

Signalbehandling och kommunikationssystem



Henrik Hallenberg



Denna bok har som huvudsyfte att användas som kurslitteratur på Yrkeshögskolan Yrgo, Göteborg. Stora av denna kursbok baseras på Wikipediaartiklar som är tillgängliga under [Creative Commons Attribution-ShareAlike License CC BY-SA 3.0](#). Materialet i denna kursbok delas under samma licens, förutom de delar som kopierats i enlighet med skolornas kopieringsavtal.

Bildmaterial från Wikipedia och andra källor delas ibland under andra licenser vilket redovisas där så behövs.

Försättssidans foto kommer från [Muhammad Syafi Al - adam](#) och [Unsplash \(fri licens\)](#).

Innehållsförteckning

1. Introduktion.....	8
1.1. Kommunikationssystem	8
1.2. Signalbehandling	9
1.3. Om innehållet i denna bok.....	9
2. Frekvensanalys.....	10
2.1. Klassificering av signaler	10
2.1.1. Vad menar vi med en signal?	10
2.1.2. Deterministiska och stokastiska signaler	10
2.1.3. Kontinuerliga och diskreta signaler	11
2.1.4. Analoga och digitala signaler.....	11
2.1.5. Periodiska och icke-periodiska signaler	12
2.1.6. Udda och jämna signaler.....	13
2.1.7. Energisignalер och effektsignalер	15
2.1.8. Exempel på signaler och deras klassificering	15
2.2. Fourierserier	16
2.2.1. Fourierserier i reell form (trigonometrisk form)	16
2.2.2. Fourierserier i komplex form	19
2.2.3. Fourierserier i amplitud-fas-form.....	21
2.2.4. Fourierserier för några vanliga periodiska signaler	22
2.3. Fouriertransformen.....	23
2.4. Diskreta Fouriertransformer och FFT	26
2.4.1. Diskret-tid Fouriertransform, DTFT	26
2.4.2. Diskret Fouriertransform, DFT	27
2.4.3. Snabb Fouriertransform, FFT	29
2.4.4. Sampling och Nyquist-Shannons samplingsteorem	29
2.5. Sammanställning av Fourierserier och -transformer	31
2.6. Bandbredd	32
3. Brus	34
3.1. Vad är brus och varför behöver det behandlas?.....	34
3.2. Olika typer av brus	34
3.3. Viktiga mått för karakterisering av brus	36
3.3.1. Brusets effektspektrum	36

3.3.2.	Ekvivalent brusbandbredd	37
3.3.3.	Bruseffekt, brusspänning och brusström.....	39
3.3.4.	Signal-brusförhållande.....	39
3.4.	Termiskt brus	41
3.4.1.	Brus i en resistor.....	41
3.4.2.	Brusberäkningar på resistiva nät.....	44
3.4.3.	Brus i en kondensator	46
3.5.	Bakgrundsbrus	47
3.6.	Brus i förstärkare.....	48
3.6.1.	Brusfaktor.....	48
3.6.2.	Brusfaktor för kaskadkopplade förstärkare	49
3.6.3.	Brustemperatur.....	51
3.7.	Brus i transmissionsledningar	52
4.	Transmissionsledningar	54
4.1.	Olika typer av transmissionsledningar	54
4.1.1.	Koaxialkabel	54
4.1.2.	Parledare	55
4.1.3.	Planär transmissionsledning	56
4.2.	Transmissionledningar och telegrafekvationen.....	56
4.3.	Vågekvationen	58
4.3.1.	Lösning för en ideal, förlustfri transmissionsledning.....	59
4.3.2.	Lösning för en allmän transmissionsledning [31]	61
4.4.	Reflektion och transmission.....	61
4.5.	Några vanliga vågutbredningsfall.....	63
4.5.1.	Anpassad last $ZL = Z_0$	63
4.5.2.	Kortsluten last $ZL = 0$	63
4.5.3.	Öppen last $ZL = \infty$	64
4.6.	Spänningsdiagrammet.....	64
4.7.	Sinusformade vågors utbredning i ideal transmissionledning	66
4.8.	Stående våg.....	68
4.9.	Smithdiagrammet.....	69
4.10.	Impedansanpassning	72
4.10.1.	Kvartsvågtransformatorn	72

4.10.2.	Stubanpassning	72
4.11.	Egenskaper hos verkliga transmissionsledningar.....	72
5.	Elektromagnetiska vågor och vågutbredning	73
5.1.	Elektromagnetiska spektrumets indelning [37]	73
5.2.	Elektromagnetiska vågors egenskaper	76
5.2.1.	Elektromagnetisk vågutbredning i fri rymd	76
5.2.2.	E- och H-fält för plan våg.....	77
5.2.3.	Polarisation	78
5.2.4.	Strålningstäthet och Poyntingvektor	80
5.3.	Elektromagnetisk vågutbredning i material.....	80
5.3.1.	Vågutbredning i dielektriska material.....	80
5.3.2.	Vågutbredning i metaller. Skineffekten.....	81
5.4.	Plana elektromagnetiska vågors reflektion mot plana ytor.....	82
5.5.	Vågutbredning i olika typer av kanaler.....	86
5.5.1.	Friis transmissionsekvation och vågutbredning i fri rymd	86
5.5.2.	Vågutbredning i reflektiv miljö. Färdning.....	86
5.5.3.	Vågutbredning i miljöer med multipelreflektion. Kanalmodeller.	87
5.5.4.	Länkbudget.....	87
6.	Antenner.....	88
6.1.	Grundläggande antennparametrar	88
6.1.1.	Antennens fjärrfält, närfält och fjärrfältsavståndet	88
6.1.2.	Strålningsdiagram och lobber	89
6.1.3.	Lobvidd.....	91
6.1.4.	Isotropa, rundstrålande och riktade diagram	91
6.1.5.	Direktivitet	92
6.1.6.	Antennvinst.....	92
6.1.7.	Reciprocitet	93
6.1.8.	Effektiv area	94
6.1.9.	Antennpolarisation. Polarisationsförluster. Kopolarisation och korspolarisation.	94
6.1.10.	Antennens ingångsimpedans	95
6.1.11.	Antennens strålningsverkningsgrad	96
6.1.12.	Antennens totala verkningsgrad	96



6.2.	Olika vanliga antenntyper.....	97
6.2.1.	Monopolantenn.....	97
6.2.2.	Dipolantenn.....	97
6.2.3.	Loopantenn	98
6.2.4.	Reflektorantenn	99
6.2.5.	Hornantenn	100
6.2.6.	Patchantenn	101
6.3.	Gruppantennen.....	101
7.	Filter	102
7.1.	Analoga filter.....	102
7.2.	Digitala filter.....	103
7.2.1.	Ett vardagligt exempel.....	103
7.2.2.	Analys av ett enkelt digitalt lågpassfiltret.....	105
7.2.3.	Vikning och vikningsfilter	107
7.2.4.	Z-transform och dess koppling till frekvensanalys	109
7.2.5.	Digitala filters överföringsfunktion $H(z)$	111
7.2.6.	FIR-filter	113
7.2.7.	IIR-filter	114
7.2.8.	Poler och nollställen	115
7.2.9.	Notchfilter	119
7.2.10.	Butterworthfilter	119
7.2.11.	Chebyshevfilter.....	119
7.2.12.	Praktisk filterimplementation	119
8.	Modulation	120
8.1.	Analog modulation	120
8.1.1.	Linjär modulation	120
8.1.2.	Modulation med dubbelt sidband	121
8.1.3.	Koherent detector	123
8.1.4.	Amplitudmodulation, AM	126
8.1.5.	Spektrum för AM	129
8.1.6.	Detektering vid AM.....	131
8.1.7.	Fasmodulation.....	134
8.1.8.	Frekvensmodulation	136



8.1.9.	Översikt av argumentmodulation	138
8.1.10.	Spektrum för argumentmodulation	140
8.1.11.	Superheterodyn mottagaren	141
8.1.12.	Spegelfrekvenser	143
8.2.	Digital modulation.....	143
	Referenser	144
	Appendix A – Tredimensionella koordinatsystem	149

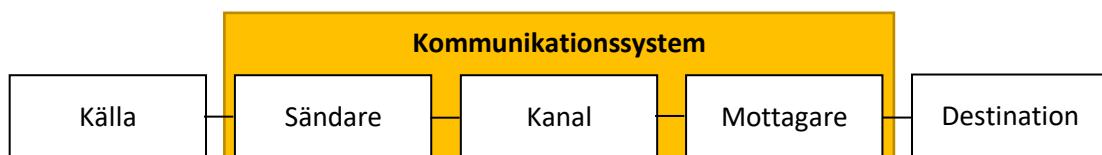
1. Introduktion

1.1. Kommunikationssystem

I en vid bemärkelse är kommunikationssystem något som möjliggör överföring av information mellan två eller flera parter, ofta separerade med ett stort avstånd. Ett gammalt men väl fungerande exempel på kommunikationssystem var de vårdkasar (signaleldar) som förr användes för att varna befolkning eller mobilisera vid ett yttert hot. Detta är ett exempel på ett optiskt kommunikationssystem med låg datatakt och överförsingshastighet.

I denna kursbok behandlas de delsystem som ingår i ett modernt kommunikationssystem som baseras på elektromagnetiska vågor för överföring. Elektromagnetiska vågor är en naturresurs med många attraktiva egenskaper för informationsöverföring, t.ex. förmågan att använda ett stort antal kanaler parallellt i samma överföringsmedium. Det gör att olika typer av information (TV, radio, mobiltelefoni, datakommunikation, satellitkommunikation, radar, ...) kan samexistera i samma miljö. Förutom den kommunikation som går via luften så används elektromagnetiska vågor även i kabelbunden kommunikation (fiberkablar, nätverkskablar, databussar, ...).

Figur 1 visar ett enkelt blockschema för ett kommunikationssystem. Varje block kan i sin tur bestå av ett antal delsystem. Med kommunikationssystem menas då den uppsättning av mjukvara, hårdvara, protokoll mm som möjliggör överföring av information från källa till destination.



Figur 1 Enkelt blockschema för ett kommunikationsystem. Information från en källa ska överföras till en destination. Informationen kan vara på en mängd olika format, t.ex. text, tal, bild, video osv.

- Sändaren innehåller en mängd delsystem som behövs för att överföra källans information till en signal som förs över till kanalen. Exempel på delsystem är A/D-omvandling, filtrering, kodning, modulation, sändarelektronik med effektförstärkare och antenn.
- Kanalen är det medium som överför signalen till mottagarsidan. Den kan vara av det deterministiska signalen som t.ex. en transmissionsledning eller en radiokanal utan reflektioner som vid satellitkommunikation, alternativt av mer komplex natur som en kanal för mobiltelefoni med multipelreflektioner och dopplerskift. Oavsett vilket så är kanalen något som behöver analyseras och



förstås eftersom den påverkar signalen med olika former av distorsion, brus, fädning mm.

- Mottagaren tar emot signalen som sänds över kanalen och ska avkoda informationen till den slutliga destinationen. I mottagaren kan ingå t.ex. antenn, mottagarelektronik, demodulation, filterering, avkodning och D/A-omvandling.

Exakt vilka delsystem som ingår i ett kommunikationssystem beror på typ av system och kommunikation. Ett klassiskt telefonisystem för fast telefoni är väldigt annorlunda än ett meshnätverk för ett hemma-WiFi. Men den generella uppbyggnaden enligt Figur 1 går som en röd tråd genom all design av kommunikationssystem.

1.2. Signalbehandling

Även signalbehandling är ett brett begrepp som kan rymma många olika saker. I en strikt tolkning så är signalbehandling begränsat till signalanalys med efterföljande filtrering. I en bredare definition så är signalbehandling teknik som används för att analysera, förändra och optimera signaler för att få mer information, för att göra signalen lättare att använda och optimera ett systems prestanda. I en mycket bred definition så är signalbehandling all den teknik som behövs för att behandla en signal så att den är överförbar mellan källa och destination. I den senare definitionen så sammanfaller signalbehandling till stor del med begreppet kommunikationssystem.

Oavsett vilken definition man väljer att använda är kunskap om signalbehandling mycket central när kommunikationssystem designas. Ett kommunikationssystem har alltid begränsningar och att veta hur man optimerar dess prestanda hänger ofta på hur bra signalbehandling som implementeras. Signalbehandlingen kan göras med analog elektronik likväld som digital elektronik eller mjukvara och i dagens moderna kommunikationssystem finner man en mix av olika teknologier och lösningar för signalbehandlingar i olika delar av informationsöverföringskedjan.

1.3. Om innehållet i denna bok

Det primära syftet med denna bok är att fungera som kursbok inom elektronikutbildningarna på yrkeshögskolan i Göteborg. Det finns ett visst överlapp mellan kurserna och det finns teoriområden som inte finns beskrivet i detalj. Ett exempel är analoga filter, som täcks i andra kurser.

Boken är under utveckling och förändras i takt med teknikutveckling samt utbildningarnas kursplanerna.

Till kursboken finns även ett övningshäftet med räkneövningar.

2. Frekvensanalys

2.1. Klassificering av signaler

2.1.1. Vad menar vi med en signal?

I signalbehandlingen och kommunikationsteorin används ofta begreppet "signal" för det mesta som är någon form av tidsvarierande funktion. En bred definition är att "signal" är en funktion som förmedlar information om ett visst fenomen. Inom elektroniken och kommunikationsteorin tänker vi ofta på signaler som någon form av tidsvarierande spänning, ström eller elektromagnetisk våg som bär information. Typiska exempel är:

- Strömmen som driver en högtalare som spelar musik
- Den elektromagnetiska vågen som förmedlar ett samtal via mobiltelefon
- Datatrafik i en fiberoptisk kabel

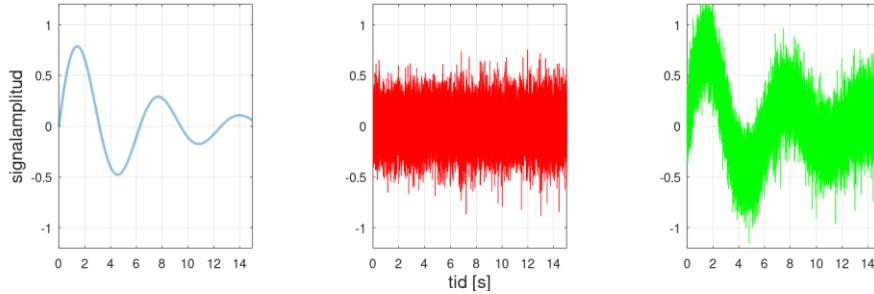
Fenomenet som representeras av en signal behöver inte vara till någon verlig nytta. T.ex. så talar man om brus, som är en form av slumpmässig signal som oftast försämrar nyttosignaler. Likväl så faller brus inom ramen för den breda definitionen av "signal", eftersom den brusiga signalen förmedlar information om ett fenomen, t.ex. effekten i en bruskälla.

En viktig sak att lägga på minnet är man normalt tänker på "signal" som en tidsvarierande funktion. Ritar vi upp en signal i en graf så är det alltså storheten tid på x-axeln. Som vi ska se i detta kapitel kan sådana tidsvarierande funktioner transformeras från detta tidsplan till ett frekvensplan och representeras av en frekvensvarierande funktion. Det är mycket vanligt att man i signalbehandling och kommunikationsteori gör denna form av analys och oftast så beskriver vi alltså signaler i en graf med storheten frekvens på x-axeln. Men det är också viktigt att komma ihåg att man ofta frångår definitionen som tidsvarierande funktion inom signalbehandlingen och breddar begreppet till att t.ex. inkludera bilder och video, som även har en variation i rumskoordinater.

2.1.2. Deterministiska och stokastiska signaler

En deterministisk signal är en signal som i varje tidpunkt kan beräknas med ett matematiskt uttryck. En stokastisk (slumpmässig) signal är en signal som i varje tidpunkt antar ett slumpmässigt värde och därför behöver modelleras stokastiskt [1]. Ett exempel på en deterministisk signal är $f(t) = e^{-t/2\pi} \sin(t)$.

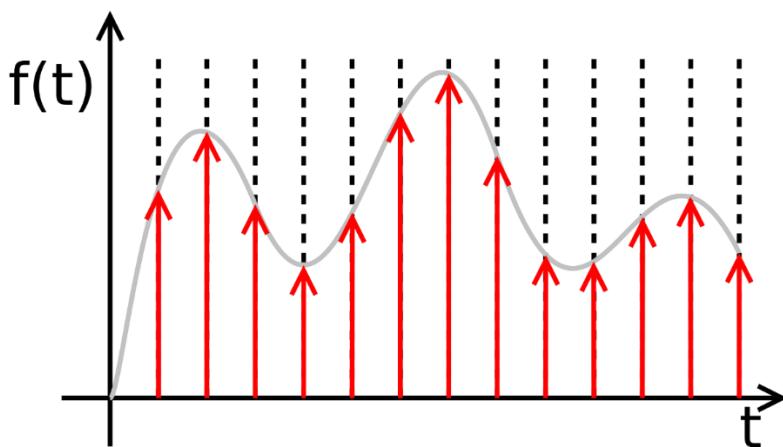
Kunskap om stokastiska signaler är också mycket centralt i signalbehandlingen och kommunikationsteori eftersom t.ex. brus modelleras som en stokastisk signal. I många fall är brus en dimensionerande faktor när man skapar ett kommunikationssystem. Se Figur 2 som visar exempel på deterministisk och stokastisk signal samt hur den senare påverkar den förra.



Figur 2 Deterministisk (blå) signal $e^{-t/2\pi} \sin(t)$, stokastisk (röd) signal i form av vitt brus samt summan (grön) av dessa två signalerna. I ett verkligt fall är den nyttiga signalen summerad med brus liknande den gröna kurvan, vilket påverkar förmågan att avgöra vad det är för meddelande som överförs.

2.1.3. Kontinuerliga och diskreta signaler

Kontinuerliga signaler är signaler som har ett värde i varje tidpunkt (inom det tidsintervall där man definierat signalen). En diskret signal har värden vid vissa diskreta tidpunkter [1]. Tidpunkterna är oftast ekvidistant fördelade. Se exempel i Figur 3.



Figur 3 Exempel på kontinuerlig signal (grå kurva) och diskret signal (röda pilar). Den kontinuerliga signalen har ett värde $f(t)$ för alla tidpunkter där signalen är definierad, medan den diskreta signalen endast har värden vid vissa tidpunkter.
By No machine-readable author provided. Rbj assumed (based on copyright claims). - No machine-readable source provided. Own work assumed (based on copyright claims)., Public Domain, <https://commons.wikimedia.org/w/index.php?curid=870308>

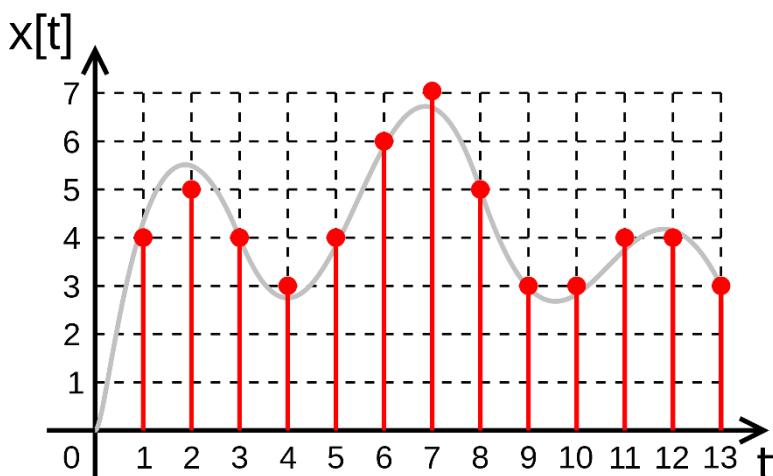
2.1.4. Analoga och digitala signaler

En analog signal är kort och gott en signal där amplituden kan ha vilket värde som helst (inom max och min-gränser) [1]. Det betyder alltså att den kan anta alla tänkbara funktionsvärden $f(t)$ inom ett givet intervall (som kan vara oändligt stort). En digital signal är då en signal som bara kan anta ett begränsat antal funktionsvärden. Man kan se det som en diskretisering även av signalamplituden.

Det är dock viktigt att påpeka att det finns något olika definitioner av analoga respektive signaler. I vissa fall så talar man om digitala signaler som är diskretiseraade både i tid och amplitud. Analoga signaler är definieras då som signaler kontinuerliga i både amplitud och tid. Att det finns olika definitioner är dock sällan ett problem, men kan leda till missförstånd.

Det ska också påpekas att det är inte alltid säkert att man håller sig till definitionen när man kommunickerar med varandra. T.ex. benämns en dataström på en databuss som en normalt som en digital signal, även om det i strikt mening är en analog signal! Detta p.g.a. att den fysiska spänningen inte är helt begränsad till vissa spänningsnivåer. Pulskanter har förstås en viss stiftid då spänningen rör sig mellan låg och hög spänning och antar därmed en oändlig mängd spänningsnivåer.

Tidsdiskreta signaler är normalt ett resultat av sampling [1], dvs. den process där en kontinuerlig signal avläses vid vissa tidpunkter. Tänk på att skilja sampling från diskretisering i amplitud, som är processen som ger en digital signal. För att få en digital signal mappas en analog signals funktionsvärdena till en definierad och diskret mängd av amplituder, se Figur 4. Detta brukar kallas A/D-omvandling eller analog-till-digital-omvandling [2].



Figur 4
Diskretisering av en analog signal (ljusgrå linje) i både tid (sampling) och amplitud (analog-till-digital-omvandling). I den senare delen kommer de samplade analoga amplitudvärdena från signalen $x(t)$ översättas till diskreta, digitala värden (röda punkter i figuren). Slutresultatet är signaler som kan representeras av tal i stället för en fysisk storhet, vilket öppnar möjligheten för behandling av signalen i en dator.
By Original: Rbj (talk · contribs)Derivative work: wdwd (talk · contribs) - This file was derived from: Digital.signal.svg:(zero-order sample&hold), Public Domain, <https://commons.wikimedia.org/w/index.php?curid=10864398>

2.1.5. Periodiska och icke-periodiska signaler

En signal är periodisk om den uppfyller villkoret $f(t) = f(t + T)$. Det minsta värdet på T som gör att villkoret är uppfyllt kallas för signalens periodtid (eller kort och gott period) [1]. Signaler som inte uppfyller detta villkor kallas för icke-periodiska signaler.

Detta villkor medför att en periodisk signal har ett upprepande mönster som pågår i all evighet. Ett viktigt samband är det mellan signalens periodtid T och grundfrekvens f_0 .

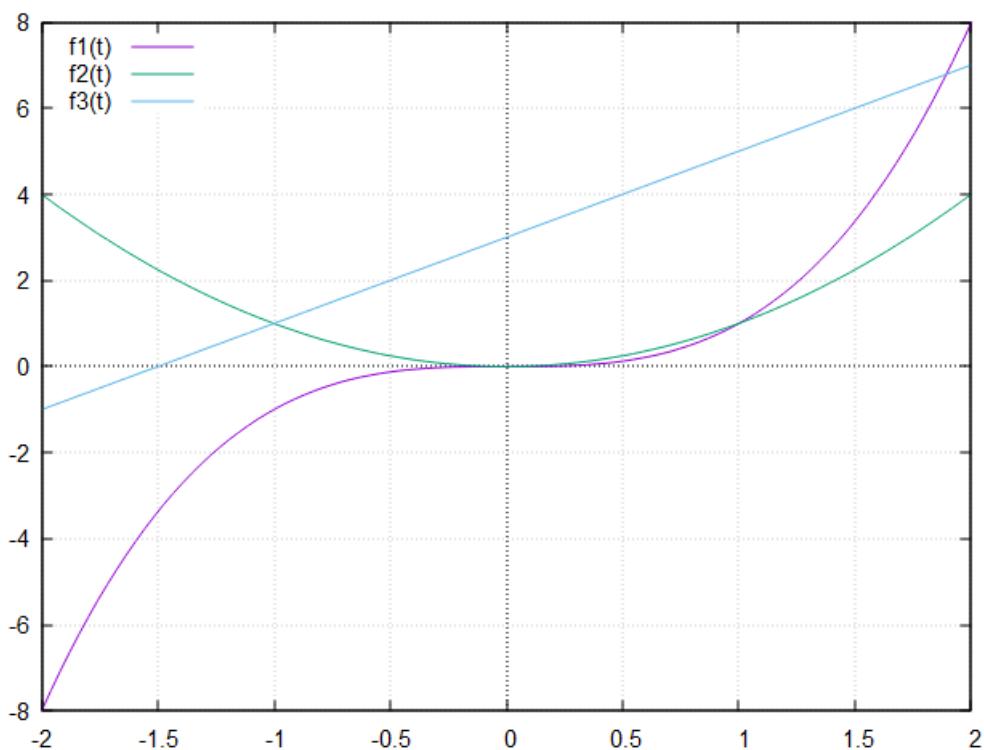
$$f_0 = \frac{1}{T} \quad \text{Ekvation 1}$$

Som visas senare så består en generell periodisk signal av fler frekvenser än bara grundfrekvensen, men Ekvation 1 ger dess längsta frekvens, därav namnet grundfrekvens.

Om signalen inte uppfyller villkoret för att vara periodisk är den icke-periodisk.

2.1.6. Udda och jämnna signaler

Jämna och udda funktioner [3] är matematiska funktioner som uppfyller vissa symmetrivillkor. En signal $f(t)$ är jämn om $f(-t) = f(t)$, udda om $f(-t) = -f(t)$. Jämna funktioners grafer är alltså symmetriska under spegling i y-axeln, medan udda funktioner är symmetriska under 180° rotation kring origo. Se Figur 5 för exempel.



Figur 5 Graf för signalerna $f_1(t) = t^3$, $f_2(t) = t^2$ och $f_3(t) = 2t + 3$. f_2 är ett exempel på en jämn funktion eftersom den speglas i diagrammets y-axel och f_1 är ett exempel på en udda funktion (den speglas i båda axlarna samtidigt). f_3 är varken udda eller jämn.

Två signaler som är viktiga i frekvensanalysen är cosinus och sinus. $\cos(t)$ är då jämn och $\sin(t)$ är udda.

Udda och jämnna funktioner har en del viktiga egenskaper:

- Den enda funktionen som är både jämn och udda är den konstanta funktionen $f(t) = 0$.

- Summan av två udda funktioner är udda, och varje multipel av en udda funktion är udda.
- Summan av två jämnna funktioner är jämnna, och varje multipel av en jämn funktion är jämn.
- Summan av en udda och en jämn funktion är varken udda eller jämn, såvida inte en av funktionerna är konstant noll.
- Produkten av såväl två udda som två jämnna funktioner är en jämn funktion.
- Produkten av en udda och en jämn funktion är en udda funktion.
- Kvoten av såväl två udda som två jämnna funktioner är jämn.
- Kvoten av en jämn och en udda funktion är udda.

De flesta signaler är varken udda eller jämnna, men varje signal $f(t)$ kan alltid delas upp i en summa $f = f_j + f_u$ av en jämn och en udda signal enligt Ekvation 2 och Ekvation 3.

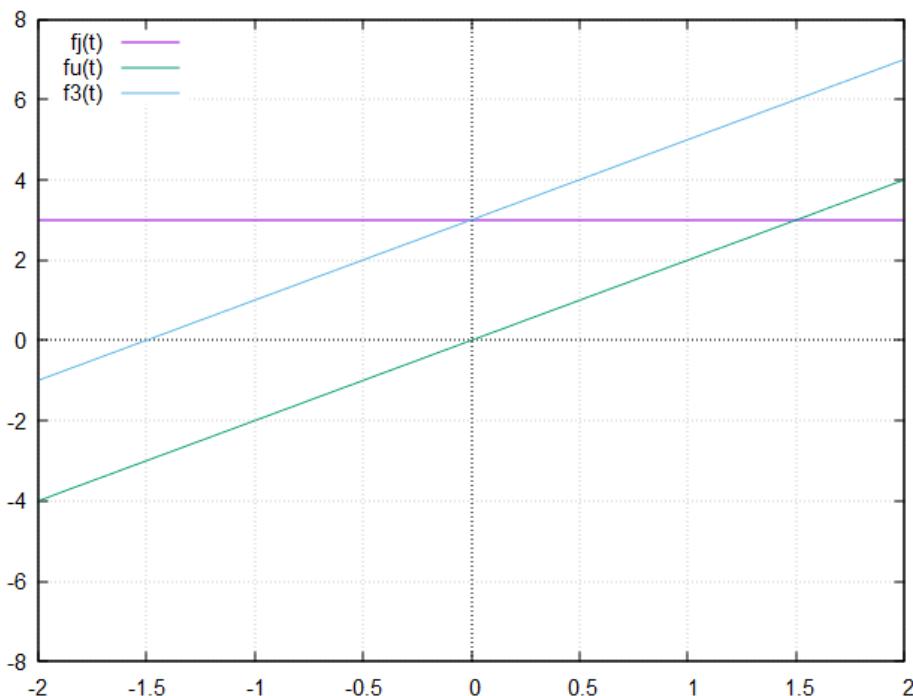
$$f_j(t) = \frac{f(t) + f(-t)}{2} \quad \text{Ekvation 2}$$

$$f_u(t) = \frac{f(t) - f(-t)}{2} \quad \text{Ekvation 3}$$

f_j är då alltid en jämn funktion eftersom den ger samma funktionsvärde för t och $-t$, dvs. grafen för f_j speglas i y -axeln. För f_u gäller att $f_u(t) = -f_u(-t)$ vilket inses om funktionsargumentet $-t$ sätts in i Ekvation 3.

$$f_u(-t) = \frac{f(-t) - f(t)}{2} = \frac{-f(t) + f(-t)}{2} = -f_u(t) \quad \text{Ekvation 4}$$

Ett exempel är uppdelningen av $f_3(t)$ från Figur 5 i en jämn och udda del, se Figur 6



Figur 6

Uppdelning av en signal i en udda och jämn del, $f_3(t) = f_j(t) + f_u(t)$.

2.1.7. Energisignaler och effektsignaler

Baserat på en signals energiinnehåll, kan den kategoriseras som antingen en energisignal eller en effektsignal. En energisignal är en signal med ändlig energi [1], vilket ges av integralen

$$E = \int_{-\infty}^{\infty} f^2(t)dt, \quad 0 < E < \infty \quad \text{Ekvation 5}$$

Detta är en generell beskrivning av signalens energi E , för en mer korrekt fysikalisk tolkning av Ekvation 5 kan man se t.ex. $f(t)$ som en spänning över en resistans 1Ω . Notera att medeleffekten för en energisignal över hela tidsintervallet $(-\infty, \infty)$ är noll, men i praktiken är man intresserad av momentaneffekten alternativt medeleffekten i ett begränsat interval $[t_0, t_0 + T]$ som då ges av

$$P = \frac{1}{T} \int_{t_0}^{t_0+T} f^2(t)dt \quad \text{Ekvation 6}$$

En effektsignal är en signal som har oändlig energi men ändlig medeleffekt. Detta definieras som ett gränsvärde

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t_0}^{t_0+T} f^2(t)dt, \quad 0 < P < \infty \quad \text{Ekvation 7}$$

2.1.8. Exempel på signaler och deras klassificering

Exempel 2.1

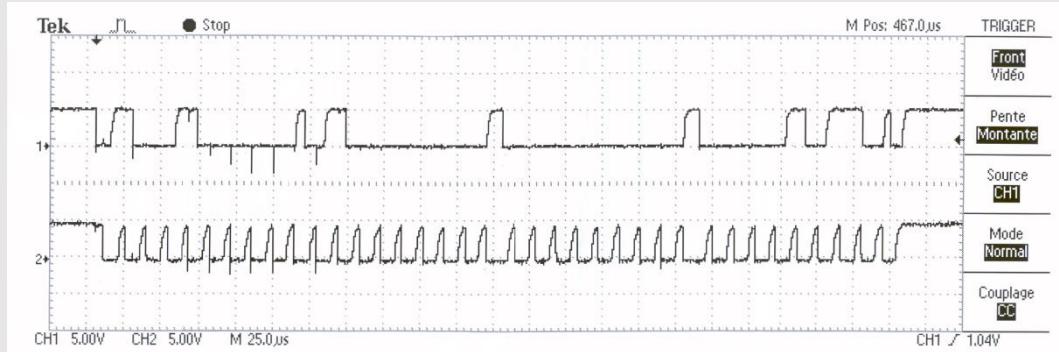
Signal: 230 V enfas nätpåspänning

Eftersom nätpåspänningen kan approximeras som $u(t) = 230\sin(2\pi \cdot 50 \cdot t + \alpha)$ där α är fasen är signalen är deterministisk, kontinuerlig och analog samt periodisk. I normalfallet är signalen varken udda eller jämn, men om $\alpha = 0$ är signalen udda och om $\alpha = 90^\circ$ är den jämn (eftersom den blir en ren cosinus). Signalen har oändlig energi och är därför inte en energisignal, men den kan klassificeras som en effektsignal eftersom har en ändlig medeleffekt.

Det är förstås viktigt att påpeka att denna klassificering gäller approximationen. Den verkliga signalen kommer t.ex. inte pågå i all evighet och har därför ändlig energi och kan egentligen klassificeras som energisignal. Dessutom varierar frekvensen och amplituden på nätpåspänningen lite och är därför inte deterministisk. De gäller alltså att ha sammanhanget klart för sig och förstå inom vilka förutsättningar som klassificeringen sker.

Exempel 2.2

Signal: Seriell I²C-kommunikation



Figur 7 I²C-signal (seriell busskommunikation). Övre grafen visar seriell data och undre visar klocksignalen.
Bilden delas som [public domain](#) av dess skapare, [Nutoj](#)

Mätningen av en I²C-signal (data respektive klocka) som visas i diagrammet ovan visar en signal som tydligt är kontinuerlig, icke-periodisk och varken udda eller jämn. Den kan klassificeras som stokastisk, dels för att den har ett visst brus, men också för att dataströmmen på en buss generellt sett inte kan förutses. När det gäller analog eller digital signal så är denna signal analog, även om man ofta talar om seriell datakommunikation som digitala signaler. Men det syns tydligt att kurvan antar alla tänkbara värden mellan sina min- och maxvärden eftersom flankerna har viss stigtid och viss falltid. Slutligen så är signalen en energisignal eftersom den har en ändlig utsträckning i tid. Medeleffekten över tid är noll.

Men sammanhanget är viktigt när klassificering sker. Om signalen hade haft försämbart brus och bättre flanker hade det varit möjligt att motivera en klassificering som digital signal. När det gäller energi och effekt så kan man se signalen som en effektsignal om den antas pågå över lång tid där energin går mot oändligheten och det då är möjligt att beräkna en medeleffekt.

2.2. Fourierserier

2.2.1. Fourierserier i reell form (trigonometrisk form)

Fourierserier [4], efter Jean-Baptiste Joseph Fourier, är en variant av Fouriertransformen för funktioner som är periodiska med periodiciteten T . Alternativt bara är definierade för ett intervall av längden T . Varje kontinuerlig periodisk funktion kan skrivas som summan av ett antal sinusfunktioner med varierande amplitud där varje sinusfunktion har en frekvens som är en heltalsmultipel av den lägsta frekvensen i den periodiska funktionen, $\frac{1}{T}$ (grundtonen).

Antag således att vi har en periodisk signal $x(t)$ med periodtiden T

$$x(t) = x(t + T)$$

Ekvation 8

$$T = \frac{1}{f_0} = \frac{2\pi}{\omega_0}$$

Ekvation 9

f_0 betecknar då signalens grundfrekvens och ω_0 är signalens grundvinkelfrekvens. Den periodiska signalen $x(t)$ kan då utvecklas i en summa som i det generella fallet har oändligt många termer

$$x(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega_0 t) + B_n \sin(n\omega_0 t)$$

Ekvation 10

Fourierkoefficienterna A_0 , A_n och B_n bestäms genom följande integraler

$$A_0 = \frac{2}{T} \int_{t_0}^{t_0+T} x(t) dt$$

Ekvation 11

$$A_n = \frac{2}{T} \int_{t_0}^{t_0+T} x(t) \cos(n\omega_0 t) dt$$

Ekvation 12

$$B_n = \frac{2}{T} \int_{t_0}^{t_0+T} x(t) \sin(n\omega_0 t) dt$$

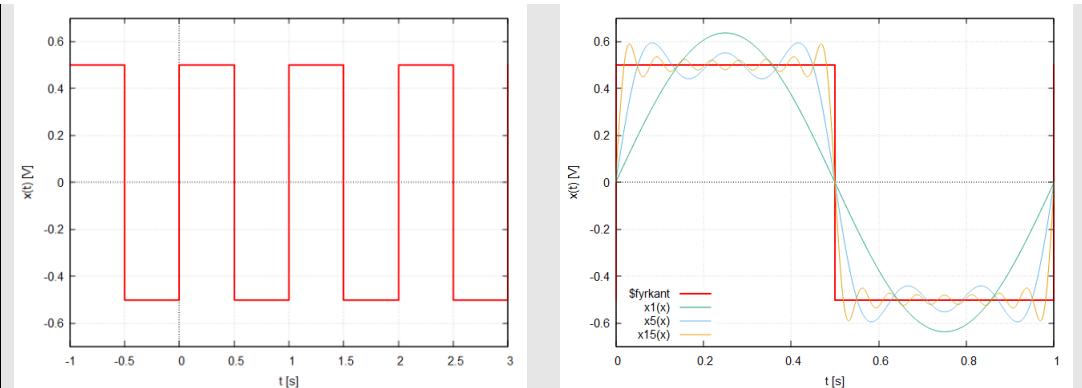
Ekvation 13

Beräkningen av koefficienten A_0 motsvarar den periodiska signalens dubbla medelvärde. Om signalen t.ex. är en spänningssignal så beräknas alltså signalens likspänningsskomponent som $\frac{A_0}{2}$, vilket gör att den första termen i Ekvation 10 blir enkel att tolka.

Beräkningen av A_n och B_n har inte en lika enkel, intuitiv tolkning. Funktionen $x(t)$ multipliceras i Ekvation 12 och Ekvation 13 med cosinus- respektive sinussignalen vars frekvenser är en multipel n av grundfrekvensen för att sedan integreras över en period. En pragmatisk tolkning är att detta ger ett mått på hur "lik" signalen är cosinus- och sinussignalerna som ingår i multiplikationen. Koefficienterna blir då ett mått på denna "likhet". Integralen upprepas sedan för olika värden på n för att finna koefficienterna för de olika vinkelfrekvenserna $n\omega_0$.

Exempel 2.3

Beräkna Fourierkoefficienterna för fyrkantsvägen (spänning) i Figur 8, vänster graf. Analysera resultatet.



Figur 8 Vänster graf visar en fyrkantsvåg med amplituden 0.5 periodtiden 1 s. Höger graf visar samma signal inzoomad till en period, samt summan för de olika Fourierkomponenterna för olika n (1, 5 och 15).

Genom att studera signalens graf ser man att medelvärdet bör vara noll, dvs. vi ska förvänta oss att A_0 ska vara noll. Antagandet stämmer, eftersom integralen i Ekvation 11 med $t_0 = 0$ och $T = 1$ ger

$$A_0 = \frac{2}{T} \int_{t_0}^{t_0+T} x(t) dt = \frac{2}{1} \int_0^{0.5} 0.5 dt + \frac{2}{1} \int_{0.5}^1 -0.5 dt = 2[0.5t]_0^{0.5} + 2[-0.5t]_{0.5}^1 = \\ = 2(0.25 - 0) + 2(-0.5 + 0.25) = 0$$

A_n och B_n beräknas med Ekvation 12 och Ekvation 13, där $\omega_0 = \frac{2\pi}{T}$.

$$A_n = \frac{2}{T} \int_{t_0}^{t_0+T} x(t) \cos(n\omega_0 t) dt = \\ = 2 \int_0^{0.5} 0.5 \cos\left(n \cdot \frac{2\pi}{T} \cdot t\right) dt + 2 \int_{0.5}^1 -0.5 \cos\left(n \cdot \frac{2\pi}{T} \cdot t\right) dt = \\ = 2 \left[0.5 \frac{T}{n \cdot 2\pi} \sin\left(n \cdot \frac{2\pi}{T} \cdot t\right) \right]_0^{0.5} + 2 \left[(-0.5) \frac{T}{n \cdot 2\pi} \sin\left(n \cdot \frac{2\pi}{T} \cdot t\right) \right]_{0.5}^1 = \\ = \frac{1}{n \cdot 2\pi} (\sin(n \cdot 2\pi \cdot 0.5) - \sin(n \cdot 2\pi \cdot 0) - \sin(n \cdot 2\pi \cdot 1) + \sin(n \cdot 2\pi \cdot 0.5)) = \\ = \frac{1}{n \cdot 2\pi} (0 - 0 - 0 + 0) = 0$$

$$B_n = \frac{2}{T} \int_{t_0}^{t_0+T} x(t) \sin(n\omega_0 t) dt =$$

$$\begin{aligned}
&= 2 \int_0^{0.5} 0.5 \sin\left(n \cdot \frac{2\pi}{T} \cdot t\right) dt + 2 \int_{0.5}^1 -0.5 \sin\left(n \cdot \frac{2\pi}{T} \cdot t\right) dt = \\
&= 2 \left[\frac{-0.5T}{n \cdot 2\pi} \cos\left(n \cdot \frac{2\pi}{T} \cdot t\right) \right]_0^{0.5} + 2 \left[\frac{0.5T}{n \cdot 2\pi} \cos\left(n \cdot \frac{2\pi}{T} \cdot t\right) \right]_{0.5}^1 = \\
&= \frac{1}{n \cdot 2\pi} (\cos(n \cdot 2\pi \cdot 0) - \cos(n \cdot 2\pi \cdot 0.5) + \cos(n \cdot 2\pi \cdot 1) - \cos(n \cdot 2\pi \cdot 0.5)) = \\
&= \frac{1}{n \cdot 2\pi} (1 - \cos(n\pi) + 1 - \cos(n\pi)) = \frac{1}{n\pi} (1 - \cos(n\pi)) = \begin{cases} 0, & n \text{ jämnt} \\ \frac{2}{n\pi}, & n \text{ udda} \end{cases}
\end{aligned}$$

B_n -termerna försvinner då n är jämnt eftersom $\cos(2\pi), \cos(4\pi), \cos(6\pi), \dots$ alla blir 1. Då n är udda blir $\cos(\pi), \cos(3\pi), \cos(5\pi), \dots$ lika med -1.

Resultatet från beräkningen av A_n - och B_n -termerna är intressant att analysera. Eftersom fyrkantsvägen i Figur 8 är en udda signal kommer Fourierserien inte innehålla några cosinustermer! Detta pga. att en udda funktion bara kan innehålla summatermer från udda funktioner eftersom udda och jämma funktioner har denna egenskap. Vi hade alltså på förhand kunnat se att A_n -termerna är noll.

Den högra grafen i Figur 8 visar hur fyrkantsvägen successivt byggs upp av summan av sinuskurvor. I grafen finns kurvorna för $n = 1, n = 5$ och $n = 15$. Även vid $n = 15$ så ser man att betydligt fler termer (dvs. högre frekvenser) behöver läggas till för att få en bra fyrkantsväg. En fyrkantsväg har alltså mycket övertoner, vilket beror på att de branta flankerna på pulserna kräver snabba svängningar och alltså höga frekvenser.

Exempel 2.1 visar hur udda signaler "byggs upp" av de udda trigonometriska funktionerna, dvs. sinussignaler. Om fyrkantsvägen hade varit jämn hade följaktligen bara cosinustermer funnits i Fourierserien. Detta uppnås genom att fasförskjuta fyrkantsvägen en kvarts period så att den blir symmetrisk kring y-axeln. Hade vi genomfört beräkningarna för en sådan fyrkantsväg hade vi alltså bara fått A_n -termerna kvar.

2.2.2. Fourierserier i komplex form

Eulers formel $e^{j\varphi} = \cos(\varphi) + j \sin(\varphi)$ [5] ger oss en möjlighet att skriva cosinus- och sinustermer på en komplex form

$$\cos(\varphi) = \frac{e^{j\varphi} + e^{-j\varphi}}{2} \quad \text{Ekvation 14}$$

$$\sin(\varphi) = \frac{e^{j\varphi} - e^{-j\varphi}}{2j} \quad \text{Ekvation 15}$$

Detta kan användas för att uttrycka Fourierserien i Ekvation 10 på en mer kompakt, komplex form med komplexa Fourierkoefficienter C_n .

$$\begin{aligned}
 x(t) &= \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega_0 t) + B_n \sin(n\omega_0 t) = \\
 &= \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \frac{e^{jn\omega_0 t} + e^{-jn\omega_0 t}}{2} + B_n \frac{e^{jn\omega_0 t} - e^{-jn\omega_0 t}}{2j} = \\
 &= \frac{A_0}{2} + \sum_{n=1}^{\infty} \frac{1}{2} (A_n - jB_n) e^{jn\omega_0 t} + \frac{1}{2} (A_n + jB_n) e^{-jn\omega_0 t} \quad \text{Ekvation 16}
 \end{aligned}$$

Eftersom $(A_n + jB_n)$ är komplexkonjugatet av $(A_n - jB_n)$ kan vi skriva om Ekvation 16 på en väldigt kompakt form med nya koefficienter $C_0 = A_0/2$, $C_n = (A_n - jB_n)/2$ och $C_{-n} = (A_n + jB_n)/2 = C_n^*$ (C_n^* betecknar komplexkonjugatet av C_n). Notera här att vi utvidgar summan till negativa n för att slutligen skriva Fourierserien som

$$x(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t} \quad \text{Ekvation 17}$$

Det är intressant att tolka Ekvation 17 eftersom summan nu innehåller negativa n , vilket på ett sätt innebär att vi har introducerat negativa vinkel frekvenser $n\omega_0$. I verkligheten har vi förstås bara positiva frekvenser, det gäller att komma ihåg att termerna C_n och C_{-n} alltid uppträder parvis enligt Ekvation 16 och att $C_n e^{jn\omega_0 t} + C_{-n} e^{-jn\omega_0 t}$ blir en positiv frekvens $n\omega_0$ bestående av en cosinus + sinus. Men tankesättet med negativa frekvenser lever kvar i signalbehandlingen eftersom det är väldigt praktiskt att arbeta med den komplexa skrivformen.

Koefficienterna C_n kan beräknas med en integral på motsvarande sätt som vi beräknar koefficienterna för den reella formen

$$C_n = \frac{1}{T} \int_{t_0}^{t_0+T} x(t) e^{-jn\omega_0 t} dt \quad \text{Ekvation 18}$$

Även denna ekvation blir alltså väldigt kompakt.

Exempel 2.4

Beräkna komplexa Fourierkoefficienterna för fyrkantsvågen i Exempel 2.1 och jämför med de reella koefficienterna.

Ekvation 18 ger med $t_0 = 0$, $T = 1$ och $\omega_0 = \frac{2\pi}{T} = 2\pi$ att

$$\begin{aligned}
 C_n &= \frac{1}{T} \int_{t_0}^{t_0+T} x(t) e^{-jn\omega_0 t} dt = \\
 &= \int_0^{0.5} 0.5 e^{-jn\omega_0 t} dt + \int_{0.5}^1 -0.5 e^{-jn\omega_0 t} dt =
 \end{aligned}$$

$$\begin{aligned}
 &= \left[\frac{0.5}{-jn \cdot 2\pi} e^{-jn \cdot 2\pi \cdot t} \right]_0^{0.5} + \left[\frac{-0.5}{-jn \cdot 2\pi} e^{-jn \cdot 2\pi \cdot t} \right]_{0.5}^1 = \\
 &= \frac{j}{n \cdot 4\pi} (e^{-jn \cdot 2\pi \cdot 0.5} - e^{-jn \cdot 2\pi \cdot 0}) - \frac{j}{n \cdot 4\pi} (e^{-jn \cdot 2\pi \cdot 1} - e^{-jn \cdot 2\pi \cdot 0.5}) = \\
 &= \frac{j}{n \cdot 4\pi} (e^{-jn\pi} - 1 + e^{-jn\pi} - e^{-jn \cdot 2\pi}) = \frac{j}{n \cdot 2\pi} (e^{-jn\pi} - 1)
 \end{aligned}$$

Termen $e^{-jn\pi}$ är 1 då n är jämnt, vilket inses enkelt med hjälp av komplexa talplanets enhetscirkel. Därmed försätter alla koefficienter med jämn n precis som i Exempel 2.1. Motsvarande, då n är udda är $e^{-jn\pi} = -1$, vilket slutligen ger koefficienterna

$$C_n = \begin{cases} 0, n \text{ jämnt} \\ -\frac{j}{n\pi}, n \text{ udda} \end{cases}$$

Jämfört med koefficienterna i Exempel 2.1 skiljer det en faktor 2 (förutom att koefficienten är komplex), vilket är en konsekvens av att vi nu har en summa som går från $-\infty$ till ∞ och termerna $-n$ och n "kommer parvis".

Sätter vi in resultatet i Ekvation 17 så får vi

$$\begin{aligned}
 x(t) &= \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t} = \sum_{n \text{ udda}} -\frac{j}{n\pi} e^{jn\omega_0 t} = \sum_{n>0, \text{ udda}} -\frac{j}{n\pi} e^{jn\omega_0 t} + \frac{j}{n\pi} e^{-jn\omega_0 t} = \\
 &= \sum_{n>0, \text{ udda}} \frac{2}{2jn\pi} e^{jn\omega_0 t} - \frac{2}{2jn\pi} e^{-jn\omega_0 t} = \sum_{n>0, \text{ udda}} \frac{2}{n\pi} \sin(n\omega_0 t)
 \end{aligned}$$

Denna summa är som väntat samma resultat som i Exempel 2.1, dvs. $x(t)$ skrivs som en serie av sinustermer $\sin(n\omega_0 t)$ med koefficienterna $\frac{2}{n\pi}$ där n är udda (och positivt).

Beräkningarna visar att den komplexa formen har tydliga fördelar. Den ger ett mer kompakt skrивsätt och alla koefficienter kan beräknas med en enda integral i.st.f. tre som för reella formen. Integralen blir i en del fall mindre arbetsam att lösa också eftersom exponentialfunktionen med talet e som bas har en enkel primitiv funktion.

2.2.3. Fourierserier i amplitud-fas-form

Eftersom summan av en cosinus- och en sinussignal med samma frekvens kan skrivas som en fasförskjuten cosinussignal (eller fasförskjuten sinussignal) går det naturligtvis att skriva Fourierserier i amplitud-fas-form

$$x(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A'_n \cos(n\omega_0 t - \varphi_n) \quad \text{Ekvation 19}$$

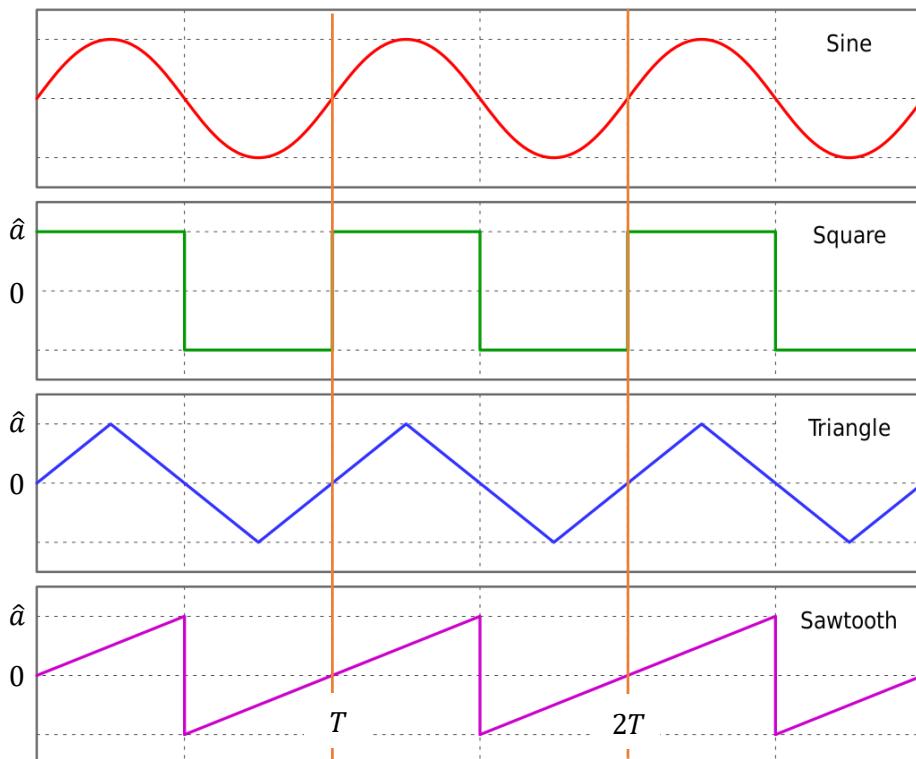
Detta skrivsätt är inte lika vanligt förekommande som reella och komplexa Fourierserier och beskrivs därför inte mer utförligt i denna litteratur.

2.2.4. Fourierserier för några vanliga periodiska signaler

Figur 9 visar några vanliga periodiska funktioner med perioden T och amplituden \hat{a} , alla med medelvärdet 0. Fourierkoefficienterna för fyrkantsvåg, triangelvåg och sågtandsvåg finns i Ekvation 20 till Ekvation 25.

Observera att om det finns ett medelvärde skilt från 0, kommer koefficienten A_0 också vara skilt från 0. Observera också att om en förskjutning med $T/4$ sker längs tidsaxeln och signalen går från jämn till udda (och vice versa), kommer A_n och B_n "byta värden". Med detta avses att koefficienternas storlek är samma för respektive heltalet n , men det kan skifta tecken.

Resonemangen gäller bara om signalen verkligen är rent udda eller jämn efter förskjutningen med $T/4$. I Figur 9 så gäller det för fyrkantsvågen och triangelvågen, men inte sågtandsvågen. Den är udda i Figur 9, men flyttas den $T/4$ är den vare sig udda eller jämn.



Figur 9
Några vanliga periodiska funktioner.
By Omegatron - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=343520>

Fourierkoefficienter, fyrkantsvåg:

$$A_0 = 0, \quad A_n = 0$$

Ekvation 20

$$B_n = \begin{cases} 0, & n = 2, 4, 6, \dots \\ \frac{4\hat{a}}{n\pi}, & n = 1, 3, 5, \dots \end{cases}$$

Ekvation 21

Fourierkoefficienter, triangelvåg

$$A_0 = 0, A_n = 0$$

Ekvation 22

$$B_n = \begin{cases} 0, & n = 2, 4, 6, \dots \\ \frac{8\hat{a}}{n^2\pi^2}(-1)^{(n-1)/2}, & n = 1, 3, 5, \dots \end{cases}$$

Ekvation 23

Fourierkoefficienter, sågtandsvåg

$$A_0 = 0, A_n = 0$$

Ekvation 24

$$B_n = \frac{2\hat{a}}{n\pi}(-1)^{n-1}$$

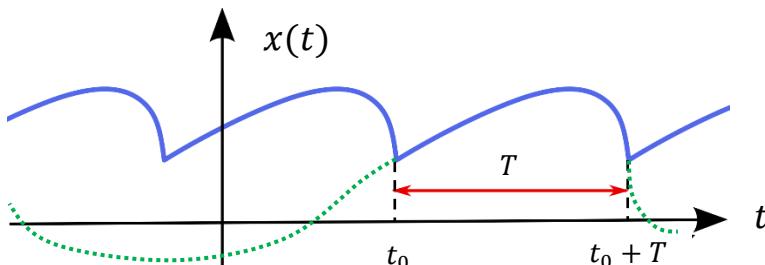
Ekvation 25

2.3. Fouriertransformen

Fourierserier hjälper oss att analysera periodiska signaler, men de flesta signaler som vi behöver behandla i kommunikationsteori och signalbehandling är icke-periodiska. Men uttrycken för en Fourierseries koefficienter hjälper oss att förstå hur vi kan expandera frekvensanalysen så att även icke-periodiska signaler kan analyseras.

Antag att vi har en godtycklig icke-periodisk signal $x(t)$ som vi analyserar i ett begränsat tidsintervall $[t_0, t_0 + T]$. Analysen kan ske med Fourierserier om vi omformar $x(t)$ till en periodisk funktion som upprepar sig utanför intervallet $[t_0, t_0 + T]$, se Figur 10.

Självklart är det så att analysen bara är giltig i intervallet $[t_0, t_0 + T]$, utanför detta interval har vi modifierat signalen så att den inte längre överensstämmer med det ursprungliga utseendet.



Figur 10

Analys av en icke-periodisk signal (streckad grön linje) över ett begränsat interval T (markerat med röd pil). Signalen kan omformas till en periodisk signal (heldragen blå linje) genom att upprepa funktionen i intervallet $[t_0, t_0 + T]$ om och om igen.

Figuren bygger på original från Wikipedia - By Oleg Alexandrov - Own work, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=2683866>

Ekvation 17 och Ekvation 18 beskriver hur $x(t)$ kan skrivas som en Fourierserie i detta interval. Antag nu att periodtiden T växer så att intervallet $[t_0, t_0 + T]$ växer och vi analyserar en större och större del av den ursprungliga funktionen. Vad händer då?

Förutom att vi kan representera en större del av den icke-periodiska funktionen med hjälp av en Fourierserie, kommer grundfrekvensen f_0 minska. Det frekvensspektrum som Fourierserierna ger kommer alltså ha tätare och tätare mellan de olika Fourierkoefficienternas frekvenser $n \cdot 2\pi f_0$. Samtidigt minskar storleken på koefficienterna, eftersom det ingår en multiplikation med $\frac{1}{T}$ i Ekvation 18. Om vi låter $T \rightarrow \infty$, kommer frekvensspektrum övergå till ett kontinuerligt frekvensspektrum eftersom $f_0 \rightarrow 0$. Vi kan se det som att vi därmed har oändligt många cosinusformade signaler som ligger oändligt tätt packade i frekvens, där amplituden på varje cosinussignal är oändligt liten (men inte noll).

Ekvation 17 och Ekvation 18 övergår då i det som kallas Fouriertransformen [6], vilket är det sambandet mellan en signal $x(t)$ och dess frekvensspektrum $X(\omega)$ (amplitudspektrum). Fouriertransformen som funktion ges i litteraturen ofta en egen beteckning \mathcal{F} .

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad \text{Ekvation 26}$$

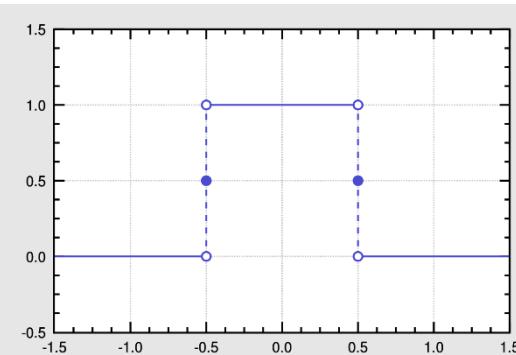
$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{j\omega t} d\omega \quad \text{Ekvation 27}$$

Ekvation 26 kallas normalt för Fouriertransform (FT) medan Ekvation 27 kallas för Invers Fouriertransform (IFT). $X(\omega)$ är inte bara begränsad till positiva vinkelfrekvenser som är den vanliga synen på frekvenser inom fysik och ellära, utan frekvensspektrum består även av negativa frekvenser. Men positiva och negativa frekvenser uppträder alltid parvis symmetriskt kring $\omega = 0$ på ett sådant sätt att de adderas till cosinus-, respektive sinussignaler. Det inses enklast med hjälp av Eulers formel liknande resonemanget i kapitel 2.2.2.

I praktiken är det få signaler där frekvensspektrum kan beräknas genom att lösa integralen. Integralen i Ekvation 34 kan bara lösas analytiskt för ett begränsat antal funktioner. Men Ekvation 34 (och Ekvation 35) kan förstås beräknas med numeriska metoder. Som vi ska se i kapitel 2.4 finns mycket effektiva algoritmer för diskreta signaler.

Exempel 2.5

Beräkna Fouriertransformen för en rektangulär puls med topp-topp-värdet 1 och längden 1 enligt Figur 11

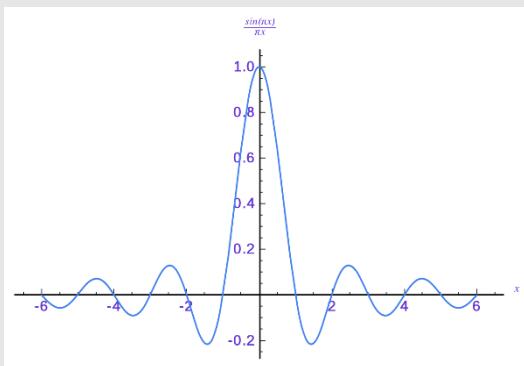


Figur 11 Rektangulär puls med längden 1 och topp-topp-värde 1.
CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=801402>

Fouriertransformen beräknas med Ekvation 26, där funktionen $x(t) = 1$ om $-0,5 \leq t \leq 0,5$ och $x(t) = 0$ för övrigt. Det ger en ganska enkel integral.

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt = \int_{-0,5}^{0,5} 1 \cdot e^{-j\omega t} dt = \left[\frac{1}{-j\omega} e^{-j\omega t} \right]_{-0,5}^{0,5} = \\ &= \frac{1}{-j\omega} (e^{-j\omega \cdot 0,5} - e^{-j\omega \cdot (-0,5)}) = \frac{1}{j\omega} (e^{j\omega/2} - e^{-j\omega/2}) = \\ &= \frac{1}{j\omega} \cdot 2j \cdot \sin(\omega/2) = \frac{\sin(\omega/2)}{\omega/2} = \frac{\sin(\pi f)}{\pi f} \end{aligned}$$

Det tredje sista ledet har skrivits om med Eulers formel och dess samband med sinusfunktionen, $\sin(x) = \frac{e^{ix} - e^{-ix}}{2j}$.



Figur 12 Fouriertransformen (frekvensspektrum) för en rektangulär puls. Diagrammet visar att frekvensspektrum
CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=801402>

Figur 12 visar att den rektangulära pulsen har ett amplitudspektrum som domineras av låga frekvenser, men att det även finns en relativt stor energi i högre frekvenser eftersom kurvan klingar av långsamt. Bandbredden är ca 1 Hz för denna puls med längden 1 s, och en tumregel är att bandbredden för en rektangulär puls med längden T är $B \approx 1/T$. Det exakta värdet beror förstås på hur man väljer att definiera bandbredd.

2.4. Diskreta Fouriertransformer och FFT

2.4.1. Diskret-tid Fouriertransform, DTFT

Vid sampling av en tidsdiskret signal $x(t)$ gäller generellt att den kontinuerliga tidsfunktionen övergår till en diskret representation av signalen. Denna representation kan ses som en vektor $x[n]$ av reella tal där n är ett heltalet $-\infty < n < \infty$. Notera övergången till en notation med hakparenteser för att tydligare skilja den kontinuerliga signalen $x(t)$ från dess samplade version $x[n]$.

Om samplingen sker med jämna mellanrum bestäms $x[n]$ av sambandet

$$x[n] = x(nT_s) \quad \text{Ekvation 28}$$

där T_s är samplingstiden. Sambandet mellan samplingstid och samplingsfrekvens är förstås $f_s = 1/T_s$.

Vektorn kan tänkas vara oändligt stor om samplingen görs över en oändlig tidsperiod, men i praktiken är $x[n]$ förstås en ändlig vektor med N stycken antal element. Av praktiska skäl låter man samplingen i det fallet inledas vid $t = 0$ vilket ger att n sträcker sig mellan $0 \leq n \leq N - 1$.

Här finns en del begreppsförvirring, eftersom Fouriertransformen för en diskretiserad signal med oändlig utsträckning i tid kallas för diskret-tid Fouriertransform [7] (eng. Discrete Time Fourier Transform, DTFT) medan Fouriertransformen för en samplad signal med ändlig utsträckning kallas för diskret Fouriertransform [8] (eng. Discrete Fourier Transform, DFT). De skiljer sig åt en del, även om den matematiska grunden är gemensam. DFT behandlas i kapitel 0.

Vid sampling och DTFT (samt även DFT) så går egentligen information om tiden förlorad. Den kontinuerliga tidsfunktionen $x(t)$ har en tidsaxel och varje tidpunkt har ett givet funktionsvärde. Men $x[n]$ är bara en sekvens av reella tal. Denna vektor skulle kunna vara resultatet av sampling av signalen $\sin(2\pi \cdot 50 \cdot t)$ lika gärna som sampling av signalen $\sin(2\pi \cdot 2500 \cdot t)$ beroende på valet av samplingsfrekvens. Vid frekvensanalys av samplade signaler så är det därför ofta praxis att göra analysen med normerade frekvenser $f_n = f/f_s$. f_n blir då en dimensionslös storhet. Behöver en översättning till verkliga frekvenser göras är det bara att räkna om f_n genom $f = f_n \cdot f_s$. Motsvarande gäller självklart vinkelfrekvenser med $\omega_n = 2\pi f/f_s$. Genom att sätta in den samplade signalen $x[n] = x(nT_s)$ i Ekvation 26 och med en övergång till normerad vinkelfrekvens kan integralen i denna ekvation skrivas om som en summa

$$\begin{aligned} X(\omega_n) &= \sum_{n=-\infty}^{\infty} x(nT_s)e^{-j\omega_n T_s} = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega_n/f_s} \\ &= \sum_{n=-\infty}^{\infty} x[n]e^{-jn\omega_n} \end{aligned} \quad \text{Ekvation 29}$$

Övergången från integral till summa sker eftersom $x[n]$ inte är en kontinuerlig funktion utan bara existerar vid enstaka tidpunkter då samplingen gjorts. Ekvation 29 är beräkningen av den samplade signalens (normerade) frekvensspektrum och den kallas för således för diskret-tid Fouriertransform, DTFT.

Här bör noteras att $X(\omega_n)$ är en kontinuerlig funktion även om $x[n]$ är diskret. Den är också periodisk i frekvensplanet med periodiciteten $\omega_s = 2\pi f_s = 2\pi/T_s$. Egenskapen att den är periodisk skiljer den från Fouriertransformen (kapitel 2.3), som generellt inte är periodisk. Men vid $-\pi < \omega_n < \pi$ (motsvarande $-0.5 < f_n < 0.5$) har $X(\omega_n)$ från Ekvation 29 samma frekvensspektrum som $X(\omega)$ från Ekvation 26, givet en viktig, fundamental förutsättning. Denna förutsättning är att samplingsfrekvensen f_s måste vara minst dubbelt så stor som signalen $x(t)$:s högsta frekvens. Detta finns beskrivet i mer detalj i kapitel 2.4.4. Detta är ett mycket viktigt resultat, eftersom det visar att det går att överföra signalanalys och filtrering till den digitala världen utan att frekvensinformationen om den analoga signalen går förlorad! Ett annat sätt att formulera denna slutsats är att en tidskontinuerlig signal helt kan återskapas från sina samplade värden förutsatt att samplingen genomförs med tillräckligt hög samplingsfrekvens.

Precis som för Fouriertransfomen i Ekvation 26 finns en inverstransform till DTFT för att beräkna $x[n]$. Den beräknas med följande integral

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega_n) e^{j\omega_n t} d\omega_n \quad \text{Ekvation 30}$$

Integrationsgränserna $-\pi$ och π i Ekvation 30 hänger samman med periodiciteten för $X(\omega_n)$. Eftersom ω_n är normerad till samplingsfrekvensen är dess periodicitet 2π .

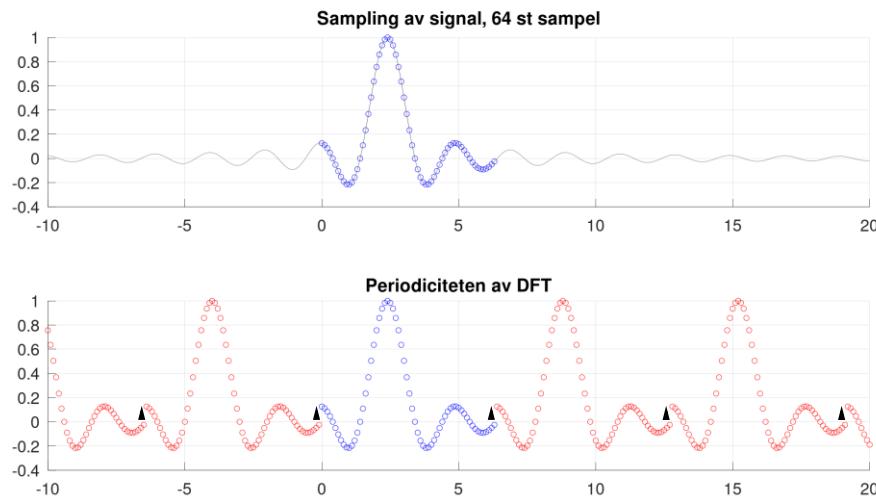
2.4.2. Diskret Fouriertransform, DFT

I praktiken består en samplad signal alltid av ett ändligt antal sampel. Därmed är DTFT mer av teoretisk nytta än praktisk nytta. För en samplad signal $x[n]$ med ändligt antal sampel definieras den diskreta Fouriertransformen DFT [8] som

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j \cdot 2\pi k n / N} \quad k = 0, 1, 2, \dots, N-1 \quad \text{Ekvation 31}$$

$X[k]$ är en vektor med frekvenser, dvs. $X[k]$ är en diskret funktion. Index k representerar då en serie av totalt N styck frekvenser. Jämför med en Fourierserie, den består av diskreta frekvenser vilket är ett resultat av att Fourierserien beräknas på en periodisk funktion. Samma sak gäller även för Ekvation 31. Den är resultatet av att en diskret-tid Fouriertransform appliceras på en samplad signal $x[n]$ med ändligt antal sampel som tänkes vara återupprepade i ett oändligt periodiskt mönster. Denna återupprepning visualiseras i Figur 13. Den samplade signalen som är markerad med blå ringar kommer i den diskreta Fouriertransformen tolkas som en periodisk funktion (under diagrammet med blå + röda ringar). I figuren syns också ett problem som

introduceras med samplingen. Den periodiska funktionen som blir resultatet av DFT:s periodisering kommer introducera icke-fysikaliska, snabba hopp (markerade med svarta pilar). Snabba förändringar innehåller förekomst av höga frekvenser. Dessa hopp kommer alltså introducera fel i det beräknade frekvensspektrumet.



Figur 13 Översta diagrammet visar sampling (64 st. sampel, blå ringar) av en signal. Eftersom samplingen är begränsad i hur många sampel som kan tas kommer den endast ta med en del av hela den tidskontinuerliga funktionen. När denna samplade signal sedan används i den diskreta Fouriertransformen motsvarar det att genomföra en diskret-tid Fouriertransform av den periodiska funktionen som visas i undre diagrammet (blå + röda ringar). Notera att den periodiska funktionen uppvisar tydliga diskontinuerliga hopp (markerade med svarta pilare) som inte finns i den ursprungliga signalen. Snabba förändringar i en periodisk funktion innehåller förekomst av högre frekvenser. DFT kan därför introducera fel i det beräknat frekvensspektrum jämfört med originalsignalens spektrum.

Självklart har DFT en inverstransform också, vilken ges av den komplexa summan i Ekvation 32 nedan.

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j \cdot 2\pi kn/N} \quad n = 0, 1, 2, \dots, N-1 \quad \text{Ekvation 32}$$

Analogt med resonemanget med sampling och att $x[n]$ är en begränsad sampling av en kontinuerlig tidsfunktion $x(t)$, kan man se $X[k]$ som en sampling i frekvensplanet av en kontinuerlig frekvensfunktion $X(\omega)$. Diskussionen kring Figur 13 och att en sampling med begränsat antal sampel motsvarar en sampling av en periodisk funktion med oändligt antal sampel kan då överföras till $X[k]$. Det

frekvensspektrumet $X[k]$ kan alltså ses som ett diskret frekvensspektrum som är periodiskt i frekvensplanet.

2.4.3. Snabb Fouriertransform, FFT

Ytterligare ett begrepp, eller snarare förkortning, som bidrar till begreppsförvirringen är FFT [9], som står för snabb Fouriertransform (eng. Fast Fourier Transform). FFT är det som normalt nämns när signalbehandlare inom olika discipliner tänker på analys av signaler och dess frekvenspektrum.

FFT är helt enkelt en snabb algoritm för DFT. Att implementera summan i Ekvation 31 i kod ger en algoritm som är tidskrävande vid långa samplingssekvenser. Antalet beräkningar växer med en faktor N^2 . FFT växer i stället med en faktor $N \cdot \log N$, vilket gör stor skillnad vid stora N . FFT visar bäst effektivitet när N är en multipel av 2 eftersom algoritmen successivt delar upp datamängden i hälften. Därför är det vanligt att antalet sampel är en multipel av 2, t.ex. 1024, 2048, 4196 osv.

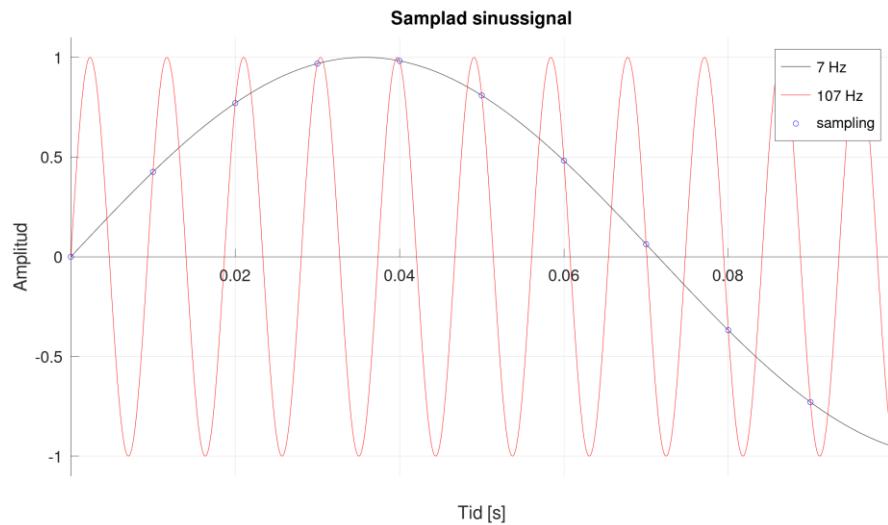
FFT finns implementerad i många olika programvaror för databehandling. Algoritmen finns förstås också färdigkodad och möjlig att ladda ner till alla möjliga programmeringsspråk.

2.4.4. Sampling och Nyquist-Shannons samplingsteorem

I kapitel 2.1.3 och 2.1.4 beskrivs kortfattat hur en analog signal kan diskretiseras i tid genom sampling. Begreppet sampling används ofta som ett samlingsnamn för en tvåstegsprocess. Först diskretiseras en tidskontinuerlig signal genom att den läses av vid vissa, diskreta tidpunkter (Figur 3), vilket ger en vektor med reella tal. I steg två så görs den tidsdiskreta signalen även diskret i funktionsvärde så att signalen också kan digitaliseras (Figur 4). I praktiken genomförs detta ofta i en och samma A/D-omvandlare, men i den teoretiska behandlingen av digitala filter väljer vi att behandla tidsdiskreta signaler, dvs. signaler som bara genomgått första steget.

Kapitel 2.1.3 och 2.1.4 nämner inget om eventuella begränsningar med att sampla och diskretisera en signal. Rent intuitivt så känns samplingsprocessen som något som kastar bort oändligt mycket data eftersom det är oändligt många av signalens funktionsvärden som ratas. För att ha en god diskret representation av den analoga, kontinuerliga signalen känns det därför naturligt att sampla tätt. Samtidigt så kommer detta generera mycket data som ska processas. För att spara på lagringsutrymme och processorkraft bör man alltså sampla så glest som möjligt. Den viktiga frågeställningen är därför hur glest man kan sampla utan att förlora signalkvalitet.

Figur 14 visar hur en 7 Hz sinussignal kan förväxlas med en signal med högre frekvens (107 Hz) efter sampling. Efter att sampling har genomförts kan dessa frekvenskomponenter inte längre skiljas åt. Tidsintervallet T_s mellan samplingspunktarna måste alltså minska så att signalerna sampelas tätare.



Figur 14 Exempel på Nyquistfrekvensen betydelse. I detta fall har två olika frekvenser samplats med en samplingsfrekvens $f_s = 100$ Hz. Den röda sinuskurvan med hög frekvens kommer få exakt samma värden efter sampling jämfört med den svarta kurvan med låg frekvens. Det går alltså inte att skilja dessa sinussignaler åt efter sampling. En signal som innehåller båda dessa frekvenskomponenter kommer alltså vara felaktig efter sampling.

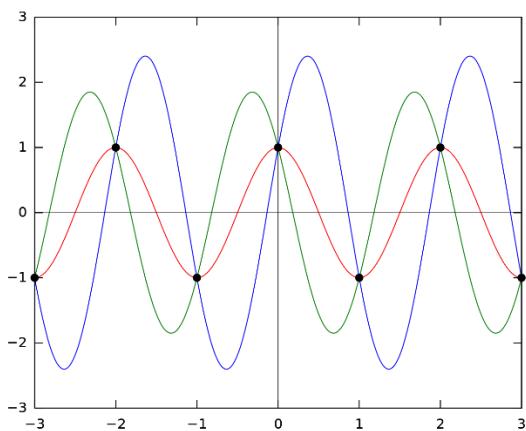
Nyquist-Shannons samplingsteorem [10] säger att samplingsfrekvensen måste vara minst dubbla samplade signalens högsta frekvens. Detta uttrycks som

$$f_s > 2f_{max} = 2B$$

Ekvation 33

där signalens bandbredd B definieras som signalens högsta frekvenskomponent. Denna frekvens brukar benämns Nyquistfrekvensen [11].

Det finns en enkel, intuitiv tolkning av Ekvation 33. En sinussignal behöver lite mer än två sampel per period för att ha en korrekt representation efter sampling. Det behövs alltså minst ett sampel vardera för den negativa respektive den positiva perioden, och samplingen behöver vara lite snabbare än två sampel per period för att undvika problemet som visas i Figur 15



Figur 15 Exempel med tre olika sinussignaler med samma frekvens som alla sampelas med en samplingsfrekvens lika med halva signalfrekvens (dvs. $f_s > f_{max}/2$). Diagrammet visar tydligt att de tre signalerna, som har olika amplitud och fas, kan få samma samplingsvärdet. Samplingsfrekvensen behöver alltså öka för att skilja dem åt.
By Qef - Created from scratch with gnuplot, based on idea from original bitmap version, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=4308333>

Fenomenet med att för höga frekvenser blir felaktigt representerade efter sampling kallas för vikning (eng. aliasing). Den resulterande störningen kallas för vikningsdistortion. Höga frekvenser som är oönskade i analysen kan förstås filtreras bort innan sampling, vilket finns beskrivet i mer detalj i kapitel 7.2.3.

2.5. Sammanställning av Fourierserier och -transformer

Vid en genomläsning av de olika verktyg som finns för frekvensanalys av signaler är det lätt att förlora den översiktliga förståelsen. Men det visar sig att det finns symmetrier. En signal kan beskrivas i tidsplanet som en tidsfunktion likaväl som i frekvensplanet som en frekvensfunktion. Tidsplanet och frekvensplanet kan sägas vara två sidor av samma mynt.

Ur alla transformer som är beskrivna i kapitel 2.1.8 till 2.4 framträder vissa generella slutsatser:

- En signal som är **kontinuerlig** i ena planet transformeras till andra planet via en **integral**.
- En signal som är **diskret** i ena planet transformeras till andra planet via en **summa**.
- En signal som är **icke-periodisk** i ena planet blir **kontinuerlig** i andra planet.
- En signal som är **periodisk** i ena planet blir **diskret** i andra planet.

Tabell 1 sammanställer alla transformer och inverstransformer enligt det mönster som framträder i punktlistan ovan.

Tabell 1 Sammanställning av olika verktyg för frekvensanalys av olika typer av signaler. Pga. utvecklingen inom det digitala området så är det diskret Fouriertransform som är det dominerande verktyget idag. Den Fouriertransformen är enkel och praktisk att implementera för sampelade data och man kan med datorer eller FPGA/ASIC genomföra frekvensanalys i realtid.

	Kontinuerlig i tidsplanet	Diskret i tidsplanet	
Icke-periodisk i tidsplanet	Fouriertransform, 2.3 $X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$ $x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{j\omega t} d\omega$	Diskret-tid Fouriertransform, 2.4.1 $X(\omega_n) = \sum_{n=-\infty}^{\infty} x[n]e^{-jn\omega_n}$ $x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega_n)e^{j\omega_n t} d\omega_n$	Kontinuerlig i frekvensplanet
Periodisk i tidsplanet	Fourierserie, 2.2.2 $C_n = \frac{1}{T} \int_{t_0}^{t_0+T} x(t)e^{-jn\omega_0 t} dt$ $x(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t}$	Diskret Fouriertransform, 2.4.2 $X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\cdot 2\pi kn/N}$ $x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k]e^{j\cdot 2\pi kn/N}$	Diskret i frekvensplanet
	Icke-periodisk i frekvensplanet	Periodisk i frekvensplanet	

2.6. Bandbredd

Med kunskap om signalens frekvensspektrum enligt Ekvation 26 kan man avgöra hur stor del av signalens effekt som finns i ett visst frekvensintervall. De signaler som vi normalt analyserar har i princip all effekt inom ett begränsat frekvensintervall. Ett exempel är mänskligt tal som har ett frekvensområde från ca 100-200 Hz (beroende på om det är en man eller kvinna) upp till 3 kHz. Se Figur 16 för ett exempel.

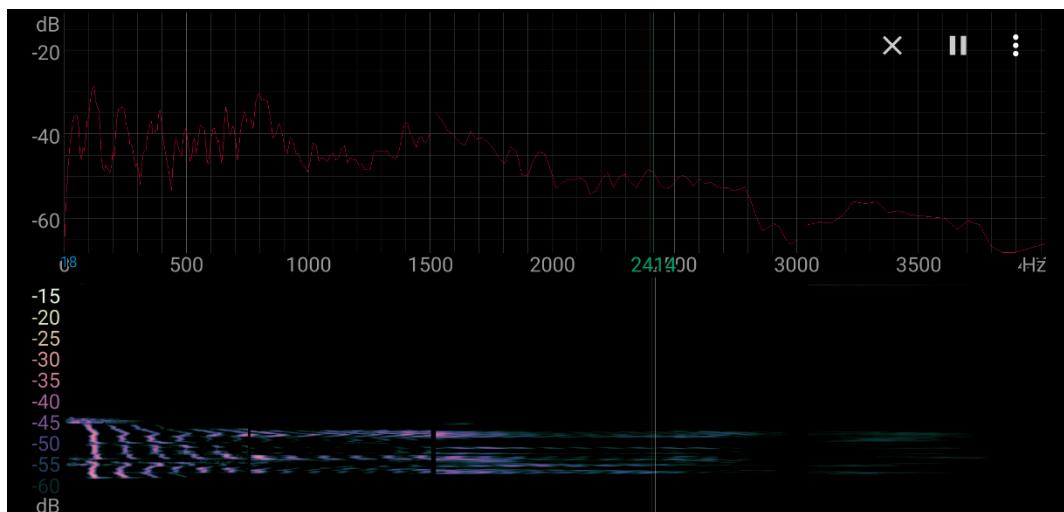
Bandbredd är inte ett enhetligt definierat begrepp, och man talar ofta om olika typer av bandbredd.

- Signalbandbredd
Ett mått på det frekvensomfång där signalens huvudsakliga effektninnehåll finns. Termen "huvudsaklig" är diffus, ofta behövs en matematisk definition där man bedömer att den signaleffekt som ligger vid frekvenser utanför bandbredden kan försummas utan att signalens kvalitet påverkas för mycket.
- Systembandbredd
Kommunikationssystemet har i sig en begränsad bandbredd som sätts av de komponenter systemet är uppbyggt av, t.ex. filter. En vanlig definition av

systembandbredd är det band som begränsas av frekvenspunkterna där systemets överföringsfunktion sjunkit 3 dB i relation till max.

- Ekvivalent brusbandbredd
Se kapitel 3.3.2.
- Bandbredd vid sampling

Vid sampling bestäms en signals bandbredd av den högsta frekvens som signalen innehåller. I praktiken bestäms denna bandbredd vid en frekvenspunkt där effektinnehållet i högre frekvenskomponenter är försumbart litet. Se mer i kapitel 2.4.4.



Figur 16

Effekttäthet för en röst (man) som säger frasen "Mary had a little lamb". Övre halvan (röd kurva) visar hur frekvensspektrum varierar med en grundfrekvens på ca. 100 Hz och frekvenskomponenter upp till ca. 2,8 kHz, där effekttätheten snabbt avtar. Signalens effekt kan därför huvudsakligen sägas vara samlad mellan dessa frekvenser. Signalbandbredden är då 2,7 kHz.

Nedre delen av grafen visar en färgkodad effekttäthet över tid (x-axeln är frekvens, y-axeln motsvarar tid). Ljusare färg innebär starkare effekttäthet.

3. Brus

3.1. Vad är brus och varför behöver det behandlas?

Brus [12] (eller distorsion) kan definieras i teoretiska termer som en slumpräglad (stokastisk) signal eller som en störning i en signal. I vardagligt språk menar man oftast ett ljud som skapas av till exempel vinddrag eller forsande vatten, eller ljudet som kommer ur en radiomottagare eller tv-apparat med dålig mottagning.

Inom elektroniken och kommunikationsteorin används brusbegreppet för att beskriva elektriskt brus som stör en elektrisk signal. Begreppet är relevant för en lång rad typer av signaler, system, fysikaliska storheter och mättekniker. Det kan till exempel vara elektromagnetiskt brus (radiovågor) som tas emot av en TV-antenn och blir till slumpvisa prickar i rutan.

Inom kommunikation sätter bruset en teoretisk gräns för hur noggrant man kan urskilja en önskad signal. Detta kallas den brusbegränsade känsligheten hos apparaten. Det är därför viktigt att kunna genomföra beräkningar och mätningar på brus för att korrekt kunna dimensionera komponenter och system.

I många kommunikationssystem är det flera användare som samtidigt använder en resurs. Ett typiskt exempel är ett hyreshus där de flesta hyresgästerna har var sin trådlösa router för WiFi och har flera olika enheter uppkopplade till sin router. Signalerna från alla dessa routrar och enheter kan behöva dela på samma frekvenskanal, vilket betyder att de stör varandra. Denna typ av störning brukar inte benämñas brus. I stället talar man om interferens. Brus och interferens har dock en del gemensamt då de stör nyttosignalen och begränsar möjligheterna till kommunikation, men de behöver behandlas på olika sätt i dimensionering och konstruktion av ett kommunikationssystem. T.ex. så kan problemet med WiFi-interferens för en användare lösas genom att routern byter till annan kanal med färre uppkopplade enheter. Detta fungerar sannolikt inte för att lösa ett brusrelaterat problem eftersom bruset normalt är lika stort på alla kanaler.

3.2. Olika typer av brus

Brus kommer från olika typer av källor och olika typer av brus har olika egenskaper och kan därmed olika påverka på ett kommunikationssystem. Vanliga typer av brus [13] som finns i ett kommunikationssystem är

- termiskt brus (thermal noise, Johnson–Nyquist noise) [14]
Brus som förekommer i all elektronik och som är orsakad av termisk påverkan på laddade partiklar (oftast elektroner) i en elektrisk ledare.
- fasbrus (phase noise) [15]
Fasbrus är slumpräglade variationer i en signals fasläge. Det är ofta en

parameter att beakta i signalgenerering i en oscillator eller klocka. I tidsplanet benämns detta ofta jitter.

- brum (mains hum) [16]

Brum är störningar som orsakas av kraftnätets växelströmmar. Det kan t.ex. vara ett problem i audiokretsar eftersom brum ger störningar i det hörbara spektrumet.

- skottbrus, hagelbrus (shot noise, Poisson noise) [17]

Skottbrus beror av att elektrisk laddningen endast uppträder i diskreta enheter (t.ex. elektroner). En ström är alltså inte helt kontinuerlig utan består av diskreta partiklar som passerar en viss punkt. Därmed kan strömmen statistiskt variera något beroende på den statistiska variationen i tid när partiklarna passerar. Skottbrus har likartade egenskaper som det termiska bruset.

- 1/f-brus, kontaktbrus (flicker noise, pink noise) [18]

Kontaktbrus uppstår pga. resistansfluktuationer, som i sin tur beror på dålig kontakt mellan t.ex. två ledare. Kontaktbruset har ett sffektspektrum som beror av $1/f$, dvs. det minskar snabbt och kan försummas då frekvensen ökar. I praktiken är detta brus endast märkbart vid så låga frekvenser som några Hz.

- Popcornbrus (burst noise, impulse noise) [19]

Popcornbrus är en typ av brus som uppkommer stötvis i halvledarkomponenter och innebär en stegvis spänningssändring. Orsaken är defekter i en PN-övergång. Namnet kommer från att om denna form av brus förstärks och kopplas till en högtalare så påminner det om ljudet när popcorn poppar.

- kosmiskt brus (Cosmic noise) [20]

Kosmiskt brus är radiofrekvent brus till följd av elektromagnetisk strålning från objekt i rymden. Källor är främst solen, månen och Vintergatan.

- kosmisk bakgrundsstrålning (Cosmic microwave background) [21]

Kosmisk bakgrundsstrålning är radiofrekvent brus till följd av universums termiska svartkroppsstrålning vid temperaturen 2,7 K. Den kosmiska bakgrundssstrålningen är ett av de viktigaste stöden för Big Bang-teorin och dess upptäckt gav Nobelpris 1978.

- atmosfäriskt brus (Atmospheric noise) [22]

Atmosfäriskt brus är radiofrekvent brus skapat av naturliga atmosfäriska processer, primärt blixturladdningar från åskväder. Atmosfäriskt brus kan observeras med i form av bakgrundsbrus från fjärran åskoväder och impulser från närliggande åskoväder.

- man made-brus

Man-made brus är ett samlingsnamn för oönskade störningar från mänsklig aktivitet, typiskt andra kommunikationssystem eller elektromagnetiska vågor som läcker från elektronik mm. Det är kan ses som en form av interferens. Idag, med så många olika typer av kommunikationssystem, så är detta en betydande störning som ofta är starkare än många andra bruskällor.

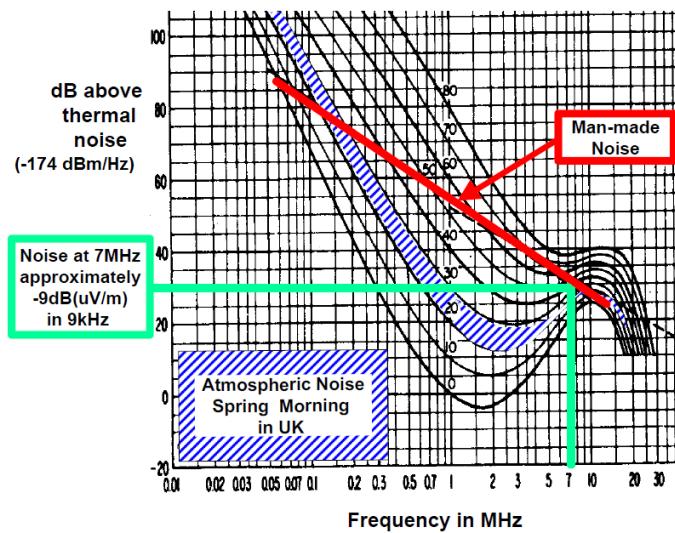
Listan är inte komplett och den visar att det finns en mängd anledningar till att brus uppkommer. Källan till bruset kan också vara olika komponenter i ett system eller komma utifrån och fångas upp av t.ex. en antenn.

3.3. Viktiga mått för karakterisering av brus

3.3.1. Brusets effektspektrum

Brus fördelar sig alltid över olika frekvenser! Kom ihåg att den enda signal som förekommer vid en enda frekvens är rena cosinus- och sinussignaler. Det skulle i.o.f.s. kunna gå att argumentera för att 50 Hz-brum är sinusformat och därmed en ren ton men i praktiken är brum inte helt rent.

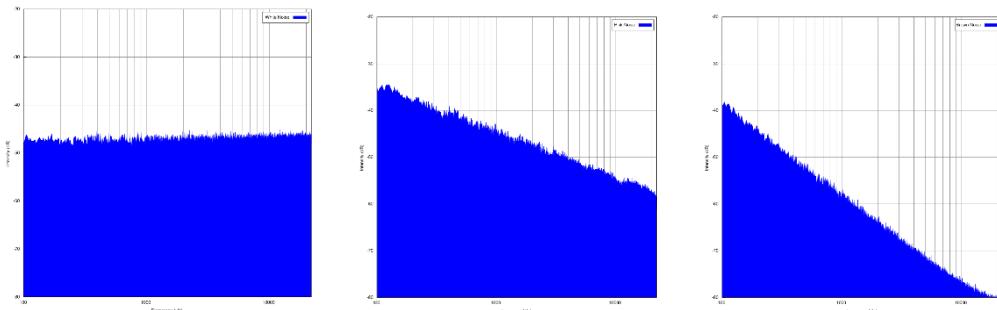
Ett mycket viktigt mått för att karakterisera brus är alltså dess effektspektrum [23], [24]. Effektspektrum talar om hur den totala bruseffekten är fördelad över olika frekvenser, och anges i W/Hz. Här kommer brusspektrum betecknas med $N_{spektrum}$ eller i vissa fall t.o.m. $N_{spektrum}(f)$ för att visa att det är en funktion av frekvensen f . För ett exempel, se Figur 17.



Figur 17

Exempel på bruskällor och dess effektspektrum. Grafen visar hur bruseffekten för termiskt brus och man-made brus är fördelad över olika delar av frekvensbandet och har således frekvens på x-axeln och någon form av effekttäthetsmått på y-axeln.
By <http://en.wikipedia.org/wiki/User:Threeme3> -
<http://en.wikipedia.org/wiki/File:Atmosphericnoise.PNG>, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=11319770>

Det är väldigt vanligt att man talar om olika färger [25] när det gäller brus. Detta är en analogi med det ljuset i synliga spektrummet, där olika frekvenser har olika färg, och vitt ljus innehåller alla färger. Några exempel är vitt brus, skärt brus, brunnt brus, violett brus osv. De har alla olika typer av effektspektrum, se Ekvation 6 för några vanliga exempel.



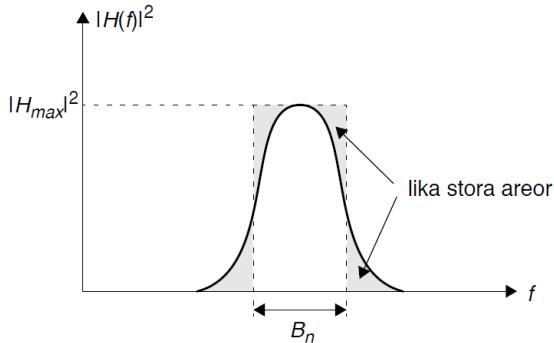
Figur 18 Olika typer av färgat brus och deras effektspektrum. I tur och ordning visas vitt brus, skärt brus och brunt brus. De har olika fördelning av bruseffekt över frekvenspektrumet, där det vita bruset har en konstant effekttäthet över frekvens.
By Warrakk - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=19273413>
<https://commons.wikimedia.org/w/index.php?curid=19273415>
<https://commons.wikimedia.org/w/index.php?curid=19273417>

Ur ett systemperspektiv så är det förstås viktigt att kunna bestämma hur stor bruseffekt totalt som stör t.ex. kommunikationen i ett kommunikationssystem. I kapitel 3.3.2 och 3.3.3 presenteras två mått som är vanligt att använda då beräkningar görs.

3.3.2. Ekvivalent brusbandbredd

Varje system (t.ex. kommunikationssystem) har en viss begränsad bandbredd given av systemets överföringsfunktion $H(f)$, vilket bl.a. betyder att en del bruseffekt från bruskällor kommer filtreras bort. För att få relevanta beräkningar behöver alltså ett mått på hur mycket brus som stör och påverkar den nyttosignal systemet använder. Från t.ex. filterteorin eller förstärkarteorin är vi bekanta med den "klassiska" definitionen av bandbredd som är frekvensbandet mellan frekvenspunkterna där t.ex. filterkurvan fallit - 3 dB från max. Den definitionen är inte helt tillämpbar i detta fall, eftersom vi önskar få ett mått på den bruseffekt som påverkar systemets funktion. Två olika filer med samma 3 dB-bandbredd kan filtrera olika mycket bruseffekt eftersom ett filter med branta filterflanker filtrerar ju mer än ett med flacka.

Ett systems ekvivalenta brusbandbredd B_n förklaras enklast i grafisk form. Figur 19 visar $|H(f)|^2$ för ett systems överföringsfunktion, vilket kan beskrivas som systemets överföringsfunktion i när det gäller effekt ($|H(f)|$ beskriver amplituden av överföringsfunktionen). Den totala arean under kurvan $|H(f)|^2$ ger då ett mått på hur mycket bruseffekt som påverkar ett sådant system, förutsatt att bruset har konstant effekttäthet. Ekvivalenta brusbandbredd B_n är då bredden på den tänkta rektangel som har motsvarande area, där höjden är maxvärdet $|H_{max}|^2$ som överföringsfunktionen har.



Figur 19 Ett systems ekvivalenta brusbandbredd. Ekvivalent brustemperatur är bredden på den tänkta rektangel (streckad linje i figuren) som har samma area som integralen av $|H(f)|^2$. Figur från [26] s. 123.

Matematiskt uttrycks detta som [26]

$$B_n = \frac{1}{|H_{max}|^2} \int_0^\infty |H(f)|^2 df \quad \text{Ekvation 34}$$

I praktiken så överensstämmer ett systems ekvivalenta brusbandbredd ganska väl med systemets 3 dB-bandbredd, eftersom $|H(f)|^2$ ofta har branta flanker. Vanligen approximeras ekvivalenta brusbandbredden med systemets bandbredd, dvs. $B_n \approx B$.

Exempel 3.1

Ett enkelt RC lågpassfilter har överföringsfunktionen $H(f) = \frac{1}{1+j \cdot 2\pi f \cdot RC}$.

Brusbandbredden kan beräknas med Ekvation 34 som (primitiva funktionen för $\frac{1}{1+x^2}$ är $\tan^{-1} x$)

$$\begin{aligned} B_n &= \frac{1}{|H_{max}|^2} \int_0^\infty |H(f)|^2 df = \frac{1}{1^2} \int_0^\infty \left| \frac{1}{1 + j \cdot 2\pi f \cdot RC} \right|^2 df \\ &= \int_0^\infty \frac{1}{1 + (2\pi f \cdot RC)^2} df = \left[\frac{1}{2\pi \cdot RC} \tan^{-1}(2\pi f \cdot RC) \right]_0^\infty \\ &= \frac{1}{2\pi \cdot RC} \left(\frac{\pi}{2} - 0 \right) = \frac{1}{4RC} \end{aligned}$$

Jämför detta resultat med den vanliga definitionen av bandbredd ($1/\sqrt{2}$ från maxvärdet eller s.k 3 dB-bandbredd) för ett RC filter som är $B = \frac{1}{2\pi RC}$. Att de olika definitionerna av bandbredd ger olika resultat är inte förvånande, men skillnaden minskar för filter med brantare filterflanker.

3.3.3. Bruseffekt, brusspänning och brusström

Även om brus (stokastisk signal) är av slumpmässig natur har det vissa mått som känns igen från karakteriseringen av deterministiska signaler. Det viktigaste är effekt. Här kommer vi beteckna bruseffekt med en egen beteckning, N .

I många fall så är brusets effektspektrum $N_{spektrum}(f)$ approximativt konstant över systemets bandbredd, vilket gör att bruseffekten beräknas med ett enkelt samband

$$N = B_n N_{spektrum}$$

Ekvation 35

$N_{spektrum}$ har som tidigare nämnts enheten W/Hz och ekvivalenta brusbandbredden Hz, vilket betyder att bruseffekten N som förväntat har enheten W.

Ofta används också begreppet brusspänning. Notera att det inte är relevant att tala om en specifik amplitud för en brusignal pga. av dess stokastiska natur. Men brus ger upphov till en spänning (och ström) så det går ändå att ha ett mått relaterat till spänning. Med brusspänning menas då brusets effektivvärde U_{eff} . Detta mått känner vi också igen från elläran och deterministiska signaler. Samma sak gäller brusström, strömmen från en bruskälla kan modelleras med dess effektivvärde I_{eff} .

Exempel 3.2

En mikrovågsförstärkare har ekvivalenta brusbandbredden $B_n = 420 \text{ MHz}$. Bruset på förstärkarens ingång har ett effektspektrum som är approximativt konstant och har ett brusspektrum $N_{spektrum}(f) = -128 \text{ dBm/Hz}$. Bruseffekten in i förstärkaren har då efter filtrering effekten

$$N = B_n N_{spektrum} = 420 \cdot 10^6 \cdot 10^{-128/10} \approx 6,7 \cdot 10^{-5} \text{ mW} \approx -42 \text{ dBm}$$

Enheten från beräkningen blir mW eftersom enheten för $N_{spektrum}$ är i dBm (dB beräknad i mW). Notera att det går att räkna i dB-skala direkt genom att räkna ut ett logaritmiskt mått för bandbredden!

$$N_{dBm} = 10 \log B_n + N_{spektrum,dBm} = 10 \log 420 \cdot 10^6 + (-128) = -42 \text{ dBm}$$

En kommentar är att förstärkaren sedan adderar brus till denna brusnivå som är på ingången, så att bruset på utgången är högre än -42 dBm. Mer om detta i kapitel 3.6.

3.3.4. Signal-brusförhållande

Ett vanligt mått som används i signalbehandling är Signal-brusförhållande [27]. Signal-brusförhållande (Signal-to-noise ratio på engelska, ofta förkortat SNR eller S/N) är en definierar förhållandet mellan signaleffekten och den bruseffekt som stör och förvanskar signalen. I mindre tekniska termer så kan man säga att SNR jämför nivån hos nyttosignalen (såsom musik) med bakgrundsbruset. Ju högre kvoten är mellan dessa desto mindre störs man av bakgrundsbruset.

SNR definieras som

$$SNR = \frac{S}{N}$$

Ekvation 36

där S och N betecknar medeleffekterna för nyttosignalen respektive brussignalen. Både signaleffekten och bruseffekten måste vara uppmätt vid samma (eller ekvivalent) punkt i ett system och inom samma systembandbredd.

Exempel 3.3

Bruseffekten i en 5 MHz-kanal i ett LTE-system (4G) har ett brusgolv (lägsta möjliga brusnivå pga. termiska bruset) som är $N = -107 \text{ dBm}$. Vid ett tillfälle är mottagen signaleffekt $S = 90 \text{ fW}$. Signal-brusförhållandet blir då (Ekvation 36)

$$SNR = \frac{S}{N} = \frac{90 \cdot 10^{-15}}{0,001 \cdot 10^{-107/10}} \approx 4,5 \text{ ggr} \approx 6,5 \text{ dB}$$

Ett signal-brusförhållande på 6,5 dB i ett mobilsystem ger en fungerande dataöverföring, men för att uppnå god datahastighet ska SNR vara åtminstone 10 dB högre.

SNR är dimensionslös och anges ofta i decibel

$$SNR_{dB} = 10 \log \left(\frac{S}{N} \right)$$

Ekvation 37

Beteckningen SNR_{dB} har här använts för att särskilja Ekvation 37 från definitionen i Ekvation 36, men nästan alltid används beteckningen SNR oavsett om linjärt mått eller decibel avses.

Två andra närbesläktade mått som används är SINR och SIR [28]. I dessa mått tar man hänsyn till interferensens medeleffekt I . SINR inkluderar då summan av interferensens och brusets effekt medan SIR endast inkluderar interferensen, se Ekvation 38 och Ekvation 39.

$$SINR = \frac{S}{I + N}$$

Ekvation 38

$$SIR = \frac{S}{I}$$

Ekvation 39

Båda dessa mått brukar ofta anges i dB, precis som SNR.

Exempel 3.4

Om det tillkommer interferens för LTE-kanalen i exempel 3.3, kommer kvaliteten på kommunikationen försämras. I ett mobiltelefonisystem är interferensen generellt högre än bruset N . Beräkna SINR om interferensen I har dubbelt så stor effekt (dvs. 3 dB högre) som bruset. Med Ekvation 38 fås

$$SINR = \frac{S}{I + N} = \frac{S}{2N + N} = \frac{90 \cdot 10^{-15}}{3 \cdot 0,001 \cdot 10^{-107/10}} \approx 1,5 \text{ ggr} \approx 1,8 \text{ dB}$$

Vid 1,8 dB är kommunikation fortfarande möjlig, men uppkopplingen kommer upplevas som långsam eftersom datatakten kommer sänkas av systemet.

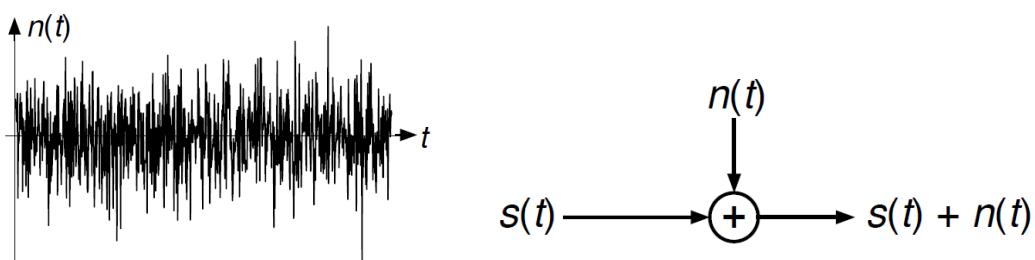
3.4. Termiskt brus

3.4.1. Brus i en resistor

Termiskt brus [14] finns i alla elektriska kretsar, och är som nämnts det fenomen när laddade partiklar påverkas av termisk energi. Termiskt brus ökar alltså med temperaturen. Denna typ av brus upptäcktes och mättes först av John B. Johnson vid Bell Labs 1926. Han beskrev sina fynd för Harry Nyquist, också på Bell Labs, som kunde skapa en modell som förklarade resultaten. Därför kallas denna typ av brus ofta för Johnson – Nyquist brus.

Bruset i en resistor ger upphov till en slumprövigt varierande spänning $n(t)$ mellan resistorns anslutningar. Den fysikaliska orsaken är att elektronerna i resistorns ledningsband kommer röra sig slumprövigt i olika riktningar p.g.a. den termiska energin. Denna slumpröviga rörelse ger upphov till en mycket liten, slumprövigt varierande växelström. Ohms lag ger sedan en spänning över resistorns anslutningar.

Om det finns en signalspänning (nyttosignal) över resistorn så kan brus och signal modelleras som en summa, se Figur 20. Denna modellering är mycket praktisk, eftersom vi ofta vill kunna behandla brus och nyttosignal separerat, t.ex. när vi beräknar signal-brusförhållande.



Figur 20 Exempel på termisk brusspänning över en resistor (vänster figur) och hur det termiska bruset $n(t)$ kan modelleras som additivt till nyttosignalen $s(t)$ (höger figur).
Figurer från [26], s. 111-112.

Bruset är slumprövigt varierande, men har ändå en del egenskaper som gör att det kan analyseras och beräknas.

- Amplitudfördelningen är normalfördelad (gaussisk), dvs. den har en täthetsfunktion $f(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{n^2}{2\sigma^2}}$. Variansen σ^2 är ett mått proportionellt mot brusets effekt medan standardavvikelsen σ är proportionellt mot brusets effektivvärde.
- Effektspektrum är konstant och frekvensberoende inom det frekvensintervall som är intressant inom elektronik och kommunikationsteori. Detta brus kallas

ofta vitt brus, i analogi med vitt ljus som är en blandning av alla synliga våglängder. Analogin haltar dock något eftersom vitt ljus egentligen inte har konstant effektspektrum.

Så även om termiskt brus är slumpmässigt varierande och oförutsägbart, kan man ändå bestämma vissa viktiga parametrar.

Det termiska brusets fördelning i frekvens för en resistor med resistansen R beskrivs av Ekvation 40 nedan [26]

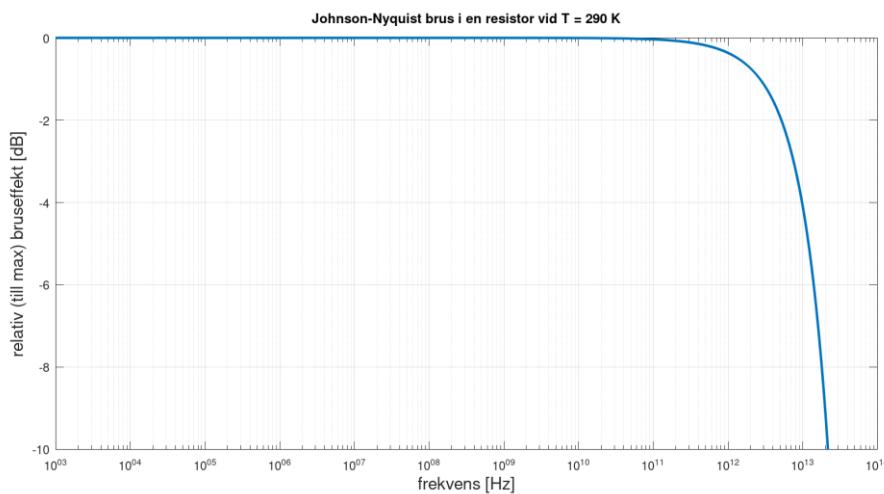
$$U_{eff,spektrum}^2(f) = \frac{4Rhf}{e^{\frac{hf}{k_B T}} - 1} \quad \text{Ekvation 40}$$

där $U_{eff,spektrum}$ betecknar brusspänningens spektralfördelning. Ekvationen innehåller två konstanter, $h \approx 6,626 \cdot 10^{-34}$ Js är Plancks konstant och $k_B \approx 1,381 \cdot 10^{-23}$ J/K är Boltzmanns konstant. T är temperaturen (i enheten K). Enheten för $U_{eff,spektrum}^2(f)$ är V^2/Hz . Notera att $U_{eff,spektrum}$ här betecknar just spektralfördelning, dvs. hur brusspänningen är fördelad över frekvens.

Figuren visar att spektralfördelningen är nästan konstant upp till ca. 100 GHz för att sedan börja avta. Det är alltså rimligt att använda en approximation vid de frekvenser där vi normalt konstruerar elektronik och kommunikationssystem. Ekvation 40 kan vid lägre frekvenser approximeras [26] med

$$U_{eff,spektrum}^2(f) = 4Rk_B T \quad \text{Ekvation 41}$$

där man kan se att termiska brusets spektralfördelning är oberoende av frekvensen f .



Figur 21 Spektralfördelning (dvs. fördelningen av effekten över olika frekvenser) för termiskt brus. Grafen är normalerad till maxvärdet vid temperaturen 290 K. Frekvensen där bruseffekten viker nedåt varierar med temperaturen, men för alla praktiska temperaturer och frekvenser (vilket är under THz-området) är kurvan i princip konstant.

För att få ett mått på brusspänningen över resistorn behöver vi definiera den över en viss brusbandbredd B_n . Ekvation 41 ger då att

$$U_{eff} = \sqrt{4Rk_BTB_n}$$

Ekvation 42

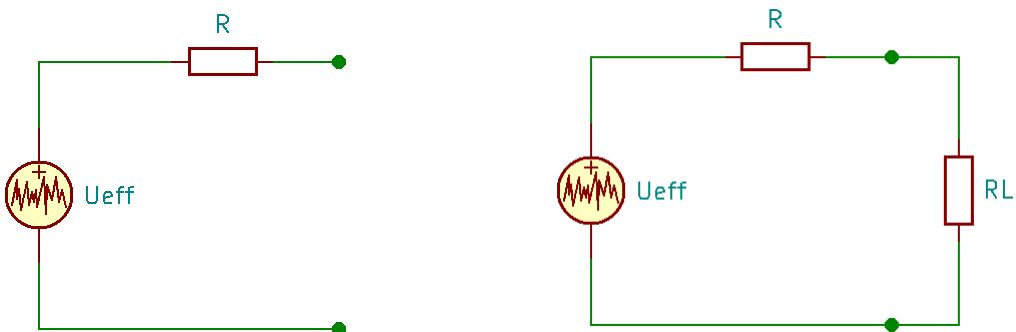
Exempel 3.6

Beräkna effektivvärdet på brusspänningen för en $1\text{ k}\Omega$ -resistor vid rumstemperatur över audiofrekvenserna $20\text{ Hz} - 20\text{ kHz}$!

Använd temperaturen $T = 290\text{ K}$. Ekvation 42 ger då

$$U_{eff} = \sqrt{4Rk_BTB_n} = \sqrt{4 \cdot 1000 \cdot 1,38 \cdot 10^{-23} \cdot 290 \cdot (20000 - 20)} \approx 0,57\text{ }\mu\text{V}$$

Denna brusspänning modelleras lämpligen som en tvåpol med en ideal, brusfri resistor R i serie med en spänningsskälla U_{eff} , se Figur 22. Modelleringen som en tvåpol med en spänningsgenerator är till stöd för att kunna genomföra beräkningar. Dels så behöver vi kunna bestämma hur mycket bruseffekt som den brusande resistorn lämnar till omgivningen, dels behöver vi kunna räkna på mer komplexa resistornätverk med kända metoder.



Figur 22

Modellering av termiskt brus i en resistor som en tvåpol. Vänster figur visar modellen, där brusspänningen är modellerad med en separat spänningsskälla. Resistorn i modellen är därför en ideal, brusfri resistans. Höger figur visar tvåpolen ansluten till en last R_L .

Med hjälp av denna modell kan bruseffekten N som levereras ut från resistorn R till en last R_L beräknas. Vi söker då det värsta fallet när bruseffekten i R_L är som störst, vilket vi från elläran vet är när $R_L = R$, dvs. då lasten är anpassad till tvåpolens inre resistans. U_{eff} kommer då spänningsdeltas så att hälften av spänningen ligger över R_L , och bruseffekten som levereras till R_L blir då

$$N = \frac{(U_{eff}/2)^2}{R_L} = \frac{U_{eff}^2}{4R} = \frac{(\sqrt{4Rk_BTB_n})^2}{4R} = k_BTB_n \quad \text{Ekvation 43}$$

Ekvation 43 visar ett mycket viktigt samband inom signalbehandling och kommunikationsteorin. Termiska bruseffekten från en resistor beror inte av resistansen R , även om dess brusspänning gör det. Den visar också att bruseffekten är proportionell mot temperaturen och bandbredden. Denna härledda bruseffekt är en teoretisk undre gräns för bruseffekten i ett kommunikationssystem. Ofta anges faktorn $k_B T$ vid en viss referenstemperatur som är 290 K . $k_B T$ motsvarar då brusets effektspektrum $N_{spektrum}(f)$ vid rumstemperatur, och vid temperaturen $T_0 = 290\text{ K}$ har detta

spektrum värdet $k_B T_0 \approx 1,381 \cdot 10^{-23} \cdot 290 \approx 4,01 \cdot 10^{-21} \text{ W/Hz}$. Mer vanligt är att detta värde anges i decibelskala och då mer specifikt som -174 dBm/Hz. Detta är ett värde att lägga på minnet eftersom det ofta förekommer i överslagskalkyler i brusberäkningar.

Exempel 3.7

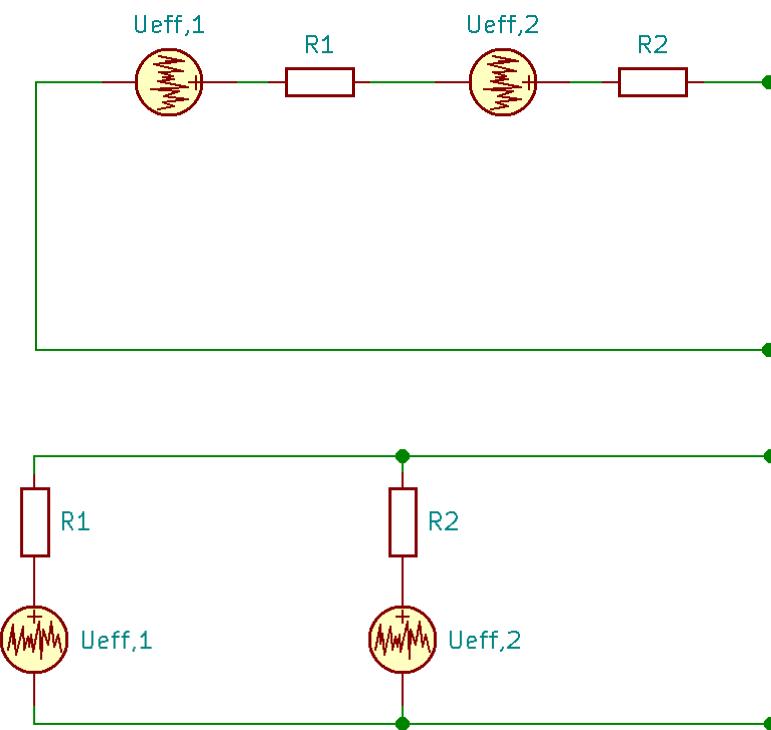
Bruseffekten vid rumstemperatur in på en anpassad förstärkning går där brusbanden är $B_n = 20 \text{ MHz}$ kan beräknas som

$$N = k_B T B_n = -174 + 10 \log 20 \cdot 10^6 \approx -101 \text{ dBm}$$

Även här går det bra att räkna i dB-skala genom att logaritmera bandbredden!

3.4.2. Brusberäkningar på resistiva nät

Modellen i Figur 22 är även till hjälp för att räkna på brus för seriekopplade och parallellkopplade resistanser. Figur 23 visar hur tvåpolsmodellen ser ut för två seriekopplade respektive parallellkopplade resistanser. Dessa kretsar kan ersättas med ekvivalenta tvåpoler enligt Figur 22.



Figur 23 Modeller för termiskt brus i seriekopplade (övre figuren) respektive parallellkopplade (undre figuren) resistorer.

För att göra beräkningar på två (eller flera) termiska bruskällor behöver behandlas som de stokastiska signaler de faktiskt är. Eftersom de har en statistisk amplitudfördelning som är normalfördelad, kommer summan av deras varianser motsvara totala variansen,

dvs. $\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2$. Detta gäller stokastiska signaler som är okorrelerade, vilket är uppfyllt för t.ex. ett resistornätverk. Den fysikaliska tolkningen är att effekterna kan summeras för två källor som summeras, men inte effektivvärdena.

För seriekopplingen i Figur 23, där spänningsskällorna ligger i serie gäller alltså att $U_{eff}^2 = U_{eff,1}^2 + U_{eff,2}^2$, dvs.

$$U_{eff} = \sqrt{U_{eff,1}^2 + U_{eff,2}^2} = \sqrt{4R_1 k_B T B_n + 4R_2 k_B T B_n} = \\ = \sqrt{4(R_1 + R_2) k_B T B_n}$$
Ekvation 44

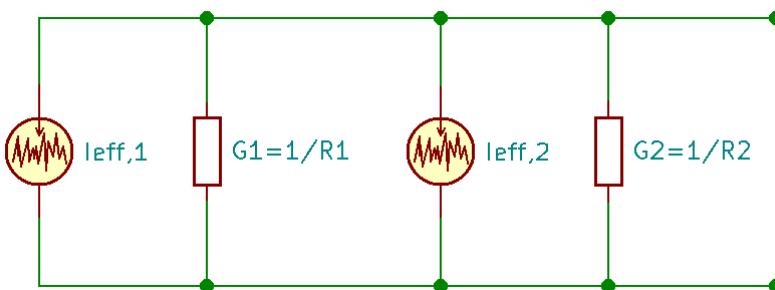
U_{eff} för den ekvivalenta tvåpolen motsvarar alltså brusspänningen från två seriekopplade resistorer $R_1 + R_2$. Eftersom ersättningsresistansen i tvåpolen i Figur 23 också är $R_1 + R_2$, så kan vi representera totala termiska bruset genererat av två seriekopplade resistorer med en tvåpolsmodell enligt Figur 22 där

$$U_{eff} = \sqrt{4(R_1 + R_2) k_B T B_n}$$
Ekvation 45

$$R = R_1 + R_2$$
Ekvation 46

Detta känns intuitivt korrekt eftersom detta är analogt med Ohms lag, men man måste komma ihåg att bruskällornas spänningar inte adderas i amplitud utan i effekt.

När det gäller parallellkopplingen så är analysen likadan men genomförs lämpligen med strömkällor i.st.f. spänningsskällor. Genom att ersätta modellen för R_1 och R_2 i Figur 23 med strömkällor enligt



Figur 24 *Ekvivalent modell för termiskt brus i två parallellkopplade motstånd där spänningsskällorna är ersatta av strömkällor och resistanserna med deras respektive konduktanser.*

På samma sätt som för de seriekopplade resistorerna kan man ersätta tvåpolen i Figur 24 med en ekvivalent tvåpol som har $I_{eff} = \sqrt{I_{eff,1}^2 + I_{eff,2}^2} = \sqrt{4(G_1 + G_2) k_B T B_n}$ och ersättningskonduktansen $G = G_1 + G_2$. Det betyder i sin tur (efter lite beräkningar) att brusspänning och resistans för ekvivalenta tvåpolen bli

$$U_{eff} = \sqrt{4 \frac{R_1 R_2}{R_1 + R_2} k_B T B_n}$$
Ekvation 47

$$R = \frac{R_1 R_2}{R_1 + R_2}$$

Ekvation 48

vilket känns igen från ersättningsresistansen för två parallellkopplade motstånd. En intressant effekt av parallellkopplingen är alltså att totala brusspänningen enligt Ekvation 47 minskar. Det kan tyckas kontraintuitivt att brus minskar när man församman två eller flera komponenter, men det är en alltså effekt av att totala motståndet minskar.

Slutligen bör nämnas att härledningarna som ger Ekvation 45 till Ekvation 48 gäller när motstånden har samma temperatur. Har de olika temperatur behöver det beaktas i beräkningarna.

3.4.3. Brus i en kondensator

En ideal kondensator är förlustfri och den slumpmässiga elektronrörelsen som ger en brusspänning i en resistor ger alltså ingen brusspänning i en kondensator [14].

Kondensatoren som komponent är alltså brusfri så länge den modelleras som en ideal kondensator. Sen så har kondensatoren förstås en liten resistans men den kan i princip alltid försummas i förhållande till resistans i andra komponenter i en kretslösning.

Men man talar ändå om s.k. kTC-brus. Detta begrepp kommer från det termiska bruset från en RC-krets. T.ex. så har ett RC lågpassfilter en resistor som ger en bruseffekt. Enligt Ekvation 41 så gäller att $U_{eff,spektrum}^2(f) = 4Rk_B T$. Detta brus kan modelleras som en spänningskälla på ingången till ett brusfritt filter, som har en ekvivalent brusbandbredd $B_n = 1/4RC$ (se Exempel 3.1). Det egengenererade bruset i filtret kommer filtreras och dämpas av filterfunktionen. Brusspänningens effektivvärde på filtrets utgång kan då beräknas som

$$U_{eff} = \sqrt{B_n \cdot U_{eff,spektrum}^2(f)} = \sqrt{\frac{1}{4RC} \cdot 4Rk_B T} = \sqrt{\frac{k_B T}{C}} \quad \text{Ekvation 49}$$

Denna ekvation visar att ett RC-filter har ett brus som inte beror av resistorns storlek. Det kan känna märkt med tanke på Ekvation 41 som visar att bruset ökar med ökande storlek på resistorn. Men samtidigt förändras filtrets egenskaper så att brusbandbredden minskar i motsvarande takt.

Eftersom närvaren av kondensatorer i en krets ger filterverkan så påverkar en kondensator bruset i en kretslösning, även om den själv inte har något brus.

Exempel 3.8

Resistorn (1000Ω) i Exempel 3.6 kopplas samman med en kondensator $C = 16 \text{ nF}$ vilket bildar ett enkelt RC lågpassfilter, vars gränsfrekvens är $f_{co} \approx 9,9 \text{ kHz}$. Beräkna brusspänningens effektivvärde för det egengenererade bruset på filtrets utgång!

Med Ekvation 49 fås

$$U_{eff} = \sqrt{\frac{1,38 \cdot 10^{-23} \cdot 290}{16 \cdot 10^{-9}}} \approx 0,50 \text{ nV}$$

Notera att filtret har en brusbandbredd $B_n = \frac{1}{4RC} = \frac{1}{4 \cdot 1000 \cdot 16 \cdot 10^{-9}} \approx 16 \text{ kHz}$ vilket skiljer sig från filtrets cut off-frekvens (och bandbredd).

3.5. Bakgrundsbrus

Med bakgrundsbrus avses det brus som kommer från naturliga källor och finns som elektromagnetisk strålning i omgivningen. Detta brus plockas upp av kommunikationssystem via antennen. Ursprunget är svartkroppsstrålning från olika källor med olika temperaturer. I kapitel 3.2 nämns bl.a. kosmiskt brus, kosmisk bakgrundsstrålning och atmosfäriskt brus. Även omgivningen runt kommunikationssystemet som består av natur och bebyggelse har en svartkroppsstrålning som beror av omgivningens temperatur. Alla dessa källor adderas till ett bakgrundsbrus som plockas upp av antennen.

Att göra en komplett analys av detta brus är ganska komplicerat. Bakgrundsbruset har nämligen olika egenskaper i olika riktningar. En antenn som riktas upp mot rymden för kommunikationssystem ser främst atmosfärens olika temperaturer och den kalla rymden medan en antenn monterad i en mast för mobilkommunikation som är riktad mot användarna på marken ser den varma marken. Något som ytterligare försvårar analysen är att en antenn är konstruerad för att ha en viss riktverkan och därmed tar emot strålning bättre i vissa rymdvinkelar.

Exempel 3.8

Nobelpriset 1978 tilldelades Kapitsa, Penzias och Wilson. Penzias och Wilson fick priset för att de med en stor hornantenn upptäckte och mätte upp den kosmiska bakgrundsstrålningen vilket gav stöd åt Big Bang-teorin. Det var när de riktade antennen mot rymden som detta bakgrundsbrus kunde detekteras.

<https://urplay.se/program/191118-vagen-till-nobelpriset-robert-wilson-och-arno-penzias>

För att modellera detta brus talar man ofta om antenntemperatur (antenna noise temperature) [29]. Antenntemperatur är den sammanviktade temperaturen från bakgrundsbruset över olika rymdvinkelar och kan då modelleras som en tänkt resistor vid denna temperatur som genererar samma bruseffekt över en viss bandbredd. Här är det viktigt att förstå att detta inte är antennens fysiska temperatur. Antenntemperatur är ett mått på omgivningens bakgrundsbrus (och temperatur) viktat med antennens förmåga att plocka upp den i olika riktningar.

Som orientering, om omgivningens brustemperatur $T(\theta, \varphi)$ över alla elevations- och azimuthvinkelar (θ, φ) är kartlagd, kan antenntemperaturen T_{ant} beräknas genom en integral över alla rymdvinkelar Ω

$$T_{ant} = \frac{1}{4\pi} \int_{4\pi} G(\theta, \varphi) T(\theta, \varphi) d\Omega \quad \text{Ekvation 50}$$

$G(\theta, \varphi)$ är då antennens strålningsdiagram. Mer om detta i kapitel 6.

3.6. Brus i förstärkare

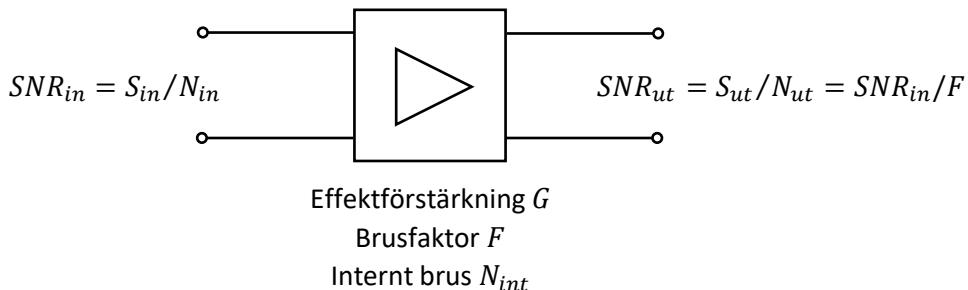
3.6.1. Brusfaktor

Att förstärka en nyttosignal betyder också att man oundvikligen förstärker det brus som finns på förstärkarens ingång lika mycket. Dessutom innehåller förstärkaren komponenter som adderar ytterligare brus. Det leder till att kvoten mellan nyttosignal och brussignal på utgången försämras i förhållande till motsvarande kvot på ingången. Med andra ord, signal-brusförhållandet SNR försämras när en signal går igenom en förstärkare.

För att få ett mått på denna förstärkning definieras en förstärkares brusfaktor [30] som

$$F = \frac{SNR_{in}}{SNR_{ut}} = \frac{S_{in}/N_{in}}{S_{ut}/N_{ut}} \quad \text{Ekvation 51}$$

Denna definition finns illustrerad i Figur 25 nedan där en förstärkare illustreras som en fyropol med in- och utgång. Notera att effektförstärkningen $G = S_{ut}/S_{in}$, så Ekvation 51 kan skrivas på den praktiska formen $F = N_{ut}/GN_{in}$.



Figur 25 Samband mellan en förstärkares signal-brusförhållande på in- och utgång samt dess brusfaktor F .

Definitionen visar att brusfaktorn alltid är $F > 1$ (dimensionlös). Ofta anges brusfaktor i dB och då gäller att $F > 0$ dB. Eftersom brus i allra högsta grad är temperaturberoende gäller definitionen vid temperaturen $T_0 = 290$ K. Det betyder också att begreppet brusfaktor behöver användas med viss kunskap och försiktighet.

Det internt genererade bruset i förstärkaren i Figur 25 adderas effektmässigt på så sätt att $N_{ut} = GN_{in} + N_{int}$. Eftersom $S_{ut} = GS_{in}$ kan Ekvation 51 skrivas som

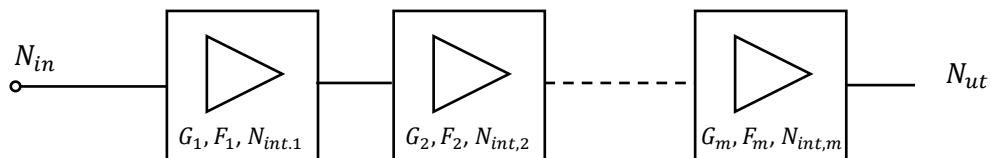
$$F = \frac{SNR_{in}}{SNR_{ut}} = \frac{S_{in}/N_{in}}{GS_{in}/N_{ut}} = \frac{N_{ut}}{GN_{in}} = \frac{GN_{in} + N_{int}}{GN_{in}} = \\ = 1 + \frac{N_{int}}{GN_{in}}$$

Ekvation 52

I Ekvation 52 har vi gjort antagandet att interna bruset N_{int} adderas till det förstärkta ingångsbruset GN_{in} för att ge totala bruset $N_{ut} = GN_{in} + N_{int}$. I modellen tänker vi oss alltså att allt intern genererat brus i förstärkarens inre delar flyttas till utgången för att summeras med GN_{in} . Detta är en praktisk black box-modell för att hantera brus på liknande sätt som att brusfaktor är ett praktiskt black box-mått att hantera förstärkarbrus.

3.6.2. Brusfaktor för kaskadkopplade förstärkare

Förstärkare kopplas ofta i kaskad för att öka den totala förstärkningen. Figur 26 visar ett blockschema för m stycken kaskadkopplade förstärkare med olika egenskaper. Eftersom varje förstärkare adderar nytt intern-genererat brus i respektive steg behöver vi ett mått på den totala brusfaktorn i kaskadkopplingen.



Figur 26 Kaskadkoppling av m stycken förstärkare med olika förstärkning och brusfaktorer.

Med modellen som definierats i kapitel 3.6.1 vet vi att det interngenererade bruset N_{int} i en förstärkare adderas till det förstärkta ingångsbruset GN_{in} så att bruset på utgången

kan skrivas som $N_{ut} = GN_{in} + N_{int}$ (alternativt ser man internbruset som en liten extra bruskälla på ingången med storleken $\frac{N_{int}}{G}$ som sedan adderas till N_{in}). I en kaskadkoppling kommer då detta brus på utgången ge ingångsbruset på nästa förstärkare. Totala bruset på utgången på kedjans sista förstärkare bli då

$$\begin{aligned}
 N_{ut} &= G_m N_{ut,m-1} + N_{int,m} = \\
 &= G_m G_{m-1} N_{ut,m-2} + G_m N_{int,m-1} + N_{int,m} = \\
 &= G_m G_{m-1} \cdots G_2 G_1 N_{in} + G_m G_{m-1} \cdots G_2 N_1 + \cdots \\
 &\quad \cdots + G_m N_{int,m-1} + N_{int,m}
 \end{aligned} \tag{Ekvation 53}$$

I Ekvation 53 har vi successivt lagt till förstärkare efter förstärkare, räknat från höger i Figur 26 och därmed summerat ihop de olika bruskällornas bidrag till totala bruset på förstärkarkedjans utgång.

Vi vet också från Ekvation 52 att brusfaktorn kan skrivas som $F = N_{ut}/GN_{in}$, där hela kedjans effektförstärkning är $G = G_m G_{m-1} \cdots G_2 G_1$. Detta ger tillsammans med Ekvation 52 den totala brusfaktorn för förstärkarkedjan

$$\begin{aligned}
 F &= \frac{N_{ut}}{GN_{in}} = \\
 &= \frac{G_m \cdots G_1 N_{in} + G_m \cdots G_2 N_{int,1} + \cdots + G_m N_{int,m-1} + N_{int,m}}{G_m G_{m-1} \cdots G_2 G_1 N_{in}} = \\
 &= 1 + \frac{N_{int,1}}{G_1 N_{in}} + \frac{N_{int,2}}{G_1 G_2 N_{in}} + \cdots + \frac{N_{int,m}}{G_1 G_2 \cdots G_m N_{in}}
 \end{aligned} \tag{Ekvation 54}$$

De första två termerna (gulmarkerade) i sista raden i Ekvation 54 kan ersättas med F_1 , eftersom $F_1 = 1 + N_{int,1}/G_1 N_{in}$. Den tredje termen (grönmarkerad) $N_{int,2}/G_1 G_2 N_{in}$ kan ersättas med $(F_2 - 1)/G_1$ osv. Detta ger slutligen uttrycket som går under namnet Friis formel eller Friis brusformel. Det senare är att föredra eftersom det finns ytterligare en formel namngiven efter Harald Friis.

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_m - 1}{G_1 G_2 \cdots G_{m-1}} \tag{Ekvation 55}$$

Denna ekvation visar på ett intressant fenomen. Den totala brusfaktorn för kaskadkopplade faktorer påverkas mest av den första förstärkarens brusfaktor. Alla övriga brusfaktorer har lägre påverkan eftersom de dels subtraheras med 1, dels divideras med förstärkning. Ju senare i kedjan, desto mindre påverkan har en förstärkares brusfaktor. Denna slutsats är i.o.f.s. ganska naturlig eftersom brus som genereras tidigt i kedjan kommer förstärkas mycket och därmed få större påverkan på det slutliga signal-brusförhållandet.

Ekvation 55 är viktig då t.ex. förstärkning i en mottagare behövs. Ofta behövs stor förstärkning vilket man löser genom att kaskadkoppla flera olika förstärkare. Ekvationen visar att den första förstärkarens brusegenskaper är mest kritisk för att få ett bra signal-brusförhållande totalt sett. Den förstärkaren brukar därför designas för att ha låg

brusfaktor, ofta på bekostnad av förstärkningen. En sådan förstärkare kallas för lågbrusförstärkare (low noise amplifier, LNA).

3.6.3. Brustemperatur

Begreppet brusfaktor för förstärkare gäller vid temperaturen $T_0 = 290$ K, se 3.6.1. I verkligheten så är förstås temperaturen inte exakt 290 K. Ofta så är skillnaden liten och sambanden som definierats och härletts i kapitel 3.6.1 och 3.6.2 stämmer ganska väl. Men i vissa tillämpningar så behöver man ta hänsyn till den verkliga temperaturen, t.ex. i lågbrustillämpningar där man kyler kretsar för att minimera bruset.

Vid anpassning så gäller att bruset på en förstärkares ingång är $N_{in} = k_B T_{in} B_n$ (maximalt överförd bruseffekt från källa till last enligt Ekvation 43). T_{in} betecknar temperaturen på förstärkarens ingång. På förstärkarens utgång har vi då bruseffekten $N_{ut} = G N_{in} + N_{int} = G k_B T_{in} B_n + N_{int}$. Den vänstra termen är då bruset på förstärkarens ingång förstärkt med faktorn G , medan den högra faktorn är det extra brus N_{int} som förstärkarens inre elektronik adderar. Detta kan omformuleras så att vi ser N_{int} som en bruskälla på ingången i stället med en viss temperatur som adderas till temperaturen T . Detta ser ut på följande vis

$$N_{ut} = G k_B T_{in} B_n + N_{int} = G k_B B_n \left(T_{in} + \frac{N_{int}}{G k_B B_n} \right) = \\ = G k_B B_n (T_{in} + T_e)$$
Ekvation 56

Termen $T_e = N_{int} / G k_B B_n$ definieras som förstärkarens effektiva brustemperatur. T_e motsvarar då ”extra temperatur” som förstärkaren adderar i förhållande till en generell temperatur T_{in} (i termer av bruseffekt på ingången). Vi har alltså nu ett mått på förstärkarens inre brus som inte är låst till definitionen av brusfaktor med $T_0 = 290$ K. Ingångsbruset T_{in} kan anta vilket värde som helst.

Sambandet mellan effektiv brustemperatur och brusfaktor är

$$T_e = (F - 1) T_0$$
Ekvation 57

$$F = 1 + \frac{T_e}{T_0}$$
Ekvation 58

Sambandet mellan signal-brusförhållande på ut- och ingång kan också formuleras med effektiv brustemperatur

$$\frac{S_{ut}}{N_{ut}} = \frac{S_{in}}{N_{in}} \cdot \frac{1}{1 + \frac{T_e}{T_{in}}}$$
Ekvation 59

Vid kaskadkopplade förstärkare blir den totala effektiva brustemperaturen

$$T_e = T_{e,1} + \frac{T_{e,2}}{G_1} + \frac{T_{e,3}}{G_1 G_2} + \dots + \frac{T_{e,m}}{G_1 G_2 \dots G_{m-1}}$$
Ekvation 60

Det är lockande att anta att brusfaktor och brustemperatur är två mått för samma sak, men det finns vissa skillnader. Brusfaktorn F är ett mått som anger degraderingen i signal-brusförhållande i en förstärkare, då temperaturen är bestämd till 290 K. Detta mått är mycket användbart vid systemberäkningar där man önskar veta kvaliteten på en signal, t.ex. vid signaldetektering i ett kommunikationssystem. Effektiva brustemperaturen T_e anger hur mycket brus som en förstärkare adderar i relation till bruset på ingången. Detta mått är användbart för brusberäkningar och nödvändigt t.ex. då temperaturen på ingångsbruset avviker mycket från T_0 . Sambandet mellan dem är dock enkelt vilket vi ser i Ekvation 57 och Ekvation 58.

3.7. Brus i transmissionsledningar

Transmissionsledningar behandlas i kapitel 4, men redan här behandlas hur en transmissionsledning påverkar bruset i ett system. En transmissionsledning är en komponent för överföring av elektromagnetiska vågor bundet till en struktur. Ett typiskt exempel är en koaxialkabel, som överför elektromagnetiska mellan två punkter där vågorna är bundna mellan kabelns centerledare och skärm mellan.

Transmissionsledningen är en icke ideal komponent som dämpar signalen som ska överföras. Dämpningen, eller förlisten, betecknas ofta L (som i engelskans Loss). Dämpning är en dimensionslös enhet som ofta anges i decibel. I datablad för kablar och andra typer av transmissionsledningar anger man ofta dämpning i förlust per längdenhet, eller dB/m.

Om alla komponenter i ett system är anpassade (mer om anpassning i kapitel 4), dvs. alla in- och utimpedanser är anpassade för maximal effektorverföring i alla steg, kommer transmissionsledningen mellan två komponenter dämpa nyttosignalen med faktorn L , vilket vi kan skriva som $S_{ut} = S_{in}/L$. Det går självtklart att se transmissionsledningen som en "dålig förstärkare" med sambandet $S_{ut} = G_{tl}S_{in}$ med en effektförstärkning $G_{tl} < 1$.

När det gäller termiskt brus så dämpas inte detta av transmissionsledningen. För det termiska bruset gäller att $N_{ut} = N_{in}$. Det kan tyckas konstigt att brus inte påverkas av kabelns dämpning och det är egentligen en förenklad modell. Mer fysikaliskt korrekt är att se transmissionsledningen som en komponent som både dämpar och genererar termiskt brus, där totalen tills slut blir att $N_{ut} = N_{in}$. Andra typer av brus kan dock dämpas på samma sätt som nyttosignalen. Till exempel kommer man made-noise som plockas upp av en antenn i ett kommunikationssystem dämpas i transmissionsledningen som ansluter antennen på samma sätt som nyttosignalen kommer dämpas.

För transmissionsledning och termiskt brus gäller då liknande samband som vi har för förstärkare. Vi kan beräkna en brusfaktor som

$$F = \frac{SNR_{in}}{SNR_{ut}} = \frac{S_{in}/N_{in}}{S_{ut}/N_{ut}} = \frac{S_{in}}{S_{ut}} = L$$

Ekvation 61

där L kan ses som en förstärkning



$$L = \frac{1}{G_{tl}}$$

Ekvation 62

Sambanden i Ekvation 61 och Ekvation 62 kan sedan användas i t.ex. Friis brusformel (Ekvation 55).

4. Transmissionsledningar

Med begreppet transmissionsledning [31] avses en struktur (t.ex. ledning) som är konstruerad för att leda elektromagnetiska vågor mellan två punkter. Teorin kring transmissionsledningar behöver beaktas då transmissionsledningens längd är tillräckligt stor i relation till våglängden på den elektromagnetiska vågen. Ett exempel, elektrikern som gör installationsjobb i ett hus behöver inte ha kunskap om transmissionsledningsteori eftersom våglängden vid frekvensen 50 Hz i ledningen är flera tusen km vilket vida överstiger längden på elkablarna. Däremot behöver nätbolagen som hanterar elnätet över en hel kontinent ibland ta hänsyn till transmissionsledningsteori. En transmissionsledning modelleras som en fyropol, se Figur 27.



Figur 27 Modell för en transmissionsledning. Parameters Z_0 kallas för ledningens karakteristiska impedans.

By Omegatron - The source code of this SVG is valid. This vector image was created with Inkscape by user Omegatron., CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=2522075>

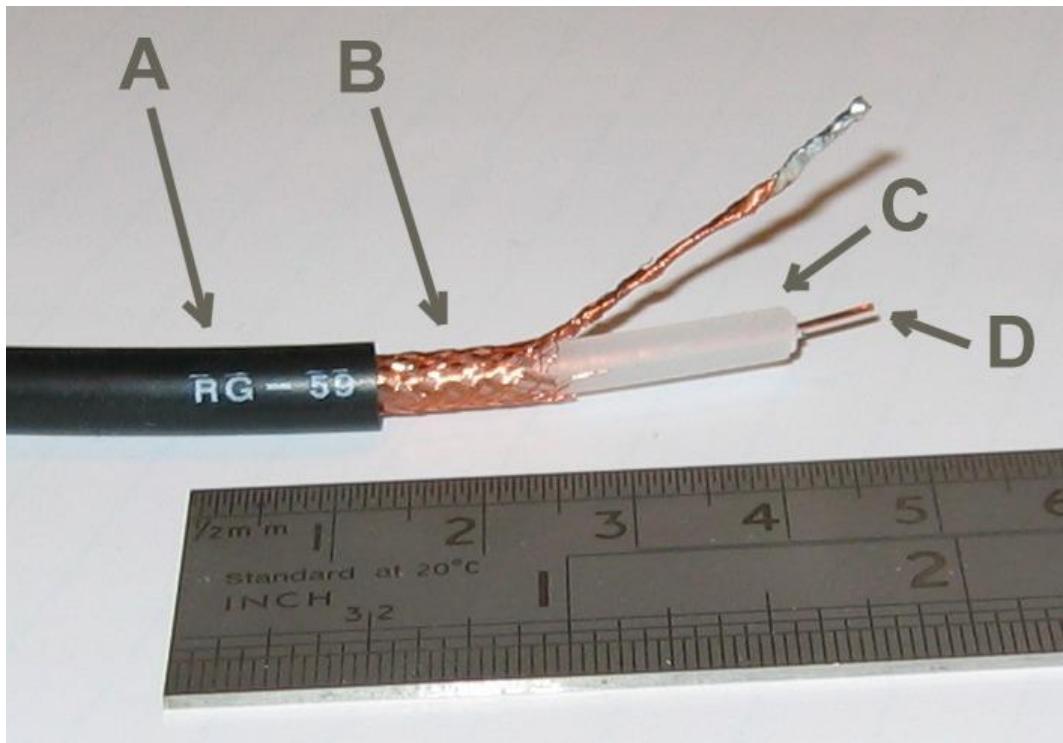
Förutom villkoret att transmissionsledningens längd är tillräckligt stor i jämförelse med våglängden, ställs ibland villkoret att ledningen ska kunna transportera elektromagnetisk energi för frekvenser från önskad frekvens hela vägen ned till frekvensen 0 Hz, dvs. leda likström. Den enda praktiska betydelse som detta villkor har är att vissa typer av ledningar ibland inte nämns som transmissionsledning, t.ex. vågledare [32].

4.1. Olika typer av transmissionsledningar

Transmissionsledningar finns i en rad olika former och typer. I detta kapitel presenteras kortfattat en del vanliga transmissionsledningar.

4.1.1. Koaxialkabel

Koaxialkablar är mycket vanliga transmissionsledningar pga. dess flexibilitet, goda skärmning och relativt låga dämpning. Koaxialkablar finns i en stor mängd olika dimensioner, kvaliteter och prisklasser och det finns flera olika typer av kontakter som är anpassade för koaxialkablar, t.ex. BNC, SMA, TNC, N-kontakt osv. Figur 28 visar uppbyggnaden av en koaxialkabel.

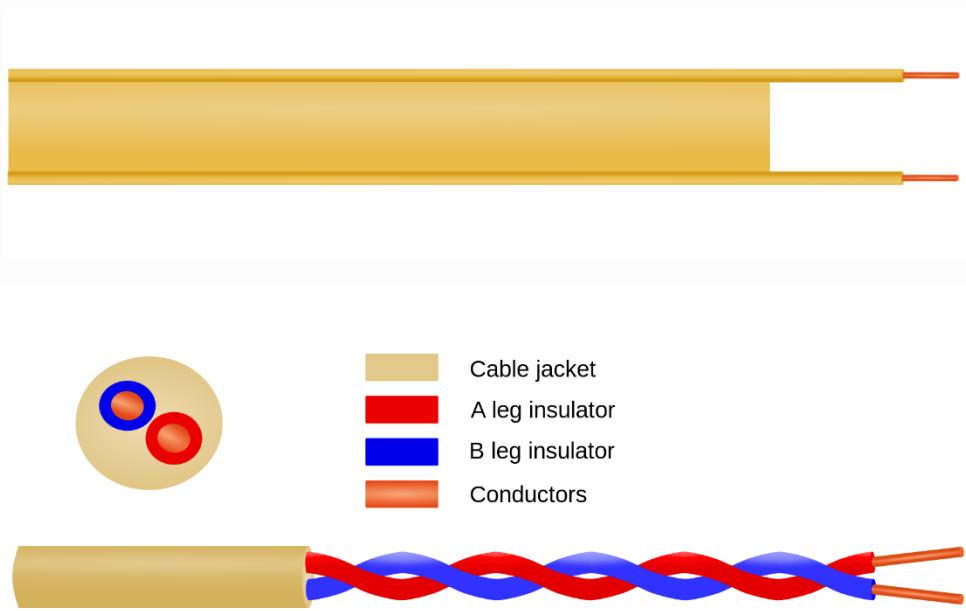


Figur 28 Koaxialkabel och dess delkomponenter yttre plasthölje (A), skärm (B), isolator (C) samt mittledare (D).
CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=50762>

I en koaxialkabel så propagerar den elektromagnetiska energin längs kabel, där det elektromagnetiska fältet är inneslutet mellan skärm och mittledare. Detta betyder att kabeln inte stör eller störs av omgivningen i någon större omfattning, eftersom skärmen ger en bra avskärmning mot omvärlden.

4.1.2. Parledare

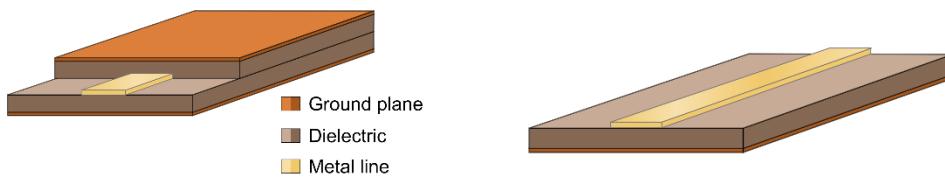
Parledare är en av de absolut mest enkla transmissionsledarna. En parledare består helt enkelt av två stycken ledare av samma typ (t.ex. koppartråd). Parledare finns i lite olika former, t.ex. balanserad parledare och tvinnad parledare [33], se Figur 29.



Figur 29 Exempel på parledare. Balanserad parledare (överst) och tvinnad parledare (underst)
 Övre bild: By Courtesy Spinningspark at Wikipedia, CC BY-SA 3.0,
<https://en.wikipedia.org/w/index.php?curid=24413398>
 Undre bild: By Courtesy Spinningspark at Wikipedia, CC BY-SA 3.0,
<https://en.wikipedia.org/w/index.php?curid=24413375>

4.1.3. Planär transmissionsledning

Planära transmissionsledningar är också mycket vanliga eftersom de kan byggas med mörnsterkortsteknologi. Två vanliga varianter av planära transmissionledningar är stripline och microstrip [34].



Figur 30 Stripline (vänster bild) och microstrip (höger bild)
 Vänster bild: By Courtesy Spinningspark at Wikipedia, CC BY-SA 3.0,
<https://en.wikipedia.org/w/index.php?curid=52984557>
 Höger bild: By Courtesy Spinningspark at Wikipedia, CC BY-SA 3.0,
<https://en.wikipedia.org/w/index.php?curid=52984582>

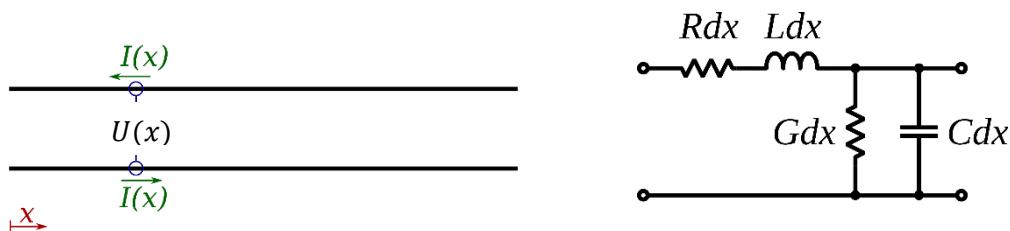
4.2. Transmissionledningar och telegrafekvationen

En enkel form av transmissionsledning som leder energi i form av en elektromagnetisk våg består av två parallella, närliggande ledare. Ledarna är åtskilda av någon form av dielektrikum. En modell över en sådan transmissionsledning och en infinitesimal

(oändligt kort) del med längden dx av transmissionsledningen finns i Figur 31. Längs ledarna finns en resistans med storleken Rdx och en induktans Ldx . Mellan ledarna finns en kapacitans Cdx och dielektrikumet har en konduktans Gdx .

I varje tvärsnitt av transmissionsledningen flyter en ström $I(x)$, som har samma storlek i de två olika ledningarna men är motriktade (ström som flyter i en riktning i ena ledaren återvänder som en returström i den andra ledaren). I tvärsnittet finns också en spänning $U(x)$ mellan ledarna.

Här är viktigt att inse skillnaden mot traditionell kretsteori där vi ser resistans, kapacitans osv. som punktformade egenskaper i ett kretsschema. Här har vi i stället en distribuerad modell, där transmissionsledningens elektriska egenskaper är utfördelade över ledningens hela längd. Från denna distribuerade modell skapar vi sedan en elektrisk modell över en infinitesimal del av transmissionsledningen. Skillnaden ligger i storleken på strukturen i förhållande till våglängden på den frekvens som vi analyserar. I kretsteorin antar vi att en komponent har en fysisk storlek som är mycket mindre än våglängden och därmed kan anses vara punktformig. I transmissionsledningteorin gäller inte detta och vi behöver därför skapa en infinitesimal, punktformig modell enligt Figur 31. I den modellen har vi då t.ex. skapat en kapacitans Cdx av ledningens tvärkapacitans C (som har enheten F/m). Motsvarande gäller de övriga parametrarna, och R , L , C och G kallas ledningens primärkonstanter. Dessa är i sin tur beroende på ledningens geometri och material.



Figur 31 Modell över en transmissionsledning som leder elektromagnetisk energi (vänster bild) och en infinitesimal del av densamma i en punkt x (höger bild). Denna infinitesimala fyrapotstående återupprepas sedan i kaskad ett oändligt antal gånger för att bygga upp hela transmissionsledningen.

Vänster bild modifierad från originalen: By Sbyrnes321 - Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=20295336>

Höger bild: By Omegatron - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=1960030>

Spänning och ström kommer förändras över den infinitesimala kretsen i Figur 31 och är därför en funktion av parametern x som löper längs transmissionsledningen längd. De är självklart också en funktion av tiden t eftersom det är växelströms- och växelspänningsförlopp som studeras. Genom att använda Kirchhoff's lagar i frekvensdomänen och $j\omega$ -metoden (se Figur 32) kan de s.k. telegrafekvationerna [35] ställas upp

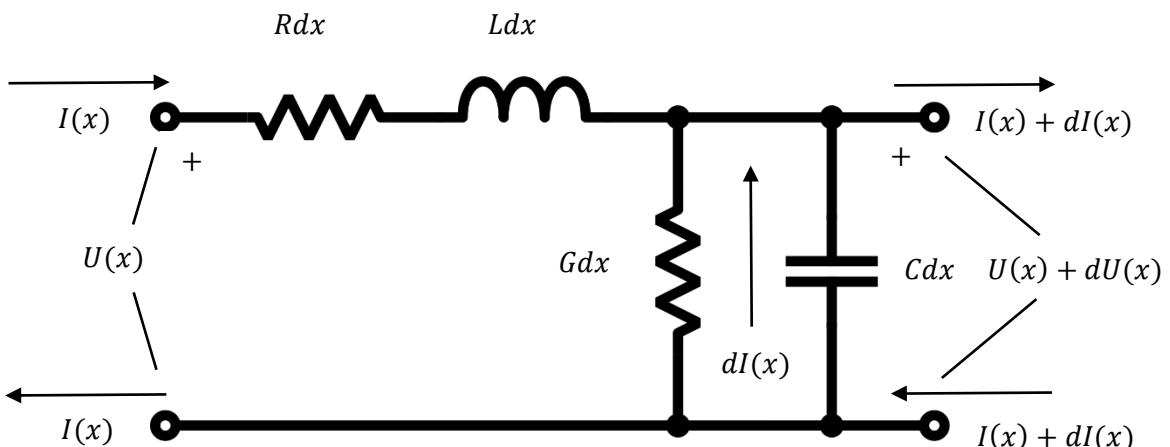
$$\frac{dU(x)}{dx} = -(R + j\omega L) \cdot I(x)$$

Ekvation 63

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

Ekvation 64

Telegrafekvationerna beskriver hur spänning och ström förändras (därför derivatorna $dU(x)/dx$ och $dI(x)/dx$) över transmissionsledningen och de är exempel på ett system med kopplade differentialekvationer. De uttrycks ofta som funktion av inte bara längden x utan också tiden t , men här har vi uttryckt dem i frekvensplanet och använt den praktiska formulering man får när man använder $j\omega$ -metoden.



Figur 32

Modell över telegrafekvationerna för en transmissionsledning.

Modifierad från: By Omegatron - Own work, CC BY-SA 3.0,

<https://commons.wikimedia.org/w/index.php?curid=1960030>

I vissa fall räcker det att studera specialfallet där ledningen är ideal (förlustfri ledning), dvs. den ledarna har $R = 0$ och dielektrikat har $G = 0$. Ekvation 63 och Ekvation 64 får då ett enklare utseende där termerna med R och G försvinner.

4.3. Vågekvationen

Genom att derivera Ekvation 63 ytterligare en gång med avseende på x , får vi

$$\frac{d^2U(x)}{dx^2} = -(R + j\omega L) \cdot \frac{dI(x)}{dx}$$

Ekvation 65

Ekvation 64 kan sättas in i Ekvation 65, vilket ger vågekvationen för spänning

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

Ekvation 66

Motsvarande manöver kan genomföras för strömmen $I(x)$ och då har vi naturligtvis vågekvationen för ström

$$\frac{d^2I(x)}{dx^2} = (R + j\omega L) \cdot (G + j\omega C) \cdot I(x) \quad \text{Ekvation 67}$$

Här är det lämpligt att återigen påminna sig om att analysen sker i frekvensplanet och med $j\omega$ -metoden. Ofta ser man analysen genomförd i tidsplanet och då ingår tidsderivator också. Men med $j\omega$ -metoden får vi en något enklare metod att komma fram till vågekvationerna eftersom en tidsderivata motsvarar en multiplikation med $j\omega$. I den mer generella formen med tidsderivator

4.3.1. Lösning för en ideal, förlustfri transmissionsledning

Ekvation 66 och Ekvation 67 är två icke kopplade differentialekvationer som kan lösas var för sig. De har uppenbarligen samma struktur och bör alltså ha samma typ av lösning. I det idealala fallet med $R = G = 0$ och en ideal, förlustfri transmissionsledning kan ekvationerna skrivas som

$$\frac{d^2U(x)}{dx^2} + \omega^2 LC \cdot U(x) = 0 \quad \text{Ekvation 68}$$

$$\frac{d^2I(x)}{dx^2} + \omega^2 LC \cdot I(x) = 0 \quad \text{Ekvation 69}$$

Den generella lösningen i tidsplanet till Ekvation 68 är en summa av två stycken spänningsvågor som propagerar längs transmissionsledningen, en i framåtriktningen och en i bakåtriktningen. Detta formuleras som

$$\begin{aligned} U(x, t) &= U^+(x, t) + U^-(x, t) = \\ &= U^+(t - x\sqrt{LC}) + U^-(t + x\sqrt{LC}) \end{aligned} \quad \text{Ekvation 70}$$

Faktorn $1/\sqrt{LC}$ motsvarar då vågens utbredningshastighet v i transmissionsledningen. Spänningen i varje punkt på transmissionsledningen är alltså en summa av två stycken spänningar som var för sig varierar över tid och längd, går i motsatt riktning och rör sig med samma hastighet genom ledningen.

Strömmen $I(x, t)$ har förstås en motsvarande lösning $I = I^+ + I^-$

$$\begin{aligned} I(x, t) &= I^+(x, t) + I^-(x, t) = \\ &= I^+(t - x\sqrt{LC}) + I^-(t + x\sqrt{LC}) \end{aligned} \quad \text{Ekvation 71}$$

Vi önskar dock finna sambandet mellan $U(x, t)$ och $I(x, t)$ uttryckt i transmissionsledningens primärkonstanter L och C . Genom att derivera Ekvation 70 med avseende på x får vi sambandet

$$\frac{dU(x, t)}{dx} = -\sqrt{LC} \frac{dU^+(t - x\sqrt{LC})}{dx} + \sqrt{LC} \frac{dU^-(t + x\sqrt{LC})}{dx} \quad \text{Ekvation 72}$$

där \sqrt{LC} är den inre derivatan. Denna derivata har vi också i Ekvation 63, och sätter vi dessa två olika uttryck för derivatan lika (med $R = 0$) så får vi

$$\begin{aligned} & -\sqrt{LC} \left(\frac{dU^+(t - x\sqrt{LC})}{dx} - \frac{dU^-(t + x\sqrt{LC})}{dx} \right) \\ &= -j\omega L \cdot \frac{dI(x)}{dx} = \\ &= -j\omega L \left(\frac{dI^+(t - x\sqrt{LC})}{dx} + \frac{dI^-(t + x\sqrt{LC})}{dx} \right) \end{aligned} \quad \text{Ekvation 73}$$

Genom att förenkla och koppla samman termer för framåtgående våg och bakåtgående våg ser vi att

$$\begin{aligned} \frac{dU^+(t - x\sqrt{LC})}{dx} &= j\omega \sqrt{\frac{L}{C}} \cdot \frac{dI^+(t - x\sqrt{LC})}{dx} \\ \frac{dU^-(t + x\sqrt{LC})}{dx} &= -j\omega \sqrt{\frac{L}{C}} \cdot \frac{dI^-(t + x\sqrt{LC})}{dx} \end{aligned} \quad \text{Ekvation 74}$$

Dessa ekvationer kan sedan integreras för att koppla samman $U(x, t)$ och $I(x, t)$ i stället för att koppla samman dess derivator. Termen $\sqrt{L/C}$ representerar då en impedans. Denna impedans kallas för transmissionsledningens karakteristiska impedans och betecknas Z_0 . Som vi kommer se under resterande delar i kapitlet om transmissionsledningar är detta en mycket viktig parameter för en transmissionsledning. I en slutlig, enklare form kan vi skriva sambandet mellan U och I som

$$\begin{aligned} U &= U^+ + U^- \\ I &= I^+ + I^- = \frac{U^+}{Z_0} - \frac{U^-}{Z_0} \end{aligned} \quad \text{Ekvation 76}$$

Minustecknet för sista termen i Ekvation 77 är en naturlig konsekvens av de definitioner vi har i Ekvation 70 och Ekvation 71. Spänningsvågens framåtgående och bakåtgående vågor har båda positiv spänning definierad i samma riktning, medan strömvågornas respektive delar har olika definition för positiv riktning (I^+ positiv åt höger i Figur 32, I^- åt vänster).

Sammanfattningsvis så är slutsatsen att den allmänna lösningen för telegrafekvationerna för en ideal transmissionsledning är att det kan existera två vågor på transmissionledningen som utbreder sig i motsatta riktningar. Hastigheten för

vågorna är densamma oavsett riktning, vilket känns intuitivt korrekt. Denna hastighet bestäms av

$$v = \frac{1}{\sqrt{LC}} \quad \text{Ekvation 78}$$

Spänning och ström är sammankopplade med ett uttryck som liknar Ohms lag via ledningens karakteristiska impedans

$$Z_0 = \sqrt{\frac{L}{C}} \quad \text{Ekvation 79}$$

Den totala spänningen respektive strömmen i en punkt på ledningen är summan av respektive framåt- och bakåtgående vågor.

4.3.2. Lösning för en allmän transmissionsledning [31]

Om ledningen inte är förlustfri går det fortfarande att härleda uttryck för den transmissionsledningens karakteristiska impedans och vågutbredningshastighet. Karakteristiska impedansen har en del likheter med Ekvation 79

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad \text{Ekvation 80}$$

När det gäller utbredningshastigheten är det något mer komplicerat. När R och G är skilda från noll kommer vågen dämpas. I den generella lösningen definieras en utbredningskonstant

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \quad \text{Ekvation 81}$$

γ är då ett komplext tal $\gamma = \alpha + j\beta$. α motsvarar då en dämpning medan β är relaterad till vågens utbredningshastighet. Beroende på frekvens och storleken på R , L , G och C kan olika approximativa lösningar härledas. Till exempel så gäller att en lågförlustledning vid höga frekvenser har

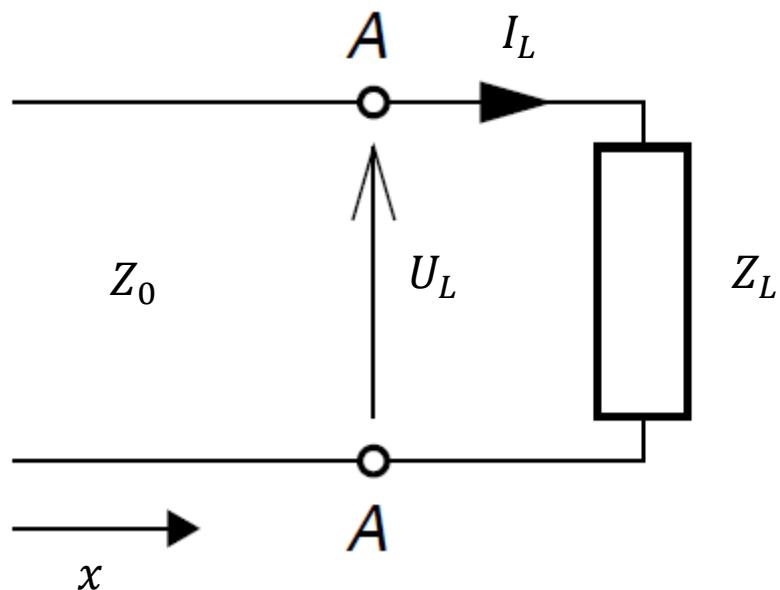
$$\alpha \approx \frac{1}{2} \sqrt{LC} \left(\frac{R}{L} + \frac{G}{C} \right) \quad \text{Ekvation 82}$$

$$\beta \approx \omega \sqrt{LC} \quad \text{Ekvation 83}$$

4.4. Reflektion och transmission

Modellen som används för vågekvationen i en transmissionsledning (se Figur 31) förutsätter att varje infinitesimal del av transmissionsledningen är likadan som de omgivande infinitesimala delarna. Med andra ord så förutsätter härledningen att transmissionsledningens egenskaper är lika över hela dess längd.

I praktiken förekommer det många fall där detta inte är uppfyllt. Några exempel är en ledning som är kortsluten eller öppen, i gränssnittet mellan transmissionskabel och kontakt, en ledning som är avslutad med en last eller en ledning som är skarvad med en annan. Lösningen på denna typ av problem är att studera vad som händer i gränssnittet där den elektromagnetiska vågen upplever olika egenskaper (impedanser). Figur 33 visar en enkel modell av en transmissionsledning där den elektromagnetiska vågen i ett snitt A-A upplever en annan impedans Z_L .



Figur 33 Modell över transmissionsledning med karakteristiska impedansen Z_0 som i snittet A-A har lasten Z_L . Lasten kan vara en verlig last eller en modell över t.ex. en kontakt eller annan transmissionsledning.
Figur från [26], s. 352.

I varje del av ledningen är den totala spänningen summan av en framåtgående våg och en bakåtgående våg, $U(x, t) = U^+(x, t) + U^-(x, t)$ (se Ekvation 70). Motsvarande gäller strömmen. I snittet A-A (över lasten Z_L) gäller då att

$$U^+ + U^- = U_L \quad \text{Ekvation 84}$$

$$I^+ + I^- = I_L \quad \text{Ekvation 85}$$

Eftersom $I^+ = U^+/Z_0$, $I^- = -U^-/Z_0$ och $I_L = U_L/Z_L$ har vi med Ekvation 85

$$I^+ + I^- = \frac{U^+}{Z_0} - \frac{U^-}{Z_0} = \frac{U_L}{Z_L} \quad \text{Ekvation 86}$$

Ekvation 84 och Ekvation 86 uttrycker spänningen U_L i snittet A-A på två olika sätt, vilket vi kan använda för att beräkna hur framåtgående vågen och bakåtgående vågen förhåller sig till varandra.

$$U^+ + U^- = Z_L \left(\frac{U^+}{Z_0} - \frac{U^-}{Z_0} \right) \quad \text{Ekvation 87}$$

Denna ekvation kan omformas för att beräkna förhållandet U^-/U^+ .

$$\frac{U^-}{U^+} = \frac{Z_L - Z_0}{Z_L + Z_0} = \rho \quad \text{Ekvation 88}$$

Ekvation 88 är ett mycket viktigt samband i transmissionsledningsteorin, och vi har där definierat och infört den s.k. reflektionskoefficienten ρ .

Den fysikaliska tolkningen av ρ och Ekvation 88 är intuitivt ganska enkel. En våg som rör sig längs en ledning kommer reflekteras då den stöter på en komponent med andra vågegenskaper. Vi ser samma fenomen i vardagen. Enkla exempel är solljusets reflektion i en glasruta, eller ekot vi hör när vi ropar högt mot en hög, vertikal bergvägg. ρ är då ett mått på denna reflektion. Notera att $\rho = 0$ då $Z_L = Z_0$ vilket stämmer väl med det specialfall då transmissionsledningen inte har en last utan fortsätter i all oändlighet. Då uppstår det inte några reflektioner, vilket Ekvation 88 också säger. $Z_L = Z_0$ motsvara även det fall där vi säger att vi har en anpassad last till transmissionsledningen.

På motsvarande sätt kan en transmissionskoefficient τ definieras och beräknas. τ är ett mått på den våg som fortsätter in i lasten (U_L) och definieras som

$$\tau = \frac{U_L}{U^+} = \frac{2Z_L}{Z_L + Z_0} \quad \text{Ekvation 89}$$

4.5. Några vanliga vågutbredningsfall

4.5.1. Anpassad last $Z_L = Z_0$

Som noterats i förra kapitlet så gäller vid anpassning att $Z_L = Z_0$. Då har vi att

$$\rho = \frac{Z_L - Z_0}{Z_L + Z_0} = 0 \quad \text{Ekvation 90}$$

$$\tau = \frac{2Z_L}{Z_L + Z_0} = 1 \quad \text{Ekvation 91}$$

Inget av den framåtgående vågen reflekteras alltså, utan vågen fortsätter in i lasten.

4.5.2. Kortsluten last $Z_L = 0$

Med $Z_L = 0$ (kortslutning) gäller

$$\rho = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{-Z_0}{Z_0} = -1 \quad \text{Ekvation 92}$$

$$\tau = \frac{2Z_L}{Z_L + Z_0} = 0 \quad \text{Ekvation 93}$$

Vågen totalreflekteras och genomgår ett teckenbyte. Inget av vågen fortsätter in i lasten.

4.5.3. Öppen last $Z_L = \infty$

Med $Z_L = \infty$ (öppen last) gäller

$$\rho = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{\infty}{\infty} = 1 \quad \text{Ekvation 94}$$

$$\tau = \frac{2Z_L}{Z_L + Z_0} = \frac{2 \cdot \infty}{\infty} = 2 \quad \text{Ekvation 95}$$

Även i detta fall sker en totalreflektion, och vågen återvänder med samma tecken och utbreder sig i transmissionsledningen bakåtriktning.

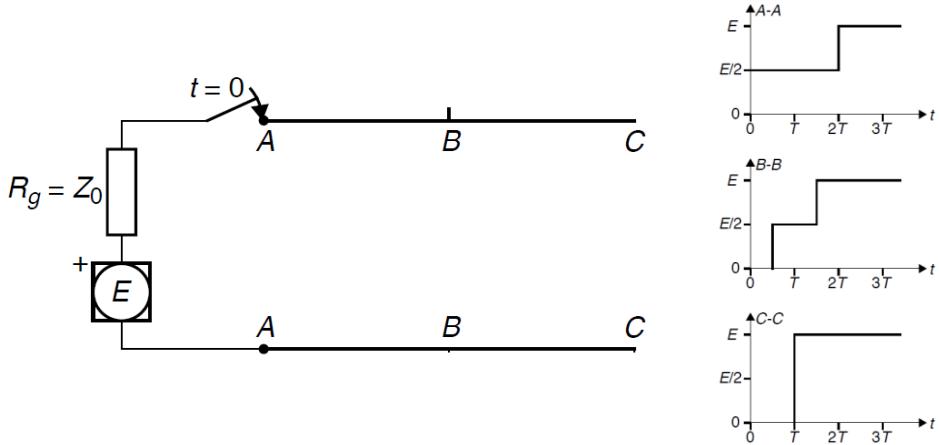
Transmissionskoefficienten $\tau = 2$ är lite mer svårtolkad, men en öppen last motsvarar en transmissionsledning utan last. Om vi mäter spänningen i det öppna gränssnittet kommer vi naturligtvis se en spänning. Den är enligt Ekvation 95 dubbelt så stor som den infallande vågen. Detta är en följd av att vi ser summan av den infallande vågen och den reflekterade vågen, där den senare är lika stor som den infallande och har samma tecken. Därför blir spänningen dubbelt så stor i det öppna snittet jämfört med U^+ .

4.6. Spänningsdiagrammet

Ett hjälpmittel för att analysera hur vågutbredningen i transmissionsledningen ser ut vid olika snitt och olika tidpunkter är spänningsdiagrammet. Figur 34 visar ett enkelt exempel med en transmissionsledning Z_0 som är ansluten till en likspänningsskälla med spänningen E och $R_g = Z_0$ i ena änden. I andra änden är ledningen öppen.

Vid tiden $t = 0$ kopplas spänningsskällan in med en switch. Då börjar vågen utbreda sig längs ledningen och figuren visar tre olika reflektionsdiagram i snitten A-A (generatoränden), B-B (mitt på ledningen) och C-C (den öppna änden). Vid $t = 0$ börjar vågen utbreda sig med spänningen $E/2$ i A-A. Anledningen till att spänningen bara är $E/2$ är att det är en spänningsdelning mellan R_g och Z_0 . Efter en viss tid når vågen B-B och efter ytterligare en stund är den framme vid C-C där den totalreflekteras med $\rho = 1$.

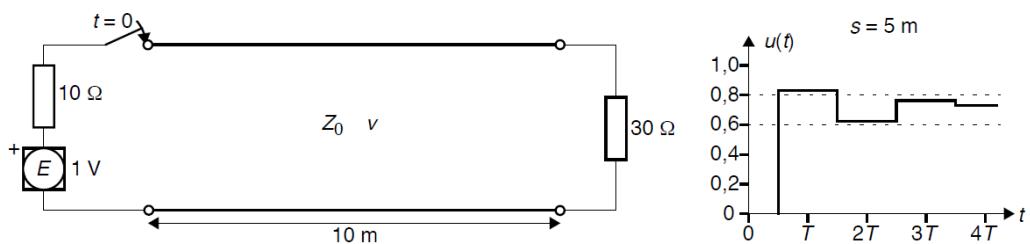
Reflektionen kommer utbreda sig i ledningens bakåtriktning. I C-C sker totalreflektionen omgående, vilket ger en total spänning $E/2 + E/2 = E$. Totalreflektionen fortsätter utbreda sig i ledningen till B-B och slutligen A-A. I dessa snitt så blir slutliga spänningen också E , men det dröjer alltså en stund



Figur 34 Vågutbredning av en stegfunktion längs en öppen ledning.
Figur från [26], s. 354.

Värt att notera, normalt förväntar vi oss att spänningen på generatorns utgång direkt når spänningen E eftersom vi inte har någon last i ledningens andra ände som drar någon ström. Men om ledningen är lång och vi måste applicera transmissionsledningsteori så dröjer det alltså en stund innan vågen har byggt upp sig till totala spänningen E . Om tiden det tar för vågen att utbreda sig längs transmissionsledningens hela längd är T , är tiden det tar för vågen att nå full spänning i generatoränden $2T$.

Ytterligare en sak som är värd att notera är att om $R_g \neq Z_0$ kommer den reflekterade vågen att reflekteras tillbaka och utbreda sig i framåtriktningen igen pga. skillnaderna i impedanser. Det blir ett oändligt antal reflektioner i båda ändarna och så småningom så har spänningen byggts upp i transmissionsledningens olika gränssnitt. Figur 35 visar ett mer komplext fall med en likspänningsgälla med $E = 1 \text{ V}$, $R_g = 10 \Omega$, $Z_0 = 50 \Omega$ och $R_L = 30 \Omega$. Reflektionsdiagrammet mitt på ledningen visar hur spänningen succesivt byggs upp av vågor som reflekteras fram och tillbaka, för



Figur 35 Vågutbredning i en transmissionsledning med missanpassning i både generator- och lastända.
Figur från [26], s. 356.

4.7. Sinusformade vågors utbredning i ideal transmissionledning

Eftersom sinusvågor har en central ställning i kommunikationssystem och signalbehandling är det naturligtvis av stort intresse att studera detta specialfall.

Den momentana spänningen från en generator skrivs som $\hat{U} \cos(\omega t)$. Om denna spänning tillförs ingången på en transmissionsledning kommer spänningen utbredas genom ledningen och variera som funktion av både tid och var på ledningen (x -koordinaten) vi befinner oss. I framåtriktningen kan vågen skrivas som

$U_f = U^+ \cos\left(\omega\left(t - \frac{x}{v}\right)\right)$ där U^+ är vågens amplitud. Eftersom det enligt kan existera en våg som utbreder sig i bakåtriktningen så har vi även $U_r = U^- \cos\left(\omega\left(t + \frac{x}{v}\right)\right)$. Jämför med Ekvation 70!

Med $j\omega$ -metoden och komplexa spänningar kan vi skriva dessa samband som

$$U_f = U^+ \cdot e^{-j\omega x/v} \quad \text{Ekvation 96}$$

$$U_r = U^- \cdot e^{j\omega x/v} \quad \text{Ekvation 97}$$

eftersom termerna $-\omega x/v$ och $\omega x/v$ kan ses som fasläget för U_f respektive U_r . Totala spänningen längs ledningen blir då

$$U = U_f + U_r = U^+ \cdot e^{-j\omega x/v} + U^- \cdot e^{j\omega x/v} \quad \text{Ekvation 98}$$

För strömmen gäller motsvarande

$$I = I_f + I_r = I^+ \cdot e^{-j\omega x/v} + I^- \cdot e^{j\omega x/v} \quad \text{Ekvation 99}$$

Eftersom $I^+ = U^+/Z_0$ och $I^- = -U^-/Z_0$ kan Ekvation 99 omformuleras till

$$I = \frac{1}{Z_0} (U^+ \cdot e^{-j\omega x/v} - U^- \cdot e^{j\omega x/v}) \quad \text{Ekvation 100}$$

Här definierar vi $\beta = \omega/v$, som kallas för transmissionsledningens faskonstant. Den har enheten rad/m och är ett slags mått hur många perioder av vågen som ryms inom en meter av transmissionsledningen. Eftersom vågen har en våglängd λ som är sträckan för en full period av vågen längs ledningen gäller sambandet $\beta\lambda = 2\pi$.

I fallet med missanpassning på en transmissionsledning, dvs. då vi har en last $Z_L \neq Z_0$ kommer reflektion att ske och en framåtgående våg U_f genererar en reflekterad, bakåtgående våg U_r . Motsvarande gäller förstås strömmen. Den totala spänningen U och strömmen I kommer då variera längs transmissionsledningen enligt Ekvation 98 och Ekvation 100. Därmed kommer också den av totala vågen upplevda impedansen variera eftersom den impedansen motsvarar kvoten U/I !

Vid lasten gäller att U^+ och U^- kan beräknas med Ekvation 88. Genom att kombinera denna ekvation med Ekvation 98 och sätta $x = 0$ (vilket ger $e^{-j\omega x/v} = e^{j\omega x/v} = 1$) vid lasten får vi reflektionsfaktorn ρ_L vid lasten som

$$\rho_L = \frac{U^-}{U^+} = \frac{Z_L - Z_0}{Z_L + Z_0} =$$

Ekvation
101

Notera att vi här valt att studera reflektionsfaktorn vid lasten och i det gränssnittet sätta $x = 0$. Om vi flyttar oss sträckan l längs transmissionsledningen i riktning mot generatorsidan (dvs. $x = -l$) kan vi studera hur kvoten U/I varierar, vilket då motsvarar den av vågen upplevda impedansen i punkten $x = -l$. Vi definierar här kretsens inimpedans som

$$\begin{aligned} Z_{in} &= \frac{U}{I} = \frac{U^+ \cdot e^{-j\omega(-l)/v} + U^- \cdot e^{j\omega(-l)/v}}{\frac{1}{Z_0}(U^+ \cdot e^{-j\omega(-l)/v} - U^- \cdot e^{j\omega(-l)/v})} = \\ &= Z_0 \frac{U^+ e^{j\omega l/v} \left(1 + \frac{U^-}{U^+} e^{-2j\omega l/v}\right)}{U^+ e^{j\omega l/v} \left(1 - \frac{U^-}{U^+} e^{-2j\omega l/v}\right)} = \\ &= Z_0 \frac{1 + \rho_L e^{-2j\omega l/v}}{1 - \rho_L e^{-2j\omega l/v}} = \{\beta = \omega/v\} \\ &= Z_0 \frac{1 + \rho_L e^{-2j\beta l}}{1 - \rho_L e^{-2j\beta l}} \end{aligned}$$

Ekvation
102

Ekvation 102 visar alltså hur inimpedansen på en transmissionsledning varierar i ett stationärfall med sinusformade vågor som utbreder sig på en transmissionsledning. Det gäller att särskilja detta från när vi studerar transinta förlopp som studerades med reflektionsdiagram i kapitel 4.6.

Om vi sätter x till hela transmissionsledningens längd, så är Z_{in} enligt Ekvation 102 den impedans som generatoren ser. Lasten på generatoren är alltså inte bara beroende på Z_L utan också transmissionsledningens egenskaper l och Z_0 ! Vid generatoren kan vi beräkna dess upplevda reflektionsfaktor ρ_{in} som

$$\rho_{in} = \frac{Z_{in} - Z_0}{Z_{in} + Z_0} = \frac{Z_0 \frac{1 + \rho_L e^{-2j\beta l}}{1 - \rho_L e^{-2j\beta l}} - Z_0}{Z_0 \frac{1 + \rho_L e^{-2j\beta l}}{1 - \rho_L e^{-2j\beta l}} + Z_0} = \rho_L e^{-2j\beta l}$$

Ekvation
103

Sambandet mellan ρ_{in} och ρ_L i Ekvation 103 visar att reflektionsfaktorn varierar längs transmissionsledningen på ett sådant sätt att den ändrar sitt fasläge med $e^{-2j\beta l}$.

Eftersom $\beta = 2\pi/\lambda$ kan Ekvation 102 och Ekvation 103 också skrivas på formen

$$Z_{in} = Z_0 \frac{1 + \rho_L e^{-j4\pi l/\lambda}}{1 - \rho_L e^{-j4\pi l/\lambda}}$$

Ekvation
104

$$\rho_{in} = \rho_L e^{-j4\pi l/\lambda}$$

Ekvation
105

4.8. Stående våg

Då två sinusformade vågor U_f och U_r rör sig i motsatta riktningar längs en transmissionsledning kommer de adderas till en total våg $U = U_f + U_r$ enligt Ekvation 98. Eftersom de rör sig med samma hastighet kommer de skapa ett stående våg-mönster. Vid vissa tidpunkter och positioner längs transmissionsledningen kommer framåtriktad och bakåtriktad sinusvåg ligga i fas och de adderas konstruktivt till det som kallas för buk. Om man rör sig en sträcka motsvarande fasförskjutningen 90° framåt eller bakåt från en buk kommer man nå punkter där sinusvågorna i stället bildar en nod.

Figur 36 visar hur ett stående våg-mönster bildas av två motriktade sinusformade vågor. Längs ledningen kommer ett mönster av spänningsbukar och -noder uppstå. Bukarna uppstår när vågor ligger i fas och i de punkter där spänningen når sitt maximala värde $U^+ + U^-$. För noderna gäller att i de tidpunkter där vågorna ligger i motfas kommer totala vågen ha en amplitud som är $U^+ - U^-$.

Även strömmen kan analyseras på detta sätt. Strömmen har naturligtvis sitt motsvarande stående våg-mönster, men den når sitt maximum (buk) i de punkter där spänningsvågen har sina noder och vice versa.

Stående våg-förhållandet [36] (VSWR, Voltage Standing Wave Ratio) definieras som kvoten mellan stående vågens (spänningen) amplitud i buken och amplituden i noden

$$S = \frac{U_{max}}{U_{min}} = \frac{U^+ + U^-}{U^+ - U^-} \quad \text{Ekvation 106}$$

Förhållandet mellan S och reflektionskoefficienten ρ kan då visas vara

$$S = \frac{1 + |\rho|}{1 - |\rho|} \quad \text{Ekvation 107}$$

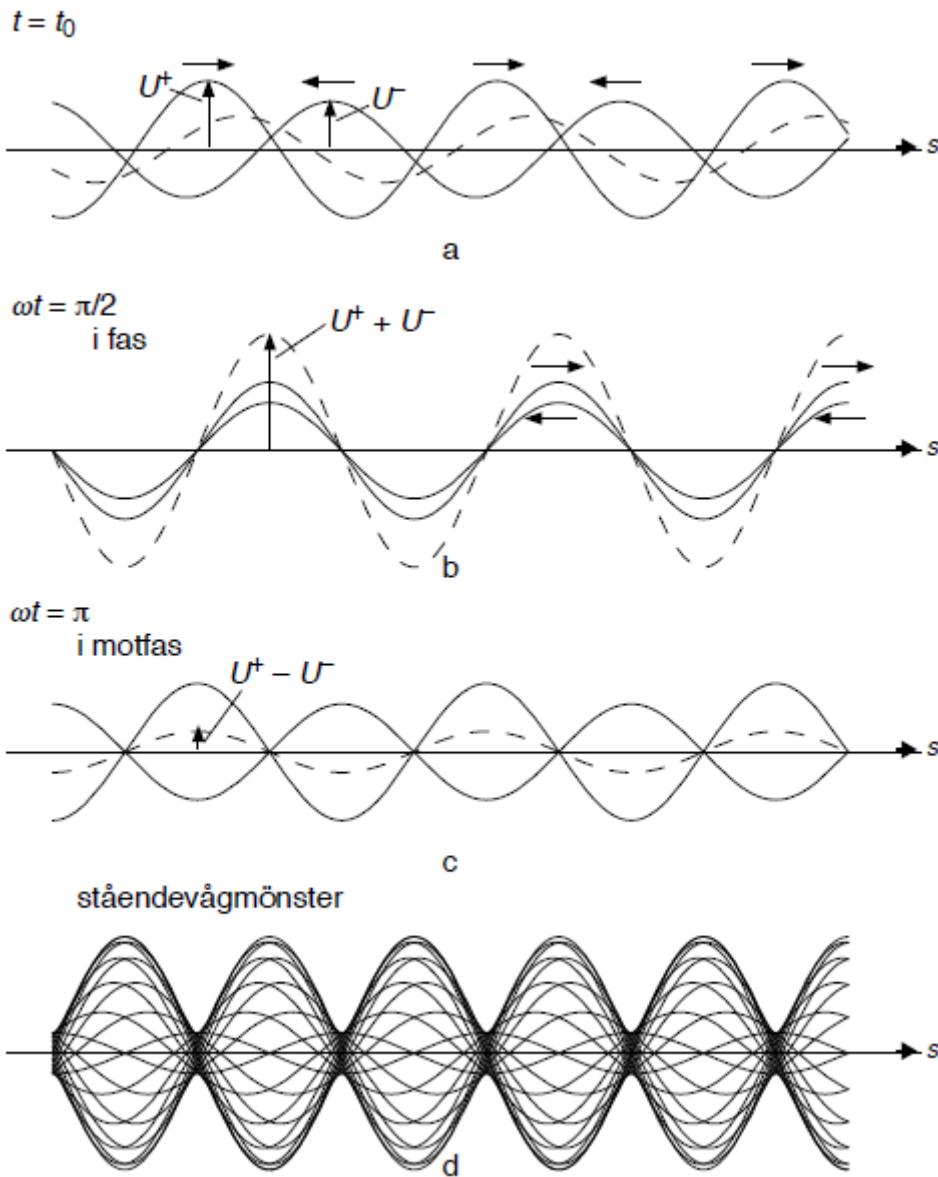
Eller

$$|\rho| = \frac{S - 1}{S + 1} \quad \text{Ekvation 108}$$

Värt att notera är att där bukar och noder uppträder i transmissionsledningen är lasten Z_{in} alltid rent reell. Detta beror på att spänningen i buk respektive nod representeras av $U^+ + U^-$ och $U^+ - U^-$ och motsvarande för strömvågen. Detta innebär att vi kan identifiera impedansens största respektive minsta värde där spänningsvågen har buk respektive nod eftersom

$$Z_{max} = \frac{U_{max}}{I_{min}} = \frac{U^+ + U^-}{\frac{1}{Z_0}(U^+ - U^-)} = Z_0 S \quad \text{Ekvation 109}$$

$$Z_{min} = \frac{U_{min}}{I_{max}} = \frac{U^+ - U^-}{\frac{1}{Z_0}(U^+ + U^-)} = \frac{Z_0}{S} \quad \text{Ekvation 110}$$



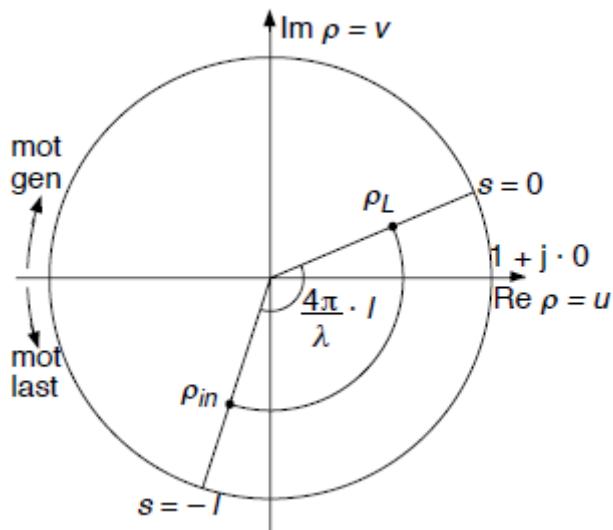
Figur 36

Figur som visar hur ett stående våg-mönster bildas längs en transmissionsledning. Två vågor med spänningssamplituderna U^+ (framåtriktningen) samt U^- (bakåtriktningen) rör sig längs ledningen i a. Diagrammet i b visar hur de vid en viss tidpunkt ligger i fas och adderas konstruktivt (streckad linje). Där summan av vågorna når sitt max kommer en spänningsbuk $U^+ + U^-$ bildas. Diagram c visar hur vågen ser ut när framåtgående och bakåtgående våg i stället ligger i motfas. Då bildas i stället spänningssnoder $U^+ - U^-$ i punkter som ligger mellan spänningssrukarna.

Figur från [26], s. 361.

4.9. Smithdiagrammet

Smithdiagrammet är ett praktiskt, grafiskt hjälpmedel för att analysera och lösa transmissionsledningsproblem. Det utgår från reflektionsfaktorns komplexa talplan enligt Figur 37.



Figur 37

Komplexa talplanet för reflektionsfaktorn ρ . När man rör sig längs en ideal, förlustfri transmissionsledningen kommer reflektionsfaktorn variera med olika faslägen, men storleken kommer vara densamma. Rörelse mot generatoren motsvarar att man roterar medurs i diagrammet och rörelse mot lasten motsvarar en moturs rotation.

Figur från [26], s. 364.

Varje tänkbar reflektionsfaktor i det komplexa talplanet i Figur 37 motsvarar en last Z_{in} enligt Ekvation 102. Detta samband är dock ganska arbetsamt att lösa och Smithdiagrammet löser det genom att lägga in kurvor för impedansen Z_{in} i samma komplexa talplan som Figur 37 visar.

För att få Smithdiagrammet så generellt som möjligt normaliseras lasten Z_{in} med ledningens karakteristiska impedans Z_0 , dvs. man beräknar $z = Z_{in}/Z_0$. Detta blir en normalerad komplex last som kan skrivas som $z = r + jx$. Detta komplexa tal hänger samman med reflektionsfaktorn enligt

$$z = r + jx = \frac{1 + \rho}{1 - \rho} \quad \text{Ekvation 111}$$

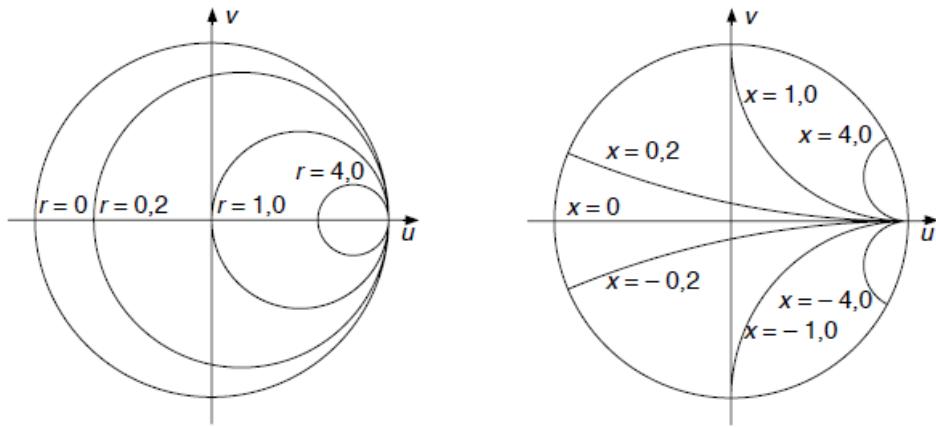
Realdelen r och imaginärdelen x för z i Ekvation 111 kan beräknas om vi delar upp även reflektionsfaktorn i sin real- och imaginärdel, $\rho = u + jv$.

Om sambandet $\rho = u + jv$ substitueras in i Ekvation 111 kan r och x identifieras

$$r = \frac{1 - (u^2 + v^2)}{(1 - u)^2 + v^2} \quad \text{Ekvation 112}$$

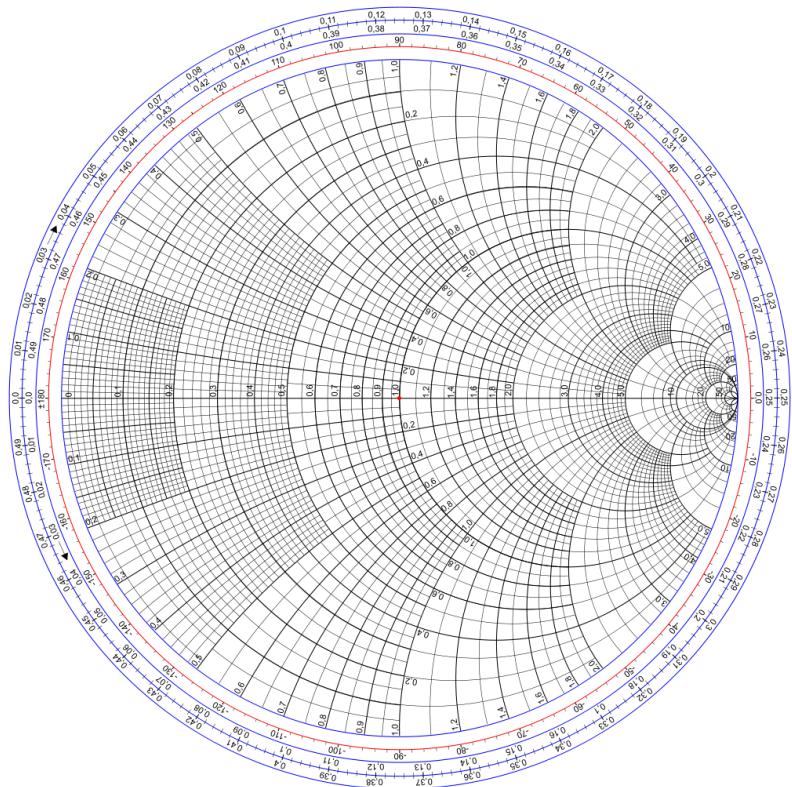
$$x = \frac{2v}{(1 - u)^2 + v^2} \quad \text{Ekvation 113}$$

Dessa ekvationer bildar i det komplexa talplanet olika cirklar för olika värden r och x , se Figur 38.



Figur 38 Lösningen till ekvationen $r + jx = \frac{1+\rho}{1-\rho}$ där $\rho = u + jv$. Konstant r respektive konstant x ger cirklar i det komplexa talplanet för ρ . Eftersom $|\rho| \leq 1$ så är endast de delar av cirklarna som uppfyller detta villkor utritade. Vänster diagram visar cirklarna för några värden på r och höger diagram visar motsvarande för x .
Figur från [26], s. 366.

De olika cirklarna för r och x i Ekvation 112 och Ekvation 113 kan ritas in i ett och samma diagram för att visa reflektionsfaktorn och impedansen i ett och samma diagram. Det är detta som benämns Smithdiagrammet, se Figur 39.



Figur 39 Smithdiagrammet.
By Wdwd - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=11539005>

4.10. Impedansanpassning

4.10.1. Kvartsvågstransformatorn

Om lasten på en transmissionsledning är rent resistiv kan en impedansanpassning ske med hjälp av en annan transmissionsledning som är en kvarts våglängd lång. Detta beror på att just en kvarts våglängd fungerar som en impedanstransformator med $\rho_{in} = -\rho_L$ enligt Ekvation 105.

För att anpassning ska ske så ska kvartsvågstransformatorns karakteristiska impedans uppfylla villkoret

$$Z_{\lambda/4} = \sqrt{Z_0 \cdot Z_L} \quad \text{Ekvation 114}$$

där Z_L är den reella lasten som ska anpassas och Z_0 den karakteristiska impedansen för transmissionsledningen som är anslutet till generatoren.

4.10.2. Stubanpassning

Skrives senare...

4.11. Egenskaper hos verkliga transmissionsledningar

Skrives senare...

5. Elektromagnetiska vågor och vågutbredning

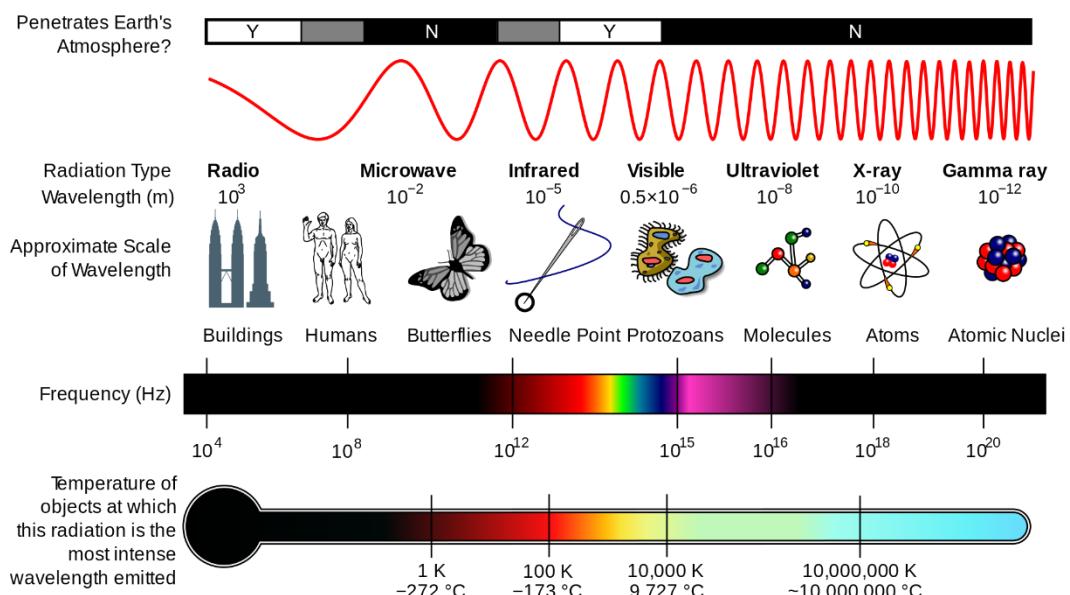
I kapitel 4 så studerades transmissionsledningar och hur vågor utbreder sig längs en ledning. Vad en våg är i detta sammanhang var lite diffust. Det som egentligen avsågs var elektromagnetiska vågor, som har den egenskapen att de kan utbreda sig och röra sig längs en ledning (t.ex. som en telegrafsignal) eller som en våg i rummet, fri från bindning.

I detta kapitel så beskrivs översiktligt egenskaperna hos en elektromagnetisk våg och det elektromagnetiska spektrumet. Elektromagnetiska vågor är kärnan i alla kommunikationssystem idag även de fall där vi inte tänker på det som t.ex. en databuss. Hur de utbreder sig i olika kanaler är viktigt att känna till för att kunna dimensionera de olika delsystemen i ett kommunikationssystem.

5.1. Elektromagnetiska spektrumets indelning [37]

Det elektromagnetiska spektrumet omfattar hela frekvensområdet för elektromagnetisk strålning med respektive, associerade fotonvåglängder. Se även Figur 40.

Det elektromagnetiska spektrumet sträcker sig från de nedre låga frekvenserna som används för modern radiokommunikation till gammastrålning i den högfrekventa änden (med korta våglängder), vilket därigenom täcker våglängder från tusentals kilometer ned till en bråkdel av storleken på en atom. Synligt ljus ligger mot den kortare änden, med våglängder från cirka 400 till 700 nanometer.



Figur 40

Diagram över det elektromagnetiska spektrumets indelning.

Av Inductiveload, NASA - self-made, information by NASABased off of File:EM Spectrum3-new.jpg från NASAThe butterfly icon is from the P icon set, File:P biology.svgThe humans are from the Pioneer plaque, File:Human.svgThe buildings are the Petronas towers and the



*Empire State Buildings, both from File:Skyscrapercompare.svg, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=2974242>*

Tabell 2 visar en sammanställning av spektrumets olika frekvensområde. Radiofrekvenser och mikrovågor (30 – 300 GHz benämns ofta mm-våg) är idag dominanta inom kommunikationssystem tillsammans med det synliga ljuset som används inom fiberoptisk kommunikation.

Ekvation 115 ger sambandet mellan frekvens f , våglängd λ och utbredningshastighet v .

$$\lambda = \frac{v}{f} \quad \text{Ekvation 115}$$

Vid vågutbredning i luft approximeras $v \approx c = 299792458$ m/s.

Tabell 2 Elektromagnetiska spektrumets indelning i olika frekvensområde [37].

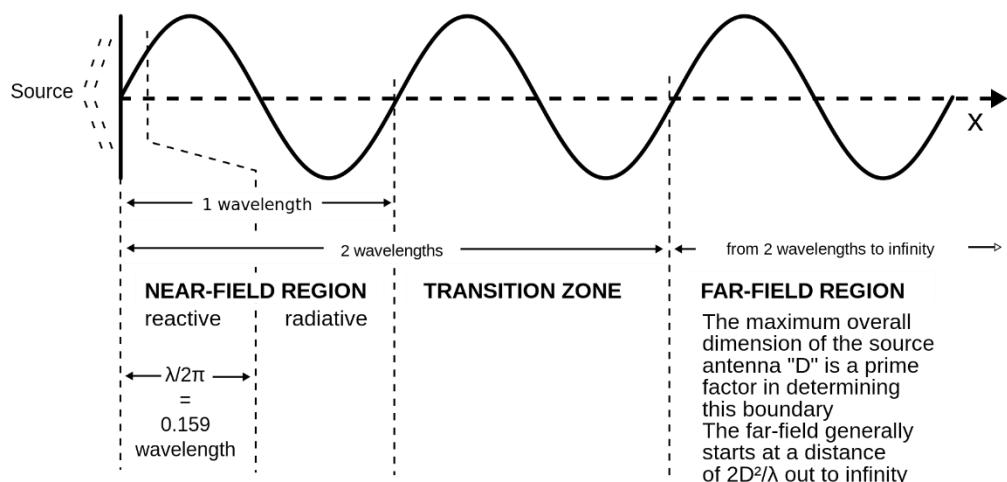
Frekvensområde	Undertyp	Frekvens	Våglängd	Energi per foton
Audiofrekvens	ELF	30–3 Hz	10 Mm – 100 Mm	
	SLF	300–30 Hz	1 Mm – 10 Mm	
	ULF	3–0,3 kHz	100 km – 1 Mm	
	VLF	30–3 kHz	10–100 km	
Radiofrekvens	LF	300–30 kHz	1–10 km	
	MF	3–0,3 MHz	100 m – 1 km	
	HF	30–3 MHz	10–100 m	
	VHF	300–30 MHz	1–10 m	
Mikrovågor	UHF	3 GHz – 300 MHz	1 dm – 1 m	
	SHF	30–3 GHz	1 cm – 1 dm	
	EHF	300–30 GHz	1 mm – 1 cm	
Infraröd (IR)	FIR	6 THz – 300 GHz	0,05–1 mm	
	MIR	100–6 THz	3–50 µm	0,4133–0,0248 eV
	NIR	405,1–100 THz	0,74–3 µm	1,9997–0,4133 eV
Synligt ljus	Röd	483,5–405,1 THz	620–740 nm ^{[3][4]}	1,9997–1,6755 eV
	Orange	508,1–483,5 THz	590–620 nm	2,1014–1,9997 eV
	Gul	525,9–508,1 THz	570–590 nm	2,1752–2,1014 eV
	Grön	576,5–525,9 THz	520–570 nm	2,3843–2,1752 eV
	Cyan	611,8–576,5 THz	490–520 nm	2,5303–2,3843 eV
	Blå	666,2–611,8 THz	450–490 nm	2,7552–2,5303 eV
	Violett	788,9–666,2 THz	380–450 nm	3,2627–2,7552 eV
Ultraviolet (UV)	UVA	951,7–788,9 THz	315–380 nm	3,9360–3,2627 eV
	UVB	1,0 PHz – 951,7 THz	280–315 nm	4,4280–3,9360 eV
	UVC	2,5–1,0 PHz	121–280 nm	10,247–4,4280 eV
	XUV	30–2,5 PHz	10–121 nm	124–10,247 eV
Röntgenstrålning (X)		30 EHz – 30 PHz	10 pm – 10 nm	124 keV - 124 eV
Gammastrålning (γ)		> 30 EHz	< 10 pm	> 124 keV

5.2. Elektromagnetiska vågors egenskaper

5.2.1. Elektromagnetisk vågutbredning i fri rymd

Teorin för elektromagnetiska vågor knyts samman av Maxwells ekvationer [38]. Även om det är James Clerk Maxwell som har fått ge namn till ekvationerna är de resultatet av en mängd olika ekvationer och samband från den intensiva forskning inom elektromagnetism som bedrevs under 1800-talet.

Om man löser Maxwells ekvationer för en ideal punktkälla som transmitterar en elektromagnetisk signal i fri rymd får man tre olika fältkomponenter vars fältstyrka avtar med faktorerna $1/r$, $1/r^2$ respektive $1/r^3$. De två senare faktorerna dämpas alltså ut snabbt då avståndet r till punktkällan ökar. En bit ifrån punktkällan är alltså bara den första av dessa tre komponenter som är möjlig att detektera. Fältet utbreder sig här som en klotformad vågfront (sfärisk vågfront) och klingar långsamt av som $1/r$.



Figur 41

Närfält och fjärrfält för en källa som utsänder elektromagnetisk strålning. Fältet består av tre olika komponenter som klingar av med olika hastighet då man fjärrmar sig från källan. Dessa komponenter har olika egenskaper som ger olika egenskaper på det elektromagnetiska fältet i olika zoner kring källan. Därför definieras olika zoner enligt bilden.
By OSHA, Dept of Labor -
http://www.osha.gov/SLTC/radiation/electromagnetic_fieldmemo/electromagnetic.html, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=20417753>

Figur 41 visar hur man delar upp området utanför en källa i olika zoner eller regioner. I närfältet är alla tre komponenter närvarande, medan det i fjärrfältet endast återstår komponenten som klingar av med faktorn $1/r$. De två andra komponenterna har dämpats ut så pass mycket att de är försumbara.

Närfältet består i sin tur av två regioner. En som ligger mycket nära källan där fältet har en stark reaktiv komponent. Att fältet är reaktivt betyder att det är en del av fältet där

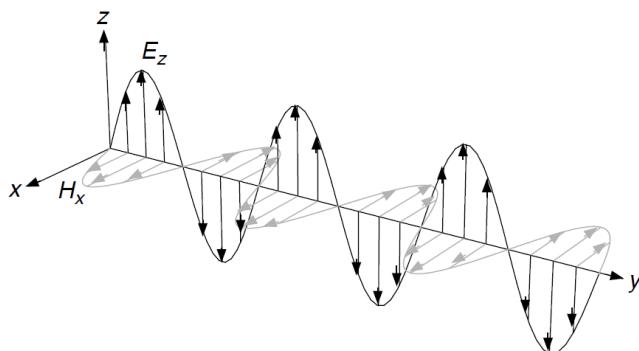
den elektromagnetiska energin pendlar mellan källan och den reaktiva regionen i källans närhet. Denna del av fältet är alltså icke-strålande eftersom denna energi inte fortsätter ut i rummet. Sen definieras även en strålande region i närfältet, där det reaktiva fältet dämpats ut och den strålande delen består av de två komponenterna som klingar av som $1/r$ och $1/r^2$. Längre ut från källan domineras fältet av $1/r$ -komponenten och då befinner man sig i fjärrfältet.

Det är självklart så att det inte är knivskarpa gränser mellan regionerna, men gränserna som definieras i Figur 41 ger en indikation på hur nära man kan vara en källa (antenn) innan man får effekter från närfältet. Objekt som placeras i närfältet har nämligen en tydlig påverkan på källan eftersom det blir en märkbar interaktion mellan källan och objektet. Detta är extra påtagligt i det reaktiva närfältet. Men befinner man sig en bit in i fjärrfältet är denna interaktion försumbar. Normalt dimensionerar och konstruerar man kommunikationssystem så att antennerna befinner sig i fjärrfältet i förhållande till varandra.

5.2.2. E- och H-fält för plan våg

Som kapitel 5.2.1 visade så är vågutbredningen för en elektromagnetisk våg i fjärrfältet sfärisk. På stort avstånd från källan är krökningen på den sfäriska vågfronten inte särskilt stor och den kan approximeras som en plan vågfront inom ett begränsat område. Dessutom är variationen i fältstyrkan inte särskilt stor om man förflyttar sig mot eller från källan. Eftersom avståndet r redan är stort har en liten förändring av r försumbar påverkan på dämpningsfaktorn $1/r$. Man kan med god approximation då anse att man har ett plant elektromagnetiskt fält.

Maxwells ekvationer får i här en relativt enkel lösning. Fältet som utbreder sig består av två ortogonala komponenter, det elektriska E-fältet och det magnetiska H-fältet. Dessa båda fältkomponenter är dessutom ortogonala mot utbredningsriktningen och bildar en s.k. TEM-våg (Transversell Elektromagnetisk våg). Figur 42 visar en ögonblicksbild av en TEM-våg som utbreder sig längs y-axeln. E-fältet (enhet V/m) och H-fältet (enhet A/m) är hela tiden ortogonala mot varandra och mot utbredningsriktningen. När utbredningen är i fri rymd så är de dessutom i fas med varandra.



Figur 42 TEM-våg som utbreder sig längs y-axeln.
Figur från [26], s. 406.

De två fältkomponenternas styrka varierar som funktion av både tid och rumskoordinat. Uttrycken för en linjärpolariserad TEM-vågs E- respektive H-fält så som den ser ut i Figur 42 kan i komplex form skrivas som

$$E_z(y, t) = \hat{E} \cdot e^{-j\omega(t-y/v)} \quad \text{Ekvation 116}$$

$$H_x(y, t) = \hat{H} \cdot e^{-j\omega(t-y/v)} \quad \text{Ekvation 117}$$

där v är vågens utbredningshastighet. Denna hastighet beror av vilket medium som vågen utbreder sig i. Det finns några specialfall som är intressanta att analysera, vilket sker i kapitel 5.3. I fri rymd, vilket egentligen är vakuum men ofta inkluderar utbredning i luft, är

$$v = c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \quad \text{Ekvation 118}$$

De två naturkonstanterna ϵ_0 och μ_0 kallas för permittivitet respektive permeabilitet för fri rymd.

Ekvation 116 och Ekvation 117 har direkta likheter med hur spänning och ström varierar längs en ideal förlustfri transmissionsledning. Likheten är inte en slump, även i fallet transmissionsledning finns en vågutbredning med E- och H-fält även om man för transmissionsledningar av praktiska skäl gör analysen baserad på spänning, ström och ekvivalenta kretskomponenter. Precis som för fallet med transmissionsledningar finns det en impedans som vågen upplever då den propagerar genom rummet. Denna impedans benämns vågimpedans [39] och dess värde ges av

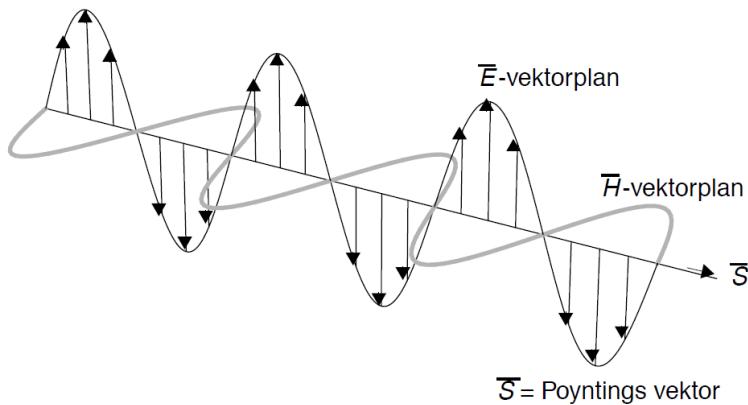
$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = \sqrt{\frac{4\pi \cdot 10^{-7}}{8,854\,187\,817 \cdot 10^{-12}}} \approx 377 \Omega \quad \text{Ekvation 119}$$

Det kan kännas kontraintuitivt att det finns en impedans skapad av "ingenting" (vakuum), men man kan se det som en faktor som kopplar samman E- och H-fältens inbördes storlekar.

5.2.3. Polarisation

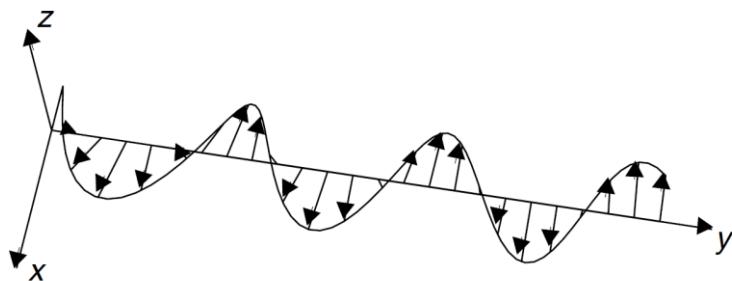
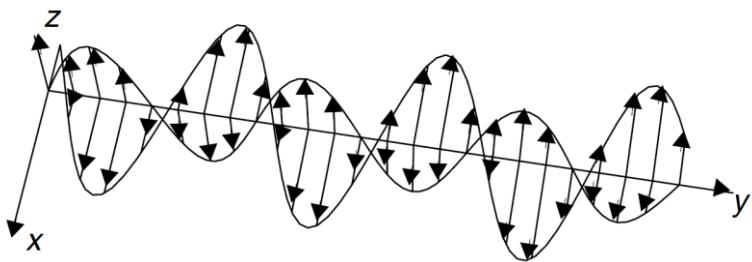
Beroende på hur fältkomponenterna är orienterade kommer fältet som utbreder sig ha olika polarisation [40]. Man talar om olika polarisationstillstånd, linjär, cirkulär och elliptisk polarisation. Den fältkomponent som anger vilken polarisation man talar om är definitionsmässigt E-fältet.

Linjär polarisation är när svänger längs en tänkt linje som är vinkelrät mot utbredningsriktningen. Linjen kan förstas vara orienterad i olika vinklar. Figur 43 visar en linjär, vertikalpolariserad våg. Om E-fältet i stället svänger i horisontalriktningen (dvs. E- och H-fältets svängningsriktning byter plats i Figur 43) så talar man om linjär, horisontell polarisation. E-fältet kan förstas ha en svängningsriktning som går i vilken vinkel som helst i förhållande till koordinatsystemets referenser. I mobiltelefonisystemen används t.ex. oftast $\pm 45^\circ$ för basstationens polarisation för den utsända signalen.



Figur 43 TEM-våg som utbreder sig med linjär, vertikal polarisation. Polarisationens riktning bestäms av E -fältets riktning.
Figur från [26], s. 436.

Cirkulär polarisation är när polarisationen roterar under vågens utbredning. Den skapas genom att ha två källor som sänder med olika linjär polarisation, inbördes ortogonala. Genom att mata källorna med samma signal där den ena är fasförskjuten en kvarts period kommer en ny polarisation skapas som roterar i en cirkulär form som en korkskruv, se Figur 47. En animering av hur detta sker finns i följande referens (Youtubefilm) [41].



Figur 44 TEM-våg som utbreder sig med cirkulär polarisation.
Figur från [26], s. 437.

Beroende på åt vilket håll rotationen sker benämns polarisation som högercirkulär eller vänstercirkulär polarisation (eng. Right Hand Circular Polarization RHCP eller Left Hand Circular Polarization LHCP).

Om de två fasförskjutna komponenterna dessutom justeras så att de får olika amplitud, kommer den roterande polarisationen inte följa en perfekt cirkel. Resultatet blir då en elliptisk polarisation.

5.2.4. Strålningstäthet och Poyntingvektor

En elektromagnetisk våg överför en effekt genom dess fältstyrkor i E- och H-fälten. Effekten överförs i vågens utbredningsriktning. Eftersom den är fördelad över en yta talar man om effekttäthet med enheten W/m^2 . Eftersom effekttätheten har en utbredningsriktning är den en vektorstorhet (ofta med beteckningen S) och kallas Poyntingvektor [42]. Storleken på effekttätheten ges av det enkla sambandet i Ekvation 120 nedan.

$$S = E_e H_e = \frac{E_e^2}{Z_0} = H_e^2 Z_0 \quad \text{Ekvation 120}$$

E_e och H_e är fältens effektivvärde. Vid sinusformade fält (som i Figur 42) är effektivvärdena $\hat{E}/\sqrt{2}$ respektive $\hat{H}/\sqrt{2}$. Notera likheterna med spänning, ström och effekt från kretsteorin!

5.3. Elektromagnetisk vågutbredning i material

5.3.1. Vågutbredning i dielektriska material

Dielektriska material [43] är material som fungerar som elektriskt isolerande material. Typiska dielektriska material är många typer av polymerer (plast), elastomerer (elastiska material av gummi-karaktär) glas, keramer, vissa vätskor och en del organiska material. P.g.a. den isolerande egenskapen är dielektriska material mycket vanligt förekommande som material i diverse komponenter. Mönsterkort, koaxialkablar och kondensatorer är sådana exempel där elektriska ledare skiljs åt av ett dielektriskt material.

För isotropa material (samma egenskaper i alla riktningar) anges en materialegenskap som benämns relativ permittivitet ϵ_r [44]. I det generella fallet är detta en frekvensberoende materialegenskap som normeras till permittiviteten för fria rymden enligt

$$\epsilon_r(\omega) = \frac{\epsilon(\omega)}{\epsilon_0} \quad \text{Ekvation 121}$$

där $\epsilon(\omega)$ är materialets permittivitet. ϵ_r är ofta den parameter som anges i datablad och tabellverk.

I sin generella form är ϵ_r ett komplext tal som skrivs

$$\epsilon_r(\omega) = \epsilon'_r(\omega) + j\epsilon''_r(\omega) \quad \text{Ekvation 122}$$

Om materialet är idealt och förlustfritt är $\epsilon_r = \epsilon'_r$ och Ekvation 118 och Ekvation 119 justeras med materialets egenskaper enligt

$$v = \frac{1}{\sqrt{\epsilon_r \epsilon_0 \mu_0}} = \frac{c}{\sqrt{\epsilon_r}}$$

Ekvation 123

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_r \epsilon_0}} \approx \frac{377}{\sqrt{\epsilon_r}} \Omega$$

Ekvation 124

Inom optiken används ofta begreppet brytningsindex [45], som är ett mått på utbredningshastigheten för en våg i ett visst material i förhållande till ljusets hastighet i vakuum. Med Ekvation 123 så kommer brytningsindex i ett dielektriskt material vara

$$n = \frac{c}{v} = \sqrt{\epsilon_r}$$

Ekvation 125

5.3.2. Vågutbredning i metaller. Skineffekten.

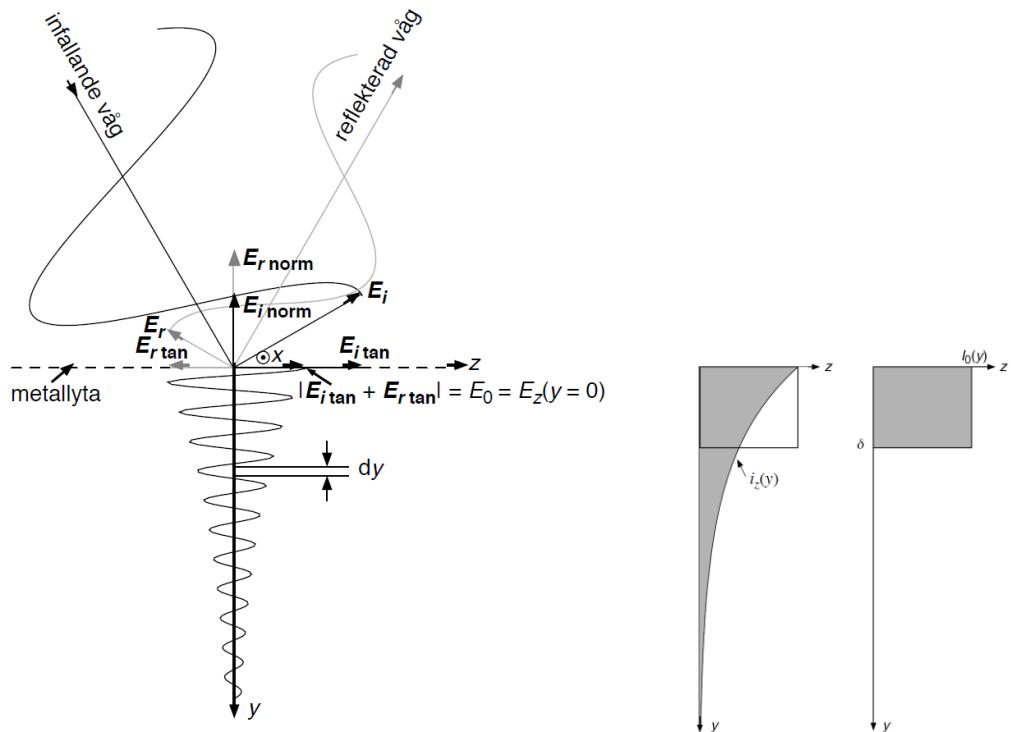
I en ideal elektrisk ledare som en metall med oändligt god ledningsförmåga (konduktivitet $\sigma = \infty$) kommer ett elektromagnetiskt fält vara noll inne i ledaren. Detta beror på att en yttre fält som ligger utanför ledarens ytskikt kommer generera en ytström på ledaren. Denna ytström genererar i sin tur ett sekundärt elektrisk längs ytskiktet som är lika stort men motriktat. Eftersom konduktiviteten är oändligt god kan strömtätheten i ytan bli oändligt stor och ingen ström "behöver" genereras längre in i materialet för att skapa det sekundära fältet. Summan av E-fälten på ytan blir då noll och inne i ledaren så existerar inget fält.

Men metaller har en ändlig konduktivitet, även om den är god (undantaget supraleddare). Därmed blir det en viss inträngning av fältet i metallen. Mest ström kommer att flyta vid ytan och sedan avklingar strömtätheten exponentiellt i takt med att avståndet från metallens ytskikt ökar, se Figur 45. Effekten kallas för skinneffekt [46]. Strömtätheten avtar alltså med en faktor e^{-y} där y är koordinaten som går in i metallen enligt Figur 45. Ett mått på inträngningen är det s.k. skinndjupet δ som är det djup där strömtätheten klingat av med faktorn $e^{-y} \approx 0,368$. Skinndjupet ges av formeln

$$\delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma}}$$

Ekvation 126

Skinndjupet är alltså beroende av frekvensen på den elektromagnetiska vågen. Vid låga frekvenser är skinndjupet stort för de flesta metaller, men för frekvenser som är typiska för dagens kommunikationssystem (GHz-området) är skinndjupet i storleksordningen μm för en god ledare. Det har viss praktisk betydelse då den koncentrationen av strömtäthet som sker vid ytan i praktiken ökar resistansen i en ledare. Stor del av ledarens tvärsnittsarea bär nämligen ingen ström, så i praktiken blir exempelvis den strömförande delen av en ledare på ett mörstorkort mindre.



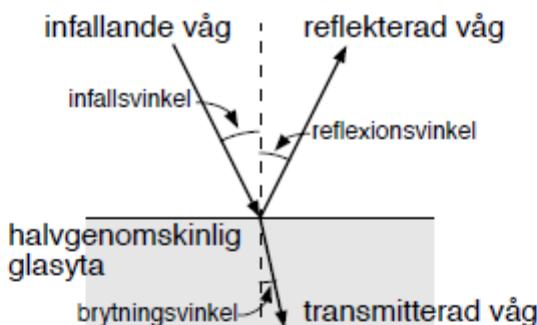
Figur 45

Fältinträngning i en icke-ideal metall (vänster bild), samt strömtäthet som funktion av inträngning (höger bild). Fältet avklingar exponentiellt med inträngningsdjupet vilket avspeglar sig i hur strömmen i ledaren fördelar sig. Skindjupet δ (eller inträngningsdjupet) definieras som det djup där strömtätheten avklingat med faktorn e^{-1} jämfört med maximala strömtätheten på ytan.

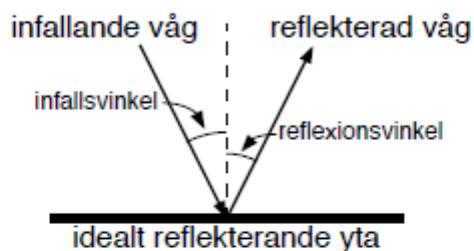
Figur från [26], s. 411 och 414.

5.4. Plana elektromagnetiska vågors reflektion mot plana ytor

En plan elektromagnetisk våg som infaller mot ett förlustfritt dielektrikums plana begränsningsyta beter sig på samma sätt som en plan ljusvåg som infaller mot en plan halvgenomskinlig glasyta, d.v.s. en del av energin reflekteras (reflexionsvinkeln = infallsvinkeln) medan resten transmitteras och bryts mot normalen, se Figur 12.3.



Figur 12.3 Reflexion och transmission vid halvgenomskinlig glasyta.



Figur 12.4 Ideal spegelreflexion.

Om dielektriket dessutom innehåller en resistiv komponent omvandlas en del av den transmitterade vågen till en ström, som ger upphov till värme.

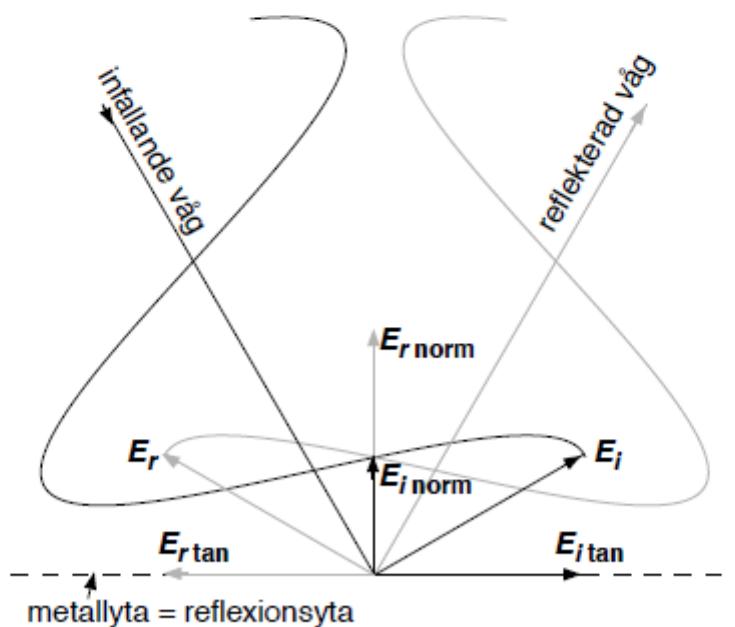
Vid ideal spegelreflexion är den transmitterade delen noll varför den reflekterade vågen innehåller lika mycket energi som den infallande, se Figur 12.4. I verkligheten finns inga sådana idealt reflekterande ytor utan ytan innehåller en liten resistiv del men ingen reaktiv (ingen dielektrikum). Energin i den reflekterade vågen är därför alltid något mindre än i den infallande vid icke ideal spegelreflexion, skillnaden utgörs av värmeförluster i reflexionsytan.

En fysikalisk bild av spegelreflexion är att den infallande vågens kraft accelererar elektronerna i reflexionsskiktet och eftersom accelererade elektroner ger upphov till elektromagnetiska vågor uppstår den reflekterade vågen.

För förståelsen av elektromagnetiska vågors transmission genom vågledare är analogin med ideal och icke ideal spegelreflexion mest

intressant. Den reflekterande ytan utgörs i vågledare av någon metall. Maxwells ekvationer anger villkoren för E och H -fälten i reflexions-(metall-)ytan. För att renodla förhållandena studerar vi först vad som händer med E -vektorn efter reflexion av en våg som infaller mot en reflekterande yta. Den infallande och den reflekterade vågens E -vektorer vid reflexionsytan uppdelas i, i förhållande till reflexionsytan, tangential- och normalkomposanter, se Figur 12.5. Det gäller då för en ideal metallyta (konduktiviteten $\sigma = \infty$) att tangentialkomposanterna är lika stora och motriktade, d.v.s. $|E_{i\ tan} + E_{r\ tan}| = 0$ [V/m] medan $\epsilon_0 |E_{i\ norm} + E_{r\ norm}| =$ laddningstätheten i ytan, som mäts i Coulomb/m² [C/m²].

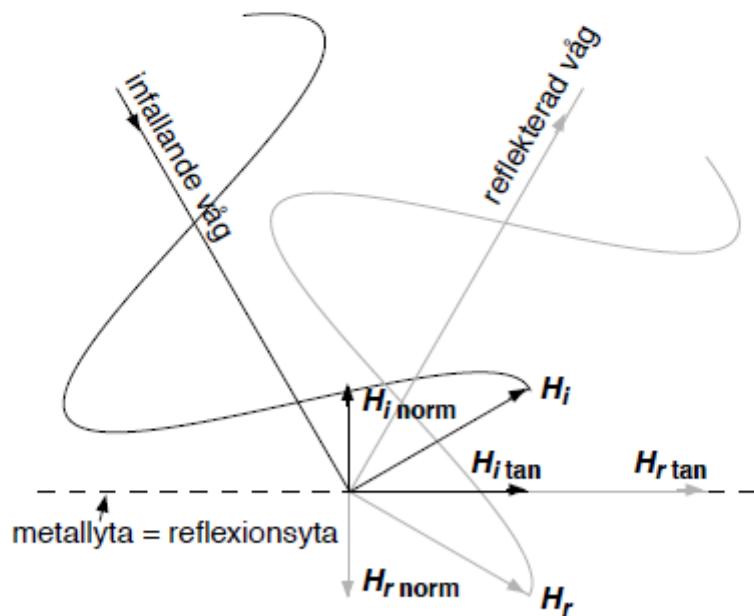
Härnäst studerar vi vad som händer med H -vektorn efter reflexion av en våg som infaller mot en reflekterande yta.



Figur 12.5 Ögonblicksbild av E -vektorer vid reflexion mot ideal ($\sigma = \infty$) metallyta. Svart representerar infallande våg medan grått representerar reflekterad våg.

Den infallande och den reflekterade vågens H -vektorer vid reflexionsytan uppdelas i, i förhållande till reflexionsytan, tangential- och normalkomposanter, se Figur 12.6. Det gäller då för en ideal metallytta (konduktiviteten $\sigma = \infty$) att normalkomposanterna av $B_{i\text{ norm}}$ och $B_{r\text{ norm}}$ är lika stora och motriktade, och resulterande $|B_{\text{norm}}| = \mu_0 |H_{i\text{ norm}} + H_{r\text{ norm}}| = |B_{i\text{ norm}} + B_{r\text{ norm}}|$ är 0 Tesla [T] medan $|H_{i\text{ tan}} + H_{r\text{ tan}}|$ = strömtätheten i ytan, som mäts i Ampère/m² [A/m²].

Dessa av Maxwell formulerade villkor gällande vid gränsen mellan två medier kallas randvillkor och används för att beräkna bl.a. vågmönster i vågledare, resonatorer och på antenner.



Figur 12.6 Ögonblicksbild av H -vektorer vid reflexion mot ideal ($\sigma = \infty$) metallyta. Svart representerar infallande våg medan grått representerar reflekterad våg.

5.5. Vågutbredning i olika typer av kanaler

5.5.1. Friis transmissionsekvation och vågutbredning i fri rymd

En central ekvation när det gäller vågors utbredning mellan sändare och mottagare är Friis transmissionsekvation [47]. Den är jämsides med Friis brusformel i kapitel 3.6.2 Harald Friis stora bidrag till teorin bakom kommunikationssystem. Friis transmissionsekvation beskriver mottagen effekt i en mottagarantenn, givet vilket avstånd man har till sändaren och vilken effekt man sänder med. Förutsättningarna för att Friis ekvation ska gälla är att kommunikationssystemet befinner sig i fri rymd, dvs. att det inte finns några reflekterande objekt i kanalen, samt att avståndet är tillräckligt stort så att antennerna är i fjärrfältet (6.1.1). Den första förutsättningen är förstås en begränsning, men för vissa system så är dessa uppfyllda. T.ex. så fungerar Friis bra för satellitkommunikation satellit-satellit eller satellit-markstation. Men i många fall så ger Friis en bra approximation och den används ofta för överslagsberäkningar.

Eftersom fältstyrkan och dess amplitud avtar med en faktor $1/r$ med avståndet till källan till den utsända elektromagnetiska strålningen, kommer effekten avta med faktorn $1/r^2$. Ökat avstånd mellan sändare och mottagare kommer ofrånkomligen ge en minskad mottagen effekt. För att motverka detta kan antenner med god riktverkan väljas. Se mer i kapitel 6, där antenners egenskap att fokusera elektromagnetisk energi går igenom. Ett mått på denna fokusering är antennvinst G . Det Friis visade är att den mottagna effekten P_m i ett system i fri rymd kan beräknas som

$$P_m = P_s G_s G_m \cdot \left(\frac{\lambda}{4\pi r} \right)^2 = P_s G_s G_m \cdot \left(\frac{c}{4\pi f r} \right)^2 \quad \text{Ekvation 127}$$

där P_s är sändareffekten, G_s och G_m är antennvinst för sändar- respektive mottagarantenn och r är avståndet mellan antennerna. Ekvation 130 ska tolkas med viss försiktighet. En vanlig missuppfattning är att utsänd effekt klingar av snabbare med ökad frekvens eftersom termen $(c/4\pi f r)^2$ ingår. Men det är en effekt av hur ekvationen är härledd, antennvinsterna G_s och G_m ökar med frekvens i motsvarande grad enligt Friis modeller. Effekten i fri rymd-utbredning avklingar som $1/r^2$ oberoende av frekvens, något annat skulle vara ofysikaliskt.

I samband med Friis och vågutbredning i fri rymd talar man ofta om begreppet EIRP. EIRP står för Effective Isotropic Radiated Power och beräknas som

$$\text{EIRP} = P_s G_s \quad \text{Ekvation 128}$$

EIRP är ett mått på hur mycket en effekt en isotrop antenn (se kapitel 6.1.4) behöver matas med för att ge samma strålningsintensitet jämfört med $P_s G_s$.

5.5.2. Vågutbredning i reflektiv miljö. Fädning.

Skrives senare.



5.5.3. Vågutbredning i miljöer med multipelreflektion. Kanalmodeller.

Skrives senare.

5.5.4. Länkbudget

Skrives senare.

6. Antenner

Detta kapitel använder material kopierat från [21] (kapitel 13) i enlighet med skolornas kopieringsavtal.

6.1. Grundläggande antennparametrar

6.1.1. Antennens fjärrfält, närfält och fjärrfältsavståndet

I kapitel 5.2 beskrivs hur ett elektromagnetiskt fält från en punktkälla (t.ex. en mycket liten antenn) utbreder sig sfäriskt. Det betyder att fasfronten för den elektromagnetiska vågen är krökt likt formen på ett klot. En betraktare som ser den elektromagnetiska vågen träffar en plan yta (t.ex. en vägg) kommer alltså uppleva att vågen har en fasvariation över den plana ytan. Även vågens amplitud kommer variera beroende på att olika punkter på den plana ytan har olika avstånd till punktkällan som utsänder den elektromagnetiska vågen.

Om man tänker sig att man placerar en mottagarantenn i denna typ av sfäriskt utbredande fält är det lätt att inse att det blir en variation av fas och även amplitud över antennens yta. En analys av denna situation blir relativt komplex, så därför genomför man normalt analys av en antenns egenskaper då den elektromagnetiska vågen är plan. Detta sker, som vi såg i kapitel 5.2, när vi befinner oss på tillräckligt långt avstånd från punktkällan. Man talar då om en antenns fjärrfält, vilket är ett viktigt begrepp inom antennteorin. Villkoret för fjärrfält för en antenn är lite förenklat att man har planvågsutbredning, vilket innebär en plan fasfront. I en antenns närfält är detta villkor inte uppfyllt.

En analogi för att få en geometrisk bild över plan respektive sfärisk fasfront är att man med ögat inte kan se jordens krökning längs horisonten när man står vid stranden och blickar ut över havet. Detta motsvarar en approximativ plan vågfront. Däremot är det enkelt att se den sfäriska krökningen på arenan Globen i Stockholm om man skulle få lov att stå uppe på toppen. Skillnaden mellan dessa två fall är att man befinner sig på väldigt olika avstånd från mittpunkten på de två klotformade objekten.

Ett vanligt mått för att avgöra om man befinner sig i fjärrfältet eller närfältet är det så kallade fjärrfältsavståndet (farfield distance, Rayleigh distance eller Fraunhofer distance). För en antenn vars apertur ("öppning") har storleken d som sitt största mått, ges fjärrfältsavståndet r av

$$r = \frac{2d^2}{\lambda}$$

Ekvation 129

där λ är den elektromagnetiska vågens våglängd. Ekvation 129 återfinns också i Figur 41 som visar gränserna mellan reaktivt närfält, strålande närfält och fjärrfält.

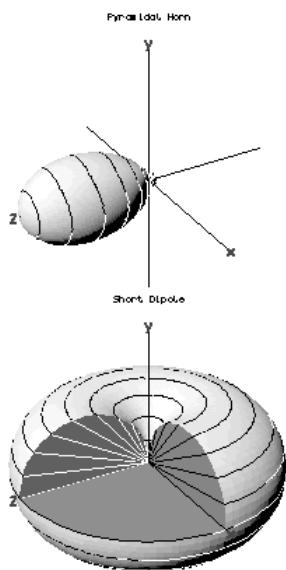
Notera att fjärrfältet är en relativt grov approximation av en plan vågsapproximation. En antenn med storleken d som placeras på avståndet r från en punktformig källa som

sänder kommer ha en fasvariation över aperturen på $22,5^\circ$, vilket i vissa applikationer kan påverka antennens egenskaper märkbart. Men det är viktigt att notera att fjärrfältsavståndet i Ekvation 166 inte på något sätt anger om en antenn fungerar eller ej. Dock så är flertalet antennparametrar som används i beräkningar härledda i fjärrfältet med plan våg. Därför så brukar man normalt kontrollera att den antenn man designat eller valt att använda i sitt system uppfyller fjärrfältsavståndet. Ekvation 129 är därför en god tumregel att ha i minnet och i de flesta fall kommer en antenn som placeras precis på fjärrfältsavståndet fungera som avsett, trots fasfelet över aperturen.

6.1.2. Strålningsdiagram och lober

En antenns uppgift kan sägas bestå av två delar. Den ska förstås omforma elektromagnetisk effekt som propagerar på en transmissionsledning till en elektromagnetisk våg som propagerar fritt genom rummet (eller tvärtom om antennen fungerar som mottagarantenn). Men den ska också fördela den utstrålade effekten så smart som möjligt i rummet, dvs. rikta den elektromagnetiska vågen enligt ett visst designval (motsvarande gäller om antennen är på mottagarsidan, då ska antennen ta upp vågor som kommer i en viss riktning så effektivt som möjligt).

Den senare funktionen med hur antennen riktar den elektromagnetiska strålningen beskrivs med antennens strålningsdiagram (radiation pattern eller antenna pattern). Strålningsdiagrammet visar hur strålningsintensitet (enhet W/m^2), elektrisk fältstyrka (V/m) alternativt elektrisk fältstyrka (A/m) varierar i olika riktningar vid ett fixt avstånd. Vanligast är att man tänker på strålningsintensitet när man talar om antennens strålningsdiagram. Se Figur 46, som visar två exempel på strålningsdiagram.



Figur 46

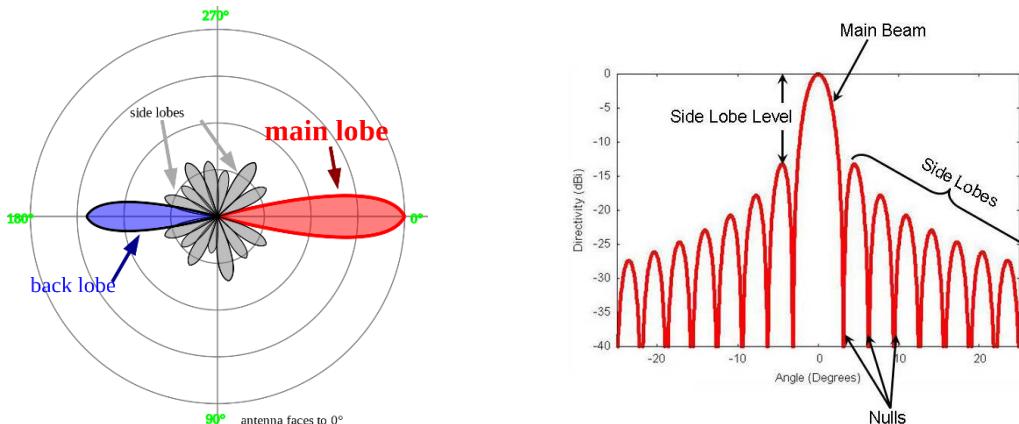
Strålningsdiagram för två olika typer av antenner. Den övre figuren visar en hornantenn, vilket är en antenntyp med god riktverkan som koncentrerar den elektromagnetiska strålningen i en viss riktning. Den undre figuren visar diagrammet (uppskuret för att få bra överblick) för en dipol, som är en enkel antenn med rundstrålande egenskaper.

Bild från By James Włodarczyk (--catslash 23:44, 31 July 2006 (UTC)) - Own work of English

*Wikipedia user Catslash, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=18113015>*

Analys av strålningsdiagram är mest praktiskt att hantera med sfäriska koordinater (se Appendix A – Tredimensionella koordinatsystem). Diagrammet blir då en funktion av elevationsvinkeln θ och azimutvinkeln φ . I det generella fallet så kan man beskriva strålningsdiagrammet som en funktion även av avstånd r , men i princip alltid tänker man sig diagrammet för ett fixt, konstant avstånd.

Ofta redovisas inte det kompletta strålningsdiagrammet, utan man gör snitt genom antenndiagrammet (typiskt genom horisontal- och vertikalplanen) och redovisar tvådimensionella diagram i polär eller rektangulär form. Se Figur 47 som visar två exempel. Skalan på diagrammens axel för strålningsintensitet är ofta någon relativ decibelskala.



Figur 47
Exempel på tvådimensionella antenndiagram som är tvärsnitt av det fulla, tredimensionella diagrammet. Vänstra diagrammet visar ett fullt azimutsnitt 360° (horisontalplanet) uppritat i en polärplot. Det högra diagrammet är ett inzoomat diagram i en rektangulär plot, som visar en mindre del av ett snitt.
*Vänster bild: By Timothy Truckle - Own work, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=4245213>*
*Höger bild: By Mr. PIM at the English Wikipedia, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=20074210>*

Figur 46 och Figur 47 visar också hur den elektromagnetiska strålningen fokuseras i olika så kallade lober (strålningslober i viss litteratur). För att skilja olika typer av lober åt har de olika benämningar:

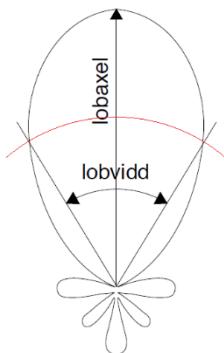
- huvudlob är den lob som innehåller maxima strålningsintensiteten. Denna lob riktas normalt mot den punkt eller det område dit man vill rikta sin kommunikation
- backlob är den lob som ligger i huvudlobens motsatta riktning
- sidlob är övriga lober, dvs. de som ligger vid sidan av huvudlob och backlob
- sidloberna tillsammans med backloben benämns ibland för bilober

På engelska blandas uttrycken beam och lobe, men det korrekta är att huvudloben är en main beam och övriga lober kallas side lobe eller back lobe. Den högra bilden i Figur 47 är alltså den som har mest korrekta engelska benämningar.

6.1.3. Lobvidd

Ett mycket vanligt förekommande mått på en antenns strålningssegenskaper är lobvidd (benämns ofta även lobbredd). Lobvidd är ett vinkelmått som anger ”spetsigheten” på huvudloben. Den mäts mellan de punkter där strålningsintensiteten fallit en viss faktor från huvudlobens maximala strålningsintensitet. Om inget annat anges är denna faktor $\frac{1}{2}$ eller -3 dB jämfört med maxpunktens intensitet.

Begreppet lobaxel är den axel som ger riktningen för antennens maximala strålningsintensitet.



Figur 48

Lobvidd för en antenns huvudlob. Lobvidden är den vinkel som visar var strålningsintensiteten sjunkit med en viss faktor (markerat med rött i figuren) i relation till maximala strålningsintensiteten.

Figur modifierad från [26], s. 435.

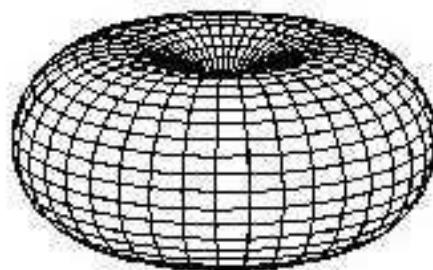
Eftersom man ofta analyserar antenner i två plan (typiskt azimut och elevation) så är det vanligt att huvudloben beskrivs med två lobvidder, en i höjd och en i sida.

6.1.4. Isotropa, rundstrålande och riktade diagram

Den isotropa antennen är en teoretisk antenn som strålar lika mycket i alla riktningar. Detta är en antenn som inte kan konstrueras, men man talar ofta om isotrop antenn eftersom den används som en jämförelseantenn. Mer om detta i kapitel 6.1.5 och 6.1.6.

En riktantenn är en antenn som riktar och fokuserar den elektromagnetiska strålningen i en viss riktning (eller tar emot strålningsintensitet bättre i vissa riktningar om det är en mottagarantenn).

Ett specialfall av riktantenn är rundstrålande antennen som sprider ut strålningen jämnt fördelat i ett plan. Figur 49 visar ett diagram från en typisk rundstrålande antenn.



Figur 49 Rundstrålande diagram för en dipolantenn. Dipolen som skapar detta diagram består av två korta vertikala antennspröt med längden $\lambda/4$. Formen på strålningsdiagrammet påminner om en donut och kallas ibland för donut-diagram.
CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=960123>

6.1.5. Direktivitet

En antenns direktivitet D [48]definieras som kvoten mellan antennens maxima strålningsintensiteten S_{max} (längs huvudlobens lobaxel) och strålningsintensiteten från en isotrop antenn som utstrålar samma totala effekt P_o . Kvoten beräknas förstås på samma tänkta avstånd r från antennen och den tänkta isotropa antennen.

Eftersom strålningsintensiteten för en isotrop antenn kan skrivas som $P_o/4\pi r^2$ (effekten fördelas jämnt över en sfär, och en sfär har ytan $4\pi r^2$) så kan direktiviteten skrivas som

$$D = \frac{S_{max}}{P_o/4\pi r^2} \quad \text{Ekvation 130}$$

Ett annat sätt att skriva Ekvation 130 som förekommer i litteraturen är att använda ett strålningsintensitetsmått U som är effekt per rymdvinkel (W/sr). Då kan Ekvation 130 formuleras som

$$D = \frac{U_{max}}{P_o/4\pi} \quad \text{Ekvation 131}$$

Ofta anges direktivitet med ett decibelmått. Eftersom kvoten är en jämförelse med en isotrop antenn, så används måttet dBi där i står för isotrop. Direktiviteten för alla antenner är större än 1 motsvarande $10 \log 1 = 0$ dBi. En tänkt, isotrop antenn har alltså direktiviteten 0 dBi medan alla fysiskt realiserabara antenner har $D > 0$ dBi.

6.1.6. Antennvinst

Direktivitet som beskrevs i kapitel 6.1.5 är kvoten mellan strålningsintensiteter för utstrålad effekt. Minst lika intressant är att göra motsvarande kvot definierad för inmatad effekt P_i . Det ger ett annat, närbesläktat mått som kallas för antennvinst G (antenna gain), vilket definieras som

$$G = \frac{S_{max}}{P_i/4\pi r^2}$$

Ekvation 132

Skillnaden mellan direktivitet och antennvinst är helt enkelt skillnaden mellan P_i och P_o , vilket motsvarar antennens effektförlust L . Eftersom verkningsgrad η generellt definieras som $\eta = P_o/P_i$ kan man med Ekvation 130 och Ekvation 132 enkelt visa att antennens strålningsverkningsgrad fås som

$$G = \eta_{str} \cdot D$$

Ekvation 133

Detta är ett enkelt och viktigt samband. Antennvinst och direktivitet är två mått som är väldigt lika och skillnaden mellan dem är helt enkelt antennens strålningsverkningsgrad. Ekvationen visar tydligt att antennvinsten ökar om man antingen gör den mer direktiv eller om man förbättrar verkningsgraden (dvs. minskar effektförlusterna).

Även antennvinst anges ofta med decibelmåttet dB_i. Här är värt att notera att en antenn kan ha en antennvinst $G < 0$ dB_i om den har tillräckligt stor effektförlust.

Antennvinst benämns ofta också antennförstärkning, men detta begrepp kan vara lite missledande eftersom det inte är en faktisk signalförstärkning utan en vinst via diagrammets riktverkan i förhållande till en isotrop antenn.

6.1.7. Reciprocitet

En viktig egenskap hos antenner är reciprocitet, vilket betyder att den har samma egenskaper i sändning och mottagning. Med egenskaper avses här att antennens parametrar såsom strålningsdiagram, antennvinst osv. är samma oberoende av om man sänder eller tar emot med antennen.

Reciprocitet gäller för passiva antenner, dvs. antenner som bara innehåller passiva element och komponenter. Men på grund av den ständigtökande graden av integration av komponenter är det nuförtiden ganska vanligt att antennen är integrerad med aktiv elektronik som förstärkare. Sådana antenner kallas för aktiva antenner och där gäller generellt inte reciprocitet.

En följd av antenners reciprocitet är att om man har ett system med sändarantenn och mottagarantenn åtskilda av ett linjärt och isotrop medium, så spelar det ingen roll vilken antenn som sitter på sändaren och vilken som sitter på mottagaren. Man kan alltså byta plats på antennerna och få samma resultat. Den mottagna effekten blir lika stor i båda fallen.

Ytterligare en följd av antenners reciprocitet är att det spelar ingen roll i vilken riktning som sändning sker. Man kan alltså även byta plats på sändarelektroniken och mottagarelektroniken. Den mottagna effekten blir lika stor i båda fallen.

En kommentar till antenners reciprocitet är att motsvarande egenskaper finns inom kretsteorin. En elektrisk fyrport med passiva komponenter har också egenskapen av att vara reciprok.

6.1.8. Effektiv area

Varje fysiskt realiseringbar antenn har en fysisk area som i mottagarfallet träffas av en infallande elektromagnetiska våg. Denna vågs effekttäthet S ger en uttagbar effekt P på antennens utgång. Kvoten mellan effekttätheten och mottagen effekt kallas för antennens effektiva area enligt följande enkla ekvation.

$$A = \frac{P}{S} \quad \text{Ekvation 134}$$

Definitionen av effektiv area har som man kan se en del likheter med definitionen av antennvinst (och direktivitet). Man kan härleda att

$$A = \frac{\lambda^2}{4\pi} G = \eta \frac{\lambda^2}{4\pi} D \quad \text{Ekvation 135}$$

Det är viktigt att komma ihåg att effektiv area och antennens fysiska area inte är samma sak. Den effektiva arean kan vara såväl mindre som större än den fysiska arean.

Vanligtvis så är de ändå i samma storleksordning.

6.1.9. Antennpolarisation. Polarisationsförluster. Kopolarisation och korspolarisation.

I kapitel 5.2.3 beskrivs en elektromagnetisk vågs polika typer av polarisationstillstånd. Eftersom en antenn är en källa till elektromagnetisk strålning används också begreppet polarisation för antenner. En antenn som t.ex. sänder vertikal polarisation benämns vertikalspolerad antenn.

Ofta kan en antenns polarisation bestämmas med blotta ögat genom att studera hur den är orienterad. En vertikalt orienterad monopol eller dipol sänder ett E-fält som är vertikalt orienterat och därför är det en vertikalspolerad antenn. Om samma antenn orienteras om 90° så att den ligger horisontell med sina spröt, då är polarisationen horisontell.

Om antennen fungerar som mottagarantenn kan den bara ta emot elektromagnetisk strålning som har samma polarisationstillstånd. Om det är skillnad mellan den infallande vågens polarisation och antennens polarisation får en effektförlust i den mottagna signalen. I allmänhet finns det alltid en viss avvikelse. Graden av polarisationsanpassning anges med PLF, polarization loss factor. Om vinkeln mellan en linjärspolerad antenns polarisationsriktning och den infallande vågens polarisationsriktning är ψ , blir polarisationsförlusten

$$\text{PLF} = \cos^2 \psi \quad \text{Ekvation 136}$$

Om $\psi = 0^\circ$ är antennen polarisationsmatchad med $\text{PLF} = 1$ och ingen effekt går förlorad. Om $\psi = 90^\circ$ är det full missanpassning och $\text{PLF} = 0$. Antennen tar i detta fall inte emot någon effekt.

I praktiken så är det svårt att konstruera en antenn där polarisationstillståndet är perfekt. Det finns ofta en liten polarisationsmissanpassning även när antennen är

korrekt mekaniskt orienterad enligt koordinatsystemet. Den effekt som antennen tar emot då den är orienterad enligt sin polarisationsriktning benämns kopolarisation, och den effekt den tar emot då den är orienterad ortogonal mot sin polarisationsriktning benämns korspolarisation. Notera att kopolarisation och korspolarisation kan variera med antennens pekriktning och de blir därmed en funktion av antenndiagrammet. Sambandet mellan korspolarisationsdiagrammet och kopolarisationsdiagrammet är ofta komplicerat och i vissa riktningar kan korspolarisationen vara större än kopolarisationen.

I vissa fall drar man nytta av att elektromagnetisk strålning och antenner har två polarisationsriktningar. Det går att sända två kanaler på samma frekvens och separera dem genom att sända dem i var sin polarisation. Det förutsätter att kanalen inte har reflektioner, eftersom reflektioner generellt vrider polarisationsriktningen. Sändar- och mottagarantenn måste då vara dubbelpolariserade och ha god undertryckning av korspolarisationen.

Moderna kommunikationssystem i miljöer med multipelreflektioner (t.ex. WiFi och mobiltelefoni) drar å andra sidan nytta av två polarisationer på ett annat sätt. Utsänd effekt från t.ex. en mobiltelefon kommer reflekteras ett antal gånger på vägen till basstationen och effekten är då utspridd ganska likafördelat i de två polarisationstillstånden. Basstationen kan dra nytta av detta genom att använda en dubbelpolariserad antenn och en mottagare som kombinerar de två polarisationerna och därmed få en effektvinst eller s.k. diversitetsvinst [49]. Detta ökar signalkvaliteten och i förlängningen datahastigheten. De senaste generationerna av mobiltelefoni och WiFi använder också polarisation för MIMO (Multiple Input, Multiple Output), som är en teknik för att sända och ta emot flera kanaler vid samma frekvens i en miljö med multipelreflektioner [50].

6.1.10. Antennens ingångsimpedans

En antenn har en fysisk ingång för montering till en transmissionsledning, alternativt direkt på sändare/mottagare om integrationsgraden är hög. Antennen har i sitt gränssnitt en ingångsimpedans Z_A som ses som en last. Ingångsimpedansen består generellt av en resistiv del R_A och en reaktiv del X_A , men det som gör antennen annorlunda som last är att den resistiva delen modelleras med två olika typer av resistanser. Dels så finns det alltid vanliga, resistiva förluster i en antenn R_L liknande en vanlig resistor. Detta genererar verkliga förluster i systemet. Men en antenns uppgift är sända ut effekten i omgivningen, vilket från transmissionsledningens perspektiv också ser ut som en effektförlust även om den är nyttig effekt i systemet. Detta modelleras som en strålningsresistans R_r . Totalt sett modelleras alltså en antennens ingångsimpedans som

$$Z_A = R_A + R_L = (R_L + R_r) + X_A$$

Ekvation 137

Antennen har därmed en reflektionsfaktor på samma sätt som en vanlig last som är ansluten till en transmissionsledning. Om systems karakteristiska impedans är Z_0 så är antennens reflektionsfaktor

$$\rho_A = \frac{Z_A - Z_0}{Z_A + Z_0}$$

Ekvation 138

Reflektionen ger upphov till en förlustfaktor $|\rho_A|^2$ i matningen av antennen (kvadraten kommer av att sambandet mellan amplitud och effekt är en kvadrering). Det är inte en verlig effektförlust i antennen eftersom det handlar om en reflekterad effekt, men denna reflektion minskar systemets effektivitet och därmed definieras en reflektionsverkningsgrad som

$$\eta_{refl} = (1 - |\rho_A|^2)$$

Ekvation 139

Att anpassa en antenn för att minimera ρ_A är av uppenbara skäl viktigt för att maximera överförd effekt i systemet. Men det finns en anledning till, vilket är att minska den stående våg som kan uppstå mellan sändare/mottagare och antennen. Om sändare alternativt mottagare också har en missanpassning till transmissionsledningen uppstår multipla reflektioner som kan skapa oönskade effekter i systemet.

6.1.11. Antennens strålningsverkningsgrad

Av den effekt som accepteras av antennen kommer en del bli effektförluster i form av värme och en del nyttig strålningseffekt. Strålningsverkningsgraden kan beräknas som

$$\eta_{str} = \frac{R_r}{R_L + R_r}$$

Ekvation 140

Med låg strålningsresistans R_r kommer kvoten i Ekvation 140 bli nära 1, dvs. accepterad effekt in i antennen kommer bli nyttig utstrålad effekt. Det finns ett annat samband som också ger antennens strålningsverkningsgrad, vilket är Ekvation 133. Förhållandet mellan direktivitet och antennvinst motsvarar alltså förhållandet mellan antennens strålningsresistans och antennens totala resistans.

6.1.12. Antennens totala verkningsgrad

I de föregående kapitlen har totalt tre olika verkningsgrader diskuterats. Dels antennens strålningsverkningsgrad, som är en sann verkningsgrad, men också reflektionsverkningsgrad och polarisationsförluster som inte är effektförluster i verlig mening utan mer en oförmåga att använda all tillgänglig nyttig effekt. Men ur ett systemperspektiv kan alla tre ses som effektförluster. En total verkningsgrad η_{ant} för en antenn är då kombinationen av alla tre enligt följande ekvation

$$\eta_{ant} = \eta_{str} \cdot \eta_{refl} \cdot PLF$$

Ekvation 141

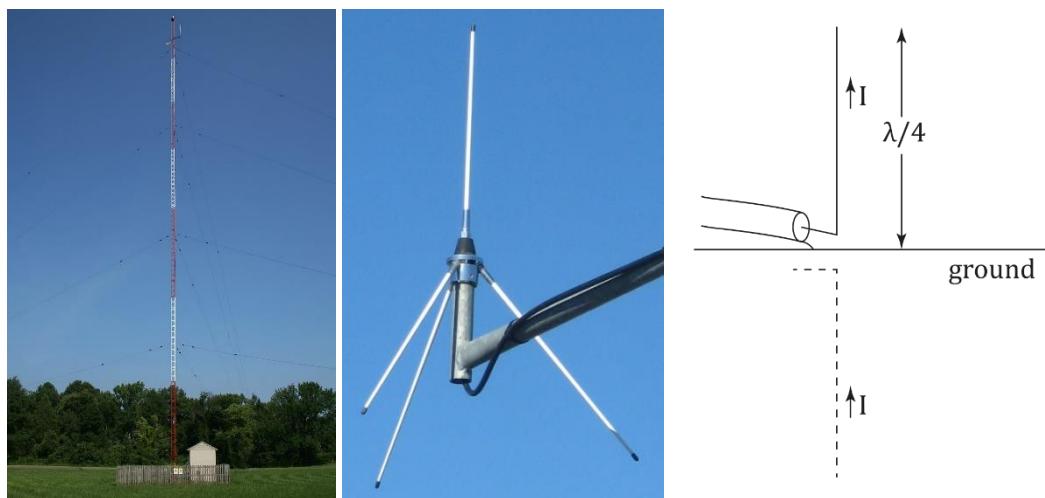
I vissa fall så särskiljer man PLF från de övriga två och då gäller att $\eta_{ant} = \eta_{str} \cdot \eta_{refl}$.

6.2. Olika vanliga antenntyper

6.2.1. Monopolantenn

Monopolantennen [51] är en väldigt enkel antenn som består av en enkel rak ledare, oftast vinkelrätt monterad mot ett ledande plan (typiskt ett jordplan). Monopolen är lik dipolantennen som finns beskriven i 6.2.2, men dipolen består av två stycken ledare.

Figur 50 visar två olika exempel på monopolantennar.



Figur 50

Exempel på monopolantennar. Vänster bild visar en monopolantenn för AM-radio. Sändaren står på marken och har en anslutning till jord och den andra till masten, som är själva antennen. Bilden i mitten visar en vertikalt monterad monopol som har ett enkelt ledande plan som består av de tre spröten som är vinklade nedåt. Bilden till höger visar en enkel konstruktion av en resonant monopolantenn från en koaxialkabal.

Vänster bild: By Ildar Sagdejev (Specious) - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=4948107>

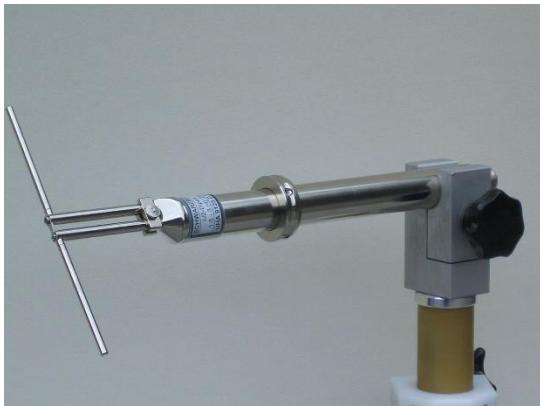
Centerbild: By Ildar Sagdejev (Specious) - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=4948107>

Höger bild: By AG Caesar - Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=67056075>

Monopolen kan ha olika längd och parametrar som strålningsdiagram, impedans och direktivitet varierar med längden. Strålningsdiagrammet är dock alltid rundstrålande pga. symmetrin i antennens konstruktion. Ofta är monopolen designad som en resonant antenn där längden på monopolens ledare är $\lambda/4$. Den har då direktiviteten $D = 5,2 \text{ dBi}$ och strålningsresistansen $R_r = 37 \Omega$.

6.2.2. Dipolantenn

Dipolantennen [52] är lik monopolen men den består som nämnts i stället av två ledare, se exempel i Figur 51. Den vanligaste varianten är en resonant dipol som har en total längd på $\lambda/2$. Den har då direktiviteten $D = 2,1 \text{ dBi}$ och strålningsresistansen $R_r = 73 \Omega$.



Figur 51 Exempel på dipolantenn
By Schwarzbeck Mess-Elektronik - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=2044470>

6.2.3. Loopantenn

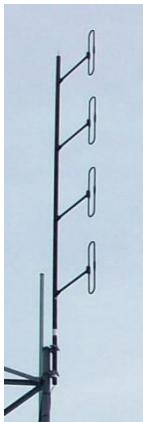
Precis som monopolen och dipolen är loopanten en mycket enkel antenn [53]. Genom att böja en metalltråd i en loop eller slinga skapas en antenn. Beroende på loopens storlek i våglängder räknat får olika egenskaper. Om loopen är liten (betydligt mindre än våglängden) så har den ett strålningsdiagram som liknar en dipols, och direktiviteten blir då ca 2 dBi.

En nackdel med loopantener är att de har ganska låg strålningsresistans och därmed impedans och är därför svåra att matcha till transmissionsledningens impedans. Ett sätt att öka strålningsresistansen är att linda tråden i fler loopar, men då ökar även förlustresistansen och verkningsgraden sjunker.



Figur 52 Loopantenn för kortvåg.
By Trixt - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=4623578>

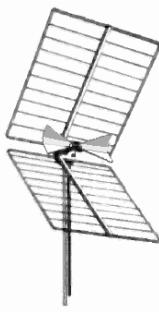
Det finns looplignande strukturer som oftast inte kallas för loopantener. Ett vanligt exempel är en vikt dipol, se Figur 53. Den vikta dipolen bildar en avlång loop som är en full våglängd. Även den vikta dipolantenennen har ett diagram med samma form som dipolen, men dess impedans är fyra gånger så stor som dipolens [52].



*Figur 53 Gruppantenn med fyra stycken vikta dipoler. Gruppantennen beskrivs mer i kapitel 6.3.
By Original image: Milonica - Derived image: Chetvorno - Derived from LocalservicesLakeMountain.jpg. Cropped out unrelated areas of image., CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=40492129>*

6.2.4. Reflektorantenn

Reflektorantennar [54] har någon form av reflektor för att reflektera och fokusera den elektromagnetiska effekten mot en fokalpunkt, där det sitter någon form av antennelement. Reflektorn används för att öka direktivitet och antennvinst.



*Figur 54 Två exempel på reflektorantennar. Vänster bild visar en parabolantenn för satellitkommunikation. I mitten finns en hörnreflektorantenn med en s.k. bowtie-antenn i fokalpunkten. Till höger syns en radarantenn för flygtrafikövervakning. Den rödvita delen är reflektordelen och den har olika mått i sidled och höjdled för att forma huvudloben smalare i azimuth och lite bredare i elevation.
Vänster bild: By en:User:Bogdangiusca, german translation Andreas -horn- Hornig - Picture taken by en:User:Bogdangiusca, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=304885>
Centerbild: By Unknown author - Retrieved August 7, 2014 from Radio and Television News magazine, Ziff-Davis Publishing Co., New York, Vol. 51, No. 6, June 1954, p. 33 on American Radio History site, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=34582966>
Höger bild: By Herr-K - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=10598519>*

En mycket vanlig typ av reflektorantenn är den cirkulära paraboliskt formade antennen som typiskt benämns parabolantenn [55]. En parabolantenn med diametern d har en direktivitet som ges av följande samband

$$D = \left(\frac{\pi d}{\lambda}\right)^2 e_A \quad \text{Ekvation 142}$$

e_A är en dimensionslös faktor som benämns apertureeffektivitet och för en parabolantenn ligger denna faktor mellan 0,55 och 0,7. Direktivitetens enkla, kvadratiska beroendet av diametern d medför att detta är en antenntyp som kan designas för en stor variation i antennvinst.

En behändig approximation för en parabolantenns lobvidd finns i Ekvation 143

$$\theta_{3dB} = \varphi_{3dB} \approx \frac{70\lambda}{d} \quad \text{Ekvation 143}$$

6.2.5. Hornantenn

Hornantennen [56] finns i flera olika former och storlekar. De har en trattliknande form och påminner i utseendet till viss del om ett blåsinstrument.



Figur 55

Olika typer av hornantener. Vänster bild visar ett pyramidformat horn. I bakändan på hornet finns kontakt för koaxialkabelanslutning. Höger bild visar ett cirkulärt horn med s.k. korrugeringar. Dessa används för att forma antennloberna. Detta horn har en annan form av anslutning i bakändan, där en s.k. vågledare kan monteras.
Vänster bild: By Schwarzbeck Mess-Elektronik - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=2049882>

Eftersom storleken och öppningsvinkeln på en hornantenn kan varieras så kan ett horn designas för olika antennvinst / lobvidd. Horn med rektangulär öppning (pyramidformat horn) kan också enkelt designas till olika lobvidder i azimuth och elevation.

Antennvisten för hornantennen beräknas med samma typ av formel som för parabolantennen. Om hornet har en cirkulär öppning (koniskt horn) gäller samma formel, se Ekvation 142. Om hornet har rektangulär öppning gäller att

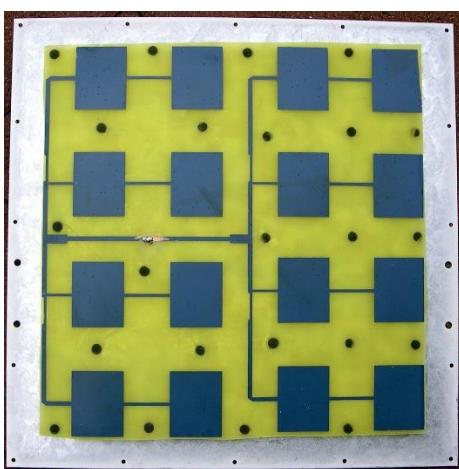
$$D = \frac{4\pi A}{\lambda^2} e_A \quad \text{Ekvation 144}$$

För hornantennar så ligger apertureffektiviteten mellan 0,4 till 0,8. Som tumregel kan $e_A = 0,5$ användas.

En stor fördel med hornantennar är att de har stor bandbredd. De används därför ofta i sammanhang där bandbredd är viktigt, t.ex. i mätsystem. De passar också mycket bra som matande element i en reflektorantenns fokalpunkt eftersom man kan dimensionera hornantennens lobvidd enkelt. Hornmatade parabolantennar är en mycket vanlig antenntyp!

6.2.6. Patchantenn

En patchantenn [57] består av en lågprofil metallyta (eng. patch) som är monterad på ett dielektriskt material över ett jordplan. Oftast är en patchantenn tillverkad av ett tvålagers mönsterkortssubstrat som en microstripantenn [58]. Att använda mönsterkort för konstruktion av en antenn medför också att matningen av antennen är enkel att implementera som en microstrip transmissionsledning.



Figur 56

Exempel på patchantenn. Denna antenn är gjord av PCB i microstripteknologi vilket är en mycket vanlig variant av patchantenn. Antennen består av totalt 16 st. patchar i en 4x4 gruppantenn (se 6.3) som är matade med ett fördelningsnät

By The original uploader was Serge Nueffer at French Wikipedia. - Transferred from fr.wikipedia to Commons by Bloody-libu using CommonsHelper., CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=16238749>

I Figur 56 syns ett exempel på en patchantenn i microstripteknologi. Matningsnätet syns också tydligt, där de olika ledningsbredderna är impedansanpassningar. Att det är totalt 16 stycken patchar som är sammansatta beror på att detta exempel är även en s.k. gruppantenn.

En patchantenn bestående av en enda patch har typisk antennvinst som ligger mellan 6-9 dBi.

6.3. Gruppantenner

Skrives senare...



7. Filter

7.1. Analoga filter

Skrives senare... Här hänvisas tills vidare till kursen Analog Elektronik alternativt Förstärkarteknik.

7.2. Digitala filter

Digitala filter är ett kunskapsområde inom digital signalbehandling. Gemensam nämnare för de olika delområden som finns inom digital signalbehandling är att de alla behandlar processning av diskreta (samplade) signaler. Själva begreppet digital signalbehandling är faktiskt något missvisande eftersom teorin ofta bygger på diskreta signaler (diskret i tid) snarare än digitala signaler (diskret i både tid och amplitud). Att diskretisera även i amplitud innebär ju en form av avrundning till ett digitalt närmevärde, vilket genererar ett litet fel. Men i de flesta så är de digitala signalerna tillräckligt närliggande de analoga värdena så att felet är försumbara.

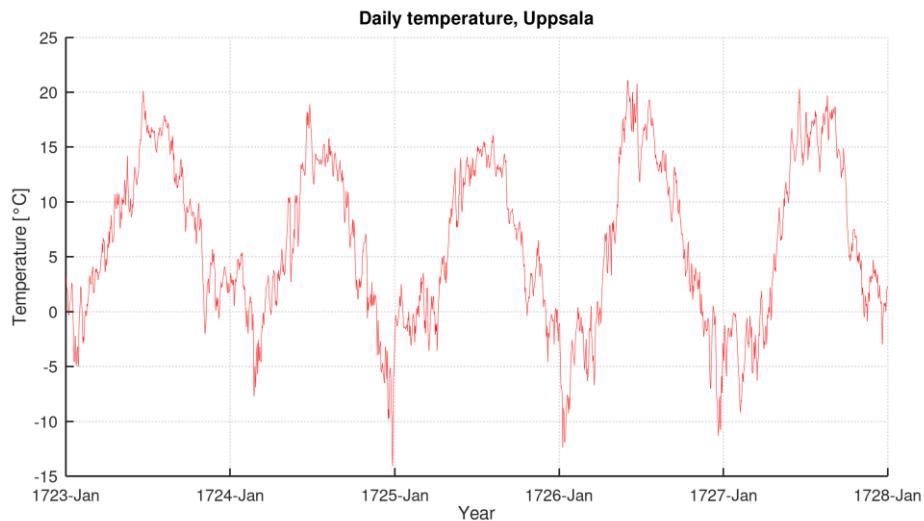
Detta kapitel behandlar den gren av digital signalbehandling som innebär filtrering där endast signalen i sig studeras utan direkt kunskap om det system som genererat signalen. Det finns andra grenar som t.ex. applicerar digital signalbehandling för att prediktera en signals framtida beteende (t.ex. talkodning) eller estimera parametrarna som styr ett dynamiskt systems beteende (t.ex. Kalmanfilter inom reglerteori). Dessa inriktningar lämnas utanför detta kapitels innehåll. När det gäller frekvensanalys av diskreta signaler, som är en viktig grund till filterteorin, hänvisas till kapitel 2.4.

7.2.1. Ett vardagligt exempel

Samplade data och även digital filtrering används ofta i vardagen utan att vi reflekterar särskilt mycket över det. Meteorologiska mätserier (t.ex. mätserier med dagliga mätningar av temperatur, lufttryck, vindstyrka och nederbörd), börskurserna på Stockholmsbörsen samt bostadens elförbrukning per timme är bara några exempel. Exemplet med elförbrukningen är resultatet av filtrerade data (integration av momentaneffekt över en timme) som sampelas en gång per timme. *Väderdata* är en mix av rådata och filtrerade data eftersom vissa data läses av vid jämta mellanrum (ex. temperatur) medan andra integreras över tid som nederbörd.

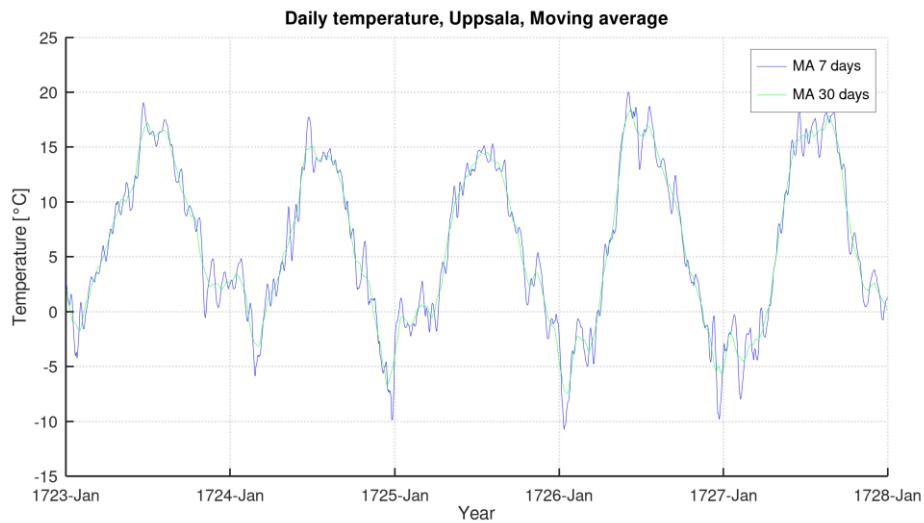
Denna typ av data upplevs ofta som brusiga och stokastiska. Börskurser går upp och ned på daglig basis på ett svårtytt sätt, men över längre tid kan trender läsas ut. Detsamma gäller meteorologiska data, dagstemperaturen kan variera väldigt snabbt från dag till dag medan årstidens temperaturväxlingar är relativt långsamma.

Figur 57 visar ett exempel från en klassisk mätserie med ca 300 år av temperaturdata uppmätt i Uppsala [59]. Figuren visar data för fem år mellan 1723-1727 och det syns tydligt att det finns en grundton med periodtiden ett år, men det finns samtidigt ett flertal svängningar med högre frekvens som ger data ett brudigt utseende.



Figur 57 Temperaturen i Uppsala, uppmätt en gång per dag under åren 1723-1727. Data är korrigerat, bl.a. för att mätplatsen flyttats några gånger.

Om data ska analyseras med avseende på t.ex. enskilda månader eller veckor är de snabba förändringarna dag till dag ett problem. Ett mycket vanligt sätt att filtrera data är då att implementera ett glidande medelvärde (eng. moving average, MA). Då görs en medelvärdesbildning för varje mätsampel genom att ett medelvärde beräknas över en mängd sampel som ligger i direkt anslutning före och efter det mätsampel som studeras. Som exempel, i detta fall kan ett 7-dagars glidande medelvärde enkelt bildas genom att beräkna medelvärde för en specifik dags mätvärde tillsammans med de tre mätsampel (dvs. tre dagar) som ligger före plus de tre mätsampel som ligger efter denna dag. Beräkningen flyttar sedan vidare till nästa dag. Det filter som blir resultatet av denna beräkning påverkar alltså sju mätsampel i taget och sägs därför ha sju tappar. Medelvärdesbildningen medför att de snabba variationerna som ger det brusiga utseendet minskar, vilket tydligt syns i Figur 58. Där har två olika glidande medelvärde implementerats, ett 7-dagars (vecka) och ett 30-dagars (månad). Det kan noteras att ju längre glidande medelvärde desto mer verkar det snabbt varierande bruset filtreras bort.



Figur 58 Filtrering med MA-7 respektive MA-30 av temperaturdata från Figur 57. Data är betydligt mindre brusigt, särskilt för ett 30 dagars glidande medelvärde.

Den enkla algoritmen som det glidande medelvärdet består av verkar således ge ett slags lågpassfilter. Filtrets gränsfrekvens verkar bero på hur lång medelvärdesbildningen är, eftersom 30 dagars glidande medelvärde filtrerar bort mer av de höga frekvenserna jämfört med ett 7-dagars glidande medelvärde. Det bör alltså vara uppenbart att filtrering kan ske på samplade data. Därmed så är den i kommunikationssystem så viktiga uppgiften att filtrera signaler inte bara något som sker på analoga signaler. Snarare tvärtom, med den processorkraft som finns tillgänglig idag så ligger mycket av filtreringen implementerad efter A/D-omvandling i programvara eller digitala kretsar som ASIC och FPGA.

Frågan är bara hur digitala filter analyseras, designas och dimensioneras. För att få en enkel introduktion i denna värld studeras i nästa kapitel hur en enskild frekvens påverkas av ett enkelt 2-tappars glidande medelvärde.

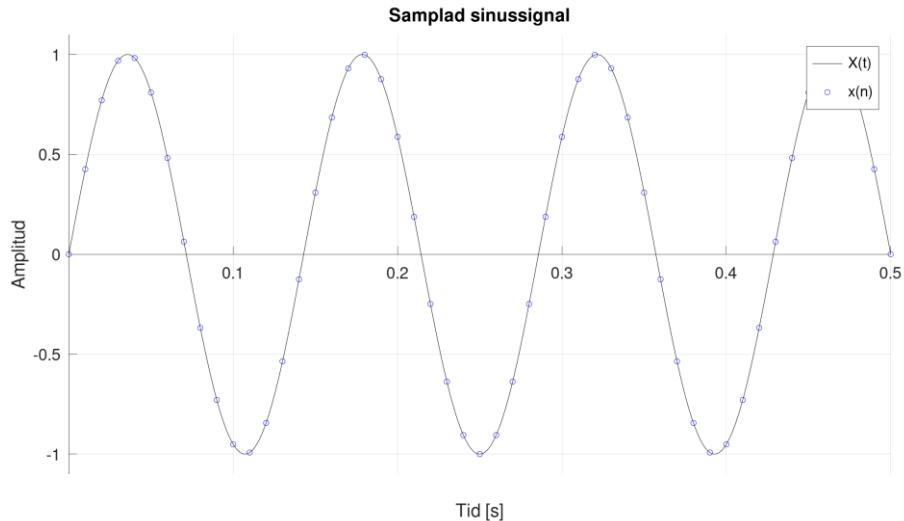
7.2.2. Analys av ett enkelt digitalt lågpassfiltret

Ett enkelt lågpassfilter kan bildas genom att ta ett glidande medelvärde med bara två tappar. Medelvärdesberäkningen görs då med det aktuella samplet plus det direkt föregående samplet. Om rådata från samplingen av den tidskontinuerliga signalen $X(t)$ ligger i en vektor $x[n]$ kan den filtrerade, diskreta signalen $y[n]$ beräknas med följande differensekvation

$$y[n] = \frac{1}{2}[x[n] + x[n - 1]] \quad \text{Ekvation 145}$$

Vektorn $x[n]$ är ju resultatet av en samplingsprocess där analogt mätdata läses av vid jämna mellanrum $t = nT_s$. Se Figur 59 som visar ett exempel med en samplad

sinussignal. T_s är tiden mellan två sampel och n är naturligtvis ett heltalet. Oftast talar man om samplingsfrekvens som då beräknas med det enkla sambandet $f_s = 1/T_s$.



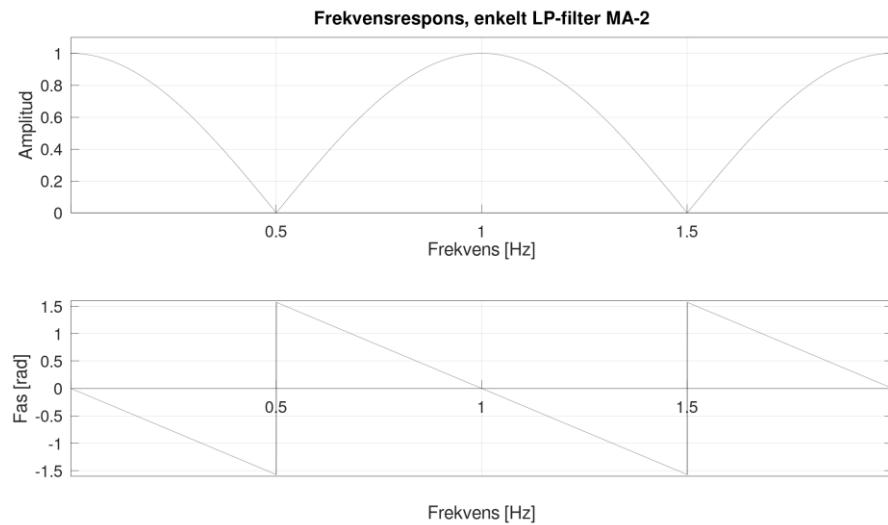
Figur 59

Exempel på en samplad signal. Den kontinuerliga signalen $X(t)$ är en sinussignal med amplituden 1 och frekvensen 7 Hz. Samplingsfrekvensen är $f_s = 100$ Hz, dvs. tiden mellan varje sampel är $T_s = 10$ ms. De blå ringarna i figuren representerar samplat data $x(n)$, vilket blir en vektor med tal.

Detta innebär att den samplade, diskreta signalen kan beskrivas med hjälp av den kontinuerliga signalen $x(t)$ som $x[n] = x(nT_s)$. För att förstå detta enkla digitala filters frekvensfunktion analyseras en sinussignal med amplituden 1 och frekvensen f . Den analoga, kontinuerliga signalen skrivs då som $x(t) = \sin(2\pi ft)$ och den samplade signalen skrivs som $x(n) = \sin(2\pi fnT_s)$. Med Ekvation 145 kan det digitala filtrets utsignal beräknas som

$$\begin{aligned}
 y[n] &= 0,5(x[n] + x[n-1]) = \\
 &= 0,5(\sin(2\pi fnT_s) + \sin(2\pi f(n-1)T_s)) = \\
 &= 0,5(\sin(2\pi fnT_s) + \sin(2\pi fnT_s - 2\pi fT_s)) = \\
 &= 0,5 \sin(2\pi fnT_s) + \sin(2\pi fnT_s - \varphi)
 \end{aligned} \tag{Ekvation 146}$$

Den sista termen $\varphi = 2\pi fT_s$ är en fasvridning som beror av frekvensen. Med komplexa metodens notation kan filtrets respons enligt Ekvation 146 skrivas som $0,5\angle 0 + 0,5\angle(-\varphi(f))$. Denna summa är en ny sinusformad diskret signal med amplitud och fas som varierar med f (och även T_s). Om frekvensen är 0 Hz är $y[n] = x[n]$. När frekvensen ökar, minskar amplituden succesivt tills $y(nT_s) = 0$, vilket sker då $\varphi = \pi$ eller $f = 1/2T_s = 0,5f_s$. Därefter kommer amplituden växa igen och den får ett periodiskt utseende enligt Figur 60.

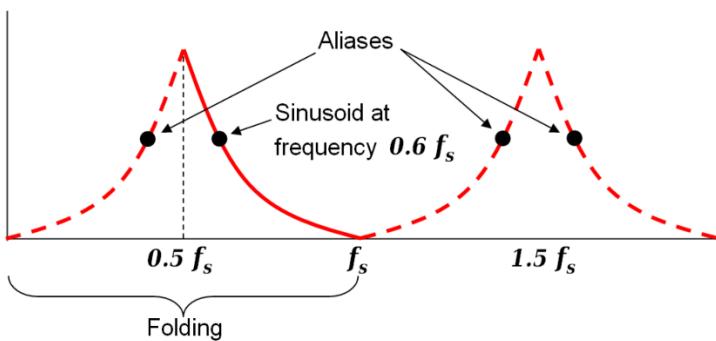


Figur 60 Frekvensresponsen (amplitud och fas) från ett enkelt glidande medelvärde med två tappar (MA-2). Samplingstiden och samplingsfrekvensen är satt till 1 s respektive 1 Hz. För låga frekvenser upp till 0,5 Hz har filtret en lågpassfiltrerande funktion. Om frekvensen är högre än 0,5 Hz kommer filtret inte längre bete sig som ett lågpassfilter.

Att det digitala filtret fungerar som ett lågpassfilter upp till halva samplingsfrekvensen är en viktig observation. Halva samplingsfrekvensen kallas för Nyquistfrekvensen [11] och hänger samman med Nyquist-Shannons samplingsteorem [10] (se kapitel 2.4.4).

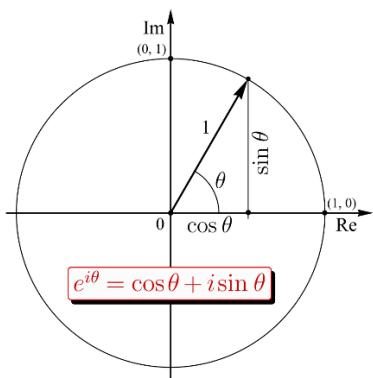
7.2.3. Vikning och vinkningsfilter

Fenomenet i Figur 14 med att för höga frekvenser blir felaktigt representerade efter sampling kallas för vikning (eng. aliasing [60]). Frekvenser över $0,5f_s$ ”viks” kring Nyquistfrekvensen $0,5f_s$ och speglas i denna frekvens, se Figur 61. Den heldragna, röda kurvan visar en signals amplitudspektrum, som är utsträckt mellan $0,5f_s$ och f_s . Eftersom samplingsfrekvensen är för låg i förhållande till signalens högsta frekvens, vilket innebär att respektive frekvenskomponent blir fel representerad efter sampling. Signalens frekvensspektrum blir speglat kring $0,5f_s$ och får ett utseende enligt den rödstreckade linjen mellan kring 0 och $0,5f_s$. Till exempel kommer frekvensen $0,6f_s$ som är markerad med en svart prick uppträda som en signal vid frekvensen $0,4f_s$.



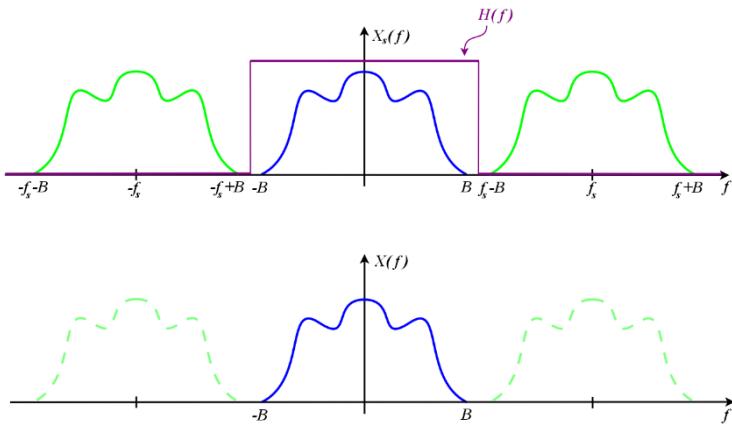
Figur 61 Vikning (aliasing) av frekvenser som ligger över $0,5f_s$.
By Bob K - Own work, CC0, <https://commons.wikimedia.org/w/index.php?curid=76446231>

Fenomenet med vikning repeteras i takt med att frekvensen ökar. I Figur 61 visas även hur frekvenser mellan kring f_s och $1,5f_s$ respektive $1,5f_s$ och $2f_s$ skulle vikas ned till ett frekvensspektrum som ligger mellan 0 och $0,5f_s$. Här är själva vikningen svårare att se och tolka eftersom en spegling kring den tänkta speglingspunkten $0,5f_s$ innebär att frekvensspektrum landar på negativa frekvenser. Men negativa frekvenser har en ganska enkel fysikalisk tolkning. Cosinus och sinus kan ses som en projicering av en enhetsvektor på x-axeln respektive y-axeln enligt Figur 62. Cosinussignal och sinussignal är då motsvarande projicering när vektorn roterar moturs med konstant vinkelhastighet, vars samband med frekvens är $\omega = 2\pi f$. Negativa frekvenser $-f$ är då helt enkelt när vektorn roterar medurs. Fenomenet med vikning till negativa frekvenser kan ses i gamla westernfilmer, där hjulen på en diligens verkar rotera långsamt bakåt isälet för verklighetens snabba rotation framåt, vilket är en konsekvens av den därtida filmbranschens standard med samplingshastigheten 24 bildrutor per sekund. En medurs rotation ger självfallet sinusformade projiceringar, och eftersom $\cos(-\omega t) = \cos(\omega t)$ samt $\sin(-\omega t) = -\sin(\omega t)$ så är enda skillnaden mellan positiva och negativa frekvenser att sinussignalen har omvänt fas.



Figur 62 Figuren visar hur cosinus och sinus kan ses som en projicering av en enhetsvektor på koordinataxlarna x respektive y. Om vektorn roterar med konstant vinkelhastighet ω kommer projiceringarna ge cosinussignal respektive sinussignal
Av Svjo - Eget arbete, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=78365163>

Vikning i samband med sampling är förstås ett problem som behöver hanteras. Lösningen är helt enkelt att införa filtrering på den analoga, kontinuerliga signalen innan sampling. Ett sådant filter är normalt ett lågpassfilter med en gränsfrekvens i näheten av Nyquistfrekvensen och det benämns vikningsfilter (eng. anti-aliasing filter [61]). Figur 63 visar principen för ett sådant filter. Den analoga signalen i figurens övre diagram filtreras med hjälp av ett lågpassfilter $H(f)$. Signalen som ska gå vidare till sampling och A/D-omvandling (blå kurva) passerar genom filtret medan högre frekvenser (grön kurva) filtreras bort.



Figur 63 Principen för ett vikningsfilter. En signal som ska sampelas för digital signalbehandling (blå kurva) innehåller även frekvenskomponenter som inte är av intresse (grön kurva). Dessa frekvenser filtreras därför bort med ett lågpassfilter $H(f)$. I figuren är $H(f)$ ett idealt filter som klipper bort alla frekvenser över filtrets gränsfrekvens. I praktiken finns det förstås alltid begränsningar och högre frekvenser dämpas, snarare är klipps bort helt.
Modifierad från: Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=1065579>

Faktum är att denna naturliga syn på negativa frekvenser är så vanlig inom digital signalbehandling att analys mm normalt sker med både negativa och positiva frekvenser närvarande i samma analys. Figur 63 är ett exempel på detta, där frekvensspektrum för signal och frekvensfunktion för filter är plottade med både positiva och negativa frekvenser.

7.2.4. Z-transform och dess koppling till frekvensanalys

I frekvensanalysen av tidsdiskreta samplade signaler används diskret-tid Fouriertransformen (kapitel 2.4.1). Transformen bestäms av Ekvation 29, vilken upprepas här

$$X(\omega_n) = \sum_{n=-\infty}^{\infty} x[n]e^{-jn\omega_n} \quad \text{Ekvation 147}$$

I analysen av digitala filter generaliseras detta samband till den s.k. Z-transformen [62], se Ekvation 148. Själva transformen betecknas ibland med en egen beteckning $Z\{ \cdot \}$.

$$Z\{x[n]\} = X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n} \quad \text{Ekvation 148}$$

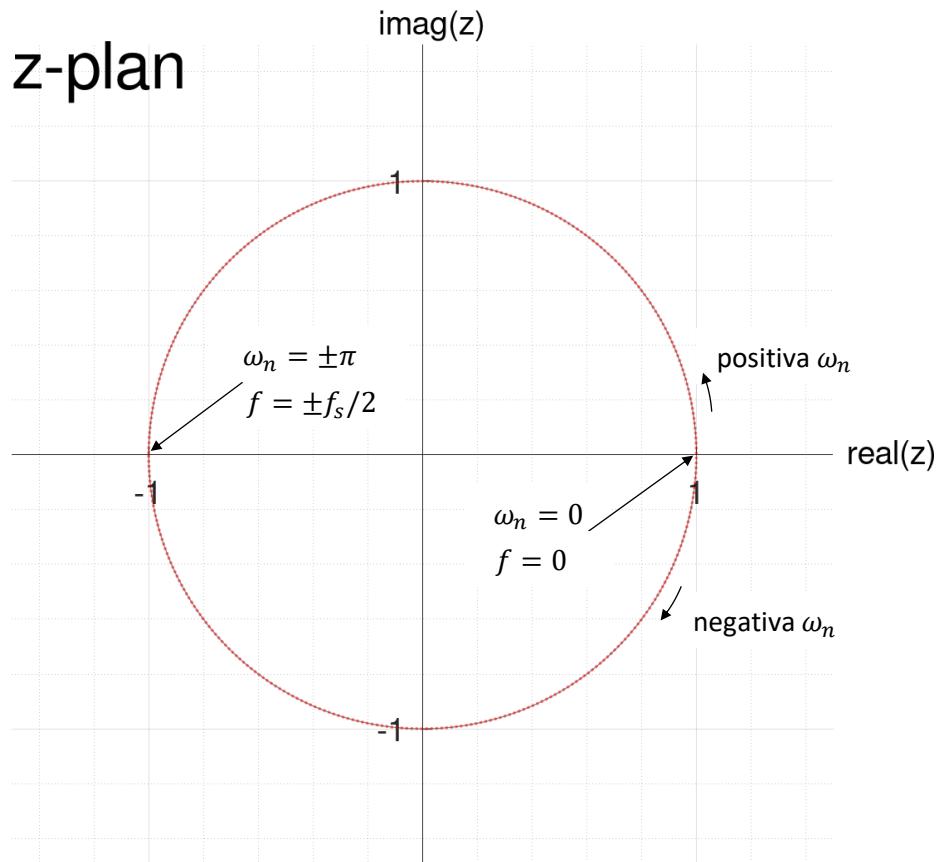
Generaliseringen består i att funktionens invärde, som är det reella ω_n i Ekvation 147 nu är ett komplex tal $z = a + jb = |z|\angle\alpha$.

I praktiken så behandlas normalt signaler som har en start vid $n = 0$, dvs. summan i Ekvation 148 skrivs ofta som

$$Z\{x[n]\} = X(z) = \sum_{n=0}^{\infty} x[n]z^{-n} \quad \text{Ekvation 149}$$

eftersom signalen $x[n]$ av praktiska skäl inte existerar för negativa n . Båda definitionerna av Z-transformen förekommer.

Kopplingen till frekvensanalys och Fouriertransform är enkel på så sätt att om man sätter $z = e^{j\omega_n}$ så övergår summan i Ekvation 148 till DTFT i Ekvation 147. $z = e^{j\omega_n}$ kan med komplex notation skrivas som $z = e^{j\omega_n} = 1\angle\omega_n$. Eftersom ω_n ligger mellan $-\pi < \omega_n < \pi$ enligt kapitel 2.4.1 så kan vi tolka vad $z = 1\angle\omega_n$ innebär i ett komplex talplan, se Figur 64.



Figur 64

Komplexa z-planet och mappningen av de reella vinkelfrekvenserna ω_n , $z = e^{j\omega_n}$ består av punkterna som ligger på enhetscirkeln (rödprickad kurva), eftersom $-\pi < \omega_n < \pi$.

Figuren visar att de normaliserade vinkelfrekvenserna ω_n består av alla punkter som ligger på enhetscirkeln i det komplexa talplanet (rödprickad kurva). De verkliga frekvenserna för den samplade signalen ligger naturligtvis på samma kurva och det kan noteras att

- Frekvensen $f = 0$ ligger i punkten $z = 1 + 0 \cdot j$
- Frekvenserna $f = \pm f_s/2$ ligger i punkten $z = -1 + 0 \cdot j$
- Övriga frekvenser mellan 0 och $\pm f_s/2$ ligger proportionellt fördelade längs cirkeln på så sätt att positiva frekvenser ligger längs övre delen av enhetscirkeln och negativa frekvenser ligger längs undre delen

När ett filter alltså analyseras och dimensioneras med Z-transformen genomförs frekvensanalysen genom att betrakta hur filtret beter sig för de z som uppfyller $z = e^{j\omega_n}$, dvs ligger på enhetscirkeln.

Z-transformen har vissa egenskaper som gör den till ett smidigt verktyg för att arbeta med digitala filter och digital signalbehandling. De viktigaste är

- Linearitet:
$$\begin{aligned} Z\{C_1 \cdot x_1[n] + C_2 \cdot x_2[n]\} &= C_1 \cdot Z\{x_1[n]\} + C_2 \cdot Z\{x_2[n]\} \\ &= C_1 X_1(z) + C_2 X_2(z) \end{aligned}$$
- Tidsförskjutning: $Z\{x[n - k]\} = z^{-k} \cdot X(z)$
- Faltning: $Z\{x_1[n] * x_2[n]\} = Z\{x_1[n]\} \cdot Z\{x_2[n]\} = X_1(z) \cdot X_2(z)$

För begreppet faltning, se vidare i kapitel 7.2.5.

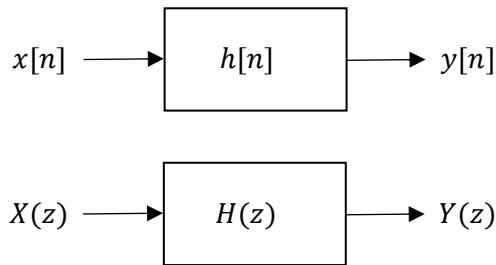
7.2.5. Digitala filters överföringsfunktion $H(z)$

Fördelen i att gå över till en mer generell transform är inte att underlätta själva frekvensanalysen av en signal, utan snarare underlätta filteranalysen. Exemplet med ett 2-tappars glidande medelvärde i kapitel 7.2.2 visar att digital filtrering ofta betyder att definiera en differensekvation och dess koefficienter. Just detta enkla filter gick att analysera rimligt enkelt utan transformverktyg, men svårigheten stiger snabbt vid större filter. Z-transformen är ett utmärkt verktyg för att förenkla design, syntes och analys av filters differensekvationer, i synnerhet om flera filter kopplas i kaskad.

En viktig förutsättning för analys av filter med hjälp av Z-transform är att filtret är tidsinvariant och linjärt. Att ett filter är tidsinvariant betyder helt enkelt att dess egenskaper är oförändrade över tid. Det uttrycks matematiskt som att om filtrering av samplade signalen $x[n]$ ger $y[n]$, så kommer en filtrering av den tidsförskjutna signalen $x[n - k]$ ge motsvarande förskjutning $y[n - k]$. Linearitet innebär att superpositionsprincipen gäller, vilket betyder två saker. Det första är att om filtreringen av $x[n]$ ger $y[n]$, så kommer filtreringen av $C \cdot x[n]$ ge $C \cdot y[n]$. C är en godtycklig konstant och kan ses som en förstärkningsfaktor eller dämpning. Det andra är att om $x_1[n]$ ger $y_1[n]$ och $x_2[n]$ ger $y_2[n]$ så kommer den summerade insignalen $x_1[n] + x_2[n]$ ge utsignalen $y_1[n] + y_2[n]$.

Ett filters överföringsfunktion är en matematisk beskrivning av hur filtret "överför" en insignal till en utsignal. Figur 65 visualiseras detta som att filtret är en slags svart låda som tar emot en insignal, bearbetar denna och skickar ut en utsignal.

Överföringsfunktionen är då en beskrivning av den svarta lådans egenskaper.



Figur 65 Enkelt blockschema som visar en grafisk representation av ett filters överföringsfunktion $H(z)$, se undre figuren. I det samplade tidsplanet har $H(z)$ sin motsvarighet $h[n]$, som kallas för filtrets impulsrespons (övre figuren). Blockschemat är en visualisering av filtret som en funktion, där indata passerar filterfunktionen och en ny sekvens av data kommer ut ur filtret som utdata.

Ett digitalt filters överföringsfunktion definieras som den funktion som uppfyller följande ekvation

$$H(z) = \frac{Y(z)}{X(z)} \quad \text{Ekvation 150}$$

Utsignalen från filtret blir helt enkelt en multiplikation $Y(z) = H(z) \cdot X(z)$, något som känns igen från ellära och analoga filter där ett filters överföringsfunktion ger sambandet $Y(j\omega) = H(j\omega) \cdot X(j\omega)$.

Sambandet mellan $x[n]$ och $y[n]$ bestäms av en s.k. faltning (eng. convolution), vilket är en matematisk operation mellan två funktioner [63]. Faltningsoperationen brukar betecknas med $*$ -tecknet, och i fallet med diskreta signal som filtreras i ett digitalt filter skrivs utsignalen som

$$\begin{aligned} y[n] &= (h * x)[n] = \sum_{k=-\infty}^{\infty} h[k] \cdot x[n-k] = \\ &= \sum_{k=-\infty}^{\infty} h[n-k] \cdot x[n] \end{aligned} \quad \text{Ekvation 151}$$

Faltning är en operation som kan vara svår att skapa en intuitiv förståelse för vad den egentligen innebär och hur den påverkar data, men glidande medelvärde ger en bra bild. I fallet med diskret faltning låter man ena vektorn med data "glida" över den andra med faktorn $n - k$, värdena i de två vektorerna multipliceras elementvis och summeras.

Faltning beskriver alltså hur filtreringen sker i tidsplanet. Men precis som för filtrering av analoga signaler, analys i tidsplanet är arbetskrävande och svår att få bra bild över vad som händer. Därför sker en transform (som är $j\omega$ -metoden alternativt Laplacetransform i analoga fallet) till z-planet för att få en enklare matematik och bättre möjlighet till förståelse.

Ett viktigt begrepp är filtrets impulsresons. Om ett digitalt filter matas med en impulsfunktion som insignal, vilken definieras som

$$x[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \quad \text{Ekvation 152}$$

så kommer faltningsfunktionen i Ekvation 151 ge $y[n] = h[n]$. Därför benämns $h[n]$ ofta som filtrets impulsrespons. Om impulsfunktionen Z-transformeras enligt Ekvation 149 får $X(z) = 1$, vilket är logiskt eftersom utsignalen då blir $Y(z) = H(z) \cdot 1 = H(z)$. Att mata ett filter med en impulsfunktion på ingången kommer alltså ge svar på filtrets karaktär och egenskaper.

I fallet med två eller flera kaskadkopplade filter så är analys i z-planet definitivt att föredra. Den totala filterfunktionen beräknas då helt enkelt som en multiplikation mellan de ingående, separata filtren. Som vi ska se i kommande kapitel så beskrivs digitala filter med rationella funktioner i z-planet, så kaskadkoppling innebär helt enkelt beräkning av polynommultiplikationer.

7.2.6. FIR-filter

Filter som kan beskrivas med en differensekvation där $y[n]$ är en linjärkombination endast av $x[n]$ -sampel kallas för FIR-filter (Finite Impulse Response). Ett filter med sådan differensekvation med ordningen N ($N + 1$ styck termer) kan skrivas enligt Ekvation 153 [64].

$$\begin{aligned} y[n] &= \sum_{k=0}^N b_k x[n - k] = \\ &= b_0 x[n] + b_1 x[n - 1] + \dots + b_{N-1} x[n - N] \end{aligned} \quad \text{Ekvation 153}$$

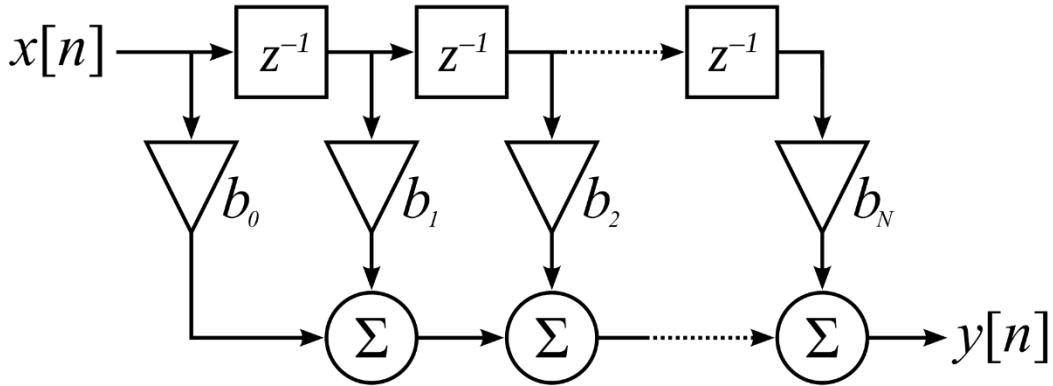
FIR-filter har vissa egenskaper. Dels så är filtret alltid stabilt, dvs. utsignalen är ändlig och kan inte börja självsvänga. En annan egenskap är att filtrets impulsfunktion är enkel att läsa ut och förstå. Matas filtret med impulsfunktionen beskriven av Ekvation 152 kommer utsignalen bli en vektor $[\dots \ 0 \ 0 \ b_0 \ b_1 \ \dots \ b_N \ 0 \ 0 \ \dots]$. Impulsresponsen blir alltså lika med filtrets koefficienter. Denna egenskap gör att man generellt kan designa ett FIR-filter som med god noggrannhet approximerar ett visst önskat impulssvar. Eftersom impulssvar för stabila filter klingar av efter tillräckligt lång tid kan man välja en tillräckligt lång del av ett impulssvar, trunkera detta och helt enkelt direkt sätta filterkoefficienterna efter det trunkerade impulssvarets värden. En nackdel är att denna typ av design tenderar att skapa långa filter som är beräkningskrävande.

Z-transformen för ett FIR-filter ges av

$$H(z) = \sum_{k=0}^N b_k z^{-k} = b_0 + b_1 z^{-1} + \dots + b_N z^{-N} \quad \text{Ekvation 154}$$

Figur 66 visar exempel på en implementationsstruktur för FIR-filter. Insignalen, som består av en kontinuerlig sekvens av sampel $x[n]$ skiftas steg för steg genom en vektor med minnen vilket motsvarar en stegvis fördröjning z^{-1} (eller T_s i det samplade

tidplanet) viktas med filterkoefficienterna $b_0 b_1 \dots b_{N-1}$ och summeras till utsignalen $y[n]$.



Figur 66 En typ av implementationsstruktur för FIR-filter.
By BlanchardJ - Own work, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=3606087>

7.2.7. IIR-filter

Genom att återkoppla utsignalen $y[n]$ tillbaka in i filtret skapas ytterligare möjligheter i filterdesignen. Återkopplingen sker genom att tidigare utsignalssampel viktas med koefficienter och summeras tillsammans med de viktade samplen från insignalen. Det kan beskrivas med följande differensekvation

$$\sum_{l=0}^Q a_l y[n-l] = \sum_{k=0}^P b_k x[n-k] \quad \text{Ekvation 155}$$

Ekvation 154 kan även skrivas på följande form

$$\begin{aligned} a_0 y[n] + a_1 y[n-1] + \dots + a_Q y[n-Q] &= \\ b_0 x[n] + b_1 x[n-1] + \dots + b_P x[n-P] \end{aligned} \quad \text{Ekvation 156}$$

eller

$$\begin{aligned} y[n] = \frac{1}{a_0} (a_1 y[n-1] + \dots + a_Q y[n-Q] + b_0 x[n] \\ + b_1 x[n-1] + \dots + b_P x[n-P]) \end{aligned} \quad \text{Ekvation 157}$$

Ofta sätts $a_0 = 1$ i ekvationerna för IIR-filter, vilket innebär att man kan se Ekvation 155 till Ekvation 157 skrivna i en något annorlunda form. Ordningen på filtret bestäms av $\max(P, Q)$.

Om filtret matas med en impulsfunktion kommer impulsresponsen aldrig klinga av helt. Denna filtertyp benämns därför IIR. Eftersom det finns en återkoppling i filtret finns risken för att filtret blir instabilt. Det är ganska enkelt att se att exemplet $y[n] =$

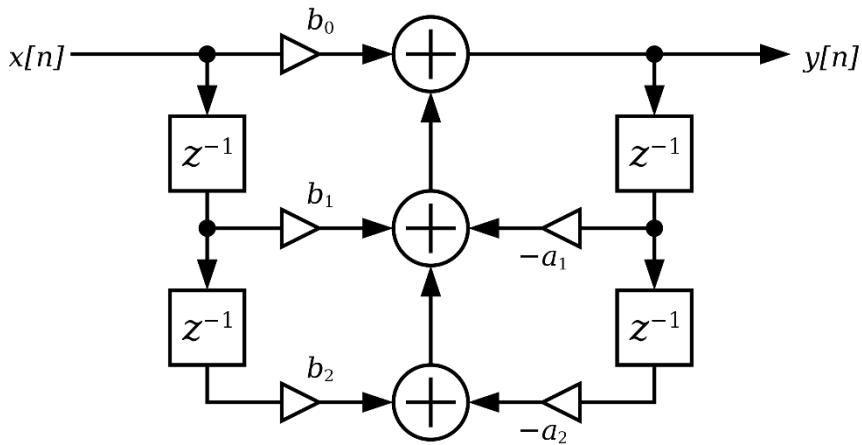
$2y[n] + x[n]$ representerar ett instabilt filter. Utsignalen kommer snabbt växa eftersom man hela tiden adderar dubbla föregående utsignalvärdet. Kriterierna för ett stabila IIR-filter beskrivs i 7.2.8.

Z-transformen för IIR-filter påminner om den för FIR-filter. Skillnaden är att $H(z)$ nu är en rationell funktion snarare än en polynomfunktion, se Ekvation 158 nedan.

$$H(z) = \frac{\sum_{k=0}^P b_k z^{-k}}{\sum_{l=0}^Q a_l z^{-l}} = \frac{b_0 + b_1 z^{-1} + \dots + b_P z^{-P}}{a_0 + a_1 z^{-1} + \dots + a_Q z^{-Q}}$$

Ekvation 158

I Figur 67 visas en typ av implementationsstruktur för IIR-filter. Denna figur visar ett s.k. biquad-filter, vilket är ett kort digitalt filter med ordningen 2. Både insignalen och utsignalen har en struktur med stegvisa fördröjningar z^{-1}



Figur 67

En typ av implementationsstruktur för IIR-filter. Filtret i figuren är ett andra ordningens filter och det benämns biquad-filter. Strukturen påminner förstås om strukturen för FIR-filter, förutom att utsignalen återkopplas tillbaka i i filtret och därmed bidrar till filtrets egenskaper.

By Akilaa - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=9984215>

7.2.8. Poler och nollställen

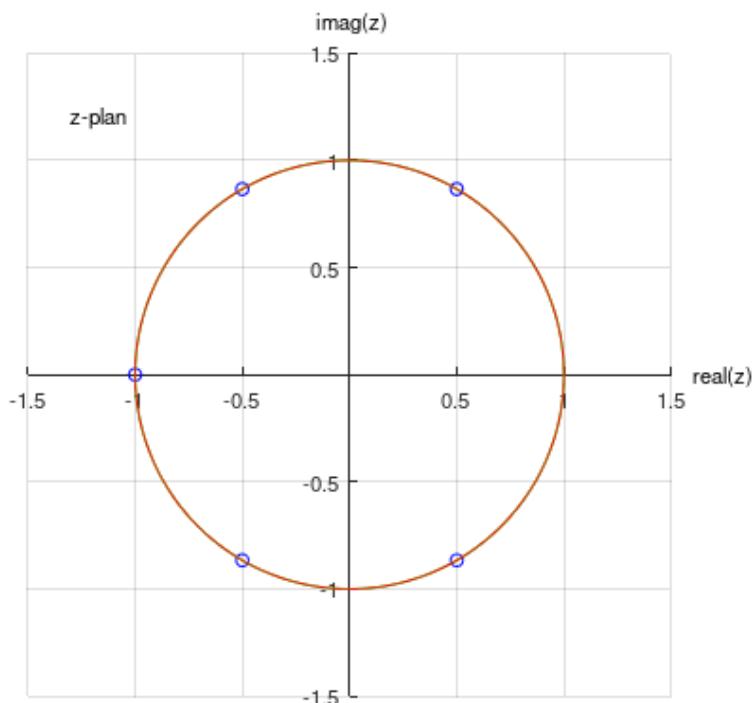
De överföringsfunktioner för FIR- respektive IIR-filter som beskrivs av Ekvation 154 och Ekvation 158 består av polynomfunktioner i variabeln z^{-1} . I fallet FIR så är det en enda polynomfunktion medan IIR består av kvoten av två polynom vilket är en rationell funktion. Polynom kan skrivas om och beskrivas av en multiplikation av faktorer beskrivna av polynomets nollställen. Dessa polynomfunktioners nollställen visar sig vara viktiga analysverktyg när digitala filter ska analyseras och designas.

Ett FIR-filter kan då skrivas om genom att göra beräkningen $H(z) = 0$. Genom att sätta $H(z) = 0$ i Ekvation 154 och förlänga med z^N fås

$$0 = b_0 z^N + b_1 z^{N-1} + \dots + b_N$$

Ekvation 159

Polynomet som ges av Ekvation 159 har N stycken nollställen, som i det generella fallet är komplexa tal. Figur 68 visar ett exempel för hur nollställena för ett glidande medelvärdes-filter med sex tappar fördelar sig. De fem rötterna till polynomekvationen $\frac{1}{6}(z^5 + z^4 + z^3 + z^2 + z + 1) = 0$ består av en reell rot och fyra komplexa rötter. Notera hur de komplexa rötterna ligger spegelsymmetriskt kring realaxeln i z -planet. Detta gäller generellt för polynomekvationer med reella koeficienter. Om det finns en komplex rot z till $H(z) = 0$ finns det även en komplex rot som är komplexkonjugatet z^* .



Figur 68 Exempel på nollställen (blå ringar) för ekvationen $\frac{1}{6}(z^5 + z^4 + z^3 + z^2 + z + 1) = 0$. Ekvationen motsvarar ett 6-tappars glidande medelvärde. Nollställena visar sig ligga på enhetscirkeln.

Ekvation 159 kan således skrivas om på formen

$$0 = b_0(z - z_1) \cdot (z - z_2) \cdot \dots \cdot (z - z_N)$$

Ekvation 160

där z_1, z_2 osv. är nollställena till ekvationen. Denna form är praktisk vid analys och design av filter. Där Z-transformen för ett FIR-filter har ett nollställe är filtrets amplitudrespons noll. Så om ett nollställe ligger på enhetscirkeln, där man finner en signals frekvenser enligt kapitel 7.2.4, kommer utsignalen bli noll vid denna frekvens. Önskar man ta bort en specifik frekvens med ett FIR-filter så kan man alltså enkelt placera ut nollställen vid dessa frekvenspunkter.

Observera att ett filters nollställe inte nödvändigtvis måste ligga på enhetscirkeln. Det är inget som hindrar att nollställe ligger var som helst i z-planet. Men det är bara när nollställen ligger precis på enhetscirkeln som en frekvens helt släcks ut av filtret.

Genom att återigen analysera det enkla lågpassfilteret som behandlades i kapitel 7.2.2 går det att skapa en enkel geometrisk förståelse för filtrets amplitud och fasrespons, dvs.

$|H(z = e^{j\omega_n})|$ och $\angle H(z = e^{j\omega_n})$ där ω_n är den normerade vinkelfrekvensen. Filtrets z-transform kan skrivas som $H(z) = \frac{1}{2}(1 + z^{-1})$ vilket har nollstället som ges av ekvationen $z + 1 = 0$. Detta nollställe ligger förstås vid $z_1 = -1$. Med hjälp av nollstället och lite algebraiska manövrar kan filtret skrivas om enligt

$$H(z) = \frac{1}{2}(1 + z^{-1}) = \frac{1}{2}z^{-1}(z + 1) = \frac{1}{2}z^{-1}(z - z_1) \quad \text{Ekvation 161}$$

Genom att sätta in $z = e^{j\omega_n}$ i Ekvation 161 kan filtrets frekvensrespons analyseras. Frekvensresponsen delas lämpligen in i amplitudrespons respektive fasrespons enligt Ekvation 162 till Ekvation 164.

$$H(e^{j\omega_n}) = \frac{1}{2}e^{-j\omega_n}(e^{j\omega_n} - z_1) \quad \text{Ekvation 162}$$

$$\begin{aligned} |H(e^{j\omega_n})| &= \left| \frac{1}{2}e^{-j\omega_n}(e^{j\omega_n} - z_1) \right| = \\ &= \frac{1}{2} \cdot |e^{-j\omega_n}| \cdot |(e^{j\omega_n} - z_1)| = \\ &= \frac{1}{2} |e^{j\omega_n} - z_1| \end{aligned} \quad \text{Ekvation 163}$$

$$\begin{aligned} \angle H(e^{j\omega_n}) &= \angle \left(\frac{1}{2}e^{-j\omega_n}(e^{j\omega_n} - z_1) \right) = \\ &= \angle e^{-j\omega_n} + \angle(e^{j\omega_n} - z_1) \end{aligned} \quad \text{Ekvation 164}$$

I Ekvation 163 betyder $|e^{j\omega_n} - z_1|$ helt enkelt avståndet i komplexa talplanet mellan de två talen $e^{j\omega_n}$ och z_1 , där $e^{j\omega_n}$ följer enhetscirkeln och varierar i takt med frekvensen. I Ekvation 164 betyder $= \angle e^{-j\omega_n}$ helt enkelt $-\omega_n$ radianer och $\angle(e^{j\omega_n} - z_1)$ är vinkelns för det komplexa talet $e^{j\omega_n} - z_1$.

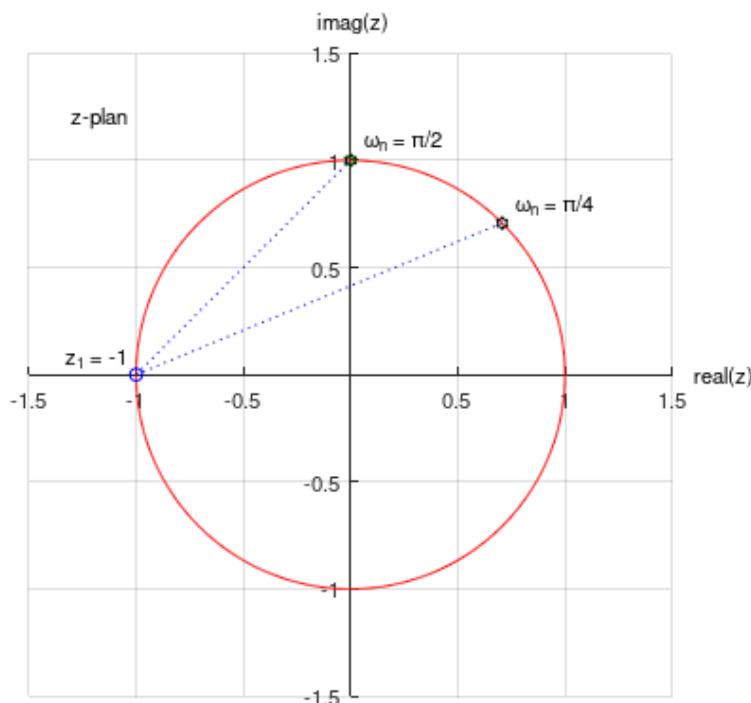
Figur 69 visar en visualisering av amplitud- och fasrespons för $H(z)$ i detta exempel. Vid $\omega_n = \pi/2$ är avståndet $|e^{j\omega_n} - z_1|$ till nollstället $\sqrt{2}$ (Pythagoras sats för en triangel med katetrarna med längden 1), vilket ger $|H(e^{j\omega_n})| = \frac{1}{2}|e^{j\omega_n} - z_1| = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} \approx 0,71$. Vinkelns $\angle e^{-j\omega_n}$ är förstås $-\pi/2$ (90°) och $\angle(e^{j\omega_n} - z_1)$ är helt tydligt $\pi/4$ radianer (45°), som då ger filtrets fas $-\pi/2 + \pi/4 = -\pi/4$ radianer (eller -45°). Detta kan jämföras med grafen i Figur 60, så ser man att amplitud och fas vid $f = 25$ Hz enligt denna geometriska analys stämmer med värdena i graferna för amplitud och fas.

Motsvarande analys för den andra punkten vid $\omega_n = \pi/4$ ger

$|H(e^{j\omega_n})| = \frac{1}{2}\sqrt{2 + \sqrt{2}} \approx 0,92$ och $\angle H(e^{j\omega_n}) = -\pi/8$ ($-22,5^\circ$). Även dessa värden stämmer med Figur 60. Det är också uppenbart att filtret har amplitudresponsen 0 när

man kommer till normerade frekvensen $\omega_n = \pi$ eftersom man då befinner sig i nollstället.

Det går alltså utmärkt att göra en analys av filtret med enkel geometri direkt i z-planet så som Figur 69 visar. Faktum är att fasen för detta filter kan läsas ut väldigt enkelt som negativa värde för vinkeln mellan realaxeln och blåprickade linjerna. Detta är en konsekvens av randvinkelsatsen. Men notera att randvinkelsatsen endast är applicerbar då nollställena ligger på enhetscirkeln!



Figur 69 Visualisering av ett enkelt filters amplitudrespons respektive fasrespons. Filtret $H(z) = \frac{1}{2}(1 + z^{-1})$ har nollstället $z_1 = -1$. Om den normerade vinkelfrekvensen ω_n varieras kommer avståndet och vinkeln till nollstället variera i takt med förändringen av ω_n , vilket visualiseras med de blåprickiga linjerna. Man ser tydligt att avståndet (absolutbeloppet) för filtret är som störst vid $\omega_n = 0$ och minskar successivt till $\omega_n = \pi$. Det är tydligt att filtret har en lågpasskaraktär. Jämför med Figur 60.

För mer komplexa filter så delas denna typ av direkt, geometrisk analys upp i delar där varje nollställe kan analyseras separat. Ekvation 160 visar ju att vi kan se filtret uppbyggt som ett polynom $b_0(z - z_1) \cdot (z - z_2) \cdot \dots \cdot (z - z_N)$. Absolutbeloppet för filtret motsvaras då av $|b_0| \cdot |(z - z_1)| \cdot |(z - z_2)| \cdot \dots \cdot |(z - z_N)|$ och fasen fås som $[\angle e^{-j\omega_n} + \angle(e^{j\omega_n} - z_1)] + \dots + [\angle e^{-j\omega_n} + \angle(e^{j\omega_n} - z_N)]$. Totala amplitudresponsen ges således av en multiplikation av de enskilda nollställenas avstånd till ω_n och fasen blir en summa av de enskilda faserna.

För IIR-filter kan motsvarande analys göras. Ekvation 158 kan skrivas om som en rationell funktion enligt följande samband.

$$H(z) = \frac{z^{-P}(b_0 z^P + b_1 z^{P-1} + \dots + b_P)}{z^{-Q}(a_0 z^Q + a_1 z^{Q-1} + \dots + a_Q)}$$

Ekvation 165

Täljaren kan analyseras på exakt samma sätt som analysen för FIR-filter med nollställen osv. När det gäller nämnaren så görs en motsvarande analys, men där kallas lösningarna till ekvationen $a_0 z^Q + a_1 z^{Q-1} + \dots + a_Q = 0$ för poler i stället. Avståndet mellan en punkt på enhetscirkeln och en pol motsvarar alltså absolutbeloppet för en funktion som står i nämnaren i ett rationellt uttryck. Det är därför uppenbart att det inte går att placera ett IIR-filters pol på enhetscirkeln eftersom det skulle ge division med 0. Det visar sig att för att bygga ett stabilt IIR-filter måste samtliga poler ligga innanför enhetscirkeln.

Analys av fas i nämnaren sker på motsvarande sätt som den för täljaren. Filtrets totala fasrespons kommer ges av fasskillnaden mellan täljare och nämnare i Ekvation 165.

Om avståndet för en pol $|(z - z_1)|$ är litet kommer $|H(e^{j\omega_n})|$ bli stort eftersom det blir en division med ett litet tal. Införandet av återkoppling och poler i ett filter ger alltså utökade möjligheter för att designa filter enligt en önskad filterfunktion.

7.2.9. Notchfilter

Skrives senare.

7.2.10. Butterworthfilter

Skrives senare.

7.2.11. Chebyshevfilter

Skrives senare.

7.2.12. Praktisk filterimplementation

Skrives senare.

8. Modulation

Detta kapitel är delvis kopierat från [21] (kapitel 5 och 7) i enlighet med skolornas kopieringsavtal.

8.1. Analog modulation

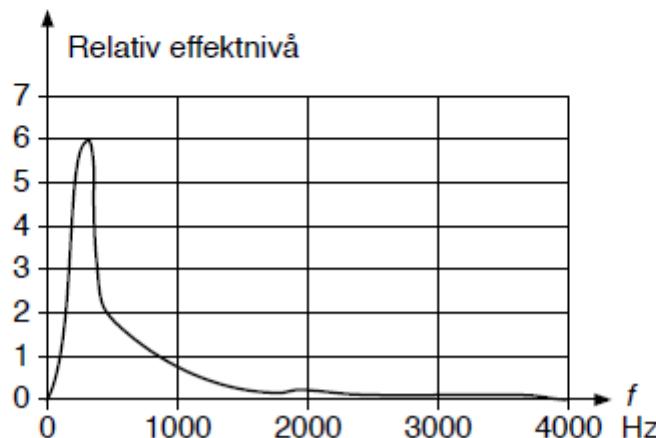
8.1.1. Linjär modulation

Vid linjär modulation frekvenstransponeras basbandssignalen $m(t)$ till passbandet genom multiplikation med en sinusformad bärväg med amplitud U_b och bärvinkelfrekvens $\omega_b = 2\pi f_b$. På så sätt bildas en modulerad bärväg:

$$u_{bm}(t) = m(t)U_b \cos \omega_b t \quad (5.1)$$

där $u_{bm}(t)$ är den modulerade bärvägens momentana spänning.

I ett allmänt fall är $m(t)$ en basbandssignal vars amplitud och frekvens varierar inom vissa gränser. Ett exempel på en sådan basbandsignal är den mänskliga rösten vars genomsnittliga frekvensfördelning visas i Figur 5.1.



Figur 5.1 Genomsnittlig frekvensfördelning för den mänskliga rösten.

Vid linjär modulation är bandpasspekret mer eller mindre likformigt med basbandsspekret och i de standardiserade linjära modulationstyper som förekommer ingår alltid en operation som kan uppfattas som multiplikation som i uttrycket (5.1). De standardiserade typer som beskrivs i det följande är:

- Modulation med dubbelt sidband (eng. DSB-modulation = = Double-Side Band modulation).
- Amplitudmodulation, AM.
- Modulation med enkelt sidband, ESB (eng. SSB-modulation = = Single-Side Band modulation).
- Modulation med stympat sidband (eng. VSB-modulation = = Vestigial-Side Band modulation).

8.1.2. Modulation med dubbelt sidband

Vi utgår till att börja med, för enkelhets skull, från en basbandssignal, $u_m(t)$ (= den modulerande signalen), med konstant amplitud = U_m och vinkel frekvens = $\omega_m = 2\pi f_m$, d.v.s.

$$u_m(t) = U_m \cos \omega_m t \quad (5.1)$$

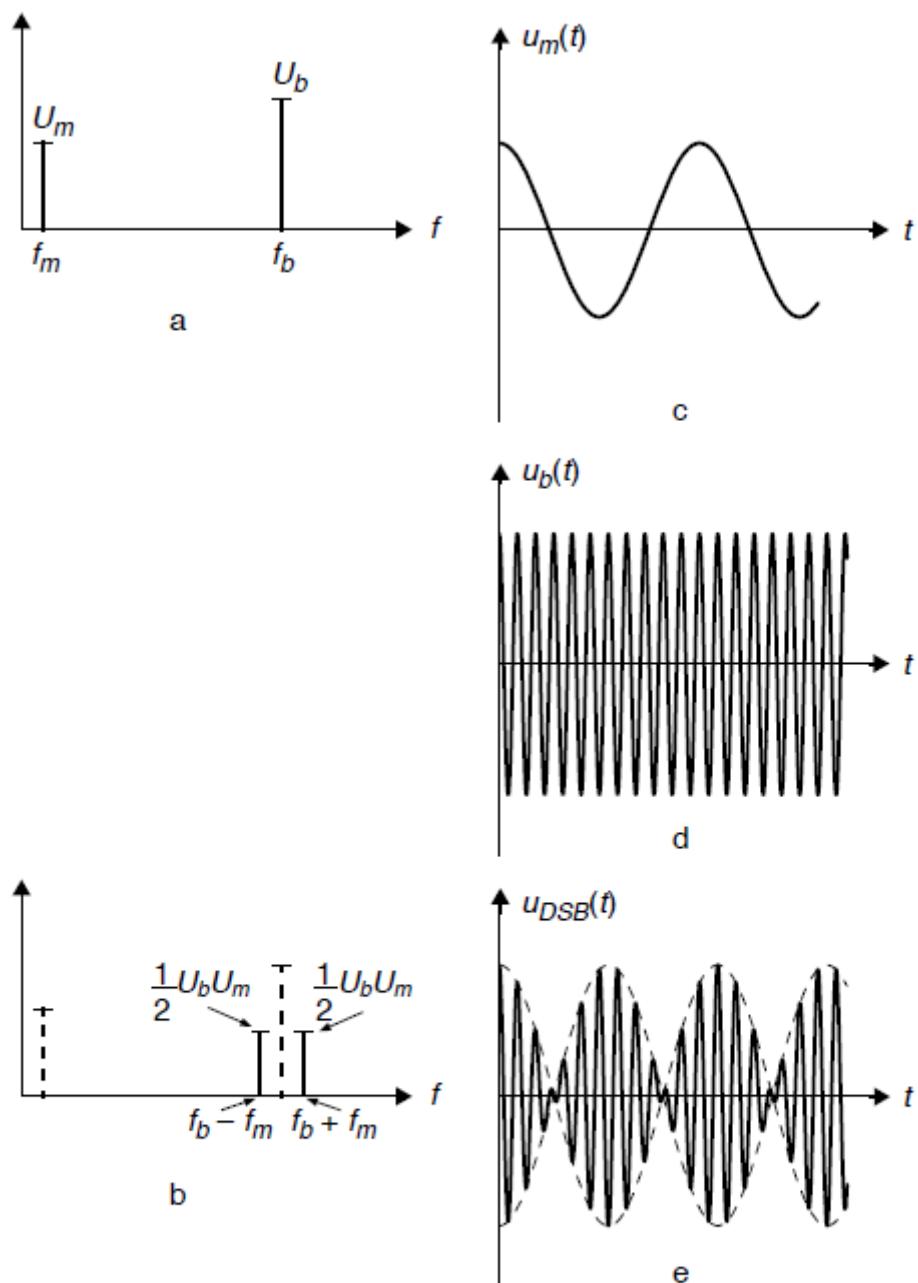
och en bär vågssignal, $u_b(t)$, med konstant amplitud = U_b och vinkel frekvens = $\omega_b = 2\pi f_b$:

$$u_b(t) = U_b \cos \omega_b t \quad (5.2)$$

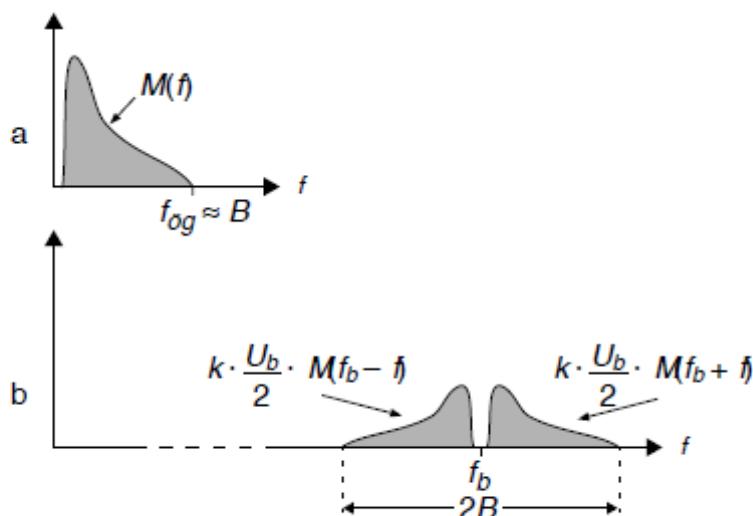
Den DSB-modulerade bär vågen, $u_{DSB}(t)$, får vi genom att multipli cera $u_m(t)$ med $u_b(t)$:

$$\begin{aligned} u_{DSB}(t) &= u_b(t)u_m(t) = U_b \cos \omega_b t \cdot U_m \cos \omega_m t = \\ &= \frac{U_b U_m}{2} [\cos(\omega_b + \omega_m)t + \cos(\omega_b - \omega_m)t] \end{aligned} \quad (5.3)$$

I Figur 5.2 a och b visas amplitudspektra och i Figur 5.2 c, d och f tidsdiagram för basbandssignalen, bär vågssignalen och den DSB-modulerade bär vågssignalen.



Figur 5.2 Amplitudspektra för a) modulerande basbandssignal U_m och b) modulerad bärvägssignal. Tidsfunktioner för c) modulerande basbandssignal $u_m(t)$, d) ommodulerad bärvägssignal $u_b(t)$, e) modulerad bärvägssignal $u_{DSB}(t) = u_b(t) \cdot u_m(t)$.



Figur 5.3 a) Basbandsspektrum, b) övre och undre sidbandsspektrum. Observera att i ett verkligt driftsfall tillkommer en förstärkningsfaktor k .

I ett allmänt fall låter vi basbandssignalen representeras av ett kontinuerligt spektrum $M(f)$ enligt Figur 5.3 a. Vid multiplikation med bärvägssignalen $u_b(t) = U_b \cos \omega_b t$ bildas ett övre sidband (eng. USB = Upper Side Band) och ett undre sidband (eng. LSB = Lower Side Band), se Figur 5.3 b, d.v.s. det bildas dubbla sidband varav benämningen DSB-modulation inses. Den modulerade signalen får genom de dubbla sidbanden en dubbelt så stor bandbredd som basbandssignalen.

Lägg märke till att det övre sidbandet är riktat åt samma håll som basbandsspektret medan det undre sidbandet är vänt i förhållande till basbandsspektret och att de båda sidbanden innehåller samma information.

8.1.3. Koherent detector

I mottagaren finns en demodulator eller detektor vars uppgift är att återbilda den mottagna DSB-signalen till ursprungssignalen = basbandssignalen.

Idealt, om kanalen inte inför någon dämpning eller fasvidning, kan vi sätta den mottagna signalen = den sända, d.v.s. = den modulerade signalen enligt uttrycket (5.1):

$$r(t) = m(t)U_b \cos \omega_b t \quad (5.4)$$

I detektorn multipliceras den mottagna signalen med en detektorsignal. Detektorsignalen utgörs av en koherent signal, $\cos \omega_b t$ (med samma fasläge och frekvens som bärvägen, här för enkelhets skull med amplituden = 1):

$$\begin{aligned} z(t) &= [m(t)U_b \cos \omega_b t] \cdot \cos \omega_b t = m(t)U_b \cos^2 \omega_b t = \\ &= \frac{U_b}{2} [m(t) + m(t) \cos 2\omega_b t] \end{aligned} \quad (5.5)$$

Den första termen i (5.5) är proportionell mot basbandssignalen medan den andra med dubbelt så hög frekvens som bärvägsfrekvensen lätt kan utfiltreras i basbandsfiltret så att vi på utgången av detta får $\frac{U_b}{2} m(t)$.

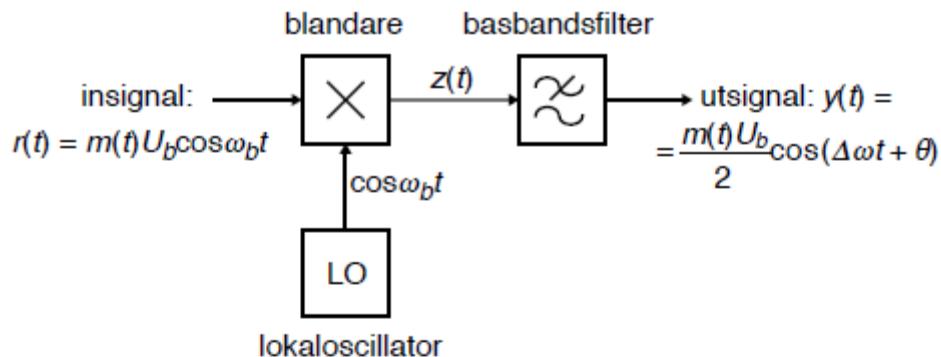
Hur stort felet blir i den detekterade signalen om detektorsignalens vinkel frekvens avviker från den riktiga med $\Delta\omega$ rad/s och fasläget avviker från det riktiga med θ radianer kan vi beräkna genom att multiplicera den mottagna signalen med (den felaktiga) detektorsignalen = $\cos[(\omega_b + \Delta\omega)t + \theta]$:

$$\begin{aligned} z_{\Delta\omega, \theta}(t) &= [m(t)U_b \cos \omega_b t] \cos(\omega_b t + \Delta\omega t + \theta) = \\ &= m(t)U_b \cos \omega_b t [\cos \omega_b t \cos(\Delta\omega t + \theta) - \sin \omega_b t \sin(\Delta\omega t + \theta)] = \\ &= m(t)U_b [\cos^2 \omega_b t \cos(\Delta\omega t + \theta) - \cos \omega_b t \sin \omega_b t \sin(\Delta\omega t + \theta)] = \\ &= \frac{m(t)U_b}{2} [(1 + \cos 2\omega_b t) \cos(\Delta\omega t + \theta) - \sin 2\omega_b t \sin(\Delta\omega t + \theta)] \end{aligned} \quad (5.6)$$

Efter basbandsfiltrering då frekvenskomponenter med dubbla bärvägsfrekvensen försvinner återstår:

$$y(t) = \frac{m(t)U_b}{2} \cos(\Delta\omega t + \theta) \quad (5.7)$$

se Figur 5.4.



Figur 5.4 Koherent detektor.

- Om den insatta detektorsignalen har rätt vinkelfrekvens , d.v.s. om $\Delta\omega = 0$ men fasfelet är θ blir den detekterade signalen

$$y_\theta(t) = \frac{m(t)U_b}{2} \cos\theta \quad (5.8)$$

vilket medför att den detekterade signalens amplitud avtar som $\cos\theta$ och därfor försvinner helt om $\theta = \pm \frac{\pi}{2}$.

- Om den insatta detektorsignalen har rätt fasläge, d.v.s. om $\theta = 0$ men vinkelfrekvensfelet är $\Delta\omega$ blir den detekterade signalen

$$y_{\Delta\omega}(t) = \frac{m(t)U_b}{2} \cos(\Delta\omega t) \quad (5.9)$$

vilket medför att den detekterade signalens amplitud varierar i takt med $\cos(\Delta\omega t)$.

Eftersom det i praktiken är omöjligt att i mottagaren skapa en detektorsignal med rätt vinkelfrekvens och fasläge skickar man tillsammans med DSB-signalen också en liten del av den omodulerade bärvägen, en pilotton. Pilottonen förstärks sedan i mottagaren innan den kan användas som detektorsignal. Modulationssorten kallas *dubbel sidband med undertryckt bärväg* (eng. DSBSC = Double Side Band Suppressed Carrier).

För äkta DSB (utan pilotton) finns det metoder att ur den mottagna signalen utvinna en detektorsignal. För sådana detektorer hänvisas till speciallitteraturen.

8.1.4. Amplitudmodulation, AM

Genom att tillsammans med de två sidbanden utsända en (odämpad) omodulerad bärvägssignal får man en amplitudmodulerad (AM) signal, som kan detekteras med en mycket enkel och billig (icke koherent) s.k. enveloppdetektor, se avsnittet **Detektering vid AM**. Vinsten med denna billiga och enkla detektor kompenseras mycket väl sändarens ganska låga verkningsgrad. Den omodulerade bärväg (utan informationsinnehåll) som också utsänds slukar nämligen mer än halva uteffekten.

Amplitudmodulation användes vid de första rundradiosändningarna på 1920-talet. Mottagaren använde en kristall som en enkel form av enveloppdetektor.

För att beskriva en amplitudmodulerad signal utgår vi från en DSB-modulerad signal, $u_{DSB}(t) = U_m \cos \omega_m t \cos \omega_b t$, med amplituden U_m och en bärvägssignal, $u_b(t) = U_b \cos \omega_b t$:

$$\begin{aligned} u_{AM}(t) &= u_{DSB}(t) + U_b \cos \omega_b t = U_m \cos \omega_m t \cos \omega_b t + U_b \cos \omega_b t = \\ &= U_b \left(1 + \frac{U_m}{U_b} \cos \omega_m t \right) \cos \omega_b t = U_b (1 + m \cos \omega_m t) \cos \omega_b t \end{aligned} \quad (5.10)$$

där

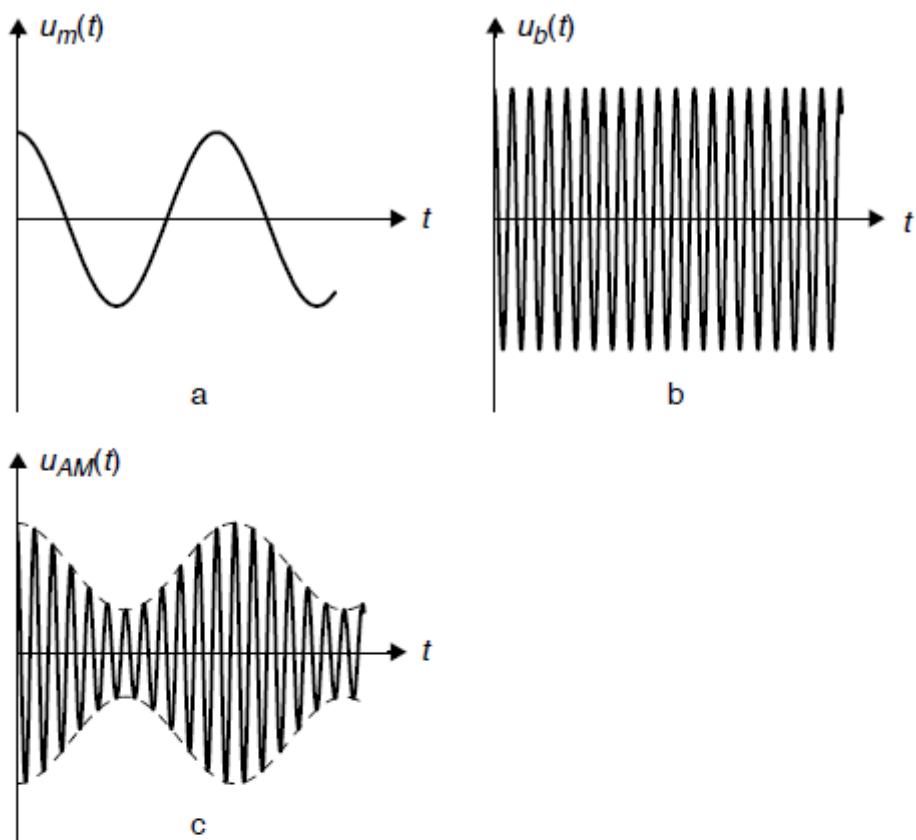
$$m = \frac{U_m}{U_b} \quad (5.11)$$

kallas modulationsgraden och

$$A(t) = U_b (1 + m \cos \omega_m t) \quad (5.12)$$

är amplitudfunktionen vars graf beskriver enveloppkurvan, se Figur 5.5 c.

Modulationsgraden är enligt (5.11) ett mått på förhållandet mellan den modulerande signalens amplitud, U_m , och den omodulerade bärvägens amplitud, U_b .



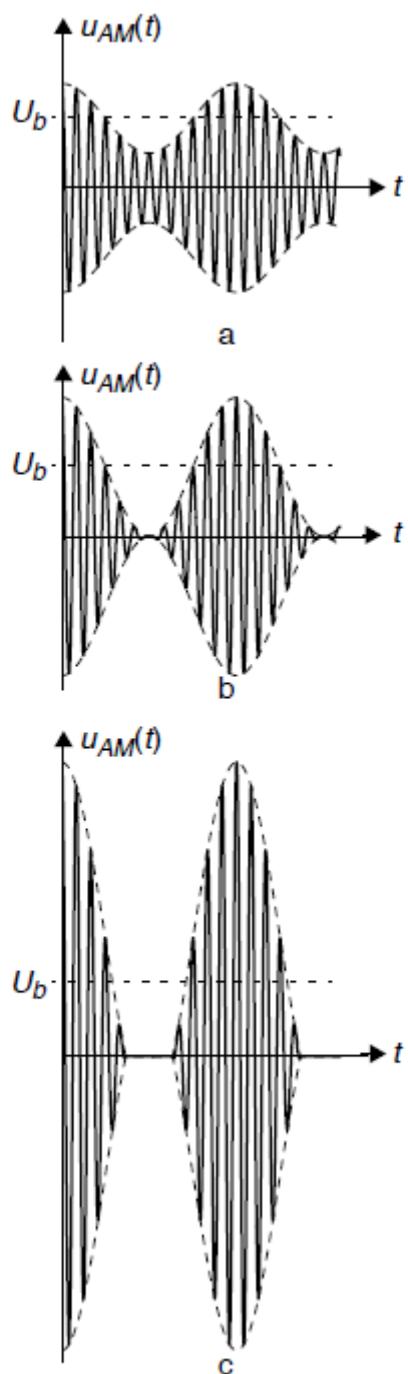
Figur 5.5
a) Modulerande basbandssignal $u_m(t)$, b) omodulerad bär- vågssignal $u_b(t)$, c) modulerad bärvägssignal $u_{AM}(t)$ med streckade enveloppkurvor.

Av Figur 5.5 c framgår att den positiva enveloppkurvan är likformig med den modulerande basbandssignalen och av Figur 5.6 c framgår att modulationsgraden, m , inte får överstiga 1 (eller 100%) för att enveloppkurvan inte skall distorderas p.g.a. övermodulering.

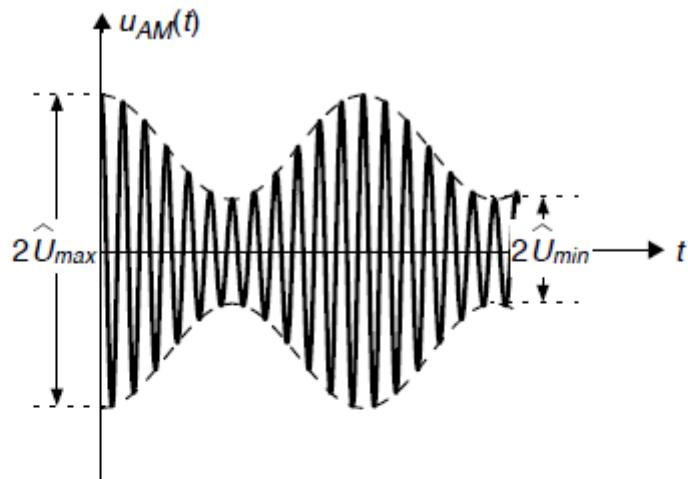
För att ur en oscilloscopebild kunna bestämma modulationsgraden avläser man enveloppens maximala topp-till-topp-värde = $2\hat{U}_{max}$ och enveloppens minimala topp-till-topp-värde = $2\hat{U}_{min}$ enligt Figur 5.7.

Med (5.12) får vi:

$$\hat{U}_{max} = U_b(1 + m) \quad (5.13)$$



Figur 5.6 AM-signal: a) $m < 1$, b) $m = 1$, c) $m > 1$.



Figur 5.7 Definition av $2\hat{U}_{max}$ och $2\hat{U}_{min}$.

och

$$\hat{U}_{min} = U_b(1 - m) \quad (5.14)$$

Med (5.13) och (5.14) kan vi teckna ett uttryck för beräkning av m :

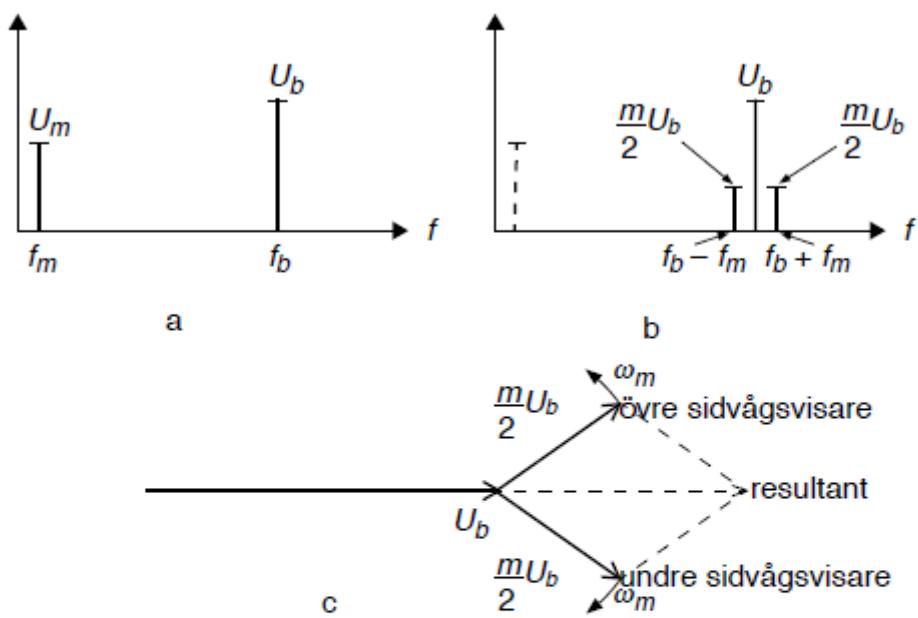
$$m = \frac{\hat{U}_{max} - \hat{U}_{min}}{\hat{U}_{max} + \hat{U}_{min}} \quad (5.15)$$

8.1.5. Spektrum för AM

För att erhålla ett AM-spektrum utvecklar vi (5.10) i sina frekvenskomponenter:

$$\begin{aligned} u_{AM}(t) &= U_b(1 + m\cos\omega_m t)\cos\omega_b t = \\ &= U_b(\cos\omega_b t + m\cos\omega_m t\cos\omega_b t) = \\ &= U_b\left(\cos\omega_b t + \frac{m}{2}\cos(\omega_b - \omega_m)t + \frac{m}{2}\cos(\omega_b + \omega_m)t\right) \end{aligned} \quad (5.16)$$

I Figur 5.8 a visas spektrum för den modulerande basbandssignalen och den omodulerade bärvägen medan Figur 5.8 b visar spektrum för AM-signalen. I Figur 5.8 c visas visardiagrammet för AM-signalen med bärvägsvisaren som riktfas.



Figur 5.8 Amplitudspektra för a) modulerande basbandssignal $u_m(t)$ och omodulerad bärvägssignal $u_b(t)$, b) AM-signal. Visardiaagram för c) AM.

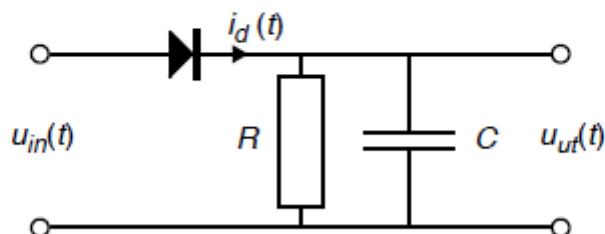
Sidvågsvisarna roterar med vinkelhastigheten ω_m rad/s mot varandra runt spetsen på bärvägssvisaren så att resultanten alltid har samma fasläge som bärvägssvisaren.

8.1.6. Detektering vid AM

För detektering av AM-signaler kan någon av de koherenta detektorer som beskrivits i samband med modulation med dubbelt sidband användas. Enklast och framför allt billigast används dock en (icke koherent) enveloppdetektor. Enveloppdetektorn lämnar en utsignal som är likformig med AM-signalens envelopp. Detektorer av denna typ används också som videodetektorer i televisionsmottagare och i toppvoltmetrar som registrerar den uppmätta spänningens toppvärde.

Figur 5.10 visar en enkel form av enveloppdetektor med den modulerade ingångsspänningen $u_{in}(t)$ och den detekterade utspänningen $u_{ut}(t)$. Valet av tidkonstanten RC är en kompromiss mellan mellanfrekvensrippel från mellanfrekvensförtäckaren och minimal distorsjon i den detekterade basbandssignalen.

Med en AM-signal ansluten till ingången visar Figur 5.11 spänningarna $u_{in}(t)$ och $u_{ut}(t)$ hos in- och utsignalerna samt diodströmmen $i_d(t)$. Dioden leder bara nära och vid toppen av varje period och lad-

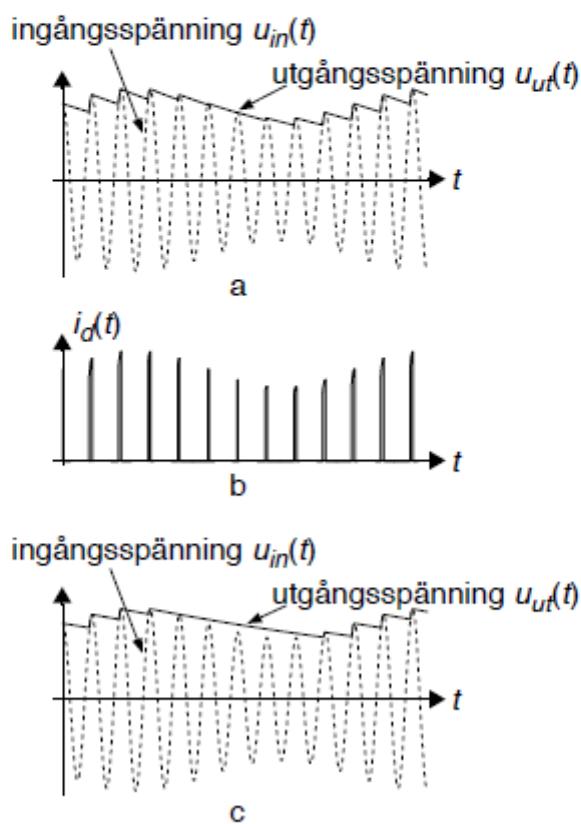


Figur 5.10 En enkel enveloppdetektor.

dar kondensatorn nästan till ingångsspanningens toppvärde. När ingångsspanningen avtar från sitt toppvärde blir dioden backspänd och kondensatorn börjar att ladda ur sig genom resistorn. Lutningen på den exponentialformade urladdningskurvan måste naturligtvis vara tillräckligt stor för att kunna följa ingångsspanningens nedåtgående envelopp, annars uppstår s.k. diagonalklippling (en slags distorsjon, se Figur 5.11 c).

För att bestämma optimal RC -tidkonstant utgår vi från AM-signalens envelopp

$$u_{env}(t) = U_b(1 + m \sin \omega_m t) \quad (5.20)$$



Figur 5.11 Vågformer: a) In- och utgångsspanningar vid optimal RC -tidkonstant, b) diodström vid optimal RC -tidkonstant, c) in- och utgångsspanningar vid för stor RC -tidkonstant ledande till diagonalklippling.

Lutningen, l_{env} , hos enveloppen får vi genom att tidsderivera uttrycket (5.20):

$$l_{env}(t) = \frac{d}{dt} u_{env}(t) = \omega_m U_b m \cos \omega_m t \quad (5.21)$$

Om ω_b är mycket större än ω_m ($\omega_b > 100\omega_m$), kommer punkten i varje bärvägsperiod, vid vilken dioden slutar leda ström och kondensatoren börja urladda, att vara en punkt på enveloppkurvan (d.v.s. på toppen av bärvägscykeln).

Låt starten på en urladdning börja vid tidpunkten t_0 , se Figur 5.12, då blir spänningen över RC -kretsen

$$u_{RC}(t - t_0) = u_{env}(t_0) \cdot e^{-(t-t_0)/RC} \quad (5.22)$$

Urladdningskurvans lutning får vi genom att derivera uttrycket (5.22) med avseende på $(t - t_0)$:

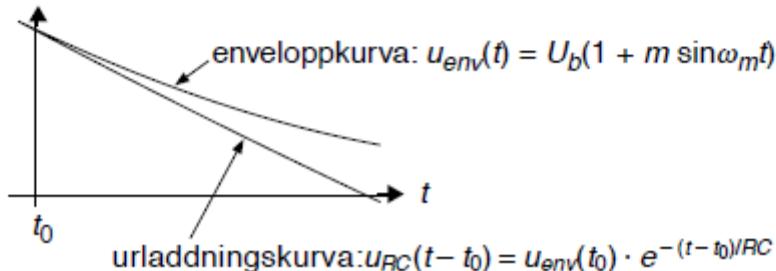
$$l_{RC}(t_0) = -\frac{u_{env}(t_0)}{RC} \quad (5.23)$$

För att utspänningen skall kunna följa enveloppkurvan måste urladdningskurvans lutning vara lika stor som eller mer negativ än enveloppkurvens lutning, d.v.s.

$$l_{RC}(t_0) \leq l_{env}(t_0) \quad (5.24)$$

som med (5.20), (5.21) och (5.23) insatt blir:

$$RC \leq \frac{1}{\omega_m} \cdot \frac{1 + m \sin \omega_m t_0}{m \cos \omega_m t_0} \quad (5.25)$$



Figur 5.12 Detaljbild av urladdningskurva och enveloppkurva vid enveloppdetektering.

Uttrycket (5.25) är nu en funktion av t_0 . "Svåraste fallet" uppstår när högerledet har minimum. Genom derivering uppstår minimet när $\sin\omega_m t_0 = -m$, som insatt i (5.25) ger kriteriet på RC -produkten:

$$RC \leq \frac{1}{\omega_m} \sqrt{\frac{1}{m^2} - 1} \quad (5.26)$$

Uttrycket (5.26) visar att för att satisfiera kriteriet skall RC -produkten väljas till noll när $m = 1$, vilket ju är omöjligt om kondensatorn skall kunna filtrera bort bärvägskomponenten. Man måste därför kompromissa och ta ett värde på RC -produkten som motsvarar en något lägre modulationsgrad och får då samtidigt godta en viss distorsion när modulationsgraden överstiger denna något lägre modulationsgrad.

8.1.7. Fasmodulation

Vid argumentmodulation kan antingen det momentana fasläget eller den momentana frekvensen hos bärvägssignalen bringas att variera i takt med basbandssignalen medan den modulerade bärvägssignalens amplitud hela tiden är konstant. Att den modulerade bärvägssignalen är konstant medför också att den utsända effektnivån är konstant - den är ju proportionell mot bärvägssignalens amplitud i kvadrat men ändå oberoende av dess frekvens och fasläge.

Den argumentmodulerade bärvägssignalens momentana spänning kan anges enligt

$$u_{arg}(t) = U_b \cos \Phi_b(t) \quad (5.33)$$

där $\Phi_b(t)$ är den modulerade bärvägssignalens momentana fasvinkel som kan skrivas

$$\Phi_b(t) = \omega_b t + \theta(t) \quad (5.34)$$

Vid fasmodulation är $\theta(t)$ [rad], som är fasens momentana avvikelse (deviation) från sitt omodulerade läge, proportionell mot den modulerande (= basbandssignalens) momentana spänning.

Utgå nämligen från en cosinusformad basbandssignal:

$$u_m(t) = U_m \cos \omega_m t \quad (5.35)$$

Fasens momentana deviation från sitt omodulerade läge blir då:

$$\theta(t) = k_\theta U_m \cos \omega_m t \quad (5.36)$$

där k_θ [rad/V] är *fasdeviationskonstanten*. Den maximala fasdeviationen, som helt enkelt kallas *fasdeviationen*, är alltså lika med $k_\theta \cdot U_m$ [rad]. Fasdeviationen kan också anges utan enhet och benämns då *modulationsindex* (i detta fall modulationsindex för fasmodulation) och betecknas m_p .

Vi kan alltså skriva:

$$m_p = k_\theta \cdot U_m \quad (5.37)$$

Det framgår av (5.37) att fasdeviationen eller modulationsindex beror av den modulerande signalen amplitud men är oberoende av dess frekvens och fas.

Det generella uttrycket för en argumentmodulerad bärvägssignal enligt (5.33) kan alltså för fallet fasmodulation med en cosinusformad modulationssignal enligt (5.35) skrivas:

$$u_{PM}(t) = U_b \cos(\omega_b t + m_p \cos \omega_m t) \quad (5.38)$$

Den momentana fasvinkeln är:

$$\Phi_b(t) = \omega_b t + m_p \cos \omega_m t \quad (5.39)$$

och den momentana vinkelfrekvensen:

$$\omega_{PM}(t) = \frac{d\Phi_{PM}(t)}{dt} = \omega_b + m_p \omega_m (-\sin \omega_m t) \quad (5.40)$$

8.1.8. Frekvensmodulation

Vid frekvensmodulation är det bärvägfrekvensens momentana avvikelse (deviation) från sitt omodulerade värde f_b som är proportionell mot den modulerande signalens momentana spänning.

För att till att börja med förstå vad som menas med *momentan frekvensavvikelse* utgår vi från uttrycket (5.34):

$$\Phi_b(t) = \omega_b t + \theta(t) \quad (5.34)$$

som ju är den modulerade bärvägssignalens momentana fas. Genom att tidsderivera detta samband får vi ett uttryck som anger bärvägssignalens momentana vinkelfrekvens:

$$\omega(t) = \frac{d\Phi_b(t)}{dt} = \omega_b + \frac{d\theta(t)}{dt} \quad (5.41)$$

Den momentana frekvensen blir då:

$$f(t) = \frac{1}{2\pi} \cdot \frac{d\Phi_b(t)}{dt} = f_b + \frac{1}{2\pi} \cdot \frac{d\theta(t)}{dt} \quad (5.42)$$

Bärvägsfrekvensens momentana avvikelse från sitt omodulerade läge definieras på följande sätt:

$$\delta f(t) = f(t) - f_b = \frac{1}{2\pi} \cdot \frac{d\theta(t)}{dt} \quad (5.43)$$

Utgå nu från en modulerande basbandssignal av cosinusform:

$$u_m(t) = U_m \cos \omega_m t \quad (5.44)$$

Enligt definitionen av frekvensmodulation blir då:

$$\delta f(t) = \frac{1}{2\pi} \cdot \frac{d\theta(t)}{dt} = k_f U_m \cos \omega_m t \quad (5.45)$$

där k_f [Hz/V] är *frekvensdeviationskonstanten*. Den maximala frekvensdeviationen kallas helt enkelt *frekvensdeviationen* och är alltså lika med $k_f \cdot U_m$ [Hz] och betecknas:

$$\Delta f = k_f U_m \quad (5.46)$$

Genom att tidsintegrera sambandet (5.45) kan vi få fram den momenta fasdeviationen vid frekvensmodulation:

$$\theta_{FM}(t) = 2\pi \int_0^t k_f U_m \cos(\omega_m \tau) d\tau \quad (5.47)$$

Den frekvensmodulerade bärvägssignalen kan då skrivas:

$$\begin{aligned} u_{FM}(t) &= U_b \cos\left(\omega_b t + 2\pi \int_0^t k_f U_m \cos(\omega_m \tau) d\tau\right) = \\ &= U_b \cos\left(\omega_b t + \frac{2\pi k_f U_m}{\omega_m} \sin \omega_m t\right) = \\ &= U_b \cos\left(\omega_b t + \frac{k_f U_m}{f_m} \sin \omega_m t\right) = \\ &= U_b \cos(\omega_b t + m_f \sin \omega_m t) \end{aligned} \quad (5.48)$$

Av (5.48) framgår dock att m_f till skillnad mot m_p inte bara beror av den modulerande signalens amplitud utan också av dess frekvens. Lägg också märke till att m_f och m_p endast är definierade för modulation med rena sinussignaler.

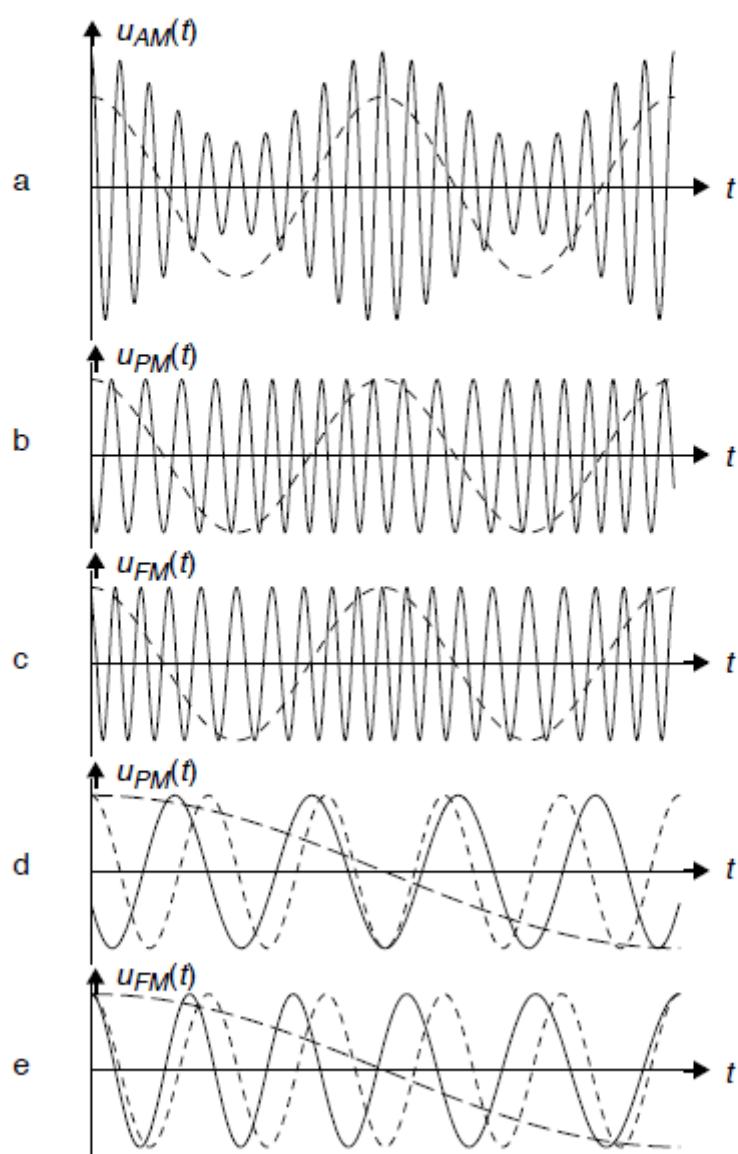
8.1.9. Översikt av argumentmodulation

Nedan följer en sammanställning av olika modulationssamband för PM och FM när den modulerande basbandssignalen är

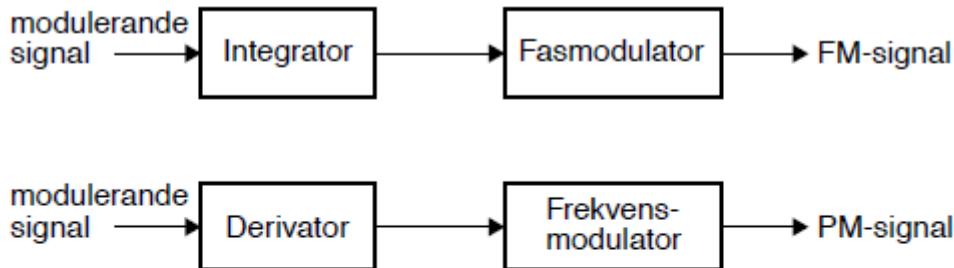
- 1) cosinusformad: $u_m(t) = U_m \cos \omega_m t$
- 2) godtycklig: $u_m(t) = m(t)$

PM	FM
<i>Omodulerad bärvägssignal</i>	
1) $U_b \cos \omega_b t$	1) $U_b \cos \omega_b t$
<i>Modulerande basbandssignal</i>	
1) $U_m \cos \omega_m t$	2) $U_m \cos \omega_m t$
2) $m(t)$	2) $m(t)$
<i>Modulerad bärvägssignal</i>	
1) $U_b \cos(\omega_b t + m_p \cos \omega_m t)$	1) $U_b \cos(\omega_b t + m_f \sin \omega_m t)$
2) $U_b \cos(\omega_b t + k_\theta m(t))$	2) $U_b \cos(\omega_b t + 2\pi \int_0^t k_f m(\tau) d\tau)$
<i>Modulationsindex (= fasdeviation)</i>	
$m_p = k_\theta U_m$	$m_f = \frac{k_f U_m}{f_m} = \frac{\Delta f}{f_m}$
<i>Momentan fas</i>	
1) $\omega_b t + m_p \cos \omega_m t$	1) $\omega_b t + m_f \sin \omega_m t$
2) $\omega_b t + k_\theta \cdot m(t)$	2) $\omega_b t + 2\pi \int_0^t k_f m(\tau) d\tau$
<i>Momentan vinkelfrekvens</i>	
1) $\omega_b + m_p \omega_m (-\sin \omega_m t)$	1) $\omega_b + m_f \omega_m \cos \omega_m t$
2) $\omega_b + k_\theta \frac{dm(t)}{dt}$	2) $\omega_b + 2\pi k_f m(t)$
<i>Momentan frekvens</i>	
1) $f_b + m_p f_m (-\sin \omega_m t)$	1) $f_b + m_f \cos \omega_m t$
2) $f_b + \frac{k_\theta}{2\pi} \cdot \frac{dm(t)}{dt}$	2) $f_b + k_f m(t)$

I Figur 5.15 a, b och c visas hur den modulerade bärvägssignalen ser ut vid AM, PM och FM. Av samma figur och av sammanställningen på föregående sida framgår att det finns stora likheter mellan PM och FM. En viktig skillnad är dock att vid FM är den momentana fasdeviationen proportionell mot modulationssignalens tidsintegral medan den vid PM är direkt proportionell, se Figur 5.15 d och e, som visar en detaljbild av hur fasavvikelsen varierar vid PM respektive FM.



Figur 5.15 a) AM, b) PM, c) FM, d) PM detaljbild, e) FM detaljbild.



Figur 5.16 Generering av FM-signal med fasmodulator och PM-signal med frekvensmodulator.

Det är också relativt enkelt att åstadkomma en FM-signal med en fasmodulator och vice versa som framgår av Figur 5.16.

8.1.10. Spektrum för argumentmodulation

En vanlig tumregel, kallad Carsons regel (eng. Carson's rule), för bandbredden hos en argumentmodulerad signal med godtyckligt modulationsindex m_θ är:

$$B_{\text{rad/s}} = 2(\Delta\omega + \omega_m) = 2\omega_m(m_\theta + 1) \quad (5.65)$$

Genom att byta $\Delta\omega$ mot Δf och $\Delta\omega_m$ mot Δf_m kan vi uttrycka bandbredden i Hertz i stället för i radianer/s:

$$B_{\text{Hertz}} = 2(\Delta f + f_m) = 2f_m(m_\theta + 1) \quad (5.66)$$

Bandbredds begränsning enligt Carsons regel skär högst bort 2% av den totala effekten.

8.1.11. Superheterodyn mottagaren

När man konstruerar en radiomottagare börjar man vanligtvis med detektorn och arbetar i båda riktningarna, d.v.s. mot ingångssteget och lågfrekvensutgången. De flesta detektorer fungerar inte bra om signalbrusförhållandet är för lågt. För att fungera kräver de dessutom en minsta spänningssnivå, i vissa fall av storleksordningen volt.

Den önskade stationens fältstyrka vid mottagarantennen kanske uppgår till några $\mu\text{V}/\text{m}$ medan andra icke önskade stationer samtidigt genererar fältstyrkor av storleksordningen V/m . Slutsatsen av detta är att det före detektorn behövs både förstärkning och filtrering.

Grovt sett kan vi strukturera svårigheterna att konstruera den högfrekventa bärvägsförstärkaren före detektorn för en fix-frekvensmottagare i följande tre punkter:

- att kunna kontrollera brusnivån i ingångssteget så att signalbrus-förhållandet blir acceptabelt,
- att kunna kontrollera olinjäriteter hos de aktiva elementen för att hålla nere distorsion och intermodulationsdistorsion (två eller flera frekvensmässigt närbelägna signaler modulerar varandra om de passerar ett olinjärt element),
- att förhindra att den högförstärkande smalbandiga förstärkaren börjar oscillera (anta t.ex. att spänningsförstärkningen är 10^6 eller 120 dB, då bli förstärkaren självsvängande om bara en miljondel av den förstärkta signalen återkopplas till förstärkarens ingång i fas med den ordinarie ingångssignalen).

Nu tillkommer den svårigheten att de flesta mottagare inte är avsedda för fix-frekvensmottaganing. Det betyder att man, förutom att vara tvungen att behärska de ovannämnda problemen också måste kunna variera avstämningssenhetens mittfrekvens inom ett frekvensområde på mer än 100 till 1. Ganska tidigt i radions utveckling insåg man att dessa problem var omöjliga att praktiskt lösa på ett tillfredsställande sätt.

För att förenkla situationen började några forskare att angripa problemen på ett nytt sätt. Man beslöt att även i fortsättningen ha en

fix-frekvens högfrekvensförstärkare före detektorn, som alltså också fick arbeta vid en fix frekvens. Antennsignalens frekvens, vare sig den var 1 eller 100 MHz, frekvenstransponerades till denna fixa frekvens, mellanfrekvensen (MF), med hjälp av en ny anordning: *blandaren*.

Den nya tekniken medförde ett genombrott för AM-radiotekniken. Mottagartypen fullkomnades 1918 av Edwin H. Armstrong och fick namnet superheterodyn-mottagaren. Sedan dess har superheterodyn-metoden dominerat marknaden och trängt ut andra s.k. raka mottagartyper, som man nuförtiden nästan bara kan hitta på museér.

Som de flesta andra uppfinningar skapar också denna en del nya problem förutom de den löser. Några av dessa nya problem är:

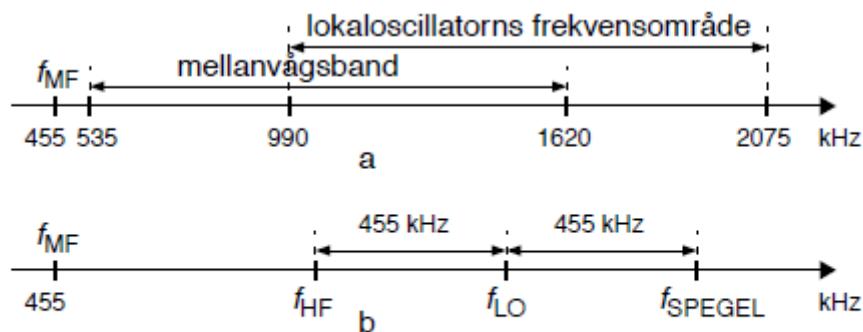
- blandaren och lokaloscillatorm (LO), som skall avstämmas till övriga avstämningseenheter före blandaren, måste också konstrueras,
- blandare alstrar mer brus än förstärkare; för att blandaren över huvud taget skall fungera som frekvenstransponator måste den innehålla olinjäriteter.

Anm.: Dessa båda nya problem påverkar också förmågan att kontrollera signalbrus-förhållandet och intermodulationsdistorionen vilket gör att det kan bli nödvändigt att lägga in fler HF-steg före blandaren.

- Lokaloscillatorm bildar inte bara *en* signal av mellanfrekvens utan också en massa andra oönskade frekvenser.

8.1.12. Spiegelfrekvenser

När en mottagare avstäms för mottagning av en bestämd station betyder det att antennkretsen avstäms till den önskade radiosignalens frekvens, f_{HF} , medan lokaloscillatoren samtidigt ställs in på frekvensen $f_{HF} + f_{MF}$ ¹, se Figur 7.4 a. Det främsta skälet för ett avstämmt HF-steg är att undertrycka spiegelfrekvenser. Dessas uppkomst förklaras av att det i antennen, förutom den önskade signalfrekvensen, $f_{HF} = f_{LO} - f_{MF}$, också finns signaler i frekvensområdet $f_{LO} + f_{MF}$. Båda dessa signaler bildar efter blandning med lokaloscillatorsignalen en signal med frekvensen f_{MF} , se Figur 7.4 b.



Figur 7.4 a) Lokaloscillatorns frekvensområde för AM-superheterodyn-mottagare, b) spiegelfrekvensbegreppet.

8.2. Digital modulation

Skrives senare.

Referenser

- [1] Wikipedia, "Signal," 20 juli 2021. [Online]. Available: <https://en.wikipedia.org/wiki/Signal>. [Använt 10 augusti 2021].
- [2] Wikipedia, "A/D-omvandlare," 19 september 2020. [Online]. Available: <https://sv.wikipedia.org/wiki/A/D-omvandlare>. [Använt 12 augusti 2021].
- [3] Wikipedia, "Jämna och udda funktioner," 16 mars 2020. [Online]. Available: https://sv.wikipedia.org/wiki/J%C3%A4mna_och_udda_funktioner. [Använt 10 augusti 2021].
- [4] Wikipedia, "Fourierserie," 04 december 2020. [Online]. Available: <https://sv.wikipedia.org/wiki/Fourierserie>. [Använt 10 augusti 2020].
- [5] Wikipedia, "Eulers formel," 16 december 2020. [Online]. Available: https://sv.wikipedia.org/wiki/Eulers_formel. [Använt 11 augusti 2021].
- [6] Wikipedia, "Fouriertransform," 25 mars 2021. [Online]. Available: <https://sv.wikipedia.org/wiki/Fouriertransform>. [Använt 17 augusti 2021].
- [7] Wikipedia, "Discrete-time Fourier transform," 6 februari 2022. [Online]. Available: https://en.wikipedia.org/wiki/Discrete-time_Fourier_transform. [Använt 1 januair 2023].
- [8] Wikipedia, "Discrete Fourier transform," 28 december 2022. [Online]. Available: https://en.wikipedia.org/wiki/Discrete_Fourier_transform. [Använt 1 januari 2023].
- [9] Wikipedia, "Fast Fourier transform," 31 december 2022. [Online]. Available: https://en.wikipedia.org/wiki/Fast_Fourier_transform. [Använt 2 januari 2023].
- [10] Wikipedia, "Nyquist–Shannon sampling theorem," 23 december 2022. [Online]. Available: https://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem. [Använt 31 december 2022].
- [11] Wikipedia, "Nyquistfrekvensen," 1 november 2020. [Online]. Available: <https://sv.wikipedia.org/wiki/Nyquistfrekvensen>. [Använt 31 december 2022].
- [12] Wikipedia, "Brus," 10 oktober 2019. [Online]. Available: <https://sv.wikipedia.org/wiki/Brus>. [Använt 12 augusti 2021].
- [13] Wikipedia, "Noise (signal processing)," 21 juli 2021. [Online]. Available: [https://en.wikipedia.org/wiki/Noise_\(signal_processing\)](https://en.wikipedia.org/wiki/Noise_(signal_processing)). [Använt 12 augusti 2021].



- [14] Wikipedia, "Johnson–Nyquist noise," 12 april 2021. [Online]. Available: https://en.wikipedia.org/wiki/Johnson%E2%80%93Nyquist_noise. [Använd 12 augusti 2021].
- [15] Wikipedia, "Phase noise," 12 maj 2021. [Online]. Available: https://en.wikipedia.org/wiki/Phase_noise. [Använd 12 augusti 2021].
- [16] Wikipedia, "Mains hum," 29 april 2021. [Online]. Available: https://en.wikipedia.org/wiki/Mains_hum. [Använd 12 augusti 2021].
- [17] Wikipedia, "Shot noise," 23 juli 2021. [Online]. Available: https://en.wikipedia.org/wiki/Shot_noise. [Använd 12 augusti 2021].
- [18] Wikipedia, "Flicker noise," 25 december 2020. [Online]. Available: https://en.wikipedia.org/wiki/Flicker_noise. [Använd 29 augusti 2021].
- [19] Wikipedia, "Burst noise," 02 oktober 2020. [Online]. Available: https://en.wikipedia.org/wiki/Burst_noise. [Använd 29 augusti 2021].
- [20] Wikipedia, "Kosmiskt brus," 28 januari 2016. [Online]. Available: https://sv.wikipedia.org/wiki/Kosmiskt_brus. [Använd 12 augusti 2021].
- [21] Wikipedia, "Kosmisk bakgrundsstrålning," 21 mars 2021. [Online]. Available: https://sv.wikipedia.org/wiki/Kosmisk_bakgrundsstr%C3%A5lning. [Använd 12 augusti 2021].
- [22] Wikipedia, "Atmosfäriskt brus," 29 januari 2021. [Online]. Available: https://sv.wikipedia.org/wiki/Atmosf%C3%A4riskt_brus. [Använd 12 augusti 2021].
- [23] Wikipedia, "Effektspektrum," 10 november 2013. [Online]. Available: <https://sv.wikipedia.org/wiki/Effektspektrum>. [Använd 13 augusti 2021].
- [24] Wikipedia, "Noise spectral density," 10 mars 2021. [Online]. Available: https://en.wikipedia.org/wiki/Noise_spectral_density. [Använd 13 augusti 2021].
- [25] Wikipedia, "Colors of noise," 10 maj 2021. [Online]. Available: https://en.wikipedia.org/wiki/Colors_of_noise. [Använd 13 augusti 2021].
- [26] C. Frank, Telekommunikation: Informationsöverföring & Överföringssystem., Knivsta, Sverige: Bokförlaget Hackspetten, 2014.
- [27] Wikipedia, "Signal-brusförhållande," 27 december 2020. [Online]. Available: <https://sv.wikipedia.org/wiki/Signal-brusf%C3%B6rh%C3%A5llande>. [Använd 12 augusti 2021].



- [28] Wikipedia, "Signal-to-interference-plus-noise ratio," 04 juni 2021. [Online]. Available: https://en.wikipedia.org/wiki/Signal-to-interference-plus-noise_ratio. [Använd 12 augusti 2021].
- [29] Wikipedia, "Antenna noise temperature," 26 december 2020. [Online]. Available: https://en.wikipedia.org/wiki/Antenna_noise_temperature. [Använd 29 augusti 2021].
- [30] Wikipedia, "Noise Figure," 24 februari 2020. [Online]. Available: https://sv.wikipedia.org/wiki/Noise_Figure. [Använd 15 augusti 2021].
- [31] Wikipedia, "Transmission line," 27 augusti 2021. [Online]. Available: https://en.wikipedia.org/wiki/Transmission_line. [Använd 28 augusti 2021].
- [32] Wikipedia, "Waveguide," 01 augusti 2021. [Online]. Available: <https://en.wikipedia.org/wiki/Waveguide>. [Använd 29 augusti 2021].
- [33] Wikipedia, "Balanced line," 02 september 2021. [Online]. Available: https://en.wikipedia.org/wiki/Balanced_line. [Använd 05 september 2021].
- [34] Wikipedia, "Planar Transmission line," 27 augusti 2021. [Online]. Available: https://en.wikipedia.org/wiki/Planar_transmission_line. [Använd 05 september 2021].
- [35] Wikipedia, "Telegrapher's equations," 27 augusti 2021. [Online]. Available: https://en.wikipedia.org/wiki/Telegrapher%27s_equations. [Använd 29 augusti 2021].
- [36] Wikipedia, "Standing wave ratio," 17 maj 2021. [Online]. Available: https://en.wikipedia.org/wiki/Standing_wave_ratio. [Använd 26 september 2021].
- [37] Wikipedia, "Elektromagnetiskt spektrum," 20 maj 2021. [Online]. Available: https://sv.wikipedia.org/wiki/Elektromagnetiskt_spektrum. [Använd 26 september 2021].
- [38] Wikipedia, "Maxwells ekvationer," 4 december 2020. [Online]. Available: https://sv.wikipedia.org/wiki/Maxwells_ekvationer. [Använd 11 augusti 2022].
- [39] Wikipedia, "Wave impedance," 17 april 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wave_impedance. [Använd 12 augusti 2022].
- [40] Wikipedia, "Polarization (waves)," 24 oktober 2022. [Online]. Available: [https://en.wikipedia.org/wiki/Polarization_\(waves\)](https://en.wikipedia.org/wiki/Polarization_(waves)). [Använd 4 januari 2023].
- [41] Youtube, "Circular polarization," 27 februari 2012. [Online]. Available: <https://www.youtube.com/watch?v=Fu-aYnRkUgg>. [Använd 4 januari 2023].



- [42] Wikipedia, "Poyntings vektor," 18 juli 2021. [Online]. Available: https://sv.wikipedia.org/wiki/Poyntings_vektor. [Använd 12 augusti 2022].
- [43] Wikipedia, "Dielectric," 20 juli 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Dielectric>. [Använd 12 augusti 2022].
- [44] Wikipedia, "Relative permittivity," 20 juli 2022. [Online]. Available: https://en.wikipedia.org/wiki/Relative_permittivity. [Använd 12 augusti 2022].
- [45] Wikipedia, "Brytningsindex," 2021 november 2022. [Online]. Available: <https://sv.wikipedia.org/wiki/Brytningsindex>. [Använd 5 januari 2023].
- [46] Wikipedia, "Skinneffekten," 3 maj 2019. [Online]. Available: <https://sv.wikipedia.org/wiki/Skinneffekt>. [Använd 5 januari 2023].
- [47] Wikipedia, "Friis transmission equation," 24 oktober 2022. [Online]. Available: https://en.wikipedia.org/wiki/Friis_transmission_equation. [Använd 5 januari 2023].
- [48] Wikipedia, "Directivity," 10 augusti 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Directivity>. [Använd 22 juni 2022].
- [49] Wikipedia, "Diversity combining," 4 december 2021. [Online]. Available: https://en.wikipedia.org/wiki/Diversity_combining. [Använd 4 januari 2023].
- [50] Wikipedia, "MIMO," 4 januari 2023. [Online]. Available: <https://en.wikipedia.org/wiki/MIMO>. [Använd 4 januari 2023].
- [51] Wikipedia, "Monopole antenna," 23 juni 2021. [Online]. Available: https://en.wikipedia.org/wiki/Monopole_antenna. [Använd 10 oktober 2021].
- [52] Wikipedia, "Dipole antenna," 11 juli 2022. [Online]. Available: https://en.wikipedia.org/wiki/Dipole_antenna#Folded_dipole. [Använd 12 augusti 2022].
- [53] Wikipedia, "Loop antenna," 3 april 2022. [Online]. Available: https://en.wikipedia.org/wiki/Loop_antenna. [Använd 12 augusti 2022].
- [54] Wikipedia, "Reflector (antenna)," 6 juni 2018. [Online]. Available: [https://en.wikipedia.org/wiki/Reflector_\(antenna\)](https://en.wikipedia.org/wiki/Reflector_(antenna)). [Använd 12 augusti 2022].
- [55] Wikipedia, "Parabolic antenna," 6 juni 2022. [Online]. Available: https://en.wikipedia.org/wiki/Parabolic_antenna. [Använd 12 augusti 2022].
- [56] Wikipedia, "Horn antenna," 11 april 2022. [Online]. Available: https://en.wikipedia.org/wiki/Horn_antenna. [Använd 12 augusti 2022].

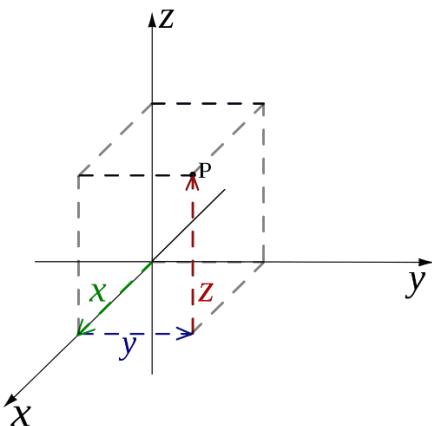


- [57] Wikipedia, "Patch antenna," 21 maj 2022. [Online]. Available: https://en.wikipedia.org/wiki/Patch_antenna. [Använd 2 januari 2023].
- [58] Wikipedia, "Microstrip antenna," 22 november 2022. [Online]. Available: https://en.wikipedia.org/wiki/Microstrip_antenna. [Använd 2 januari 2023].
- [59] H. M. A. Bergström, "Daily air temperature and pressure series for Uppsala (1722-1998)," *Climate Change*, vol. 53, nr April 2022, pp. 213-252, 2002.
- [60] Wikipedia, "Aliasing," 16 december 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Aliasing>. [Använd 1 januari 2023].
- [61] Wikipedia, "Anti-aliasing filter," 24 maj 2022. [Online]. Available: https://en.wikipedia.org/wiki/Anti-aliasing_filter. [Använd 1 januari 2023].
- [62] Wikipedia, "Z-transform," 2 maj 2021. [Online]. Available: <https://sv.wikipedia.org/wiki/Z-transform>. [Använd 2 januari 2023].
- [63] Wikipedia, "Convolution," 3 januari 2023. [Online]. Available: <https://en.wikipedia.org/wiki/Convolution>. [Använd 4 januari 2023].
- [64] Wikipedia, "Finite impulse response," 28 januari 2023. [Online]. Available: https://en.wikipedia.org/wiki/Finite_impulse_response. [Använd 10 oktober 2022].
- [65] Wikipedia, "Koordinatsystem," 14 januari 2021. [Online]. Available: <https://sv.wikipedia.org/wiki/Koordinatsystem>. [Använd 08 augusti 2022].
- [66] Wikipedia, "Spherical coordinate system," 1 augusti 2022. [Online]. Available: https://en.wikipedia.org/wiki/Spherical_coordinate_system. [Använd 10 augusti 2022].

Appendix A – Tredimensionella koordinatsystem

För att hantera koordinater (punkter), vektorer, ytor mm i ett tredimensionellt rum används olika typer av tredimensionella koordinatsystem [65]. Vanligaste varianterna är kartesiska (rektagulära), cylindriska och sfäriska koordinater.

Figur 70 visar exempel på det kartesiska koordinatsystemet. Detta består av tre ortogonala axlar (ofta betecknade x , y och z). Absolut vanligast är att använda ett s.k. högerorienterat system. Namnet högerorienterat kommer från att axlarnas ordning fås genom att spreta med och inbördes vinkla högerhandens tumme (x), pekfinger (y) och långfinger (z). Dessa ger då axlarnas riktning i förhållande till varandra.

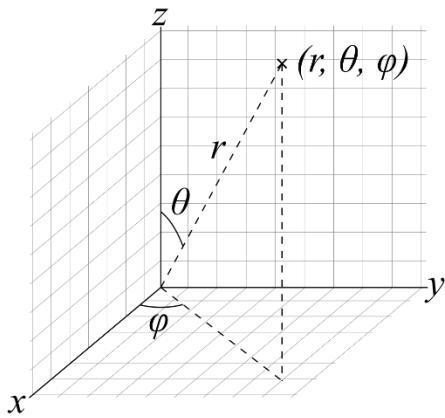


Figur 70 Kartesiskt, högerorienterat koordinatsystem.
By Cronholm144 - Own work, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=2277021>

Läget för punkten P i det tredimensionella rummet (se Figur 70) anges med punktens x -, y - respektive z -koordinat. Dessa tre tal är helt enkelt punktens projicering på de tre axlarna. Denna taltrippel skrivs ofta på formen (x, y, z) . Punkten $(0, 0, 0)$ kallas för koordinatsystemets origo. Jämför med hur en punkt anges med sina x - och y -koordinater i ett vanligt tvådimensionellt, kartesiskt koordinatsystem!

I många situationer så är det mer praktiskt att ange en koordinat genom att tala om vilket avstånd och riktning den har i förhållande till koordinatsystemets origo, se Figur 71. Figuren visar hur avståndet r samt de två vinkelarna θ, φ kan användas för att ange en punkts koordinater. För dessa tre sfäriska koordinater gäller att $r \geq 0$, $0^\circ \leq \theta \leq 180^\circ$ och $0^\circ \leq \varphi < 360^\circ$ [66].

Avståndet r kallas ofta för radie och de två vinkelarna θ, φ har olika benämningar inom olika områden inom matematik och fysik. I denna bok används sfäriska koordinater inom antennteorin där benämningen för θ är elevation och φ är azimut. Förutom att det finns andra benämningar på θ, φ finns det också andra definitioner av utsträckningen av vinkelarna θ och φ ! För att ytterligare spä på förvirringen brukar man inom matematiken kasta om beteckningarna på vinkelarna. Det gäller således att kontrollera definitioner, beteckningar och namnkonventionen noga när man behandlar sfäriska koordinatsystem.



Figur 71

Sfäriska koordinater.

By Andeggs - Own work, Public Domain,

<https://commons.wikimedia.org/w/index.php?curid=7478049>

De rektangulära koordinaterna kan med trigonometri beräknas från sfäriska koordinater som

$$x = r \sin \theta \cos \varphi$$

$$y = r \sin \theta \sin \varphi$$

$$z = r \cos \theta$$

Ekvation 166

och de sfäriska koordinaterna beräknas från de rektangulära som

$$r = \sqrt{x^2 + y^2 + z^2}$$

$$\theta = \begin{cases} \tan^{-1}\left(\frac{\sqrt{x^2 + y^2}}{z}\right) & , \text{om } z > 0 \\ 90^\circ & , \text{om } z = 0 \text{ och både } x, y \neq 0 \\ \tan^{-1}\left(\frac{\sqrt{x^2 + y^2}}{z}\right) + 180^\circ & , \text{om } z < 0 \text{ och både } x, y \neq 0 \\ 180^\circ & , \text{om } x = 0, y = 0 \text{ och } z < 0 \\ \text{odefinierad} & , \text{om } x = 0, y = 0 \text{ och } z = 0 \end{cases}$$

Ekvation 167

$$\varphi = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & , \text{om } x > 0 \\ \tan^{-1}\left(\frac{y}{x}\right) + 180^\circ & , \text{om } x < 0 \text{ och } y \geq 0 \\ \tan^{-1}\left(\frac{y}{x}\right) - 180^\circ & , \text{om } x < 0 \text{ och } y < 0 \\ 90^\circ & , \text{om } x = 0 \text{ och } y > 0 \\ -90^\circ & , \text{om } x = 0 \text{ och } y < 0 \\ \text{odefinierad} & , \text{om } x = 0 \text{ och } y = 0 \end{cases}$$