsimpedans, $Z_L = R_L + jX_L$, således at der afleveres mest

mulig realeffekt i denne. Vi skal altså maksimere

$$\begin{aligned} \operatorname{Re}(VI^*) &= \operatorname{Re}\left(\frac{EZ_L}{Z_L + Z_G} \cdot \frac{E}{(Z_L + Z_G)^*}\right) = E^2 \frac{R_L}{|Z_L + Z_G|^2} = E^2 \frac{R_L}{(R_L + R_G)^2 + (X_L + X_G)^2} \\ &\leq E^2 \frac{R_L}{(R_L + R_G)^2} \end{aligned}$$

idet den maksimale værdi nemt ses at blive opnået for $X_L = -X_G$. Tilbage står at maksimere over R_L . Ved at differentiere fås at maksimum opnås for $R_L = R_G$, så samlet kan man konkludere at den maksimale effekt afsættes i Z_L , når denne vælges som $Z_L = Z_G^*$, altså den kompleks konjugerede til Z_G .

Dette kaldes **impedanstilpasning**. Den maksimale effekt kaldes rådighedseffekten og bliver i øvrigt $E^2/(4R_G)$, hvis E er en effektivværdi.

Eksempel 4.4 - Thevenin ækvivalent

Vi ser på det simple kredsløb i Figur 4.3. Generatorspændingen E er nem at finde for den ses som udgangssignal når der ikke løber noget strøm ud ("tomgang") og det var netop situationen da vi beregnede overføringsfunktion:

$$E = V_3 = \frac{X(f)}{j2\pi fRC + 1} = H(f)X(f)$$

Generatorimpedansen kan vi finde ved at slukke for E , dvs. $X(f) = 0$, og finde den impedans man ser ind i fra udgangen. Når $X = 0$ er R og C parallelkoblet og hvis man husker en formel for parallelkoblede modstande fra skoletiden får man

$$Z_G = R \parallel \frac{1}{j2\pi fC} = \frac{R \frac{1}{j2\pi fC}}{R + \frac{1}{j2\pi fC}} = \frac{R}{j2\pi fRC + 1}$$

Hvis man ønsker det vil man så kunne finde den tilpassede impedans. \square

4.6 Tidsdiskrete lineære systemer

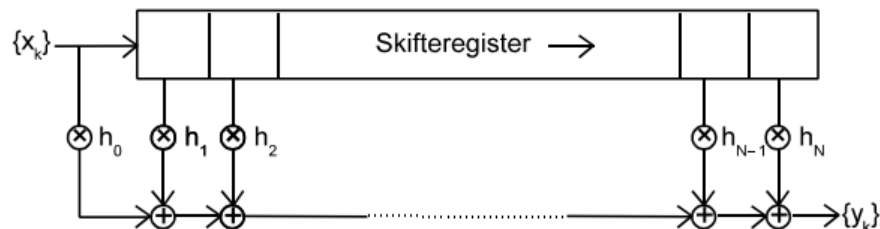
Moderne realiseringer af lineære systemer vil ofte operere på samplede værdier som altså findes på adskilte (diskrete) tidspunkter kT_S . Vi indførte frekvensdomænet for disse i Afsnit 3.3.

Et (kausalt) tidsdiskret lineært system med indgangssignal x_k og udgangssignal y_k til tiden kT_S kan beskrives som

$$y_k = h_0 x_k + h_1 x_{k-1} + h_2 x_{k-2} + \dots = \sum_{n=0}^{\infty} h_n x_{k-n} \quad (4.5)$$

Denne operation kaldes som for det tidskontinuerte tilfælde en **foldning** og vektoren (h_0, h_1, h_2, \dots) kaldes også her **impulsresponsen**.

I figuren vises et sådant system med endelig impulsrespons, dvs. afhængigheden af x går kun et endeligt antal skridt tilbage i tiden. I MATLAB kan funktionen udføres ved $y = \text{conv}(h, x)$, (foldning er *convolution* på engelsk).



Figur 4.5: Tidsdiskret lineært system med endelig impulsrespons.

Det tidsdiskrete systems virkning kan som det tidskontinuerte system beskrives ved en **overføringsfunktion** $H(f)$:

$$Y(f) = H(f)X(f) \quad \text{hvor} \quad H(f) = \sum_{k=0}^{\infty} h_k e^{-j2\pi k f T_s} \quad (4.6)$$

Vi giver et eksempel i den tilknyttede øvelse. Som for det tidskontinuerte tilfælde er $H(f)$ en kompleks funktion. $|H(f)|$ angiver ændringen i amplitude for et harmonisk signal med frekvensen f og $\angle H(f)$ giver faseforskydningen. Bemærk en meget stor forskel til det tidskontinuerte tilfælde: **Overføringsfunktionen er periodisk med perioden $1/T_s$** (som alle spektre i øvrigt for tidsdiskrete systemer). Vi har ikke umiddelbart defineret harmonisk signal i dette tilfælde, men tænk på en samplet version af et tidskontinuert harmonisk signal. Læg mærke til at signalerne ikke umiddelbart har nogen fysisk størrelse, men de kan eventuelt være samplede tidskontinuerte signaler som spændinger eller strømme.

KAPITEL 5

Transmission med pulser

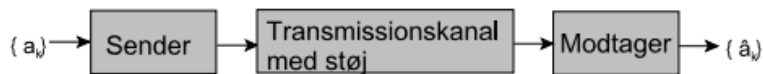
De fysiske forbindelser i et kommunikationssystem kan benytte mange forskellige medier som vi skal se i Kapitel 7. Der er dog kun to hovedtyper for selve kommunikationen:

- **Basisbåndstransmission** hvor signalerne transmitteres uden at blive flyttet i frekvens. Dette kapitel vil beskæftige sig med en form for basisbåndskommunikation hvor der anvendes pulser til repræsentation af digital information.
- **Modulerede systemer** hvor signaler flyttes til en højere frekvens for at kunne tilpasses til transmissionsmediet, fx radio. Kapitel 6 vil beskæftige sig med dette.

I dette kursus vil vi først og fremmest beskæftige os med **digital kommunikation**, da alle nyere kommunikationssystemer baserer sig på dette. Med digital kommunikation mener vi at informationen vi ønsker at transmittere er repræsenteret som **symboler**, a_k valgt ud af et såkaldt **alfabet**. I disse noter angiver M antallet af symboler i alfabetet og digital betyder at der netop et endeligt antal af disse - ikke fx værdier i et kontinuert interval. Indekset k har vi tilføjet for at sige at vi vil transmittere a_k til tiden kT , hvor T kaldes **symboltiden**, den tid der går mellem hvert symbol. Fra tidligere i kurset kendes det grundlæggende binære alfabet $\{0, 1\}$, $M = 2$, hvis elementer vi som tidligere kalder **bit**. Som det vil fremgå af dette og det næste kapitel, er dette alfabet helt utilstrækkeligt i den fysiske kommunikation og vi vil fx anvende alfabeter som $\{-7, -5, -3, -1, 1, 3, 5, 7\}$ som har $M = 8$. Den helt præcise sammenhæng mellem dette alfabet og det grundlæggende binære $\{0, 1\}$ -alfabet kommer vi ind på nedenfor, men bemærk allerede nu at 3 bit kan danne 2^3 forskellige kombinationsmuligheder og det er netop M for det viste alfabet.

Årsagen til at alle systemer nu anvender digital kommunikation er først og fremmest at de transmitterede signaler under transmissionen påvirkes af omgivelser og transmissionsmedium - i nedenstående Figur 5.1 vist som transmissionskanal med støj - så de ser meget anderledes ud ved modtagelsen. Når der kun er et endeligt antal mulige symboler kan modtageren vælge det mest sandsynlige, noget der ville være helt umuligt hvis vi anvendte et alfabet med uendeligt mange værdier, et analogt signal. Vi har vist dette på figuren som \hat{a}_k for at markere at det er modtagerens gæt på a_k . Som det vil fremgå af dette kapitel tager modtageren selvfølgelig fejl engang imellem, men som vi skal

se, er der metoder til at gøre risikoen for dette så lille som muligt og derfor vil digital kommunikation være analog kommunikation overlegen.

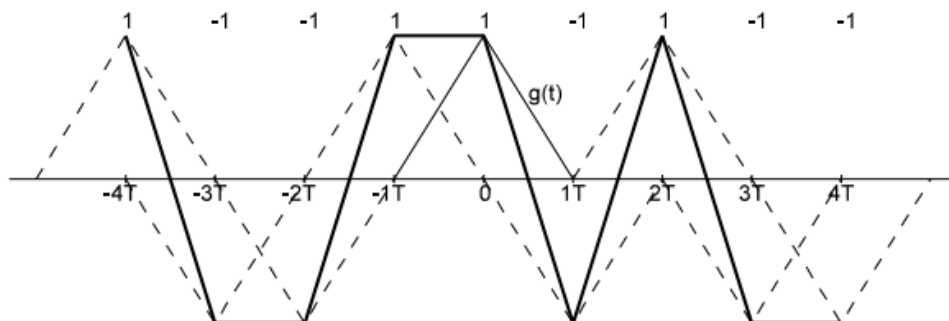


Figur 5.1: Transmissionssystem.

I Kapitel 1 viste vi **kommunikationssignaler**, hvor det samlede signal var opbygget af en række ens pulser (Figur 1.2 og 1.3). Hver af disse pulser blev multipliceret med værdien af et symbol som fx kunne være ± 1 , og på denne måde kommunikerer det digitale signal til en modtager, der ud fra det modtagne signal finder symbolerne. Umiddelbart giver denne form for kommunikation ikke nogen forskydning i det anvendte frekvensområde (basisbåndstransmission) men den anvendes også som grundlag for digital modulation som vi skal se i Kapitel 6. Vi kan udtrykke det transmitterede signal matematisk som

$$x(t) = \sum_{k=-\infty}^{\infty} a_k g(t - kT) \quad (5.1)$$

Dette beskriver et signal, hvor symbolet a_k transmitteres til tiden kT ved hjælp af en **puls** (også kaldet **pulsform**), $g(t)$, som er forskudt til tiden kT og skaleret med faktoren a_k . Denne form for kommunikation kaldes **PulsAmplitudeModulation** eller blot **PAM**.



Figur 5.2: PAM-signal (fuldt optrukket) bestående af overlappende pulser med formen $g(t)$ og en tilfældig sekvens fra alfabet $\{-1, +1\}$.

Som vi kan se på figuren og også så i Figur 1.3 kan varigheden af den enkelte puls være større end T , og summationen ovenfor beskriver da også at de overlappende dele så adderes i stedet for blot at blive sat efter hinanden. Man kunne tænke sig andre metoder til at repræsentere symbolerne fysisk. Vi kommer ind på nogle i kapitlet om modulation, men hvis vi holder os til basisbåndstransmission kunne man foreslå **PPM**, **PulsPositionModulation**, hvor symbolets værdi angives som en placering af en smal puls i det længere interval T (beskrives lidt nærmere i Afsnit 6.3), eller man kunne foreslå

PWM Pulsbreddemodulation, hvor symbolets værdi angives med bredden (engelsk *width*) af pulsen. Ingen af disse anvendes i samme udstrækning som PAM og vi skal som sagt også i Kapitel 6 se at PAM er grundlaget for de fleste digitale modulationsformer.

Da der transmitteres ét symbol a_k i hver tidsperiode T (symboltiden), kan vi tale om en hastighed for dette: **symbolhastighed**. Enheden for symbolhastighed er **baud**:

$$\text{symbolhastighed} = \frac{1}{T} \text{ baud} \quad (5.2)$$

Det er klart at kommunikation med meget store alfabet overfører mere information end små alfabet, fx det binære alfabet $\{-1, +1\}$, selv om symbolhastigheden er den samme. Man har derfor brug for et mål for information. Hvis man forestiller sig alle M symboler beskrevet som en følge af valg med to udfald vil en følge af n valg kunne beskrive 2^n symboler. De to udfald kan angives ved værdierne 0 og 1, dvs. som **bit**, og man kan sige at n bit kan beskrive 2^n symboler eller omvendt: M symboler giver $\log_2 M$ bit, hvor \log_2 betyder logaritmen med grundtal 2. Man kan derfor definere en hastighed der beskriver informationsoverførslen, og denne kaldes **signaleringshastighed** eller **bitrate**:

$$\text{Signaleringshastighed eller bitrate } R = \frac{1}{T} \cdot \log_2 M \text{ bit/s} \quad (5.3)$$

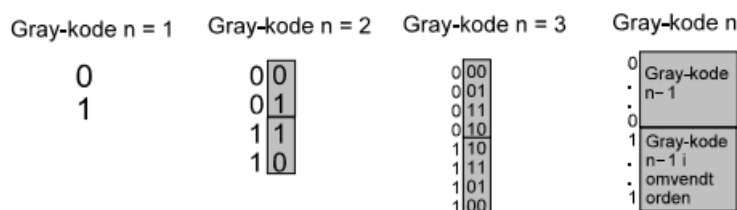
Læg mærke til forskellen på de to hastigheder, idet mange ganske ukritisk angiver baud, når de mener bit/s. Lidt senere vil vi se at symbolhastigheden (i baud) kan være begrænset af egenskaber ved det (lineære) system som transmissionsmediet udgør, mens alfabetets størrelse (og dermed bitraten) er begrænset af hvor god en opløsning man kan skelne alfabetets værdier fra hinanden og her er begrænsningen noget andet som vi vil se.

Det viser sig senere at det ikke er helt ligegyldigt hvordan sammenhængen mellem de M symboler og de tilhørende bitsekvenser er. Dette skyldes at transmissionssystemer, som beskrevet i næste afsnit, sommetider opfatter et modtaget symbol forkert og man er derfor interesseret i at en sådan fejl ikke skal give anledning til mange fejl i bitsekvensen. Oftest sker fejlen til “nabosymbolet” og det kunne være interessant hvis man kunne lave en sammenhæng hvor dette kun gav anledning til én bitfejl. Dette illustrerer vi det følgende eksempel.

Eksempel 5.1 Gray kodning af symboler til bit

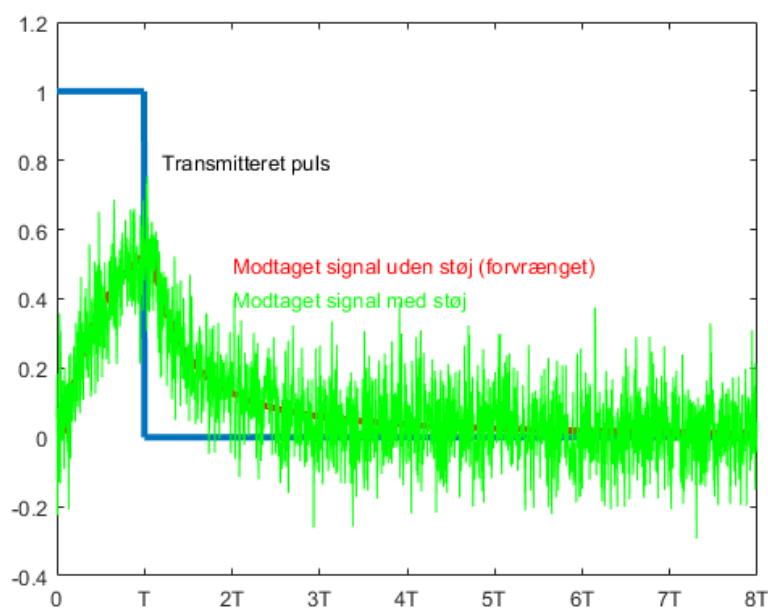
Et lille eksempel illustrerer problemet: Lad $M = 4$ og alfabetet være $\{-3, -1, +1, +3\}$. Der skal altså $\log_2 4 = 2$ bit for hvert symbol. Hvis vi tildeler bitsekvenserne 00, 01, 10 og 11 til disse kan man se at man får to fejl ved overgangen mellem -1 og $+1$. Dette problem løses ved den såkaldte **Gray-kode** som for eksemplet tildeler 00, 01, 11 og 10

som ses kun at give én fejl ved overgang til nabosymbolerne. (Det gælder endog cyklisk, så hvis -3 og $+3$ er naboer kommer der også kun én fejl her). Gray-koder kan konstrueres systematisk som vist på nedenstående Figur 5.3. \square



Figur 5.3: Systematisk konstruktion af Gray-kode.

I Kapitel 7 vil vi komme tilbage til forskellige transmissionsmedier og hvad de har af effekt på det transmitterede signal. Det er klart at det må være af betydning hvordan den enkelte puls slipper igennem systemet og som eksempel viser vi Figur 5.4, hvor man kan se, hvordan en firkantet puls kan se ud ved modtagelsen efter at have gennemløbet en transmissionsledning bestående af et par metalliske ledere, fx en abonnentledning til en almindelig telefon.



Figur 5.4: Firkantet puls ved afsendelsen og modtagelsen.

På Figur 5.4 ser vi to fænomener:

- Den modtagne puls bliver **forvrænget** da transmissionsledningen ikke overfører alle frekvenser ens. Den modtagne puls kan netop anes på figuren og vi har kunstigt vist

den som en fuldt optrukket streg. Forskellige frekvenser ankommer til modtageren på forskellige tidspunkter. Dette kaldes dispersion og dispersionen ændrer pulsens form kraftigt.

- Det modtagne signal udviser kraftige udsving fra den forvrængede form på grund af **støj** der er uvedkommende signaler som udefra forstyrrer pulsen.

Vi behandler disse to fænomener hver for sig i de efterfølgende afsnit, først støjen.

5.1 Støj i transmissionssystemer

Som defineret ovenfor er **støj udefrakommende og uvedkommende signaler**. Det er karakteristisk at vi normalt ikke kan sige noget præcist om disse signaler, så vi må ty til en **stokastisk beskrivelse** som vi meget kort skrev det i Kapitel 1. Det ligger langt uden for dette kursus' rammer at gøre dette med den rette matematik, men alligevel vil vi nedenfor vise en matematisk beskrivelse af støjen i Figur 5.4 og i et senere Eksempel 5.2 omtale hvordan en sådan frembringes i en simulering.

I et givet transmissionssystem er der flere kilder til støj og det er ikke på forhånd helt nemt at vide hvad der er det dominerende. Følgende kilder kan komme på tale:

- Baggrundsstøj der kommer ind i (radio)systemet via antenner og kredsløb der utilsigtet virker som antenner. Når vi kalder det baggrundsstøj er det fordi det stammer fra forskellige naturlige støjkilder som giver signaler i forskellige frekvensområder. Under 30 MHz stammer støjen fra ionosfæren hvor den er opsamlet fra lynudladninger under tordenvejr. Mellem 30 MHz og 1 GHz stammer støjen fra rummet og over 1 GHz stammer støjen fra de termiske svingninger af atmosfærens molekyler.
- Termisk støj. Elektroner der transporterer de elektriske strømme har tilfældige bevægelser som funktion af temperaturen. Dette giver anledning til små tilfældige strømme som over en modstand giver en spænding (Ohms lov). Denne støj findes jævnt fordelt over alle frekvenser op til i hvert fald 1000 GHz. Somme tider afkøler man kredsløb for at reducere denne støj.
- Et lignende fænomen opleves på grund af at strømme er sammensat af individuelle ladningsbærere (elektroner) og ikke bare er en jævn strøm med konstant intensitet. De små udslag dette giver anledning til kaldes haglstøj (engelsk *shot noise*) og dette findes fx også i lysledere, hvor signalerne transmitteres ved fotoner. På én måde adskiller denne sidste form for støj sig fra den almindeligt termiske, nemlig ved at være afhængig af signalet. Hvis der ikke er nogen ladning/lysmængde at transportere er der jo ikke noget der kan variere.

- Støj fra omgivelserne. Her tænkes på kunstigt frembragte signaler i modsætning til baggrundsstøjen ovenfor. Der er mange konkurrerende transmissionssystemer, fx kan man tænke på et telefonkabel der går ind i et lejlighedskompleks hvor alle lejligheder har hver deres telefon med måske ISDN eller ADSL. En anden mulighed er radiotransmission på samme frekvens. Også uden for kommunikationsverdenen genereres generende støj, fx ved start og stop af motorer og ved gnister der fremkommer i motorers kommutatorer eller andre steder hvor der overføres store strømme.
- Støj fra signalet selv. I visse transmissionssystemer kan der danne sig kopier af et transmitteret signal, og kopierne ankommer til modtageren på forskudte tidspunkter. Det er fx tilfældet med ekkoer der kan opstå i transmissionsledninger (se Kapitel 7) og radiosignaler der har gennemløbet forskellige udbredelsesveje (også i Kapitel 7). I princippet kan disse fjernes hvis man er i stand til at finde ud af hvad der sker, så det er nærmere en slags tilfældig forvrængning.

Som det fremgår af beskrivelsen af de forskellige støjkilder kan støjen meget sjældent beskrives som et deterministisk signal - og i de tilfælde kan den principielt fjernes. "Rigtig" støj er et stokastisk signal og kan kun beskrives med metoder fra sandsynlighedsregning og stokastiske processer. Det kan vi ikke lige klare her, men på Figur 5.5 viser vi hvordan støjsignalet værdier fordeler sig i den model for støj vi anvendte i Figur 5.4. I denne model gør vi forskellige antagelser som viser sig at være meget realistiske for ovennævnte støjkilder:

- Støjsignalet, $n(t)$, antages at kunne **adderes** til det signal, $x(t)$, vi er interesserede i, dvs. det modtagne signal er $r(t) = x(t) + n(t)$. Det er som regel tilfældet.
- Støjsignalet antages at være **uafhængigt** af det signal vi er interesseret i. Det er også som regel tilfældet, men det er klart ikke tilfældet for ovennævnte haglstøj, hvor støjens "styrke" er afhængigt af signalet. Som regel kan denne afhængighed klares ved at beskrive støjen som havende forskellige styrker (σ nedenfor), mens de enkelte støjværdier ikke afhænger af signalet. Det er også klart at det er umuligt at beskrive den sidste form for støj (fra ens eget signal) som uafhængig af signalet, så modellen kan ikke anvendes til denne form for støj.
- Støjens værdier beskrives ved sandsynligheden for at støjværdien, n , falder i intervallet $a \leq n \leq b$ er

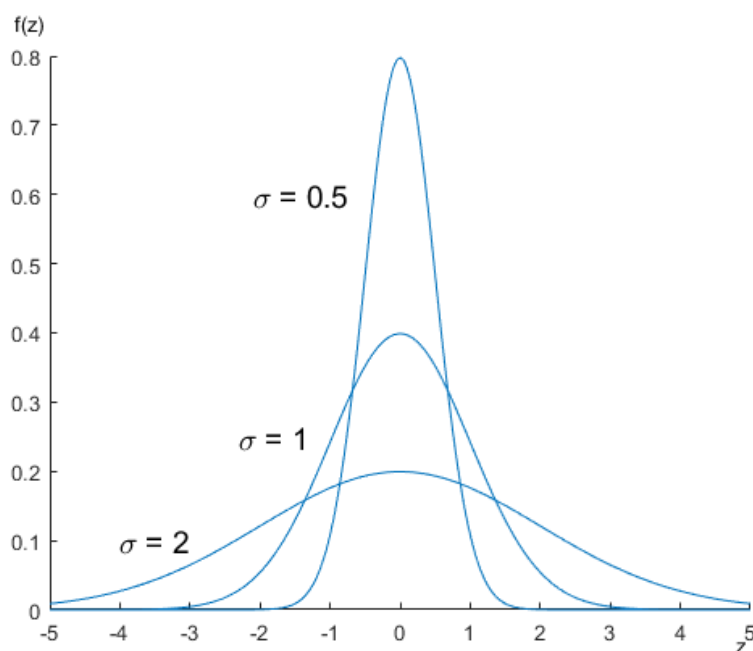
$$P(a \leq n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz \quad (5.4)$$

Vi har ikke defineret sandsynlighedsbegrebet, men et uformelt forsøg på dette er: Det er et tal mellem 0 og 1 som angiver den brøkdel af et stort antal samples af

n som forventes at falde i det givne interval. I sandsynlighedsregningen kaldes den funktion der integreres for sandsynlighedstæthedsfunktionen $f(z)$, altså

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} \quad (5.5)$$

og denne vises for forskellige σ (sigma) i Figur 5.5.



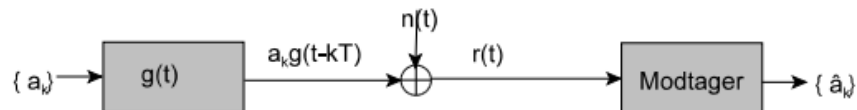
Figur 5.5: Fordelingen af amplituder i normalfordelt (gaussisk) støj.

Denne form for stokastisk signal kaldes **normalfordelt** eller **gaussisk** og den er i en vis forstand den mest uforudsigelige form for støj. Læg fx mærke til at vilkårligt høje værdier kan forekomme; de er bare ikke særligt sandsynlige. Som det kan ses af figuren er σ et mål for “styrken” af støjen. I Figur 5.4 anvendte vi $\sigma = 0,1$. Læg i øvrigt mærke til at middelværdien af normalfordelt støj er 0, da tætheden er symmetrisk om $z = 0$.

- Fordelingen af værdierne siger ikke noget om de frekvenser støjen findes ved. Normalt regner vi med at støjen i det frekvensområde, der interesserer, er **hvid**, dvs. støjen har lige stor styrke for alle frekvenser. Som det ses af beskrivelsen af forskellige støjkluder ovenfor er dette rimeligt realistisk for et begrænset frekvensområde.

Sammenfattende anvender vi altså gerne **additiv, hvid, normalfordelt støj** i vores modeller af transmissionssystemer. Figuren nedenfor viser i blokdiagram en sådan model af et transmissionssystem som transmitterer pulser med pulsformen $g(t)$ og der adderes

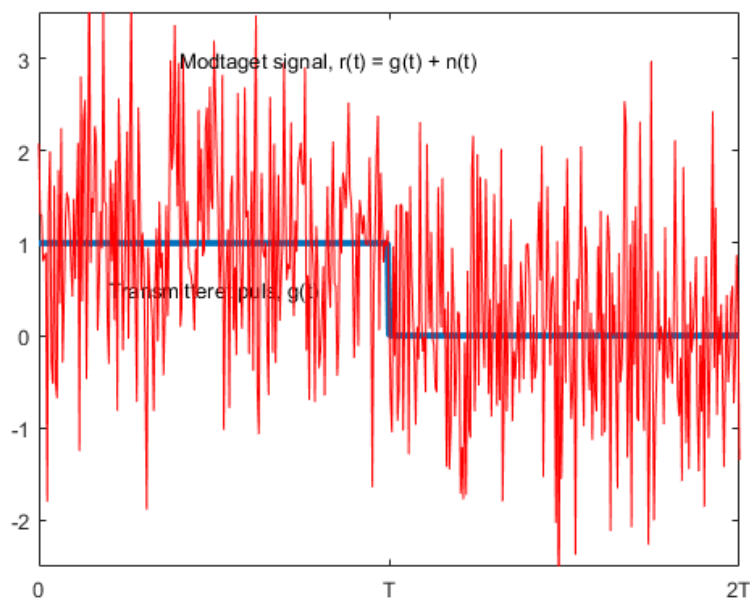
støj, $n(t)$, til. I næste afsnit ser vi på hvordan vi skaber den optimale modtager for disse pulser, dvs. den modtager der kan bestemme \hat{a}_k med mindst mulig sandsynlighed for afvigelse fra den sendte a_k .



Figur 5.6: Transmission af pulser $g(t)$ med additiv støj, $n(t)$.

5.1.1 Optimal modtager for signal i støj

Hvis vi ser bort fra forvrængningen, vil en firkantpuls, $g(t)$, med støj, $n(t)$, se ud som på Figur 5.7.



Figur 5.7: Transmitteret firkantsignal, $g(t)$, med støj lagt til under transmission.

Når signalet modtages skal man finde et eller andet sted at sample det for at kunne bestemme den transmitterede værdi. Fx kunne det være at figuren viser transmission af $+1$, men at -1 også er en mulig symbolværdi i systemet, og problemet er at skelne disse. Hvis man sampler på et eller andet sted i løbet af firkanten vil nogle samples give negative værdier og man ville kunne tro at den transmitterede værdi var -1 . I middel er støjen 0, så det vil nok være fornuftigt at udregne middelværdien over hele intervallet fra

0 til T af det modtagne signal $r(t)$:

$$s(T) = \int_0^T r(t)dt = \int_0^T g(t)dt + \int_0^T n(t)dt = T + \int_0^T n(t)dt$$

Det sidste integral er middelværdien af støjen og den vil være 0 hvis man integrerer fra $-\infty$ til ∞ . For vores interval kan den give noget forskelligt fra 0, men det synes alligevel fornuftigt. For signalet i figuren er anvendt støj med $\sigma = 1$ og udregning af $s(T)$ gav $0,989 \cdot T$, hvilket klart indikerer at værdien $+1$ blev transmitteret.

Man kan vise at denne metode faktisk er den **optimale modtager** for den firkantede puls, dvs. den modtager der giver **mindste sandsynlighed for fejl**. Hvis man har en mere generel pulsform $g(t)$ med varighed T_p , udregner man på lignende måde

$$s(T_p) = \int_0^{T_p} r(t)g(t)dt = \int_0^{T_p} g(t)^2dt + \int_0^{T_p} n(t)g(t)dt = E + \int_0^{T_p} n(t)g(t)dt$$

Det er sådan set meget naturligt at man skal multiplicere det modtagne signal med pulsformen $g(t)$ idet man så komme til at vægte de “stærke” steder i signalet højere end de “svage” steder. I et mere avanceret sprog betyder dette at man udregner den såkaldte **korrelation** mellem det modtagne signal og pulsformen. Hvis vi som ovenfor har transmitteret $+1$, fås uden støj værdien E (der kaldes pulsformens **energi**):

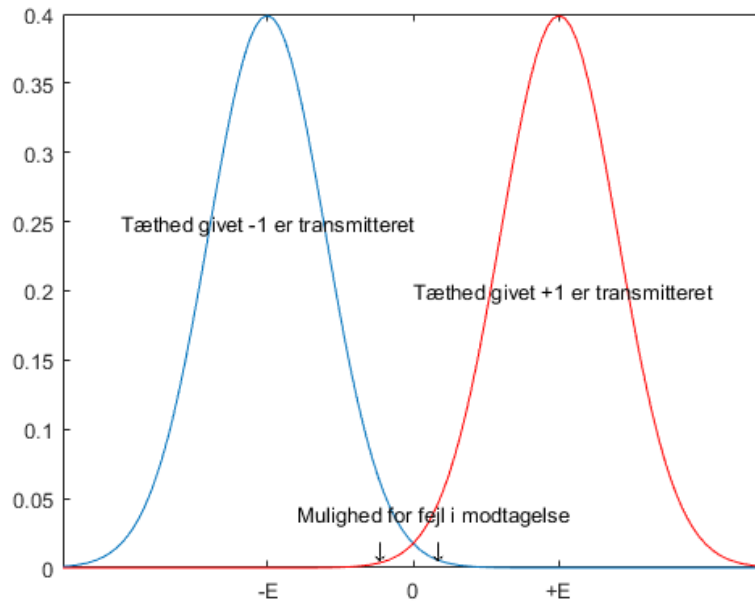
$$E = \int_0^{T_p} g(t)^2dt$$

og det sidste integral giver en normalfordelt værdi med $\sigma = \sigma_{\text{støj}}\sqrt{E}$. Hvis vi har andre symbolværdier end $+1$ får vi for symbolet a_k :

$$s(T_p) = \int_0^{T_p} r(t)g(t)dt = \int_0^{T_p} a_k g(t)^2dt + \int_0^{T_p} n(t)g(t)dt = a_k E + \int_0^{T_p} n(t)g(t)dt$$

På Figur 5.8 viser vi tæthedsfunktionerne for den støjfyldte korrelation for to mulige transmitterede symboler -1 og $+1$. På figuren ser vi et meget vigtigt karakteristikum for transmission af signaler i støj: Der er mulighed for at tage **fejl** selv med en optimal modtager. Vi vil ikke her beregne en **fejlsandsynlighed**, men den kan findes som arealet hvor de to tæthedsfunktioner overlapper.

Det er klart at fejlsandsynligheden afhænger af forholdet mellem signalets energi E og støjens styrke udtrykt ved σ^2 . Når energien øges, flyttes de to kurver længere fra hinanden og overlappet bliver mindre. Når støjen omvendt øges, bliver de to kurver bredere og



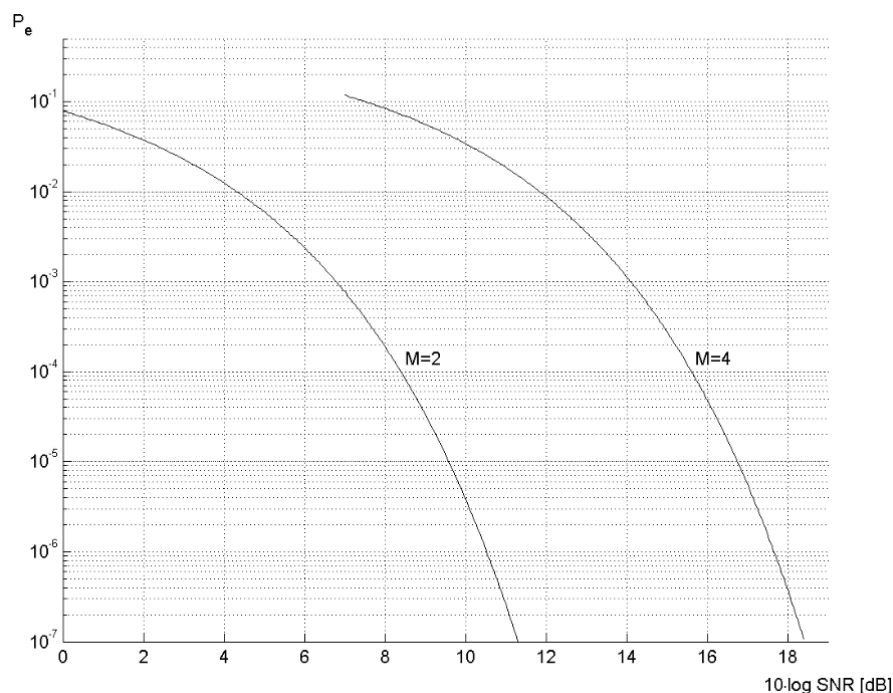
Figur 5.8: Sandsynlighedstæthed for korrelationen givet to forskellige transmitterede symboler.

overlappet større. Dette forhold kaldes **signal/støjforhold** (SNR) og det er almindelig praksis at angive det i et logaritmisk mål, nemlig **dB** defineret (jf. Afsnit 4.2) ved

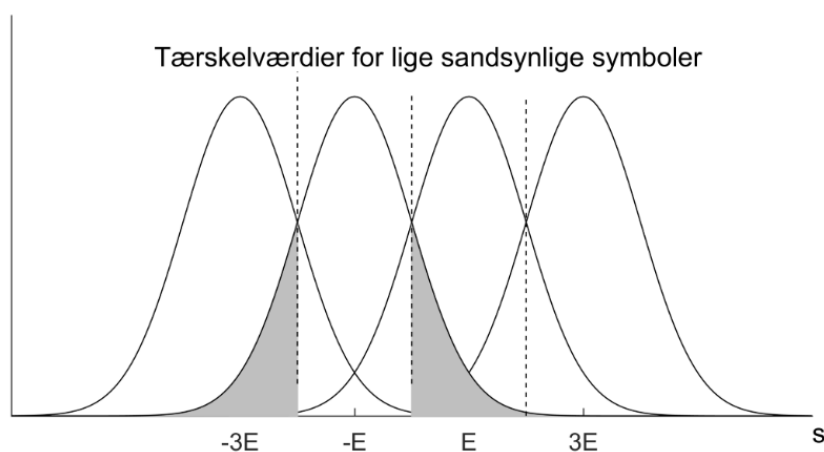
$$\text{SNR}_{\text{dB}} = 10 \cdot \log_{10}(\text{SNR})$$

I Figur 5.9 viser vi typiske sammenhænge mellem fejlsandsynlighed P_e og SNR_{dB} . Næsten alle systemer har en sådan “vandfaldskurve” for denne sammenhæng. Det er ikke helt præcist $\text{SNR}_{\text{dB}} = \log_{10}(E/\sigma^2)$ der ses på figuren, men SNR her er forholdet mellem den gennemsnitlige energi og $2\sigma^2$. På Figur 5.9 er vist fejlsandsynligheder for to systemer, nemlig med 2 og 4 symboler i alfabetet. Ved flere end to symboler må modtageren indlægge nogle tærskelværdier så man kan skelne symbolerne. I første eksempel var tærskelværdien 0, midt i mellem de modtagne værdier og som regel lægges tærsklerne også midt i mellem de andre værdier. Som det ses, kræves der mere energi i signalet for at kunne skelne fire værdier med samme fejlsandsynlighed og samme støj. Det er altså **signal/støjforholdet der afgør, hvor mange værdier det er muligt at skelne og dermed bestemmer bitraten**.

I Figur 5.10 viser vi et eksempel på sandsynlighedstætheder for 4 symboler: -3 , -1 , $+1$, $+3$.



Figur 5.9: Fejlsandsynlighed som funktion af signal/støjforhold.



Figur 5.10: Sandsynlighedstæthed for korrelationer givet fire forskellige transmitterede symboler. De grå felter viser fejlmuligheder hvis -1 er transmitteret (og arealet er sandsynligheden for disse fejl).

Umiddelbart kunne man tro at på grund af støjens stokastiske karakter ville der intet være at gøre med hensyn til at fjerne fejlene i modtagelsen og sådan troede man også i mange år - indtil 1948. Dette år publiceredes et grundlæggende værk om kommunikation af C. E. Shannon. Her beviste han at man kan opnå en vilkårlig lille fejlsandsynlighed uden at ændre på signal/støjforholdet, når blot man anvender **fejlkorrigerende koder**.

Fejlkorrigerende koder er matematiske konstruktioner, hvor man blandt alle mulige sekvenser (fx 2^N binære sekvenser af længde N) vælger nogle der adskiller sig på så mange positioner som muligt. Hvis der er 2^K sådanne sekvenser - kaldet kodeord - kan man transmittere K bit i en sådan blok og koden siges at have hastighed K/N . Fidusen består i finde et valg hvor "afstanden" mellem kodeordene bliver så stor som mulig, idet man så ud fra en modtaget sekvens kan finde det nærmeste kodeord. Det lyder som en tvivlsom fidus, da man jo spilder energi på at transmittere de ekstra $N - K$ symboler, men man kan vise at der alligevel kan opnås en forbedring blot K/N er mindre end den såkaldte kanalkapacitet, der kan udregnes fra signal/støjforholdet. Den ultimative fejlsandsynlighed 0 kan kun opnås ved uendeligt lange blokke, men praktiske systemer med begrænsede N giver en reduktion i transmitteret energi pr. symbol. Fx findes et meget anvendt system med $N = 2040$ bit og $K = 1784$, hvor man kan opnå en fejlsandsynlighed på 10^{-6} ved $\text{SNR} = 6,4$ dB. Sammenlignes med Figur 5.9 ses der at være en **kodningsgevinst** på ca. 4 dB. Sammenhængen mellem SNR og fejlsandsynlighed er en vandfaldskurve som på figuren, blot noget stejlere. I kurset 34311 blev fejldetektion og retransmission demonstreret som et værktøj til at opnå (næsten) fejlfri transmission, men dette værktøj er kun brugbart ved meget små fejlsandsynligheder - langt mindre end $1/\text{rammelængde}$ - da der ellers spildes energi ved retransmission af alt for mange rammer. Koder til fejldetektion og fejlkorrektion er dog ofte de samme.

Moderne modtagere er næsten altid implementeret med sampling af det modtagne signal og behandling af de modtagne værdier digitalt. Denne **tidsdiskrete, optimale modtager** sampler med tidsafstanden T_S og udregner altså (for det transmitterede a_k)

$$s = \sum_{i=0}^{L-1} r(iT_S)g(iT_S)T_S = \sum_{i=0}^{L-1} a_k g_i^2 T_S + \sum_{i=0}^{L-1} n_i g_i T_S = a_k E + \sum_{i=0}^{L-1} n_i g_i T_S$$

hvor vi har indført en kortere skriveform med indeks for de samplede værdier og L for pulsens længde. E er som før pulsformen $\mathbf{g} = (g_i)$'s energi:

$$E = \sum_{i=0}^{L-1} g_i^2 T_S$$

en tilnærmelse til den tidligere præsenterede energi med et integral. Den sidste sum giver også her en normalfordelt variabel med $\sigma = \sigma_{\text{støj}}\sqrt{E}$ og situationen er ganske som på Figur 5.8. Omskriver vi udregningen lidt fås

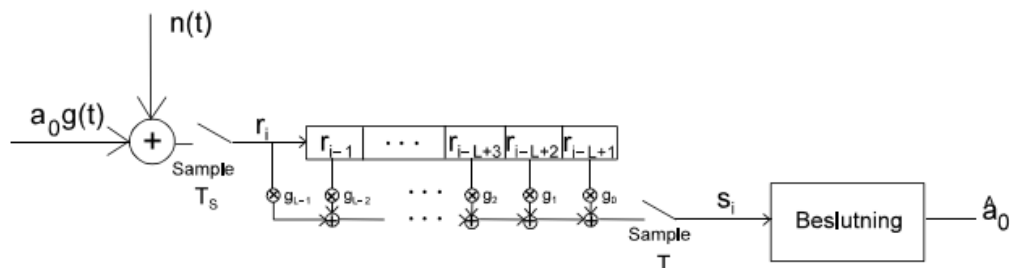
$$s = \sum_{i=0}^{L-1} r_i g_i T_S = \sum_{j=0}^{L-1} r_{L-1-j} g_{L-1-j} T_S = \sum_{j=0}^{L-1} r_{L-1-j} g_j^{(M)} T_S$$

som ved sammenligning med udtrykket (4.5) ses at være en foldning af \mathbf{r} og vektoren $\mathbf{g}_M = (g_0^{(M)}, g_1^{(M)}, \dots, g_{L-1}^{(M)})$. Det er altså resultatet af at sende det modtagne signal gen-

nem et tidsdiskret lineært system med impulsrespons

$$\mathbf{g}_M = (g_0^{(M)}, g_1^{(M)}, \dots, g_{L-1}^{(M)}) = (g_{L-1}, g_{L-2}, \dots, g_0),$$

hvilket er pulsen \mathbf{g} vendt om i tid. Dette system kaldes et **tilpasset filter** (engelsk: *matched filter*), så korrelationen mellem det modtagne signal og pulsformen udregnes altså ved at sende signalet gennem det tilpassede filter med impulsrespons \mathbf{g}_M og se på filterets udgang når hele pulsen er kommet ind. Det tilpassede filter er, som man kan se, pulsen vendt i tid og vi vil senere anvende $g(-t)$ som en kort skriveform for dette selvom der egentlig også er en tidsforskydning på $(L-1)T_S$. Den optimale modtager kan altså siges at bestå af et **tilpasset filter** efterfulgt af en **sampling** på det rette tidspunkt (her $(L-1)T_S$) og en **beslutning** om hvad det samplede signal er tættest på, fx ved eksemplet i Figur 5.8 at afgøre om det samplede signal er positivt eller negativt. Figur 5.11 viser en sådan optimal modtager.



Figur 5.11: Tidsdiskret, optimal modtager.

De fleste modtagere er realiserede ved hjælp af sådanne filtre, hvad enten de er tidskontinuerte eller tidsdiskrete. Et interessant problem er hvor tæt samplingen skal foregå for at opnå de ønskede egenskaber. I modtageren skal samplingen være tæt nok til at samplingssætningen er opfyldt for alle betydende frekvenskomponenter i det modtagne signal. Vi vil nedenfor se at frekvensindholdet for de mere avancerede pulsformer er af samme størrelsesorden som symbolhastigheden (udtrykt i Hz i stedet for baud). Vi har ikke beskæftiget os meget med senderen, men ofte frembringes pulsformen her af en tidsdiskret impulsrespons $\mathbf{g} = (g_i)$ svarende det samplede modtagne signal uden støj. Et lavpasfilter kan være nødvendigt for at danne det transmitterede tidskontinuerte signal og samplingsfrekvensen i senderen er bestemt af de egenskaber der kræves af pulsen under transmission, fx vil man ofte foreskrive et meget lavt frekvensindhold uden for de relevante frekvenser for ikke at genere andre transmissionssystemer (som beskrevet ovenfor som støj fra omgivelser). I Afsnit 5.2 vil vi komme lidt nærmere ind på hvordan pulsformen \mathbf{g} helt præcist opstår og hvilke egenskaber den skal have.

Eksempel 5.2 - Simulering af transmission med pulser

Når man skal simulere et transmissionssystem i MATLAB eller teste en realisering er det vigtigt at de transmitterede data er tilfældige, så systemet ikke “afspores” ved et bestemt forløb eller et sådant forløb måske ikke opdages da man fx kun testede med 01010101... Helt tilfældige data kan man jo ikke frembringe, men man kan komme tæt på. I MATLAB er der forskellige tilfældighedsfunktioner (der dog baserer sig på samme grundfunktion). En af disse er `randi` der giver tilfældige heltal, fx som anvendt nedenfor fås en tilfældig sekvens af 0 og 1.

```
% Lav 100 tilfældige +-1 symboler
N=100;
a=2*randi([0 1],1,N)-1;
```

De genererede symboler skal anvendes til et PAM-signal med en bestemt pulsform $g(t)$. For ikke at komplicere dette eksempel vælger vi en pulsform af længde T som altså ikke overlapper nabopulser. Vi bruger en savtakformet puls og sampler den $m = 8$ gange. Som i tidligere øvelser findes det transmitterede signal v ved at folde g med en version af a -sekvensen hvor symbolerne er anbragt med $m-1$ 0'er mellem sig (det kaldes ofte en opsamlet version af a og MATLAB har en funktion `upsample` der kunne frembringe denne, men for ikke at få for mange funktioner i kurset gør vi det på den simple måde):

```
% Anbring symbolerne med m-1 = 7 0'er i mellem
m=8;
a0=[a; zeros(m-1,N)];
a0=reshape(a0,1,N*m);

% Savtakpuls, højde 1
g=(0:m-1)/m;
v=conv(g,a0);
```

I MATLAB genereres en sådan støj nemt med funktionen `randn` så det støjfyldte signal r kan genereres som

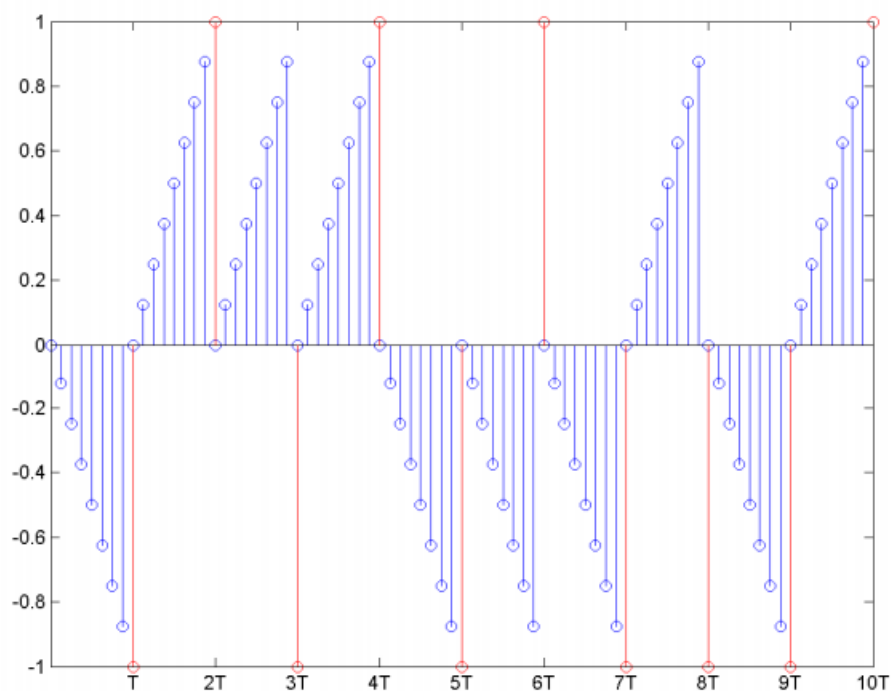
```
% Tilføj støj
sigma=2;
r=v+sigma*randn(1,length(v));
```

Størrelsen σ er vist i programmet som `sigma`. Støjen bliver automatisk hvid, idet det er en følge af at de enkelte elementer i støjen er uafhængige af hinanden.

Den optimale modtager skal folde med impulsresponsen for det tilpassede filter, g_M , som er g vendt om i tid (`fliplr`). Foldningen samples så på de rigtige tidspunkter med afstand T dvs. m i programmet. Det er ret simpelt her når pulserne ikke overlapper; der skal simpelthen samples i hver m 'te værdi af foldningen og man starter ved m . Ved andre pulsformer skal man se lidt mere på hvordan man starter, men det er stadig hver m 'te.

```
% Optimal modtager med tilpasset filter
g_M=fliplr(g);
s=conv(r,g_M);
% Sampling i slutningen af hver puls af længde T (=8)
a_hat=sign(s(m:m:end-m));
```

Som man kan se, findes det bedste gæt, \hat{a} , på transmitterede symbol a , ved at se på fortegnene af de samplede værdier da a kan have to værdier (\pm) som på Figur 5.8. De første 10 symboler og de resulterende modtagne værdier kunne illustreres ved



Figur 5.12: De 10 første pulser med tilhørende beslutninger i rødt. Der ses 3 fejlagtige beslutninger ud af de 10.

```
stem(0:79,v(1:80));
hold on
stem(m:m:80,a_hat(1:10),'r')
set(gca,'XTick',m:m:80)
```

```
set(gca,'XTickLabel',{'T','2T','3T','4T','5T','6T','7T',  
'8T','9T','10T'})  
hold off
```

Resultatet for en bestemt kørsel af programmet er vist ovenfor. Når man har længere simuleringer må man lade MATLAB udregne antallet af fejl:

```
sum(a~=a_hat)
```

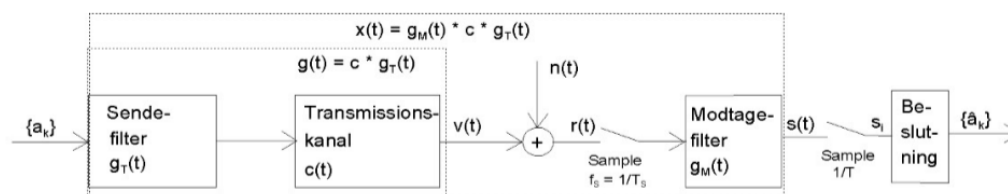
Her var der 19 fejl i 100 symboler, ret mange, men vi skulle jo se noget. \square

5.2 Transmissionssystem som lineært system

I et virkeligt kommunikationssystem vil den modtagne pulsform $g(t)$ være en kombination (foldning) af den transmitterede puls, $g_T(t)$, og impulsresponsen, $c(t)$, for transmissionskanalen:

$$g(t) = c * g_T(t)$$

Kanalen er et fysisk system og den har derfor nogle egenskaber som kan gøre at pulsen $g(t)$ ikke kan vælges frit. Som et eksempel kan vi se på Figur 5.4. Hvis vi ser “gennem” støjen kan vi skimte en kraftigt forvrænget puls og især kan vi se en lang “hale” på pulsen som vil blande sig med efterfølgende pulser, hvis man forestiller sig en sekvens af firkantpulser. Dette skyldes impulsresponsen for kanalen og der er ikke så meget at gøre ved det som vi skal se i Kapitel 7. Man kan eventuelt anvende filtrering og andre metoder til at opnå en såkaldt **udligning** af kanalen og det er ofte brugt. I kurset “Digital Kommunikation” kommer vi nærmere ind på sådanne metoder, men der er stadig ikke frit valg, så det vil være interessant at finde ud af hvordan pulsformerne skal vælges for at der ikke sker en sådan blanding af pulserne med hinanden. Et blokdiagram for det samlede system er vist på figuren



Figur 5.13: Basisbåndstransmissionssystem med total impulsrespons $x(t)$.

Det samlede system har impulsresponsen

$$x(t) = g_M * c * g_T(t)$$

hvor $*$ betyder foldning, da hver af de viste impulsresponser skal foldes med signalet. Ved udgangen af modtagefilteret har vi altså

$$s(t) = \sum_{k=-\infty}^{\infty} a_k x(t - kT) + \sum_{k=-\infty}^{\infty} g_M * n(t - kT)$$

og den optimale mdtager skal ud fra dette (samplet hver T) beslutte sig for de transmitterede symboler. Det er svært fordi der på udgangen af filteret ikke alene er et bidrag fra det symbol man gerne vil se, men også et eller andet bidrag fra de omgivende symboler. Man kalder dette fænomen **intersymbolinterferens** - ofte forkortet **ISI**. Hvis vi udtrykker dette lidt mere præcist i en formel fås, når man sampler for symbol nr. k , a_k :

$$s_k = \sum_{i=-\infty}^{\infty} a_{k-i} x_i$$

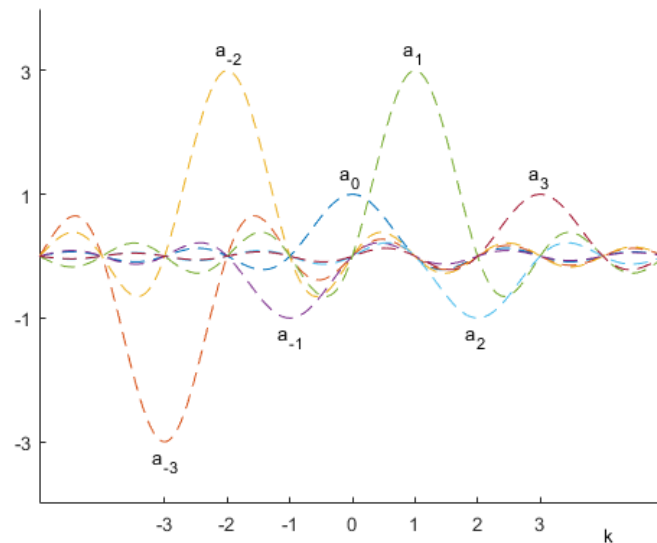
hvor $x_i = x(iT)$ angiver de vægtningsfaktorer de enkelte symboler indgår med i resultatet (og vi har set bort fra støjen). Det er sjældent at der indgår uendeligt mange symboler i ISI. Hvis fx det kun er de to omgivende symboler, bliver det noget enklere: $s_k = a_{k-1}x_1 + a_kx_0 + a_{k+1}x_{-1}$. Det vi gerne ville have er at der kun findes $s_k = a_kx_0$ og ikke alle de omgivende $x_i \neq 0$, da det så vil være meget enklere at finde det transmitterede symbol a_k . Denne problemstilling blev behandlet allerede i 1928 af Nyquist som opskrev det ret indlysende **Nyquist-kriterium for undertrykkelse af ISI**

$$x_i = \begin{cases} c & \text{for } i = 0 \\ 0 & \text{for } i \neq 0 \end{cases} \quad (5.6)$$

hvor c er en eller anden konstant. Et lille udsnit af et signal og de tilhørende pulser er vist på Figur 5.14, hvor pulsformen sikrer at der ikke er nogen interferens, idet værdierne for a_k som det ses ikke forstyrres af de omliggende da pulsformen $x(t)$ netop har nuller ved symbolerne.

Som tidligere er det mere operationelt at vide hvordan dette udtryk er i frekvensdomænet. Problemstillingen minder om den vi havde i Kapitel 2, hvor vi skulle rekonstruere et analogt signal ud fra dets samplede værdier. I samplepunkterne skulle signalet rekonstrueres præcist som den pågældende samplede værdi uden indblanding fra andre sampleværdier, mens der mellem punkterne var bidrag fra mange samplinger lagt oven på hinanden via interpolationsfunktionerne. I Kapitel 4 nåede vi frem til en formulering i frekvensdomænet med et ideelt lavpasfilter (jf. Afsnit 4.4.1) med afskæringsfrekvens $1/(2T)$. Hvis filteret helt præcis skal give x_i som i (5.6) skal det være

$$X(f) = \begin{cases} cT & \text{for } |f| \leq \frac{1}{2T} \\ 0 & \text{for } |f| > \frac{1}{2T} \end{cases} \quad (5.7)$$



Figur 5.14: Syv symboler i et PAM signal med tilhørende pulser uden ISI. Det samlede signal er ikke vist, men fås ved addition.

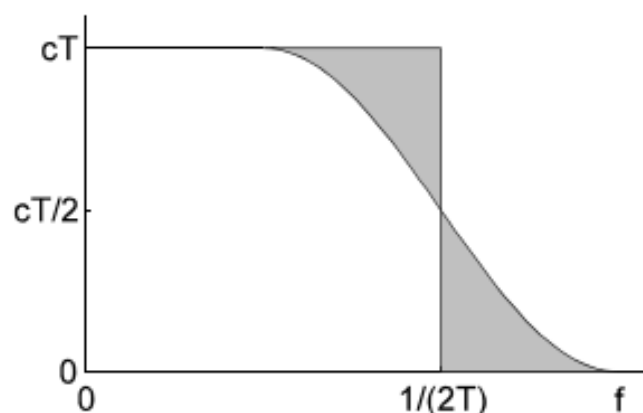
Nyquist (første) kriterium i frekvensdomænet kan altså udtrykkes som at overføringsfunktionen mellem indgangssymbolerne a_k og de samplede værdier efter det tilpassede filter, s_k , skal være det ideelle lavpasfilter $X(f)$ givet ovenfor.

Man kan fx sige det således: For at overføre $B = 1/T$ baud skal der anvendes frekvenser op til $1/(2T) = B/2$ Hz eller anderledes udtrykt B baud kræver en båndbredde på $B/2$ Hz.

Signalet vi viste i introduktionen på Figur 1.3 anvendte netop sådan nogle pulser som var 0 de rigtige steder og altså opfyldte Nyquist-kriteriet. Læg i øvrigt mærke til at disse pulser ideelt set er uendeligt lange og det transmitterede signal beskrevet i (5.1) dannes ved at lægge disse **oven på hinanden** (med passende forskydninger svarende til afstanden T mellem symbolerne som illustreret på Figur 5.14) og ikke bare efter hinanden som fx ved en firkantet pulsform. Den uendelige længde af disse pulser og diskussionen i Kapitel 4 om muligheden af at konstruere et ideelt lavpasfilter viser, at det ikke lader sig gøre i praksis at opfylde Nyquists 1. kriterium. Man skal vente uendeligt længe på resultatet, og hvis man skærer pulsen af for kort virker det ikke helt godt som vi fx så i øvelsen til Afsnit 2.2, når samplingsfrekvensen nærmede sig den kritiske værdi. Dette forhold var Nyquist selvfølgelig også klar over så han formulerede også en udvidet, blødere version af Nyquist-kriteriet, hvor der anvendes en lidt større båndbredde.

Man kan vise, at kravet (5.6) om nulgennemgange de rigtige steder kan opfyldes hvis

overføringsfunktionen $X(f)$ opfylder en særlig symmetribetingelse omkring $f = 1/(2T)$. Denne betingelse (**Nyquists 2. kriterium**) er illustreret på Figur 5.15.



Figur 5.15: Symmetribetingelse for $X(f)$ i Nyquists 2. kriterium.

De to grå områder skal være ens. Nogle meget anvendte overføringsfunktioner med netop denne egenskab er de såkaldte **cosinus-roll-off** karakteristikker, hvis formeludtryk er

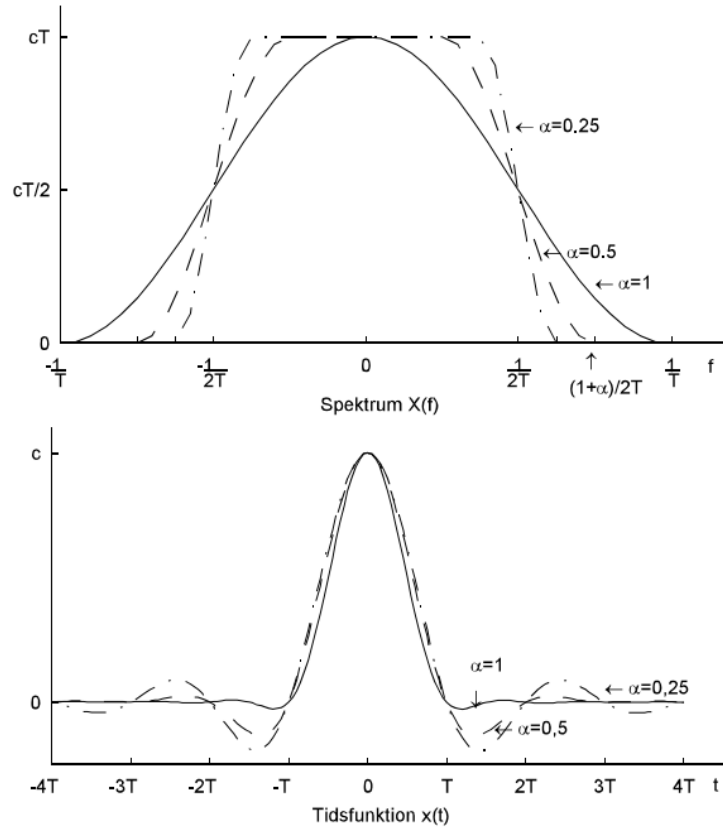
$$X(f) = \begin{cases} cT & 0 \leq |f| \leq (1 - \alpha)\frac{1}{2T} \\ c\frac{T}{2} \left(1 - \sin \frac{2\pi|f|T - \pi}{2\alpha}\right) & (1 - \alpha)\frac{1}{2T} < |f| \leq (1 + \alpha)\frac{1}{2T} \\ 0 & |f| > (1 + \alpha)\frac{1}{2T} \end{cases} \quad (5.8)$$

Definitionen indeholder parameteren α , roll-off-faktoren, som giver den **ekstra båndbredde** ($0 \leq \alpha \leq 1$) der bruges ud over $1/(2T)$ fra Nyquists 1. kriterium. $X(f)$ og den tilsvarende tidsfunktion er vist for $\alpha = 0, 25, 0, 5$ og 1 på Figur 5.16. Formeludtrykket for tidsfunktionen er

$$x(t) = c \cdot \text{sinc} \left(\frac{t}{T} \right) \frac{\cos(\alpha\pi t/T)}{1 - 4\alpha^2 t^2/T^2} \quad (5.9)$$

Parameteren α gives somme tider som en procent. Det maksimale roll-off, $\alpha = 1$ eller 100%, giver den dobbelte båndbredde sammenlignet med det ideelle lavpasfilter i Nyquists 1. kriterium ((5.7) svarende til roll-off $\alpha = 0$). For sinus-roll-off funktionerne går $x(t)$ også til $\pm\infty$, men de aftager hurtigere end for det ideelle filter så en tilnærmelse med afskæring i en vis længde virker langt bedre.

Ovennævnte Nyquist-karakteristikker kan ikke anvendes direkte som pulsformer og modtagefiltre, da de udtrykker den samlede overføringsfunktion $X(f)$ i systemet, mens modtagefilteret skal være tilpasset pulsformen. Har den transmitterede pulsform $\mathbf{g}_T = (g_0^{(T)}, g_1^{(T)}, \dots, g_{L-1}^{(T)})$ den Fouriertransformerede $G_T(f)$, jf. (3.6), vil den transformerede, $G_M(f)$, af



Figur 5.16: Cosinus-roll-off som tidsfunktion $x(t)$ og spektrum $X(f)$.

det tilpassede filter $\mathbf{g}_M = (g_0^{(M)}, g_1^{(M)}, \dots, g_{L-1}^{(M)})$ i modtageren blive

$$\begin{aligned} G_M(f) &= \sum_{i=0}^{L-1} T_S g_i^{(M)} e^{-j2\pi i f T_S} = \sum_{i=0}^{L-1} T_S g_{L-1-i}^{(T)} e^{-j2\pi i f T_S} \\ &= e^{-j2\pi(L-1)fT_S} \sum_{i=0}^{L-1} T_S g_{L-1-i}^{(T)} e^{j2\pi(L-1-i)fT_S} = e^{-j2\pi(L-1)fT_S} G_T^*(f) \end{aligned}$$

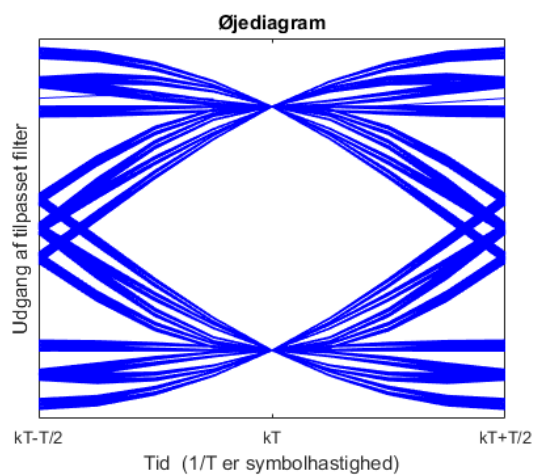
og den samlede karakteristik

$$\begin{aligned} X(f) &= G_T(f) G_M(f) = G_T(f) e^{-j2\pi(L-1)fT_S} G_T^*(f) = e^{-j2\pi(L-1)fT_S} |G_T(f)|^2 \\ &\Rightarrow |G_T(f)| = |G_M(f)| = \sqrt{|X(f)|} \end{aligned} \quad (5.10)$$

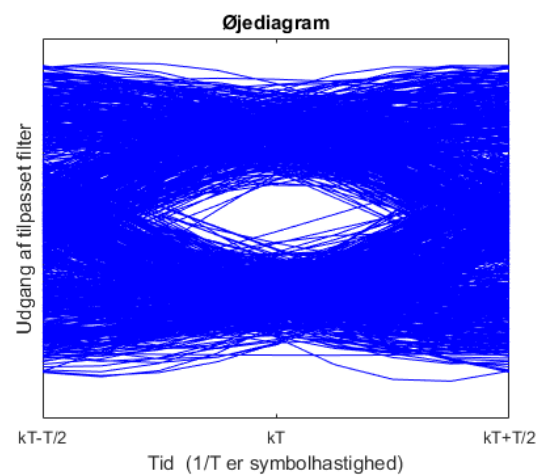
Pulsformen og modtagefilteret skal altså have amplitudespektrum som kvadratroden af $|X(f)|$ for at modtagelsen i støj kan fungere optimalt. Man skal huske på at det er i frekvensdomænet at det er kvadratroden og det er langt fra det samme i tidsdomænet. For eksempel er tidsfunktionen for kvadratroden af cosinus roll-off karakteristikken

$$rc(T) = \sqrt{\frac{c}{T}} \cdot \frac{4\alpha}{\pi} \cdot \frac{\cos \frac{(1+\alpha)\pi t}{T} + \frac{\pi(1-\alpha)}{4\alpha} \operatorname{sinc} \frac{(1-\alpha)t}{T}}{1 - 16\alpha^2 t^2 / T^2} \quad (5.11)$$

I praksis anvendes ofte en simpel metode til visualisering af transmissionssystemer med intersymbolinterferens og støj, det såkaldte **øjediagram**. Dette viser alle mulige forløb af signalet set over én eller nogle få symboltider (T). Et eksempel er vist på figurerne 5.17 og 5.18, hvor der ses et perfekt forløb af et signal og et signal med støj og intersymbolinterferens. Signalet betragtes på udgangen af det tilpassede filter over én symboltid. Det er nemt at se hvor man skal sample signalet for at få den bedst mulige adskillelse af de to amplituder og man kan se hvordan øjet lukker når der er problemer. Signalerne i figurerne er genereret i MATLAB ved anvendelse af `eyediagram`-funktionen, men i et realiseret system vil man opnå samme billede ved anvendelse af et oscilloskop, hvor der tegnes et nyt forløb over skærmen for hvert T og skærmen husker de gamle forløb. Dette anvendes ofte i eksperimentelle og professionelle systemer.



Figur 5.17: Øjediagram for signal uden støj og intersymbolinterferens.



Figur 5.18: Øjediagram for signal med både støj og intersymbolinterferens.

KAPITEL 6

Modulation

I dette kursus vil vi anvende ordet **modulation** om den proces der flytter frekvensen for et transmitteret signal op til en eller anden frekvens der passer til transmissionsmediet. Mange andre steder anvendes betegnelsen modulation i en lidt bredere betydning, fx i betegnelsen PAM. Det mest almindeligt forekommende er flytning til frekvenser hvor radiokommunikation er mulig. Flytning til endnu højere frekvenser forekommer i avanceret optisk kommunikation, såkaldt kohærent optisk kommunikation. Vi kommer tilbage til de mulige frekvensområder i Kapitel 7, men vi kender alle FM radiostationer der anvender området 87 MHz til 108 MHz, og mange andre er nok også kendte.

For at kunne modulere er det en rimelig antagelse at man skal skaffe sig et signal med en frekvens i nærheden af det man ønsker. Det foregår ved at have en såkaldt **bærebølge** med **bærefrekvensen** f_c :

$$v(t) = A \cdot \cos(2\pi f_c t + \phi) \quad (6.1)$$

Hvis vi kalder det informationsbærende basisbåndssignal - analogt eller digitalt - for $x(t)$ kan vi tænke os flere metoder til **modulation af bærebølgen med $x(t)$** :

- Signalet $x(t)$ kan ændre amplituden A , simplest $A = ax(t)$, hvor a er en konstant. Dette kaldes **amplitudemodulation**, **AM**. Vi ser mere på denne modulationsform nedenfor. I sin simpleste form optager AM en båndbredde omkring frekvensen f_c der er det dobbelte af båndbredden for basisbåndssignalet $x(t)$.
- Signalet $x(t)$ kan ændre frekvensen, simplest $f = f_c + ax(t)$, hvor a er en konstant. Dette kaldes **frekvensmodulation**, **FM**. Som antydnet ovenfor er det kendt af enhver radiolytter. Der er en ulineær sammenhæng mellem $x(t)$ og det resulterende $v(t)$, hvilket gør det uhyre svært at udtrykke noget om FM ved simple metoder. I forhold til AM anvendes en betydeligt større båndbredde, men til gengæld er systemet ret immunt over for støj. Vi vil ikke gå nærmere ind på FM.
- Signalet $x(t)$ kan ændre fasen ϕ , simplest $\phi = ax(t)$. Dette kaldes **fasemodulation**, **PM** (efter engelsk *phase modulation*). I almindelig brug kendes systemet kun for digitale signaler, som vi skal se nedenfor.

I de følgende afsnit vil vi først se på amplitudemodulation med et analogt signal og dernæst betragte modulation med digitale signaler.

6.1 Analog amplitudemodulation

Som ovenfor indeholder basisbåndsignalet $x(t)$ den information vi vil transmittere og vi antager at det tilhørende spektrum er $X(f)$. Signalet flyttes til et højere frekvensområde ved at multiplicere det med bærebolegen:

$$v(t) = x(t) \cdot \sqrt{2} \cdot \cos(2\pi f_c t + \phi) \quad (6.2)$$

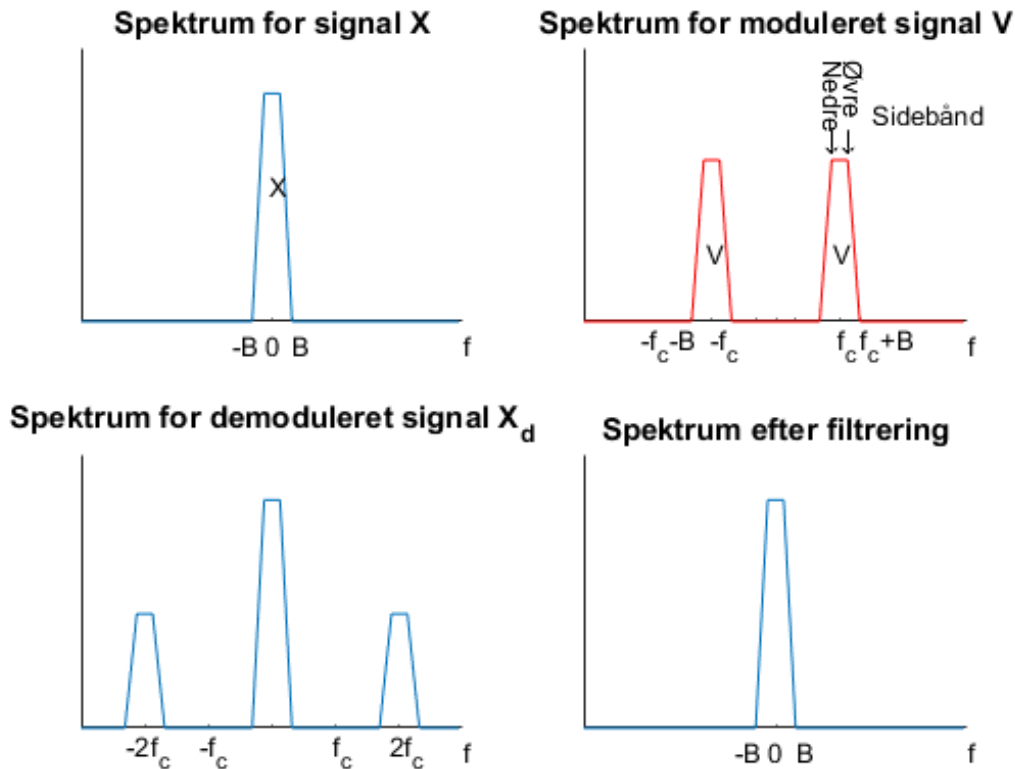
Denne proces kaldes også **blanding**. Faktoren foran cosinus er indført for at beholde signalets størrelse ved den modsatte proces: demodulation (se nedenfor). Det er ikke helt nemt at lave elektroniske eller optiske komponenter til en sådan modulation, men matematisk virker det meget simpelt og i digital signalbehandling kan det også gøres simpelt, hvis f_c er moderat. Ved anvendelse af Eulers formel kan vi omskrive til

$$v(t) = \frac{\sqrt{2}}{2} e^{j\phi} x(t) e^{j2\pi f_c t} + \frac{\sqrt{2}}{2} e^{-j\phi} x(t) e^{-j2\pi f_c t}$$

som gør det nemt at finde signalets spektrum (diverse regneregler for Fouriertransformation anvendes, men det er ikke så interessant):

$$V(f) = \frac{\sqrt{2}}{2} e^{j\phi} X(f - f_c) + \frac{\sqrt{2}}{2} e^{-j\phi} X(f + f_c) \quad (6.3)$$

Hvis $X(f)$ er 0 for frekvenser over en vis størrelse B (dvs. båndbegrænset) og $B < f_c$, vil første led kun have positive frekvenskomponenter, og det er et skift af det originale spektrum op til frekvensområdet $f_c - B$ til $f_c + B$. Båndbredden er altså nu $2B$, dobbelt så stor som den er for basisbåndsignalet $x(t)$. Det synes umiddelbart at være et spild af båndbredde, men som vi skal se nedenfor kan man udnytte det dobbelte bånd til at transmittere to uafhængige signaler. Den del af spektret der er over f_c kaldes det **øvre sidebånd** og den anden del kaldes tilsvarende det **nedre sidebånd**. Det andet led som kun findes på den negative del af akse er den kompleks konjugerede af den positive del og har den samme fysiske betydning. Læg mærke til at det her får betydning at vi indførte negative frekvenser i Kapitel 3, da det giver en nem forklaring af sidebåndene. Et eksempel på modulation og den omvendte proces, demodulation (se nedenfor), er vist på Figur 6.1.



Figur 6.1: Spektrum for AM-moduleret signal og det tilsvarende demodulerede signal.

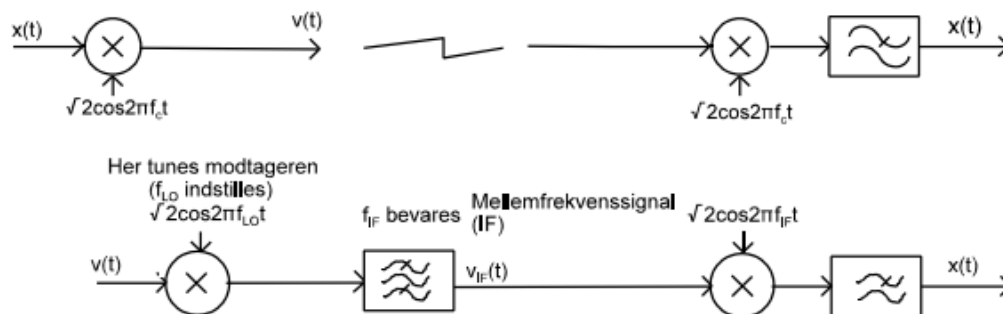
For at gendanne det oprindelige basisbåndsignal $x(t)$ skal $v(t)$ **demoduleres**. Det gøres ved at gentage multiplikationen af et signal, nu $v(t)$, og bærebolegen. For at se hvordan det virker, kan vi udregne resultatet $x_d(t)$ af denne multiplikation

$$\begin{aligned} x_d(t) &= v(t) \cdot \sqrt{2} \cos(2\pi f_c t + \phi_d) = x(t) \cdot 2 \cos(2\pi f_c t + \phi) \cdot \cos(2\pi f_c t + \phi_d) \\ &= x(t) \cdot \cos(\phi - \phi_d) + x(t) \cdot \cos(2\pi 2f_c t + \phi + \phi_d) \end{aligned}$$

hvor vi har taget hensyn til at bærebolegen i demodulatoren ikke nødvendigvis har samme fase, ϕ_d , som den i modulatorens (ϕ). Vi ser at vi får det oprindelige basisbåndsignal $x(t)$ multipliceret med en faktor samt en komponent ved den dobbelte frekvens $2f_c$. Det kan man fx nemt se ved at regne i frekvensdomænet. Vi bruger modulationsformlen (6.3) igen:

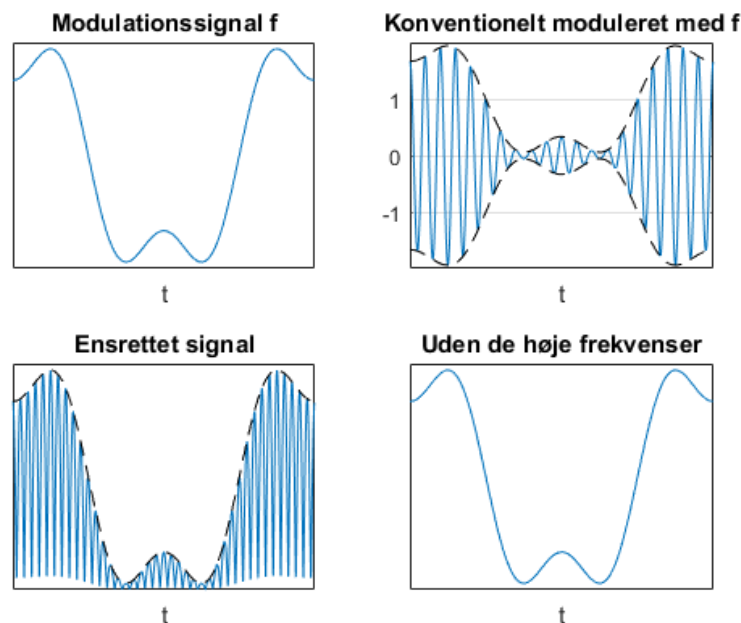
$$\begin{aligned} X_d(f) &= \frac{\sqrt{2}}{2} e^{j\phi_d} V(f - f_c) + \frac{\sqrt{2}}{2} e^{-j\phi_d} V(f + f_c) \\ &= \frac{1}{2} e^{j(\phi_d + \phi)} X(f - 2f_c) + \frac{1}{2} X(f) (e^{j(\phi_d - \phi)} + e^{-j(\phi_d - \phi)}) + \frac{1}{2} e^{-j(\phi_d + \phi)} X(f + 2f_c) \\ &= \cos(\phi_d - \phi) X(f) + \frac{1}{2} e^{j(\phi_d + \phi)} X(f - 2f_c) + \frac{1}{2} e^{-j2(\phi_d + \phi)} X(f + 2f_c) \end{aligned}$$

og vi ser igen at vi får det oprindelige basisbåndsignal samt en komponent ved den dobbelte frekvens $2f_c$. Vi viser spektrene på Figur 6.1. Denne sidste komponent fjernes med et lavpasfilter. I virkelige modtagere kan demodulationen godt være implementeret forskelligt fra modulation. I demodulatoren er der behov for på en eller anden måde at finde bærefrekvensen f_c ud fra signalet (bærebølgesynkronisering). Det er ikke nok med frekvensen, også $\phi_d - \phi$ skal være 0 eller i hvert fald en konstant så man har styr på faktoren på $x(t)$. Demodulationen foregår oftest i flere trin. Dette såkaldte superheterodynprincip er vist i Figur 6.2. Først skabes en mellemfrekvens (IF) ved en multiplikation af det modtagne signal med et lokalt signal med frekvensen f_{LO} og udfiltrering (med det viste båndpasfilter) af mellemfrekvenssignalet. Mellemfrekvensen, $f_{IF} = f_{LO} - f_C$, vælges uafhængigt af den oprindelige modulationsfrekvens f_C , da flere forskellige signaler så vil kunne demoduleres med det samme kredsløb, den egentlige demodulator. Denne udfører de mere komplicerede operationer såsom at holde styr på faseforskellen og den helt præcise frekvens for senderen (f_{IF} vil variere med f_C og upræcis f_{LO} , men det er kun små afvigelser). Dette er fx smart ved radiomodtagning, hvor tunerens, f_{LO} , og det efterfølgende filter udvælger det signal der er af interesse. Demodulatoren og eventuelle forstærkere før denne skal arbejde ved en noget lavere frekvens end hvis demodulationen foregik direkte og de er derfor nemmere at implementere.



Figur 6.2: Modulation og demodulation. Superheterodynmodtager. Boksene med filtre er beskrevet i teksten.

Amplitudemodulation anvendes i forskellige versioner. Den såkaldte **konventionel AM** anvender i stedet for $ax(t)$ multiplikationsfaktoren $1 + ax(t)$, hvilket tilføjer en periodisk komponent med frekvensen f_c til signalet, og dette gør det nemt for modtageren at finde demodulationsfrekvensen. Denne form for AM anvendes i de almindelige AM radiostationer. Demodulationen kan gøres meget enkelt, da $1 + ax(t)$ altid er en positiv størrelse (a kan vælges passende så $|ax(t)| < 1$), og ensretter man signalet og derefter lavpasfilterer det, fås et signal der følger $1 + ax(t)$ og fratrækning af jævnspændingen 1 giver $ax(t)$ uden at man behøver at kende de præcise faser for ϕ og ϕ_d . Dette illustreres på Figur 6.3.



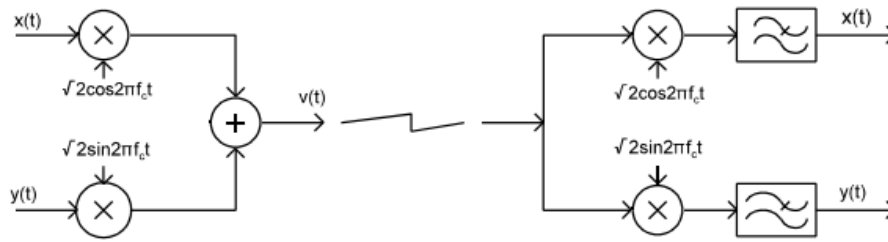
Figur 6.3: Spektrum for AM-moduleret signal og det tilsvarende demodulerede signal.

I modsætning til konventionel AM har den i (6.2) viste form ikke en periodisk komponent og den kaldes derfor **AM med undertrykt bærebølge** eller med engelske forkortelser **AM-DSB-SC** (dvs. *AM, double sideband, suppressed carrier*). I nogle vigtige anvendelser fjerner man det ene sidebånd ved filtrering og får så **enkeltsidebånds AM** med betegnelsen **AM-SSB**. Informationen i de to sidebånd er jo det samme da de stammer fra de samme frekvenser i basisbåndsignalet. AM-SSB anvendes, hvor det er vigtigt at have plads til mange forskellige kanaler, fx i lidt ældre kabelsystemer til overførsel af mange telefonsamtaler (**frekvensmultiplexning**, *FDM*). Her er kanalerne placeret med en frekvensafstand på 4 kHz og mindre kan det næsten ikke være, når telefonkanaler er båndbegrænsede til 3,4 kHz. AM-SSB findes også i visse radiokommunikationssystemer og generelt for sådanne systemer gælder det at det kan være ganske svært at finde den rigtige frekvens til demodulation.

6.1.1 Kvadraturmodulation

Hvis vi har to signaler, $x(t)$ og $y(t)$, kan vi lade hver af dem modulere en bærebølge, begge med frekvensen f_c , men forskudt $\pi/2$ eller anderledes udtrykt: cosinus og sinus med samme frekvens. Dette udvider ikke frekvensområdet, da spektrene for hvert af de to bidrag optager de samme frekvenser:

$$v(t) = x(t) \cdot \sqrt{2} \cos 2\pi f_c t + y(t) \cdot \sqrt{2} \sin 2\pi f_c t \quad (6.4)$$



Figur 6.4: Kvadraturmodulation.

I modtageren behandler vi signalet i to parallelle grene med de pågældende bærebølger. Negligerer vi eventuelle faseforskelle for at gøre det lidt enklere, får vi i de to grene $i(t)$ og $q(t)$:

$$\begin{aligned}
 i(t) &= v(t) \cdot \sqrt{2} \cos 2\pi f_c t = x(t) \cdot 2 \cos^2 2\pi f_c t + y(t) \cdot 2 \sin 2\pi f_c t \cdot \cos 2\pi f_c t \\
 &= x(t) + x(t) \cdot \cos 4\pi f_c t + y(t) \cdot \sin 4\pi f_c t \\
 q(t) &= v(t) \cdot \sqrt{2} \sin 2\pi f_c t = x(t) \cdot 2 \cos 2\pi f_c t \cdot \sin 2\pi f_c t + y(t) \cdot 2 \sin^2 2\pi f_c t \\
 &= y(t) - y(t) \cdot \cos 4\pi f_c t + x(t) \cdot \sin 4\pi f_c t
 \end{aligned}$$

Efter lavpasfiltering fås de to originale signaler $x(t)$ og $y(t)$. Dette system vises på Figur 6.4. Denne form for amplitudemodulation kaldes **kvadraturmodulation** og de to signaler kaldes på engelsk ofte *In-phase and Quadrature (**I** and **Q**) components*. Ofte anvendes betegnelsen **QAM**, *Quadrature Amplitude Modulation* om det viste system. Man kan altså transmittere to signaler med én bærefrekvens og på den måde udnytter man den dobbelte båndbredde i det modulerede signal. Modulationen af farvesignalet i det analoge TV-system var et vigtigt eksempel på en analog anvendelse af dette og vi skal senere se at det er et meget vigtigt begreb i digital modulation.

6.2 Digital modulation

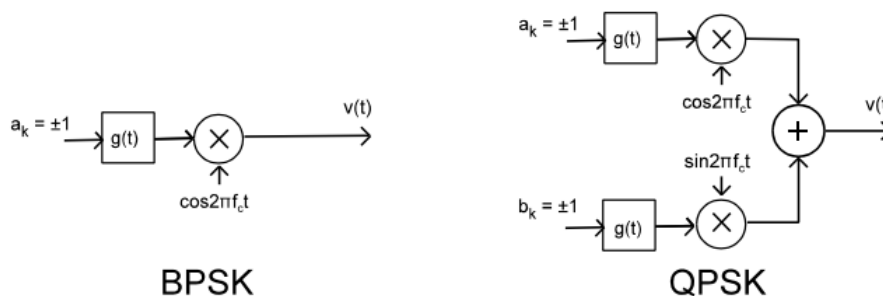
I introduktionen til de tre modulationsformer, **AM**, **FM** og **PM** skelnede vi ikke om signalet $x(t)$ var et almindeligt analogt signal eller et basisbåndsignal der indeholdt digital information i lighed med PAM signalerne fra Kapitel 5. Vi vil nu beskæftige os nærmere med de sidste under navnet digital modulation. Der har været tradition for at have nogle lidt afvigende betegnelser for de digitalt modulerede signaler: **ASK**, **FSK** og **PSK**, hvor SK står for *shift keying* idet *keying* traditionelt har været anvendt om repræsentation af digitale signaler. A, F og P er som tidligere for amplitude, frekvens og fase (*phase*). I det efterfølgende vil vi se at der måske ikke er så meget grund til at skelne mellem disse forskellige typer og vi vil vise at kvadraturmodulation anvendt rigtigt også kan omfatte PSK.

I Kapitel 5 diskutererede vi basisbåndsignaler bestående af en følge af pulser multiplie-

rede med symbolværdier (PAM). Anvender vi et sådant signal hvor tiden mellem symbolerne, symboltiden, er T , med det simplest mulige alfabet $a_k = \pm 1$ og pulsen $g(t)$ som signal for en amplitudemodulator, fås et 2ASK signal:

$$\sum_{k=-\infty}^{\infty} g(t - kT) a_k \cos 2\pi f_c t = \sum_{k=-\infty}^{\infty} g(t - kT) \cos(2\pi f_c t + \phi_k), \quad \phi_k = 0 \text{ eller } \pi \quad (6.5)$$

hvor den sidste omskrivning viser at det lige så godt - og oftest - kan kaldes **Binary Phase Shift Keying, BPSK**, idet det kan beskrives som at symbolerne skifter fasen i det modulerede signal med π . Det er ofte tilfældet at et digitalt modulationssystem kan fortolkes som forskellige analoge modulationssystemer og det giver ofte anledning til nogen forvirring, idet forskellige betegnelser kan referere til et og samme system. Vi viser BPSK-modulatoren på Figur 6.5.



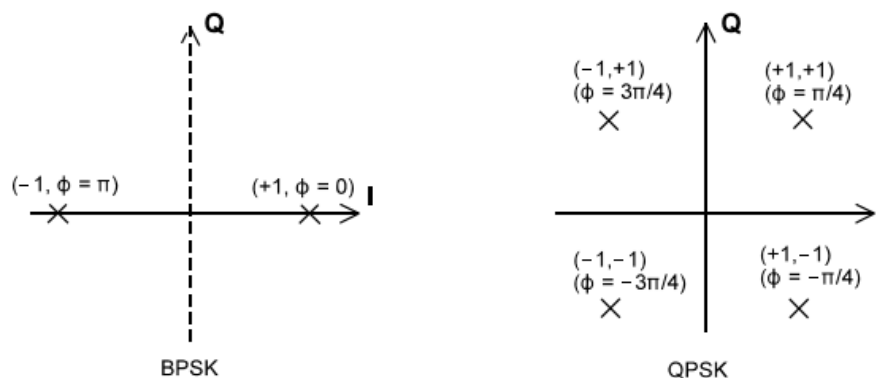
Figur 6.5: BPSK og QPSK modulation.

6.2.1 To-dimensionale modulationssystemer

Når vi bruger to digitale signaler, a_k og b_k , og en kvadraturmodulator i stedet for den enkelte modulator ovenfor, kan vi frembringe et **to-dimensionalt signalsæt** i et bestemt frekvensbånd. Vi anvender samme pulsforn, $g(t)$, og samme T i de to grene (I og Q som før) og får da

$$v(t) = \sum_{k=-\infty}^{\infty} g(t - kT) (a_k \cos 2\pi f_c t + b_k \sin 2\pi f_c t) \quad (6.6)$$

Det simplest eksempel er **QPSK, Quadrature Phase Shift Keying**, som fremkommer når de to sekvenser er binære ± 1 symboler. Vi viser QPSK-modulatoren på Figur 6.5.



Figur 6.6: BPSK og QPSK som 2-dimensionale konstellationer. Fasen ϕ forklares i teksten.

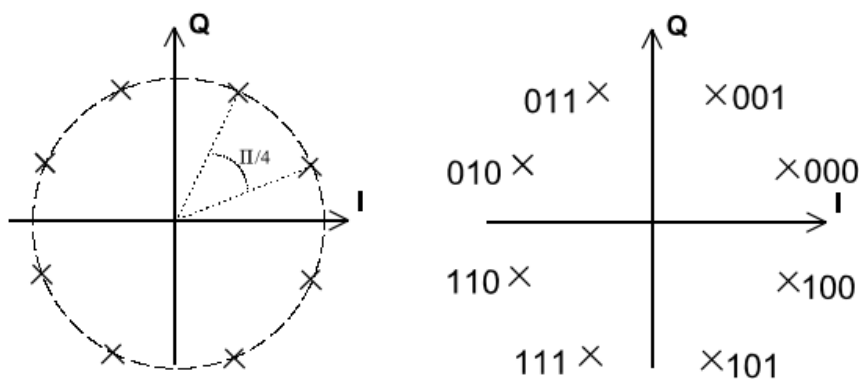
Nedenfor vil vi give nogle mere avancerede eksempler, hvor der ikke altid er uafhængighed mellem a - og b -værdierne, men hvor hvert par (a_k, b_k) repræsenterer en vis mængde information. Hvis vi underforstår selve modulationen og kun beskriver signalsættene i de såkaldte **konstellationsdiagrammer**, dvs. (a_k, b_k) placeret i to dimensioner, får man BPSK- og QPSK-systemerne som vist på Figur 6.6 med $a_k = \pm 1$ og $b_k = \pm 1$.

Flere punkter kunne placeres på en cirkel og sådanne systemer kaldes generelt **PSK**, *phase shift keying*. Fx **8PSK**, hvor de 8 punkter fordeles jævnt på en cirkel og derfor nemt fortolkes som 8 forskellige faser. 8PSK anvendes fx i den nyeste datatransmissionsstandard i GSM, EDGE, hvor hvert tidsinterval T transmitterer 3 bit, hvilket er en øgning i forhold til den almindelige standard, GMSK (se senere), som kun har 1 bit pr. tidsinterval. BPSK er 2PSK i denne beskrivelse som vi så i 6.5, og QPSK er 4PSK, idet de 4 faser er $\phi = \pi/4$, $3\pi/4$, $-3\pi/4$ og $-\pi/4$:

$$\sqrt{2} \cos(2\pi f_c t + \phi) = \sqrt{2} \cos \phi \cdot \cos 2\pi f_c t + \sqrt{2} \sin \phi \cdot \sin 2\pi f_c t$$

og de fire faser giver netop ± 1 i faktorerne foran bærebølgerne. Somme tider er QPSK med faserne 0 , $\pi/2$, π og $3\pi/2$, som frembringes af en anden form for modulator. Det betyder ikke noget for systemets egenskaber. De 3 nævnte PSK-systemer er vist på Figur 6.6 og 6.7.

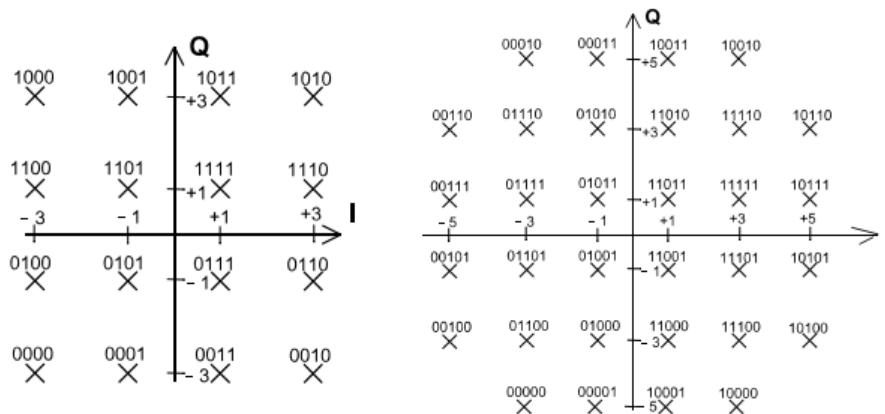
Konstellationspunkterne i PSK-systemer skal tilknyttes binære symboler for almindelig informationsudveksling. Ligesom for PAM systemer vil det være smart hvis kun én binær fejl indtræffer hvis man fejler til det nærmeste nabopunkt, og som vi skal se længere nede afhænger sandsynligheden for fejl af afstanden, så fejl til de umiddelbare nabopunkter er mest hyppige. Gray-kodning kan også her anvendes til dette. Her kan man udnytte at Gray-koden er cirkulær så der også kun er én fejl ved overgangen fra den ene ende til den anden, jf. Figur 5.3. Man kan altså anvende 00, 01, 11 og 10 rundt på cirklen for QPSK



Figur 6.7: 8PSK som 2-dimensional constellation med Gray-kodning for binære symboler.

og de på Figur 6.7 viste for 8PSK.

En anden placeringsmulighed er en gitterstruktur. Disse kaldes som regel **QAM** - *Quadrature Amplitude Modulation* eller i nogle systemer **QASK**. Som regel er der 2^b punkter. Hvis b er et lige tal fås en kvadratisk constellation som vist for **16QAM** på Figur 6.8, hvor der anvendes 4 mulige symboler i hver gren (± 1 og ± 3 , 4PAM). Læg i øvrigt mærke til at QPSK på Figur 6.6 også kan opfattes som 4QAM. For ulige b er der flere mulige constellationer og Figur 6.8 viser et muligt 32QAM-system.



Figur 6.8: Konstellationsdiagrammer for 16QAM og et muligt 32QAM-system.

Systemer med virkelig mange punkter findes. Fx op til 2^{14} i en vis del af det velkendte ADSL-system. Konstellationspunkterne skal også her tilknyttes binære symboler for almindelig informationsudveksling og det gælder om at minimere de binære fejl som indtræffer hvis man fejler til det nærmeste nabopunkt. For QAM-systemer er det mere indviklet end for PSK da der er op til fire nærmeste naboer og fire andre der er $\sqrt{2}$ læn-

gere væk. Ved almindelige signal/støjforhold kan man se bort fra disse sidste, så man har kun de op til fire nærmeste naboer at tage hensyn til. Hvis konstellationen er kvadratisk (ovennævnte b lige) er det nemt, da man kan anvende en $b/2$ Gray-kode uafhængigt i hver af de to dimensioner som vist for 16QAM på Figur 6.8. For ikke-kvadratiske konstellationer, fx 32PAM, er det umuligt, men selvfølgelig kan man finde en fordeling der minimerer antallet af binære fejl. En sådan er vist for 32QAM-systemet på figuren.

Formlerne for symbolhastighed, (5.2), og bitrate, (5.3), gælder selvfølgelig også for de to-dimensionale systemer. Da der transmitteres ét to-dimensionalt symbol (a_k, b_k) for hver symboltid T fås:

$$\text{symbolhastighed} = \frac{1}{T} \text{ baud}$$

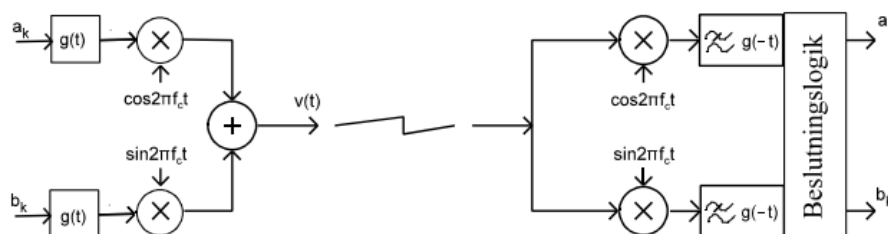
og signaleringshastigheden (bitraten) bliver:

$$\text{Signaleringshastighed eller bitrate } R = \frac{1}{T} \cdot \log_2 M \text{ bit/s}$$

De to-dimensionale systemer har grundlæggende spektre som de én-dimensionale spektre de er sammensat af. Imidlertid kan der være en vis afhængighed mellem de to grene som kan påvirke spektret og hvis signalerne er deterministiske kan man få næsten hvilket som helst spektrum inden for den båndbredde de én-dimensionale systemer har. Er signalerne derimod stokastiske og uafhængige af hinanden i de to grene kan de én-dimensionale spektre adderes. Et 16QAM system vil altså få et spektrum ganske som vist i Figur 6.1, hvis signalerne på udgangen af g -filtrene har et sådant spektrum.

6.2.2 Modtagelse af kvadraturmodulerede signaler

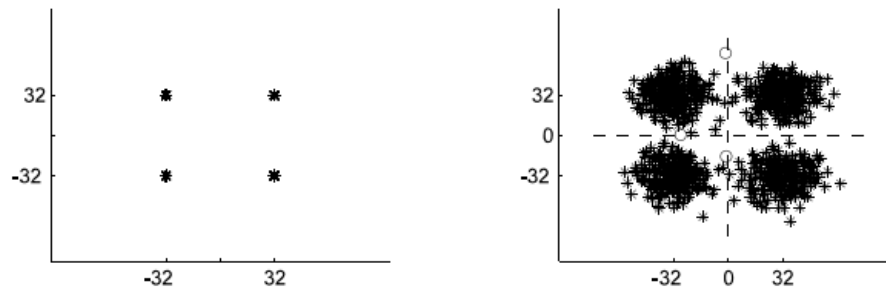
En modtager for de digitale, kvadraturmodulerede signaler frembringes ved at udvide systemet i Figur 6.4 med tilpassede filtre i hver gren efter lavpasfiltrene (eller kombineret med dem, da det tilpassede filter som regel også er lavpas) efterfulgt af sampling og kvantisering (A/D-omsætter) eller A/D-omsætteren efterfulgt af et digitalt, tilpasset filter. Hele systemet er vist på Figur 6.9.



Figur 6.9: Sender og modtager for kvadraturmoduleret digitalt signal.

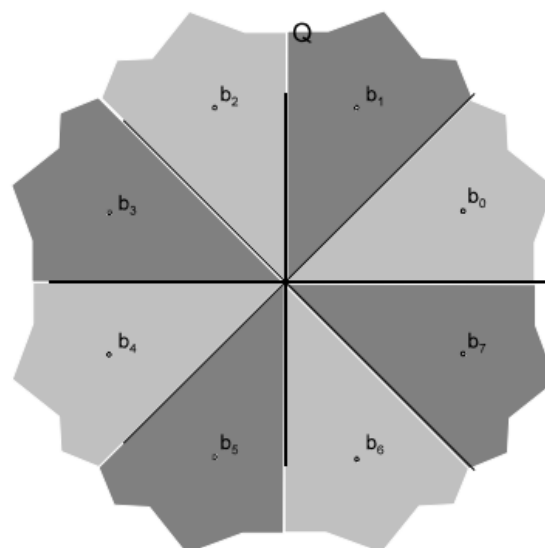
Modulation og demodulation bliver ofte realiserede i flere trin ligesom på Figur 6.2, men det ændrer ikke det overordnede princip. Støjen er som i Kapitel 5 gaussisk fordelt med

“styrken” σ (jf. (5.5)) og adderes til signalet $v(t)$. På udgangene af de to tilpassede filtre, $g(-t)$, får vi koordinaterne for et modtaget punkt i to dimensioner ligesom vi illustrerede systemerne i Figur 6.6 til Figur 6.8. Det viser sig at støjen i den to-dimensionale plan fordeler sig i en “sky” omkring de rigtige konstellationspunkter med styrken σ i hver retning som vist på Figur 6.10 for et simuleret QPSK system.



Figur 6.10: Simuleret QPSK system med og uden støj. Cirklerne viser fejlagtige beslutninger.

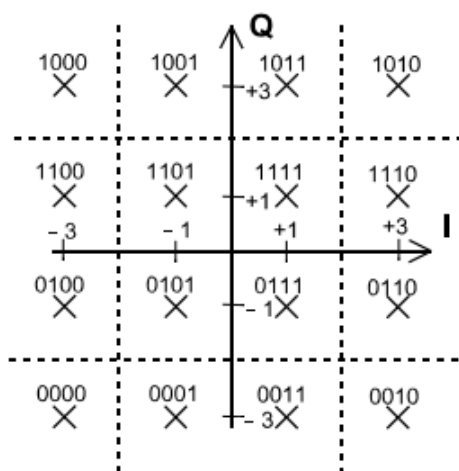
Den **optimale modtager** finder det nærmeste signalpunkt ved at udregne afstande eller ved forskellige geometriske betragtninger. Dette foregår i den ovenfor viste beslutningslogik og nedenfor ser vi på forskellige forenklinger ved simple signalkonstellationer. I én dimension i Kapitel 5 klarede vi beslutningerne om de transmitterede værdier ved at sammenligne de samlede værdier på udgangen af det tilpassede filter med nogle tærskelværdier, men her ser det en smule mere indviklet ud.



Figur 6.11: Beslutningsområder for 8PSK

Den to-dimensionale IQ-plan deles op i **beslutningsområder** hvis grænser udgøres af rette linier vinkelret på forbindelseslinien mellem nabopunkter. En illustration for 8PSK

findes på Figur 6.11 og for QPSK kan områderne ses afgrænset af de stiplede linjer i Figur 6.10. I de fleste tilfælde udgør signalpunkterne en ret regulær struktur, så det er simpelt at foretage beslutningerne. For 8PSK kan man finde fasevinklerne som $\arctan(Q/I)$. Disse er uafhængige af en fælles faktor i signalernes størrelse fra fx en forstærker; en meget stor fordel ved PSK-systemer. Det kan endda forenkles yderligere, idet I - og Q -signalets fortegn bestemmer hvilken kvadrant man er i, og man kan vælge mellem de to punkter i kvadranten ved at se på om $|I| > |Q|$, idet man i det tilfælde har modtaget punktet nærmest ved I -aksen. I 16QAM kan man beslutte for hver dimension for sig om signalet er < -2 , mellem -2 og 0 , mellem 0 og $+2$ eller $> +2$ (jf. Figur 6.8). Det giver nogle rektangulære beslutningsområder som vist på Figur 6.12.

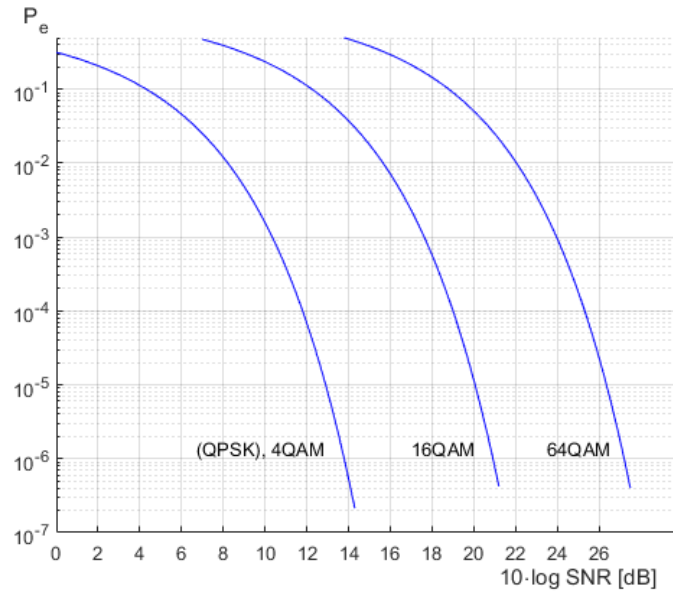


Figur 6.12: Beslutningsområder for 16QAM (afgrænset af stiplede linjer og akserne).

Dette er også tilfældet på Figur 6.10, hvor de “cirklede” modtagepunkter netop er faldet ind i et forkert beslutningsområde, idet grænserne mellem områderne følger koordinatsystemets akser gennem $(0,0)$ som vist med de stiplede linjer. Da de kvadratiske QAM-system som QPSK og 16QAM således kan ses som to uafhængige PAM-systemer og modulation/demodulation ikke ændrer fejlsandsynligheder i de to systemer (hvis man sørger for at det modulerede signal er kompenseret med $\sqrt{2}$ som tidligere skrevet) kan fejlsandsynligheder findes ud fra PAM-systemerne, Figur 5.9. Vi viser resultatet for forskellige QAM-systemer på Figur 6.13. Som på Figur 5.9 er SNR forholdet mellem den gennemsnitlige energi og $2\sigma^2$.

6.2.3 FSK, Minimum Shift Keying (MSK) og GMSK

FSK vil vi ikke gå meget ind i. Systemet anvender to frekvenser til at repræsentere binære signaler og endnu flere hvis alfabetet er større. Meget tidlige modemer for telefonlinjer (300 bit/s!) anvendte FSK, men nu er det ikke meget anvendt.



Figur 6.13: Fejlsandsynlighed for forskellige QAM-systemer som funktion af signal/støjforhold.

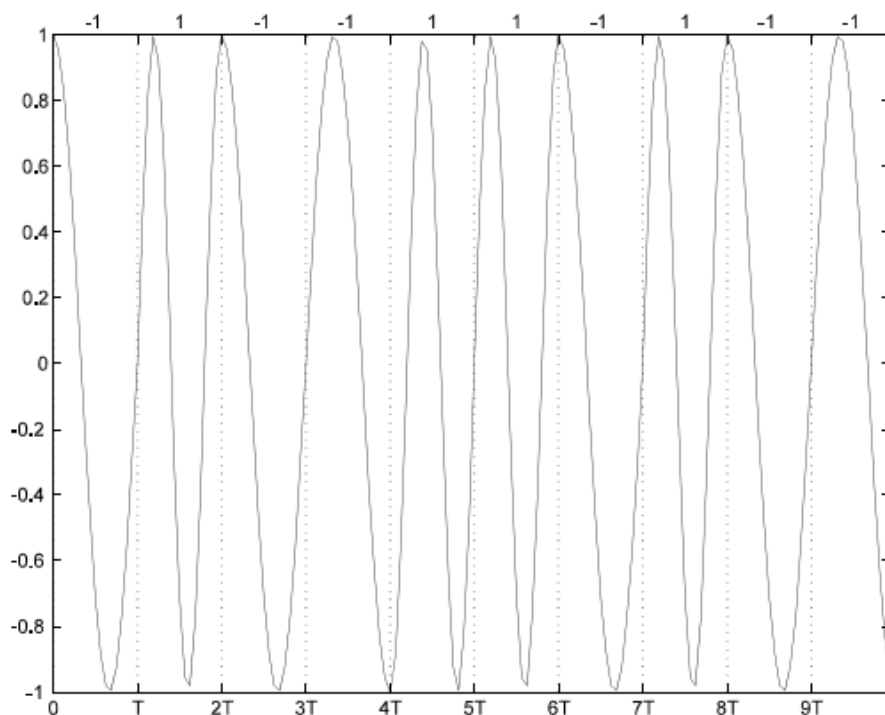
Der er et problem i FSK, når der skiftes mellem de to frekvenser, idet de i almindelighed giver en diskontinuitet i signalet, eller anderledes udtrykt, fasen springer pludseligt. Dette problem løses i **MSK** (***M**inimum **S**hift **K**eying*) modulation, hvor de to frekvenser f_{-1} og f_{+1} er tilpasset til symbolintervallet T således at deres forskel er $1/(2T)$, dvs. $f_{+1} = f_{-1} + 1/(2T)$. Hvis de to mulige signaler har samme fase ved $t = 0$, vil faserne altså adskille sig med $2\pi \cdot 1/(2T) \cdot T = \pi$ ved $t = T$, og man vælger så at fortsætte fra denne fase i næste tidsinterval således at diskontinuiteten undgås. Hvis de to frekvenser var vilkårlige ville dette ikke være muligt uden mindst et “knæk” i signalet. Systemet er altså FSK med en ret lille frekvensforskel (deraf minimum i navnet). Nogle få tidsenheder af et MSK-signal er vist på Figur 6.14 og det ses tydeligt at der ikke er fasespring. Ved hvert “symbolpunkt” fortsættes blødt med næste symbols signal.

Hvis fasen måles i forhold til middelfrekvensen $f_c = 1/2(f_{-1} + f_{+1})$ kan den siges at være 0 eller π for lige multipla af T og $\pm\pi/2$ for ulige multipla. Vi viser dette i Figur 6.15. MSK signalet for sekvensen $\{a_k\}$ ($a_k = \pm 1$) kan altså skrives som

$$v(t) = a \cos \left(2\pi f_c t + a_k \frac{\pi t - kT}{2T} + \phi_k \right), \quad kT \leq t \leq (k+1)T \quad (6.7)$$

hvor $\phi_k = \sum_{i=0}^{k-1} a_i \frac{\pi}{2}$

Læg mærke til at fasen vokser/aftager lineært ($a_k \cdot (\pi/2T) \cdot t$) i hvert tidsinterval og en

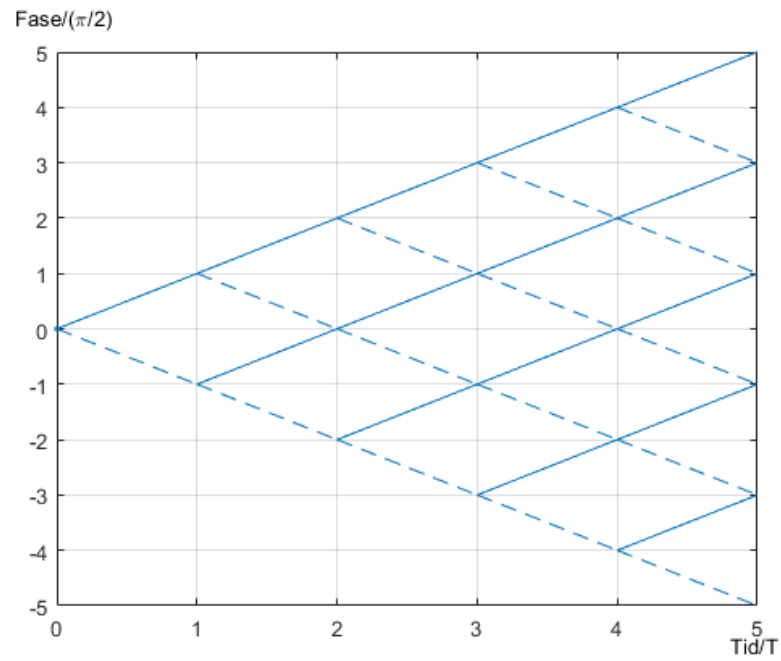


Figur 6.14: MSK-signal.

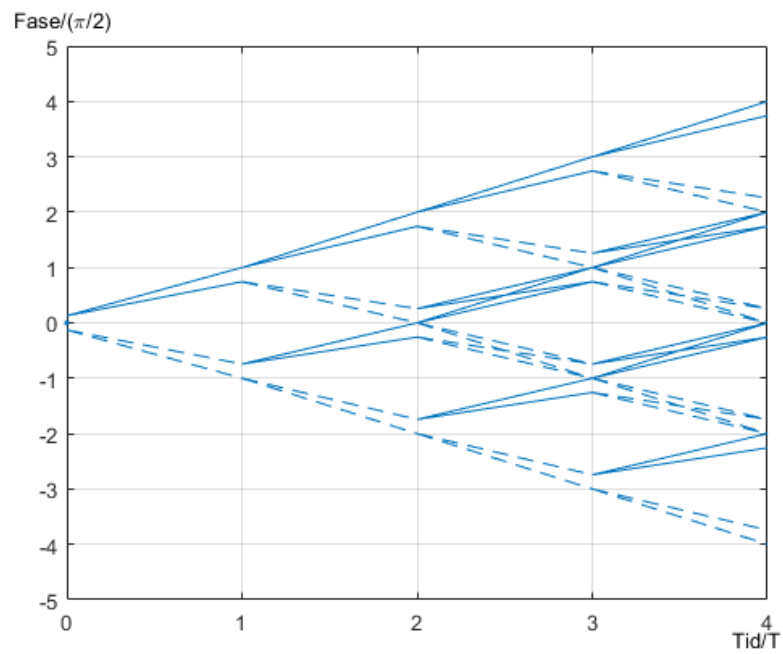
sådan lineær afhængighed i fasen er det samme som en forskellig frekvens. Fasen kan beskrives ved hjælp af et fase træ som starter med rod i $t = 0$ og så vælger mellem to grene for hvert T afhængigt af symbolet a_k (vist som fuldt optrukket for $a_k = +1$ og stiplet for $a_k = -1$). Faseskiftet i løbet af T -intervallet er lineært og i alt $\pm\pi/2$. I fase træet lader vi fasen vokse uhæmmet i stedet for som normalt at betragte fasen modulo 2π som er perioden for bære bølgen. Figur 6.15 viser et fase træ for 5 symboler i et MSK-signal.

Som man kan se på Figur 6.14 er signalets maksima og minima konstante. Man siger at indhyllingskurven er konstant. Dette har en stor betydning for forstærkere og andre komponenter i systemerne. Fejlsandsynligheden er som for BPSK, hvilket gør systemet bedre end traditionel FSK. Imidlertid bruger et MSK et bredere frekvensbånd end f.eks. QPSK med båndbegrænsede (cosinus roll-off) pulser, da hver frekvens varer T og i den forstand svarer til en firkantet puls.

Frekvensområdet kunne blive mere koncentreret hvis man udglatter faseskiftene. Et sådant system er **Gaussian MSK (GMSK)** som bliver brugt i DECT og GSM. I stedet for skiftet på $\pm\pi/2$ i løbet af tidsintervallet T bliver skiftet fordelt over $3T$ i GMSK og Gaussian stammer fra at man tilnærmer en pulsform svarende til normalfordelingens tæthedsfunktion. Tilnærmelsen har 3 værdier, så et enkelt symbol giver anledning til et faseskift på



Figur 6.15: Fasetræ for MSK signal. Fuldt optrukken linje er faseskift for kodning af +1, mens stipling er for -1.



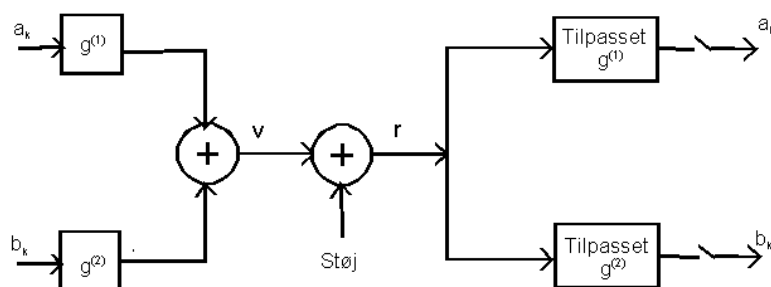
Figur 6.16: Fasetræ for GMSK signal.

$$\pm (\alpha, \pi/2 - 2\alpha, \alpha)$$

Da fasen er en sum af alle disse skift vil de mulige værdier ved $t = kT$ kunne være de originale MSK-værdier, $0, \pm\pi/2$, eller π , eller disse $\pm 2\alpha$. I DECT og GSM er afvigelserne omkring $\pm\pi/9$. I Figur 6.16 viser vi et fasetræ for GMSK med samme symbolik for ± 1 som for MSK-træet. GMSK giver et smallere spektrum, hvilket er af betydning i mobilkommunikation med mange kanaler tæt på hinanden. Imidlertid betyder den blødere overgang at der indføres en vis intersymbolinterferens så fejlsandsynligheden forøges i forhold til MSK.

6.3 Multidimensionale signaler

Ved kvadraturmodulation kunne vi transmittere to signaler uafhængigt af hinanden. Denne uafhængighed opnåede vi ved at bruge cosinus og sinus som bærebølger og derefter filtrere bidragene ved den dobbelte bærefrekvens fra, jf. Figur 6.9. Denne teknik har været anvendt i mange år, men med moderne signalbehandling kan man opnå uafhængig transmission af flere signaler på en mere direkte måde. For at illustrere dette vil vi først betragte nedenstående system



Figur 6.17: Sender og modtager for ortogonale signaler.

og stille os selv det spørgsmål: Kan det lade sig gøre at modtage a_k og b_k uafhængigt af hinanden som figuren antyder og også med støj? Det øverste tilpassede filter udregner

$$s^{(1)} = \sum_{i=0}^{L-1} r_i g_i^{(1)} T_S = \sum_{i=0}^{L-1} (a_k g_i^{(1)} + b_k g_i^{(2)} + n_i) g_i^{(1)} T_S = a_k E^{(1)} + \sum_{i=0}^{L-1} n_i g_i^{(1)} T_S$$

hvis og kun hvis

$$\sum_{i=0}^{L-1} g_i^{(1)} g_i^{(2)} = 0$$

Dette kaldes at de to pulser er **ortogonale**. Selvfølgelig kan samme udregning laves for det nederste filter for modtagelse af b_k og betingelsen for ortogonalitet er den samme. Med ortogonale signaler kan flere informationsstrømme altså dele samme transmissionsvej og støjens virkning reduceres så meget som muligt gennem de tilpassede filtre. Et eksempel er

anvendelse af **CDMA** i mobilkommunikation som beskrevet i “Introduktion til Netværk og Protokoller, Bind 2, Afsnit 11.2”. Ved at generalisere formelen for **ortogonalitet** kan vi altså nu have mere end de to informationsstrømme vi havde ved kvadraturmodulation:

$$\sum_{k=0}^{L-1} T_s g_k^{(n)} g_k^{(p)} = \begin{cases} 0 & \text{for } n \neq p \\ E^{(n)} & \text{for } n = p \end{cases} \quad (6.8)$$

skal være opfyldt for alle par (n,p) af pulser $g^{(n)}$ og $g^{(p)}$.

Eksempel 6.1 - Simple ortogonale pulser

Lad os se på pulser af længden $L = 4$. Fire meget simple ortogonale pulser kunne findes ved lade 1 (eller en simpel firkantet puls hvis vi ser på det som tidskontinuerte pulser) forskyde sig gennem alle fire placeringer: $\mathbf{g}^{(1)} = [1, 0, 0, 0]$, $\mathbf{g}^{(2)} = [0, 1, 0, 0]$, $\mathbf{g}^{(3)} = [0, 0, 1, 0]$ og $\mathbf{g}^{(4)} = [0, 0, 0, 1]$. Beslutning om hvad der modtages skal så gøres i et 4-dimensionalt rum, hvilket ikke er helt nemt at illustrere. Man kan eventuelt pålægge den restriktion at man kun må anvende én af de fire ortogonale pulser ad gangen hvilket selvfølgelig gør beslutningerne lidt simplere. Et sådant system kaldes af indlysende årsager **PPM**, **PulsPositionModulation**. Det er en dårlig udnyttelse af de fire dimensioner, men beslutninger er simple da man kun behøver at finde hvilket af de fire filtre der giver det højeste udgangssignal. Eventuelt kan man benytte $\pm \mathbf{g}^{(n)}$ og detektere det numerisk højeste signal samt fortegnet af dette. Dette kaldes “biortogonal signalering”. For begge disse simple systemer er det et problem at den transmitterede energi er koncentreret i et ret kort interval ud af den samlede symboltid T men resten af tiden er der ingenting. Det vil være meget bedre at få energien spredt mere ud og det kan man opnå ved følgende fire pulser:

$$\mathbf{g}^{(1)} = [+1, +1, +1, +1]$$

$$\mathbf{g}^{(2)} = [+1, -1, +1, -1]$$

$$\mathbf{g}^{(3)} = [+1, +1, -1, -1]$$

$$\mathbf{g}^{(4)} = [+1, -1, -1, +1]$$

Man kan ikke finde mere end fire der er indbyrdes ortogonale. (Hvorfor?)□

Hidtil har vi i kurset kun beskæftiget os med signaler som funktioner af tiden med reelle værdier, men for det efterfølgende afsnit om OFDM er vi nødt til at modificere betingelsen for ortogonalitet lidt. Man kunne også modificere mange andre formler, men vi har kun brug for denne. Lad $g^{(n)}$ og $g^{(p)}$ være pulser med komplekse værdier. Betingelsen for **ortogonalitet** er da:

$$\sum_{k=0}^{L-1} T_S g_k^{(n)} g_k^{(p)*} = \begin{cases} 0 & \text{for } n \neq p \\ E^{(n)} & \text{for } n = p \end{cases} \quad (6.9)$$

skal være opfyldt for alle par (n,p) af pulser $g^{(n)}$ og $g^{(p)}$.

Modifikationen består altså i den komplekse konjugering af ét af signalerne - det er ligegyldigt hvilket - og det er nemt at indse at det er nødvendigt, for gjorde man ikke det, ville energien E blive en kompleks størrelse i stedet for summen af energierne i den reelle og imaginære dimension.

Eksempel 6.2 - Ortogonale, komplekse harmoniske funktioner

Vi vil her vise at komplekse harmoniske funktioner fra (3.1) udgør et sæt af ortogonale funktioner hvis frekvenserne vælges på en bestemt måde. Vi vil se på disse funktioner samplet med samplingtiden T_S og længde L . Den n 'te af de udvalgte funktioner er da

$$\mathbf{g}^{(n)} = [e^{j2\pi \frac{n}{LT_S} 0T_S}, e^{j2\pi \frac{n}{LT_S} 1T_S}, \dots, e^{j2\pi \frac{n}{LT_S} (L-1)T_S}]$$

Det er nemt at se at T_S går ud i eksponenterne så det k 'te element af $\mathbf{g}^{(n)}$ er $g_k^{(n)} = e^{j2\pi \frac{n \cdot k}{L}}$. Ortogonaliteten ses af

$$\sum_{k=0}^{L-1} T_S g_k^{(n)} g_k^{(p)*} = \sum_{k=0}^{L-1} T_S e^{j2\pi \frac{n \cdot k}{L}} e^{j2\pi \frac{-p \cdot k}{L}} = \begin{cases} T_S \frac{\left(e^{j2\pi \frac{(n-p)}{L}}\right)^L - 1}{e^{j2\pi \frac{(n-p)}{L}} - 1} = 0 & \text{for } n \neq p \\ \sum_{k=0}^{L-1} T_S \cdot 1 = T_S L & \text{for } n = p \end{cases}$$

□

6.3.1 Ortogonal frekvensmultipleks, OFDM

I dette afsnit vil vi beskrive en anvendelse af multidimensionale signaler som er meget moderne, fx anvendes **OFDM** i de digitale transmissionssystemer for radio og TV, DAB og DVB og i 4. generations mobilsystemet LTE ligesom det anvendes i ADSL.

I OFDM transmitteres der i hver af L kanaler et komplekst symbol X_n ($n = 0, \dots, L-1$) i løbet af tidsenheden T . De enkelte kanaler skabes af ovennævnte komplekse, harmoniske funktioner med frekvenserne $0/T, 1/T, \dots, (L-1)/T$ og da de samplede værdier ved en samplingtid $T_S = T/L$ giver ortogonale signaler som vist i Eksempel 6.2 bliver de L kanaler uafhængige af hinanden. De kan derfor adderes som i Figur 6.17 og stadig skilles fra hinanden i modtageren. I hver tidsperiode T transmitteres altså et såkaldt OFDM-symbol som indeholder de L værdier der fremkommer af

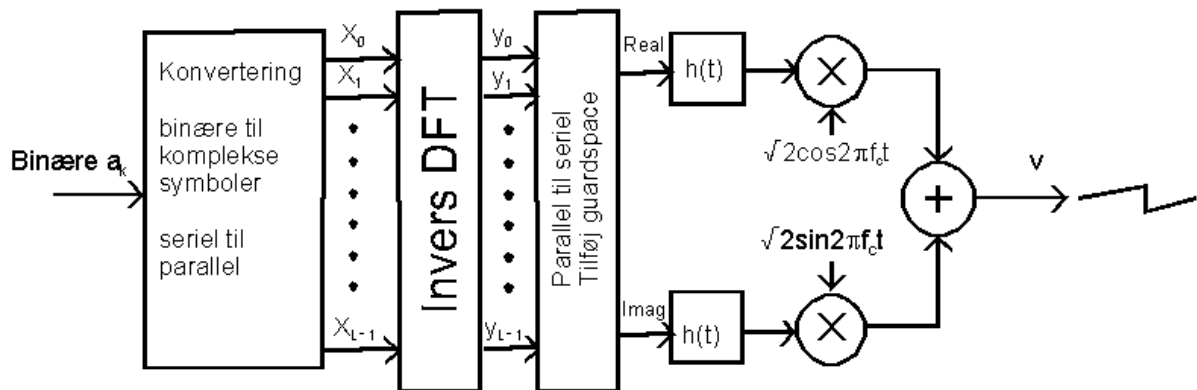
$$y_m = \frac{1}{\sqrt{L}} \sum_{n=0}^{L-1} X_n e^{j2\pi \frac{m \cdot n}{L}}, \quad \text{for } m = 0, \dots, L-1 \quad (6.10)$$

Denne udregning synes ret kompliceret, men den ses nemt at være den inverse Fourier-transformation for periodiske signaler (DFT) svarende til Fourierrækken i (3.5) og det er en af årsagerne til at OFDM er blevet meget populært da udregningen for $L = 2^{\text{potens}}$ kan udføres meget effektivt ved den såkaldte Fast-Fourier-Transform-algoritme (fft). Faktoren foran summen er indført for at skalere de komplekse, harmoniske eksponentialfunktioner således at $\sum_{k=0}^{L-1} g_k^{(n)} g_k^{(n)*} = 1$ som vist i Eksempel 6.2. På den måde bevarer symbolerne X_n deres størrelse ved modtagelse. Da $e^{j2\pi \frac{m \cdot (L-n)}{L}} = e^{j2\pi \frac{m \cdot (-n)}{L}}$ er de højeste frekvenser blot udtryk for negative frekvenser så det virkelige frekvensindhold i OFDM symbolerne er frekvenser fra 0 op til $L/(2T)$. Det viser sig i øvrigt at effektspektret er ret konstant over hele dette område.

OFDM-systemer anvendes selvfølgelig til at transmittere binære sekvenser, så der skal først ske en konvertering fra bits til de komplekse symboler X_n . Dette gøres som for almindelige konstellationer i digitale modulationssystemer, fx som QAM (Figur 6.8 hvor I-aksen er realdelen af X_n og Q-aksen den imaginære del) eller for de simpleste som QPSK, dog som regel drejet lidt så symbolerne er $X_n \in \{1, j, -1, -j\}$ som ved Gray-kodning svarer til bitparrene $\{00, 01, 11, 10\}$.

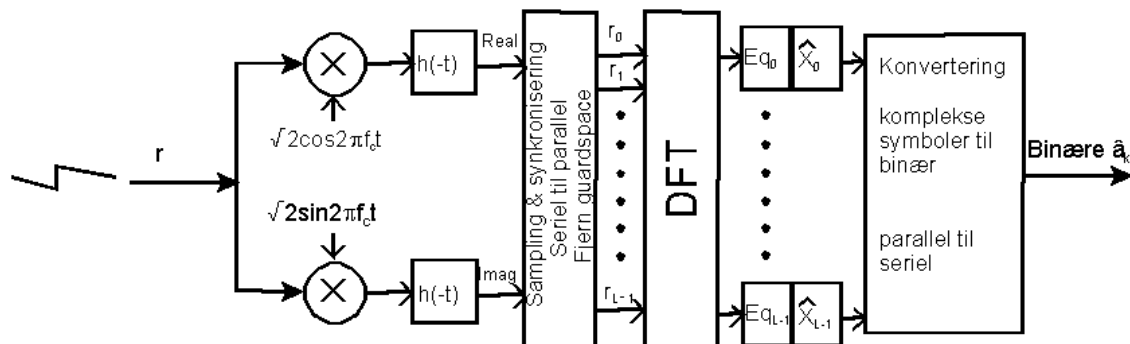
Når OFDM-symbolet transmitteres som en seriel sekvens vil transmissionskanalen forvrænge det transmitterede så varigheden af sekvensen hos modtageren er længere end T . Det skyldes at nogle transmissionskanaler som fx metalliske ledere ikke transmitterer alle frekvenser lige hurtigt eller at signalet udbreder sig ad forskellige veje mellem sender og modtager som i et radiosystem. Transmissionskanaler beskrives mere udførligt i Kapitel 7, men for at nabo-OFDM-symboler ikke skal overlappe hinanden sendes de med et vist mellemrum, det såkaldte *guardspace*, og det viser sig at modtagerens implementering forenkles hvis der til hver sekvens tilføjes et stykke af begyndelsen af OFDM-symbolsekvensen som udfyldning af mellemrummet.

Transmissionen af OFDM-symbolsekvenserne afhænger af den kanal der er til rådighed. Hvis fx kvadraturmodulation er mulig som ved radiotransmission kan real- og imaginærdelene transmitteres i hver deres kanal og man får et samlet system som vist i Figur 6.18 hvor kvadraturmodulationen svarer til Figur 6.4. Vi har forøget bæreølgerne med $\sqrt{2}$ så energien bevares efter bortfiltrering i modtageren af frekvenser omkring den dobbelte bærefrekvens, f_c , helt som i Figur 6.4. Figuren viser også filtre der danner pulsformen $h(t)$. Når der ikke anvendes kvadraturmodulation må man sikre at sekvenserne kun har reelle værdier. Det beskrives i afsnittet nedenfor om ADSL.



Figur 6.18: Sender for OFDM system med kvadraturmodulation.

Modtageren vises på Figur 6.19. Det modtagne signal, $r(t)$, er forvrænget og udsat for støj. Efter det tilpassede filter, $h(-t)$, samples signalet med samplingfrekvensen $1/(LT)$ og der synkroniseres så man finder starten på OFDM-symbolerne. Man kan så fjerne mellemrummene (*guardspace*). DFT-blokken udregner



Figur 6.19: Modtager for OFDM system med kvadraturmodulation.

$$R_n = \frac{1}{\sqrt{L}} \sum_{m=0}^{L-1} r_m e^{j2\pi \frac{m \cdot n}{L}}, \quad \text{for } n = 0, \dots, L-1$$

som er præcis svarende til modtagelsen i Eksempel 6.2. Værdierne R_n ville nu svare til de transmitterede X_n hvis der ikke var forvrængning og støj. Forvrængningen reduceres som tidligere beskrevet ved den såkaldte udligning (*equalization*) og det er meget nemmere at gøre dette hver kanal for sig ($Eq_0 \dots Eq_{L-1}$) som vist, da forvrængningen i det frekvensområde, som en enkelt kanal dækker, ikke varierer særligt meget, så man kan nøjes med at multiplicere $A_n e^{j\phi_n} R_n$ hvor A_n og ϕ_n er bestemt ved målinger af den pågældende kanals overføringsfunktion. Efter udligningen skal det mest sandsynlige symbol, \widehat{X}_n , findes

som i fx Figur 6.12.

Den store fordel ved OFDM ligger i muligheden for at tilpasse konstellationerne for X_n til de faktiske forhold ved den pågældende frekvens. Man kan så vælge at have store konstellationer hvor signal/støjforholdet er godt og mindre konstellationer ved dårlige signal/støjforhold. På denne måde får man udnyttet kanalen bedst muligt. Udligningen er som beskrevet ovenfor også enklere.

6.3.1.1 ADSL

ADSL betyder **A***symmetric* **D***igital* **S***ubscriber* **L***ine*, dvs. en transmissionslinje for telefonabonnenter der tillader en forholdsvis stor kapacitet mod brugeren og en noget mindre kapacitet den anden vej. Det er meningen at det anvendes på de almindelige telefonabonnentlinjer, så der ikke skal etableres nye kabler o.lign. ADSL findes i forskellige versioner med forskellige bithastigheder, men her tager vi det grundlæggende system (ITU-T standard G992.1). Systemet kendes også som **DMT**, Diskret Multitone Transmission, pga. opdelingen i kanaler. ADSL i sin grundform anvender OFDM med $L = 512$ og frekvensafstanden mellem kanalerne er $1/T = 4312,5$ Hz. Frekvensområdet for OFDM-symbolerne er altså 0 til $L/(2T) = 1104$ kHz. I ADSL anvendes basisbåndstransmission så kvadraturmodulation er udelukket. OFDM-symbolerne skal derfor være reelle og når man tænker på at et reelt signal ved DFT har et spektrum hvor bidrag ved positive og negative frekvenser er hinandens kompleks konjugerede er det nemt at indse at den inverse DFT giver reelle OFDM-symboler hvis $X_n, n = 0, \dots, L-1$ som funktion af en blok informationssymboler, $a_k, k = 0, \dots, L-1$ vælges som:

$$\begin{array}{ll} X_0 = a_0 & X_0 e^{jm \cdot 0 \cdot \frac{2\pi}{L}} \text{ reel} \\ X_{\frac{L}{2}} = a_{\frac{L}{2}} & X_{\frac{L}{2}} e^{jm \cdot \frac{L}{2} \cdot \frac{2\pi}{L}} = a_{\frac{L}{2}} (-1)^m \text{ reel} \\ X_k = a_k + ja_{L-k} & X_k e^{jm \cdot k \cdot \frac{2\pi}{L}} + X_{L-k} e^{jm \cdot (L-k) \cdot \frac{2\pi}{L}} = \\ X_{L-k} = a_k - ja_{L-k} & 2a_k \cos mk \frac{2\pi}{L} - 2a_{L-k} \sin mk \frac{2\pi}{L} \text{ reel} \end{array}$$

De binære informationer kodes som PAM-symboler (jf. Kapitel 5) med Gray-kodning som i Eksempel 5.1. Afhængig af forholdene i den enkelte kanal kan man have fra 4 (2 bit) til 2^{15} (15 bit) symboler. Kanal 0 bruges ikke da frekvenser omkring 0 Hz ikke kan komme gennem de transformatorer der som regel er på en abonnentlinje. Hvis ADSL forbindelsen deles med almindelig telefoni anvendes kanal 1 til 6 (og de tilsvarende negative frekvenser, kanal 511 ned til 506) ikke. I retningen fra brugeren til nettet (*upstream*) anvendes et lignende OFDM system med $L = 64$, så det optager kanalerne op til nr. 31. Fra nettet (*downstream*) anvendes så kanalerne 32 til 480. De to kommunikationsretninger muliggør at kanal anvendelsen kan ændres i løbet af kommunikationen, når dårligere eller bedre forhold indtræffer på nogle kanaler. Under alle omstændigheder sættes allokeringen af bit op ved starten af en kommunikation ud fra de på den tid herskende forhold.

Som beskrevet under OFDM anvender ADSL et *guardspace* og det er valgt til 32 symboler. Yderligere anvendes hver 69. OFDM-symbolblok til synkroniseringsformål, så det effektive antal blokke à 512 symboler er $1/T \cdot 512/(512 + 32) \cdot 68/69 = 4000$ blokke/s. Bitraten bliver altså et multiplum af 4 kbit/s. Det er selvfølgelig ikke alle kanaler der kan anvende det fulde antal (15) bit, så den virkelige største kapacitet er 768 kbit/s mod nettet og 8,4 Mbit/s mod brugeren.

KAPITEL 7

Kanalmodeller

Indtil nu har vi ikke så præcist beskæftiget os med selve det medium der bærer de transmitterede signaler. Vi har kort nævnt metalliske transmissionsledninger, radiokanaler og optiske fibre, men ikke sådan for alvor. Dette vil vi gøre i dette afsluttende kapitel. Afsnit 7.2 og 7.3 er lettere redigerede udgaver af tilsvarende afsnit i Kapitel 6 i “Digital Communication, 34230, Lecture Notes” af Jørn Justesen og Knud J. Larsen. Af tidsmæssige årsager har vi været nødt til at bevare originalsproget. Danske begreber er dog angivet i *kursiv*. Notationen er en lille smule forskellig fra den øvrige del af noterne, idet vi i dette kapitel anvender vinkelfrekvensen $\omega = 2\pi f$ som frekvensmål. Vinkelfrekvensen måles i rad/sek. Dette fjerner en del 2π i formler rundt omkring. De anvendte noter er ret detaljerede, så de bør læses med en vis overfladiskhed. I den danske del af kapitlet forekommer en del af begreberne oversat til engelsk i *kursiv*.

En **transmissionskanal** er et system som overfører et (elektrisk) indgangssignal (fx udgangssignalet fra et pulsformende filter) til en udgang hvor behandlingen af det modtagne signal starter. Der indgår således noget elektronisk udstyr, fx forstærkere, og et fysisk medie, fx metalliske ledninger eller det frie rum, og måske udstyr for at tilgå mediet som fx antenner eller lasere. Transmissionskanalen introducerer støj i systemet. Denne del af kurset har fokuseret på den signalbehandling der foregår i sender og modtager og vi ser transmissionskanalen som en omgivelse med nogle naturligt givne egenskaber. I dette kapitel vil forsøge at beskrive disse, så læseren forstår den fysiske og tekniske baggrund for dem.

Vi vil interessere os for følgende kanaler

- Metalliske transmissionsledninger (*Metallic transmission lines*)
- Optiske fiberforbindelser (*Optical transmission lines*)
- Radiokanaler (*Radio channels*)

Disse tre hovedtyper er beskrevet i afsnittene 7.1 - 7.3.

Beskrivelsen er noget forskellig fra type til type, men for nogle er overføringsfunktionen (jf. Kapitel 4) for kanalen givet:

$$C(\omega) = A(\omega)e^{j\theta(\omega)} \quad (7.1)$$

hvor $A(\omega)$ kaldes **amplitudekarakteristikken** (*amplitude response*) eller $1/A(\omega)$ **dæmpningen** (*attenuation*) og $\theta(\omega)$ **fasekarakteristikken** (*phase response*) eller $-\theta(\omega)$ **fasedrejningen**. Kanalen er ideel eller **forvrængningsfri** (*distortionless*) når $A(\omega)$ er uafhængig af frekvensen og $\theta(\omega)$ proportional med ω : $\theta(\omega) = -\omega \cdot \tau$ fordi denne overføringsfunktion viser at indgangssignalet bliver bevaret ved udgangen (dog måske skaleret med en konstant lig $A(\omega)$) og forsinket tiden τ (den tidligere nævnte “forsinkelsessætning” for Fouriertransformation).

Som det ses i det følgende, eksisterer forvrængningsfri transmission kun i meget begrænset omfang. Der er imidlertid metoder til at fjerne eller undertrykke forvrængningen så kanalen opfører sig som forvrængningsfri. Disse metoder kendes som **kanaludligning** eller *channel equalization*, men det er lidt for indviklet for dette kursus. I kurset 34230 gives en kort introduktion til emnet.

7.1 Metalliske transmissionsledninger

7.1.1 Definition og model for transmissionsledning

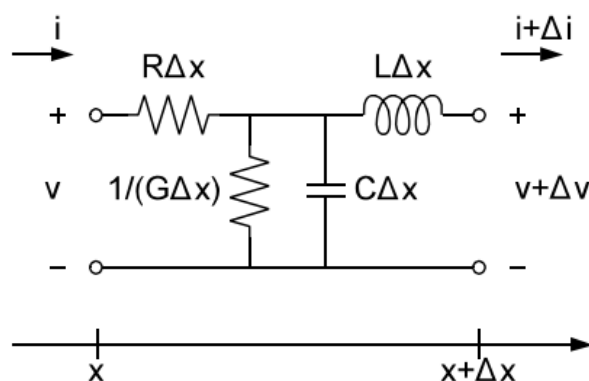
En **transmissionsledning** (*transmission line*) kan defineres præcist, men temmelig bredt, som enhver struktur, der kan lede en elektromagnetisk bølge fra ét punkt til et andet. Da elektromagnetiske fænomener beskrives ved Maxwell’s ligninger [7.1], [7.2], kan man anderledes udtrykt altså betragte en transmissionsledning som et sæt af randbetingelser, der tillader beskrivelse af en én-dimensional elektromagnetisk bølge mellem to punkter.

Almindeligvis anvendes begrebet lidt mere indskrænket. For det første kræves, at længden af transmissionsledningen er mindst en ikke forsvindende brøkdel af bølgelængden for den højeste frekvens af interesse for det system, man betragter. Er ledningen nemlig kortere, er de betragtede felter såkaldt kvasistationære, og den almindelige kredsløbsteori fra Kapitel 4 finder anvendelse. Bølgelængden behandles i Afsnit 7.1.1.2. For det andet vil vi i dette kursus udelukke de såkaldte bølgeledere fra begrebet transmissionsledninger, idet vi kræver, at alle frekvenser fra den nævnte høje frekvens og helt ned til frekvensen 0 skal kunne overføres på en transmissionsledning, og at karakteristika for ledningen varierer blødt og kontinuert over hele dette frekvensområde. Fremgangsmåden ved behandling af bølgeledere er dog ikke meget forskellig.

Vi er mest interesserede i homogene transmissionsledninger, der består af to metalliske ledere med samme egenskaber i hele ledningens udstrækning og omgivet af en god isolator (dielektrikum), der også er homogen i ledningens længderetning. Som eksempler haves parsnoede ledninger, der fx anvendes til at forbinde en telefonabonnent, og koaksialledninger, der fx anvendes til transmission af TV-signaler. Vi vil se lidt nærmere på eksempler i Afsnit 7.1.2.

7.1.1.1 Primære konstanter

I [7.1] opstilles de grundlæggende differentiaalligninger for homogene transmissionsledninger ud fra Maxwell's ligninger. Vi vil her anvende en simplere metode, der bygger på anvendelse af almindelig kredsløbsteori på et lille udsnit af ledningen. Dette kan gøres, fordi de elektriske og magnetiske felter inden for dette udsnit kan betragtes som værende kvasistationære, når udsnittet er lille nok, og dets udstrækning lader vi alligevel gå mod nul. Figur 7.1 viser en model for et lille stykke, Δx , af en ledning.



Figur 7.1: Lille stykke af transmissionsledning.

Det af strømmen i lederne frembragte magnetfelt fremkalder et induktivt spændingsfald langs ledningen. Hvis lederne ikke er ideelle, tabes effekt i selve ledningen, dvs. man får i så tilfælde tillige et resistivt spændingsfald. På tilsvarende måde vil potentialforskellen mellem lederne give anledning til en forskydningsstrøm mellem dem som i en kondensator, og tillige en ledningsstrøm, hvis dielektrikumet ikke er tabsfrit. Af disse betragtninger fås det på Figur 7.1 viste kredsløb.

Ledningen kan altså karakteriseres ved fire parametre: **resistans** R (Ω/km), **induktans** L (H/km), **konduktans** G (S/km) og **kapacitans** C (F/km) alle gældende pr. længdeenhed. Bemærk, at man i telekommunikation sædvanligvis anvender km som lændeenhed i modsætning til den rigtige SI-enhed m. Udseendet af kredsløbet i Figur 7.1 må ikke lede til den antagelse, at resistans og induktans kun findes i den ene leder; de er selvfølgelig

fordelt på de to ledere, dog ikke nødvendigvis ligeligt. Ligeledes har potentialet på nederste leder intet at gøre med fx jordpotentialet; faktisk anvendes transmissionsledninger ofte **balanceret**, dvs. potentialet på de to ledere er symmetrisk om 0. R , L , G og C kaldes under ét ledningens **primære konstanter** (*primary parameters*), uagtet at de faktisk er frekvensafhængige. De primære konstanter bestemmes ud fra geometrien og materialeparametre samt frekvensen. I Figur 7.6 - 7.8 ses nogle værdier fra datablade. I de følgende fire mindre afsnit omtales forskellige fysiske forhold i forbindelse med denne bestemmelse for at pege på forskellige afhængigheder, fx af frekvensen.

Resistansen R

Resistansen kan for lave frekvenser beregnes ud fra ledernes tværsnitsareal, A , og materialets resistivitet, ρ . I hver leder fås ρ/A . Typiske materialer er kobber og aluminium. Ved højere frekvenser kan man ikke regne med det fysiske tværsnitsareal, men må tage hensyn til, at strømtætheden aftager eksponentielt med afstanden fra overfladen, dvs. strømmen søger mod overfladen af lederen, den såkaldte **strømfortrængning** (*skin effect*). I et koaksialrør er den omtalte overflade indersiden af det yderste rør og overfladen af centerlederen. Man kan vise, at man kan tage hensyn til strømfortrængning ved at beregne resistansen som ovenfor, men som om strømmen var begrænset til og jævnt fordelt i et overfladelag af tykkelse, [7.1]:

$$\vartheta = \sqrt{\frac{\rho}{\mu\pi f}}$$

hvor ρ er ovennævnte resistivitet, og μ er permeabiliteten (ca. $\mu_0 = 4\pi \cdot 10^{-4}$ H/km i ikke-magnetiske materialer). Størrelsen ϑ (theta) kaldes **indtrængningsdybden** (*penetration depth* eller *skin depth*). En vigtig konsekvens af ovenstående formel er, at R vokser med \sqrt{f} , når strømfortrængning er indtrådt. Man kan omtrent regne med, at dette er tilfældet, når ϑ er sammenlignelig med radius eller materialets dimension i øvrigt.

Induktansen L

Induktansen er sammensat af to bidrag, nemlig bidraget L_y stammende fra feltet uden for lederne, og bidraget L_i fra det indre felt. Når ledernes tværudstrækning er lidt større end indtrængningsdybden, dvs. for højere frekvenser, gælder $2\pi f L_i \approx R$. Da R vokser med \sqrt{f} , aftager L_i i dette frekvensområde med \sqrt{f} , hvilket bevirker, at induktansen L aftager mod L_y for højere frekvenser.

Kapacitansen C

Kapacitansen C er nærmest uafhængig af frekvensen. Det viser sig, at med god tilnærmelse gælder $L_y C = \mu\epsilon$, hvor μ er permeabiliteten, og ϵ er permittiviteten ($\epsilon_r\epsilon_0 = \epsilon_r \cdot 8,85 \cdot 10^{-9}$ F/km). Den relative permittivitet ϵ_r afhænger af dielektriket og er 1 for vakuum, 2,3 - 2,5 for polyetylen (PE og PEX) og 2 - 2,5 for (tørt) papir.

Konduktansen G

Konduktansens frekvensafhængighed er bestemt af dielektrikets egenskaber. Ved jævnstrøm her de fleste dielektrika en meget høj isolationsmodstand, men ved vekselstrøm forøges G stærkt pga. dielektriske tab. For mange dielektrika gælder, at **tabsfaktoren**, som man ofte betegner $\tan \delta$ og definerer ved $\tan \delta = G/2\pi fC$ (δ , *angle of loss*), er næsten frekvensuafhængig i store frekvensområder, dvs. G vokser proportionalt med frekvensen, da C som omtalt er stort set uafhængig. En typisk værdi for $\tan \delta$ er 10^{-3} , så strømmen, der skyldes de dielektriske tab, vil være relativt lille i forhold til den, der skyldes C . Yderligere gælder, at disse tab normalt er små i forhold til tabene i lederne.

7.1.1.2 Telegraf ligningen og de sekundære konstanter

Som i Kapitel 4 lader vi nu strøm og spænding være stationære vekselstørrelser, $v(x, t) = \operatorname{Re}(V(x)e^{j\omega t})$ og $i(x, t) = \operatorname{Re}(I(x)e^{j\omega t})$, men i modsætning til den almindelige kredsløbsteori afhænger de af hvor vi er på ledningen. Betragtes Figur 7.1, ses at spændingsstigningen i forhold til længden over det lille udsnit ved stedet x og frekvensen ω når $\Delta x \rightarrow 0$ bliver:

$$\frac{\Delta V}{\Delta x} \rightarrow \frac{dV}{dx} = -R \cdot I(x) - L \cdot j\omega I(x) = -(R + j\omega L) \cdot I(x)$$

og

$$\frac{dI}{dx} = -G \cdot V(x) - C \cdot j\omega V(x) = -(G + j\omega C) \cdot V(x)$$

og ved at sætte disse sammen fås den såkaldte **telegraf ligning** (*telegraphist's equation*):

$$\frac{d^2 V}{dx^2} = (R + j\omega L)(G + j\omega C) \cdot V(x) = \gamma^2 V(x) \quad (7.2)$$

Denne form for differentialligning kaldes en bølgeligning, da løsningerne er bølger, som vi skal se nedenfor. Størrelsen γ er et komplekst tal, som kan spaltes i real- og imaginærdel

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta \quad (7.3)$$

Bemærk at α og $\beta \geq 0$, da R, G, L og $C \geq 0$. γ kaldes **udbredelseskonstanten** eller **vandringsekspONENTEN** (*propagation coefficient*). Disse navne retfærdiggøres sammen med betegnelsen for differentialligningen ved at betragte løsningen $V(x)$ til differentialligningen:

$$V(x) = K_1 e^{-\gamma x} + K_2 e^{\gamma x}$$

hvor konstanterne K_1 og K_2 bestemmes af randbetingelserne. De to løsninger kan fortolkes som to spændingsbølger, der bevæger sig i henholdsvis den positive og den negative retning. Dette indses ved at skrive den reelle tidsfunktion for $e^{-\gamma x}$:

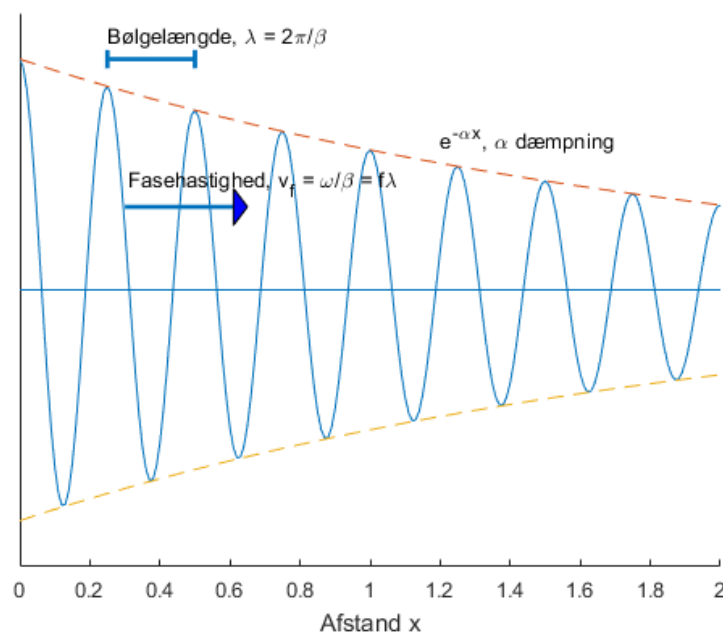
$$v_+(x, t) = \operatorname{Re}(e^{-\gamma x} e^{j\omega t}) = e^{-\alpha x} \cdot \cos(\omega t - \beta x)$$

Denne tidsfunktion er vist på Figur 7.2. For fastholdt t ses det, at punkter med samme fase forekommer i afstanden λ , hvor

$$\lambda = \frac{2\pi}{\beta} \quad (7.4)$$

kaldes **bølgelængden** *wavelength*. β kaldes normalt slet og ret beta, men betegnelsen **fasekonstanten** (*phase change coefficient*) anvendes undertiden. Enheden for β er egentlig længde^{-1} , men for at huske, hvordan argumentet til cos-funktionen er udtrykt, anvendes som regel enheden **rad/km**. For voksende t vil et punkt med konstant fase forskydes i x -aksens positive retning med hastigheden

$$v_f = \frac{\omega}{\beta} = f \cdot \lambda \quad (7.5)$$



Figur 7.2: Dæmpet bølge med positiv udbredelsesretning.

fasehastigheden (*phase velocity*), hvor f er frekvensen udtrykt i Hz. Samtidigt med at bølgen skrider frem, vil dens amplitude aftage eksponentielt på grund af faktoren $e^{-\alpha x}$, da $\alpha > 0$, hvis R og/eller $G > 0$ (ledninger med tab). α kaldes ledningens **dæmpningskonstant** (*attenuation coefficient*). Enheden for α er også længde^{-1} , men det er nødvendigt at præcisere, at den naturlige eksponentialfunktion anvendes, så derfor skal enheden for α være **Np/km**. Det er ikke svært at se forholdet mellem amplituderne på to steder på ledningen med afstanden ℓ er $e^{-\alpha \ell}$, så $\alpha \ell$ er dæmpning i (naturlig) logaritmisk måleenhed. Det er almindeligt at udtrykke dæmpningen i **dB/km**, og man kan omregne til Np/km

ved at dividere med 8,686. Vi indførte dB i Kapitel 4.

På samme måde som ovenfor indses, at $e^{\gamma x}$ svarer til en dæmpet bølge, der udbreder sig i den negative retning. Hvordan forholdet mellem disse er, afhænger af randbetingelserne som vi skal se nedenfor. Spændingen på et givet sted fås ved superposition af de to bølger (interferens).

Ved udregning af dV/dx ud fra udtrykket $V(x)$ og indsættelse i differentiaalligningen for dV/dx fås løsningen for strømmen:

$$I(x) = \frac{K_1}{Z_0} e^{-\gamma x} - \frac{K_2}{Z_0} e^{\gamma x}$$

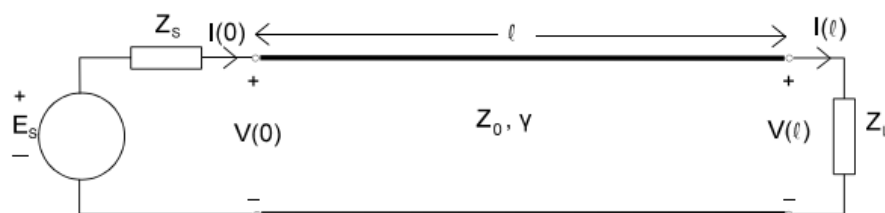
hvor

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (7.6)$$

kaldes ledningens **karakteristiske impedans** (*characteristic impedance*). Denne angiver altså på ethvert sted forholdet mellem spænding og strøm i begge de to løsninger. Bemærk fortegnet for det sidste led i strømudtrykket; strømmen løber jo i den negative retning svarende til bølgens udbredelsesretning. Enheden for Z_0 er selvfølgelig Ω . Tilsammen kaldes γ og Z_0 for ledningens **sekundære konstanter** (*secondary parameters*). De afhænger begge af frekvensen både direkte gennem ω og gennem de primære konstanter frekvensafhængighed (Afsnit 7.1.1.1).

7.1.1.3 Overføringsfunktion for en transmissionsledning

Vi betragter nu en transmissionsledning af endelig længde, ℓ , belastet med impedansen Z_L og tilsluttet en spændingskilde E_S med generatorimpedans Z_S som vist på Figur 7.3.



Figur 7.3: Transmissionsledning med generator og belastning.

Løsningsudtrykkene forenkles en del ved at indføre **refleksionskoefficienten** (*reflection coefficient*) ρ_L som forholdet mellem den reflekterede og den indfaldende spændingsbølge ved $x = \ell$:

$$\rho_L = \frac{K_2 e^{\gamma \ell}}{K_1 e^{-\gamma \ell}} = \frac{K_2}{K_1} e^{2\gamma \ell} = \frac{Z_L - Z_0}{Z_L + Z_0} \quad (7.7)$$

hvor det sidste udtryk findes ved brug af randbetingelsen $V(\ell) = I(\ell)Z_L$ ved $x = \ell$. Den fuldstændige løsning er altså sammensat af en kontinuert bølge, der udbreder sig fremad langs ledningen og rammer belastningen ved enden, og en ligeledes kontinuert bølge, der stammer fra refleksion af den fremadgående bølge ved belastningen i overensstemmelse med refleksionskoefficienten ρ_L . Den resulterende spænding opstår som superposition af disse to bølger.

Regnes videre fås spændingen på stedet x :

$$V(x) = E_S \frac{Z_0}{Z_0 + Z_S} \cdot \frac{e^{-\gamma x} + \rho_L e^{\gamma x - 2\gamma \ell}}{1 - \rho_L \rho_S e^{-2\gamma \ell}}$$

hvor vi også har introduceret en refleksionskoefficient, ρ_S , ved ledningens indgang. Vi finder af dette udtryk overføringsfunktionen $C(\omega)$ som viser sig at være uafhængig af generatorimpedansen:

$$C(\omega) = \frac{V(\ell)}{V(0)} = \frac{(1 + \rho_L)e^{-\gamma \ell}}{1 + \rho_L e^{-2\gamma \ell}} \quad (7.8)$$

Betragter vi en ledning belastet med sin karakteristiske impedans bliver $\rho_L = 0$ og vi får altså overføringsfunktionen mellem indgangen og stedet x :

$$C(x, \omega) = \frac{V(x)}{V(0)} = e^{-\gamma x} = e^{-\alpha x} e^{-j\beta x}$$

Vi kan også finde ledningens **indgangsimpedans** Z_i :

$$Z_i = \frac{V(0)}{I(0)} = Z_0 \frac{1 + \rho_L e^{-2\gamma \ell}}{1 - \rho_L e^{-2\gamma \ell}} = Z_0 \frac{1 + |\rho_L| e^{-2\alpha \ell} e^{j(\angle \rho_L - 2\beta \ell)}}{1 - |\rho_L| e^{-2\alpha \ell} e^{j(\angle \rho_L - 2\beta \ell)}} \quad (7.9)$$

Indgangsimpedansen afhænger af frekvensen på en meget kompliceret måde når man ser på alle de foregående udtryk. Bemærk, at indgangsimpedansen ligesom overføringsfunktionen kun afhænger af ledningens længde og belastning, mens den selvfølgelig er uafhængig af generatorimpedansen. Det ses, at hvis $\alpha \ell$ er stor, bliver $|\rho_L| e^{-2\alpha \ell}$ lille, så betydningen af belastningen bliver ringe, eller anderledes udtrykt, man ser ind i Z_0 ved indgangen i dette tilfælde. Det er også tilfældet når $Z_L = Z_0$, da ρ_L så bliver 0.

Eksempel 7.1 - Forskellige typiske tilfælde

Vi vil nu kort se på forskellige specialtilfælde af den netop udledte generelle teori for ledninger i frekvensdomænet. Disse specialtilfælde er: Den tabsfri ledning og ledninger med tab. Ledninger med tab opfører sig forskelligt afhængig af den anvendte frekvens: Tabene kan være relativt små eller de kan være af væsentlig betydning. Der er grund til at understrege, at en fysisk ledning sagtens kan have væsentlige tab i et frekvensområde, mens tabene er relativt små i et andet område. Der er ingen grund til at anvende formler fra disse eksempler, hvis man kender en lednings parametre; med en lommeregner eller

MATLAB er det nemmere at regne på de generelle udtryk, vi netop har udledt.

En **tabsfri** ledning har $R = 0$ og $G = 0$. Ved anvendelse af formlerne fra foregående afsnit fås

$$Z_0 = \sqrt{\frac{L}{C}}, \quad \gamma = \sqrt{-\omega^2 LC} \Rightarrow \\ \alpha = 0 \text{ og } \beta = \omega\sqrt{LC}, \quad v_f = \frac{1}{\sqrt{LC}}$$

Vi ser at Z_0 og v_f er uafhængige af frekvensen. Indsætter man fra betragtningerne om induktans og kapacitans i Afsnit 7.1.1.1 fås

$$v_f = \frac{1}{\sqrt{LC}} \approx \frac{1}{\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{\epsilon_r}\sqrt{\mu_0\epsilon_0}} = \frac{c}{\sqrt{\epsilon_r}}$$

hvor c er lysets hastighed. Fasehastigheden er altså kun lidt under lysets, typisk $2/3$. Spændingsbølgen i den positive retning,

$$v_+(x, t) = \cos(\omega t - \beta x)$$

er en simpel periodisk funktion (af x) med perioden λ , bølgelængden.

For ledninger med tab vil vi først se på de såkaldte **små tab** hvorved man forstår, at spændingsændringen pr. længdeenhed hovedsageligt skyldes induktansen, og strømændringen hovedsageligt skyldes kapacitansen. Eller anderledes udtrykt

$$R \ll \omega L \text{ og } G \gg \omega C$$

Som det vil fremgå af Afsnit 7.1.2 findes adskillige eksempler på sådanne ledninger med små serietab og en lille tabsfaktor for dielektriket. Den første relation vil for øvrigt altid være opfyldt blot frekvensen er høj nok, da R ved høje frekvenser kun vokser med $\sqrt{\omega}$, og L er nogenlunde konstant (jf. Afsnit 7.1.1.1).

Når disse relationer er opfyldt, fås følgende tilnærmede (ved anvendelse af $\sqrt{1+x} \approx 1 + x/2$ og $1/(1-x) \approx 1+x$) udtryk:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}}$$

og

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \approx \left(\frac{R}{2} \sqrt{\frac{C}{L}} + \frac{G}{2} \sqrt{\frac{L}{C}} \right) + j\omega\sqrt{LC}$$

Fra γ -udtrykket fås

$$\alpha \approx \frac{R}{2} \sqrt{\frac{C}{L}} + \frac{G}{2} \sqrt{\frac{L}{C}} \approx \frac{R}{2Z_0} + \frac{G}{2} Z_0$$

og

$$v_f \approx \frac{1}{\sqrt{LC}}$$

Sammenfattende kan man altså sige, at den karakteristiske impedans Z_0 og fasehastigheden v_f i denne tilnærmelse er lig de tilsvarende værdier for den tabsfri ledning, mens dæmpningen er forskellig fra 0. Der er altså i denne tilnærmelse ingen **dispersion** - alle frekvenser udbreder sig med samme hastighed, men der er dæmpningsforvrængning, da α ændrer sig med frekvensen, som beskrevet nedenfor.

Hvis resistansen pga. strømfortrængning vokser med kvadratroden af frekvensen og tabsfaktoren er nogenlunde konstant over frekvensområdet (jf. Afsnit 7.1.1.1), fås

$$\alpha \approx A_1 \sqrt{f} + A_2 f$$

Ved ikke alt for store frekvenser er leddet $A_1 \sqrt{f}$ som regel overvejende, dvs. α vokser proportionalt med \sqrt{f} . Tabene i sig selv kan altså blive betydelige, men tabene forbliver relativt små, idet R (tab i ledere) vokser langsommere end ωL , og G (tab i dielektrikum) vokser som ω , men $G \ll \omega C$.

Spændingsbølgen i den positive retning er

$$v_+(x, t) = e^{-\alpha x} \cdot \cos(\omega t - \beta x)$$

Denne funktions maksima aftager med x , men det er karakteristisk, at afstanden mellem disse maksima er konstant λ , bølgelængden. Tilsvarende vokser minima med x og har konstant afstand ganske som på Figur 7.2.

For de fleste fysiske ledninger vil der for lave frekvenser gælde at tabene er i lederne og ikke i dielektriket. Eller anderledes udtrykt

$$R \gg \omega L \text{ og } G \ll \omega C$$

Dette tilfælde kaldes normalt **væsentlige tab**. Moderne systemer anvender ret høje frekvenser så vi vil ikke regne på dette tilfælde som var ret relevant i analog telefoni, hvor telefonen blot er tilsluttet en lang parsnoet ledning. Vi vil dog lige se på et par resultater:

$$\alpha \approx \beta \approx \sqrt{\frac{\omega RC}{2}}$$

og

$$v_f \approx \sqrt{\frac{2\omega}{RC}}$$

Da v_f er frekvensafhængig optræder der **dispersion**, dvs. forskellige frekvenser udbreder sig med forskellig hastighed på en ledning af denne type. Der er altså både dæmpningsforvrængning og faseforvrængning i en ledning, hvor tabene er relativt store.

I dette eksempel har vi vist at for en given fysisk transmissionsledning vil der være tre karakteristiske frekvensområder. De fleste karakteristika ændres i disse områder og som en illustration ser vi på dæmpningen α :

- Ved relativt lave frekvenser vokser α med kvadratroden af frekvensen og de fleste andre parametre er også stærkt frekvensafhængige.
- Tabene får mindre betydning ved højere frekvenser i relation til den virkning induktans og kapacitans har. Det får α til at blive nærmest konstant og de andre parametre viser en lignende stabilitet.
- For højere frekvenser hvor strømforvrængning indtræder vokser α igen (med kvadratroden af frekvensen og for rigtigt høje frekvenser endog proportionalt med denne pga. tab i dielektriket). De andre parametre forbliver stabile.

De tre frekvensområder ses fx tydeligt på Figur 7.7. \square

7.1.1.4 Tidsdomænebeskrivelse

Ovenstående behandling af transmissionsledninger er lidt urealistisk da periodiske signaler med én frekvens ω kun sjældent eksisterer - i hvert fald slet ikke i digital kommunikation. Vi var imidlertid nødt til at opstille modellerne i frekvensdomænet fordi det er den eneste måde hvorpå vi kan komme frem til en telegrafligning vi kan løse. Parametrene er også frekvensafhængige og deres værdi er blevet bestemt ved målinger med sinussignaler med bestemte frekvenser, så en realistisk model kan kun gøres ved frekvensbetragtninger.

Imidlertid kunne vi sagtens ved MATLAB beregne udgangssignalet fra en given ledning ved at transformere indgangssignalet til frekvensdomænet, multiplicere med overføringsfunktionen og transformere det tilbage til tidsdomænet. Det kan man ikke lige gøre analytisk da udtryk for transformationen af mange funktioner er ganske ukendte. Man kan imidlertid ved passende tilnærmelser få et resultat der ligner det faktiske for et trinsignal eller et firkantsignal. Disse tilnærmelser fører til en udbredelseskonstant af formen

$$\gamma \approx A\sqrt{j\omega} + B \cdot j\omega$$

hvor A og B er konstanter, uafhængige af frekvensen. Som tidligere set vil B -leddet beskrive forsinkelsen i ledningen så ser vi bort fra den, er der kun A -leddet som giver forvrængningen af signalet. Man kan så vise at det såkaldte **trinsvar** (*step response*) dvs.

det signal $v(x, t)$ der er resultat af påtrykning af et indgangssignal der springer til 1 ved tiden 0 bliver

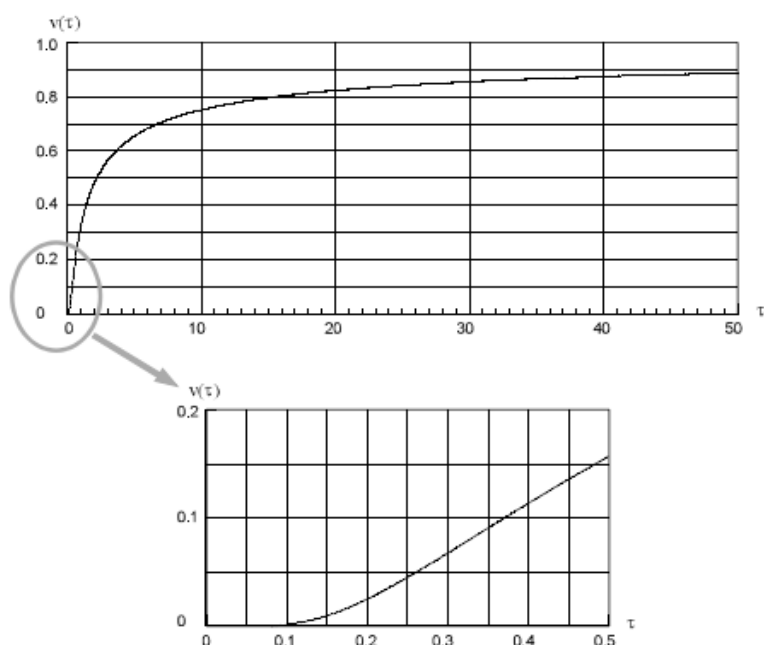
$$v(x, t) = 2 \int_{Ax/\sqrt{2t}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 2 \cdot Q(Ax/\sqrt{2t}) \quad \text{for } t > 0$$

hvor funktionen der integreres over svarer til sandsynlighedstæthedsfunktionen for en normalfordelt variabel, jf. (5.4), og $Q(\cdot)$ er en kort betegnelse for dette integral. Hvis tidsskalaen gøres parameteruafhængig gennem at indføre substitutionen $\tau = 2t/A^2x^2$ får vi:

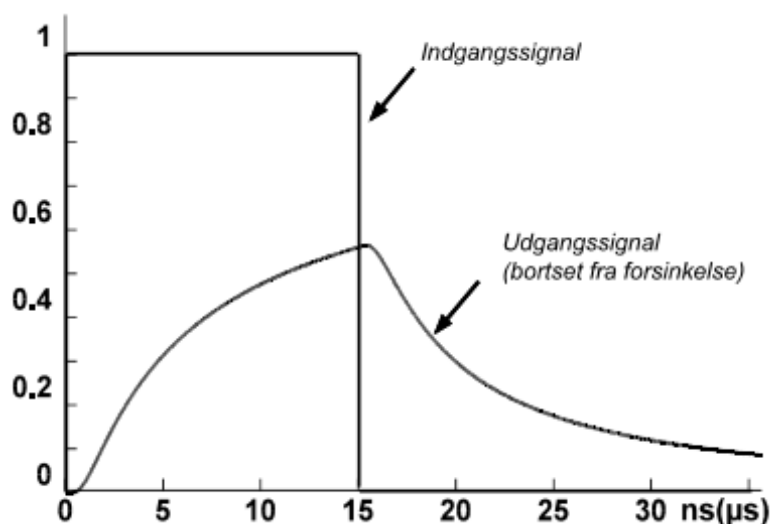
$$v(\tau) = 2 \cdot Q(1/\sqrt{\tau}) \quad \tau > 0$$

Denne funktion vises i Figur 7.4. For $\tau = 0$ er $Q(\infty) = 0$, så der sker ikke noget spring ved tiden 0; det tager sin tid. For $\tau < 0$ bliver $v(\tau)$ selvfølgelig 0 selvom det ikke kan ses af udtrykket ovenfor.

På basis af dette kan man også finde svaret på en firkantet puls som vist i Figur 7.5. Vi har tidligere i Figur 5.4 set denne kraftigt forvrængede puls. For en kort ledning vil den stadig se firkantet ud da τ “går hurtigt” i forhold til tiden, mens lange ledninger giver det forvrængede udseende. Dette fører som beskrevet i Kapitel 5 til **intersymbolinterferens**.



Figur 7.4: Trinsvar (med parameteruafhængig tidsskala) for en transmissionsledning med \sqrt{f} afhængig dæmpning.



Figur 7.5: Svar på et firkantet indgangssignal for en transmissionsledning med \sqrt{f} afhængig dæmpning (tid i μs : $R = 176\Omega/\text{km}$, $C = 50 \text{ nF}/\text{km}$, $x = 1,08 \text{ km}$ og tid i ns: $\alpha_{500 \text{ kHz}} = 11 \text{ dB}/\text{km}$, $x = 0,1 \text{ km}$).

7.1.2 Eksempler på transmissionsledninger i telekommunikation

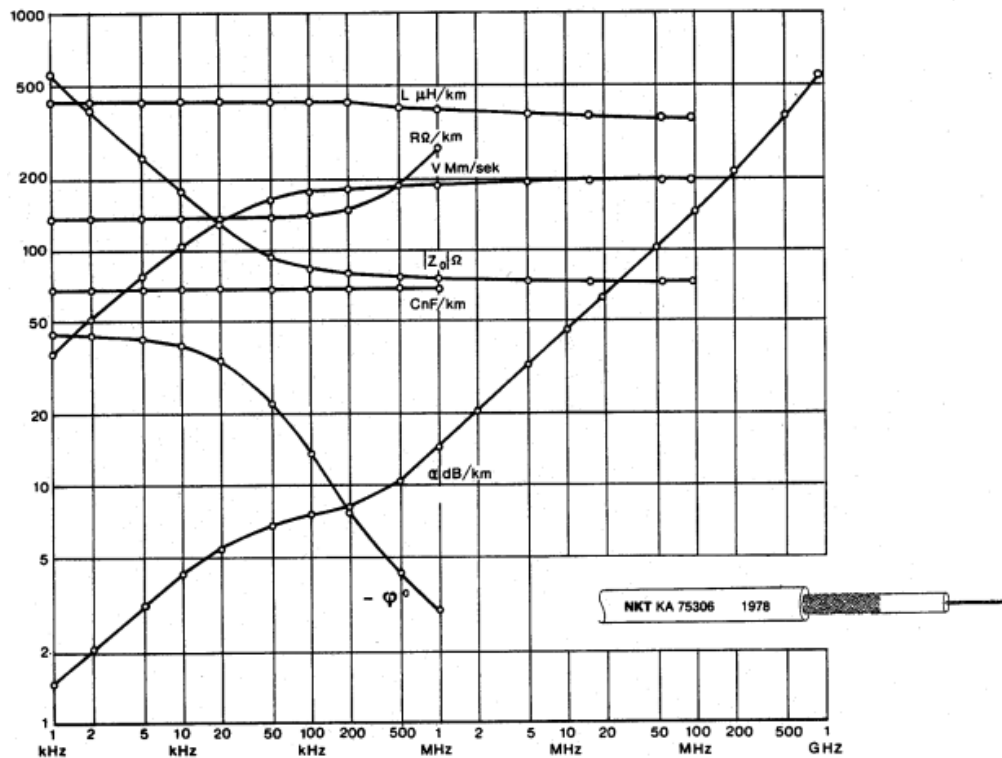
Det viser sig at de modeller, der just er præsenteret, beskriver de faktiske forhold for de anvendte transmissionsledninger ret godt. Man skelner mellem to typer af ledninger

- Koaksialledninger, eksempel fra [7.3] i Figur 7.6.
- Parledninger, eksempler på parsnoede ledninger fra [7.3] i Figur 7.7 - 7.8.

Koaksialledninger (*coaxial lines*) anvendtes tidligere på de højere niveauer af telekommunikationsnet, men er nu mest udbredt til distribution af kabel-TV og som antennekabel som det fremgår af Figur 7.6.

Parledninger (*pair lines*) består af to parallelle tråde og de anvendes i telekommunikationsnet først og fremmest til at abonnentforbindelser. Moderne anvendelser som ISDN og ADSL indebærer betydeligt højere frekvenser end de 4 kHz som de oprindeligt var udviklet til. I vore tider udføres forbindelserne som **parsnoede ledninger** (*twisted-pair lines*) samlet i **kabler** med op til 1000 par (jf. Figur 7.8 med 400 par). Snoningen anvendes for at reducere interaktionen mellem forskellige par, den såkaldte krydstale, se Afsnit 7.1.3. Tråddiametrene er fra 0,4 til 0,7 mm og materialet er kobber eller aluminium. Forskellige isolationsmaterialer har været anvendt (PE, PVC og papir). Eksempler fra [7.3] vises i Figur 7.7 - 7.8.

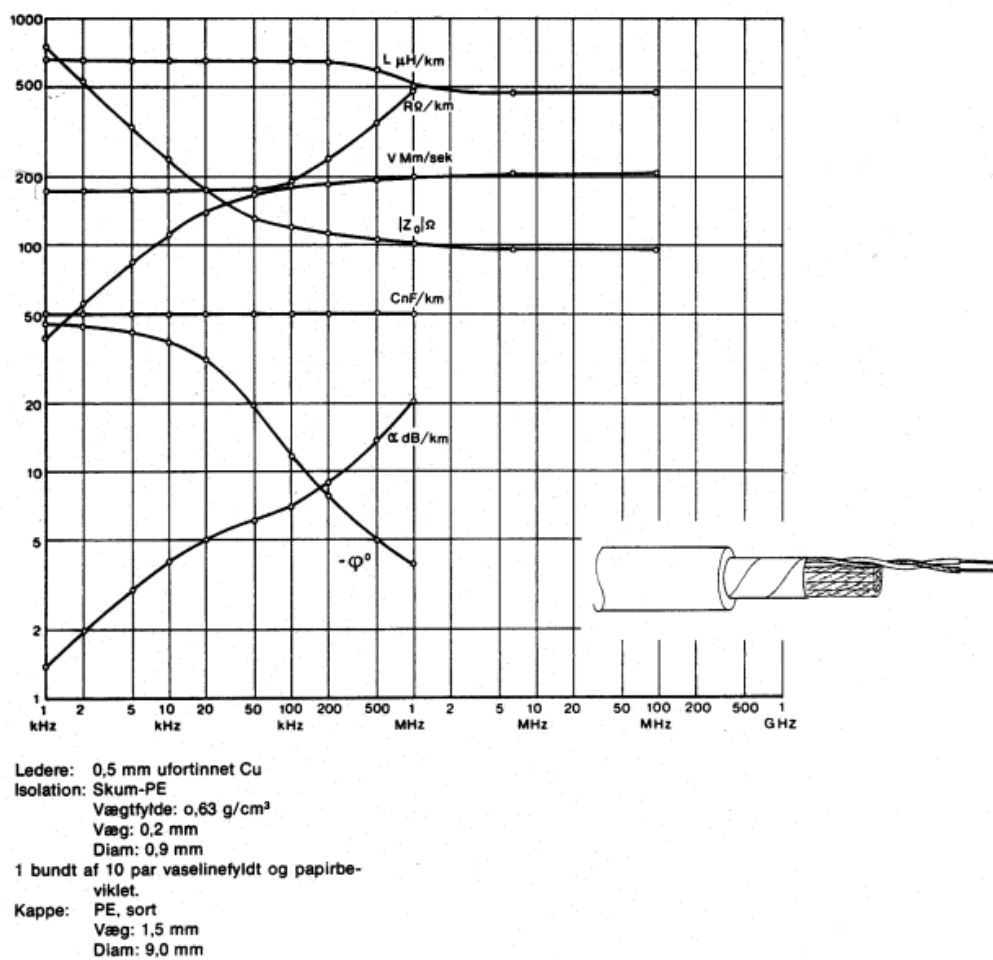
Datablad nr. 30-4 af 8-2-78
FS 82 E 07
75 Ω koaksialledning, type KA 75306



Leder: 7x0,16 mm forselvet Cu
 Isolation: PE
 Diam: 2,95 mm
 Skærm: 66 tråde
 0,15 mm forselvet Cu
 Fyldningsfaktor: 0,49
 Dækning: 74%
 Diam: 3,55 mm
 Kappe: PVC, grå
 Diam: 5,0 mm
 Forkortningsfaktor: 0,67
 Godkendt til nærføring under nr.
 117.9990.29

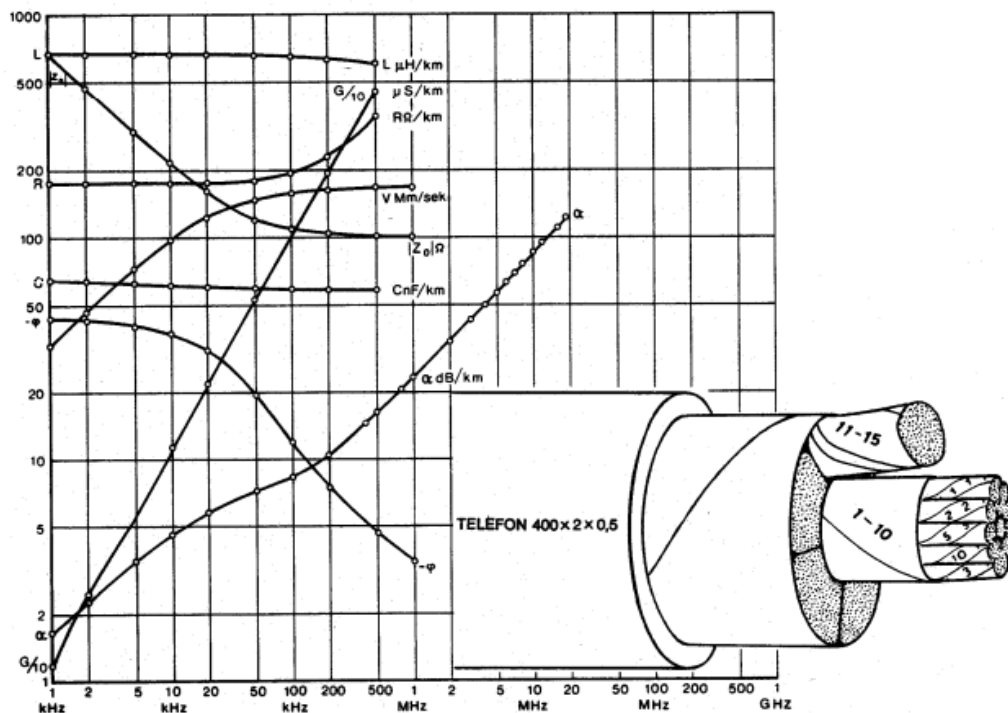
Figur 7.6: Datablad for TV antennekabel fra [7.3].

Datablad nr.: 150-2 af 9-8-76
 FS 73 K 03
 10x2x0,5 mm CU/SPE/VAS/PAP/PE



Figur 7.7: Datablad fra [7.3] for abonnentkabel af høj kvalitet.

Datablad nr.: 160-5 20-1-1976
 VS 19 L 87 rev. 24-6-1977
 400x2x0,5 mm Cu-TIN/PVC/PVC



Ledere: 0,5 mm forlignet Cu
 med 0,23 mm PVC-væg
 Parsnoet (stigning 75 mm for par 1).
 Bundtsnoet med 10 par i hvert bundt
 Par 1.: blå/hvid.
 Bundtets udv. diameter ca. 8 mm.

Midte: kordel af 10 bundter
 Lag: 6 kordeller af 5 bundter hver
 Kappetykkelse: 2,8 mm PVC
 Diameter over kappe: 44½ mm

Monteringskabel til indendørs installation.
 Type 41 efter TU cirkulære 9 af Jan. 74.

Figur 7.8: Datablad fra [7.3] for kabel til kortrækkende anvendelser. Bemærk at tabene ved høje frekvenser er større for PVC end for PE.

Bemærk i øvrigt at G er så lille ($< 1\mu\text{S}/\text{km}$) at den almindeligvis ikke vises. Kun for Figur 7.8 med et kabel til kortrækkende anvendelser er isolationen af dårlig kvalitet og G vises.

7.1.3 Krydstale

Indtil nu har vi kun betragtet forholdene for en enkeltstående transmissionsledning, men det kan ikke undgås at flere par i et kabel interagerer med hinanden. Dette fænomen kaldes **krydstale** (*crosstalk* eller blot *xtalk*) selv i moderne systemer hvor man ikke umiddelbart kan opfattes hvad der kommunikeres og hvor de andre ledninger blot udgør uvedkommende **støj**.

Den gensidige påvirkning kan skyldes tre forskellige slags kobling:

- **Galvanisk kobling** (*galvanic coupling*) skyldes en resistans der er fælles for de to ledninger. Et eksempel er systemer hvor der er en fælles returtråd for flere signaltråde.
- **Kapacitiv kobling** (*capacitive coupling*) skyldes en kapacitans mellem trådene i de to ledninger.
- **Induktiv kobling** (*inductive coupling*) skyldes magnetfeltet som omgiver ledningerne (gensidig induktans).

Den galvaniske kobling er meget generende og tidligt i telekommunikationens historie opdagede man at den skulle undgås ved altid at fremføre ledninger i par så der ikke var nogen fælles returleder. Man står så tilbage med de to andre koblingstyper. Nedenfor vil vi beskrive hvad man gør for at reducere disse i praksis.

De signaler der kobles fra en ledning til en anden udbreder sig her i begge retninger fra koblingspunktet. Krydstale kan derfor ses i to varianter:

- **Nærkrydstale** (***NEXT***, *near-end crosstalk*) som er krydstale i samme ende som senderen for den forstyrrende ledning.
- **Fjernkrydstale** (***FEXT***, *far-end crosstalk*) som er krydstale i den modsatte ende af senderen for den forstyrrende ledning.

De to varianter findes selvfølgelig altid, men hvad der er generende afhænger af anvendelsen af ledningen. Det er klart at krydstale forstyrrer modtagelsen af de som regel svage signaler, så hvis anvendelsen af ledninger i et kabel er ensrettet generer kun fjernkrydstalen. Er anvendelsen derimod dobbeltrettet vil de forholdsvis kraftige sendersignaler nemt forstyrre modtagelsen gennem nærkrydstale. Krydstalen skal altså ses i forhold til signalet på den ramte ledning og man kan udtrykke dette som et specielt signal/støj forhold, **krydstaleafstanden** (*signal-to-crosstalk ratio*).

Koaksialledningers struktur hvor det meste af feltet er mellem yderrøret og inderlederen eliminerer praktisk talt krydstale.

For en parsnoet ledning bevirker snoningen at et forstyrrende magnetfelt skiftevis inducerer positive og negative spændinger i hver "sløjfe" langs ledningen så den totale inducerede spænding er nærmest forsvindende. Samtidigt vil det magnetfelt, ledningen

forstyrrer andre med, reduceres da retningen af det vender for hver “sløjfe”. Den kapacitive kobling reduceres ved at anvende ledningerne balanceret, dvs. spændingen på de to ledere er symmetrisk om 0. Hvis kapacitansen fra disse til den forstyrrede ledning er rimeligt ens vil de symmetriske spændinger bevirke lige store, men modsat rettede strømme som altså opvejer hinanden. Ved anvendelse af snoning og balancering er der således ikke megen induktiv og kapacitiv kobling mellem forskellige par, men der er lidt tilbage som skyldes skævheder i strukturen, så det er stadig et alvorligt problem i parsnoede kabler. Man skal i øvrigt huske at der også er kobling til andre former for ledninger, fx højspændingsledninger.

Erfaringer med krydstale i parsnoede ledninger viser

- Fjernkrydstalen vokser med $\sqrt{\ell}$ og ikke ℓ som man kunne tænke sig ud fra simple overvejelser.
- Fjernkrydstalen vokser med $\omega^{3/4}$ i hvert fald for ikke alt for høje frekvenser.

7.2 Optical transmission lines

Communication by light through the atmosphere has been known for a long time, e.g. smoke signals or optical telegraphs. The main obstacle was the atmospheric disturbances that prevent communication or at least limits the range very much. Another obstacle is the source of light which has to be very concentrated, a problem that was first solved by the advent of lasers about 1960, but at that time optical communication through the atmosphere was outdated by telegraphy and telephony that had been using metallic transmission lines for more than 120 years. What is needed, is a similar transmission medium that can guide the light, and **optical fibers** (*optiske fibre* or *lysledere*) were suggested in 1966. At that time such fibers were used for medical viewing equipment and similar purposes, but the attenuation was far too high to be of any value for even moderate distances. During the next 20 years fiber technology provided fibers with a reasonable attenuation and the installation of such systems began.

Below we shall give a short description of the optical fiber as a communication channel. Most of the material is from [7.4] which is suggested for further reading.

7.2.1 Principle of guided light propagation and single mode fibers

In its simplest form an optical fiber consists of a cylindrical **core** (*kerne*) of silica glass (*kvarsglas*) with refractive index n_1 (*brydningsindeks*) surrounded by a **cladding** (*kappe*) whose refractive index n_2 is lower than that of the core. Because of the abrupt change at

the core-cladding interface such fibers are known as **step-index fibers**. Another type is the **graded-index fiber** where the refractive index decreases gradually inside the core from n_1 down to n_2 . We shall not deal with this type of fiber.

The simplest explanation of the light-guiding property of the fiber is that of conventional geometric optics, but it is only valid as long as the core radius $a \gg \lambda$, the wavelength of the light. Traditionally light is described by wavelengths and the interesting wavelengths for communication purposes are $0.8 \mu\text{m}$ to $1.6 \mu\text{m}$ (in the near infrared region of the spectrum, the visible range is from $0.4 \mu\text{m}$ to $0.75 \mu\text{m}$) corresponding to frequencies of about 200 THz. For the explanation, consider a ray of light which is inside the core. When it strikes the cladding at an angle (with the normal to the cladding) θ_1 it will be transmitted to the cladding forming the angle θ_2 with the normal, and the relation is the well known law of refraction

$$\frac{\sin \theta_2}{\sin \theta_1} = \frac{n_1}{n_2}$$

Now, if a θ_1 gives $\theta_2 > 90^\circ$, the ray is reflected back to the core (with the same angle, θ_1 , as the incident angle). This happens at an incident angle of $\theta_1 > \theta_{\text{crit}}$ given by

$$\frac{1}{\sin \theta_{\text{crit}}} = \frac{n_1}{n_2} \Rightarrow \sin \theta_{\text{crit}} = \frac{n_2}{n_1}$$

which is less than 1 since $n_2 < n_1$. When the reflected ray strikes the cladding again it will be with the same angle and thus it is reflected again and so on. Thus the ray stays entirely inside the core. With $n_1 = 1.503$ and $n_2 = 1.500$, θ_{crit} becomes 86.4° . Now it is easily seen that if light is sent into the beginning of the fiber at several angles $> \theta_{\text{crit}}$ it will travel (with velocity c/n_1) through different paths, and thus a light pulse will undergo **dispersion** and a pulse sequence will have intersymbol interference. Such fibers are known as **multi-mode fibers**. A rough estimate of the possible bit rate for this case may be made by assuming that the time difference of the different paths has to be less than the symbol length T at the end of the fiber (length ℓ). This gives

$$T > \frac{n_1}{c} \left(\frac{\ell}{\sin \theta_{\text{crit}}} - \ell \right) = \frac{\ell n_1}{c n_2} (n_1 - n_2)$$

Continuing with n_1 and n_2 as above, 10 km of such a fiber would allow only a bit rate of 10 Mbit/s. Thus the main fiber of interest will be a fiber where the different propagation paths are not found, the **single-mode fiber**. It is clear that the number of possible paths is decreased by reducing the difference between n_1 and n_2 and it is also intuitively clear that for a thinner fiber the possibility of different incident angles is smaller such that we would expect that a very thin fiber may have the possibility of providing only one path of propagation. However, the simple geometrical-optical considerations above do not hold since the radius of single-mode fibers turns out to be approximately equal to

the wavelength. Involving Maxwell's equations it is shown in [7.4] that the condition for single-mode is

$$\frac{a}{\lambda} \sqrt{n_1(n_1 - n_2)} < 0.271$$

With n_1 and n_2 as above, the fiber will be single-moded for $\lambda > 1.2\mu\text{m}$, if the radius is $a < 4.8\mu\text{m}$ and most communication fibers are about this size.

7.2.2 Dispersion and attenuation for single-mode fibers

Even for single-mode fibers dispersion is found. The **dispersion parameter** characterizes the dependence of the group delay on a small change in wavelength. Some processing of the signal has to be done to compensate for the dispersion. If P_{in} is the optical power launched into a fiber of length ℓ the output power P_{out} at the end is given by

$$P_{\text{out}} = P_{\text{in}} e^{-2\alpha\ell}$$

As we remember from the metallic transmission lines the **attenuation** or **fiber loss** α is normally given in dB/km and it is 8.686 times α in the power equation above. Fiber loss depends on the wavelength of the transmitted light. In Figure 7.9 we show a typical example of the loss variation with wavelength.

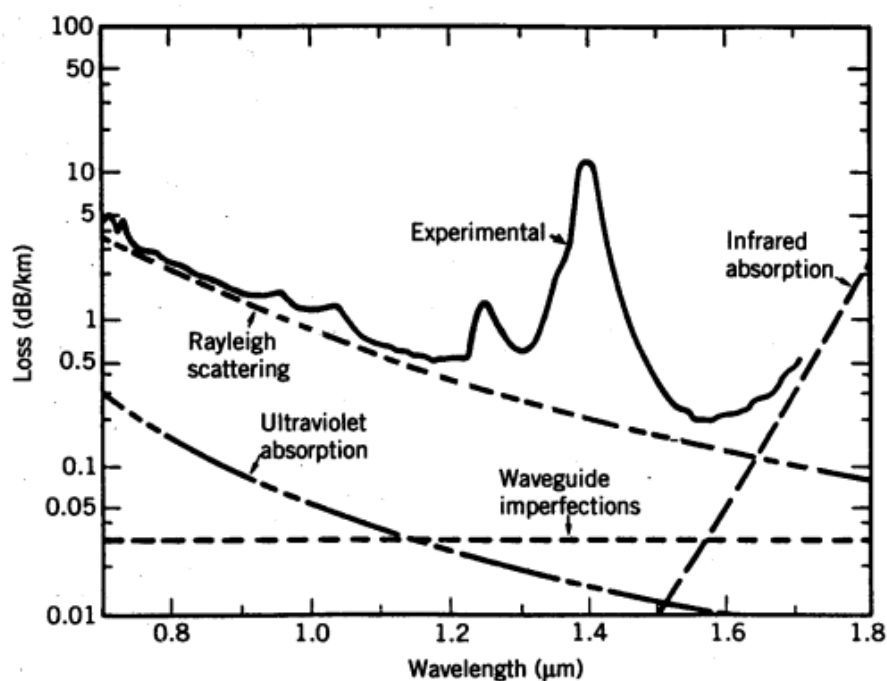


Figure 7.9: Spectral loss profile of a single-mode fiber. Wavelength dependence of fiber loss for several fundamental loss mechanisms is also shown. From [7.4]

The fiber in the example is close to the fundamental limit of 0.15 dB/km for silica fibers in the wavelength region near 1.55 μm . A number of peaks is also seen. These peaks are

due to absorption in impurities e.g. residual water vapor in the silica whereas the silica itself has a relatively little absorption in the range in question, but some loss outside (tails of which are shown as ultraviolet and infrared absorption in the figure). The figure also shows other loss mechanisms, but we shall not discuss these here but refer to [7.4].

It is interesting to note from Figure 7.9 that communication with little loss seems to be found around $1.3\mu\text{m}$ and $1.55\mu\text{m}$ and these are interesting since optical sources for these wavelengths are found (cf. next section). Older systems also used $0.8\mu\text{m}$, but with multi-mode fibers.

In addition to the linear effects introduced briefly, a number of non-linear effects are also found, some of which is about to become of technical use.

It should be noted that **optical amplifiers** exist. These are described in [7.4].

7.2.3 *Optical transmitters and receivers*

The optical transmitter consists of several components:

- An optical source which may be an **LED**, light-emitting diode, or a **semiconductor laser** (sometimes called a laser diode). We shall discuss the optical sources briefly below.
- A modulator for converting electrical data into optical form. In many systems this is not a separate component, since the optical source may be modulated directly by varying the current through it, **direct modulation**. For high speeds, direct modulation provides a problem with chirp, i.e. the wavelength changes slowly along the pulse. For these speeds, so-called external modulators are used. As for direct modulation, these still provide intensity modulation, but the speeds may be 2.5 Gbit/s or even 10 Gbit/s. A certain class of optical communication systems with **coherent modulation** increases the performance by use of modulation techniques as the techniques described in Chapter 6 and they often require a separate modulator. We shall not discuss such systems here, but refer to [7.4].
- An electrical driving circuit for supplying the current to the optical source.

A number of differences between the two types of optical sources exists. We shall not go into detail, but point out some of the differences, [7.4]:

- The light emitted from an LED is incoherent, i.e. the phase of each photon is random, whereas the semiconductor laser emits photons all having about the same phase.

- The emitted light from an LED has a relatively wide spectral width (30 - 60 nm) and a relatively large angular spread ($\sim 100^\circ$). In contrast, the spectral width of the laser is narrow (~ 1 nm) and the narrow emitting angle makes it easier to launch a significant portion of the power into the fiber.
- The semiconductor laser is capable of emitting high powers (~ 10 mW) compared to about 1 mW for the LED. The less efficient coupling of LEDs further reduces the launched power.
- The semiconductor laser may be modulated with very high frequencies, 24 GHz is reported. The maximum modulation frequency for LEDs is of the order of a few hundreds MHz. Thus both devices have a lowpass characteristic.
- Semiconductor lasers are more expensive than LEDs.

Both devices need an input current consisting of a constant bias current and an added current related to the modulated signal in order to operate in a fairly linear way. The emitted, modulated power is then proportional to the modulation current which means that the optical power is proportional to the square root of the electrical power, or 1 dB of optical power corresponds to 2 dB of electrical power.

The optical receiver consists of several components:

- A coupler for presenting the received signal to the photodetector. Similar to the same item for the transmitter.
- A photodetecting device which is some kind of semiconductor diode. Two types are currently in use: **p-i-n photodiode** and **APD, avalanche photodiode**. We shall discuss these devices briefly below.
- A preamplifier following the photodiode.
- For non-coherent systems we have the so-called **direct detection**, which has a receiver filter, synchronization and detection units as for normal baseband systems (cf. Chapter 5), e.g. using cosine roll-off pulses. For coherent systems, the receiver is more complex like the demodulators described in Chapter 6.

The photodiode transforms the received optical power P_r into a current I_r which is directly proportional to the power

$$I_r = RP_r$$

where R is known as the responsivity of the diode. The main difference between the two types of photodiodes is the larger responsivity of the APD, which is due to an avalanche

effect.

The relation above assumes the conversion to be noise free, but as you would expect, this is not the case. Two fundamental noise mechanisms are found: **shot noise** (*haglstøj*) and **thermal noise**. The relation between power and current holds if the variables are interpreted as average values. The shot noise is a result of the fact that electrical current consists of a stream of electrons that are generated at random times. The thermal noise is a result of electrons moving randomly in any conductor, and this kind of noise is also Gaussian and approximately white (up to about 1 THz). In most cases, the thermal noise by far dominates the shot noise. In an optical communication system, other types of noise are found in addition to the two cases of receiver noise degrading performance further, [7.4].

7.3 Radio channels

7.3.1 Principles, propagation, frequency bands

Electromagnetic waves may be radiated from conducting (metallic) structures, and similarly absorbed by such structures, when the wavelength is comparable to the physical size of the structure. These structures are known as **antennas** (*antenner*). As the size ranges from ~ 10 km to 1 mm, the useful frequency range is ~ 30 kHz - 300 GHz. Clearly the required technology is very different for the various parts of this frequency range, but in addition the propagation of the field is quite different.

The waves propagate in principle along straight lines, but deviations may be found due to three phenomena affecting the propagation of the waves:

- Refraction (*brydning*). The refractive index of the atmosphere decreases with altitude and thus the wave is refracted towards the ground.
- Reflection (*refleksion*). Surfaces like water, but also the ionosphere, reflect the waves similar to a mirror.
- Diffraction (*diffraktion*) on obstacles. An obstacle located on the path of a wave becomes the source of a secondary emission which interferes with the original emitted waves.

At low frequencies, refraction allows the signal to be received at large distances, even though the transmitter is below the horizon. Reflection of radio waves in the ionosphere contributes significantly to this propagation in certain frequency intervals. The possibility of obtaining virtually global coverage has been of great importance in the first half of the

Wavelength	Frequency band	Use	Frequency
0.1 μm - 1 μm	Ultraviolet light Visible light Infrared light	Experimental	$3 \cdot 10^{14} - 3 \cdot 10^{15}$ Hz
100 μm - 1 mm	Terahertz	Spectroscopy, tomography Short range comm.	300 GHz - 3 THz
1 mm - 10 mm	Extremely high frequency (EHF)	Radio astronomy, wireless links, radio navigation	30 GHz - 300 GHz
1 cm - 10 cm	Super high frequency (SHF)	Satellite comm, Microwave link, radar Satellite broadcast Wireless LAN	3 GHz - 30 GHz
0.1 m - 1 m	Ultra high frequency (UHF)	Radar, microwave ovens Wireless LAN Mobile comm. Bluetooth, GPS, DECT Walkie-talkie TV broadcast	300 MHz - 3 GHz
1 m - 10 m	Very high frequency (VHF)	FM, DAB & TV broadcast Nautical Aeronautical Mobile comm.	30 MHz - 300 MHz
10 m - 100 m	High frequency (HF)	AM broadcast (SW) Amateur radio Aeronautical, RFID	3 MHz - 30 MHz
100 m - 1 km	Medium frequency (MF)	AM broadcast (MW)	300 kHz - 3 MHz
1 km - 10 km	Low frequency (LF)	AM broadcast (LW) Aeronautical navigation, RFID	30 kHz - 300 kHz
>10 km	ELF, SLF, ULF, VLF	Submarine & subterrestrial comm.	3 Hz (!) - 30 kHz

Tabel 7.1: The frequency ranges for different radio applications.

20th century, but we shall not discuss such systems here. At higher frequencies the signal cannot be received much outside the range of a direct line-of-sight. Some diffraction may take place, and reflections from structures on the ground may influence the reception. Thus communication is limited to short distances from transmitters on the ground, but global coverage can be achieved with a small number of satellites having direct line-of-sight to the transmitters/receivers.

In order to limit interference between different users, and possibly also for other reasons, the use of the electromagnetic spectrum is regulated by national and international bodies. Table 7.1 indicates some applications of the spectrum, but it cannot show the more precise allocations. An always updated version of the frequency allocation in Denmark may be seen at the website www.erhvervsstyrelsen.dk.

7.3.2 Line-of-sight channels

7.3.2.1 Link budget

The power generated by the transmitter amplifier and transferred to the antenna, P_T , is radiated in different directions in proportions described by the radiation pattern of the antenna. The power radiated in a particular desired direction is expressed by the **antenna gain**, G_T , relative to omni-directional radiation over the 4π spatial radians. The power received by an antenna with effective area A_R at distance d is

$$P_R = \frac{P_T G_T A_R}{4\pi d^2}$$

An omni-directional receiving antenna may be proven, [7.1], to have an effective area depending on the wavelength as

$$\frac{\lambda^2}{4\pi}$$

The **gain of the receiving antenna** expresses the efficiency of the actual antenna compared to the omni-directional one so we obtain

$$G_R = \frac{4\pi A_R}{\lambda^2} \quad (7.10)$$

If we introduce the **free space loss** as

$$L_S = \left(\frac{4\pi d}{\lambda} \right)^2 \quad (7.11)$$

the expression for the received power simply becomes

$$P_R = \frac{P_T G_T G_R}{L_S L_A} \quad (7.12)$$

where we have also introduced one more loss factor, L_A , that may account for **atmospheric losses** (rain) and **technological losses** in parts of the receiver. The power calculation

(7.12) is often expressed in dB and referred to as the **link budget**, since it allows the designer to evaluate the margin for acceptable performance and to consider alternatives for improving the system.

Note that the refraction described above sometimes allows a line-of-sight along the surface of the earth longer than it could be expected from the curvature of the earth.

7.3.2.2 Receiver losses

The received signal is corrupted by noise, some of which is related to fundamental physical fluctuations that depend on the temperature. The noise has a component from the received electromagnetic field, and one from the receiver amplifier. These are often combined into an equivalent **noise temperature**, τ . Without going into the thermodynamics involved, we note that the noise is Gaussian, and its spectral density is related to the equivalent temperature as

$$N_0 = k\tau \quad (7.13)$$

where k is Boltzmann's constant, 1.3810^{-23} Joule/Kelvin. N_0 is related to σ^2 of the Gaussian noise distribution.

7.3.3 Fading channels

Many radio channels suffer some degradation due to **multipath** (*flervejsudbredelse*) **effects**. This may be very severe as in the case of mobile communication in a city environment, where the actual line of sight is often obstructed by buildings, and the signal is reflected in multiple ways. In other cases the receiver may have a clear path to the transmitter on a satellite or a high mast, but the signal is reflected from the ground (even more so from the sea), and arrives along a secondary path. In contrast to the channels dealt with up to now, this channel is **time-variant**, since the positions of the antennas are moving and so are some of the obstacles causing reflections and attenuation.

7.3.3.1 Outline of a general model

We shall express the received signal as a sum of several components, with α_n representing the attenuation of the n th component, and $\phi_n = \omega\tau_n$ representing the delay expressed as a phase relative to the carrier frequency. If we are looking of the response to a cosine wave, $\cos \omega t$ (an unmodulated carrier), we get

$$\sum_n \alpha_n(t) \cos(\omega(t - \tau_n(t))) = \operatorname{Re} \left(\sum_n \alpha_n(t) e^{-j\omega\tau_n(t)} e^{j\omega t} \right) = \operatorname{Re} (z(t) e^{j\omega t})$$

where we have introduced the channel response $z(t)$ in complex notation as

$$z(t) = \sum_n \alpha_n(t) e^{-j\phi_n(t)} \quad (7.14)$$

Usually, the attenuations and delays may be assumed to vary slowly, but even a slight change in the signal path can cause significant change in the phase because $\phi_n = \omega\tau_n$ is with a high frequency ω . The phase changes cause the components to sometimes add in phase, while at other times they tend to cancel. This is the phenomenon known as **signal fading**. Notice that the time varying channel response usually introduces new frequencies in the signal in contrast to the time invariant response that we have worked with until now.

Although it is in principle possible to describe a fading signal in a deterministic way by all its components as in (7.14), the physical environment changes in a random way and there are many paths to keep track of. Thus we need a stochastic description which will be developed below.

Studying the delays τ_n we see two phenomena: 1) They change over time, so the channel is time-variant. 2) At a given time, a number of different τ_n may be seen corresponding to the actual paths found at that time. The difference in delay between the first and the last of these is called the multipath delay spread. This is of course changing over time so to characterize a channel some average has to be done leading to the definition of **multipath delay spread** or just **multipath spread** denoted T_m , the maximum difference in delays in the channel. T_m may be found experimentally by modulating the carrier mentioned above with short pulses (as physical realizations of delta functions) and measuring the delays of the channel output. We are now in a position to model the effect on a general signal $x(t)$ with limited bandwidth W . While small changes ($\Delta\tau < 1/W$) in the received phase are naturally modeled by linear transfer functions ($e^{-j\omega\Delta\tau}$), large phase differences ($\Delta\tau > 1/W$) must be interpreted as delayed multipath components since $e^{-j\omega\Delta\tau}$ changes its phase more than 2π over the bandwidth W . Thus the received signal consists of several delayed components added, the n th being delayed n/W and multiplied by a complex number $c_n(t)$ which describes small differences in delays for a group of components around the delay of n/W . The number of components is determined by the multipath delay spread as $L + 1 = T_m W$ (rounded suitably) + 1:

$$\sum_{n=0}^L c_n(t) x\left(t - \frac{n}{W}\right) \quad (7.15)$$

Further Gaussian noise is also added to the signal, but usually this is of minor importance.

The $c_n(t)$ for the $L + 1$ taps are considered to be mutually independent Gaussian variables since they may represent a group of components arriving around n/W with a random phase. If there is a dominant direct path, we can have a nonzero mean of the real and imaginary part and a preferred angle. In this case the amplitude follows a Rice distribution, and the channel is known as a **Ricean fading** channel. We shall consider only

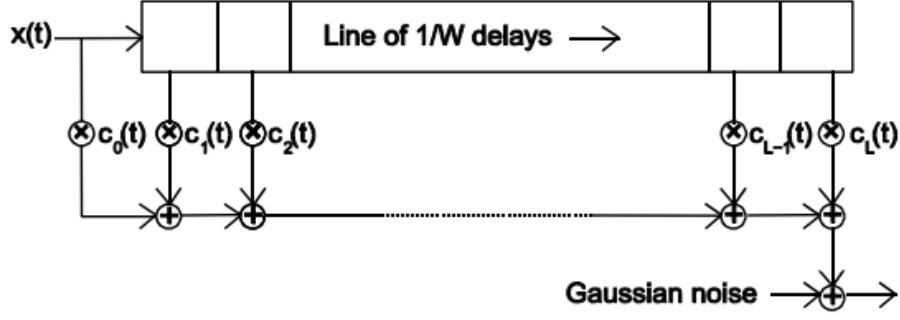


Figure 7.10: Model for multipath channel.

the simpler case, **Rayleigh fading**, where the real and imaginary parts are zero mean Gaussian variables with standard deviation σ . The resulting amplitude distribution is known as the Rayleigh distribution:

$$f(\alpha) = \frac{\alpha}{\sigma^2} e^{-\frac{\alpha^2}{2\sigma^2}}, \quad \alpha > 0 \quad (7.16)$$

In this case the phase is evenly distributed between 0 and 2π . More detail and examples may be found in [7.5].

The inverse of the multipath delay spread T_m is a frequency that indicates the bandwidth where all frequency components fade simultaneously. This frequency, $B_c = 1/T_m$, is known as the **coherence bandwidth**. We consider the fading to be **frequency nonselective** if $W < B_c$ (i.e. only one component is found, $L = 0$), and otherwise **frequency selective**, i.e. fading may affect different parts of the frequency band at different times.

7.3.3.2 Channel models for mobile communication

Whereas some simple examples of fading channels can be analyzed using the models outlined above, the performance of mobile communication systems, which are designed to operate in a complex environment, are often evaluated with channel data collected from extensive measurements. In some cases these data are condensed into a series of complex models representing typical environments.

So far we have assumed that attenuation α varied slowly with time. However, for a vehicle moving at a significant speed, v , there will be a **doppler shift** of the frequency in the range $\pm f_d$, where $f_d = v/\lambda$. For each component of the multipath signal a doppler spectrum may be specified.

For GSM several models are specified representing different types of terrain [7.6]. For the typical urban area 12 taps are specified, each with a Rayleigh distribution. The relative time delays range from 0 to 5 μs and the relative power from 0 to -11 dB. The model for

hilly terrain is similar, but has larger delays, whereas in rural areas a component with a Ricean distribution represents a direct line of sight.

7.4 References for Chapter 7

- [7.1] David K. Cheng. *Fundamentals of Engineering Electromagnetics*. Addison-Wesley, 1994, ISBN: 0-201-56611-7.
- [7.2] S. Ramo, J.R. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics, Third Edition*. John Wiley & Sons, 1994. ISBN: 0-471-58551-3.
(Anvendes i kursus 34120 Elektromagnetisme for kommunikation).
- [7.3] NKT. *Kabelteknisk Håndbog, Svagstrøm*. Nordiske Kabel og Traadfabriker, 1981.
- [7.4] Govind P. Agrawal. *Fiber-Optic Communication Systems, 2nd Edition*. John Wiley & Sons, 1997. ISBN 0-471-17540-4.
(Anvendes i kursus 34130 Introduktion til optisk kommunikation)
- [7.5] John G. Proakis, and Masoud Salehi. *Communication Systems Engineering, Second Edition*. Prentice Hall Inc, 2002. ISBN: 0-13-095007-6.
- [7.6] ETSI. *GSM Recommendation 05.05: Transmission and Reception*. Version 3.11.0, 1992.

INDEKS

På grund af det engelske sprog i dele af Kapitel 7 vil indekset indeholde mange engelske ord, men de vigtigste danske termer fra dette kapitel skulle også være med. Der er lidt problemer med sortering af æ, ø og å, fx er de første ord uden for den alfabetiske rækkefølge.

øjediagram, 79

øvre sidebånd, 82

16QAM, 89

8PSK, 88

A/D-konvertering, 22

additiv støj, 65

ADSL, 101

alfabet, 59

aliasering, 18, 41

aluminium, 106, 115

AM, 81, 86

AM-DSB-SC, 85

amplitude, 14

amplitudekarakteristik, 104

amplitudemodulation, 81

amplitudespektrum, 26, 32

analog/digital-konvertering, 22

analogt signal, 7

angle of loss, 107

antialiaseringsfilter, 18

APD, 124

ASK, 86

Asymmetric Digital Subscriber Line, 101

atmospheric loss, 127

attenuation coefficient, 108

avalanche photodiode, 124

bærebølge, 81

bærefrekvens, 81

bølgelængde, 108

båndbegrænset signal, 16, 17

båndbegrænset spektrum, 33

båndbredde, 76, 77

balanceret ledning, 106

basisbåndstransmission, 59

baud, 61

beslutning, 71

beslutningsområde, 91

bit, 18, 59, 61

bitrate, 61

blanding, 82

blokdiagram, 11

BPSK, 87

C (kapacitans for ledning), 105

capacitive coupling, 119

CDMA, 97

characteristic impedance, 109

coaxial lines, 115

coherence bandwidth, 130

coherent modulation, 123

cosinus-roll-off, 77

crosstalk, 118

dæmpning, 104

dæmpningskonstant, 108

dB, 48, 68

demodulation, 83

deterministisk signal, 8

DFT, 36
 digital kommunikation, 8, 59
 digitalt signal, 8
 direct detection, 124
 direct modulation, 123
 Diskret Fourier-Transformation, 36
 diskret tid, 7
 dispersion, 112, 113, 121, 122
 DMT, 101
 doppler shift, 130

 effektspektrum, 42
 effektværdi, 53
 elektrisk ladning, 47
 elektrisk spænding, 47
 elektrisk strøm, 47
 elektriske effekt, 48
 energispektrum, 42

 fading, 129
 fase, 14
 fasedrejning, 104
 fasehastighed, 108
 fasekarakteristik, 104
 fasekonstant, 108
 fasemodulation, 81
 fasespektrum, 26, 32
 fejl, 67
 fejlkorrigerende koder, 69
 fejlsandsynlighed, 67
 $\text{fft}(v)$, 29
 firkantpuls, 27
 fjernkrydstale, 119
 flervejsudbredelse, 128
 FM, 81, 86
 foldning, 46, 57
 forstærker, 50
 forvrænget, 47
 forvrængning, 62

 forvrængningsfri kanal, 104
 Fourierkoefficienter, 24
 Fourierrække, 24
 Fouriertransformation, 26, 32
 frekvens, 13, 23, 24
 frekvensdomæne, 23, 24
 frekvensmodulation, 81
 frekvensmultipleksning, 85
 frekvensopløsning, 35
 frequency nonselective, 130
 frequency selective, 130
 FSK, 86

 G (konduktans for ledning), 105
 galvanic coupling, 119
 galvanisk kobling, 119
 gaussisk, 65
 GMSK, 94
 Gray-kode, 61
 grundfrekvens, 24
 guardspace, 99

 haglstøj, 125
 harmonisk funktion, 14
 hvid støj, 65

 idelt lavpasfilter, 52
 $\text{ifft}(V)$, 30
 ikke-periodisk signal, 8, 37, 42
 impedans, 51, 55
 impedanstilpasning, 56
 impulsrespons, 46, 57
 indgangsimpedans, 110
 indgangssignal, 45
 indtrængningsdybde, 106
 inductive coupling, 119
 induktans, 105
 induktiv kobling, 119
 information, 7, 8
 interpolation, 16

intersymbolinterferens, 75, 114
invers Fouriertransformation, 30
irreversibel proces, 20
ISI, 75

kabel, 115
kanaludligning, 104
kapacitans, 105
kapacitiv kobling, 119
karakteristisk impedans (Z_0), 109
kausalt system, 45
Kirchhoffs strømlov, 54
koaksialledning, 115
kobber, 106, 115
kodningsgevinst, 70
kommunikationssignal, 60
kompleks harmonisk funktion, 24
kompleks repræsentant, 51
konduktans, 105
konstellationsdiagram, 88
konventionel AM, 84
korrelation, 67
krydstale, 118
krydstaleafstand, 119
kvadraturmodulation, 86
kvantisering, 7, 13, 18, 20
kvantiseringsfejl, 20
kvantiseringsinterval, 20

L (induktans for ledning), 105
lavpasfilter, 52
LED, 123
lineær kvantisering, 20
lineært system, 46
link budget, 128
logaritmisk kvantisering, 21

midtpunkt, 20
modulation, 81
moduleret signal, 10

moduleret system, 59
MSK, 93
multidimensionale signaler, 96
multipath, 128
multipath delay spread, 129
multipath spread, 129

nærkrydstale, 119
nedre sidebånd, 82
noise temperature, 128
normalfordelt, 65
nummerere, 20
Nyquists 1. kriterium, 75
Nyquists 2. kriterium, 77

OFDM, 98
Ohms lov, 49, 51
optical amplifier, 123
optimal modtager, 67, 70, 91
Ortogonal frekvensmultipleks, OFDM, 98
ortogonale pulser, 96
ortogonalitet, 97
overføringsfunktion, 47, 57
overstyring, 20

p-i-n photodiode, 124
pair lines, 115
PAM, 60
parledninger, 115
parsnoede ledninger, 115
PCM, 21
penetration depth, 106
periodisk signal, 8, 13, 24, 29, 36, 42
periodisk spektrum, 37
phase change coefficient, 108
phase velocity, 108
PM, 81, 86
PPM, 60, 97
primære konstanter, 106

primary parameters, 106
 propagation coefficient, 107
 PSK, 86, 88
 puls, 9, 60
 PulsAmplitudeModulation, 60
 Pulsbreddemodulation, 61
 pulsform, 60
 PulsPositionModulation, 60, 97
 PWM, 61

 QAM, 86, 89
 QPSK, 87

 R (resistans for ledning), 105
 Rayleigh fading, 130
 realeffekt, 53
 reflection coefficient, 109
 refleksionskoefficient, 109
 rekonstruktion, 13, 20
 resistans, 105
 Ricean fading, 129

 sampling, 13, 71
 samplingfrekvens, 15, 17
 samplingsætning, 16, 52
 samplingtid, 13
 secondary parameters, 109
 sekundære konstanter, 109
 semiconductor laser, 123
 shot noise, 125
 signal, 7
 signal/støjforhold, 21, 68
 signaleringshastighed, 61
 $\text{sinc}(x)$, 15
 skin depth, 106
 skin effect, 106
 skineffekt, 54
 spektrum, 26
 støj, 63, 118
 step response, 113

 stokastisk signal, 8, 43, 63
 strømfortrængning, 106
 symbol, 59
 symbolhastighed, 61
 symboltid, 59
 system, 45

 tabsfaktor, 107
 tabsfri ledning, 111
 talesignal, 9
 technological loss, 127
 telegrafligning, 107
 telegraphist's equation, 107
 thermal noise, 125
 Thevenin's sætning, 55
 tidsdomæne, 23, 24
 tidsinvariant, 46
 tilpasset filter, 71
 time-variant channel, 128
 todimensionalt signalsæt, 87
 transistor, 50
 transmissionskanal, 103
 transmissionsledning, 104
 trinsvar, 113
 twisted-pair lines, 115

 udbredelseskonstant, 107
 udgangssignal, 45
 udligning, 74
 udstyringsområde, 19
 ulineær kvantisering, 20

 v_f , 108
 vandringssekspont, 107
 vinkelfrekvens, 14

 wavelength, 108

 xtalk, 118

 Z_0 , 109