

PROSJEKTARBEID

FY1013 ELEKTRISITET OG MAGNETISME II

NOVEMBER 2005

- *En fysisk betraktning av nervecellen*
Mariann Suh Jøssang, Reidun Sletten, Miriam Bøgh Meling, Nina Bjørk Arnfinnsdottir
- *Batteriladeren*
Gunnar Skjervold, Magnus Nordling, Magnus Berg Johnsen
- *Filterkretser*
Jon Alm Eriksen, Lodve Brastad
- *Elektromagnetiske bølger*
Morten Ingvaldsen, Katrine Alne, Anfrid Dahlen
- *Spektroskopi*
Ingebjørg Berglie, Lene Grøterud Leer
- *Superledere*
Torill Lekve, Håvard Myge, Ole Henrik von Munthe af Morgenstjerne
- *Bygging av DC-motor med utredning av de underliggende fysiske prinsipper*
Lars Erik Klemetsen, Henrik Sivertsen, Jørgen Tande
- *Vindkraften utnyttes*
Olav Andreas Sæther, Nils Bjarne Gunnufsen
- *Energibånd i faste stoffer med anvendelser*
Einar Stiansen, Eivind Tore Hansen, Morten Klungervik, Anders Langseth
- *Dioder*
Astrid-Sofie Vardøy, Christina Aas, Audun Eskeland Rimehaug, Are Opstad Sæbø
- *Datamaskinen - Historie, oppbygning og virkemåte*
Andreas Wahl, Patrick Slåttnes, Simen Eidsmo Barder

En fysisk betraktning av nervecellen

Laget av:

Mariann Suh Jøssang

Reidun Sletten

Miriam Bøgh Meling

Nina Bjørk Arnfinnsdottir

FY 1013 – Elektrisitet og Magnetisme

Abstrakt

Vårt nevralt nettverk er et veldig komplekst system. I hovedsak er det elektriske impulser som forplanter seg gjennom nerveceller og viderefører informasjon. Dette kan veldig forenklet beskrives ved modeller av hvordan nervecellen er oppbygd og dets bestanddeler oppgaver. Selv om dette er et organisk system kan vi beskrive nervecellens funksjon ved hjelp av enkle fysiske modeller.

Nervecellen har et overskudd av negative ioner som gjør den mer negativ enn væsken den befinner seg i. Dette medfører en potensialforskjell over cellemembranen. Membranpotensialet endres når nervecellen mottar elektriske impulser, og når denne endringen er stor nok utløses et aksjonspotensial som gjør at signalet sendes videre.

Når aksjonspotensialet har oppstått, vil nervecellene sende informasjonen videre med enten elektrisk eller kjemisk overføring. Ved hjelp av denne kjemiske og elektriske overføringen kan de elektriske impulsene forplante seg videre gjennom nervebaner slik at informasjonen blir overført til det relevante stedet. En reaksjon i samsvar med påvirkningen kan da inntreffe.

Hele dette systemet tilpasset oss slik at vi kan fungere og reagere optimalt.

Innhold

Abstrakt	s2
Innholdsfortegnelse	s3
Forord	s4
1. Innledning	s6
1.1 Nervesystemet.....	s6
1.2 Nerveceller.....	s6
2. Membranpotensialet	s9
2.1 Kort om cellemembranen.....	s9
2.2 Hvilepotensialet – det konstante membranpotensialet.....	s10
2.3 Nernst likningen.....	s13
2.4 Cellemembranen som en elektrisk kondensator.....	s15
3. Aksjonspotensialet	s19
3.1 Innledende redegjørelse av et akson, sammenligning med en krets.....	s19
3.2 Aksjonspotensial.....	s21
3.3 Antall ioner.....	s23
3.4 Impuls eller ikke impuls.....	s23
3.5 Impulsens forplantning.....	s24
3.6 Hvor fort skulle vi vente at signalet i et akson dør ut? Tilbake til kretsen.....	s26
4. Den voltstengte natrium- og kaliumkanalen	s30
4.1 Voltstengt natrium kanal	s30.
4.2 Voltstengt kalium kanal	s32
5 Hvordan nervecellene kommuniserer - Synapsetransmisjon	s34
5.1 Elektrisk synapse.....	s34
5.2 Signaloverføring i kjemisk synapse.....	s36
5.3 Litt om nevrontransmittere.....	s38
5.4 Transmittor reseptorer.....	s38
Sammendrag	s40
Etterord	s41
Vedlegg	s43
Referanseliste	s45

Forord

Fysikken har lenge vært en vitenskap som holdt seg til å beskrive ikke-levende objekter. Ikke bare fordi det har vært og fortsatt er vanskelig å studere de komplekse prosessene i levende systemer; en har rett og slett ikke vært sikker på om de fysiske lovene en hadde var gyldige for levende organismer. Denne tvilen gjaldt da ikke på det makroskopiske nivå; det er temmelig opplagt at en hest vil falle med like stor sikkerhet som en stein om man kaster dem begge ut fra et hustak. Problemet gjaldt de mer grunnleggende mekanismene på cellenivå og under. Levende vesener utviser helt andre typer egenskaper enn de objektene en hadde studert for å utlede de fysiske lovene. De har det med å vokse og utvikle seg, reprodusere seg, for etter hvert å forfalle og dø. Dette var en viktig bakgrunn for at forskere helt frem til et stykke ut i det 19. århundre diskuterte muligheten for at levende organismer var styrt helt andre typer lover enn de man til da hadde kjent til. Selv opphavet til organiske molekyler var det uvisshet om, og det var mange som mente at de bare kunne settes sammen ved hjelp av en mystisk livskraft.

Dette synet ble først utfordret da Friedrich Wöhler i 1828 lyktes i å sette sammen grunnstoffet urea fra uorganiske materialer. Etter dette gikk utviklingen raskt fremover, men forløpet av det er en annen historie. I dag anvendes fysiske lover til å beskrive levende vesener på alle nivåer. Å utvide fysikken til områder som tidligere har vært forbeholdt biologi, har vist seg å være veldig fruktbart; en har nådd kunnskap om organismers oppbygning og funksjon som kanskje ikke ville vært mulig på noen annen måte. Fortsatt er det store områder innen dette fagfeltet hvor det mangler nøyaktig viten om hva som skjer og hvorfor. Til nå har en imidlertid ikke funnet et eneste område innen det biologiske fagfeltet hvor fysikkens lover ikke kan benyttes.

Hvordan nervene i kroppen kommuniserer med hverandre er et godt eksempel på et felt fysisk forskning har bidratt mye til å kaste lys over; i en slik grad at feltet gjerne omtales med et eget navn: nevrofysikk. Denne oppgaven ønsker å vise noen av sidene ved dette. Hovedformålet har vært å ta opp noe av grunnlaget for at det kan gå en elektrisk impuls gjennom en nervecelle, samt forløpet av impulsen. Der det har vært mulig er det gitt

eksempler på hvordan fysiske lover er brukt til å beskrive fenomener i nervecellene. Nevrofysikk er imidlertid et så stort og komplekst fagfelt at vi har funnet det nødvendig å presentere en del grunnleggende fakta om nervesystemet for å få en helhetlig sammenheng.

De verdier en har for spenning, tiden det tar for en impuls å gå gjennom cellen, og flere andre forhold som gjelder nervesystemet er resultater av direkte målinger på nerveceller, ikke utregninger fra konstante verdier. Fordi det er snakk om en så liten skala er det vanskelig å få måleresultatene nøyaktige. Verdiene som oppgis varierer derfor til en viss grad fra bok til bok. Vi har valgt verdiene vi benytter i oppgaven på grunnlag av hvilke verdier som går igjen hyppigst, men har ingen garantier for at disse er de mest riktige. Til denne oppgaven vil de likevel være tilstrekkelige.

Begrepene hvilepotensial, potensial og spenningsforskjell brukes til en viss grad om hverandre teksten. Så lenge det ikke uttrykkes bestemt at det er aksjonspotensialet det handler om, vil begrepene vise til potentialet en finner ved å måle spenningen på innsiden av en celle i forhold til spenningen på utsiden. Utsiden vil alltid regnes som nullpunkt.

1 Innledning

1.1 Nervesystemet [1,2,3]

Det er vanlig å dele nervesystemet i to deler: sentralnervesystemet og det perifere nervesystemet. Sentralnervesystemet består av hjernen og ryggraden, det perifere nervesystemet består av serier av nerver som går parvis ut fra ryggmargen og hjernestammen.

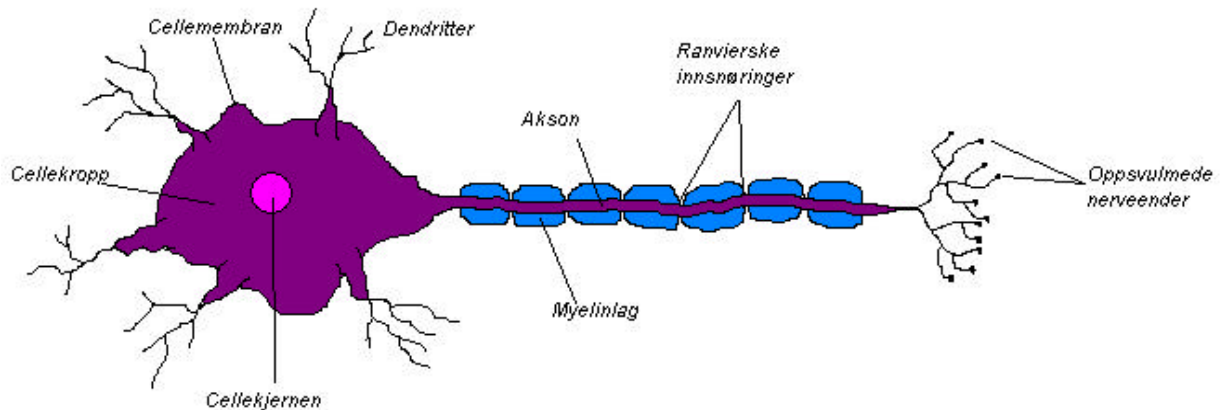
Sentralnervesystemet består av en grå og en hvit substans. Den grå substansen er nervecellenes kropp og deres korte utløpere (dendritter). Den hvite substansen består av nervefibrene som er nervecellenes lange utløpere (aksoner). I både den grå og hvite massen finnes det støtteceller, glialceller, som utgjør ca 50 % av nervesystemets volum. Glialcellene verken sender eller mottar nerveimpulser, men de omgir nervecellene og holder dem på plass, samtidig som de gir nervecellene næring og renses bort giftstoffer.

Det vi vanligvis kaller nerver er bunter av aksoner som er omgitt av glialceller og fibret bindevev. Nervene, som forgrenes ut fra sentralnervesystemet, blir tynnere og er i kontakt med kroppens sanseceller, muskelceller og kjertelceller. De gjør kroppen i stand til å fungere som den skal.

1.2 Nerveceller [1,2,3]

I en menneskehjerne finnes det flere hundre milliarder nerveceller. Og disse danner ca 10^{10} sammenkoblinger. Nervecellene (nevroner) består som andre celler av en cellemembran og cytoplasma med organeller, inkludert kjernen. Cytoplasmaet som også kalles celsaften består av vann, elektrolytter (elektrisk ladde molekyler og ioner) og næringsstoffer (se tabell x på side x). Cellemembranen er en barriere mellom cytoplasma og den ekstracellulære væsken og gjør cellen i stand til å oppføre seg som en selvstendig del av kroppen. Membranen regulerer også transport av stoffer inn og ut av cellen. Den styrer altså strømmen av ioner inn og ut av cellen, og da også spenningsforskjellen mellom cellens inn- og utside. Dermed styrer cellemembranen sendingen av nerveimpulser.

Nervene cellene deles inn i grupper etter antall utløpere som går ut fra cellekroppen, og også innenfor disse gruppene finnes det store forskjeller. I denne oppgaven vil vi derimot fokusere på fellestrekkene til alle disse gruppene.



Figur 1. En skjematisk fremstilling av en nervecelle.

Cellekroppen, med cellens kjerne i midten, holder nervecellen i live. Dendrittene som kommer ut av cellekroppen er fiber som virker som antenner som får beskjeder fra andre nerveceller.

Aksonet er spesialisert for å lede de elektriske nerveimpulsene. I mennesker er aksonet alt fra noen få millimeter til over 1m langt. Aksonet er en god elektrisk leder i den betydning at det fungerer utmerket i samspill med resten av nervesystemet. Sammenlignet med for eksempel metallet kobber er aksonet derimot en dårlig leder. Stor motstand i væsken inne i aksonet gjør at ledningsevnen blir dårlig, opptil 100 millioner ganger dårligere enn ledningsevnen til kobber. Hvordan kroppen har løst dette problemet kommer vi tilbake til senere i oppgaven.

Aksonet er hos noen av nervecellene er omgitt av et fettete hvitaktig myelinlag. Denne myelinkjeden dannes på fosterstadiet ved at glialceller kveiler seg rundt aksonet. Med 1-2mm avstand oppstår det innsnevring av myelinkjeden, såkalte ranvierske innsnøringer der aksonmembranen ligger naken i 1-2 μ m. Disse innsnøringene er viktige for ledningen

av impulser langs aksonet. Skader på myelinlaget vil ødelegge timingen for nerveimpulsene.

På enden av aksonet forgrener det seg og ender i oppsvulmede nerveender. Disse oppsvulmede nerveendene sender signaler til ulike celler i kroppen.

Nerveimpulsene som går fra nervecelle til nervecelle gjennom hele kroppen danner et intrikat nettverk. Nervecellene er imidlertid aldri i direkte kontakt med hverandre. De smale mellomrommene mellom nervecellene der impulsen videreføres fra nerveendene til dendrittene kalles synapser. Her sendes impulsen videre på en av to måter; via en elektrisk eller en kjemisk synapse. De elektriske synapsene kan sammenlignes med en tennplugg i en bil, her hopper impulsen over det smale mellomrommet og føres videre. De kjemiske synapsene er mer kompliserte, her er det spesielle transmitterstoffer (nevrotransmittere) som overfører informasjon til den andre cellen.

2. Membranpotensialet [4,7,10]

2.1 Kort om cellemembranen [4]

Nervecellen består av store deler væske, denne kalles ICF – ”intracellular fluid”. Væsken som omgir nervecellen kalles ECF – ”extracellular fluid”. Celleveggen som skiller disse to typer væskene kalles cellemembranen. Stoffene som er oppløst i ECF og ICF er både av organisk og uorganisk opprinnelse, men komposisjonen innen de to gruppene er forskjellig.

	Konsentrasjon i ICF (mM)	Konsentrasjon i ECF (mM)	Mulig å krysse cellemembranen? (mM)
K ⁺	125	5	Ja
Na ⁺	12	120	Nei* (vil komme tilbake til dette)
Cl ⁻	5	125	Ja
A ⁻	108	0	Nei
H ₂ O	55000	55000	Ja

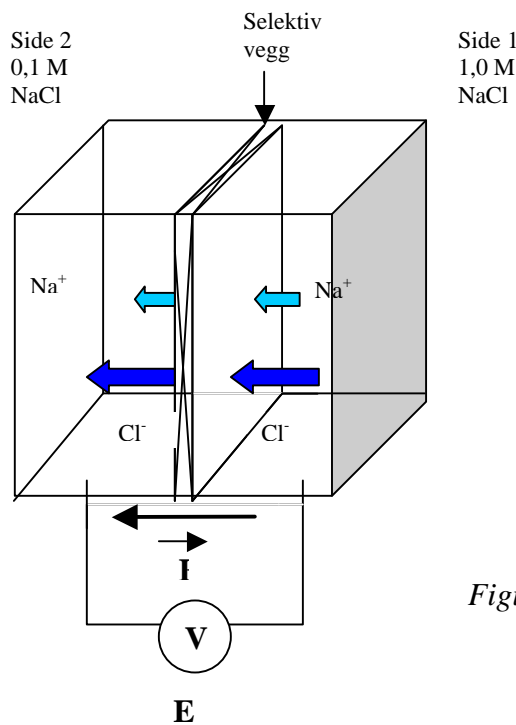
Tabell1. Komposisjonene som er vist i tabell 1 er forenklet, ved å kun inkludere de stoffene som er viktige med hensyn til de basiske osmose- og elektriske egenskapene til cellen. Her er A- en samlbetegnelse på diverse anioner. Benevningen mM=mol/mm³

Cellemembranen fungerer som en porøs selektiv skillevegg. Den lar noen stoffer passere, men forhindrer andre. Denne egenskapen skyldes proteinmolekylene som membranen delvis er bygget opp av. Noen proteinmolekyler er knyttet til den indre eller ytre overflaten av membranen, mens andre penetrerer gjennom membranen og skaper en bro mellom innsiden og utsiden. Det er disse broene som utgjør ionekanalene. Ionekanalene er åpne for noen stoffer og lukket for andre – selektive. De selektive egenskapene er varierende for forskjellige typer nerveceller. Denne reguleringen av ionetransporten, gjennom membranen, muliggjør en spenningsforskjell mellom cellens inn- og utside, og dermed også sending av nerveimpulser.

2.2 Hvilepotensialet – det konstante membranpotensialet [4]

Membranpotensialet er en potensialforskjell mellom innsiden og utsiden av en nervecelle. Vi skal nå konsentrere oss om hvordan det konstante membranpotensialet, også kalt hvilepotensialet, oppstår. Hvilepotensialet vil variere for forskjellige typer nerveceller, i og med at det avhenger av hvilke ionekanaler som er konstant åpne. For en generell nervecelle sier man at det ligger på ca -70mV . -70mV kommer av et vist negativt overskudd av anioner på innsiden av cellemembranen. På utsiden av membranen er potensialet definert lik null. For å forstå hvordan dette potensialet oppstår skal vi begynne med en enkel fiktiv modell, for så å nærme oss virkeligheten ved å gjøre den mer kompleks.

Vi starter ved å sammenligne nervecellen med en boks, som er delt inn i to avdelinger. La oss si, hypotetisk sett, ytre del og indre del av cellen. For cellen fungerer membranen som en selektiv skillevegg, derfor må vi ha en porøs, selektiv vegg i boksen også, som skiller avdeling 1 fra 2.

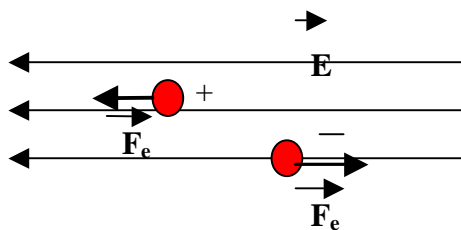


Figur2

Figur 2. En fiktiv modell av en celle, med forskjellig konsentrasjon av saltet NaCl på hver side av en porøs, selektiv barriere. Størrelsen $M(\text{molar}) = \text{mol}/\text{dm}^3$

Som vi ser fra figuren er det forskjellig mengde løsning av saltet NaCl i de to avdelingene. Vi lar i første omgang den selektive skilleveggen være permeabel for både Na⁺, Cl⁻ og vann.

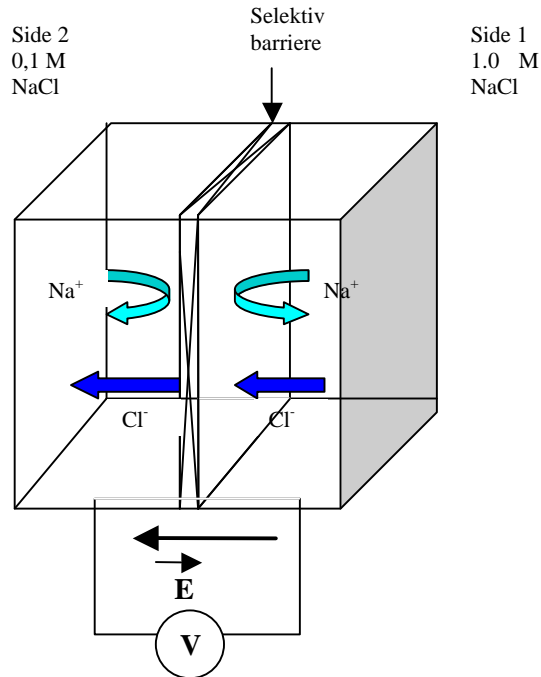
Na⁺ og Cl⁻ vil starte å bevege seg fra side 1 til side 2, som følge av konsentrasjonsforskjellene i de to avdelingene. Denne diffusjonen vil opphøre når konsentrasjonen av NaCl er lik i begge avdelingene. Før likevekten innstiller seg oppstår det imidlertid et interessant fenomen. En viktig faktor å ta i betraktning er at i løsninger, vil Na⁺ og Cl⁻ ikke bevege seg med samme hastighet. Cl⁻ er mer mobil og vil bevege seg fra side 1 til side 2 raskere enn Na⁺. Dette fører til at vi får et overskudd av negative ioner på side 2, før likevekten har innstilt seg. Det settes opp et midlertidig E-felt fra side 1 til side 2. Den elektriske kraften, F_e , vil da virke på henholdsvis positive og negative partikler slik:



Figur 3. Hvordan elektriske krefter virker på henholdsvis positive og negative partikler i et elektrisk felt, E .

Den elektriske kraften vil altså redusere farten til kloridionene og øke farten til Natriumionene. Spenningsforskjellen som voltmeteret viser nå, se figur 1, kalles likevektspotensialet. Dette potensialet vil bygge seg opp til de to ionene beveger seg med samme hastighet. Likevekt oppnås når det er like store mengder med NaCl på begge sider av boksen. Da vil konsentrasjonskraften som er opphavet til den elektriske kraften opphøre, og diffusjonspotensialet forsvinne. Dette stemmer for så vidt dårlig med en

virkelig celle. I en reel celle er det et konstant potensiale - hvilepotensialet som i innledningen er gitt til å være ca -70mV. Vi må gjøre noen endringer. I denne nye modellen lar vi kun Cl⁻ få krysse den porøse veggen.



Figur 4. Samme boks, som i figur 1, delt inn i to avdelinger vha en porøs barriere.

Situasjonen blir enda at kloridionene vil bevege seg fra høyre mot venstre, men nå vil det bygge seg raskere opp et negativt overskudd på venstre side. Likevekt vil nås når det er like store mengder med kloridioner på begge sider. Dette skjer når de elektriske kreftene, som driver Cl⁻ fra side 2 til side 1, balanserer konsentrasjonsgradienten som driver Cl⁻ motsatt vei, slik at vi har en netto bevegelse av kloridioner gjennom barrieren lik 0.

$$\Sigma F_e = \Sigma F_k \quad (2.2.1)$$

Jeg har i (2.2.1) valgt å bruke et kraftbegrep på konsentrasjonsgradienten, i og med at virkningen av gradienten på den ioniske løsningen vil utarte seg som en slags sum av krefter på kloridionene. Det er en upresis definisjon, men hensiktsmessig for å gi et klart bilde på hva som skjer.

I dette tilfellet vil altså ikke de elektriske kreftene forsvinne i og med at konsentrasjonskreftene aldri slutter å opphøre. Dette vil igjen si at likevektspotensialet holder seg konstant når likevekten har innstilt seg – vi har et hvilepotensiale!

2.3 Nernst likningen [4,7]

Hvilepotensialet, vi kom frem til i figur 4 kan også kalles *Nernst potensialet*, V_x , til ionet membranen er permeabel for, i vårt tilfelle Cl^- .

$$V_{\text{Cl}} = (RT/ZF) \cdot \ln([\text{Cl}^-]_1/[\text{Cl}^-]_2), \quad (2.3.1)$$

Her er R gass konstanten, T er temperaturen, Z er ladningen til det permatimle ionet (-1 for klor), F er Faraday's konstant (96500C/mol), $[\text{Cl}^-]_1$ og $[\text{Cl}^-]_2$ er konsentrasjonen av klorid i henholdsvis avdeling 1 og 2. Setter man inn tallsvar for disse størrelsene får man at $V_{\text{Cl}} = -58\text{mV}$.

Hadde Na^+ vært det aktuelle, ville vi hatt V_{Na} . Merk at likningen kun gjelder for et ion om gangen.

Selve likningen kommer vi frem til ved å betrakte likevektssituasjonen. Ved dette tidspunktet må vi ha at netto strømning av ioner gjennom membranen som er lik null. Man kan her trekke en klar parallell til pn-ledere. For i likhet med celle-membranen har p-n-ledere en diffusjonsstrøm som skyldes konsentrasjonsgradienten, og en driftsstrøm som skyldes de elektriske kreftene. Ved likevekt må disse strømmene være like store og motsatt rettet:

$$I(\text{diffusjon}) + I(\text{drift}) = 0 \quad (2.3.2)$$

Vi kan uttrykke diffusjonsstrømmen slik:

$$I(\text{diffusjon}) = A * Z * F * \phi \quad (2.3.3)$$

Her er ϕ = fluks av ioner, Z ladningen til det aktuelle ionet og F er faraday's konstant på $F = 96500 \text{ C/mol}$. FZ omformer fluksen av ioner til fluks av ladning – elektrisk strøm. A er arealet ionene strømmer gjennom.

Fluksen for et gitt ion, Y -ionet, kan beskrives ved:

$$\phi = D_y(dC/dx) \quad (2.3.4)$$

Her er $D_y = uRT$, som er diffusjonskonstanten. Denne bestemmes av u =mobiliteten til ionet i membranen, R =gasskonstanten og T =temperaturen. dC/dx er rett og slett konsentrasjonsgradienten, der variablene varierer med tykkelsen på membranen.

Driftsstrømmen kan uttrykkes ved lik argumentasjon:

$$I(\text{drift}) = A * Z * F * \phi \quad (2.3.5)$$

Fluksen for en ladet partikkel kan uttrykkes ved spenningsgradienten, $\nabla V = dV/dx$:

$$\phi = u * Z * F * C * (dV/dx) \quad (2.3.6)$$

Faktoren $Z * F * C$ er konsentrasjonen av ladning i en bestemt posisjon. Vi ser en tydelig parallell til Gauss lov:

$$j = \sigma * E = \sigma * (-\nabla V) \quad (2.3.7)$$

Her er j =strømtetthet og σ = ladningstetthet

Kombinerer nå (2.3.6) med (2.3.5) og (2.3.4) med (2.3.3). Resultatet puttes inn i (2.3.2):

$$uAZF(RTdC/dx + ZFCdV/dx) = 0$$

Løser differensial ligningen:

$$(RTdC/dx = (- ZFCdV/dx)$$

$$(- RT/ZF)(dC/C) = dV$$

$$(- RT/ZF) \int_{[Y]_u}^{[Y]_i} dC/C = \int_{V_u}^{V_i} dV$$

Firkantklammene markerer konsentrasjonen av Y-ioner u = utenfor membranen, i = innenfor membranen. V er henholdsvis spenningen på innsiden og utsiden.

Løsningen blir da:

$$RT/ZF (\ln[Y]_i - [Y]_u) = V_i - V_u$$

Eller

$$RT/ZF (\ln [Y]_u / [Y]_i) = V_i - V_u = \Delta V \quad (2.3.8)$$

Her er ΔV = Nernstpotensialet, eller hvilepotensialet

Nernst - eller hvilepotensialet for figur 4 ble tidligere oppgitt til å bli -58mV, ved bruk av Nernst ligningen. Ved utregning av dette er det imidlertid ikke tatt hensyn til at startkonsentrasjonene på 1,0M og 0,1M for $[Cl^-]_1$ og $[Cl^-]_2$ vil forandres når vi oppnår likevekt – Burde man ikke ta hensyn til at dette? Svaret er nei.

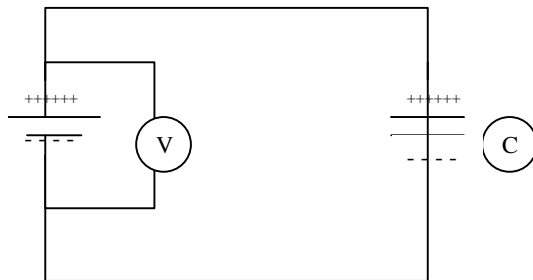
2.4 Cellemembranen som en elektrisk kondensator [4]

Slik situasjonen utarter seg i figur 4, skal det ikke mange klorid-ioner til før vi registrerer en potensialforskjell mellom avdeling 1 og 2. For å finne ut nøyaktig hvor mange

kloridioner som kreves for å få hvilepotensialet på -58mV , kan vi sammenligne cellemembranen med en elektrisk kondensator.

En kondensator er en ladningslagrer som består av to ledende plater, separert fra hverandre med et isolerende medium. I figur 4 blir de to saltløsningene i avdeling 1 og 2 de ledende ”platene” og den porøse veggen imellom den isolerende delen. For en reell celle blir ICF og ECF de ledende ”platene”, og deler av cellemembranen den isolerende barrieren.

Når en kapasitans kobles opp mot et batteri som vist i figur 5, vil spenningen fra batteriet sørge for at elektronene blir fjernet fra den ene ledende platen og oppsamlet på den andre – kondensatoren lades opp.



Figur 5. Elektrisk krets som inneholder en kapasitans.

Når kondensatoren er fullt oppladet vil spenningen over kondensatoren være lik spenningen fra batteriet. Da vil de to platene ha like store og motsatte elektriske ladninger, $+Q$ og $-Q$. Kapasitansen C er et mål på kondensatorens evne til å holde på ladning. Mengden ladning, Q , på hver av konduktorplatene er definert som:

$$Q = C \cdot \Delta V \quad (2.4.1)$$

Kondensatorens kapasitans er dessuten direkte proporsjonal med arealet av konduktorplatene og invers proporsjonal med distansen mellom de to platene, $C = (\epsilon_0 * A)/d$. Dette finner vi på følgende måte:

Det elektriske feltet, E , mellom platene kommer fra Gauss' lov:

$$E = \sigma/\epsilon_0 \quad (2.4.2)$$

Vi antar to parallelle flater som er uniformt ladet med motsatt og like store ladninger. Her er σ = flateladningstettheten og ϵ_0 = permittiviteten til luft (vakum). Potensialforskjellen mellom platene er dermed:

$$\Delta V = E*d = Q/\epsilon_0*d/A \quad (2.4.3)$$

Der d = plateavstanden og A = arealet av platene. Dermed følger det at kapasitansen, $C = Q/\Delta V$, er

$$C = (\epsilon_0 * A)/d \quad (2.4.4)$$

Kapasitansen avhenger også av det isolerende mediet mellom platene. Den biologiske membranen's kapasitans er regnet ut til å være 10^6 F. Med denne verdien, kan vi regne ut ladningsmengden som trengs for å oppnå et hvilepotensialet på $V_{Cl} = -58$ mV:

$$Q = C*\Delta V$$

$$Q = 10^6\text{F}*(-58\text{mV}) = 5,8*10^{-8}\text{C}$$

For å finne ut hva denne ladningsmengden svarer til i antall kloridioner, må vi omforme svaret fra coulomb av ladning til mol av ioner. Dette gjøres ved å dividere svaret vårt med Faraday's konstant = 10^5 C/mol:

$$5,8 \cdot 10^{-8} \text{ C} / 10^5 \text{ C/mol} = 5,8 \cdot 10^{-13} \text{ mol}$$

Når 1 mol = $6,022 \cdot 10^{23}$, får vi:

$$5,8 \cdot 10^{-13} \text{ mol} \cdot 6,022 \cdot 10^{23} = 3,49 \cdot 10^{11}$$

Vi trenger altså $3,49 \cdot 10^{11}$ kloridioner for å oppnå likevektspotensialet på -58mV. Hvis volumet av hver avdeling i figur 2 inneholder 1 ml, vil side 2 i figur 2 inneholde ca $6 \cdot 10^{20}$ klor- og natriumioner. Dette gir:

$$((3,49 \cdot 10^{11}) / (6 \cdot 10^{20})) \cdot 100\% = 5,82 \cdot 10^{-8}\%$$

Vi ser at kun $5,82 \cdot 10^{-8}\%$ av de opprinnelige kloridionene trenger å forflyttes for å få et hvilepotensiale på -58mV!

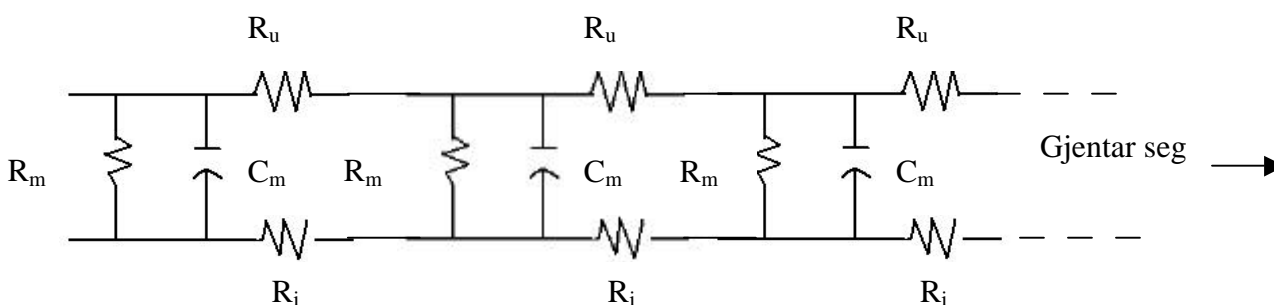
Merk at potensialet vi er kommet frem til ikke er det reelle *hvilepotensiale*. For å oppnå det korrekte hvilepotensiale på ca -70mV er det flere faktorer vi må ta hensyn til. I Modellen vi har regnet på i figur 2 har vi tatt hensyn til konsentrasjonsgradienten og de elektriske kreftene, samt membranens permeabilitet til kloridionet. I virkeligheten er det, som vist i tabell 1, mange flere ioner og ta hensyn til, samt innflytelsen natrium-kalium-pumpen har. Denne pumpen vil vi komme nærmere innpå senere i oppgaven, men ikke i forbindelse med hvilepotensialet. Osmotisk balanse er også en essensiell faktor. Se vedlegg 1.

3. Aksjonspotensialet

3.1. Innledende redegjørelse av et akson, sammenligning med en krets. [11, 10]

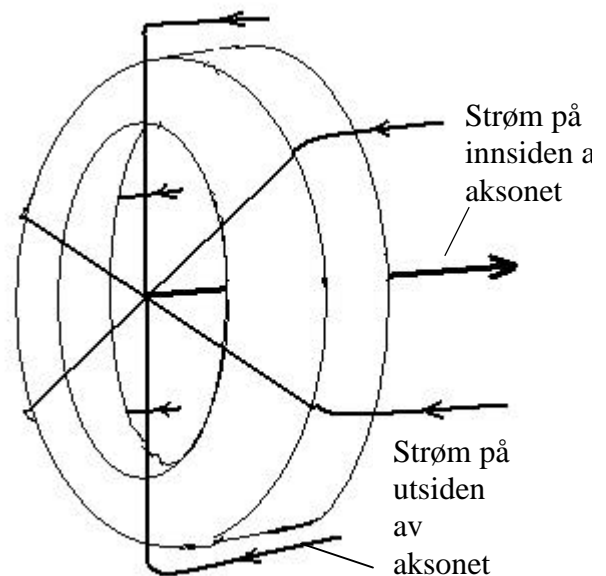
Aksonet er den delen av en nervecelle som leder den elektriske impulsen. Det kan dermed være kort vei til å sammenligne den med en isolert ledningsbit nedsenket i en elektrisk ledende væske. Vi vil ta bildet videre og forsøke å gjøre rede for noen av aksonets fysiske egenskaper ved å tegne det opp som en elektrisk krets. I en slik analyse må vi ta hensyn til motstanden i væsken både på innsiden og på utsiden av membranen. Fordi membranen ikke er helt tett men lekker strøm ut i væsken som omgir den, vil membranen karakteriseres av både motstand og kapasitans. Vi trenger altså fire elektriske parametere til kretsen vår. Da kapasitans og motstand er jevnt fordelt nedover hele lengden av aksonet, vil ikke en enkel krets med fire komponenter være tilstrekkelig beskrivelse. En kommer nærmere det riktige bildet ved å forstå aksonet som en svært lang krets med mange små kretsbitar heftet i hverandre.

I motsetning til elektriske kretser er det i nervecellen ikke bare én ladningsbærer, men minst tre forskjellige, nemlig kalium, natrium og klor. En kunne tro at kretsen måtte bygges opp med tre forskjellige typer ledninger, omtrent som separate rør for varmtvann og kaldtvann i et hus, for å kompensere for dette men det er heldigvis ikke nødvendig. Alle ladede partikler opplever samme kraft per ladning, og bidrar dessuten til potensialet på samme måte. Kretsen vil derfor bestå av bare én type ledning. Vi vil også operere med en samlet motstand i cellemembranen, istedenfor å ta hensyn til at membranen har ulik motstand for de enkelte ionene. I praksis vil derfor modellen vår tilsvare en krets med én type ladningsbærer. Dette er en tilnærming, men for denne teksten vil bildet være tilstrekkelig.



Figur 6. Aksonet som en uendelig lang krets satt sammen av mange små kretser. [11]

Motstanden i væsken på utsiden av aksonet er R_u og i væsken på innsiden er den R_i . Membranen har motstand R_m og kapasitans C_m . Verdiene av de enkelte er som det fremgår av tabellen, avhengig av hvorvidt aksonet er kledt i et lag myelin eller ikke. Som nevnt er myelin er et hvitt og fettete stoff som ligger som en isolasjon rundt aksonet, kun avbrutt i små punkter kalt ranvierske knutepunkt. På hvilken



Figur 7. Strømmen som går i et tverrsnitt av aksonet. [11]

måte det bedrer transporten av nerveimpulsen kommer vi nærmere tilbake til senere.

Egenskap	Akson uten myelin	Akson med myelin
Radius akson	$5.00 \cdot 10^{-6}$ m	$5.00 \cdot 10^{-6}$ m
Motstand per lengdeenhet i væsken på utsiden og på innsiden av aksonet (r)	$6.37 \cdot 10^9 \Omega/m$	$6.37 \cdot 10^9 \Omega/m$
Ledningsevne per lengdeenhet av aksonmembranen (gm)	$1.25 \cdot 10^{-4}$ mho/m	$3.00 \cdot 10^{-7}$ mho/m
Kapasitans per lengdeenhet av aksonet (C)	$3.00 \cdot 10^{-7}$ F/m	$8.00 \cdot 10^{-10}$ F/m

Tabellen viser størrelser for ulike egenskaper hos aksoner med og uten myelin. [11]

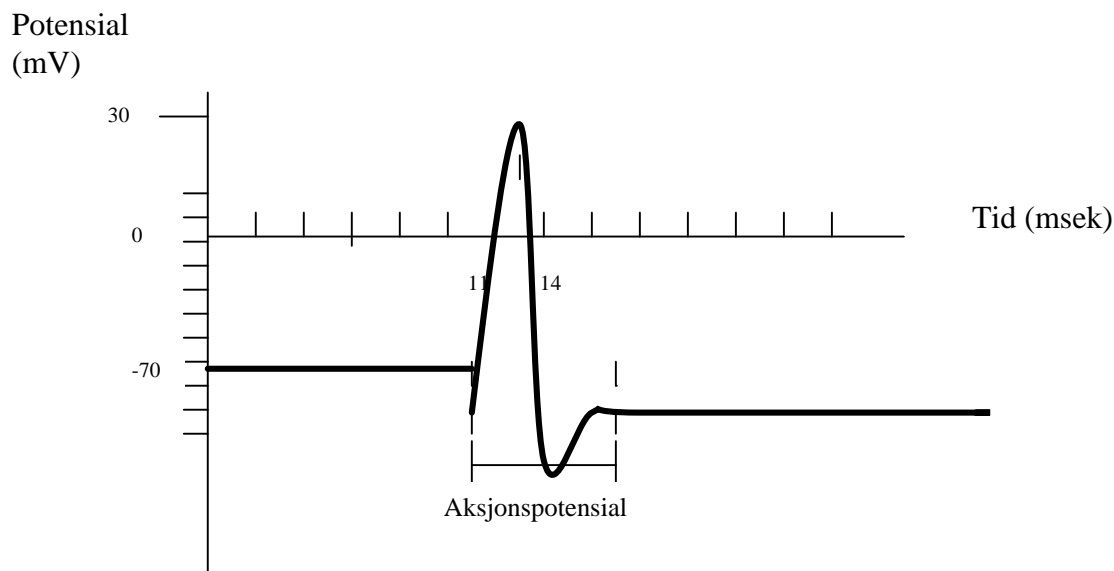
Kretsen som her er presentert kan i flere tilfeller være hensiktsmessig som et bilde av aksonet. Ser en nærmere på den viser det seg likevel at den ikke kan være korrekt. I en

krets bygget opp på denne måten vil den elektriske impulsen dø ut etter kort tid. Senere i teksten vil det vises at signalet i en nervecelle ikke kommer lenger enn 16 mm før det forsvinner. Hadde dette vært tilfelle for nervecellene våre ville vi hatt store problemer med simpelthen å bevege føttene, ettersom de lengste aksonene er opp til en meter lange. Kretsen vår lykkes i å vise på hvilken måte nervecellens oppbygning gir et dårlig utgangspunkt for transport av elektriske signaler. På tross av dette vet vi at impulsene går gjennom selv de lengste aksoner uten å miste verken amplitude eller form. Det interessante her blir dermed hvilke måter nervecellen jobber for ikke bare å omgå disse begrensningene, men også utnytter omgivelsene i best mulig grad. For å få en mer korrekt forståelse av nevronet, og da spesielt aksonet, som leder av elektriske impulser må en være klar over at cellen tar en svært aktiv rolle i transporten ved blant annet hele tiden å fornye signalet den skal føre. Hvordan den gjør dette står i nær sammenheng med mekanismene som setter i gang den elektriske impulsen i en nervecelle.

3.2. Aksjonspotensial [4, 5, 10, 11]

Det elektriske signalet i en nervecelle er karakterisert av noe vi omtaler som et aksjonspotensial. I korte trekk kan en beskrive dette som en hurtig spenningsendring i nervecellen som forplanter seg nedover hele lengden av aksonet. Grunnleggende for hele prosessen er den tidligere omtalte elektriske spenningen eller potensialet, som levende celler opprettholder over membranen. Med utsiden som nullpunkt er potensialet som nevnt målt til rundt -70mV . Blir nervecellen utsatt for riktig stimuli vil den imidlertid reagere med å endre dette potensialet i henholdsvis positiv eller negativ retning. Stimuli kan være mekanisk trykk, kjemiske substanser eller en tilført spenning, fra ytre kilder ved at eksempelvis en muskel strekker seg, eller som følge av at andre nevroner viderefører et signal. I tilfellene der potensialet går mot en mer negativ verdi vil ingenting skje. Det er først når membranen slipper gjennom tilstrekkelig mange natriumioner til at potensialet endres i positiv retning, at det blir interessant. I det membranpotensialet når en bestemt terskelverdi, vil dette utløse en ytterligere spenningsendring som forplanter seg gjennom aksonet. Denne spenningsendringen omtales ofte som aksjonspotensialet, og er svært kraftig sett i forhold til hvilke spenningsforhold som pleier å være over en cellemembran. Nervecellen er unik blant celler i sin evne til å generere det.

Nøyaktig hva er det som skjer i aksonet når det genereres et aksjonspotensial? Vi har sett at det er forholdet mellom negative og positive ioner som er opphavet til spenningen over cellemembranen. Noen av de viktigste mekanismene som er med på å kontrollere dette forholdet er voltstengte kalium- og voltstengte natriumkanaler, i tillegg til natriumkalium-pumpen. Disse vil vi vente med å gå nærmere inn på funksjonen av. Tilstanden hvor innsiden av aksonet har en omtrentlig spenningsverdi på -55mV i forhold til væsken på utsiden får membranen til plutselig å åpne kanalene for natrium slik at disse ionene, tiltrukket av negativ ladning, strømmer inn. Potensialet spretter opp til en verdi på 30mV , før nervecellen reagerer på nytt med å åpne portene for kalium og lukke for natrium. Prosessen i et gitt punkt er over på et tusendels sekund. Potensialet synker videre til en verdi på -90mV , før det stiger tilbake til utgangsverdien på -70mV . I løpet av de tusendels sekunder det tar før hvilepotensialet er gjenopprettet er aksonet ikke i stand til å generere en ny impuls. Dette setter en grense for frekvensen av impulser i en nervecelle. Hos mennesker har man funnet maksimalt antall elektriske impulser per sekund til å være 100.



Figur8. Forløpet av et enkelt signal i et nevron.[11]

3.3. Antall ioner [11]

Det er på sin plass å bemerke at antallet ioner som strømmer gjennom membranet i løpet av et aksjonspotensial er for lite til å endre tettheten av ioner i særlig stor grad. Fra sammenhengen

$$\Delta Q = C\Delta V \quad (3.3.1.)$$

der ΔQ er endringen i ladning, C er kapasitansen per meter lengde av aksonet (for verdi se tabell x) og ΔV er endringen i spenning over membranen, kan en regne ut omtrent hvilket antall det er snakk om. Spenningen går fra -70mV til 30mV , noe som gir en ΔV på 100 mV . Ved å dele på ladning per ion, som er nær $1.6 \cdot 10^{-19}$ Coulomb, får vi at i et akson uten myelin vil $1.87 \cdot 10^{11}$ natriumioner passere membranen og komme inn i aksonet. Samme antall kaliumioner forsvinner ut av aksonet like etterpå. Av ligningen kan vi også regne ut at antall natriumioner på innsiden av en meter med akson er rundt $7 \cdot 10^{14}$ når aksonet har hvilepotensial, mens antall kaliumioner er rundt $7 \cdot 10^{15}$. Antall ioner ved hvilepotensial er altså av en størrelsesorden 10^3 ganger større enn det antall som går ut og inn gjennom membranen. Målinger viser at antallet ioner som strømmer inn og antallet ioner som strømmer ut i virkeligheten er tre ganger høyere enn det våre beregninger kommer frem til, men konklusjonen er likevel den samme.

3.4. Impuls eller ikke impuls [5, 11]

En ting som er viktig å merke seg er at nervecellen jobber etter en "alt eller ingenting" - lov. Så lenge spenningen kommer over den gitte terskelen, som for de aller fleste nevroner ligger på -55mV , vil impulsen gå i nervecellen med full intensitet. Hvis spenningen ikke når denne verdien vil ikke aksjonspotensialet finne sted i det hele tatt. Det blir det samme som når en fyrer av en pistol. Presser du ikke hardt nok på avtrekkeren vil ikke kulen sendes av gårde. Men i det avtrekkeren trykkes langt nok inn skytes kulen ut med en fast hastighet. Du kan ikke skyte noen forsiktig ved å trykke svakt på avtrekkeren, slik du heller ikke kan skyte noen ekstra hardt ved å klemme avtrekkeren inn med alle krefter. Årsaken til at vi oppfatter sanseintrykk av varierende styrke ligger dermed ikke i styrken på nerveimpulsen. Den er tettere knyttet til frekvensen av impulser

som går gjennom nervecellen, eller antallet nerveceller som sender impulsen. Som vi skal se ligger det en naturlig begrensning på hvor mange impulser som kan gå gjennom et akson i løpet av en gitt tid.

Det faktum at spenningsverdien aksjonspotensialet når er den samme for alle impulsene som går gjennom aksonet, tyder på at informasjon som utveksles mellom og lagres av nervene i kroppen er binær¹, eller bygget opp av bits, i likhet med hos en datamaskin. Det er trolig en sammenheng mellom verdien for potensialet og mengden av energi som kan lagres i aksonet ved hjelp av dens kapasitanssegenskaper. En nerveimpuls lader nemlig fullstendig ut kapasitansen i et akson, og den må lades opp på nytt før neste impuls kan gå.

Energien som kreves til denne oppladningen kan beregnes gjennom formelen

$$E = \frac{1}{2} C \cdot (\Delta V)^2 \quad (3.4.1.)$$

der C er kapasitans per meter akson og ΔV er endring i spenning. Utledning av formelen vil sløyfes. For en meter med akson gir dette

$$E = \frac{1}{2} \cdot 3 \cdot 10^{-7} \cdot (0.1)^2 = 1.5 \cdot 10^{-9} \quad (3.4.2.)$$

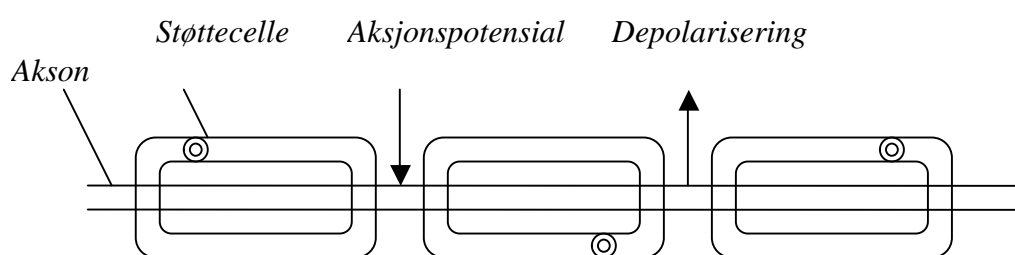
For å lade opp igjen en meter med akson trenger en altså $1.5 \cdot 10^{-9}$ J/m. Da en impuls er over etter omtrent 10^{-2} sekunder og et akson kan lede omtrent hundre impulser per sekund, trenger aksonet, selv når det har maksimalt å gjøre, bare $1.5 \cdot 10^{-7}$ W/m for å lade kapasitansen opp på nytt.

3.5. Impulsens forplantning [4, 10, 11]

Aksjonspotensialet som her er forklart finner sted i et lite område om gangen. Det plutselige hoppet i spenning i ett punkt av aksonet vil imidlertid utløse samme reaksjon i neste punkt, og slik forplanter impulsen seg som en kjedereaksjon nedover hele lengden

¹ Totallssystem; en datamaskin kan når det kommer til stykket kun skille mellom høy spenning og lav spenning, og informasjon er dermed representert med lange rekker av 0 og 1. Disse minste mulige bitene informasjon kalles bits.

av aksonet. En bølge er et godt bilde på bevegelsen. Det eneste som egentlig flytter seg i lengderetningen av aksonet er spenningshoppet; ionene som inngår beveger seg kun inn og ut gjennom membranen og væsken som leder impulsen forblir i ro. I et akson uten myelin går impulsen i en sammenhengende bølgebevegelse inntil den når enden og der for eksempel setter i gang impulsen i neste nervecelle. Dette gjelder blant annet for nervecellene i hjernen. De fleste andre nerveceller, spesielt de store cellene vi finner ute i kroppen, har sine aksoner dekket med et lag myelin. Fordi myelinet hindrer ionene i å bevege seg ut og inn i aksonet, kan ikke aksjonspotensialet finne sted andre steder enn ved de ranvierske innsnevringene. Her vil altså bølgen vår, om vi fortsetter å betrakte forplantningen av aksjonspotensialet på denne måten, bli brutt opp og ikke tillates å bevege seg mer enn de 1-2 μm som utgjør lengden av innsnevringene. En annen side ved myelinet sin effektive isolasjonsegenskap er at da strømmen hindres i å lekke ut gjennom membranen, går det elektriske signalet mye fortere gjennom de delene av aksonet som er dekket av dette.

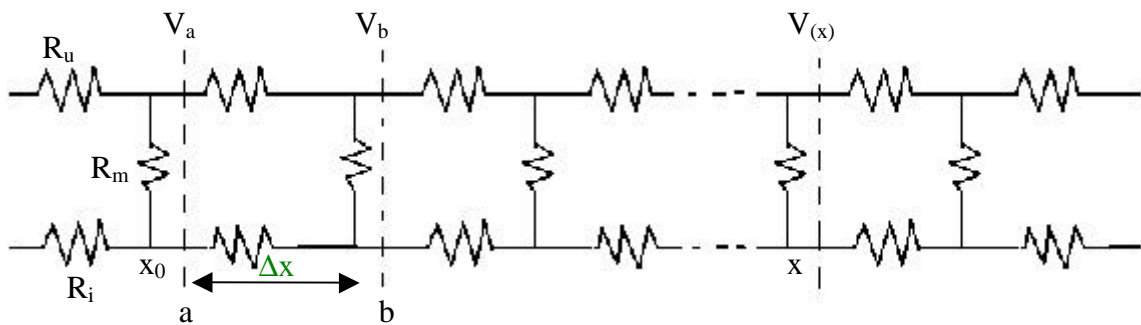


Figur 9. Hvordan aksjonspotensialet beveger seg fremover i et akson .[4]

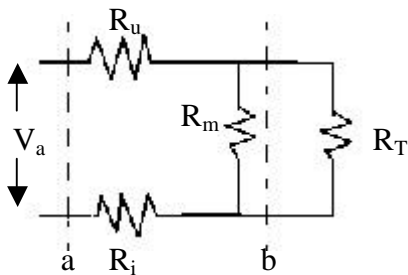
Vi begynner å nærme oss nøkkelen til hvordan nervecellen kan være en så vellykket leder av elektriske signaler på tross av de dårlige forutsetningene den har. Strømmen går mye fortere gjennom aksonet når det er dekket av myelin. Men hadde det vært dekket av myelin hele veien ville signalet blitt svekket og dø før det nådde enden av aksonet. Dette er løst med å åpne myelinet i innsnevringene og dermed tillate signalet å bli fornyet. Som i så mange andre tilfeller har vi en gylden middelvei hvor det beste fra to løsninger er kombinert.

3.6. Hvor fort skulle vi vente at signalet i et akson dør ut? Tilbake til kretsen. [11]

La oss vende tilbake til kretsen som ble brukt som representasjon av aksonet til å begynne med. Den største feilen med denne var at den ikke tok i betraktning hvordan nervecellen hele tiden vedlikeholder signalet som går i aksonet. Vi kan derfor søke å forbedre vår tilnærming ved å føye til spenningsgeneratorene, som vil være en parallell til den stadige fornyingen av aksjonspotensialet. Generatorene plasseres med jevne mellomrom i kretsen som vist i figur 10:



Figur 10a). Den uendelig lange kretsen hvor kapasitansen er sett bort fra. [11]



Figur 10 b). Motstander til høyre for linjen b er erstattet med R_T . [11]

For å analysere denne kretsen vil det gjøres nok en forenkling ved å se bort fra kapasitansen. Dette for å forhindre at utregningen blir for komplisert. Så lenge kapasitansen er fullstendig oppladet er dette en gyldig tilnærming, og det vil la seg gjøre å beskrive svekningen, eller dempningen av spenningen langs aksonet. Aksonets tidsavhengige atferd vil vi dessverre ikke få kunnskap om med denne modellen.

Vi står igjen med følgende problem: Hvordan skal vi gå frem for å beregne spenningen $V(x)$ i et punkt x når det tilføres en spenning V_a i et punkt x_0 ? Vi begynner med å undersøke spenningsfallet over et lite segment av kretsen Δx som befinner seg mellom linjene a og b. Vi antar at kretsen er uendelig lang og at den totale motstanden i kretsen på høyre side av b er R_T . Da kretsen er uendelig lang innebærer dette at den totale motstanden i kretsen til høyre for enhver vertikal linje som tilsvarer b også er R_T . Motstanden til høyre for a er dermed R_T . Dette gir oss mulighet til å uttrykke R_T gjennom kjente størrelser.

$$R_T = R_o + R_i + (R_T \cdot R_m)/(R_T + R_m) \quad (3.6.1.)$$

Målinger har vist at motstanden på innsiden og motstanden på utsiden av aksonet er tilnærmet like. Vi setter dermed $R_i = R_o = R$ slik at ligningen kan forenkles til

$$R_T = 2R + (R_T \cdot R_m)/(R_T + R_m) \quad (3.6.2.)$$

Løsningen av denne blir

$$R_T = R + [R^2 + 2R \cdot R_m]^{1/2} \quad (3.6.3.)$$

En analyse av kretsen gir

$$V_b = V_a / (1 + [2R \cdot (R_T + R_m) / (R_T \cdot R_m)]) = V_a / (1 + \beta) \quad (3.6.4.)$$

der β er størrelsen innenfor firkantklammen. R og R_m er verdier som tilhører en svært liten del av aksonet med lengde Δx . Vi får dermed sammenhengene

$$R = r \cdot \Delta x \quad \text{og} \quad 1/R_m = g_m \cdot \Delta x \quad \text{eller} \quad R_m = 1/g_m \cdot \Delta x$$

der g_m er ledningsevnen til membranen. Bruker vi dette samtidig som vi gjør Δx tilstrekkelig liten, får vi fra ligning (3.6.4.)

$$R_T = r \cdot \Delta x + [(r \cdot \Delta x)^2 + 2(r \cdot \Delta x) \cdot 1/g_m \cdot \Delta x]^{1/2} \sim [2r/g_m]^{1/2} \quad (3.6.5.)$$

og

$$\beta = [2 \cdot r \cdot \Delta x \cdot ([2r/g_m]^{1/2} + [1/g_m \cdot \Delta x]) / ([2r/g_m]^{1/2} \cdot 1/g_m \cdot \Delta x)] \sim (2r \cdot g_m)^{1/2} \cdot \Delta x = \Delta x / \lambda \quad (3.6.6.)$$

hvor

$$\lambda = [1/2rg_m]^{1/2} \quad (3.6.7.)$$

Vi går nå tilbake til ligning IV. En svært liten Δx innebærer en liten β og leddet $1/(1 + \beta)$ er derfor omtrent lik $1 - \beta$. Spenningen V_b i b som ligger i en avstand Δx fra a kan dermed uttrykkes

$$V_b = V_a [1 - \Delta x / \lambda] \quad (3.6.8.)$$

For å få spenningen i en avstand x fra a deler vi denne avstanden opp i et antall n av Δx slik at $n \cdot \Delta x = x$. Ligningen kan da benyttes på følgende form nedover aksonet:

$$V(x) = V_a [1 - \Delta x / \lambda]^n \quad (3.6.9.)$$

Gjør vi denne om på rekkeform får vi

$$V(x) = V_a [1 - \Delta x / \lambda + (n(n-1)/2!) \cdot (\Delta x / \lambda)^2 - (n(n-1)(n-2)/3!) \cdot (\Delta x / \lambda)^3 + \dots] \quad (3.6.10.)$$

Da Δx er forsvinnende liten, må n være veldig stor. Dette gjør at rekken nærmer seg rekkeutviklingen av en eksponentialfunksjon, og vi kan skrive

$$V(x) = V_a e^{-x/\lambda} \quad (3.6.11.)$$

Signalet nedover aksonet svekkes altså eksponensielt. For et akson uten myelin får vi en λ på 0.8 mm, med andre ord vil impulsen minke til 37 % av opprinnelig verdi etter 0.8 mm. Med myelin er λ lik 16 mm. Merk at disse resultatene forteller hva som ville skje om ikke aksjonspotensialet stadig ble fornyet. Stykkene med myelin er ikke mer enn 2 mm lange, så med en λ på 16 mm svekkes potensialet med kun 13 % innen det når neste ranvierske knutepunkt. Denne spenningsverdien er fortsatt stor nok til å sette i gang et nytt aksjonspotensial.

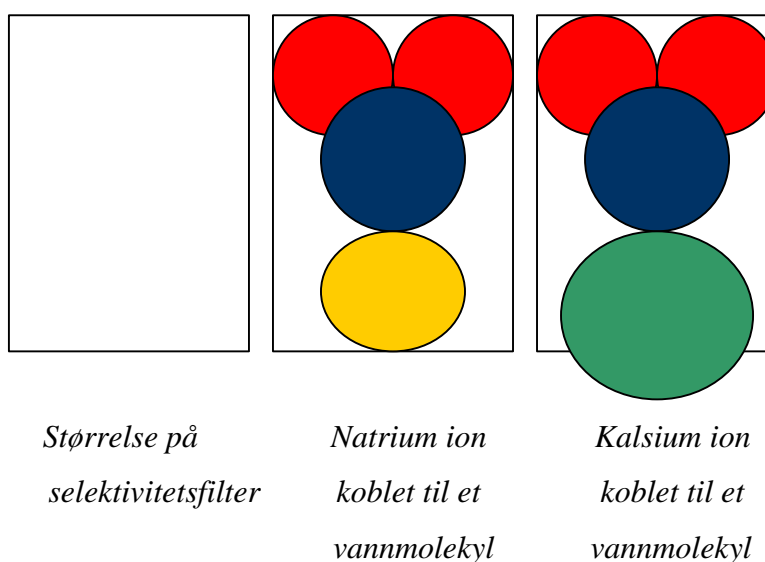
4. Den voltstengte natrium- og kaliumkanalen [8,9]

Membranpotensialet regulerer om natrium- og kaliumkanalene er åpne eller lukket, og derfor sier vi at de er voltstengte.

4.1 Voltstengt natriumkanal [8,9]

Natriumkanalen består av fire hoveddeler, der hver del inneholder seks spiraler. Se figur 12. Disse fire delene former en slags pore mellom dem, og denne poren i membranen er høyt selektiv for Na^+ ioner. Dette er fordi den har et selektivitetsfilter som gjør natriumkanalen 12 ganger mer permeabel for Na^+ enn til K^+ . Se figur 11.

Figur 11 *Illustrerer hvordan et natriumselektivitetsfilter fungerer*

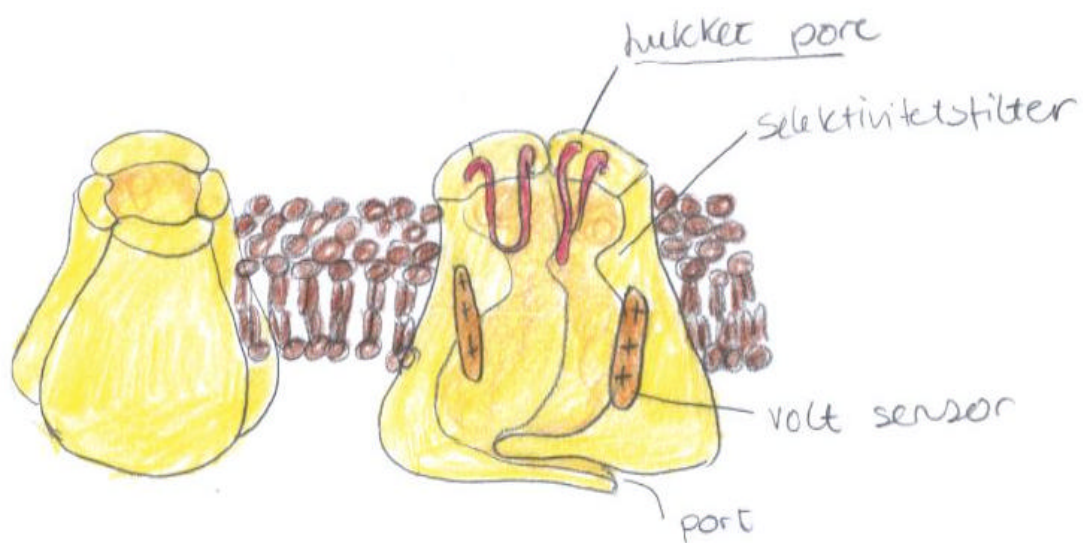


Vi ser at vannmolekyler transporterer Na^+ ioner gjennom selektivitetsfilteret. Når Na^+ skal igjennom kanalen løsrives de fra vannmolekylet. Fra figuren ser vi at dimensjonen av natriumselektivitetsfilteret passer til Na^+ og ikke K^+ , og dette gjør at K^+ blir ekskludert fra å passere gjennom poren.

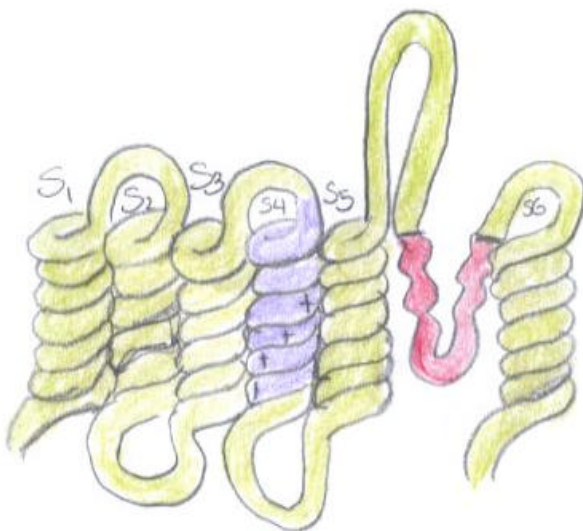
Det er forandringer i det elektriske potensialet i membranen som gjør at poren er enten åpen eller lukket.

I spiral S4, (se figur 13) sitter en spenningsensor. I denne spiralen er det positivt ladet aminosyre som er regelmessig fordelt langs spiralen. Ved et aksjonspotensial, som nevnt tidligere, blir S4 dyttet bort fra innsiden av membranen og denne forandringen i molekylet får kanalen til å åpne seg. Se figur 14.

Denne natriumkanalen åpner seg etter en liten forsinkelse og lukker seg igjen etter ca 1 ms, og den kan ikke åpnes igjen før membranpotensialet har returnert til en negativ terskelverdi.



Figur 12 Lukket natriumkanal



Figur13 utsnitt av natriumkanal.



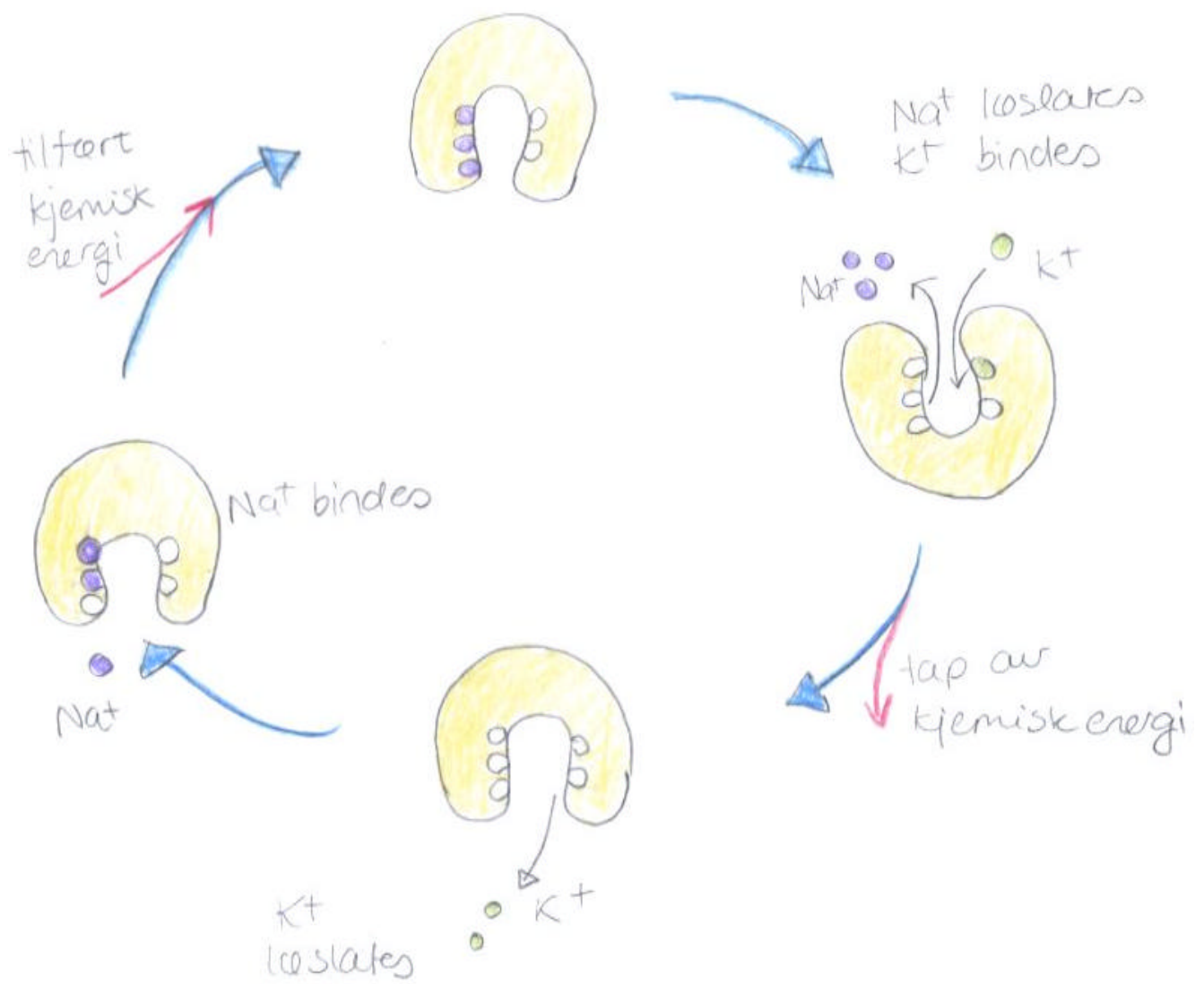
Åpen pore

Figur14 . Åpen natriumkanal

4.2 Voltstengt kaliumkanal [8,9]

Vi har sagt at natriumkanalene åpnes ved en viss forandring i membranpotensialet. Det gjør også kaliumkanalene, men de åpner seg ikke like umiddelbart, og på grunn av denne forsinkelsen reverseres membranpotensialet like etter. Denne utstrømmingen av K^+ ioner kan sammenlignes med en likeretter. Kaliumkanalen har samme struktur som natriumkanalene, og det samme skjer når porene åpnes.

Når membranpotensialforskjellen har ført til at natrium har strømmet inn og kalium ut, blir dette reversert ved at en pumpe, kalt *natrim-kaliumpumpe*, transporterer ionene gjennom membranen mot deres konsentrasjonsgradient. Denne syklusen blir drevet av kjemisk energi. Se figur 15.



Figur15 Natrium kalium pumpe. Syklusen er drevet av kjemisk energi.

5. Hvordan nervecellene kommuniserer; synapsetransmisjon

Synapse; punktet hvor aktiviteten blir overført fra en nervecelle til en annen

Det er i synapsen aktiviteten blir overført fra et nevron til en annen. Vi har to forskjellige typer, den elektriske- og kjemiske synapsen. Enkelt sett kan man sammenligne disse to synapsene med e-post og vanlig brevsending. I begge synapsene skal informasjon videreformidles fra et nevron til et annet, men det skjer på forskjellige måter og med forskjellig hastighet. Den elektriske synapsen kan sammenlignes med e-post, ved at der er tilnærmet umiddelbar informasjonsoverføring og ingen forsinkende mellomledd som ved vanlig brevsending. Den kjemiske synapsen kan sammenlignes med et "sender-postkontor-postmann-mottaker system". Her bestemmer senderen innholdet i brevet og hvem brevet spesifikt skal være til. Dette brevet havner på et postkontor og venter på å bli transportert av postmannen til den rette mottakeren. Mottakeren åpner brevet og kan handle etter informasjonen den har fått.

5.1 Elektrisk synapse [8,9]

Vi sammenlignet denne synapsen med en sending av e-post. Det er fordi det går en direkte strøm av ioner gjennom kanaler fra den ene nervecellen til den andre. Strukturen til en elektrisk synapse er illustrert i fig (16). Senderen i dette systemet kalles det *presynaptiske* nevronet og er kilden til strømmen. Det er her informasjonen som skal videreformidles kommer fra og kan sammenlignes med en e-post som skal frem til en mottaker. Mottakeren kalles det *postsynaptiske* nevron, og det er hit strømmen går til. Når informasjonen har nådd frem kan nevronet reagere ut fra den informasjonen den har mottatt.

Grunnen til at strømmen av ioner kan gå direkte gjennom kanaler mellom nevronene er at membranene til disse to kommuniserende nevronene kommer så nær hverandre i synapsen at de er koblet sammen og danner en "kanalåpning". Ut av figuren ser vi at denne forbindelsen danner en pore hvor ionene kan strømme igjennom. Disse

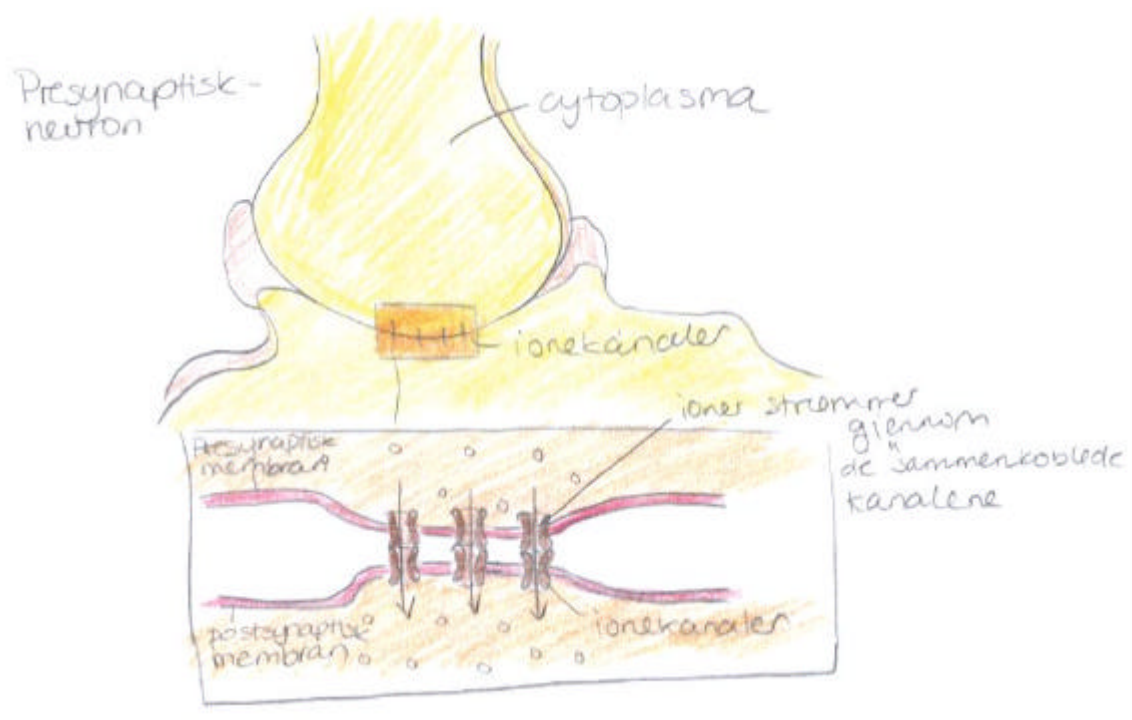
porene er mye større enn natrium- og kaliumkanalene og her kan alle hovedcellulære ioner og mange små organiske molekyler *diffundere* mellom cytoplasma i de pre- og postsynaptiske nevronene.

Kilden til denne ionestrømmen er potensialdifferansen som er lokalt generert av aksjonspotensialet.

Denne overføringen er "bidirectional"; strømmen kan flyte i begge retninger gjennom poren, avhengig av hvilke av de to nevronene aksjonspotensialet ble generert i.

Transmisjonen er også ekstraordinær rask fordi ioner strømmer gjennom poren øyeblikkelig, slik at kommunikasjonen kan skje uten forsinkelse.

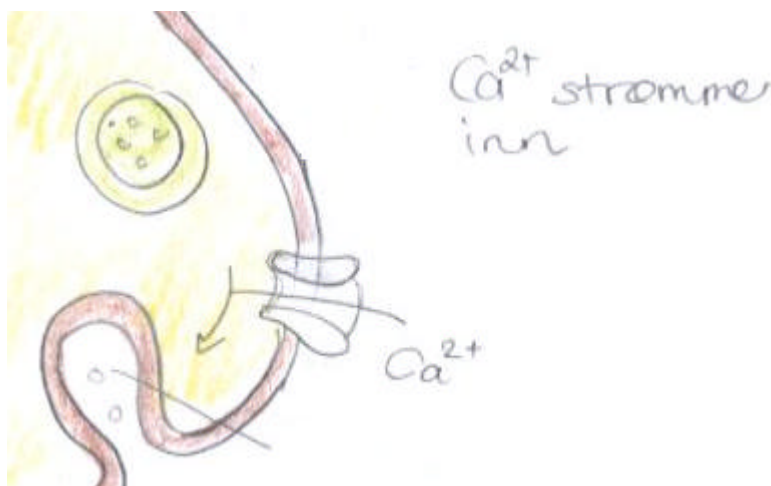
En mer generell oppgave den elektriske synapsen har er å synkronisere elektrisk aktivitet blant forskjellige nevroner. For eksempel har vi hjernestammenevroner som genererer rytmisk elektrisk aktivitet som ligger til grunn for blant annet pusting, og er synkronisert av elektriske synapser.



Figur16 Elektrisk synapse

5.2 Signaloverføring i kjemisk synapse [8,9]

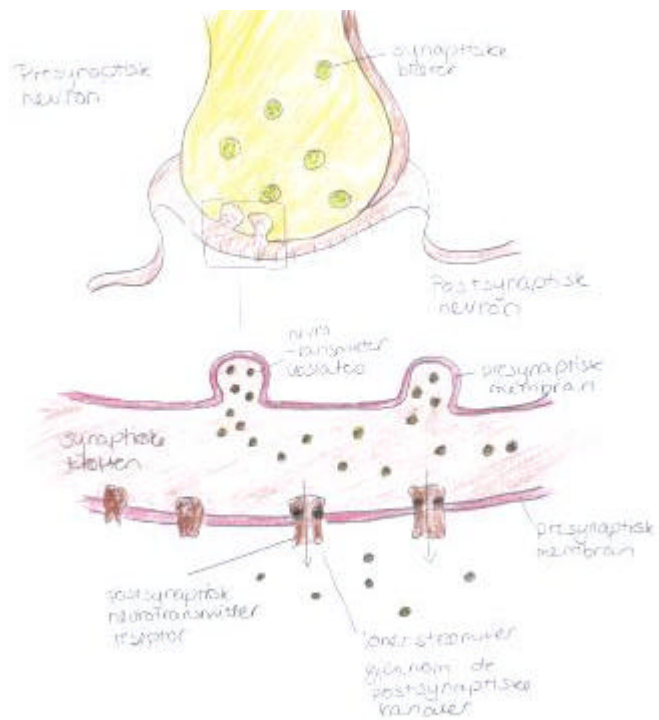
Forskjellen på en elektrisk- og en kjemisk synapse er at avstanden mellom det pre- og postsynaptiske nevronet er mye større. Denne avstanden kalles den synaptiske kløften, og den gjør at kommunikasjonen blir forsinket. Det er her postkontoret og postmannen kommer inn i bildet. I sendernevronet er det kuleformede organeller kalt *synaptiske blærer*, og de fungerer som et brev. Se figur 18. Inni disse blærene er det informasjon til mottakernevronet, og denne informasjonen er satt sammen av transmitterstoffkombinasjoner. Forskjellige transmitterstoffkombinasjoner kan sammenliknes med forskjellig innhold i brev til forskjellige mottakere. Prosessen er satt i gang når sendernevronet blir utsatt for et aksjonspotensial. Forandringen i membranpotensialet forårsaket av ankomsten til aksjonspotensialet, fører til at kalsiumkanalene i membranen til sendernevronet åpner seg. Se figur (17). Denne kalsiumkanalen er også "voltstengt" slik vi har beskrevet natrium- og kaliumkanalene tidligere.



Figur 17. Calsiumkanal

Her har vi at Ca^{2+} konsentrasjonen er mye lavere innvendig i sendernevronet enn utenfor i kløften. Åpningen av disse kanalene medføre en strøm av Ca^{2+} inn i sendernevronet, og dette resulterer i at Ca^{2+} konsentrasjonen i cytoplasma øker til et mye høyere nivå som igjen tillater blærene med transmitterkombinasjoner å smelte sammen med

plasmamembranen til sendernevronet. Denne sammensmeltingen er Ca^{2+} avhengig og har den viktigste oppgaven; den løslater transmitterkombinasjonene til den synaptiske kløften. Vi får en påfølgende eksocytose hvor transmitterstoffene diffunderer gjennom kløften og bindes til spesielle reseptorer som sitter på mottakernevronet. Se figur 18. Det er denne diffunderingen gjennom kløften som gjør at informasjonen blir forsinket. Akkurat som postmannen som skal bringe brevet til mottakeren, noe som heller ikke skjer umiddelbart. Disse reseptorene som sitter på mottakernevronet kan sammenliknes med hus som har forskjellige adresser. Informasjonen når mottakernevronet ved at transmitterstoffene binder seg til reseptorer. Det viktige her er at ikke hvilke som helst transmitterkombinasjoner kan bindes til hvilke som helst reseptor, slik som et brev må leveres til riktig adresse og ikke til hvilke som helst hus for å nå rett mottaker. Da har informasjonen betydning når den når den riktige mottakeren som medfører at riktig handling blir utført. Når transmitterstoffene har kommet til riktig reseptor vil mottakernevronet åpne (og noen ganger) lukke kanaler. Dette forandrer ionenes evne til å flytte inn- eller ut av mottakercellen. Den resulterende induserte transmitterstrømmen forandrer konduktansen og vanligvis membranpotensialet til mottakernevronet ved å øke eller minke sannsynligheten for at nevronet vil avfyre et aksjonspotensial. På denne måten blir informasjonen overført fra ett nevron til et annet.



Figur18 . Kjemisk synapse

5.3 Litt om nevrontransmittere [8,9]

Det finnes mer enn 100 forskjellige transmittere som kan klassifiseres inn i to hovedkategorier: småmolekyl-transmittere og neuropeptider (lange kjeder). Multiple transmittere kan produsere forskjellig type respons på individuelle mottakernevroner alt etter transmitterkombinasjonen. For eksempel kan et nevron bli igangsatt av en type transmitterkombinasjon og hemmet av en annen type. Farten til mottakernevronresponsen som er produsert av forskjellige transmitterstoffer varierer, og tillater kontroll av elektriske signal over forskjellige tidsskalaer. Generelt er småmolekyl-transmitterne hurtigere meglere mens neuropeptidene er tregere.

5.4 Transmitter reseptorer [8,9]

Reseptorene er proteinmolekyler, og disse forklarer de spesifikke og kraftige handlingene til visse transmitterkombinasjoner på muskler og nerveceller. Reseptor molekylene har betydning for evnen til transmittere, hormon og medikamenter til å forandre de funksjonelle områdene til nevronet. Vi har for eksempel vårt naturlige smertestillende medikament i kroppen kalt *endorfiner* (produseres blant annet under samleie og trening).

Disse endorfinene kan oppta en reseptorplass som gjør at en nervecelle ikke får formidlet smerte til en annen, altså vi føler ingen smerte som vi ellers ville ha gjort.

Sammendrag

Vi har i denne oppgaven sett på hvordan vi ved hjelp av fysikkens lover kan beskrive en nervecelle og hvordan den sender elektriske signaler. Nervecellene bruker elektriske impulser, nerveimpulser, til å sende informasjon til hverandre og til resten av kroppen. Disse signalene kommer av en endring i cellens membranpotensial og forplanter seg langs cellens akson før de overføres til andre celler.

Membranpotensialet oppstår på grunn av forskjell i elektriske ladninger på ioner innenfor og utenfor cellemembranen. Et overskudd av negative ioner inne i cellen gir et membranpotensial på ca -70 mV. Når det kommer inn elektriske signaler til nervecellen vil membranpotensialet endres, og når denne endringen er tilstrekkelig stor vil det utløses et aksjonspotensial.

Aksjonspotensialet fører til at kanaler (natriumkanaler) som sitter i cellemembranen åpnes og positive ioner strømmer inn. Dette med hensikt i å videreføre signalet. Når signalet er sendt, vil andre kanaler (kaliumkanaler) åpne seg for å reversere potensialendringen, og positive ioner strømmer ut. Ionene befinner seg nå på motsatt side av membranen enn det de var i utgangspunktet, og blir tvunget tilbake av en pumpe drevet av kjemisk energi.

Aksjonspotensialet fører altså til at nervecellene videreformidler informasjon slik at de kan kommunisere med hverandre. Dette skjer i synapsene der informasjon blir sendt, enten ved direkte ionestrøm mellom nervecellene (elektriske synapser) eller ved at kanaler åpner seg på cellemembranen som fører til at blærer med informasjon blir sendt videre til den neste nervecellen (kjemisk synapse).

Etterord

I denne oppgaven er det gjort en mengde forenklinger. På langt nær alle aspekter ved transporten av den elektriske impulsen i en nervecelle er heller presentert. Dette er det flere årsaker til. Fysisk forskning på nervesystemets funksjon har de siste hundre årene utvidet kunnskapsfeltet i en så stor grad at å ta for seg alt ned i minste detalj ville resultert et oppslagsverk på flere tusen sider. Mangel på plass og tid er derfor en grunn. En annen og kanskje viktigere årsak er at en svært detaljert beskrivelse ville mistet den store helheten og oversikten vi ønsket å få frem i vår oppgave. Spesielt i en fremstilling rettet mot en mottaker uten store forhåndskunnskaper om nervernes oppbygning og kommunikasjonsevne, er det viktig å formidle en oversiktlig sammenheng og prøve å unngå at budskapet drukner i en mengde formler tatt ut av sammenheng.

Vi håper leseren gjennom vår presentasjon likevel har fått et innblikk i hvor omfattende det fysiske fagfeltet er. Fysikk handler ikke utelukkende om store tunge ting som beveger seg eller om elektroners ferd gjennom et magnetfelt, for å komme med noen eksempler. Som nevnt i forordet var det mange som i lang tid trodde fysikken var fullstendig ute av stand til å beskrive den levende verden, og at dyr og planter, amøber og fisk, samt alt annet som måtte finne på å være i live, var styrt av en ukjent livskraft. Denne tanken har blitt kraftig tilbakevist, blant annet ved hjelp av oppdagelser og sammenhenger som er presentert i denne oppgaven. Likevel er det ett som er viktig å merke seg. De kunnskaper en har fått om biologiske systemer fra fysisk forskning er utelukkende av kvantitativ art. En vet for eksempel svært mye om forløpet av en nerveimpuls gjennom et akson. Det vil si hvor fort den beveger seg, hvilke spenningsforhold det er snakk om, hvor lang tid det hele tar og så videre. Kunnskaper om *hvordan* fenomener utarter seg finnes det nær sagt ingen ende på. Akkurat *hvorfor* de finner sted gang etter gang er et spørsmål som viser seg vanskeligere å besvare.

Et eksempel som vi var inne på i oppgaven gjelder menneskets hjerne. Det er som vi husker mye som tyder på at informasjon lagres og bearbeides på samme måte her som hos en datamaskin. I datamaskiner er man i stand til å analysere informasjonen ned til de

minste bestanddelene uansett hvor kompliserte sammenhenger det er snakk om. Det eneste man til slutt sitter igjen med er en lang rekke med 0 og 1. Hvordan noe tilsynelatende enkelt som begrepet av en firkant er satt sammen i den menneskelige bevissthet har man imidlertid ingen kunnskap om. Foreløpig er det lite som peker i retning av at man noensinne vil få en klar forståelse av fenomener som tanker og følelser. Eller for den del, kunnskap om et underliggende prinsipp bak livet selv. Kanskje er dette viten som for all tid vil være utenfor rekkevidde. Det er vanskelig å si. Forskningen er uansett fortsatt i full gang.

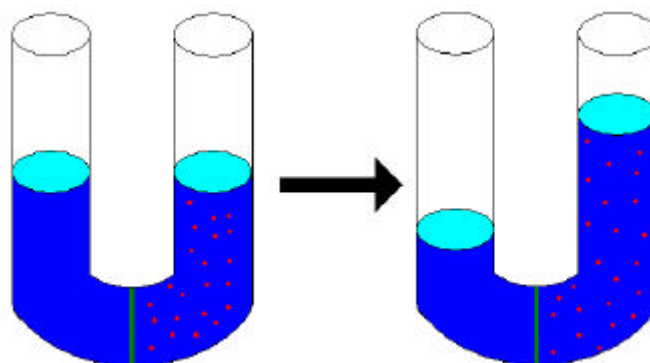
Vedlegg 1

Osmose [4]

I en løsning vil det finnes frie vannmolekyler i tillegg til vannmolekyler som er bundet til det løste stoffet. Dette betyr at konsentrasjonen av vannmolekyler i en løsning er mindre enn konsentrasjonen av vannmolekyler i en ren væske. Konsentrasjonen av vannmolekyler vil minske når konsentrasjonen av løsningen øker.

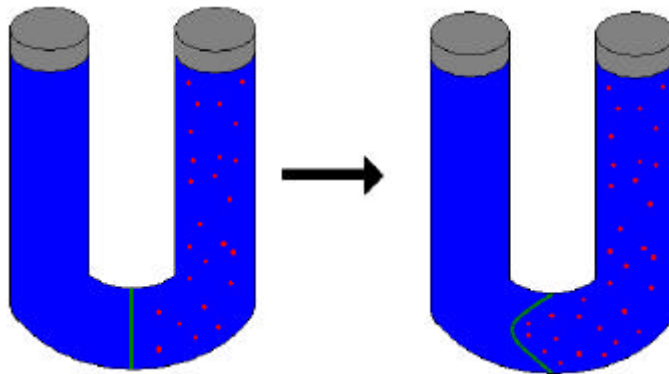
Hvis man deler et U-formet rør i to med en halvgjennomtrengelig membran (som slipper igjennom vannmolekyler, men ikke et løst stoff). Og fyller venstre side av U-røret med rent vann og høyre side med en løsning av stoffet, vil det nå være større konsentrasjon av vannmolekyler på venstre enn på høyre side. For å oppnå likevekt vil vannmolekyler diffundere til høyre side av membranen helt til konsentrasjonen av vann er lik på begge sider.

Denne transporten av vann gjennom halvgjennomtrengelige membraner kalles osmose.



Figur 1: et U-formet rør fylt med vann på høyre side og en løsning på venstre side, skilt av en membran som slipper gjennom vannmolekyler men ikke det løste stoffet.

Osmosen fører til, som vist på figuren, at væskehøyden stiger på høyre side og synker på venstre side. Hadde vi i stedet fylt røret helt opp på hver side og satt på lokk slik at væsknivået ikke kunne stige, og i tillegg byttet ut membranen med en som var elastisk ville vannmolekylene diffundere mot høyre og presse membranen mot venstre.



Figur 2: U-formet rør med lokk, der rent vann og en løsning er skilt av en elastisk halvt gjennomtrengelig membran som slipper gjennom vannet men ikke det løste stoffet.

Cellemembraner fungerer som slike elastiske halvgjennomtrengelige membraner. Inne i cellen finnes organiske molekyler med elektrisk ladning (for det meste proteiner) som ikke slipper ut gjennom cellemembranen. Utenom disse molekylene har væsken inne i og utenfor cellen mer eller mindre samme kjemiske sammensetning. Dette medfører en ubalanse i vannkonsentrasjonen inni og utenfor cellen. Vannmolekyler vil diffundere inn i cellen, for å prøve å utligne denne forskjellen, og cellens volum øker.

Vanndiffusjon er et problem for alle celler pga molekylene inne i cellen som ikke slipper ut. For å oppnå likevekt vil celleveggen utvide seg til det uendelige eller til den sprekker, for først da er det like stor konsentrasjon av de organiske molekylene utenfor og inni cellen. Løsningen for cellene hos levende vesener er å gjøre cellemembranen ikke-permeabel for enkelte molekyler i væsken som omgir cellen. Vannet vil være i likevekt så lenge konsentrasjonen av ikke-permeable molekyler utenfor cellen er lik konsentrasjonen av ikke-permeable indre molekyler. Cellene kan derfor øke og minke i volum for å opprettholde likevekten, men de står ikke i fare for å sprekke.

Hos mennesker er det spesielt natriumioner som nektes adgang til cellen. (Natriumioner slippes imidlertid inn i nerveceller i forbindelse med at nerveimpulser blir sendt, de blir da fraktet inn i cellen via natrium-kalium-pumpen)

Referanseliste

Bøker:

- [1] Biological Psychology-an introduction to behavioral and cognitive neuroscience. M.R.Rosenzweig, S.M.Breedlove og N.V. Watson, Inc. Sunderland Massachusetts, 2005.
- [2] Menneskekroppen, fysiologi og anatomi, Bjålle, Haug, Sand, Sjaastad og Toverud, Gyldendal Norsk Forlag AS, 2004.
- [3] The world of the cell, Becker, Reece og Poenie.
- [4] Cellular physiology of nerve and Muscle, Gary G. Matthews, Blackwell publishing, 2003
- [5] Psychology, Michael W. Passer, Ronald E. Smith, McGraw-Hill, 2001.
- [6] Generell fysikk, Egil Lillestøl, Ola Hunderi og Jan R. Lien, Universitetsforlaget, 2001.
- [7] Lånt av støvneng
- [8] Neuron science, Purves, Augustine, Fitzpatrick, Hall, LaMantia, McNamara, Williams. Inc. Sunderland Massachusetts. 2004.
- [9] Exploring the brain. Bear, Connors, Paradiso. Lippincott Williams & Wilkins, 2001.
- [10] Biological physics, Philiip Nelson, W. H. Freeman ad Company, 2004.
- [11] Physics in Biology and Medicine, Paul Davidovits, Harcourt AcademicPress,2000.

Internett:

- [12] <http://hyperphysics.phy-astr.gsu.edu/hbase/biology/mempot.html>

Batteriladeren

av

Gunnar Skjervold, Magnus Nordling og Magnus Berg Johnsen

**En prosjektoppgave i FY1013 Elektrisitet og magnetisme II
høsten 2005**

Sammendrag

Dette prosjektet tar for seg en batteriladers oppbygning og sammenlikner forskjellige design av batteriladere. Felles for alle batteriladere er likeretting av vekselstrøm og eventuell transformering av vekselstrømmen i forkant. Prosjektet går gjennom flere måter å gjøre dette på og kommer etter hvert fram til to modeller for batteriladere som det gjøres målinger og analytiske betraktninger av.

Innholdsfortegnelse

Sammendrag	2
Innholdsfortegnelse	3
Bakgrunnsteori	5
Transformatoren.....	5
Selvinduktans	5
Jernkjernen	7
Gjensidig induktans	10
Likeretterkretsen	12
Halvledere	12
pn-overgang, diode	13
Likeretterkretser	15
Analytiske betraktninger	19
Ekspirement, målinger	23
Konklusjon	31
Kilder	32

Innledning

Et batteri har lagret kjemisk energi og denne kan hentes ut som elektrisk energi. Dette foregår ved en redoksreaksjon, en kjemisk reaksjon som innebærer overføring av elektroner fra ett stoff til et annet. Når disse kjemiske reaksjonene skjer på to adskilte steder forbundet med en ledning vil elektronene vandre i ledningen mellom de to stedene og dette gir en elektrisk strøm i ledningen. Dette beskriver utladningen av et batteri. I noen batteri er de kjemiske reaksjonene reversible og batteriet kan lades opp igjen og brukes flere ganger. For å lade opp batteriet må vi "tvinge" elektronene til å gå motsatt vei og slik reversere reaksjonen. Dette gjøres ved å sette på en ytre spenning i motsatt retning av spenningen over batteriet. Nettspenningen i Norge er en sinusbølge med frekvens 50Hz og maksimal amplitude $A = \sqrt{2} \cdot 230V \approx 325V$. Gjennomsnittsspenningen fra dette signalet er null, den presser altså like mye på i begge retninger. Ved å omforme vekselspenningen til likespenning i motsatt retning av batterispenningen får den påtrykte spenningen et positivt gjennomsnitt og batteriet lades opp. Mange batterier (blant annet de fleste mobilbatterier) tåler imidlertid ikke så høy spenning som nettspenningen leverer. En batterilader består derfor av to hoveddeler: en transformator og en likeretterkrets. Transformatoren justerer ned spenningen til ønsket verdi og likeretterkretsen omformer vekselspenning til likespenning.

Bakgrunnsteori

Transformatoren

Spenning som blir sendt gjennom høyspentledninger i el-nettet har en spenning på ca 400000V. Skulle denne spenningen komme gjennom en elektrisk maskin som kan ses på som en enkel motstand på ca 2000Ω, så vil det føre til at den får en strøm gjennom seg på 200A. En så kraftig strøm gjennom det fleste apparater vil rett å slett svidd til fillebiter. En transformator brukes til å hindre at slik skjer ved å få endret spenningen slik at den blir mer egnet. En transformator består grovt av tre ting. En primærkrets, en sekundærkrets og en jernkjerne. Primærkretsen består av en spenningskilde og en spole (eventuelt en motsand). Spolen er kveilet omkring en jernkjerne. På den andre siden av kjernen vil det være enda en spole. Denne spolen vil være en del av den sekundære kretsen.

Selvinduktans

Siden spenningskilden i primærkretsen bruker vekselstrøm vil dette føre til at spenningen i kretsen varierer. Dette påvirker spolen slik at det blir lagd elektromagnetisk induksjon eller selvinduktans. Dette kan best forklares hvis en ser på hva som i en spole som går fra en fast strømstyrke til en annen fast strømstyrke. Anta at det går likestrøm gjennom en spole, så vil det være et magnetfelt inne i spolen som blir generert p.g.a. strømmen som går gjennom kretsen. Når strømmen så endres, vil også magnetfeltet inne i spolen endres. Men spolen vil ikke bli endret. Dette fører til at det blir indusert en strøm og retningen avhenger om strømmen i kretsen øker eller minsker. I følge Lenzs lov vil alle induserte effekter motvirke det som genererte induksjonen. Det betyr at indusert strøm vil være i samme retning og som kretsstrømmen om kretsstrømmen minker, og motsatt rettet om kretsstrømmen øker. Men den induserte strømmen vil kun skje så lenge det skjer en endring. Så fort strømdringen slutter, så vil den induserte strømmen forsvinne. Når det er vekselstrøm i kretsen vil det være induksjon hele tiden fordi kretsstrømmen endres hele tiden.

Dette fører også til at det blir indusert en spenning (ems) i spolen. Denne emsen vil være:

$$\mathbf{e} = -L \cdot \frac{dI}{dt} \quad 1.1$$

Og den kan utledes fra Faradays lov og definisjonen av selvinduktans som er:

$$L = \frac{N \cdot \mathbf{f}_B}{I}$$

$$\Rightarrow L \cdot I = N \cdot \mathbf{f}_B \quad 1.2$$

Hvis en så deriverer den magnetiske fluksen og strømmen på tid, så en få:

$$L \cdot \frac{dI}{dt} = N \cdot \frac{d\mathbf{f}_B}{dt} \quad 1.3$$

Faradays lov for spoler er:

$$\mathbf{e} = -N \cdot \frac{d\mathbf{f}_B}{dt} \quad \text{eller} \quad \mathbf{e} = -L \cdot \frac{dI}{dt} \quad 1.4$$

For å finne magnetfeltet inne i spolen kan man bruke Amperes lov som er:

$$\oint \vec{B}_0 \cdot d\vec{l} = B_0 l = \mu_0 n I l \quad 1.5$$

$$\Rightarrow B_0 = \mu_0 n I = \mu_0 I \frac{N}{l}$$

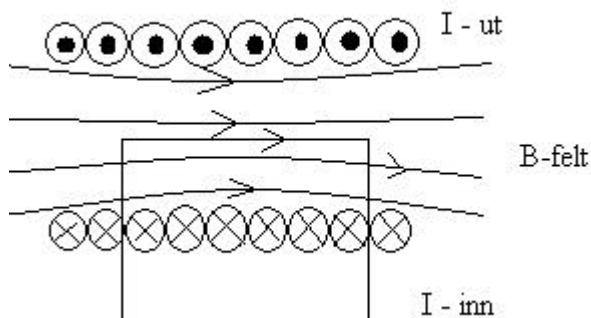


Fig.1.1 Spole med en Amperekurve som en integrerer over

Hvis en så setter det en fikk om magnetfeltet inn i likn.1.1 og vet at den magnetiske fluksen \mathbf{f}_m er lik magnetfeltet multiplisert med arealet av spolen, så vil en finne hva selvinduktansen er:

$$LI = N\mathbf{f}_m = NBA = NA \cdot \mu_0 I \frac{N}{l} \quad 1.6$$

$$\Rightarrow L = \mu_0 N^2 \frac{A}{l}$$

Den effekten som blir dannet i spolen og som det som blir tilført til sekundærkretsen er:

$$P = \mathbf{e}_1 I_1$$

Jernkjernen

Siden vi har en jernkjerne inne i spolen, vil de magnetiske egenskapene til kjernen også påvirke hvor stort magnetfeltet blir.

Disse egenskapene kalles magnetisering, har M som tegn og er definert som:

$$[M] = \frac{\text{magnetisk dipolmoment}}{\text{volum}}$$

og har samme enhet som B/μ_0 , som er A/m.

Det totale magnetfeltet blir da:

$$\vec{B} = \vec{B}_0 + \mu_0 \vec{M} \quad 1.7$$

Men siden magnetfeltet blir forskjøvet så må en bruke en hjelpestørrelse H . Denne hjelpestørrelsen er kun avhengig av magnetfeltet B_0 og er kalles den magnetiske intensiteten.

$$\vec{H} = \frac{\vec{B}_0}{\mu_0} \quad 1.8$$

Der $\mu_0 = 4\pi \cdot 10^{-7}$.

Magnetiseringen, M , være proporsjonalt med den magnetiske intensiteten og er:

$$\vec{M} = c_m \vec{H} \quad 1.9$$

c_m kalles magnetisk susceptibilitet, og er en konstant som avhenger av hva slags materiale som magnetfeltet går igjennom. Ved å nå sette disse to likningene inn i likningen for det totale magnetfeltet så vil en få:

$$\vec{B} = \mu_0 (1 + c_m) \vec{H} \quad 1.10$$

Der $1 + c_m = \mu_r$, som er den relative permeabiliteten.

Grunnen til at en bruker en jernkjerne er at jern er et ferromagnetisk materiale. Jern har da naturlig magnetiske domener inne i seg. Inne i et slikt domene er alle magnetiske dipoler (molekyler som har et positivt overskudd av ladninger i den ene enden og et negativt overskudd av ladninger i den andre enden) rettet i samme retning. Hvert av disse domenene dekker et lite område og er i en tilfeldig retning. Når jernet så blir påvirket av et ytre magnetfelt, i dette tilfellet magnetfeltet fra primærspolen, så starter domenene å snu seg i samme retning som magnetfeltet.

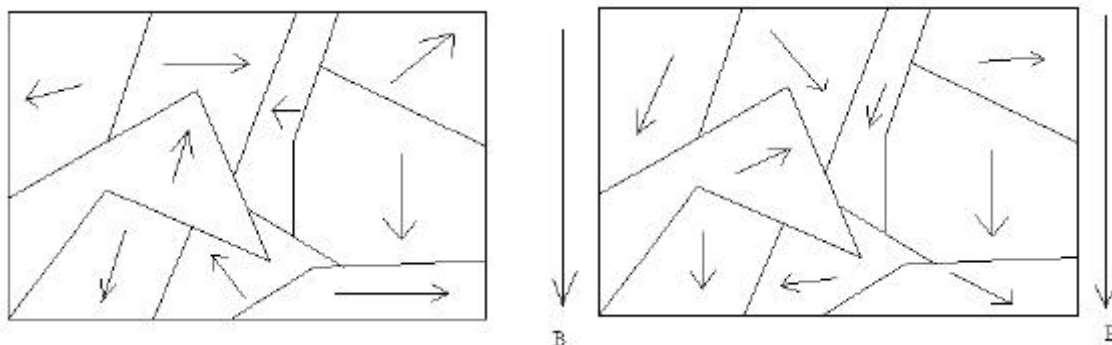


Fig.1.2 Domener som retter seg inn etter magnetfeltet over tid

Til slutt vil alle domene være rettet i samme retning som det påtrykte magnetfeltet B_0 . Siden alle dipolene er nå i samme retning så vil en få et indre magnetisk felt i jernkjernen. Det er dette som er magnetiseringen M , som vil øke styrken på det totale magnetfeltet B som blir lagd.

Siden primærkretsen går på vekselstrøm, så har man det et påtrykt magnetfelt B_0 som endrer seg over tid. Dette gjør at domene også vil endre retning avhengig av tiden slik at det totale magnetfeltet B endres. Hvis en nå ser på sammenhengen mellom magnetfeltet B og den magnetiske intensiteten H når vi slår på strømmen og den går i positiv retning¹, så vil vi se at de to vil vokse til et gitt punkt.

¹. Vi kunne like godt ha sagt at den går mot høyre eller venstre. Vi vil bare ha noe som beskriver hvilken retning strømmen går i ved et gitt tidspunkt siden strømmen i kretsen er $I=I_0\sin(\omega t+\alpha)$

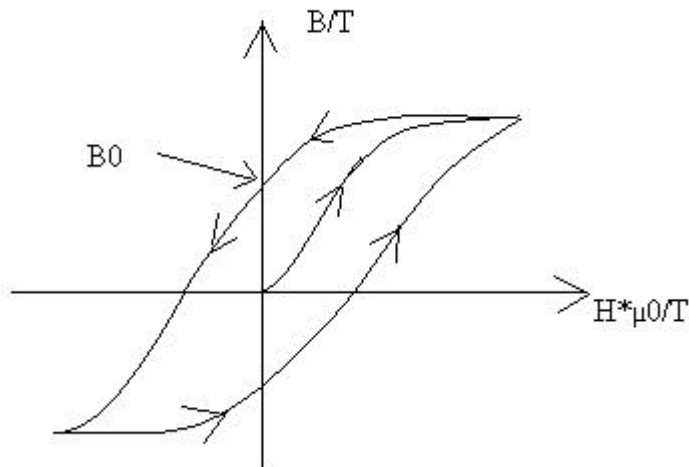


Fig.1.3 Hysteresekurve. Viser sammenhengen mellom magnetisk intensitet og det totale magnetfeltet inne i jernkjernen.

Ved dette punktet vil alle dipoler inne i kjernen være rettet i samme retning slik at alle bidrag til B kommer fra nå kun fra H siden $H=B_0/\mu_0$. Når så strømmen så starter å gå motsatt vei, begynner B_0 å bli mindre. Den magnetiske intensiteten H vil også bli mindre som en følge av dette.

Når så strømmen er blitt null igjen, så er det ingen påtrykt magnetfelt som påvirker jernkjernen. Jernkjernen er nå blitt magnetisk og vil fortsette å være slik hvis den ikke blir påvirket av noe magnetfelt. Men når strømmen starter å gå i negativ retning, så vil også det påtrykte magnetfeltet bli negativ. Det vil føre til at dipolene inne i jernkjernen starter å snu seg i den nye retningen. Når strømmen så er blitt maksimalt negativt så vil alle dipolene ha snudd. Sammenhengen mellom H og B vil heretter oscillere slik figuren viser. Når alle dipolene snur så trengs det energi for få det til. Dette energitapet fører til at det går mindre spenning/strøm gjennom sekundærkretsen, og tilsvarer areal inne i hysteresekurven (fig.1.2).

En annen ting som skjer i kjernen er at det går strømmer inne i den. Disse strømmene, som heter Eddystrømmer; forekommer på grunn av magnetfeltet. Siden magnetfeltet hele tiden endrer seg, fører det til at elektronene vil bli påvirket og starter å bevege seg. Disse strømmene fører til at det blir dannet varme inne i kjernen og dette fører til energi tap. Dette sammen med energitapet p.g.a. magnetisk hysterese fører til at effekten² som kommer til den sekundære spolen er:

$$P_s = P_p - [\text{magnetisk hysterese}] - [\text{Eddystrømmer}]$$

² Effekt = energi per tidsenhet

Gjensidig induktans

Når magnetfeltet inne i jernkjernen går gjennom den sekundære spolen, vil det bli også her bli induisert fluks. Denne fluksen er avhengig av hva antall viklinger i sekundærspolen og strømmen i primærkretsen og er:

$$N_s \mathbf{f}_{21} = M I_p \quad 1.11$$

Dette fører til at emsen som blir produsert er:

$$\mathbf{e}_s = -N_s \cdot \frac{d\mathbf{f}_s}{dt} = -M \frac{dI_p}{dt} \quad 1.12$$

Der M stå for den gjensidige induktansen og er:

$$M = \frac{\mu \cdot A \cdot N_p \cdot N_s}{l_p} \quad 1.13$$

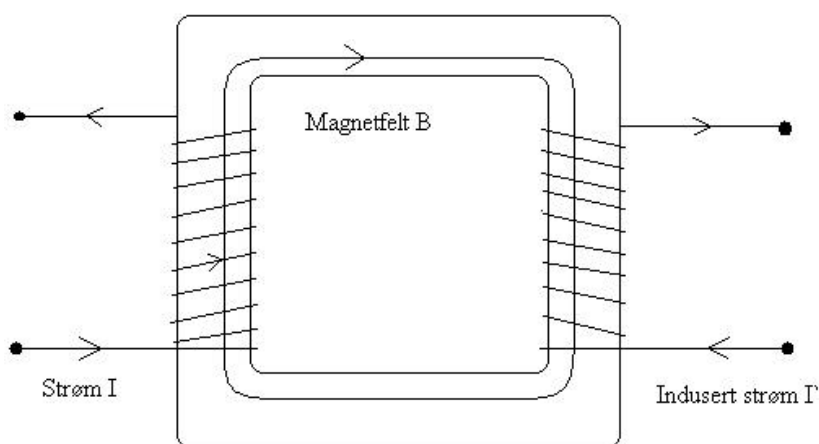


Fig.1.4 Transformator med en strøm I gjennom primærkretsen. Strømmen fører til et magnetfelt som går gjennom jernkjernen og inducerer en strøm Γ i sekundærkretsen.

En ting til som også skjer i sekundærkretsen er at når den gjensidig induisert strømmen går gjennom kretsen, så vil denne strømmen og magnetfeltet føre til selvinduktans inne i sekundærkretsen. Dette gjør at sekundærkretsen tilfører et liten bit til det totale magnetfeltet B, og som igjen påvirker primærkretsen med en gjensidig induktans fra sekundærkretsen. Hvis det er stor forskjell mellom strømmene i de to kretsene (f. eks. 1A i prim. og 10mA i sek.) så kan en se bort ifra disse to induktansene.

Siden kretsen ikke har egen strøm for induksjon, så vil størrelsen til emsen være avhengig av strømmen i primærkretsen. Dette gjør at emsen blir:

$$\mathbf{e}_s = -M \cdot \frac{dI_p}{dt} \quad 1.14$$

Når en nå har likninger for emsen i de to spolene kan en bruke dette til å finne forholdet mellom antall viklinger en trenger for å få ønsket spenning i sekundærkretsen.

$$\frac{\mathbf{e}_s}{\mathbf{e}_p} = \frac{-M \cdot \frac{dI_p}{dt}}{-L \cdot \frac{dI_p}{dt}} = \frac{M}{L} = \frac{\frac{\mu \cdot A \cdot N_p \cdot N_s}{l_p}}{\frac{\mu \cdot A \cdot N_p^2}{l_p}} = \frac{N_s}{N_p} \quad \frac{\mathbf{e}_s}{\mathbf{e}_p} = \frac{N_s}{N_p} \quad 1.15$$

Siden begge spenningene har samme tidsavhengighet, så gjør dette en får følgende forhold:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p} \quad 1.16$$

Likeretterkretsen

Det karakteristiske kretselementet for likeretterkretsen er dioden som er en anvendelse av halvledere.

Halvledere

Halvledere er en gruppe materialer som har en ledningsevne mellom isolator og leder. I en leder er det mange valenselektroner som er løst bundet til krystallen og dermed lett settet i bevegelse av et elektrisk felt, mens hos en isolator er det få eller ingen slike løst bundne elektroner. Et elementært halvledermateriale har typisk kovalente bindinger mellom atomene i stoffet og dette er avgjørende for ledningskapasiteten. Når temperaturen øker begynner atomene å vibrere og med tilstrekkelig kinetisk energi kan valenselektroner løsriveres fra gitterstrukturen og bli tilgjengelige som ledningselektroner slik at materialet kan lede strøm. Man ser altså at ledningsevnen for en halvleder øker med temperaturen i motsetning til ledningsevnen for en leder som minker når temperaturen øker.

Det er imidlertid flere mekanismer som bidrar til strømledningsevnen i en halvleder. Når et elektron løsriver seg fra gitterstrukturen etterlater den et "hull", en mangel på et elektron, i den kovalente bindingen den løsrev seg fra. Dette hullet kan sees på som en positiv ladning siden et annet elektron kan fylle hullet som igjen gir opphav til et hull der dette elektronet var osv. slik at hullet forflytter seg. Når et elektrisk felt påtrykkes stoffet vil da hullet bevege seg med strømretningen mens elektrontransporten går i motsatt retning. Hullene bidrar imidlertid mye mindre til den elektriske ledningsevnen til halvlederen enn de løsrevne elektronene. For at hullet skal bevege seg må et nytt elektron bryte den kovalente bindingen med gitterstrukturen og "hoppe" over i hullet for hver forflytning mens de frie elektronene står mye friere til å bevege seg langs materialet. Hull kan også fylles av frie elektroner som da igjen bli bundet til gitterstrukturen og dermed frata materialet to ladningsbærere. Denne effekten kalles rekombinasjon. Pga at andre elektroner løsriver seg er imidlertid et bestemt antall frie ledningsbærere tilgjengelig pr. volumenhet av et materiale for en bestemt temperatur. Dette tallet kalles materialets intrinsiske konsentrasjon n_i . I en elementær halvleder vil det være like mange hull pr. volumenhet som frie elektroner pr. volumenhet siden hvert frie elektron gir opphav til et hull.

Ved å forurense den rene halvlederen med atomer med ett ekstra valenselektron i forhold til de atomene i den rene halvlederen gir de forurensende atomene ett ekstra fritt elektron hver. Dette øker ledningsevnen betraktelig da tallet på frie ladningsbærere pr. volumenhet øker. Antall hull øker imidlertid ikke i forhold til om lederen hadde vært kun av ett grunnstoff. Denne prosessen kalles doping og halvlederen betegnes som dopet. Analogt kan man dope en ren halvleder med atomer som har ett færre valenselektron enn grunnstoffet halvlederen består av. På den måten får man økt mengden hull pr. volumenhet og dermed også økt halvlederens ledningsevne. Antall frie ladningsbærere pr.

volumenhet fra doping er typisk mye større enn materialets intrinsiske konsentrasjon. Når man doper for å øke mengden fri ladningsbærere får man en n-type halvleder mens doping for å øke antall hull gir en p-type halvleder.

pn-overgang, diode

En diode består av et p-type materiale og et n-type materiale som er satt sammen slik at de danner en pn-overgang, basisen i en halvlederdiode. Materialene er i kontakt over en flate hvor det på p-siden er et overskudd av hull og på n-siden et overskudd av ladning. I nærheten av denne flaten vil da ladningene fylle hullene (rekombinasjon) slik at i nærheten av kontaktflaten er p-siden negativt ladet og n-siden positivt ladet, mens resten av materialet er nøytralt. Området ved kontaktflaten hvor p- og n-siden er ladet kalles sperresonen. Denne skjevheten i ladning gir opphav til et kontaktpotensiale over kontaktflaten og et elektrisk felt i nærheten av kontaktflaten fra n-siden til p-siden. Dette potensialet er materialavhengig men ligger vanligvis på 0,6 til 0,7 Volt. Dette feltet motvirker flyt av elektroner fra n-siden over til p-siden og hull fra p-siden over til n-siden (diffusjonsstrømmen). På n-siden vil det på tross av at det er mange flere fri elektroner enn hull likevel være noen hull som vandrer mot kontaktflaten og blir dyttet over på p-siden av det elektriske feltet. Tilsvarende vil noen elektroner på p-siden bli dyttet over på n-siden av samme felt. Ser altså at feltet som virker mot majoritetsbærerne (de vanligste ladningsbærerne, elektroner i n-type og hull i p-type) virker med minoritetsbærerne (hull i n-type og elektroner i p-type). Denne mekanismen kan observeres som en strøm, pn-overgangens bakovermetningsstrøm eller driftsstrøm, og går motsatt retning av diffusjonsstrømmen. Driftsstrømmen er temperaturavhengig i større grad enn den er avhengig av potensialet over kontaktflaten siden effekten av at de ekstra ladningsbærerne for denne strømmen (hull på n-siden og elektroner på p-siden) øker i antall når temperaturen øker har mer å si enn styrken på feltet.

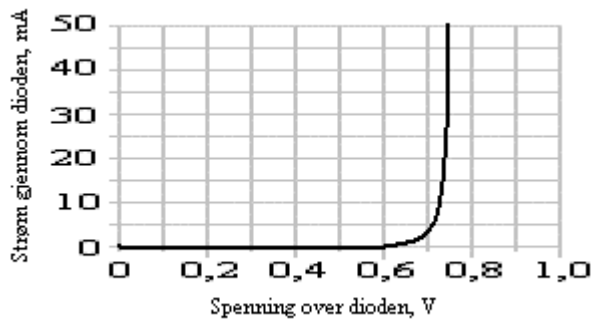
I en krets er oppførselen til halvlederdiode avhengig av hvilken retning spenningen trykkes på. Hvis p-siden av pn-overgangen kobles til batteriets negative pol og n-siden av halvlederen til den positive er pn-overgangen negativt forspent. Resultatet er at sperresonen blir bredere, potensialet over kontaktflaten øker og majoritetsbærerne klarer ikke å komme seg over sperresonen slik at diffusjonsstrømmen ikke kan gå. Den eneste strømmen i diode er da driftsstrømmen. I motsatt tilfelle når pn-overgangen er positivt forspent, dvs. n-siden er koblet til negativ pol på spenningskilden og p-siden til positiv pol, er spenningen fra kilden påtrykt i motsatt retning av spenningsforskjellen over kontaktflaten slik at den påtrykte spenningen reduserer potensialforskjellen over kontaktflaten og hjelper flere majoritetsbærere over sperresonen. Diffusjonsstrømmen er gitt i likning 2.1 hvor v_D er spenningen over pn-overgangen, k er Boltzmanns konstant ($=1.381E-23$ J/K), q er elementærladningen ($1,6E-19$ Coulomb) og T er materialets temperatur i grader kelvin.

$$I_d = I_0 e^{\frac{qv_D}{kT}} \quad (2.1)$$

Diodelikningen, likningen for total strøm gjennom diode når den er forspent (likning 2.2), tar også hensyn til driftsstrømmen I_0 som virker i motsatt retning av

diffusjonsstrømmen. Grafen for strømmen gjennom dioden som funksjon av spenningen for en diode er gitt i figur 2.1. Det går fram av grafen at dioden leder svært godt når spenningen kommer over terskelspenningen på omtrent 0,75V og for lavere spenninger leder svært dårlig.

$$i_D = I_d - I_0 = I_0 \left(e^{\frac{qv_D}{kT}} - 1 \right) \quad (2.2)$$

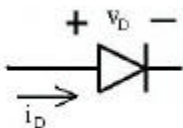


Figur 2.1: Strøm som funksjon av spenning gjennom en diode

Driftsstrømmen er imidlertid veldig liten (størrelsesorden 10^{-9} A til 10^{-15} A) slik at for diodespenninger v_D større en et par tidels volt kan likning 2.3 brukes med tilfredsstillende nøyaktighet.

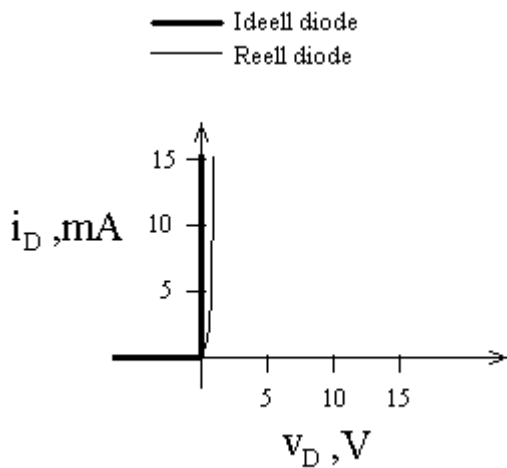
$$i_D = I_0 e^{\frac{qv_D}{kT}} \quad (2.3)$$

En halvlederdiode består av en pn-overgang og en ohmsk kontakt i hver ende (en på n-siden og en på p-siden). Egenskapene til dette kretselementet er, som det framgår av informasjonen om pn-overgangen, at det bare leder strøm når strømmen går én bestemt vei (når dioden er positivt forspent). Vel og merke med forenklingen at den marginale driftsstrømmen i det negativt forspente tilfellet ses bort fra. Kretselementet til dioden med strømretning og diodespenning i positivt forspent tilfelle er tegnet i figur 2.2.



Figur 2.2: Kretselement for diode med diodespenning og strømretning i positivt forspent tilfelle.

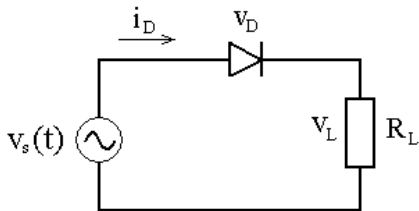
En diode er altså et kretselement som skal lede strøm best mulig når det er en spenning som gjør at dioden er positivt forspent og ikke lede noe strøm når den er negativt forspent. Når dioder modelleres er dermed en modell av en tenkt perfekt diode et nærliggende valg. Dette kalles den ideelle diodemodellen og en slik ideell diode kan betraktes som et brudd i kretsen når $v_D < 0$ og en kortslutning når $v_D \geq 0$. Strømmen gjennom dioden som funksjon av spenningen for en ideell diode og en reell diode er gitt i figur 2.3.



Figur 2.3: Strøm som funksjon av diodespenning for ideell og reell diode

Likeretterkretser

Likeretterkretser kan lages på flere måter. Kretstegning for en veldig enkel likeretterkrets er gitt i figur 2.4.



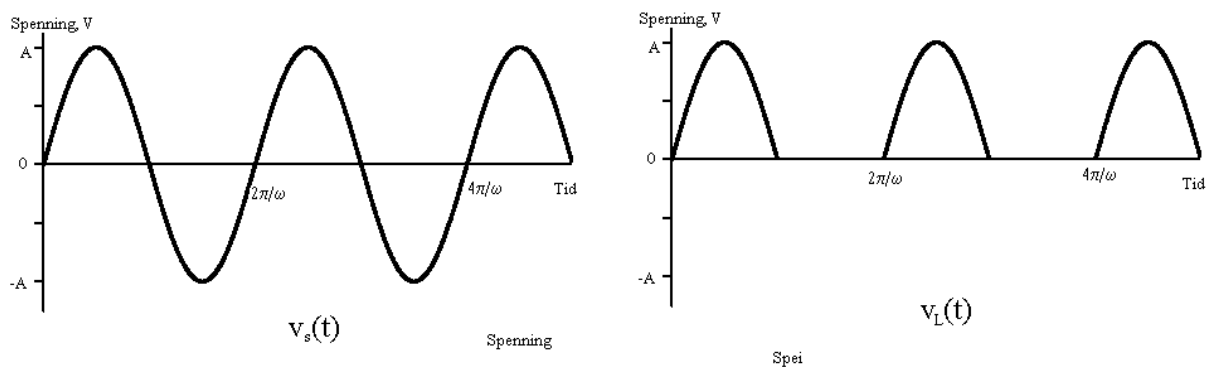
Figur 2.4: Kretstegning for enkel likeretter

Ved hjelp av ohms lov, $V=RI$, og å benytte ideell diodemodell kan strømmen i_D og spenningen v_L regnes ut når $v_s(t)$ er kjent. Den ideelle diodemodellen leder bare strøm når $v_D \geq 0 \Rightarrow v_s \geq 0$ slik at

$$i_D = 0 \text{ når } v_s < 0$$

$$v_s = R_L \cdot i_D \Rightarrow i_D = \frac{v_s}{R_L} \text{ når } v_s \geq 0$$

Siden den ideelle dioden er en kortslutning for $v_D \geq 0$ og brudd i kretsen for $v_D < 0$ blir $v_L = 0$ for $v_s < 0$ og $v_L = v_s$ for $v_s \geq 0$.

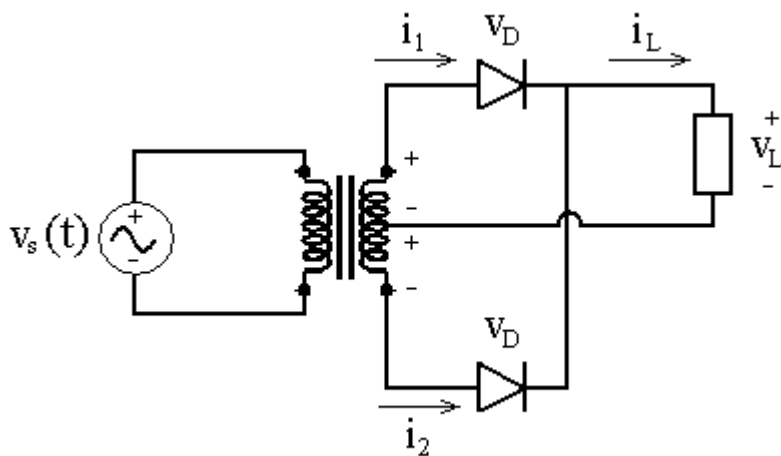


Figur 2.5: Spennning over motstand og utgangsspenning $v_s(t) = A \sin(\omega t)$

Gjennomsnittsspenningen finner man ved å integrere over en periode, f. eks. fra $t = 0$ til $t = 2p / \omega$ og dele på lengden av perioden. Det går klart frem av figur 2.5 at gjennomsnittet for $v_s(t)$ er null. Gjennomsnittet for $v_L(t)$ er gitt ved:

$$v_L(t) = \frac{\omega}{2p} \int_0^{p/\omega} A \sin(\omega t) dt = \frac{A}{\pi}$$

Selv om kretsen likeretter signalet er den ikke spesielt effektiv siden energien i den negative delen av sinussignalet ikke utnyttes. For å også utnytte den negative delen må man bruke en helbølgeretter, en type helbølgeretter er vist i figur 2.6.



Figur 2.6: Helbølgeretter.

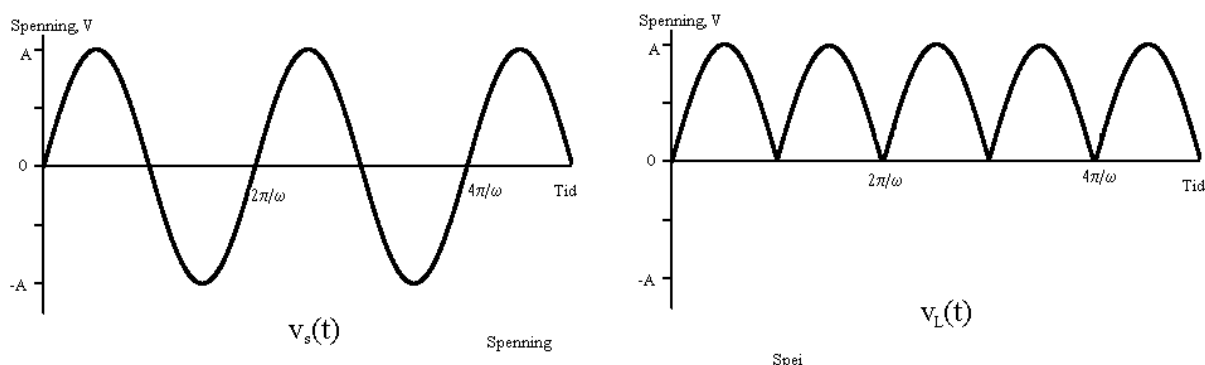
Denne kan deles opp i to kretser med en transformator mellom. Primærkretsen er kretsen hvor spenningskilden og en spole fra transformatoren inngår. Sekundærkretsen består av to transformatorspoler i serie hvor kontaktpunktet mellom de to spolene har en leder som går ut til resten av sekundærkretsen. Diodene er plassert på en slik måte at når strømmen går én vei gjennom transformatoren blokkerer en av diodene strømmen mens den andre leder strømmen inn på lastmotstanden mens når strømmen bytter de to diodene rolle. Dermed blir strømmen inn på lastmotstanden alltid positiv. De to spolene i sekundærkretsen har likt

antall viklingstall og det er gitt som 1:N i forhold til spolen i primærkretsen. Strømmen i_L gjennom motstanden blir da i de to tilfellene gitt som:

$$i_L = i_1 = \frac{Nv_s}{R_L} \text{ for } v_s \geq 0$$

$$i_L = i_2 = \frac{-Nv_s}{R_L} \text{ for } v_s < 0$$

retningen for i_L er likevel alltid positiv på grunn av plasseringen av diodene (se figur 2.6) Spenningene $v_s(t)$ og $v_L(t)$ er gitt i figur (2.7)



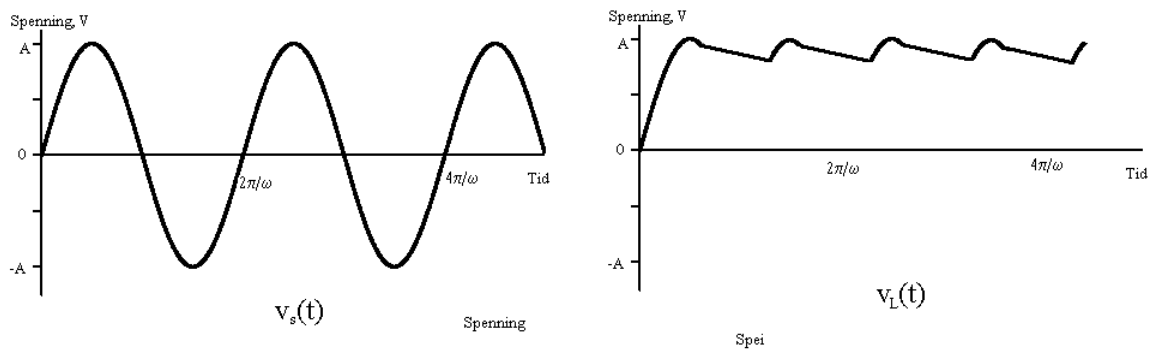
Figur 2.7: Spennning over spenningskilde $v_s(t)$ og over lastmotstand $v_L(t)$ for helbølgeretterer

Intuitivt bør denne likeretteren gi mye bedre energiutnyttelse enn halvølgeretteren og ved å betrakte gjennomsnittsspenningen bekreftes dette:

$$v_L(t) = \frac{w}{2p} \left(\int_0^{p/w} A \sin(wt) + \int_{p/w}^{2p/w} A \sin(wt - p/w) \right) = 2 \frac{A}{p}$$

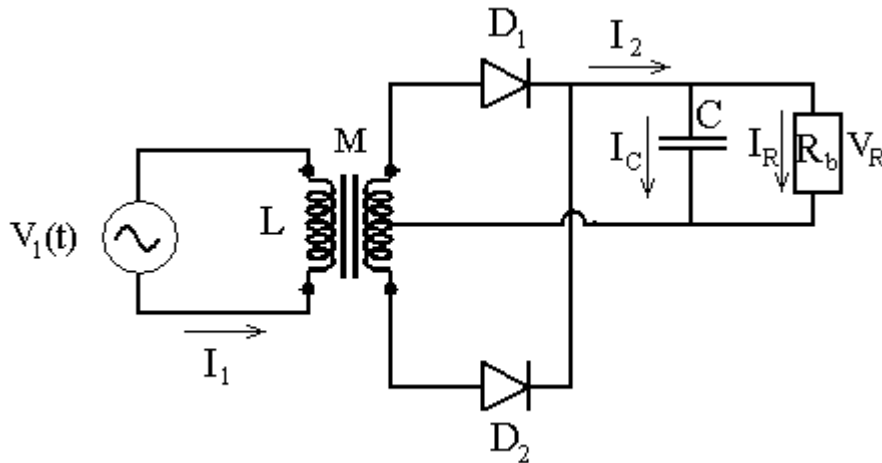
Det går fram av resultatet at helbølgeretteren gir dobbelt så høy gjennomsnittsspenning som halvølgeretteren.

Fra figur 2.7 går det fram at selv om signalet har fått en positiv gjennomsnittsspenning og aldri blir negativt alternerer signalet likevel mye. Det er derfor ønskelig å redusere denne variasjonen og få jevnest mulig likespenning som resultat. En måte å gjøre dette på er å inkludere en kondensator i likeretterkretsen som et enkelt lavpassfilter. Da lades kondensatoren opp når spenningen fra transformatoren er større enn kondensatorspenningen og ut når spenningen fra transformatoren er lavere enn kondensatorspenningen. Dermed blir ripplespenningen, variasjonen om gjennomsnittsverdien for dc-strømmen, jevnet betraktelig ut (se figur 3.1 for kretstegning for en slik krets og figur 2.8 for spenningskarakteristikk).



Figur 2.8: Spennning over spenningskilde $v_s(t)$ og over lastmotstand $v_L(t)$ for helbølgeretter med filter

Analytiske betraktninger



Figur 3.1: Kretstegning for referanse under utregninger

For primærkretsen (se figur 3.1) gjelder følgende likhet;

$$V_{1,0}e^{i\omega t} - L\dot{I}_1 - R_1I_1 = 0,$$

hvor $V_{1,0}$ er signalgeneratorens amplitude, L er selvinduktansen i spole 1, R_1 er resistansen i primærkretsen, ω er fasehastigheten og I_1 er strømmen i primærkretsen. Den fysiske spenningen er da;

$$Fysisk \quad V_1 = \text{Im}\langle V_1 \rangle = V_{1,0} \sin \omega t.$$

Dette skulle resultere i en strøm I_1 gitt ved;

$$I_1 = \frac{V_{1,0}}{Z} e^{i\omega t},$$

hvor impedansen Z er;

$$Z = R_1 + i\omega L.$$

Den fysiske strømmen finnes ved;

$$Fysisk \quad I_1 = \text{Im}\langle I_1 \rangle = \frac{V_{1,0}}{|Z|} \sin(\omega t - \mathbf{a}),$$

hvor

$$|Z| = \sqrt{R_1^2 + (\omega L)^2} \text{ og}$$

$$\mathbf{a} = \arctan\left(\frac{\omega L}{R_1}\right).$$

For $\omega L \gg R_1$ vil $\mathbf{a} \rightarrow \frac{\mathbf{p}}{2}$.

Den gjensidige induktansen fra primærkretsen til sekundærkretsen genererer en induisert spenning i sekundærkretsen, V_2 , som uttrykkes ved;

$$V_2 = M \dot{I}_1 = Mi\omega I_1,$$

noe som skulle resultere i en fysisk spenning;

$$\text{Fysisk } V_2 = \text{Im}\langle Mi\omega I_1 \rangle = M\omega \frac{V_{1,0}}{|Z|} \sin(\omega t - \mathbf{a} + \frac{\mathbf{p}}{2}).$$

Med $\omega L \gg R_1$ og $\mathbf{a} \rightarrow \frac{\mathbf{p}}{2}$ vil vi da kunne gjøre følgende forenkling;

$$\text{Fysisk } V_2 = \frac{M}{L} V_{1,0} \sin \omega t = \frac{N_2}{N_1} \cdot \text{Fysisk } V_1,$$

hvor N_i er viklinger i spole i , noe som skulle stemme bra.

For kretssløyfen med kondensatoren skulle vi da få denne sammenhengen etter å ha derivert med hensyn på tiden t ;

$$M \dot{I}_1 - \frac{1}{C} I_C = i\omega V_2 - \frac{1}{C} I_C = 0 \Rightarrow I_C = i\omega C V_2 = -M\omega^2 C \frac{V_{1,0}}{Z} e^{i\omega t}.$$

$$\text{Fysisk } I_C = \text{Im}\left(-M\omega^2 C \frac{V_{1,0}}{Z} e^{i\omega t}\right) = \frac{N_2}{N_1} \omega C V_{1,0} \cos \omega t.$$

Lignende for sløyfen med lastmotstanden;

$$M \dot{I}_1 - R_b I_R = V_2 - R_b I_R = 0 \Rightarrow I_R = \frac{V_2}{R_b} = \frac{Mi\omega \frac{V_{1,0}}{Z} e^{i\omega t}}{R_b},$$

hvor R_b er lastmotstanden. Videre;

$$\text{Fysisk } I_R = \text{Im}\left(\frac{Mi\omega \frac{V_{1,0}}{Z} e^{i\omega t}}{R_b}\right) = \frac{1}{R_b} \frac{N_2}{N_1} V_{1,0} \sin \omega t.$$

Dette skulle totalt gi en strøm I_2 ;

$$\text{Fysisk } I_2 = \frac{N_2}{N_1} V_{1,0} \left(\frac{1}{R_b} \sin \omega t + \omega C \cos \omega t \right)$$

Når det ikke lenger går strøm gjennom dioden, vil man få følgende sammenheng;

$$I_C + I_R = 0 \Rightarrow \tan \omega t_0 = -\frac{\omega}{\omega_0}, \quad \omega_0 = \frac{1}{R_b C},$$

$$t_0 \rightarrow \frac{P^+}{2\omega} \text{ når } \omega R_b C \text{ stor.}$$

Når $t = t_0$ vil det ikke lenger gå strøm gjennom dioden og vi vil få utladning av kondensatoren gjennom lastmotstanden. Dette gir følgende likning;

$$R_b I_{\text{utl}} = -\frac{Q}{C},$$

hvor I_{utl} er utladningsstrømmen. Dette gir løsning;

$$V_R(t) = V_{2,0} e^{-\omega_0 t},$$

hvor $t = t - t_0$.

Strømmen som denne spenningen gir opphav til vil derfor ta over når strømmen fra spolen forsvinner. Når spenningen over spolene igjen blir større enn kondensatorspenningen vil det igjen være denne som driver kretsen, strøm vil passere gjennom dioden og vi vil få samme situasjon som vi først hadde.

Tiden det tar før denne situasjonen igjen oppstår finner vi ved å konsultere de forskjellige grafene; intervallet mellom et toppunkt og der spenningen over lastmotstanden igjen stiger med en trigonometrisk form skulle gi tidsrommet t_{utl} hvor det ikke går strøm gjennom dioden.

Vi ønsker å benytte en helbølgeretter. Dette gir en litt annen situasjon; da vil utladningsstrømmen raskere bli erstattet av strøm fra spolene; tidsintervallet med utladning vil bli betydelig kortere. Dette gir dermed en mindre ripple-spenning.

Ved små ripple-spenninger kan denne t_{utl} settes lik halve perioden til signalet fra

$$\text{signalgeneratoren } t_{\text{utl}} \approx \frac{T}{2}.$$

Et uttrykk for spenningen over lastmotstanden er³;

$$V_R \approx V_{2,0} \left(1 - \frac{t_{\text{utl}}}{R_b C} \right) \approx V_{2,0} \left(1 - \frac{T}{2R_b C} \right).$$

Ripple-spenningen blir da;

$$V_{\text{ripple}} = V_{2,0} \frac{T}{2R_b C},$$

³ <http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/rectct.html#c4>

og rms-verdien blir for en slik sagtannet funksjon;

$$V_{\text{ripple-rms}} = V_{2,0} \frac{T}{2 \cdot 2 \cdot \sqrt{3} R_b C} = V_{2,0} \frac{T}{4 \cdot \sqrt{3} R_b C}.$$

Den likestrømmen vi da bør kunne få er;

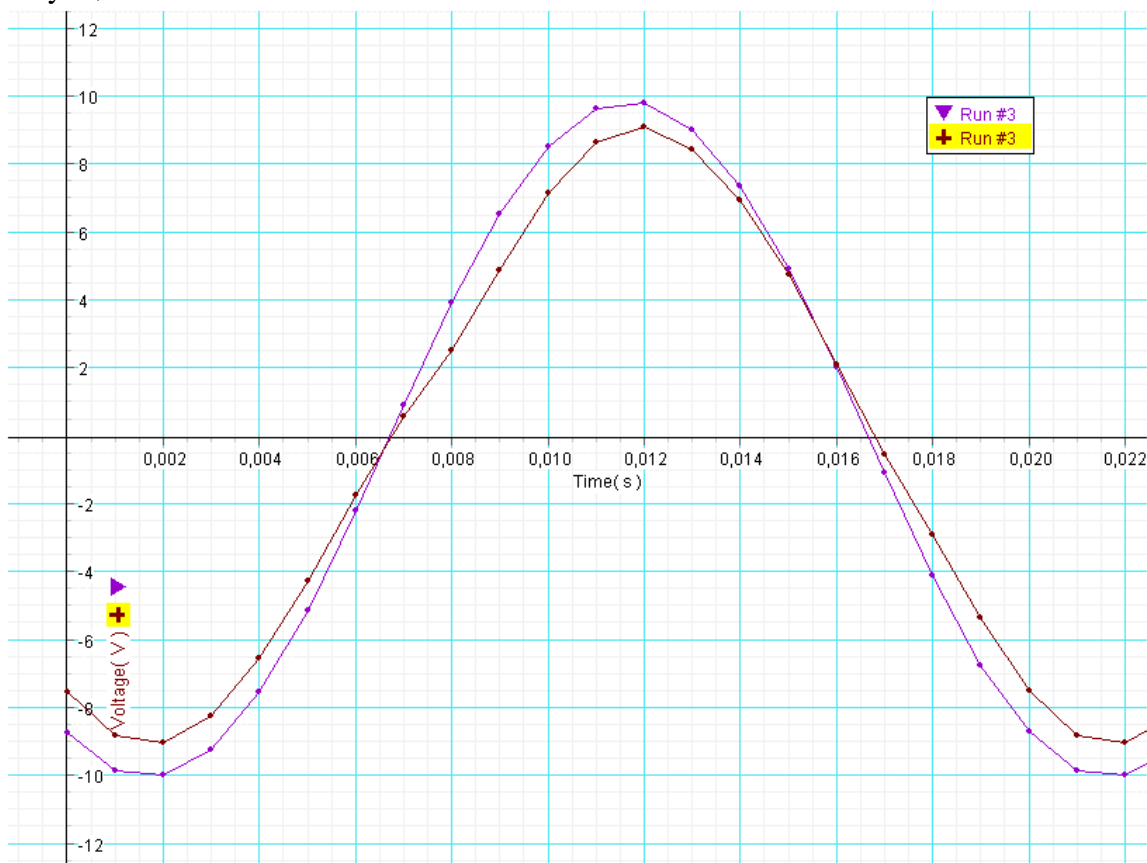
$$V_{DC} = V_{2,0} - \frac{V_{\text{ripple}}}{2} = V_{2,0} \left(1 - \frac{T}{4 R_b C}\right).$$

I disse utregningene gjør vi flere forenklinger;

- vi ser bort fra selvinduktans i spole 2 og gjensidig induktans fra sekundærkretsen til primærkretsen,
- vi betrakter dioden som ideell; intet spenningsfall når det går strøm og det går strøm så fort spenningsforskjellen er større enn 0..

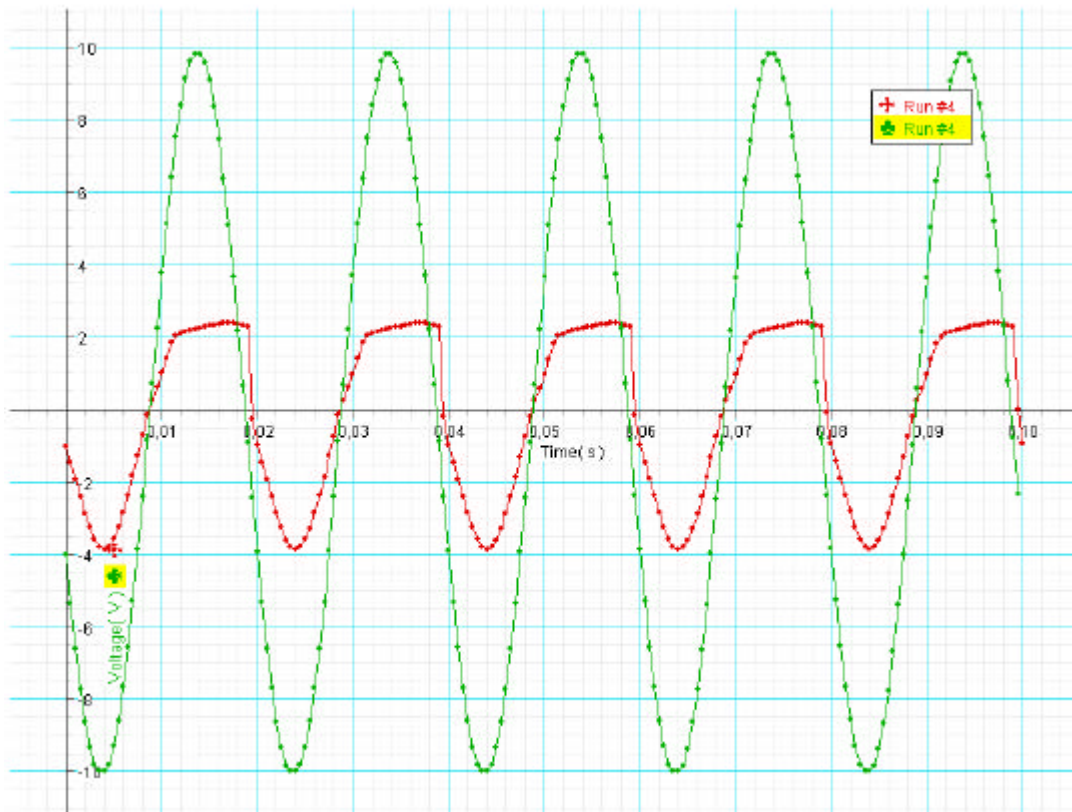
Eksperiment, målinger

Under forundersøkelsene så vi på effekttap i en transformator, figur 4.1 skulle vise at det er et lite tap når vi har en primærkrets og sekundærkrets med like mange viklinger på de to spolene. Dette kan komme av mange ting; flukstap, magnetisk hysteresis og eddystrømmer er blant disse.

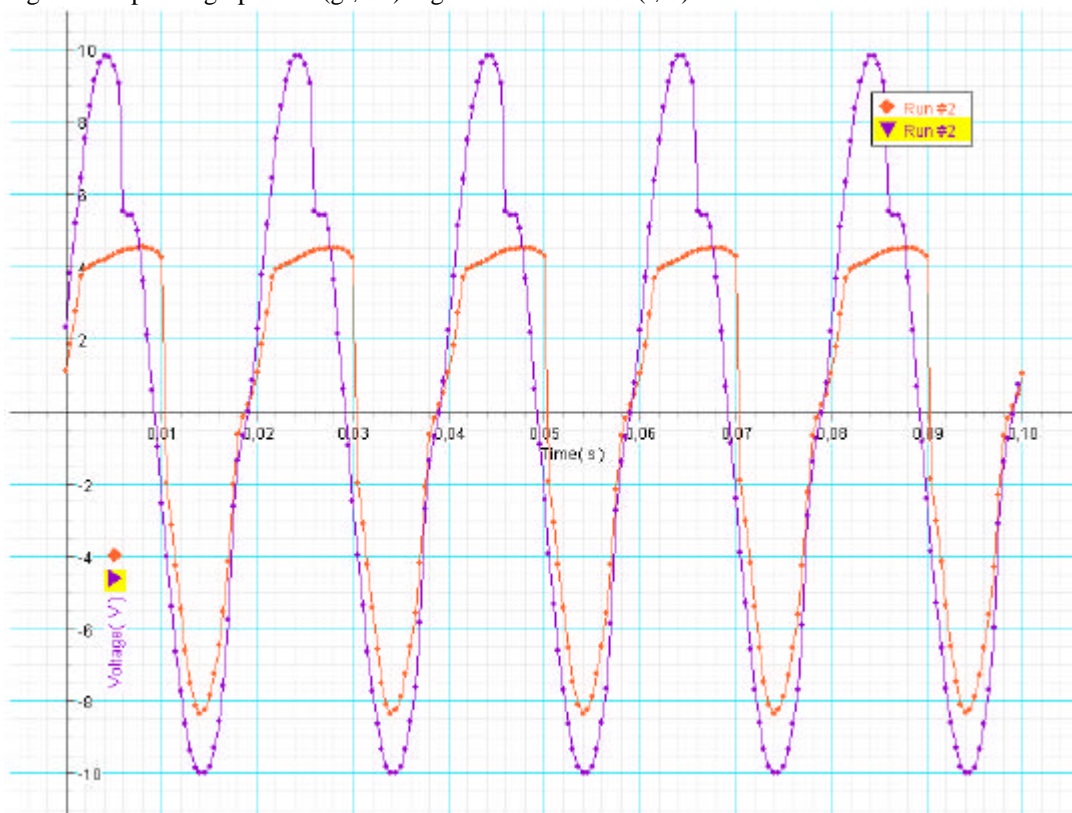


Figur 4.1: Spenning i primær (lilla)- og sekundærkretsen (brun)

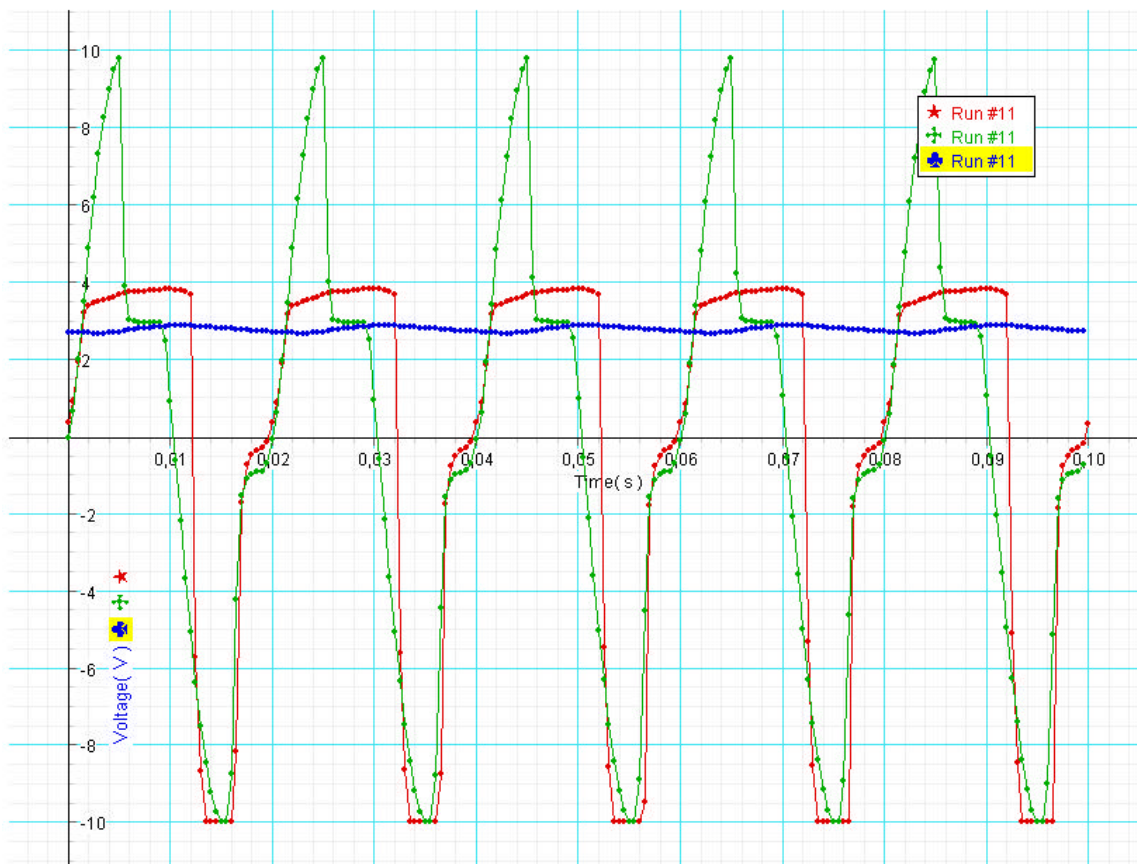
En viktig antakelse vi gjør i eksperimentene våre med transformatorer er at den gjensidige induktansen fra sekundærkretsen til primærkretsen er neglisjerbar; med andre ord at strømmene og spenningene i primærkretsene ikke blir påvirket av hva som foregår i sekundærkretsen. Av figur 4.2 til 4.4 kan vi se at sannhetsgehalten i dette er avhengig av hvor mye vi transformerer ned eller opp; for sterke spenninger i sekundærkretsen ser vi at primærkretsens signal blir forstyrret, dette må skyldes gjensidig induktans fra sekundærkretsen.



Figur 4.2: Spenning i primær (grønn)- og sekundærkretsen (rød)

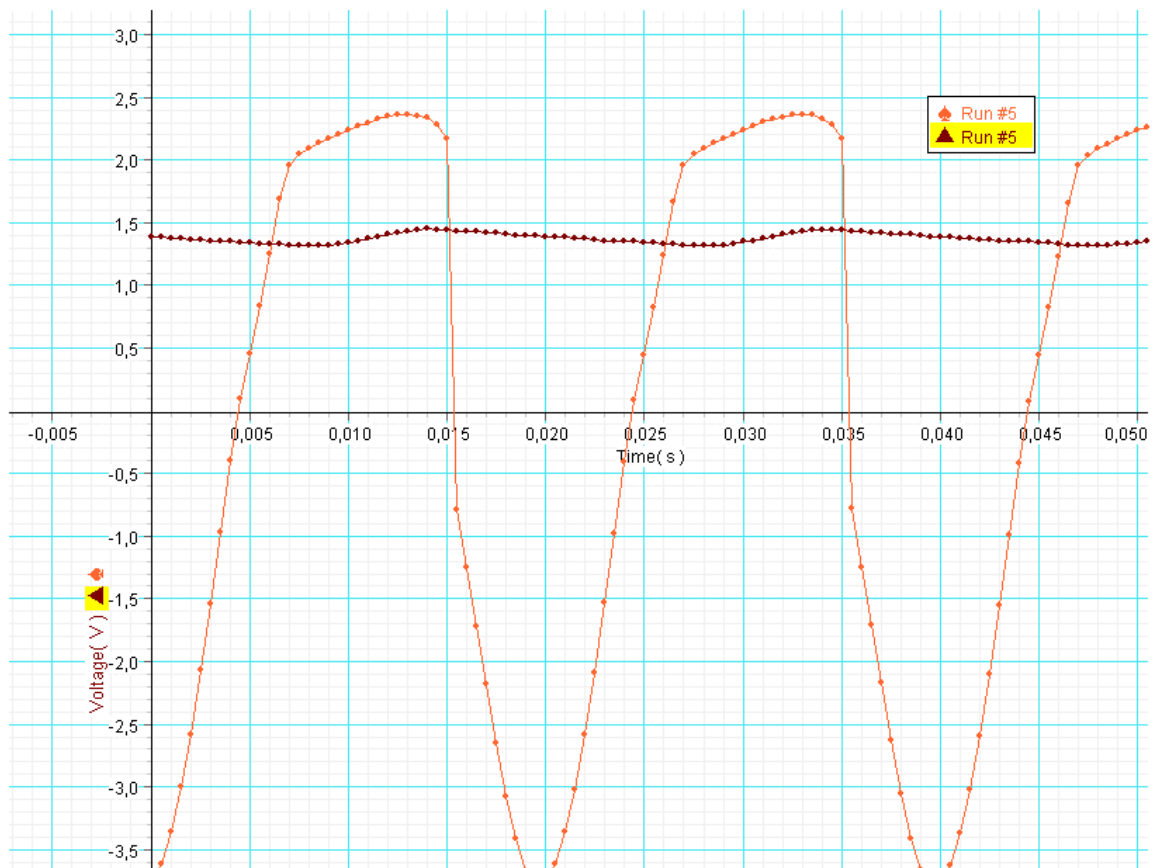


Figur 4.3: Spenning i primær (lilla)- og sekundærkretsen (oransje)

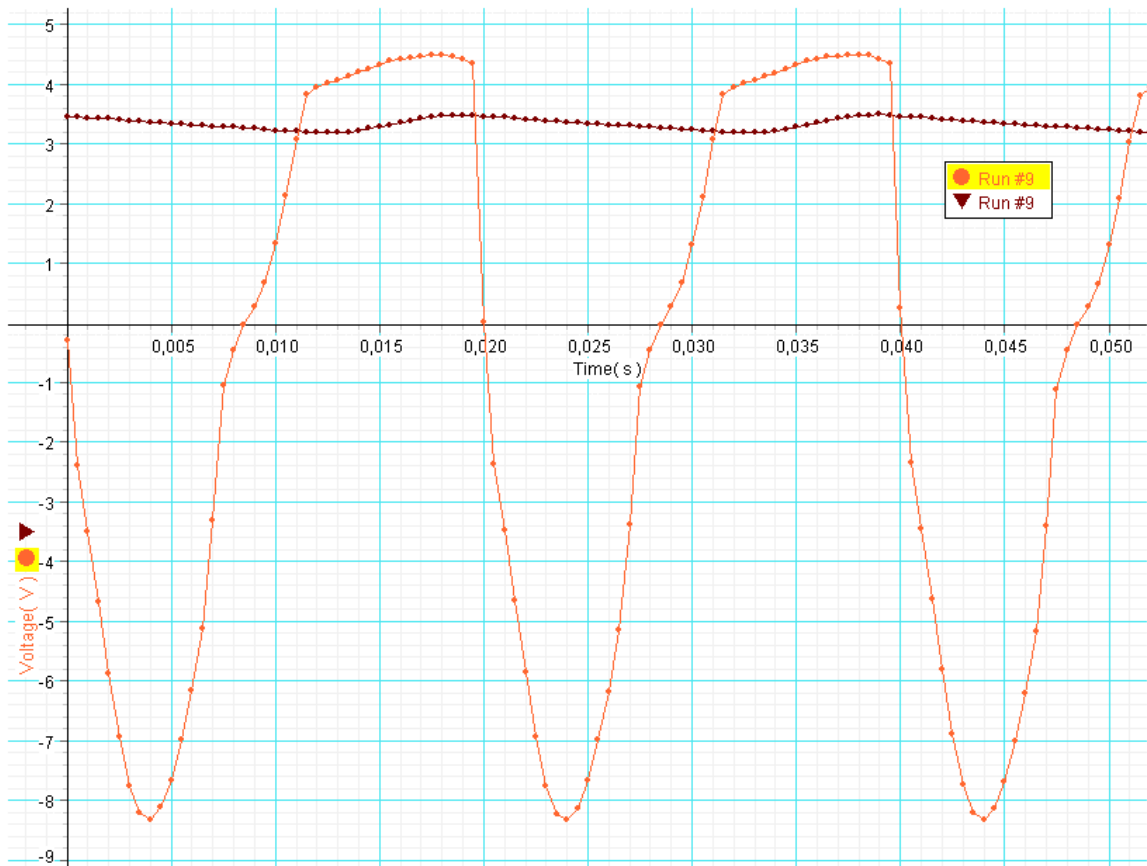


Figur 4.4: Spenning i primær (grønn)- og sekundærkretsen (rød). I tillegg vises her spenningen (blå) over lastmotstanden.

Det at vi ikke opererer med ideelle dioder gjør at spenningen i batteriet ikke blir like stort som forventet; vi observerer et spenningsfall. Fra figurene 4.5 og 4.6, som viser spenningen over batteri og spoler for to forskjellige spolespenninger, ser vi at dette fallet er rimelig konstant lik 0,7-1,0V, noe som skulle stemme godt overens med at dioder ikke leder strøm før en terskelverdi for spenningen.



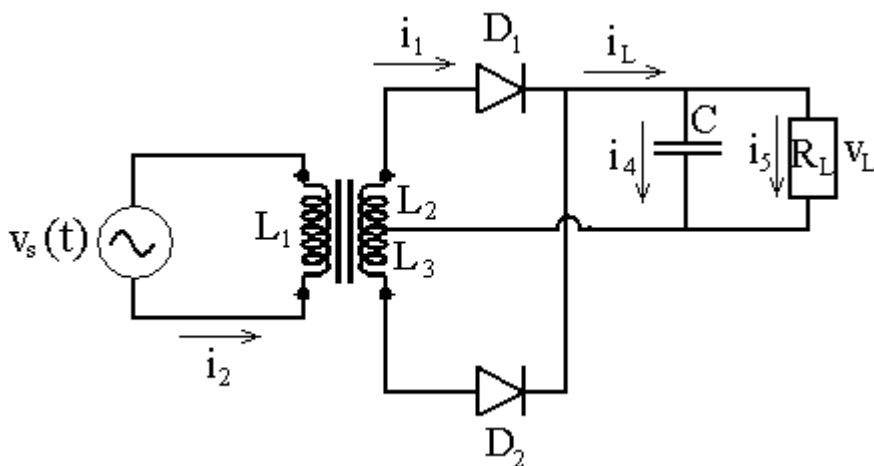
Figur 4.5: Spenning over spole 2 (rød) og lastmotstanden (brun) med lav spolespenning



Figur 4.6: Spenning over spole 2 (rød) og lastmotstanden (brun) med høyere spolespenning

Ut fra de teoretiske betraktningene ble to ladekretser testet. Ladekrets A består av en transformator og en helbølgelikeretter med et antiripplefilter. Ladekrets B består av en transformator og en halvølgelikeretter også den med samme antiripplefilter.

Ladekrets A ble koblet opp som vist på figur 4.7 med verdier gitt i tabell 4.8.



Figur 4.7: Ladekrets A, transformator og helbølgelikeretter med filter

Tabell 4.8: Verdier på kretselementer for krets i figur 4.1

$v_s(t)$ Volt	R_L Ohm	C Farad	L_1 Viklinger	L_2 Viklinger	L_3 Viklinger	Kjerneareal m^2
$v_s(t) = \sqrt{2} \cdot 7 \cos(100\pi t)$	15	$10E-3$	400	200	200	$3,6E-4$

De to diodene, D_1 og D_2 , er like. Pasco dataloggingsutstyr ble brukt for å utføre målinger og datastudio ble brukt for å framstille målingene grafisk. Spenningen over lastmotstanden (v_L) og spenningen over de to spolene L_2 og L_3 ble målt, fremstilt grafisk (figur 4.9) og sammenliknet med de teoretisk forventede verdiene.

Vi ser umiddelbart at det er visse avvik fra forventede verdier. For det første er amplituden til spenningen over motstanden mindre enn spenningen over spole 2 og 3. Det andre avviket sees i formen på spenningen over de to spolene; denne er ikke helt trigonometrisk.

Det første kan forklares med at dioden ikke er ideell; det at den ikke leder strøm før visse spenningsverdier (se figur 2.1) gjør at det i praksis blir et spenningsfall over dioden (her må også andre spenningsfall i kretsen også ta sitt ansvar), her på ca 0,9V.

Det andre kan forklares med at selvinduktansen i spole 2 og 3 ikke er helt neglisjerbar, samt at kondensatoren motvirker det elektriske feltet.

Vi vurderte så strømmene i primær-og sekundærkretsen. I primærkretsen går det kontinuerlig strøm, generert av signalgeneratoren. I sekundærkretsen går det bare strøm når kondensatoren lades opp. Strømmen til kondensatoren er ganske beskjeden; den er nesten oppladet. Relativt store deler av I_2 går derfor gjennom batteriet/lastmotstanden.

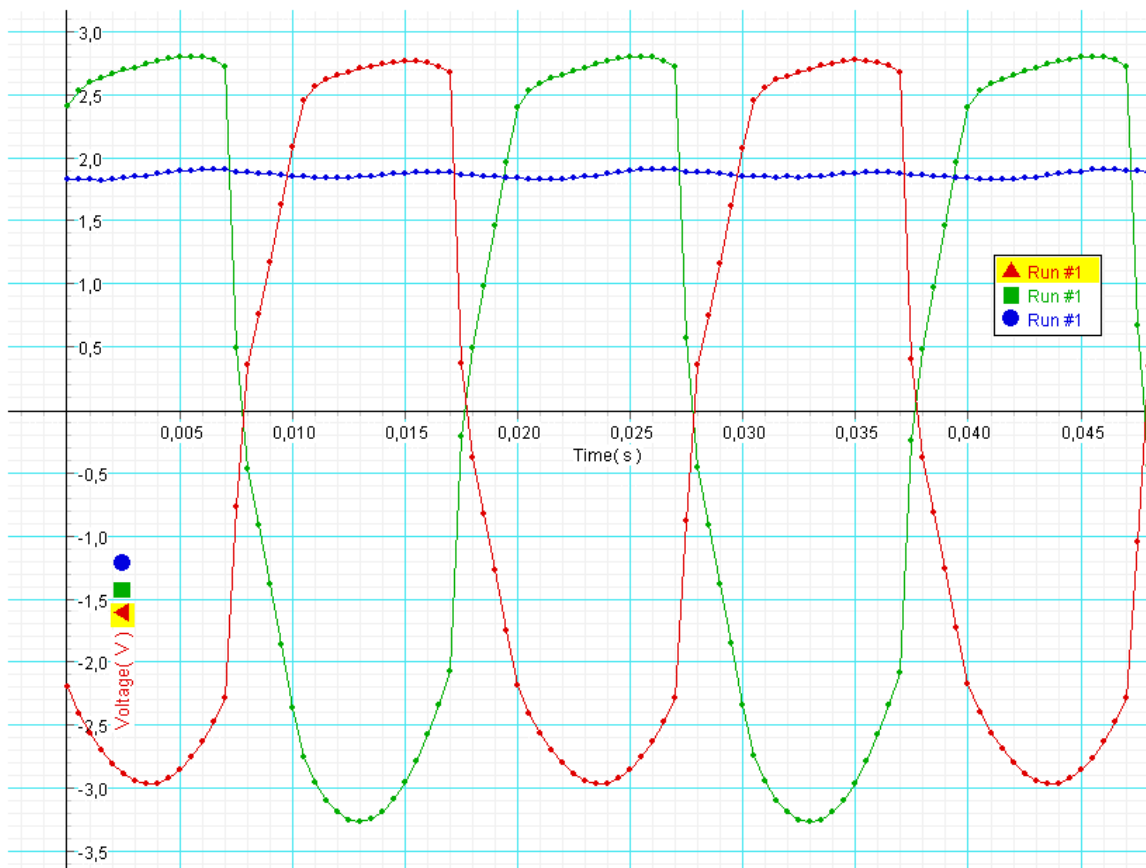
Strømmen i sekundærkretsen blir derfor rimelig lav, selv om uttrykk ovenfor ikke skulle gi dette. Men, som sagt, flatheten til figur 4.2 skulle indikere et visst selvinduktans-bidrag som reduserer spenningen over spoler 2 og 3.

For denne kretsen skulle beregnet ripple-spenning, $V_{ripple-rms}$, bli;

$$V_{ripple-rms} = \frac{N_2}{N_1} V_{1,0} \frac{T}{4\sqrt{3}R_b C} = \frac{200}{400} 7,0V \frac{\frac{s}{50}}{4\sqrt{3} \cdot 15\Omega 10mF} = 0.096V$$

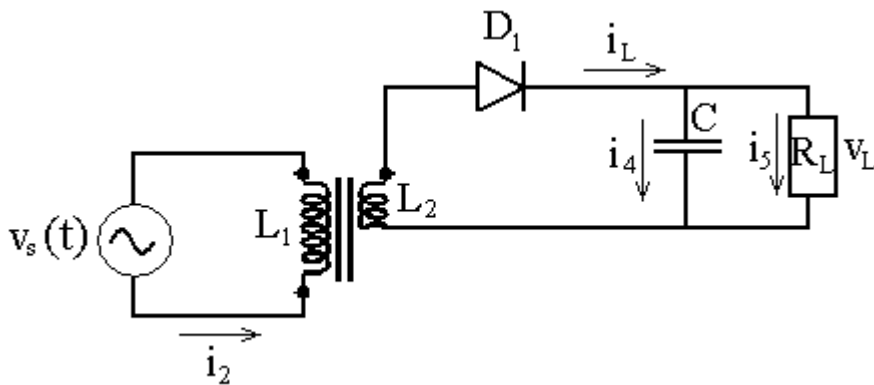
Denne verdien er på ingen måte uakseptabel i forhold til målte verdier som vi kan lese ut fra figur 4.9; denne ser man er i den samme størrelsesorden som den beregnede; vi vil anslå den til å være omtrent 0,06V. I tillegg må man ta hensyn til at den reelle spenning over batteriet er noe mindre enn den beregningene gir på grunn av fall over dioden.

Forholdet mellom spolespenningen og spenningen over batteriet er omtrent 1,5; dermed skulle de teoretiske verdier og de faktiske ligge svært nær hverandre.

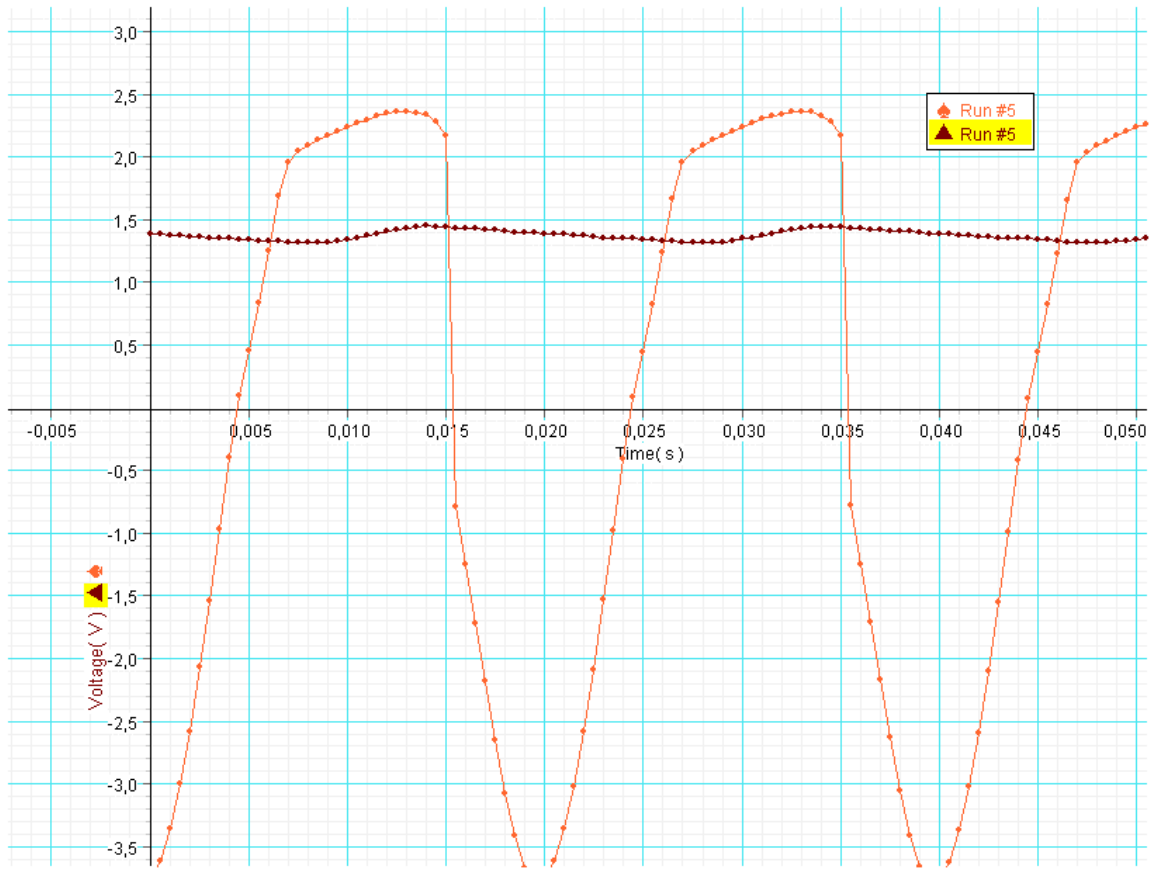


Figur 4.9: Spenning over motstand (blå), spenning over spole2 (rød), spenning over spole3 (grønn) som funksjon av tid

En enklere ladekrets, figur 4.10, ble koblet opp (Ladekrets B) med samme verdier på komponentene som ladekrets A gitt i tabell 4.8. Også her ble spenning over lastmotstanden (v_L) og spenningen over spole L_2 målt og framstilt grafisk (figur 4.11). Vi ser her en liknende graf med den forskjell at utladningstiden er langt større. Dette fører også til at ripple-spenningen blir større.



Figur 4.10: Ladekrets B, enklere ladekrets, halvbølgeretter med filter



Figur 4.11: Spenning over lastmotstand, spenning over spole2 som funksjon av tid

Konklusjon

Av de likeretterkretsene vi har undersøkt, gir helbølgelikeretteren best utnyttelse av spenningen og jevnest ripplespenning. Dette er også forventet fra teorien. Enkelte teoretiske utregninger for kretsen hadde ikke like stort samsvar, mye på grunn av at vi regner med en ideell diodemodell og ser bort fra selvinduktans i spolene i sekundærkretsen og gjensidig induktans fra sekundær til primærkrets. Disse avvikene kunne imidlertid blitt redusert om vi hadde brukt nettspenning som har mye høyere amplitude (~230V rms) enn signalgeneratoren vi brukte (7V rms) og dermed ville spenningen blitt mer drastisk transformert ned og forholdet mellom strømmene i sekundær og primær kretsen blitt mer gunstig med hensyn på å redusere gjensidig induktans.

Kilder:

Bind 2 av Generell fysikk av Egil Lillestøl, Ola Hunderi og Jan R. Lien
University Physics 11.ed av Freedman og Young
Principles and Applications of Electrical Engineering av Giorgio Rizzoni
General Chemistry av Raymond Chang

Internettsider:

Forelesningsnotater i FY1013 (høst 2005) av Jon Andreas Støvneng
http://web.phys.ntnu.no/~stovng/FY1013_2005/fy1013.htm#notater

Georgia State University, Hyperphysics, center-tapped full wave rectifier
<http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/rectct.html#c4>

Filterkretser

Prosjekt i faget FY1013

Høsten 2005

av

Jon Alm Eriksen og Lodve Brastad

Abstract

Denne oppgaven presenterer teori rundt passive filterkretser, samt forsøk med passive filterkretser for å teste teori i praksis. Oppgaven presenterer også hvordan operasjonsforsterkeren fungerer, og hvordan den kan benyttes til å forsterke signal i filterkretser.

Innholdsfortegnelse

1. Formål.....	3
2. Filter.....	3
3. Analoge filterkretser.....	5
3.1 Passive analoge filterkretser.....	5
3.1.1 RC-Krets.....	6
3.1.2 RLC-krets.....	7
3.1.3 Kaskadefilter.....	9
4. Forsøk med filterkretser.....	11
4.1 Metode.....	11
4.2 Resultater.....	12
4.2.1 Spenningsfall over R i RLC-krets.....	12
4.2.2 Spenningsfall over C i RLC-krets.....	12
4.2.3 Spenningsfall over L i RLC-krets.....	13
4.2.4 Spenningsfall over LC i RLC-krets.....	14
4.2.5 Spenningsfall over C i RC-krets.....	16
4.2.6 Spenningsfall over C i RC-krets koblet i kaskade.....	17
5. Forsterkning av signal i analoge filterkretser.....	18
5.1 Operasjonsforsterkeren.....	18
5.2 Dioden og Transistoren.....	21
5.2.1 Dioden.....	21
5.2.2 Transistoren.....	22
5.3 Eksempel på kobling av operasjonsforsterker i en filterkrets.....	24
6. Konklusjon.....	25
Litteraturliste.....	26

1. Formål

Formålet med dette prosjektet er å forstå hvordan elektroniske komponenter i filterkretser benyttes for å filtrere og forsterke signal i en krets. Fokus for oppgaven vil derfor være å få innsikt i hvordan motstander, kondensatorer og spoler benyttes til å filtrere signaler, og hvordan operasjonsforsterkeren og dens komponenter fungerer for å forsterke signaler i filterkretser.

2. Filter

Helt generelt kan vi si at et filter er en form anordning som forsterker, demper eller stopper frekvenser når et signal slippes igjennom. En mer formell definisjon er at filteret fungerer som en lineær prosess som omformer et frekvensspekter. Mer matematisk kan det uttrykkes:

$$Y(\omega) = H(\omega)X(\omega) \quad (\text{Lign. 2.1})$$

Her er $X(\omega)$ den Fouriertransformerte av det innkommende signalet, $Y(\omega)$ den Fouriertransformerte av det utgående signalet og $H(\omega)$ er overføringsfunksjonen (eng. transferfunksjon). En viktig observasjon er at overføringsfunksjonen $H(\omega)$ er konstant for en gitt verdi av ω , dette gjør prosessen lineær.

Overføringsfunksjonen $H(\omega)$ er mer presist en kompleks funksjon og kan skrives på formen

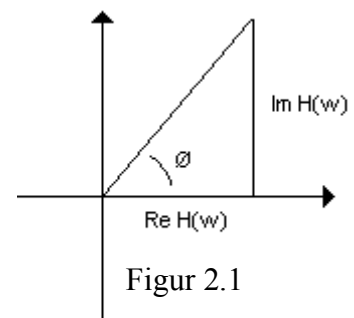
$$H(\omega) = |H(\omega)|e^{i\phi(\omega)}$$

Denne formen viser to viktige aspekter ved et filter, omforming av amplitude og omforming av fasen. For en gitt frekvens ω er amplituden på utgående signal lik amplituden på innkommende signalet ganger $|H(\omega)|$. Fasen på det utgående signalet er lik fasen på innkommende signal pluss faseforandringen $\phi(\omega)$ i filteret. Et filter kan altså forandre amplitude og fase til et signal. Det kan derimot ikke forandre frekvenser, eller legge til frekvenser som ikke er innhold i det innkommende signalet..

Av figur 2.1 ser vi at vi finner faseforandringen $\phi(\omega)$ ved

$$\phi(\omega) = \arctan\left(\frac{\text{Im}(H(\omega))}{\text{Re}(H(\omega))}\right)$$

Hvor $\text{Im}(H(\omega))$ er imaginærdelen til $H(\omega)$, og $\text{Re}(H(\omega))$ er realdelen til $H(\omega)$.

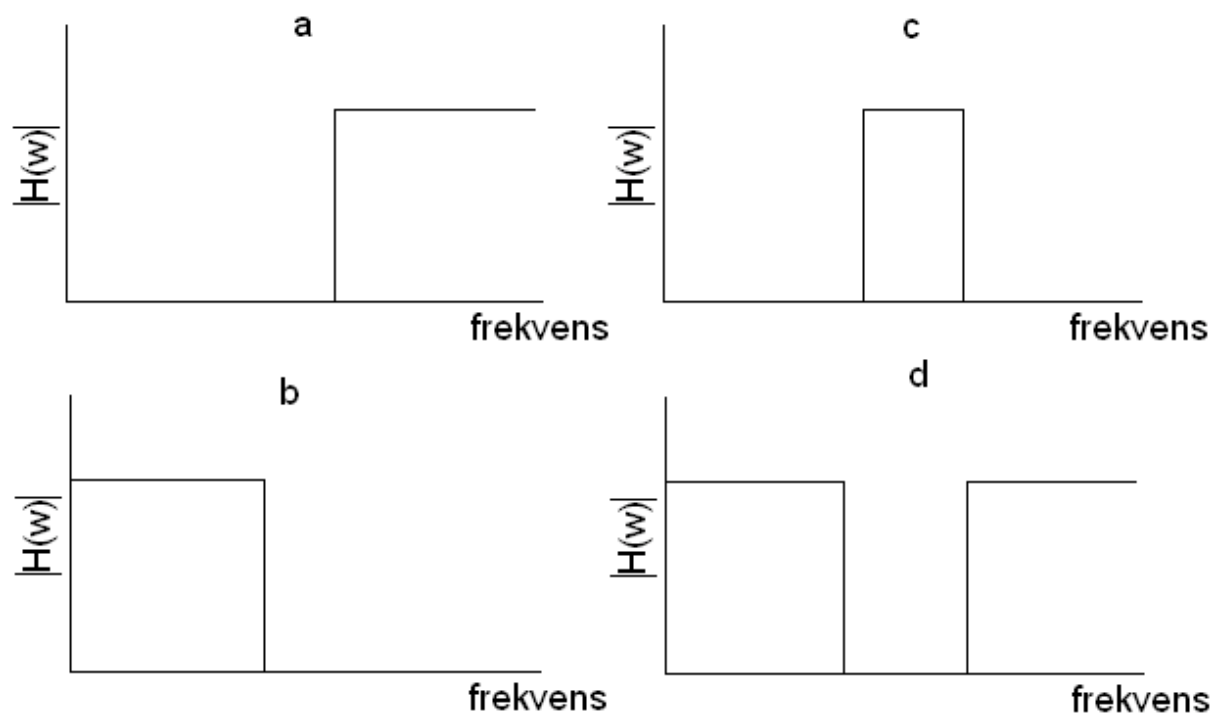


Vi vet fra Fouier teorien, at alle funksjoner kan skrives som en sum av enkle sinus og cosinusfunksjoner. Dette innebærer at selv et stygt signal kan beskrives som en superposisjon av harmoniske signaler.

Dette innebærer videre at selv om signalkilden inneholder et spekter av forskjellige frekvenser kan vi beregne $H(\omega)$ for de forskjellige frekvensene signalet er bygget opp av. Amplitude på innkommende signal $|X(\omega)|$, amplituden på utgående signal $|Y(\omega)|$ og faseforandringen $\phi(\omega)$ er derfor funksjoner av ω .

Begreper om filter

Et høypassfilter (figur 2.2a) er et filter som slipper igjennom høye frekvenser og demper lave, et lavpassfilter (figur 2.2b) slipper gjennom lave frekvenser og demper høye, mens et bandpassfilter (figur 2.2c) er et filter som slipper gjennom et område av frekvensspekteret og demper andre. I tillegg har vi bandstopfilter (figur 2.2d) som demper en hvis del av et frekvensspekter og slipper igjennom alle andre. I alle tilfellene kaller vi de frekvensene som slipper igjennom for filterets passband og de frekvensene som ikke slipper igjennom for filterets stoppband.



Figur 2.2 Filtertyper

Grensefrekvensen mellom stoppband og passband er det vi kaller cutoff frekvensen

I realiteten er det ikke snakk om skarpe enkle frekvenser som blir forsterket eller forminsket, men et område rundt den frekvensen som blir forsterket/forminsket mest. En klarer derfor ikke å lage et skarpt skille mellom stoppband og passband. Vi definerer cutoff frekvensen ω_c ut fra overføringsfunksjonen som,

$$H(\omega_c) = \frac{1}{\sqrt{2}} H(\omega_m)_{maks} \quad (\text{lign. 2.2})$$

Hvor ω_m er den/de frekvensene som gir $H(\omega)_{maks}$

Vi skal ikke gå nøye inn på hvorfor nettopp dette er definisjonen, men kan nevne at det henger sammen med at intervallet mellom ω_c og ω_m svarer til halvverdibredden til maks intensitet av et signal som funksjon av frekvensen.

For bandpassfilter (og bandstopfilter) får vi to cutoff frekvenser ω_{c1} og ω_{c2} . Intervallet mellom disse kaller vi $\Delta\omega_c$. Vi skal bruke dette til å definere Q-verdien (eng. Quality value)

$$Q = \frac{\omega_0}{\Delta\omega_c}$$

Det er ofte ønskelig med en lav halvverdbredde og en så høy Q- verdi som mulig. Med en stor Q-verdi kan man fjerne eller forsterke veldig skarpe deler av et frekvensspekter.

3. Analoge filterkretser

Omtrent overalt hvor vi har elektronikk finner vi analoge filterkretser. De brukes i alt fra radio, fjernsynsapparater, telefoner og musikkanlegg til avanserte computere i alle former. Selv der en konverterer signaler fra analoge til digitale, bruker en analoge filter.

Vi kan dele alle elektroniske analoge filterkretser inn i to hovedgrupper. Aktive og passive. Det som kjennetegner aktive filter er at de inneholder en eller flere forsterkere som kan forsterke signalet når det sendes gjennom filterkretsen. En passiv filterkrets består derimot kun av kondensatorer, spoler og motstander. Signalet blir ikke forsterket når det passerer gjennom kretsen. En passiv filterkrets kan, som vi skal se, brukes til å forsterke enkelte frekvenser, men vil alltid virke som et intensitetsdempende element i en større krets.

Passive filter trenger ikke strømtilførsel og er derfor brukt i lavenergissystemer. Aktive filtre er generelt mindre og mer kompakte enn passive filtre, og er derfor ofte mer hendige. De har også en stor fordel i at de både kan dempe eller forsterke intensiteten til et signal. Et problem med aktive filtre er at når en bruker forsterkere får en alltid inn støy i det opprinnelige signalet.

Videre skal vi ta for oss analoge passive filterkretser, bygget opp av kondensatorer, spoler og motstander, og se hvordan vi kan bruke kombinasjoner av disse elementene til å lage filter beskrevet ovenfor.

3.1 Passive analoge filterkretser

Først vil vi definere impedansen Z som forholdet mellom spenningen V og strømmen I .

$$Z = \frac{V}{I}$$

Impedansen til de forskjellige kretselementene definerer vi som:

$$Z_R = R \quad Z_L = i\omega L \quad Z_C = \frac{1}{i\omega C}$$

Hvor R er resistans i motstand, L er induktans i spole og C er kapasitans i kondensator. i er her den imaginære enheten $\sqrt{-1}$

Videre definerer vi overføringsfunksjonen.

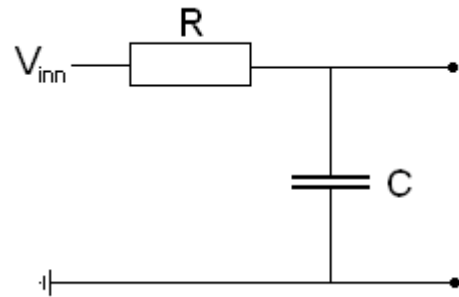
$$H(\omega) = \frac{Z_A}{Z_K}$$

Hvor Z_A representerer impedansen til kretselementet(/elementene) vi taper utgående signal over, og Z_K er impedansen til hele kretsen.

3.1.1 RC-Krets

La oss først se på en enkel seriekoblet RC-krets, hvor vi taper utgående signal over kondensatoren, se figur 3.1.

$$H(\omega) = \frac{1}{R + \frac{1}{i\omega C}}$$



Figur 3.1

Amplituden på det utgående signalet blir da,

$$V_{oC} = |H(\omega)|V_0 = \left| \frac{1}{R + \frac{1}{i\omega C}} \right| V_0 = \frac{\frac{1}{\omega C} V_0}{\sqrt{R^2 + \frac{1}{\omega^2 C^2}}} = \frac{V_0}{\sqrt{\omega^2 C^2 R^2 + 1}} \quad (\text{Lign. 3.1})$$

$$\lim_{\omega \rightarrow 0} V_{oC} = V_0$$

$$\lim_{\omega \rightarrow \infty} V_{oC} = 0$$

Som vi ser er amplituden på spenningsfallet avhengig av frekvensen. For høye verdier av ω går V_{oC} mot 0, mens for lave verdier av ω går V_{oC} mot V_0 . Tapper vi spenningen over kondensatoren har vi med andre ord en filterkrets som slipper lave frekvenser gjennom, og demper høye frekvenser. Et lavpassfilter.

Sender vi hvit støy (like stor amplitude for alle frekvenser) gjennom kretsen beskrevet ovenfor dempes de høye frekvensene og vi sitter igjen med lave frekvensene.

Ser vi derimot på spenningsfallet over resistansen, se figur 3.2, blir overføringsfunksjonen:

$$H(\omega) = \frac{R}{R + \frac{1}{i\omega C}}$$

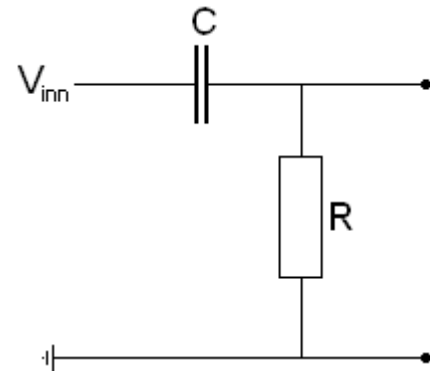
Vi finner amplituden av utgående signal på samme måte som ovenfor.

$$V_{oR} = |H(\omega)|V_0 = \left| \frac{R}{R + \frac{1}{i\omega C}} \right| V_0 = \frac{RV_0}{\sqrt{R^2 + \frac{1}{\omega^2 C^2}}} = \frac{V_0}{\sqrt{1 + \frac{1}{R^2 \omega^2 C^2}}} \quad (\text{Lign. 3.2})$$

$$\lim_{\omega \rightarrow 0} V_{0C} = 0$$

$$\lim_{\omega \rightarrow \infty} V_{0C} = V_0$$

Her har vi altså et filter som virker motsatt av det forrige. Et høypassfilter.



Figur 3.2

3.1.2 RLC-krets

Vi kobler nå på en spole i kretsen vår, slik at vi får en RLC krets. Når vi ser på amplituden av spenningsfallet over motstanden, se figur 3.3, får vi:

$$V_{0R} = |H(\omega)|V_0 = \left| \frac{R}{R + i\omega L + \frac{1}{i\omega C}} \right| V_0 = \frac{RV_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \quad (\text{Lign. 3.3})$$

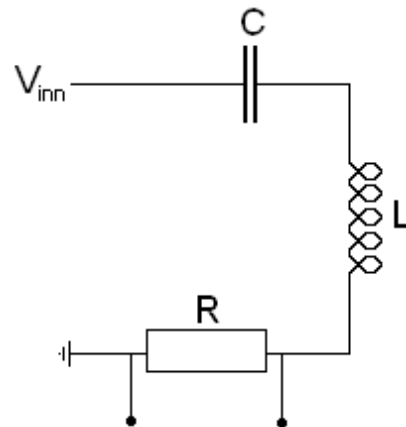
$$\lim_{\omega \rightarrow 0} V_{0R} = 0$$

$$\lim_{\omega \rightarrow \infty} V_{0R} = 0$$

$$\lim_{\omega \rightarrow \omega_0} V_{0R} = V_0$$

$V_{0R \max}$ får vi når $L\omega = \frac{1}{\omega C}$, eller $\omega = \omega_0 = \frac{1}{\sqrt{LC}}$. Denne kalles også resonansfrekvensen til dette systemet og gir oss $V_{0R \max} = V_0$. Denne resonansfrekvensen er uavhengig av inngangs og utgangssignal, og gjelder for alle RLC-kretser koblet i serie.

Både høyere og lavere verdier av ω (enn ω_0) gir lavere verdier av V_{0R} . Dette er et enkelt eksempel på et båndpassfilter, som slipper igjennom vinkelfrekvenser rundt $\omega = \omega_0$ og demper lavere og høyere verdier av ω .



Figur 3.3

Tapper vi signalet over kondensatoren i samme krets (figur 3.4), får vi

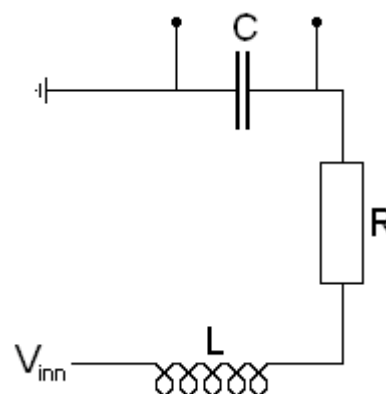
$$V_{0C} = |H(\omega)|V_0 = \left| \frac{\frac{1}{i\omega C}}{R + i\omega L + \frac{1}{i\omega C}} \right| V_0 = \left| \frac{\frac{1}{LC}}{\frac{i\omega R}{L} - \omega^2 + \frac{1}{LC}} \right| V_0 = \frac{\omega_0^2 V_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + \frac{\omega^2 R^2}{L^2}}} \quad (\text{Lign. 3.4})$$

$$\lim_{\omega \rightarrow 0} V_{0C} = V_0$$

$$\lim_{\omega \rightarrow \infty} V_{0C} = 0$$

$$\lim_{\omega \rightarrow \omega_0} V_{0C} = V_0 \frac{\omega_0 L}{R}$$

Her har vi et lavpassfilter med en resonansfrekvens. Det er også interessant å påpeke at vi med et passivt filter har klart å forsterke signalet ved resonansfrekvensen, uten å øke intensiteten til signalet.



Figur 3.4

La oss se på samme krets men nå tappe utgående signal over Spolen (figur 3.5):

$$V_{oL} = |H(\omega)|V_0 = \left| \frac{i\omega L}{R + i\omega L + \frac{1}{i\omega C}} \right| V_0 = \frac{\omega L}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} V_0$$

$$= \frac{1}{\sqrt{\frac{R^2}{\omega^2 L^2} + \left(1 - \frac{1}{\omega^2 LC}\right)^2}} V_0 = \frac{1}{\sqrt{\frac{R^2}{\omega^2 L^2} + \left(1 - \frac{\omega_0^2}{\omega^2}\right)^2}} V_0$$

(Lign. 3.5)

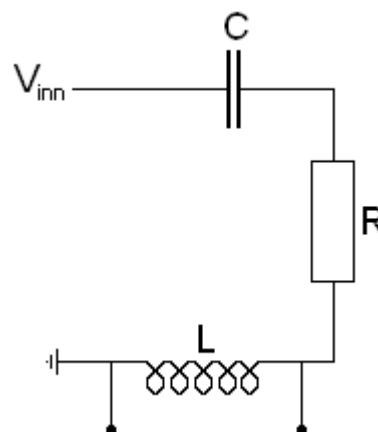
$$\lim_{\omega \rightarrow 0} V_{oL} = 0$$

$$\lim_{\omega \rightarrow \infty} V_{oL} = V_0$$

$$\lim_{\omega \rightarrow \omega_0} V_{oL} = V_0 \frac{\omega_0 L}{R}$$

Vi får et høypassfilter med resonansfrekvens.

La oss til sist måle spenningsfallet over kondensatoren og spolen sammen, se figur 3.6.



Figur 3.5

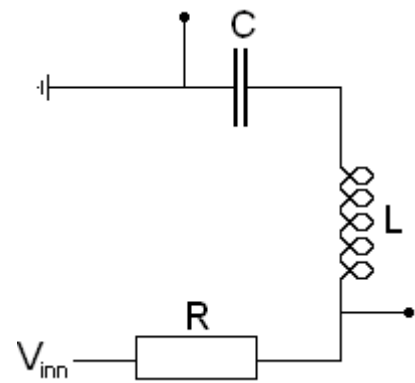
$$V_{oLC} = |H(\omega)|V_0 = \frac{\left| i\omega L + \frac{1}{i\omega C} \right|}{\left| R + i\omega L + \frac{1}{i\omega C} \right|} V_0 = \frac{\left(\omega L - \frac{1}{\omega C}\right) V_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} = \frac{V_0}{\sqrt{\frac{R^2}{\left(\omega L - \frac{1}{\omega C}\right)^2} + 1}} \quad (\text{Lign. 3.6})$$

$$\lim_{\omega \rightarrow 0} V_{0LC} = V_0$$

$$\lim_{\omega \rightarrow \infty} V_{0LC} = V_0$$

$$\lim_{\omega \rightarrow \omega_0} V_{0LC} = 0$$

Vi har altså et Båndstoppfilter. Dette virker motsatt av RLC-filteret der vi taper spenningen over motstanden. Kanskje ikke like overraskede da summen av spenningsfall over komponentene i en krets er lik påtrykt spenning.



Figur 3.6

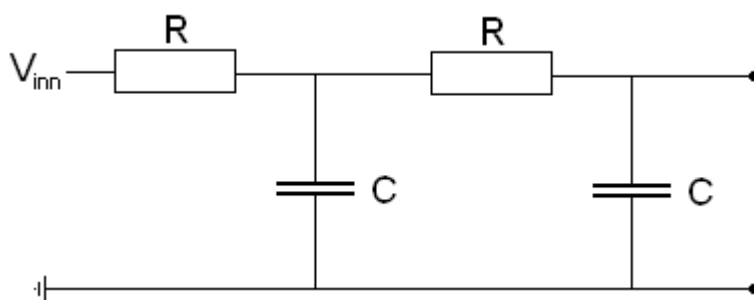
3.1.3 Kaskadefilter

Det er ingenting i veien for å filtrere et signal flere ganger. Vi kobler filtre i kaskade. Spesielt enkelt blir det å beskrive matematisk hvis vi sender signalet gjennom to like filtre etter hverandre.

La oss ta det første lavpassfilteret vi beskrev, hvor vi taper signalet over en kondensator i en enkel seriekoblet RC-krets. La oss sende et signal gjennom to slike filter, se figur 3.7.

$$V_{0R2} = |H(\omega)|V_{0R1} = |H(\omega)||H(\omega)|V_0 = |H(\omega)|^2 V_0$$

$$V_{0R2} = \left| \frac{R}{R + \frac{1}{i\omega C}} \right|^2 V_0 = \frac{R^2 V_0}{R^2 + \frac{1}{\omega^2 C^2}} = \frac{V_0}{1 + \frac{1}{R^2 \omega^2 C^2}} \quad (\text{Lign. 3.7})$$



Figur 3.7

Sammenligner vi dette spenningsfallet med spenningsfallet vi fikk over kondensatoren i en enkel RC-krets, ser vi at når vi kobler i kaskade får vi brattere kurver og en skarpere frekvens som slipper igjennom. Vi ser det kanskje enda tydeligere hvis vi kobler to like RLC-kretser i kaskade, og taper spenningen over resistansen.

$$V_{0R2} = \left| \frac{R}{R + i\omega L + \frac{1}{i\omega C}} \right|^2 V_0 = \frac{R^2 V_0}{R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2} \quad (\text{Lign. 3.8})$$

$$\lim_{\omega \rightarrow 0} V_{OR} = 0$$

$$\lim_{\omega \rightarrow \infty} V_{OR} = 0$$

$$\lim_{\omega \rightarrow \omega_0} V_{OR} = V_0$$

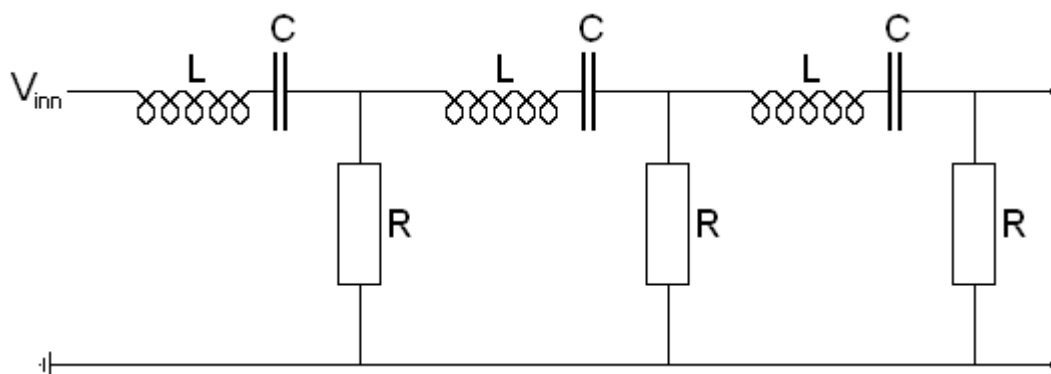
Her får vi samme grenseverdier som i en enkel RLC-krets, men en skarpere halvverdibredde og en høyere Q-verdi. Kobler vi mange slike kretser i kaskade (figur 3.8) vil vi få et utgangssignal som tilslutt domineres av et veldig lite frekvensspekter.

$$V_{OR2} = \left| \frac{R}{R + i\omega L + \frac{1}{i\omega C}} \right|^2 V_0 = \frac{R^2 V_0}{R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2}$$

$$V_{OR3} = \left| \frac{R}{R + i\omega L + \frac{1}{i\omega C}} \right|^3 V_0 = \frac{R^3 V_0}{\left(R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right)^{3/2}} \quad (\text{Lign. 3.9})$$

$$V_{OR4} = \left| \frac{R}{R + i\omega L + \frac{1}{i\omega C}} \right|^4 V_0 = \frac{R^4 V_0}{\left(R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right)^{4/2}}$$

osv...



Figur 3.8

Vi har nå kun presentert de aller enkleste kombinasjonene av analoge filterkretser. Alle de forskjellige kombinasjonene av komponenter, og de forskjellige måtene å koble disse på, i kaskade med like eller helt andre kretser, gir oss et vell av mulige filterkretser med sine respektive egenskaper. Dette gjør oss i stand til, i hvert fall i teorien, til å formatere et inngangssignal til det vi måtte ønske.

4. Forsøk med filterkretser

For å teste teori om filterkretser i praksis, gjennomførte vi forsøk der vi sendte hvitt støy inn på filterkretser beskrevet i kapittel 3.

4.1 Metode

For hvert forsøk sender vi hvitt støy inn på filterkretsen og datalogger spenningsfallet over utgangen av filterkretsen. Vi ender opp med et plot som viser spenningsfallet som funksjon av tiden. Resultatet herfra gir vi navnet datalogg1. Tilsvarende datalogger vi spenningsfallet over hele kretsen. Resultatet herfra gir vi navnet datalogg2.

Datalogg1 og datalogg2 er egentlig gjennomsnittet av 10 loggninger, hvert med et intervall på 1 sekund. Dette gjøres for at målingene skal bli så nøyaktige som mulig fra en støykilde som ikke er ideell.

Som nevnt i kap. 2 kan alle funksjoner beskrives som en sum av enkle sinus- eller cosinusfunksjoner. En enkel sinus- eller cosinusfunksjon svinger med en gitt frekvens og en gitt amplitude. Vi benytter derfor fourieranalyse i Mathlab til å presentere formen på vårt dataloggede spenningsfall som et spekter av frekvenser med ulik amplitude.

Spenningsfallet over hele filterkretsen, Fouriertransformert av datalogg2, vil representere amplituden på det opprinnelige frekvensspekteret i støyen. Fouriertransformert av Datalogg1 vil representere amplituden for frekvensspekteret målt over en utgangen av filterkretsen.

Hvis vi dividerer amplituden på en gitt frekvens i datalogg1 med amplituden på den samme frekvensen i datalogg2, vil vi oppnå et amplitudeforhold. Størrelsen på amplitudeforholdet forteller dermed hvor mye amplituden har endret seg i forhold til den opprinnelige amplituden, og amplitudeforhold mindre enn 1 medfører derfor at amplituden for denne frekvensen har blitt dempet i filterkretsen. Som vi ser av ligning 2.1 er dette amplitudeforholdet det samme som amplituden til overføringsfunksjonen:

$$\frac{|Y(\omega)|}{|X(\omega)|} = |H(\omega)|$$

Vi lar Mathlab tegne en graf over amplitudeforholdet for et frekvensspekter. Dermed kan vi se hvor mye ulike frekvenser i filterkretsen blir dempet, og evt. forsterket.

BM: En ideell hvit støykilde har samme amplitude for alle frekvenser. Dette tilsier at hvis en vet amplituden på støyen, så er datalogg2 unødvendig. Da kan en presentere en graf med frekvensspekteret og tilhørende amplitude fra datalogg1, og om ønskelig en referanselinje som viser den opprinnelige amplituden i støyen. Støykilden vår er derimot ikke ideell, og en divisjon med frekvensspekteret i støyen og presentasjon av amplitudeforholdet gir derfor et mer nøyaktig resultat.

For å oppnå en variasjon i amplitude over frekvensspekteret som tydelig gir inntrykk av effekten av filtreringen, måtte vi prøve oss frem med ulike verdier på resistans, kapasitans og induktans. Verdien på disse komponentene vil derfor variere noe i oppsettet av de forskjellige kretsene som blir presentert i resultatdelen.

Sammenhengen mellom vinkelfrekvens og frekvens er $f = \frac{\omega}{2\pi}$

4.2 Resultater

4.2.1 Spenningsfall over R i RLC-krets

Komponentverdier

R=10 Ω

L= 39,7 mH

C= 212 nF

Forventet resonansfrekvens,

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} = 1738,93 \text{ Hz}$$

Denne kretsen er beskrevet i ligning 3.3, og figur 3.3. Amplitudeforholdet vi fikk ved det dataloggede spenningsfallet over motstanden og hele kretsen er gjengitt nedenfor:

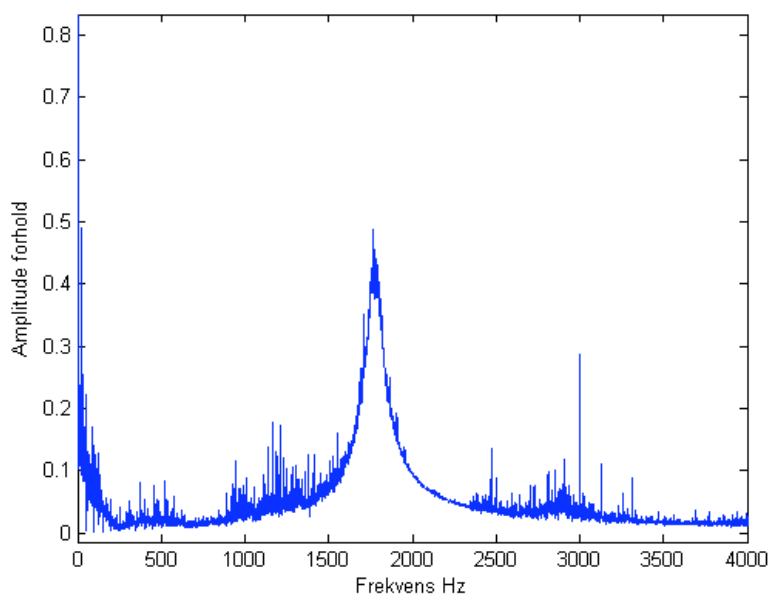


Diagram1

Som forventet ser vi av diagram1 at filteret vårt slipper gjennom frekvenser rundt resonansfrekvensen, og demper de andre. Vi hadde likevel forventet en høyere verdi for amplitudeforholdet enn 0,5 ved resonansfrekvensen, da amplitudeforholdet ifølge 3.3 burde gå mot 1. Vi observerer også at filterkretsen ikke demper de laveste frekvensene < 200 Hz.

4.2.2 Spenningsfall over C i RLC-krets

Komponentverdier

R=10 Ω

L= 39,7 mH

C= 212 nF

Forventet resonansfrekvens,

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} = 1738,93Hz$$

Denne kretsen er beskrevet i ligning 3.4, og figur 3.4. Amplitudeforholdet vi fikk ved det dataloggede spenningsfallet over motstanden og hele kretsen er gjengitt nedenfor.

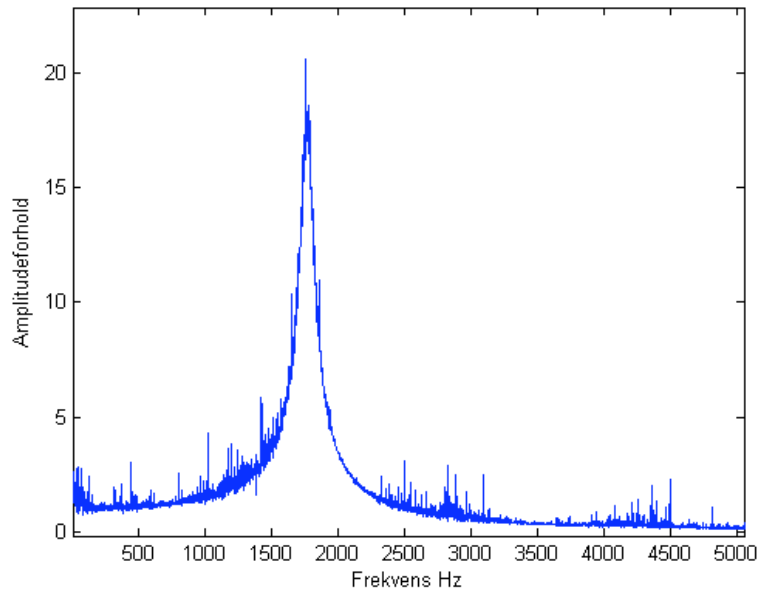


Diagram2

Av diagram2 ser vi at vi har klart å forsterke signalet rundt resonansfrekvensen 20 ganger. Vi ser antydninger til at filteret vårt gjennom frekvenser lavere enn resonansfrekvensen, og demper de som er høyere, som vi kan forvente ifølge ligning 3.4.

4.2.3 Spenningsfall over L i RLC-krets

Komponentverdier

R=10 Ω

L= 12,6 mH

C= 212 nF

Forventet resonansfrekvens,

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} = 3086,69Hz$$

Denne kretsen er beskrevet i ligning 3.5, og figur 3.5. Amplitudeforholdet vi fikk ved det dataloggede spenningsfallet over motstanden og hele kretsen er gjengitt nedenfor:

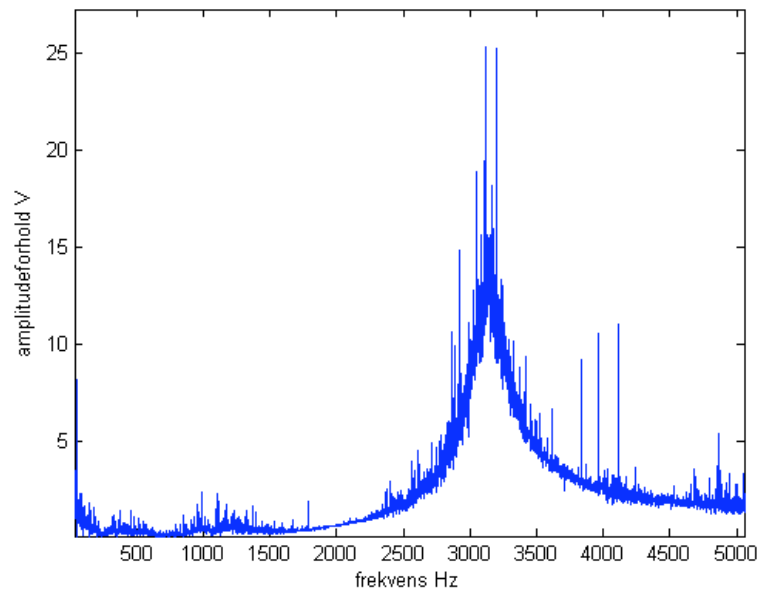


Diagram3

For å se forventet tendens bedre måtte vi for dette filteroppsettet måtte bruke en annen spole med et lavere viklingstall og induktans. Dette har følgelig gitt oss en høyere resonansfrekvens. Av diagram3 ser vi at denne resonansfrekvensen ligger rundt den forventede verdien. Vi har også her klart å forsterke resonansfrekvensen.

4.2.4 Spenningsfall over LC i RLC-krets

Komponentverdier

$R = 42 \Omega$

$L = 39,7 \text{ mH}$

$C = 212 \text{ nF}$

Forventet resonansfrekvens,

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} = 1738,93 \text{ Hz}$$

Denne kretsen er beskrevet i ligning 3.6, og figur 3.6. Amplitudeforholdet vi fikk ved det dataloggede spenningsfallet over motstanden og hele kretsen er gjengitt nedenfor:

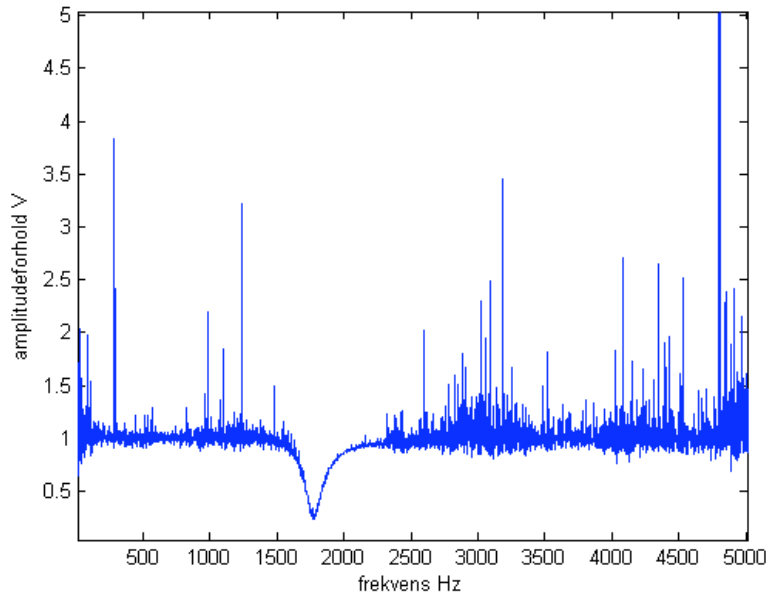


Diagram4

Vi ser av diagram4 at filteret slipper alle frekvenser gjennom, unntatt i et intervall mellom ca. 1600 og 2200 Hz; vi har et båndstopfilter.

For å se tendensen måtte vi øke motstand i kretsen til 42Ω , dette for å få en lavere Q-verdi og et bredere bandstop.

I dette tilfellet er $H(\omega)_{maks} = 1$, (se ligning 2.2 og 3.6)

Vi ser på $\Delta\omega_c$

$$\frac{1}{\sqrt{2}} H(\omega)_{maks} = H(\omega_c)$$

$$\frac{1}{\sqrt{2}} = \frac{\omega_c^2 L - \frac{1}{C}}{\sqrt{R^2 \omega_c^2 + \left(\omega_c^2 L - \frac{1}{C}\right)^2}}$$

$$\frac{1}{2} = \frac{\left(\omega_c^2 L - \frac{1}{C}\right)^2}{R^2 \omega_c^2 + \left(\omega_c^2 L - \frac{1}{C}\right)^2}$$

$$\omega_c R = \pm \left(\omega_c^2 L - \frac{1}{C}\right)$$

$$\omega_c^2 L \pm \omega_c R - \frac{1}{C} = 0$$

(Det er bare de positive verdiene for ω_{c1} og ω_{c2} vi er ute etter)

$$\omega_{c1} = \frac{R \pm \sqrt{R^2 + 4L/C}}{2L} \rightarrow \omega_{c1} = \frac{R + \sqrt{R^2 + 4L/C}}{2L}$$

$$\omega_{c2} = \frac{-R \pm \sqrt{R^2 + 4L/C}}{2L} \rightarrow \omega_{c2} = \frac{-R + \sqrt{R^2 + 4L/C}}{2L}$$

$$\Rightarrow \Delta\omega_c = |\omega_{c1} - \omega_{c2}| = \frac{R}{L}$$

Vi ser her at $\Delta\omega_c \sim R$, derfor valgte vi en høyere verdi på resistansen for å bedre se tendensen i filtreringen som vises diagram4.

4.2.5 Spenningsfall over C i RC-krets

Komponentverdier

R= 42 Ω

C= 212 nF

Denne kretsen er beskrevet i ligning 3.1, og figur 3.1. Amplitudeforholdet vi fikk ved det dataloggede spenningsfallet over motstanden og hele kretsen er gjengitt nedenfor:

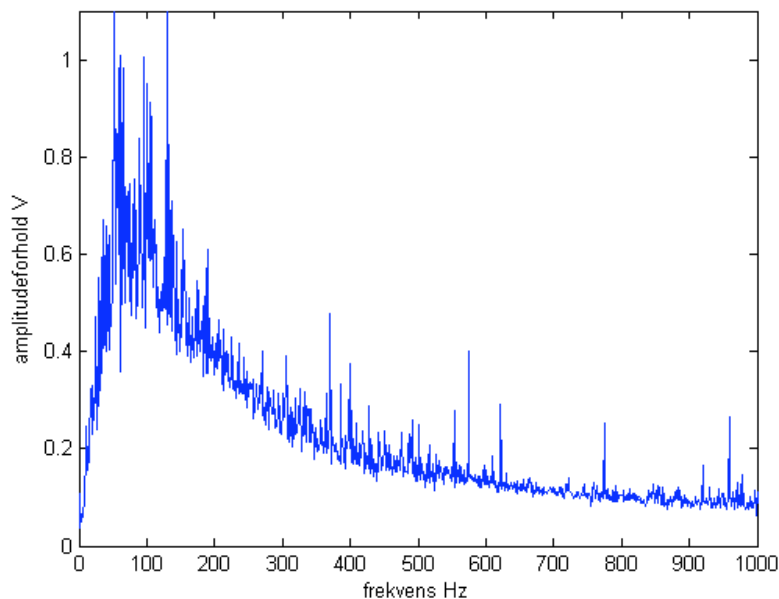


Diagram5

Når frekvensen går mot null skal amplitudeforholdet gå mot 1 ifølge ligning 3.1. Av diagram5 ser vi at dette ikke er tilfellet for frekvenser under 50 Hz. Dette klarer vi ikke gi noen god forklaring på. Sett bort fra dette er resultatene i tråd med forventede resultater.

4.2.6 Spenningsfall over C i RC-krets koblet i kaskade

Komponentverdier

$R = 42 \Omega$

$C = 212 \text{ nF}$

Denne kretsen er beskrevet i ligning 3.7, og figur 3.7. Amplitudeforholdet vi fikk ved det dataloggede spenningsfallet over motstanden og hele kretsen er gjengitt nedenfor:

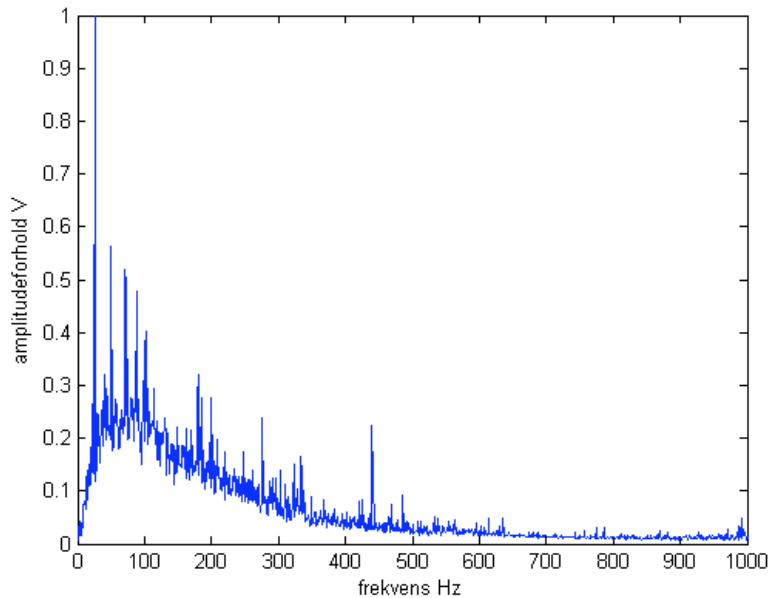


Diagram5

Sammenligner vi diagram5 med diagram4 ser vi at samme signalet har blitt filtrert to ganger. Vi har dermed fått påvist forventet oppførsel av filter koblet i kaskade. Forøvrig ser vi også her at amplitudeforholdet går mot null for frekvenser mindre enn 50 Hz. Ifølge ligning 3.7 skal amplitudeforholdet gå mot 1 når frekvensen går mot null.

5. Forsterkning av signal i analoge filterkretser

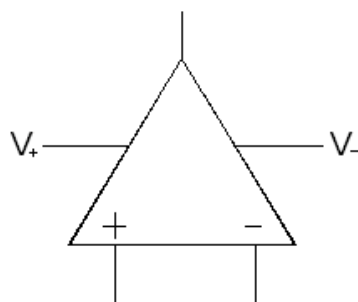
5.1 Operasjonsforsterkeren

Filterkretser kan være oppbygd som aktive eller passive filterkretser. Forskjellen på et aktivt og en passiv filterkrets er at den aktive filterkrets også inneholder operasjonsforsterker, som forsterker amplituden til signaler i filterkretsen. Martinsen skriver følgende om operasjonsforsterkeren, ofte forkortet opamp:

”Nå brukes som en generell byggeblokk for konstruksjon av forsterkere, oscillatorer, signalprosessering- signalgenererende kretser. De erstatter også kretser basert på diskrete komponenter. En opamp inneholder et stort antall transistorfunksjoner til å oppnå følgende data: Meget høy forsterkning, direktekobling, høy inngangsimpedans og differensialinnganger”

(Martinsen, URL)

Operasjonsforsterkeren er altså en viktig komponent med mange anvendelsesmuligheter. Siden vårt fokus er filterkretser, vil vi kun se på operasjonsforsterkeren som signalforsterker.



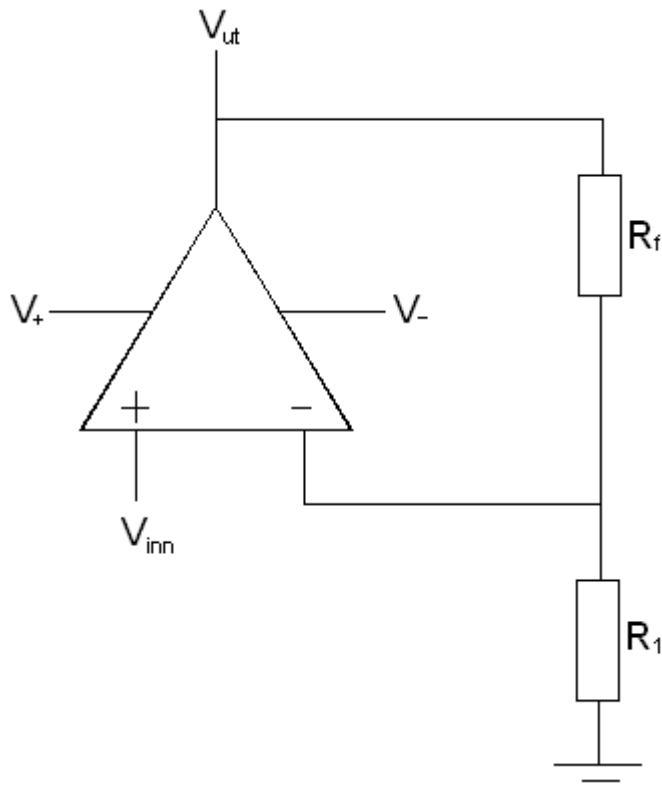
Figur 5.1

Figur 5.1 viser symbolet til operasjonsforsterkeren. Den har to innganger og en utgang, kalt terminaler. På toppen av trekanten er utgangen. På undersiden av trekanten er de to inngangene. Inngangen markert ”-” er inverterende inngang. Inngangen markert ”+” er ikke-inverterende inngang. For resten av kapittel 4 er inngang ”-” referert til som minus, og inngang ”+” referert til som pluss. Signalet som skal forsterkes kobles enten på minus eller pluss, avhengig av om en ønsker et invertert forsterket utgangssignal (et forsterket signal i motfase av det som blir sendt inn), eller et ikke-invertert forsterket utgangssignal. Operasjonsforsterkeren benytter et påtrykt

potensial koblet til V-pluss og V-minus for å forsterke signalet.

Operasjonsforsterkeren opererer etter følgende regel: Hvis potensialet på minus er høyere enn potensialet på pluss, vil den redusere amplituden på utgangssignalet til pluss og minus har samme potensial. Hvis minus har lavere potensial enn pluss vil den øke amplituden på utgangssignalet til potensialet er likt med potensialet på pluss. Denne prosessen går svært raskt.

Eksempel: Vi ønsker et ikke-invertert forsterket signal. Da kobler vi inngangssignalet til pluss. Se figur 5.2 nedenfor:



Figur 5.2

Eksempelet på oppkobling er laget etter koblingsskjema fra Hyperphysics opamp, URL

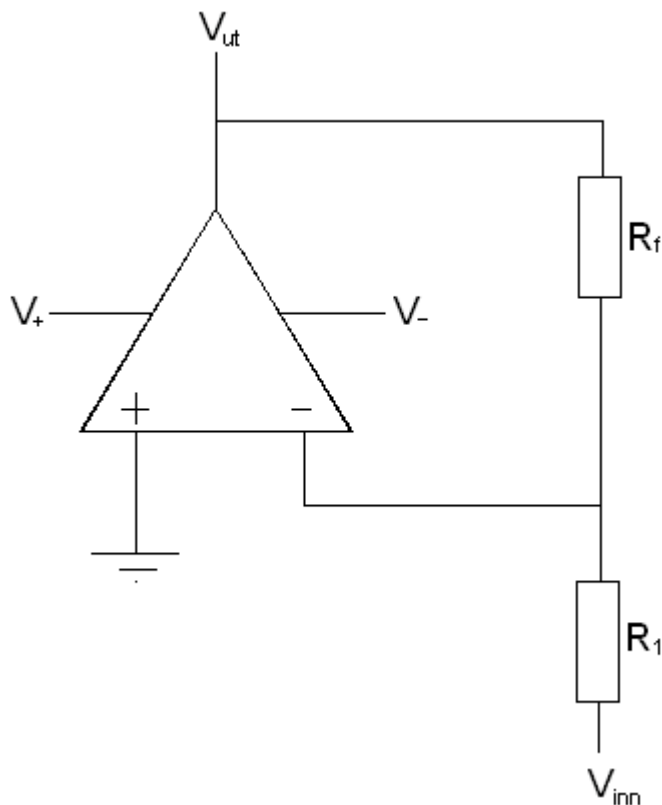
På minus har vi jord som potensiale i starttilstanden. Hvis signalet V-inn har høyere potensial enn jord, vil opampen øke spenningen på utgangssignalet. Vi ser at dette vil øke spenningen på minus, slik at pluss og minus etterhvert får samme potensial.

Hvis V-inn avtar, vil minus bli større enn pluss, og opampen vil redusere utgangssignalet. Dette vil redusere spenningen på minus, slik at pluss og minus får samme potensial.

Hvis spenningen over V-inn er negativ, vil opampen redusere spenningen på V-ut. Da blir spenningen på minus også negativ, og pluss og minus får samme potensial. Vi ser at V-ut hele tiden er i fase med V-inn.

Ifølge Hyperphysics Opamp, URL, er forsterkningsforholdet V_{ut}/V_{inn} gitt ved $1 + R_f/R_1$

Vi ønsker nå et invertert forsterket signal. Da kobler vi inngangssignalet til minus. Se figur 5.3 nedenfor:



Figur 5.3

Eksempel på oppkobling er laget etter koblingsskjema fra Hyperphysics opamp, URL

På pluss har vi jord som potensiale i starttilstanden. Hvis spenningen på V-inn er større enn pluss, vil opampen redusere spenningen på utgangssignalet. Den reduserte spenningen på utgangssignalet vil redusere spenningen på minus, slik at pluss og minus får samme potensiale. Dette medfører at V-ut signalet kommer i motfase til V-inn.

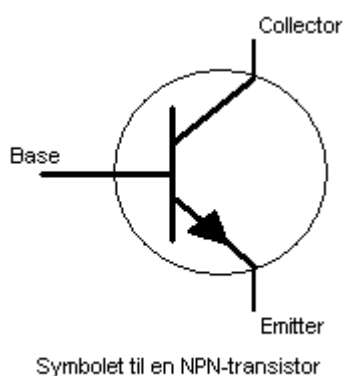
Hvis V-inn avtar, vil minus bli mindre enn pluss, og opampen vil øke spenningen på utgangssignalet. Utgangssignalet er derfor også nå i motfase til inngangssignalet. Den økte spenningen på utgangssignalet vil øke spenningen på minus, slik at pluss og minus får samme potensiale.

Hvis spenningen over V-inn er negativ, vil opampen øke spenningen på V-ut til minus og pluss igjen har samme potensial. Vi legger merke til at endring i V-ut hele tiden er i motfase til V-inn. Derfor har minus fått navnet inverterende inngang.

Ifølge Hyperphysics Opamp, URL, er forsterkningsforholdet V_{ut}/V_{inn} gitt ved R_f/R_1

Operasjonsforsterkeren er bygd opp av motstander og transistorer. Det er utenfor vårt fokus å gi en detaljert innføring i hvordan operasjonsforsterkeren er programmert til å utføre det som er beskrevet, men ved å forstå hvordan transistoren fungerer, vil vi kunne oppnå en intuitiv innsikt i hvordan operasjonsforsterkeren kan benytte sin driftsspenning til å forsterke et inngående signal.

5.2 Dioden og Transistoren



Symbolet til en NPN-transistor

Figur 5.4

Transistoren er et svært viktig kretselement. Den har blant annet vært en forutsetning for utviklingen av den moderne datamaskinen, hvor transistorer er den viktigste komponenten i datamaskinens prosessor. Isolert sett kan en transistor benyttes på to måter: Den kan benyttes som en bryter som slår av og på et signal, eller den kan brukes til å forsterke et signal i en krets. Ved å kombinere transistorene på ulike måter på en krets er transistoren svært anvendelig. Transistoren gjør det mulig å styre hvilken respons en krets skal gi på ulike inngangssignaler. Operasjonsforsterkeren er sammensatt av transistorer og motstander for å utføre sin funksjon.

Forenklet kan transistoren betraktes som to dioder knyttet anode mot anode. For å forstå hvordan transistoren fungerer vil det derfor være nyttig å først se på hvordan dioden fungerer.

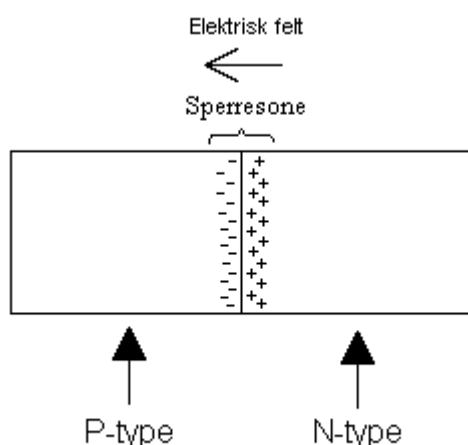
5.2.1 Dioden



Dioden

Figur 5.5

En diode (figur 5.5) er delt inn i to regioner kalt anode og katode. Anoden kan bestå av dopet silisium av P-type. Kathoden kan bestå av dopet silisium av N-type. Dopet P-type silisium har overskudd av hull. Dopet N-type silisium har overskudd av frie elektroner. Både N-type silisium og P-type silisium er nøytralt ladet. Når N- og P-type silisium får kontakt, vil følgende skje i grenseovergangen mellom de to silisiumtypene:



Figur 5.6

Elektroner i N-regionen vil på grunn av sin termiske energi diffundere over i P-type og falle i ledige hull i P-type valensbånd. I grenseområdet til N-type vil dermed P-type få overskudd av elektroner og bli negativt ladet. Tilsvarende vil N-type i grenseområdet få overskudd av hull. Hullene i N-type valensbånd oppstår fordi det kontinuerlig pågår en generering og rekombinasjon av hull og elektroner mellom ledningsbåndet og valensbåndet. Teori sier at produktet av antall elektroner i ledningsbåndet og antall hull i valensbåndet skal være konstant for en gitt temperatur. For å holde dette produktet konstant vil en diffusjon av elektroner bort fra N-type ledningsbånd medføre at antall hull øker i N-type valensbånd.

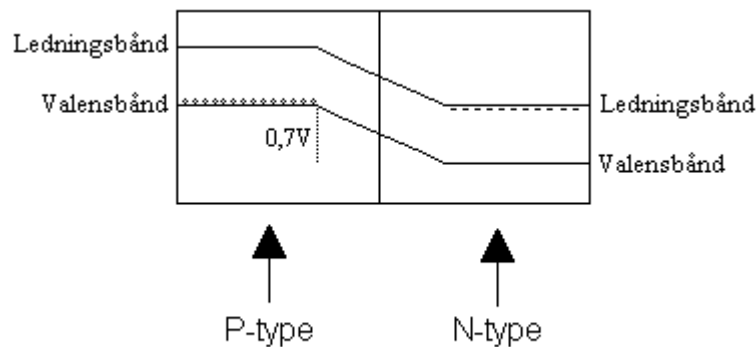
Således vil en elektronstrøm fra N-type til P-type i ledningsbåndet også kunne betraktes som en strøm av hull fra P-type til N-type i valensbåndet.

Som vi ser av figuren 5.6 vil dette resultere i et elektrisk felt som motvirker vandringer av elektroner mot ledige hull i P-type silisium. De elektronene som fremdeles har nok kinetisk energi til å passere det elektriske feltet vil okkupere nye ledige hull i P-type, og størrelsen på

det elektriske feltet vil vokse. Det elektriske feltet vil øke til elektronene ikke kan passere det elektriske feltet. Dette feltet blir kalt sperresone.

Fordi absoluttverdien av potensialdifferansen $V = \int E dx$, vil størrelsen på det elektriske feltet i figur 5.6 representere en potensialforskjell mellom P-type og N-type. For NP-dopet silisium er denne potensialforskjellen ca. 0,7 volt.

Figur 5.7 viser en kvalitativ skisse av denne potensialforskjellen. Elektronene i ledningsbåndet i N-type møter en potensialforskjell på 0,7 V når de forsøker å entre ledningsbåndet i P-type.



Figur 5.7

basert på figur fra Hyperphysics diode, URL

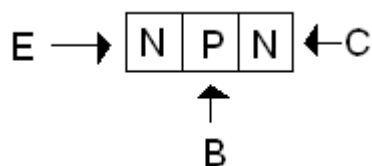
Hvis vi kobler et batteri med høyest potensial på N-type, vil dette generere et elektrisk felt som peker i samme retning som det elektriske feltet i figur 5.6. Dette vil dermed øke den potensialforskjellen som er illustrert i figur 5.7 mellom P-type og N-type. Det går derfor ikke noe strøm gjennom en diode hvis høyest potensial kobles til N-siden (kathoden).

Hvis vi kobler et batteri med høyest potensial på P-type, vil dette generere et elektrisk felt som peker i motsatt retning av det elektriske feltet i figur 5.6. Dette vil dermed motvirke potensialforskjellen som er illustrert i figur 5.7, og strøm kan passere gjennom dioden.

Konklusjon: Strøm går gjennom en diode hvis høyest potensial kobles til P-siden (anoden).
En diode er dermed et elektrisk kretselement som bare lar strømmen gå en vei.

5.2.2 Transistoren

Det finnes flere typer transistorer. Vi har valgt å se på en type transistor i detalj. Dette er en npn-transistor av BJT-type (bipolar junction transistor).



Figur 5.8

En NPN-transistor, se figur 5.8, består av tre regioner: en emitter (E), en base (B) og en collector (C). Ved å regulere potensialet på basen B, kan en regulere strømmengdens om passerer fra collector til emitter.

I denne transistoren er emitter og collector N-type-regioner av dopet silisium, mens basen er en P-type-region av dopet silisium. Vi ser at basen grenser både mot emitter og collector, og som en forenkling kan en derfor betrakte denne transistoren som to dioder koblet anode mot anode. Som vi har sett fra kap. 5.2.1, medfører dette at det er etablert en potensialforskjell mellom base og emitter, og en potensialforskjell mellom base og collector.

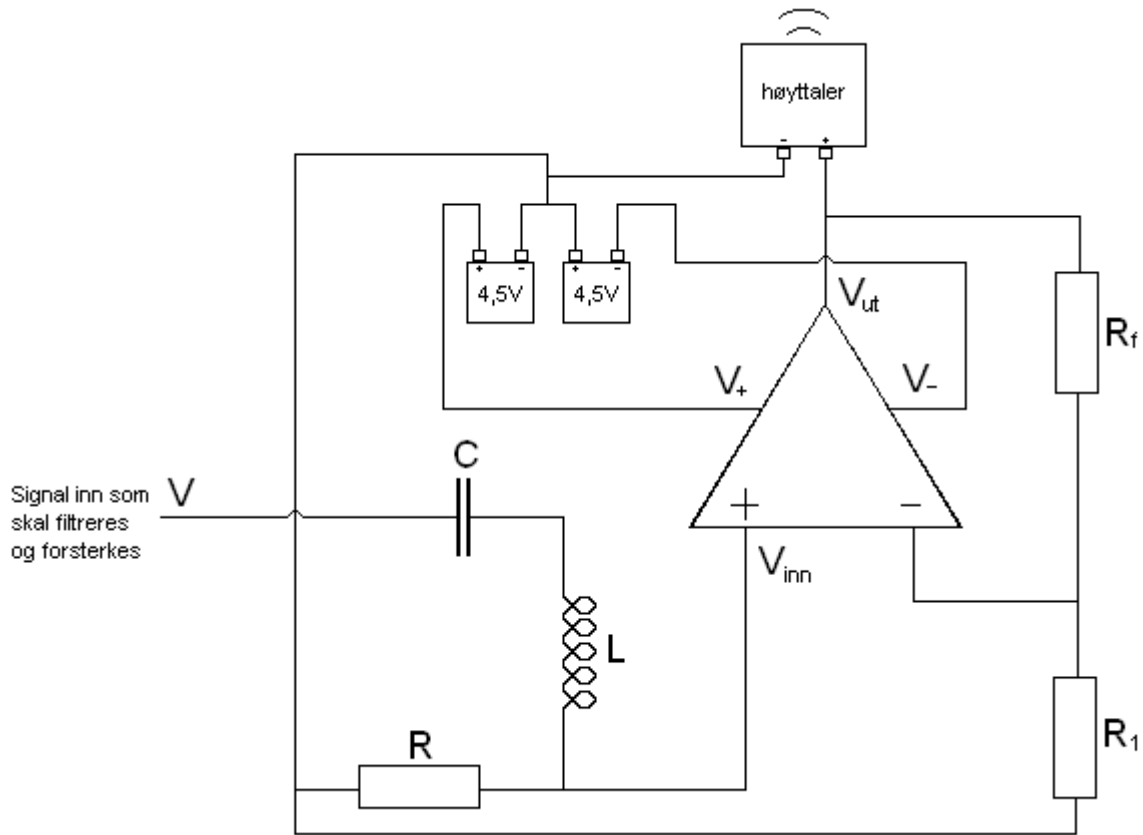
Vi kobler en potensialforskjell på f.eks. 9V mellom E og C fra en ekstern spenningskilde; positiv pol på C. Videre antar vi at basen har samme potensial som emitter. Hvis vi velger å se på emitter/base som en diode, og base/collector som en annen diode, vet vi fra kap. 5.2.1 at sperresonen mellom base og collector vil øke fra 0,7 til 9,7 V. Potensialet i basen vil dermed være uendret, og slik vil sperresonen mellom emitter og base være uendret. En uendret sperresone mellom emitter og base medfører at det ikke går strøm mellom emitter og base, og dermed ingen strøm mellom emitter og collector.

Hvis vi øker spenningen på basen til f.eks. 0,7V, vet vi fra kap. 5.2.1 at dette vil redusere potensialbarrieren mellom emitter og base, og elektroner kan passere fra emitter til base. Spenningsøkningen i basen vil også redusere potensialforskjellen/sperresonen mellom base og collector til 9V. Basen er laget smal for å unngå at elektroner fra emitter okkuperer hull i basen, men istedet beveger seg inn i det elektriske feltet mellom base og collector. Vi ser av figur 5.6 at elektroner som beveger seg inn i dette feltet ikke blir frastøtt, men blir akselerert av det elektriske feltet mot collector, og vandre videre mot batteriets plusspol. Noen elektroner som entrer basen fra emitter vil okkupere hull i basen før de entrer dette elektriske feltet, men disse elektronene blir fraktet bort av basens påtrykte potensiale på 0,7V, slik at basens potensial forblir uendret.

En liten spenningsendring på basen er nok til å gi en stor økning i strøm mellom emitter og collector. Strømmen mellom emitter og collector øker med økende spenning på basen B, forutsatt at denne spenningen ikke overstiger potensialet i collector.

Stor strøm gjennom transistoren fra collector til emitter vil medføre et lavt spenningsfall over transistoren. I henhold til Kirchoffs spenningsregel vil dette medføre at andre komponenter i kretsen vil få en større andel av spenningsfallet fra spenningskilden. Liten strøm gjennom transistoren vil medføre høyt spenningsfall over transistoren, og dermed mindre spenningsfall over andre komponenter i kretsen. Kobler vi inngangssignalet på basen B vil vi dermed kunne forsterke dette inngangssignalet ved en påtrykt spenning over collector og emitter.

5.3 Eksempel på kobling av operasjonsforsterker i en filterkrets



Figur 5.9

Figur 5.9 er et koblingsdiagram av en filterkrets som først filtrerer et signal, og deretter forsterker det og sender det ut gjennom en høyttaler. Vi definerer nullpunkt for potensialet til høyttalerens minusutgang. Følgende skjer med signalet som skal filtreres og forsterkes:

1. En ser av koblingsdiagrammet at signalet går inn på en krets hvor kondensatoren og spolens impedans vil gi en filtrering av signalet. Kondensatoren vil, i tillegg til å bidra med filtrering av signalet, justere spenningen på inngangssignalet slik at det varierer rundt nullpotensialet.
2. Spenningen på det filtrerte signalet sendes inn på operasjonsforsterkeren. Vi har ikke behov for å invertere signalet, så vi velger pluss-inngangen på operasjonsforsterkeren.
3. Driftsspenningen på operasjonsforsterkeren ser vi er $\pm 4,5V$ i forhold til nullpotensialet. Driftsspenningen benyttes til å forsterke signalet. Siden inngangssignalet vil svinge rundt nullpotensialet, vil maksimal forsterkning på utgangssignalet være differansen mellom spenningen på inngangssignalet og batteriets maksimum/minimumspotensiale, dvs en maksimal forsterkning på ca. $4,5V$. Signalet går inn på høyttaleren og en får lyd.

Vi ser at jord i filterkretsen er halvparten av potensialfallet til driftsspenningen over operasjonsforsterkeren. Årsaken til dette valg av potensiale for jord i filterkretsen er som følger:

Inngangssignalet til operasjonsforsterkeren består av en spenning som varierer i positiv/negativ retning rundt et middelpotensiale. Av koblingsdiagrammet ser en at hvis operasjonsforsterkeren skal oppnå samme potensial på operasjonsforsterkerens to innganger, må den også variere sin utgangsspenning rundt samme potensiale som inngangssignalets middelpotensiale. Operasjonsforsterkerens vil kunne oppnå maksimal amplitudeforsterkning når den kan variere sin utgangsspenning tilsvarende halvparten av potensialfallet til operasjonsforsterkerens driftsspenning. Høyest amplitudeforsterkning får vi derfor ved et middelpotensiale på filterkretsen som tilsvarer halvparten av potensialfallet over operasjonsforsterkeren, dvs. 4,5V i eksempelet i koblingsdiagrammet.

Forøvrig kan en innenfor denne maksimalverdien for amplituden, regulere amplituden gjennom å justere størrelsen R_f/R_1 .

6. Konklusjon

Med arbeidet bak denne oppgaven har vi fått en god innsikt i hvordan aktive og passive filterkretser fungerer. Vi har også sett dette i praksis på laben gjennom målinger i filterkretser. Disse målingene har i stor grad vært i samsvar med teori. En naturlig videreføring av dette prosjektet ville vært å sett på hvordan aktive filterkretser fungerer i praksis.

Litteraturliste

- Alonso & Finn (1992) Physics (kap. 23.11)
- FAQ archives, URL http://www.faqs.org/docs/electric/Semi/SEMI_4.html
- Hartmann, William (1998) Signal, Sound and Sensation (kap. 9)
- Hyperphysics diode, URL <http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html>
- HyperPhysics opamp, URL <http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/opampvar.html#c2>
- Hyperphysics
Transistor, URL <http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html>
- Martinsen, URL <http://ri.hive.no/martinsen/fagside-01+02/met60/underlag%20uke%2003.doc>
- Nilsson&Riedel (2005) Electric Circuits (kap. 14, 15)
- Støvneng, Jon Andreas Notat om filterkretser
- Transistor, URL <http://www.lucent.com/minds>
- Winder, Steve (2002) Analog and Digital filterdesign (kap. 1, 2)

Elektromagnetiske bølger

Morten Ingvaldsen, Katrine Alne og Anfrid Dahlen

Prosjekt i FY1013 Elektrisitet og magnetisme II, NTNU, høst 2005

1 Innholdsfortegnelse

1 Innholdsfortegnelse

2 Abstract

3 Innledning

4 Teoretisk del

4.1 Historikk og dualisme

4.2 Utledning av bølgeligningen

4.3 Dispersjon

4.4 Det elektromagnetiske spekteret

4.5 Lyshastigheten

4.6 Inteferens

4.7 Diffraksjon

4.8 Polarisasjon

4.9 Anvendelser

5 Praktisk del, FM-sender

5.1 Sender og mottaker

5.2 Frekvensmodulasjon

5.3 LC-kretser

5.4 Transistor

5.5 Oscillator

5.6 Multivibrator

5.7 Forsterkere

5.8 Antenne

6 Fremgangsmåte og resultater

7 Konklusjon

8 Kilder

2 Abstract

I dette prosjektet har vi valgt å fordype oss innen elektromagnetiske bølger.

Prosjektet kan grovt sett deles inn i to deler. Den første delen består av teori rundt elektromagnetiske bølger. Vi ser først på de i et historisk perspektiv. Deretter utleder vi bølgeligningen for det elektriske og det magnetiske feltet. Videre ser vi på kjente bølgefenomener som dispersjon, diffraksjon, interferens og polarisasjon. I tillegg ser vi på det elektromagnetiske spekteret og lyshastigheten. Under avsnittet om lyshastigheten har vi gjennomført et forsøk, hvor vi målte lyshastigheten. Tilsist i den første delen tar vi for oss hvilke anvendelser elektromagnetiske bølger har.

Den andre delen er en praktisk del. Her har vi sett på anvendelser av elektromagnetiske bølger, og bygget en FM-sender for å vise et praktisk eksempel på hvordan man kan nyttegjøre seg en type elektromagnetiske bølger, nemlig radiobølger. I denne delen tar vi for oss oppbygging og virkemåte av FM-senderen. Vi betrakter hver av delene i kretsen for seg og forklarer oppbygging og virkemåte. I tillegg har vi tatt med en drøftings- og resultatdel for senderen.

3 Innledning

Det er en helt vanlig dag. Du står opp, skrur på radioen og lytter til siste nytt mens du spiser frokost. Deretter drar du på skolen på forelesning. Halveis ute i forelesningen er du lei, og tar opp mobilen for å taste på den i stedet. Kanskje finnes det noen andre likesinnede i en forelesningssal ikke så langt unna, som gladelig vil tekste litt med deg? Vel hjemme igjen tiner du middagsrester, sendt med fra Mamma, i mikrobølgeovnen. Du spiser foran tv'en så du får med deg siste episode av yndlingsserien. Deretter setter du deg ned foran pc'en og prøver å få gjort noe fornuftig, selv om det er vanskelig når man simpelthen ikke kan la være å koble seg på msn. Høres dette kjent ut?

Elektromagnetiske bølger er en naturlig del av hverdagen for alle av oss. Over er det kun nevnt et fåtall av de anvendelsene elektromagnetiske bølger har. I tillegg kan vi blant annet nevne røntgenbilder, kreftbehandling, solceller og fotosyntesen. Alt er eksempler på anvendelser av elektromagnetiske bølger!

I dette prosjektet har vi valgt å fordype oss innen nettopp elektromagnetiske bølger. Vi har jobbet ut fra problemformuleringen:

Hva er elektromagnetiske bølger, hvilke egenskaper har de, og hvordan kan de anvendes?

4 Teoretisk del, Elektromagnetiske bølger

4.1 Historikk og dualisme

I dag definerer vi lys som både bølger og partikler (fotoner), avhengig av hvilken sammenheng vi betrakter det i. Tidligere var situasjonen en annen. Fram til midten av 1800-tallet, da Newtons klassiske fysikk fortsatt var den dominerende, ble lys sett på som partikler. At lys kunne være bølger var likevel ikke helt utenkelig for flere kjente fysikere. På 1600-tallet, ble det oppdaget flere forhold ved lys som ikke kunne forklares ved å betrakte lys kun som partikler. Christian Huygens (1629-1695) og Pierre de Fermat (1601-1655) var blant de som mente at lys var bølger, og de framsatte teorier som i dag er grunnlaget for geometrisk optikk. Det tok likevel over 200 år før det ble lagt fram bevis for lysets bølgenatur. På begynnelsen av 1800-tallet beviste Thomas Young (1773-1829) lysets bølgenatur ved sitt velkjente dobbelspalteforsøk.

Heinrich Hertz (1857-1894) var den første som ved hjelp av eksperiment viste at elektromagnetiske bølger brer seg i vakuum med lysets hastighet. I 1887 klarte Hertz å produsere elektromagnetiske bølger med makroskopisk bølgelengde ved hjelp av LC-kretser. Han fant deretter en formel for hastigheten til bølgene, gitt ved $v = \lambda \cdot f$. Han fant også ut at denne hastigheten for elektromagnetiske bølger var lysfarten.

Maxwell hadde i 1873 argumentert for at elektromagnetiske bølger eksisterte, og at de måtte bevege seg med lysfarten. Dette hadde han kommet fram til gjennom betraktninger av det som senere har blitt kalt Maxwells ligninger.

Hertz bekreftet det Maxwell hadde funnet ved hjelp av Maxwells ligninger. Ved å kombinere ligningene kunne han utlede bølgeligningen for elektriske felt og for magnetiske felt. Dermed var det bevist at lys var elektromagnetiske bølger. Selv om det var Hertz som viste at lys er elektromagnetiske bølger, var det Maxwell som la fram ideen. SI-enheten til frekvens, hertz (Hz), er oppkalt etter Heinrich Hertz. Hertz tilsvarer s^{-1} .

Med Hertz og Maxwells bevis for at lys var elektromagnetiske

bølger ble partikkelteorien knust. Det skulle likevel vise seg at det ikke var så galt å betrakte lys som partikler. Einsteins fotoelektriske effekt (1905) viste at også lys har partikkelnatur. Først på 1930-tallet, med framveksten av kvante-elektrodynamikken, ble det lagt fram en teori som omfattet både partikkel- og bølgenaturen til lyset. Lys består av fotoner, som er masseløse partikler. Mange vil spørre seg hvordan fotoner kan være partikler når de ikke har masse. Selv om fotoner ikke har masse, har de likevel de aller viktigste egenskapene til partikler. De har impuls (massefart) og de har energi.

Dualisme er et begrep som ofte blir brukt i forbindelse med lys. For mange kan det være vanskelig å forstå hvordan lys kan være partikler og samtidig bølger. Hvorfor er det slik at noen fenomener ved lys kan forklares ved å se på lys som partikler, mens andre kan vi bare forstå ved å betrakte lys som bølger? Dette var en av grunnene til at vi ikke fikk en teori som inkluderte begge sidene til lyset før på 1930-tallet. Med Maxwells bevis på lysets bølgenatur, ble partikkelteorien knust. For datidens fysikere var det helt umulig at lys kunne være partikler i noen sammenhenger og bølger i andre tilfeller. For dem var det sort/hvitt; lys var enten bølger eller partikler.

Lysets dualisme gjør at vi kan forklare alle fenomener ved lys. Noen kan kun forklares ved å se på lys som partikler, mens andre framkommer når vi betrakter lys som elektromagnetiske bølger. Det er også fenomener som kan forklares både ved partikkelnaturen og bølgenaturen. Når vi betrakter hvordan lyset forplanter seg, må vi se på lysets bølgenatur. Under følger en oversikt over hvilke fenomener som forklares ved bølgenatur og hvilke fenomener som forklares ved hjelp av partikkelnatur.

Fenomen	Bølgenatur	Partikkelnatur
Diffraksjon	Ja	Nei
Interferens	Ja	Nei
Polarisasjon	Ja	Ja
Brytning	Ja	Nei
Farge	Ja	Ja
Compton-effekten	Nei	Ja
Fotoelektrisk effekt	Nei	Ja

Tabell 4.1.1

Kilder: [18], [19], [20].

4.2 Utledning av bølgeligningen

4.2.1 Maxwell's ligninger

And He said "let there be light"

$$\oint_A \mathbf{E} \cdot d\mathbf{A} = \frac{Q}{\varepsilon_0} \quad (1)$$

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = -\frac{\partial\Phi_B}{\partial t} \quad (2)$$

$$\oint_A \mathbf{B} \cdot d\mathbf{A} = 0 \quad (3)$$

$$\oint_S \mathbf{B} \cdot d\mathbf{S} = \mu_0 \mathbf{I} + \varepsilon_0 \mu_0 \frac{\partial\Phi_E}{\partial t} \quad (4)$$

Dette er Maxwells ligninger på integralform. \mathbf{E} beskriver det elektriske feltet, \mathbf{B} det magnetiske feltet og \mathbf{I} strømmen. \mathbf{A} er arealet vi integrerer over, \mathbf{S} er linjen vi integrerer rundt, Q er total ladning innenfor arealet, ε_0 er permittiviteten til vakuum, μ_0 er permeabiliteten til vakuum, Φ_B er magnetisk flukstetthet gjennom arealet omsluttet av s , Φ_E er elektrisk flukstetthet gjennom arealet A . Alle bokstaver i fete typer er vektorstørrelser.

Vi vil nå, med utgangspunkt i Maxwells ligninger på differensialform, utlede bølgeligningene for det elektriske feltet \mathbf{E} og det magnetiske feltet \mathbf{B} . På differensialform blir ligningene over som følger:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0} \quad (5)$$

$$\nabla \times \mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \quad (6)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (7)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \varepsilon_0 \mu_0 \frac{\partial\mathbf{E}}{\partial t} \quad (8)$$

Disse ligningene er gyldige for medier som beskrives som ladede punktpartikler i vakuum elektromagnetisk. \mathbf{J} er strømtettheten mens ρ er ladning per volumenhet. ∇ er den partiellderiverte med hensyn på x , y og z .

Vi vil her utlede bølgeligningene for vakuum, men de vil også med god tilnærming være gyldig for andre medier som elektromagnetisk kan betraktes som vakuum. I vakuum har vi ingen ladninger, og dermed ingen strøm. Det vil si at $\mathbf{J} = 0$ og $\rho = 0$. Vi får da fra ligningene over

$$\nabla \cdot \mathbf{E} = 0 \quad (9)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (10)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (11)$$

$$\nabla \times \mathbf{B} = \varepsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} \quad (12)$$

4.2.2 Utledning av bølgeligningen for det elektriske feltet \mathbf{E}

Vi har fra vektorrelasjoner at

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = -\mathbf{c}(\mathbf{a} \cdot \mathbf{b}) + \mathbf{b}(\mathbf{a} \cdot \mathbf{c})$$

Dermed kan vi skrive

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla^2 \mathbf{E} + (\nabla(\nabla \cdot \mathbf{E}))$$

Vi setter inn fra ligning (9) og får at

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla^2 \mathbf{E}$$

Altså er, i følge (10),

$$-\nabla^2 \mathbf{E} = \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right)$$

Her kan vi på høyresiden bytte om de partiellderiverte slik at

$$\nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right) = \frac{\partial}{\partial t}(\nabla \times \mathbf{B})$$

Deretter setter vi inn fra ligning (12)

$$\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = \frac{\partial}{\partial t} \varepsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} = \varepsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

Dermed har vi at

$$\nabla^2 \mathbf{E} = \varepsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (13)$$

Dette er bølgeligningen for det elektriske feltet \mathbf{E} i tre dimensjoner. Den generelle formen av bølgeligningen er

$$\frac{\partial^2 \mathbf{D}(x, t)}{\partial t^2} = \frac{1}{v^2} \frac{\partial^2 \mathbf{D}(x, t)}{\partial t^2} \quad (14)$$

hvor v er hastigheten bølgen forplanter seg med. Ved sammenligning av de to ser vi at

$$v = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \equiv c \quad (15)$$

altså er v nøyaktig lyshastigheten i vakuum.

4.2.3 Utledning av bølgeligningen for det magnetiske feltet \mathbf{B}

Nå vil vi utlede bølgeligningen for det magnetiske feltet \mathbf{B} . Vi starter på samme måte som for det elektriske feltet \mathbf{E}

$$\nabla \times (\nabla \times \mathbf{B}) = -\nabla^2 \mathbf{B} + \nabla(\nabla \cdot \mathbf{B})$$

Vi bruker ligningene (11) og (12) og får at

$$\nabla \times (\nabla \times \mathbf{B}) = -\nabla^2 \mathbf{B} = \nabla \times \left(\varepsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}\right) = (-\varepsilon_0 \mu_0) \frac{\partial^2 \mathbf{B}}{\partial t^2}$$

Ordnet gir dette at

$$\nabla^2 \mathbf{B} = \varepsilon_0 \mu_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} \quad (16)$$

som er bølgeligningen for det magnetiske feltet \mathbf{B} i tre dimensjoner. Ved sammenligning med (14) ser vi at også her er

$$v = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} = c$$

4.2.4 Plane, harmoniske elektromagnetiske bølger i vakuum

Vi betrakter harmoniske planbølger på formen

$$\mathbf{E} = \mathbf{E}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_E) \quad (17)$$

$$\mathbf{B} = \mathbf{B}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_B) \quad (18)$$

hvor $\omega = c|\mathbf{k}| = ck$. Disse to ligningene er en mulig løsning av (13) og (16). Det finnes også flere andre mulige løsninger. I tillegg til å oppfylle bølgeligningene må \mathbf{E} og \mathbf{B} også oppfylle Maxwells 4 ligninger (9-12).

Vi har fra (9) at divergensen til \mathbf{E} -feltet skal være 0, $\nabla \cdot \mathbf{E} = 0$. Vi undersøker divergensen til den harmoniske planbølgen vi valgte

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{E}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_E) = -\mathbf{E}_0 \cdot \mathbf{k} \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_E)$$

For at dette skal være lik 0 for alle \mathbf{r} og t må

$$\mathbf{k} \cdot \mathbf{E}_0 = 0 \Rightarrow \mathbf{k} \perp \mathbf{E}_0 \Rightarrow \mathbf{k} \perp \mathbf{E}$$

På tilsvarende måte finner vi at

$$\mathbf{k} \perp \mathbf{B}$$

Altså står bølgens forplantningsretning vinkelrett på retningen til bølgeutslaget. Det vil si at elektromagnetiske bølger er transversale.

Fra ligning (10) har vi at

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

Vi undersøker hvilke betingelser som må oppfylles for at dette skal stemme for bølgen vi har valgt. Venstre side i ligning (10) blir da

$$\nabla \times \mathbf{E} = (-\mathbf{k}) \times \mathbf{E}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_E)$$

Høyre side blir som følger

$$-\frac{\partial \mathbf{B}}{\partial t} = -\omega \mathbf{B}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_B)$$

Ordnet gir dette at

$$\mathbf{k} \times \mathbf{E}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_E) = \omega \mathbf{B}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_B) \quad (19)$$

Vi kan anta at både \mathbf{k} og \mathbf{E}_0 begge er $\neq 0$, siden det motsatte ikke ville gitt noen bølge. Vi ser dermed at $\mathbf{B}_0 \perp \mathbf{E}_0$. Fra før har vi i tillegg at

$$\mathbf{k} \perp \mathbf{B}_0$$

og

$$\mathbf{k} \perp \mathbf{E}_0$$

Vi bruker dette sammen og finner at

$$kE_0 \frac{\mathbf{B}_0}{B_0} \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_E) = \omega \mathbf{B}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi_B)$$

Vi ser at dersom denne skal gjelde for alle \mathbf{r} og t må $\phi_E = \phi_B$. Dette gir at (19) reduseres til

$$\mathbf{k} \times \mathbf{E}_0 = \omega \mathbf{B}_0$$

som igjen gir, ut fra ligningene (17) og (18), at

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B} \quad (20)$$

Oppsummert kan vi nå si at for plane, harmoniske elektromagnetiske bølger i vakuum står \mathbf{k} , \mathbf{E} og \mathbf{B} normalt på hverandre, og dersom vi bruker dette i ligning (20) får vi at

$$E = \frac{\omega}{k} B = cB \quad (21)$$

Fordi både \mathbf{E} og \mathbf{B} oppfyller bølgeligningen, som vist tidligere, følger det at ligning (4) er oppfylt. Dermed har vi vist at alle Maxwells ligninger er oppfylt, samtidig som bølgeligningen er oppfylt.

Kilder: [2], [4], [19], [20], [22], [23], [27].

4.3 Dispersjon

Dersom vi betrakter elektromagnetiske bølger i stoff vil vi se at disse ikke forflytter seg med samme hastighet som elektromagnetiske bølger i vakuum. Dette vil forekomme selv om det ikke finnes noen frie ladninger eller en strøm i mediet. Grunnen er at de elektromagnetiske bølgene induserer svingninger i de ladde partiklene i mediet, og dermed vil vi få nye bølger som danner en superposisjon sammen med den opprinnelige bølgen. Vi får ut en resultantbølge som forplanter seg med en hastighet $\neq c$ og som avhenger av frekvensen og bølgelengden til bølgen. Dette fenomenet kalles dispersjon.

I et materiale som er homogent og isotropt, det vil si ensartet og likt i alle retninger, kan man erstatte konstantene ϵ_0 og μ_0 med henholdsvis ϵ , som er permittivitetskonstanten til mediet, og μ , som er permeabilitetskonstanten til mediet. Den relative permittiviteten er da

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} \quad (22)$$

mens den relative permeabiliteten er

$$\mu_r = \frac{\mu}{\mu_0} \quad (23)$$

Bølgefarten i mediet blir nå

$$v = \frac{1}{\sqrt{\epsilon\mu}} \quad (24)$$

Den relative hastigheten til vakuum i forhold til mediet defineres som brytningsindeksen n .

$$n = \frac{c}{v} = v = \frac{1}{\sqrt{\epsilon_0\mu_0}} \sqrt{\epsilon\mu} = \sqrt{\epsilon_r\mu_r} \quad (25)$$

Kilder: [2], [4], [19], [20], [22], [23], [27].

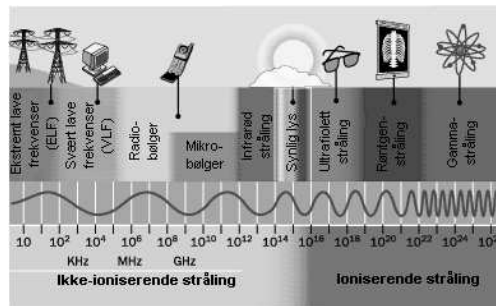
4.4 Det elektromagnetiske spekteret

Det finnes elektromagnetiske bølger med mange forskjellige frekvenser og energier. Energien til en enkelt lyspartikkel avhenger av frekvensen,

og derfor har lys med forskjellig bølgelengde forskjellige egenskaper. Vi deler vanligvis bølgene inn i sju kategorier.

Første kategori inneholder den strålingen som transporterer minst energi, radiobølger. Radiobølger har en frekvens fra bare noen få Hz til 10^9 Hz. Bølgene i denne kategorien er mye brukt innen teknologien for å sende trådløs informasjon gjennom lufta. Grunnen til at det er akkurat disse bølgene man bruker er at de er enkle å lage, og at de ikke er helsefarlige. De har rett og slett for lav energi til å kunne gjøre stor skade på naturen.

Mikrobølger har litt kortere bølgelengde. De har en frekvens fra 10^9 Hz til 3×10^{11} Hz. Ved siden av oppvarming av mat blir disse bølgene brukt i blant annet radar og andre kommunikasjonsmidler.



Figur 4.4.1 Det elektromagnetiske spekteret, kilde: [7]

Videre har vi infrarødt lys. De har en frekvens fra 3×10^{11} Hz til 4×10^{14} Hz, og blir produsert av molekyler og varme legemer, for eksempel et menneske med normal kroppstemperatur.

Synlig lys innebærer all stråling fra 4×10^{14} til 8×10^{14} Hz, altså et relativt lite intervall. Sammen med radiobølgene er det bare disse bølgene som klarer å trenge uforstyrret gjennom jordens atmosfære. Det er også verdt å legge merke til at det er denne typen stråling solen sender ut mest av.

Den mest energirike strålingen finner man i de tre siste kategoriene. Det er ultrafiolettstråling (UV-stråling), røntgenstråling og gammastråling. Ultrafiolett stråling omfatter alle bølger fra 8×10^{14}

til 3×10^{17} Hz, og blir produsert blant annet av eksiterte atomer og i elektriske utladninger. Energien fra dette lyset er stor nok til å ionisere atomer og spalte opp molekyler. Dette skjer når UV-lys treffer atmosfæren vår. Da blir energien brukt til å spalte opp ozongass. På denne måten hindrer ozonlaget at UV-lys skal komme seg ned på jordoverflaten.

Røntgenstråling blir produsert ved at elektroner blir akselerert mellom en katode og en anode. Ved anoden blir elektronene bremsert opp, og røntgenstråling blir sendt ut ved at elektronene vekselvirker med atomkjærnene. Denne bremsestrålingen er uavhengig av hvilket metall anoden er laget av.

Gammastråling blir sendt ut av radioaktive substanser. De har en frekvens i intervallet mellom 3×10^{18} og 3×10^{22} Hz, og er til stede i store kvanta i for eksempel kjernereaktorer. Siden gammastrålene har så stor energi kan de gjøre betydelig skade dersom de blir absorbert av celler, selv i små mengder. Likevel blir denne typen elektromagnetiske bølger brukt som behandling mot enkelte former for kreft.

Kilde: [20]

4.5 Lyshastigheten

4.5.1 Historisk perspektiv

Lenge var det en stor usikkerhet rundt lyshastigheten. Euklid (300-tallet f.Kr) var en av de første store til å uttale seg om den. Han antok at lyshastigheten måtte være veldig stor eller uendelig. Galilei (1564-1642) forsøkte å måle lyshastigheten, men uten hell. Basert på dette antok han at lyshastigheten måtte være uendelig. Descartes (1596-1650) uttalte seg også om lyshastigheten. Han antok at den måtte være uendelig.

Det stod en danske bak den aller første ordentlige målingen i 1676, nemlig Ole Rømer (1644-1710). Han beregnet den på grunnlag av tidspunktene for når Jupiter skjulte sin måne Io, og kom ved hjelp av dette meget nær den virkelige verdien for c . Både Huygens (1629-1695) og Newton (1642-1727) beregnet på grunnlag av Rømers

målinger verdien for lysets hastighet. Huygens beregnet den til å være lik $2,3 \times 10^8 \text{ m/s}$, mens Newton fant at den måtte være $2,4 \times 10^8 \text{ m/s}$

Debatten sluttet imidlertid ikke med dette, den pågikk frem til James Bradley i 1729 la frem sine resultater hvor han hadde oppdaget uregelmessigheter i lyset fra stjerner som en følge av jordens banebevegelse. På grunnlag av dette konkluderte han med at lyshastigheten måtte være endelig.

I 1849 foretok Armand Hippolyte L. Fizeau de første laboratoriemålingene av c . Fizeau målte også lyshastigheten i andre medier enn luft, og fant at disse var vesentlig mindre enn lyshastigheten i luft.

I dag er c en av våre fundamentale størrelser. Den ble definert til den verdien den har i dag i 1983 på grunnlag av målinger. Man definerer i dag meteren ut fra lyshastigheten ved å si at en meter er den strekningen lyset går på $1/c$ sekund. Lyshastigheten brukes også ofte for å beskrive astronomiske avstander ved at man kaller den avstanden lyset går i vakuum på et år for et lysår. Et lysår er ca. 10^{16} m. Lyshastigheten er i dag definert som

$$c = 299792458 \text{ m/s}$$

Kilder: [19], [20], [27]

4.5.2 Et smakfullt forsøk

I dette forsøket skal vi ved hjelp av en mikrobølgeovn, en pose marshmallows og en linjal forsøke å måle lysets hastighet.

Mikrobølger er en type elektromagnetiske bølger. De har en bølgelengde fra 0.3 m ned til 10^{-3} m. Frekvensen varierer mellom 10^9 og 3×10^{11} Hz. Mikrobølger brukes blant annet i mikrobølgeovner for å varme mat.

I en mikrobølgeovn har vi såkalt stående bølger. Stående bølger oppstår ofte når en bølge treffer et annet medium, slik at bølgen igjen blir reflektert. Summen av den innfallende og den reflekterte bølgen vil igjen se ut som en bølge. Denne bølgen vil imidlertid ikke bevege seg, men kun svinge frem og tilbake på samme sted, derav navnet stående bølge.

Når flere stående bølger når sin maksimalverdi, sin bølgetopp, på samme sted sier vi at vi får konstruktiv interferens. Disse bølgenes amplitude vil da adderes sammen, og man får ut en ny og større amplitude. Andre steder vil man få destruktiv interferens. Dette vil forekomme når det møtes bølger hvor noen har en bølgetopp mens andre har en bølgebunn. Også her vil resultantamplituden bli summen av de enkelte amplitudene. Resultatet blir et utslag som er mye mindre enn de opprinnelige. Dersom summen av utslagene i den positive retningen er nøyaktig likt summen av utslagene i den negative retningen vil vi få utsløkning, altså ikke noe bølgeutslag. De stedene hvor det forekommer konstruktiv interferens, vil maten bli mer oppvarmet. Mikrobølgeovnen vil altså ikke varme maten jevnt. For å forsøke å motvirke dette, setter man i dag inn en tallerken som kan rotere i mikrobølgeovnen. Slik vil resultatet bli jevnere. Avstanden mellom punktene hvor man finner konstruktiv interferens, altså hvor det blir mest brent, vil være en halv bølgelengde.

Hastigheten til en bølge er gitt ved

$$v = \lambda \cdot f$$

hvor λ er bølgelengden og f er frekvensen, eller svingninger per tidsenhet.

For å måle lyshastigheten fjernet vi først den roterende tallerkenen i mikrobølgeovnen, slik at alt skulle ligge helt i ro. Deretter dekket vi hele bunnen i mikrobølgeovnen med marshmallows. Vi kjørte så marshmallowsene på middels effekt i noen minutter. Det dannet seg da et mønster, marshmallowsene ble mer stekt noen steder enn andre. Vi målte avstanden mellom disse punktene, som er en halv bølgelengde. Til slutt leste vi av frekvensen til mikrobølgene i mikrobølgeovnen. Figur 4.5.2.1 viser marshmallowsene slik de så ut etter endt forsøk.



Figur 4.5.2.1 Marshmallowsene etter endt forsøk

Vi fant at

$$\lambda \approx 0,125 \text{ m}$$

og vi leste av frekvensen til å være

$$f = 2450 \text{ MHz}$$

Dermed kan vi regne ut hastigheten v til bølgene

$$\lambda \cdot f = 0,125 \text{ m} \cdot 2450 \times 10^6 \text{ Hz} = 3,06 \times 10^8 \text{ m/s}$$

Vi finner den relative feilen

$$\frac{v - c}{c} = 0.0215 \approx 2\%$$

Vi ser at vi kom svært nært den korrekte verdien for c . Usikkerheten ligger i avlesing av avstanden mellom bølgetoppene, samt i at marshmallowsene este mye ut og dermed flyttet bølgetoppene litt.

Kilde: [6]

4.6 Interferens

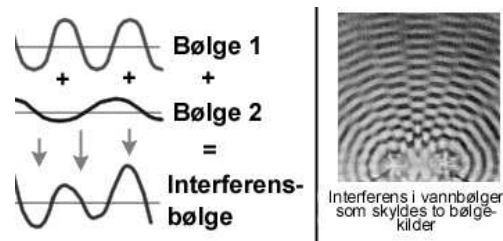
Interferens er et fenomen som oppstår når to bølger befinner seg på samme sted til samme tid. Dette gjelder ikke bare for vannbølger og lydbølger, men også for elektromagnetiske bølger. Når to bølger

overlapper hverandre kan vi finne resultantbølgen ved å superponere de to bølgene.

Betingelsen for at superposisjonsprinsippet skal gjelde, er at bølgeligningen er oppfylt, noe som vi nå vet gjelder for elektromagnetiske bølger. Vi vil da se at amplituden til resultantbølgen fås ved å addere amplituden til de to enkelte bølgene. På grunn av en forskjell i veilengde fra hver av bølgekildene til et gitt punkt, vil bølgene komme ut av fase. Resultantbølgen vil da få en amplitude som er ulik amplitudene til de to opprinnelige bølgene. Dette kaller vi interferens. Hvis begge bølgene oppfyller bølgeligninga, er det lett å finne resultantbølgen.

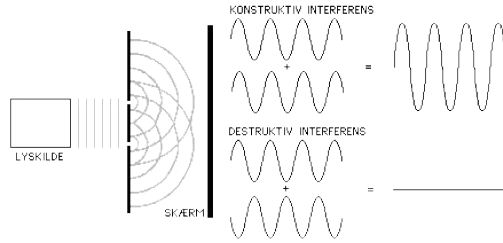
Utslaget til resultantbølgen finner vi ved å summere utslagene til de to bølgene på samme sted. Hvis de to bølgene har samme amplitude og motsatt fase, vil de to bølgene utslukke hverandre overalt. Dette kalles destruktiv interferens. Destruktiv interferens forekommer når en av bølgene har en bølgetopp mens den andre har en bølgebunn på akkurat samme sted. Bølgene har en veiforskjell gitt ved

$$\Delta r = r_1 - r_2 = \left(n + \frac{1}{2}\right) \cdot \lambda \quad (26)$$



Figur 4.6.1 Superposisjon av elektromagnetiske bølger, kilde: [9]

Dersom bølgene har bølgetopp på samme sted (det vil si at forskjellen i veilengde mellom de to bølgene er et helt antall bølgelengder), vil de to bølgene forsterke hverandre. Vi har da konstruktiv interferens. Mellom punkter med destruktiv og konstruktiv interferens, får vi delvis konstruktiv eller delvis destruktiv interferens.



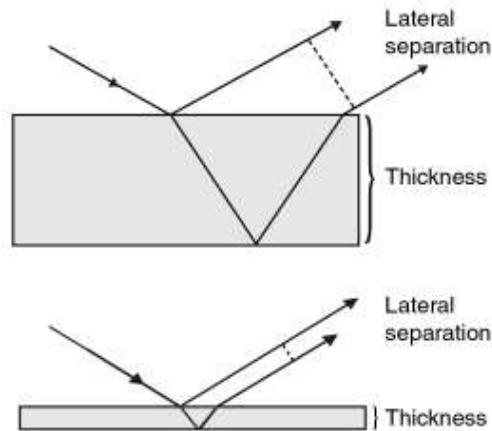
Figur 4.6.2 Konstruktiv og destruktiv interferens, kilde: [10]

Interferens kan oppstå på grunn av lysbrytning i tynne sjikt eller filmer. Dette kan vi for eksempel få dersom vi sender lys inn mot et tynt oljelag på vann. Siden vi nå har tre ulike stoff (luft, olje og vann), vil vi også ha tre ulike brytningsindekser. Lyset vil altså reflekteres fra to flater med ulik brytningsindeks. Når brytningsindeksen n_1 er mindre enn n_2 (se figur 4.6.3), så vil refleksjonene fra de to flatene ha motsatt fase. Hvis vi har

$$AB + BC = n \cdot \lambda \quad (27)$$

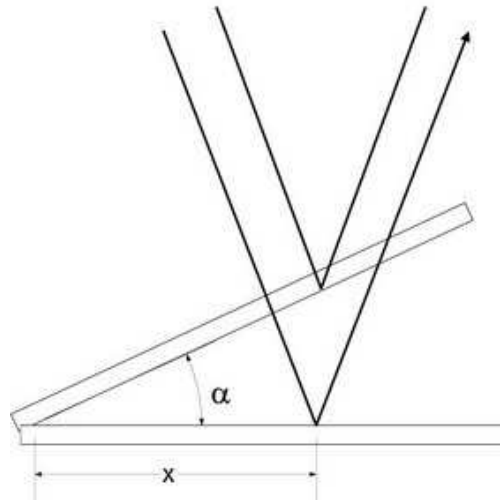
$$n = 0, \pm 1, \pm 2, \dots$$

vil vi få konstruktiv interferens gitt at $n_1 < n_2 < n_3$. Vi vil få destruktiv interferens dersom $n_1 < n_2$ og $n_3 < n_2$. n_1 , n_2 og n_3 er her brytningsindeksen til de ulike mediene.



Figur 4.6.3 Interferens på grunn av lysbrytning i tynne filmer, kilde [12]

Interferens fås også ved refleksjon. Vi sender en bølge inn mot ei flate med reflekterende egenskaper. Da vil en del av bølgen fortsette inn i det andre mediet mens resten av bølgen blir reflektert. Hvis det befinner seg ei reflekterende flate nedenfor den første flata vil den transmitterte bølgen treffe denne flata og bli delvis reflektert. Vi har nå to reflekterte bølger, og disse kan sammen skape interferens.



Figur 4.6.4 Interferens ved refleksjon fra to ikke-parallele flater, kilde [11]

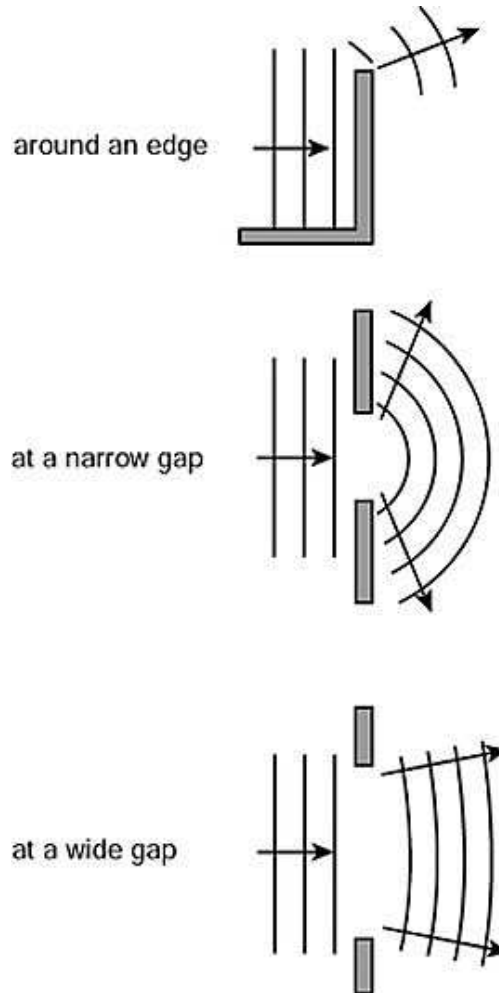
Et fenomen som kan oppstå når bølger møtes innenfor et avsperrert område, er stående bølger. Dette har vi for eksempel i mikrobølgeovner. Stående bølger er ikke vandrende bølger. Det vil si at det ikke er mulig å følge bølgen i en fase, da den ikke beveger seg fremover, men kun opp og ned i ethvert punkt. I slike tilfeller kaller man bølgen en svingning eller en stående bølge.

Kilder: [2], [18], [19], [20], [27]

4.7 Diffraksjon

Diffraksjon, eller spredning, er de mønstrene som dannes ved interferens når lys møter et hinder. Dette kan for eksempel være en spalte eller et objekt som hindrer deler av bølgen. Disse objektene er av størrelsesorden tilnærmet lik bølgelengden til lyset. Lyset vil

gå rundt, eller gjennom, disse hindrene, men de vil også endre retning eller bevegelsesmønster. Diffraksjon er altså spredning av lys rundt hjørner. Jo større forholdet mellom bølgelengde og åpning eller hindring er, jo større blir diffraksjonen eller bøyningen.



Figur 4.7.1 Diffraksjon, kilde [13]

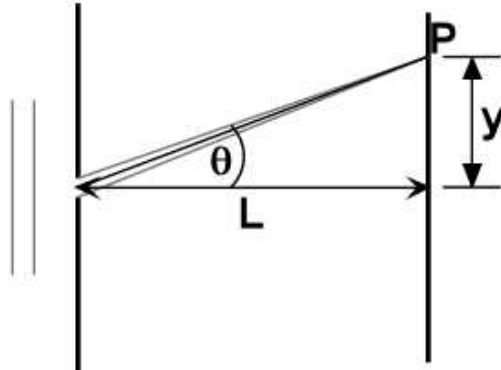
Diffraksjonsgitter blir brukt til å skille ut enkelte farger eller bestemte bølgelengder. Det finnes to typer diffraksjonsgitter; refleksjonsgitter og transmisjonsgitter. Et transmisjonsgitter består av mange parallelle spalter. Avstanden mellom spaltene kalles git-

terkonstanten d . Maksimal intensitet er gitt ved

$$d \sin \theta = n\lambda \quad (28)$$

$$n = 0, \pm 1, \pm 2, \dots$$

Forutsetninga for at denne formelen skal gjelde er at $L \gg \lambda$.



Figur 4.7.2 Diffraksjon, kilde [14]

Et refleksjonsgitter består av mange parallelle riper i ei reflekterende, metallisk flate. Ripene sørger for at lyset spres i alle ulike retninger, mens de reflekterende områdene mellom ripene fungerer på samme måte som spaltene i et transmisjonsgitter. Vi får maksima for intensiteten der faseforskjellen mellom reflektert lys fra to tilstøtende reflekterende flater er gitt ved

$$\phi = 2\pi n \quad (29)$$

$$n = 0, \pm 1, \pm 2, \dots$$

Kilder: [2], [18], [19], [20], [27].

4.8 Polarisasjon

Polarisasjon finner vi i alle transversale bølger. Siden alle elektromagnetiske bølger er transversale, vil hver enkelt elektromagnetisk

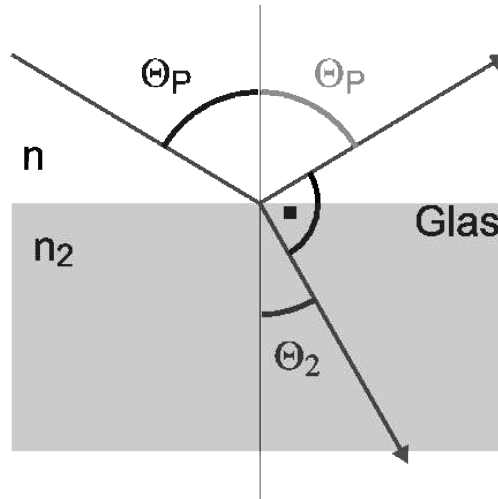
bølge alltid være polarisert. Det finnes flere ulike typer polarisasjon; lineær polarisasjon, sirkulær polarisasjon og elliptisk polarisasjon. Hva slags type polarisasjon vi har, er avhengig av hvordan dipolene i lyskilden svinger/vibrerer. Polarisasjonsretningen til lyset som sendes ut, vil da ha samme retning som vibrasjonsretningen til dipolene.

Naturlig lys er upolarisert. Det vil si at lyskilden består av molekyler med vilkårlig orientering. Lyset fra lyskilden vil da være lineært polarisert i alle mulige retninger. For å omforme upolarisert lys til polarisert lys, kan vi benytte det som kalles et polarisasjonsfilter. Polarisasjonsfilteret virker på den måten at det filtrer bort alt lys som ikke har den polarisasjonsretningen vi ønsker. Polaroid er det mest kjente polarisasjonsfilteret. Det finner man for eksempel i solbriller. Polaroid består av lange polymerkjeder med en bestemt retning, som er elektrisk ledende på langs men ikke på tvers. Dette gjør at de absorberer alt lys om er polarisert langs kjeden, og slipper i gjennom lys som har polarisasjonsretning på tvers av polymerkjedene.

Vi kan også oppleve polarisasjon ved refleksjon. Hvis vi sender inn lys mot en reflekterende overflate, og innfallsstrålen danner en vinkel θ med innfallsloddet, så vil en del av lyset reflekteres med samme vinkel, mens resten av lyset vil transmitteres ned i det andre mediet. Hvis innfallsvinkelen θ_1 er lik polarisasjonsvinkelen, eller Brewstervinkelen, θ_p til den gitte overflaten, så vil det reflekterte lyset være fullstendig polarisert. Hvis innfallsvinkelen ikke er lik polarisasjonsvinkelen, så vil det reflekterte lyset være delvis polarisert. Polarisasjonsvinkelen finner vi ved hjelp av Snells brytningslov

$$\begin{aligned} \frac{\sin \theta_1}{\sin \theta_2} &= \frac{n_2}{n_1} \\ \frac{\sin \theta_p}{\sin(\frac{\pi}{2} - \theta_p)} &= \frac{n_2}{n_1} \\ n_1 \cdot \sin \theta_p &= n_2 \cdot \sin(\frac{\pi}{2} - \theta_p) \\ n_1 \cdot \sin \theta_p &= n_2 \cdot \cos \theta_p \\ \tan \theta_p &= \frac{n_2}{n_1} \end{aligned} \tag{30}$$

Her er n_1 og n_2 brytningsindeksen til de to ulike mediene.



Figur 4.8.1 Polarisasjonsvinkel, kilde [15]

Polarisasjonen forteller generelt om i hvilken retning vi har forflytning. Hvis har forflytning, eller utslag, i y-retning, er polarisasjonen i y-retning. Polarisasjonen av elektromagnetiske bølger avhenger av retningen på det elektriske feltet. Polarisasjonen har samme retning som det elektriske feltet og svinger i henhold til det.

Kilder: [2], [18], [19], [20], [27].

4.9 Anvendelser

Siden elektromagnetisk stråling har så mange egenskaper, blir de brukt på forskjellige måter innen teknologi, kommunikasjon, energiproduksjon og i naturen. Noen eksempler er: solceller, sende trådløs informasjon (TV, radio, telefon), røntgenbilder og fotosyntese.

4.9.1 Solceller

Før produksjon av elektrisk strøm kan man bruke solceller. Prinsippet går ut på at lys treffer en pn-overgang på en halvleder (for eksempel silisium). Da vil energien fra lyset absorberes og brukt til å eksitere elektroner fra valensbåndet til ledningsbåndet. Elektronet i ledningsbåndet og hullet i valensbåndet vil dermed være fritt til å bevege seg og bidra med å generere strøm i solcellepanelet. På grunn av det naturlige elektriske feltet mellom p- og n-områdene, vil hullene bevege seg mot p-siden mens elektronene vil bevege seg mot n-siden. Dette vil resultere i en spenningsforskjell mellom n- og p-siden av halvlederen, og solcellepanelet kan derfor brukes som en spenningskilde.

4.9.2 Fotosyntese

I naturen finner man mange fine grønne planter som strekker seg etter sollyset for å overleve. Det som skjer er at lys fra sola treffer noen grønne flekker på bladene, som heter klorofyll. Disse flekkene absorberer energien som lyset gir. Denne energien blir brukt til å omdanne vann og karbondioksid til sukker og oksygen. Oksygenet blir skilt ut, mens sukkeret blir til mat for plantene.

4.9.3 Røntgenbilder

Røntgenstråling er bærere av så stor energi at den lett kan trenge seg igjennom materie. Det er denne egenskapen som blir utnyttet når man tar røntgenbilder. Siden beinvev og andre vev i kroppen absorberer denne strålingen forskjellig, vil vi få et ganske bra bilde av hvordan kroppen ser ut innvendig på den fotografiske filmen som tar imot den gjennomgående strålingen. Man bør imidlertid ikke utsette seg for mer stråling enn nødvendig på grunn av skadevirkningene når disse høyfrekvente bølgene blir absorbert av cellene i kroppen. Siden røntgenstråling også har den egenskap at den kan ødelegge celler, blir den ofte brukt til behandling av kreftpasienter der strålingen blir rettet mot kreftsvulster.

4.9.4 Trådløs informasjon

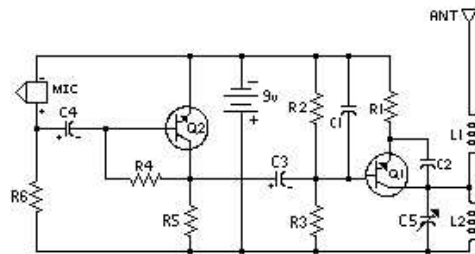
Når man sender trådløs informasjon ved hjelp av en mobiltelefon eller en fjernkontroll, utnytter man at elektromagnetiske bølger kan indusere spenning i en mottakende krets. Det er forskjellige måter å sende informasjon på. En godt brukt måte er amplitudemodulasjon (AM). En gammel AM-radio bygger på dette prinsippet. Det går ut på at en sender ut radiobølger med en fast frekvens og varierende amplitude. I radioen vil signalene bli likerettet slik at man bare står igjen med amplitudevariasjonene. Disse amplitudevariasjonene vil da generere lyd i radioens høyttaler.

Kilde: [19], [20], [24], [25].

5 Praktisk del, FM-sender

5.1 Sender og mottaker

I dette prosjektet har vi laget en sender, som en illustrasjon på hvordan man kan utnytte elektromagnetiske bølgers egenskaper. Senderen består ganske enkelt av en krets som gjør likestrøm om til vekselstrøm, en forsterker og en antenne. Når senderen er slått på, vil vi da få et varierende elektrisk felt i antenne. Dette varierende elektriske feltet vil da generere elektromagnetiske-bølger, som vil bre seg gjennom lufta til de treffer mottakerens antenne. Der vil bølgene indusere et elektrisk felt, som igjen vil generere en elektrisk strøm.



Figur 5.1.1 Koblingsskjema til FM-senderen vår, kilde: [5]

Av skjemaet over kan vi se at alle komponentene som ligger til venstre for kondensator C3 er forsterkerdelen. Nederst til høyre har vi en LC-krets som bestemmer frekvensen til bærebølgen. Transistoren som er koblet til LC-kretsen, skrur seg av og på i takt med svingningene. Dette vil hindre at vekselstrømmen i LC-kretsen dør ut.

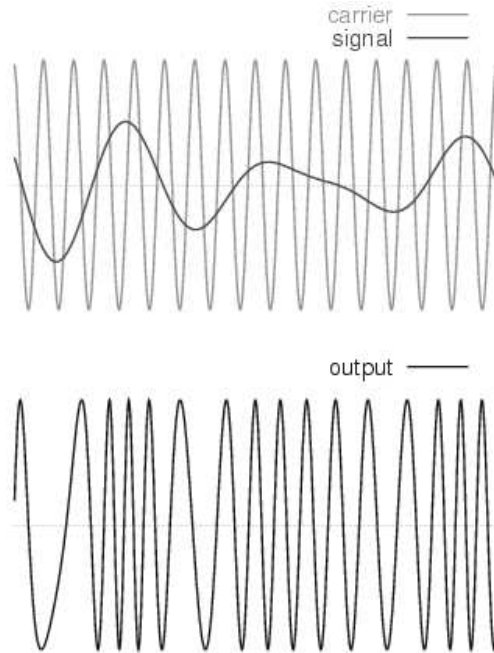
Komponent	Verdi
R1	207 Ω
R2, R5, R6	4,7 k Ω
R3	10 k Ω
R4	100 k Ω
C1	1 nF
C2	5,6 pF
C3, C4	10 μ F
C5	3 – 18 pF
L1	5 viklinger, 4 mm i diameter
L2	3 viklinger 1cm i diameter
T1, T2	2N3904

Tabell 5.1.1 Komponentverdier, kilde: [5]

5.2 Frekvensmodulasjon

I Norge er det FM-bånd som brukes til kringkasting. Dette er fordi FM-signaler tåler mer støy og forstyrrelser enn det AM-signaler gjør. FM står for frekvensmodulasjon, mens AM står for amplitudemodulasjon. Frekvensmodulasjon vil si at vi modulerer signalet slik at variasjonen i frekvens blir stor. Utgangssignalet vil da ha varierende frekvens. Vi påvirker på denne måten oscillatorfrekvensen direkte.

Ved hjelp av en mikrofon blir lavfrekvente signaler (det vil si lavfrekvent i forhold til frekvensen FM-senderen sender ut) sendt inn til forsterkeren, der signalet forsterkes. Derfra sendes signalene videre til oscillatoren. De lavfrekvente svingningene fra mikrofonen gjør at radiosignalets frekvens varierer. Kringkastet frekvens vil øke eller minske i takt med lydsvingningene. Når vi snakker med lavfrekvent lyd (mørk stemme) vil frekvensen svinge lite, men når vi snakker med høyfrekvent lyd (lys stemme) svinger utsendt frekvens mer. Det vil si at for 'høye' lydssignaler er frekvenssvingen større enn for 'lavere' lyder. Frekvenssvingen er altså forskjellen mellom ytterpunktene til frekvensen. Frekvensmodulasjonen bestemmer hvor mye frekvensen skal svinge ut til siden, mens lydssignalets tonefrekvens bestemmer hvor fort utsendt frekvens varierer.

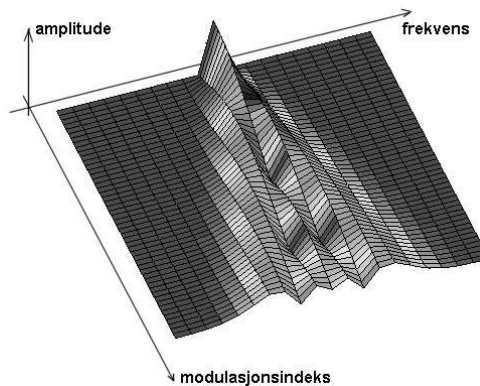


Figur 5.2.1 Frekvensmodulasjon, kilde [17]

Frekvensmodulasjon blir brukt når vi har frekvenser på over 30MHz. Dette er fordi FM-signaler trenger større båndbredde enn AM-signaler. I frekvensmodulasjonen har vi teoretisk sett et uendelig antall sidebåndkomponenter per modulasjonsfrekvens. Modulasjonsindeksen M forteller oss hvor kraftig modulasjonen er. Modulasjonsindeksen er gitt ved

$$M = \frac{\Delta f}{f_{LF}} \quad (31)$$

der Δf er halve frekvenssvingen og f_{LF} er det lavfrekvente inngangssignalet fra forsterkeren.



Figur 5.2.2 Frekvensmodulasjon og modulasjonsindeks, kilde [16]

Båndbredden avhenger av frekvenssving og frekvensen som sendes. Den forteller om hvor fort signalet varierer. Båndbredden er gitt ved

$$B = 2 \cdot (\Delta f + f_{LF}) \quad (32)$$

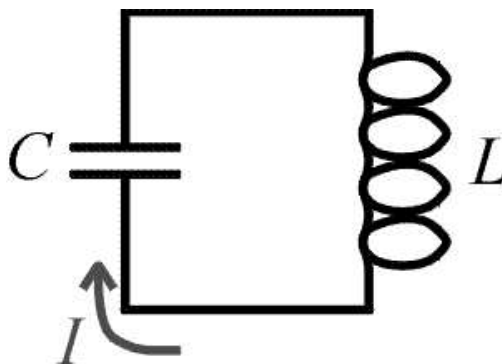
PM eller fasemodulasjon kalles ofte for indirekte frekvensmodulasjon. I dette tilfellet har oscillatoren stabil frekvens. I PM-sendere er det koblet inn et nettverk som påvirker signalet fra oscillatoren ved å endre fasen. En FM-mottaker kan ikke skille FM- og PM-signaler, fordi utsendt signal fra FM- og PM-sendere ser helt like ut.

Kilder:[2], [21]

5.3 LC-kretser

LC-kretser består ganske enkelt av en kondensator og en spole. De blir ofte brukt i både sendere og mottakere. Det som er spesielt med LC-kretser er at strømmen kan svinge harmonisk med en frekvens bestemt ut fra kapasitansen til kondensatoren og induktansen til spolen.

Bevis for at en LC-krets svinger harmonisk



Figur 5.3.1 LC-krets, kilde: [6]

Kirchhoffs andre lov gir:

$$\begin{aligned} \varepsilon_L - \varepsilon_C &= 0 \\ -L \frac{dI}{dt} - \frac{q}{C} &= 0 \\ \frac{d^2q}{dt^2} + \frac{q}{LC} &= 0 \end{aligned}$$

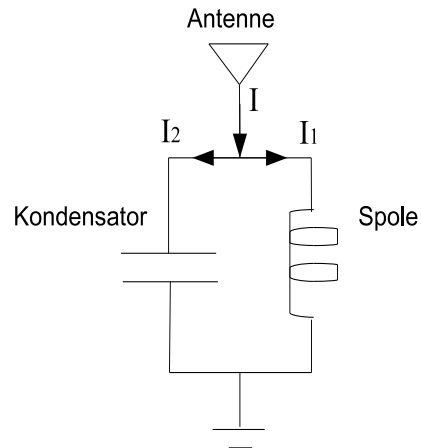
som har den generelle løsningen

$$q = Q_0 \cos\left(\frac{t}{\sqrt{LC}} + \phi\right)$$

Det vil si at dersom det går strøm i en LC-krets, så vil den svinge harmonisk med en vinkelfrekvens på $\omega=1/\sqrt{LC}$. Dette tilsvarer en frekvens på $f=1/(2\pi\sqrt{LC})$. I virkeligheten vil det alltid være en motstand i kretsen. Da har vi en RCL-krets. På grunn av motstanden ville frekvensen vært mindre og strømmen dødd ut eksponentielt.

Dersom vi har lyst til å bruke LC-kretsen som en oscillator i en sender, er vi nødt til å ta hensyn til dette. Da kan man lage en ny krets i tillegg, som hele tiden mater LC-kretsen med ny strøm til rett tid. Dette vil da opprettholde de harmoniske svingningene. I en mottaker utnytter man LC-kretsens egenskaper på en annen måte. Da utnytter man at en LC-krets, som blir utsatt for en ytre spenning bestående av mange forskjellige frekvenser, vil filtrere ut nesten alle frekvensene unntatt de som skaper resonans i kretsen. I

mottakere kobler man ofte en slik krets til antenna slik at bare de signalene man er interessert i slipper igjennom. Da kan man se på antenna som en spenningskilde som blir koblet inn i parallell med kretsen.



Figur 5.3.2 LC-krets med antenne koblet i parallell

Kirchhoffs første lov gir

$$I = I_1 + I_2$$

Vi setter spenningen fra antenna til $V_0 e^{i\omega t}$

Kirchhoffs andre lov gir

$$V_0 e^{i\omega t} = \frac{q}{C}$$

$$V_0 e^{i\omega t} = L \frac{dI}{dt}$$

$$\frac{q}{C} = L \frac{dI}{dt}$$

Vi lar $q = Q_0 e^{i\omega t}$ være en løsning. Det gir

$$V_0 e^{i\omega t} = \frac{Q_0 e^{i\omega t}}{C}$$

$$V_0 = \frac{Q_0}{C}$$

$$Q_0 = V_0 C$$

Det gir strømmen

$$I_{01} = \frac{dq}{dt} = iV_0 C \omega$$

Vi finner strømmen I_2 ved å la $I_2 = I_{02} e^{i\omega t}$

$$V_0 e^{i\omega t} = iL I_{02} \omega e^{i\omega t}$$

$$V_0 = iL I_{02} \omega$$

Det gir strømmen

$$I_{02} = \frac{V_0}{iL\omega}$$

Den totale strøamplituden er

$$I_0 = I_{01} + I_{02}$$

$$I_0 = V_0 \left(iC\omega + \frac{1}{iL\omega} \right)$$

Man kan se at strømmen i kretsen er avhengig av vinkelfrekvensen til den påtrykte spenningen. Ved å derivere dette uttrykket kan vi finne ut av hvilken frekvens som skaper resonans i kretsen.

$$\frac{dI}{d\omega} = V_0 \left(iC + \frac{-iL}{(iL\omega)^2} \right) = 0$$

$$\omega = \frac{1}{\sqrt{LC}}$$

Vi har et toppunkt i $\omega = 1/\sqrt{LC}$. Dette kan man forøvrig teste ved å derivere en gang til og se at den andrederiverte i punktet er negativ. Det vil si at dersom en antenne tar imot signaler med mange forskjellige frekvenser, så er det bare de signalene med frekvens nærmest $\omega = 1/\sqrt{LC}$ som slipper igjennom kretsen. Resten blir filtrert bort.

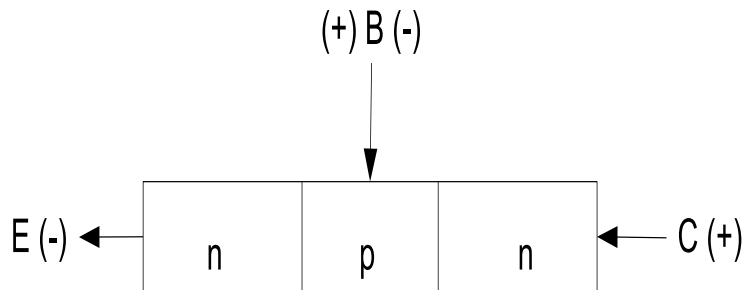
Kilde: [19]

5.4 Transistor

En transistor inneholder tre halvledere som er satt sammen.

Energien til ladningsbærerne i halvledere, ledere og isolatorer kan grovt deles inn i to områder. Det er valensbånd (E_V) og ledningsbånd (E_C). I ledningsbåndet finner vi elektronene med høyest energi, mens i valensbåndet finner vi elektroner med lavere potensiell energi. Det vil si at elektronene i valensbåndet ligger nærmere atomkjernen.

For halvledere er det slik at ledningsevnen øker for økende temperatur. Da vil den tilførte termiske energien bli overført til elektroner i valensbåndet slik at de kommer seg opp i ledningsbåndet. I valensbåndet vil det da bli en ledig plass. Disse plassene blir kalt for hull, og blir sett på som positive ladningsbærere. Dersom vi setter på en spenning over halvlederen, vil denne ledige plassen bevege seg mot minuspolen samtidig som elektronet i ledningsbåndet vil trekkes i motsatt retning. Etterhvert som temperaturen øker, vil det da bli flere elektroner i ledningsbåndet og dermed vil ledningsevnen øke. I transistorer er halvlederne dopet. Det vil si at det er enten tilført flere elektroner i ledningsbåndet, eller at det er blitt fjernet elektroner i valensbåndet. På denne måten har de økt ledningsevnen til halvlederen.

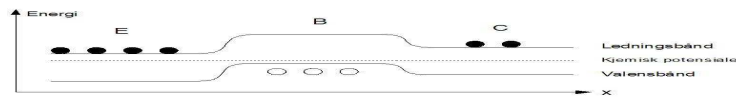


Figur 5.4.1 Pnp-transistor

Inne i en transistor finner vi tre dopete halvledere som er satt sammen. Halvlederen i midten er dopet med motsatt type ladningsbærere enn de på kantene. Det gir oss to hovedtyper transistorer, npn- og pnp-transistorer. En npn-transistor er dopet med negative ladningsbærere i hver ende mens den er dopet med positive i midten.

For en pnp-transistor er det motsatt. Siden komponenten består av tre områder av forskjellig dopete halvledere, finnes det normalt tre tilkoblingspunkter. Det er collector (C), base (B) og emitter (E).

Mellom de forskjellige dopete områdene vil det normalt være elektriske felter. Det er fordi elektronene nærmest grenseovergangen på n-siden vil falle ned i ledige tilstander på p-siden. De dopete feltene er i utgangspunktet elektrisk nøytrale, og det vil derfor være et overskudd av positiv ladning på n-siden mens det er overskudd av negativ ladning på p-siden. Dette resulterer i et lite elektrisk felt i grensene mellom n- og p-halvlederne. Av den grunn vil et elektron ha forskjellig potensiell energi i de forskjellige områdene. Energien som funksjon av lengden på transistoren vil da se slik ut:



Figur 5.4.2 Energibånd

Vi setter på en ytre spenning over over en npn-transistor, slik at vi har en stor spenning mellom collector og emitter, og moderat spenning over base og emitter. For at det skal gå strøm i komponenten, må vi ha en smal base. Da vil de fleste elektronene fra emitter komme over til collector uten å bli rekombinert med hull.

Dersom vi kaller strømmen inn i collector for I_C , inn i base for I_B og strømmen ut av emitter for I_E , kan vi nevne noen sammenhenger mellom dem. Kirchhoffs første lov eller knutepunkt-regelen gir:

$$I_C + I_B = I_E$$

Det vil si at vi har størst strøm ut av emitter. Det finnes også en nyttig sammenheng mellom I_C og I_B :

$$I_C = \beta_0 I_B$$

Strømmen inn i collector er altså proporsjonal med strømmen inn i base. På grunn av denne egenskapen egner transistoren seg ypperlig til å forsterke signaler. Dersom det går en positiv strøm inn i base, vil det også gå en positiv strøm inn i collector. Collector-strømmen vil variere i takt med base-strømmen, men er β_0 ganger større. På den måten blir signalene som kommer inn i base forsterket. En transistor kan også virke som en bryter. Det gjør den i en multivibrator, som blir diskutert senere.

Kilde: [26]

5.5 Oscillator

En FM-sender består av to hoveddeler; forsterker og oscillator. Det er oscillatoren som sender signalet til antenna. Frekvensen til det utsendte signalet på FM-senderen, blir bestemt av oscillatoren. En oscillator kan bestå av en spole med induktans L og en kondensator med kapasitans C . I tillegg har vi også en transistor. LC-kretsen skaper harmoniske svingninger gitt ved ligninga $q = Q_0 \cos(\omega t + \phi)$. Transistoren skaper resonans ved å svinge i takt med LC-kretsen og sørger dermed for at signalet ikke dør ut. Oscillatoren danner altså utgangssignalet, og sørger for frekvenssvingen. Frekvensen fra oscillatoren vil variere på bakgrunn av frekvensen til innsendt signal fra forsterkeren. Frekvensen til utgangssignalet er gitt ved:

$$f = \frac{\omega}{2\pi} \quad (33)$$

ω er her vinkelfrekvensen til svingningen og er per definisjon gitt ved $\omega = 1/\sqrt{LC}$.

$$f = \frac{1}{2\pi\sqrt{LC}} \quad (34)$$

Induktansen L til de selvlagde spolene finner vi på følgende måte:

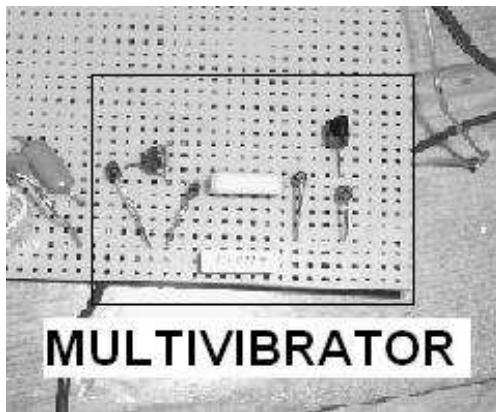
$$L = \frac{\mu_0 N^2 A}{l} \quad (35)$$

der N er antall viklinger, A er tverrsnittarealet og l er lengden til spolen. Tallverdien til induktansen blir da

$$L = \frac{4\pi \cdot 10^{-7} \cdot 3^2 \cdot \pi \cdot 0,005^2}{0,005} H$$
$$L = 1,78 \times 10^{-7} H$$

Kilder:[18], [19], [20], [21]

5.6 Multivibrator



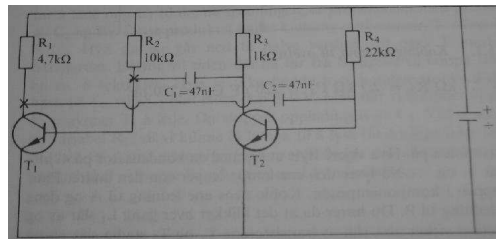
Figur 5.6.1 Multivibrator

En multivibrator er en elektrisk krets som har to ulike tilstander, og som brukes til å iverksette enkle systemer. Eksempel er transistorer eller tidtakere. En multivibrator fungerer slik at når den ene delen av kretsen er på, er den andre av. Multivibratoren er en slags bryter, og de to delene skrur hverandre av og på ved at den ene lader opp den andres kondensator, som går til base til den andre transistoren. Når den første skrur på, vil base til den andre transistoren gå til jord. Den andre kretsen, som er av, vil da lade seg ut.

Den vanligste typen multivibrator er en astabil multivibrator, og det er denne typen vi har benyttet i vår krets. En astabil multivibrator er ikke stabil i noen av tilstandene den varierer mellom, den

svinger kontinuerlig fra en tilstand til en annen. Den vil generere en firkantpuls. Andre typer multivibratore er monostabile og bistabile multivibratore.

En multivibrator vil svinge med en slik frekvens at det skapes et signal i kretsen som vi kan høre. Dersom man ønsker å endre denne frekvensen kan man bare endre komponentenes verdi. Å øke kondensatorens kapasitans fører for eksempel til at frekvensen til lyden synker. Vi har brukt multivibratoren som mikrofon i vår FM-sender.



Figur 5.6.2 Koblingsskjema for multivibrator, kilde [24]

Siden vi skulle bruke multivibratoren som mikrofon, måtte vi endre på noen verdier. Dette er de komponentene vi brukte:

Komponent	Verdi	Enhet
R_1, R_3	2, 2	$k\Omega$
R_2, R_4	33	$k\Omega$
C_1, C_2	1	μF
T_1, T_2	BC549C	

Tabell 5.6.1

Kilder: [1], [2], [3], [4], [21]

5.7 Forsterkere

En transistor kan forsterke en strøm. Man sender da både strøm inn gjennom kollektoren og inn gjennom basen, slik at strømmen som taes ut ved emmitter er summen av disse to. Vi definerer forholdet

mellom kollektorstrømmen og basestrømmen som strømforsterknin-
gen β_0 .

$$\beta_0 = \frac{I_C}{I_B}$$

Med strømforsterkning menes transistorens evne til å forsterke et signal. En liten basestrøm gir opphav til en mye større kollektorstrøm.

Kilder: [1], [2], [3], [4], [21]

5.8 Antenne

Kilden til elektromagnetiske bølger er elektriske ladninger i bevegelse. Derfor må det være elektriske ladninger i bevegelse i vår antenne. To vanlige antennetyper er elektrisk dipolantenne og magnetisk dipolantenne (spole).

5.8.1 Elektrisk dipolantenne

I denne typen antennene oscillerer dipolmomentet, derav navnet dipolantenne. Dersom det elektriske dipolmomentet hadde vært konstant, ville det bare vært et elektrisk felt rundt antenna, og dermed hadde det ikke blitt generert elektromagnetiske bølger. For at det skal bli sendt ut elektromagnetiske bølger må det altså også være et magnetisk felt tilstede. Det får vi dersom det elektriske feltet oscillerer. Da vil det gå en vekselstrøm i antenna, og Amperes lov sier da at det også må være et magnetisk felt tilstede. Bølgene vil bre seg utover med lysets hastighet, og med samme frekvens som de varierende feltene. Dersom vi plasserer en observatør i en avstand mye større enn antennes utstrekning, vil vi oppleve at intensiteten vil avhenge av retningen strålingen har. Det er fordi det amplituden til det elektriske feltet er proporsjonal med $\sin\theta$, der θ er vinkelen bølgeretningen danner med dipolen. Dette vil resultere i at vi får størst intensitet i planet vinkelrett på antenna, og null intensitet på aksene langs med dipolen.

Kilde: [19], [20]

5.8.2 Magnetisk dipolantenne

Som for en elektrisk dipolantenne må vi også her ha et oscillerende dipolmoment. Rundt en spole med konstant dipolmoment finnes det bare et magnetisk felt, og vi får ikke generert elektromagnetiske bølger. Dersom dipolmomentet varierer, vil fluksen inne i spolen også variere. Dette betyr at det også må være et oscillerende elektrisk felt tilstede, i følge Faraday - Henrys lov. Inne i spolen lar vi det magnetiske dipolmomentet peke langs z-aksen. Det vil si at det induerte elektriske feltet ligger i plan parallelt med xy-planet, nærmere bestemt i ringer rundt spolen. Intensiteten fra antenna vil også her avhenge av retningen i forhold til z-aksen. Dermed vil det være størst intensitet i xy-planet, og null intensitet på z-aksen.

Kilde: [20]

6 Fremgangsmåte og resultater

I den praktiske delen av dette prosjektet har vi blant annet lært at selv om en krets fungerer i teorien, så vil den ikke nødvendigvis fungere i praksis. Vi visste at dersom vi kunne få en krets til å oscillere ville vi kunne generere elektromagnetiske bølger, og det var dette vi gikk ut ifra da vi startet opp med den praktiske delen av prosjektet. Det å lage en oscillerende krets var imidlertid lettere sagt enn gjort. Det ble mye prøving og feiling. En lang stund var det lite som svingte i kretsen og mye som svingte i hodene våre!

Etter en tid kom vi over et kretsskjema for en multivibrator. Vi lagde en slik krets, og fant at multivibratoren genererte en vekselspenning. Deretter ville vi undersøkte om den genererte elektromagnetiske bølger. For å finne ut dette lot vi multivibratoren nærme seg radioens antenne samtidig som radioen stod på AM-modus. Resultatet ble en høyfrekvent lyd i radioens høyttaler. Vi hadde laget vår første sender!

Neste utfordring var å lage en mottaker. Etter å ha sett på flere kretsskjemaer og tenkt mye, kom vi frem til en skisse av en mulig mottaker. Tanken var at signalene som ble mottatt i antennen skulle filtreres i en LC-krets. Deretter skulle de ønskede signalene bli

forsterket, for så å bli brukt til og lage lys i en pære. Disse planene ble feid raskt av banen etter første møte med Nils Kristian Rossing. Nils Kristian Rossing er ansatt ved NTNU og ved Vitensenteret i Trondheim, og hadde sagt ja til å gi oss litt starthjelp på den praktiske delen av prosjektet. Han har lang erfaring med denne typen praktisk arbeid. Han mente vi hadde skissert en alt for komplisert krets, og vi la derfor nye planer sammen: Vi skulle lage en FM-sender med mikrofon.

Planen var altså å lage en sender som skulle sende ut bølger med en frekvens som kunne fanges opp av en vanlig radio. Etter å ha skaffet alle komponentene, begynte vi å lodde. Resultatet ble heller dårlig. Radioen fanget ikke opp noen signaler! Vi plasserte komponentene tettere sammen slik at det skulle bli mindre induktans i kretsen. Dette fordi kretsens induktans vil påvirke frekvensen til bærebølgen, noe som har en betydelig innvirkning på høyfrekvente LC-kretser. Dette ga imidlertid ikke noe resultat, vi hørte fortsatt ingenting i radioen.

Etterhvert kom vi på å regne ut hvilken frekvens signalene ble sendt ut med. Det ga oss feilen. Senderen sendte ut bølger med en frekvens på over 230MHz, mens en radio bare kan ta imot signaler med frekvenser mindre enn 108 MHz. Spolen i senderkretsen, som hadde fem viklinger og en diameter på 4mm, ble byttet ut med ny spole. Denne hadde tre viklinger og en diameter på 1 cm. Mikrofonen vår virket heller ikke, antageligvis fordi det er en mikrofon beregnet på datamaskiner. Vi erstattet den derfor med multivibratoren vi hadde laget tidligere. Neste forsøk på å sende signaler ble vellykket; senderen virket!

Induktansen til den nye spolen regnet vi til å være

$$L = 1,78 \times 10^{-7} H$$

Vi anslo at trimmekondensatoren hadde en kapasitans på omkring

$$C = 10 \times 10^{-12} F$$

Bærebølgen ble altså sendt ut med en gjennomsnittlig frekvens f :

$$f = \frac{1}{2\pi\sqrt{LC}}$$

$$f = \frac{1}{2\pi\sqrt{1,78 \times 10^{-7}H \times 10 \times 10^{-12}F}}$$

Det ga en frekvens på $f = 119$ MHz.

I praksis måtte radioen være stilt inn på 97 MHz for å høre signalene senderen sendte ut. Avvik mellom teori og praksis er derfor på 18.5%. Dette avviket skyldes at selve kretsen også har en induktans, ikke bare spolen. Det er også stor utsikkerhet i forbindelse med kapasitansen til kondensatoren, da det er vanskelig å vite nøyaktig hva trimmekondensatoren er justert til.

7 Konklusjon

Vårt prosjekt består av to deler: En teoretisk del, hvor vi har sett nærmere på elektromagnetiske bølger, og en praktisk del, hvor vi har bygget en FM-sender.

Ved å bruke enkle hjelpemidler de fleste har hjemme, en mikrobølgeovn, en linjal og en pose marshmallows, beregnet vi lyshastigheten. Dette forsøket gjorde vi mest for moro skyld og for å komme i gang med oppgaven. Vi fant at lysfarten var

$$c = 3,06 \times 10^8 \text{ m/s}$$

med en feil på ca. 2 %.

Vi bygde en FM-sender ut fra et koblingsskjema vi fant på Internett. Ved hjelp av en vanlig radio detekterte vi elektromagnetiske bølger generert av senderen. De utsendte bølgene hadde en frekvens på 97MHz. Induktansen til de hjemmelagde spolene vi brukte i den endelige senderen var

$$L = 1,78 \times 10^{-7} \text{ H}$$

I tillegg satte vi oss inn i bruk av teksteditoren LaTeX. Vi ønsket å skrive prosjektet i LaTeX av flere grunner. Først og fremst fordi det er et anvendelig verktøy når man skal fremstille matematiske formler, men også fordi det endelige resultatet blir mye penere enn dersom man bruker programmer som for eksempel Word. Dessuten ønsket vi å lære oss å bruke LaTeX fordi vi antageligvis vil få stor nytte av dette senere i studiet. Selv om det var en god del å sette seg inn i til å begynne med, angret vi ikke på at vi valgte denne løsningen.

8 Kilder

Internettsider

<http://www.tpub.com> [1]

<http://www.wikipedia.org> [2]

<http://www.uio.no/studier/emner/matnat/fys> [3]

<http://www.caplex.net> [4]

<http://www.geocities.com/tomzi>.

[geo/2-transistor/2-transistor.htm](http://www.geocities.com/tomzi/geo/2-transistor/2-transistor.htm) [5]

<http://www.physics.umd.edu/>

[ripe/icpe/newsletters/n34/marshmal.htm](http://www.physics.umd.edu/ripe/icpe/newsletters/n34/marshmal.htm) [6]

<http://www.energifakta.no/documents/Miljo> [7]

<http://www.pa.msu.edu/courses/1998spring/>

[PHY232/lectures/accircuits/lccircuits.html](http://www.pa.msu.edu/courses/1998spring/PHY232/lectures/accircuits/lccircuits.html) [8]

<http://per.jensen10.person.emu.dk/Virtuelle.htm> [9]

<http://users.cybercity.dk/kam1966/kwiat.htm> [10]

<http://cnx.rice.edu/content/m12911/latest> [11]

http://plus2physics.com/optical_interference [12]

<http://www.tiscali.co/uk/> [13]

<http://www.wfu.edu/physics/pira/PhysicsDrawings.htm> [14]

<http://wwwex.physik.uni-ulm.de/lehre/>

[gk3a-2003/img468.gif](http://wwwex.physik.uni-ulm.de/lehre/gk3a-2003/img468.gif) [15]

<http://fag.grm.hia.no/olata/infoteo/lydeffekter.htm> [16]

http://en.wikipedia.org/wiki/Frequency_modulation [17]

Bøker

Young og Freedman *University Physics*, Addison-Wesley Publishing Company, Inc. 1996 [18]

Egil Lillestø, Ola Hunderi og Jan R. Lien *Generell fysikk for universiteter og høyskoler Bind 2, Varmelære og elektromagnetisme*, Universitetsforlaget 2001 [19]

Marcelo J. Alonso og Edvard J. Finn *Physics*, Addison Wesley Longman Ltd. 1992 [20]

Norsk Radio Relæ Liga *Alfa Bravo Charlie for radioamatører*, Universitetsforlaget 1972 [21]

Hefter

Karl Rottmann *Matematisk formelsamling*, Spektrum Forlag 2003 [22]

Carl Angell og Bjørn Ebbe Lian *Fysiske størrelser og enheter*, Universitetsforlaget 2004 [23]

Innføring i elektronikk Vedleggshefte til elektronikkstartpakke [24]

Notater

Notat om solceller (Fått av Jon Andreas Støvneng) [25]

Forelesningsnotater, transistor (Fra FY1013) [26]

Forelesningsnotater, elektromagnetiske bølger (Fra FY1002) [27]

Prosjektrapport i FY1013

Spektroskopi



**Av Ingebjørg Berglie og
Lene Grøterud Leer**

1 Abstrakt

Allerede tidlig på 1800-tallet oppdaget fysikeren og optikeren Joseph von Fraunhofer de mørke linjene i solens spektrum, seinere kalt de fraunhoferske linjene, og dette dannet grunnlaget for den astronomiske spektroskopien. Det gikk imidlertid ikke lang tid før flere vitenskapsmenn, som arbeidet innenfor andre områder, fattet interesse for dette nye fenomenet, og etter 200 års forskning har vi i dag en rekke spektroskopiske metoder som vi kan benytte i vår søken etter kunnskap om verdenen rundt oss. En felles beskrivelse av de spektroskopiske metodene er at de alle bruker elektromagnetisk stråling til å analysere et materiale, og for kjemikere er disse metodene så å si uerstøttelige i hverdagen. Selv om det nesten utelukkende er fysikere som har arbeidet med å utvikle spektroskopien til det vi kjenner den som i dag, er det er det først og fremst de fysikerne som jobber med faste stoffer og biofysikk som benytter seg av spektroskopi i det daglige arbeidet.

2 Innholdsfortegnelse

<u>1</u>	<u>ABSTRAKT</u>	2
<u>2</u>	<u>INNHOLDSFORTEGNELSE</u>	3
<u>3</u>	<u>INNLEDNING</u>	4
<u>4</u>	<u>HISTORIE</u>	5
<u>5</u>	<u>ELEKTROMAGNETISK STRÅLING</u>	8
<u>6</u>	<u>SPEKTROSKOPISKE METODER</u>	10
6.1	<u>Spektralanalyse</u>	10
6.2	<u>Massespektroskopi</u>	11
6.3	<u>Infrarød spektroskopi</u>	14
6.4	<u>Ramanspektroskopi</u>	16
6.5	<u>UV/Synlig spektroskopi</u>	18
6.6	<u>Mikrobølgespektroskopi</u>	18
6.7	<u>Røntgenspektroskopi</u>	19
6.8	<u>NMR</u>	21
<u>7</u>	<u>SPEKTROSKOPI I PRAKSIS</u>	24
7.1	<u>Studiet av verdensrommet</u>	24
7.2	<u>Våre spektroskopiske opplevelser</u>	26
<u>8</u>	<u>OPPSUMMERING</u>	27
<u>9</u>	<u>AVSLUTNING</u>	28
<u>10</u>	<u>TAKK TIL</u>	28
<u>11</u>	<u>LITTERATURLISTE</u>	29
11.1	<u>Bøker</u>	29
11.2	<u>Internett</u>	29

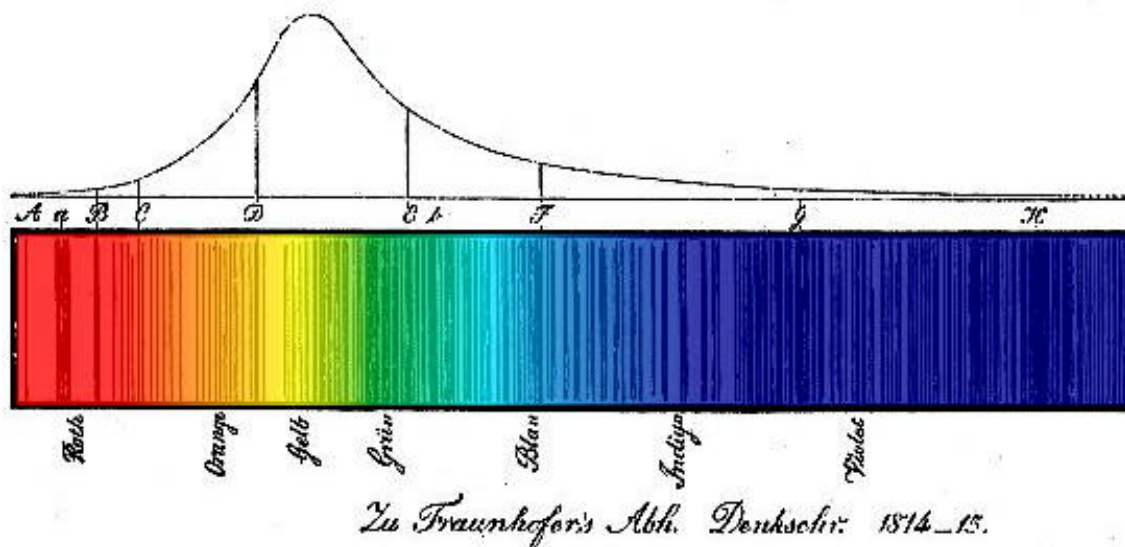
3 Innledning

Spektroskopi er en fellesbetegnelse for studiet av vekselvirkningene mellom elektromagnetisk stråling og materie, og de eksperimentelle dataene som de spektroskopiske studiene gir er bølgelengdene eller frekvensene til strålingen. I dette prosjektet skal vi ta for oss forskjellige spektroskopiske metoder som har hatt mye å si for nåtidens forskning innenfor fysikk, kjemi, medisin og astronomi. Opp igjennom tidene har forskere over hele verden vært på jakt etter nye metoder for å skaffe seg ny kunnskap om den verdenen vi lever i, enten det er på atomnivå eller om universet. Spektroskopi har siden tidlig på 1800-tallet hatt en sentral rolle i denne utviklingen, og metoden er fortsatt svært viktig innenfor dagens forskning. I dette prosjektet, vil vi ta for oss noen av de historiske begivenhetene som har lagt grunnlaget for den spektroskopiske utviklingen.

Siden den elektromagnetiske strålingen er avgjørende for spektroskopiske studier, skal vi i tillegg se litt nærmere på den. Vi skal også prøve å finne ut mer om bruksområdene til de forskjellige metodene, samt sette oss inn i teorien bak noen utvalgte metoder. Spektroskopi er en eksperimentell analysemetode som er svært viktig innenfor flere vitenskapelige forskningsområder, og for å få et inntrykk av nytten av de forskjellige metodene, skal vi forsøke å få til et besøk hos noen som bruker spektroskopi i det daglige arbeidet. Spektroskopi er også en svært viktig teknikk for flere ulike forskningsområder, og vi vil derfor ta en nærmere kikk på disse.

4 Historie [14-19]

Allerede i 1802 observerte kjemikeren og fysikeren William Hyde Wollaston (1766-1828) mørke linjer i solspekteret. Dette tolket han som mellomrom som separerte fargene i sollyset, og han fattet ikke mer interesse for disse. I 1814 gjenopplaget så den tyske fysikeren og glassliperen Joseph von Fraunhofer (1787-1826) disse linjene mens han jobbet med å undersøke de optiske egenskapene til forskjellige typer glass. Disse linjene kjenner vi i dag som fraunhoferske linjer, og det var denne oppdagelsen som la grunnlaget for spektralanalysen av stoffer i gasstilstand. Fraunhofer beskrev i alt 574 linjer i solspekteret, og de sterkeste fikk tilegnet en bokstav, fra A til Z. Han skjønte etter hvert at fargeskilleteorien til Wollaston ikke kunne være riktig, siden det ikke var en diskontinuitet i fargespekteret der de mørke linjene oppstod. Seinere oppdaget han også at spekteret fra forskjellige stjerner manglet noen av de mørke linjene som han observerte i solspekteret og motsatt. Den engelske, tyskfødte astronomen, Sir Friedrich Wilhelm Herschel (1738-1822), kom tidlig på 1800-tallet med en teori om at fraunhofer linjene enten kunne bli dannet som følge av absorpsjon av en kald gass i jordens eller solens atmosfære, og i dag vet vi at alle lysende gasser absorberer fotoner som har den samme bølglengden som de eksiterte atomene sender ut. De mørke linjene i absorpsjonsspekteret til en gass tilsvarer bølglengdene til fotonene som er absorbert, og alle stoffer har et eget spekter ved visse forutsetninger. Denne oppdagelsen revolusjonerte den klassiske fysikken siden man hittil hadde trodd at elektroner kunne absorbere en hvilken som helst energimengde, men nå viste det seg at dette ikke kunne være tilfellet.

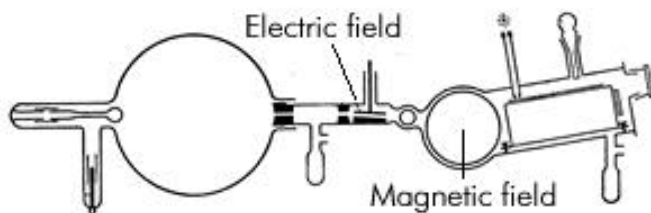


Figur 1: Fraunhoferske linjer i solspekteret [15]

Herschel hadde noen år tidligere, i år 1800, studert temperaturen til de forskjellige fargene i sollyset ved å sende lyset gjennom et glassprisme, og han oppdaget da at temperaturen til fargene økte fra den lilla til den røde delen av spekteret. Etter å ha oppdaget dette mønsteret bestemte han seg for å måle temperaturen like utenfor den røde delen av spekteret, og til hans store overraskelse hadde dette området den høyeste temperaturen av alle. Herschel gjorde så flere forsøk på denne "varmeprodukerende strålingen" og fant ut at den oppførte seg akkurat som synlig

lys. Dette var første gang det ble påvist at det finnes stråling som vi ikke kan se med øynene våre (bølgelengde utenfor området mellom 400 nm og 700 nm), og strålingen Herschel hadde oppdaget kalles i dag infrarød stråling. Infra betyr nedenfor, noe som kommer av at strålingen ligger under det synlige røde lyset. I dag brukes infrarød stråling i den infrarøde spektroskopien, samt innenfor medisin, geologi, etterforskning av kriminalsaker, brannsløkking, militæret, registrering av varmetap i hus og mye mer. Senere har det blitt oppdaget flere typer usynlig stråling, blant annet radiobølger, mikrobølger, ultrafiolett stråling, røntgenstråling og gammastråling. Alle disse typene har fått fellesbetegnelsen elektromagnetisk stråling, og de kjennetegnes ved at de er bølger som forplanter seg ved lysets hastighet i rommet med en elektrisk og en magnetisk komponent. Skotten James Clerk Maxwell (1831-79) antok allerede i 1864 eksistensen av elektromagnetiske bølger da han samlet de fire ligningene (Maxwells ligninger) som beskriver hvordan elektriske og magnetiske felt oppfører seg, men den elektromagnetiske strålingen ble ikke påvist før i 1887 da den tyske fysikeren Heinrich Rudolph Hertz (1857-94) klarte å lage et apparat som kunne sende ut radiobølger. Spektroskopien har tatt i bruk flere av de elektromagnetiske strålingstypene for å analysere og gjøre avbildninger av materialer, og det arbeides fortsatt med å utvikle instrumenter som kan utnytte de resterende typene.

I løpet av 1900-tallet stod begivenhetene i kø for spektroskopien, og spektroskopiske metoder ble tatt i bruk innenfor de fleste naturvitenskapelige områder. I 1909 oppdaget den britiske fysikeren Joseph John Thomson (1856-1940), under jakten på partikkelen som fraktet positiv ladning, at alle grunnstoff hadde forskjellige ioner og til



Aston's design for the mass spectrograph.

Figur 2: Astons første modell for spektrografen [18]

et elektromagnetisk felt klarte å ionisere og separere atomene i neongassen. De kunne da observere en liten forskjell i massen og kom fram til at dette måtte bety at grunnstoffet neon har to forskjellige atomkjerner, kalt isotoper. Ved å undersøke en rekke andre grunnstoffer oppdaget Aston i alt 212 naturlig forekommende isotoper, og for dette arbeidet fikk han seinere nobelprisen i kjemi.

Thomson hadde tidligere laget et apparat for å bestemme massen til atomkjerner, kalt "Thomsons parabol apparat". I dette apparatet ble en partikkelstrøm (ioner) utsatt for et elektromagnetisk felt, og han kunne da se at partikkelstrømmen ble avbøyd. Han beregnet så farten til partiklene til å være lik styrken på det elektriske feltet dividert på styrken på det magnetiske feltet. Thomson kunne også beregne forholdet mellom massen og ladningen til partikkelstrømmen med apparatet sitt, ved å slå av det elektriske feltet og måle graden av avbøyning. Selv om parabolapparatet til Thomson ga en verdi for forholdet mellom massen og ladningen til partiklene i strømmen, var ikke metoden så veldig nøyaktig. Aston ville derfor prøve å lage et

hans store overraskelse hadde neon to. Sammen med den britiske fysikeren og kjemikeren Francis William Aston (1877-1945) begynte han å undersøke dette fenomenet, og prosessen førte til utviklingen av massespektrografen (massespektrometeret). Aston, som jobbet som assistent for Thomson, stod for utviklingen av dette apparatet som ved hjelp av

apparat som var mer nøyaktig, og den første forandringen han foretok var å skille det elektriske og det magnetiske feltet slik at partiklene først ble avvært av det elektriske feltet før de kom inn i det magnetiske feltet. I Thomsons modell ble partiklene beveget både til høyre og venstre, samt opp og ned av det elektromagnetiske feltet, men ved å rotere det elektriske feltet nitti grader ville partiklene bare kunne bevege seg opp og ned. Disse justeringene gjorde at hastigheten til partiklene ikke lengre hadde betydning for hvor de traff fotopapiret Aston brukte til å registrere målingene (detektorskjerm), og alle partikler med samme masse ville da treffe detektoren i samme punkt. Aston hadde laget den første massespektrografen (seinere kalt massespektrometer) til å beregne massen til atomkjerner, og han fortsatte med dette arbeidet fram til 1937 da han kunne vise fram sin endelige massespektrograf. Denne modellen var omtrent hundre ganger mer nøyaktig enn den første utgaven, og noen av endringene Aston hadde gjort var å bruke bøyde plater til å produsere det elektriske feltet, magnetfeltet var blitt mye kraftigere og spaltene til å fokusere partikkelstrømmen kunne nå justeres. I dag brukes massespektrometeret først og fremst innenfor kjemien



Figur 3: Astons tredje og siste spektrograf [18]

Et område hvor spektroskopien har hatt en spesielt stor betydning er innenfor astronomien, vitenskapen om verdensrommet. Dette er den eldste naturvitenskapen vi har, og den første astronomiske observasjonen ble registrert allerede 4000 f. Kr. Studiet av universet og stjernene er svært utfordrende siden disse elementene ikke er noe vi kan ta med inn på laben til observasjon, og i den forbindelse har de spektroskopiske metodene vært svært viktige. Etter at de fraunhoferske linjene i solspekteret ble oppdaget tidlig på 1800-tallet begynte flere astronomer og fysikere å studere lyset fra stjernene og sola med spektroskop, og en av disse var nordmannen Carl Frederik Fearnley (1818-1890). Han observerte sola med et spektroskop og klarte etter hvert å lage detaljerte tegninger av protuberanser eller gasskyer i solens atmosfære. Den britiske astronomen William Huggins (1824-1910) klarte i 1864 å registrere det første stjernespekteret på en fotografisk plate. Denne registreringen gjorde det mulig å bestemme stjernenes kjemiske sammensetning, siden alle grunnstoffer har et eget spekter, og hastighet. For å bestemme hastigheten ble teorien den østerrikske fysikeren Christian Andreas Doppler (1803—1853) kom fram til i 1842 tatt i bruk, den såkalte dopplereffekten. Denne sier at bølgelengden til lyset (eller lyden) fra en bevegelig kilde endres når kilden og observatøren beveger seg i forhold til hverandre, og dermed kan hastigheten til stjernene bestemmes ved å måle forskyvningen av tilhørende spektrallinjer. Spektroskopi ble seinere tatt i bruk i studiet av dobbeltstjerner og stjernenes rotasjon. Dobbeltstjerner er stjerner som står meget nær hverandre på himmelen, og disse kan da avsløres på grunn av variasjon i det tilhørende spekteret. Vitenskapsmenn kunne ved hjelp av spektralanalysen fastslå at stjernene inneholder mange av de samme grunnstoffene vi finner på jorda, men et grunnstoff, nemlig helium (det gresk ordet for sol) ble faktisk funnet i verdensrommet før det ble funnet her på jorda.

5 Elektromagnetisk stråling [16, 20-21]

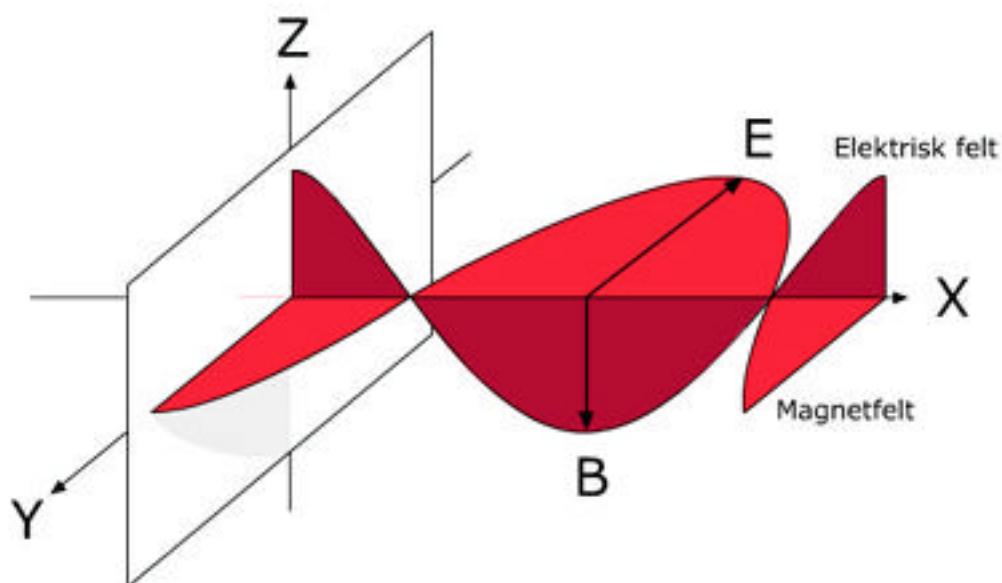
Alle spektroskopiske metoder baserer seg på det faktum at atomer kan ta opp (absorbere) eller sende ut (emittere) energi i form av elektromagnetisk stråling, og målet med metodene er å finne ut noe om det materialet som analyseres.

Elektromagnetisk stråling er i realiteten overføring av energi i form av elektromagnetiske bølger, hvor strålingen består av "energipakker" eller fotoner som hver har en energi som er avhengig av bølgelengden eller frekvensen. I vakuum eller luft gjelder følgende:

$$E = \frac{h \cdot c}{\lambda} = h \cdot f, \text{ der } c = \lambda \cdot f \quad (1)$$

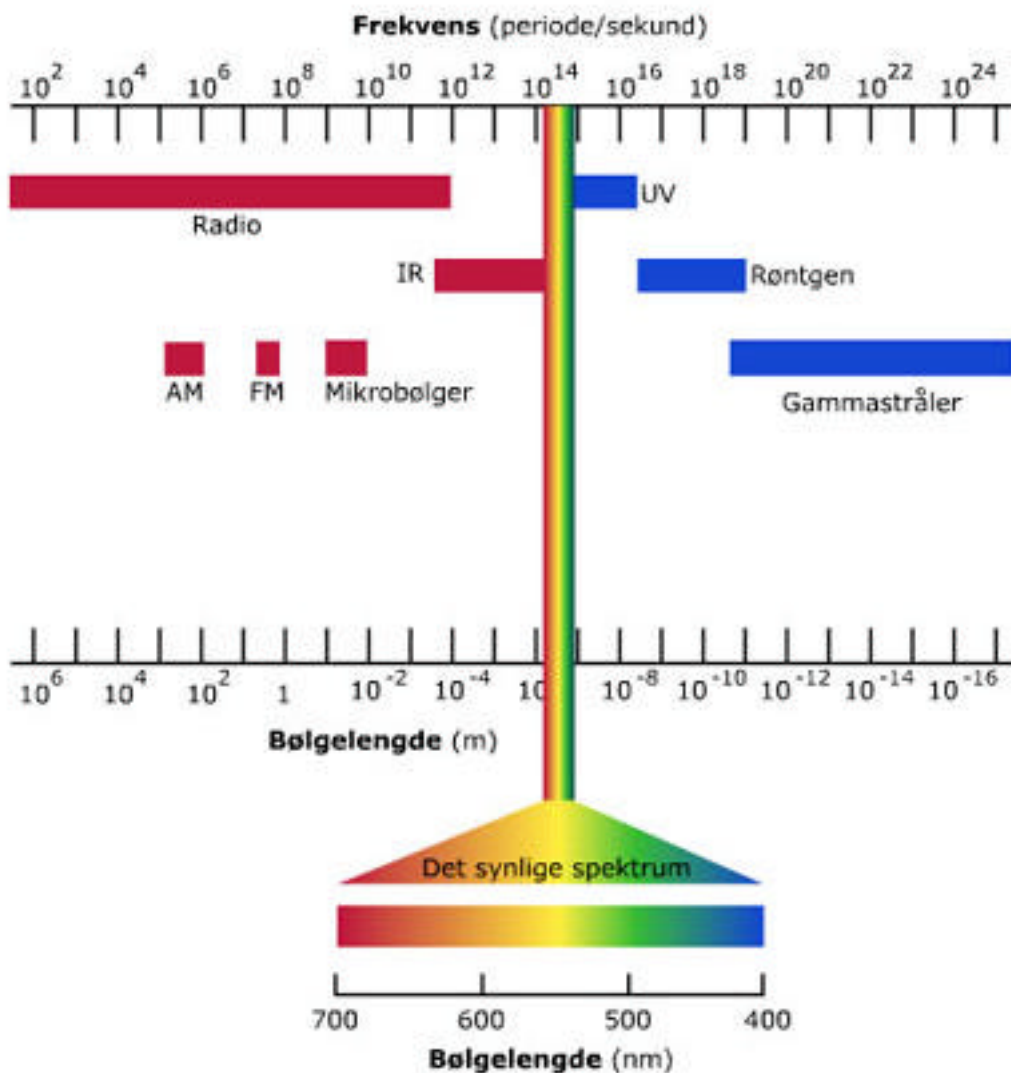
I denne ligningen er E er energien i ett foton (J), h er Plancks konstant ($6,63 \cdot 10^{-34} \text{ J} \cdot \text{s}$), c er lysfarten i vakuum (299 792 458 m/s), λ er bølgelengden (m) og f er frekvensen (s^{-1}) til strålingen. Denne ligningen viser at jo kortere bølgelengde (eller høyere frekvens), desto mer energirik er den elektromagnetiske strålingen. Man deler den elektromagnetiske strålingen inn etter minkende bølgelengde (eller økende frekvens) (se figur 5): radar- og radiobølger, mikrobølger, infrarød stråling, synlig lys, ultrafiolett stråling, røntgenstråling og gammastråling. De forskjellige strålingstypene blir skapt på forskjellige måter, men de er alle helt frie fra kilden de kommer fra. Lange radiobølger blir for eksempel skapt som følge av elektriske svingninger i store antenner, synlig lys kommer av bevegelse mellom elektrontilstander i atomet, mens gammastråling kommer av endringer i strukturen til atomkjernen.

Maxwells ligninger beskriver elektromagnetismen som interaksjonen mellom oscillerende (varierende) elektriske og magnetiske felt i bevegelse. Begge disse feltene står vinkelrett på hverandre, samt på retningen til den elektromagnetiske strålingen. I de spektroskopiske metodene er det vanligvis enten den magnetiske eller den elektriske komponenten av strålingen som deltar aktivt i analyseprosessen.



Figur 4: Elektromagnetisk bølge som brer seg i positiv x-retning med hastighet lik lysets [20]

Elektromagnetisk stråling innehar både partikkel- og bølgenatur, noe som har skapt hodebry for mange vitenskapsmenn opp igjennom tidene. Partikkelnaturen ligger i at strålingen består av fotoner, som fungerer som små energipakker, mens bølgenaturen kommer av at den elektromagnetiske strålingen brer seg med lyshastigheten i vakuum. Lyshastigheten er en naturlig fartsbarriere som det, i følge kvantefysikken, er umulig å overskride, og den blir redusert når strålingen går igjennom faste stoffer og væsker. Et eksempel på dette er diamant hvor lyshastigheten er 125 000 000 m/s, altså mer enn halvert i forhold til den ideelle hastigheten for lyset. Det ligger også i bølgenaturen at elektromagnetisk stråling kan reflekteres ved hjelp av metalloverflater, at det kan oppstå interferens mellom to koherente elektromagnetiske bølger (bølger med samme frekvens og konstant eller ingen faseforskjell), samt at stråling med ulik bølgelengde spres forskjellig når det treffer på atomer eller molekyler. Sistnevnte er grunnen til at himmelen ser blå ut. Sola sender ut lys med en rekke farger, men siden blått lys har kortere bølgelengde enn alle andre farger, vil det blå lyset spres mye mer effektivt. Dette gjør at himmelen (fra bakken) ser blå ut, selv om himmelrommet egentlig er helt svart, med unntak av lyset fra sola, stjerner og lignende.



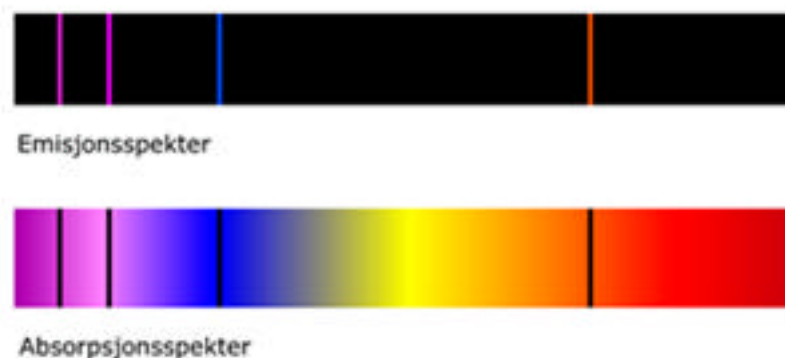
Figur 5: Det elektromagnetiske spektrum [22]

6 Spektroskopiske metoder

6.1 Spektralanalyse [22]

Spektralanalysen er den eldste av de spektroskopiske metodene, og det var oppdagelsen av de mørke linjene i solspekteret som dannet grunnlaget for denne teknikken. I dag blir spektralanalysen først og fremst brukt innen kjemi, fysikk og astronomi, men metoden har også blitt et viktig instrument for lydteknikere. Alle grunnstoffer vil under gitte omstendigheter sende ut et bestemt spekter, og siden hvert grunnstoff har forskjellige spektre kan man finne ut hva materialet som undersøkes består av ved å undersøke lyset materialet sender ut. Et spektralapparat inneholder et prisme eller et gitter som deler opp lysstrålen fordelt på bølgelengdene lyset består av, og dette gir et fargebånd eller spektrum hvor fargene kommer i samme rekkefølge som i en regnbue. Slike spektre blir vanligvis delt inn i tre grupper, henholdsvis kontinuerlige, emisjons- og absorpsjonsspektre. Kontinuerlige spektre inneholder alle bølgelengder, og er kanskje det viktigste spekteret i denne sammenhengen.

Emisjonsspektra består av bølgelengdene til lyset som kommer direkte fra lyskilden, og viser dermed hvilke bølgelengder som blir emittert (sendt ut) av materialet lyset ble sendt igjennom. En lysende gass har vanligvis et linjespektrum som emisjonsspektrum, mens et glødende, fast stoff har et kontinuerlig emisjonsspektrum. Dette betyr at glødende, faste stoffer sender ut lys med alle bølgelengder, mens lysende gasser bare sender ut lys med bestemte bølgelengder. Det motsatte av emisjonsspektrum er absorpsjonsspektrum som i stedet viser hvilke bølgelengder som er blitt absorbert (tatt opp) av materialet. En lysende gass har et absorpsjonsspektrum som ser ut som et kontinuerlig spektrum med noen mørke linjer, og disse linjene tilsvarer da linjene i emisjonsspekteret til gassen. Siden vi bare kan se lys med en bølgelengde mellom 400 og 700 nm kan vi bare se de mørke linjene i absorpsjonsspekteret som ligger mellom disse verdiene, men grunnstoffer kan også absorbere bølgelengder som ligger utenfor det synlige området, og disse trenger vi da egne analyseapparater til å bestemme. Sammenligner man spekteret til en ukjent gass med spektrene til kjente grunnstoffer målt i et laboratorium kan man med 100 % sikkerhet vite hvilke grunnstoffer den ukjente gassen består av, og siden ioner og molekyler også har karakteristiske spektra kan denne metoden også benyttes på disse.



Figur 6: Emisjons- og absorpsjonsspektra til hydrogen i den synlige delen av det elektromagnetiske spekteret [22]

6.2 Massespektroskopi [1, 23]

Massespektrometri er den mest nøyaktige metoden for å bestemme atom- og molekylærmasse til et stoff, og massespektrometre er svært anvendelige og blir i dag også brukt i karbondatering, radioaktive dateringsprosesser og undersøkelse av spor etter forurensing. Satellitter og romskip er også vanligvis utstyrt med massespektrometre for å kunne identifisere partikler i verdensrommet.

Hovedprinsippet til alle massespektrometrene er at de på en eller annen måte separerer ionene i partikkelstrømmen basert på forholdet mellom ladningen og massen. Ionene treffer så en detektor hvor de blir registrert, og spekteret som dannes kalles da et massespektrum. Man deler vanligvis massespektrometre inn i statiske og dynamiske. I statiske massespektrometre blir et stoff i gasstilstand bombardert med en elektronstråle som har tilstrekkelig energi til å dele opp partiklene (atomer eller molekyler). Kollisjonene mellom elektronene og partiklene produserer positive ioner ved å fjerne et elektron fra hver partikkel, og disse positive ionene (masse m , ladning q) blir så akselerert av to parallelle og motsatt ladde plater, altså et elektrisk felt. Hvis potensialforskjellen mellom de to platene er U , vil ionene forlate det elektriske feltet med en kinetisk energi gitt ved:

$$K = q \cdot U = \frac{1}{2}mv^2 \quad (2)$$

For å sørge for at alle ionene som ankommer magnetfeltet hvor massesepareringen foregår har samme hastighet (v), blir de sendt gjennom et område med et uniformt elektrisk og magnetisk felt. Det elektriske feltet (E) lages på samme måte som i akselereringsfasen og står vinkelrett på farten til ionene, mens det magnetiske feltet (B_0) står vinkelrett på farten til ionene og det elektriske feltet. Kraften det elektriske feltet utøver på ionene (qE) virker i motsatt retning av kraften fra det magnetiske feltet (qvB_0), og disse balanseres når:

$$qE = qvB_0 \quad (3)$$

Ionene med for stor hastighet i forhold til den foretrukne hastigheten vil bli separert fra ionestrømmen fordi den magnetiske kraften overskrider den elektriske, og tilsvarende vil de med for lav hastighet også separeres fra strømmen, siden den elektriske kraften er sterkere enn den magnetiske. Dermed vil alle ionene ankomme separasjonsfasen med en hastighet lik:

$$v = \frac{E}{B_0} \quad (4)$$

Når ionene ankommer magnetfeltet (B) har de samme hastighet og ladning, og siden dette feltet står vinkelrett på farten vil ionene bli avbøyd i en sirkulær bane, før de treffer en detektorskjerm. Kraften fra magnetfeltet (qvB) er da lik sentripetalakselerasjonen (mv^2/r), og radiusen (r) til banen ionene følger blir da følgende:

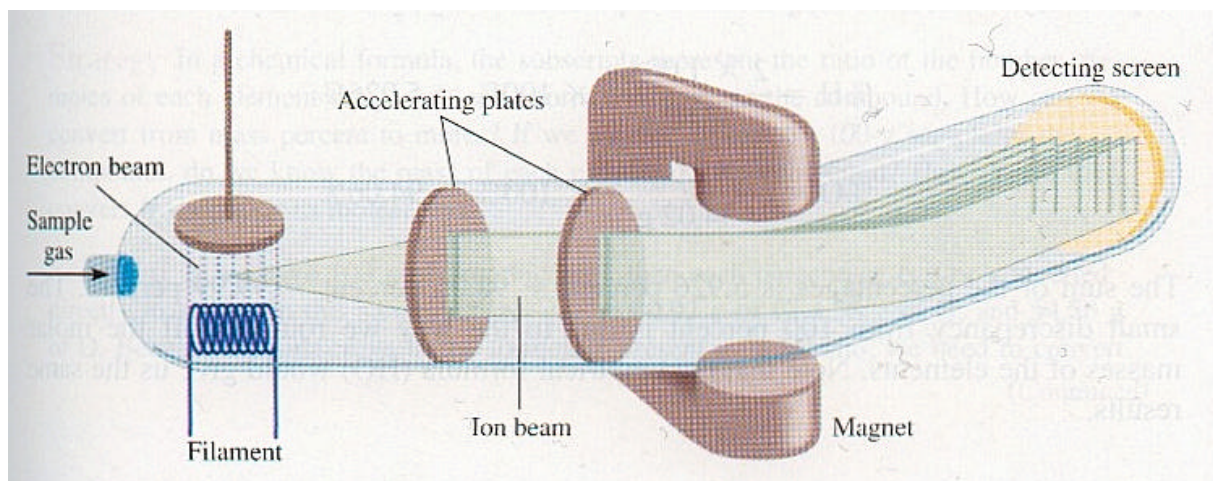
$$r = \frac{mv}{qB} \quad (5)$$

Siden radiusen til denne banen er avhengig av forholdet mellom massen og ladningen til partikkelen (altså q/m), vil partikler med liten q/m følge en stor kurve mens partikler med stor q/m følger en liten kurve. Hvis ionestrømmen består av partikler med samme ladning kan massen til hvert ion bestemmes ved at detektorskjermen registrerer strømmen hver type ion skaper. Denne er da direkte proporsjonal med antall ioner, og dermed kan den relative mengden av de forskjellige isotopene bestemmes. Siden hastigheten til ionene er gitt ved ligning 4, kan radiusen også skrives slik:

$$r = \frac{mE}{qBB_0} \quad (6)$$

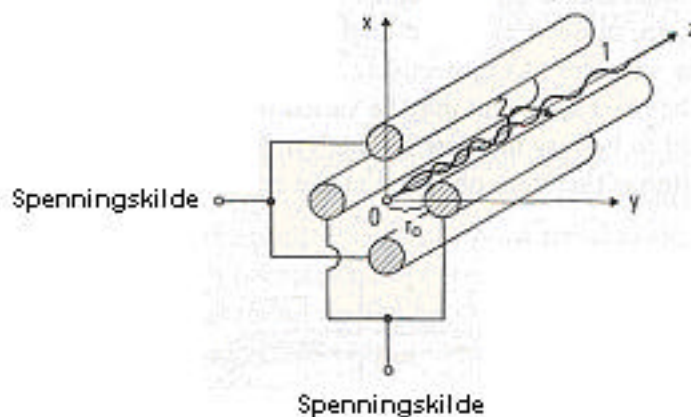
Sammenhengen mellom potensialet til det akselererende, elektriske feltet og den kinetiske energien i ligning 2 gir også et tredje uttrykk for radiusen.

$$r = \frac{1}{B} \sqrt{\frac{2mU}{q}} \quad (7)$$



Figur 7: Et statisk massespektrometer (uten hastighetsseparator) [13]

I dynamiske massespektrometre blir masseseparasjonen gjort ved at ioner blir oscillert av et elektrisk felt generert av fire sirkulære elektroder og to spenningskilder bestående av en direkte og en høyfrekvent, varierende spenning. Ionene ankommer feltet i den ene enden av elektrodeoppsettet, og bare de ionene som har en stabil oscillering, det vil si at de har en begrenset amplitude i begge retninger, vil nå detektoren i den andre enden av elektrodesystemet. Når radiusen til systemet og den direkte og den høyfrekvente spenningen er bestemt, vil stabiliteten til de oscillerende ionene bare være avhengig av forholdet mellom masse og ladning. Dette betyr at dersom alle ionene har samme ladning, er stabiliteten bare avhengig av massen til ionene, som så kan bestemmes idet ionene treffer detektoren (se figur 8, bane 1). Ioner som ikke har en stabil oscillering vil ikke nå detektoren, men i stedet treffe elektrodene og bli eliminert (se figur 8, bane 2). Denne typen massespektrometer kalles dynamiske fordi variasjon feltparameterne (direkte spenning, høyfrekvent spenning eller radius) er helt essensielt for å få til en masseseparasjon. Slike massespektrometre kan, i motsetning til statiske massespektrometre, ikke brukes til å analysere molekyler fordi sammensetningen til et molekyl også er avhengig av bindingsenergien til atomene det består av, og ikke bare massetallet.



Figur 8: Dynamisk massespektrometer [1]

6.3 Infrarød spektroskopi [2-3, 24-27]

Infrarød spektroskopi er en mye anvendt metode innen forskning og industri. Metoden blir mest brukt til å finne ut noe om vibrasjonstilstandene til et stoff, men også for å finne ut hvilke stoffer en prøve består av eller hvor rent et stoff er. Selv enkle organiske molekyler kan gi komplekse spekter. Molekyler blir påvirket av omgivelsene sine. I en ideell tilstand er molekylet alene i verden, men i realiteten befinner molekylet seg gjerne i en blanding av et eller annet slag, noe som gjør at vi får et lite avvik i bindingsforholdet. Selv en liten endring i disse bindingsforholdene, vil utgjøre en forandring i molekylvibrasjonene. Vi har to forskjellige typer molekylære vibrasjoner, strekking og bøying. En vibrasjon må medføre en forandring i dipolmomentet for at et molekyl skal kunne bli oppdaget med infrarødstråling. Det er dette som gjør at for eksempel karbondioksid (lineært molekyl), som har ett symmetrisk strekk, ikke alltid vil vise så mange bånd på et infrarødt spekter som man ville anta, siden et symmetrisk molekyl ikke vil medføre noe endring i dipolmomentet. Med andre ord kan vi si at en betingelse for eksitasjon av infrarød stråling er molekylær asymmetri.

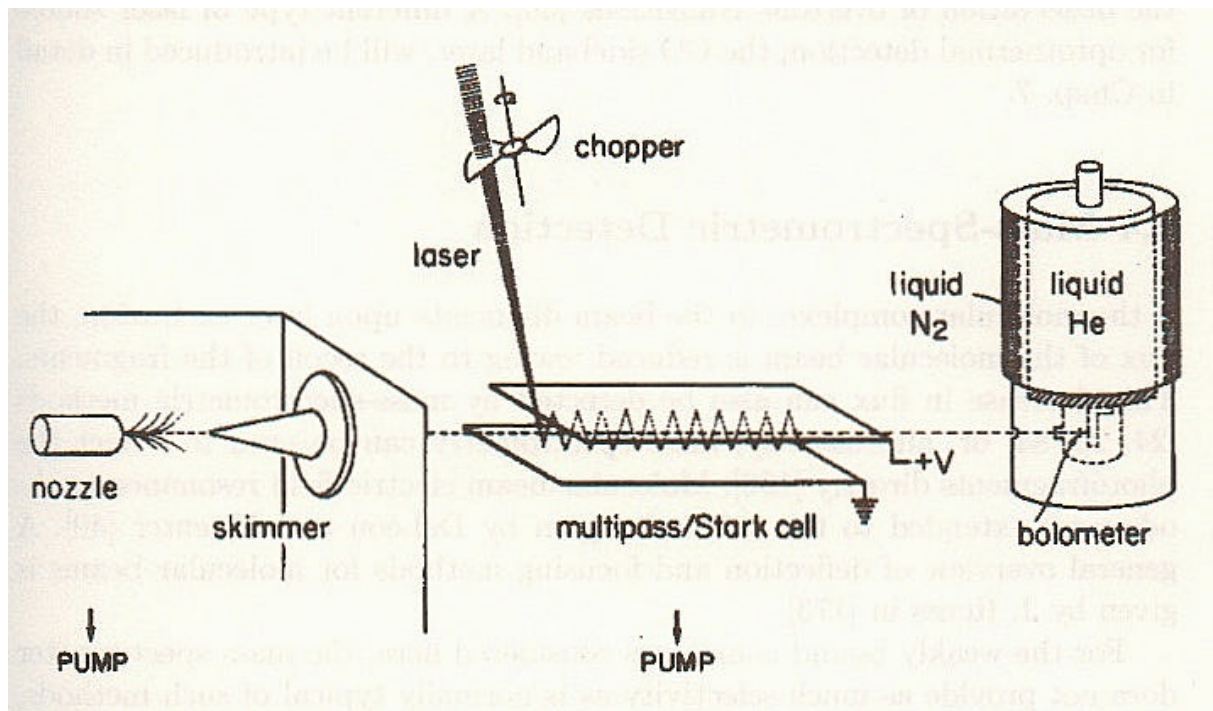
Energien som blir absorbert ved en bestemt frekvens, korresponderer til frekvensen av en vibrasjonsbevegelse i en binding. Vi kan finne ut hvilke typer bindinger en forbindelse inneholder, og dermed identifisere de funksjonelle gruppene, ved å måle frekvensen der det absorberes infrarød stråling. Vi kan analysere det infrarøde spekteret ved å sammenligne det med spektra av kjente stoff, og til denne formen for identifikasjon ser man på området der man har svært mange komplekse vibrasjoner. Dette området refereres ofte til som fingeravtrykkregionen. Flertallet av de funksjonelle gruppene har karakteristiske infrarøde absorpsjoner som er like i forskjellige molekyler. Absorpsjonsfrekvensen (f) er avhengig av to ting, styrken på bindingen og massen av atomene i bindingen. Bølgetallet ($\bar{\nu}$) øker dess sterkere bindingen er, og minker ved økende masse. Dette tallet finner vi ved å ta den inverse av bølgelengden til absorpsjonsfrekvensen (λ), og vi har følgende sammenheng, der c er lysfarten og f er frekvensen:

$$l = \frac{c}{f} = \frac{1}{\bar{\nu}} \quad (8)$$

Infrarød spektroskopi utføres ved å sende mange infrarøde lysbølger mot en prøve på samme tid. Endringene i lysbølgene, etter at de har blitt reflektert av eller trengt gjennom prøven, blir så målt. Deretter gjennomfører man en Fouriertransformasjon, som er en omregning fra bølgelengdeområdet til frekvensområdet. For infrarøde studier av molekylære forbindelser, benytter man seg i dag spesielt av to metoder: direkte absorpsjon spektroskopi og optotermisk bestemmelse. I direkte absorpsjon spektroskopi har vi at en laserstråle passerer gjennom molekylstrålen, og en kan da måle absorpsjonen som funksjon av frekvensen. I dag finner vi som oftest et dobbelstrålespektrometer i laboratorier, og dette instrumentet benytter stråling i det midt- og nærinfrarøde området. Optotermisk bestemmelse benytter seg derimot av termiske detektorer, ofte et bolometer. Bolometeret har en elektrisk motstand som er sterkt avhengig av temperaturen, og det kan oppdage temperaturforandringer helt ned til $3 \cdot 10^{-8}$ K med stor nøyaktighet. Senker man temperaturen ned mot det absolutte nullpunkt (0 K), øker nøyaktigheten ytterligere. Energien fra den infrarøde

strålingen sørger for en temperaturforandring, og denne kan registreres ved at man sender strålingen mot bolometeret.

Figuren under viser hvordan et optotermisk spektroskop fungerer. Trykkforskjell mellom kammeret molekylene er plassert i og kollisjonskammeret gjør at molekylene trekker inn i kollisjonskammeret, og der blir molekylstrålen eksitert ved hjelp av en infrarød laserstråle. Molekylstrålen gjennomgår så en rekke kollisjoner med laserstrålen, og dette gir en temperaturforandring (energiforandring), som tilslutt blir registrert av bolometeret.



Figur 9: Oppsett for infrarød optotermisk spektroskopi [3]

Infrarøde frekvenser er for høye til at de kan digitaliseres, derfor må frekvensene flyttet til et frekvensområde man kan arbeide med. En slik modulering finnes for eksempel i radioer. Denne flyttingen gjøres med et interferrometer, siden det ikke er mulig å foreta denne prosessen elektronisk med så høye frekvenser. I et slikt instrument blir det infrarøde lyset kastet tilbake på seg selv ved hjelp av to speil, og dette fører til at lysbølger enten blir ødelagt eller forsterket. Vi får med andre ord det en kaller destruktiv (ødeleggende) eller konstruktiv (forsterkende) interferens.

Forskere har også funnet ut at infrarød spektroskopi kan benyttes for å bestemme hvor mye vann tørrfisk inneholder. Dette gjøres ved at man sender en stråle med infrarødt lys ned mot fisken. Fisken lyses dermed opp der strålen treffer, fordi lyset forplanter seg i vevet som om det skulle være en bunt med optisk fiber. Av lyset som kommer ut av fisken blir det gjennomført en spektralanalyse som kan fortelle oss hvor mye vann som er til stede.

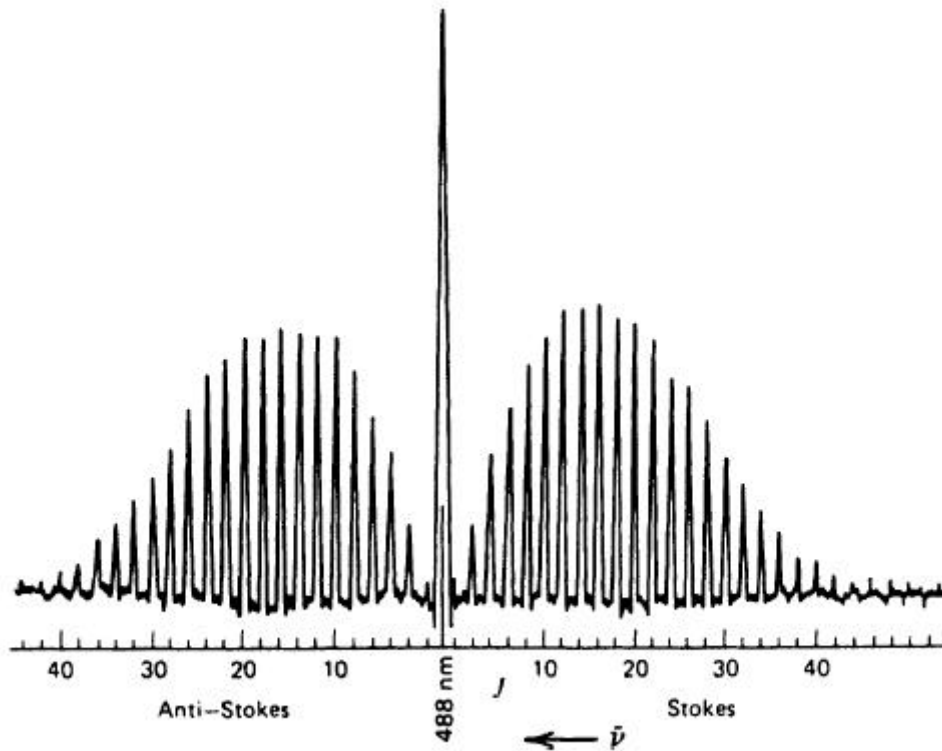
6.4 Ramanspektroskopi [4-6, 27-31]

Ramanspektroskopi bygger på prinsippet om molekylær spredningseffekt, som tilsvarer uelastisk støt mellom kuler (energien er ikke bevart). Når materie blir utsatt for intens stråling med en bestemt frekvens utenfor området det klarer å absorbere, vil en del av stråling blir spredd ut i alle retninger, og dette fenomenet kalles raman-effekten. Effekten ble oppdaget i 1928 av den indiske fysikeren Chandrasekhara Venkata Raman (1888-1970), og kan beskrives som forekomsten av ekstra linjer i spekteret til monokromatisk lys (lys bestående av bare en bølgelengde), for eksempel fra en laser. I kjemi brukes ramanspektroskopi til å studere molekylvibrasjon, rotasjon, endringer i bindingsenergi og til identifisering av molekyler, mens i fysikk brukes metoden til å studere egenskapene til kondenserte stoffer, altså faste stoffer og væsker hvor vekselvirkningene mellom partiklene er sterke. Metoden er en komplementærmetode til den infrarøde spektroskopien, det vil si at de to teknikkene utfyller hverandre, og de brukes ofte sammen for å studere vibrasjonen i molekyler.

Når et molekylært system blir utsatt for stråling skjer det to ting: mesteparten av energien (fotoner med samme frekvens som systemet) blir absorbert av systemet, men en liten del av strålingen klarer ikke systemet å absorbere, og denne blir spredd ut i alle retninger med en frekvens som vanligvis er lavere enn frekvensen til den innkommende strålingen. Linjespekteret til det spredde lyset vil ha en framstående linje som svarer til den opprinnelige bølgelengden til strålingen, samt ekstra linjer på hver side, som da svarer til bølgelengdene til det utsendte lyset. Forskjellen i energi mellom de innkommende og de utsendte fotonene er lik vibrasjonsenergien til det molekylære systemet. Ramanspekteret framkommer ved å sette opp denne energidifferansen som funksjon av intensiteten til den utsendte strålingen. Siden den absorberte strålingsenergien er mye høyere enn den utsendte består oppgaven til et ramanspektrometeret i å separere disse fra hverandre. Dette gjøres ved at bølgelengdene som fører til absorpsjon blir filtrert ut ved hjelp av en monokromator. Ramanspredningen kan så registreres med en detektorskjerm, og hele prosessen skjer uten at prøven som undersøkes blir ødelagt. Når en prøve skal undersøkes med et ramanspektrometer, trenger man heller ikke gjøre noe forbehandling av prøven.

For å forstå hva som egentlig skjer i denne prosessen må vi ned på molekylnivå. Et molekyl består av atomer som igjen består av en kjerne og elektroner. Elektronene svever rundt kjernen, og de er i stand til å ta opp energi som er nøyaktig lik energiforskjellen mellom to (eventuelt flere) energitilstander. Som vi tidligere har sett består elektromagnetisk stråling av fotoner med bestemte energier, og sender man slik stråling mot et molekyl har vi tre mulige utfall. Dersom energien til fotonet som kommer inn er lik energien til fotonet som sendes ut, blir energien først absorbert og elektronet går over i en høyere energitilstand. Men siden elektroner søker lavest mulig energi, vil det raskt falle ned til nivået det var på, og da sendes det ut et foton med like stor energi som det innkommende fotonet. Dette kalles Rayleigh spredning (eller elastisk spredning) og skjer når den elektromagnetiske strålingen er i resonans med overgangsenergien til elektronet. Hvis den elektromagnetiske strålingen ikke er i resonans med overgangsenergien, og fotonenergien til den innkommende strålingen er større enn den utsendte, betyr det at en del av energien blir brukt til å endre vibrasjons-, rotasjons- eller den elektriske energien til molekylet og resten av

energien sendes ut som et foton. Denne strålingen kalles Stokes lys og er en form for ramanspredning. I motsatt tilfelle, altså når energien til den innkommende strålingen er mindre enn energien til den utsendte, endres også vibrasjons- rotasjons- eller den elektriske energien til molekylet. Det sender da ut den innkommende energien samt forskjellen mellom energitilstandene som et foton, og denne typen ramanspredning kalles anti-Stokes lys. Vi ser dermed at ramanspektroskopi ikke er avhengig av at den elektromagnetiske strålingen som brukes er i resonans med molekylets energinivåer, slik som flere av de andre spektroskopiske metodene.



Figur 10: Ramanspektrum av CO₂ [30]

6.5 UV/Synlig spektroskopi [7, 32-33]

UV/Synlig spektroskopi er en metode som fortrinnsvis brukes i analytisk kjemi, hvor formålet er å finne konsentrasjonen til en løsning. Når stråling vekselvirker med molekyler kan det skje flere prosesser: absorpsjon, refleksjon, spredning, fluorescens/fosforescens og fotokjemiske reaksjoner (absorpsjon som fører til bryting av kjemiske bindinger). Fluorescens er når atomet stråler ut synlig lys som følge av at atomet blir eksitert av en type kortbølget stråling, og effekten opphører når den eksiterende strålingen opphører. Dersom fluorescensen fortsetter en liten stund etter at den eksiterende strålingen opphører, kaller vi det fosforescens. I UV/Synlig spektroskopi er det vanligvis absorpsjonen vi ønsker å studere. Prøven fylles i en kuvette (en liten glassbeholder) og denne plasseres så i et spektrometer. Inne i spektrometeret blir lys, i dette tilfellet UV og synlig lys, sendt gjennom prøven og mengden absorbert stråling blir målt. Forskjellige stoff vil da absorbere lys med ulike bølgelengder, og denne absorbansen er proporsjonal med konsentrasjonen til stoffet.

Elektromagnetisk stråling har en energi som avhenger av bølgelengden (eller frekvensen), og når et molekyl absorberer stråling øker energien til molekylet. For visse atomer og molekyler har fotoner med bølgelengder i UV og det synlige området av det elektromagnetiske spekteret tilstrekkelig energi til å medføre overganger mellom ulike elektrontilstander. Bølgelengden som absorberes har akkurat den energien som kreves for å flytte et elektron fra en lavere elektrontilstand til et høyere nivå.

6.6 Mikrobølgespektroskopi [27, 34-36]

Mikrobølger er elektromagnetiske bølger med en bølgelengde mellom 1 mm og 100 mm, eller med andre ord lengre enn infrarødt lys og kortere enn radiobølger, og disse danner grunnlaget for mikrobølgespektroskopien. Denne teknikken brukes hovedsakelig til å undersøke molekylstruktur i kjemien, men den kan også brukes til å identifisere molekyler i verdensrommet. Utgangspunktet for mikrobølgespektroskopien er at forandringer i et molekyls rotasjonsenergi gir et spektrum som består av skarpe absorpsjons- eller emisjonslinjer. Bølgelengden til disse linjene kan bestemmes med stor nøyaktighet, og dette gjør det igjen mulig å finne hovedtreghetsmomentene til molekylet med like stor nøyaktighet. Når treghetsmomentene er bestemt kan denne informasjonen brukes til å finne den geometriske strukturen til molekylet, noe som er et av de viktigste forskningsområdene innen kjemi. Man kan også finne dipolmomenter og rotasjonsbarrierer ved hjelp av denne teknikken.

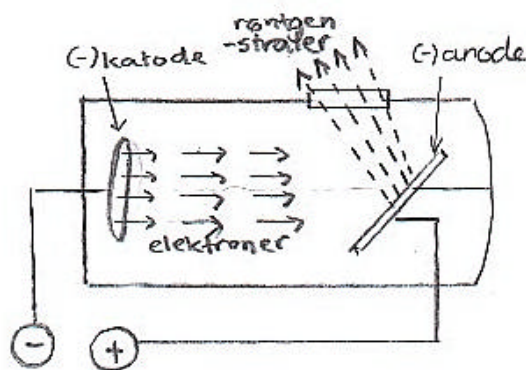
Metoden kan også brukes til å overvåke kjemiske reaksjoner ved at stoffene som skal reagere blir plassert i en beholder med en mikrobølgesender på den ene siden, og en mottaker på den andre siden. Beholderen må være av et materiale som tillater mikrobølgene å passere, eller ha vinduer som strålingen kan gå igjennom. Mikrobølgene kan da stråle ut fra senderen, gjennom prøven og registreres av mottakeren, noe som i neste omgang kan gi oss et bilde av reaksjonsprosessen.

6.7 Røntgenspektroskopi [8-9, 37-40]

Røntgenstråling er elektromagnetisk stråling med kort bølgelengde, kortere enn for eksempel synlig lys. Strålingen oppstår når elektron med stor energi treffer et metall, og den kjennetegnes ved at den er gjennomtrengbar, ioniserende, og den avbøyes ikke av elektriske eller magnetiske felt. Strålingen oppstår ved oppbremsing av elektroner, eller ved kvantesprang av elektroner i atomer. Røntgenspekteret består av to deler, det kontinuerlige spekteret og linjespekteret. Det er det sistnevnte som er karakteristisk for røntgenstrålingen. Linjespekteret i røntgenstrålingen ligner på linjespektra til hydrogen og andre gasser. Egenskapene til røntgenstråling gjør at den blir brukt i mange sammenhenger, og for folk flest er nok medisin og tannmedisin mest kjent. Strålingen kan også brukes i behandlingen av kreft, men dessverre kan røntgenstrålene også skade de friske cellene, og ikke bare kreftcellene. I industrien blir den brukt til å finne feil på faste stoff ved røntgendiffraksjon. Molekylstrukturen til DNA ble også fastlagt ved hjelp av røntgenstråling.

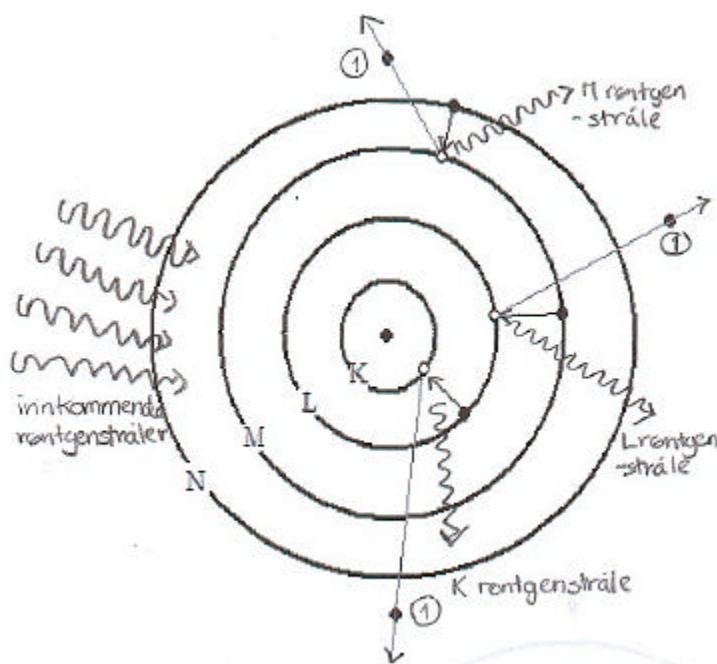
I røntgen absorpsjon spektroskopi blir et elektron eksitert til et ledig elektronnivå ved fotonabsorpsjon. Om en da måler absorpsjonskreftene som en funksjon av fotonenergien, kan en bestemme terskelenergien for en eksitasjon fra et bestemt kjernenivå til den lavest ledige tilstanden. En vanlig metode for å utføre røntgen absorpsjon spektroskopi, er å se på elektronutbyttet som en funksjon av fotonenergien. Ved røntgen fotoelektron spektroskopi finner en størrelsen til den kinetiske energien til elektronene ved hjelp av en fiksert elektronkilde som eksiterer elektron fra prøven. Dette gjør en i stand til å bestemme hvilke stoff en har i prøven, siden den kinetiske energien avhenger av bindingsenergien.

Røntgen fluorescens spektroskopi er en av de vanligste spektroskopi metodene, og den kan beskrives som en ikke nedbrytende analytisk metode som blir brukt for å finne og fastlegge konsentrasjonene til grunnstoffene (fra det 4. til det 92. atomet i periodetabellen) i et stoff, pulver eller væsker. En kan få et røntgen spekter fra den eksiterte prøven dersom den blir utsatt for stråler med tilstrekkelig kort bølgelengde. Denne metoden lønner seg særlig med høyintensitets røntgenrør, følsomme detektorer og en passende røntgen optikk. Kort oppsummert går denne metoden ut på at man sender røntgenstråler (med en energi på flere elektronvolt) mot prøven. Når strålingen sendes mot prøven, får vi en vekselvirkning mellom elektronene, og dette gjør at prøven sender ut stråling. Vi kan så registrere og analysere strålene som blir utsendt, og dette kan gi oss informasjon om hva prøven inneholder.



Figur 11: Skjematisk oppstilling av et røntgenrør

Figuren over viser en skisse av et vanlig røntgenrør. Røntgenstrålene blir dannet ved at en frigjør elektroner og akselerer dem mot en metallskive (anoden) med en energi på mange tusen elektronvolt. Når disse elektronene treffer metallskiven, får vi dannet røntgenstråling (stiplet linje).



Figur 12: Innkommende røntgenstråler slår ut elektroner, og energiforskjellen mellom det utslåtte elektronet og det som erstatter det, sendes ut som røntgenstråling.

Denne figuren skal prøve å illustrere hvordan røntgenstrålene slår ut elektroner i underskallene og danner tomrom. Dette er ikke en ønskelig tilstand for atomet, og resten av elektronene vil dermed prøve å kompensere for det utslåtte elektronet. Elektroner fra skall utenfor vil dermed fylle opp tomrommet, for at atomet skal komme tilbake til sin stabile tilstand. Siden disse elektronene ikke vil være bundet til atomet med like mye energi som elektronet som ble slått ut, vil energiforskjellen (altså forskjellen i bindingsenergien mellom skallene) sendes ut som røntgenstråler. Denne strålingen er ulik for forskjellige stoff.

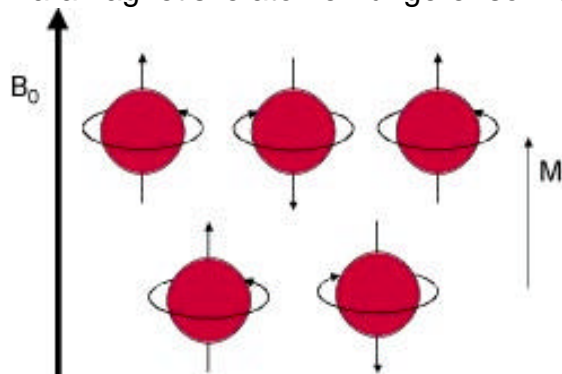
6.8 NMR [10-11, 41-43]

En av de mest kjente spektroskopiske metodene heter NMR, og denne forkortelsen står for "Nuclear Magnetic Resonance", som på norsk blir kjernemagnetisk resonans. Utviklingen av NMR til bruk i analyse begynte på starten på 1950-tallet, og i dag blir metoden i hovedsak brukt i studiet av molekylstruktur og –sammensetning i kjemi og for å gjøre avbildninger av kroppens organer i medisin, såkalt magnetresonanstomografi eller MR. En MR-maskin består av en stor magnet formet som en ring, flere radiosendere og –mottakere og datautstyr for å tolke signalene. I et sterkt magnetfelt vil ulike kroppsvev magnetiseres i forskjellig grad, og dermed sende ut forskjellig stråling når de påvirkes av radiobølger. Denne informasjonen blir tolket av datamaskinen som så kan gi detaljerte tredimensjonale bilder av kroppens indre organer og blodflyt. Ved å injisere kontrastmidler i kroppen kan organer med unormal blodsirkulasjon, som svulster, trange blodårer eller hjerneskaner, identifiseres siden kontrastmidlene i hovedsak vil følge blodet rundt i kroppen. Kontrastmidler inneholder paramagnetiske metall (for eksempel mangan) som øker magnetiseringsevnen til kroppsvevet, og dermed også radiostrålingen som blir sendt ut og registrert av MR-maskinen.

NMR baserer seg på at visse atomer er paramagnetiske, og de vil da innrette seg parallelt eller antiparallelt med et påtrykt magnetisk felt. Atomer er paramagnetiske hvis de inneholder netto uparede elektronspinn, eller sagt med andre ord: atomkjernen inneholder et ulikt antall protoner og/eller et ulikt antall nøytroner.

Paramagnetiske atomer fungerer som små magneter med en sør- og en nordpol, og

dette kommer av at kjernen med uparede spinn spinner rundt sin egen akse, og denne bevegelsen danner et lite magnetfelt. Det sterkeste paramagnetiske atomet er hydrogen, som bare består av et proton og et elektron. Siden forekomsten av hydrogen er stor i mange forskjellige materialer er bruken av den mest vanlig i NMR - spektroskopi, men ^{31}P (fosfor-31), ^{13}C (karbon-13), ^{23}Na (natrium-23) og ^{19}F (fluor-19) kan også brukes. Når paramagnetiske atomer ikke er utsatt for noen ytre påvirkning, har spinn en vilkårlig retning,



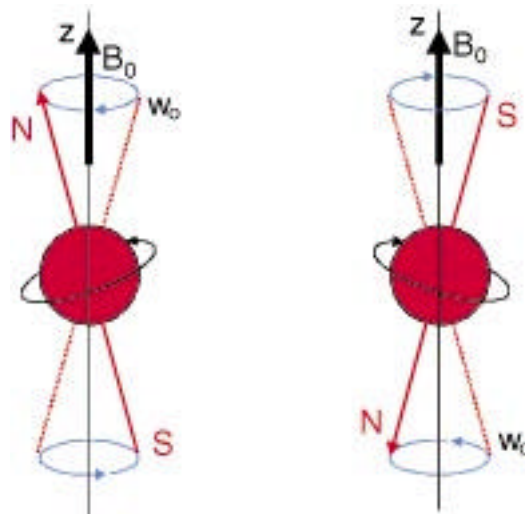
Figur 13: Innretting av paramagnetiske atomer i et påtrykt felt [43]

men dersom et sterkt magnetfelt blir skrudd på vil spinnnet innrette seg med eller mot det påtrykte feltet. De fleste vil imidlertid være parallelle med feltet siden den retningen gir det laveste energinivået, og dette er det foretrukne for atomer. Det påtrykte magnetfeltet fører også til at de parallelle og antiparallele atomspinnene blir tvunget til å rotere rundt det ytre feltets retning, en bevegelse som kalles precesjon. Denne bevegelsen skyldes mekaniske krefter mellom atomkjernens magnetiske moment og det ytre magnetfeltet, og rotasjonsfrekvensen til bevegelsen er gitt ved Larmors ligning:

$$\omega_0 = \gamma \cdot B_0 \quad (9)$$

Her er ω_0 lik rotasjonsfrekvensen (radianer pr sekund), γ er den gyromagnetiske konstanten og B_0 er styrken på det ytre feltet. Den gyromagnetiske konstanten er

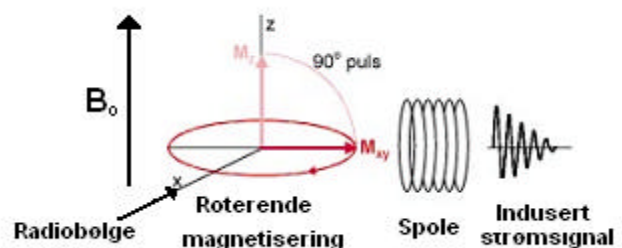
avhengig av hva slags atom man bruker i NMR-prosessen, men for det vanligste atomet, nemlig hydrogen, er denne konstanten lik 42,5781 MHz/T.



Figur 14: Presesjonsbevegelsen til atomspinn som er henholdsvis parallelt og antiparallelt med det ytre magnetfeltet [43]

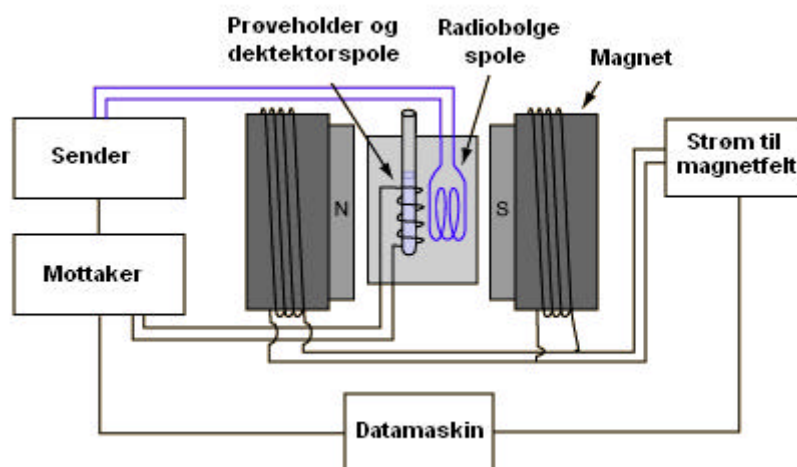
Innrettingen av atomspinn vil gi en nettomagnetisering som peker i samme retning som og er proporsjonalt med det ytre magnetfeltet. Denne netto magnetiseringen kan bestemmes ved å la den indusere en målbar strøm i en spole, og for at dette skal skje må magnetiseringen bevege seg i forhold til spolens åpning. Dette kan gjøres ved å sende elektromagnetisk stråling vinkelrett på magnetfeltet som atomspinnet lager. Radiostrålingen som benyttes har en energi som tilsvarer energiforskjellen mellom parallelt og antiparallelt atomspinn, og den sendes inn mot materialet som skal undersøkes i korte pulser.

Elektromagnetisk stråling består av en elektrisk og en magnetisk komponent som står vinkelrett på hverandre og på strålingens retning, og når atomspinn utsettes for denne strålingen vil det dreies vekk fra det påtrykte feltets retning. I NMR benyttes radiobølger med en flippvinkel på 90 grader (også kalt radiobølgens puls) til denne prosessen, og hvis man da plasserer spolens akse vinkelrett på det påtrykte feltet, vil den roterende magnetiseringen kunne indusere en strøm i spolen. Den induserte strømmen har da en styrke som er proporsjonal med styrken til netto magnetiseringen, og siden to atomer av samme grunnstoff vil oppleve et litt forskjellig magnetfelt når de befinner seg i ulike kjemiske miljø, kan denne sammenhengen blant annet brukes til å skille et hydrogenatom i et alkan (mettet hydrokarbon med enkeltbindinger mellom alle atomene, for eksempel CH_4 eller C_2H_6) fra et hydrogenatom i ei aromatisk gruppe (kjemisk forbindelse med en eller flere benzenringer, for eksempel C_6H_6).



Figur 15: Roterende magnetisering som induserer en strøm i spolen [43]

For å sørge for at denne spektroskopiske metoden gir gode spektre, kreves det et sterkt, stabilt og homogent magnetfelt, og til dette trengs en magnet. I NMR er det tre forskjellige magnettyper som er aktuelle, nemlig en permanent magnet, elektromagnet eller superledende magnet. De tidligste NMR spektrometrene brukte permanente magneter til å lage magnetfeltet, men dessverre er ikke slike magneter i stand til å lage magnetfelt på mer enn ca 1,4 T, noe som er for lite for dagens bruk. Elektromagneter er mye dyrere enn permanente magneter, både i innkjøp og bruk, og de krever også god avkjøling for å sikre et stabilt felt siden det trengs en meget sterk strøm for å generere magnetfeltet (noe som fører til høy varmeutvikling på grunn av elektrisk motstand). Slike magneter kan imidlertid lage magnetfelt med en styrke på inntil 34 T, noe som er det sterkeste som noen gang er oppnådd. Det mest vanlige i dagens NMR spektrometre er superledende magneter, som egentlig er en undergruppe av elektromagnetene. Hver superledende magnet er laget for å generere et felt med en bestemt styrke, fra 6 til 18 T, og et NMR spektrometer som bruker denne magnettypen består av tre skall der det midterste består av en superledende legering (niobtinn). Det superledende materialet har ingen elektrisk motstand når det avkjøles til 4 K, og for å oppnå denne temperaturen blir området mellom det ytterste og midterste skallet fylt med flytende nitrogen og området mellom det midterste og innerste fylt med flytende helium. Så lenge temperaturen holdes på 4 K er det ingen motstand i legeringen, og hvis man kobler legeringen til en sterk spenningskilde og så deretter kobler av spenningskilden vil strømmen fortsette å gå gjennom legeringen og generere et sterkt og stabilt magnetisk felt.

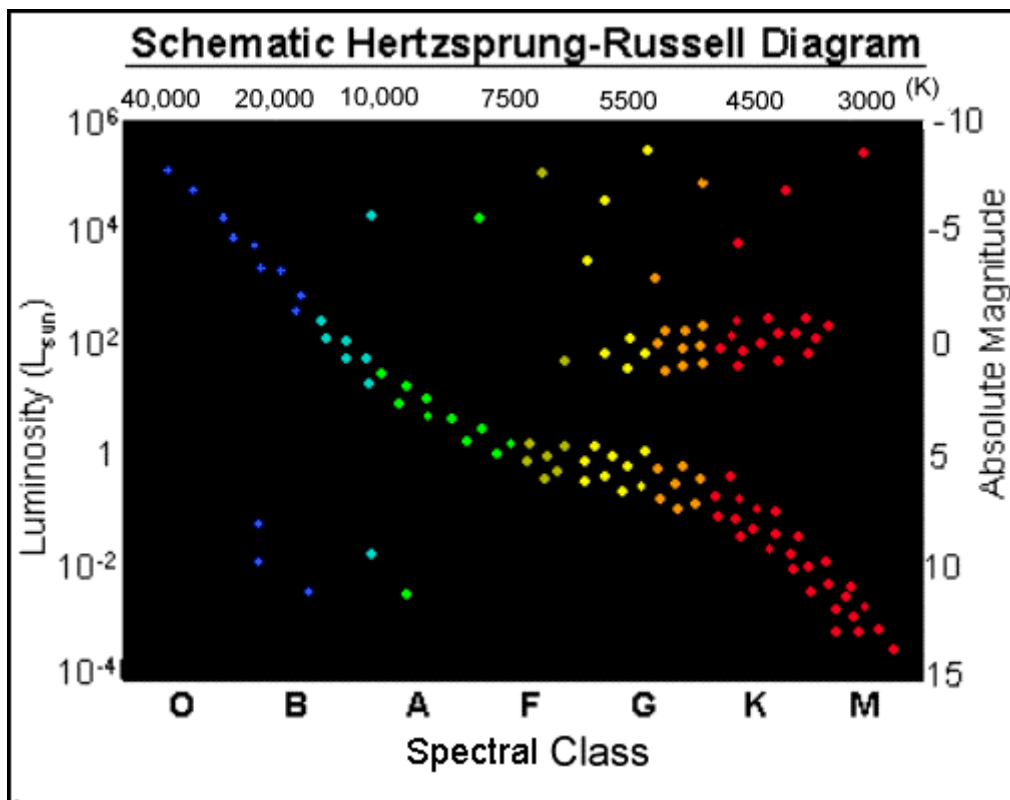


Figur 16: Enkel oppstilling for en NMR-maskin

7 Spektroskopi i praksis

7.1 Studiet av verdensrommet [12, 44-51]

Spektroskopi er en viktig forskningsmetode innenfor astronomien, og teknikken førte til oppdagelsene av flere nye grunnstoffer. Etter som spektroskopene ble forbedret, begynte man å klassifisere informasjonen disse observasjonene gav, og det ble da raskt et behov for å klassifisere observasjonene. William Huggins (1824-1910) var en av de første som begynte å sortere denne informasjonen, og han var med på å utvikle et enkelt, men samtidig viktig klassifikasjonssystem, som lenge var i bruk blant astronomer. I dag bruker man Harvard Classification (OBAFGKM), som er et system der man gir stjernene en bokstav etter forekomsten av spektrallinjer (som igjen sier noe om stjernens overflatetemperatur). Klassifiseringen går fra O til M, der de blå stjernene (O) er de varmeste og de røde stjernene (M) de kaldeste. Denne klassifiseringen lever videre i Hertzsprung-Russell diagrammet, hvor posisjonen til en stjerne er avhengig av stjernens masse og alder.



Figur 17: Et eksempel på HR-diagram [51]

Stjerner flest er består av 75 % hydrogen og 20 % helium (resterende 5 % er tyngre metaller). Helium blir til gjennom en fusjon av to hydrogenatomer, og denne kjernefysiske fusjonen er et resultat av den høye temperaturen. I varmere stjerner kan også tyngre atomer fusjonere. Vi kan bestemme sammensetningen til en stjerne ved å se på absorpsjonsspekteret til stjernen. Siden alle grunnstoffer har et eget absorpsjonsspekter kan man sammenligne stjernesppekteret med spektrene til kjente grunnstoffer, og dermed identifisere grunnstoffene stjernen består av. Det er blitt utviklet datamodeller til dette arbeidet, og disse ser både på temperaturen og

tettheten til stjernen, samt hvor sterk den observerte linjen er. Vi kan da få informasjon om mengden av de forskjellige grunnstoffene i stjernens atmosfære. Det er nettopp det faktum at stjernene har ulike temperaturer, som gjør at vi kan se forskjeller i stjernespektrene. I hver av ionisasjonstilstandene vil hvert kjemisk element i en stjerne ha en karakteristisk temperatur, der elementet mest effektivt produserer absorpsjonslinjer i sin del av spekteret. Stjernens temperatur og bevegelsesretning kan også bestemmes ved å se på absorpsjonsspekteret. Dersom bølgelengden til linjene er forskjøvet mot rødt (lengre bølgelengde), er stjernen på vei vekk fra oss, men har vi derimot en blåforskyvning, kommer den mot oss. Dette fenomenet er koblet opp mot dopplereffekten, som sier at bølgelengden endres når kilden og mottaker beveger seg i forhold til hverandre. Spektrallinjene vil også være bredere for roterende stjerner. Dette kommer av at stjernens nærmeste del vil representere den vanlige hastigheten til stjernen (i forhold til jorda), mens sidene vil ha hver sin hastighet, der den ene er litt større og den andre litt mindre enn den vanlige hastigheten. Denne forskjellen i hastighet gir de roterende stjernene bredere spektrallinjer.

I rommet mellom himmellegemene finnes det molekyler, ioner og radikaler i ørsmå konsentrasjoner. Noen steder i rommet finner vi opphopninger av molekyler, bedre kjent som tåker. Disse kan være flere lysår i diameter og har svært lavt trykk, men på grunn av sine enorme dimensjoner vil de kunne absorbere alt synlig lys fra stjernene bak tåken. Mikrobølger og infrarødt lys på sin side vil derimot ikke bli absorbert, og kan derfor slippe gjennom tåken og danne grunnlag for spektroskopiske observasjoner. Studier og observasjoner har vist at tåkene har en komplisert kjemi, og ved hjelp av mikrobølgespektroskopi har man funnet over 120 ulike molekyler i disse tåkene. Noen forskere hevder at de interstellare molekylene (interstellare betyr mellom stjernene) er opphavet til livet på jorden. Stjernesystemer med planeter, stjerner og kometer blir dannet ved at tåkene fortettes, og kometene har vist seg å inneholde de samme interstellare molekylene som tåkene. En kollisjon mellom jorden og en komet ville frigjøre tonnevis av en rekke ulike organiske molekyler, og en bedre egnet start på livsprosessen enn dette er det vanskelig å tenke seg.

7.2 Våre spektroskopiske opplevelser

Da vi begynte med dette prosjektet, var noe av det første vi foretok oss å prøve og få tak i noen som bruker spektroskopi innen forskning eller som en del av det daglige arbeidet. Takket være Jon Andreas Støvneng, kom vi i kontakt med Ole Kristian Eide som for tiden er doktorstipendiat ved institutt for materialteknologi (faggruppe uorganisk kjemi på NTNU), og vi var så heldige å få lov til å besøke han på laben hans. Eide forsker til vanlig på metallocene/methylaluminoksan katalysatorer for polymerisering av olefiner, og han er en av de som bruker infrarød spektroskopi i forskningsarbeidet. Da vi besøkte han, fikk vi lov til å være med å sette på flere prøver med forskjellige kjemiske forbindelser, og selv om analyseringen ikke gikk helt som det skulle, var det veldig lærerikt! Ole Kristian Eide viste oss også hvilken informasjon vi kan få ut av spekteret det infrarøde spektroskopet gir.

Vi tok også kontakt med Sintef for å høre om det var noen der vi kunne være med for å se på hvordan de bruker spektroskopiske metoder i forskningsprosjektene sine, men dessverre var det ingen der som skulle gjennomføre analyser i perioden prosjektet vårt varte. De henviste oss imidlertid videre til medisinstudenten, Silje Alvestad (NTNU, det medisinske fakultet, nevrovitenskap) som inviterte oss ned til St. Olavs hospital for å være med å sette på en NMR-prøve. Silje Alvestad forsker på epilepsi og bruker i den forbindelse NMR til å studere rottehjernen. Dette gjøres ved at et kontrastmiddel som inneholder mangan blir injisert i rotta like før den blir avlivet, deretter dissekeres hjernen ut før prøven plasseres i en beholder med tungtvann (deuterium) og settes inn i NMR-maskinen. Forskningsavdelingen til NTNU på St. Olavs hospital har tre NMR-maskiner på henholdsvis 300 MHz (7,1 Tesla), 500 MHz (11,7 Tesla) og 600 MHz (14,1 Tesla), og vi fikk se på når maskinen på 600 MHz skulle benyttes.

8 Oppsummering

Spektroskopi er en fellesbetegnelse på analytiske metoder som ser på hvordan elektromagnetisk stråling vekselvirker med et stoff, og ut i fra disse resultatene kan man skaffe seg informasjon om stoffet. Ved å bruke spektroskopiske metoder i det vitenskapelige arbeidet, kan man skaffe seg kunnskap om den kjemiske sammensetningen til et stoff, massen, strukturen, vibrasjonene, rotasjonstilstandene og mye mer. Spektralanalysen omfatter studiet av emisjons- og absorpsjonsspektrene til forskjellige stoffer, og baserer seg på at alle grunnstoffer og molekyler har et unikt spekter ved gitte forutsetninger. Massespektroskopien ser på hvordan ladede partikler oppfører seg i elektriske og magnetiske felt, og denne informasjonen kan si noe om forholdet mellom massen og ladningen til partiklene som studeres. Hvis man er interessert i å finne ut noe om vibrasjonstilstandene og strukturen til molekyler, kan man bruke infrarød spektroskopi. Molekylet vil absorbere den energien som samsvarer med frekvensen til vibrasjonsbevegelsene i molekylets bindinger, og dette kan vi benytte for å finne strukturen til molekylet. I ramanspektroskopi blir stoffet som skal undersøkes utsatt for intens stråling, og noe av strålingen vil da bli spredt ut i vilkårlige retninger. Ved å studere denne spredningseffekten kan man blant annet få informasjon om molekylets vibrasjons- og rotasjonsbevegelser. Raman- og infrarød spektroskopi er såkalte komplementære metoder, det vil si at de utfyller hverandre, og ved kombinere disse kan man få svært detaljert kunnskap om det stoffet som undersøkes.

NMR benytter seg av at visse atomer er paramagnetiske, noe som fører til at de innretter seg på en bestemt måte når de blir utsatt for et magnetisk felt. Atomene får da et magnetisk dipolmoment, og ved å la dette indusere en strøm i en spole, kan man finne ut mer om prøven. Denne teknikken er flittig brukt innenfor kjemi, medisin og biologi, men også fysikere har tatt i bruk metoden, i studiet av faste stoffer. Astronomien er kanskje det vitenskapelige området som har hatt størst utbytte av den spektroskopiske utviklingen. Ved å ta i bruk de tidligere omtalte teknikkene, har man fått kunnskap om deler av verdensrommet som mennesket aldri vil nå (iallfall ikke på veldig lang tid).

9 Avslutning

Dette prosjektet var vært svært lærerikt, men også veldig krevende. Da vi begynte med prosjektarbeidet, visste vi ikke stort mer enn at det fantes noe som het spektralanalyse og noe som het massespektroskopi. Men etter en tur innom diverse nettsider, fant vi raskt ut at spektroskopi er mye mer enn det. Siden forkunnskapene våre på dette området ikke var de beste, måtte vi bruke mye tid på å sette oss inn i de forskjellige metodene, og det er ikke få timer som har gått med til lesing av bakgrunns- og tilleggslitteratur om emnet. Vi oppdaget fort at det ikke var så lett å finne god og oversiktlig informasjon om dette emnet, så vi var derfor nødt til å leite gjennom de bøkene vi kom over, og deretter trekke ut det som var relevant for vårt prosjekt. Siden spektroskopi først og fremst brukes innenfor kjemi og medisin, var det heller ikke så lett å finne anvendelser av de omtalte spektroskopiske metodene innenfor fysikk.

Et av målene våre med prosjektet var å se hvordan spektroskopi brukes i praksis, og takket være flere hjelpsomme personer fikk vi arrangert to besøk til folk som benytter seg av spektroskopi i det daglige arbeidet. Disse besøkene var noe av beste med hele prosjektet, og de gjorde at vi ble enda mer motiverte! Alt i alt har vi lært veldig mye i løpet av disse ukene som prosjektet har vart, og selv om det har vært veldig tidkrevende, har vi fått et mye større utbytte enn vi hadde trodd på forhånd!

10 Takk til

Jon Andreas Støvneng for all hjelp underveis i prosjektet, og for at han hjalp oss å komme i kontakt med Ole Kristian Eide.

Ole Kristian Eide for at vi fikk lov til å komme og se på laben hans og se hvordan infrarød spektroskopi fungerer i praksis.

Silje Alvestad som viste oss hvordan hun bruker NMR i sine forskningsprosjekter, og Øystein Risa som gav oss grundige og utfyllende forklaringer underveis i prosessen.

11 Litteraturliste

11.1 Bøker

1. *Particle beam microanalysis, Fundamentals, methods and applications*, E. Fuchs, H. Oppolzer og H. Rehme, VCH Verlagsgesellschaft, Weinheim i Tyskland, 1990
2. *Modern spectroscopy*, J. Michael Hollas, 4. utgave, John Wiley & Sons, England, 2004
3. *Infrared Spectroscopy of Molecular Clusters*, M. Havenith, Springer Verlag 2002
4. *Surface infrared and raman spectroscopy, methods and applicaton*, W. Suetaka, Plenum press, New York, 1995
5. *Modern techniques in applied molecular spectroscopy*, Francis M. Mirabella, John Wiley & Sons, Canada, 1998
6. *Vibrational spectroscopy spectroscopy, methods and application*, Alois Fadini og Frank-Michael Schnepel, Ellis Horwood Limited, England, 1989
7. *UV-VIS spectroscopy and its applications*, Perkampus, Springer Laboratory, Springer-Verlag 1992
8. *Photoelectron Spectroscopy of High Temperature Superconductors and Related Compounds*, Henrik Nylén, Lund 1998
9. *Rom Stoff Tid*, Jerstad, Sletbak, Auen Grimenes, J. W. Cappelens Forlag A.S., Oslo 2003
10. *A complete introduction to modern NMR spectroscopy*, Roger S. Macomber, John Wiley & Sons, Canada, 1998
11. *Fourier transforms in NMR, optical and mass spectroscopy, A user's handbook*, Alan G. Marshall og Francis R. Verdun, Elsevier science publishers, Nederland, 1990
12. *Practical amateur spectroscopi*, Stephen F. Tonkin, Springer Verlag, London Limited 2002
13. *General Chemistry*, Raymond Chang, 4. utgave, Mc-Graw-Hill, Boston, 2005

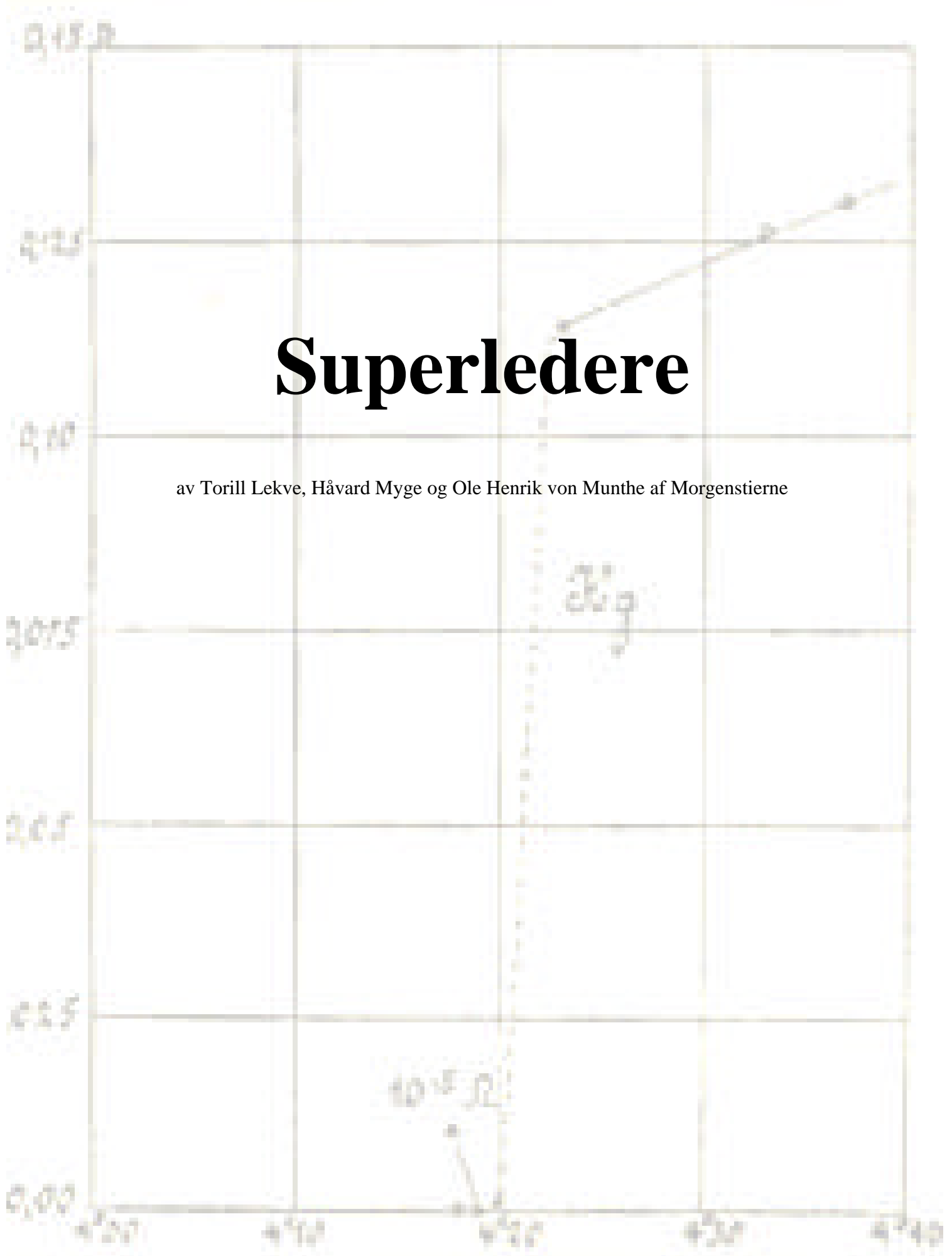
11.2 Internett

14. <http://www.geocities.com/CapeCanaveral/9454/overview.htm>
15. <http://chem.ch.huji.ac.il/~eugeniik/history/wollaston.html>
16. <http://www.caplex.net>
17. <http://www.observatoriemuseet.uio.no/Utstilling/astronomi.html>
18. <http://www-outreach.phy.cam.ac.uk/camphy/index.htm>
19. <http://www.plicht.de/chris/35fraunh.htm>
20. <http://romteknologi.no/books/38/4.html>
21. http://nn.wikipedia.org/wiki/Elektromagnetisk_str%C3%A5ling
22. <http://romteknologi.no/books/38/3.html>
23. <http://hyperphysics.phy-astr.gsu.edu/hbase/magnetic/maspec.html#c1>
24. http://www.iu.hio.no/~hannet/organisk_kjemi/org_tekn/IR-spekt.htm
25. <http://orgchem.colorado.edu/hndbksupport/irtutor/tutorial.html>
26. <http://www.kj.uib.no/ft-ir-norsk>
27. http://www.kjemi.uio.no/09_spekt

28. <http://www.answers.com/topic/raman-effect?hl=raman>
29. <http://www.kosi.com/raman/resources/tutorial/>
30. http://www.chem.au.dk/~www/weeknotes/week_914.pdf
31. http://www.kjemi.uio.no/09_spekt
32. http://www.nmas.no/Produkter/Analytisk_kjemi/UV-VIS_Spektroskopi/785/UV_VIS_teoripdf
33. http://www.kjemi.uio.no/09_spekt
34. http://www.spectroscopyeurope.com/Process_16_4.pdf
35. <http://folk.uio.no/haraldm/>
36. http://www.kjemi.uio.no/09_spekt
37. http://www.dartmouth.edu/~soilchem/XAS_descript.htm
38. <http://www.kornsirkler.org/Sistenytt/Index.html>
39. <http://www.amptek.com/xrf.html>
40. http://ion.eas.asu.edu/descript_xrf.htm
41. http://www.phys.ntnu.no/brukdef/undervisning/tfy4310_lab/ovinger/mbl_nmr.pdf
42. <http://hyperphysics.phy-astr.gsu.edu/hbase/nuclear/larmor.html>
43. http://www.tidsskriftet.no/pls/lts/PA_LT.VisSeksjon?vp_SEKS_ID=16690
44. www.aftenposten.no/viten/article772189.ece
45. www.nvg.org/org/taf/publikas/klasnova.htm
46. www.nvg.org/org/taf/tema/orion1.htm
47. http://home.online.no/~taninfo/undervisning/fysikk/3fy/3fy_testoppgaver/stjerne_lys.html
48. www.nvg.org/org/taf/tema/orion1.htm
49. http://home.online.no/~taninfo/undervisning/fysikk/3fy/3fy_testoppgaver/stjerne_lys.html
50. www.nvg.org/org/taf/publikas/klasnova.htm
51. http://astrosun2.astro.cornell.edu/academics/courses/astro201/hr_diagram.htm

Superledere

av Torill Lekve, Håvard Myge og Ole Henrik von Munthe af Morgenstierne



Innholdfortegnelse

	Side:
Innledning	2
1) Oppdagelsen av superlederen	3
2) Magnetisering	5
• Atomære og nukleære magnetiske momenter	7
• Diamagnetisme	8
• Paramagnetisme	9
• Ferromagnetisme	10
• Hysterese	10
3) Superleder type I	13
• Kritisk magnetfelt	14
• Meissnereffekten	15
• Penetrasjonsdybde	17
• Magnetisering	19
4) BCS teori	22
• Cooper par	22
• Energigap	23
• Koherenslengde	25
5) Superleder type II	27
• Virvelfase	27
• Pinning	31
• Høytemperatur superleder	32
6) Flukskvantisering	37
7) Demonstrasjon av pinning og meissnereffekt	39
8) Anvendelser	44
• Superledende strømkabler	44
• Elektriske supermotorer	45
• Levitasjon	46
• Måling av hjernebølger	47
9) Oppsummering	48
10) Kilder	50

Innledning

Ingen på gruppa vår visste mye om superledere da vi valgte dette tema. Vår kunnskap begrenset seg stort sett til det vi har lest om svevetog i blader som illustrert vitenskap. Vår nysgjerrighet om fysikken bak svevetoga, samt at vi har hatt om vanlige ledere og isolatorer i dette semesteret, gjorde at vi ville skrive om dette. Det viste seg at det var et interessant tema og at det omhandlet mye mer enn svevetog!

I de første kapitlene vil vi se på fysikken og historien bak superledere, så vil vi se på fenomenet i praksis gjennom en demonstrasjon vi fikk på laben, før vi avslutter med å skrive om anvendelser.

Tekstene er skrevet med inspirasjon av kildene det er henvist til på slutten av hvert kapittel. Vi har skrevet kildehenvisning under illustrasjoner vi har kopiert fra andre skriv. De illustrasjonene det ikke er kildehenvisning til har vi lagd selv, med inspirasjon fra teksten. I kapittel 10 har vi en fullstendig referanseliste over kildene.

1) Oppdagelsen av superlederen

For å finne ut hva en superleder er, og forstå hvorfor den er så viktig, finner vi det naturlig å skrive om oppdagelsen av superlederen.

Den nederlandske fysikeren Heike Kamerlingh Onnes klarte i 1908 å fremstille flytende helium. Ingen hadde noensinne klart dette før pga. at helium kondenserer ved en temperatur på 4,2 K. I 1911 prøvde Onnes å fryse ned kvikksølv med flytende helium, med håp om å bevise at motstanden ville forsvinne helt ved det absolutte nullpunkt. Det han oppdaget var meget merkverdig. Som kjent sank resistansen proporsjonalt med temperaturen, men ved en bestemt temperatur (4,15 K) ble motstanden plutselig lik null. Altså måtte man ikke fryse ned metallet til 0 K. Det var åpenbart at dette fenomenet hadde stor praktisk betydning, selv om man på den tida ikke hadde noen teoretisk forklaring. Onnes kalte dette fenomenet for ”superledning”, og fikk i 1913 nobelprisen for oppdagelsen. Denne oppdagelsen var begynnelsen av et fysisk/kjemisk tema som har engasjert fysikere over hele verden i alle år og frem til i dag, og vil sannsynligvis fortsette å gjøre det fremover også.

Det var ingen som kunne komme med en god forklaring på hvorfor superlederen oppførte seg slik som den gjorde ved den kritiske temperaturen T_C . Hans Meissner og Robert Ochsenfeld studerte lenge på superlederens magnetiske egenskaper og fant ut i 1933 et nytt fenomen knyttet til superlederen, den såkalte Meissnereffekten. Dette gikk ut på at når superlederen nådde T_C så skapte den et magnetfelt som var tilsvarende det påtrykte magnetfeltet, men motsatt rettet, slik at magnetfeltet ble lik null inne i superlederen.

Ikke før 24 år senere klarte fysikere å finne en god teoretisk forklaring på superleder-mysteriet. John Bardeen, Leon Cooper og Robert Schrieffer var de som klarte dette ved å fremstille BCS-teorien. Teorien sier at noen av de negativt ladde elektronene i en superleder danner elektronpar, såkalte ”cooperpar”. Disse cooperparene kan ”flyte” langs kanaler som er dannet av den regelmessige strukturen av positivt ladde ioner i materialet. Det er denne ”elektron-væsken” som gjør at metallet er superledende.

Etter BCS –teorien gikk interessen for superlederen ned. Dette kom av at det var veldig dyrt å fremstille flytende helium. I 1986 kom det et nytt gjennombrudd. Georg Bednorz og Alex Müller fant ulike materialer som ble superledere ved 35 K, og i 1987 fant de et keramisk materiale som ble superleder ved 92 K. Etter denne oppdagelsen begynte fysikere igjen å forske på superledere, siden man med så ”høye” temperaturer kunne bruke flytende nitrogen som er mye billigere, i kjølingsprosessen i stedet for helium. Disse nye superlederne ved så mye høyere temperatur, kalles ”høytemperatur superledere”.

Etter 1987 har man forsket mye på superledere, og funnet ut veldig mye om fenomenet. I dag bruker man superledere i forskjellige legeinstrumenter og andre teknologiske innretninger. Dagens ”varmeste” superleder har en kritisk temperatur på 138 K, men det er blitt observert et ”motstandsdropp” ved 150 K på et ukjent keramisk materiale. Man kan derfor håpe på at man snart får en superleder med kritisk temperatur opp mot 150 K.

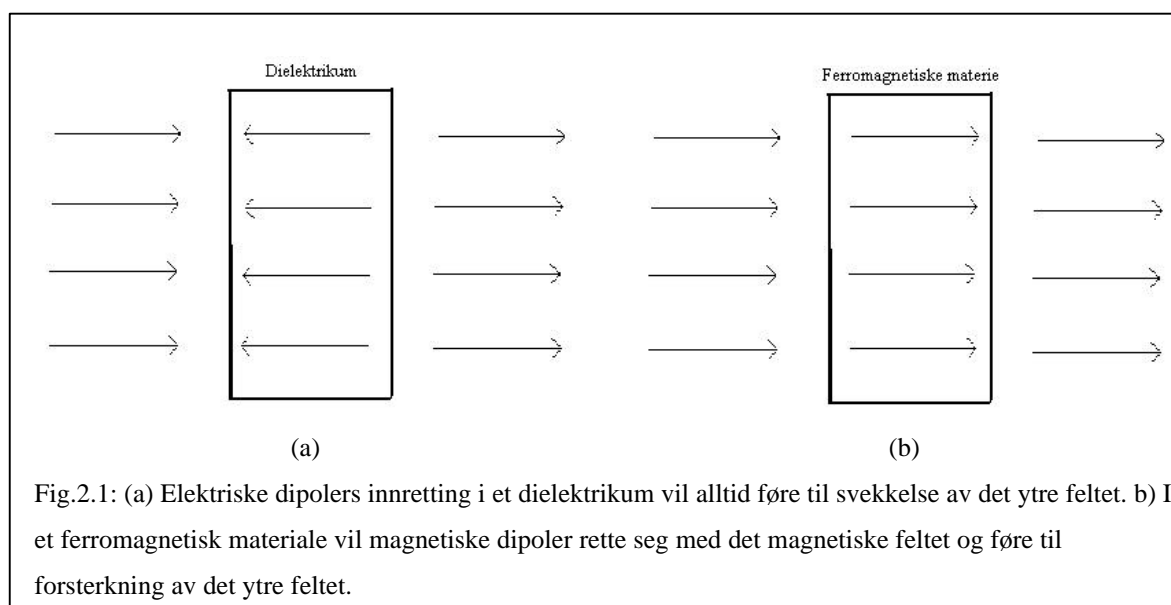
Kilder:

- http://www.nrk.no/programmer/tv/schrodingers_katt/
- Serway, Moses og Moyer; ”Modern Physics”, 1997

2) Magnetisering

De magnetiske egenskapene til superledere er noe av det som gjør superledere så interessante. Vi skal derfor gi en innføring i magnetisme generelt før vi fordyper oss i superledere.

Vi vet at et elektrisk felt fører til innretning av atomære/molekylære elektriske dipoler i et dielektrikum, noe som igjen fører til en svekkelse av det elektriske feltet (fig. 2.1a). Ved å plassere et ferromagnetisk materiale i et magnetisk felt får vi en forsterkning av feltet (fig. 2.1b).



For de fleste materialer vil vi få en svak svekkelse (diamagnetiske materialer) eller en svak forsterkning (paramagnetiske materialer) av det magnetiske feltet. Men med enkelte stoffer (ferromagnetiske) vil vi få en dramatisk forsterkning av feltet. Dette vil vi komme tilbake til senere.

Materialets magnetiske egenskaper karakteriseres ved deres magnetisering (\vec{M}), magnetisk dipolmoment per volumenhet. I likning 2.1 kan vi se at dette er det samme som strømmen delt på lengden av materialet (lengden av materialet i strømmens retning);

$$M = \frac{dm}{V} = \frac{IdA}{ldA} = \frac{I}{l} \quad \text{likning 2.1}$$

Sammenhengen mellom magnetisk felt uten magnetisk materiale (B_0) og magnetisk felt med magnetisk materiale (B) er;

$$\vec{B} = \vec{B}_0 + \mu_0 \vec{M} \quad \text{likning 2.2}$$

Ved hjelp av uttrykket for den magnetiske intensiteten (H) i likning 2.3 kan vi skrive om uttrykket for feltet med magnetisk materiale (likning 2.2) til en sammenheng mellom intensiteten og magnetiseringen (likning 2.4).

$$\vec{H} = \frac{\vec{B}_0}{\mu_0} \quad \text{likning 2.3}$$

$$\vec{B} = \mu_0 (\vec{H} + \vec{M}) \quad \text{likning 2.4}$$

For et isotropt homogent materiale vil sammenhengen mellom magnetiseringen og intensiteten være proporsjonal. Dette kan vi se i likning 2.5. Proporsjonalkonstanten definerer vi til å være c_m . Denne er forskjellig fra material til material og blir kalt materialets magnetiske susceptibilitet.

$$\vec{M} = c_m \vec{H} \quad \text{likning 2.5}$$

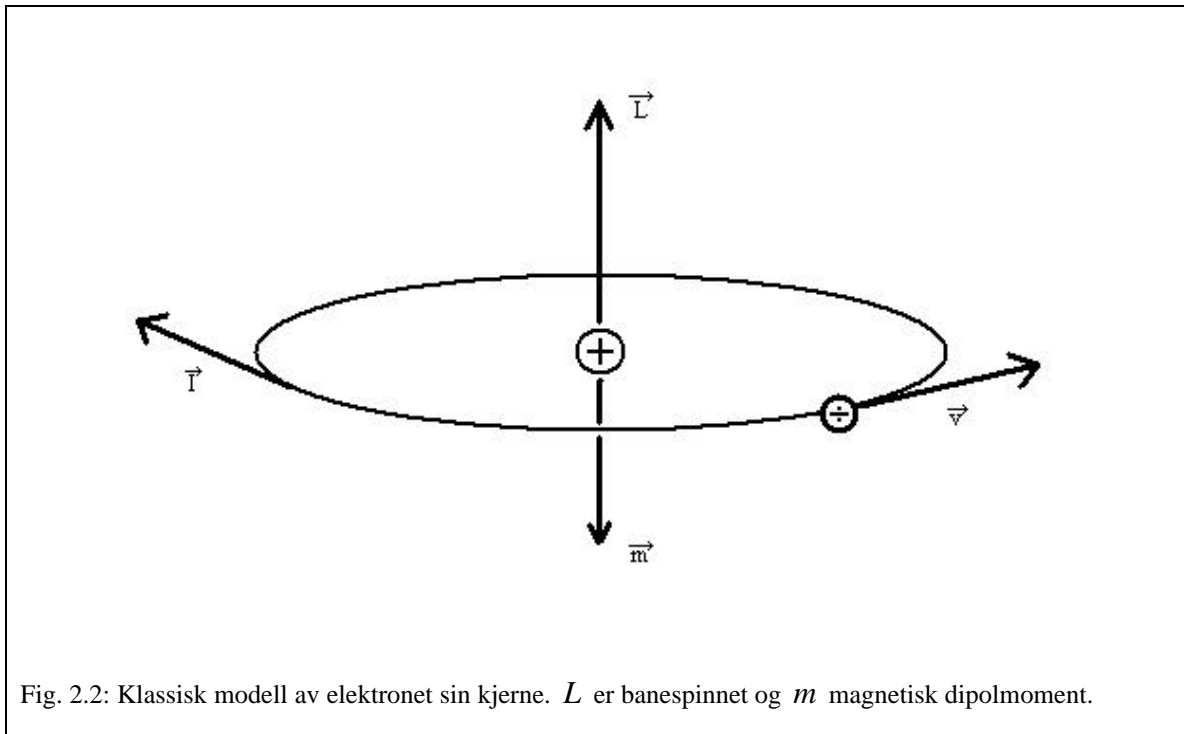
Med sammenhengen mellom intensitet og magnetisering (likning 2.5) kan vi skrive om uttrykket for totalt magnetfelt (likning 2.2) slik at en ser at det er en proporsjonal sammenheng mellom totalt magnetfelt og intensitet;

$$\vec{B} = \mu_0 (\vec{H} + \vec{H} c_m) = \mu_0 \vec{H} (1 + c_m) = \mu \vec{H} \quad \text{likning 2.6}$$

For superledere er $c_m = \pm 1$, altså blir $B = 0$. Dette tilsier at magnetisk felt ikke trenger inn i superledere. Vi kommer tilbake til dette senere i oppgaven.

Atomære og nukleære magnetiske momenter

Vi ser først på den klassiske modellen for elektronet rundt sin kjerne (fig. 2.2).



Omløpstiden (T) til et elektron rundt sin kjerne er gitt i likning 2.7. r er radiusen til banen til elektronet og v er farta til elektronet.

$$T = \frac{2\pi r}{v} \quad \text{likning 2.7}$$

Strømmen rundt atomkjernen vil være ladning delt på omløpstiden;

$$I = \frac{e}{T} = \frac{ev}{2\pi r} \quad \text{likning 2.8}$$

Ved å se på hva banespinn (L) og det magnetiske dipolmomentet (m) er gitt ved kan en se at det er en lineær sammenheng mellom disse som avhenger av ladningen (e) og elektronets masse (m_e). Denne sammenhengen kommer fram i følgende likninger;

$$L = r \times p = r \times m_e v \quad \text{likning 2.9}$$

$$m_L = IA = \frac{\div e v}{2pr} p r^2 = \frac{\div e v r}{2} = \frac{\div e}{2m_e} L \quad \text{likning 2.10}$$

Vi har nå regnet klassisk på det, men også hvis vi regner kvantemekanisk på det vil vi få tilsvarende svar, men størrelsen på banespinn er da kvantisert, og av likning 2.10 ser en da at også det magnetiske dipolmomentet er kvantisert.

Elektronet har også egenspinn (S) som bidrar til det magnetiske dipolmomentet, elektronet er altså en liten magnet i seg selv;

$$m_s = \div \frac{e}{m_e} S \quad \text{likning 2.13}$$

Også protonet og nøytronet har egenspinn, men på grunn av at deres masse er mye større enn elektronets, er dette ≈ 2000 ganger mindre enn elektronets, og kan derfor neglisjeres.

Fra kvantemekanikken har vi at elektroner opptrer parvis, så for et atom i grunntilstand kan elektronenes magnetiske momenter oppheve hverandre, men har vi atomer med oddeantall elektroner vil vi alltid få et magnetisk dipolmoment.

Diamagnetisme

Diamagnetiske stoffer kjennetegnes som tidligere nevnt ved at de gir en svak minking av det påtrykte feltet. Vi har ingen netto dipolmoment.

I slike stoffer vil det elektriske dipolmomentet til to og to elektroner oppheve hverandre. Men når det settes på et påtrykt magnetisk felt vil elektronene med magnetisk

dipolmoment mot feltet få en svak økning og elektroner med moment med feltet få en svak minkning. Dette skjer fordi systemet vil motsette seg den økende magnetiske fluksen.

Dette kan forklares med hjelp av paringsreglen (kvantemekanikk) og lenzs lov.

Alle stoffer har en slik diamagnetisk effekt, men denne effekten er veldig liten for vanlige materialer sammenlignet med den paramagnetiske og ferromagnetiske effekten.

Paramagnetisme

Atomer og molekyler med ett eller flere uparede elektroner vil ha permanente magnetiske dipolmomenter, men uten ytre felt vil innrettingen av disse være tilfeldig. Dermed vil det totalt ikke dannes noe magnetisk felt, men med et ytre felt vil disse dipolmomentene innrette seg med det magnetiske feltet, noe som fører til en økning av det totale feltet. Dette kan forklares ut fra å se på hva den potensielle energien til en magnetisk dipol er gitt ved;

$$U = -\vec{m} \cdot \vec{B} \qquad \text{likning 2.14}$$

Som vi vet vil alt i naturen "søke mot" lavest mulig potensiell energi, og av likning 2.14 ser vi at vi vil få minst mulig potensiell energi når det magnetiske feltet og magnetiseringen peker samme vei ($\vec{m} \parallel \vec{B}$). Altså vil vi få en innretting av de magnetiske momentene som fører til forsterkning av feltet.

Termiske bevegelser og kollisjoner vil imidlertid føre til en forstyrning av innrettingen. Jo høyere temperatur, jo mer forstyrning, noe som fører til en minkende magnetisering.

Ved høye temperaturer vil denne forstyrringen bli så dominerende i forhold til magnetiseringen, noe som fører til at den diamagnetiske effekten, som ikke er temperaturavhengig med mindre vi snakker om superledere, vil bli større enn den paramagnetiske. Vi vil altså få en svak svekkelse av feltet når temperaturen blir høy nok.

Ferromagnetisme

Med ferromagnetiske materialer får vi en kraftig magnetisering når vi har et ytre felt, og også etter at feltet er slått av vil vi ha magnetisering, riktignok noe lavere. Vi snakker altså om permanente magneter.

Et eksempel på et ferromagnetisk materiale er jern, som også er opphavet til betegnelsen ferromagnetisme (ferro betyr jern).

Dette fenomenet lar seg ikke forklare klassisk. Vi må altså bruke kvantemekanikk. Vi skal derfor ikke gå i detalj, men gi et overflatisk bilde av hva som skjer.

To elektroner i et begrenset område av materialet kan ifølge Paulis eksklusjonsprinsipp ikke oppta nøyaktig samme energitilstand. Dette fører til at uparede elektroner med tilnærmet parallell banespinn vil ordne seg i baner slik at den midlere avstanden blir størst mulig, for å redusere coulombfrastøtingen, og dermed redusere den potensielle energien mest mulig. Elektroner vil også ordne seg slik at spinn og magnetisk dipolmoment blir parallelle, for å få den laveste energitilstanden. Dette blir omtalt som en spontan innretting, som altså skjer i ferromagnetiske materialer når de blir utsatt for ytre magnetfelt. Fordi denne spontane innrettingen har lavest mulig potensiell energi vil den bli opprettholdt også etter at det ytre magnetfeltet er borte.

Hysteres

For ferromagnetiske stoffer er ikke sammenhengen mellom B og H entydig. Vi kan studere denne sammenhengen eksperimentelt ved hjelp av en "rowlandring" (fig. 2.3).

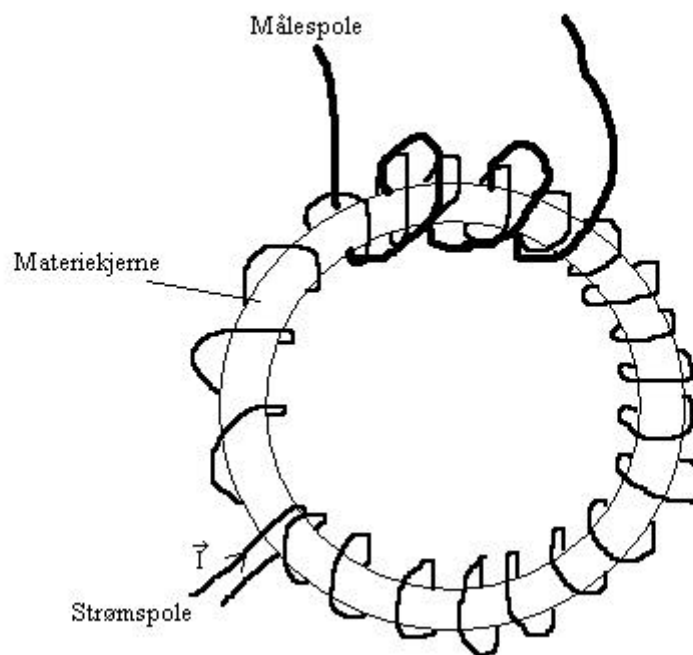


Fig. 2.3: En rowlandring. Ringen er et ferromagnetisk materiale, og vi har en strømførende spole rundt hele ringen som skaper et påtrykt felt (B_0) og en spole øverst til å måle det magnetiske feltet (B).

Ved å se på følgende fem punkter samt fig. 2.4, kan vi forstå sammenhengen mellom B og H til et ferromagnetisk materiale:

1. Vi starter med et umagnetisert materiale og ingen strøm, og altså null påtrykt felt. Vi har da $B = 0$.
2. Vi øker strømmen og får en magnetisering og et magnetisk felt.
3. Når vi skrur av strømmen har magnetiseringen i materialet ført til at vi ikke har $B = 0$! Vi har fått et permanent magnetfelt i ringen.
4. For å få $B = 0$ må vi sende en "negativ" strøm gjennom spolen, altså en strøm motsatt vei.
5. Når vi fortsetter slik, noe man for eksempel kan gjøre med en vekselstrøm, ser vi fort at vi vil få en graf som i fig. 2.4.

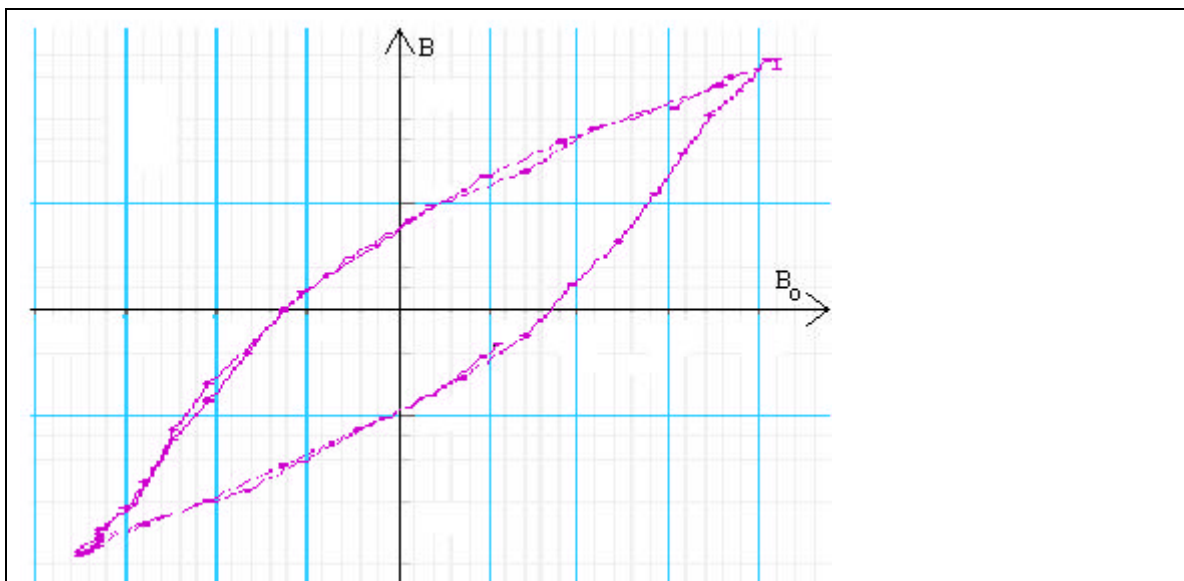


Fig. 2.4: Magnetisk hysteresese. Ved å sende vekselstrøm gjennom en spole rundt en rowlandring vil en få en graf av denne typen.. Totalt magnetisk felt (B) er logget som funksjon av påtrykt magnetisk felt (B_0).

(Kilde: Labrapport; "Magnetisk hysteresese", av Torill Lekve og Ole Henrik von Munthe af Morgenstjerne).

Kilder:

- Egil Lillestøl, Ola Hunderi og Jan R. Lien; "Generell fysikk – bind 2", 2001
- Serway, Moses og Moyer; "Modern Physics", 1997

3) Superleder type I

Superledere av type I består stort sett av rene metaller som er normalt ledende ved normale temperaturer, men får motstand lik 0 når de kjøles ned til en viss temperatur. Denne temperaturen kalles den kritiske temperaturen til stoffet (T_C), og varierer fra stoff til stoff.

Motstanden i metallet minker gradvis ettersom temperaturen synker, helt til $T = T_C$, da motstanden stuper bratt, som vist på figur 3.1..

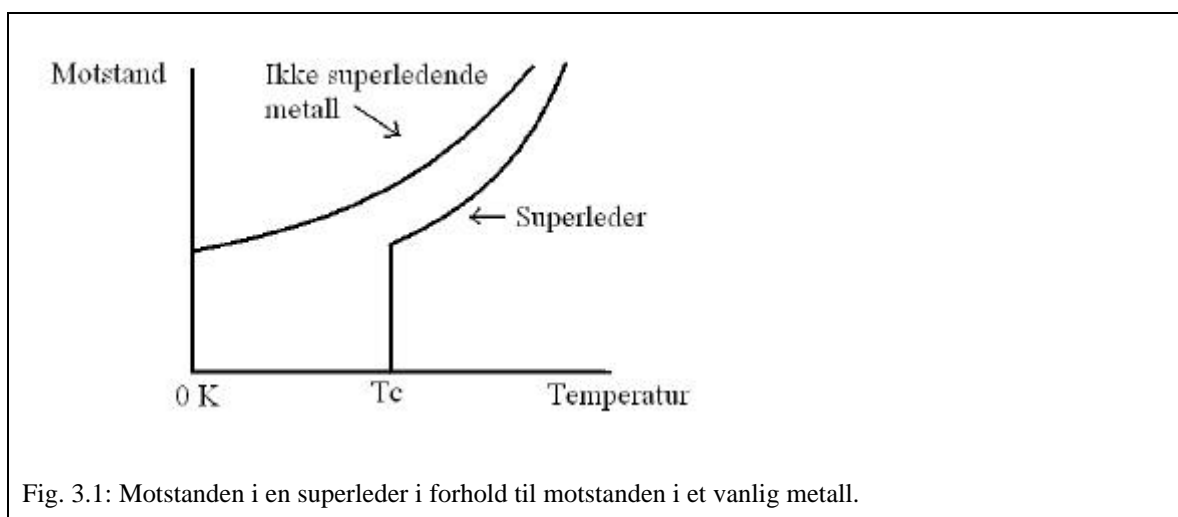


Fig. 3.1: Motstanden i en superleder i forhold til motstanden i et vanlig metall.

I tillegg til å være superledende oppfører disse stoffene seg som perfekte diamagneter. Når et stoff er en diamagnet vil det si at om stoffet blir utsatt for et ytre magnetfelt, vil stoffet sette opp et motsatt rettet magnetfelt. Dette feltet vil svekke det påtrykte feltet. Dersom en superleder blir utsatt for et påtrykt magnetfelt ved $T < T_C$, induseres det en overflatestrøm i superlederen, og denne strømmen setter opp et motsatt rettet magnetfelt slik at det totale magnetfeltet i stoffet blir lik null. Superlederen er da en *perfekt diamagnet*.

Superledere av type I ble oppdaget i 1911, da Heike Kamerlingh-Onnes oppdaget at motstanden i kvikksølv plutselig sank til omtrent 0 ved 4,17 K.

Type I superledere krever lavere temperatur for å oppnå superledende tilstand enn type II superledere, som vi skal omtale senere.

Kritisk magnetfelt

Når en superleder av type I blir utsatt for et magnetfelt (B), vil verdien for den kritiske temperaturen (T_c), synke samtidig med at magnetfeltet øker. Hvis magnetfeltet overskrider stoffets kritiske magnetfelt (B_c), vil den superledende tilstanden til stoffet bli ødelagt. Stoffet vil da aldri bli superledende, uansett temperatur, men ha normalt ledende egenskaper med endelig motstand. Det kritiske feltet avhenger av temperaturen omkring superlederen og stoffets kritiske temperatur.

$$B_c(T) = B_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \quad (\text{likning 3.1})$$

Som vi kan se av likningen over er $B_{c \max}$ ved 0K.

Det kritiske feltet begrenser også hvor mye strøm en superleder tåler før den går vekk fra sin superledende tilstand. Ifølge Amperes lov vil en strøm (I) i en leder danne et magnetfelt utenfor lederen, og denne sammenhengen kan vi se i likning 3.2. Denne likninga gjelder en sylinderformet, rett, strømførende leder.

$$B = \frac{\mu_0 I}{2\pi r} \quad (\text{likning 3.2})$$

$$I = \frac{2\pi r B}{\mu_0} \quad (\text{likning 3.3})$$

I likning 3.2 og 3.3 er r avstanden fra sentrum av lederen til overflaten. Det vil altså si at en superleder ikke kan føre uendelig mye strøm, men har en øvre grense som varierer fra stoff til stoff.

For type I superledere er verdien for de kritiske feltene svært lave. På grunn av dette er ikke superledere av type I brukt til å produsere såkalte "høyfeltsmagneter", også kalt superledende magneter. Dette fordi strømmen generert av magnetfeltet ødelegger den superledende tilstanden.

Meissnereffekten

Før 1933 var det antatt at superledning var en demonstrasjon av perfekt ledning. Hvis en perfekt leder ble avkjølt til en lavere temperatur enn sin kritiske temperatur, hele tiden utsatt for et magnetfelt, ville magnetfeltet bli fanget inne i superlederen, selv etter at det påtrykte feltet var fjernet. Den siste tilstanden til en perfekt leder ville være avhengig av hva som fant sted først.

- 1) Hvis lederen ble utsatt for et magnetfelt etter at den var blitt nedkjølt til en temperatur lavere enn T_C , ville feltet bli støtt ut av superlederen.
- 2) Hvis feltet ble påtrykt før nedkjølingen derimot, ville ikke feltet bli ekskludert etter at temperaturen nådde en lavere temperatur enn T_C .

Da eksperimentene for å undersøke den magnetiske oppførselen til superlederne ble gjennomført i 1930-årene, viste resultatene noe helt annet. I 1933 oppdaget Meissner og Ochsenfeld at når et metall ble superledende i nærvær av et svakt magnetfelt, ble feltet ekskludert, slik at $\vec{B} = 0$ overalt inne i superlederen. Den samme siste tilstanden ble oppnådd uansett om feltet ble påtrykt før eller etter at stoffet ble nedkjølt til en lavere temperatur enn sin kritiske temperatur. Dette medfører at en superleder av type I er mer enn bare en perfekt leder med resistivitet $\mathbf{r} = 0$, men også en perfekt diamagnet med $\vec{B} = 0$. Fenomenet med utelukkingen av det magnetiske feltet fra superlederens indre kalles *meissnereffekten*.

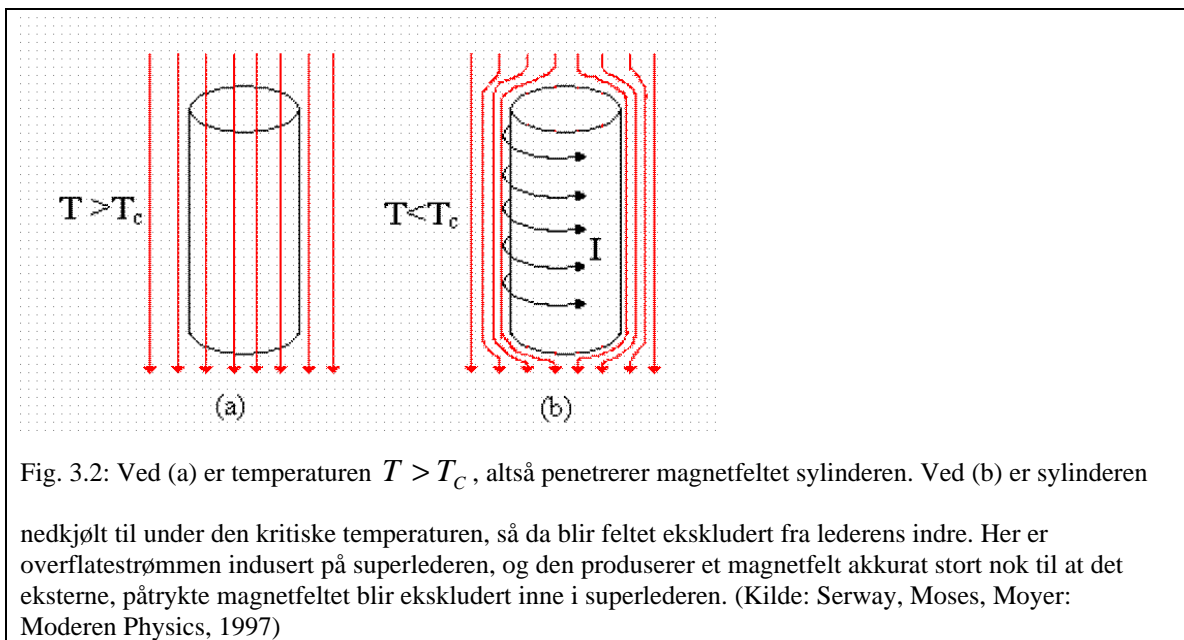


Fig. 3.2: Ved (a) er temperaturen $T > T_c$, altså penetrerer magnetfeltet sylinderen. Ved (b) er sylinderen nedkjølt til under den kritiske temperaturen, så da blir feltet ekskludert fra lederens indre. Her er overflatestrømmen induisert på superlederen, og den produserer et magnetfelt akkurat stort nok til at det eksterne, påtrykte magnetfelt blir ekskludert inne i superlederen. (Kilde: Serway, Moses, Moyer: Modern Physics, 1997)

Egenskapen ved at $\vec{B} = 0$ i en superleder er faktisk like fundamental som egenskapen ved ingen motstand og viser den viktige rollen magnetisme spiller i superledning.

Fra læren om elektrisitet har vi at en god leder utelukker statisk elektrisk felt ved å flytte ladninger til overflaten. Faktisk produserer overflateladningene et elektrisk felt som akkurat kansellerer det eksterne påtrykte feltet inne i ledere. På liknende vis ekskluderer superlederen det magnetiske feltet ved å danne overflatestrøm. Dette er vist på fig. 3.2. Overflatestrømmen vil forsvinne når det eksterne magnetfeltet er fjernet.

En kan bruke enkle argumenter basert på lover for elektrisitet og magnetisme for å vise at magnetfeltet inne i en superleder ikke kan forandres med tiden.

- Ifølge Ohms lov er det elektriske feltet i en leder proporsjonalt med motstanden til ledere. Siden $R = 0$ i en superleder, *må det elektriske feltet inne i en superleder være 0!*
- Faradays lov for induksjon kan uttrykkes som følgende

$$\oint \vec{E} d\vec{s} = -\frac{d\Phi_B}{dt} \quad (\text{likning 3.4})$$

Linjeintegralet av det elektriske feltet rundt en lukket kurve er den negative tidsderiverte til fluksen gjennom sløyfa.

Siden $\vec{E} = 0$ overalt inne i superlederen, er integralet over enhver lukket vei i superlederen lik 0. Dette medfører at

$$\frac{d\mathbf{f}_B}{dt} = 0, \text{ som sier oss at } \textit{den magnetiske fluksen i en superleder ikke kan forandres!}$$

Fra dette kan vi konkludere med at B må være konstant i en superleder!

Fordi en superleder er en perfekt diamagnet virker den frastøtende på en permanent magnet. Dette fikk vi demonstrert på laben, og vil bli omtalt senere.

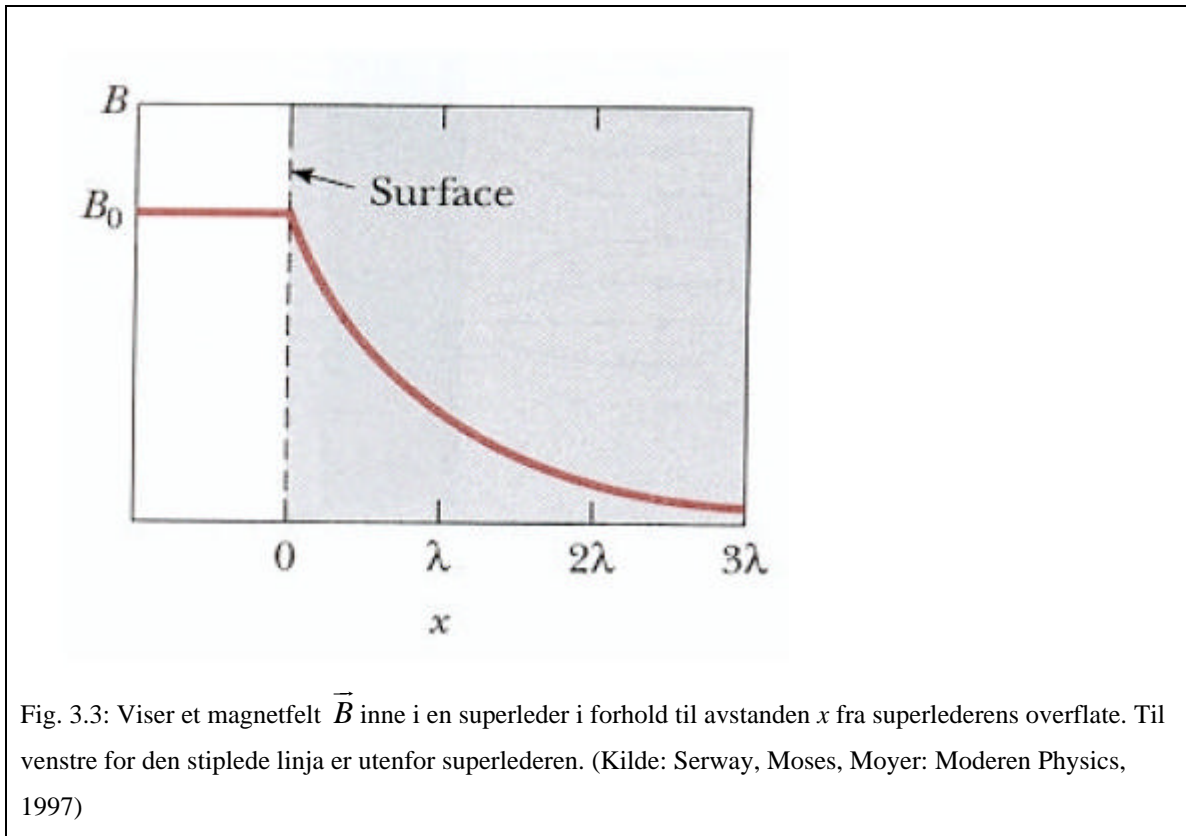
Penetrasjonsdybde

Som vi har sett på tidligere, vil magnetfeltet bli stengt ute fra type I superlederes indre på grunn av overflatestrømmer. I virkeligheten befinner ikke disse strømmene seg i et uendelig tynt lag på overflaten, men trenger et lite stykke inn i superlederen. I dette laget, som er omtrent 100 nm tykt, avtar magnetfeltet eksponentielt fra verdien til det ytre feltet og til 0. Dette er gitt ved følgende uttrykk

$$B(x) = B_0 e^{-\frac{x}{l}} \quad (\text{likning 3.5})$$

Her er det antatt at det ytre feltet er parallelt med overflaten til superlederen. I denne likninga (3.5) er B_0 verdien av magnetfeltet på overflaten, x er avstanden fra overflaten til et indre punkt, og l en parameter kalt penetrasjonsdybden.

Variasjonen i magnetfeltet i forhold til dybden feltet trenger inn i superledere av type I, er vist på figur 3.3.



Som vi ser på figur 3.3, blir magnetfeltet veldig lite på dybde "et par λ " under overflaten. Penetrasjonsdybden varierer med temperaturen i samsvar med formelen

$$l(T) = l_0 \left[1 - \left(\frac{T}{T_c} \right)^2 \right]^{\frac{1}{2}} \quad (\text{likning 3.6})$$

hvor l_0 er penetrasjonsdybden ved 0 K.

Fra dette likning 3.6 ser vi at l nærmer seg uendelig når T nærmer seg T_c . Videre, når T nærmer seg T_c , og stoffet er i en superledende tilstand, vil magnetfeltet trenge dypere og dypere inn i stoffet. Til slutt vil feltet penetrere hele stoffet (l blir uendelig), og stoffet viker vekk fra den superledende tilstanden og tilbake til sin normale tilstand.

Magnetisering

Når et stoff blir utsatt for et ytre magnetfelt (\vec{B}_0), vil det forekomme en magnetisering (\vec{M}). Magnetfeltet (\vec{B}) i stoffet er relatert til \vec{B}_0 og \vec{M} gjennom forholdet $\vec{B} = \vec{B}_0 + \mu_0 \vec{M}$. Når stoffet er i en superledende tilstand, er $B = 0$, ergo følger;

$$\vec{M} = -\frac{\vec{B}}{\mu_0} = c_m \vec{H} \quad (\text{likning 3.7})$$

Her er $c_m (= -1)$ magnetisk susceptibilitet (mottakelighet). Når som helst når materialet er superledende, vil dets magnetisering motvirke det eksterne magnetfeltet, og den magnetiske susceptibiliteten har sin maksimale negative verdi. Igjen ser vi at type I superledere er perfekte diamagneter.

På figur 3.4 ser vi forholdet mellom magnetfeltet inne i en type I superleder og det eksterne feltet (parallelt med en lang sylinder) ved $T < T_c$. På figur 3.5 ser vi magnetiseringen i forhold til det påtrykte feltet ved en satt temperatur.

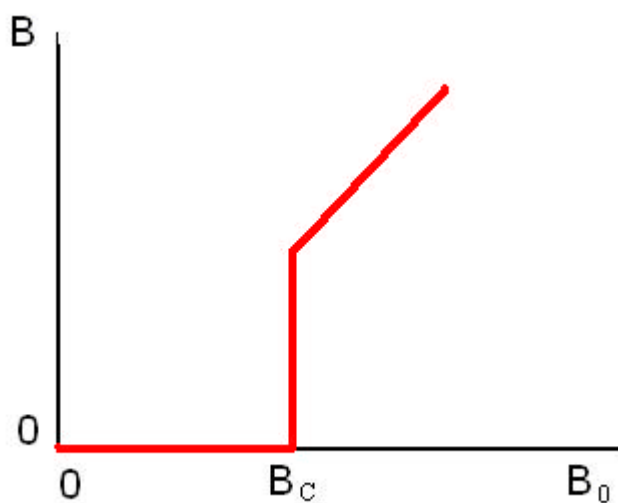


Fig. 3.4: Viser et magnetfelt inne i en type I superleder i forhold til et eksternt felt (parallelt til en lang sylinder) ved $T < T_C$. (Kilde: Serway, Moses, Moyer: Modern Physics, 1997)

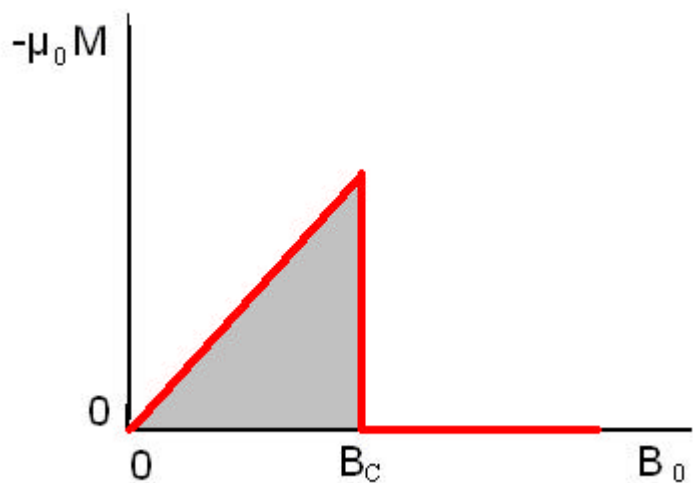


Fig.3.5: Her ser vi magnetisering i forhold til et eksternt felt ved en konstant temperatur. Vi ser at når $B_0 > B_C$ er magnetiseringen tilnærmet 0: $\bar{M} \approx 0$ for $B_0 > B_C$. (Kilde: Serway, Moses, Moyer: Modern Physics, 1997)

Etter at meissnereffekten ble oppdaget, var Fritz og Heinz London i stand til å utvikle likninger basert på likevekts termodynamikk. De kunne forklare det kritiske magnetfeltet ved hjelp av en energiøkning i den superledende tilstanden, en økning på grunn av ekskluderingen av fluks fra superlederens indre. Et system foretrekker å være i den tilstanden som har minst mulig fri energi, altså må den superledende tilstanden ha lavere fri energi enn i den normale tilstanden. Hvis E_s representerer energien i en superledende tilstand per volumenhet, og E_g energien i en normal tilstand per volumenhet, så er $E_s < E_g$ når temperaturen er under T_C , og materialet er superledende. Utelukkingen av et felt (B) fører til at den totale energien i den superledende tilstanden øker med $\frac{B^2}{2\mathbf{m}_0}$ per volumenhet. Verdien til det kritiske feltet er da definert med likningen

$$E_s + \frac{B_c^2}{2\mathbf{m}_0} = E_g \quad (\text{likning 3.8})$$

Fordi londonteorien også gir temperaturavhengighet av E_s , kunne ikke et eksakt uttrykk for $B_c(T)$ gis.

Merk at feltekskluderingsenergien $\frac{B^2}{2\mathbf{m}_0}$ er arealet under kurven i fig. 3.5.

Kilder:

- Serway, Moses, Moyer: "Moderen Physics", 1997
- <http://www.angstrom.uu.se/solidstatephysics/education/kvant/pdf/supraledning.pdf>
- Lillestøl, Hunderi, Lien: "Generell fysikk – bind 2", 2001

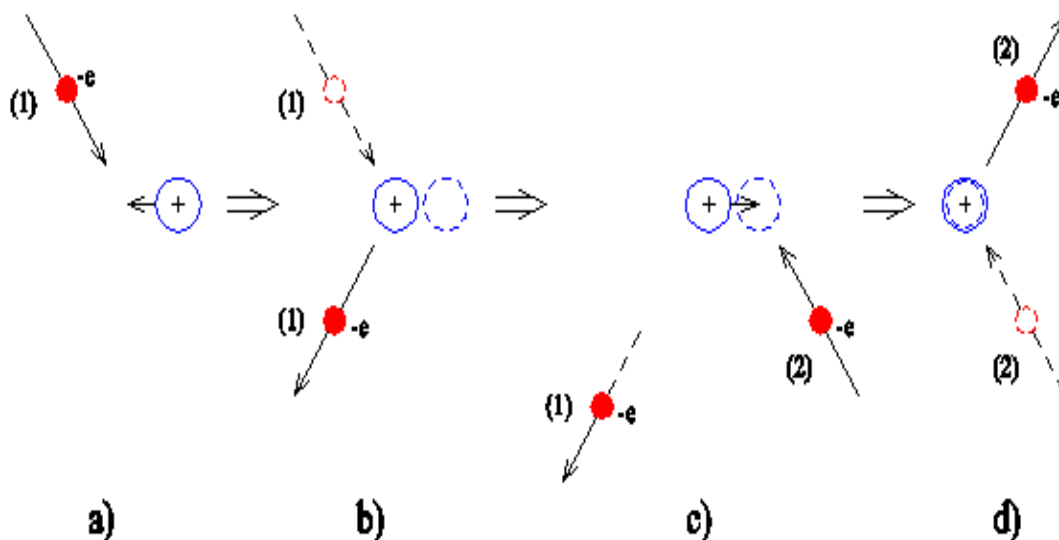
4) BCS– teori

Fra klassisk fysikk vet man at i en leder så kolliderer elektroner med hverandre hele tiden. Dette skaper resistans i lederen. Etter oppdagelsen av superledere fant forskerne ut at man ikke kan forklare superledende tilstander ut fra den klassiske modellen. Dette kommer av at det alltid vil være noen elektroner som kolliderer, ergo vil resistiviteten aldri bli lik null. Fenomenet kan heller ikke beskrives med den kvantefysiske modellen, der man ser på et enkelt elektron som en enkel bølge som går gjennom materialet. I 1957 ble BCS – teorien som forklarte fenomenet fullstendig på et mikroskopisk plan, publisert. Den blir ofte omtalt som den viktigste oppdagelsen innenfor teoretisk fysikk i det 20. århundre. Det er viktig å merke seg at denne teorien bare tilfredsstillende type I superledere som metaller og legeringer, og ikke type II superledere som består av forskjellige oksider.

Cooperpar

Det sentrale med BCS – teorien er at to elektroner i superlederen går sammen og danner et cooperpar. Intuitivt så motstrider dette coulombvekselvirkningen, som sier at 2 elektroner som har samme ladning, vil frastøte hverandre. Cooper fant ut at elektronparene likevel ikke stred i mot coulombvekselvirkningen.

Elektronene befinner seg inne i et metall som kan ses på som et gitter av positive ioner, og dermed er det andre krefter som påvirker elektronene. Det var en grunnleggende ustabilitet i metallet siden det var forskjellige krefter som motarbeidet hverandre. Tegningen på neste side illustrerer den lille, men netto tiltrekningen mellom to elektroner.



Figur 4.1: netto tiltrekning av 2 elektroner.

I figur a) ser vi elektron 1, som er på vei mot et positivt ion. Pga. motsatte ladninger vil begge partiklene gå mot hverandre. Figur b) viser systemet etter at elektronet har krasja inn i ionet og er på vei bort igjen. Ionet har flyttet seg litt bort i fra sin likevektsstilling. Pga. elektronets høye hastighet, ca. 100 000 m/s, vil ionet på denne tidsskalaen bevege seg veldig sakte mot likevektsstillingen sin. I det elektron 2 kommer (figur c)) med samme fart som elektron 1, men motsatt rettet, vil også dette elektronet kollidere med ionet og fortsette som vist i figur d). Elektronene har da kommet litt nærmere hverandre tilsvarende forflytningen av ionet. For å få en superleder, må alle elektronene oppføre seg slik som vist ovenfor. Det er faktisk akkurat det som skjer. To og to elektroner går sammen, og vi får "elektronvæsken" som gjør at strømmen går kontinuerlig. Det som også er viktig å legge merke til, er at elektronene i cooperparene har motsatt rettet impuls og spinn, men hvis alle parene er like, vil man få total impuls lik null, og dermed ingen total strøm.. Derfor må elektronparene være ulikt rettet for å få en superledende strøm. Elektron 1 og 2 må også selvsagt ikke komme på likt inn mot det positive ionet siden de da vil frastøte hverandre.

Spørsmålet man kan stille seg: Hvorfor dannes det plutselig cooperpar ved T_c ? BCS – teorien sier at når alle elektronene i en superleder har gått sammen i bundne par, er det nådd en grunntilstand. I denne grunntilstanden kan man se på superlederen som et kvantemekanisk system med ulike energinivåer, der alle cooperparene er låst i samme kvantetilstand. Hvis man skal endre impulsen på ett cooperpar, må man altså forandre impulsen på alle cooperparene. For å få strøm i superlederen, kan man sette en spenning over den, og strømmen vil gå kontinuerlig i all ”evighet” (målt til minst 100 000 år) fremover selv om man skrur av spenningen. Dette kommer som nevnt tidligere av at cooperparene er låst i samme kvantetilstand, og for å endre impulsen på et cooperpar må man endre impulsen på alle parene. For at strømmen da skal stoppe, må alle cooperparene stoppe opp samtidig, og det skjer ikke.

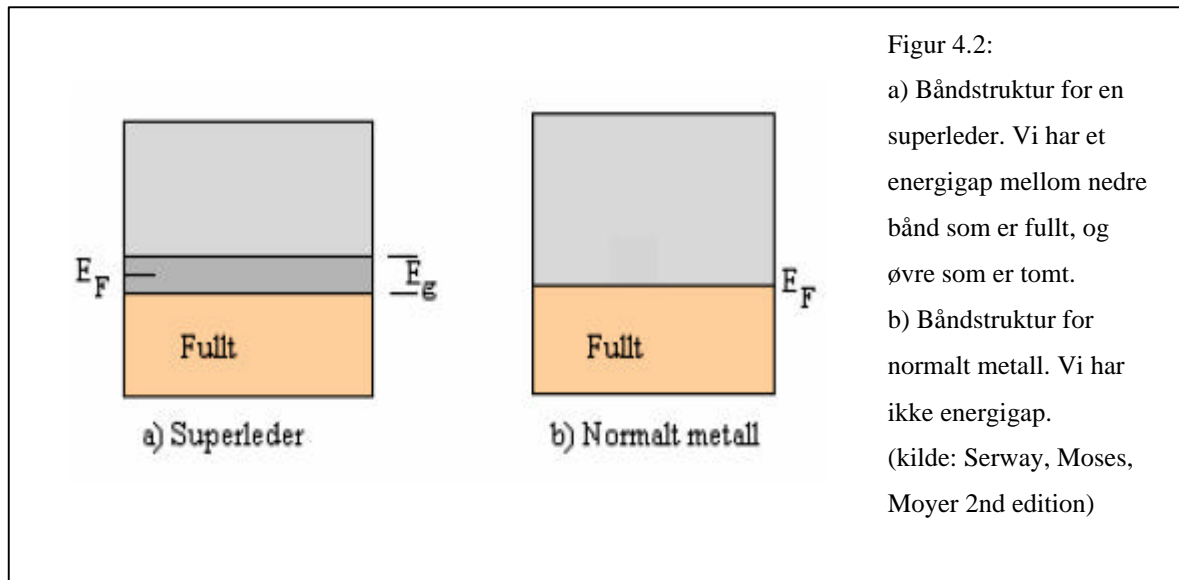
Koherenslengde

En annen viktig parameter forbundet med superledning er koherenslengden, λ , som er gjennomsnittsdistansen mellom elektronene i et Cooper par. Dette er altså avstanden for hvor langt unna elektronene i et par har kontakt med hverandre. Koherenslengden er den minste dimensjonen hvor den superledende tilstanden kan bli ødelagt.

En superleder er av type I om $\lambda < l$, og de fleste av de rene metallene faller under denne kategorien. En økning i forholdet $\frac{l}{\lambda}$ er mest gunstig for en type II superleder. En detaljert analyse viser at koherenslengden og penetrasjonsdybden begge er avhengig av den ”gjennomsnittlige frie veien” til elektronene i den normale tilstanden. Den ”gjennomsnittlige frie veien” til elektronene i et metall kan bli redusert ved å dope metallet, noe som forårsaker at penetrasjonsdybden øker, mens koherenslengden minker.

Energigap

I en superledende tilstand er stabiliteten avhengig av sterk korrelasjon mellom cooperparene. BCS – teorien forklarer superledende oppførsel ved hjelp av energinivåer, og sier at det finnes et energigap mellom øvre og nedre energibånd.



E_g er den nødvendige energien som skal til for å bryte opp et cooperpar. Energigapet oppstår pga. at det er en stabiliserende energi assosiert med elektronene i cooperparene. Gapet er altså et mål på hvor mye termisk energi som må til for å flytte elektronene fra superledende tilstand til eksitert tilstand. Energigapet (E_g) ved $T = 0$ K er i følge BCS – teorien $3,53 k_B T_c$, der k_B er boltzmans konstant.

Superledere med høy T_c vil ha store energigap, men størrelsen på energigapet avhenger også av det påtrykte magnetfeltet. Et magnetfelt påvirker elektronene i cooperparene. Hvis magnetfeltet er kraftig nok, vil det kunne bryte opp parene pga. at elektronene har motsatt spin. Da blir superlederen ødelagt, og vil da bli et vanlig metall igjen. Akkurat den størrelsen på magnetfeltet som bryter cooperparene kalles det kritiske magnetfeltet B_c .

Fermienergien E_f er den største kinetiske energien som et fritt elektron kan ha ved $T = 0$ K i et normalt metall. E_f er av størrelsesorden 5 eV, og E_g i en superleder er på $\sim 10^{-3}$ eV. Altså energigapet i en superleder er veldig lite i forhold til fermienergien til et metall.

Det var faktisk nordmannen Ivar Giaever som klarte å bevise dette energigapet

som BCS – teorien forklarte, gjennom et tunnelingsforsøk. Gapet ville hindre elektronene i å tunnelere i et visst spenningsområde som tilsvarer størrelsen på gapet. Gapet vil dermed sette igjen en tydelig "signatur" i måledataene. Dette fikk Giaever nobelprisen i fysikk for. Han er den eneste nordmannen som har fått denne prestisjetunge prisen.



Figur 4.3: Ivar Giaever med glassklokka der han la grunnlaget for nobelprisarbeidet.
(kilde: Fysikkåret 2005 portrett Ivar Giaever av Kristian Fossheim)

Kilder:

- Serway, Moses og Moyer; "Modern Physics", 1997
- www.superconductors.org

5) Superleder type II

I løpet av 1960-åra ble forskere klar over en ny type superleder, en superleder de kalte superleder type II.

Superleder type II har to kritiske magnetfelt, B_{c1} og B_{c2} . Oppførselen til en type II superleder er som følger, for ulike størrelser av det påtrykte magnetiske feltet (B);

- $B < B_{c1}$; som for type I superleder, men B_{c1} er lavere enn B_c for type I superleder.
- $B > B_{c2}$, ikke superledende, og vanligvis ikke ledende i det hele tatt. For type II superledere består stort sett av keramer og andre materialer som ikke er ledende når den kritiske strømverdien (I_c), den kritiske temperaturen (T_c) eller det kritiske magnetfeltet (B_{c2}) overskrides (se fig. 5.2).
- $B_{c2} > B > B_{c1}$, her får vi noe som blir kalt en virvelfase. B_{c2} er høyere enn B_c for type I superledere. Det er denne virvelfasen som er unik for type II superledere. Det er denne tilstanden hos en type II superleder vi vil beskrive egenskapene til i dette kapitlet.

Virvelfase

Når en superleder befinner seg i en såkalt virvelfase, vil det være energetisk gunstig å slippe gjennom magnetiske flukslinjer. Der hvor disse flukslinjene går gjennom vil materialet ikke være superledende, men rundt disse er materialet superledende, noe som gjør at det oppstår virvler av superstrøm rundt de magnetiske flukslinjene (fig. 5.1). De magnetiske flukslinjene kan sees på som "tuber" av magnetisk fluks som går gjennom superlederen. Strømvirvlene blir ofte omtalt som "abrikosovvirvlene", etter den sovjetiske fysikeren A.A. Abrikosov som teoretisk forutsa disse i 1957.

Når vi øker det påtrykte feltet, øker antall "tuber", og til slutt overlapper de hverandre, og vi får ikke lenger en superleder. Dette skjer ved B_{c2} .

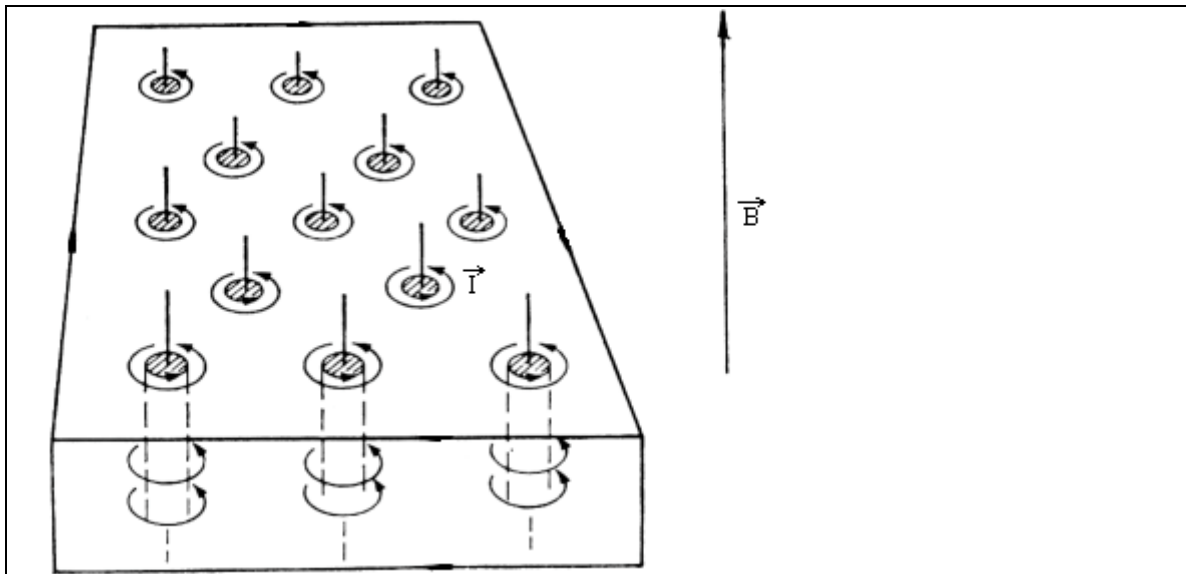


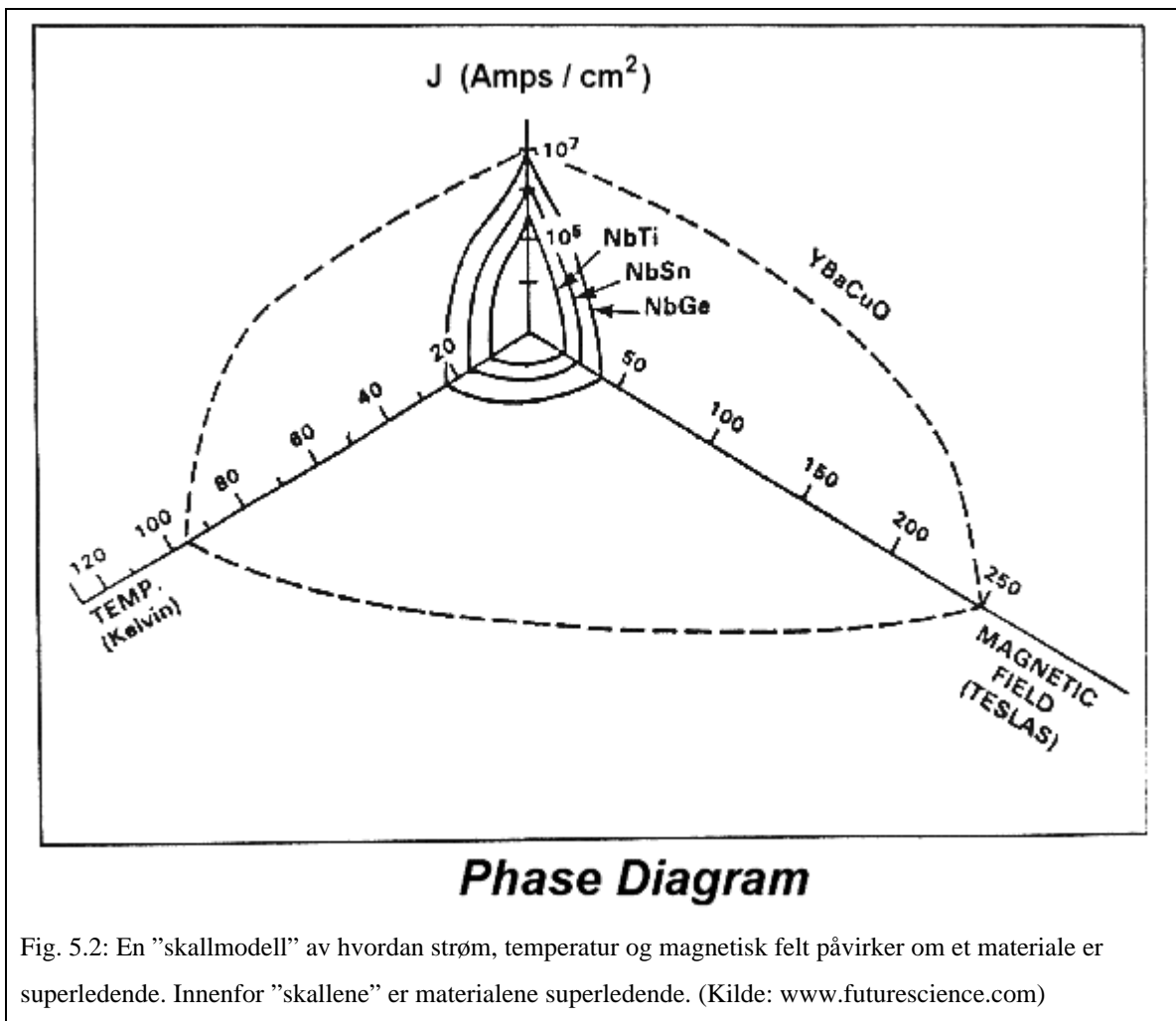
Fig. 5.1: Magnetisk fluks går gjennom "hull" eller "tuber" i superlederen (steder hvor superlederen ikke er superleder). Rundt disse magnetiske flukslinjene går det superstrømmer. (Kilde: http://lanoswww.epfl.ch/studinfo/courses/cours_supra/TypeI-II/default.htm)

I tabell 5.1 er en oversikt over kritisk temperatur hos noen type II superledere.

Superleder:	Kritisk temperatur:
$(\text{Hg}_{0.8}\text{Tl}_{0.2})\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{8.33}$	138 K* (record-holder)
$\text{HgBa}_2\text{Ca}_2\text{Cu}_3\text{O}_8$	133-135 K
$\text{HgBa}_2\text{Ca}_3\text{Cu}_4\text{O}_{10+}$	125-126 K
$\text{HgBa}_2(\text{Ca}_{1-x}\text{Sr}_x)\text{Cu}_2\text{O}_{6+}$	123-125 K
$\text{HgBa}_2\text{CuO}_{4+}$	94-98 K

Tabell 5.1: En oversikt over kritisk temperatur hos noen type II superledere. (Kilde: <http://www.superconductors.org/Type2.htm>)

Ved hjelp av en "skallmodell" (fig. 5.2) kan vi også få et bilde av sammenhengen mellom kritisk strøm, magnetfelt og temperatur. Inni skallene er materialene superledende.



Magnetiseringen (M) til en typisk type II superleder i forhold til det påtrykte feltet (B_o) er vist i figur 5.3.

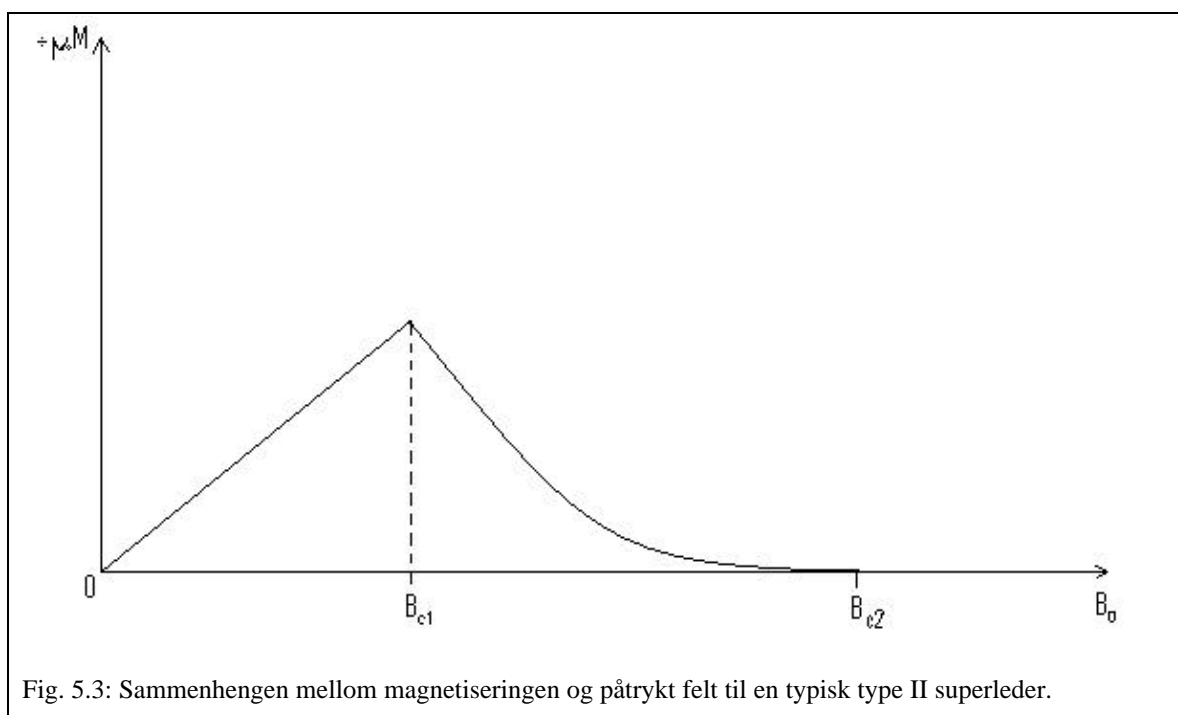
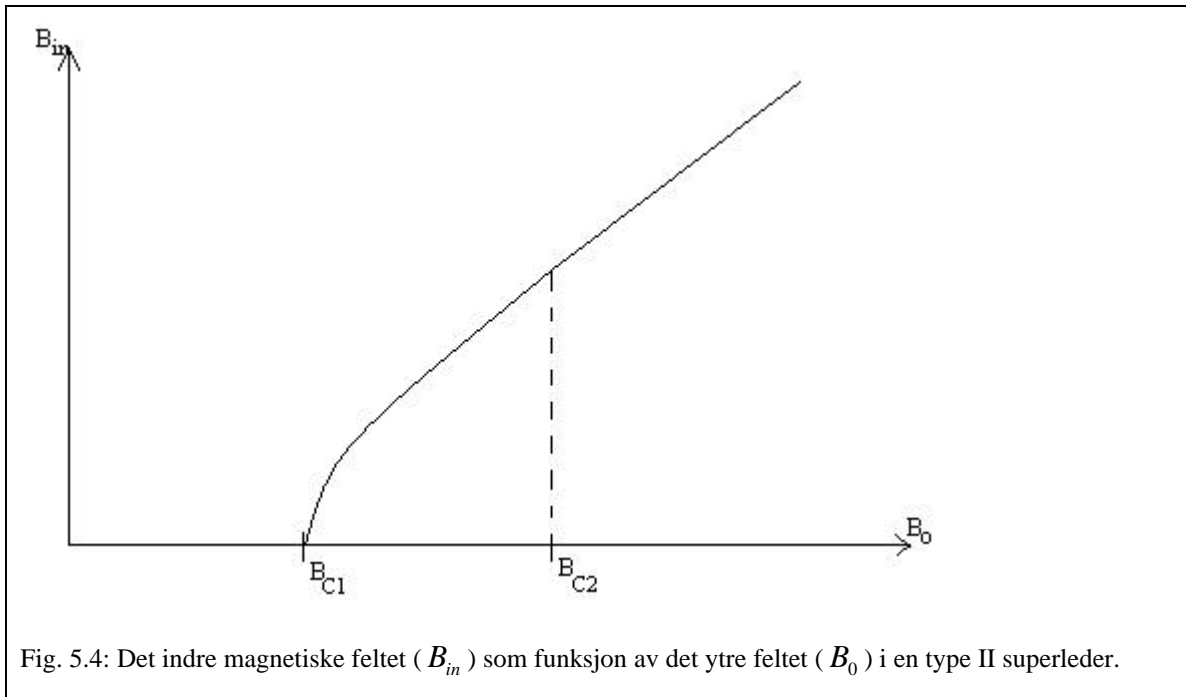


Fig. 5.3: Sammenhengen mellom magnetiseringen og påtrykt felt til en typisk type II superleder.

I figuren ser vi at magnetiseringen vil øke fram til B_{c1} og avta derfra og fram til B_{c2} , hvor den er svak negativ. De fleste type II superledere vil følge denne grafen. Det at materialer som er en type II superleder gir en svak svekkelse av det påtrykte feltet når den høyeste kritiske feltverdien (B_{c2}) overskrides, må bety at disse materialene er diamagnetiske over denne kritiske feltverdien. Dette gjelder selvsagt også over den kritiske temperatur og strømverdien, altså utenfor "skallet" i fig. 5.2.

Når magnetiseringen er null, slik den er når $B_0 > B_{c2}$, i superlederen, vil det påtrykte/ytre feltet være det samme som inni. Vi kan her tenke oss at "tubene" med magnetisk fluks, som vi har omtalt tidligere, overlapper hverandre. All magnetisk fluks går altså gjennom. Men når det påtrykte feltet er mellom B_{c1} og B_{c2} ser vi at magnetiseringen gradvis øker når vi går fra B_{c2} , dette er fordi færre og færre "tuber" med magnetisk fluks går inn i superlederen. Ved $B_0 = B_{c1}$ vil ingen "tuber" trenge inn i superlederen. Dermed øker magnetiseringen proporsjonalt med det påtrykte feltet, slik vi har sagt i kapittelet om superleder type I. Denne sammenhengen mellom påtrykt felt (B_0) og indre felt (B_{in}) i en type II superleder er vist i figur 5.4.



Pinning

Når en type II superleder er i virveltilstand kan en lorentzkraft føre til frastøting mellom tubene (likning 5.1). I figur 5.5 ser vi dette når vi sender en strøm gjennom en superleder, men også når det ikke sendes strøm gjennom vil vi ha en lorentzkraft. Denne kommer av virvlestrømmene.

$$\vec{F} = e\vec{v} \times \vec{B} \qquad \text{likning 5.1}$$

Denne kraften fører til en bevegelse, som fører til fluksendring i tiden, som igjen fører til motstand. Denne motstanden er meget liten sammenlignet med motstanden i en vanlig leder, men den er uønsket uansett, for vi vil ikke ha noe motstand i en superleder!

Ved å tilsette urenheter i superledermaterialet på en bestemt måte, kan man imidlertid låse, eller ”pinne”, tubene fast. Vi får da null resistans helt til strømmen blir så sterk at lorentzkraften blir for sterk for pinningen.

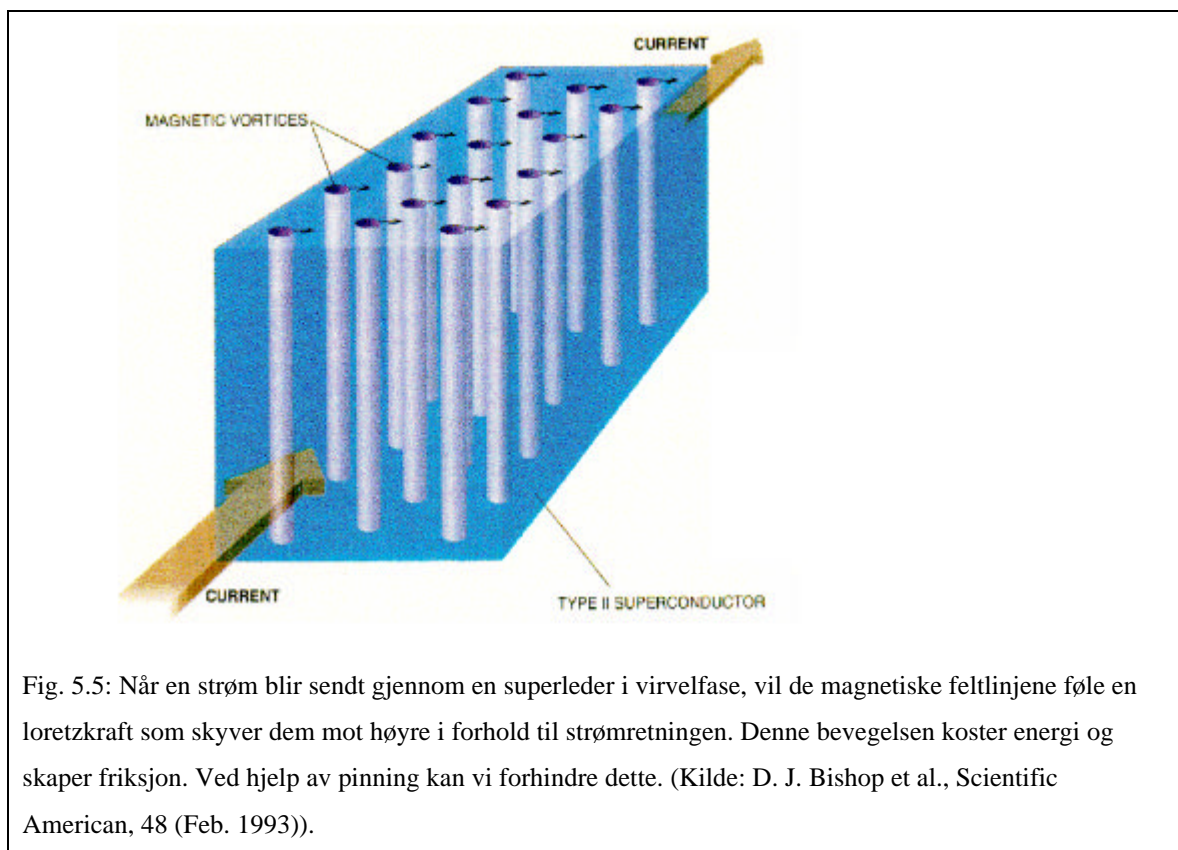


Fig. 5.5: Når en strøm blir sendt gjennom en superleder i virvelfase, vil de magnetiske feltlinjene føle en loretzkraft som skyver dem mot høyre i forhold til strømretningen. Denne bevegelsen koster energi og skaper friksjon. Ved hjelp av pinning kan vi forhindre dette. (Kilde: D. J. Bishop et al., Scientific American, 48 (Feb. 1993)).

Det forskes mye på hvordan man kan pinne tubene best mulig i forskjellige superledermaterialer. Dette er selvsagt et meget viktig tema innenfor superledersforskningen, men dessverre altfor komplisert til at vi kan gå nærmere inn på det her. Vi kan nevne at den mikroskopiske strukturen til materialet er meget viktig for hvor sterk den kritiske strømmen kan være.

Høytemperatur superledere

Dette er en "sjanger" innenfor type II superledere.

Lenge var Nb_3Ge den superlederen man kjente til med høyest kritisk temperatur, med sine 23,2 K. Men i 1986 kom et stort gjennombrudd innenfor superledningsforskningen. J. George Bednorz og Karl Alex Müller fant ut at en sammensetting av lantan, barium, kobber og oksygen ($La_{2-x}Ba_xCuO_4$), som er en isolator (keramisk) ved romtemperatur, var superledende ved 30 K. Samme år fant forskere ut at ved å bytte ut barium med strontium kunne en får en kritisk temperatur på 36 K. Og året etter fant en

forskningsgruppe fra USA ut at en sammensetting av yttrium, barium, kobber og oksygen ($YBa_2Cu_3O_{7-x}$) kunne bli superledende ved 96 K! Det helt store med dette er at det er over kokepunktet til nitrogen (77K), vi kan altså benytte oss av flytende nitrogen istedenfor flytende helium som kjølingsmiddel. Flytende nitrogen er mye billigere og lettere å håndtere enn flytende helium. Dette var noe vi dro nytte av da vi så på pinning til en superleder av samme sammensetting som forskningsgruppen fra USA kom fram til i 1987.

Gjennombruddet som Bednorz og Müller skapte i 1986 skapte en helt ny æra innenfor høytemperatursuperledere, og da det den 8. mars 1987 skulle holdes en konferanse om dette i New York var interessen mildt sagt enorm! Over 3000 mennesker møtte fram, hendelsen blir omtalt som "The Woodstock of Physics", og en forskningseksplisjon innenfor temaet startet rundt om i verden. I skrivende stund er rekorden på rundt 138K, men det vil ikke sjokkere noen om den snart blir brutt, for dette er et populært og viktig forskningsfelt innenfor moderne fysikk.

I figur 5.7 er "evolusjonen" til høytemperatursuperledere vist. Legg merke til den enorme utviklingen som har skjedd siden 1986! Det er vanlig å klassifisere superledninger inn i tre grupper. Disse gruppene er illustrert i figur 5.8.

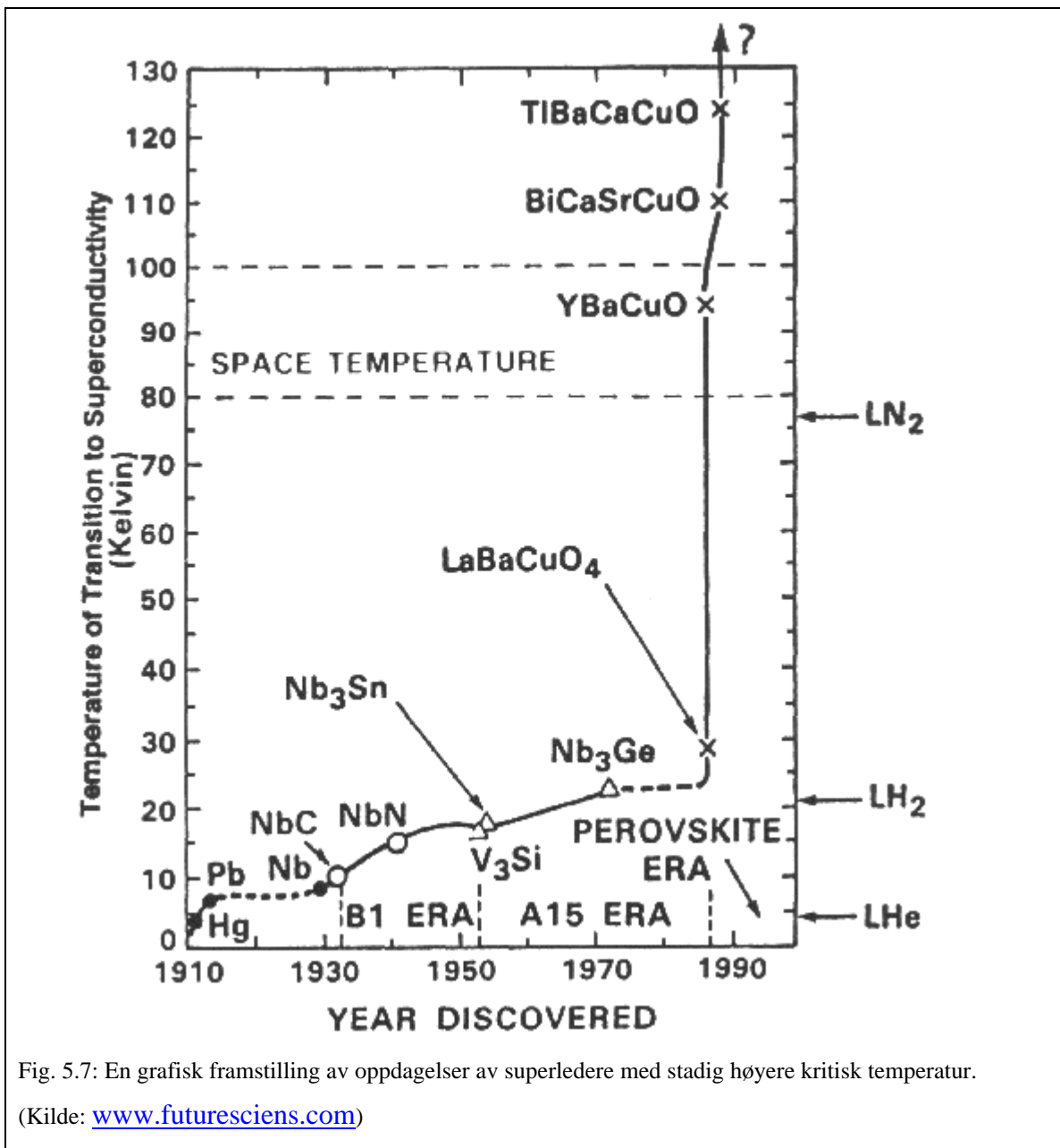


Fig. 5.7: En grafisk framstilling av oppdagelser av superledere med stadig høyere kritisk temperatur.

(Kilde: www.futuresciens.com)

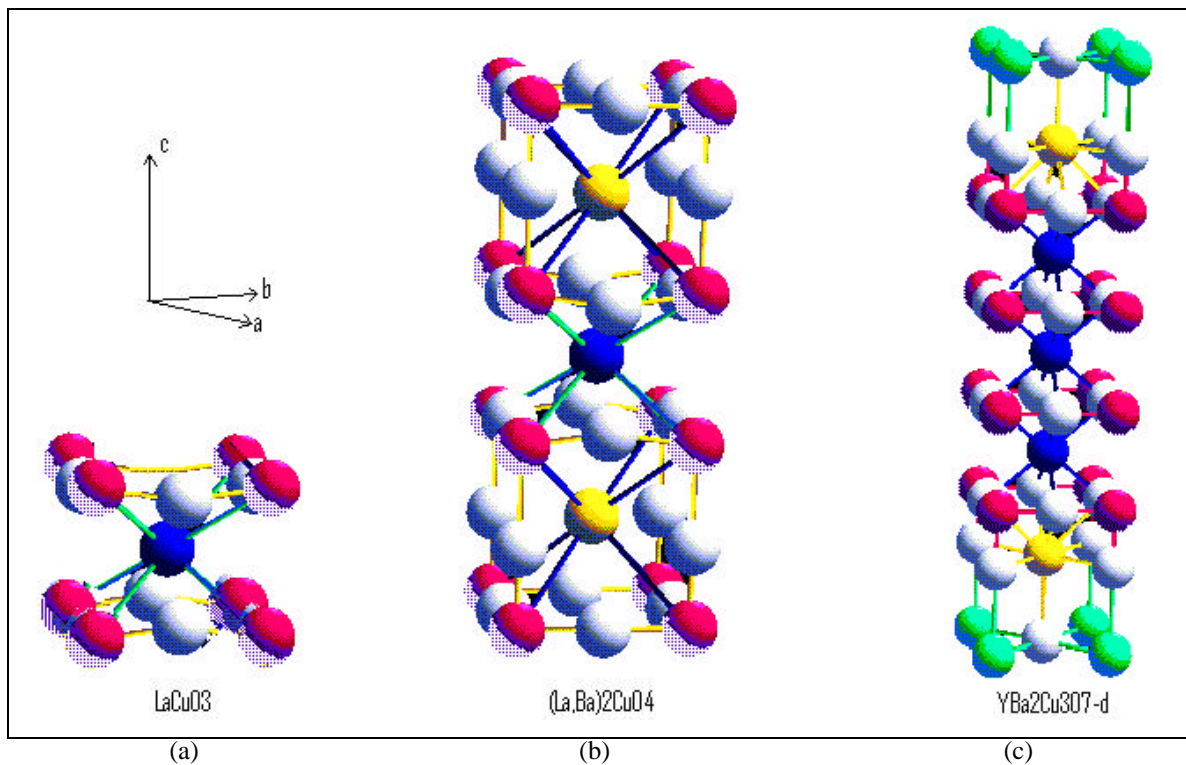


Fig. 5.8: De tre gruppene det er vanlig å dele superledninger opp i. Det er illustrert eksempler på a) fundamental perovskite, b) enlags perovskite og c) dobbellags perovskite.

Den første gruppa består av romsymmetriske strukturer, slik som for eksempel $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$, som er en av de første ”høytemperatursuperlederne” (med en kritisk temperatur på 10 K). I den andre gruppa finner vi strukturer som er symmetriske i a-b-planet. I denne gruppa finner vi blant annet $\text{La}_2\text{Ba}_2\text{CuO}_4$ som har en kritisk temperatur på 38 K. Den tredje gruppa blir omtalt som en mangelagsstruktur. Strukturen til molekylende i denne gruppa er ikke symmetrisk i a, b eller c-planet. Her finner vi $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ som vi har snakket om tidligere. Det er viktig å være klar over at x-en er meget avgjørende for i hvilken grad materialet er en superleder. Er x null er det nærmest ikke snakk om noe superleder i det hele tatt.

I mer kompliserte kobberstrukturer enn de som er illustrert i figur 5.8 kan en se en direkte sammenheng mellom antall kobberoksygenlag og den kritiske temperaturen. Jo flere kobberoksygenlag vi har, jo høyere blir den kritiske temperaturen. På bakgrunn av dette er det grunn til å tro at vi snart kan klare å sette sammen superledere med høyere

kritisk temperatur. Og vi kan jo drømme om, og kanskje en dag få, kritiske temperaturer over romtemperatur!

Den kritiske strømstyrken i den sistnevnte superlederstrukturen er mye høyere gjennom kobberoksygenlagene enn det den er vinkelrett på disse. Vi kan altså så på ledningsevnen som 2-dimensjonal.

Kilder:

- Serway, Moses og Moyer; "Modern Physics", 1997
- <http://online.itp.ucsb.edu/online/lnotes/balents/node18.html>
- <http://physics.ucsd.edu/~jorge/hole.html>
- <http://www.math.psu.edu/qdu/Res/Pic/gallery1.html>

6) Flukskvantisering

Hvis vi har en superledende ring som i fig. 5.6 og setter på et magnetisk felt, for så å slå det av igjen, har vi fanget flukslinjene som går gjennom hullet! Dette skjer fordi magnetiske materialer ikke er energetisk tjent med fluksforandring, så ringen ”vil” beholde flukslinjene som går gjennom når vi har et påtrykt felt, og det kan den ved at det oppstår en strøm (I) som vist på figur 5.6. Ettersom det ikke er motstand i ringen, vil denne strømmen teoretisk fortsette i evig tid! Det kan nevnes at S.S. Collins gjorde dette forsøket, men etter to og et halvt år måtte han stoppe. Det ble trailerstreik, så da fikk han ikke mer flytende helium til å holde superlederen avkjølt. Målinger fra dette og andre forsøk viser at vi antageligvis har absolutt null motstand i en superleder, og selv etter at alle feilmarginene er tatt med, er resistansen så liten at den for alle praktiske formål kan neglisjeres (ca. $10^{-26} \Omega m$).

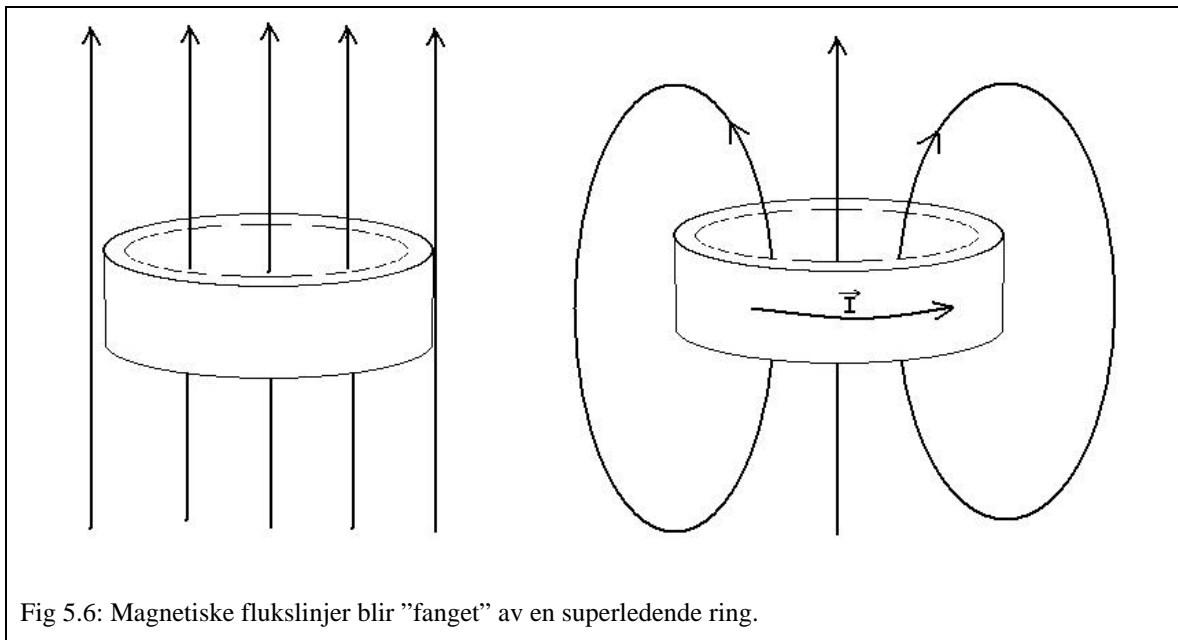


Fig 5.6: Magnetiske flukslinjer blir ”fanget” av en superledende ring.

Fritz London mente at denne fangede fluksen måtte være et fundamentalt kvantefenomen, på samme måte vi har snakket om at spinn er det. Han mente den

magnetiske fluksen måtte være kvantisert i enheter av h/e . Elementærladningen er plassert i nevneren fordi han mente strømmen består av enkle elektroner som ladningsbærere.

Gjennom forsøk viste det seg imidlertid at fluksen er gitt ved halvparten av det London mente (likning 5.2). Dette stemmer jo bra med BSC-teorien som sier at vi har to elektroner som ladningsbærere.

$$\Phi = \frac{nh}{2e} = n\Phi_0 \quad \text{likning 5.2}$$

I likningen er n et positivt heltall og $\Phi_0 = \frac{h}{2e} = 2,0679 \cdot 10^{-15} \text{ Tm}^2$ den magnetiske flukskvantiseringen.

Kilder:

- Egil Lillestøl, Ola Hunderi og Jan R. Lien; "Generell fysikk – bind 2"
- <http://www.apollon.uio.no/vis/art/2003/2/superled>

7) Demonstrasjon av pinning og meisnereffekt

Vi fikk en demonstrasjon av et meget spennende fenomen, som kan oppstå takket være meisnereffekt og pinning. Vi fikk se en vanlig magnet sveve over en superleder. Dette er tilsynelatende ikke noe ekstraordinært, for man kan se noe lignende hvis man plasserer en magnet med syd-/nordpolen ned over en magnet med nord-/sydpolen opp, men det er en meget ustabil tilstand. Det vil i praksis si at et en bitte liten kraft på en av magnetene vil føre til at den svevende magneten vil falle ned på bakken. Men i demonstrasjonen vi fikk se var magneten låst fast i en posisjon over superlederen! Det vil si at det måtte meget store krefter til for å fjerne magneten permanent ut av sin posisjon over superlederen (fig. 1a). Og hvis vi hadde løftet superlederen opp og holdt den på skrå, opp ned (fig. 1b) eller på hvilken som helst annen måte, ville magneten fortsatt vært i samme posisjon i forhold til superlederen! (Dette forutsetter selvsagt at superlederen er under sin kritiske temperatur, så det er litt vanskelig å få til i praksis).

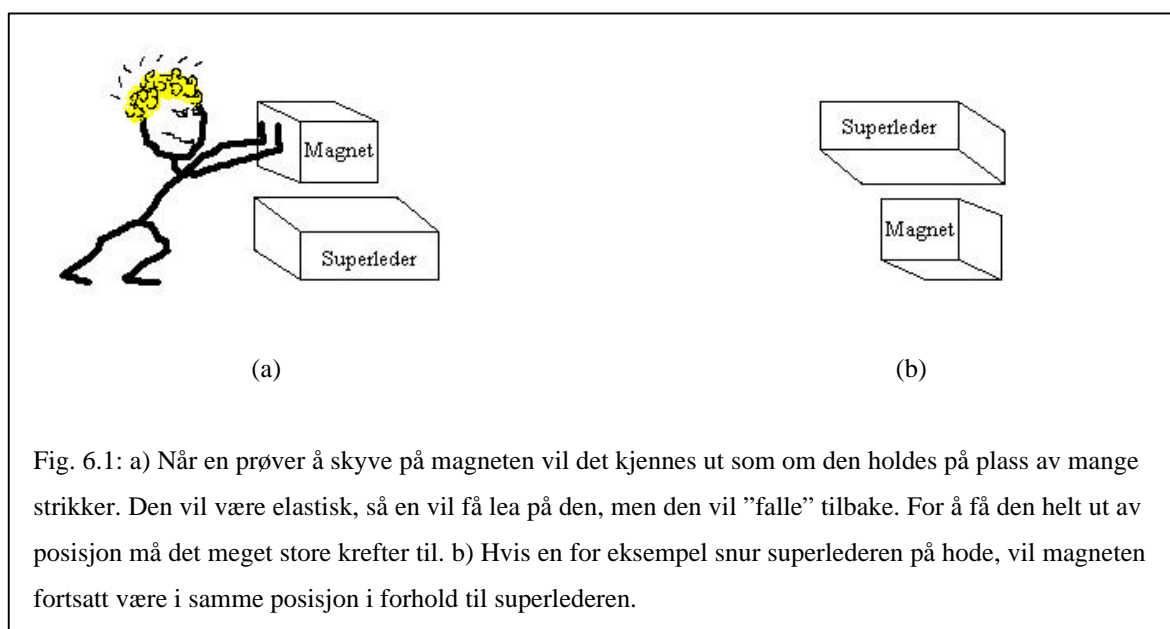


Fig. 6.1: a) Når en prøver å skyve på magneten vil det kjennes ut som om den holdes på plass av mange strikker. Den vil være elastisk, så en vil få lea på den, men den vil "falle" tilbake. For å få den helt ut av posisjon må det meget store krefter til. b) Hvis en for eksempel snur superlederen på hode, vil magneten fortsatt være i samme posisjon i forhold til superlederen.

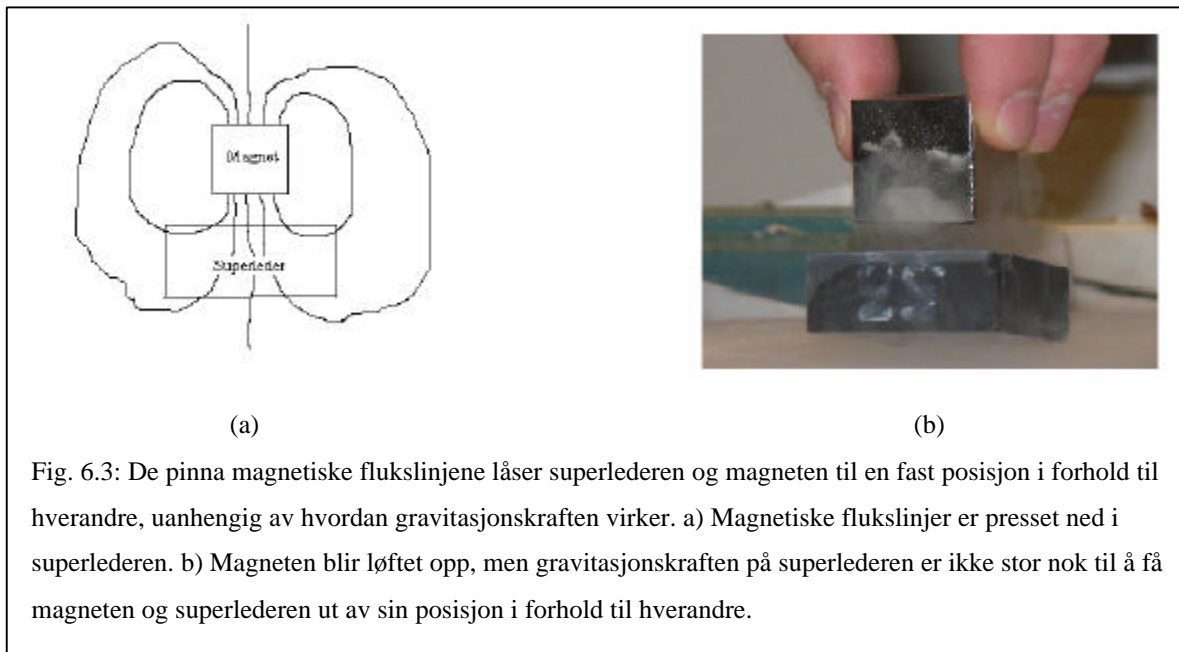
"Magi!", vil nok mange si, men en bedre forklaring skal vi gi.

Vi plasserte en type II superleder, som var en sammensetning av $YBa_2Cu_3O_{7-x}$ (keramisk), i en papptallerken. Ved å helle flytende nitrogen oppi tallerkenen kom superlederen under sin kritiske temperatur, og keramet ble superledende.

Når vi slapp små magneter oppå superlederen føyk de til siden før de kom ned til superlederen. Dette gjorde de fordi superlederen oppretter et usynlig "skjold" mot magneter, noe som er forklart bedre under avsnittet om meissnereffekt. Men når vi presset en magnet ned mot superlederen ble den låst fast ca. to centimeter over ledere. Dette skjer fordi vi har med en superleder av type II å gjøre, og når en plasserer en magnet i nærheten av en type II superleder vil en kunne presse magnetiske flukslinjer gjennom superlederen fra magneten. Hvis magnetlinjene er pinna i superlederen, som de var i vårt tilfelle, vil magneten være låst fast i et magnetisk "gitter". Dette magnetiske "gitteret" vil være der selv om vi fjerner magneten igjen, for som vi har sagt tidligere er ikke superlederen energetisk tjent på fluksforandring.

Når magneter kom i nærheten av superlederen ble det opprettet et ytre magnetfelt slik at feltet inni superlederen ble null (meissnereffekt). Dette så vi da vi slapp de små magnetene ned mot superlederen. Men siden det var en type II superleder var det som sagt mulig å presse magnetiske flukslinjer inn, og når slike flukslinjer først er inne, vil magneten vi har presset flukslinjene inn med være låst i en posisjon. Dette kommer av, som vi har sagt før, at superlederen ikke er energetisk tjent med fluksforandring.

Siden den magnetiske kraften i en type II superleder enten kan virke tiltrekkende eller frastøtende, vil pinningen føre til at magneten er låst i samme posisjon i forhold til superlederen selv om vi for eksempel snur superlederen opp ned (fig. 6.1b) eller dytter på magneten. De magnetiske kreftene vil hele tiden tilpasse seg slik at de opphever gravitasjonskreftene slik at det ikke blir magnetisk fluksforandring i superlederen (forutsatt at gravitasjonskreftene ikke er større en den magnetiske kraften superlederen klarer å lage). Dette kan skje fordi det er null motstand i ledere, dermed vil det ikke koste energi å øke eller minke det magnetiske feltet.



Da nitrogenet i papptallerkenen var fordampet ble superlederen varmere enn sin kritiske temperatur, og den var ikke lenger superledende. Vi plasserte da en isolator oppå superlederen, og en magnet oppå der igjen (fig. 6.4). Deretter helte vi flytende nitrogen oppi og ventet på at superlederen vår skulle komme under sin kritiske temperaturen igjen.

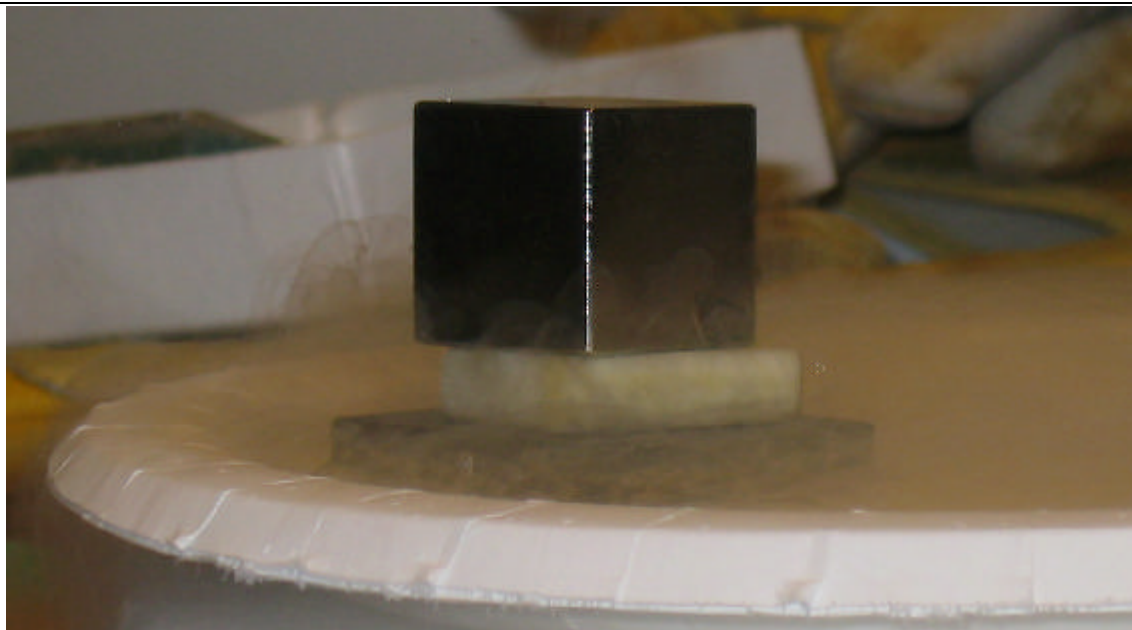


Fig. 6.4: Superlederen med en isolator mellom seg og magneten.

Da superlederen var kald nok fjernet vi isolatoren og magneten vår svedde igjen. Vi fikk nå omtrent samme situasjon som da vi først frøs ned superlederen, for så å plassere magneten over. Den eneste forskjellen er at vi nå hadde ”fryst inn” de magnetiske feltlinjene.

Før superlederen kom under den kritiske temperaturen var den en isolator. Da gikk alle de magnetiske feltlinjene fra magneten rett gjennom superlederen. Når vi så fryser ned lederen mens magnetlinjene går gjennom vil disse ”fryses fast”, og der vi ikke har magnetiske feltlinjer blir materialet superledende, som gjør at vi får meissnereffekt. Intuitivt skulle en tro det er det samme om en ”fryser fast” feltlinjene eller presser de inn etterpå slik vi gjorde først. Men i praksis klarer vi ikke å presse inn like mange feltlinjer som vi klarer å fryse fast. Dette gjør at posisjonen er mer stabil når feltlinjene er ”fryst fast” enn når de er presset inn.

Når magneten svedde over superlederen, som i de to tilfellene over, ga vi den et puff slik at den spant rundt. Den spant da nærmest friksjonsfritt rundt. Eneste friksjonen som virket var en liten forandring av magnetisk fluks samt luftmotstand. Hadde vi med andre

ord hatt en perfekt kule kunne vi i vakuum fått den til å spinne rundt bokstavelig talt i evigheter.

Kilder:

- www.superconductors.org
- <http://www.fys.uio.no/super/levitation/>
- <http://my.execpc.com/~rhadley/magsuper.htm>
- <http://physics.ucsd.edu/~jorge/hole.html>
- Egil Lillestøl, Ola Hunderi og Jan R. Lien; ”Generell fysikk – bind 2”, 2001
- <http://www.fys.uio.no/super/mo/#meissner>

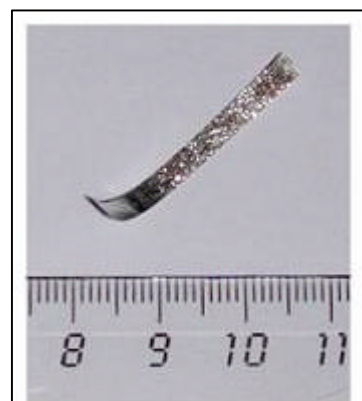
8) Anvendelser

Det er som sagt forsket i årevis på superledere. Har man kommet så langt som å kunne bruke dette fenomenet til noe nyttig? Svaret på det spørsmålet er definitivt ja! Det at man kan føre strøm gjennom en superleder uten energitap har stor praktisk og ikke minst økonomisk betydning. Superlederen er som nevnt tidligere også en sterk og stabil elektromagnet. Disse to egenskapene drar man nytte av i dagens teknologi, og det brukes til mange forskjellige formål. Vi skal se litt nærmere på disse anvendelsene:

1. Superledende strømkabler
2. Elektriske supermotorer
3. Levitasjon
4. Måling av hjernebølger

Superledende strømkabler

I praktiske anvendelser er det naturlig å bruke høytemperatursuperledere, siden man da kan bruke flytende nitrogen til å kjøle ned superlederen. Det materialet som har vist seg best egnet som superleder i et strømførende bånd, kalles BSCCO. Dette materiale blir superledende ved ca. 110 K. Strømførende bånd lages i dag med en metode som kalles "pulver i rør". Et pulver av små krystallkorn med BSCCO fylles i et sølvrør og settes under trykk. Deretter dras røret gjennom et hull, slik at det blir mer enn ti ganger så langt. Mange slike rør plasseres inne i et nytt rør, som på nytt dras ut i lengden. Etter dette vales røret til et tynt bånd. I figur 8.1: kan man se et slikt bånd. I teorien vil en strømkabel lagd av slike bånd kunne føre ca. 140 ganger mer strøm enn en kobber ledning med samme dimensjon. Når man regner med kjøling av superlederen vil strømkabelen kunne transportere ca. 3-5 ganger så mye strøm som kobberledningen.



Figur 8.1: Et bånd av BSCCO superleder (kilde: www.fysikknett.no)

Elektriske supermotorer

Elektriske motorer er basert på elektromagneter. Magnetfeltet gjør slik at aksene inne i motoren begynner å snurre rundt. Superledning gir da en mulighet til å øke effekten, og dermed få høyere motorytelse. Superledende motorer kan bli nesten 3 ganger mindre i fysisk størrelse og gi samme effekt.



Figur 8.2: En motor basert på superlederteknologi.
(Kilde: www.fysikknett.no)

I dag er det vanlig med elektrisk drevne båtmotorer som er veldig store. Dette er et problem når man skal reparere motoren. Hvis vi kan bytte ut motoren med en motor som er 3 ganger mindre så vil det ha store fordeler. Som visst i figur 8.3. er motoren satt på utsiden av båten, og vil da være lett å flytte dem uten at hele båten må i dokk.



Figur 8.3: Små
(super)elektromotorer på undersiden
av et skip

En annen fordel er jo at man bruker flytende nitrogen til nedkjøling av superlederen. Nå trenger man ikke olje, og dermed er motoren mye mer miljøvennlig

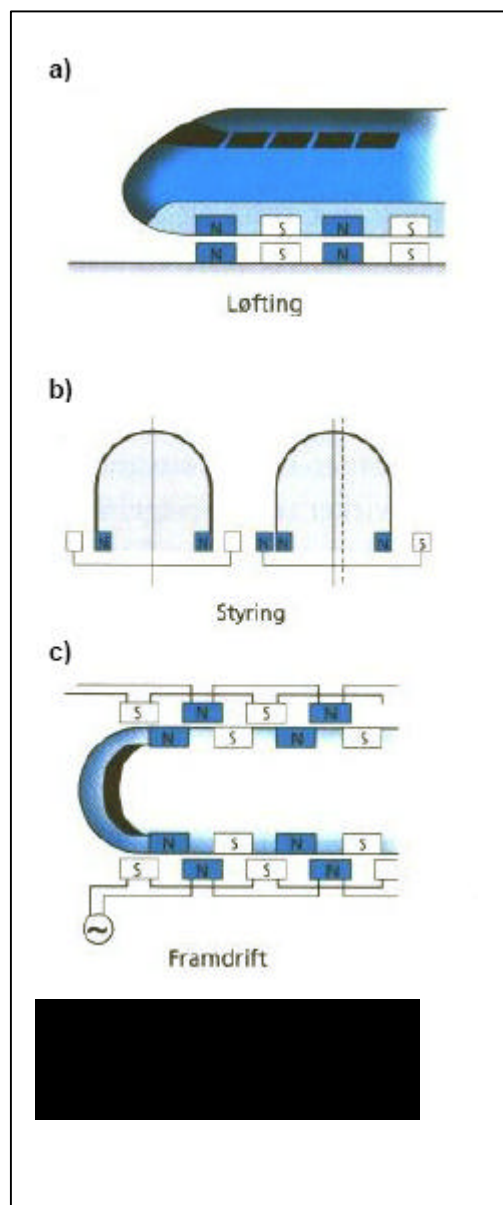
Levitasjon

Å få ting til å sveve har vel vært et ganske omdiskutert tema opp gjennom tidene. Det har alltid kostet mye energi, og dermed penger å få gjenstander til å fly. Etter superlederens magnetiske egenskaper ble påvist ble det nye spekulasjoner rundt emnet levitasjon. Japanerne klarte først å bygge et vellykket svevetog.

På figur 8.4 a) ser vi at det er plassert mange superledende elektromagneter. I skinnegangen er det plassert spoler med jernkjerner. Dette skaper et veldig stort og stabilt magnetisk felt som holder toget på plass i en bestemt posisjon over bakken hele tiden. I figur 8.4 b) er det illustrert hvordan styremekanismen til toget er.

Hvis toget kommer litt for langt til siden i banen, vil magnetiske krefter bringe det tilbake på plass igjen, slik som forklart i pinning teorien.

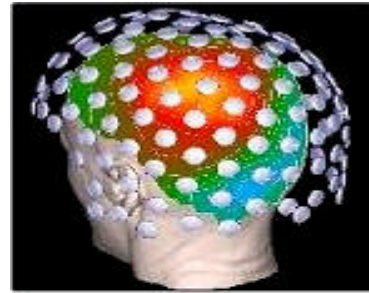
I figur 8.4 c) vises det hvordan man får toget til å gå fremover. Langs sidene av skinnegangen er det plassert elektromagneter der en kan synkronisere polariteten med framdriften av toget. På denne måten kan en hele tida oppnå en situasjon der magnetene på toget blir tiltrukket av skinnegangmagnetene som er rett foran, og frastøt av skinnegangmagnetene som er rett bak dem. Takten i polskiftene til skinnegangmagnetene regulerer farten til toget. Ved å gjøre synkroniseringen dårligere kan en bremse toget.



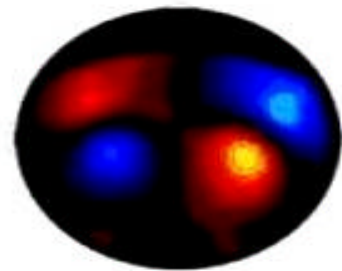
Måling av hjernebølger

Et måleinstrument som kalles SQUID kan måle magnetfelt nøyaktig helt ned til $T = 1 \text{ f T}$ som er mer enn 1 milliard svakere enn jorda sitt magnetfelt. Kroppen vår skaper magnetfelt hele tiden på mange forskjellige måter. Disse magnetfeltene er som regel sterke nok til at en SQUID kan oppdage dem. Dermed kan man kartlegge og måle i detalj en hver funksjon av kroppen vår. Det er som regel bare hjertet og hjernen som er interessante og viktige kilder til magnetfelt i kroppen vår. Det er faktisk mulig å måle hjerneaktiviteten ved hjelp av å feste mange SQUIDer på en "hette" som vist på bilde 1 i figur 8.5. På bilde 2 ser man magnetfelt som dannes når en person hører et lydsignal. Rød farge viser magnetfelt som går oppover, og blå farge magnetfelt som går nedover. I bilde 3 ser man hvilke deler av hjernen som er aktiv når man hører et lydsignal.

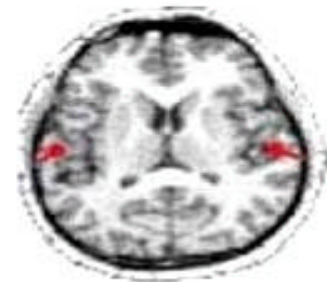
Slik som dette kan forskere finne de forskjellige sentrene i hjernen. Når du ser et bilde, bruker du et område i hjernen, et annet når du hører en lyd, og et annet når du beveger foten din etc. Med denne vitenskapen kan man f.eks. finne ut nøyaktig hvor i hjernen et epileptisk område befinner seg, slik at en hjernekirurg kan gå inn på riktig sted i hjernen og operere bort det syke område uten å skade noe som skal være der.



Bilde 1



Bilde 2



Bilde 3

Figur 8.5: Måling av hjerneaktivitet
(kilde: www.fysikknett.no)

9) Oppsummering

Vi har lært mye om superledere, og vi har lært at det er et veldig avansert og veldig omfattende tema, så det er utrolig mye mer å lære også.

Superlederteknologien brukes i flere sammenhenger, ikke bare til magnetisk levitasjonstog, eller svevetog, som vi innledningsvis nevnte at var det vi først og fremst forbandt med superledere. Superledende strømkabler, elektriske supermotorer og måling av hjernebølger er andre anvendelser superlederteknologien kan brukes til. Denne teknologien er veldig ny, for selv om superlederen ble oppdaget i 1911, var det først på midten av 80-tallet, i forbindelse med oppdagelsen av type II superleder, at det ble et satsningsområde innenfor moderne fysikkforskning. Det er derfor grunn til å tro at det finnes mye mer å lære om superledere enn det vi vet i dag, og fortsetter den enorme fremgangen som har pågått siden åttitallet, vil hverdagen vår helt sikkert bli mer preget av denne teknologien. Kanskje vi for eksempel om noen år kan reise med svevetog fra Trondheim til Oslo på et par timer.

Vi har først og fremst sett på teknologien bak superledere. Vi vil nå gi en kort oppsummering av det vi mener er det viktigste.

Innledningsvis skrev vi en kort innføring i magnetisme. Vi sa at ferromagnetiske og paramagnetiske materialer har et magnetisk dipolmoment, som innretter seg med det ytre magnetiske feltet og gir en økning av dette. Diamagneter har ingen permanent magnetisering, men med et ytre felt vil materialet motsette seg dette feltet ved å opprette en svak magnetisering motsatt vei av det påtrykte feltet.

Superledere er materialer som har null resistans under en viss temperatur. Denne temperaturen blir kalt superlederens kritiske temperatur (T_c). Type I superleder, som den nederlandske fysikeren Heike Kamerlingh Onnes oppdaget i 1911, oppfører seg som en perfekt diamagnet. Under den kritiske temperaturen slippes det ikke inn noe magnetisk fluks i slike materialer. Dette fluksekskluderingsfenomenet kalles meissnereffekt. Hvis en superleder av type I blir utsatt for et ytre felt som er over den kritiske feltverdien (B_c) blir den superledende tilstanden ødelagt. Denne feltverdien er mindre enn 0,2T for alle type I superledere. Det finnes også en kritisk strømverdi (I_c) for superledere. Sendes det mer

strøm gjennom en superleder enn denne verdien, blir også den superledende tilstanden ødelagt.

I løpet av 1980-åra ble det oppdaget en ny type superleder. Denne kalles superleder type II, og karakteriseres ved at den har to kritiske magnetfelt. Når det påtrykte feltet er under den laveste kritiske feltverdien (B_{c1}) vil det ikke trenge inn noe magnetisk fluks, akkurat som for en type I superleder. Når feltverdien er over den høyeste kritiske feltverdien (B_{c2}) vil den superledende tilstanden bli ødelagt og det ytre magnetiske feltet vil gå fullstendig gjennom materialet. Men når en type II superleder blir utsatt for et felt som ligger mellom den laveste og den høyeste kritiske feltverdien vil den være i en såkalt virveltilstand. Det vil da være hull i superlederen som ikke er superledende, og rundt disse vil det være virvler av superstrøm. Det er viktig å forankre, eller ”pinne”, disse flukslinjene fast slik at det ikke blir fluksforandring i materialet, noe som skaper friksjon.

BSC-teorien sier at superledere av type I befinner seg i en bundet tilstand med såkalte cooperpar. Disse består av to elektroner som har motsatt spinn og magnetisk moment. I grunntilstanden til en superleder er alle elektronene bundet i cooperpar. Elektronene går friksjonsfritt gjennom superlederen i en bevegelse som lar seg beskrive med en enkel bølgefunksjon. For å bryte opp cooperparene trengs det en kraft av størrelsesorden 1meV . Denne teorien holder ikke for type II superledere, og det finnes i dag heller ingen teori som gir en skikkelig beskrivelse av hva som skjer i en type II superleder.

Høytemperatursuperledere er en type II superleder som har de høyeste kritiske temperaturene. I dag er rekorden på rundt 138 K, men denne rekorden vil trolig bli slått i løpet av de nærmeste åra. Alle disse superlederne består av kobberoksider, og det ser ut til at det er en direkte sammenheng mellom antall kobberoksygenlag og den kritiske temperaturen. Jo flere slike lag, jo høyere kritisk temperatur. Disse superlederne har også et høyt kritisk magnetfelt og høyere kritisk strømverdi enn andre superledere. Men de består av keramer, og er derfor ganske sprø og går fort i stykker.

Det gjenstår altså mye forskning før det vil komme en fullstendig forståelse av superledere, men det er ingen tvil om at det finnes et enormt potensial i dette fenomenet, som vi drar nytte av i moderne teknologi og antageligvis vil dra enda mer nytte av i fremtiden.

10) Kilder

- Egil Lillestøl, Ola Hunderi og Jan R. Lien; "Generell fysikk – bind 2", 2001
- Serway, Moses og Moyer; "Modern Physics", 1997
- Alonso-Finn; "Fundamental University Physics"

Internettsider:

- www.superconductors.org
- www.apollon.uio.no/vis/art/2003/2/superled
- <http://www.fys.uio.no/super/levitation/>
- <http://my.execpc.com/~rheadley/magsuper.htm>
- <http://physics.ucsd.edu/~jorge/hole.html>
- <http://www.fys.uio.no/super/mo/#meissner>
- <http://online.itp.ucsb.edu/online/lnotes/balents/node18.html>
- <http://physics.ucsd.edu/~jorge/hole.html>
- <http://www.math.psu.edu/qdu/Res/Pic/gallery1.html>
- <http://www.apollon.uio.no/vis/art/2003/2/superled>
- www.fysikknett.no
- <http://www.angstrom.uu.se/solidstatephysics/education/kvant/pdf/supraledning.pdf>
- <http://www.forskning.no/Artikler/2003/oktober/1065526989.54>

Bygging av DC-motor med utredning av de underliggende fysiske prinsipper

En prosjektoppgave i FY1013



Forfattere

Lars Erik Klemetsen

Henrik Sivertsen

Jørgen Tande

Innholdsfortegnelse

1. Sammendrag.....	s. 3
2. Kort historikk.....	s. 4
3. Bakgrunnsstoff fysiske fenomener i DC-motor.....	s. 5 - 17
▪ 3.1 Magnetiske felt	
• 3.1.1 Strømførende leder.	
• 3.1.2 Strømførende sløyfe.	
• 3.1.3 Strømførende spole	
▪ 3.2 Hvorfor magnetiseres jern?	
• 3.2.1 Beregning	
▪ 3.3 Strøm og magnetisering	
▪ 3.4 Dipolmoment og spinn	
▪ 3.5 Reluktans	
4. Bygging av DC motor og dens komponenter.....	s. 18 - 25
▪ 4.1 Komponenter	
• 4.1.1 Ytre skall og permanentmagnet	
• 4.1.2 Bunn- og topplokk.	
• 4.1.3 Rotor og aksling.	
• 4.1.4 Modell av jernet i spolen	
• 4.1.5 Spole	
• 4.1.6 Kommutator	
• 4.1.7 Når enden er god er allting godt	
5. Effekttap.....	s. 26 - 35
• 5.1 Virvelstrømmer	
• 5.2 Kobbertap	
• 5.3 Luftgaptap og gyroeffekt	
• 5.4 Hysterese	
• 5.5 Effekttap av selvinduktans	
6. Målinger og resultat.....	s. 36 - 40
7. Noen elektromotorer i dag.....	s. 41 - 42
8. Kommentarer.....	s. 43
9. Kilder.....	s. 44

1. Sammendrag:

I dette prosjektet gikk vi inn på de fysiske prinsippene som ligger bak en DC-motor. Vi tok da utgangspunkt i en enkel prototyp motor som vi laget selv og brukte teori fra elektromagnetismen til å forklare de ulike elementene som spiller inn. For å gi leseren tilstrekkelig bakgrunnsteori startet vi med å gi en kort innføring i begreper og sammenhenger som man ser igjen i motoren. Videre beskrev vi byggeprosessen for vår egen motor for at leseren skulle kunne orientere seg om det byggetekniske aspektet ved DC-motoren. Tanken var å vise at man, med litt oppfinnsomhet, tilgang til diverse materialer og kjennskap til feltteori og magnetisering, skulle kunne ”oppfinne” en motor basert på induksjon. Det neste som stod for tur var å gjøre fysiske målinger på komponentene og motoren i ett. Særlig var effekt og effekttap interessant i testingen, og vi har lagt ekstra vekt på teorien rundt dette. I industrien er effektanalysen avgjørende for at en prototyp skal få livets rett og masseproduseres. Med høy reluktans, store virvelstrømmer, hysteresetap, selvinduktans og en lite optimalisert konstruksjon, ble nok motoren vår lite effektiv. Likevel fungerte motoren relativt bra uten ekstra belastning med en omdreingshastighet på 3000 RPM (rounds per minute) ved ca 8 Volt.

Prosjektet har virkelig vært svært lærerikt og motiverende; mye fordi vi har fått kombinert teori med praksis og at vi valgte å skrive innenfor et felles interesseområde.

2. Kort historikk:

Hans Christian Ørsted var den første som oppdaget en sammenheng mellom elektrisitet og magnetisme, rundt år 1820. Allerede året etter kunngjorde Ampère en rekke resultater fra de første forsøk han utførte med utgangspunkt i Ørsteds arbeid. Tidlig på 1830 tallet oppdaget Michael Faraday og Joseph Henry uavhengig av hverandre sammenhengen mellom det å endre på magnetiske felt og det å indukere elektrisk energi i elektriske kretser. Dette ble kjent som elektromagnetisk induksjon.

De tidligste elektromotorene var en del annerledes enn de som brukes i dag. Datidens elektromotorer brukte ikke rotasjonsprinsippet, men forskerne hentet heller sin inspirasjon fra den horisontale dampmaskinen. De fleste elektromotorene på denne tiden var essensielt solenoider som dro et metallstykke inn i spolen når de ble strømførende. Metallstykket var festet til en stang og et drivhjul, akkurat som i en dampmotor. Joseph Henry utviklet blant annet, i 1831, det som antas å være den første automatiske elektromotoren av denne typen, en som ikke manuelt trengtes å omstilles eller nullstilles. Han viet mye av sin tid til elektromagnetisme og i 1893 ble navnet hans gitt til den elektriske enheten induktans, Henry.

Amerikaneren Frank J. Sprague(1857-1934) var en pioner i utvikling og bruk av likestrømsmotorer slik vi kjenner dem i dag. I 1884 installerte han en motor til å drive en heis i Boston, og rundt år 1900 var elektromotorene allerede godt i gang med å fortrenge turbiner og dampmaskiner som drivkraft i industrien.

3.1 Bakgrunnsstoff:

3.1.1 Magnetiske felt:

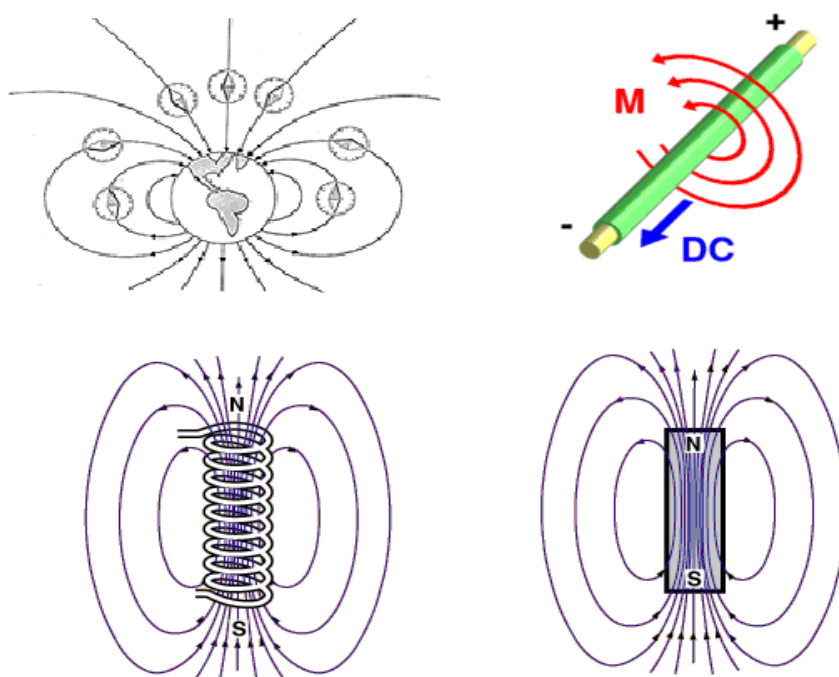
For å skjønne de fysiske prinsippene som ligger til grunn for at den elektromagnetiske motoren faktisk fungerer, er det essensielt å ha en viss forståelse av magnetiske felt. I vår elektromotor danner vi et magnetisk felt ved å føre strøm gjennom en spole, og det er nettopp dette som er drivkraften.

Magnetiske felt dannes ved at elektrisk ladde partikler beveger seg. Dette kan f.eks være en strøm av elektroner i en ledning, eller elektroner i sirkulære baner rundt atomer. Magnetfeltet er tenkte linjer som viser retningen til kraften på en "løs nordpol". Tette magnetfeltlinjer betyr sterkt magnetfelt og spredte linjer betyr svakt magnetfelt. For en strømførende leder er magnetfeltet gitt ved retningen på strømmen. For å finne retningen på feltet har vi noen tommelfingerregler, eller høyrehåndsregler, som vi kommer tilbake til.

De magnetiske feltlinjene oppstår rent skematisk svært ulikt de to andre fjernkreftene i naturen, elektrisk kraft og gravitasjonskraft. De sistnevnte går f.eks i linjer radielt ut fra kuler med masse m og/eller ladning q , eller i linjer mellom objekter, mens magnetfeltet alltid oppstår i lukkede kurver.

Illustrasjon: Typiske magnetfelt for henholdsvis jorda, en strømførende ledning, en spole og en stavmagnet:

Figur 3.1:

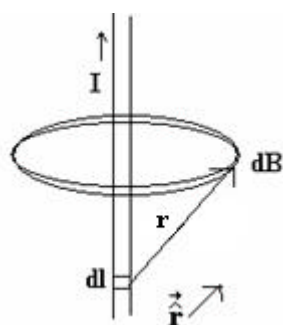


Når man skal «oppfinne» elektromotoren, trenger man i utgangspunktet bare vite hvordan magnetfeltet oppfører seg rundt en strømførende leder for videre å kunne forutse hvordan spolens magnetfelt ville se ut. Siden oppdagelsen av elektisitet går hånd i hånd med utviklingen av generatoren, kan det være vanskelig å påstå at vi ikke visste om **B**-feltet i en spole når vi forutsetter at elektrisitet er oppfunnet. Generatoren er nemlig bygget på samme prinsipp som elektromotoren, så man kan ironisk si at vi oppfinner en el.mag.-motor ved å bruke en el.mag.-motor.

3.1.2 Strømførende leder:

Magnetfeltet til en strømførende ledning kan utledes fra **Biot-Savarts lov**, som gir en sammenheng mellom punktladninger i bevegelse og magnetfelt. Vi ser da på magnetfeltet rundt en infinitesimal del av en uendelig lang strømførende leder, i en avstand **r**:

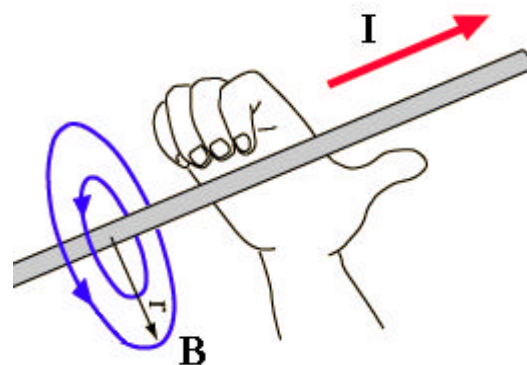
Figur 3.2



$$d\vec{B} = \frac{\mu_0}{4\pi} \frac{I d\vec{l} \times \hat{r}}{r^2}.$$

Det totale magnetiske feltet får vi ved å summere alle bidrag **dl** langs lederen. Av symmetriårsaker får vi kun bidrag til **B**-feltet i sirkulær retning. Med høyrehåndsregelen blir retningen på **B** og styrken:

$$B = \frac{\mu_0}{2\pi} \cdot \frac{I}{r}$$

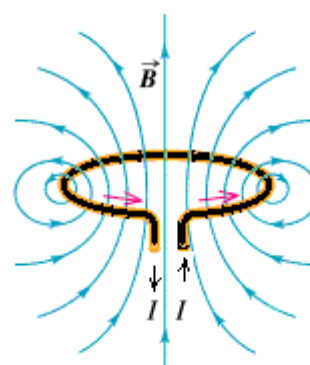


Figur 3.3

3.1.3 Strømførende sløyfe:

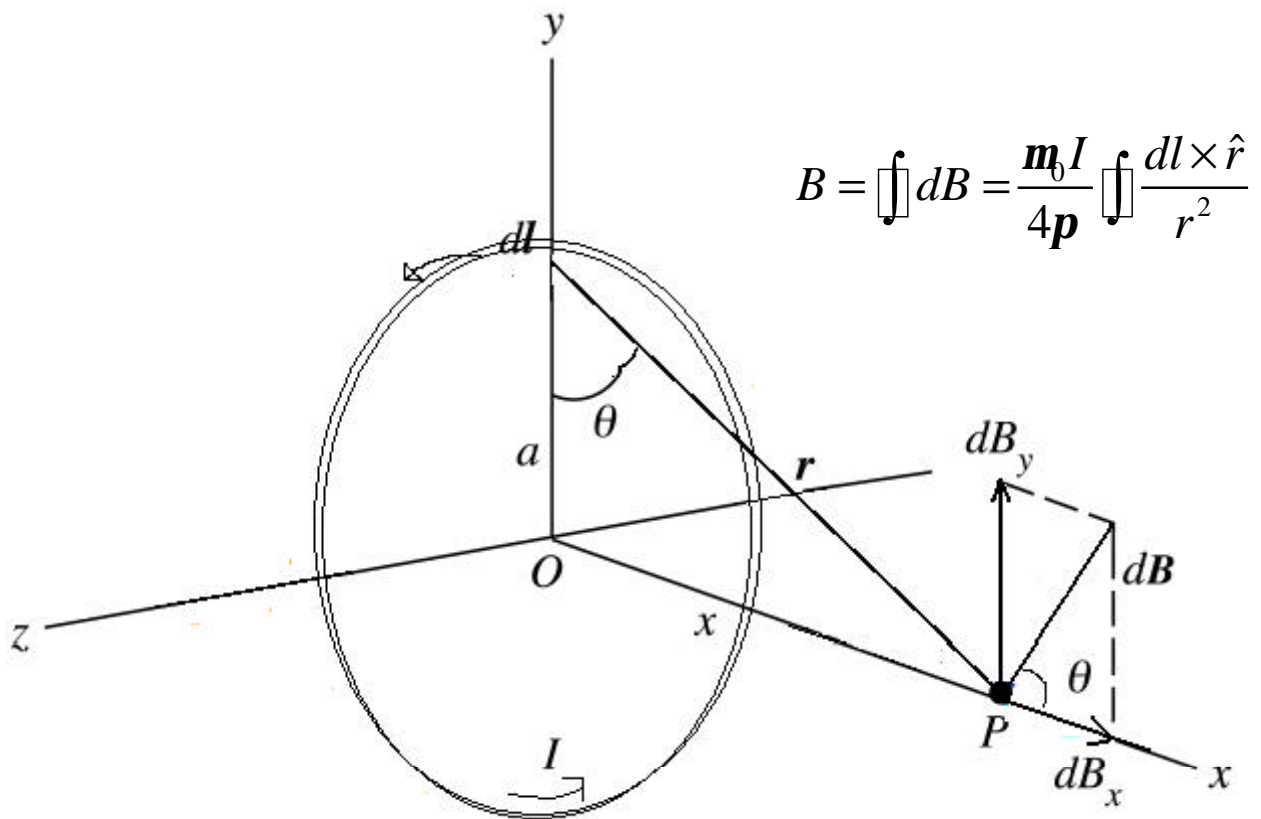
Hvis vi bøyer ledningen til å bli sirkulær, vil vi få en strømførende sløyfe. Magnetfeltet i denne vil intuitivt se slik ut:

Figur 3.4



Det totale magnetfeltet \mathbf{B} på aksen til en slik lukket strømsløyfe finner vi ved å summere opp bidraget fra alle differensielle lederstykker som til sammen utgjør den lukkede strømsløyfa, dvs ved å integrere uttrykket for $d\mathbf{B}$ rundt sløyfa:

Figur 3.5



Symmetrien når vi skal integrere $d\mathbf{l}$ rundt sløyfen gjør at vi bare får bidrag i x-retningen:

$$dB_x = dB \cdot \cos \theta = dB \cdot \frac{a}{r} = dB \cdot \frac{a}{(a^2 + x^2)^{1/2}}$$

$$= \frac{\mu_0 I dl}{4\pi} \frac{a}{(a^2 + x^2)^{3/2}}$$

Det totale feltet vi får ved å integrere rundt kurven blir da

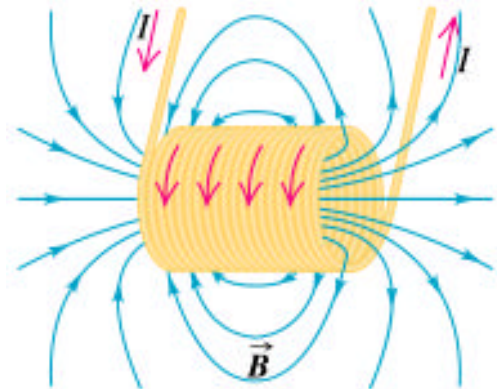
$$B_{total} = \oint dB_x = \frac{\mu_0 I}{4\pi} \frac{a}{(a^2 + x^2)^{3/2}} \oint dl$$

$$= \frac{\mu_0 I}{4\pi} \frac{a}{(a^2 + x^2)^{3/2}} \cdot 2\pi a = \frac{\mu_0 I}{2} \frac{a^2}{(a^2 + x^2)^{3/2}}$$

Når vi beveger oss videre til spolen vil vi også få en tilsvarende symmetri langs aksene, og tilnærmet homogent med hele innsiden av spolen avhengig av lengden på spolen, viklingstetthet, relativ permeabilitet inne i spolen og radius.

3.1.4 Strømførende spole:

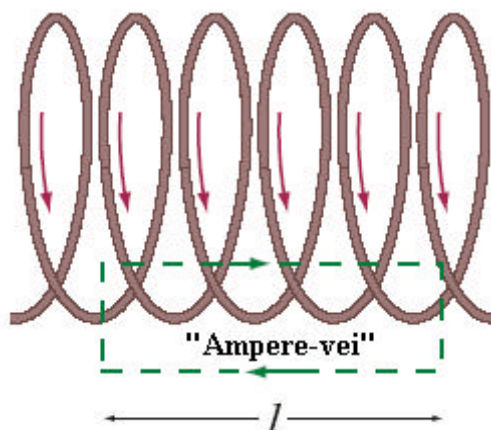
Vi går så videre med å sette sammen mange slike strømsløyfer som i forrige side, slik at vi danner en spole, eller solenoide. Magnetfeltet blir da som på figuren:



Figur 3.6

La oss si at den har n sløyfer pr lengde og strøm I . Magnetfeltet kan finnes ved å bruke Ampères lov og en integrasjonsvei som vist i figuren. Integrasjonsveien er valgt slik at strømmene i viklingene har riktig retning ut fra høyrehåndsregelen. Når spolen er lang i forhold til diameteren kan vi gjøre tilnærmingen at magnetfeltet er homogent og parallelt med aksene inne i solenoiden og at feltet er null på utsiden.

Figur 3.7



Ampères lov sier:

$$\oint B dl = \mu_0 I_{(\text{gjennom } A)}$$

Der $I_{\text{gjennom } A}$ er lik $l \cdot n \cdot I$

$$B = \mu_0 n I$$

Denne tilnærmingen blir enda bedre når man fyller spolen med et ferromagnetisk materiale. Magnetiseringen av materialet er proporsjonalt med det ytre feltet og kan forsterke B-feltet inne i en spole med flere hundre størrelsesordener. Les om ferromagnetisme på side 9 til 11.

Vi skal i dette prosjektet lage en spole som omsluttet et slikt ferromagnetisk materiale, og vi vil på side 39 og 40 gjøre en tilnærming til hvor stort magnetfelt vi får i denne hjemmesnekrede elektromagneten.

3.2 Hvorfor magnetiseres jern:

Elektroner i baner, uten påtrykt magnetfelt, rundt atomkjerner er mikroskopiske strømsløyfer som vil sette opp magnetfelter rundt seg. Denne magnetismen slukker seg ofte ut, enten fordi to elektroner går i samme bane, men i motsatt retning, eller fordi naboatomer er tilfeldig orientert slik at magnetismen fra hver mikroskopiske strømsløyfe kansellerer hverandre. Materialer med en slik innstilling på atomært nivå, vil være magnetiske nøytrale. For mer om spinn se egen del side 15 og 16.

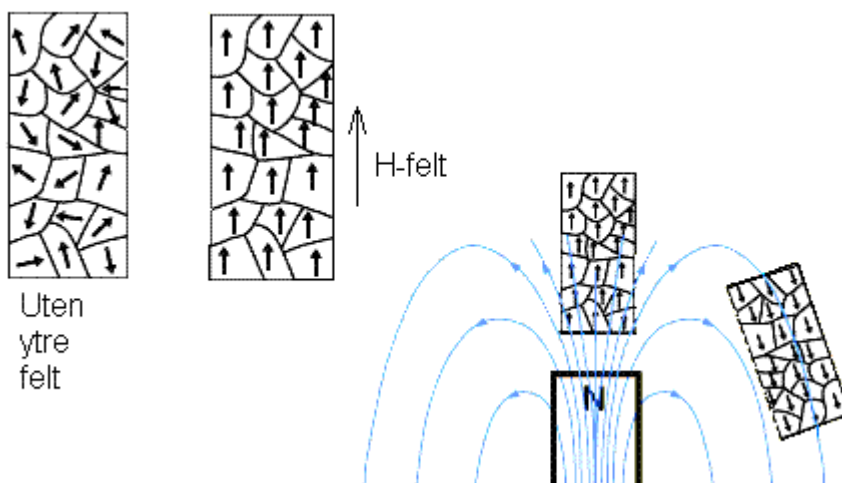
Ferromagnetiske materialer har den egenskapen at når de utsettes for et ytre magnetfelt vil dipolene i materialet rette seg inn etter retningen til det ytre magnetiske feltet og forbli slik etter den ytre påvirkningen er fjernet. I tillegg til å bevare sin magnetisering vil slike materialer ha en veldig sterk magnetisering når de settes i et ytre magnetfelt. Dette gjør sitt til at total feltet blir mange ganger forsterket. Jern, kobolt og nikkel er slike ferromagnetiske materialer. I vår motor bruker vi bløt jern i spolen.

For å forklare dette må vi innføre noe som kalles Paulis eksklusjonsprinsipp og noen av Werner Heisenbergs tanker. Eksklusjonsprinsippet går ut på to elektroner i en begrenset mengde av et materiale ikke kan oppta samme energitilstand samtidig. Heisenberg mente

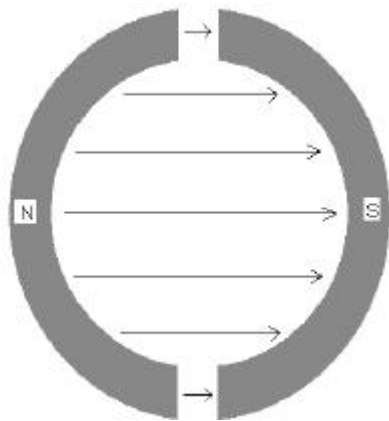
at som en konsekvens av dette ville vi ha at to uparede elektroner med tilnærmet parallelle spinn vil innrette seg i baner slik at midlere avstand mellom dem blir størst mulig. Dette fordi coulombfrastøtningen mellom dem reduseres og dermed blir den potensielle energien minst mulig. Naturen foretrekker den laveste energitilstanden og dermed vil elektronene ordne seg slik at spinnene deres og dermed deres magnetiske dipolmoment \mathbf{m}_s er parallelle. I analogi med den elektriske dipolen er energien til en magnetisk dipol \mathbf{m} i et uniformt magnetisk felt \mathbf{B} gitt ved $U = -\mathbf{m} \cdot \mathbf{B}$, slik at dipolen har sin laveste energi når \mathbf{m} og \mathbf{B} er parallelle.

Sagt på en enklere måte vil alle de tre ytterste elektronene i jern gå i samme retning. Det gjør at jern, når alle jernatomene ligger parallelt, gir et meget sterkt magnetfelt. Denne spontane innretningen skjer ikke gjennom hele materialstykket, men innenfor såkalte magnetiske områder eller domener, hvor alle de magnetiske dipolene er parallelle. Domenene varierer i volum, men har ofte en lineær utstrekning på 0,1 og 1mm og inneholder mellom 10^{17} og 10^{21} atomer. I en stor, uniform krystall kan lengden av et magnetisert område være opptil flere centimeter. Innretningen på det magnetiske feltet varierer fra område til område. Domenene kan forandre seg på to måter ved at et ytre magnetfelt påvirker dem: Enten kan områder med magnetisering tilnærmet parallelle med det ytre feltet øke i størrelse på bekostning av nabo domener. Eller så kan områder få en ny spontan innretning mer lik det ytre feltet, men uten å endre fysisk størrelse.

Figur 3.8

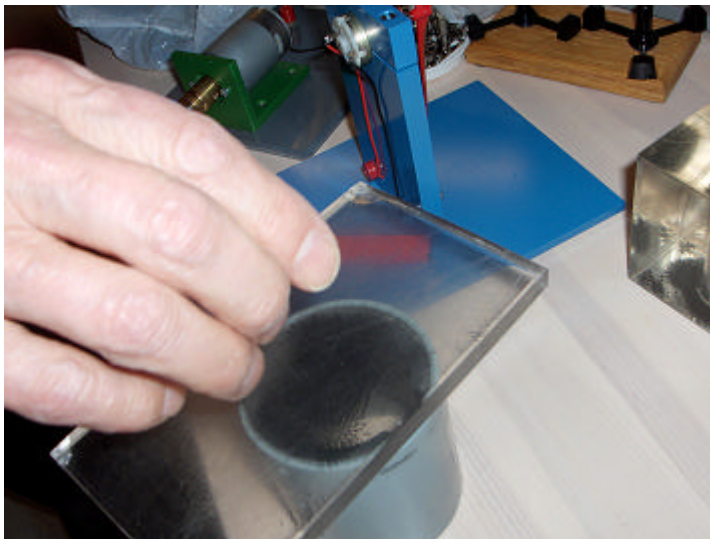


I den permanente magneten vi brukte i vår elektromotor var feltet rettet horisontalt i hele magneten.



Figur 3.9

Dette fant vi ut ved å plassere en glassplate med jernspon i seg over magneten, se bildet nedenfor. Inne i glassplaten fløt det en væske som inneholdt jernspon. Disse jernsponene vil magnetiseres og rette seg inn etter det påtrykte feltet, i dette tilfellet vår permanentmagnet.



Elektriske felt fører til innretning av atomære eller molekylære elektriske dipoler i et dielektrikum. Det samme foregår med magnetiske dipoler i et magnetfelt. I et dielektrikum vil dette føre til en svekkelse av det elektriske feltet, men i magnetisk felt vil innrettingen av dipoler derimot styrke det magnetiske feltet. Dette er fordi magnetfeltlinjene alltid danner lukkede sløyfer.

Det finnes flere materialer som påvirkes av ytre magnetfelt enn de ferromagnetiske. I paramagnetiske materialer vil magnetfeltet bare få en liten økning. I diamagnetiske materialer vil faktisk det magnetiske feltet svekkes.

3.2.1 Beregning:

Magnetisering, \mathbf{M} , er størrelse vi kan bruke for å bestemme stoffenes magnetiske egenskaper. Denne størrelsen er definert som: *magnetisk dipolmoment per volumenhet*. Enheten for magnetisering er dermed den samme som (strømstyrke x areal)/volum, det vil si A/m. Dette er den samme enheten som for B/μ_0 . Magnetfeltet kan skrives som summen av virkningen fra det påtrykte feltet og magnetiseringen, slik at

$$\mathbf{B} = \mathbf{B}_0 + \mu_0 \mathbf{M} \quad (1)$$

Vi innfører en størrelse \mathbf{H} som bare avhenger av det ytre feltet, for eksempel den påtrykte strømmen, og er uavhengig av magnetiseringen. Denne vektoren kalles *magnetisk intensitet*, har samme enhet som \mathbf{M} og er definert ved

$$\mathbf{H} = \mathbf{B}_0/\mu_0 \quad (2)$$

Vi kan nå omskrive ligning (1) til $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ (3)

For et isotropt (et materiale som har samme fysiske egenskaper i alle retninger) og homogent materiale vil \mathbf{M} være proporsjonal med \mathbf{H} og i tillegg ha samme retning. Dermed får vi

$$\mathbf{M} = x_m \mathbf{H} \quad (4)$$

Hvor størrelsen x_m kalles *magnetisk susceptibilitet* (mottakelighet) og er en materialkonstant uten benevning. Ved hjelp av denne størrelsen kan vi skrive

$$\mathbf{B} = \mu_0(\mathbf{H} + x_m \mathbf{H}) = \mu_0(1 + x_m)\mathbf{H} = \mu \mathbf{H} \quad (5)$$

Hvor

$$\mu = \mu_0(1 + x_m) \quad (6)$$

kalles *permeabiliteten* for mediet og har same enhet som μ_0 , altså Tm/A eller H/m.

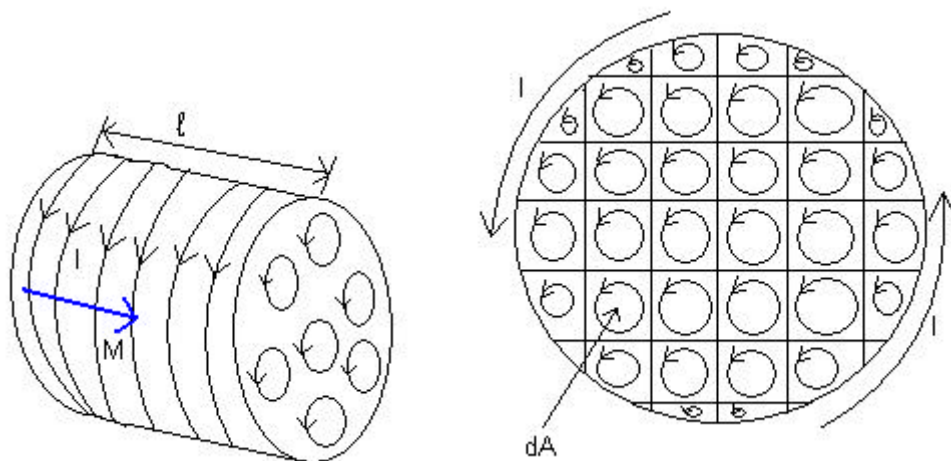
$$\mu_0 = 4\pi \cdot 10^{-7} \text{ Henry} \cdot \text{m}^{-1} \quad \text{magnetisk permeabilitet}$$

Den *relative permeabiliteten* er definert som

$$\mu_r = \mu/\mu_0 = 1 + x_m \quad (7)$$

og er en materialkonstant uten benevning. Stor x_m betyr stor magnetisering. I ferromagnetiske materialer er den relative permeabiliteten stor (størrelsesorden 1000). Sammenhengen mellom det totale magnetfeltet og det påtrykte feltet er ikke lineær og kurven man da får kalles hysteresekurve. Dette fenomenet kommer vi nærmere inn på senere.

3.3 Strøm og magnetisering:



Figur 3.10

Vi ser på små, men makroskopiske volumelementer av sylindere som vist på figur 3. Volumelementene har samme lengde l og, bortsett fra langs sylindrens ytterside, samme tverrsnitt dA , som vist på figuren. Hver dipol tenkes som en strøm I som går i en sløyfe rundt elementenes yttersider og vinkelrett på sylinderaksen. Dipolene vil derfor ha dipolmomenter $dm = IdA$, alle med retning langs sylinderaksen. Ved å addere alle disse dipolene, får vi et totalt dipolmoment

$$m = \int dm = I \int dA = IA$$

Dette gir en magnetisering

$$M = \frac{\text{dipolmoment}}{\text{volum}} = \frac{IA}{Al} = \frac{I}{l}$$

Når vi setter sammen elementene, vil strømmen langs yttersiden av ett element inne i sylindere oppheves av en like stor, men motsatt rettet strøm i naboelementene. Den eneste nettostrømmen vil derfor være strømmen I rundt sylindrens krumme ytterflate. Sylindere kan dermed sees på som en "spole" med tverrsnitt A og lengde l . \mathbf{B} -feltet for en uendelig lang spole med konstant \mathbf{B} -felt er gitt ved

$$B = \mu_0 \frac{I}{l}$$

Her har vi erstattet n med $N/l = 1/l$. Vi ser at dette stemmer overens med resultatet over når $\mathbf{B} = \mu_0 \mathbf{M}$. Der vi har erstattet n med $N/l = 1/l$. Vi ser at dette stemmer med resultatet over når $\mathbf{B} = \mu_0 \mathbf{M}$.

Dermed blir magnetiseringens bidrag til det totale magnetfeltet er derfor gitt ved ligning

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$$

Beregning av magnetfelt i materie fra for eksempel strømmer kan nå gjøres forholdsvis enkelt ved å starte med magnetfeltet i vakuum. Vi kan erstatte Ampères lov for \mathbf{B} -feltet med en tilsvarende lov for \mathbf{H} -feltet, slik at

$$\oint \vec{H} \cdot d\vec{r} = I_{fri}$$

Her ser vi at proporsjonalitetskonstanten ikke μ_0 er med, dermed ser vi at \mathbf{H} -feltet uavhengig av materialet. Når vi har funnet \mathbf{H} -feltet, finner vi \mathbf{B} -feltet fra ligning

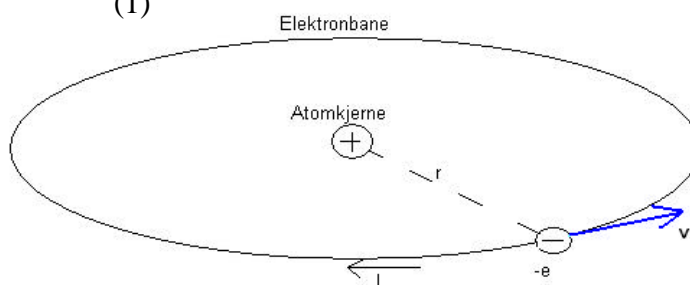
$$\mathbf{B} = \mu_0(\mathbf{H} + x_m\mathbf{H}) = \mu_0(1 + x_m)\mathbf{H} = \mu\mathbf{H}$$

3.4 Dipolmoment og spinn:

I en klassisk modell av atomet vil elektronene bevege seg i sirkulære baner rundt atomkjernen, og vil dermed sette opp en strøm med retning motsatt baneretningen;

$$I = \frac{e}{T} = \frac{ev}{2\pi r} \quad (1)$$

Figur 3.11



Magnetiske dipolmoment er en egenskap til en magnet som indikerer hvor raskt magneten vil rette seg inn etter et ytre magnetfelt.

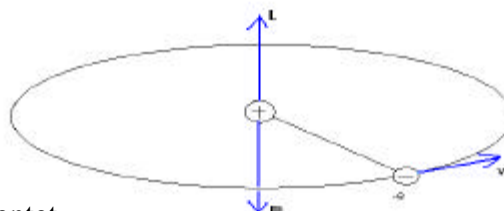
Et magnetisk moment vil oppstå pga av denne sirkulasjonsstrømmen, gitt ved

$$m_L = IA = \frac{ev}{2\pi r} \pi r^2 = \frac{evr}{2} \quad (2)$$

Banespinnet for elektronet i den sirkulære banen er

$$L = rm_e v \quad (3)$$

Figur 3.12



Dermed kan vi skrive det magnetiske momentet

$$\mathbf{m}_L = -\frac{e}{2m_e} \mathbf{L} \quad (4)$$

Minustegnet i ligningen skyldes at for elektronet er det magnetiske momentet og banespinnet i motsatt retning. Sammenhengen mellom \mathbf{m}_L og \mathbf{L} er utledet under forutsetning av at den elektriske ladningen beveger seg i sirkelbane, men dette gjelder også for andre periodiske banebevegelser selv om det ikke er en sirkelbane. Et prinsipp fra kvantemekanikken sier at størrelsen på banespinnet er kvantisert, slik at det alltid er et helt tall multiplisert med konstanten $\hbar = h/2\pi$ der h er Plancks konstant. Dermed vil mulige verdier av banespinnet være $L=l\hbar$ der l er null eller et positivt heltall.

I tillegg til banebevegelsen har elektronet en egenrotasjon eller et egenspinn \mathbf{S} . Også dette vil bidra til det magnetiske momentet, slik at elektronet kan oppfattes som en liten magnet;

$$\mathbf{m}_s = -\frac{e}{m_e} \mathbf{S} \quad (5)$$

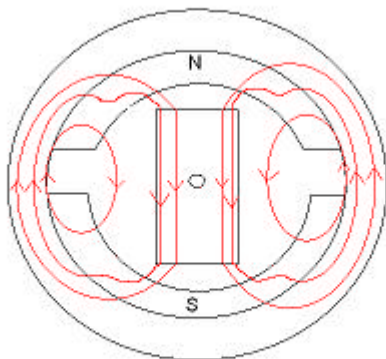
For et proton eller nøytron vil det magnetiske momentet fra spinnnet være nesten 2000 ganger mindre enn for elektronet. En kan dermed konkludere at for et atom er det totale magnetiske momentet dominert av elektronets magnetiske momenter. Det magnetiske momentet for et atom er derfor lik summen av m_L og m_s for alle elektronene i atomet. I klassisk fysikk er det ingen regler for hvordan de forskjellige elektronene beveger seg i forhold til hverandre, slik at de forskjellige dipolmomentene ville ha vilkårlige retninger i forhold til hverandre. Men kvantemekanikken sier noe helt annet, elektronene i et atom og i kovalente molekylbindinger opptrer parvis. Hvilket betyr at for et atom i grunntilstanden vil elektronenes magnetiske momenter oppheve hverandre parvis. Vi kan dermed trekke følgende konklusjon: Atomer med et odde antall elektroner må ha et netto magnetisk dipolmoment.

3.5 Reluktansmoment:

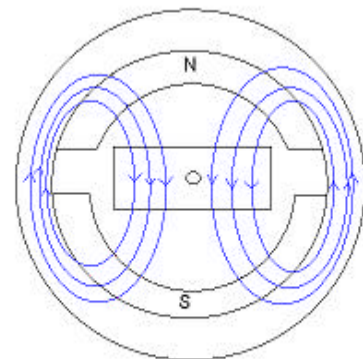
For å forklare dette fenomenet skal jeg bruke energiligningen for et magnetfelt. Den ser slik ut

$$W_{magn} = \frac{1}{2} \int_V (\vec{B} \cdot \vec{H}) dv$$

Siden magnetfeltet vil få et større volum av jernet å gå gjennom når den er dreid 90⁰ grader enn når den står vertikalt i magneten, vil den også få større energi i denne posisjonen. Naturen søker etter lavest energi vil rotoren stå vertikalt inne i motoren. Se figurer neste side. Den kraften vi må bruke på å dreie rotoren og dermed øke energien, kalles reluktansmomentet. For å få motoren vår til å gå av seg selv måtte vi ha minst 2 ampère og da dreie den i gang med menneskekraft. Når vi hadde fått den i gang måtte vi ha minst 1,5 ampère for at den skulle opprettholde bevegelsen sin.



Figur 3.13 vertikal



Figur 3.14 horisontal

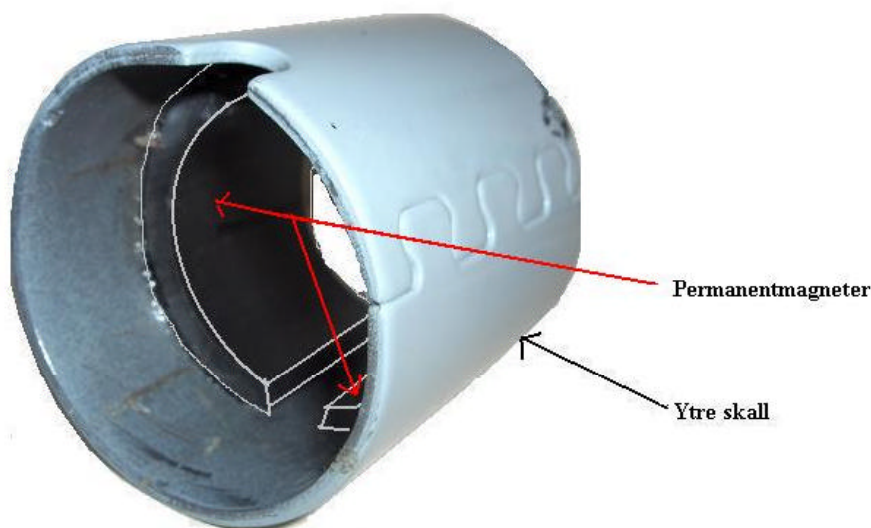
4.0 Bygging av DC-motoren og dens komponenter:

Den mest tidkrevende delen av dette prosjektet har den praktiske. Da vi fikk ideen om å skrive prosjekt om elektromotoren, falt det oss naturlig å også bygge en. Tanken var videre at vi skulle kunne komme frem til en elektromotor prototyp kun på teoretisk grunnlag og uten kjennskap til andre byggetekniske løsninger. Med andre ord var planen å «oppfinne» elektromotoren på nytt.

Etter hvert som vi gikk nærmere inn på byggeprosessen og fant ut hva som var tilgjengelig av utstyr og materialer på diverse laber og verksteder på NTNU, kom vi frem til at en fullt ut egenprodusert motor ville kreve veldig mye tid på verksted og «produktutvikling og design». Vi ble enig om at det ville bli mer gjennomførbart å bruke deler fra andre motorer for å sette sammen vår egen. Etter mye diskusjon med stipendiater, amanuensiser og professorer samt endeløs trasking mellom fakulteter, fikk vi tilslutt samlet nok deler til å kunne bygge noe som i teorien skulle kunne fungere. Da var det i hovedsak på elektrobygget vi høstet de fleste komponentene. Enkelte deler måtte justeres noe for å passe inn i modellen vår, og andre deler måtte spesiallages ved å sende skjematiske tegninger til de ansatte ved teknisk verksted nede i U3 ved Institutt for Fysikk.

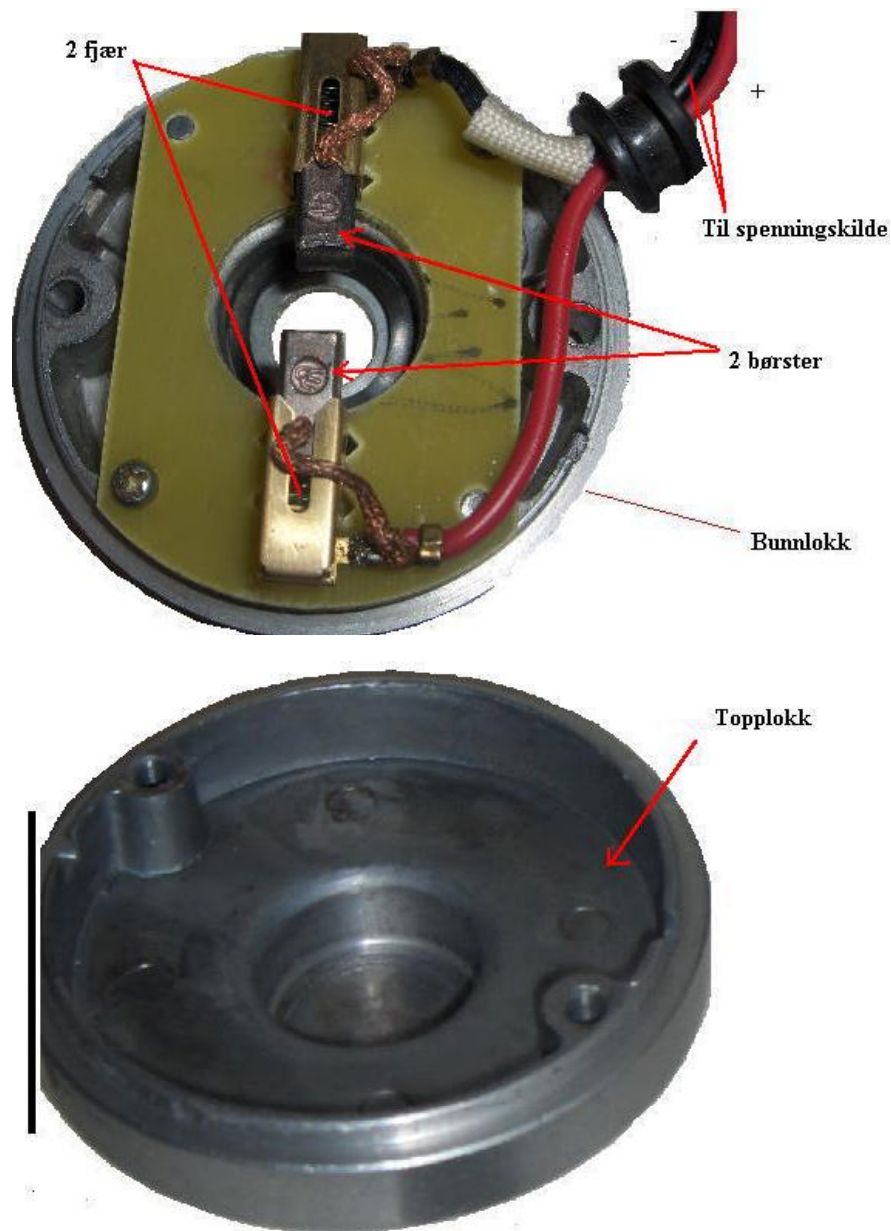
4.1 Komponenter:

4.1.1 Ytre skall og permanentmagneter:



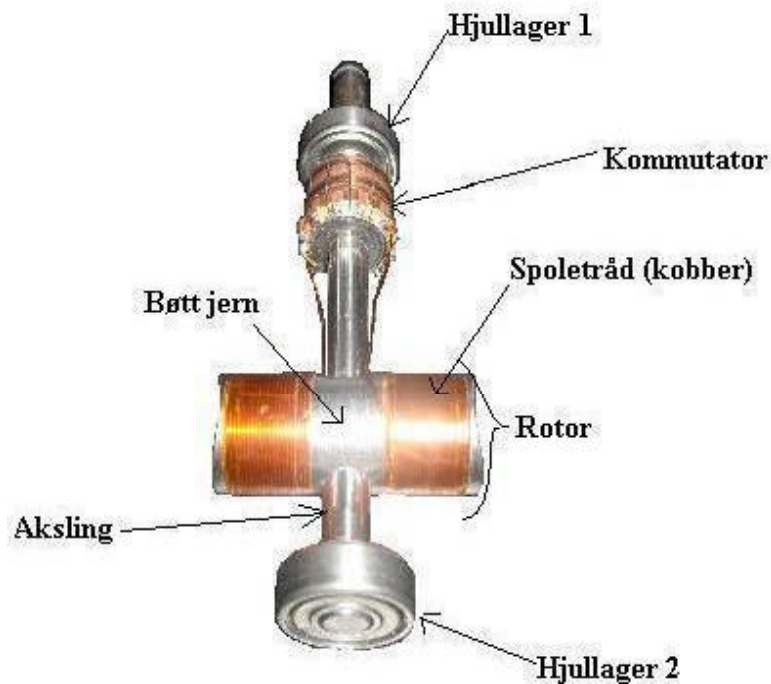
Det første vi anskaffet var de permanente magnetene. Vi ville ha sirkulære magneter slik at vi fikk minst mulig luftgap mellom dem og en sirkulerende rotor, og dette fant vi i en el-motor for sparksykkel som vi fikk lov til å demontere. Denne var vi heldige med siden permanentmagnetene var festet til en metallsylinder akkurat slik vi hadde lagt opp til i skissene våre. Vi testet de også mot en stavmagnet og bekreftet av de var av to motsatte poler. Bunn- og topplokk tok vi også med oss for å bruke i egen motor.

4.1.2 Bunn- og topplokk:



Bunnlokket bestod av 2 ledende børster som kunne kobles enkelt til en strømkilde ved at vi byttet ut de originale uttakene. I midten var der også et hull som kunne brukes som utgang til akslingen. Børstene presses mot sentrum (eller kommutatoren) ved hjelp av to fjærer. Bunnlokket var formet med en innhuling i midten. Dette slik at akslingen kunne stå stabilt inne i motoren og for at permanentmagneter og rotor ikke skulle skrape sammen. Når den enden av akslingen som hadde hjullager var satt inn i innhulingen ble det helt optimalt og lite friksjon. Sammen med metallsylinderen dannet da bunn- og topplokk statoren.

4.1.3 Rotor og aksling:



Her er motorens «innmat» med tilhørende topp- og bunnlokk på henholdsvis bunn og topp. Særlig har denne delen har voldt oss mye arbeid. Vi brukte også her noen deler fra sparksykkelmotoren, men måtte gjøre litt nye innstillinger på kommutator etter at vi kvittet oss med den eksisterende rotoren. Vi tok da i bruk heftig utstyr fra 2.verdenkrig under Knut Gjervans ekspertise for å gjøre de nødvendige inngrepene;

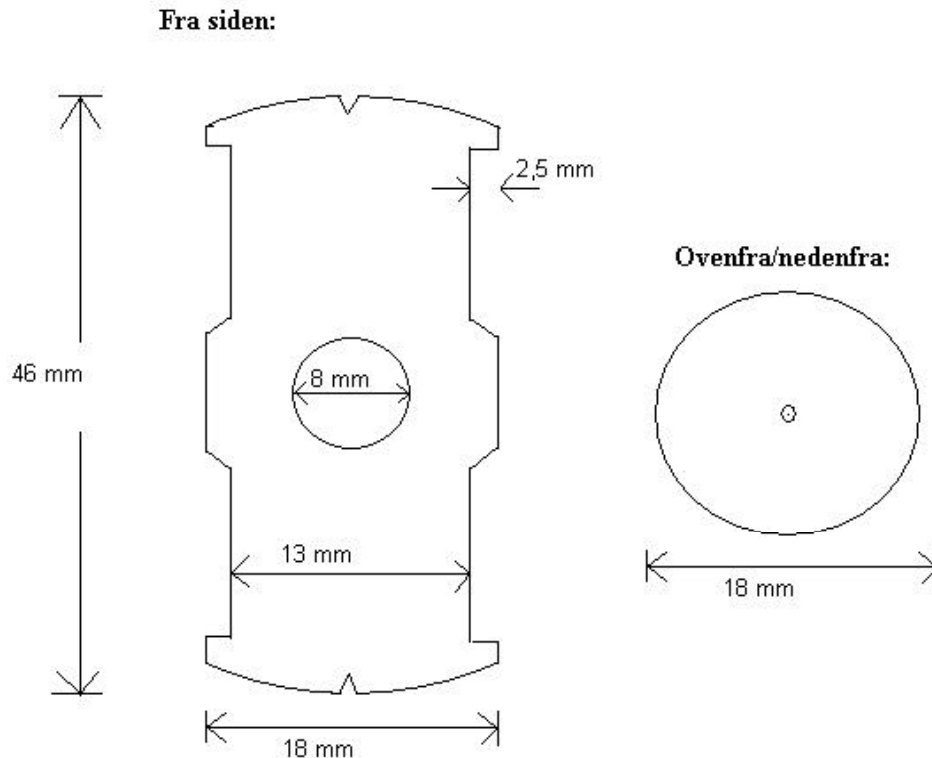


Med tanke på at vi i utgangspunktet skulle lage en motor som skulle være mulig å regne på, så valgte vi innlysende nok å forme det magnetiserbare materialet inne i kobberviklingene som en sylinder. Samtidig måtte vi prøve å lage den slik at den induserte så kraftig magnetfelt som mulig og overgikk styrken på feltet til de permanente magnetene.

Siden vi bare hadde plass til en 4,6 cm lang rotor så måtte vi lage en elektromagnet med så mange viklinger som mulig på et veldig lite metallstykke for å få til dette. Se tegning av vårt bløte jern nedenfor:

4.1.4 Modell av jernet i spolen:

Figur 4.1

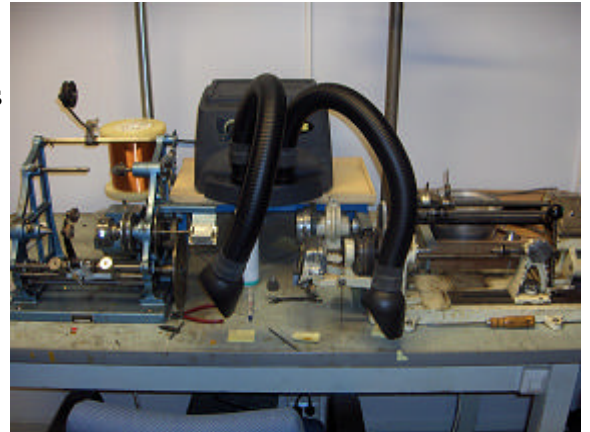


Ovenfor er utformingen som ga plass til flest mulig viklinger på metallet og hadde minst mulig luftgap mellom permanentmagnetene og rotoren for vår én-spulers DC-motor.

- ∞ De to innrissene på endene av metallet var for å senere kunne bruke en viklingsmaskin til å vikle spolen. Her ble det gjort en liten feil i utformingen pga. en litt misvisende tegning fra vår side (vi tegnet ikke ovenfra). Vi fikk altså et langt hakk på tvers av begge endene av metallet i stedet for kun et innriss. Men dette fikset vi fort.
- ∞ Grunnen til innskjæringene på 2,5mm på sidene er for å kunne vikle flere lag med kobber rundt metallet. Hullet i midten er for akslingen. Dette måtte også være pinlig nøyaktig for at vi skulle kunne banke inn akslingen slik at den stod fast og rett. Heldigvis tok mekanikerne våre disse kravene seriøst, og vi fikk tilbake en overraskende symmetrisk og korrekt metallbit, som bestilt.
- ∞ Luftgapet mellom metallet og magnetene var kun 1,5mm, så symmetrien i metallet måtte være nærmest perfekt, særlig med tanke på at den senere skulle gå 50 runder i sekundet

4.1.5 Spole:

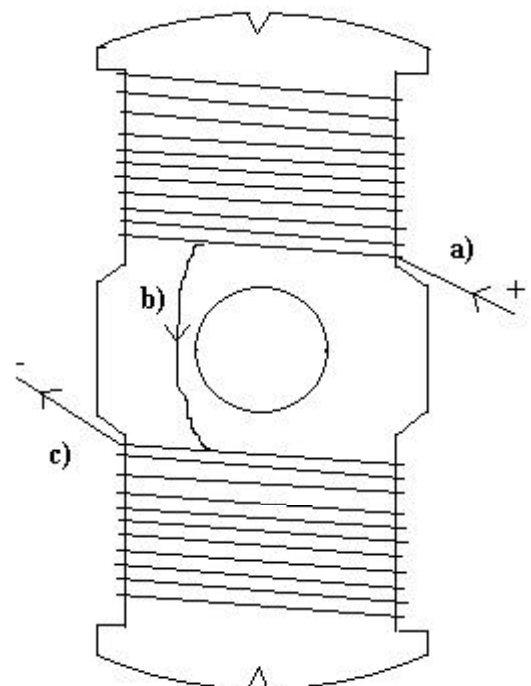
Etter at det bløte jernet var ferdig utformet stod spolen for tur. Vi hadde allerede vært flere ganger på besøk hos Arne Moholt ved det elektriske verkstedet i 3.egt på realfagsbygget, og diskutert gjennomførelse mang en gang med han tidligere, så det var da bare å sende over stafettpinnen. Viklemaskinen de hadde der fikk vi dessverre ikke prøve uten opplæring, men for orienteringens skyld såg den slik ut:



Vi hadde tidligere kommet frem til at å bruke 0,65mm kobbertråd og vikle den i fire lag utenpå metallet ville gi best utslag for indusering av magnetfelt i spolen. Det var altså denne diameteren som best ville kunne snurres i fire lag rundt metallet.

Forklaring av vikling:

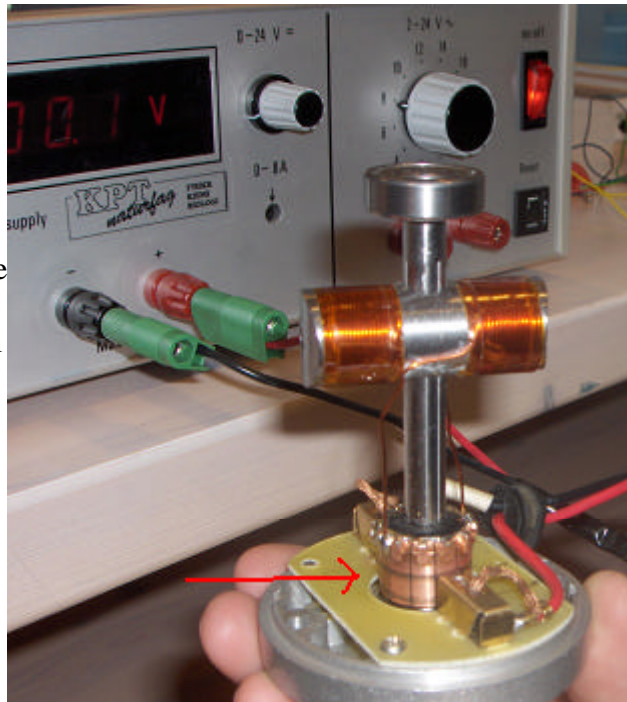
Strøm fra spenningskilde sendes inn gjennom **a)** og vikles oppover og nedover langs metallet til det er fire lag med kobber rundt det. Deretter går kobberlederen **b)** over til andre siden og vikles som et speilbilde av den motstående siden, og går så ut i **c)**.



Figur 4.2

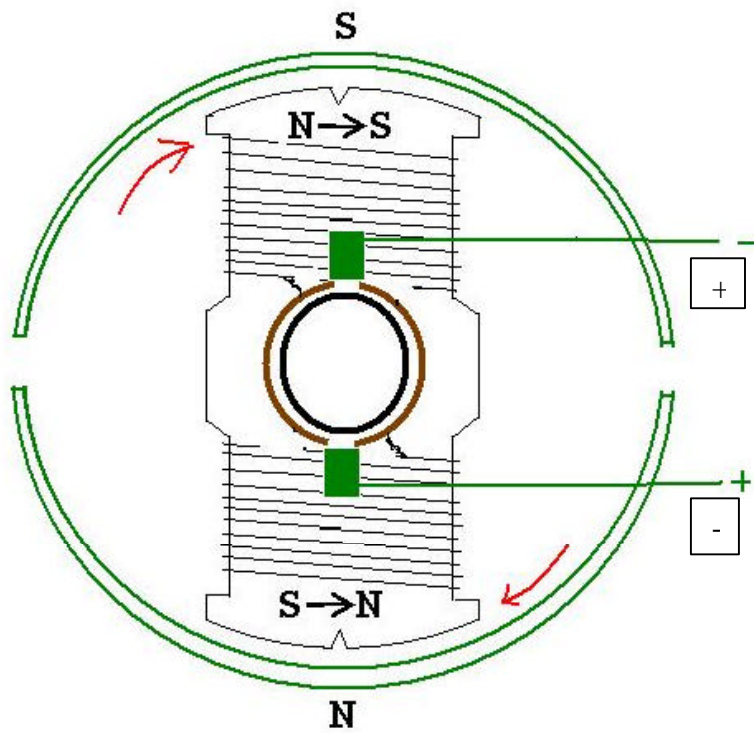
4.1.6 Kommutator:

Kommutatoren er en glidende strømførende kontakt mellom spenningskilde og spole. Kommutatoren skal i en DC-motor med én spole gjøre slik at kretsen hele tiden er lukket og at strømrretningen skifter retning når det er mest hensiktsmessig. I motorer med flere spoler (eller ingen avgrensede spoler) varierer da hvilke spoler som sendes strøm gjennom til enhver tid. For å forklare nærmere vår egen løsning og kommutatorens funksjon har vi illustrert el-motoren i ved to ulike tidspunkt (se neste side).

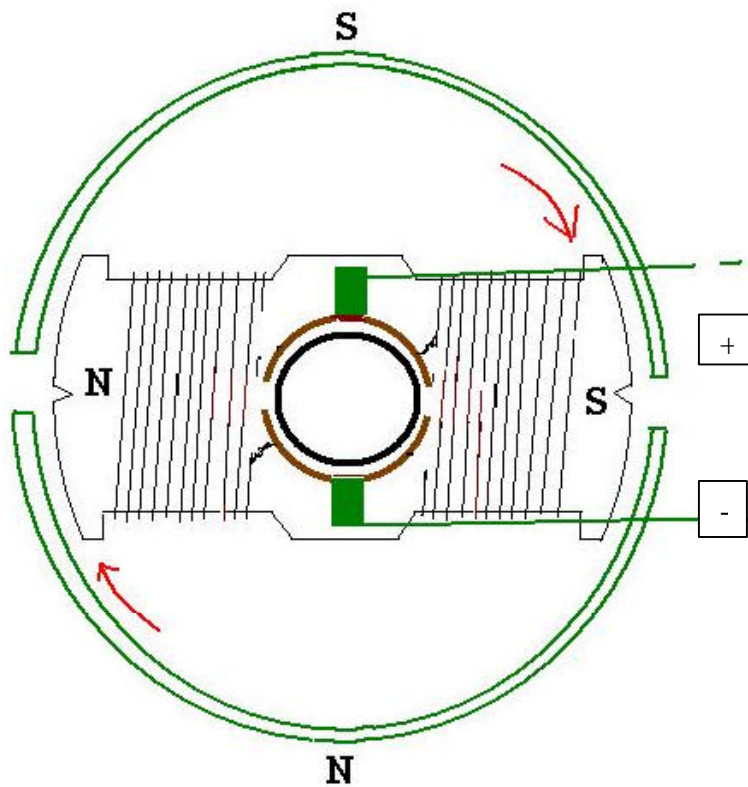


På figuren på neste side ser vi hvordan det induerte feltet i spolen er tidsbestemt og avhengig av hvilken posisjon rotoren står i. Ved høyrehåndsregelen ser vi at det induerte feltet veksler mellom å tiltrekke og frastøte permanentmagnetene hver gang rotoren står vertikalt og horisontalt. Denne løsningen kjennetegner DC-motoren og gjør slik at induert og permanent magnetfelt alltid skaper en drivende kraft i rotasjonsretningen.

Kommutatoren vår hentet vi fra en annen DC-motor, med tolvarmet rotor. For at strømvekslingen skulle virke i til vår motor, måtte vi seriekoble kommutatorplatene i en seksgang, slik at vi fikk to like, sammenhengende kommutatorplater, som på figuren. Dette hjalp Erling Kristiansen ved elektronikkverkstedet oss med. Han loddet sammen platene med tinn uten at det ødela noe for den glidende kontakten mellom kommutator og børster.



Figur 4.3: Likevektsposisjon ved $t = t_0$



Figur 4.4: Vertikal posisjon ved $t = t_0 + (1/4)T$

4.1.7 Når enden er god er allting godt:

Når alt dette var gjort kunne vi endelig teste ut motoren og gjøre fysiske målinger. Til stor jubel fungerte motoren etter to ukers planlegging, diskutering og masing på NTNU-ansatte!



Her er alle komponentene til motoren vår samlet. Monteringen går veldig fort og enkelt, så det er bare å løsne på lokket for de skuelystne.

5. Effekttap:

Vi har en del forskjellige momenter som spiller inn når vi skal regne på effekttap i en elektromotor. Blant annet jerntap, koppertap, luftgaptap, selvinduktans, friksjonstap og gyroeffekttap. Jerntap består av hysteretap og virvelstrømstap. Virvelstrømstapet oppstår i jernet til en elektrisk motor, generator eller transformator. Koppertap er tapet som oppstår på grunn av resistivitet i vindingene til en elektrisk motor, generator eller transformator. Luftgaptap oppstår i mellomrommet mellom stator og rotor. Friksjonstapet oppstår blant annet på grunn av treghet i kulelager i motoren eller mellom børstene og kommutatoren. Utenom luftgaptap og gyroeffekttap vil alle disse energitapene omsettes til varme.

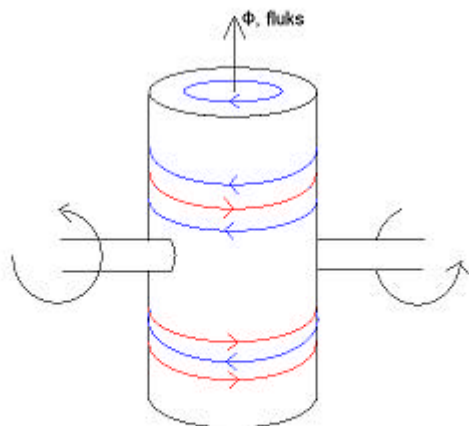
Til sammen utgjør virvelstrøms- og hysteretapene maskinens jerntap. Siden disse er knyttet til frekvensen, vil de være konstante for en gitt hastighet og konstant strøm. Hystereser er ommagnetisering av jernet, dess større frekvens en elektromagnet får dess oftere vil den ommagnetiseres.

5.1 Virvelstrømmer:

Virvelstrømmer, eller eddycurrents, oppstår i vår motor i det bløte jernet omhyllt av spolen. I jernemnet induserer magnetfeltet kraftige virvelstrømmer, som virker som kortsluttede sekundærspoler. De eksisterer bare så lenge det ytre magnetfeltet endrer feltstyrken. Dette er en konsekvens av Lenz' lov som sier at den induserte strømmen motsetter seg fluksforandringen. Mange tenker seg at virvelstrømmene oppfører seg som små tornadoer inne i materialet. Virvelstrømmene setter opp magnetfelter som er rettet mot det ytre magnetfeltet. Strømstyrken i virvelstrømmene er størst helt ute i overflaten av jernstykket vårt, men avtar forholdsvis raskt innover. Inntrengningsdybden blir definert som dybden inn til det stedet der strømstyrken har sunket til 37 % av verdien helt ute i overflaten. Inntrengningsdybden avtar etter som frekvensen øker. Inntrengningsdybden til eddystrømmer kan regnes ut i fra

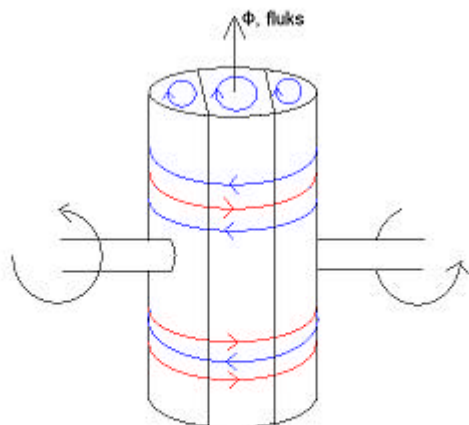
$$d = \sqrt{\frac{2}{\omega\mu_0\sigma}} \quad (1)$$

hvor $\omega = 2\pi f$ og f er frekvensen. σ er materialets ledningsevne. Varmeutviklingen blir størst der strømstyrken er størst. Siden vi har et massivt stykke bløt jern er energitapet som kommer av virvelstrømmer betydelig. Det merket vi da vi lot motoren gå over noe tid for å finne ulike vinkelhastigheter ved forskjellige strømstyrker. Da økte temperaturen til jernstykket vårt betraktelig.



Figur 5.1

Virvelstrømmene, blå linjer på figur 1 og 2, vil bevege seg i y- og x-retning som små virvler. De vil bore seg inn i materialet. De røde linjene er kobbertråden med retning på den påtrykte strømmen. I de moderne elektromotorene er spolen delt opp i mange små plater som er laminert på utsiden slik at de ikke fører strøm seg i mellom. Dette gjør sitt for å redusere virvelstrømmene siden de får mindre areal å boltre seg på, se figur 2.



Figur 5.2

Siden vår spole er laget av en eneste jernklump vil virvelstrømmene ha et stort vrimleområde og dermed vil mye av tilført energi i motoren gå tapt på grunn av virvelstrømmenes bevegelse. Vi har lett mye i bøker, på internett og spurt en del ansatte ved forskjellige institutt ved NTNU, og alle kommet med samme svar. Det å tallfeste slikt

tap som virvelstrøm er nesten ugjennomførbart. Hvis vi hadde visst hvor mye effekt vi får ut av motoren kunne vi ha prøvd å tallfeste noen av de andre tapene, men både luftgaptapet og gyroeffekttapet lar seg vanskelig tallfeste.

5.2 Kobbertap:

Strømgjennomgangen i et materiale forårsaker spenningsfall i samsvar med ohms lov ($V = R \cdot I$) med tilhørende joulsk effekttap, dette er med på å varme opp motoren vår. Dette innebærer videre at viklingene selv varmes opp. Viklingene vi brukte har en maksimal temperatur på 120° . Ledere med positiv temperaturkoeffisient får økt resistivitet i et proporsjonalt forhold til temperaturøkningen. Ved konstant strømgjennomgang betyr dette at det oppstår en selvforsterkende temperaturstignings-/resistivitetsøkningssprosess mens motoren går, noe som igjen øker av maskinens ohmske tap, gjerne kalt «kobbertap». Formelen for sammenhengen mellom resistivitet og temperatur ser slik ut

$$R = R_0 [1 + a(T - T_0)] \quad a = \text{temperatur koeffisient (for kobber} = 0,0068)$$

R= resistivitet, T= temperatur

5.3 Luftgaptap og gyroeffekt:

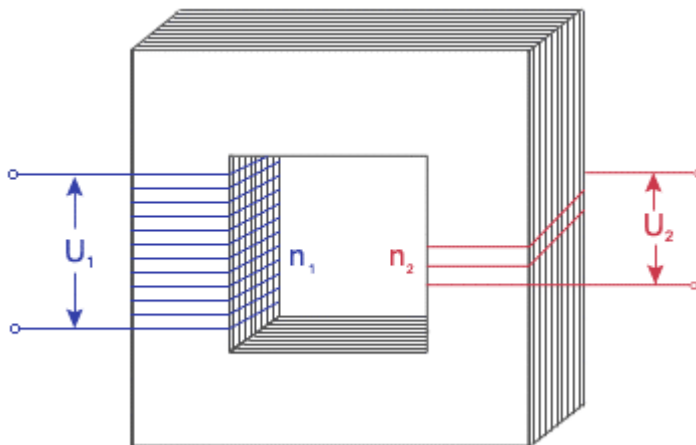
Permeabiliteten til luft er mye mindre enn til gode ferromagneter, rundt størrelsesorden 1 til 1000. Det betyr at evnen til å bli magnetisert og forsterke det totale magnetfeltet er mye større for ferromagneter enn for luft. Det er dette som ligger til grunn når vi snakker om luftgaptap. Siden luftgapet som må være i en elektromotor av bygningsmekaniske årsaker genererer et svakere magnetfelt enn det vi hadde fått hvis det hadde vært magnetiske materialer der, vil det gå ut over dreiemomentet til spolen vår. Det vil bety at vi får mindre kraft (N) av motoren og dermed også mindre effekt (W).

Gyroeffekttapet vil være et tap i form av kinetisk energi. Hvis ikke aksling eller rotor har en masse som er uniformt fordelt vil denne ujevnheten i balansepunkt i enheten skape en gyroeffekt og dermed vibrasjoner i motoren. Dette kunne vi kjenne da motoren gikk. Det er her viktig å ha massesentrum til rotor og aksling nærmest mulig rotasjonsaksen.

5.4 Hysterese:

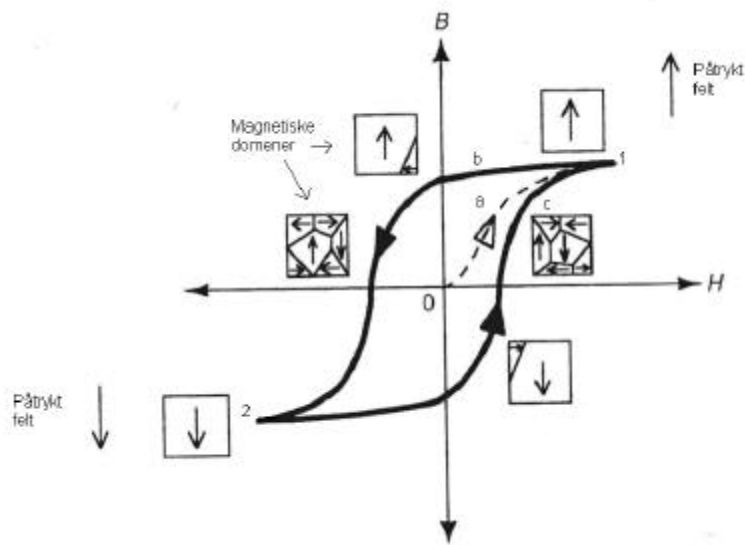
Fenomenet hystereser er ommagnetisering av ferromagnetiske materialer som påvirkes av et ytre magnetfelt. Hystereser er vanskelig å forklare inngående ved bare bruk av ord, så vi skal billedliggjøre fenomenet med et eksempel. Vi kan studere sammenhengen mellom den magnetiske feltstyrke \mathbf{H} og den magnetiske flukstetthet \mathbf{B} inne i et magnetisk materiale, i vårt tilfelle bløt jern, ved hjelp transformator-komponent av bløt jern. Når man skal regne på effekt og ta med energitapet av hystereser med i betraktningen vil det tapet være arealet under hystereser-kurven. Dette vil merkes som varme i det bløte jernet.

Figuren under består av en primærspole U_1 som er viklet rundt en av sidene til materialet vi ønsker å studere. Rundt komponenten er det også viklet en sekundærspole U_2 som har som formål å måle magnetfeltet i firkanten. \mathbf{H} - og \mathbf{B} -feltet varieres ved å variere ved å variere strømmen \mathbf{I} i primærspolen. Den magnetiske intensiteten \mathbf{H} er uavhengig av materialet og direkte proporsjonalt med strømstyrken \mathbf{I} i primærspolen. For denne sammenhengen se egen del om strøm og magnetisering side 13. Sammenhengen mellom \mathbf{B} og \mathbf{H} vil dermed gi oss sammenhengen mellom \mathbf{B} og \mathbf{I} .

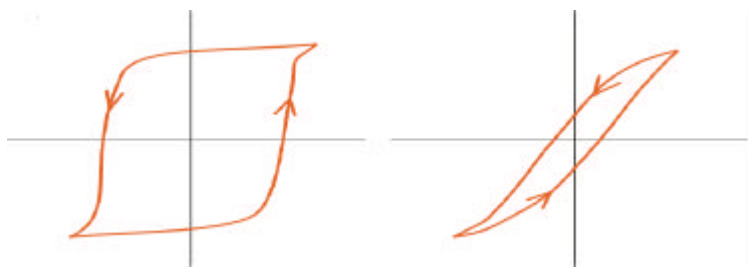


Figur 5.3

Som vi ser av graf 1 vil vi i origo ha \mathbf{I} , \mathbf{B} og $\mathbf{H} = 0$. Vi øker strømmen opp til en viss verdi for deretter å redusere den tilbake til null. Selv om magnetfeltet i utgangspunktet var null, har vi nå fått en permanent magnet med $B \neq 0$, men fortsatt med $\mathbf{I} = 0$. For å redusere \mathbf{B} feltet til null må vi sende en strøm med motsatt retning gjennom primærspolen. Sammenhengen mellom \mathbf{B} - og \mathbf{H} -feltene for et ferromagnetisk materiale er vist på figuren nedenfor. På figuren ser vi at \mathbf{B} -feltet øker raskt langs kurven a opp til en "asymptotisk" verdi (punkt 1). Ved denne verdien har vi oppnådd maksimal magnetisering og får deretter bare en svak økning av \mathbf{B} -feltet som funksjon av \mathbf{H} . For å redusere feltstyrken til $\mathbf{B} = 0$ må det sendes en strøm i motsatt retning, og dette svarer til en verdi $\mathbf{H} = -\mathbf{H}_0$. Kurven fortsetter til vi får maksimal magnetisering i punkt 2. Om strømmen snus igjen, får vi de samme forholdene som fra punkt 1, men med motsatt fortegn. Sammenhengen følger nå kurven c tilbake til punkt 1. Denne kurven, hvor sammenhengen mellom \mathbf{B} og \mathbf{H} avhenger av forhistorien, kalles en hystereser-kurve.



Figur 5.4



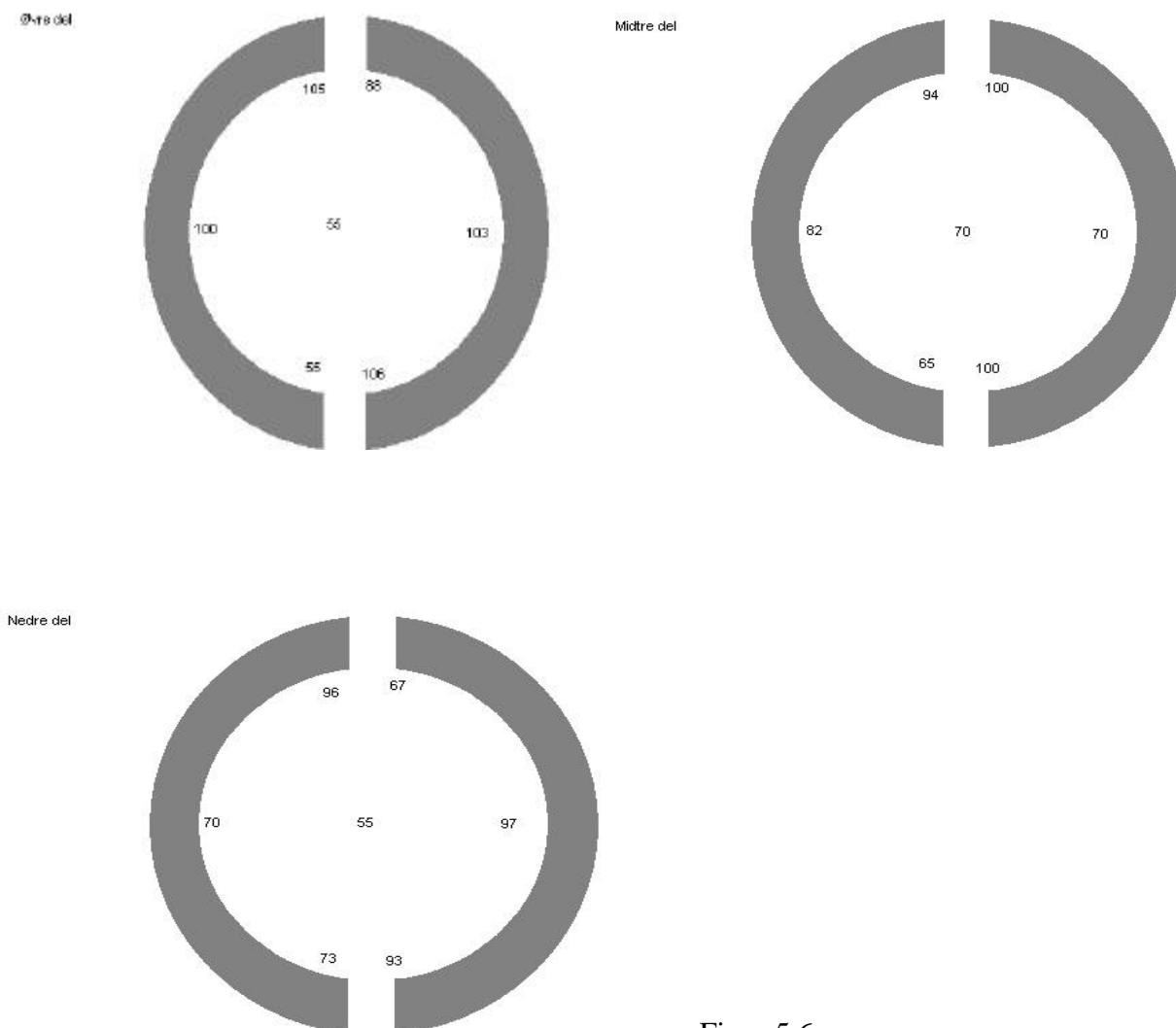
Figur 5.5

Hardt magnetisk materiale har en bredere hysteresekurve enn et bløtt magnetisk materiale. Det vil si at det skal stor endring i \mathbf{H} -feltet, eller sterk strømstyrke, til for å endre magnetfeltet i et hardt ferromagnetisk materiale. Det betyr at permanentmagneter må lages av hardt ferromagnetisk materiale og at kjernen i en elektromotor må lages i bløtt.

En permanent magnet kan avmagnetiseres ved hjelp av ytre magnetfelt, men dette kan også gjøres ved å varme materialet opp til en kritisk temperatur som varierer fra ett materiale til ett annet. Denne temperaturen kalles curie-temperaturen. For jern er $T_c = 1043 \text{ K}$. Når et ferromagnetisk materiale varmes opp til over sin curie-temperatur, blir det paramagnetisk. Denne overgangen finner sted når den termiske bevegelsen blir så stor at den dominerer over de kreftene som opprettholder den spontane innretningen. Paramagneter er materialer hvor dipolene vil ha tilfeldig orientering. Uten ytre påvirkning vil det totale magnetiske feltet i disse materialene være null.

I tillegg er det verdt å nevne at de magnetiske domenene i harde magnetiske materialer over tid vil gjennomgå spontane forandringer også ved romtemperatur. Dette så vi da vi målte magnetfeltet til vår permanentmagnet som er noen år gammel. Det varierte stort i styrke, helt fra 70mT til 97mT i lengderetning. Vi ble fortalt at magneten skulle stråle rett

ut i horisontal retning fra nord til sør og i teorien skulle være homogen. Da vi målte feltstyrken B viste det seg at den var langt fra homogen, se figurer på neste side.



Figur 5.6

Hvis vi skulle ha målt hysteresefeltet i vår motor hadde det vært med veldig mange antakelser i regnestykket. Vi hadde blitt nødt til å ta elektromagneten i motoren og målt hysteresekurven den produserer for deretter å multiplisere det med en ukjent konstant for å få riktig verdi. Dette siden ommagnetiseringen inne i elektromagneten er annerledes enn utenfor. Det kan også tenkes at det permanente magnetfeltet påvirker hysteresekurven slik at en eventuell måling av hysteresen utenfor motoren ville vært annerledes enn inne i motoren, slik forholdene er i realiteten. Andre svakheter i en slik beregning er at vi har inhomogent felt, at vi må ta utgangspunkt i en "uendelig lang spole" og at den er fylt med en geometrisk perfekt sylinder av magnetiserbart materiale med kjent permeabilitet. Vi bare antok at relativ permeabilitet i vår elektromagnet lå rundt verdien for jern, altså μ_r lik ca 200. Videre kommer usikkerhetsmoment i målinger og utstyr.

5.5 Effekttap av selvinduktans:

Når man skal snakke om effekttap er det også viktig å nevne "generatoreffekten" eller selvinduktansen. Ettersom de elektriske motorene bygges på samme prinsipp som generatoren, kan det tenkes at den elektromagnetiske motoren også inducerer en spenning. For å forstå dette må vi se litt på Faraday's lov.

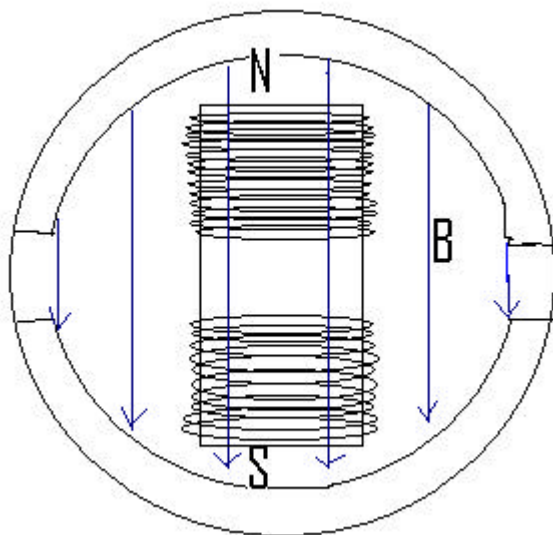
Faraday's lov gir en fundamental sammenheng mellom elektrisitet og magnetfelt. Ved å endre på magnetiske felt rundt strømførende sløyfer, kan det induseres en spenning (Emf). Den induerte spenningen i en sløyfe er altså lik den tidsderiverte av fluksen gjennom den. Videre er den induerte spenningen for en solenoide, lik den tidsderiverte av den totale fluksen gjennom alle "sløyfene".

Formelen lyder: $V_{(indusert)} = -N \frac{d\Phi_B}{dt}$, der N er lik ant. viklinger.

Eksempel: Vår elektromotors selvinduktans ved 3000 RPM:

Magnetisk fluks, Φ , gjennom en lukket sløyfe er definert som $B \cdot A$, der B er et ytre felt og A er arealet av sløyfen. Anta at det permanente ytre magnetfeltet i vår motor går i helt rette linjer fra nordpol til sørpol og at feltet er homogent med verdi 0,075 Tesla overalt. (Dette blir ifølge våre målinger middelverdien rundt området der spolen roterer – se måleresultat på forrige side) I tilstanden det går mest fluks gjennom spolen, er fluksen lik NBA:

Figur 5.7



Vi har da:

$$N = 160$$

$$D = 0,017 \text{ m}$$

$$A = \pi \left(\frac{D}{2}\right)^2 \text{ m}^2$$

$$B = 0,075 \text{ T}$$

Merk: På figuren ovenfor sees det bort ifra at vi sender spenning gjennom spolen og inducerer et felt som overgår det permanente når det bløte jernet er til stede. Om vi hadde tegnet inn dette mens motoren

var i gang, ville den totale fluksen (påtrykt + induert) alltid pekt i retning med vertikal komponent imot den permanente. Det vil gjøre det enklere å forestille seg endringen av

de permanente feltlinjene gjennom spolen om vi ser bort ifra dette. Den samme endringen i fluks vil være tilstede uavhengig av om vi sender strøm gjennom spolen eller ei.

Fluksen gjennom spolen i denne tilstanden blir da:

$$\Phi = N \cdot B \cdot A = N B p \left(\frac{D}{2}\right)^2 = 160 \cdot 0,017 \cdot 3,14 \cdot (0,017/2)^2 = 0,0001542682,$$

altså ca 154 **mWb**

Straks vi kobler til en spenningskilde og setter i gang motoren, vil spolene begynne å rotere i magnetfeltet og fluksen vil endres periodisk:

Figur 5.8

-Vi definerer vinkelen mellom positivt B-felt, sløyfeplatene.



q til å være vinkelen og normalen til

Fluksen til enhver tid intuitivt:

gjennom spolen blir da

$$\Phi(t) = \Phi_0 \cdot \cos(\omega t)$$

og den induserte spenningen blir videre: $V_{(indusert)} = \frac{d\Phi(t)}{dt}$

$$= -\omega \Phi_0 \cdot \sin(\omega t)$$

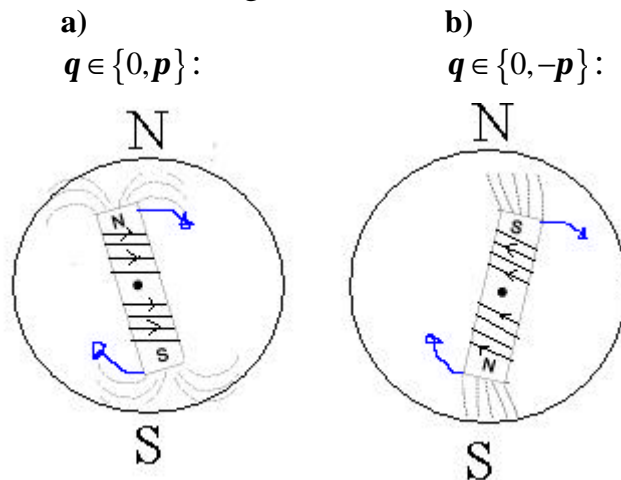
Hvis vi da bruker vår høyeste testede vinkelfrekvens på 3000 omdreiinger i minuttet ved 7,9 Volt som påtrykt spenning, altså $\omega = 100\pi$ rad/sek, kan vi beregne denne spenningen som induseres i spolen i motsatt retning av den påtrykte som funksjon av tiden.

$$V_{(indusert)} = -0,048 \cdot \sin(100\pi \cdot t)$$

Og tatt i betraktning at positiv og negativ retning endres for hver halve periode; akkurat idet spolen står vinkelrett på magnetfeltet og fluksen er lik null, så blir spenningen alltid med samme fortegn.

Retningen på spenningen blir da ifølge Lenz' Lov den retningen som gir en kraft som igjen motvirker endringen (se neste side:)

Figur 5.9

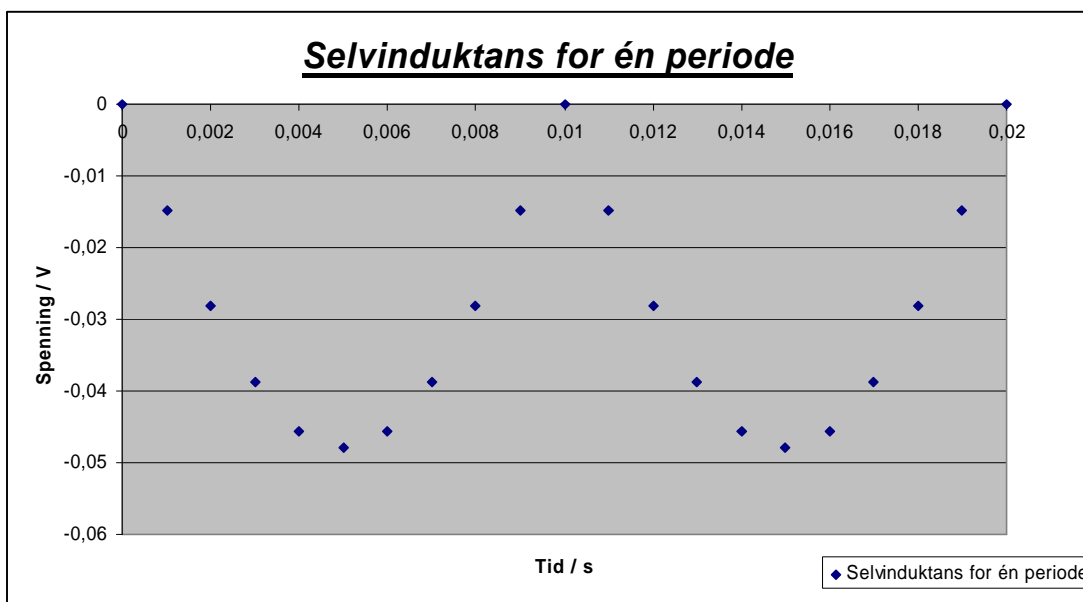


Av tegningen ser vi at spenningen må skifte fortegn akkurat idet spolen står vinkelrett på det permanente magnetiske feltet. Akkurat ved dette tidspunktet vil også kommutatoren på akslingen gjøre slik at strømmen som sendes ut av spolen skifter fortegn. Med andre ord: Spenningen vil ikke skifte fortegn.

Spenningen vil da variere over tid, og imot den påtrykte spenningen når vi kobler til en likestrømskilde.

Vi forutsetter at spenningskilden kjører en spenning på 7,9 og at den ikke kompenserer for ytre påvirkninger på spenningen. Vi forutsetter også at selvinduktansen er den eneste ytre spenningspåvirkningen. Da kan vi fremstille en tenkt endring av spenningen gjennom spolen i en graf:

$$V_{\text{(indusert)}} = -0,048 \cdot |\sin(100p \cdot t)|$$



Graf 5.1

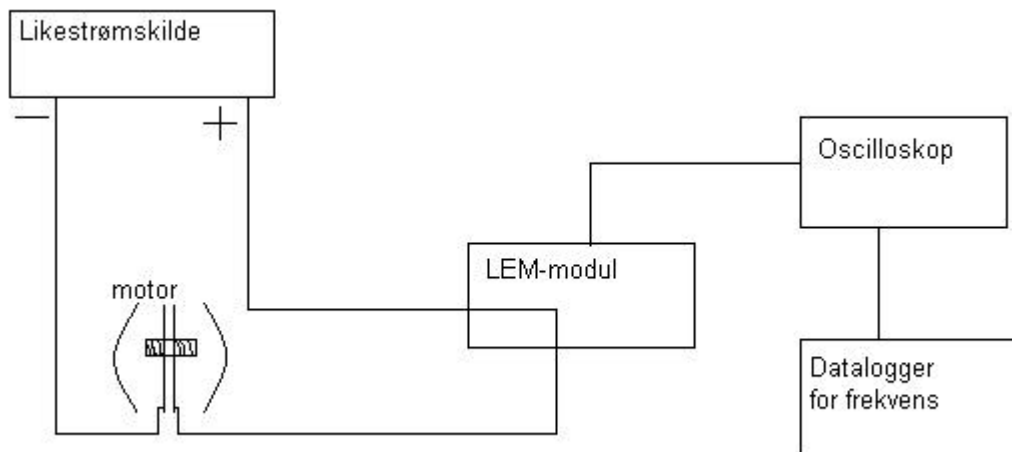
Vi ser at spenningsendringen 0,048 Volt ikke blir betydelig i forhold til den påtrykte spenningen på 7,9 Volt. Strømmen gjennom spolen blir da ikke særlig svekket (max 0,6%), og dette blir igjen ikke noe særlig tap av effekten. For at selvinduktansen virkelig skal ha noe å si må man ha en mer effektiv motor enn vår. En motor som bruker lite spenning og har stor vinkelfrekvens blir i større grad påvirket av denne induserte motspenningen. Ta for eksempel en tilsvarende motor som kjører på 2 Volt og har da ti ganger så høy omdreingshastighet som vår. Da vil den i følge formelen vår: $V_{(indusert)} = -\omega\Phi_0 \cdot \sin(\omega t)$, få ca en halv Volt induert og en maksimal svekkelse på nesten 25%.

Å regne nøyaktig på selvinduktansen er svært komplisert. I vår elektromotor måtte vi nok tatt hensyn til bl.a. økningen i resistansen i lederne når temperaturen økte og virvelstrømmene som også danner en mer virkningsfull "selvinduktans". For en ideell motor uten resistans i komponentene, ingen friksjon og med en spole fylt med høyt magnetiserbart stoff som ikke produserer virvelstrøm (eddy-current), så ville selvinduktansen spilt en mye større rolle for effektetapet. I et slikt tilfelle ville selvinduktansen hindret motoren i å øke sin vinkelfrekvens i det uendelige. Selvinduktansen øker proporsjonalt med vinkelfrekvensen og vil til slutt nå igjen den påtrykte spenningen. Det kan da tenkes at de som lager de mer avanserte elektriske motorene må ta mer hensyn til selvinduktans i sine beregninger.

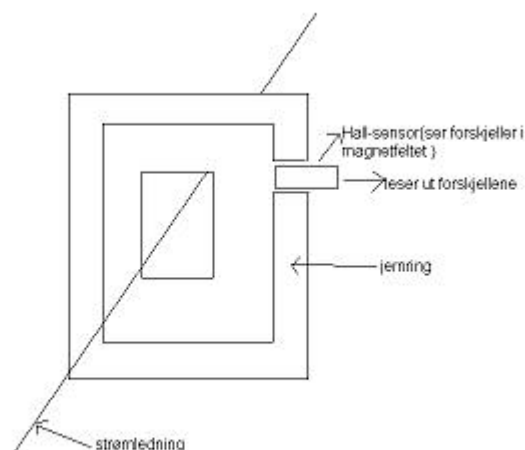
6. Målinger og resultat:

6.1 Utregninger av sirkelfrekvenser:

Figur 6.1

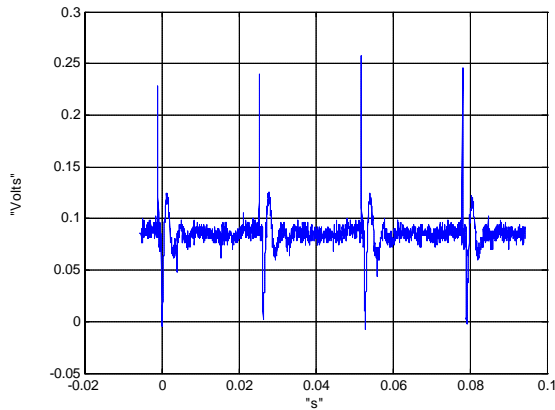


Slik koblet vi utstyret sammen da vi skulle finne farten til vår motoren. Ledningen er her ikke fysisk koblet til LEM-modulen, den går derimot gjennom et rør som detekterer forandringen i Hall-feltet. Dette forsøket gjennomførte vi på en lab på elektrobygget med hjelp av stipendiat Eirik Forr. Nedenfor har vi laget en forenklet modell av LEM-Modulen. LEM-Modulen fungerer slik at magnetfeltet som settes opp rundt lederen vil variere på grunn av kortslutningen i motoren vår som oppstår to ganger per periode. Hallsensoren oppfatter disse forandringene i magnetfeltet. Hallsensoren er koblet til et oscilloskop som igjen er koblet til en datalogger.

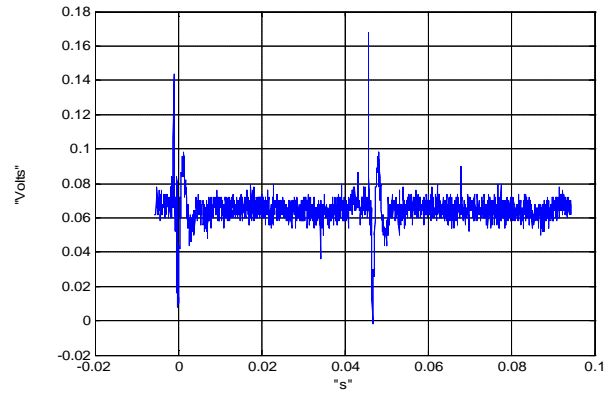


Figur 6.2

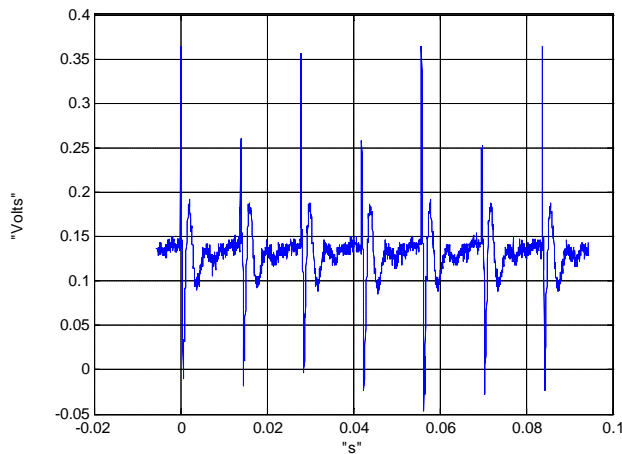
Toppene vi ser på grafene er hver gang kretsen i motoren kortsluttes, dette vil skje to ganger pr omdreining, så en periode er tiden mellom annenhver topp. Teksten under grafene sier hvor mye strøm og spenning vi puttet inn i motoren. Verdiene på y-aksen representerer den induerte spenningen fra ledningen gjennom LEM-modulen.



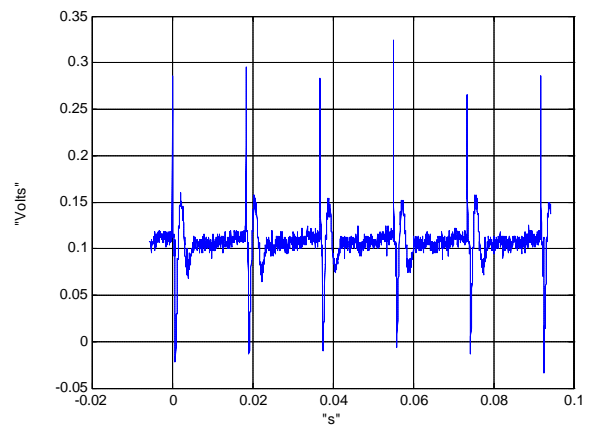
Graf 6.1: 1,96 A og 3,4 V



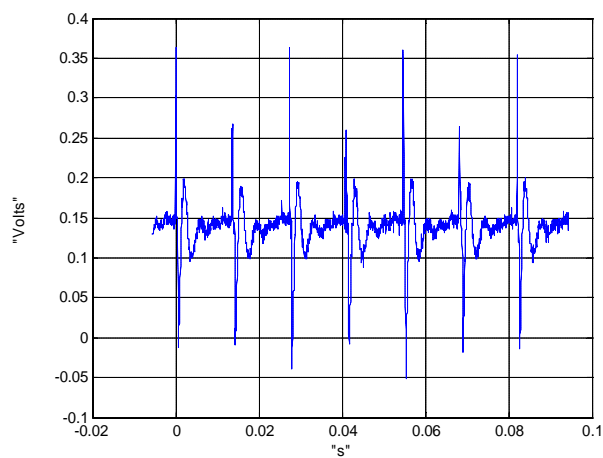
Graf 6.2: 1,54A og 2 V



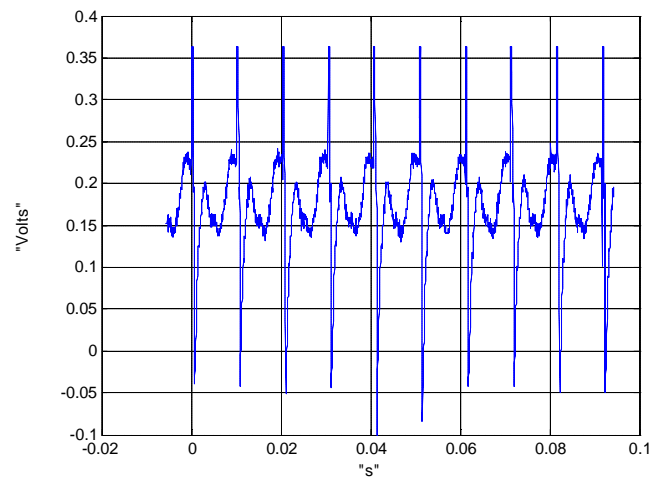
Graf 6.3: 3 A og 5,5 V



Graf 6.4: 2,5 A og 4,5 V



Graf 6.5: 3,25 A og 6V



Graf 6.6: 4,15 A og 7,9 V

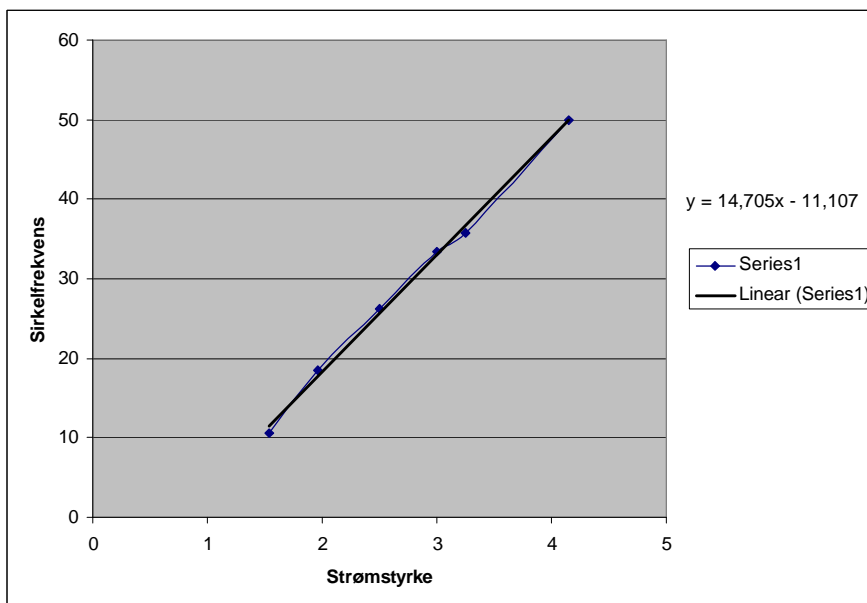
Strømstyrke[A]	Spenning[V]	Periode[s]	Sirkelfrekvens[s ⁻¹]	r/min
1,54	2	0,094	10,638	638,298
1,96	3,4	0,054	18,519	1111,111
2,5	4,5	0,038	26,316	1578,947
3	5,5	0,030	33,333	2000,000
3,25	6	0,028	35,714	2142,857
4,15	7,9	0,020	50,000	3000,000

Tabell 6.1

Etter å ha lest av periodene til de enkelte strømstyrkene førte vi disse inn i tabell 2 og kunne dermed regne ut sirkelfrekvensen til motoren fra:

$$f = \frac{1}{T}$$

På det meste klarte vi å få ut en fart på 50 omdreininger i sekundet, eller 3000 r/min, men dette kunnet vært høyere hadde vi hatt kraftigere utstyr som kunne levert mer enn 4,15 A. Denne farten kan derimot ikke være noen arbeidshastighet på motoren, da isolasjonsbelegget på koppertråden vil smelte ved ca 120 grader, og kretsen vil da kortslutte. Vi ble anbefalt å ikke bruke en strømstyrke på mer enn 2 A over lengre tid.



Graf 6.7

Vi kan se av denne grafen, som viser sirkelfrekvensen som funksjon av strømstyrken, at det er en lineær sammenheng mellom dem. Grunnet vår noe enkle oppbygning av

motoren måtte vi starte den manuelt, og da måtte vi ha en strømstyrke på min 2 A. Vi måtte i tillegg kjøre en strøm på over 1,5 A inn i motoren for at den skulle fortsette å dreie. Siden vi ble anbefalt å ikke tilføre inn mer enn 2A, vil denne motoren ikke kunne produsere mye effekt.

6.2 Tilnærming av magnetfelt fra spolen:

Uten jern i solenoiden vil spolen sette opp et magnetfelt, og vi antar her under utregningene at vi har en uendelig lang spole med jevn viklingstetthet.

Fra Amperes lov har vi:

$$\oint_l B \cdot dl = \mu_0 I$$

$$\rightarrow Bl = \mu_0 n l I_0$$

$$B = \mu_0 n I_0$$

Hvor n er antall viklinger pr lengdeenhet og I_0 er strømmen vi putter inn. I spolen vår har vi til sammen 160 viklinger. Og hele spolen vår er 46 mm lang. Under utregningene her regner vi på magnetfeltet med en strøm på 4,15 A.

$$\rightarrow B = 4\pi \cdot 10^{-7} \cdot \frac{160}{0,046} \cdot 4,15 = 0,018T$$

Regner så ut det totale magnetfeltet vi får fra spolen med jern i kjernen. Dette svaret vil da bli lik summen av feltet fra strømmen i viklingene, og feltet fra overflatestrømmen i jernet.

Igjen, bruker Amperes lov;

$$\oint_l B \cdot dl = \mu_0 I$$

$$Bl = \mu_0 (n l I_0 + l j_m)$$

Hvor j_m er flatestrømstettheten(strøm pr lengdeenhet) til overflatestrømmen til jernet.

$$B = B_0 + \mu_0 M$$

B_0 er magnetfeltet fra strømmen i viklingene og M er et begrep vi innfører kalt magnetisering. Her er størrelsen på magnetiseringen ikke noe annet enn størrelsen på overflatestrømstettheten til jernet inni spolen.

Vi kan skrive

$$M = c \frac{B_0}{\mu_0}$$

hvor c er stoffets magnetiske susceptibilitet, og vi får da videre

$$B = B_0 + \mu_0 M = (1 + c)B_0$$

Vi har innført en størrelse relativ permeabilitet μ_r definert ved:

$$(1 + c) = \mu_r$$

Vi får da et uttrykk for det totale magnetfeltet fra spolen og jernet:

$$B = \mu_r B_0 = 200 \cdot 0,018 = 3,6T$$

Noe som i og for seg er et meget kraftig magnetfelt, større enn magnetfeltet fra permanentmagneten, men det må det jo også være for at motoren vår skal kunne fungere.

En kommentar til disse resultatene er at de muligens i en liten grad stemmer med vår motor, de er kun teoretiske verdier. For eksempel med viklingstettheten som i teorien regnes ut i fra at det kun er et lag viklinger, har vi i vår motor fire lag. I tillegg var det meget vanskelig for oss å kunne måle noe magnetfelt i sammensatt motor, noe som kunne ha bekreftet/avkreftet våre utregninger.

7. Noen elektromotorer i dag:

Det finnes flere typer elektromotorer i dag, alt avhengig av hva den skal brukes til. Motorene kan være like utvendig, men er som regel forskjellige i måten de er bygd opp på innenfra og hvor mye effekt man kan få ut av dem. Fordelen med en DC-motor kontra en AC-motor er at DC-motoren er meget enkel å kontrollere farten på. Den egenskapen at energikilden er koblet direkte til feltet i motoren tillater en meget presis spenningskontroll, hvilket er nødvendig med fart og momentavhengige oppgaver. DC-motoren brukes ofte på kjøretøyer, i for eksempel en golfbil. I biler brukes DC-motorer med permanentmagneter for eksempel i vindusheiser, vindusviskere og startmotorer. Denne typen elektriske motorer produserer et høyt dreiemoment ved lav fart, og er selvbremsende ved frakobling av strøm. En ulempe med denne type motor er at de ikke kan operere over lengre tid med høy fart, fordi de da vil overopphete og dermed destruere seg selv.

Elektromotorer er i stadig utvikling for å kunne være et fullverdig alternativ til brenselmotorer. Det tyske firmaet Siemens har klart å bygge den første høytemperatur-superledende motoren. Den superledende effekten har vært kjent lenge, hvor den elektriske motstanden forsvinner når temperaturen i visse materialer senkes til en bestemt temperatur. Resultatet er store strømtettheter i de strømførende lederne og dermed oppnås mye større magnetiske feltstyrker. I dag ligger strømtettheten i superledere typisk ti ganger høyere enn i kopperledere. På grunn av den store strømtettheten kan rotoren til en slik motor bygges vesentlig mindre enn for en konvensjonell motor. Prototypmotoren til Siemens har en maksimal strømtetthet på ca 75 A/mm^2 , mens koppertråd normalt blir belastet med ca $4\text{-}8 \text{ A/mm}^2$. Superledere tåler kun små vekselstrømfelt, dermed er det per i dag bare mulig å oppnå superledning i likestrømskretser, og for å oppnå et tilstrekkelig magnetiseringsfelt må lederen holdes avkjølt ned mot 25 K . Selv om Siemens motor ikke var designet for høy virkningsgrad, viser testresultater at den har lavere tap enn konvensjonelle motorer, og det også når en tar med den nødvendige nedkjølingen i regnestykket.

I bilindustrien har de største selskapene sett alvoret i forurensingsproblemene til brenselmotorer, og innsett at disse motorene er mindre aktuelle når vi tar miljøspørsmålet med i betraktning. Toyota har prøvd å være i forkant av utviklingen av hybridbiler. De har laget en bil med navn Prius som har både bensin- og elektromotor. Ved vanlig kjøring bruker bilen kun elektromotoren, men hvis en gir på vil også bensinmotoren hjelpe til. Prius bruker $0,4$ liter på mila, og har en meget bra energiutnyttelse. For eksempel vil all bremsing lade batteriene, hjulbremsene er elektriske så bremsekraften kjøres direkte i retur som elektrisitet.

En av verdens største likestrømsmaskiner på 143 tonn står i verkstedet til Component Service på Brakerøya i Drammen. Der gjennomgår den fullstendig overhaling



Rotoren har en diameter på 4,1 meter. Ytelsen er på 3080 kilowatt og motoren roterer med 60 omdreininger i minuttet.

Vår er ikke fult så stor...

8. Kommentarer:

Vi er godt fornøyde med motoren og prosjektet vårt. Det eneste vi er misfornøyd med er at vi ikke fikk målt effekten motoren vår kunne yte. Da kunne vi blant annet prøvd å fastsette energitapet vi får av virvelstrømmer i rotoren vår. Vi forhørte oss ved mange fakulteter og ved mange verksteder på NTNU og fikk noen ideer om hvordan effektmålingen kunne gjøres, men ingen hadde noen apparaturer som kunne hjelpe. En av ideene var å feste en hyssing fra akslingen vår med et masselement ved enden og løfte det opp, for så å ta tiden motoren brukte på løfte massen en viss lengde. Ut fra dette kunne vi ha sammenlignet vår effekt med 1 hestekraft, altså effekt nok til å kunne heise 75 kg 1 meter opp i luften på ett sekund. En annen idé gikk ut på å feste en drill til akslingen og måle strømmen vi får generert. Problemet da ville vært å vite eksakt hvor mye effekt drillen gir oss, og om det er direkte sammenheng mellom effekttapet vi får om vi bruker motoren som en generator, eller nettopp som en motor. En tredje mulighet var å bruke en momentnøkkel til å finne dreiemomentet ved en viss spenning. Men vi fikk aldri tak i en momentnøkkel. Uansett så føler vi at teorien vi nå kan gir oss nok oversikt til å kunne diskutere effekttap kvalitativt, med eller uten målinger.

9. Kilder:

Bøker:

- Bok: Generell fysikk Bind 2 Varmelære og elektromagnetisme (LHL)
- Bok: Electronic Motor Data Handbook, Martin Clifford
- Bok: Modeling and High – Performance Control of Electric Machines, John Chiasson
- Bok: Brushless Permanent – Magnet and Reluctance Motor Drives, T. J. E. Miller
- Internett

Nett:

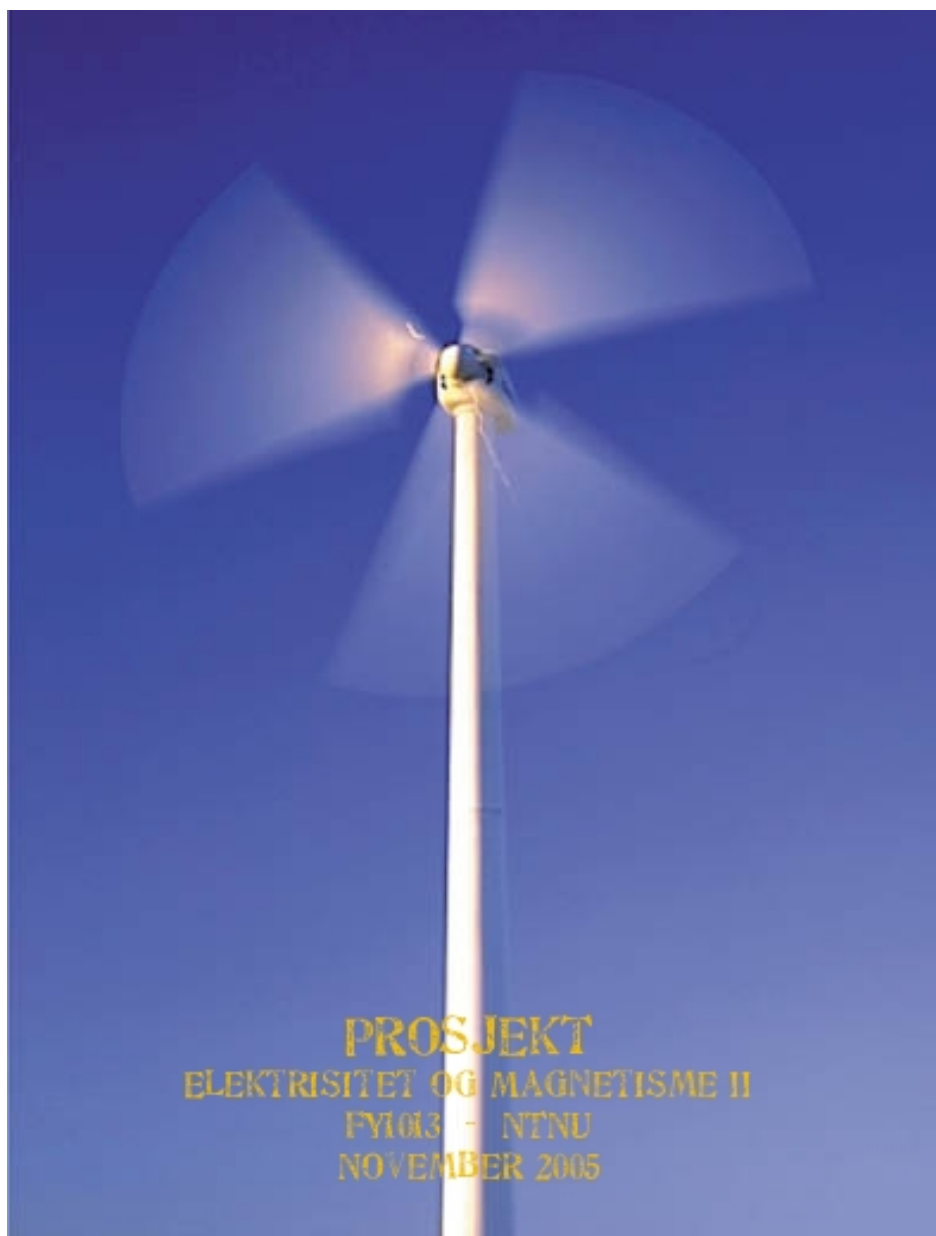
- <http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html>
- www.fys.uio.no/studier/kurs/fy040/teori/magnetisme.ppt
- <http://www.elsiden.no/eltekn/magn/Kap%2054/5.HTM>
- <http://www.reliance.com/mtr/mtrthrmn.htm>
- <http://www.iet.aau.dk/~por/teaching/emsd7e11/mm3.pdf>
- http://folk.uio.no/ogm/fys1120/Oblig2_05.pdf
- <http://www.dself.dsl.pipex.com/MUSEUM/POWER/eleceng/eleceng.htm>
- http://www.museumsnett.no/ntm/no/utstillingene/stroem_lys/elmot.htm
- <http://www.howstuffworks.com>

med mer.....

Takk til

- Elektroverksted Institutt for Fysikk NTNU ved Arne Moholt, Knut Gjervan og Erling Kristiansen
- Mekanisk verksted Institutt for Fysikk NTNU Arnulf Bjøldstad
- Elektroverksted Institutt for Elkraft NTNU
- Professor Roy Nilsen Institutt for Elkraft NTNU
- Stipendiat Erik Forr Institutt for Elkraft NTNU
- Professor Thor Bernt Melø Institutt for Fysikk NTNU
- Amanuensis Jon Andreas Støvneng Institutt for Fysikk NTNU

VINDKRAFTEN UTNYTTES



OLAV ANDREAS SÆTHER OG NILS BJARNE GUNNUFSEN



Abstrakt

Vinden bærer med seg store mengder energi. De siste drøye tusen år er denne energien blitt utnyttet av mennesker, i størst grad i kornmøller, pumper, skipsfart og i den senere tid også til elektrisitetsproduksjon. Frem til forrige århundre var Norge en stor vindnasjon med sin store flåte av seilskuter.

Vindenergien blir utnyttet ved å overføre den til mekanisk bevegelsesenergi. Dette blir gjort ved hjelp av seil som gir vinden stor motstand. I vindmøller festes seilene til en roterende aksling, og det finnes vindmøller til alle formål som roterer med både horisontal og vertikal akse. Den horisontale er i dag mest utbredt.

I elektriske vindmøller gjøres kraften i rotasjonen om til elektrisitet i generatoren ved magnetisk induksjon. En roterende magnet omgis av elektromagneter tilknyttet det elektriske nettet. Når denne roterer endres magnetfeltet rundt den, slik at det blir induert elektrisk spenning ut på elektrisitetsnettet. Det vanlige er å bruke en generator som produserer strøm i tre forskjellige faser, og da i de aller fleste tilfeller den asynkrone varianten, som inneholder en kortslutningsrotor. Dette fordi vinden yter en varierende effekt. Derfor bruker man også i stor grad indirekte tilknytning til elektrisitetsnettet.

Historisk er vindmøller et nytt fenomen i det Norske landskapet. Mens Nederland og andre land på kontinentet tok i bruk vindmøller i kverning av av korn foretrakk nordmenne å temme vannets krefter og bygge små møllehus i elver og bekkefar. Dette tok vi med oss videre inn i den elektriske tidsalder, og 99% av vår elektriske kraft kommer i dag fra vandkraftanlegg. Mens Norge har bygd vannkraft har land som Danmark og Tyskland måtte satse på metoder som ikke stilte krav til fjell og fossefall. Vindkraft ble et naturlig valg.

I Norge opplever vi nå at kraftbehovet øker, og det er ikke ønskelig med flere store oppdemninger av elver og fossefall. Vi blir altså nødt til å satse på andre former for miljøvenlig energiproduksjon, og vindkraft er et av flere alternativer. Den tekniske utviklingen har komst langt, og prisnivået sunket betraktelig. Offentlige rapporter antyder at det finnes potensial til å doble Norges kraftproduksjon ved integrasjon av vindkraft alene. Investeringslysten hos kraftselskapene er stor, men kapasiteten på det eksisterende kraftnettet setter begrensninger. Skal hele potensialet hentes ut trengs det politisk vilje til statlige utbygginger.

Videre medfører den norske geografien med holmer, skjær og fjell betydelig større kostnader til fundamentering og utbygging forøvrig, sammenlignet med land som har åpne slettelandskaper. Det antydes at en utbygging av vindkraft i dag vil gi en pris på om lag 27-34 øre pr kWh, noe som ligger over dagens spotpris. Men forventninger i markedet er at utbyggingskostnadene vil fortsette å synke de kommende år, ettersom man får mer erfaring og ytterligere ny teknologi.

Innholdsfortegnelse

Abstrakt.....	2
Innledende.....	4
Historikk.....	4
Vikinger.....	4
Persere.....	5
Kinesere, tibetanere og europeere.....	6
Amerikanske vindroser.....	7
Teknisk.....	7
Rotorbladene.....	7
Generatoren.....	9
Trefaset synkrongenerator.....	12
Trefaset asynkron kortslutningsgenerator.....	14
Indirekte nettilkobling.....	14
Transformator.....	15
Kjøling.....	15
Anemometer.....	15
Lokale vindforhold.....	16
Vindens effekt.....	17
Betz' lov.....	18
Vindmølleparker.....	19
Norske tilstander.....	19
Potensialet for vindkraft i Norge.....	19
Eksisterende vindkraft, og planer for vindkraft i Norge.....	20
Offshore vindturbiner.....	23
Integrasjon av vindkraft i eksisterende nett.....	24
Vindkraft og energilagring.....	26
Konkluderende.....	28
Kildehenvisninger.....	32

Innledende

En åker med halvmodne havrekorn. Vinden feier gjennom den. Mennesket ser hvordan kornene bøyer seg etter tur, og danner seg i bølgeformasjoner i stråmassivene. I det ene øyeblikket befinner en slik bølge seg ett sted. I det neste et annet. Med en annen fart. I ny retning.

Vinden er veldig inkonsekvent både når det gjelder styrke, hastighet, tidspunkt og retning. Det eneste man visste om den, var at den alltid var der, at den var sikker og kraftful. Og man fant etterhvert ut flere måter å gjøre seg nytte av energien den bar på.

Med dette prosjektet ønsker vi å se på hvordan mennesket gjennom tidene har benyttet seg av kreftene som finnes i vinden, og hvordan vinden i dag brukes til å lage elektrisitet. Vi vil også vurdere Norges fremtid som vindkraftsprodusent.

Historikk

Vikinger

Naturfenomenet vind har dermed betydd mye for menneskenes sivilisasjoner. Seil har fraktet folk i båter til helt ukjente havner og kontinenter. Et folkeslag fra Skandinavia var rundt år 1000 beryktet på verdenshavet for sine for samtiden svært så stabile seilbåter som de brukte på plyndringstoktene sine. I tillegg til motivasjonen fra sin klokkeetro på lovdordene i mytologien, var vikingene godt bereiste i Europa, og hadde god kunnskap om båtkonstruksjon og navigering.

Litt lenger frem i tid, for et par hundre år siden, fraktet de stolte seilskutene folk og fe over havene med last og passasjerer. I seilskipsnæringen var Norge en av de virkelig store aktørene. Særlig langs kysten sør og øst i landet hadde man ypperlige forhold for å drive skipsverft. Forekomstene av tømmer var store, og det var enkelt å føre tømmeret til kysten med de mange vannveiene. En stor prosentandel av seilskipene som seilte på sjøen på verdensbasis hadde derfor hjemhavn i Norge.

Tiden med seilskutene var over da dampmaskinen gjorde innpass. Men det betydde ikke slutten på å utvinne vindenergi, slik mange dengang trodde. Allerede på da skuta Fram gjorde sin banebrytende polferd, var vinden tilbake på banen her til lands. Skipet hadde fått installert en elektrisitetsproduserende vindturbin som sørget for belysningen ombord. Selv den dag i dag spekulerer man i å la kontainerskip ha seil for å spare dyrebart drivstoff til motorene.

At det var på det åpne hav at man først utnyttet vindkraften her til lands var nok mer tilfeldig. Men det kan være nyttig å tenke over hva som skiller vinden til havs fra den på land. Det er nemlig store friksjonskrefter ute og går når vinden blåser. Vinden bremses av alt som står i dens vei, være seg skjærgård, trær, busker, hus, vindmøller, fjell osv. , og de stedene på jordens overflate hvor man har minst friksjon over større områder er da nettopp sjøene. Med andre ord er det der de kraftmessige forutsetningene for å utnytte vind i teorien er best.

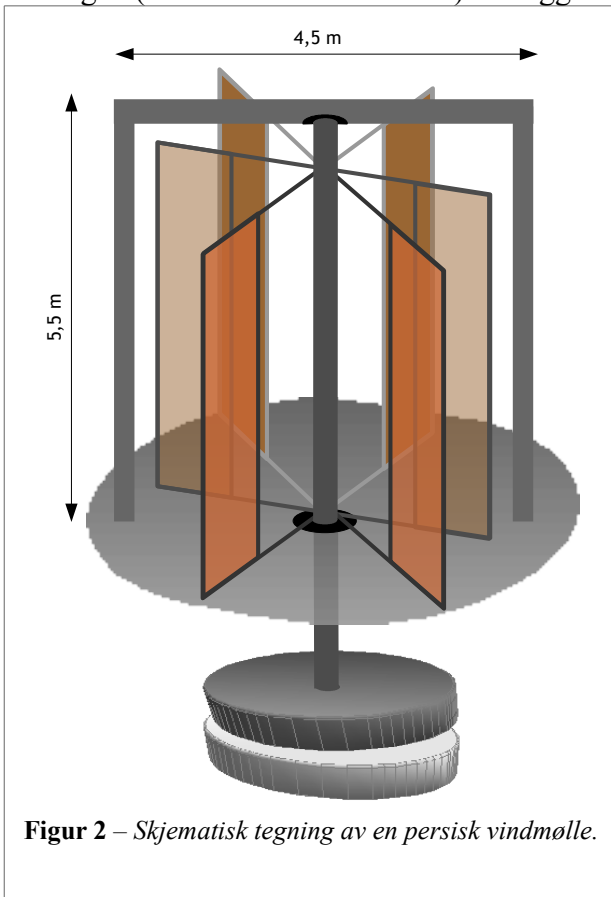
Persere

Men det blåser da riktignok på land også. De første sporene på bruk av vindenergi har vi fra Persia rundt år 950. Flere geografer fra den tiden har i sine notater nevnt hvordan folket i Seistan lot vinden drive møller som hentet vann opp fra kanaler under jorda.

Det ble rapportert om flere måter å konstruere møller på i Persia allerede den gang. Den ene typen forsvant med tiden, mens den andre eksisterte frem til vår tid (Man regner med at den siste ble satt ut av drift en gang på 1970-tallet). Mølletypen var et vanlig syn i et stort område som blant annet omfatter Iran, Herat og Afghanistan. Det mest kjente mølleområdet i regionen er Neh, hvor man hadde en linje på hele 75 slike møller.

Mølla (se figur 1) hadde til hensikt å male opp

mais. Den ble satt sammen slik at møllestenen var plassert under skovlene. Dette innebar at akslingen (som er i senteret i aksen) i tillegg til å bære vekta av seilene også gikk gjennom et hull i



steinsenga(som den hvilte på) og også bar vekta til den øvre møllestenen. Stenene kunne være inntil halvannen meter i diameter og spindelen var tilrettelagt slik at når man skulle reparere eller bytte ut en slitt møllesten, kunne man huke den av uten å måtte ta ned seilene.

Seilene på de persiske møllene var om lag fem-seks meter høye, og diameteren varierte fra tre til noe over fire meter. Som vist i figur 3 dekker seilene bare det ytterste området på bladene. Seilene er ganske enkelt laget av flere tynne tømmerstokker som er blitt bundet sammen.

Rundt seilkonstruksjonen måtte man oftest bygge vegger som sendte vinden gjennom en åpning som ble trangere og trangere mot seilene for å få mest mulig kraft (og mølla krevde likevel meget sterk vind for å fungere). Se figur 2 til høyre. Noen møller ble bygget med en åpning på hver av de fire veggene, slik at den kunne brukes i vind fra flere retninger.



Figur 1 -Mann ved siden av en klassisk persisk vindmølle. Denen typen møller var operative fra rundt år 950 til på 1970-tallet.¹

¹ Figur 1 er fra boken *History of Windpower*, Cambridge University Press 1994

Med tanke på at hver mølle hadde ganske kort levetid, og man stadig måtte bygge nye, har en del sett på det som litt bemerkelsesverdig at det ikke oppstod endringer i byggetradisjonen på møller i regionen før på slutten av nittenhundretallet, nesten tusen år etter den første er konstatert at ble bygd.

Kinesere, tibetanere og europeere

Det påstås at det var korsfarererne som introduserte vindmøllene for den europeiske sivilisasjonen, men det hersker en del usikkerhet rundt akkurat dette. Likeså er man skeptiske til teorier om at Tibet hadde en stor vindmølletradisjon allerede på 400-tallet, dette bygget på nedskrevne tekster som nevner vindmøller samt de vinddrevne bønnehjulene man har tradisjon på å bruke i landet. De fleste regner med at det er snakk om feiloversettelser av tekstene. Det har i alle fall ikke blitt funnet noen spor som tyder på at det fantes vindmøller der så tidlig. Uansett tilegnet man seg kunnskapen fra Persia, og laget møller som betegnes som “strømlinjeformede” ettersom seilene var konstruert slik at man fanget vinden på den ene siden, mens den andre var glattere eller laget i kurve for å gi mindre motstand enn den andre.

I Kina tok man til seg lærdommen fra Persia etter at en minister (som var en meget god matematiker) under Chinghi Khan kom med en persis beskrivelse av møllene, som han riktingok som en feiltakelse presenterte som europeiske, og kommenterte hvor godt det skulle være hvis Kina kunne begynne å bruke den. Kina fikk sin egen vindmølletradisjon, og møllene ble hovedsakelig brukt for å heise opp vann. Disse tok også en annen form enn de persiske, og kalles av historikere for velutrustede.

Europeerne tok den beste lærdommen fra alle tre. De laget etterhvert svært driftsikre, glidende og velutrustede møller. I stor grad ble møller med horisontale akser brukt, og oftest med fire vinger/



Figur 3 - Tradisjonell europeisk vindmølleutforming.²

seil. Se figur 3. Disse viste seg å være mer effektive enn de vertikale. Man unngikk blant annet problemene med å måtte konstruere vingene slik at de både måtte kunne bevege seg med og mot vinden, slik de stående møllene måtte. Nøyaktige håndverkere laget tannhjul som styrte rotasjonen fra å gå horisontalt til vertikalt inne i mølla, og ned til møllehjulet. I utviklingen av bedre byggeteknologi hadde vindmøllene god drahjelp fra land med mange småelver, hvor samme teknologien ble brukt i møller som var drevet av fossekraft.

² Figur 3 er fra <http://www.biopix.dk/>

Amerikanske vindroser

Olje ble med tiden et innbringende stoff å pumpe opp av jorda. Mange vindmøller ble bygd til dette formålet. Et typisk syn fra den amerikanske landsbygda er rotorene som blir brukt til slike pumper, for ikke snakke om vannpumper. Disse har mange blader, og kalles vindrosetypen (figur 4).

Denne typen rotor ble også brukt til elektrisitetsproduksjon i begynnelsen. Det var først en danske, Poul la Cour, som innså at vindmøllene blir mer effektive når de har få rotorblader, og dermed beveget seg mye hurtigere. De kjente vindkraftverkene med tre blader i propellen regnes som dansk standard, men de fleste land har tatt etter. Danskene var tidlig ute med sin elektrisitetsproduksjon med vindmøller. Allerede før 1920 hadde mer enn hundre lokale elverk installert en vindmølle. Disse dekket den gang hele 3 % av det danske energiforbruket.



Figur 4 - Vindroser i California roterer saktere enn de danske propellene, med flere blader.³

Interessen for vindkraft internasjonalt har gått i berg og daler, med en stor økning i interessen de siste år, i stor grad som følge av alt fokuset på miljøvern og menneskeskapt klimaforandringer.

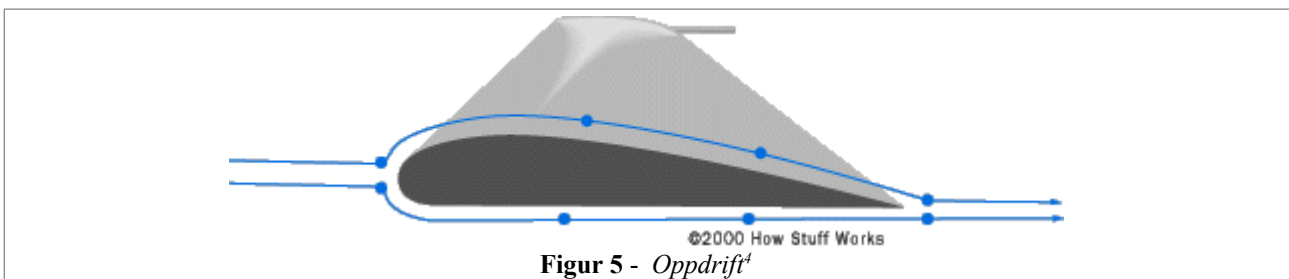
Teknisk

Energi er den mengde arbeid et fysisk system kan utføre. Arbeid defineres som kraft gange strekning og får da dimensjonen $nM = J$

Energi kan ikke oppstå, og kan ikke forsvinne. Altså kan den heller ikke forbrukes, den kan bare omgjøres til en annen form. Vindmøllas oppgave er å omdanne luftmolekylenes bevegelsesenergi om til rotasjonsenergi på rotorsystemet i vindmølla. (Rotoren er betegnelsen på alt i vindmølla som dreier som direkte følge av vinden).

I elektriske vindmøller skal denne rotasjonsenergien igjen omdannes til elektrisk energi. De neste delkapitlene ser vi på hvordan dette gjøres.

Rotorbladene



Se nå på figur 5. Luftmolekylene som passerer på oversiden av vingen må bevege seg hurtigere enn

³ Figur 4 er fra <http://www.qmarkstudio.com/Kunst.htm>

⁴ Figur 5 er hentet fra <http://travel.howstuffworks.com/airplane6.htm>

luftmolekylene på undersiden, ettersom de har en lengre vei å forflytte seg enn de under. Dette medfører at lufttrykket på undersiden av vingen vil være høyere enn det over, og vi får slik en kraft som trekker vingen oppover, vinkelrett på den virkende vindretningen.

Et annet fysisk faktum som må få følger for utarbeidelsen av propellbladene kan illustreres med reklameflaggene til enkelte busselskap. På taket av bussen har noen funnet ut at de kan feste en liten vimpel. Når bussen er i fart, flagerer flagget bakover i motsatt retning av den bussen kjører. Det er ikke vind, men luftmotstanden som oppstår når bussen kjører. Hvis bussen nå kjører med en konstant fart, og vi lar det komme inn en vind vinkelrett på bussens kjøreretning, vil flagget slå skrått ut. Vindkreftene som virker på det vil være vektorsummen av "kjørevinden" og vinden som vi sendte fra siden. Resultantkraften vil være et sted mellom bakover og vinkelrett på kjøreretningen i samme retning som vinden blåser. Hvis vi endrer vindhastigheten, vil vinkelutslaget til flagget variere.

Det samme skjer når rotorbladene beveger seg. Det er ikke nok at vinden virker på dem, de får også en vindkraft på dem takket være rotasjonsbevegelsen. Hvis man ser på to punkter på et rotorblad, ser man at et punkt helt ytterst på propellbladet, beveger seg mye raskere i forhold til luften rundt punktet, enn et punkt valgt lengre inne på propellbladet. Derfor vil resultantkraften være forskjellig fra de to stedene selv om den ytre vinden har samme hastighet mot begge punkter.

Når vi nå slår sammen disse to elementene over, skjønner vi at rotorbladet er formet med en form som ligner på det i figur 5, men at denne formen er vridd, slik at resultantkraften som virker på bladet og slik lager en oppdrift, vil være i samme retning for hele bladet.



Figur 6 - Rotorbladene er vridd og kan dreies for å øke ytelsen.⁵

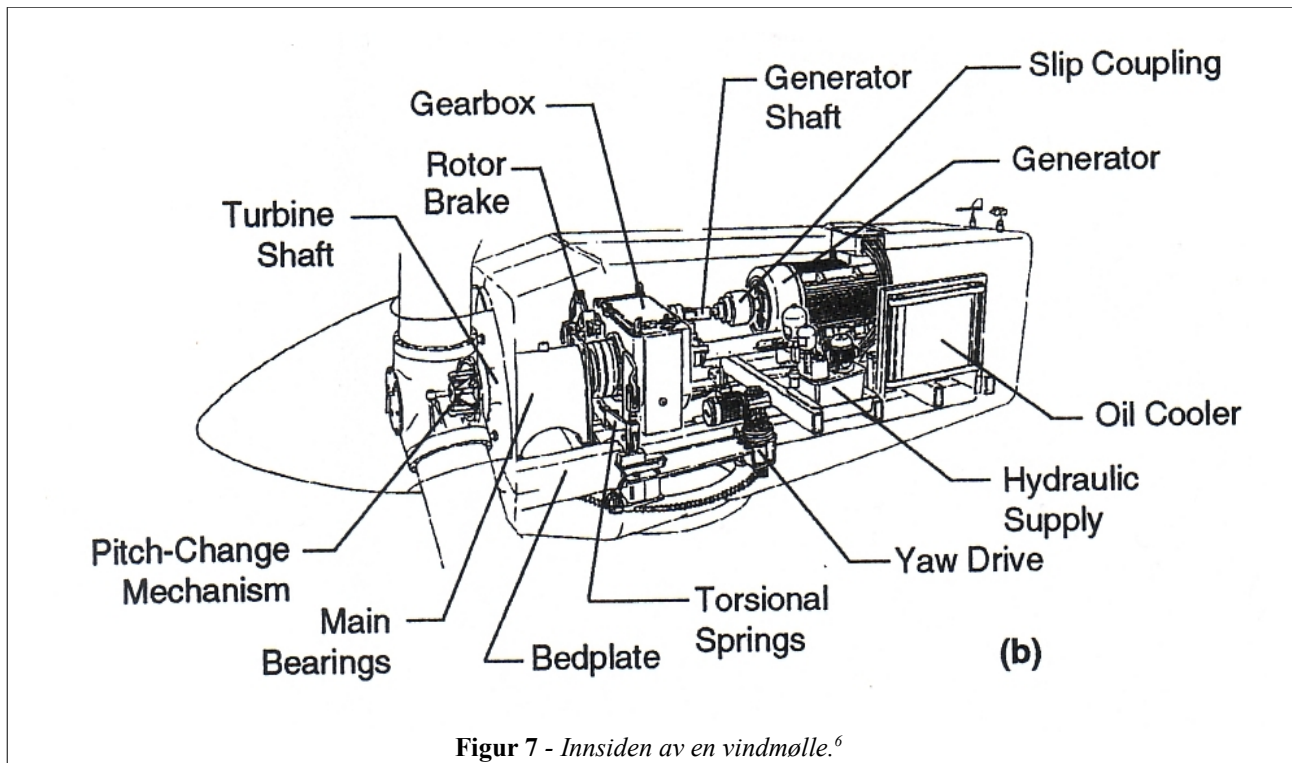
Rotorbladene var ofte konstruert i metall. I tillegg til å veie uønsket mye, lå problemer som metalltretthet og lurte. Det er store krefter som virker på bladene, særlig turbulens sliter på konstruksjonene. Man har et eksempel på en tysk vindmølle med propelldiameter på 100 m som ikke tålte belastningen, og det utviklet seg revner. Og det var etter bare knappe tre ukers drift. Det er dramatisk når så store deler bryter av.

Man konstruerer derfor de fleste moderne, store vindmøller (slik som på figur 6) av polyester som blir forsterket med glassfiber, karbonfiber eller aramid. Treforsterkninger forsøkes også brukt i

⁵ Figur 6 er fra www.windpower.org

voksende grad.

Rotorbladene har også en annen viktig finess. På moderne møller er de nemlig konstruert slik at de kan dreies rundt seg selv, slik at man alltid får mest mulig oppdrift. (Tenk på flymaskiners flapper på vingene som de bruker under take-off eller sving). På figur 7 er denne funksjonaliteten beskrevet som "Pitch-Change Mechanism". Dette er pirkete ingeniørkunst, da bladet bare må rotere inntil en grad om gangen. Det er som regel hydrøyliske systemer som styrer vinkelen. Disse er igjen styrt av en datamaskin som er koblet til vindmåleren (beskrevet senere).



Som det synes på figur % er det ikke bare rotorbladene som kan vrís på. Selve vindmølla er lagget slik at den kan roteres rundt horisontalt. (markert Yaw Drive på figur 7). Dette er da også for å stille mølla til den vinkelen i forhold til vinden som den yter best i, og er også styrt av en datamaskin.

Generatoren

Etter girkassa ender rotasjonen fra propellen opp i generatoren (se figur 8). Dette er den komponenten i vindmølla som produserer elektrisiteten. For å forstå hva som skjer inne i generatoren, må man kjenne til magnetisk induksjon. Dette vil vi vise ved et eksempel der vi genererer strøm i en spole ved hjelp av en stavmagnet.

⁶ Figur 7 er fra boken *Wind Turbine Technology* av David A Spera, 1994 Asme Press

På figur 9 og 10 nedenfor har vi en elektromagnet (lederspole) og en stavmagnet (permanent magnet). Elektromagneten (spolen som er plassert øverst i de to figurene) er fastmontert, mens den permanente magneten kan rotere. De grønne linjene markerer magnetfeltet fra stavmagneten fra magnetens nordpol til sydpol.

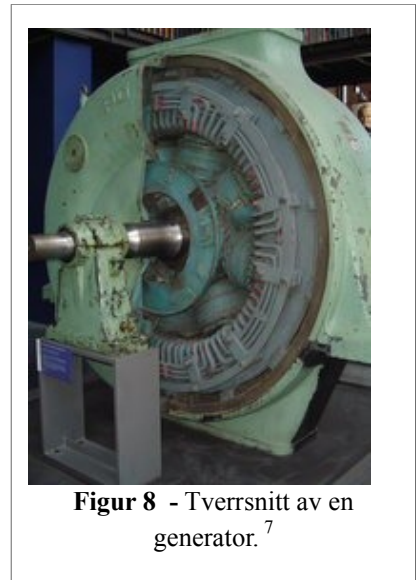
Magnetisk fluks Φ er målet på magnetfeltet \vec{B} som går gjennom en tenkt flate. For enkelhets skyld antar vi at magnetfeltet har en konstant absoluttverdi $|\vec{B}|=B_0$. Når vi har en ledersløyfe som det går et tidsavhengig magnetfelt gjennom (fluks), vil det oppstå en elektrisk strøm i kretsen; det blir *indusert* en elektrisk spenning. Matematisk er den magnetiske fluksen gitt ved:

$$\Phi = \int \vec{B} \cdot d\vec{A}$$

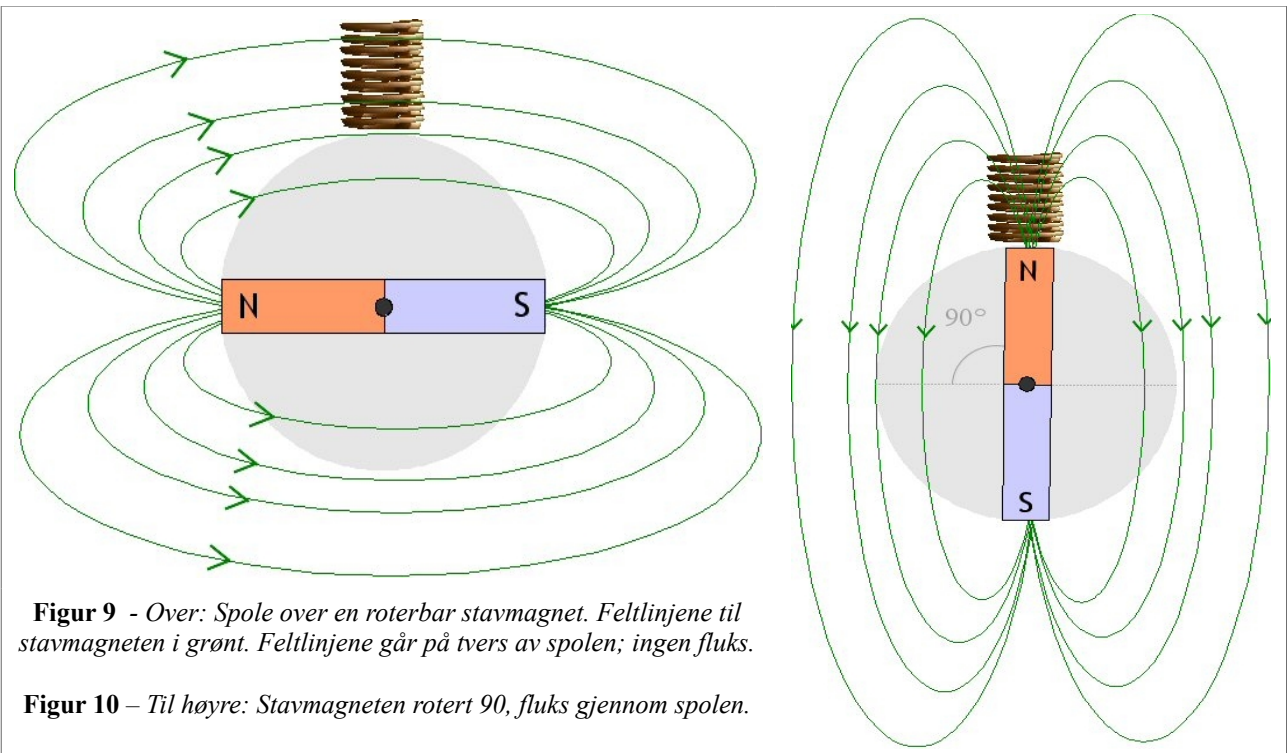
Spolen omslutter gaussflaten som vi skal finne fluksen gjennom. Vi må altså finne arealet av flaten for å kunne integrere magnetfeltet over det. Spolen består av et antall n viklinger/sløyfer med radius r som er forbundet. Altså er det flere flater, hver med areal $\pi \cdot r^2$, oppå hverandre, og vi får arealet $A = n \cdot \pi \cdot r^2$

Vi definerer θ som vinkelen magnetfeltet \vec{B} danner med figurenes horisontalplan. Fluksen gjennom spolen i figur 1 og 2 er da gitt ved

$$\Phi = n \cdot \pi \cdot r^2 \cdot B_0 \cdot \sin \theta$$



Figur 8 - Tversnitt av en generator.⁷



Figur 9 - Over: Spole over en roterbar stavmagnet. Feltlinjene til stavmagneten i grønt. Feltlinjene går på tvers av spolen; ingen fluks.

Figur 10 – Til høyre: Stavmagneten rotert 90, fluks gjennom spolen.

⁷ Figur 8 er fra http://de.wikipedia.org/wiki/Elektrischer_Generator

Nå kan vi se på de to situasjonene i figur 9 og figur 10. På figur 9 går magnetfeltet tilnærmet parallellt med flatene i spolen. I denne situasjonen har vi ingen fluks:

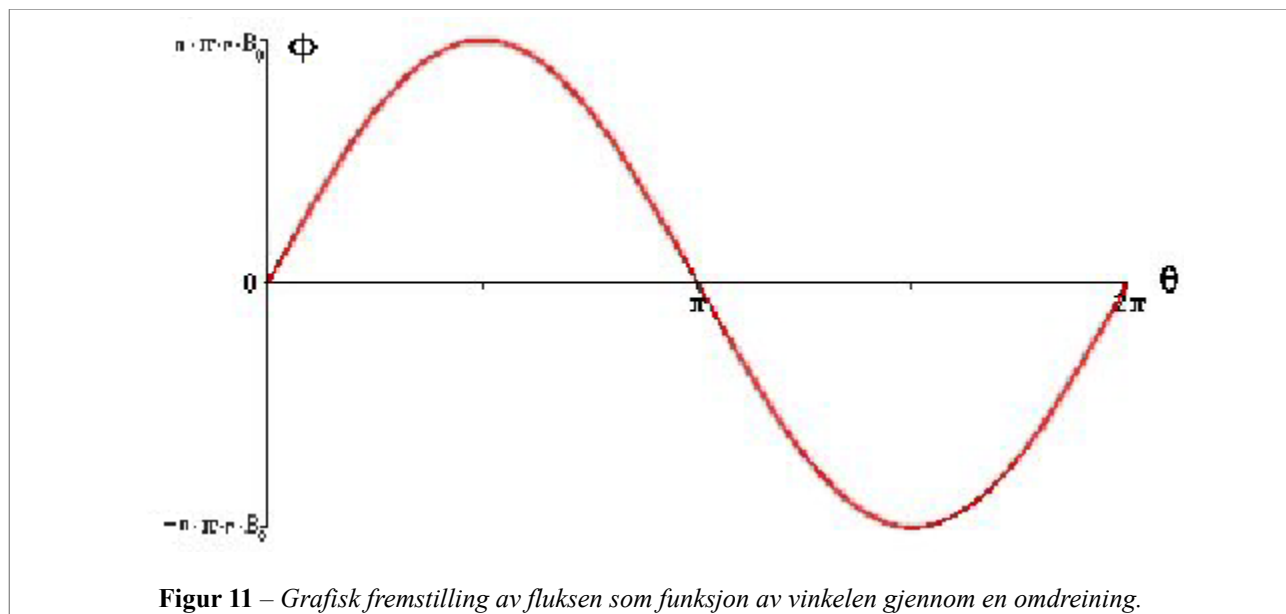
$$\Phi_0 = 0$$

Se nå på figur 2. Vi har nå dreiet stavmagneten 90° med klokka. Vi ser at feltlinjene nå går tilnærmet vinkelrett gjennom flatene våre, og vi har dermed den største fluksen.

Når stavmagneten roterer videre, vil fluksen gjennom spolen synke. Etter 180 graders rotasjon, vil fluksen være tilbake på null. Rotasjonen fortsetter, og vi får motsatt fortegn på verdier for magnetfeltet som går gjennom flaten ettersom vi nå har snudd om stavmagneten (i forhold til på de to figurene) slik at magnetfeltet peker den andre veien. Vinkelen varierer med tiden slik at $\theta(t) = \omega \cdot t$, der ω er vinkelhastigheten til magneten og t er tiden. (vi behandler den her som om den var konstant, men i en vindmølle vil denne også variere med tiden). Vi ser altså at

$$\Phi(t) = n \cdot 2 \cdot r^2 \cdot B_0 \cdot \sin(\omega \cdot t)$$

der B_0 er den konstante størrelsen på magnetfeltet fra den permanente magneten.



Figur 11 – Grafisk fremstilling av fluksen som funksjon av vinkelen gjennom en omdreining.

Som nevnt induseres det elektrisk spenning i en sløyfe av strømførende materiale, når det er endringer i den magnetiske fluksen gjennom sløyfa. Verdien på induisert spenning er gitt ved den negative deriverte av fluksen:

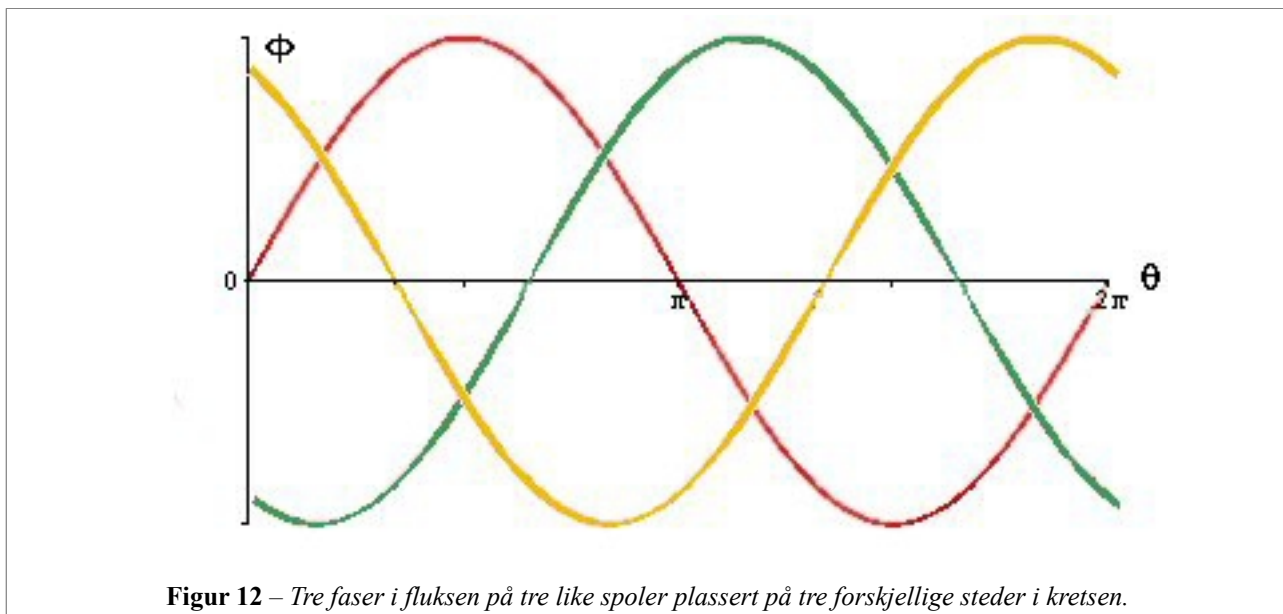
$$\epsilon = -\dot{\Phi}$$

$$\epsilon = -n \cdot \pi \cdot r^2 \cdot B_0 \cdot \omega \cdot \cos(\omega \cdot t)$$

Vi ser da at vi får induisert en sinusformet vekselspenning i spolen, se figur 11.

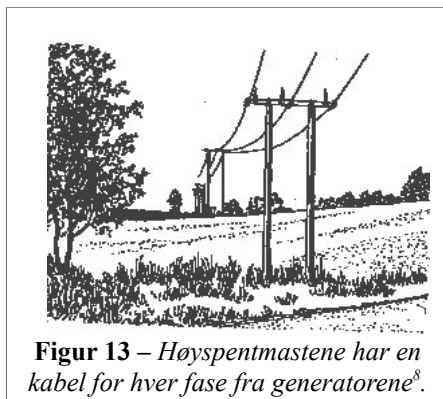
Generatorer i elektrisitetsverk er oftest bygd opp med en roterende magnet i sentrum, med spoler tilknyttet elektrisitetsnettet plassert rundt denne. La oss tenke oss tre like spoler plassert rundt en

roterende stavmagnet. Disse tre spolene vil hver for seg få indusert spenning slik som i eksemplet over. Men de er plassert forskjellig i forhold til magneten. De vil dermed gi samme spenning, men til forskjellig tid. De ligger i hver sin fase, $1/3$ av perioden bak den foregående magneten. Dette blir illustrert i figur 12. (Vi forutsetter at de er symmetrisk plassert rundt den roterende magneten.)



Trefaset synkrongenerator

Det er utbredt med trefasegeneratorer i elektrisitetsverk. Disse generatorene sender ut vekselspenning i tre forskjellige faser. At bruken er omfattende i Norge konstaterer vi lett når vi titter på høyspentmastene. (Figur 13). Disse har tre linjer, en for hver fase.

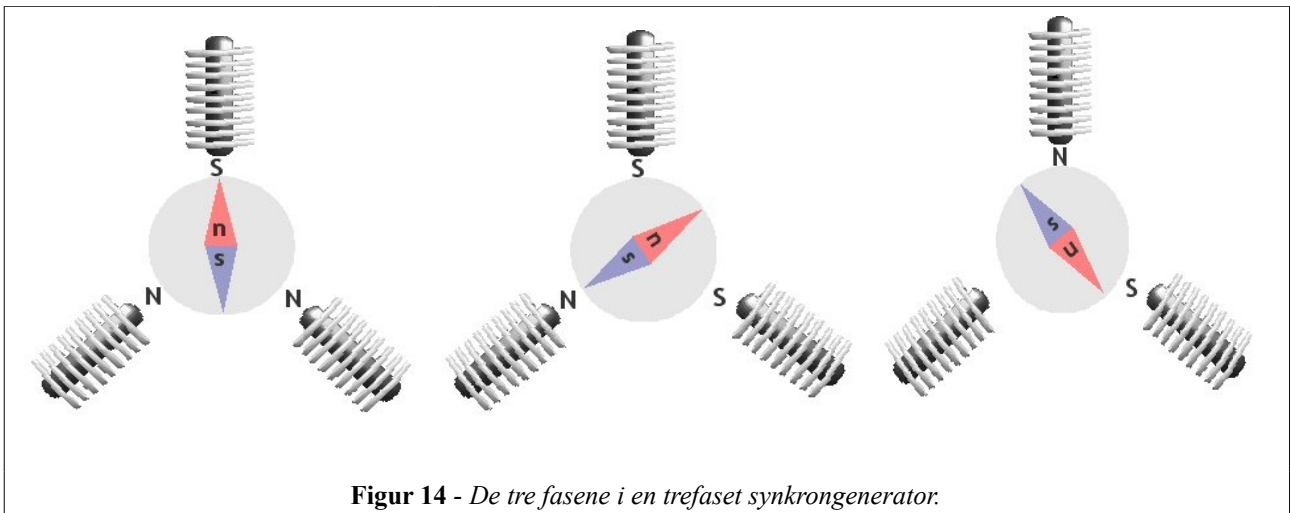


En trefaset generator fungerer ved at man har tre elektromagneter (spoler) plassert i en sirkel slik vist på figur 14. Konstruksjonen med de tre fastmonterte spolene kalles generatorens *stator*; ettersom den holder seg i ro. Magnetten i midten roterer. Den kalles *rotor*, og er (oftest) en permanent magnet.

Når rotoren dreier, svinger magnetfeltet til de tre elektromagnetene i statoren i fase med rotasjonen slik vi beskrev det i forrige avsnitt. Videre vil den induserte spenningen svinge i fase med hver enkelt av dem.

Fluksen for hver av de fastmonterte elektromagnetene når sin maksimale og minimale verdi når rotorens poler peker parallelt med dem. For de to andre magnetene i statoren vil fluksen være halv verdi av maksimalt, og hver for seg være i ferd med å øke og minke (i hver sin retning). Når magneten i rotoren står vinkelrett på en av dem, vil den spolen være magnetisk nøytral, og dermed ikke ha noen indusert spenning.

⁸ Figur 13 er fra www.be.no



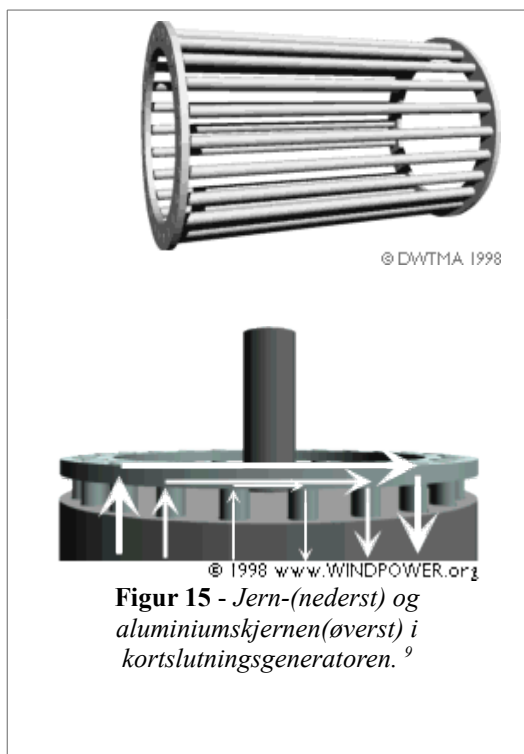
Vekselstrømmen i elektrisitetsnett i Europa har en frekvens på 50 Hz. Med dette som bakgrunn, tar vi utgangspunkt i et tenkt tilfelle der rotoren jobber med konstant 50 runder i sekundet. På figuren har vi tre magneter i statoren, og én i rotoren. Hvis vi dobler antall poler i både statoren og rotoren, vil bare halve rotasjonen være nødvendig for å generere samme frekvens på den induerte spenningen. For å få en frekvens på 50 Hz vil det med seks elektromagneter i statoren og to i rotoren bare nødvendig med 25 runder pr. sekund. Med 12 elektromagneter i statoren og 8 poler i rotoren behøves bare 12,5 roteringer pr. sekund. Generatorer med mange magneter i rotor og stator kalles lavhastighetsgeneratorer.

Høyhastighetsgeneratorene er mindre i volum og mye billigere å konstruere enn lavhastighetsgeneratorene. Dette er hovedargumentene som legges til grunn når man vurderer hvilken type generator som skal benyttes i kraftverket. I de fleste vindturbiner brukes fire eller seks poler (to eller tre magneter) i rotoren.

Rotoren er koblet til propellens rotasjon via en girkasse. Den permanente magneten i rotoren må være meget kraftig når man skal produsere mye elektrisitet. Jo mer kraft man tilfører rotoren, dess mer strøm produseres. Omdreiningshastigheten må holdes konstant, slik at man får en jevn frekvens på vekselstrømmen. Av den grunn er man oftest tvunget til å bruke indirekte tilslutning til elektrisitetsnettet når man velger å bruke synkrongeneratorer i vindmøller.

En annen faktor med synkrongeneratorene er at permanente magneter blir avmagnetiserte med tiden hvis de befinner seg i veldig sterke magnetfelt. Det er sterke magnetfelt i generatorer. Metallene (f.eks. neodymium) som kraftige magneter lages av forekommer også ganske sjeldent i naturen, og prisene på dem er dermed i tillegg veldig høye. I vindmøller med synkrongeneratorer er det av disse grunnene vanlig at man i steden har en elektromagnet i rotoren, og at man forsyner denne med likestrøm via et børstesystem.

Trefaset asynkron kortslutningsgenerator



I tillegg til småskala vannenergiverk, er det stort sett bare innen utnyttelse av vindenergi at man bruker kortslutningsgeneratorene. Det som gjør at den asynkrone generatoren skiller seg fra den synkrone (den som er beskrevet i de forrige avsnittene) er rotoren.

Rotoren i en asynkron kortslutningsgenerator består av to metallsirklene som bindes sammen av mange stenger av strømledende kobber eller aluminium. Se figur 15. Rotoren består videre av en jernkjerne, og isolerende lag med hull til de strømførende stengene. Den er plassert midt i statoren, på samme måte som den permanente magneten i den synkrone generatoren. En vesentlig forskjell fra den synkrone generatoren er også at elektromagnetene i statoren må være tilknyttet en spenningskilde, for at det skal kunne induseres spenning i rotoren slik beskrevet under.

Denne rotoren dreies. (Det er ikke magnetfeltet som driver rotoren rundt, det er vinden). Når rotasjonen foregår svært raskt rundt, dvs. mer enn 50 omdreininger i sekundet, vil magnetfeltet i rotoren rotere raskere enn magnetfeltet i statoren roterer. Da blir det induert spenning i hver av de tversgående aluminiumsstengene som forbinder de to aluminiumssirklene, dette på grunn av magnetfeltene til elektromagnetene i statoren. Siden de to aluminiumssirklene kortslutter stengene, vil det øyeblikkelig gå en strøm i hver av stengene, og rotoren danner slik sine egne magnetiske felt, som roterer med rotoren. Jo kraftigere vi dreier rotoren, jo mer energi blir overført via induksjon i statorspolen og til elektrisitet som ledes ut til elnettet.

Asynkrone generaterer kan man koble rett opp mot elektrisitetsnettet. Forskjellen mellom rotasjonshastigheten ved maksimal elektrisitetsproduksjon og rotasjonshastigheten ved tomgang er et sted rundt 1 %. Det er på denne prosentforskjellen at asynkrongeneratoren yter. Hvis vi tenker oss at den er forbundet til en strøm med frekvens på 50 Hz, vil generatoren gå på tomgang ved 25 omdreininger i sekundet (vi tenker oss at den er firepolet). $25 \cdot 60 = 1500$ omdreininger i minuttet. Den maksimale ytelsen vil da oppnås ved 1515 omdreininger i minuttet.

Dette glippet fører til at den er mer fleksibel for variasjoner enn den synkrone generatoren, og girkassen blir mindre belastet. En annen egenskap til generatoren er at rotoren selv tilpasser seg antall poler i statoren. Dermed kan samme rotoren brukes til mange forskjellige antall poler i stator. Den største fordel er uansett at den kan kobles direkte til elektrisitetsnettet, da den holder stabil frekvens.

Indirekte nettilkobling

Hvis man bruker en synkrongenerator i en vindmølle kan denne i motsetning til den asynkrone ikke

⁹ Figurene hentet fra <http://www.windpower.org/da/tour/wtrb/async.htm>

tilknyttes direkte til strømnettet. Frekvensen kan ikke variere i takt med vindhastighetens variasjon.

Første steg i å få en jevn frekvens på spenningen fra en synkrongenerator er å likerette strømmen. Dette gjøres ved hjelp av effekttransistorer. En effekttransistor er en stor halvlederkontakt som kan slutte og bryter strømmen uten bruk av mekanikk. Deretter kan strømmen omdannes med en inverter til vekselstrøm med en fast frekvens som matcher elektrisitetsnettet.

Det er noen små ulemper ved bruk av denne type nett-tilkobling.

Indirekte nett-tilkobling har selvsagt noen fordeler også. Ved vindstøt (altså kraftige vinder som ikke varer) vil den ekstra energien i vinden lagres som bevegelsesenergi, med hjelp av kompliserte styringssystemer. Dette vil ikke skje med en asynkrongenerator. En annen fordel er at man kan regulere faseforkyvningen til strømmen, slik at den er i samsvar med den på elnettet.

Transformator

Når man snakker om store vindmøller, det vil si vindmøller som produserer 150 kW genererer vindmøllen normalt en 690 V trefaset vekselstrøm. Strømmen føres så i kabler ned tårnet til en transformator som oftest plasseres helt i bunnen, eller i et eget hus ved siden av mølla.

I transformatoren økes gjerne spenningen til et sted mellom 10 000 V og 30 000 V. (avhenger av standarden for det lokale elnettet)

Kjøling

Man har et visst energitap i vindmøllens rotoranlegg som følge av friksjon i kulelagrene, som gjør at deler av rotasjonsenergien som ideelt sett skulle blitt brukt til elektrisitetsproduksjon blir omdannet til termisk energi. I vindmøller er derfor kjølesystemer nødvendig. Det vanligste er luftkjøling. Enkelte vindmøller konstrueres med vannkjøling. Dette gir enkelte tekniske fordeler for generatoren, men til gjengjeld blir det vanskeligere å få varmen ut av tårnet.

Anemometer

For å kunne utnytte vindenergien mest mulig effektivt er det viktig at møllen er stilt slik at vinden treffer mest mulig vinkelrett på propellen. Moderne vindmøller snur seg automatisk ved hjelp av



Figur 16 – Vindhastigheten måles med et anemometer.¹⁰

datautstyr som leser av måleapparater festet på møllen. Den mest vanlige måten å måle vindens hastighet og retning på er fortsatt å bruke et anemometer. Dette er et apparat som har tre kopper plassert slik at de roterer når det blåser, se figur 16. Rotasjonen leses av elektronisk i m/s. Værhanen, som i de fleste tilfeller er plassert sammen med anemometeret, er en flat innretning som stiller seg parallellt med vindretningen, og slik måler av denne. Ved hjelp av målene fra anemometeret avgjør vindmøllerenes computer om det er fornuftig å dreie rotoren mot vindhastigheten og sette rota

¹⁰ Figur klippet fra pakistansk frimerkeutgivelse 23. mars 1961 med tittel "Meteorologisk dag".

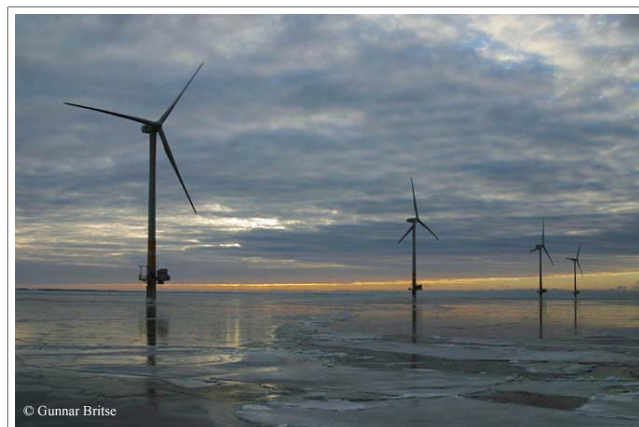
I blant annet arktiske forhold kan det bli problematisk å ha et vanlig anemometer i drift, da det kan fryse fast. En måte å unngå dette på er å koble til utstyr som holder den på en tilstrekkelig høy temperatur til ikke å fryse. En annen mulighet er å benytte et anoetertype som baserer seg på ultralyd eller laser-teknologi. Disse registrerer lysets eller lydets faseforskyvelse ved refleksjon i luftmolekylene. Det finnes også en variant som kalles varmetrådsanometer. Denne opererer ved å måle temperaturforskjellen mellom to varmetråder som er plassert henholdsvis på vindsiden og den siden av mølla som er i le. Grunnen til at det fungerer er at jo kraftigere det blåser, jo mer avkjøles den varmetråden som er mest utsatt for vind.

Lokale vindforhold

Vinden varierer kraftig. Mellom Afrikas østkyst og havnene i India la man til sjøs på bestemte tider av året når vinden blåste riktig retning mellom havnene. I britiske møller ble det snakket om møllere som i ukesvis hadde gått og drevet dank, da de plutselig i høstssesongen fikk masse vind, slik at de måtte jobbe opp mot tre døgn i strekk uten de store pausene.

Vinden om natten er oftest mye roligere enn den om dagen. Den er roligere både i styrke, hvor ofte den skifter retning og i hvor mye turbulens det er i den. Hovedårsaken til dette er at temperaturforskjellen mellom havet og lufta er større om dagen enn natta. At det er mer vind om dagen er ikke et stort hinder for kraftselskapene, ettersom

Som nevnt i den historiske biten så går vinden knirkefritt gjennom landskapet når det ikke er stort å finne av friksjon. Det beste landskapet er da å finne langs havoverflaten. Vann regnes som det best egnede landskapet for vindmøller (fig 17), selv om det er en del friksjon forbundet med dannelsen av bølger. Men så snart bølgene er dannet, beveger de seg i samme retning som vinden, og vil slik være skape mindre ujevnheter for vinden.



Figur 17 - Optimal effekt der friksjonen er lav. Bilde fra et britisk offshore vindmølleanlegg.¹¹

Lanskaper med mye trær eller bebyggelse har mye ujevnheter. Slike områder som har mye turbulens er de minst attraktive når man skal bygge vindkraftverk. For det første vil man kunne utnytte mest mulig av energien i vinden når man er lengst vekk fra ujevnheter. For det andre vil turbulens fra f.eks. klipper føre til større belastning på kraftstasjonen, slik at man kan regne med å ende opp med kortere levetid

Objekt som står i veien for vinden demper ofte vinden i et område med fem ganger større objektets dimensjon, og det i hver retning. Når man kommer opp i høyder rundt en kilometer over bakken, finner man vind som er helt upåvirket av forhold på bakkenivå. Jo høyere opp fra bakken vindmøllas propell er montert, jo mer effektivt vil de altså virke. Det er viktig å huske på at i de fleste tilfeller snakker man om at det er høyden over bakken som gjelder, og ikke over laveste punkt i nærheten. Det vil i mange tilfeller være naturlig å tro at å plassere en mølle på toppen av en skrent

¹¹ Figur 17 er fra <http://www.fmft.net/archives/000099.html>

vil føre til økt effekt, ettersom vinden må over kanten, men det motsatte vil i stor grad kunne være tilfellet, da skrenter og lignende ofte fører til turbulens i lang avstand før vinden når skrenten. Dessuten fører turbulens ved bakker og skrenter til mer slitasje enn en mer stabil vind på en åpen slette. Enkelte mykt formede åser kan likevel være å foretrekke til plassering av vindmøller.

Vindens effekt

Det er hastigheten til vinden som har mest å si for hvor mye energi vi kan hente ut fra vinden med en vindturbin.

Vi tenker oss nå et eksempel med et massivt objekt, f.eks. en bil, som er i fart. Fra Newtons 2. lov har vi at dersom objektets hastighet fordobles, så kreves det $2^2 = 4$ ganger mer energi å stanse objektet. Vi kan tenke oss vinden som en lang rekke objekter som beveger seg (det letteste er da å se vinden for seg som mange skiver som beveger seg etter hverandre).¹²

Når vindhastigheten dobles vil hver av skivene behøve fire ganger mer energi for å stanses. I tillegg vil det gå dobbelt så mange slike skiver gjennom propellen. Det som skjer når vi henter energi fra vind med en propell er jo at propellen bremser vinden ved å overføre dens bevegelsesenergi til seg selv. Vi ser dermed at hvis vindens hastighet dobles, vil energien som kan hentes ut med en propell bli $2^2 \cdot 2 = 8$ ganger høyere enn sin opprinnelige verdi.

Energien i et legeme er proporsjonal med legemets masse. Bevegelsesenergien i vinden avhenger altså av luftens massetetthet. ($1,225 \text{ kg/m}^3$ ved vanlig atmosfærisk tryk og 15 grader celsius.)

Den totale effekten av vind som passerer gjennom tversnittet av en sirkelformet flate (som en propell) er

$$P = 1/2 \cdot \rho \cdot v^3 \cdot A = 1/2 \cdot \rho \cdot v^3 \cdot \pi \cdot r^2$$

P er den totale effekten gjennom flaten [W]

ρ er massetettheten til lufta [kg/m^3]

A er arealet som rotorbladene sveiper over på en omdreining [m^2]

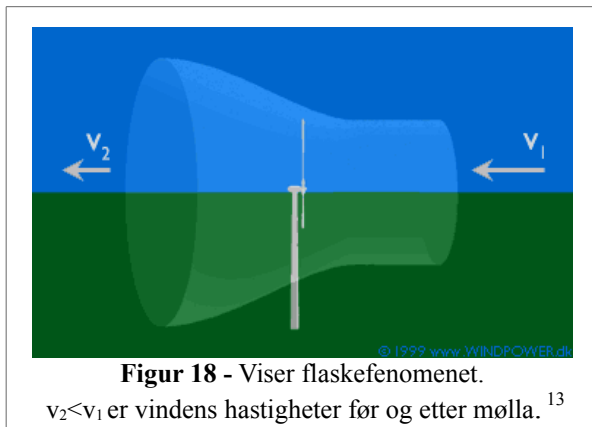
r er rotorbladenes lengde [m]

Energien som kan hentes ut avhenger også av A, og dermed med faktor r^2 . Dobling av rotorbladets lengde vil altså gi en firedobling av energien.

Det er derimot mange måter å optimalisere og spesialisere møllene til lokale forhold; større rotorblader betyr ikke nødvendigvis økning i den elektriske kraftproduksjonen. F.eks. vil man i områder med lav gjennomsnittlig vindhastighet kunne maksimere strømproduksjonen ved å bruke en mindre generator til en gitt lengde på propellbladene, eller lengre blader til en annen gitt generator. Dette fordi en liten generator vil kunne medføre økt *driftstid* (dvs. timer generatoren

¹² www.windpower.org/da/tour/

er i drift pr. år) mens en større vil kunne yte mer når det blåser som verst.



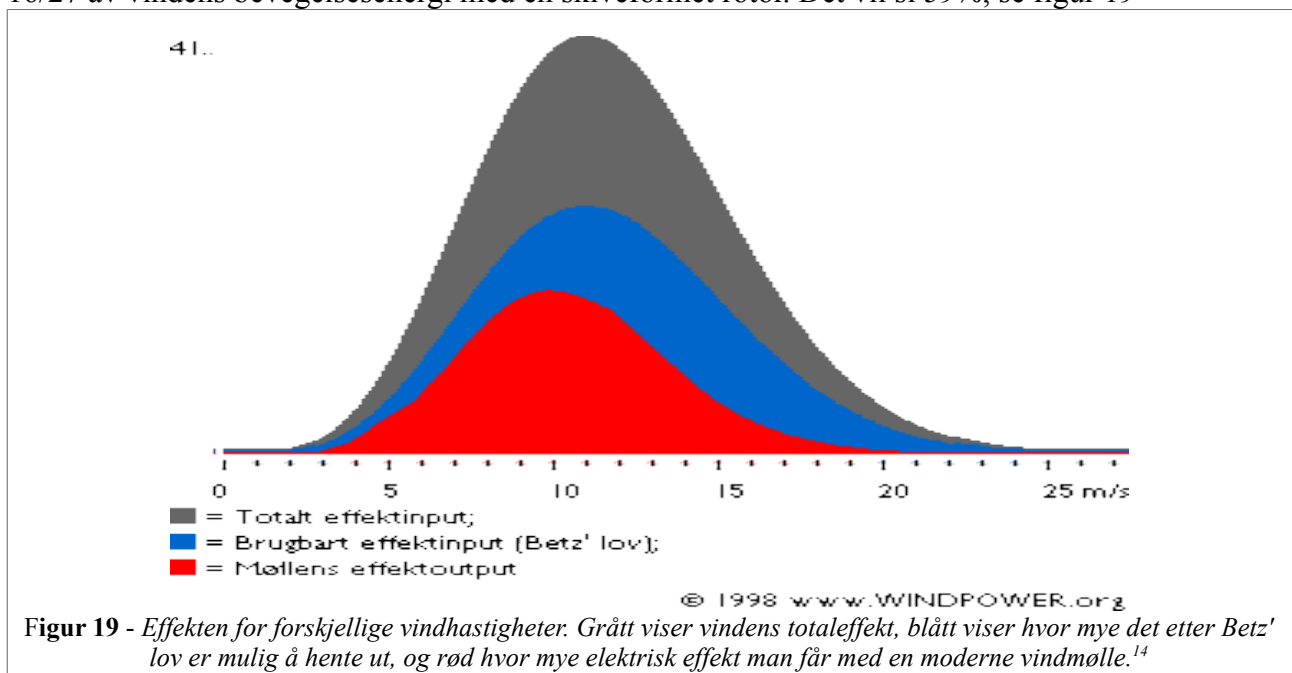
Ettersom vindmøllen henter sin energi ved å bremse vinden, må vindhastigheten være lavere på baksiden av møllen. Videre må mengden luft som går gjennom propellen være den samme både før og etter mølla. Hvis vi tenker på vinden som passerer gjennom arealet som propellen feier over som et vindrør, ser vi dermed at luften som har passert gjennom møllen må utfylle et større areal enn før. Dette er kjent som flaskefenomenet. En typisk avbøyning er illustrert i figur 18.

Betz' lov

En av de mest grunnleggende lover i aerodynamikken er Betz' lov. Den bygger på flaskehalsen som ble beskrevet i forrige avsnitt. Vi tenker oss den ideelle vindmølla. Møllas oppgave er å omdanne vindens kinetiske energi (bevegelsesenergi) om til rotasjonsenergi. I det ideelle tilfellet omdannes all energien vinden bærer på om til rotasjonsenergi. Hva skjer da?

Vel, energien omdannes til rotasjon ved at rotoren bremser vinden. Vinden mister sin bevegelsesenergi. Altså kan den ikke ha noen hastighet. Den vil stoppe opp. Men hvis vinden stopper opp, vil kan den ikke bevege seg bort fra mølla, og mølla vil igjen ikke kunne utnytte bevegelsesenergien fra vinden som "passerer" gjennom, ettersom den vinden som passerer gjennom står stille.

Loven som ble formulert av Albert Betz i 1919 sier at det ikke lar seg gjøre å omdanne mer enn 16/27 av vindens bevegelsesenergi med en skiveformet rotor. Det vil si 59%, se figur 19



¹³ Figur 18 er fra www.windpower.org

Vindmøllekraftverk opererer stort sett med vinder fra 3 m/s til 25 m/s. De stoppes automatisk ved hastigheter over og under dette, for å unngå å volde seg selv og omgivelsene skader. Som man kan se av grafen i figur 19 så er det ingen store ulemper ved dette, ettersom man ikke kan hente ut den store effekten ved disse vindhastighetene.

Vindmølleparker

Også selve vindmølla lager turbulens. Når man konstruerer store vindmølleparker der det dermed ikke vanlig å plassere møller i kjølvannet alt for nært opptil en annen mølle. Det vanlige er å ha om lag tre ganger rotordiameteren til vindmøllene som avstand vindmølletårnene imellom. I den dominerende vindretningen har man gjerne inntil ni ganger rotordiameter som avstand mellom møllene. Som følge av avbøyningen vi ser illustrert i figur 1 over, er det også vanlig å ikke plassere en mølle rett bak en annen. Selv om man tar disse forhåndsreglene, må man regne med at andre rad med vindmøller i en vindmøllepark vil være 5% mindre effektiv enn den ytterste.

Norske tilstander

Potensialet for vindkraft i Norge.

Norge har stadig økende kraftbehov. Det er ønskelig å dekke dette uten å øke utslippene av CO₂.

De økonomiske kostnadene for utbygging av vindkraft har sunket kraftig de siste 20 årene. I alle ledd fra design til produksjon har man hatt forbedringer; mer effektiv fabrikasjon av vindturbiner, økt effektivitet i vindturbinens ytelse, redusert materialforbruk, forbedret design, økt vindhastighet, høyere tårn og større rotor, bedre beregningsverktøy for plassering, lavere kostnader for installasjon og nettilkobling.

Det forventes at kostnadene skal synke ytterligere. Danske eksperter regner med at man i 2020 har et kostnadsnivå som er 50 % av nivået i 1995.

Tall fra 1992/94 antyder et potensial for vindkraft i Norge på 32-73 TWh/år, og en kraftpris på 35-65 øre/kWh, mens tall fra 2005 er oppjustert og antyder et potensial på 250 TWh/år og 27-40 øre/kWh.

Nesten 70 % av dette potensialet ligger i Finnmark. I landet for øvrig er det Troms og Sør-Trøndelag som markerer seg som aktuelle. Noen av årsakene til at Finnmark utgjør en så stor andel er gode vindforhold kombinert med store åpne landområder. Derimot er det lavt eget forbruk, så en utstrakt utbygging her er betinget av at andre landsdeler importerer kraft fra Finnmark. Men det eksisterende kraftnettet i Finnmark har ikke kapasitet for den overføringsmengde vi her snakker om. Det konkluderes med at kun en brøkdel av potensialet er realiserbart å bygge ut. Statnett på sin side

14 *Figur 19 er fra www.windpower.org/da/tour/wres/powdensi.htm*

vurderer fortløpende hvorvidt det er samfunnsøkonomisk å anbefale utbygginger av nettet. Så langt holder de igjen hva angår økt kapasitet fra Finnmark og sørover, og anbefaler heller utbygging av vind- og gasskraft i Midt-Norge. Kraftnettet i Midt-Norge har kapasitet til eksport videre sørover, samt at det er kraftunderskudd i landsdelen, noe som gir grunnlag for økt utbygging her.

Eksisterende vindkraft, og planer for vindkraft i Norge.

Tall fra NVE, september 2005 gir at det i dag er eksisterende 274 MW installert effekt. Videre er det gitt konsesjon for ytterligere 845 MW installert effekt fordelt på totalt 22 vindparker. (Se tabell 1 og 2). Det er levert konsesjonssøknad på 1766 MW fordelt på 20 parker, og forhåndsmelding på 47 parker. (Av de forhåndsmeldte inngår Havsul-prosjektet som spesifikt nevnes senere.)

Nr	NAVN	TILTAKSHAVER	MW
1	FJELDSKÅR	AGDER ENERGI PRODUKSJON AS	4
2	HARØY	SANDØY VINDKRAFT A/S	4
3	HAVØYGAVLEN	ARCTIC WIND AS	40
4	HITRA (ELDSFJELLET)	STATKRAFT ENERGI AS	55
5	HUNDHAMMERFJELLET-1	NORD-TRØNDELAG ELEKTRISITETSVERK	2
6	HUNDHAMMERFJELLET-2	NORD-TRØNDELAG ELEKTRISITETSVERK	3
7	MEHUKEN I	KVALHEIM KRAFT AS	4
8	SANDHAUGEN	NORSK MILJØKRAFT TROMSØ AS	9
9	SMØLA	STATKRAFT ENERGI AS	150
10	UTSIRA VIND- OG HYDROGENANLEGG	HYDRO ENERGI	1
11	VIKNA HUSFJELLET	NORD-TRØNDELAG ELEKTRISITETSVERK	2
			274

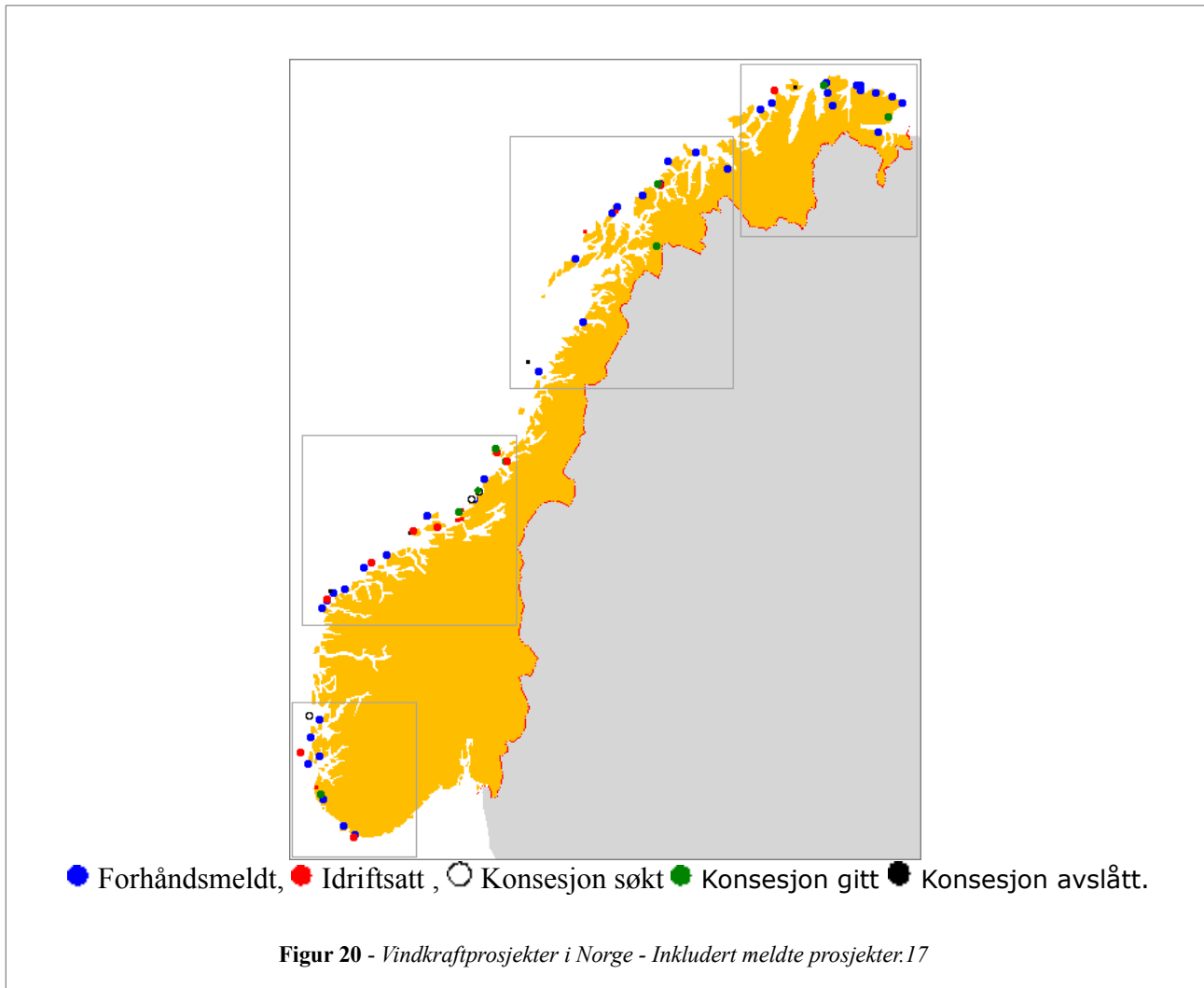
Tabell 1 - Idriftsatte anlegg, kilde NVE. ¹⁵

Nr	NAVN	TILTAKSHAVER	MW
12	BESSAKERFJELLET	TRØNDERENERGI KRAFT AS	51
13	HARBAKSFJELLET	NORSK HYDRO ASA	90
14	HUNDHAMMERFJELLET-3	NORD-TRØNDELAG ELEKTRISITETSVERK	45
15	HØG-JÆREN	JÆREN ENERGI AS	80
16	KJØLLEFJORD PÅ GARTEFJELLET	STATKRAFT DEVELOPMENT AS	40
17	KVITFJELL	NORSK MILJØKRAFT TROMSØ AS	200
18	NYGÅRDSFJELLET	NORDKRAFT VIND A/S	7
19	SKALLHALSEN	STATKRAFT ENERGI AS	65
20	VALSNESET	TRØNDERENERGI KRAFT AS	12
21	VALSNESET TESTSTASJON	VIVA AS	6
22	YTRE VIKNA	NORD-TRØNDELAG ELEKTRISITETSVERK FKF	249
			845

Tabell 2 - Konsesjon gitt, men ikke satt i drift. ¹⁶

15 http://www.nve.no/modules/module_111/netbasNVE.asp?script=8

16 http://www.nve.no/modules/module_111/netbasNVE.asp?script=8



På Norgeskartet (figur20) ser vi at det er en rekke forhånds meldte prosjekter i Finnmark. Dette skyldes de allerede nevnte forhold, samt at man har 25 % investeringsstøtte som medfører en stor interesse for å få sendt inn forhånds meldinger, og utarbeide konsesjonssøknader. Mange av prosjektene her vil nok bli skrinlagt.

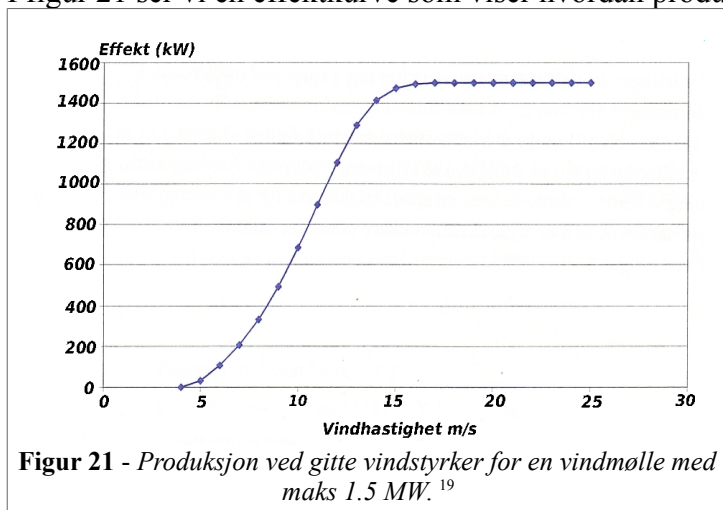
Noen flere nøkkeltall er tatt med i tabell 3

Anlegg	Status	Installert effekt [MW]	Andre nøkkeltall.	Årlig produksjon [Gwh]	Husstander som får dekt forbruk.
Smøla 1	20 møller i drift	40		120	6000
Smøla 2	48 møller i drift	110		330	16500
Hitra	24 møller i drift	55	450 mill. kr. Investert.	150	7500
Hammerfest	Konsesjonssøknad sendt.	110	ca. 1 mrd. kr.	320	16000
Kjøllefjord Finnmark	Har konsesjon, ikke utbygd.	40	ca. 320 mill. kr.	155	7750
Skallhalsen Vadsø	Konsesjon delvis innvilget	40-65	320-500 mill. kr.	ca. 200	ca. 10000
Selbjørn Austevoll	Konsesjonssøknad sendt.	40	ca. 320 mill. kr.	155	7750
Fræna	Konsesjonssøknad sendt.	95	500-700 mill. kr.	155-240	Opptil 12000
Havøygavlen	16 møller i drift.	40	2,5 MW pr. mølle	118	5900
Utsira	2 møller i drift.	1,2	0,6 MW pr. mølle	3,4	170
Båtsfjord	Konsesjon søkt 70-330 møller.	50-1000	0,8-3,5 MW pr. mølle	160-3200	Maksimalt 160000
Bugøynes	Forhåndsmeldt 150 - 300 møller	50-100	1,5-3,5 MW pr. mølle	160-275	Maksimalt 13500
Snøfjord	Forhåndsmeldt 70-150 møller	50-300	0,8-3,5 MW pr. mølle	160-960	48000
Harbaksfjellet	Konsesjon gitt. 33 møller.	90	2-3 MW pr. mølle	235	11750

Tabell 3 - Nøkkeltall, vindkraftprosjekter fra Hydro og Statkraft. Installert effekt er angitt for noen av prosjektene.¹⁸

Hva er forskjellen på installert effekt årlig produksjon? Den oppgitte effekten på en vindmøllepark er gitt i MW, og angir maksimal mulig produksjon pr time på de installerte møllene. Et år består av 8760 timer, mens det er kalkulert at «brukstid» på et vindkraftanlegg er 3000 timer. Med brukstid menes en stipulert tid med full effekt. I virkeligheten har man ikke 3000 timer med full effekt, og 5700 timer avslått, men jevnt over variert effekt. Møllene slås derimot automatisk av ved vindstyrke over 25 m/s og under 5 m/s.

I figur 21 ser vi en effektkurve som viser hvordan produsert elektrisitet varierer ved vindhastigheten



på en mølle med maks produksjon på 1.5 MW. Tilsvarende kurver vil man få for andre møller. Da med utflating på den gitte møllens maksimum, og omtrentlig lik av krumning fra 0 til maks.

1 MW installert effekt vil dermed tilsvare 3000 MWh = 3GWh årsproduksjon. Brukstid på 3000 timer utgjør et landssnitt. For enkelte vindkraftprosjekter kan det brukes en annen faktor.

18 www.statkraft.no Derunder konsesjonssøknader for vindkraftanlegg. Pdf filer,

Tallene for Smøla 1, 2 og Hitra er i helhet hentet fra Statkrafts nettsider. Tall for de øvrige prosjektene er kun delvis oppgitt, (de fleste steder kun oppgitt installert effekt), og de øvrige tall er beregnet, ved bruk av faktorer basert på tallene for Smøla 1, 2 og Hitra.

Hva gjelder antall møller på de ennå ikke utbygde anlegg kan dette variere. Konesjonssøknadene baserer seg i utgangspunktet på total installert effekt, ikke antall møller. Flere av disse anleggene kan få møller på inntil 5 MW, mens de største på Smøla er på 2.3 MW. Lokale forhold kan medføre at det velges å installere mindre møller enn hva som er teknisk tilgjengelig. Andre relevante tall kan være at de fleste vindkraftanlegg vil medføre utbygging 5-10 km ny vei inne i området, og 5-20 km kraftledning, og anleggsvirksomheten før parken er klar til drift er fra ett til tre år. Anlegget på Smøla båndlegger totalt 18 km² Tilsvarende areal vil bli båndlagt ved utbygging andre steder.

Konesjonene gis for 25 år. Hva angår lønnsomhet må vindparkene tilfredsstillende lønnsom inntjening med planlagt avvikling når konesjonsperioden utløper.

Offshore vindturbiner

Offshore vindturbiner er ny og spennende teknologi som åpner mange muligheter. Kostnadene er høyere enn ved montering på land, i hovedsak knyttet til økte kostnader ved montering, fundamentering, frakt, vedlikehold m.m. Men på sikt vil det likevel kunne være lønnsomt på grunn av bedre vindressurser, lavere turbulens, og større muligheter for storskala turbiner.

Firmaet Havgul AS har utviklet planer for verdens største offshore vindparker (fire områder) på Mørkekysten utenfor Molde/Ålesund/Hustavika. Dette er et meget spennende prosjekt.

For Havsul-prosjektet viser foreløpige målinger en middel-vindstyrke gjennom hele året på over 8 m/s. Vanndybdene i området er fra 4 til 30 meter, og egnet for utbygging. Ved å flytte turbinene til sjøs reduseres eller utelukkes mange av miljøhensynene man må ta ved utbygging til lands, det være seg visuell forurensning, «blitz-effekt» for naboer, båndlagt utmark, m.m. Andre hensyn må derimot tas, blant annet til fiskeri og skipsfart. Sjøfugler vil også utgjøre en faktor som må tas hensyn til.

Havsul-prosjektet er blitt sagt å være fullt ut realiserbart. Av viktige faktorer nevnes at Møre og Midt-Norge for øvrig har kraftunderskudd i dag. (Det er netto import av kraft fra andre landsdeler.) Videre er det godt utbygd kraftnett for transport i distriktet. Videre vil Ormen Lange prosjektet medføre ytterligere kraftbehov, samt oppjustering av kraftnettet. Man må koble seg på det oppjusterte kraftnettet med en sjøkabel. Men en grov helhetsvurdering gjør at dette prosjektet virker fornuftig. Man kan eventuelt også argumentere med at en slik stor offshorepark vil plassere Norge langt fram som vindkraftnasjon og bli en viktig aktør i forskning og videreutvikling. Danmark har flere offshoreparker som er i drift. Den eldste (Vindeby) sto på plass allerede i 1991, den største hittil var ferdig i 2002. Det er vedtatt ytterligere utbygginger som skal føre til totalt 13,5 TWh/år produsert av havparker alene, innen 2030. Også vindkraftnasjonen Tyskland er i gang med dette.

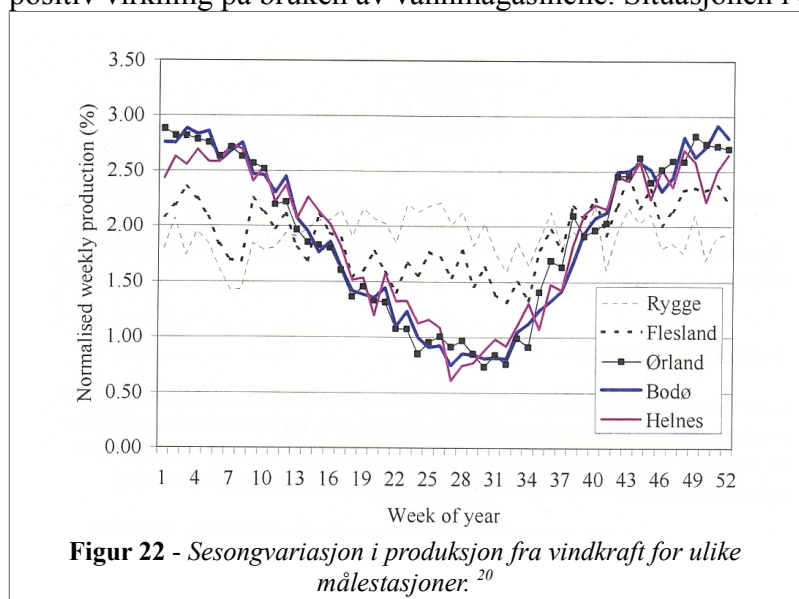
19 *Figur 21 er fra Teknisk rapport. Integrasjon av vindkraft.*

Bunnforholdene i Norge er derimot vanskeligere enn i Danmark (jevnt over dypere). Det er i dag tilgjengelig teknologi for å realisere Havsulprosjektet, men en storstilt utbygging av ytterligere offshore vindparker i Norge forutsetter ny teknologi.

Integrasjon av vindkraft i eksisterende nett

Plasseringen av vindparkene vil påvirke effektflyten i kraftnettet. Små vindkraftparker (<10 MW) kan integreres direkte i distribusjonsnettet og kan ha en positiv virkning på effektflyten, redusere kraftunderskudd og være plassert nærmere sluttbruker. Noe som også reduserer effektap i transportnettet. Større vindkraftparker må tilknyttes regional og sentralnett.

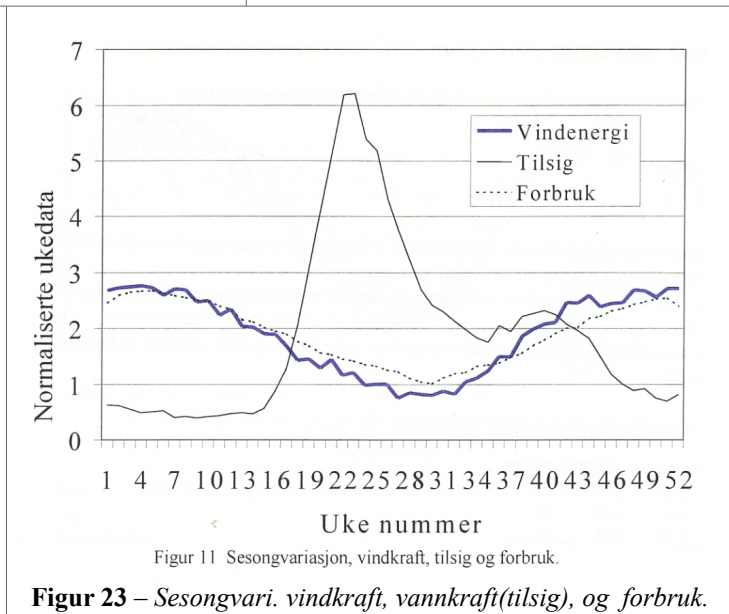
Sesongvariasjonen i vindkraften viser seg å samstemme meget bra med forbruket av elektrisitet. (Mest produksjon om vinteren, når det er mest forbruk). Integrering av vindkraft kan derfor ha en positiv virkning på bruken av vannmagasinene. Situasjonen i dag er at man i løpet av vinteren får



tomme vannmagasin ved vannkraftverkene. Dette er fordi man ikke har tilsig av nytt vann, og nedbøren kommer i form av snø, samtidig som man tapper mye vann for å øke kraftproduksjonen. Vindforholdene er derimot best om vinteren. Dette ser vi i figur 22, som baserer seg på faktiske målinger ved utvalgte målestasjoner. Mens figur 23 viser en gjennomsnittlig sesongvariasjon sett i forhold til sesongvariasjon i forbruk, og vannkraft produksjon.

Tilsiget til vannmagasinene kan variere så mye som 30 % fra år til år. Tilsvarende variasjon for vindkraft er 20 %. Man har ikke påvist noen sammenheng mellom variasjoner i vindstyrke og tørkeperioder.

Det norske kraftsystemet er 99 % basert på vannkraft. Variasjoner av tilsig både geografisk og sesongmessig gjør at det er lønnsomt å samkjøre de ulike verkene. Dette har tidligere skjedd lokalt, regionalt og senere nasjonalt, koordinert av forskjellige organisasjoner. I dag er dette underlagt det felles nordiske



Figur 23 – Sesongvari. vindkraft, vannkraft(tilsig), og forbruk.

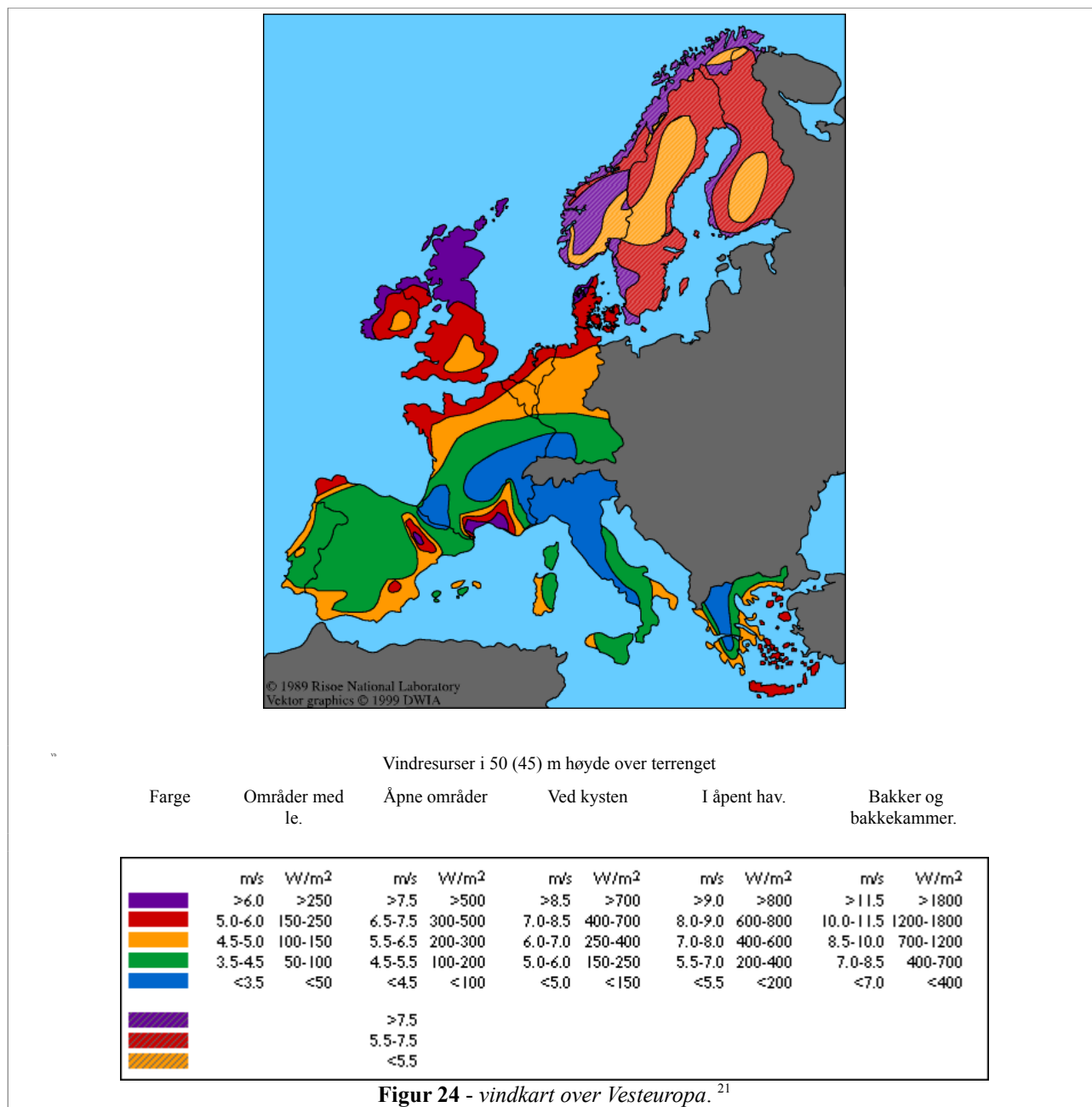
kraftmarkedet Nordpool. Også utbygginger, både eksisterende og framtidig, inngår i det nordiske

20 Teknisk rapport. Integrasjon av vindkraft

kraftmarkedet.

De nordiske landene har omfattende planer for utbygging av vindkraft. Danmark på sin side har igangsatt handlingsplan for utfasing av fossil energiproduksjon, hvor en vesentlig del erstattes med vindkraft. Målsettingen er 10 % vindkraft innen 2005 (oppnådd allerede i 2001), og 50 % innen 2030. Danmark har i dag en total elproduksjon på ca 35 TWh.

Tilsvarende tall for resten av Norden er: Sverige, 4 TWh vindkraft innen 2010, Norge 3 TWh vindkraft innen 2010, og Finland 1 TWh.



Figur 24 viser hvordan vindressursene varierer i Vesteuropa. Som vi ser ligger Norge svært godt an. Den eneste forklaringen på at Norge ikke allerede har blitt en stor vindkraftnasjon er at vi også har vært velsignet fra naturens side med stort potensiale for vannkraft og rike olje og gass resurser. Men

²¹ Figur 24 hentet fra www.windpower.org

nye behov krever nye løsninger. Danmark er et foregangsland. På tross av dårligere vindforhold enn Norge har de satset stort, og hele 20% av elektrisiteten produseres ved bruk av vindkraft. Norge har mer komplisert landskap som vil gi dyrere utbygginger enn i Danmark. Men vi vil også ha større effektivitet. (Den oppgitte driftstiden på møller i Danmark og Tyskland er 2600 timer. Mot nevnte 3000 i Norge.)

Teknologien finnes, og vil stadig videreutvikles. Kostnadene er synkende. De vedtatte 3 TWh innen 2010 er absolutt oppnåelige, og i bransjen håpes det at dette målet oppgraderes.

Selv om vindkraft er antatt/simulert å ha en produksjonsfordeling gjennom året som samstemmer bra med forbruksfordelingen vil det være nødvendig med energilagring i perioder med overproduksjon. Det finnes en rekke typer batterier som vil være egnet til dette formålet. Man kan si at batteriet tilsvarer vannmagasinene i vannkraft-sammenheng. Helt sammenlignbart er det likevel ikke, da man ved vannmagasin lagrer energien i form av vann med høy potensiell energi, i et lager før produksjon av elektrisk energi, mens man med vindkraft og batterier lagrer elektrisk ladning (man kan ikke «lagre» vinden).

Vindkraft og energilagring:

En annen måte å utføre energilagring er ved å bruke den elektriske strømmen til hydrogenproduksjon, i form av elektrolyse.

Hydro åpnet sommeren 2004 et forsøksanlegg på Utsira. Se figur 25. Anlegget består av 2 vindturbiner, hver med tårn på 46 m. Rotorblad med 40 m i diameter, og maksimal ytelse på 600 kW. Den ene møllen produserer kun for det åpne markedet, mens den andre produserer for et lukket «forsøkssamfunn». Dette har 10 husstander, som primært forsynes med hele sitt kraftbehov fra dette anlegget. (De var kun tilknyttet det ordinære kraftnettet i tilfelle stans på forsøksanlegget.) Overskuddskraften kjøres inn i en elektrolyserør. Her produseres det hydrogen og oksygen ved å spalte vannmolekyler. Elektrisk strøm (dvs. overskuddsstrøm fra vindmøllen) føres igjennom vann iblandet lut (elektrolytt, for å gi vannet lederegenskaper), fra en anode til en katode. Hydrogenet lagres så i en trykktank. Ved dårlige vindforhold eller økt kraftbehov kan hydrogenet brukes til å produsere elektrisk kraft ved bruk av brenselcelle. Forsøksanlegget har eget utstyr for dette, brenselcelle og hydrogenmotor som genererer strøm.

Under gode vindforhold produserer vindturbinen nok til å forsyne både husholdningene, og en elektrolyserør. Når det blåser for mye eller for lite og møllene står stille, brukes lagret hydrogen til å produsere strøm. Når det blåser litt, men ikke tilstrekkelig til å forsyne husholdningene, får vindturbinene støtte til kraftproduksjon fra brenselcelle og hydrogenmotor. Anlegget har utstyr til å sikre stabil kraftproduksjon.



Figur 25 - Bilde av forsøksanlegget på Utsira.²²

Hydro har signalisert at de ønsker å stenge dette forsøksanlegget i løpet av første halvdel av 2006. Erfaringene har imidlertid vist at et slik småskala anlegg er nok til å gjøre små øysamfunn selvforsynte med elektrisk kraft. Det er mange fordeler med dette. Ikke minst vil man ved utbygging av slike anlegg sikre stabil strømforsyning i områder med lite utbygd infrastruktur. Små kystkommuner kan gjøres helt uavhengige av det ordinære kraftmarkedet. Tanken er at det i enkelte tilfeller vil være billigere å bygge ett småskala anlegg som forsyner kommunen framfor å bygge ut det nasjonale hovedkraftnettet.

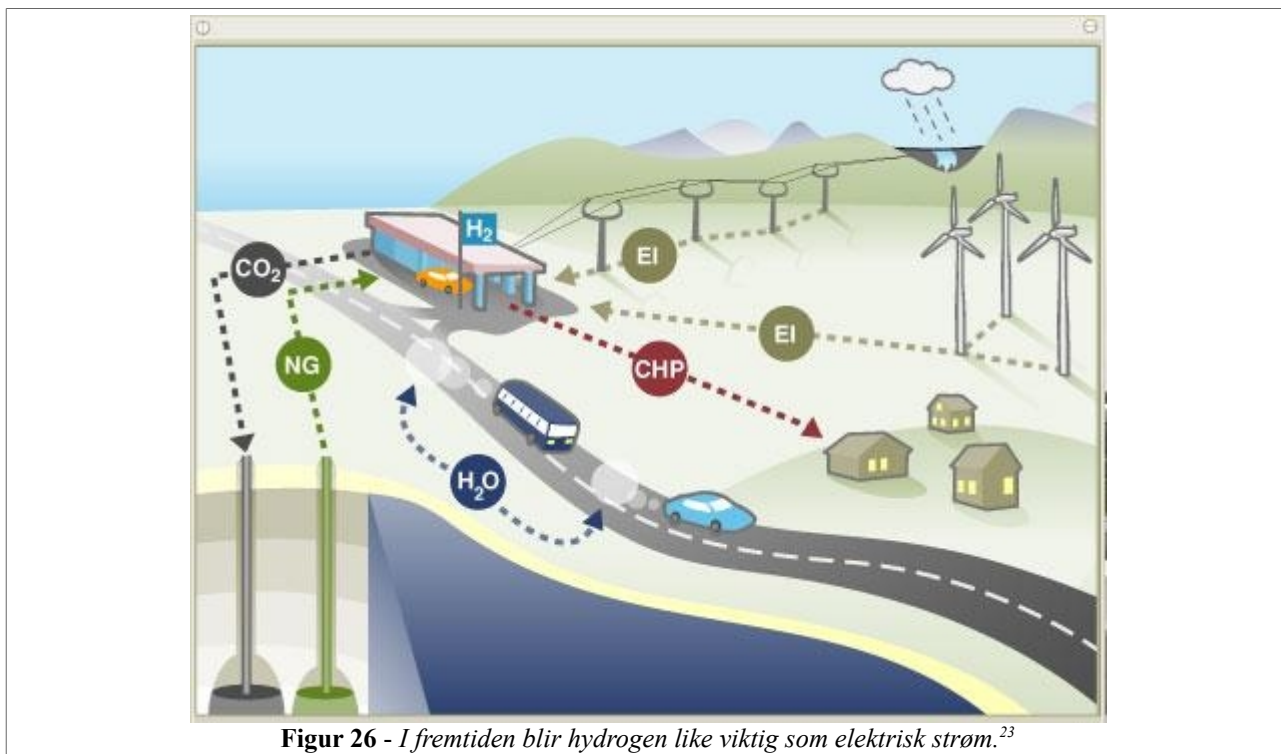
Det er et stort energitap ved overgangen fra elektrisitet med elektroner som energibærere, via hydrogen, (samt nytt til) bruk av hydrogen i produksjon av elektrisitet. Dette er derfor ikke en særlig egnet metode (med dagens teknologi) for energilagring ved store vindparker. Derimot går man mot et samfunn hvor hydrogen som energibærere vil bli brukt istedenfor fossilt brennstoff i biler, busser, båter osv. Man kan tenke seg at det anlegges små elektrolyseanlegg rundt omkring ved de forskjellige vindkraftverkene, og hvor hydrogenet her blir transportert til lokale

«bensin»stasjoner ol. Hydrogen kan også lages ved nedbryting av hydrokarboner (olje, naturgass, kull). Men når man etter hvert får hydrogendrevne kjøretøy og maskiner som allemannseie, nettopp for å spare forbruket av fossilt brennstoff, vil det være lite nyskapende og et skritt i feil retning om ikke hovedandelen av den økte hydrogenproduksjonen kommer av elektrolyse. Vindkraft er bare en av mange måter å lage strøm til elektrolysen. Men kombinert med nettopp denne form for energilagring åpner det seg mange muligheter for framtidig satsing på vindkraft.

Som allerede kommentert er det gode vindforhold i Finnmark, mens hovedkraftnettet for eksport av kraft fra Finnmark og sørover er dårlig utbygd. Her vil nok veldig mange prosjekter bli skrinlagt, nettopp på grunn av manglende kraftnett. Men man kan eventuelt filosofere over hvorvidt Finnmark

²² Figur 25 fra www.hydro.com/library/attachments/our_activities/new_energy/wind/50851_Utsira_N.pdf

er et egnet sted for storskala hydrogenproduksjon tilknyttet vindkraft, når vi i overskuelig framtid får et samfunn hvor hydrogen utgjør den viktigste kraftbærer.



Konkluderende

Gjennom alle tider har mennesket gjort seg nytte av kreftene i vinden. I Norge går linjen tilbake til seilskutetiden og handel, og videre til vikingene og deres skip. At vinden er kraftigst til havs er noe som med ny teknologi kan komme vindkraft utbyggingen til gode, da det gjør mulig bygging av havparker med store møller som gir høy produksjon.

Men vindmøller er et nytt fenomen i norsk historie. Røttene går til andre himmelstrøk og langt tilbake i tid. Perserne utviklet sine vertikalkslede vindmøller til maling av mais rundt år 950 e.Kr. Til Europa kom møllene senere, og da muligens som følge av at korsfarerne hadde møtt den persiske teknikken. Kina tok også erfaring fra Persia, men utviklet denne videre.

Energi kan ikke oppstå eller forsvinne, men kun omdannes fra en form til en annen. Vindmøllens oppgave er å omforme bevegelsesenergien i vinden til en rotasjonsenergi i vindmøllens rotorsystem. (Her brukes ordet rotor om alt som roterer som direkte følge av vinden.) I elektriske omdannes rotasjonsenergien videre til elektrisk energi.

Det dannes et lufttrykk under rotorbladet som presser bladet oppover, dette skaper rotasjonsbevegelsen. Tuppen på rotorbladet vil ha en høyere hastighet i forhold til vinden rundt tuppen, sammenlignet med et punkt lengre inne. Dette gir at de to punktene har forskjellig resultantkraft. Rotorbladene er dreid for å sikre at retningen av resultantkraften er lik over hele bladet.

De gamle rotorbladene var lagd av metall og var svært tunge. I moderne møller brukes det

²³ Figur 26 er fra http://statkraft.no/pub/annen_miljoanlig/hydrogen/kort_forklart/index.asp

polyester, glassfiber og karbonfiber. Moderne møller har også en rekke teknikk som styres automatisk av datasystemer, og som gjør at møllen alltid er innstilt optimalt til vindforholdene.

Strømmen lages i en generator.

Dette er som regel en trefaset vekselstrømsgenerator. Denne fungerer ved at tre elektromagneter (lederspoler) er fastmontert, med innbyrdes like vinkler (utgjør hjørnene i en likesidet trekant med vinkler hver på 60°) og en permanent magnet som roterer i midten, (sentrum av trekanten). Når den permanente magneten roterer vil dette føre til endringer i magnetfeltet gjennom lederspølene, og dermed indusere elektrisk spenning i disse. Spenningen vil være en sinusformet vekselspenning, med tre faser. Hver av dem $1/3$ fase bak den forrige.

De fastmonterte spolene kalles *stator* siden de er statiske. Den roterende permanente magneten kalles *rotor*. Hver av fasene har maksimum og minimum når Rotorens poler er parallell med spolen(e).

Den elektriske strømmen beholdes i 3 faser gjennom hele høyspentnettet. Man kan lett se at en en høyspent mast bærer tre kraftige ledere. Disse fører strømmen i de tre fasene.

Nettet i Europa har en frekvens på 50 Hz. Dette vil tilsvare at rotoren dreier 50 runder pr sekund. Man kan derimot redusere antallet omdreininger ved å øke antallet poler. Både på stator og rotor. Har man f.eks. 12 elektromagneter som stator, og en rotor med 8 poler vil man bare trenge 12.5 omdreininger pr. sekund. Generatorer med mange magneter i stator og rotor kalles lavhastighetsgeneratorer. Høyhastighetsgeneratorer er derimot mindre i volum og billigere å bygge. I de fleste vindturbiner bruke 4-6 poler på rotoren.

Rotoren er koblet til propellens rotasjon via en girkasse. Girkassen har til hensikt å hjelpe til med stabil høy frekvens. Permanente magneter blir fort avmagnetisert når de befinner seg i magnetfelt. (Spolene som det induseres strøm i vil på sin side også omdanne seg med magnetfelt.). Derfor er det vanlig å istedenfor bruke en elektromagnet som rotor, og hvor denne forsynes med likestrøm via et børste system.

Den overstående metoden kalles for synkrongenerator, og er nødvendig bakgrunnskunnskap. Det er denne teknikken som brukes i de fleste vannkraftverk. Derimot er det en annen teknisk løsning som er mer vanlig brukt i vindmøller. Denne kalles asynkron kortslutningsgenerator. Her består rotoren av to ringer som er bundet sammen av mange stenger av kobber eller aluminium. Videre en kjerne av jern. I denne typen generator er statorene tilknyttet en spenningskilde, og dette induserer spenning i rotoren.

Det er her vinden som driver rotoren rundt, (ikke magnetfeltet). Når frekvensen er høy (omlag 50HZ) roterer magnetfeltet i rotoren fortere enn magnetfeltet i statoren. Det induseres strøm i stengene. Sirklene kortslutter stengene og det går strøm i stengene. Rotoren danner så sitt eget magnetfelt. Denne generatoren holder stabil frekvens og kan kobles direkte til elektrisitetsnettet.

Generatorer avgir mye varme. Den vanligste kjølemetoden i vindmøller er luftkjøling. Vannkjølingssystemer finnes.

Nyere teknologi gjør det mulig å bygge vindmøller til havs. Dette anses som en kjempefordel da friksjonen vinden møter her er minimal, og det er følgelig et stort potensiale og hente ut med vindmøller. På land bremses vinden av mange andre objekter i tillegg til vindmøllene. Dette skaper også uønsket turbulens.

Når vindmøllen roterer bremser den luften, det er vindens reduksjon i bevegelsesenergi som får møllebladene til å rotere. Energien i vinden er proporsjonal med vindhastigheten i tredje potens. Effekten som som kan tas ut av en vindmølle avhenger videre av arealet som rotorbladene sveiper over. Effekten er gitt ved $P = \frac{1}{2} \dots$ *lim inn* Dobling av lengden på rotorbladene vil altså gi en firedobling av energien. Lokale forhold kan likevel føre til at det er ønskelig med små møller framfor store.

Nedbremsingen fører til at vinden bak møllen dekker et større areal enn den dekket før den traff møllebladene, dette gir avbøyning av vindretningen.. Betz lov sier at det er maksimalt mulig å omdanne 59% av vindens energi.

Vindmøllene lager også selv turbulens, og dette sammen med avbøyningen påvirker hvordan man bør plassere vindmøllene i en vindmøllepark i forhold til hverandre. Det er vanlig med inntil 9 ganger rotordiameter mellom møllene, og man monterer normalt ikke en mølle bak en annen.

Store møller leverer normalt en 690V trefaset vekselstrøm. I transformator transformeres den opp til høyspent på 10 000 – 30 000 V.

Kostnadene på utbygging av vindkraftparker har sunket kraftig. Det er ventet at den fortsetter å synke. Rapporter av nyere dato antyder et vindkraftpotensiale på 250 TWh til en pris på 27-40 øre/kWh. Omlag 70% av potensialet ligger i Finnmark. Dårlig utbygd kraftnett og kraftoverskudd i regionen gjør at mesteparten av dette ikke er realiserbart. I Midt Norge har man gode vindforhold, kraftunderskudd og godt utbygd nett. Her vil det være lønnsomt å bygge ut vindkraft.

Det er i dag bygd ut 274 MW installert effekt, innvilget konsesjon på 845 MW, og levert konsesjonsøknad på 1766 MW samt mange forhåndsmeldte prosjekter.



Figur 27 – Simulert bilde av planlagte Kjøllefjord vindpark.²⁴

Installert effekt svarer til maksimal produksjon pr. time. Det opereres med en driftstid på 3000 timer på vindkraftanlegg i Norge. Stipulert årlig produksjon blir dermed 3GWh pr. 1MW installert effekt.

Utbygging av vindkraftanlegg medfører båndlegging av relativt store arealer av utmark. Det trengs også veier som må tåle lange og tunge kjøretøy, og strekking av flere km. Kraftledning.

Danmark og Tyskland satser nå stort på havparker. Bunnforholdene

i Norge gjør dette vanskeligere å realisere her. Eneste planlagte havpark så langt er Havsul prosjektet på Mørekysten. Dette er likevel realiserbart, da allerede påstartede utbyggingen av Ormen Lange feltet medfører utbygging av kraftnettet, det er gode vindforhold, og kraftunderskudd i regionen.

Små vindparker kan integreres direkte i distribusjonsnettet nærme sluttbruker. Store parker må kobles på region og sentralnettet. Vindkraftens sesongvariasjon følger i stor grad forbruket. Dette er en positiv effekt i et land hvor 99% av strømmen lages med vannkraft. Vannkraften er på sin side i motfase med forbruket. (Størst produksjon når det er minst forbruk.) I Norge er det vedtatt

24 Figur 27 er fra http://statkraft.no/pub/vindkraft/prosjekter_og_soknader/kjoellefjord/index.asp

utbygging av vindkraft tilsvarende 3 TWh innen 2010. Dette målet vil bli nådd, og potensialet er mye høyere selv uten store utbygginger av kraftnettet.

I vindkraft lagrer man elektrisk ladning i batterier ved overproduksjon. Man kan bruke energien til å produsere hydrogen ved elektrolyse. Da er det hydrogenet som er energibæreren. Hydrogenet kan senere brukes til strømproduksjon ved bruk av brenselcelle. Denne teknologien kan åpne for at små øysamfunn kan bli selvforsynt av elektrisk kraft uten store utbygginger av kraftnettet. Hvilket videre kan bety at endel av det «ikke realiserbare» potensialet i Finnmark likevel kan realiseres. Det er lettere å lagre hydrogen på en tank, enn å lagre elektrisk ladning i et batteri. Tanken har konstant kapasitet, (volum), mens batteriets evne til å lagre elektrisk ladning blir svekket over tid. Derimot er det store effekttap på først lage hydrogen, for så videre å lage strøm ved bruk av hydrogen igjen.

I stor skala tenkes denne teknologien brukt ved anlegg hvor produksjonen av hydrogen er selve hovedformålet. Dette kan være ønskelig etterhvert som vi går mot et samfunn hvor hydrogen vil brukes som drivstoff i biler, busser og andre større maskiner. Dette kan kanskje åpne for utbygginger i Finnmark.

En totalvurdering ut i fra de forhold vi har kartlagt, og oppsummering av konklusjoner i rapporter fra Sintef, NVE, m.f. vi har lest gjør at vi mener Norge har et reelt potensial for å bli en stor vindkraftnasjon.

Kildehenvisninger

Forsideillustrasjon

Bilde av en mølle i Vikna vindmøllepark
www.safeinherit.net/nord_trondelag.htm

Figurer

Kilder til figurer er gitt fortløpende i fotnoter. Figurer hvor ikke kilde er oppgitt er eget produkt.

Historikk

Bøker og publikasjoner

Power from wind: A history of windmill technology
1994 Cambridge University Press
Richard L. Hills
Kap. 1 og 2

Nettressurser

Danske Vindmølleindustrien
<http://www.windpower.org/da/tour/>

Teknisk

Bøker og publikasjoner

Wind Turbine Technology
1994 Asme Press
David A. Spera
Kapittel 1,2,3 og 5

Nettressurser

Danske Vindmølleindustrien
<http://www.windpower.org/da/tour/>

Wikipedia.org
http://no.wikipedia.org/wiki/Elektrisk_motor
http://de.wikipedia.org/wiki/Elektrischer_Generator

Tidligere års prosjektoppgaver
[http://web.phys.ntnu.no/~stovneng/FY1303NY/prosjekt/van
nkraft.pdf](http://web.phys.ntnu.no/~stovneng/FY1303NY/prosjekt/van
nkraft.pdf)

Norske tilstander

Bøker og publikasjoner

Teknisk rapport. Integrasjon av vindkraft i det norske kraftsystemet.
2001 Sintef Energiforskning.
K.O.Vogstad, M.M. Belsnes m.f.
Side 9, 10-11, 14-15, 23-24, 27.

Nettressurser

Havsulprosjektene,
Eksisterende offshoreparker, Danmark:
<http://www.havsul.no/Web?id=207.2.3.214.214>
Om forholdene og områdene:
<http://www.havsul.no/Web?id=207.2.3.210.210>
Forhåndsmeldingen til prosjektet:
<http://www.havsul.no/Web?id=207.2.3.211.211>

Norsk Vassdrags og energidirektorat.
Rapport 17-05
<http://www.nve.no/FileArchive/337/Rapp%2017-2005.pdf>
Idriftsatte og planlagte anlegg
[http://www.nve.no/modules/module_111/netbasNVE.asp?sc
ript=8](http://www.nve.no/modules/module_111/netbasNVE.asp?sc
ript=8)

Norsk Hydro.
www.hydro.no Derunder konsesjonsøknader for
vindkraftannlegg. Pdf filer.

Statkraft.
www.statkraft.no Derunder konsesjonsøknader for
vindkraftannlegg. Pdf filer.

Offentlige meldinger

Danmark, Energi og miljøministeriet, 1996. Sverige,
Svenska Energimyndigheter, 1999. Norge,
Stortingsmelding 29, 1999. Og Finland, Finenergy, 2000.
Gjengitt i: *Teknisk rapport. Integrasjon av vindkraft i det
norske kraftsystemet.*

Personer

Samtale med, Post. Doc. Magnus Korpås. Institutt for
kraftteknikk. NTNU / Sintef.

Energibånd i faste stoffer med anvendelser

Prosjektrapport i FY1013

**Einar Stiansen
Eivind Tore Hansen
Morten Klungervik
Anders Langseth**

Innholdsfortegnelse

- s.3 Abstract
- s.4 Innledende ord
- s.4 Kvantefysisk beskrivelse av elektronet
- s.4 Hydrogenatomet
- s.4 Kvantisering
- s.5 Atomer med flere elektroner
- s.6 Eksklusjonsprinsippet
- s.6 Energibånd
- s.8 Kronig-Penney modellen
- s.10 Elektrisk ledningsevne
- s.12 Halvleder
- s.12 Doping; n-type og p-type
- s.14 P-n diode
- s.16 Innebygd potensial
- s.16 Energibånddiagram for en p-n overgang
- s.18 Bias over en p-n diode
- s.19 Matematisk beskrivelse av en ideell diode
- s.21 Grenseflaten som en kapasitans
- s.22 LED-diode
- s.23. Fotodetektorer og solceller
- s.24. Transistorer; bipolare
- s.26. Strøm i en transistor
- s.27. Energibånd i en transistor
- s.28. Operasjonsmodi
- s.29 Anvendelser av transistor (forsterker og bryter)
- s.31 Konklusjon
- s.32 Referanseliste

Abstract

I klassisk fysikk blir elektroner beskrevet som punktladninger som går i bane rundt atomkjerner. I en kvantefysisk beskrivelse betraktes elektroner derimot som bølger beskrevet ved en bølgefunksjon Ψ . I kvantefysisk teori stilles det spesielle krav til denne bølgefunksjonen, noe som fører til kvantisering. For å kunne beskrive elektronet må ha kjennskap til elektronets fire såkalte kvantetall. I følge eksklusjonsprinsippet kan ikke to elektroner i samme system ha samme kvantetall, slik at tilstander utover grunntilstanden blir påtvunget.

Resultatet av de kvantefysiske prinsippene er at elektroner i et atom kun kan inneha visse energiverdier og man kan snakke om tillatte og forbudte energitilstander. Når atomer går sammen i et fast stoff vil man få splitting av tillatte energinivåer, og dermed energibånd. Mellom energibåndene vil man få båndgap der det ikke finnes tillatte energinivåer. Størrelsen på dette båndgapet og fordelingen av elektroner for et bestemt stoff bestemmer stoffets ledningsevne og dermed om det kan klassifiseres som leder, halvleder eller isolator.

Halvlederens ledningsevne kan forsterkes vesentlig ved å forurense halvlederen med andre typer atomer. Dette kalles doping, og man skiller mellom p-type og n-type dopede halvledere. Når man doper to ulike deler av samme halvlederstykket med ulik type doping får man det man kaller pn-overganger. Slike pn-overganger viser seg å ha interessante egenskaper som utnyttes i elektriske komponenter som dioder og transistorer.

Transistorer og dioder brukes på mange ulike områder og inngår i så å si alt av moderne elektriske artikler. Dioder brukes for eksempel som likerettere i blant annet batteriladere, og enkelte dioder har den egenskapen at de kan brukes som lyskilder. Transistorer har egenskaper som gjør at de kan brukes i forsterkere og i alle slags logiske kretser.

Innledende ord

Omtrent halvparten av alle fysikere i verden er faststoff fysikere. Nyvinninger innen nettopp dette feltet har revolusjonert det teknologiske samfunnet, oppfinnelser som for eksempel transistoren er grunnmuren i dagens IT hverdag. Vi ser i dette prosjektet på hvordan vi får såkalte bånd av tillatte energier for et elektron i et fast stoff, og hvilke implikasjoner dette har.

Kvantefysisk beskrivelse av elektronet

Båndteori for faste stoffer er et produkt av en kvantemekanisk beskrivelse av elektroner. Det faller derfor naturlig å introdusere grunnleggende kvantemekaniske prinsipper som er nødvendige for å forstå disse elektronenes natur.

I klassisk fysikk ser man på elektronet som en punktladning som følger bestemte baner rundt atomets kjerner. I kvantefysikken beskrives elektronene som bølger, gitt ved en

bølgefunksjon; $\Psi(\mathbf{r}, t)$. Der $|\Psi^2|$ er sannsynlighetstettheten for å finne elektronet i en bestemt

posisjon ved tiden t . Som en konsekvens av denne materiebølgeteorien sier Heisenbergs usikkerhetsrelasjon at posisjon og impuls er omvendt proporsjonale størrelser, dermed lar det seg ikke gjøre å gi en skarp beskrivelse av et elektrons posisjon og impuls samtidig. For en

retning i rommet, si x-aksen har vi altså; $\Delta x \Delta p_x \geq \frac{\hbar}{2}$, dermed har man ingen mulighet til å

bestemme posisjonen til elektronet, langt mindre kan vi snakke om bestemte baner.

Hydrogenatomet

Vi betrakter i første omgang et nøytralt, enkeltstående hydrogenatom. I dette tilfellet beveger elektronet seg i potensialfeltet (Coulombfeltet) fra kjernen. Verdien av dette er veldefinert, Schrödingers likning lar seg derfor løse analytisk for dette spesialtilfellet.

Den tidsuavhengige Schrödingers likning i tre dimensjoner;

$$-\frac{\hbar}{2m} \nabla^2 \Psi(\mathbf{r}) + U(\mathbf{r})\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (1)$$

Løsningen av denne partielle differensiallikningen kan finnes ved å separere variable og er på formen $\Psi(r, \mathbf{q}, \mathbf{f}) = R(r)\Theta(\mathbf{q})\Phi(\mathbf{f})$, altså som en kombinasjon av tre likninger der hver er avhengig kun av en variabel i det sfæriske koordinatsystemet. Vi skal se at en kvantisering av Ψ fremkommer naturlig som en konsekvens visse krav man må stille løsningene.

Kvantisering

Den semiklassiske Bohrteorien innførte kvantetallet n for å beskrive energinivåene i ett atom. Den bølgemekaniske beskrivelsen av elektronet fører med seg ytterligere kvantisering av beskrivelsen av elektronet.

- N-kvantetallet fra Bohr-teorien er å finne igjen i den moderne kvantefysikken, og gir tillatte energitilstander for elektronet. For eksempel vil energinivåene for hydrogen være gitt ved;

$$E = -\frac{m_r e^4}{2(4\pi\epsilon_0)^2 \hbar^2} \left(\frac{1}{n^2}\right), n=1,2,3\dots$$

Legg merke til at m_r ikke er elektronmassen, men den reduserte massen.

- I Bohrteori gir hver verdi for banedreieimpulsen en tilhørende verdi for energi. Men Bohr-teorien begrenser elektronets bevegelse til sirkelbevegelse, straks denne begrensningen er fjernet kan man vise at hver n kan ha flere verdier av banedreieimpuls. Disse verdiene er gitt ved;

$$L = \sqrt{l(l+1)}\hbar, \quad l = 0, 1, 2, \dots, (n-1)$$

Verdiene fremkommer ved å kreve at $\Theta(\mathbf{q})$ tar endelige verdier ved $\mathbf{q} = 0$ og $\mathbf{q} = \mathbf{p}$ (\mathbf{y} må være normaliserbar).

Av historiske årsaker er det vanlig å tilegne en bokstav til de første verdiene av l :

Tabell 1. Tilordning av bokstav for banedreieimpuls

l	0	1	2	3	4
Bokstav	s	p	d	f	g

- I tillegg til at størrelsen på banedreieimpulsen er kvantisert har vi også romlig kvantisering det vil si at også retningen til \mathbf{L} kvantisert. Dette fremkommer på grunn av kravet om at $\Phi(\mathbf{f}) = \Phi(\mathbf{f} + 2\mathbf{p})$. \mathbf{L} kan dermed kun ta visse orienteringer i rommet. Hvis vi ser på de tillatte verdiene av \mathbf{L} langs z-aksen er de gitt ved;

$$L_z = m_l \hbar, \quad m_l = 0, \pm 1, \dots, \pm l$$

Vi ser nå at mens den semiklassiske Bohrteorien gav et kvantetall, gir den bølgemekaniske beskrivelsen av elektronet tre. Ved å kreve sfærisk harmoniske bølgefunksjoner tvinger ytterligere kvantisering seg fram.

- Gjennom eksperimenter er det vist at elektronet selv også har et "spinn". Dette spinn er konstant og er en egenskap ved elektronets natur. Retningen er derimot kvantisert i nøyaktig to retninger, "spinn-opp" og "spinn-ned". Likningen for størrelsen på spinn har samme form som for banedreieimpulsen; $S = \sqrt{s(s+1)}\hbar$, og siden banedreieimpulsen gav $2l+1$ ulike tillatte komponenter og vi her krever nøyaktig to løsninger må vi ha $s = \frac{1}{2}$.

$$\text{Altså } S = \frac{\sqrt{3}}{2} \hbar$$

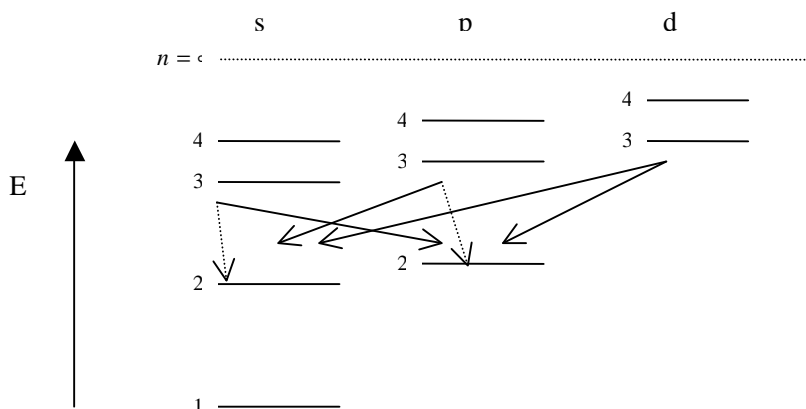
$$\text{Retningen er kvantisert i henhold til; } S_z = m_s \hbar, \quad m_s = \pm \frac{1}{2}.$$

Vi har altså hele fire kvantetall n, l, m_s, m_l som må oppgis for å beskrive et elektron i et hydrogenatom.

Atomer med flere elektroner

Vi har så langt bare sett på beskrivelsen av elektronet i hydrogenatomet. I ett atom med flere enn ett atom beveger ikke lenger elektronet seg i et sentralfelt, men også i feltet induisert av de andre elektronene. En nøyaktig og analytisk løsning av Schrödingers ligning for slike tilfeller er umulig. Men, ved å gjøre den såkalte sentralfelttilnærmingen kan vi igjen basere oss på en "to-legeme-problemstilling". I denne tilnærmingen ser vi på hvordan elektronet beveger seg i potensialet fra kjernen og gjennomsnittsbidraget til potensialet fra alle andre elektroner. Like fullt vil totalenergien til elektronet i et mangeelektronatom være avhengig av

både n og l . Uavhengigheten av bandedreimpulsen var en konsekvens av Coulombkraftfeltet og forsvinner når vi introduserer flere elektroner. På denne måten for man en fininddeling av tillatte energinivåer selv for en gitt verdi av hovedkvantetallet n . På figuren er det visst at et elektron som deeksiterer fra $n=3$ til $n=2$ ikke bare har en mulig energidifferanse, men hele seks. Deeksitasjonene med $\Delta l = 0$ er sagt å være ulovlige, selv om de opptrer en sjelden gang. Regelen er at spranget må tilfredsstillte $\Delta l = \pm 1$.



Figur 1,
deeksitasjon i et
mangeelektronato
m. Større verdi for
 l gir en høyere
totalenergi.
"Forbudte"
sprang med stippet
pil

Elektronene fordeler seg altså i hovednivåer bestemt av n , og undernivåer av disse bestemt av l . Men naturen tilstreber gjerne tilstander som minimaliserer energien, hvordan kan det da ha seg at ikke alle elektronene legger seg i grunntilstanden? Svaret ble formulert av Wolfgang Pauli, og har senere blitt kalt eksklusjonsprinsippet.

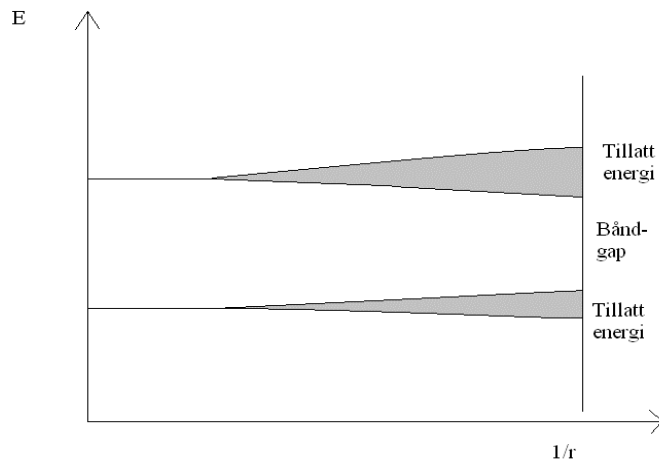
Eksklusjonsprinsippet

Grunnen til at elektroner blir tvunget opp i høyere energinivåer enn grunntilstanden er at innenfor et system kan to elektroner ikke ha samme sett kvantetall. Slik blir nå antallet elektroner på et bestemt energinivå begrenset. For en gitt verdi av n og l vil det som vi har sett være tilgjengelig $2l+1$ tilstander som følge av variasjon av m_l . I hver av disse tilstandene kan det være to elektroner som følge av to mulige retninger av elektronets egenspinn. Totalt har vi dermed $N=2(2l+1)$ tilgjengelige tilstander på hvert gitte nl -orbital ($l \leq n-1$).

Energibånd

Vi ser på atomer i en krystallstruktur med et stort antall atomer. Hvis vi nå antar at vi kunne "strekke ut" gitteravstanden, slik at atomene ikke påvirker hverandre på noen slags måte, da ville energinivåene i atomene tilsvare de diskrete nivåene beskrevet for enkeltatomet. Men ettersom denne gitteravstanden minskes vil en vekselvirkning mellom atomene føre til en splitting av energinivåene. Vi snakker om et stort antall atomer ($\square 10^{22}$) og hvert av disse vil gi opphav til et nivå innenfor det vi nå ser framstår som et

bånd av tillatte energitilstander. Bredden på båndet ΔE er bestemt av antallet atomer som er nær nok til å sterkt vekselvirke. Dette er et lite antall sett i forhold til antallet i systemet som helhet, dermed vil ikke båndbredden øke med antall atomer, men snarere vil alle de tillatte nivåene legge seg innenfor intervallet ΔE . Nivåene vil ligge så tett at vi essensielt kan se på det som en kontinuerlig fordeling av tilstander.



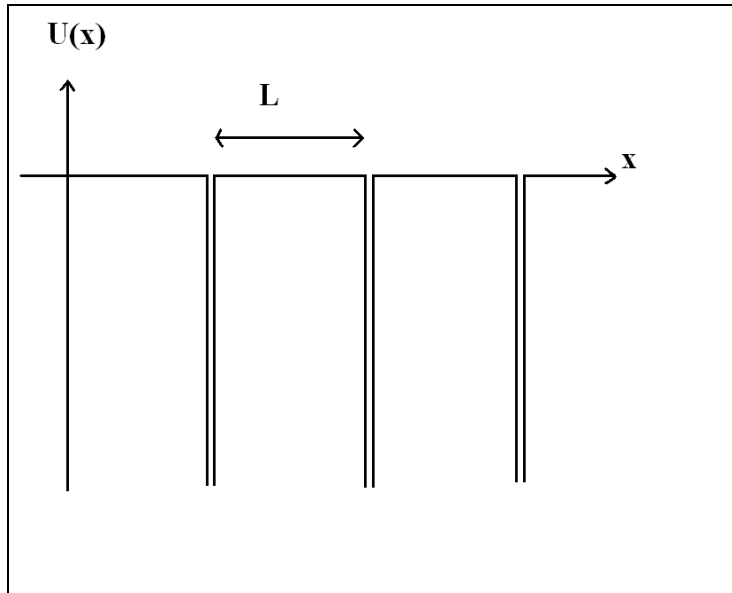
Figur 2. Ved å la gitteravstanden bli mindre og mindre, vil de diskrete energinivåene breseg ut i bånd

Merk området mellom de tillatte energinivåene, som vi skal se er bredden på dette båndgapet essensielt for å forstå et materiales elektriske egenskaper. Som vi har sett blir gjerne elektronet beskrevet ved en bølgefunksjon i kvantefysikken. Vi skal se at dette gir oss en mer matematisk beskrivelse av hvorfor energinivåsplittingen oppstår.

Kronig-Penny modellen

(-drøftingen baserer seg på referanse 5 og 6.)

Modellen tar utgangspunkt i at elektroner beveger seg i et krystallgitter og opplever derfor et periodisk potensial. For enkelhets skyld betrakter vi ett èn-dimensjonalt tilfelle.



Figur 3. Periodisk potensial for en èn-dimensjonal gitterstruktur med L som atomær avstand.

For å lage oss et analytisk, relativt enkelt løsbart problem lar vi potensialbrønnene i ved de positive kjernene være beskrevet ved en deltafunksjon:

$$U(x) = -U_0 L \sum_{N=-\infty}^{\infty} \delta(x - NL)$$

Vi ser altså på en krystall av uendelig utstrekning for å se bort fra effekter fra randa til krystallen.

Det at potensialet og elektronets potensielle energi er periodisk betyr at en translasjon langs x -aksen i en avstand L gir oss samme situasjon vi startet med, den kvadrerte bølgefunksjon som gir oss sannsynlighetsdistribusjonen må derfor ha samme periodisitet som gitteret selv;

$$|\Psi(x + L)|^2 = |\Psi|^2$$

Noe som innebærer at bølgefunksjonen, i henhold til Blochs teorem, bare kan variere med en reell konstant fasefaktor K ;

$$\Psi(x + L) = e^{iKL} \Psi(x)$$

Den generelle løsningen av schrødingers likning (1) i områdene med konstant null potensial

er: $\Psi = A \sin kx + B \cos kx$ med $k = \frac{\sqrt{2mE}}{\hbar}$. Ved å kreve kontinuitet i grenseområdene, $x=nL$,

og at diskontinuiteten for den deriverte i samme området er proporsjonal med styrken på

deltafunksjonen ($\Delta(\frac{\partial \Psi}{\partial x}) = -\frac{2mL}{\hbar^2} U_0 \Psi(nL)$) kan man med *mye* algebra eliminere vekk

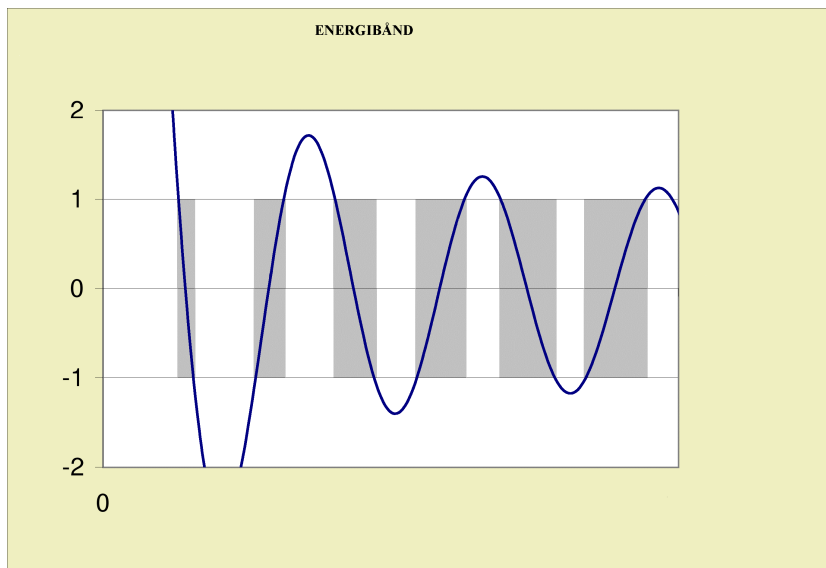
konstantene A og B . Konklusjonen for Kronig-Penny modellen er at den gir legger begrensninger på k for at vi skal få analytisk løsning. Løsning eksisterer hvis k oppfyller følgende relasjon:

$$\cos KL = \cos kL - \frac{mLU_0}{\hbar^2 k} \sin kL \quad (2)$$

Vi ser nå at siden $\cos KL$ bare tar verdier fra -1 til 1 vil det være områder hvor likningen ikke er oppfylt. Mens det er områder, bånd, av k -verdier der det er tillatte løsninger og dermed tillatte energier.

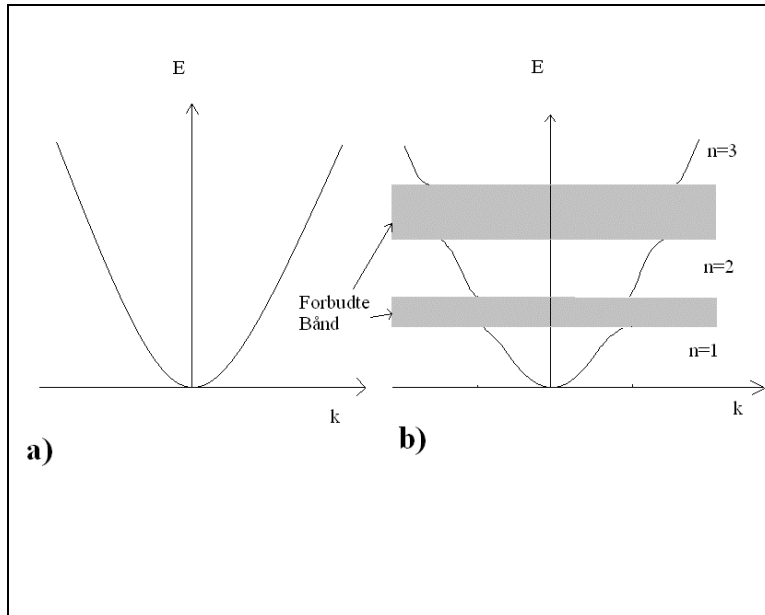
Vi tenker oss at x -aksen er "bøyd rundt i en ring" slik at vi kan sette randkravene, $\Psi(x + NL) = \Psi(x)$ vi får da videre fra Blochs teorem at $\Psi(x + NL) = \Psi(x) = e^{iKLN} \Psi(x)$. Fordi Ψ må ta entydige verdier må e^{iKLN} være røtter av 1.

Dermed; $e^{iKLN} = 1 \Rightarrow KL = \frac{2\pi n}{N}$ der n er heltall ($n = 0, \pm 1, \dots, (N-1)$), og N er antallet potensialbrønner. Da sistnevnte antallet typisk er meget stort, vil en heltallig variasjon i n kun gi en neglisjerbar differanse. Dermed har vi innenfor de tillatte båndene kontinuerlig tillatt energi i grensen $N \rightarrow \infty$.



Figur 4. $\cos kL + \frac{mLV_0}{\hbar^2 k} \sin kL$ plottet i mørkt, tillatte k -verdier befinner seg i intervallene farget i grått der $\cos KL$ ligger mellom -1 og 1.

Ved å plote det tilhørende E - k kurven med diskontinuiteten i henhold til de tillatte verdiene i figuren over kan dette se noe ut som i figuren under:



Figur 5. Energi som funksjon av k i a) for et fritt elektron og i b) for elektron som opplever et periodisk potensial

Som nevnt vil de kvantetilstandene som har lavest assosiert energi fylles opp først, deretter tvinges elektronene opp i høyere energinivå i samsvar med Paulis eksklusjonsprinsipp. Det høyeste energibåndet som ved det absolutte nullpunkt ($T=0$ K) er fullstendig fylt kalles valensbåndet. Et elektron som deltar i ledingsevnen til et fast stoff befinner seg i ledingsbåndet som er det ”neste” båndet over valensbåndet. Som vi har sett er de tillatte energiverdiene til et elektron under påvirkning av krefter begrenset til bånd. Men det holder ikke å vite at disse båndene eksisterer, vi må også vite om tilstandene er besatt av elektroner. Sannsynligheten for at en tilstand er besatt av et elektron er beskrevet av Fermi-Dirac distribusjonen;

$$f(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1} \quad (3)$$

som i korte trekk sier at sannsynligheten for at et elektron har nok energi til å ”hoppe” fra valensbåndet til ledningsbåndet (eg. sannsynligheten for at et elektron har en gitt verdi relativt til en referanseenergi; Fermienergien, E_F) avhenger av temperaturen. Vi ser at

$$f(E_F) = \frac{1}{2},$$

noe som betyr at sannsynligheten for at tilstand med fermienergien er besatt

med et elektron er $1/2$. Størrelsen på dette spranget bestemmer også i hvor stor grad elektroner vil befinne seg i valens- eller ledningsbåndet ved aktuelle temperaturer. Vi ser at sannsynligheten for å finne elektroner i energinivåer over fermienergien faller eksponentielt. Derfor er sannsynligheten for å finne et elektron i ledningsbåndet forsvinnende liten hvis båndgapet er stort.

Elektrisk ledningsevne

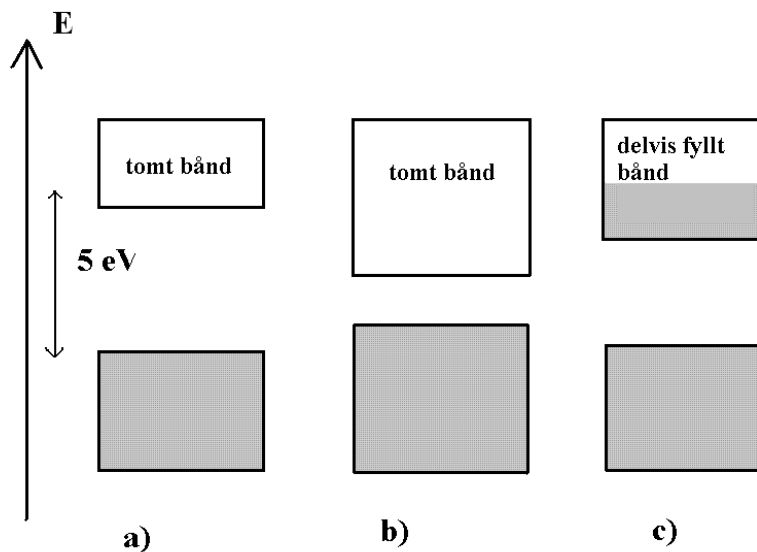
Det er vanlig å kategorisere faste stoffer etter deres evne til å lede strøm, det være seg elektronenes mulighet til å vandre i stoffet under påvirkning av et (svakt) elektrisk felt. Hvilke forhold må så være oppfylt for at et elektron skal delta i den elektriske ledningsevnen? Et energibånd som er tomt for elektroner kan selvsagt ikke bidra til ledningsevnen da det nettopp ikke har noen ladningsbærere. At et fullt bånd ikke heller bidrar til elektrisk ledning er derimot ikke like intuitivt klart. Hvis et elektron skulle bidra til ledning ville det drive mot det påtrykte feltet og således ha mottatt energi fra

feltet. Men dette kan ikke skje fordi alle tilstander i det tillatte båndet er besatt. Skal elektronet eksiteres til et energinivå der det kan delta i en elektrisk ledning må den tilførte energien tilsvare minst energigapet, E_g , det forbudte båndet representerer. Ved en gitt temperatur er det, som vi så ovenfor, en sannsynlighet for at et elektron får tilført tilstrekkelig energi fra gittervibrasjoner til å eksitere til ledningsbåndet.

Det som karakteriserer en *isolator* er nettopp at valensbåndet er fullt og båndgapet er stort, for stort til at den aktuelle termiske energien vanligvis er tilgjengelig. Et eksempel på isolator er diamant med et båndgap på 6eV. Et elektron kan derimot "presses" over energibarrieren ved å bli eksitert av et foton, høy temperatur eller å påtrykke et sterkt elektrisk felt. Sistnevnte er det vi kjenner som et dielektrisk sammenbrudd. Dette gjør seg derimot bare gjeldene kun for meget kraftige elektriske felt.

Akkurat som for en isolator har også *halvlederne* et fullt okkupert valensbånd ved $T=0$ K, og et tomt ledningsbånd. Forskjellen er her at båndgapet er av en mye mindre størrelsesorden. Slik at den termiske eksiteringen av elektroner bidrar i større grad til å skape mobile ladningsbærere. Vi skjønner at den elektriske ledningsevnen til halvledere er sterkt avhengig av temperatur. Hvis temperaturen blir hevet i tilstrekkelig grad vil vanligvis isolatorer oppnå samme egenskaper som en halvleder, og motsatt; ved lave temperaturer vil halvledere oppføre seg som isolatorer.

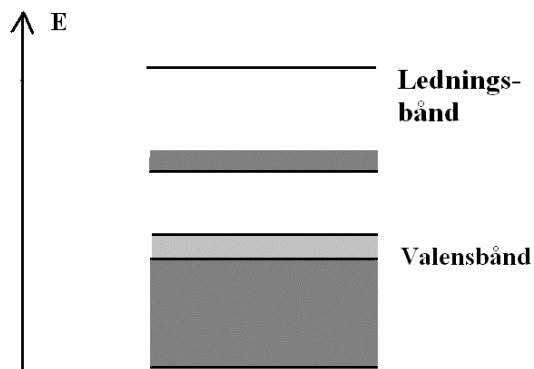
For metaller er bildet et annet, her er valensbåndet fullt, men også ledningsbåndet er delvis fylt ved $T=0$ K. Et elektron kan dermed lett eksiteres til et litt høyere energinivå som respons på et påtrykt felt.



Figur 6. skjematisk fremstilling av energibånd for a) isolator, b) halvleder og c) metall (leder). Alle ved $T=0$ K. Legg merke til det lille båndgapet for halvledere.

Halvleder

I en halvleder er atomene bundet sammen i middels sterke bindinger. At bindingene ikke er sterkere enn de er innebærer at det selv ved romtemperatur er nok termisk energi tilstede til at et betydelig antall bindinger brytes og elektroner blir frigjort til ledningsbåndet. Et eksempel på halvleder er silisium med et båndgap på 1,12eV. Når et elektron eksiteres til ledningsbåndet og forlater bindingen vil den etterlate en ledig posisjon. Et valenselektron fra et naboatom kan ta denne plassen og dermed flytte den ledige bindingen. På denne måten vil et slikt "hull" vandre gjennom en halvleder under påvirkning av et felt. Elektronene vandrer mot feltet slik at hullet vandrer med feltet og oppfører seg som en positivt ladet partikkel. I en ren (intrinsikk) halvleder vil hver eksitasjon medføre et hull i valensbåndet, vi har dermed to typer ladningsbærere og bidragsytere til elektrisk ledningsevne.



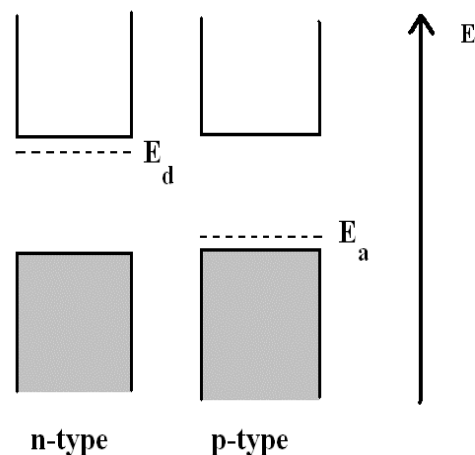
Figur 7. halvleder ved $T > 0$ K. Legg merke til øverste lag i valensbåndet i lyst er tomt for elektroner. De er eksitert til ledningsbåndet og har etterlatt seg hull.

Doping

En intrinsikk halvleder opptrer sjelden, selv små urenheter i krystallen i form av gitterfeil eller forurensningsatomer vil påvirke ledningsevnen i stor grad. Det å tilføre en halvleder en mengde forurensningsatomer kalles doping. For gruppe IV halvledere vil de viktigste dopingsatomene være type III og type V atomer. Vi ser først på type V doping som gir opphav til det vi kaller n-type doping

N-type

Ved å tilføre en type V atom til en type IV halvleder vil fire av de fem valenselektronene til type V atomet delta i kovalente bindinger med omkringliggende type IV atomer. Det resterende elektronet vil således være løst bundet til positive kjernen. La oss si vi tilførte fosfor til en silisiumkrystall, da kan det vises eksperimentelt at ioniseringsenergien til fosforatomet er 0,05eV. Det vil si at det kun kreves 0,05eV for å eksitere dette løst bundet elektronet til ledningsbåndet. Halvlederen har nå et overskudd av elektroner, og sies nå å være n-type (negativ-type) halvleder. Fosforatomet gav fra seg et elektron og kalles et donatoratom. På grunn av den lille energidifferansen nødvendig for å eksitere elektronet er det vanlig å si at alle donatoratomene er ionisert når man skal beregne ledningsevne. Ved en slik doping skal det kun meget små konsentrasjoner med forurensningsatomer til for å øke ledningsevnen betraktelig. Elektroner blir den desidert største bidragsyteren til ledning, faktisk så dominerende at vi kan neglisjere ledning pga hull.



Figur 8. Introduserte tillatte energinivåer som følge av doping.

P-type

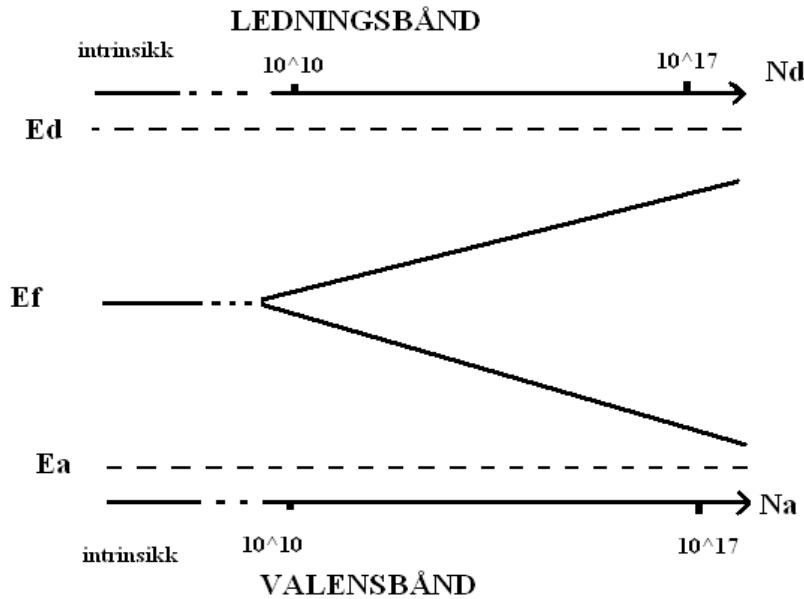
Man kan også tilføre et atom som har færre valenselektroner enn det som er nødvendig for å danne en fullstendig kovalent krystallstruktur. Ved å tilføre en type III atom f. eks. bor til silisium vil det kun dannes tre kovalente bindinger i stedet for fire. Denne ”mangelen” på et elektron som deltar i bindinger framstår som en ledig tilstand rett over toppen på valensbåndet. Det kreves således lite energi for at et valenselektron i et annet atom eksiteres til denne tilstanden. Et elektron som gjør dette vil etterlate seg et hull (ledig tilstand) i valensbåndet som kan delta i ledningsevnen. Vi ser at boratomet har mottatt et elektron og er således en akseptor. Silisiumet har en overvekt hull som bidrar til ledningsevnen og kalles en p-type (positiv-type) halvleder. For at det sistnevnte skal gjelde må antall akseptoratomer være større enn antall elektroner i ledningsbåndet. Det er viktig å legge merke til at mens en eksitasjon i en intrinsikk halvleder gav opphav til to ladningsbærere, ett elektron i ledningsbåndet og ett hull i valensbåndet, gir doping kun en ladningsbærer for hvert forurensningsatom.

Et annet viktig aspekt ved doping av halvledere er det såkalte np-produktet. Ved en gitt temperatur er produktet av konsentrasjonen av elektroner i ledningsbåndet og konsentrasjonen av hull en konstant.

$$n \cdot p = K(T)$$

Dvs. at ved å dope en halvleder til å f. eks. øke konsentrasjonen av elektroner med en faktor 10, vil konsentrasjonen av hull bli redusert med samme faktor. På denne måten vil ledningsegenskapene til en n-type halvleder være hovedsakelig av elektroner og kun neglisjerbart av hull. Det er derfor mulig å styre ledningsevnen til halvlederen i meget stor grad ved å tilføre forurensningsatomer i varierende mengde.

Doping medfører også at ferminivået forandrer posisjon. For en intrinsikk halvleder befinner dette nivået seg midt i båndgapet, men den skyves opp mot ledningsbåndet eller ned mot valensbåndet avhengig om doping er henholdsvis n-type eller p-type. Som vi har sett beskriver dette nivået den energien som har 0,5 sannsynlighet for å være besatt. Dersom dette nivået ligger nærme ledningsbåndet vil sannsynligheten for å finne elektroner i ledningsbåndet være stor. Vi vil da ha mange elektroner i ledningsbåndet, og som vi har sett desto færre hull, slik at vi den dominerende ladningsbæreren er elektroner. I motsatt fall ligger ferminivået nærmere valensbåndet slik at få elektroner eksiteres opp til ledningsbånd. Vi har i så fall ledning i dominerende grad av hull og altså en p-type halvleder.



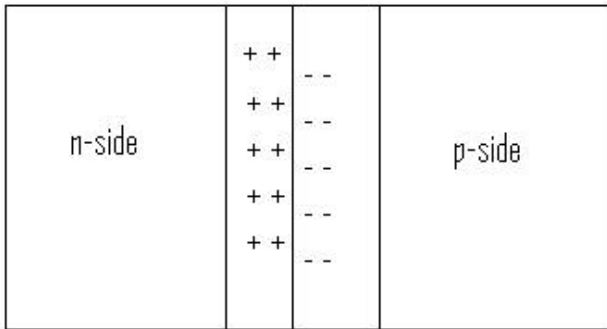
Figur 9. Skjematisk fremstilt posisjon av ferminivået for en intrinsikk halvleder (ladningskonsentrasjon $\approx 10^{10}$), og som funksjon av dopingskonsentrasjon. Konsentrasjoner oppgitt i antall pr. cm^3

P-n diode

(Drøftingen baserer seg på referanse 9)

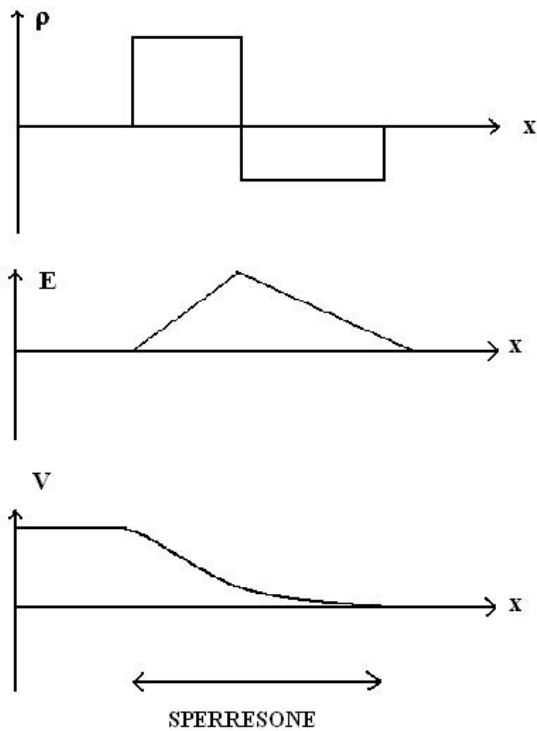
En p-n overgang består av en halvleder som er dopet med akseptorer på den ene siden og donorer på den andre. Dens hovedoppgave er å *likerette* strømmen i kretsen slik at den kun går en vei. Den har også andre anvendelser som bla fotodiode (lysfølsom og brukes i solceller), LED (diode som sender ut lys) og varaktor diode, (brukes som en variabel kapasitans i en krets)

Dersom vi tar utgangspunkt i en halvleder av silisium får vi en p-n diode ved å dope de to ulike sidene med f.eks. fosfor og bor. På hver av de to sidene har vi da ulike konsentrasjoner elektroner og hull som kan bevege seg fritt, og naturlig nok så vil det oppstå en diffusjonsprosess der elektronene og hullene ønsker å jevne ut konsentrasjonen slik at konsentrasjonen av elektroner og hull vil være lik overalt. Men denne prosessen kan ikke vare evig siden hullene vil etterlate seg negativt ladete akseptorkjerner og elektronene vil etterlate seg positivt ladete donorkjerner fordi kjernene ikke kan bevege seg. Det vil nå settes opp et elektrisk felt som hindrer den videre diffusjonsprosessen, og det vil innrette seg en likevekt. Denne likevekten baserer seg på at ladningene som diffunderes vil rekombinere slik at vi fortsatt får at $n \cdot p = n_i^2$ på begge sider av dioden. Litt enkelt kan vi si at akseptorene ønsker å beholde hullene, mens donorene vil beholde elektronene.



Figur 10. P-n kontakt i likevekt. Området med ladningene kalles for sperresone eller depleksjonsområde.

Hvis vi ser på figuren over kan vi begynne å gjette oss til hvordan det elektriske feltet og det innebygde potensialet ser ut. Det elektriske feltet vil være øke der det er positiv ladning og minke der det er negativ ladning, det er også klart at feltet må være null i de nøytrale områdene langt unna grenseflaten. For potensialet har vi at det vil ha motsatt fortegn av det elektriske feltet slik at det vil være konstant i de nøytrale områdene, mens det vil minke i grenseflaten.



Figur 11. Egenskaper ved en idealisert pn-kontakt. Øverst er ladningsfordelingen, i midten det elektriske feltet og nederst potensialet.

Om vi snur figuren og tenker på det motsatte tilfellet ser vi lett at vi får et negativt elektrisk felt og dermed en økning i potensialet.

Innebygd potensial i en p-n diode

For å finne det innebygde potensialet kan vi benytte to relativt enkle metoder. I den ene metoden må vi bruke Poissons ligning og integrere over ladningene to ganger. Dette forutsetter at vi må se på ladningene i sperresonen til å være konstante på begge sider av grenseflaten slik som figur 8 viser. Den andre metoden inneholder ingen integrering, men vi må til ha kjennskap til energibåndteori for en p-n overgang og konsentrasjon av ladningsbærere i ledningsbåndet og valensbåndet.

Ved å bruke Poissons ligning;

$$\frac{d^2V}{dx^2} = -\frac{\mathbf{r}(x)}{\mathbf{e}},$$

der $\mathbf{r}(x)$ er ladningsfordelingen får vi:

$$E(x) = -\frac{dV}{dx} = -\frac{d}{dx} \int dV = \frac{1}{\mathbf{e}} \int p(x) dx$$

$$E_n(x) = \frac{eN_D x}{\mathbf{e}} + K_2 = \frac{eN_D x}{\mathbf{e}} - \frac{eN_D x_n}{\mathbf{e}} = \frac{eN_D}{\mathbf{e}} (x - x_n)$$

$$E_p(x) = -\frac{eN_A x}{\mathbf{e}} + K_1 = -\frac{eN_A x}{\mathbf{e}} - \frac{eN_A x_p}{\mathbf{e}} = -\frac{eN_A}{\mathbf{e}} (x + x_p)$$

Der E_n og E_p er det elektriske feltet på henholdsvis n-siden og p-siden. Dette gir for det innebygde potensialet:

$$V_{ib} = -\int_{-x_p}^{x_n} E(x) dx = \frac{eN_D}{\mathbf{e}} \left(\frac{x_n^2}{2} \right) + \frac{eN_A}{\mathbf{e}} \left(\frac{-x_p^2}{2} \right)$$

Når vi vet at ladningene på hver side av grenseflaten er like store, dvs.: $eN_A x_p = eN_D x_n$ får vi:

$$V_{ib} = \frac{eN_D^2 x_n^2}{2\mathbf{e}} \left(\frac{N_A + N_D}{N_A N_D} \right) = \frac{eN_A^2 x_p^2}{2\mathbf{e}} \left(\frac{N_A + N_D}{N_A N_D} \right)$$

Den andre metoden forutsetter at vi vet at konsentrasjonen av hull i valensbåndet er

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right) \text{ på p-siden og konsentrasjonen av elektroner i ledningsbåndet på n-siden}$$

er $n = n_i \exp\left(\frac{E_F - E_i}{kT}\right)$. Fra energibåndteori om p-n overgangen skal vi snart vise at

$|V_{ib}| = |V_p| + |V_n|$ der $qV_p = -(E_i - E_F)$ og $qV_n = -(E_i - E_F)$. Ved å substituere for $E_i - E_F$ og

sette $q=e$ får vi: $|V_{ib}| = \frac{kT}{e} \ln \frac{N_A N_D}{n_i^2}$

Energibånddiagram for en p-n overgang

Når vi skal tegne opp et energibånddiagram for en p-n overgang er det viktig å huske på at det kjemiske potensialet (fermienergien) er konstant i hele dioden. Som vi har sett i figur 9. avhenger fermienergien av dopingskonsentrasjonen. Vi vil da få forskjellige energinivåer for valensbåndet og ledningsbåndet på hver side av sperreområdet.

Hvis vi ser på hva som skjer matematisk er det viktig å huske på at det har innstilt seg en dynamisk likevekt, og at prosessene ikke har stoppet opp. I tillegg til diffusjonsstrømmen vil

vi ha en driftsstrøm. Denne driftsstrømmen oppstår fordi det termisk genereres hull nær kanten av sperresonen på n-siden og elektroner på p-siden. På grunn av det elektriske feltet vil disse ladningene bli transportert til den andre siden, og vi får altså to nye strømmer i motsatt retning av de to diffusjonsstrømmene. Altså vil vi ha at summen av diffusjonsstrømmen og driftsstrømmen vil være null. Hvis vi setter opp de aktuelle ligningene for de positive ladningsbærerne får vi:

$$J_p = J_p(\text{Drift}) + J_p(\text{diffusjon}) = 0$$

Her er $J_p(\text{diffusjon})$ knyttet til strømmen fra diffusjonsprosessen mellom hull og elektroner, mens $J_p(\text{drift})$ er knyttet til den termiske genereringen av elektroner og hull nær kantene av sperresonene. For å løse denne ligningen må vi bruke uttrykk for driftsstrømmen og diffusjonsstrømmen, og en del relasjoner mellom disse.

$$J_p = q\mathbf{m}_p p E_x - qD_p \frac{dp}{dx} = 0$$

$$J_p = \mathbf{m}_p \left(p \frac{dE_i}{dx} - kT \frac{dp}{dx} \right) = 0$$

Her er \mathbf{m}_p hullmobiliteten, og D_p diffusjonskoeffisienten.

Ved å substituere for dp/dx får vi når:

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right)$$

$$\frac{dp}{dx} = \frac{p}{kT} \left(\frac{dE_i}{dx} - \frac{dE_F}{dx} \right)$$

Som da gir oss:

$$J_p = \mathbf{m}_p \left(p \frac{dE_i}{dx} - kT \cdot \frac{p \left(\frac{dE_i}{dx} - \frac{dE_F}{dx} \right)}{kT} \right) = \mathbf{m}_p p \frac{dE_F}{dx} = 0$$

Siden vi nødvendigvis vil få de samme ligningene for de negative ladningsbærerne, vil vi uavhengig av fortegn få at $\frac{dE_F}{dx} = 0$, nemlig at den kjemiske energien må være konstant gjennom hele dioden.

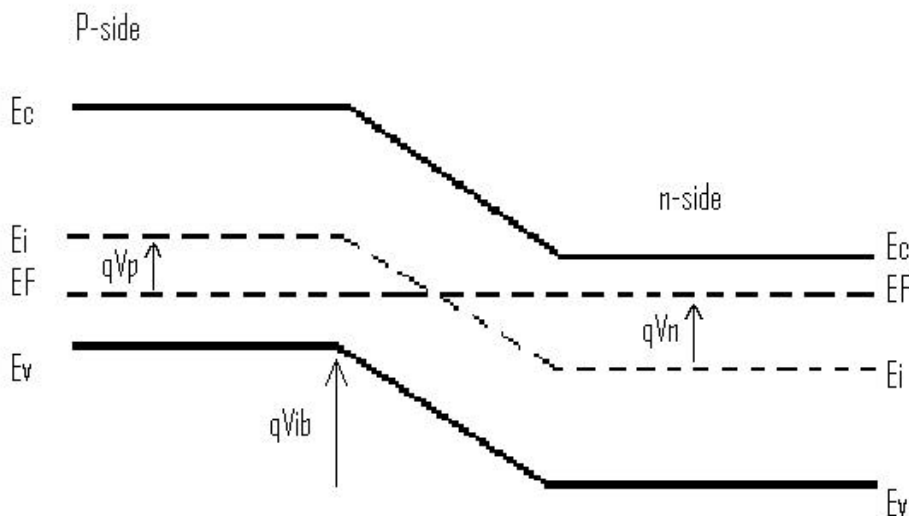
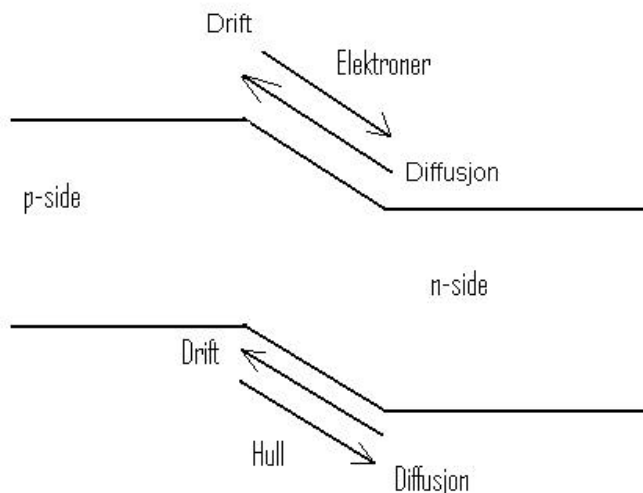


Fig 12. Energibåndfigur med potensialforskjeller. Vi ser E_F er konstant gjennom hele dioden, og at p-siden og n-siden ligger på forskjellig energetisk nivå

Bias over en p-n diode

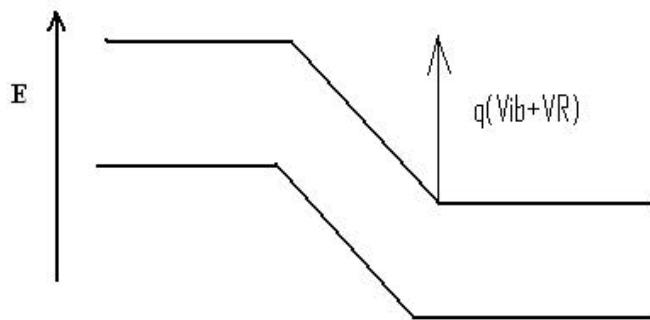
Når vi setter spenning over en diode kaller vi den for bias. Den viktigste egenskapen til en p-n diode er som sagt å være en likeretter i en krets. Dette ser vi ganske lett ved å studere ett *strøm-spenning* diagram. Der ser vi at strømmen går lett i den ene retningen, og at det knapt går noen strøm når vi snur biasen.

For å forklare hva som skjer med strømmen når vi anvender en bias på dioden må vi se på hva som skjer med de to diffusjonsstrømmene og de to driftsstrømmene. Siden driftsstrømmene kommer fra ladningsbærerne som genereres termisk nær kanten av sperresonen vil det være naturlig å anta at denne prosessen vil være den samme som når vi ikke anvendte bias på dioden. Dette vil vi også anta ved å se på et energidiagram, fordi driftsstrømmene ikke har noen potensialbarriere å passere vil ikke en endring av denne barrieren ha noen å si for driftsstrømmene.



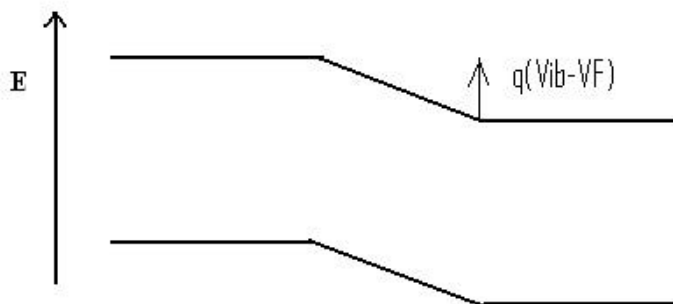
Figur 13. energibåndmodell med de fire strømmene indikert

En interessant egenskap oppstår når vi setter på en bakover bias. Den påtrykte spenningen vil da sette opp ett elektrisk felt som har samme retning som feltet inne i dioden. Siden det er dette feltet som hindrer diffusjonsstrømmen, vil vi intuitivt anta at dette vil redusere strømmen gjennom dioden når det vil innstille seg en ny likevekt. Dersom vi ser på hva som skjer med et energibånddiagram når setter på en bakover bias, ser vi at potensialforskjellen mellom ledningsbåndene øker. Dette gjør at det vil bli vanskeligere for ladningsbærerne å bevege seg gjennom grenseflaten, og det vil da gå mindre strøm gjennom dioden.



Figur 14. Pn-kontakt med påtrykt bakover-bias. VR representerer den påtrykte spenningen.

Når vi setter på en forover bias vil den påtrykte spenningen sette opp et elektrisk felt som er motsatt rettet av det som allerede eksisterer i dioden. Dette vil da bety at vi svekker det som hindrer diffusjonsstrømmen, og vi vil da anta at strømmen gjennom en diode vil øke med økende påtrykt spenning. Når vi ser på ett energibånddiagram som viser hva som skjer med en forover bias, vil vi få at potensialforskjellen ladningsbærerne må passere i grenseflaten vil bli mindre, og dermed mulighet for at flere ladningsbærere vil bevege seg gjennom dioden.



Figur 15. Pn-kontakt med påtrykt forover-bias

Matematisk beskrivelse av en ideell diode

For å avgjøre formen på *strøm-spenning* diagrammet må vi vite litt mer om hvordan diffusjonsstrømmene oppfører seg på hver side av sperresonen. Diffusjonsprosessen er et forsøk på å få samme konsentrasjonen av henholdsvis hull og elektroner på hver side av grenseflaten. Om vi ser på likevektskonsentrasjonene av hull og elektroner på hver side av grenseflaten får vi:

n_{n0} konsentrasjon av elektroner på n-siden

n_{p0} konsentrasjon av elektroner på p-siden

p_{n0} konsentrasjon av hull på n-siden

p_{p0} konsentrasjon av hull på p-siden

Siden vi vet at $n \cdot p = n_i^2$ på hver side av grenseflaten, får vi det innebygde potensialet:

$$V_{ib} = \frac{kT}{q} \ln \frac{n_{n0}}{n_{p0}} = \frac{kT}{q} \ln \frac{p_{n0}}{p_{p0}}$$

Som kan skrives om slik at vi får:

$$n_{n0} = n_{p0} \exp\left(\frac{qV_{ib}}{kT}\right) \text{ og } p_{p0} = p_{n0} \exp\left(\frac{qV_{ib}}{kT}\right)$$

Hvis vi ser på hva som skjer om vi anvender en bias V på dioden, og antar at forholdene mellom elektronkonsentrasjonene er de samme som under likevekt får vi:

$$n_n = n_p \exp\left(q \frac{V_{ib} - V}{kT}\right)$$

Hvis vi i tillegg kun ser på når en liten bias blir anvendt, har vi at: $n_n \approx n_{n0}$ som gir:

$$n_n = n_p \exp\left(q \frac{V_{ib} - V}{kT}\right) = n_{n0} = n_{p0} \exp\left(\frac{qV_{ib}}{kT}\right)$$

$$n_p - n_{p0} = n_{p0} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]$$

Nå må vi i tillegg huske på at ladningene som diffunderer gjennom grenseflaten vil rekombinere med de motsatte ladningene på den andre siden. Det viser seg at for elektroner har vi at konsentrasjonen et stykke l_p fra kanten på sperresonen inn i p-siden er:

$$dn(l_p) = dn(0) \exp\left(-\frac{l_p}{L_n}\right)$$

Der $dn = n_p - n_{p0}$, $dn(0) = n_{p0} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]$ og L_n er diffusjonslengden til elektronet. Nå

bruker vi igjen at $J_n = qD_n \frac{dn}{dx}$ for diffusjonsstrømmen. Ved å se på strømtettheten ved kanten på sperresonen får vi:

$$J_n(l_p = 0) = -\frac{qD_n n_{p0}}{L_n} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]$$

En tilsvarende analyse for hullene gir:

$$J_p(l_n = 0) = -\frac{qD_p p_{n0}}{L_p} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]$$

Dette gir en total strømtetthet:

$$J = J_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]$$

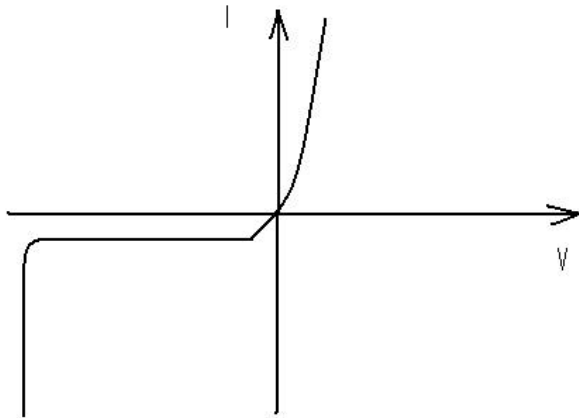
Og for strømmen får vi da den ideelle diode ligningen:

$$I = I_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]$$

Vi ser at dersom vi anvender en så stor negativ bias at eksponentialleddet går mot null vil vi få en konstant strøm på $-I_0$. I følge det vi tidligere har skrevet må dette da representere driftsstrømmen, noe som igjen betyr at eksponentialleddet representerer diffusjonsstrømmen. Altså:

$$I_{diff} = I_0 \exp\left(\frac{qV}{kT}\right)$$

$$I_{drift} = -I_0$$



Figur 16. Strøm-spenning forhold for ideell diode. Kun foroverbias gir stor strøm, men ved stor bakoverbias kan et dielektrisk sammenbrudd inntreffe. Dette er indikert med stor negativ strøm ved stor negativ bias på figuren.

Som indikert i figur 16 vil dioden få et sammenbrudd ved stor bakoverbias. Dette kommer av at den elektriske feltstyrken i sperresonen kommer over en viss verdi (ca 10^5 V/cm) og vi får et dielektrisk sammenbrudd. Men vi vil også få problemer med stor foroverbias. Dersom vi anvender en foroverbias som er større enn det innebygde potensialet, vil vi også ødelegge dioden. Dette illustreres enkelt ved å tenke på hvordan et energibånddiagram ville sett ut om vi anvender en foroverbias sterkere enn det innebygde potensialet. Det er ganske innlysende at en slik situasjon må føre til en ødeleggelse av dioden siden potensialforskjellene gjør at strømmene ønsker å skifte retning, noe som er i kontrast med egenskapene til n og p siden. Eventuelt kan vi tenke på sperresonen som i en slik situasjon da vil få negativ utstrekning.

Grenseflaten som en kapasitans

Så langt har vi avdekket at en diode har en sperresone som inneholder en viss mengde ladning, og at ladningen i sperresonen varierer med spenningen over diode. Hvis vi sammenligner med en vanlig kapasitans ser vi helt klare likhetstrekk, bare at vi nå vil ha en kapasitans som varierer med spenningen. En diode som blir laget for å utnytte dette kalles for en varaktor.

Det viser seg at kapasitansen til en slik diode er gitt ved:

$$C = \frac{e_{si}}{W}$$

Som er nøyaktig det samme som for en plate kondensator. Hvis vi ser på en diode der den ene siden har en høy dopingkonsentrasjon i forhold til den andre, en **abrupt one-side junction**, har vi at sperresonen kun ligger på den ene siden av grenseflaten. Der har vi at:

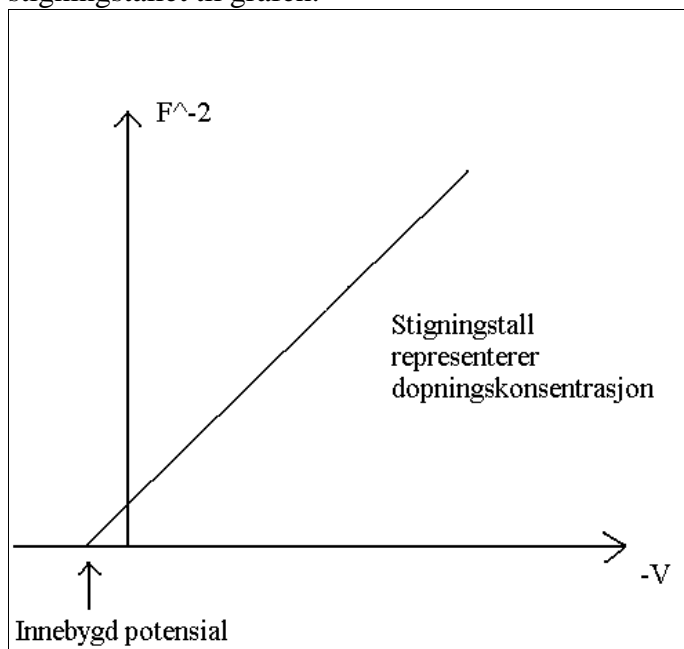
$$W = \left(\frac{2e[V_{ib} - V]}{qN} \right)^{\frac{1}{2}}$$

Der V er anvendt bias og N er dopingkonsentrasjonen til siden med lav dopingkonsentrasjon.

Hvis vi nå utfører et forsøk der vi plotter $\frac{1}{C^2}$ som en funksjon av V , kan vi finne ut både det innebygde potensialet og dopingkonsentrasjonen.

$$\frac{1}{C^2} = \frac{2(V_{ib} - V)}{qeN}$$

Ut fra formelen over ser vi at V_{ib} er gitt fra skjæringspunktet med førsteaksen og N fra stigningstallet til grafen.



Figur 17. kapasitans-spenningsdiagram for en varaktor diode

LED-diode (light emitting diode)

I det vi anvender en forover bias på en p-n diode, vet vi at diffusjonsstrømmen øker. Et resultat av dette er at også rekombinasjonen av elektroner og hull øker. Under denne rekombinasjonen vil det frigjøres energi ved at elektroner faller tilbake til valensbåndet og fyller opp hull. Fra Bohrs teori om atommodeller og tillatte energinivåer vet vi at når et elektron faller til en lavere energitilstand vil det sende ut et foton med en gitt bølgelengde. Derfor vil det sendes ut lys fra en p-n diode når hull og elektroner rekombinerer.

Ved å bruke ulike grunnstoffer og variere mengden av de kan vi fikserer størrelsen på båndgapet slik at det vil emitteres lys i ønskede bølgelengder, også infrarøde og ultrafiolette spekteret. Det er derimot umulig å lage en diode som sender ut hvitt lys siden hvitt lys er en samling av alle bølgelengdene i det synlige spekteret. Også enkelte deler av det synlige spekteret har vist seg vanskelig å lage klart lys av. Dette kommer av at det er kvantefysisk større sannsynlig med eksitasjoner over et direkte båndgap, dvs når valensbåndet og ledningsbåndet ikke overlapper hverandre. Spesielt for blått har dette vært et problem, det viser seg vanskelig å lage et direkte båndgap med tilsvarende energi som blått lys. Vanligvis baserer LED-diodene som lyser blått seg på silisium som bare har indirekte båndgap, og som dermed ikke er den mest energetisk gunstige løsningen.

Selv om vi har et indirekte båndgap fins det måter å effektivisere lysemittasjonen på. Dette gjøres best ved å innføre en isoelektrisk kjerne, dvs. å erstatte en dopingkjerne med en annen kjerne fra samme gruppe. En annen metode som igjen er litt mindre effektiv er å innføre andre urenheter i halvlederen. Om det ikke finnes kjente isoelektriske kjerner eller urenheter som gjør lysdioden mer energieffektiv for de ønskede lysfargene, må vi bruke en halvleder med indirekte båndgap.

Et eksempel på en LED-diode er et AlGaAs/GaAs system der vi har ulike materialer på hver side av p-n overgangen, en såkalt heteroovergang. Ved å variere konsentrasjonen av aluminium kan vi bestemme om vi ønsker et indirekte eller direkte båndgap. Så lenge vi ønsker et direkte båndgap i delen som skal emittere lyset må aluminiumskonsentrasjonen være på under 44%, noe som gjør at vi vil få lys med bølgelengder fra 880nm-663nm, altså i det røde og infrarøde spekteret. Som vi ser av *figur 18* vil vi ha en n-side med 70% aluminium og to ulike typer p-sider, der delen med 40% Al vil fungere som lyskilde. Denne n-siden har et større båndgap enn p-siden noe som gjør den til en effektiv injektor av elektroner, mens det samtidig er større sjans for rekombinasjoner i et mindre båndgap. I tillegg vil også fotonene som produseres lettere kunne passere n-siden pga. det store båndgapet. Alt i alt gjør dette at et slikt system vil levere klart rødt lys veldig energieffektivt.

n-side	AlGaAs	70%Al
p-side	AlGaAs	40%Al
p-side	GaAs	

Figur 18. hetero pn-overgang

I tillegg til dioder som lyser rødt og blått er også gult, grønt og oransje vanlige farger på lysdioder. Alle disse diodene har vanligvis indirekte båndgap, men bruker en isoelektrisk kjerne for å gjøre lyset klarere og mer energieffektivt.

Fotodetektorer og solceller

Når vi så på en LED-diode brukte vi at når elektroner faller ned til en lavere energitilstand vil det sende ut et foton. Fra kvantefysikken vet vi at det motsatte også er gyldig, at et foton kan eksitere et elektron gitt at fotonet har nok energi. Det er dette prinsippet som vi bruker i en fotodetektor. I motsetning til en LED-diode trenger en fotodetektor kun riktig båndgap og er ikke avhengig av om det er direkte eller indirekte.

Mekanismen vi bruker *for å detektere* fotoner er at et elektron som eksiteres nær eller i sperresonen vil gi opphav til en strøm som følger av det elektriske feltet i området. Dersom elektronet blir eksitert i de nøytrale områdene av en p-n overgang vil de rekombinere med hull før de når sperresonen, og kan være opphav til en strøm. Derfor vil en god fotodetektor ha en stor sperresone, slik at flest mulig av de eksiterte elektronene vil være opphav til en strøm gjennom dioden. Fra teorien om p-n dioden vet vi at å sende på en stor bakover bias vil vi kunne maksimere sperresonen.

Fra kvantefysikken vet vi at om fotonet ikke har nok energi, vil det ikke være i stand til å eksitere elektroner, men også fotoner med for høy energi vil vi få problemer med å *detektere*. Dette har sammenheng med at de mest energirike fotonene er de som absorberes først i en

halvleder. For å øke effekten til detektoren vil det derfor være naturlig å sørge for en tynt lag på toppen, slik at minst mulig lys absorberes før de kommer til sperresonen.

Dersom vi lar være å anvende en bakover-bias på en fotodetektor og istedenfor plasserer den i en krets der den vil fungere som en spenningskilde vil vi ha en solcelle. Den spenningen en slik diode vil bidra med kalles for fotovoltaisk effekt og den kan vi finne ved å bruke den ideelle diode ligningen.

$$I = I_0(\exp(\frac{qV}{kT}) - 1)$$

Dersom vi løser denne med hensyn på V får vi:

$$V = \frac{kT}{q} \ln(1 + \frac{I}{I_0})$$

Størrelsen av V vet vi må være mindre enn det innebygde potensialet, som vi igjen vet er mindre enn potensialet i båndgapet. Vi kan dermed avgjøre sånn noenlunde hvor mye spenning en slik solcelle på bakgrunn av egenskapene til materialet vi velger å bruke. En standard silikon solcelle vil produsere ca 1 volt og 20mA. Det er derfor vanlig å koble sammen flere celler i serie og parallell for å sørge for en tilfredsstillende energiproduksjon.

Et av de største problemene med en solcelle er at de ønsker å utnytte sollyset som består av et stort spekter frekvenser og bølgelengder, og som vi har sett er det vanskelig å utnytte et slikt stort spekter. Dette gjør at man får en teoretisk virkningsgrad som ligger langt unna 100 %. Det er vanskelig å angi nøyaktig hva som er den teoretiske virkningsgraden, siden det forskes mye på solceller og man hele tiden kan finne nye og bedre teoretisk metoder. Men de høyeste eksperimentelle verdiene som er oppnådd er på rundt 40 %. Et annet problem er at for å få transportert strøm ut av solcellene. For å gjøre dette trenger vi en ohmsk metallkontakt på begge sider av solcella. En slik kontakt vil hindre sollyset i nå ned til selve dioden. Et kompromiss har vært å lage tynne metallfingre på tvers av solcella og knytte disse til en større metallbit på kanten av cella.

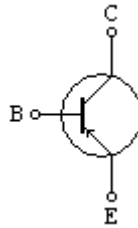
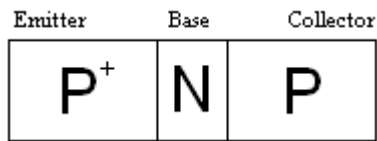
Transistorer

En transistor er en elektrisk komponent som ble utviklet ved Bells Laboratorier i USA. Den første transistoren ble laget i 1948, og tre personer som har fått æren for denne revolusjonerende oppfinnelsen er John Bardeen, Walter Brattain, og William Shockley. Ordet transistor er satt sammen av de engelske ordene transfer og resistor. Direkte oversatt vil ”transfer resistor” bety overførbar elektrisk motstand. Det finnes flere typer transistorer som for eksempel bipolare transistorer (BJT) og felteffektransistorer (FET).

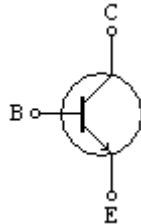
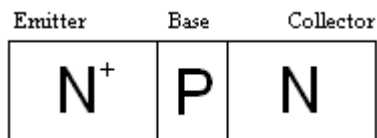
Bipolare transistorer

Bipolare transistorer består av to pn-overganger hvor de to overgangene har en felles p eller n region. Denne felles regionen ligger mellom to regioner med motsatt type doping slik at vi enten har en pnp-transistor eller en npn-transistor.

Den midterste regionen, som er vesentlig smalere enn de andre, kalles for transistorens base og de to andre regionene kalles emitter og kollektor. I de aller fleste tilfellene har emitterregionen betydelig større dopingkonsentrasjon enn basen, som igjen har større dopingkonsentrasjon enn kollektorregionen. I elektriske kretser har pnp-transistorer og npn-transistorer hvert sitt skjematisk symbol.



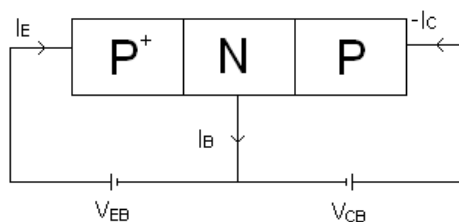
Figur 19a): Pnp-transistor og dens symbol i elektriske kretser.



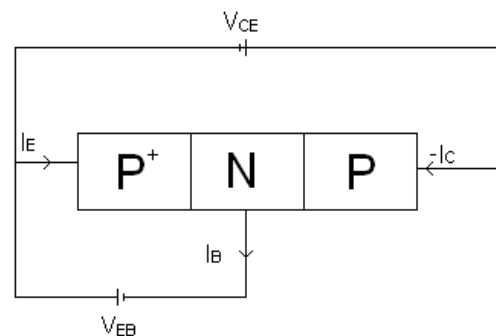
Figur 19b): Npn-transistor og dens symbol i elektriske kretser.

Konfigurasjoner

Når en transistor koples til to forskjellige spenningskilder kan dette gjøres på forskjellige måter og vi skiller blant annet mellom felles base konfigurasjon og felles emitter konfigurasjon. I felles base konfigurasjon koples en spenningskilde over emitter og base, mens en annen spenningskilde koples over base og kollektor. Det eneste felles kopleingspunkt er altså basen, derav navnet felles base konfigurasjon. I felles emitter konfigurasjon koples da spenningskilder over emitter og base, og over emitter og kollektor.



Figur 20a): Felles base konfigurasjon.



Figur 20b): Felles emitter konfigurasjon.

Strøm i en transistor

Vi ser på felles base konfigurasjon av en pnp-transistor. Vi kaller spenningen over emitter og base for V_{EB} der vi for positiv spenning har høyest potensial på emitter. Spenningen over base og kollektor kaller vi V_{CB} der vi for positiv spenning har høyest potensial på kollektor.

Strøm i emitteren er sammensatt av strøm av hull fra emitter til base og strøm av elektroner fra base til emitter: $I_E = I_{Ep} + I_{En}$

Strøm i kollektor er sammensatt av strøm av hull fra base til kollektor og strøm av elektroner fra kollektor til base: $I_C = I_{Cp} + I_{Cn}$

Kirchoffs knutepunktregel gir oss strøm i basen: $I_B = I_E - I_C$

Vi definerer felles base strømfaktor: $\mathbf{a}_0 \equiv \frac{I_{Cp}}{I_E}$

Kollektorstrømmen blir da: $I_C = \mathbf{a}_0 I_E + I_{Cn} = \mathbf{a}_0 I_E + I_{CB_0}$

I_{CE0} kalles kollektor-base lekkasjestrøm med åpen emitter. Det er altså den strømmen som går mellom kollektor og base når $I_E = 0$.

Strømfaktoren kan også skrives $\mathbf{a}_0 = \frac{I_{Cp}}{I_{Ep} + I_{En}} = \frac{I_{Ep}}{I_{Ep} + I_{En}} \cdot \frac{I_{Cp}}{I_{Ep}} = \mathbf{g} \cdot \mathbf{a}_T$

der vi definerer to nye størrelser: $\mathbf{g} \equiv \frac{I_{Ep}}{I_E}$ og $\mathbf{a}_T \equiv \frac{I_{Cp}}{I_{Ep}}$ som er henholdsvis emittereffektivitet

og basetransportfaktor. Emittereffektiviteten forteller oss hvor stor andel av emitterstrømmen som skyldes vandring av hull, mens basetransportfaktoren forteller hvor stor andel av hull som vandrer fra emitter til base, som også vandrer videre til kollektor.

Dersom vi ser på felles emitter konfigurasjon av den samme transistoren vil strøm i emitteren i følge Kirchoffs knutepunktregel bli: $I_E = I_B + I_C$

Uttrykket for Kollektorstrømmen blir da: $I_C = \mathbf{a}_0 I_E + I_{CB_0} = \mathbf{a}_0 (I_B + I_C) + I_{CB_0}$

som gir: $I_C = \frac{\mathbf{a}_0}{1 - \mathbf{a}_0} I_B + \frac{I_{CB_0}}{1 - \mathbf{a}_0} = \mathbf{b}_0 I_B + I_{CE0}$

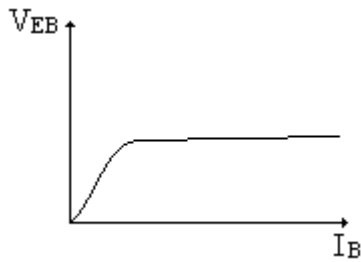
der β_0 er en størrelse kalt felles emitter strømforsterkning og som er definert ved:

$$\mathbf{b}_0 \equiv \frac{dI_C}{dI_B} = \frac{\mathbf{a}_0}{1 - \mathbf{a}_0}$$

I_{CE0} kalles kollektor-emitter lekkasjestrøm med åpen base. Det vil si den strømmen som går mellom kollektor og emitter når $I_B = 0$

Ut i fra felles base konfigurasjonen kan vi konkludere med at α_0 er mindre enn, i "beste fall" nesten lik 1, siden I_C må være mindre enn I_E . (Fordi noen av hullene vil rekombinere å bidra til I_B) Vi ser at når α_0 går mot 1 så vil β_0 være mye større enn 1 noe som vi skal se er avgjørende for transistorens virkning. β_0 er typisk i størrelsesorden 10-100.

Med felles emitter konfigurasjon vil vi ha en sammenheng mellom strøm inn på basen og spenningen over emitter og base som skissert nedenfor.

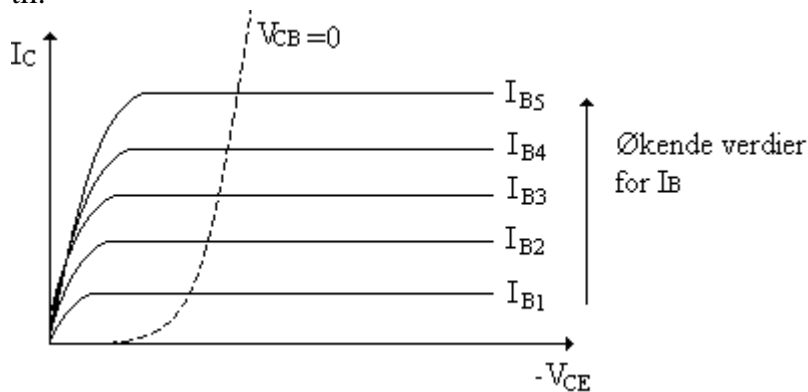


Figur 21: Sammenheng mellom basestrøm og emitter-base spenning.

Sammenheng mellom kollektorstrøm og spenningen over kollektor og emitter kan vi tenke oss ved å sammenligne base-kollektorkovergangen med en ideell diode. I det som vi senere vil kalle det aktive område for transistoren, vil vi ha bakover biasert denne pn-overgangen og strøm ut i fra kollektoren. Når vi skisserer strøm-spenning kurven tar vi derfor og speiler strøm-spenning grafen fra den ideelle dioden om begge aksene.

Vi ønsker å skissere hvordan basestrømmen påvirker Kollektorstrømmen og tegner derfor flere kurver i samme koordinatsystem for ulike verdier av I_B .

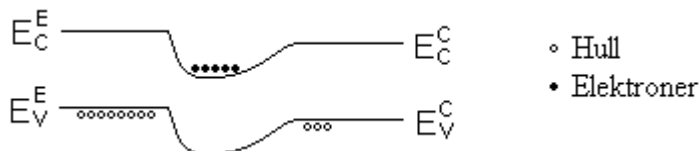
I figuren nedenfor har vi i tillegg tegnet inn grafen for når kollektor-emitter spenningen er like stor som emitter-base spenningen (stiplet linje). Betydningen av dette skal vi komme tilbake til.



Figur 22: Sammenheng mellom kollektorstrøm og emitter-kollektor spenning For ulike verdier av I_B .

Energibånd for en transistor

Dersom vi betrakter en pnp-transistor som ikke er koplet inn i noen krets, slik at alle tre tilkoplingspunkter ligger på samme potensial, kan vi se på de to pn-overgangene hver for seg og bruke samme argumenter som for den enkle pn-overgangen til å tegne energibåndene for transistoren. Vi må nå ta hensyn til det innebygde potensialet i både emitter-base overgangen og base-kollektor overgangen.



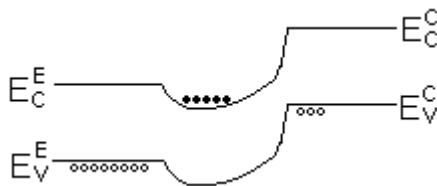
Figur 23: Pnp- transistor i likevekt utenfor krets.

Operasjonsmodi

Vi fortsetter å se på pnp-transistoren, men denne gangen koplet til to spenningskilder i en felles base konfigurasjon.

Vi kan veksle mellom å ha V_{EB} og V_{CB} positiv og negativ slik at vi har fire ulike kombinasjonsmuligheter som gir fire ulike operasjonsmodi for transistoren. Ved å bruke det vi vet om de enkelte pn-overgangene kan vi se hvordan de forskjellige kombinasjonene av fortegn på spenningene vil påvirke energibåndene til transistoren og dermed vil vi kunne si noe om hva som karakteriserer de fire operasjonsmodiene.

- Med positiv V_{EB} og negativ V_{CB} vil emitter/base-overgangen være forover biasert, mens kollektor/base-overgangen vil være bakover biasert. Dette betyr at vi får en lavere energibarriere for hull mellom emitter og base, slik at hull vil begynne å vandre fra emitter til base. I basen vil vi nå kunne få rekombinasjon av elektroner og hull, men med tilstrekkelig smal base vil mange hull kunne passere basen og ”falle ned” i kollektoren. Det vil altså begynne å gå en relativt stor strøm fra emitter til kollektor. Vi sier nå at transistoren er i aktivt modus.



Figur 24: Energibånd for aktiv transistor.

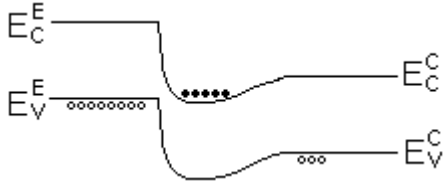
- Med både V_{EB} og V_{CB} positiv vil begge pn-overgangene i transistoren være forover biasert. Det vil si at vi får lavere energibarrierer både mellom emitter og base og mellom kollektor og base. En liten økning i spenningene i forhold til likevekt vil føre til store strømmer, så transistoren oppfører seg som et knutepunkt og vi sier at vi har metning.



Figur25: Energibånd for transistor ved metning.

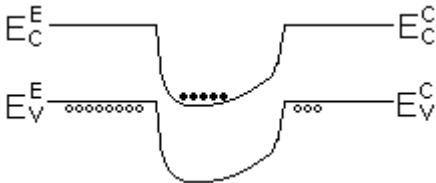
- Med negativ V_{EB} og positiv V_{CB} vil emitter/base-overgangen være bakover biasert, mens kollektor/base-overgangen vil være framover biasert. Vi ser at situasjonen er veldig lik situasjonen for en aktiv transistor der vi nå får en lav energibarriere for hull mellom kollektor og base, slik at hull kan begynne å vandre. Forskjellen ligger her i dopingkonsentrasjonen for de ulike lagene. Kollektoren har en vesentlig mindre

konsentrasjon av doping enn emitteren slik at strømmen gjennom transistoren blir betydelig lavere. Vi kan betrakte situasjonen som om vi har koplet en aktiv transistor ”feil” vei, og sier at vi har invertert modus.



Figur 26: Energibånd for invertert transistor.

- Med både V_{EB} og V_{CB} negativ vil begge pn-overgangene i transistoren være bakover biasert, og vi får høye energibarrierer både mellom emitter og base, og mellom kollektor og base. Resultatet er små strømmer og vi sier at transistoren er i cut-off.



Figur 27: Energibånd for transistor i cut-off.

På samme måte vil vi kunne finne ut at for en npn-transistor får vi:

- Med V_{EB} negativ, V_{CB} positiv: aktiv transistor
- Med V_{EB} negativ, V_{CB} negativ: metning
- Med V_{EB} positiv, V_{CB} negativ: invertert transistor
- Med V_{EB} positiv, V_{CB} positiv: cut-off

Det samme vil også gjelde om vi ser på en felles emitter konfigurasjon, men da må vi huske at spenningsforskjellene over emitter og base og over kollektor og base, henholdsvis V_{EB}^* og V_{CB}^* , er avhengige av fortegnene til både V_{EB} og V_{CE} som på figuren representerer spenning fra spenningskildene. I figuren som viser strøm-spenning karakteristikk for transistoren har vi tegnet inn grafen for $V_{EB}^* = V_{CE}$ eller ekvivalent $V_{CB}^* = 0$. Til høyre for denne kurven vil da den aktive regionen for transistoren befinne seg, mens vi til venstre har metning.

Anvendelser av transistorer

Når en transistor anvendes i en elektrisk krets har den egenskaper som gjør at den kan brukes enten som en bryter, eller som en forsterker av elektriske signal. Den vanligste konfigurasjonen er den vi har kalt felles emitter, og vi skal derfor se på hvilke sammenhenger denne kan benyttes og hvordan bruken av denne samsvarer med de transistoregenskapene vi allerede har nevnt.

Strømmen i basen I_B kalles i praktiske situasjoner ofte for kontrollstrømmen da det er denne strømmen vi kontrollerer og som bestemmer hva som skal foregå i transistoren når det er en kjent spenning V_{CE} over emitter og kollektor. At vi kan bruke en strøm til å kontrollere hva som skjer, når det som vi tydelig har sett er spenninger som bestemmer transistorens oppførsel, skyldes at strømmen I_B vil øke i takt med spenningen V_{EB}^*

Transistor som forsterker

Da vi så på sammenhengen mellom strømmer i transistoren så vi at vi hadde en sammenheng mellom kontrollstrømmen I_B og den kontrollerte strømmen I_C gitt ved: $I_C = \beta_0 I_B + I_{CE0}$.

Under visse forutsetninger er β_0 en konstant slik at vi har en lineær sammenheng mellom kontrollstrøm og kontrollert strøm. Det vi da kan gjøre er å sende signalet vi ønsker å forsterke som kontrollstrøm. Strømmen gjennom transistoren vil da være en forsterket utgave av kontrollstrømmen der mønsteret i signalet er bevart.

Et eksempel på en situasjon der det kan være ønskelig å forsterke signal er dersom vi ønsker å måle lysintensiteten fra en svak lyskilde ved hjelp av et amperemeter koplet til en solcelle der sammenhengen mellom lysintensitet og generert strøm er kjent. Dersom amperemeteret vi bruker er beregnet for sterkere strømmer vil det være mulig å kople strømmen fra solcellen til basen på en transistor, mens vi lar amperemeteret måle kollektorstrømmen. Amperemeteret vil da gjøre et større utslag. Dersom vi kjenner sammenhengen mellom endring i I_B og I_C for transistoren, det vil si at β_0 er kjent, kan vi beregne styrken på det opprinnelige signalet og regne om til lysintensitet.

Dersom vi ønsker ytterligere forsterkning av et elektrisk signal, kan man ganske enkelt lede Kollektorstrømmen fra en transistor A inn på basen til en annen transistor B. Samlet strømforsterkning blir da produktet $\beta_{\text{tot}} = \beta_{0A} * \beta_{0B}$.

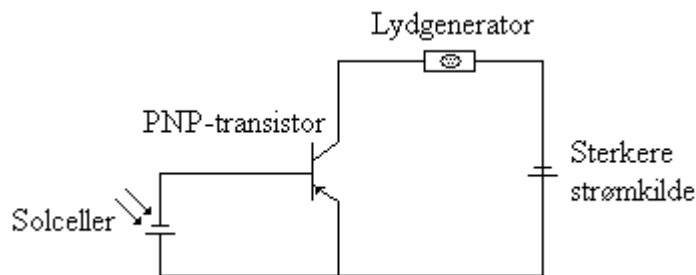
Transistorer brukes, på lignende måte som i det enkle eksempelet, i alle slags forsterkere. Transistorer brukes da riktignok i kombinasjon med flere komponenter i langt mer avanserte kretser for å oppnå ønsket kvalitet på det forsterkede signalet.

Transistor som bryter

Som vi har sett vil det ikke kunne gå mye strøm gjennom en transistor når den settes i "cut off". Når en transistor brukes som bryter vil derfor denne tilstanden være ekvivalent med at bryter står i av-posisjon. For å skru på "bryteren" må vi gjøre noe med transistoren slik at strøm kan passere gjennom transistoren. Vi kan da sette den i aktivt modus eller i en tilstand med metning. Dersom vi antar at det er en kjent negativ spenning V_{CE} , kan vi ved å sørge for at $I_B=0$ (og dermed $V_{EB}^*=0$) skru "bryteren" av. For å skru "bryteren" på er det bare å sende en kontrollstrøm inn på basen. Dersom denne kontrollstrømmen er så stor at V_{EB}^* blir større enn V_{CE} vil V_{CB}^* bli positiv slik at vi får metning.

Strømmen vi ønsker å skru på er typisk betydelig større enn kontrollstrømmen.

Vi kan, i et tenkt eksempel, forestille oss at vi har lyst til å bli vekket av en eller annen elektrisk lydgenerator ved soloppgang. Vi har tilgang til et lite solcellepanel som ikke kan generere nok strøm til å drive denne generatoren, men vi er også i besittelse av en sterkere strømkilde som er i stand til å få generatoren til å virke. Vi kan da kople inn en transistor slik at solcellepanelet styrer strømmen inn på basen, mens den sterkere strømkilden er koplet over kollektor og emitter. Lydgeneratoren kople vi i serie med Kollektorstrømmen. Når sola står opp, vil solcellepanelet generere en liten strøm som åpner for at den sterkere strømmen kan gå gjennom transistoren og få den hjemmelagede vekkerklokka til å ringe.



Figur 28: Hjemmelaget vekkerklokke.

De fleste forbinder trolig ordet transistor med datamaskiner. I datamaskiner er det transistorens bryteregenskaper som utnyttes i såkalte logiske kretser. I logiske kretser opereres det med verdiene 0 og 1, representert av transistorer som er skrudd av og på. Transistorer veksler da mellom cut-off og metning.

Konklusjon

Vi har sett hvordan denne teknologien har sitt utspring i kvantefysisk argumentasjon, hvordan kvantiserte energitilstander for elektroner i atomer fører til den såkalte båndteorien, og igjen sett hvordan dette utnyttes i elektroniske kretser.

Oppfinnelsen av den integrerte kretsen på slutten av 50-tallet er en hendelse som har revolusjonert teknologien og preget hele samfunnet vi lever i. En integrert krets er enkelt sagt en samling av sammenkoplete transistorer, dioder, kondensatorer og elektriske motstander fabrikkert på ett enkelt stykke silikon, også kalt en chip. En slik chip inneholder flere millioner komponenter pr kvadratcentimeter. Integrerte kretser, og dermed også transistorer, brukes over alt rundt oss, som for eksempel i datamaskiner, mobiltelefoner, biler og kameraer. Dersom man klarer å forstå hovedprinsippene bak halvlederkomponenter som transistorer og dioder, vil man derfor samtidig forstå mye av den moderne verden vi lever i.

Referanseliste

1. L.Solymar, D. Walsh: "Electrical properties of materials"
2. B. Tuck, C. Christopoulos: "Physical electronics"
3. H. Young, R. Freedman: "University Physics"
4. Serway, Moses, Moyer: "Modern Physics"
5. http://quantummechanics.ucsd.edu/ph130a/130_notes/node156.html
6. P.C. Hemmer; "Kvantemekanikk"
7. P.C. Hemmer: "Faste stoffers fysikk"
8. C. Kittel: "Introduction to solid state physics"
9. G. Parker: "Introductory semiconductor device physics"
10. G. Glasford: "Analog electronic circuits"
11. A. Ferendeci "Solid state and electron devices"
12. www.allaboutcircuits.com
13. <http://www.britneyspears.ac/lasers.htm>

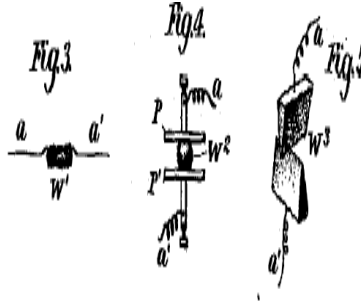
Dioder

Astrid-Sofie Vardøy, Christina Aas, Audun Eskeland Rimhaug og Are Opstad Sæbø¹
Institutt for fysikk, NTNU, Trondheim, Norge

14.11.05

¹Prosjekt i FY1013 Elektrisitet og Magnetisme 2

Sammendrag Dioder er i sin enkleste form et kretselement som bare leder strøm en vei. Denne egenskapen er svært nyttig i flere sammenhenger, og benyttes i dag i stor grad i pc, radio, tv, laser og mye mye mer. I denne oppgaven ser vi på dagens virkemåter og hvordan en ser for seg at diodene kommer til å bli brukt i fremtiden. Samtidig tar vi for oss hvordan forskjellige dioder fungerer, og ser på hvordan energibåndstrukturen i forskjellige metaller og halvledere kan føre til denne nyttige diodeeffekten. Oppgaven inneholder i tillegg et praktisk forsøk der vi tar for oss det historiske aspektet ved dioden. Vi prøvde å lage en diode av helt enkle gjenstander som minner om de aller første diodene.



Figur 1: Diverse patenter på historiens første dioder

Kapittel 1

Innledning

I 1874 ble det oppdaget at kontakt mellom noen metallsulfider og spisse metallgjenstander gav forskjellig motstand avhengig av hvilken polaritet den påtrykte spenningen hadde. Slike kretselementer er historiens første eksempler på bruk av dioder, og ble brukt i radioer og likerettere. Disse krystallene var imidlertid både inkonsekvente og uforutsigbare så etter hvert foretrakk en å benytte vakumtuber (radiatorer). Vi vil prøve å lage en krystall (diode) av et barberblad og et blyantbly etter en oppskrift funnet på internett(27). Først etter at Russell Ohl i 1939 oppdaget hva som fikk krystaller til å virke ble det åpnet for en storstilt industri. I dag finner man pn-dioder i alle sammenhenger. De blir brukt som likeretterer og pn-overgangen kan virke som lysemitterende dioder og solceller (fotodioder). Kobler man sammen to dioder kan man få en transistor. Her går vi inn på teorien bak halvledere, doping, pn overganger, metall-halvleder overganger og punktdioder. Deretter beskriver vi de viktigste bruksområdene for dioder i dag og i nærmeste fremtid.

Kapittel 2

Teori

2.1 Metall, halvledere og isolatorer

En halvleder er et materiale som forenklet sagt verken har god ledningsevne eller isolasjonsevne. Atomer med flere elektroner vil fordele disse på flere energinivåer. I de høyere energinivåene er det også flere forskjellige energitilstander, og energinivået kan da refereres til som et energibånd. I hver energitilstand kan det maksimalt være to elektroner, en for hver spinn-tilstand. Dette er kjent som Paulis eksklusjonsprinsipp. Det høyeste energinivået hvor det befinner seg elektroner kalles gjerne valensbåndet, og energien til dette båndet er kjent som Fermienergien. I ledere er dette nivået ikke fylt helt opp, slik at elektroner kan bevege seg fra det ene atomets valensbånd til det neste atomets valensbånd. I isolatorer og halvledere er derimot valensbåndet ved absolutt temperatur null grader Kelvin fylt helt opp (3). For å lede strøm må disse da få elektroner opp i et høyere bånd. Dette båndet blir da et ledningsbånd, og definisjonen av Fermienergien mister litt av sin verdi, siden vi nå ikke lenger har helt kontroll på hvor hvert enkelt atom har sine mest energirike elektroner. Energiavstanden fra ytterste kant av valensbånd til innerste kant av ledningsbåndet kalles gjerne båndgapet mellom disse to. Siden energitilstandene mellom båndene er forbudte, kan ikke denne energinivåøkningen skje gradvis. Derimot må elektroner som skal opp i ledningsbåndet få tilført all energien på en gang, slik at de kan hoppe opp i båndet over (3). Denne energien kan tilføres for eksempel ved å varme opp materialet, utsette det for et elektrisk felt eller ved å bombardere stoffet med fotoner. Og her kommer forskjellen på isolatorer og halvledere inn. Isolatorer har et båndgap som er for stort til at vi kan oppnå dette for et betydelig antall elektroner. Halvledere har et mindre båndgap, slik at vi lettere kan få elektroner opp i ledningsbåndet, og gjøre materialet i stand til å lede strøm. I tillegg til at elektronene i ledningsbåndet fungerer som negative ladningsbærere, vil det også bli hull i valensbåndet der disse elektronene var. Hullene kan så fungere som positive ladningsbærere som går motsatt vei, og fremmer ledningsevnen ytterligere.

2.1.1 Forskjeller mellom halvledere og ledere

Det er et par viktige forskjeller mellom halvledere og ledere i måten de reagerer på eksterne påvirkninger. Ledere vil generelt få redusert ledningsevnen med økende temperatur. Dette kan klassisk forklares med at atomenes økte translasjonsenergi gjør det vanskeligere for elektronene å bevege seg fra atom til atom. I halvledere vil imidlertid den økte termiske energien føre til at flere elektroner får nok energi til å komme seg over i ledningsbåndet, og ledningsevnen vil dermed øke.

Et elektrisk felt vil ha samme effekt. Elektronene får mer energi, og det blir flere av dem i ledningsbåndet.

2.2 Doping av halvledere

N- og p-halvledere er halvledere med overskudd av henholdsvis negativ og positiv ladning. Dette oppnås ved en prosess kalt doping, hvor halvlederen tilsettes urenheter i form av små mengder av et annet stoff. For å oppnå n-halvledere vil vi bruke tilsetningsstoffer som har flere valenselektroner enn halvlederstoffet vi starter med. Har halvlederen vår fire valenselektroner, som for eksempel silisium har, vil det egne seg bra å dope den med et stoff fra gruppe V. Stoffene i gruppe V har fem valenselektroner, hvorav fire i hvert atom vil brukes til å inngå kovalente bindinger med halvlederatomene rundt seg. Det femte vil da bli løsere bundet til donoratomet (2), og refereres gjerne til som donorelektronet. Ved lave temperaturer vil selvfølgelig donorelektronet forbli bundet til donoratomet. Men energien som kreves for å få donorelektronet opp i ledningsbåndet er vesentlig lavere enn for elektronene i halvlederen. Ved en liten tilførsel av energi, for eksempel ved en temperaturøkning, vil derfor donorelektronet frigjøres tidlig, og etterlate donoratomet som et positivt ion. Donorelektronet kan nå bevege seg fritt rundt i halvlederen, og bidra til å lede strøm. Donoratomet, som nå er et ion, vil forbli fiksert i halvlederen, men har ikke skapt noen huller i valensbåndet. Materialet vi nå står igjen med, er en n-halvleder.

Hvis en nå i stedet tilsatte et atom fra gruppe III til den samme halvlederen, ville det motsatte skje. Et gruppe III-atom har tre valenselektroner, hvor alle tre i hvert atom vil inngå i kovalente bindinger. Det er allikevel ett for lite i forhold til halvlederatomene, og det blir skapt et hull i halvlederens valensbånd. Et slikt tilsatt atom blir da nødt til å ta imot elektroner, og refereres derfor gjerne til som et akseptoratom (3). Ved en energibetraktning kan vi si at hvis et elektron skulle oppta dette hullet, måtte det hatt en høyere energi enn den valenselektronene har, siden donoratomet netto energi nå ville vært negativ (2). Elektronet i hullet har imidlertid ikke tilstrekkelig energi til å være i ledningsbåndet, så dette kan ikke være tilfellet. Dermed har vi et hull i valensbåndet som kan bevege seg rundt i halvlederen og fungere som en positiv ladningsbærer, forutsatt at elektronene i halvlederen har nok energi til å bevege seg rundt i valensbåndet. Dette blir som en slags boble-i-vann-effekt, en analogi flere læreverker benytter. Hullet fungerer som en positiv ladningsbærer ved at det forflyttes når elektroner ved siden av glir over i hullet. Tilsvarende vil en boble i vann bevege seg ved at boblen fylles med vann, og dermed tvinges videre. Det nå negativt ladde akseptoratomet vil imidlertid forbli i ro. I sum har akseptoratomet generert et hull i valensbåndet uten å gi et elektron til ledningsbåndet, og materialet vi står igjen med er en p-halvleder.

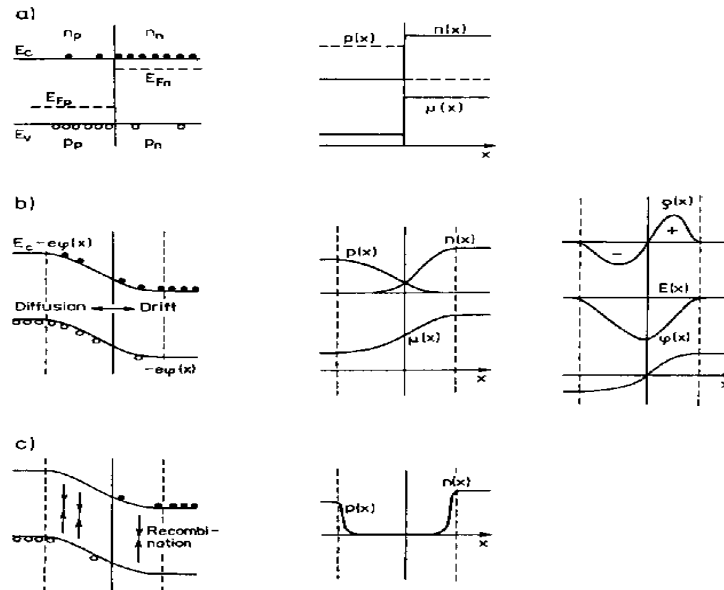
For at doping skal ha noen betydelig effekt, må antallet elektroner fra donoratomer i n-halvledere eller antallet hull fra akseptoratomer i p-halvledere, være betydelig større enn det halvlederen selv bidrar med. Eksempelvis vil en silisiumhalvleder ved romtemperatur ha ett elektron i ledningsbåndet per 10^{13} atomer. Silisium dopes derfor gjerne med ett donoratom per 10^6 til 10^{10} silisiumatom for å få en brukbar n-halvleder (3).

2.3 pn-overgang

Når vi så setter sammen en n-halvleder med en p-halvleder, får vi det vi kaller en pn-overgang. Nivået av doping på n-delen i forhold til p-delen angis som overgangens dopingprofil. Denne dopingprofilen har stor betydning for overgangens egenskaper, og forskjellige dopingprofiler anvendes derfor

til forskjellige formål. Den enkleste formen å beskrive er imidlertid den uniformt dopede overgangen; hvor p-delen har like mange hull per halvlederatom som n-delen har frie donorelektroner.

pn-overgangen består da altså av en del med akseptoratomer, og en del med donoratomer. Selve grensesnittet mellom disse to delene refereres gjerne til som den metallurgiske overgangen eller grenseflaten (3). For enkelhets skyld beskriver vi her en overgang hvor dopingkonsentrasjonen er uniform i hver halvdel, og hele overgangen mellom p- og n-delen finner sted i grenseflaten, i stedet for å se på en gradvis overgang. I det de to delene settes sammen har vi en stor tetthetsgradient på tvers av overgangen både for hull og elektroner. Dette vil da føre til at elektroner vil begynne å diffundere fra n-delen over i p-delen (1), og hullene vil begynne å diffundere motsatt vei. Forutsatt at det ikke er koblet noen spenningskilde til pn-overgangen, vil denne diffusjonen etter hvert opphøre. Dette fordi elektroner som diffunderer inn i p-delen etterlater seg positivt ladde donoratomer i n-delen, og på samme måte vil hullene som diffunderer inn i n-delen etterlate seg negativt ladde akseptoratomer i p-delen. Resultatet blir en region med netto positiv ladning i n-delen, og en region med netto negativ ladning i p-delen. Disse to regionene sammen refereres gjerne til som romladningsregionen (1). Her får vi et industert elektrisk felt perpendikulært på grenseflaten som har retning fra n-delen til p-delen. I romladningsregionen står vi til slutt igjen med praktisk talt ingen hull eller donorelektroner. Det vil fortsatt være en tetthetsgradient i grensesnittene mellom p-delen og romladningsregionen, og mellom n-delen og romladningsregionen, men diffusjonskreftene vil motvirkes av det elektriske feltet. I termisk likevekt vil kraften fra det elektriske feltet og diffusjonskreftene akkurat utligne hverandre (1), og vi får en stasjonær tilstand. Hele denne prosessen er fremstilt i figur 2.1

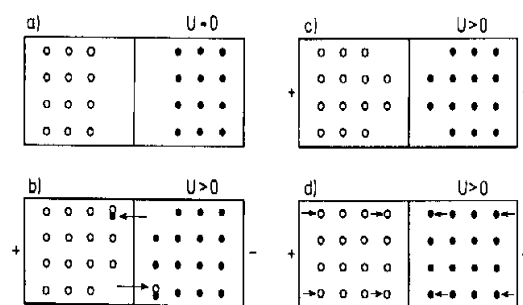


Figur 2.1: pn-overgangen går fra å ha en betydelig potensialforskjell i a), til likevekt i c) via b). Figur hentet fra (1)

2.3.1 Positiv bias

Hvis vi nå utsetter en pn-overgang for et eksternt potensiale, kan vi igjen få en forflytning av ladninger. Kobles denne spenningen slik at plusspolen på spenningskilden går til p-delen av dioden, og minuspolen går til n-delen, vil både hullene og donorelektronene bringes nærmere grensesnittet. Den totale potensialforskjellen mellom de to delene vil da reduseres fra $|V_0|$ til $|V_0| - |V|$, hvor V_0 er den opprinnelige potensialforskjellen, og V er verdien av den påtrykte spenningen (4). Dette gjør at elektronene som går fra n-delen til p-delen og hullene som går motsatt vei, alle får en lavere potensialbarriere å passere. Potensialet for elektronene og hullene har naturligvis motsatt fortegn, men det har også ladningene, og de må derfor begge overvinne den samme barrieren (4). Etter hvert som V øker, øker også diffusjonen, og vi får en strøm gjennom dioden fra p-delen til n-delen. Det er imidlertid fortsatt en svak strøm fra n-delen til p-delen av de hullene og elektronene som ikke har en potensialbarriere å overvinne, men når V blir sterk nok går netto strøm fra p til n.

Denne typen overgang refereres vanligvis til som positiv bias eller foroverbias.



Figur 2.2: Strøm i en positiv bias. Høyre side av hver figur er n-delen, venstre side er p-delen. Ved a) er pn-overgangen i likevekt uten noen spenningskilde koblet til. Området rundt grenseflaten er evakuert for hull og elektroner. Bilde b) illustrerer hvordan elektronene og hullene diffunderer på tvers av grenseflaten, c) viser situasjonen etter, mens atomene nærmest grenseflaten etterfylles i d). Figur hentet fra (1).

2.3.2 Negativ bias

Kobler vi nå spenningskilden i motsatt retning, altså med plusspolen på n-delen og minuspolen på p-delen, vil vi få stikk motsatt effekt.

Potensialforskjellen på tvers av grenseflaten øker fra $|V_0|$ til $|V_0| + |V|$ (4). Både elektronene fra n-delen og hullene fra p-delen får dermed en enda høyere barriere å passere, og strømmen blir tilnærmet lik null. Den svake strømmen fra n-delen til p-delen nevnt over vil riktignok fortsette, upåvirket av den eksterne spenningskilden (4), men i praktisk betydning har vi ingen strøm. Denne typen oppkobling kalles en negativ bias eller en bakoverbias. En enda mer kvalitativ beskrivelse vil være at plusspolen på spenningskilden trekker elektronene i n-delen bort fra grensesnittet, og tilsvarende vil minuspolen trekke hullene i p-delen i motsatt retning. Siden n-delen selv ikke kan være kilde til huller, og p-materialet ikke kan gi elektroner (3), vil det ikke være noen huller og elektroner som møtes i grenseflaten, og vi får ingen strøm.

2.3.3 Strøm gjennom en ideell pn-overgang

I et metall vil elektrontettheten være gitt av Fermi-Dirac-fordelingen (2)

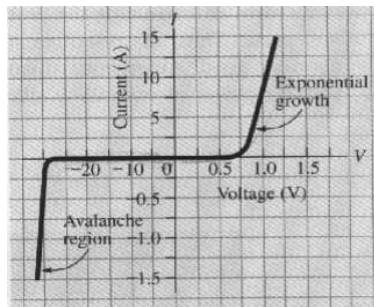
$$f_F(E) = \frac{1}{1 + \exp\frac{E-E_F}{k_B T}} \quad (2.1)$$

hvor k_B er Boltzmanns konstant og T er temperaturen. Dette er altså en sannsynlighetsfordeling som gir oss sannsynligheten for å finne et elektron med energien E ved en gitt fermienergi og temperatur. Når $E - E_F \gg k_B T$ kan ligning 2.1 tilnærmes med Boltzmannfordelingen

$$f(E) = \exp-\frac{E - E_F}{k_B T}. \quad (2.2)$$

Vi ser at antall elektroner i en energitilstand avtar eksponensielt med energien. Når det settes på en negativ spenning på n-siden vil fermienergien øke, og antall elektroner som har tilstrekkelig energi til å strømme over til p-siden, vil øke eksponensielt.

Ved negativ bias vil fermienergien på n-siden bli lavere enn på p-siden. Hvis det skal gå strøm i dette tilfellet må elektronet på p-siden ha en energi som gjør at det kan passere barrieren. I følge ligning 2.2 vil det alltid være en liten sannsynlighet for å finne elektroner med tilstrekkelig energi til å strømme over barrieren. Denne sannsynligheten vil ikke forandre seg hvis vi øker den påtrykte spenningen. Derfor vil vi alltid ha en liten strøm ved negativ bias. Enten ved sterk doping av p- og n-siden eller stor nok negativ bias, vil det likevel kunne strømme elektroner fra p- til n-siden som ikke har nok energi til å komme over barrieren. Dette kan forklares ved tunnelering av elektroner gjennom den nå smale barrieren. Tunneleringssannsynligheten er sterkt avhengig av bredden på barrieren. Derfor vil man få en sterk økning av strømmen ved en liten økning av den påtrykte spenningen, etter at tunnelering først har oppstått. Ved denne spenningen sier vi at dioden kollapser. Strøm som funksjon av spenning for en ideell diode er fremstilt i figur 2.3.



Figur 2.3: Strøm s.f.a spenning for en ideell pn-diode

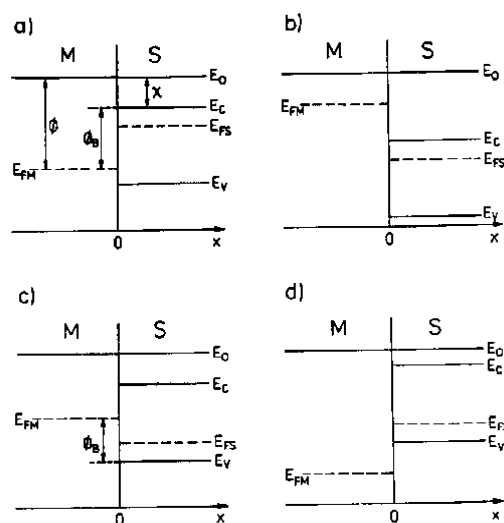
2.4 Metall-halvlederoverganger

Metall-halvlederoverganger blir laget ved oppvarming av metall til damp som deretter kondenserer på den kaldere halvlederoverflaten. Motstanden i slike kontakter er avhengig av polariteten på den påtrykte spenningen, noe som gjør at også disse kontaktene kan virke som en diode. Barrieren på grenseflaten mellom metallet og halvlederen kalles ofte en Schottky-barriere, og dioder basert på en slik overgang kalles derfor for en Schottky-diode.

2.4.1 Flatekontakt

Med flate kontakter menes det her at det er kontakt mellom metallet og halvlederen over en hel flate, ikke bare i et punkt.

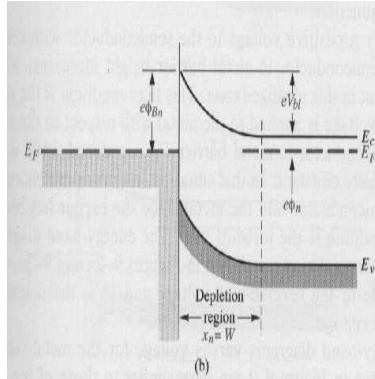
På figur 2.4 ser vi en oversikt over energinivåene i en metall-halvlederovergang *før* likevekten har



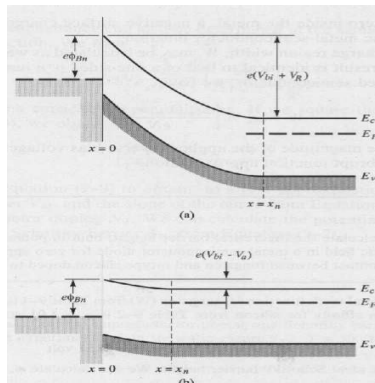
Figur 2.4: (a) Energinivåene i et metall med fermienergi E_{Fm} ligger under fermienergien E_{Fc} i en n-dopet halvleder før likevekt. $E_{Fm} < E_{Fc}$ (b) N-dopet halvleder og $E_{Fm} > E_{Fc}$ c) P-dopet halvleder og $E_{Fm} > E_{Fc}$ d) P-dopet halvleder og $E_{Fm} > E_{Fc}$. Figuren er hentet fra (1)

innstilt seg. De fire forskjellige mulighetene er å sette sammen en n-dopet halvleder med et metall med $E_{Fm} < E_{Fc}$ eller $E_{Fm} > E_{Fc}$. Her er fermienergien i metallet, E_{Fm} , energien til det høyeste energinivået som er okkupert i metallet (vi forutsetter altså null temperatur), og E_{Fc} er fermienergien til halvlederen. I tillegg kan vi sette sammen en p-dopet halvleder med et metall med $E_{Fm} < E_{Fc}$ eller $E_{Fm} > E_{Fc}$. Hvis en setter sammen et metall og en lavt- til moderat dopet halvleder, kan en få en barriere inne i metallet som på mange måter minner om den barrieren som oppstår i en pn-diode. Hvis for eksempel en n-dopet halvleder settes sammen med et metall der $E_{Fm} < E_{Fc}$, vil det strømme elektroner fra halvlederen til metallet og hull fra metallet til halvlederen helt til fermienergien i halvlederen E_{Fc} er lik fermienergien i metallet E_{Fm} . Metallet tar altså plassen til p-halvlederen i en pn-overgang. I halvlederen vil en nå få positivt ladede donoratomer, mens det i metallet er overskudd av elektroner. Dette fører til et elektrisk felt i grenseflaten mellom metallet og halvlederen, og dermed en potensialforskjell. Elektroner som prøver å strømme fra metallet til halvlederen vil derfor møte en barriere på samme måten som i pn-overgangen. Dette er illustrert i figur 2.5. Ved å sette på en negativ spenning på halvledersiden og en positiv spenning på metallsiden, vil vi oppnå en positiv bias av samme grunn som vi fikk en positiv bias i en pn-overgang. Hvis vi snur spenningen vil det føre til en negativ bias, og det vil ikke gå strøm før barrieren blir så "tynn" at elektronene kan tunnelere gjennom. Positiv- og negativ bias for en metall-halvlederovergang er vist i figur 2.6.

I en metall-halvlederovergang kan vi få strømmen til å bli lineært avhengig av spenningen. Vi sier da at vi har en ohmsk kontakt. Dette kan oppnåes på to forskjellige måter. Den ene måten er å velge et metall med fermienergi lik energien til de elektronene som har lavest energi i ledningsbåndet i halvlederen. Da vil det ikke oppstå en Schottky-barriere i overgangen mellom metallet og halvlederen. Den andre måten å oppnå dette resultatet på, er å dope halvlederen så sterkt at Schottky-barrieren blir veldig ”tynn”. Da vil vi få tunnelering av elektroner og hull gjennom barrieren ved negativ bias. Fordi barrieren blir tynnere og tynnere ettersom den påtrykte spenningen øker, vil tunneleringssannsynligheten også øke. Hvis kontakten skal være ohmsk, må antallet elektroner som tunnelerer være proporsjonal med den påtrykte spenningen.



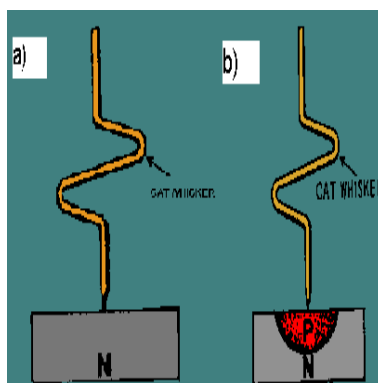
Figur 2.5: Figuren viser potensialet i en metall-halvlederovergang ved termisk likevekt. Figuren er hentet fra(2)



Figur 2.6: Figurene viser energinivåene i en metall-halvlederovergang, øverst ved negativ bias og nederst ved positiv bias. Den øverste linje på halvledersiden er ledningsbåndet, den stiplede linjen er fermienergien og den nederste linjen er valensbåndet. Figuren er hentet fra(2)

2.4.2 Punktkontaktdiode med pn-overgang

Punktkontaktdioden er avhengig av kontaktrykket mellom et punkt og en halvlederkrystall for å fungere. Halvlederkrystallen kan være en rektangulær n-type silikon eller germanium krystall, mens



Figur 2.7: Figuren viser en punktkontakt-diode med en n-dopet halvleder og en metallstang (en catwhisker). I a) har ingen elektroner diffundert over i metallet, mens i b) har dette skjedd og det har oppstått en pn-overgang rundt kontaktpunktet

punktet vanligvis er en tynn metalltråd kalt "catwhisker" (24). Metalltråden kan for eksempel være berylliumkobber, bronsefosfor eller wolfram (tungsten). Uansett bør metallet være fra gruppe tre i periodesystemet slik at det er tre elektroner i ytterste skall. I prosessen for å lage punktkontakt-dioden presses metalltråden mot halvlederkrystallen mens det kjøres en relativt høy strøm gjennom metalltråden. På grunn av varmen som oppstår vil elektronene diffundere fra halvlederen til metalltråden, og dermed konvertere halvlederen til p-type materiale. Denne konverteringen skjer i et svært lite område akkurat omkring kontaktpunktet mellom materialene. Grunnen til at man bruker en punktkontakt metalltråd framfor en flat metallplate er for å produsere et høyintensitets elektrisk felt ved kontaktområdet uten å måtte bruke en stor ekstern spenningskilde. Hvis man skulle satt en så høy spenning på en vanlig halvleder ville den overflødig varme ødelagt metallet. Den elektriske kretsen går nå gjennom metalltråden og via halvlederkrystallen. Derfor må metalltråden ha liten resistans i sin videre kontakt med kretsen. Halvlederkrystallen må ha en ohmsk metall-halvlederkontakt i sin videre kontakt med kretsen. En av forskjellene mellom overgangsdioden og punktkontakt-dioden er at ved positiv bias vil resistansen til punktkontakt-dioden være høyere enn overgangsdioden fordi kontaktflaten mellom metallet og halvlederen er mye mindre. Derimot vil ikke punktkontakt-dioden bli like uavhengig av ytre påtrykt spenning som overgangsdioden ved negativ bias. Punktkontakt-dioden har en fordel ovenfor overgangsdioden; kapasitansen mellom metalltråden og halvlederen er mindre enn kapasitansen mellom de to sidene av overgangsdioden. Dette gjør at en mindre del av strømmen med høy frekvens, generert av kondensatoren, vil gå i kretsen til punktkontakt-dioden. Punktkontakt-dioder brukes i dag i radioer (både AM og FM detektering) og TV-er for å detektere bilde. Historisk eskalerte utviklingen deres i løpet av andre verdenskrig da behovet for bedre radiomottakere var sterkt.

2.4.3 punktkontakt-diode med metallsulfid

Kapittel 3

Forsøk

Sammendrag De første diodene som ble brukt i radioer og likerettere var lagd av metallsulfider og spisse metallgjenstander. På internett fant vi en oppskrift på hvordan en slik punktkontaktdiode kunne lages ved hjelp av helt enkle gjenstander som en finner i et vanlig norsk hjem. Vi ville prøve å lage denne dioden og anvende den i en likeretter. For å ha noe å sammenligne med ønsket vi også å lage en likeretter med en moderne diode som er laget av en pn-overgang.

3.1 Teori

3.1.1 Likeretter

En likeretter er som navnet tilsier en krets som likeretter strømmen. Den gjør vekselspanning om til likespenning. Det finnes mange forskjellige kretser som likeretter strøm. Det de alle har felles er at de består av en diode som kun leder strøm en vei. På figur 3.1 er det vist en type som består av en diode og en kondensator. Sammen vil disse to komponentene føre til en likespenning over resistansen som er koblet i parallell med kondensatoren. Når vekselspanningen er positiv lades kondensatoren opp, samtidig som det går en positiv strøm gjennom resistansen. I det spenningen over kondensatoren blir større en vekselspanningen vil dioden stenge av spenningkilden fra resten av kretsen og vi har essensielt en lukket RC-krets der kondensatoren lades ut eksponensielt

$$I = \frac{Q_0}{RC} \exp\left(\frac{-t}{RC}\right). \quad (3.1)$$

Her er Q_0 ladningen ved $t = 0$, R er resistansen og C er kapasitansen. Også i dette tilfellet er strømmen gjennom resistansen positiv. Etter at vekselspanningen har vært negativ i en halv periode, blir den etter hvert like stor som spenningen over kondensatoren igjen. Med en gang den blir større, vil kondensatoren lades opp igjen og strømmen gjennom resistansen øker. Spenningsamplituden til vekselspanningen minus spenningen over kondensatoren i det den blir ”tatt igjen” av vekselspanningen, kalles ofte for ripplespenningen. Den sier hvor mye spenningsfallet over resistansen varierer. I en god likeretter bør ripplespenningen være liten. Dette får en til ved å ha en stor tidskonstant i RC-kretsen. Tidskonstanten er gitt ved

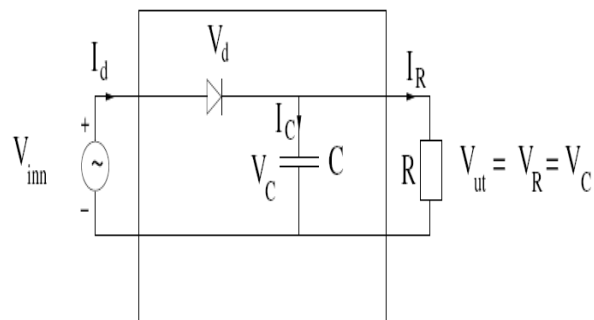
$$\tau = RC \quad (3.2)$$

og forteller hvor lang tid det tar før strømmen ved $t = 0$ er redusert med en faktor $\frac{1}{e}$.

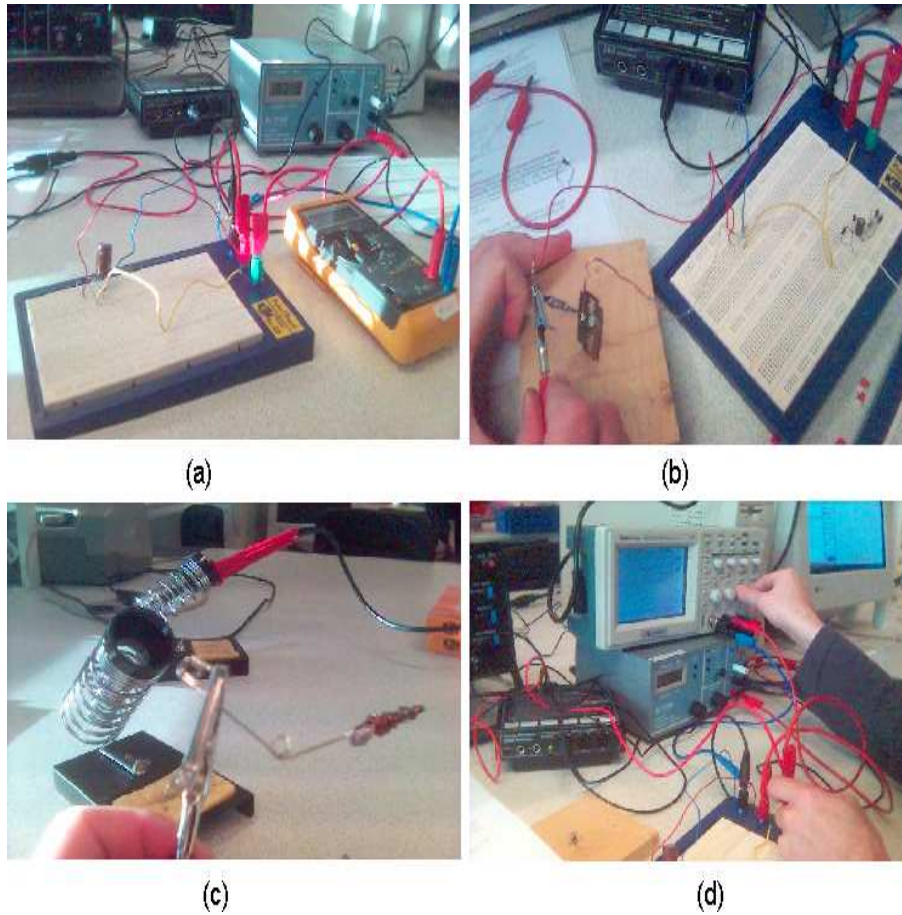
3.2 Utstysrliste

- Loddebolt med våt svamp
- Oppsuger for loddetinnet
- Loddetinn med fluss
- Holder for å holde på komponentene man lodder
- Barberblad med rust på (oksidert)
- Bly (fra blyant)
- Kobbertråder, en tykk og en tynn
- Sikkerhetsnål til å feste blyet på
- Treplate til å lodde på
- Saks og tang for å klippe metallet med
- Datamaskin med matlab dataloggingsprogram installert (DataStudio)
- Digitalt oscilloskop
- Masse ledninger
- Koblingsbrett
- Kondensator -50V maks, $39 \mu\text{F}$
- Kjøpt diode
- Vekselstrømsspenningskilde -50Hz, først 37,8Hz
- Multimeter
- Resistans 100Ω (koblet i kretsen med kondensator)
- Datalogger (svart boks med 3 input-kanaler)
- Metalltråd
- Rusten jernstang

3.3 Bilder og figurer



Figur 3.1: Skjematisk skisse av likeretteren som ble brukt i forsøket. Figuren er hentet fra(12)



Figur 3.2: a) Koblingsbrettet b) hjemmelaget diode koblet inn i resten av kretsen c) sikkerhetsnålen med grafittstang d) apparatoppstilling

3.4 Metode

3.4.1 Likeretteren

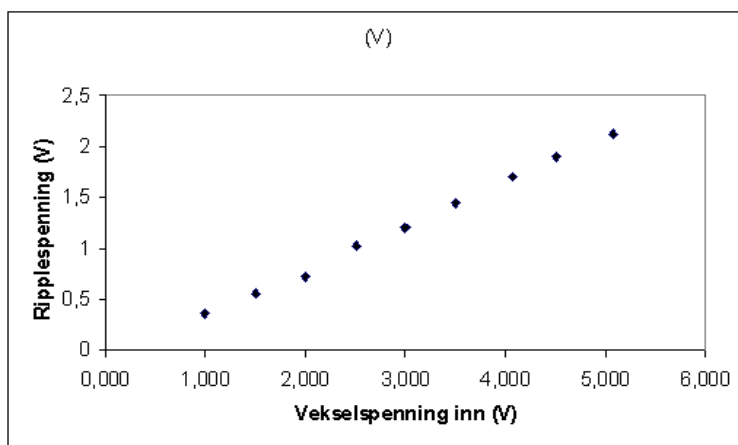
I dette forsøket ble kretsene som ble brukt (se fig 3.1) koblet på et koblingsbrett. På koblingsbrettet ble det koblet slik at spenningsfallet over resistansen kom ut på den ene utgangen (dvs. spenning etter resistansen i forhold til felles jord, men dette var lik spenningsfallet over resistansen), og denne ble koblet opp mot både et digitalt oscilloskop og en datalogger som igjen var koblet til en datamaskin. Dataprogrammet het Data Studio og var i stand til å logge bl.a. spenning over to punkter med jevne mellomrom. Fordi innspenning var 50 Hz, ble Data Studio stilt inn til å logge spenningen over resistansen med en frekvens på 200 Hz - altså fire ganger pr. svingning. Spenningskilden som ble brukt viste ikke hvilken spenning den gav. Derfor ble det koblet på et voltmeter over spenningskilden. En annen viktig detalj er at i likeretteren ble det brukt en kondensator som bare tålte spenning den ene veien. Den siden med en gul strek på skulle være den siden med negativ ladning. Dette var viktig å huske, ellers ville kondensatoren blitt ødelagt. I kretsen ble det satt inn en kondensator som hadde

en kapasitans på $39\mu\text{F}$, og en på $330\mu\text{F}$. Den største av dem var så stor at det ble vanskelig å gjøre skikkelige målinger med den. Likeretterkurven ble så godt som konstant. For lettere å kunne gjøre målinger ble derfor kondensatoren på $39\mu\text{F}$ benyttet. Da det var i orden kom det frem en nydelig likeretterkurve på oscilloskopet (se figur 3.4). Plottet av spenningen er fremstilt i graf 3.5. I tillegg målte vi ripplespenningen for ulike innspenninger, og resultatet er fremstilt i graf 3.3. Målingene her ble gjort ved å bruke cursor-funksjonen på oscilloskopet.

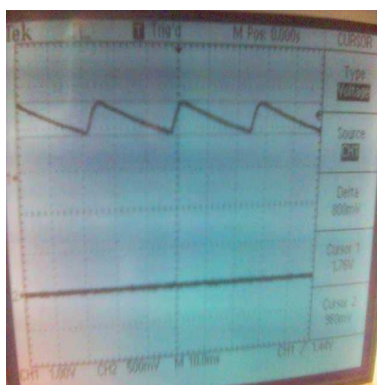
3.4.2 Hjemmelaget punktkontakt diode

Dioden vi leste om på nettet var, i tillegg til ledninger, satt sammen av et rustent barberblad, en sikkerhetsnål og en grafittstang (sånn som er inni blyanten). Tre dager før forsøket ble barberbladet pusset med fil og lagt i ramsalt vann. I tillegg ble det kokt i saltvann 1-2 ganger om dagen. Grunnen til dette var at det skulle få et rustlag på utsiden. Kokingen skulle føre til at prosessen gikk fortere. På labben ble det loddet på en ledende metalltråd til barberbladet, og grafittstangen ble surret på sikkerhetsnålen med en ledende metalltråd. Ved å koble sikkerhetsnålen og barberbladet til kretsen og presse grafittstangen mot rusten på barberbladet, skulle vi nå merke en bedre ledningsevne ved polarisering av spenningen den ene veien i forhold til den andre. En slike diode ville sannsynligvis ikke tåle en stor negativ bias. Det ble derfor brukt små spenninger i området $50\text{mV} - 1000\text{mV}$. Etterpå ble barberbladet erstattet av en rusten metallstang (armeringsjern fra byggeplass) for å sjekke om den også kunne virke som en diode.

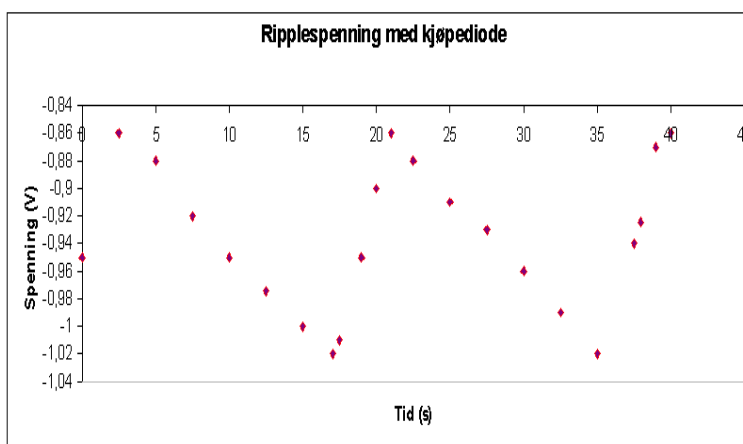
3.5 Resultat og beregninger



Figur 3.3: Ripplespenning s.f.a innspenning



Figur 3.4: Likeretterkurven som kom opp på oscilloskopet



Figur 3.5: Graf som viser likeretterkurven

Når vi plasserte grafittstangen på barberbladet oppstod det aldri noen diode-effekt. Den ledet strøm like godt den ene veien som den andre veien. Selv om grafittstangen ble plassert alle mulige steder på barberbladet, og med forskjellig trykk, gikk strømmen bare rett gjennom. Barberbladet fungerte rett og slett bare som en leder.

Når barberbladet ble erstattet med en rusten metallstang ble resultatet litt annerledes, men ikke ønskelig. Her ble effekten at det enten gikk strøm gjennom som i en leder, eller så gikk det ikke strøm i det hele tatt. Om det gikk strøm eller ikke, var avhengig av om hvor mye rust det var mellom grafittstangen og metallet. Det ble ikke observert noen forskjell i strømmen om grafittstangen ble presset hardt eller svakt mot stangen.

3.6 Konklusjon og diskusjon

Figur 3.5 burde egentlig vist likeretterkurven sammen med innspenningen. Det som ble problemet med å få dette til, var at Data studio gav en veldig merkelig graf. Det gav en rotete kurve der spenningen varierte mye, men med en periodisk variasjon og der aksene var helt på jordet. Plottet var

innstilt på å ta fire målinger pr periode, og det skulle være 50 perioder på ett sekund. På figur 3.5 ser en klart og tydelig at dette ikke stemmer overens med grafen. Løsningen ble å plote den periodiske variasjonen til plottet i Excel. Selv om aksene ikke er som de skal gir den en fin illustrasjon på hvordan spenningen så ut på oscilloskopet. Det er først og fremst derfor den er tatt med her. En ser at når spenningen er på topp og spenningen over kondensatoren blir større enn spenningskilden, vil det gå en strøm som avtar eksponensielt slik utladning av en kondensator vil gjøre. Når spenningskilden blir positiv og større enn spenningen over kondensatoren, vil dioden slippe strøm gjennom og lade opp kondensatoren igjen. Hvor mye spenningen faller før den lades opp igjen, kalles ripplespenning, og denne målte vi som sagt s.f.a innspenningen. Her ble resultatet rimelig. Ripplespenningen økte når innspenningen økte, noe som kommer av at jo større spenningen over kondensatoren er, jo større blir strømmen ut av kondensatoren når innspenningen kuttes.

Resultatet var altså tilfredsstillende for den butikkjøpte dioden. Likeretteren fungerte som den skulle. Neste punkt var nå å erstatte den butikkjøpte dioden med den hjemmelagde. Det ble som sagt ingen suksess.

Grunnen til at det ikke ble suksess med barberbladet var trolig at det ikke var nok rust på barberbladet. Det ble som sagt pusset før det ble lagt til rusting for å fjerne eventuelt beskyttende lag. Det kan hende at det ble pusset for lite, og at dette var med på å forhindre rusteprosessen. Ellers mistenkes det at barberbladene til Gillette består av en legering som ikke rustet så lett, noe en i etterpåklokskapens navn kan si at vi burde tenkt på på forhånd. Hvorfor ikke metallstangen virket vet vi ikke, men det kan ha sammenheng med at kontaktflaten ble for stor mellom grafitstangen og metallstangen. Når det er sagt skal det sies at dioder av dette slaget aldri har vært spesielt pålitelige. I oppskriften på internett (27) står det at en må føre grafitstangen frem og tilbake på barberbladet og med forskjellig trykk for å få en diodeeffekt: "If the pressure is too hard or not hard enough, the diode will not work, so experiment. The exact spot on the razor is also critical, since some spots will have too much or too little oxide on them to make the diode. Dette vil si at det også er en del tilfeldigheter som spiller inn og at en må være tålmodig og heldig for å klare å lage en diode. En annen ting vi kanskje gjorde feil, var at vi holdt grafitstangen med hendene. Hvis dioden er følsom for trykk, burde sikkerhetsnålen vært festet på en måte som gjorde at den virket som en fjær. Den ville dermed presset grafitstangen ned på barberbladet med konstant trykk og det ville vært mye lettere å oppdage diodeeffekten.

Kapittel 4

Å lage en punktkontakt diode med pn-overgang

Vi prøvde lenge å få tak i en Si-krystall med en ferdiglaget ohmsk kontakt på den ene siden av krystallen. En Ohmsk metall-halvlederkontakt på den ene siden av krystallen er viktig når en skal lage en punktkontakt diode med pn-overgang. Uten det vil en i tillegg til diodeeffekten i pn-overgangen i krystallen, også få en diodeeffekt i overgangen mellom metallet og halvlederen. Slike kontakter lages ved å dampe halvledermetallet på en oksidfri metallflate, og må derfor skje i vakum. Å prøve på noe sånt ville med andre ord vært å gape for høyt. Hvis vi derimot hadde fått tak i en slik del, kunne vi prøvd å lage en punktkontakt diode med pn-overgang, som beskrevet i teoridelen. Vi fikk desverre ikke tak i en slik del i tide, men vi fikk tenkt litt på hvordan vi skulle utført forsøket. Med god hjelp fra sakkynndige fysikere fra Institutt for Fysikk, kom vi fram til en relativt enkel metode som kanskje ville virket. Mye av problemet med å lage en slik punktkontakt diode ligger i å unngå oksidlaget som dannes på overflaten av en Si-krystall når den er i kontakt med luft. Et oksidlag vil føre til at det ikke blir kontakt mellom metallet og krystallen. En mulighet ville ha vært å pusse vekk oksidlaget i et lukket rom uten oksygen. Eventuelt så kunne vi pusset halvlederen i en inert gasstråle (i en gass som ikke reagerer med Silisium) og fått et brukbart resultat. Vi skulle altså klart å få vekk oksidlaget, men å få utført forsøket før det ble dannet et nytt oksidlag virket vanskelig. Den endelige planen ble å lodde på en ledning på metallsiden av krystallen, sette krystallen på en drill (en stor drill som står på gulvet) og erstatte borret med en spiss metallstang. Ved høy omdreining og passelig press mot krystallen var håpet at oksidlaget skulle bli slipt bort og at friksjonen mellom metallstangen og halvlederen skulle gi stor nok varme til at vi ville få diffusjon av elektroner fra den n-dopede halvlederen og inn i metallstangen. På denne måten ville vi fått en pn-overgang i halvlederen i området rundt kontaktpunktet. Ved å holde metallstaven på plass ville vi sannsynligvis unngått dannelse av nytt oksid i kontaktpunktet. Dermed kunne vi koblet dioden i serie med en resistans og satt på en vekselspanning. Hvis vi samtidig hadde koblet på et oscilloskop kunne vi avdekket om vi hadde klart å lage en diode eller ikke.

Kapittel 5

Bruk av dioder

5.1 Nåtidens bruk av dioder

5.1.1 Likerettarar

Definisjonen på ein diode er at det berre kan gå straum igjennom den ein veg. Alle diodar vil difor meir eller mindre vera ein likerettar. Dette er til dømes nyttig for å gjera vekselstraum om til likestraum.

5.1.2 Lysemitterande diodar (LED)

LED (Light-emitting diode) er ein diode som sender foton i den delen av det elektromagnetiske spekteret som er synleg for det menneskelege auga. Fotona vert danna når elektrona dett ned i tomme hol. Høgda på fallet avgjer energien til fotonet og dermed frekvensen på den utsendte strålinga. Alle diodar sender ut energi når elektron rekombinerer med hol, men ikkje alle diodar sender ut detekterbar stråling(8), som for diodar er infraraudt eller synleg lys (VLED). For synleg lys må ein ha større avstand mellom leiingsbandet og valensbandet.

Fargen på det synlege lyset vert avgjort av den eksakte avstanden mellom banda. Denne avstanden er materialbestemt. Til dømes gjev Aluminium Gallium arsenat raudt og infraraudt lys, Gallium Aluminium fosfat grønt lys og Gallium nitrat grønt og blått lys. Ein kan og få rosa og lilla lys ved å bruka lagdelte LED. Lagdelte LED vert og brukt for å få kvitt lys, då dekker ein til dømes ein blå Gallium nitrat LED med ei gulaktig fosforkappe(9). Gult lys stimulerer dei raude og grøne motakkarane i augo, og blandinga av gult og blått lys vil dermed framstå som kvitt lys for oss.

Sidan lyset frå LED'ar allereie er farga slepp ein å tapa effekt gjennom å filtrera kvitt lys. Til forskjell frå ei farga lyspære er difor funksjonen til hylsteret rundt ein LED å beskytta halvleiaren og å fokusera lyset som kjem ut av dioden.

Medan Silisium og Germanium ikkje er effektive i forhold til å gje ut lys, er generelt sett samansette halvleiarar betre(7), til dømes GaAS, GaP og GaAlAs. I ein vanleg diode ender halvleiaren sjølv opp med å absorbera ein del av desse fotona(8), medan ein i ein lysemitterande diode prøver å maksimera kor mange foton som vert frigjevne til omgivnadane.

Lysemitterande diodar bruker lite energi i forhold til vanlege lyspærer og fluorescerande lys (lysrør og liknande). På grunn av dette vert LED'ar stadig meir brukt i lommelykter, trafikklys og leikar (t.d. leikerobotar) i staden for glødelampar. Lysemitterande diodar har og den fordel at dei varer særse lengje (10 år), omlag dobbelt så lengje som dei beste lysrøra og 20 gangar så lengje som dei beste glødelamppane(9).

I dag er nok desse bruksområda for lysemitterande diodar nokre av dei mest vanlege(9):

- tynne, lette informasjonstavler. T.d. på flyplassar og togstasjonar.
- infraraude diodar i fjernkontrollar.
- blinklys og sykkellykter.
- raude eller gule LED er t.d. brukt i skilt der det er viktig å ha nattsyn. T.d. i flykabinar, brua på skip og ubåtar.
- i dyre fargeprintarar (LED printarar).
- generell opplysing.

5.1.3 Laserdiodar

Ein laserdiode minner på mange måtar om ein LED, der den største forskjellen er at laseren gjev ut koherent lys (retningsbestemt, same bylgjelengd og i fase). Teoretisk vert ein diode ein laserdiode, i staden for LED, under visse, gunstige forhold der elektronet og holet eksisterer saman ei lita stund før dei rekombinerer. Dermed kan rekombinasjonen kontrollerast med hjelp av eit foton (som vert reflektert frå laserstrålen). Under rekombinasjonen vil eit nytt foton bli emittert, med eksakt same retning, polarisasjon og fase som det stimulerande fotonet. Denne stimulerede emitteringa vil forsterka lyset og ein kan få ein laserdiode.

I den enklaste laserdioden, homooverganglaserdioden (homojunction laser diode), polerer ein endefasettane på overgangsdioden slik at dei fungerer som speglar(10) for fotona og gjer det mogeleg for laseroscillasjon. Sideflatene bør vera ujevne for å hindra lekkasje av lys frå laserkammeret. Desse homooverganglaserane er veldig ineffektive. Dei krev så mykje energi at dei berre kan operera i korte pulsar for å hindra at halvleiaren smelter(9). Så sjølv om desse laserdiodane er viktige historisk sett, og lette å forklara, er dei ikkje særleg praktiske og veldig lite brukt.

Ei forbetring av desse laserdiodane er dei doble heterostrukturelle laserane(10). Her bruker ein grenseflata mellom to enkle krystallhalvleiingar med forskjellige bandgapenergiar. Ein legg eit materiale med lavt bandgap mellom to materiale med høgt bandgap. Namnet, doble heterostrukturelle laserar, kjem av at ein dermed får to grenseflater mellom material med ulike bandgap (Døme: GaAs i samband med AlGaAs). Dette gjer både dei optiske eigenskapane og bæreigenskapane til laseren betre.

Ein diodelaser har ein tilbakekoplingsmekanisme der den sender noko av lyset (fotona) tilbake til ein fotodiode som optisk registrerer kor mykje lys som kjem og gjer signal tilbake til kretsen som driv laserdioden(11). Ved hjelp av dette regulerer kretsen kor mykje straum som går igjennom laserdioden. Dette er viktig, dioden kortsluttar fort ved for sterk straum.

Dei største fordelane med diodelaserar er at dei kan vera veldig små, krev lite energi og er billige å produsera. Dette har gjort dei til viktige elektroniske komponentar som er mykje brukt i telekommunikasjon, måleinstrument, strekkodelesarar, CD-spelarar, laserprintarar og dei kjente laserpeikarane. I tillegg har dei viktige bruksområde innan medisin. I telekommunikasjon vert laserdiodar stort sett

brukt som lett modulerte og lett kopla lyskjelder for kommunikasjon ved hjelp av fiberoptikk. Typiske måleinstrument er avstandsmålarar og laserspektroskopi, det siste kan ein til dømes bruka for å måla konsentrasjonar av ulike ion i blodprøvar eller for å finna informasjon om molekyla i ein prøve. Innan medisin er vel laserkirurgi det største bruksområdet. Ein kan både operera grå stær, fjerna uønska hårvekst, sletta ut rynker og fjerna kreftsvulstar ved hjelp av laser. I seinare tid har ein og klart å få diodelaserar med så god Koherenslengda og monokromati at dei eignar seg til aktivitetar som krev særst nøyaktige laserar t.d. holografi.

5.1.4 Fotodiodar

I fotodiodane skjer omtrent det motsette som i LED'ar og laserdiodar. Fotodioden absorberer foton eller lada partiklar og skaper ein straum i ein ekstern krins. Dette kan den gjera med å absorbera eit foton som får eit elektron i valensbandet til å verta eksitert til leiingsbandet, dermed vil det byrje å gå ein straum. Dei fleste fotodiodar er laga av Silisium, Germanium eller Indium Gallium arsenid(13).

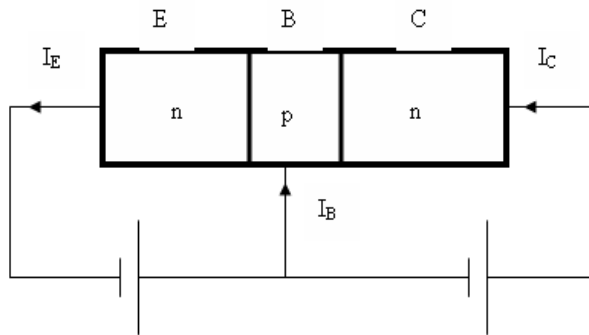
Fotodiodar kan anten verta brukt i negativ bias eller utan påtrykt spenning (null bias). I null bias vil fotona som treff dioden forårsaka ein spenningsforskjell over p-n overgangen og det vil oppstå ein straum i positiv bias; frå p-sida til n-sida i halvleiaren. Dette er den fotoelektriske effekten som er utgangspunktet for solceller.

Dei fleste pn-diodar har ekstremt høg resistans mot å la straum gå frå n- til p-sida(negativ bias). Når lys med passande frekvens skin på overgangen vil denne resistansen verta redusert, dermed kan ein bruka ein negativt biasa diode som ein detektor ved å registrera straumen som går gjennom den. Denne effekten gjer oss i stand til å detektera veldig små kvanta med lys og er nyttiggjort i spektroskopi, fotografering, optiske posisjonssensorar, overflatekarakteristikk, laseravstandsmålarar, optisk kommunikasjon og medisinske instrument for avbildning.

5.1.5 Transistorer

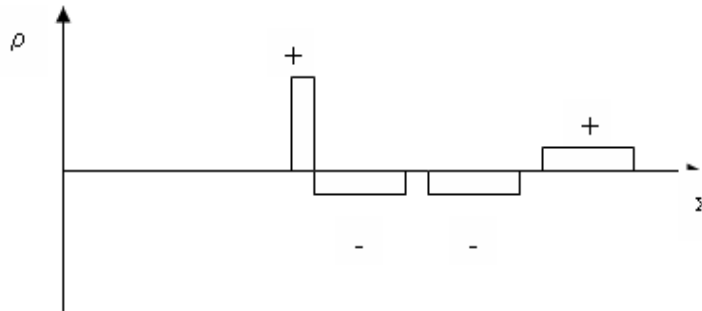
En av de aller første typene som ble utviklet var transistorer med bipolare overganger (1948) (6). Den andre typen transistor heter felteffekttransistor. En slik transistor fungerer på prinsippet om tiltrekning og frastøtning mellom ladninger på grunn av påtrykt elektrisk felt. Signalforsterkningen skjer fordi i noen deler av transistoren vil ladninger bli frastøtt, mens i andre deler vil ladninger bli tiltrukket. Dette kan brukes til å styre mengden strøm som kommer gjennom komponentene. Felteffekttransistorer fungerer mer effektivt enn bipolare fordi et stort signal kan bli kontrollert av en veldig liten mengde energi.

Vi har valgt å fordype oss i hvordan den bipolare transistoren fungerer. Det finnes to former bipolare overganger; n-p-n eller p-n-p (5). I begge formene benytter man seg av to pn-overganger, altså to dioder. Hvis man for eksempel ser på en n-p-n transistor, som er den enkleste å lage og standardmessig bruker silisiummateriale, vil den ene n-typen kalles emitter (E) og den andre n-typen collector (C). P-typen imellom kalles base (B). Når transistoren settes inn i en krets kobles den slik at man har positiv bias ved np-overgangen og negativ bias ved pn-overgangen.



Figur 5.1: Skjematisk figur av transistor

På grunn av Kirchhoffs løkkeregler vil $I_B = I_E - I_C$. Ved likevekt er $V_{BE} = V_{BC} = 0$. Da har vi en ladningstetthet som ser slik ut:



Figur 5.2: Ladningstetthet i diode-overgangene

Denne ladningstettheten viser hvordan ladningene fordeler seg akkurat i diode-overgangene. Årsaken bak fordelingen ligger i at en diode ved likevekt vil ha likt kjemisk potensial (Fermi-energi) over hele dioden. Dermed må noen ladninger forflytte seg over overgangen for å senke/heve potensialet på hver side, og man oppnår et indre elektrisk felt ved hver overgang. Dette indre elektriske feltet skaper en barriere som enten kan senkes ved å sette på et ytre påtrykt felt, eller som kan tunneleres gjennom når overgangen er smal nok.

Dersom basen er bred vil emitterte hull fra E rekombinere i stor grad med elektroner i B , dvs. at hullene ikke "collectes" (samles) av $C \Rightarrow I_C$ blir liten og transistoren oppfører seg i realiteten som to separate pn-overganger. Hvis basen derimot er smal får man stor strøm mot C og transistorvirkning. Fordi begge de påtrykte spenningene kan både være positive og negative får vi fire ulike operasjonsmodi for transistoren:

1) Aktiv modus: $I_C \simeq I_E$ (begge er store) og $I_B \simeq 0$, positiv V_{EB} og negativ V_{CB} .

- 2) Metning: de påtrykte spenningene er små, mens alle tre strømmene er store \Rightarrow lukket krets/knutepunkt.
- 3) Cut-off: Negativ spenning på begge pn-overgangene \Rightarrow små strømmer.
- 4) Invertert: Omtrent som aktiv modus, men mindre strømfaktor i basen på grunn av mindre emitter-effektivitet fordi tettheten av hull i C er mindre enn tettheten av elektroner i B .

Man kan beregne en strømovertføringsrate ved å dele IC på IE , og beregne en strømvinning ved å dele IC på IB . Disse beregningene vil vise om transistoren faktisk klarer å forsterke liten forandring i strømmen fra emitter til base, til stor strømforandring inn i collectoren. Da får man en forsterker som gjør et lite signal inn til overgangen med positiv bias om til et stort signal (strømforandring) i overgangen med negativ bias. Dette skjer fordi strømmen i den positive overgangen henger direkte sammen med strømmen i den negative overgangen.

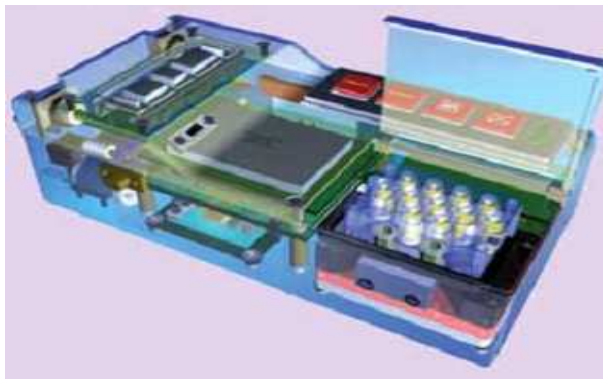
Transistorenes lave pris, fleksibilitet og pålitelighet har gjort den til en nøkkelkomponent innen moderne elektronikk. I tillegg til forsterkere er de mye brukt som svitsjer, og digitale datamaskiner av alle typer inneholder transistorer.

5.2 Framtidig bruk av dioder

På bakgrunn av nyere utvikling har man fått blått laserlys, med kortere bølgelengde, som brukes i HD-DVD og Blu-Ray teknologi til optisk lagring og lesing. På grunn av den kortere bølgelengden tillater blått laserlys høyere tetthet på lagringen og dermed flere bytes på diskene. I mars 2003 annonserte det japanske selskapet SANYO at de hadde utviklet en høy-effekts blå-lilla laserdiode med lavt støynivå, for neste generasjons optiske disk (15). Den blå-lilla dioden er laget av gallium nitrat og ble utviklet fordi etterspørselen krevde større lagringskapasitet på diskene. Dette vil snart innta det store kommersielle markedet, overta for den nåværende DVD-en og muliggjøre dobbelt-lag lagring i større skala. Allerede i mai 2005 annonserte Toshiba at de har laget en trelags disk med lagringskapasitet på 45 gigabytes. Men Blu-Ray holder på å utvikle tolags disk med kapasitet 50GB og trelags disk med kapasitet 100GB. Dette er en hard formatskonkurranse mellom to nesten like, framtidige DVD-formater som begge baserer seg på blå laser teknologi. I det lange løp er denne konkurransen muligens svært sunn for teknologiutviklingen, men inntil videre skaper den bare publikumsforvirring og irritasjon.

Et annet bruksområde for blått laserlys er innenfor plantedyrking og -overvåking (16). Blått lys kan nemlig måle mengden klorofyll i planter, og mengden av grønne korn forteller om plantens vekst og sunnhetstilstand. Steen Grüner Hanson ved avdelingen for Optikk og Fluid Dynamikk forteller at den blå laserstrålen vil bli absorbert av klorofyllet i planten. Der vil det omdannes til rødt lys som blir emittert av planten. Ved å analysere det emitterte røde lyset kan man se om planten mangler vann eller gjødsel. Man vil også kunne måle mengden oksygen planten tar opp fra jorda. Denne overvåkingen kan gjøres fra fly, traktorer eller håndholdte instrumenter. Laserlyset kan også brukes i roboter for å fjerne ugress. I dag har man to muligheter i kampen mot ugress: luke det vekk som krever mye fysisk tungt arbeid, eller å sprøyte på giftstoffer som dreper ugresset, men skaper en fare for forurensning av grunnvannet. Det miljøvennlige, framtidige alternativet er da roboter som overvåker jordene, finner ugresset og skjærer over stilkene med laserstråler. Problemet forskerne har oppdaget gjennom uttestinger av systemet så langt har vært at de infrarøde diodene de har brukt for å kutte stilkene er for dårlig fokusert. Her er løsningen gitterlasere, som er under utvikling på forskningssenteret Risø. Gitterlasere fungerer slik at et filter velger ut et utsnitt av lyskjeglen som gjør at man oppnår en skarp fokusering. Denne tynne laserstrålen sendes deretter på et optisk gitter som returnerer strålen til laserdioden hvor den blir forsterket og sendes ut enda smalere og kraftigere. Blått laserlys vil igjen være enda mer effektivt ettersom det blå lyset blir absorbert av plantene og

man dermed ikke trenger en så kraftig stråle for å kutte over stilkene.



Figur 5.3: Bio Detector, et eksempel på et minilaboratorium.

Hele laboratorier på en brikke er også en del av den nye utviklingen. Dette laboratoriet kan bestå av et system av kanaler som man kan sende en prøvewæske gjennom, f.eks. blod eller urin (16). Kantene i kanalene er dekket av forskjellige antistoffer som selektivt binder seg til ulike kjemiske stoffer i prøvewæsken, spesielt de man da ønsker å påvise. For å sjekke om man har funnet et særskilt stoff kan man sende laserlyset gjennom brikken og se om lysets brytningsindeks er endret eller ikke. Jo flere kanaler man kan pakke sammen i brikken, desto flere stoffer kan man analysere på en gang. Det er bølgelengden på laserlyset som setter grensen for hvor tynne kanalene kan være. Det er her fordelene av blått laserlys kommer inn. Hvis man i nærmeste framtid kan klare å gjøre blått laserlys billig nok, kan man masseprodusere disse små laboratoriene og bruke dem i veldig mange varierte bruksområder, f.eks. som en del av feltarbeid i den tredje verden. Inntil videre brukes de for å oppdage luftbårne biologiske agenter brukt i terrorisme, DNA testing og narkotika oppdagelser. Man har både håndholdte systemer, som Sandia's MicroChemLab Bio Detector, og laboratoriebaserede systemer, som Caliper Life Sciences' LabChip 3000 -brukt for narkotikaoppdagelser (21). Sandia National Laboratories holder for tiden på å utvikle MicroChemLab Bio Detector videre. Dette minilaboratoriet består av en 10 cm lang (total lengde) isolert separasjonskanal som er etsset inn i en silikon mikrodatabrikke. Ved hjelp av trykk får du presset prøvematerialet inn på brikken der elektriske felter brukes for å manipulere bevegelsen til fluidene i mikrokanalene. Blållilla laserdioder trengs for å indusere fluorescens fra de kjemisk merkede proteinene inne i kanalene. Deretter kobler man til en detektor, gjerne lysabsorberende diode, for å måle mengden av fluorescerende emisjon. Framtidsplanene til Caliper Life Sciences består av å legge flere funksjoner til labbrikkene, blant annet polymere kjedereaksjoner brukt på DNA prøver. De jobber også med å videreutvikle metodene for å bevege fluidene over brikken. Sandia forteller at "Lab-on-a-chip" fremdeles er veldig tidlig i utviklingskurven og har et enormt potensial videre. De spår økt bruk av minilaboratorier i områder som vannanalyse, helsesystemer for hjemmebruk og sykdomsdeteksjon.

5.2.1 Organiske dioder

Thomas Garm Pedersen, ved Instituttet for Fysikk og Nanoteknologi ved Aalborg Universitet, har forsket på organiske nanolydioder (17). Dette er lydioder basert på organiske plastmaterialer, kalt OLED (Organic Light Emitting Diode). Plast består jo av kjedeformede molekyler, såkalte polymerer. Mettede polymerer, altså kjeder uten dobbeltbindinger, er glimrende elektriske isolatorer og

er i dag brukt i ledninger for å beskytte. Men i umettede polymerer er elektronene i de annenhver plasserte dobbeltbindingene løst bundet og kan dermed lede en elektrisk strøm gjennom materialet, liknende en halvleder. Gjennom å koble et batteri til kan man få plasten til å lyse. Denne nye forskningen har enorme utviklingsmuligheter og mange veldige positive trekk: Plastteknologi er mye enklere enn mange andre teknologiske områder. I tillegg kan plastlagene bli svært tynne og fleksible, slik at OLED kan bøyes og rulles sammen. Plast kan produseres i store arealer og fargen på det utsendte lyset bestemmes av den kjemiske sammensetningen av polymerene. Ytelsesevnen til de polymere diodene er lavere enn de silisiumbaserte diodene på grunn av mindre bevegelighet for ladningsbærerne, men dette oppveies helt klart av bøyeligheten til materialet og den mye lavere prisen. Allerede nå holder store firmaer som Philips på med å utvikle skjermer basert på OLED piksler til bruk i data-, mobiltelefon- og tv-skjermer. Og i framtiden kommer kanskje OLED tapeten til å lyse opp i stua, og OLED avisen med digitale nyheter. Problemer som først må overvinnnes er å kunne lage tynne nok plastlag for at strømmen skal kunne gå, og å lage riktig typer plastikk med best effektivitet.

OLED er bygget opp med en glassplate som underlag (17). Oppå glasset ligger et lag av Indium-Tinn-Oksid (ITO), et gjennomsiktig metallag som leder strømmen inne i plasten og samtidig lar lyset slippe ut. Det siste laget er plasten, oppå ITO-laget. Plasten legges gjennom såkalt spincoating, dvs. at polymeren dryppes på underlaget for så å bli satt i hurtig rotasjon for å danne ultratynne lag. Tykkelsen kan kontrolleres gjennom rotasjonshastigheten. Oppå plastlaget kommer en topp-elektrode, som legges i et vakuumkammer der metaller varmes opp til fordampning. Når lysdioden plasseres i vakuumkammeret, legger de fordampede metallatomene seg på plasten og danner en elektrisk kontakt. Men ved å bruke "fotoluminescens" framfor "elektroluminescens" trenger man ikke å lage metallkontakter for å lede strøm gjennom dioden fordi UV-lyset man da sender inn mot dioden blir absorbert og utsendt på nytt i form av lys med lavere energi. Et siste bruksområde for organiske nanolysdioder er plastikk-solceller. Eneste problemet er at effektiviteten er svært lav fordi elektronene som absorberer lyset, ikke frivillig vil flytte på seg og skape elektrisk strøm. Dette har man klart å forbedre gjennom å tilsette C60-molekyler til plasten fordi disse molekylene virker tiltrekkende på elektronene. Selv at forskerne er svært flinke til å løse problemene de støter på, er det fremdeles et par utfordringer for neste generasjon: vann kan faktisk ødelegge de organiske diodene, og levetiden til de blå organiske diodene er bare på 1000 timer- i forhold til 10 000- 40 000 timer for røde og grønne OLEDs (22). Men tross problemer, dette er spennende teknologi for framtiden! OLED kan oppfriskes 1000 ganger raskere enn LCD-skjermer, det gjør at du nesten kan vise noe i "real-time" (22). På grunn av formbarheten kan du i framtiden sy små OLED skjermer inn i klærne dine, blant annet for utendørs overlevelsesutstyr. Og de framtidige TV-ene vil kunne rulles opp og fungere som gardiner!

5.2.2 Nanoteknologi

Halvledermaterialer med nanostrukturer er en stor del av framtiden (14). Et optimalt mål er å lage elektroniske komponenter med så små dimensjoner at de kan reguleres av ett enkelt elektron, såkalte ett-elektron komponenter. Et annet forskningsområde er halvledere bestående av kvantepunkter istedenfor energibånd. Disse halvlederne er nanometer store halvledende krystaller innbakt i et annet krystallinsk halvledende materiale med større båndgap. Dermed blir lednings- og valenselektronenes (hullenes) bevegelse begrenset i alle retninger og kvantisert i likhet med elektronenes bevegelser i atomer. Selv at kvantepunktene inneholder opp mot en million atomer, har deres elektroniske og optiske egenskaper likheter med tilsvarende egenskaper hos enkeltatomer. I tillegg er det ytre krystallinske materialet lett å integrere i vanlig faststoffselektronikk. Fordelen med kvantepunktene er at de har diskrete energitilstander som gir mange flere muligheter. Problemet hittil er manglende



Figur 5.4: Et eksempel på en skjerm laget av organisk lysemitterende dioder, fra Universal Display Corp.

kontroll på den geometriske plasseringen og utformingen av de enkelte punktene. Når man får kontrollert koblingen mellom kvantepunktene kan man lage nye klasser av komponenter for bruk i kvantedatamaskiner, optiske hukommelselementer og enkeltfoton lasere. I en kvantedatamaskin vil kvantepunktene fungere som logiske elementer (qubits) (26). Kvantedatamaskinen utnytter da at et kvantemekanisk to-nivå system har uendelig mange tilstander som svarer til alle koherente superposisjoner av de to tilstandene. To-nivå systemene kan enten basere seg på elektronenes spinn eller på ladede elektron-hull par (excitoner som er et elektron og et hull i en midlertidig stabil bane rundt hverandre). Utfordringen er å holde systemet koherent lenge nok til å kunne utføre et større antall operasjoner. En av ideene er å bruke laserpulser for å gjøre disse operasjonene. Excitonene kan være i den kvantemekaniske superposisjonstilstanden i nanosekunder, til og med mikrosekunder. Dette er tusen og millioner ganger lenger tid enn en laserpuls, og tillater dermed tusenvis av laserpulser innenfor et operativt tidsrom for å gjøre logiske operasjoner.

Men det er ikke bare i kvantedatamaskiner man finner kvantepunkter. Fordi kvantepunkter kan designes til å inneholde ett enkelt kvantisert elektron om gangen, kan dette elektronet sende ut ett enkelt foton om gangen. Dette kan utnyttes til å sende kodede signaler, kalt kvantekryptering (14). Enda et spennende forskningsområde er Spintronics. Spintronics (spinnbasert elektronikk) handler om å regulere elektroniske komponenter gjennom vekselvirkning med elektronenes magnetiske moment (spinn). Målet er å styre en strøm ved å påvirke et enkelt spinn med et magnetfelt (eller polarisert lys). Elektronenes spinn inneholder nemlig et magnetisk moment som kan bli oppdaget og kontrollert ved hjelp av magnetiske felt. Denne forskningen brukes i første omgang til å lage hukommelselementer med ultra stor kapasitet (14 og 18). Nyeste resultater er 1TB på en vanlig 3,5" diameter disk (19). En av de vanligste måtene å lage spinnpolarisert strøm på er å kjøre strømmen gjennom et ferromagnetisk materiale. En slik GMR (Giant MagnetoResistance) komponent består av to lag av ferromagnetisk materiale, separert av tomt rom. Hvis de to magnetiseringsvektorene i de ferromagnetiske lagene er parallelle vil strømmen nærmest strømme fritt, mens hvis magnetiseringsvektorene er antiparallelle oppnår systemet størst motstand. Forskjellen i resistans mellom disse to orienteringene kan være så høy som 200 % (19). Det finnes to varianter av GMR; "current-in-plane" der strømmen går parallelt med lagene, og "current-perpendicular-to-the-plane" der strømmen går vinkelrett på lagene (19). I framtiden vil vi kanskje få se MRAM (Magnetic Random Access Memory) og spinnbaserte transistorer som krever at det utvikles mag-

netiske halvledere der ferromagnetismen oppstår ved romtemperatur. Disse transistorene vil bestå av magnetiske dioder som dannes av to lag; et magnetisk halvlederlag (p-laget) og et elektronlag der spinn kan induseres. Dermed får vi en likeretter. Det er ikke bare vanlige likerettere som lages, men også Spin-LEDs. Her blir en strøm av spinnpolariserte elektroner skutt inn i en magnetisk halvleder. Når de spinn-polariserte elektronene rekombinerer med de ikke-polariserte hullene får vi sirkulært-polarisert emittert lys (20).

Til sist holder forskningen på med å lage molekylær elektronikk, altså elektroniske kretsløp basert på enkeltmolekyler. I de seneste år har man klart å vise at enkeltmolekyler kan fungere som både motstander, dioder og transistorer i nanoskala komponenter. Forskningen har også laget diode-pumpede lasere, istedenfor lasere drevet av vanlig lamper. Dermed oppnår man høy elektrisk-til-optisk effektivitet og avkjølingsbehovet er blitt drastisk mindre. Levetiden er mye lengre og hele det nye lasersystemet er mer kompakt i størrelse, altså krever mindre plass.

Kapittel 6

Konklusjon

Gruppen har gjennom dette fordypningsarbeidet de siste ukene oppnådd en mye bedre innsikt i dioders virkemåter og funksjoner i samfunnet. Vi har i tillegg satt oss grundig inn i den teoretiske bakgrunnen og forsøkt å forstå hvordan en diode kvantefysisk fungerer. Mye av den praktiske forståelsen ville vi heller oppnå gjennom et forsøk. Vi bestemte oss for å bygge to dioder og sammenligne begge med en kjøpt, fungerende diode. Den ene av diodene vi skulle bygge fant vi på en internettside som handlet om å bygge ens egen radio av husholdningsartikler. Den andre dioden skulle være en punktkontakt diode med metall-halvlederovergang. Dette materialet kom dessverre altfor sent fra universitetet i Oslo, og vi var nødt til å skrive teoretisk om hvordan vi ville gjennomført forsøket i stedet. Selv om den hjemmelagde dioden tilslutt ikke fungerte, og vi aldri fikk bygget den andre dioden, følte gruppen selv at vi har lært veldig mye gjennom dette prosjektet. Den eneste virkelige feilen er den vi ikke lærer noe av”, John Powell.

Kapittel 7

Litteraturliste

1. Fundamentals of semiconductors and devices av Rolf Enderlein og Normann J. M. Horing
World Scientific
2. Semiconductor physics and devices-basis principles av Donald A. Neamen
3. Generell fysikk for universiteter og høyskoler, bind 2 varmelære og elektromagnetisme, av Egil Lillestøl, Ola Hunderi og Jan R. Lien
4. Physics for scientists and engineers av Fishbane, Gasiorowicz og Thornton
5. Notater fra Elektrisitet og Magnetisme 2- faget, om transistorer.
6. MSN Encarta (internett leksikon): http://encarta.msn.com/encyclopedia_761566928/Electronics.html
7. Lasers by John Watson: <http://vcs.abdn.ac.uk/ENGINEERING/lasers/semi.html>
8. HowStuffWorks: <http://electronics.howstuffworks.com>
9. Wikipedia-the free encyclopedia: http://en.wikipedia.org/wiki/Main_Page
10. Toshiba sin produktguide for synlege laserdiodar:
<http://www.semicon.toshiba.co.jp/eng/prd/opto/doc/pdf/bce0026b.pdf>
11. Semiconductor Laser Diodes presentert av Ryan LoIacono:
http://www.ee.eng.buffalo.edu/faculty/cartwright/teaching/ee494s01/Presentations/Semiconductor_Laser_Diodes.pdf
12. Øving 4 i FY1013: <http://web.phys.ntnu.no/~stovneng/FY1013.2005/oving4.pdf>
13. Photodiode Characteristics and Applications by UDT Sensors Inc.:
<http://www.udt.com/Datasheets/Other/PhotodiodeCharacteristics.pdf>
14. En rapport om det teknologiske fremsyn om nanoteknologi av Thomas Bjørnholm, NSC-KU, Ole Hansen, MIC-DTU, Kent Mattson, NKT-Research og Kristian Stubkjær, DTU (koordinator).
15. En pressemelding fra SANYO om deres utvikling av høy-effekts blålilla laserdiode for neste generasjons optiske disk: www.sanyo.co.jp/koho/hypertext4-eng/0303news-e/0326-e.html

16. En artikkel fra RisøNyt nr. 1 i 2002, "Fremtidens lasere".
17. "Organiske nanolydioder" av lektor Thomas Garm Pedersen ved Instituttet for Fysikk og Nanoteknologi ved Aalborg Universitet.
18. "The new era of spintronics" av Albert Fert, Jean-Marie George, Henri Jaffrès, Richard Mattana og Pierre Seneor, publisert i Europhysics News (2003) Vol. 34 No. 6.
19. En definisjon av Spintronics fra Wikipedia: <http://en.wikipedia.org/wiki/Spintronics>
20. Et abstrakt av artikkelen "Bipolar Spintronics", skrevet av Jaroslav Fabian ved Karl-Franzens Universitet, presentert ved MRS høstmøtet i 2002.
21. www.memagazine.org/backissues/may04/features/shrinklab/shrinklab.html
22. <http://science.howstuffworks.com/oled.htm>
23. Et bilde av praktisk bruk av myke OLEDs, fra [Google.no](http://www.google.no)
24. www.fnrf.science.cmu.ac.th/theory/microwave/Microwave
25. <http://modernworldview.net/diode/>
26. En definisjon av kvantedatamaskiner fra Wikipedia: http://en.wikipedia.org/wiki/Quantum_computer
27. Enkel oppskrift på radio: http://sci-toys.com/scitoys/radio/homemade_radio.html

**Fysikkprosjekt i Elektrisitet og Magnetisme 2
Høsten 2005**

Datamaskinen

Historie, oppbygning og virkemåte



Av

Andreas Wahl, Patrick Slåttnes og Simen Eidsmo Barder

Innholdsfortegnelse

	Side
Innledning	4
Datamaskinens historie	4
Abakusen og de første regnemaskinene	4
Transistoren og mikroprosessen	5
Datamaskinens oppbygning	8
Kabinettet	9
Hovedkort og koblinger	9
Formfaktor	10
Tastaturet	10
Skrivemaskinen	10
Dovrak-oppsettet	11
Tastaturets oppbygning	11
Fordeler og ulemper	13
Gummikuppel-tastaturet	14
Datamusen	14
Datamusens historie	14
Datamusens oppbygning	15
Behandling av data	16
Transistorer	17
Halvledere	17
Dioder	18
”Bipolar junction transistor”, BJT	18
”Field-effect transistor”, FET	19
Mikroprosessor	20
Primærlagring	23
RAM	23
Hurtigbuffer	25
ROM	26
Sekundærlagring	27
Magnetisk lagring	27

Optisk lagring	27
Harddisk	28
Virkemåte	28
CD og DVD	30
Skjerm/monitor	31
Signal	32
Bilderørskjermer – CRT	33
Visningen av bilder	34
Flatskjermer – LCD	35
Flytende krystaller	35
LCD-skjermens sammensetning	35
Passive rutenett	36
Aktive rutenett	37
Lys	37
Farge	37
Andre typer monitorer og projektorer	37
Datamaskinens fremtid	38
Refereanser	40

Innledning

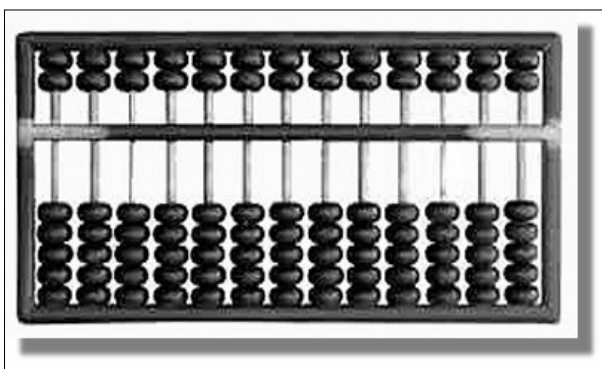
Datamaskiner utgjør en del av hverdagen til nær sagt alle mennesker i vår del av verden. De finnes overalt rundt oss og for mange er de et nyttig og viktig verktøy. Ingenting tyder på at datamaskiner kommer til å miste sin posisjon i samfunnet, snarere tvert i mot vil det bli både flere av dem og de vil bli viktigere. Mange tenker på en Personal Computer, PC, når de hører ordet datamaskin. Det kan enten være en bærbar laptop eller en stasjonær skrivebordsmaskin.

De fleste grunnkomponentene i andre datamaskiner finnes igjen i PC-er. I denne oppgaven har vi derfor sett på hvordan PC-en er bygget opp og hvordan komponentene bygger på elektriske og magnetiske egenskaper.

Datamaskinens historie [20-27]

Abakusen og de første regnemaskinene

En datamaskin kan defineres som en maskin som tar inn en form for data, prosesserer denne informasjonen, og gir ut behandlede data. Mennesket har alltid ønsket å forenkle det rutinemessige arbeidet som vanlig regning er, og oppdaget tidlig at man ikke kom langt nok ved å bare regne på fingrene. Småstein ble derfor brukt som hjelpemiddel ved større utregninger, derav ordet ”kalkulere”, som kommer fra det latinske ordet for småstein; ”calculus”. Kuleramma, eller abakusen, er en avansert form for regning med steiner, figur 1.



Figur 1; Kuleramma var den første ”datamaskinen”.

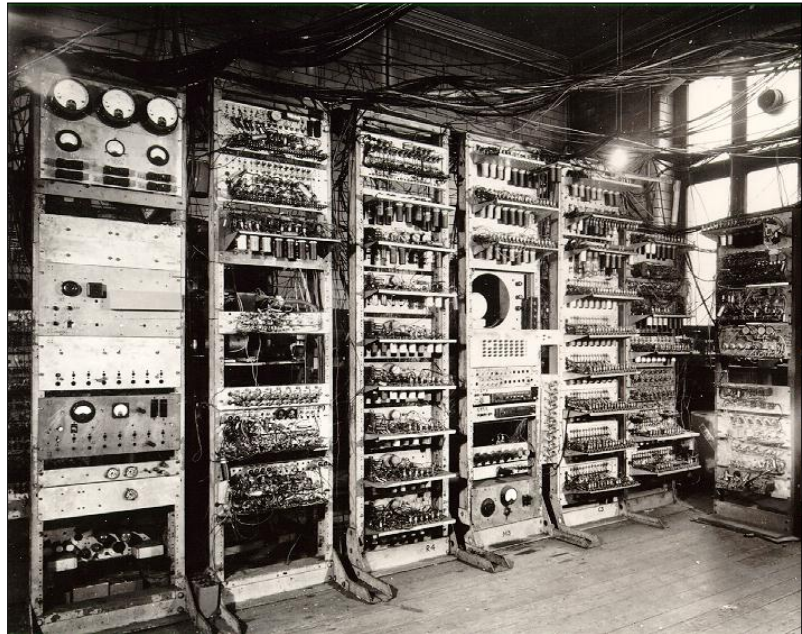
Abakusen ble trolig oppfunnet av babylonerne rundt 2400 f. Kr, og kan med sitt primitive input/outputsystem kalles den første datamaskinen. Abakusen var det raskeste regneredskapet helt frem til 1642, da matematikeren Blaise Pascal bygget den første mekaniske regnemaskinen. Pascal får vanligvis æren for denne oppfinnelsen, selv om Leonard da Vinci tegnet skisser av en lignende innretning på 1500-tallet.

Charles Babbage var en engelsk matematiker som i 1822 utviklet ”The Difference Engine”, verdens første automatiske, mekaniske regnemaskin. Maskinen ble aldri fullført, blant annet fordi de mekaniske elementene han trengte for å bygge den ikke var ferdig utviklet på den tiden. Charles Babbage blir likevel ofte omtalt som datamaskinens far, fordi hans arbeider forandret definisjonen på hva en datamaskin skal bestå av;

- en input/outputenhet
- en innretning til å styre maskinen
- en lagringsenhet til både mellomlagring og lagring av ferdig resultat
- en innretning til å automatisk utføre selve regneoperasjonene
- en skriverenhet som visualiserer det ferdige resultatet

Den dag i dag omfatter denne definisjonen i prinsippet de aller fleste datamaskiner.

Inspirert av Babbage's "Difference Engine", ble datamaskinen "Mark 1" bygget ved Harvard-universitetet i 1937-43, figur 2. Maskinen ble bygget av studenten Howard H. Aiken med finansiell støtte fra IBM, "International Business Machine Corporation", som senere skulle bli verdens største dataselskap. På grunn av 2. verdenskrig skulle den flere tonn tunge maskinen hovedsakelig brukes til å regne ut baner for kanonkuler. Da den ble lansert tok det imidlertid ikke lang tid før den var utdatert, i og med at de mekaniske komponentene den bestod av, var i ferd med å bli erstattet av elektriske.



Figur 2; Datamaskinen Mark 1, en utrolig stor regnemaskin i forhold til dagens standarder. Avanserte lommeregnerne kan i dag utføre samme operasjoner som Mark 1.

Bare 2 år etter lanseringen av Mark 1, ble maskinen "ENIAC" ferdigutviklet av John Mauchly og John Presper Eckert ved ingeniørhøyskolen i Pennsylvania. ENIAC, "Electronic Numerical Integrator and Calculator", var verdens første elektriske datamaskin, og med mulighet for å utføre 5000 operasjoner per sekund, var den over 1000 ganger raskere enn Mark 1, noe som tilsvarer 60-125 kHz. Til tross for forbedringen, var ENIAC fortsatt en førstegenerasjons datamaskin, som tok opp 170 kvadratmeter, veide hele 30 tonn og som hadde problemer med overoppheting.

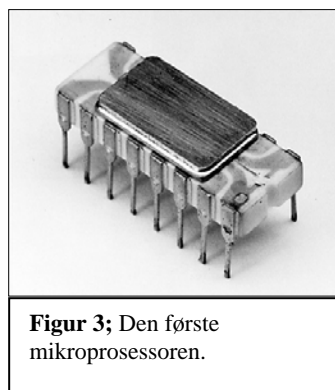
Transistoren og mikroprosessen

Med den nobelprisvinnende oppfinnelsen av transistoren, ved Bell Laboratories i USA i 1947, startet datarevolusjonen for alvor. I 1959 ble de første databrikkene utviklet ved Texas Instruments, noe som åpnet opp for en helt ny generasjon datamaskiner, basert på transistorer og integrerte kretser. Transistorene fungerte som på/avbrytere eller forsterkere til de elektriske kretsene i databrikken, og man kunne dermed "programmere" dem til å frakte med seg informasjon i form av elektrisk strøm. Førstegenerasjons-datamaskinene måtte benytte seg av såkalte *radiatorer* til å gjøre samme jobben, og disse tok opp en enorm plass i forhold til transistorene.

I 1964 ga IBM ut sin maskin "IBM 360", som var den første hvor kundene til en viss grad selv kunne velge maskinens spesifikasjoner. Dette året hadde prisen på en databrikke sunket fra 1000 dollar til 10 dollar på grunn av stadig nye teknologiske framskritt. Gordon Moore, en av stifterne av Intel Corporation, forutsa i 1965 at antallet transistorer på en integrert krets ville dobles hver 18. måned. Denne forutsigelsen fikk betegnelsen "Moore's lov", og har vært en bra tilnærming helt frem til i dag.

Neste generasjon datamaskiner kom med Ted Hoff's oppfinnelse av mikroprosessen i 1971, figur 3.

Året etter ga Intel ut sin første mikroprosessor, som var liten nok til å få plass i en fyrstikkeske, men samtidig kunne yte like mye som ENIAC. Med mikroprosessen kom også den første PC (Personal Computer); "MITS Altair 8800". I motsetning til sine forgjengere, som var beregnet på næringslivet og det offentlige, var denne maskinen rettet mot privatpersoner. På denne tiden ble også det første kommersielle dataspillet, "Pong", gitt ut.



Figur 3; Den første mikroprosessen.

Et par år senere lanserte Steven Jobs og Stephen Wozniak den første datamaskinen med farger; "Apple II", og Bill Gates & Paul Allen stiftet "Microsoft Corporation", eller "Microcomputer Software Corporation", figur 4.

Etter å ha inngått en svært gunstig avtale med IBM, utviklet Microsoft operativsystemet til deres første PC i 1981. Like etter ga de ut operativsystemet "Windows 1.0". Dette ble ingen stor suksess, men Windows-serien var grunnlaget for at de senere skulle bli den ledende leverandøren av operativsystemer.



Figur 4; Fra venstre: Paul Allen og Bill Gates, stifterne av Microsoft Corporation. Bill Gates er i dag en av verdens rikeste menn.

I 1993 ga Intel ut prosessoren "Pentium", og med sine 3,1 millioner transistorer, muliggjorde den krevende multimedieoperasjoner som film, musikk og avanserte spill. Tabell 1 under viser utviklingen til Intels prosessorer, som lenge var de ledende på markedet.

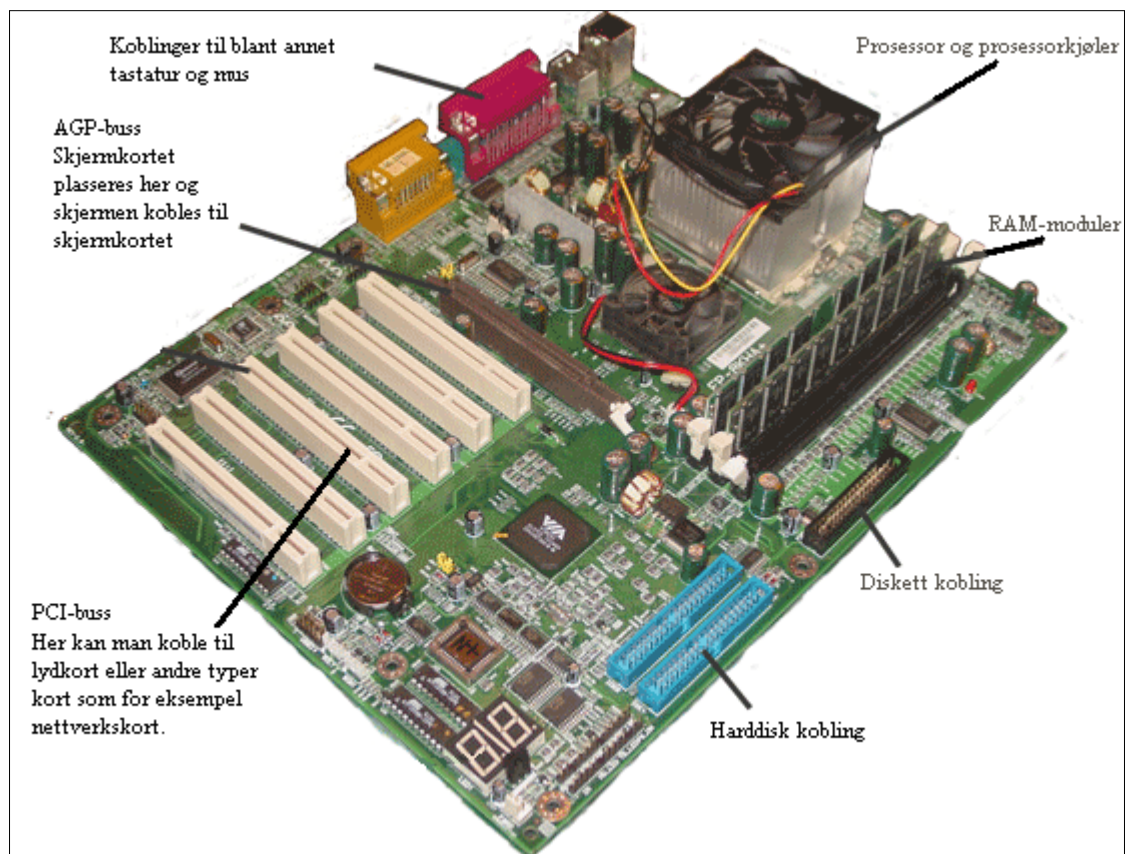
Som nevnt tidligere følger denne utviklingen Moores lov til en viss grad, og på grunn av den teknologiske utviklingen og konkurranse mellom bedriftene, kommer fremtidens datamaskiner til å bli både mindre, raskere og billigere.

Prosessortype	Lanseringsår	Antall transistorer	MIPS
8080	1974	6'000	0,64
8088	1979	29'000	0,33
80286	1982	134'000	1
80386	1982	275'000	5
80486	1989	1'200'000	20
Pentium	1993	3'100'000	100
Pentium 2	1997	7'500'000	300
Pentium 3	1999	9'500'000	510
Pentium 4	2000	42'000'000	1700
Pentium 4 "Prescott"	2004	125'000'000	7000

Tabell 1; Utviklingen av Intels prosessorer. MIPS = Millioner instruksjoner som kan utføres per sekund. Dette kan sees på som prosessorens totale ytelse.

Datamaskinens oppbygning [30-33]

Dagens datamaskiner kommer i veldig mange forskjellige størrelser og fasonger, men de har alle sammen likhetstrekk. ”Arkitekturen” i en datamaskin, hvordan komponentene skal settes sammen og hvordan de skal kommunisere med hverandre, har forandret seg enormt siden den første datamaskinen. Det er veldig vanskelig å ta for seg alle disse forskjellige arkitekturene, men vi skal ta for oss den mest vanlige per 2005.



Figur 5; Her kan man se et ganske vanlig hovedkort som har de fleste og de viktigste funksjonene som man finner i en datamaskin i 2005.

Det første man legger merke til når man ser en datamaskin er ofte skjermen, musen, tastaturet og kabinettet, se forside og figur 6. Skjermen er den delen som man bruker mest, det er den delen som viser oss all informasjon i form av tekst, bilde, video og alle andre former for visuell informasjon. Tastaturet og musen er de komponentene man bruker for å kommunisere med datamaskinen, uten et tastatur kunne vi aldri ha skrevet en tekst, og musen gjør ting utrolig mye lettere for oss når vi vil navigere mellom, for eksempel, programmer.

Kabinettet

Kabinettet er bare en metallkasse som beskytter alle de sensitive komponentene som en datamaskin består av. Man kan sammenligne kabinettet med menneskets kropp; huden fungerer på samme måte som kabinettet for den skal bare beskytte våre sensitive indre organer. Uten kabinettet kan de sensitive komponentene mye lettere utsettes for statisk elektrisitet og støv/skitt som til slutt kan ødelegge disse. Mange privatpersoner har som hobby å lage egne kabinett. Siden kabinetter bare har som funksjon å beskytte komponentene og samle disse på et sted, er de ikke særlig vanskelig å bygge. Det er ikke uvanlig at personer har lagd kabinett av små kommoder eller små kjøleskap! Kabinettet skjuler også en strømforsyning, en komponent som alle datamaskiner må ha. Denne strømforsyningen er laget spesielt for datamaskiner og kan derfor levere ganske stabile spenninger og strømmer, som også er nødvendig på grunn av de sensitive kretsene i for eksempel en prosessor. Alle komponentene i en datamaskin trenger elektrisitet for å fungere.



Figur 6; Kabinetter kommer i utrolig mange forskjellige farger og fasonger. Dette er et ganske vanlig kabinett; enkelt å oversiktlig.

Hovedkort og koblinger

Når man åpner et kabinett og ser på hva som befinner seg der inne, vil man kanskje først legge merke til hovedkortet. Dette er det største sammenhengende kretskortet som finnes i en datamaskin, og det har som oppgave å koble alle de andre komponentene sammen. Noen komponenter kobles direkte på hovedkortet, dette inkluderer prosessoren, RAM'en og diverse kort som for eksempel skjermkort og lydkort, se figur 5. Skjermen kobles ofte til skjermkortet som igjen er koblet til hovedkortet, men noen ganger kan skjermen også kobles direkte til hovedkortet hvis det finnes et integrert skjermkort der. Høytalere kan kobles til et lydkort som gjør at man for eksempel kan høre på musikk som spilles av i datamaskinen. Mus og tastatur kobles direkte til hovedkortet. Komponenter som harddisker og CD-stasjoner kobles til hovedkortet ved hjelp av diverse kabler som skal overføre data mellom disse og andre komponenter som sitter på hovedkortet. I realiteten er det "kabler", også kalt for "busser", mellom alle komponentene på et hovedkort og selve hovedkortet, men man velger å si at de komponentene som ligger veldig nært hovedkortet er koblet på direkte. Et eksempel på en "buss" er PCI-bussen, som man ser på figur 5, en kobling som gjør det mulig å feste på for eksempel et lydkort.

Formfaktor

Hovedkortet som man kan se på figur 5 er laget etter en bestemt arkitektur; den har en viss størrelse og inneholder en viss mengde av standard "buss"-typer, men hvordan de forskjellige hovedkortprodusentene plasserer disse "buss"-typene og andre tilkoblinger kan variere mye, poenget er at de holder en viss standard på arkitekturen, som for eksempel størrelsen på selve kortet. Noen standarder, også kalt for formfaktor, blant hovedkort er ATX, MicroATX, BTX, MicroBTX, PikoBTX, ITX, LPX og WTX. Formfaktoren man ser på figur 1 er av typen ATX og er mest vanlig i dagens datamaskiner. ATX formfaktoren kan ikke være større enn 305 * 244 millimeter stor.

Tastaturet [40-43]

Skrivemaskinen

Når man skal se nærmere på tastaturets historie, er det naturlig å begynne med skrivemaskinen. Den første mekaniske skrivemaskinen ble oppfunnet av amerikaneren Christopher Latham Sholes i 1867. I 1877 ble den forbedret og markedsført av Remington, USAs eldste våpenprodusent. På begynnelsen av 1900-tallet begynte skrivemaskinen å bli et vanlig verktøy i forretningslivet. Opprinnelig var tastene plassert alfabetisk, og intensjonen var at man skulle skrive kun med pekefingerne. Etter hvert som folk begynte å benytte seg av alle fingrene, økte skrivehastigheten betraktelig. Skrivemaskinen var ikke beregnet for slike hastigheter, og det endte ofte med at de mekaniske delene hang seg fast. Sholes løste dette problemet ved å arrangere tastene på en ulogisk måte, slik at tastene som ble brukt oftest, ble plassert lenger vekk fra sentrum av skrivemaskinen. Dette oppsettet fikk betegnelsen QWERTY, etter de seks første tegnene øverst på venstre side, figur 7, og ble umiddelbart tatt i bruk.



Figur 7; QWERTY-tastaturet.

Dvorak-oppsettet

August Dvorak forsket på ergonomi ved Universitetet i Washington, og i 1936 utviklet han et nytt oppsett på tastene, som hovedsakelig innebar at vokalene og konsonantene ble plassert på hver sin side. Dette gjorde at man oftere brukte annenhver hånd, noe som igjen medførte høyere skrivehastighet og bedre ergonomi. På denne tiden hadde skrivemaskinene en forbedret mekanikk, slik at de i teorien var i stand til å fungere også under høye hastigheter. Dessverre hadde QWERTY-oppsettet blitt så vanlig i forretningslivet at ingen turte å lansere noen ny skrivemaskin med Dvorak-oppsett.

Tastaturet erstattet raskt skrivemaskinen idet den kommersielle datamaskinen kom på markedet, og selv om det i dag er mulig å få tak i tastaturer med Dvorak-oppsett, benytter de aller fleste seg av QWERTY-tastaturet.

Tastaturets oppbygning

Et vanlig tastatur er bygd opp av rundt 105 forskjellige taster, avhengig av hvilket språk det tilhører. Man deler vanligvis tastene inn i fire forskjellige deler:

- alfabetet med tall og tegn
- funksjonstaster
- kontrolltaster
- numerisk del

Den numeriske delen av tastaturet har blitt plassert til høyre for å gjøre inntasting av numeriske data mer effektiv, og følger samme oppsett som en kalkulator. Funksjonstastene F1–F12 er plassert øverst på tastaturet, og tilegnes forskjellige kommandoer avhengig av hvilket program man kjører. Kontrolltastene brukes ofte til å navigere i dokumenter, slik som tastene "Home" og "End" som plasserer markøren på henholdsvis starten eller slutten av linja. Mange av de nyere tastaturene kalles *Windows-tastaturer*, siden de har to egne taster som er beregnet for å få frem menyer i operativsystemet Windows.

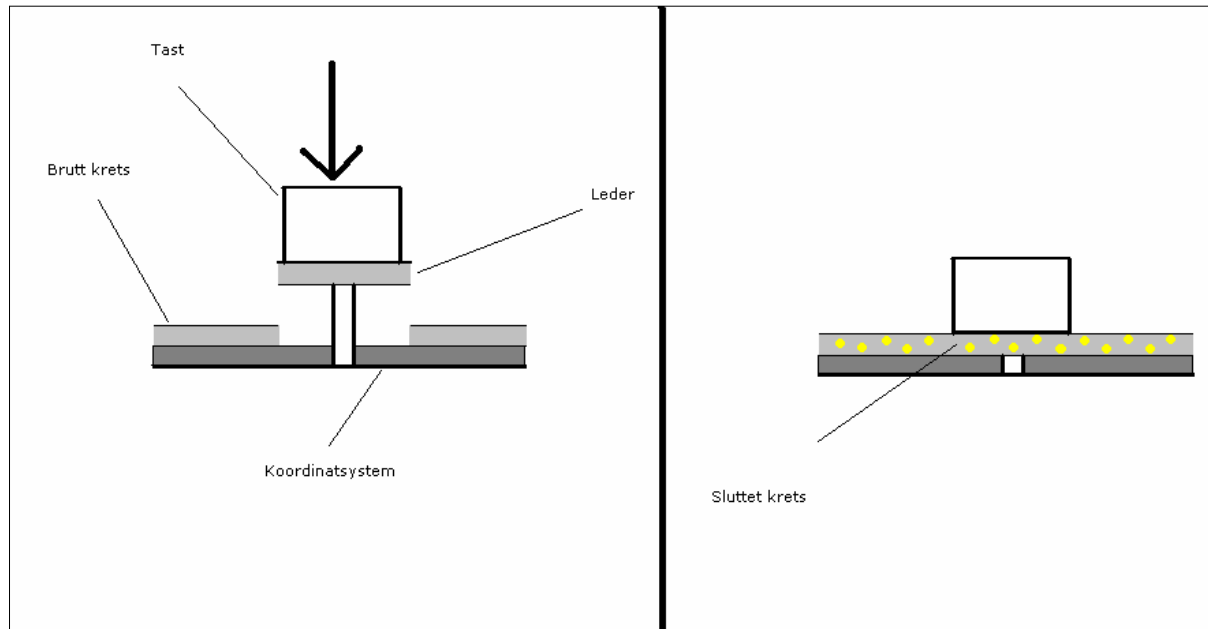
Når det gjelder tastaturets innvendige oppbygning finnes det flere typer, og hovedskillet går mellom mekaniske og kapasitansbaserte tastaturer. Felles for dem begge er at tastene er plassert oppå et koordinatsystem, noe som gjør at hver tast har sine helt unike (x, y)-koordinater, figur 8. Hvilke koordinater som hører til hvilke taster, ligger lagret i en minnebrikke av typen ROM, "Read-Only-Memory". I tillegg ligger det en mikroprosessor inne i tastaturet, noe som gjør at tastaturet egentlig er en slags datamaskin i seg selv.



Figur 8; Koordinatsystemet inne i et tastatur.

I mekaniske tastaturer, som er mest utbredt, er det plassert en liten brutt krets under hver eneste tast. Kretsen er brutt akkurat i punktet under tasten, slik at når tasten presses ned, vil en liten leder plassert på undersiden av tasten føre til at kretsen blir sluttet, figur 9. Prosessoren

registrerer at det går en svak strøm gjennom en bestemt krets i koordinatsystemet, og henter dermed informasjon ut fra ROM-minnet om hvilken eller hvilke taster som hører til denne kretsen.

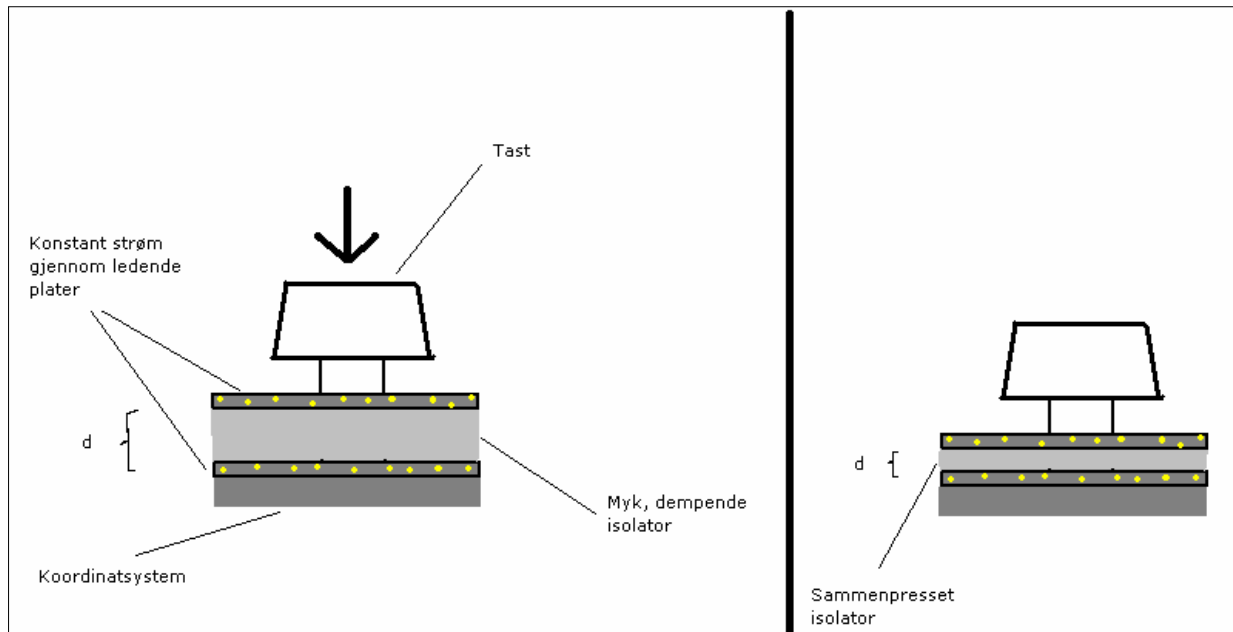


Figur 9; Før og etter et tastetrykk i et mekanisk tastatur

Det kapasitans-baserte tastaturet har fått navnet sitt siden hver tast er konstruert som en kondensator. I et kapasitansbasert tastatur går det en konstant strøm gjennom alle delene av koordinatsystemet. På undersiden av selve tastene sitter det fast en plate (leder) som beveger seg ned mot koordinatsystemet når man trykker ned tasten. En tilsvarende plate sitter fast i selve koordinatsystemet, mens en myk, dempende isolator hindrer platene i å komme i fysisk kontakt med hverandre. Hver enkelt tast vil dermed fungere som en kondensator, og kapasitansen vil øke i området under tasten når denne blir presset ned. Dette kan vi se siden kapasitansen C er omvendt proporsjonal med avstanden d mellom platene:

$$C = \epsilon_0 A / d,$$

(ϵ_0 er permittiviteten i vakuum og A er arealet til platene). C øker altså når d minker, se figur 10.



Figur 10; Før og etter et tastetrykk i et kapasitans-basert tastatur

På samme måte som med mekaniske tastaturer, er det prosessoren som merker at kapasitansen øker ved noen bestemte koordinater, og som sjekker med ROM-minnet hvilke taster disse tilsvarer.

Hvis man for eksempel trykker ned tasten "a", vil prosessoren registrere de tilhørende (x, y)-Koordinatene og "spørre" ROM-minnet om hvilken tast dette tilsvarer, for deretter å sende beskjed til datamaskinen om at det var "a"-tasten som ble trykket ned. Operativsystemet bestemmer dermed hvilken handling som skal utføres, avhengig av hvilket program man kjører. Kjører man et skriveprogram, vil operativsystemet gi beskjed om å skrive bokstaven "a", som dermed kommer opp på skjermen. Hvis man i det samme tilfellet vil skrive den store bokstaven "A", gjør man dette ved å holde nede "SHIFT"-tasten samtidig som man trykker ned "a"-tasten. Kombinasjonen "SHIFT" + "a"-tasten blir dermed sendt til operativsystemet, som bestemmer at dette tilsvarer en "A". Andre alternative kombinasjoner kan være systemkommandoer som "CTRL"+"ALT"+"DELETE", som skruer av maskinen og starter den på nytt.

Hvis man holder flere taster inne samtidig vil det generelt sett alltid være operativsystemet som bestemmer hvilken kommando dette tilsvarer, i forhold til hvilket program man kjører. Unntaket er hvis man holder inne for mange taster; prosessoren i tastaturet gir vanligvis ikke beskjed til operativsystemet om at flere enn 3 taster holdes inne samtidig, men dette antallet varierer fra tastatur til tastatur.

Fordeler og ulemper

På samme måte som en ball spretter hvis man slipper den i gulvet, vil en tast også "sprette" idet man trykker den ned, selv om man ikke merker det selv. Dette skaper problemer for et mekanisk tastatur, i og med at kretsen under tasten dermed vil sluttes og brytes mer enn én

gang per tastetrykk. Dette fenomenet forekommer i alle innretninger med mekaniske knapper, og er kjent som "bounce", eller byks på norsk. Hadde prosessoren i tastaturet ikke tatt hensyn til dette, ville det kommet opp flere hundre "a"-er, selv om man bare trykket ned tasten én gang. Kretsen i tastaturet inneholder derfor en "logisk enhet", som omformer signalet slik at prosessoren tolker det som ett trykk. I et kapasitansbasert tastatur er det ingen fysisk kontakt mellom platene som leder den elektriske strømmen, så "bounce" er ikke noe man trenger å ta hensyn til.

Mekaniske tastaturer blir lettere utsatt for slitasje enn kapasitans-baserte tastaturer, og har derfor kortere levetid. Til gjengjeld er kapasitans-baserte betydelig mye dyrere.

Gummikuppel-tastaturet

Gummikuppel-tastaturet, "the rubber dome keyboard", er det vanligste mekaniske tastaturet, og har fått navnet sitt på grunn av en føyelig gummikuppel med en hard karbonkjerne som sitter under hver enkelt tast. Når man trykker ned tasten, vil gummien bli presset sammen, og gi en lett motstand mot fingeren. På undersiden av tasten sitter den lille lederen festet til karbonkjernen, og slutter som nevnt kretsen så lenge tasten holdes nede. Idet man slipper opp tasten, vil gummien gå tilbake til sin opprinnelige form, slik at tasten dyttes tilbake til utgangspunktet. Det er med andre ord hardheten i gummien som bestemmer hvor hardt man må trykke på hver enkelt tast. Gummikuppeltastaturet er ikke like utsatt for korrosjon eller støv som andre mekaniske tastaturer, nettopp på grunn av den beskyttende gummikuppelen.

Datamusen [50-52]

Datamusens historie

I 1968 ble den første datamusen utviklet av Douglas Engelbart. Den skulle brukes på Stanford Research Institute i California, og kunne bare bevege seg i x- eller y-retning, altså ikke diagonalt. Treklossen med hjul fikk offisielt navnet "XY Position Indicator For A Display System", men dette ble raskt erstattet av det noe enklere kallenavnet "mus", på grunn av den halelignende ledningen som stakk ut fra enden. Før datamusen ble utviklet, brukte man piltastene på tastaturet til å navigere med, noe som var både upraktisk og tidkrevende. Likevel var folk lite interesserte i musen, siden de aller fleste dataprogrammene baserte seg utelukkende på tekst.

I 1970 ble datamusen sterkt forbedret, da Bill English erstattet de to retningshjulene med en ball som kunne bevege seg i alle retninger. Denne modellen holdt seg relativt uforandret i nesten 25 år, og ble en stor suksess.

Gjennombruddet kom likevel ikke før i 1984, da Apple lanserte datamaskinen "Macintosh", som hadde et av de første operativsystemene basert på "pek og klikk"-metoden. Noe senere, da

Microsoft ga ut Windows 3.1, et tilsvarende operativsystem til PC, var suksessen et faktum. Datamusen som ble brukt på Apples "Macintosh", hadde bare én knapp, noe som har vært maskinens kjennetegn helt frem til i dag. Microsoft baserte derimot operativsystemet sitt på at PC-musen skulle ha to knapper. I motsetning til Apples datamus, har PC-musen gjennomgått store forandringer når det gjelder antall knapper. Selv om de fleste PC-musene hadde tre knapper, ble det på slutten av 1990-tallet produsert mus med opptil fem knapper, noe som blant annet var praktisk til dataspill.

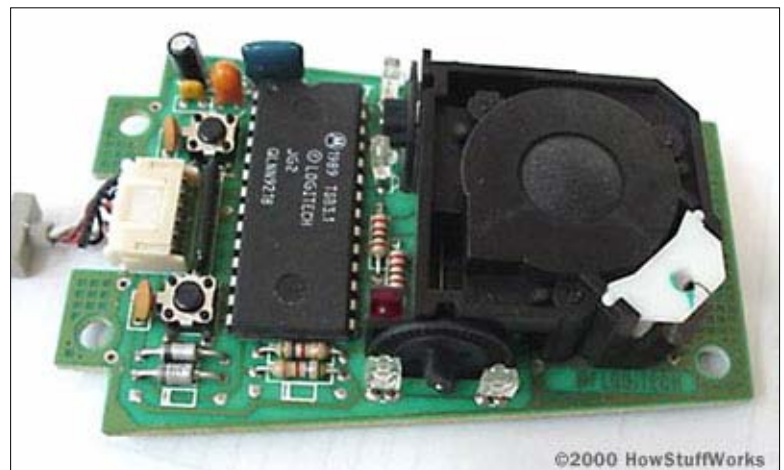
På slutten av 2000-tallet ble det såkalte "rullehjulet" oppfunnet. Dette ble plassert sammen med den midterste knappen, og hensikten var å kunne rulle på det uten å bevege mer enn en finger. Dermed kunne man bla gjennom dokumenter og sider på internett, uten å måtte bevege på hele armen.

Det hadde lenge vært en annen type datamus på markedet, som benyttet seg av optikk istedenfor mekanikk når det gjaldt manøvreringen. Denne måtte imidlertid brukes oppå en spesiell gittermønstrer musematte, og måtte holdes på en spesiell måte for å fungere. Den optiske musen ble derfor ingen stor suksess, før en kraftig oppgradert versjon kom i 1999. Siden den ikke trengte noen spesiell matte og var tilnærmet vedlikeholdsfri, overtok den raskt markedet og er i dag den desidert mest solgte datamusen. En datamus basert på laserteknologi er også å få tak i, men på grunn av den høye prisen er den ikke like konkurransedyktig.

Datamusens oppbygning

Den såkalte "optisk-mekaniske" musen bruker i likhet med den mekaniske en ball for å angi hvor den beveger seg, men har optiske sensorer som registrerer bevegelsen. Den mekaniske musen baserer seg utelukkende på mekanikk og er i dag lite brukt.

Når man beveger en optisk-mekanisk mus over en bordflate, vil ballen rulle langsmed flaten, figur 11, og samtidig bevege på to små hjul inne i musen. Disse hjulene er plassert slik at den ene registrerer bevegelse i x-retning og den andre i y-retning. To skiver med 36 hull jevnt fordelt rundt ytterkanten, er festet til hjulenes akslinger og beveger seg dermed i samsvar med hjulene. Begge skivene har to par optiske sensorer, som består av en infrarød LED, "Light Emitting Diode", og en infrarød sensor. Det er altså totalt fire LED'er og fire sensorer, figur 12. En LED er kort fortalt en liten lyspære som ved å bruke den samme teknologien som en transistor, varer mye lenger enn en vanlig lyspære. Siden LED'en og sensoren er plassert på hver sin side av skiven med 36 hull i, vil sensoren registrere om skiven beveger på seg og eventuelt hvor fort.

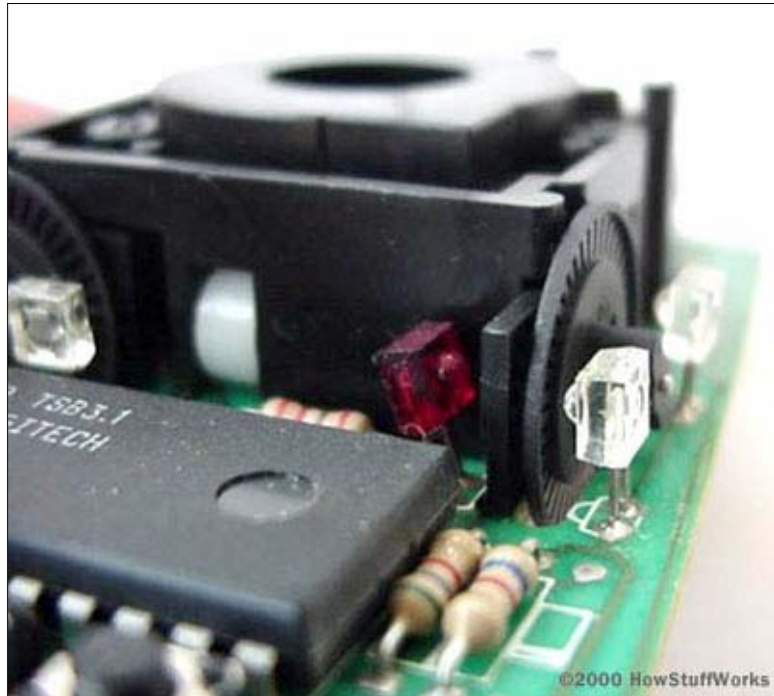


Figur 11; Her ser man tydelig ballen inne i en optisk-mekanisk mus.

I likhet med tastaturet har datamusen sin egen innebygde mikroprosessor. Denne mottar signaler fra de fire sensorene og kan ut i fra dette fastslå nøyaktig hvor fort og i hvilken retning

musen beveger seg. Deretter blir denne informasjonen og informasjon om eventuelle museklikk sendt videre til datamaskinen, som sørger for at musen beveger seg over skjermen.

Den forbedrede optiske musen fra 1999, benytter seg også av LED-teknologien, men på en litt annen måte enn den optisk-mekaniske. På undersiden av musen sender en LED ut et rødt lys, som reflekteres av bordflaten. Et lite kamera tar 1500 bilder av bordflaten per sekund ved å absorbere det reflekterte LED-lyset. En spesiell type mikroprosessor klarer å kjenne igjen forandringer i mønsteret på bordflaten, slik at den registrerer hastighet og retning på musebevegelsen. Som i den optisk-mekaniske musen blir deretter signalet om bevegelse og eventuelle museklikk sendt videre til datamaskinen.



Figur 12; På dette bildet ser man tydelig det ene paret av en hvit LED og en rød sensor på hver side av den hullete skiven.

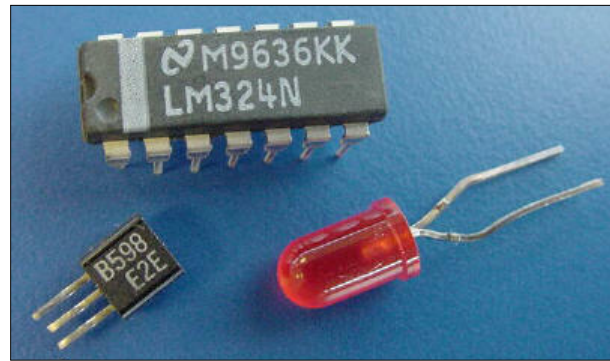
Den største fordelen med den optiske musen, er at den kan brukes uansett type underlag, fordi den ikke er avhengig av friksjon fra underlaget, slik som den optisk-mekaniske. Samtidig har den lenger levetid og blir heller ikke hemmet av støv, siden den ikke inneholder noen bevegelige deler. Den er riktignok noe dyrere, og enkelte spillentusiaster hevder at den ikke er like følsom som en mekanisk mus.

Behandling av data [59-77]

For at en datamaskin skal kunne kalles en datamaskin må den utføre en jobb, regne på noe, eller ”tenke” som noen liker å si. Til dette bruker maskinen en eller flere prosessorer, CPU, også kjent som en ”Central Processing Unit” på engelsk. En prosessor er i dag en integrert krets med mange hundre millioner transistorer, kapasitanser og andre typer kretskomponenter. Så å si alle komponentene i en datamaskin, alt fra skjermkort til RAM-minnet, er bygd opp av transistorer.

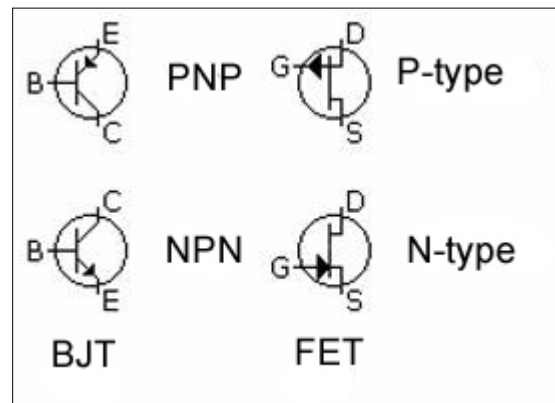
Transistorer

Transistor er satt sammen av to ord; "transfer" og "resistor". Siden patenteringen av transistoren i 1947 har utviklingen vært enorm. I dag regnes den som en av de største oppfinnelsene i moderne historie, sammen med trykkekunsten, bilen og telefonen. Grunnen til det er at den finnes i nær sagt all elektronikk vi omgir oss med, noe som igjen skyldes at den kan masseproduseres til forsvinnende liten pris per transistor. I dag lages det knapt enkelttransistorer, de er i stedet integrert i små kretser sammen med dioder, motstander og kondensatorer. Kostnadene knyttet til utviklingen av en slik integrert brikke, eller "Chip" på engelsk, kan være ganske høye, men fordi de deretter gjerne produseres i milliontall blir kostnadene fortsatt veldig lave. Mens de enkleste integrerte kretsene inneholder rundt 20 transistorer, kan en avansert prosessor i dag inneholde mellom ca 300 millioner og 1,7 milliarder transistorer!



Figur 13; Fra toppen og med klokken: en enkel integrert krets, en LED eller "Light Emitting Diode" og en enkelt transistor.

Det finnes to hovedkategorier av transistorer: "Bipolar Junction Transistor", BJT, og "Field Effect Transistor", FET. Felles for begge er at de har 3 koblinger, hvor spenningen på den ene inngangen, "base", styrer strømmen mellom de to andre, "emitter" og "collector", se figur 14. Transistorer kan brukes til å forsterke radiosignaler og lyden i et musikkanlegg, men i datamaskiner brukes de stort sett som elektriske brytere i digitale kretser. Faktisk er det på grunn av transistorenes, og dermed digitale datamaskiners, lave pris at vi ser en tydelig trend i å digitalisere informasjon, som f. eks for TV. Transistorer kan kategoriseres på mange forskjellige typen transistor, (BJT eller FET), polariteten, maksimum strømstyrke og/eller frekvens. Vi vil ikke gå inn på alle de forskjellige kombinasjonene og egenskapene her, men kun ta for oss de vanligste.



Figur 14; En skjematisk oversikt over hovedkategorien av transistorer, BJT og FET. Bokstavene B, E og C står for "Base", "Emitter" og "Collector"

Halvleder

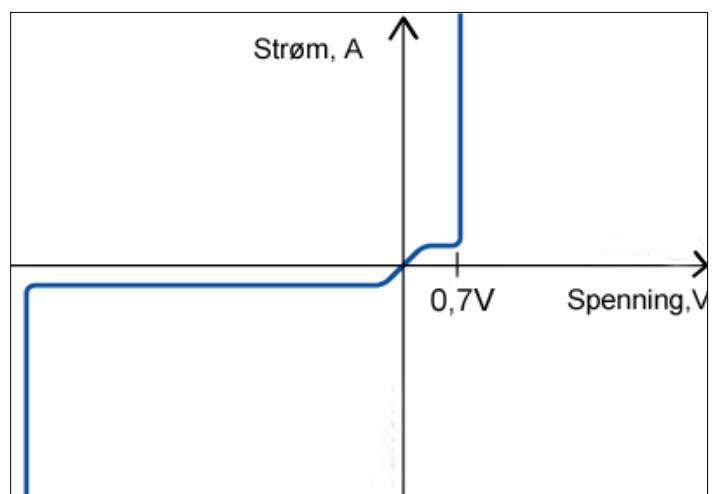
En transistor består i hovedsak av tre lag med halvledere. En halvleder er et materiale som leder strøm "halventusiastisk", det vil si at den befinner seg et sted midt i mellom en leder og en isolator. De første transistorene ble laget av germanium, (Ge), men i dag brukes mest silisium, (Si) eller "silicon" på engelsk. Silisium finnes ikke fritt i naturen, men det er likevel det nest mest vanlige grunnstoffet i jordskorpa. Hele 27,72 % av jordskorpa består av dette stoffet. Silisium finnes naturlig i kjemiske forbindelser med mange andre grunnstoffer, og opptrer blant annet som stoffet silisiumoksid, som vi kjenner som sand, kvarts og bergkrystall. Sammen med germanium og karbon, har silisium en unik egenskap i elektronstrukturen; det har fire

elektroner i det ytre elektronskallet og kan derfor forme fine krystaller. Silisium i denne formen har bare perfekte kovalente bånd og fungerer nesten som en isolator.

Ved å dope silisiumet kan man derimot gjøre den ledende. Med doping menes, som i sportsverden, å tilsette litt av et annet stoff. Dersom man tilsetter fosfor, (P), eller arsen, (As), får man n-type doping med overskudd av elektroner og dermed negativ ladning. De frie elektronene gjør at silisiumet nå kan lede strøm. Positiv, p-type, doping får man ved å tilsette bor, (B), eller gallium, (Ga). Det fører til et overskudd av hull, det vil si underskudd av elektroner, og igjen har isolatoren blitt til en elektrisk leder. Det er viktig å presisere at silisium ikke blir en god leder av doping, derfor kalles det halvleder.

Dioder

Dersom en n-type halvleder og p-type halvleder settes sammen får man en diode. Alle de frie elektronene i n-delen vil da legge seg inntil p-delen, og omvendt for hullene i p-delen. Dette fører til et ganske interessant fenomen; nå kan nemlig dioden bare føre strøm den ene veien. Dersom man prøver å føre strøm fra n-typen til p-typen, vil kun en veldig liten strøm passere, men dersom spenningen snus slik at strømmen går fra p-typen til n-typen fungerer dioden som en leder. Det skyldes at hullene da blir trukket mot den andre, negative siden av p-delen og elektronene mot den positive på motsatt side. Det trengs riktig nok en liten spenning for å løsrive hullene og elektronene fra hverandre, for silisium er denne på ca. 0,7 V. Dersom man setter en stor nok spenning "feil vei" over dioden vil den vil den naturligvis kollapse og lede strøm. Dette kan sees i figur 15.



Figur 15; En skjematisk beskrivelse av egenskapene til en diode, strøm som funksjon av spenningen.

"Bipolar junction transistor", BJT

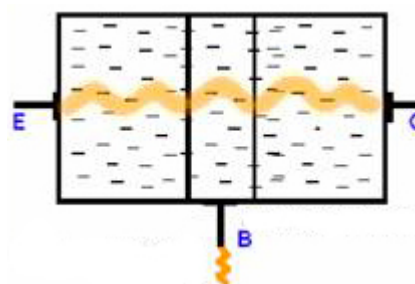
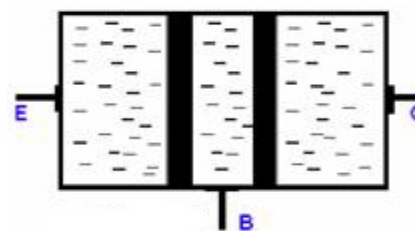
En transistor består av tre lag med halvledermateriale istedenfor to. Det kan da enten være npn eller pnp, som uansett blir to dioder som står "rygg mot rygg". Da vil strømmen blokkeres begge veier, men ved å tilføre en liten strøm til den midterste komponenten kan en større strøm gå igjennom hele transistoren. Forholdet mellom disse er vanligvis rundt 100. Man sier ofte at transistorer er strømstyrt, men i realiteten styres de av spenningen mellom "basen" og "emitteren". "Emitterstrømmen" er gitt ved Ebers-Moll formel som:

$$I_e = I_{es} \left(e^{\frac{V_{be}}{V_T}} - 1 \right)$$

, hvor I_e er emitter strømmen, I_{es} er den inverse av maksimalstrømmen for "base-emitter" dioden, V_T er spenningen gitt av temperaturen og V_{be} er base-emitter spenningen. Det finnes også transistorer som styres av lys og disse kalles fototransistorer. Faktisk vil også vanlige transistorer kunne påvirkes av lys.

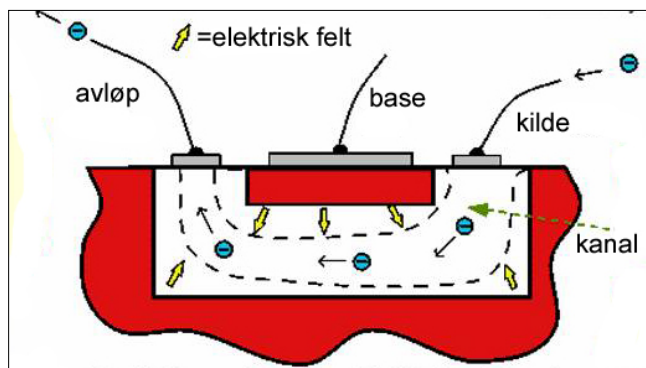
"Field-effect transistor", FET

På slutten av 1960-tallet gikk produsentene av transistorer over fra BJT til FET og i dag består de fleste integrerte kretser og chips av FET. Navnet kommer av at det svake elektriske signalet er koblet til en elektrode som skaper et elektrisk felt i resten av transistoren. FET blir også kalt unipolare fordi de bare bruker enten hull eller elektroner som ladningsbærere. Når det svake elektriske signalet snur fra positivt til negativt snues også feltet i transistoren og er det ikke noe signal er det heller ikke noe elektrisk felt, figur 17.



Figur 16; En skisse på hvordan strøm kan opprettes mellom E og C ved å forandre B!

Den mest brukte typen FET i dagens PC-er er metalloksid FET, MOSFET. Den ser ut som en litt flattrykket U, der selve U-en er en n-type halvleder. Området rundt kalles base og er av en p-type halvleder. Hvilken type halvleder som er hvor er egentlig ikke så viktig, forskjellen er at det vil være hull som er ladningsbærere istedenfor elektroner. Den enden av U-en hvor elektronene kommer inn, kalles kilden og den andre avløpet. Strømmen går altså gjennom U-en, men på grunn av halvlederens egenskaper vil en del av elektronene være "bundet" opp mot kanten. Strømmen går derfor bare gjennom en kanal i sentrum av U-en. Dersom det nå påtvinges en spenning på basen forandres dette. Dersom det tilføres elektroner vil elektronene i U-en bevege seg



Figur 17; En skjematisk oversikt over en FET

lenger unna basen å minske arealet der strømmen kan passere helt til den er helt blokkert. Dersom basen tilføres en positiv ladning vil arealet derimot bli større. En slik transistor er derfor veldig godt egnet som analog forsterker, så vel som digital bryter.

Formelen for avløps-strømmen er gitt ved:

$$I_d = I_{dss} \left(1 - e^{-\frac{V_{gs}}{V_p}}\right)^2$$

der I_d er avløps-strømmen, I_{dss} er avløps-strømmen ved null spenning over basen, V_{gs} er spenningen over basen og V_p er spenningen som skal til for å blokkere strømmen.

Mikroprosessor

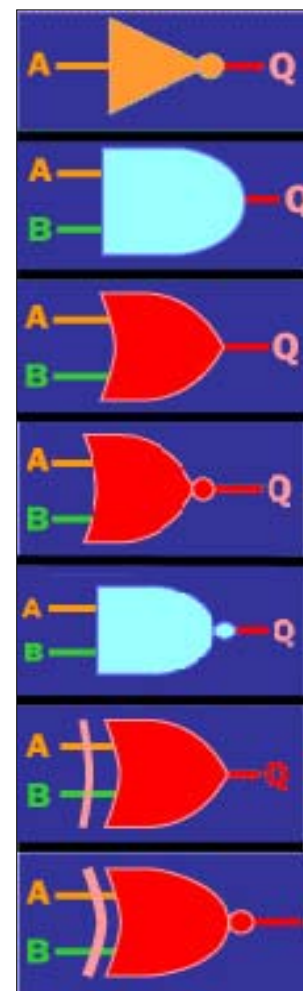
Uansett hva slags PC du måtte bruke har den en prosessor som er selve hjernen i maskinen. Siden 70-tallet har denne vært integrert i én chip og kalt mikroprosessor eller ground processing unit (CPU). Med en chip menes en liten, tynn brikke av silikon med en integrert krets av motstander, kondensatorer og ikke minst transistorer. En CPU kan være to-tre centimeter lang og kan inneholde flere titals millioner transistorer, eller noen tusen transistorer på noen få kvadratmillimeter.

For å gjøre noe veldig komplisert enkelt kan man si at en CPU er instruert til å gjøre tre ting:

- Utføre matematiske operasjoner som addisjon, subtraksjon, multiplikasjon og divisjon ved å bruke sin Arithmetic/Logic Unit, ALU.
- Flytte data fra en lagringsenhet til en annen.
- Ta en avgjørelse for deretter å hoppe til en ny oppgave basert på den avgjørelsen.

En mikroprosessor kan gjøre langt mer kompliserte ting enn dette, men dette er de mest grunnleggende oppgavene.

Ved hjelp av transistorene og boolsk logikk utfører CPU'en tre, fem eller syv enkle operasjoner som igjen kan settes sammen til mer avanserte operasjoner. Om det er tre, fem eller syv operasjoner som ligger til grunn avhenger av hvordan man ser det, de fire siste bygger nemlig på de tre første. Disse består av et par transistorer som danner en såkalt port og denne porten kan forandre signalet avhengig av hvordan transistorene er koblet. Det digitale signalet kan forandres i en av følgende tre type porter:



Figur 18; En skjematisk oversikt over porter.

- ”IKKE”: En ”ikke-port” er en som inverterer signalet slik at utgangssignalet er det motsatte av inngangssignalet. 0 blir til 1, og 1 blir til 0.
- ”OG”: En ”og-port” følger logisk av ordet og; dersom både A og B er 1 vil også Q være 1.
- ”ELLER”: En ”eller-port” gir 1 dersom A eller B (eller begge to) er 1.

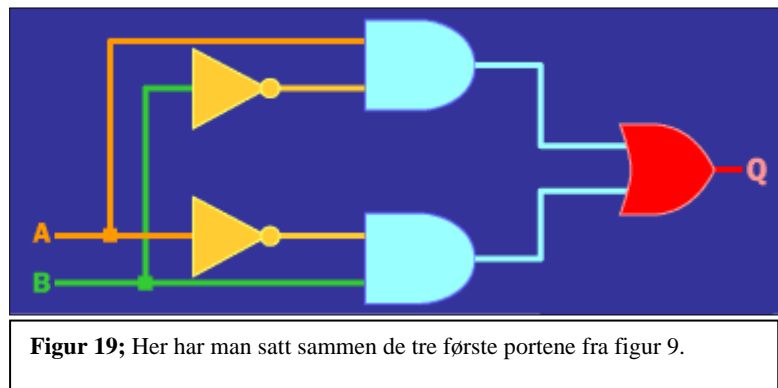
Det er de tre grunnleggende portene, men det er vanlig å ta med to til, nemlig:

- ”HVERKEN”: Denne porten er en kombinasjon av ”ELLER” og ”IKKE”, og gir kun 1 dersom hverken A eller B er 1.
- ”IKKE BEGGE”: Denne er en kombinasjon av ”OG” og ”IKKE”, og gir 1 unntatt når A og B er 1.

I tillegg til dette tar noen også med:

- ”ENTEN”: En ”enten-port” er en som gir 1 kun når enten A eller B er 1, ikke begge.
- ”LIKE”: Denne porten gir 1 kun når A og B er like.

Grunnen til at disse ofte ikke taes med er at de lett kan settes sammen av de tre første fra figur 18, som vist på figur 19. Disse portene er faste fra produksjonen av, hvilke porter som skal brukes og i hvilke rekkefølge bestemmes av CPU'en. På den måten kan den ved til dels kompliserte kombinasjoner regne og ”tenke”. Men addisjon krever for eksempel bare en ”enten-port”.



Figur 19; Her har man satt sammen de tre første portene fra figur 9.

Tabell 2 viser en del spesifikasjoner for mikroprosessorer, samt utviklingen av Intels mikroprosessorer gjennom tidene.

- Datoen forteller hvilket år prosessoren ble lansert.
- Transistors viser antallet transistorer i brikken.
- ”Microns” er bredden på den minste lederen i brikken, målt i mikrometer. Til sammenligning er et hårstrå ca 100 mikrometer tykt.
- ”Clock speed” er maksimum klokke hastighet, som vil si den maksimale hastigheten på prosessorens ”puls”.
- ”Data width” er databredden, eller hvor mange ”bit” ALU'en kan håndtere samtidig. En ekstern data buss kan som vist hente data med en annen bredde enn den prosessorens ALU opererer med.
- MIPS står for millioner av instruksjoner per sekund og er et tilnærmet tall for hvor godt en prosessor fungerer.

Name	Date	Transistors	Microns	Clock speed	Data width	MIPS
8080	1974	6,000	6	2 MHz	8 bits	0.64
8088	1979	29,000	3	5 MHz	16 bits 8-bit bus	0.33
80286	1982	134,000	1.5	6 MHz	16 bits	1
80386	1985	275,000	1.5	16 MHz	32 bits	5
80486	1989	1,200,000	1	25 MHz	32 bits	20
Pentium	1993	3,100,000	0.8	60 MHz	32 bits 64-bit bus	100
Pentium II	1997	7,500,000	0.35	233 MHz	32 bits 64-bit bus	~300
Pentium III	1999	9,500,000	0.25	450 MHz	32 bits 64-bit bus	~510
Pentium 4	2000	42,000,000	0.18	1.5 GHz	32 bits 64-bit bus	~1,700
Pentium 4 "Prescott"	2004	125,000,000	0.09	3.6 GHz	32 bits 64-bit bus	~7,000

Tabell 2; Denne tabellen viser oss den generelle utviklingen av prosessorer over ca 30 år.

Som tidligere nevnt er det en enorm oppgave å designe en chip. Det er veldig viktig at transistorene og portene settes sammen på den riktige måten for å få mikroprosessoren til å yte maksimalt. På fagspråket kaller man dette prosessorarkitektur. Jo flinkere arkitekten er, jo bedre yter prosessoren i forhold til de gitte spesifikasjonene som størrelse, antall transistorer osv. Ironisk nok er det mikroprosessorer i form av andre datamaskiner som gjør mesteparten av jobben med å regne ut hva som er den mest hensiktsmessige arkitekturen for nye mikroprosessorer.



Figur 20; En moderne Intel Pentium 4 prosessor, en integrert krets.

Primærlagring [80-87]

En datamaskin kan sammenlignes med menneskenes hjerne på mange områder, og et av disse områdene er minnet; en datamaskin må kunne "huske" ting, ellers ville vi aldri ha kunnet skrive dokumenter, eller kjørt flere programmer samtidig. På samme måte som vi mennesker har korttidshukommelse og langtidshukommelse, har en datamaskin minnetyper som kan lagre data i korte eller lengre tidsintervaller. En harddisk er et eksempel på "langtidshukommelsen" og er laget slik at vi kan lagre dokumenter, musikk eller andre datafiler i mange år uten at det blir borte. CD-plater, DVD-plater og disketter er også laget med det samme formålet. Som "korttidshukommelse" bruker datamaskinen forskjellige typer RAM, hurtigbufferer og ROM, og disse minnetypene er viktigere enn, for eksempel, en harddisk for at datamaskinen, som vi kjenner den til daglig, skal fungere skikkelig.

RAM

RAM står for "Random Access Memory" og er det minnet der alt av programmer som man kjører ligger i. Det finnes mange typer RAM-minne, men den viktigste i datamaskiner i dag er DRAM, Dynamic Random Access Memory. Denne minnetypen er ikke integrert på hovedkortet, og må kjøpes separat hvis man bygger en datamaskin selv. Denne minnemodulen kan se ut som den på figur 21. En slik minnemodul er bygd opp av flere mindre minnebrikker. Det er de svarte rektangulære formene på bildet. Før disse minnemodulene ble vanlig, festet man selve minnebrikkene på hovedkortet. Men etter hvert fikk man plassproblemer fordi minnemengdene ble så store. Derfor ble minnemodulene oppfunnet. Nederst på figur 21 kan man se gullfargede "pins", kontaktene som man fester til hovedkortet og som kommuniserer med resten av datamaskinen.

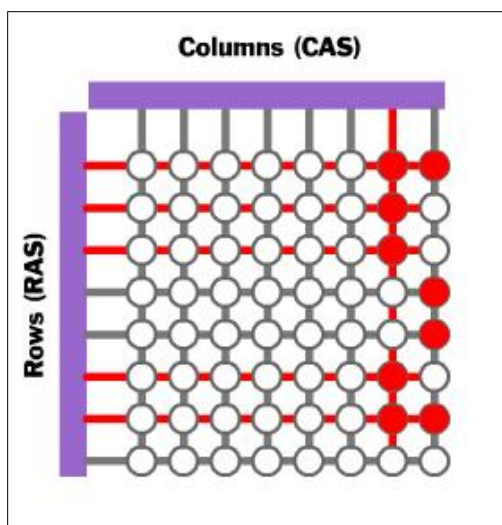


Figur 21 : her er et bilde av en vanlig DRAM-modul som de fleste datamaskiner i dag bruker.

"Random" refererer til et prinsipp i minneteknologien som gjør at prosessoren kan hente/skrive data fra et hvilket som helst sted i minnet. Dette er i motsetning til SAM, "Serial Access Memory", der data bare kan hentes/skrives i en bestemt rekkefølge. Fordelene med RAM'en er derfor klar; hvis ikke prosessoren kunne lese/skrive i hvilken som helst rekkefølge ville det ta mye lengre tid, og datamaskinen ville oppføre seg mye tregere. Hvis man ikke hadde RAM/SAM i det hele tatt, men brukte harddisken som primærminne, ville datamaskinen ha blitt enda tregere. Hvis man skriver et dokument i "word" og hører på musikk samtidig gjennom et musikkprogram, vil faktisk begge disse ligge i RAM'en til man stenger programmene. Da vil prosessoren skjønne at man har stengt de, og vil fjerne de fra minnet slik at man har ledig minne til andre programmer. Operativsystemet som man kjører, for eksempel Microsoft Windows, vil alltid ligge i dette minnet, siden det i realiteten er et ganske avansert program som man alltid bruker når datamaskinen er slått på. Med en gang man slår av datamaskinen, vil også all digital informasjon som ligger i RAM'en forsvinne. Neste gang man slår på datamaskinen vil prosessoren "hente" operativsystemet fra harddisken og legge det i primærminnet/RAM'en.

RAM'en er, på samme måte som en prosessor, en integrert krets som består av millioner eller milliarder av transistorer og kapasitanser som opererer i par. Hvert av disse parene vil fungere som en minnecelle, altså at den kan holde en "bit" av data, enten en 1'er eller en 0'er. Resistoren fungerer som en bryter, mens kapasitansen er den som holder en "bit". Hvis for eksempel kapasitansen er full av elektroner, så vil dette representere en 1'er, men hvis den er tom vil den representere en 0'er. I virkeligheten er det ikke så svart/hvitt; kapasitansen kan holde en 1'er så lenge den er mellom 50-100 % full av elektroner. Problemet med disse kapasitansene er at de mister elektronene meget fort, en full kapasitans mister nesten alle elektronene på noen få millisekunder. Det er på grunn av dette at all informasjon fra primærminnet vil forsvinne når man slår av datamaskinen! Derfor må RAM'en oppdateres flere tusen ganger i sekundet for å holde på de samme data over lengre tid mens datamaskinen er slått på og det her herfra ordet "Dynamic" kommer inn. Dynamisk minne er rett og slett minne som må oppdateres jevnlig for at datamengder ikke skal bli borte.

Hvordan kan prosessoren vite hvor den skal hente datamengder fra primærminnet? Minnecellene som ble beskrevet ovenfor blir nemlig ordnet i rader og kolonner. Når en kolonne og en rad skjærer hverandre får man en minneadresse. For eksempel hvis en minnecelle ligger på rad 2 og kolonne 3, vil denne minnecellen ha en spesifikk adresse. Det er denne adressen prosessoren går etter når den vil vite hva slags "bit" minnecellen holder. Kolonnene kalles for CAS, Column Access Strobe, og radene for RAS, Row Access Strobe. Se figur 22. Hvis man vil gi to minneceller, en i CAS 2/RAS 1 og en i CAS 2/RAS 4, sender man strøm gjennom CAS 2 først, så sender man strøm i RAS 1 og RAS 4 samtidig. Da vil bare disse to minnecellene få ladning verdien 1, altså "på". Det er transistorene i hver minnecelle som blir aktivert når strøm blir sendt gjennom en kolonne, så blir strøm sendt gjennom de radene som er ønskelig, og kapasitansene i de minnecellene som skal ha verdi 1 eller "på" blir ladet opp. Minnecellene opererer ikke alene, de trenger "hjelp" fra mange forskjellige kretser som har spesifikke oppgaver. En krets har ansvaret med å lade opp hver minnecelle som trenger det mange tusen ganger i sekundet, mens en annen krets har som oppgave å fortelle prosessoren hvilken minnecelle som har en 1'er og hvilken som har en 0'er.



Figur 22; Her en skjematisk oversikt over hvordan minnecellene er ordnet i rader og kolonner. Rødt er 1 eller "på", og hvitt er 0 eller "av"

Når en prosessor vil ha data fra minnemodulen, må den sende et signal til minnekontrolleren, som er en av disse kretsene med spesifikke oppgaver. Så må minnekontrolleren sjekke om det er klart for lesing av data på minnebrikkene før den sender et signal tilbake til prosessoren om at det er klart og at den nå kan begynne å lese av datamengdene. Denne prosessen defineres ofte som farten på en minnemodul og måles i frekvens eller aksessetid. Denne tiden er på moderne minnemoduler ofte lavere enn 50 nanosekunder. En vanlig minnemodul per 2005 har ca 512 Megabytes med plass. Siden det er 8 bits i en byte:

$$512 * 10^6 * 8 = 4\,096\,000\,000$$

Altså 512 Megabytes er ca 4,1 milliarder bits, og siden det er en transistor og en kapasitans som representerer en bit, så får vi ca 4,1 milliarder transistorer og kapasitanser på

minnemodulen! Når vi tenker på størrelsen på en minnemodul så er dette ikke så rart; en moderne prosessor i dag vil ha ca 200-700 millioner transistorer på den integrerte kretsen, og en minnemodul er ca 6-7 ganger større i areal enn en prosessor.

Noen ganger kan det skje feil; en minnecelle kan bli påvirket av en magnetisk eller elektrisk kraft, og på grunn av dette kan den skifte verdier, fra 1 til 0 eller omvendt. For å forhindre at prosessoren tar i mot feil data fra minnet, har man opprettet feilsjekkingsrutiner i dagens DRAM brikker. Som man kan se på figur 22, er minnecellene ordnet i 8 stykk per rad/kolonne. Ved å legge til en ekstra rad er det mulig å sjekke om hver kolonne har riktige verdier. Når prosessoren skriver data til minnet, summerer den opp alle de første 8 minnecellene i hver kolonne. Hvis denne summen blir et oddetall vil den siste minnecellen i den respektive kolonnen få verdien 1, og motsatt hvis summen blir et partall. Når prosessoren skal lese fra minnet på et senere tidspunkt leser den av den siste minnecellen først, for å sjekke om summen av de første 8 ble et oddetall eller partall. Så summerer den opp de første 8 minnecellene på nytt igjen, og hvis denne summen ikke stemmer med den siste minnecellen kvitter prosessoren seg bare med de ukorrekte kolonnene, for så å prøve å skrive disse på nytt. Slike minnefeil er egentlig sjelden i dagens minnebrikker, men systemer som virkelig trenger et stabilt minne har spesialbygde feilsjekkingsrutiner.

Det finns mange typer DRAM-brikker; FPM DRAM, EDO DRAM, SDRAM, DDR SDRAM og RDRAM er noen. De fleste datamaskiner i dag bruker SDRAM eller DDR SDRAM. Dette er bare varianter av den første DRAM-brikken, og forskjellene er relativt små; noen varianter opererer på høyere fart, er synkronisert med andre komponenter i datamaskinen eller er laget for å brukes som videominne på for eksempel skjermkort.

Hurtigbufferne

Hurtigbufferne er mest kjent som "cache" på engelsk. Man finner disse i mange komponenter i en datamaskin, men de viktigste er nok hurtigbufferne i prosessoren og i harddisken. Formålet er å gjøre datamaskinen enda raskere i å "huske" enn det den hadde vært uten hurtigbufferne. Det er ganske billig å gjennomføre og lage, og denne teknologien har derfor blitt meget viktig i en datamaskin i dag.

Prinsippet går ut på at man beholder datamengder som man bruker ofte i små minnebrikker integrert på prosessorbrikken eller på harddiskens kretskort. Det viser seg nemlig statistisk sett at datamaskinen, eller programmene datamaskinen, bruker en spesifikk type data oftere enn andre typer data. Dette er ofte små algoritmer eller deler av programmer som ligger i operativsystemet, siden disse brukes jevnlig uten at brukeren av datamaskinen vet det. Dette kan være programmer som har ansvaret for at internettforbindelsen er stabil eller kanskje rett og slett et skriveprogram. Man kan teste ut prinsippet av hurtigbufferne i en datamaskin hvor som helst; hvis man bruker en diskett som inneholder en "word"-fil, og åpner denne filen for første gang siden man startet datamaskinen, vil det ta ganske lang tid. Men hvis man lukker denne filen etter at den har blitt åpnet og så åpner den samme filen på nytt igjen, vil det ta mye kortere tid. Datamaskinen har da brukt en hurtigbuffer på denne filen fordi den trodde at den kanskje ble åpnet på nytt igjen. Hurtigbufferen i dette eksempelet var RAM'en som ble beskrevet tidligere. Prosessoren har integrert veldig små mengder av veldig dyrt minne. En måte å klassifisere primærminnet på er etter "pris per byte". Hvis prisen er høy, er minnet sannsynligvis meget raskt, og motsatt hvis prisen er lav. Man kan i prinsippet utstyre en datamaskin med store mengder av det dyreste minnet, men dette blir for dyrt for forbrukerne.

Derfor bruker man det dyreste minnet i meget små mengder i form av hurtigbufferne, og det dyreste finner man integrert på prosessoren. Disse hurtigbufferne er laget slik at prosessoren kan kjøre/regne gjennom en kode, som inneholder en datamengde som gjentar seg ofte, meget raskt.

ROM

ROM står for "Read-Only Memory", prosessoren kan altså bare lese av en ROM-brikke og ikke skrive noe på den, slik som den kan med RAM-minnet. ROM-brikker har derfor ganske lite med lagringsplass og inneholder som oftest informasjon om en komponent som prosessoren trenger å vite for at den skal fungere på riktig måte. Komponenter som har en ROM-brikke integrert kan være tastatur, mus, hovedkort, skjermkort og lyd kort, figur 23.



Figur 23; Her er en ROM-brikke som finnes i alle datamaskiner i dag på hovedkortet. Den har navnet "BIOS"

ROM-brikken er laget på samme måte som en RAM-brikke når en tenker på rader og kolonner, og at det er en enkelt minnecelle per "bit". Prosessoren kan også lese av data på samme måte som i RAM, men ROM-brikken bruker ikke en transistor og en kapasitans per minnecelle, men i stedet en diode. I tillegg vil ikke en ROM-brikke miste noe av sin informasjon når strømmen i en datamaskin slås av. Hvis man ser på figur 22, så kan man se radene, kolonnene og minnecellene som lå akkurat der radene og kolonnene krysset hverandre. Hvis en minnecelle i en ROM-brikke skal holde en "bit" med verdi 1, har man rett og slett plassert en diode der. Hvis man vil vite hva slags verdi en minnecelle i rad 1/kolonne 2 har, setter man bare rad 1 til "jord" og sender en strøm gjennom kolonne 2. Hvis det befinner seg en diode i denne minnecellen vil strømmen sendes til "jord" og minnecellen vil bli oppfattet som en 1'er. Hvis det ikke befinner seg en diode i denne minnecellen, vil ikke kolonnen og raden være koblet sammen, og strømmen vil dermed ikke gå gjennom til "jord".

ROM-brikker er altså designet til å holde en viss datamengde som bestemmes av selve maskinvaren. Å lage en ROM-brikke er derfor ganske tidkrevende, hver enkelt diode må plasseres helt nøyaktig i den minnecellen som skal ha verdien 1, og ikke i noen av minnecellene som skal ha verdi 0. Produksjonen av slike brikker består derfor ofte av en "prøv og feil" periode der mange brikker og oppsett kastes til fordel for en som fungerer bedre, men når designet av brikken er i boks tar det veldig kort tid å lage mange slike brikker. ROM-brikker er meget billig i produksjon, og det er nok derfor man finner forskjellige varianter av den i alt som trenger en bestemt datamengde som skal brukes om igjen mange ganger. I leker for barn kan man finne slike ROM-brikker som inneholder musikk/lyd eller kontrollkoder for en bevegelse.

Sekundær lagring [90-96]

Sekundær lagring er et uttrykk som brukes for alle typer å lagre data på som er vesentlig tregere enn for eksempel RAM, eller de primære lagringsenhetene. Sekundære lagringsenheter omfatter blant annet cd, dvd, datatape, diskett og harddisker. Disse kan igjen klassifiseres i 2 grupper; magnetisk lagring og optisk lagring.

Magnetisk lagring

Magnetisk lagring noe av det viktigste som finnes når en tenker på hvordan vi har lagret data de siste 100 årene og revolusjonen innen digital informasjon som vi opplever i dag. Uten magnetisk lagring hadde kanskje aldri informasjonsalderen slik vi kjenner den funnet sted.

Mange forskjellige komponenter bruker prinsippene i magnetisk lagring; harddisker, forskjellige disketter, videotape, lyd-kassetter og gamle datataper. Sammenhengen mellom disse er at de bruker en eller annen slags manipulator som forandrer det magnetiske substratet som ligger på en plastisk film eller en plate/skive. Dette kan være digital informasjon eller lyd/video. Den samme manipulatorene er også en sensor, og kan måle endringene som den selv har laget på det magnetiske substratet. Dette gjøres så om til lyd/video eller en annen form for data.

Det magnetiske substratet er et ferromagnetisk materiale, ofte jernoksid. Hvis man utsetter dette materialet for et magnetisk felt vil det være magnetisert i lang tid hvis det ikke utsettes for et nytt magnetisk felt. Derfor mister ikke disketter, harddisker og andre komponenter informasjonen man har lagret på disse hvis de ligger ubrukt over lang tid. Hvis man vil forandre informasjonen som ligger på for eksempel en diskett, kan man bare forandre det magnetiske feltet igjen, og dermed substratet, og det er dette som gjør at vi kan lagre og slette data mange ganger om igjen. Det er manipulatorene som lager dette magnetiske feltet når noe blir skrevet på en diskett eller på andre magnetiske lagringsmedium. Grunnen til at slike lagringsmedier har blitt så populære er at de er veldig billige å lage og enkle i bruk.

Prinsippet er enkelt: når manipulatorene, ofte kalt et skrive/lese hode, beveger seg over det magnetiske substratet, forandrer den hvilken retning de magnetiske polene på en samling av små ferromagnetiske partikler ligger. Dette vil igjen representere en 1 eller en 0, også kalt en "bit", og all digital informasjon er lagret slik; i 1'ere og 0'ere.

Optisk lagring

Optisk lagring er et ganske nytt fenomen sammenlignet med magnetisk lagring, og gjennombruddet kom med CD'en tidlig på 1980-tallet. CD står for "Compact Disc" og ble først og fremst laget fordi man ville ha et nytt medium til å lagre musikk og lyd på i digital form, i stedet for å bruke kassetter som baserer seg på magnetisk lagring, eller LP-plater som baserer seg på analog lagring. Siden CD-teknologien er digital tok det ikke lang tid før den første CD-ROM'en kom; en komponent man kunne feste til datamaskinen for å lese av CD'er som inneholdt programmer, bilder eller annen digital informasjon. CD-ROM står for "Compact Disc Read-Only Memory".

Harddisk [90-92]

I dag er nok harddisken den mest ”populære” lagringskomponenten, figur 15. En harddisk er den komponenten i datamaskiner over hele verden som lagrer all data over lang tid, i motsetning til RAM’en som mister all data når strømmen slås av. På harddisken lagres alle musikkfilene, filmene, bildene, dokumentene, programmene og masse mer som vi bruker til daglig på en pc. Uten denne delen ville vi aldri kunne ha lagret så mye som vi gjør i dag. Denne ”vitale” delen bruker prinsipper innen elektrisitet og, ikke minst, magnetisme for å lagre data. Ved bruk av magnetisme kan en harddisk per 2005 lagre opptil 500 GB (Gigabytes). Til sammenligning bruker en mp3-sang ca 3-5 MB (Megabytes) av den totale harddisk plassen. Det vil si at vi kan få plass til mellom 100000-200000 sanger på en slik harddisk. Mange datamaskiner i dag, spesielt servere, har ofte flere harddisker, og det er ikke uvanlig at privatpersoner ofte har flere TB (Terrabytes) med harddiskplass. Har man først lagret noe på en harddisk kan det ligge der i 10-20 år uten at det forsvinner, hvis man ikke sletter det selvfølgelig.

Virkemåte

De enkle prinsippene bak en harddisk er veldig lik andre magnetiske lagringsmedier, men en harddisk er allikevel unik. Først å fremst bruker den glass- eller metallplater dekket med et magnetisk materiale som det er enkelt å kontrollere og manipulere. Disse platene, som ligner veldig på en CD, spinner rundt i en utrolig høy fart. En vanlig plate per 2005 spinner med en fart på 7200 RPM, (revolutions per minute), men det er ikke uvanlig med plater som spinner med 10000 RPM eller mer. Den høye roteringsfarten gjør at lese- og skrivehodene kan lagre og lese av data mye fortere sammenlignet med eldre harddisker.

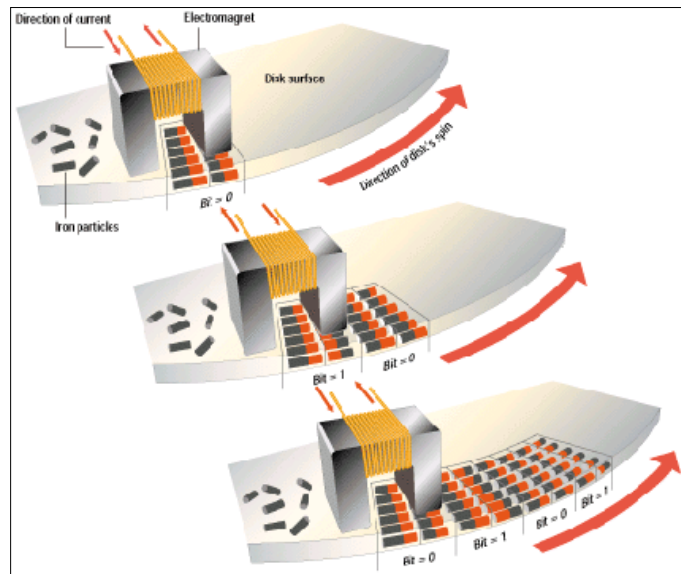
Platene var ofte laget av finbørstet aluminium, men etter som mer og mer data skulle lagres på mindre områder, måtte lese- og skrivehodet ”sveve” bare noen få nanometer over overflaten til platen. Dermed måtte også platene bli jevnere, og derfor er mesteparten av disse i dag laget av finpusset glass. Dette er dog ikke uproblematisk siden glass ofte brister lettere enn andre materialer. Og siden utviklingen har vist at platene i fremtidige harddisker spinner fortere og fortere, har harddiskindustrien investert store summer i forskning på materialer som er så sterke som mulig, har veldig jevne overflater og som samtidig er billig å framstille.

Gamle harddisker brukte ofte jernoksid som det magnetiske materialet, også kjent som rust. Dette ble også brukt i kassetter og andre tape-baserte medier. Det er billig å framstille og bruke, men etter hvert som lagringskapasiteten økte ble også jernoksid for grovt som magnetisk



Figur 24; Dette er en helt vanlig harddisk. Man kan se lese- og skrivehoder og at det er 6 av dem. Dette betyr at det ligger til sammen 3 plater/skiver oppå hverandre.

materiale. Hvis man skal klare å lagre mye data på et lite område, må også partiklene som manipuleres bli mindre. Derfor bruker harddiskene i dag et helt annet substrat som ligger som et veldig tynt lag over og under selve platen som spinner rundt. Dette substratet er ofte en slags koboltblanding med en tykkelse på noen titals nanometer. Siden det ferromagnetiske materialet er veldig finkornet, må også lese- og skrivehodene bli mindre og de må klare å manipulere disse, i virkeligheten, små ansamlingene av partikler. I tillegg lagrer harddisken data på både over- og undersiden på en plate, så hvis man har 3 plater trengs det 6 lese- og skrivehoder, figur 24. Ved å øke antall plater som ligger i en harddisk, kan man øke lagringskapasiteten dramatisk.



Figur 25; Her ser man overflaten på en plate i en harddisk og hvordan retningen som de magnetiske partiklene ligger i representerer "bits". All digital informasjon er representert i "bits", 1'ere og 0'ere.

Lese- og skrivehodet er selve manipulatoren. Denne har som oppgave å magnetisere de små ferromagnetiske partiklene som ligger på platene. Disse kan igjen manipuleres til å ligge i et bestemt mønster, og det er disse mønstrene som omgjøres til "bits", 1'ere eller 0'ere, figur 16. Et lese- og skrivehode er i prinsippet en hesteskoformet metallgjenstand. Denne er surret inn med en elektrisk leder, og når strøm sendes i en bestemt retning i lederen, vil et magnetisk felt induseres. Skifter man retningen på strømmen, vil også det magnetiske feltet skifte retning. På denne måten kan lese- og skrivehodet bestemme hvilken vei de ferromagnetiske partiklene skal ligge. Man polariserer partiklene slik at de ligger gruppevis. Hver gruppe består av 2 rader med partikler som er magnetisert. Hvis 2 rader i en gruppe har samme pol vendt mot hverandre, vil dette representere en 1'er, og motsatt for en 0'er. Når den skal lese av data, svever den bare over de områdene som allerede er manipulert til å representere data. Da vil det induseres en liten strøm i lederen som er surret rundt "hesteskoen", og dette gjøres så om til "bits", figur 25.

Det er en grunn til at lese- og skrivehodet må sveve over platen. Hvis den var borti overflaten, ville den skrape løs små partikler som videre kan ødelegge data. En harddisk er derfor også meget partikkelfri innvendig. Mange tror faktisk at man har dannet et vakuum inne i harddisken for å klare å holde støvpartikler og andre partikler borte fra platene, men dette er ikke tilfellet. I stedet har man klart å filtrere bort de fleste partikler ved produksjon, i tillegg er harddiskene utstyrt med filtre som kan rense bort alle partikler som kan være skadelige for platene etter at de er ferdig produsert.

Det er utrolig å tenke på hvordan de forskjellige komponentene i en harddisk må fungere sammen og hvor fort dette skjer. Man kan forestille seg at datamengdene ligger på platene i et fullstendig kaos, men ingeniørene har laget et system som gjør at datamengdene lettere kan finnes av lese- og skrivehodene. Man har rett og slett delt platene inn i små områder som man kaller spor og sektorer. På disse områdene ligger det en bestemt mengde med "bits". Gamle harddisker hadde kanskje bare 20 spor og 16 sektorer som man lett kunne se på platens overflate, men dagens plater har så mange at man må bruke et kraftig mikroskop for å kunne se disse. Nye harddisker er også delt inn i soner og sektorer som varierer i antall og størrelse med

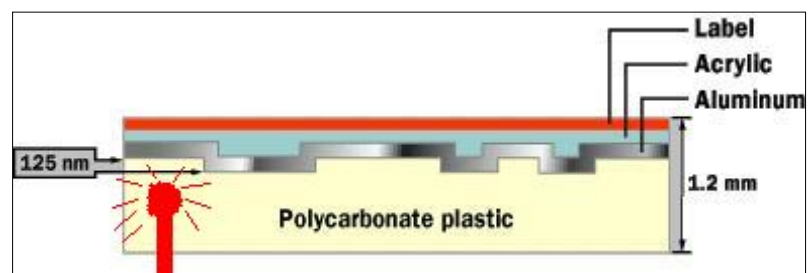
henhold til plasseringen på platen. Innerst på platen er det ikke like mange spor og sektorer som på den ytterste delen av platen. Dette gjør at man kan fordele datamengder på en mye mer effektiv måte og lagringskapasiteten øker derfor også.

Harddiskene må, som mange andre komponenter ha en "hjerne" som forstår signalene som blir registrert av lese- og skrivehodene. Derfor har alle harddisker et kretskort som er festet på undersiden. Dette kretskortet gjør for eksempel om strømsignalene som hodene registrerer til "bits". Det har også som oppgave å formidle disse videre til RAM'en, slik at prosessoren kan dra nytte av dem.

CD og DVD [92-95]

CD, DVD og alle andre former for optisk lagring baserer seg på den samme teknologien som ble utviklet på begynnelsen av 1980-tallet, og da denne teknologien først kom ut kunne den lagre enorme mengder data sammenlignet med andre lagringsmedier. Alle vet hvordan en CD ser ut, men hvordan den er satt sammen og hvordan den fungerer er det ikke like mange som vet.

En CD er bare ca 1.2 millimeter tykk, men består av flere lag med forskjellige materialer, figur 26. Det første laget, som ligger underst av alle lagene, er en type plastisk polykarbonat som er gjennomsiktig. Det er på dette laget at den digitale informasjonen blir "presset" på i form av topper. Så blir et



Figur 26; Her kan man se de forskjellige lagene på en CD. Laserstrålen sendes altså nedenfra og oppover på bildet.

tynt, reflekterende lag av aluminium lagt over plastikken. Deretter legger man et lag av det beskyttende materialet akryl, som er en ganske hard plasttype. Deretter kan man trykke på et bilde eller en annen form for kunst som beskriver innholdet på CD'en. Alle "bitene" på en slik CD blir skrevet i en sammenhengende spiral, fra sentrum av platen og gradvis lenger ut. Denne sammenhengende "veien" med data er ca 0,5 mikrometer bred, og hver av toppene er ca 125 nanometer høye. Når data skal leses fra CD'en bruker man laser som har en bølgelengde på ca 780 nanometer. Når lyset registrerer en forandring mellom en topp og bunn, altså topp/bunn eller bunn/topp forandringen, blir dette gjort om til en 1'er. Hvis det ikke er noen topp/bunn forandring over en strekning blir dette forstått som en 0'er. Det er forskjellen i fasen på lyset som reflekteres fra toppene og bunnene som registreres av en fotodiode og som gjøres om til "bits".

Andre CD-typer, som for eksempel CD-R plater og CD-RW plater, baserer seg på samme prinsipp, men her skal det også være mulig å lagre data på en slik plate fra en vanlig datamaskin. På en CD-R plate bruker man et organisk fargemateriale som kan manipuleres av laserens i en CD-spiller som er laget slik at man kan lagre data også. Dette fargematerialet er egentlig reflekterende, men laserens kan "brenne" dette materialet slik at det oppstår mørke

flekker som ikke er reflekterende. Disse mørke flekkene oppfattes som topper, og forandringen i brent/ikke brent, eller reflektert/ikke reflektert lys vil oppfattes som en 1'er, på samme måte som på en vanlig CD.

En CD-RW plate, samt diverse typer DVD-plater, bruker samme prinsipp; man bruker laser av forskjellige frekvenser til å skrive og lese datamengder på en overflate av forskjellig oppbygning. Ved å bruke kortere bølgelender, kan man også forminske størrelsen på toppene på overflaten til platen, og da kan man også pakke disse tettere enn på en vanlig CD. I tillegg har man klart å utvikle så sensitive fotosensorer og algoritmer som kan se forskjell på opptil 2 lag med "bits". En vanlig CD har ca plass til 750 Megabytes, figur 27, mens en vanlig DVD i dag kan lagre mellom 4,4-17 Gigabytes med data.



Figur 27; Her ser man en vanlig CD med operativsystemet "Windows Me".

Skjerm / Monitor [100-113]

Alle datamaskiner som skal styres eller kontrolleres av mennesker er avhengige av å ha en form for output som gir oss informasjon og tilbakemelding. I de aller fleste datamaskiner er dette i form av en visualisering. Det kan være alt fra et lite rødt lys eller en viser på et speedometer til de mest informasjonsrike skjermer og displays. Når det gjelder datamaskiner som brukes av vanlige forbrukere, på skoler og i hjemmet, bruker vi skjermer, og i all hovedsak bilderør (CRT) og LCD skjermer. Datamaskiner kan naturligvis også kobles til andre type skjermer, projektorer og andre komponenter for visualisering. Uansett baserer disse seg på to grunnleggende egenskaper i menneskehjernen. Den første er hjernens evne til å sette sammen en mengde fargede prikker til et meningsfylt bilde. Dette ser du tydelig hvis du går veldig nær en veldig stor skjerm, for eksempel på fotballkamp. Du vil da kun se en mengde fargede pærer, eller områder, der du på avstand så et bilde. Disse områdene eller prikkene kaller vi på dataspråket piksels. Den andre egenskapen er evnen til å sette sammen en mengde stillbilder som vises raskt etter hverandre til én bevegelse. Minimum antall bilder som må vises for at vi skal oppfatte dette som sammenhengende er 15 per sekund.

Signal

For st du skal se noe som helst på PC-skjermen er den avhengig av strømforsyning og signaler om hva den skal vise. Dette signalet kan være analogt, VGA, eller digitalt, DVI, og det finnes derfor to forskjellige typer ledninger. Gamle CRT-skjermer bruker analogt, nyere har gjerne tilkoblingsmuligheter for begge, mens LCD-skjermer har digitalt som standard. Digitalt finnes enten som "Single Link" eller "Dual Link", avhengig av hvilken oppløsning som ønskes. Det finnes også digitale ledninger som takler analoge signaler i tillegg. Disse kalles DVI-I, der I-en står for "Integrated", og har som vist i figur 28 fire ekstra pins for å føre det analoge signalet.

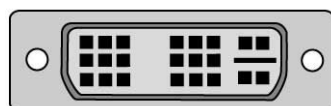
Analog skjermkontakt



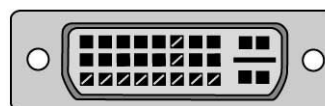
Signalfordeling i analog skjermledning

1: Rød ut	6: Rød retur (jord)	11: Skjerm ID 0
2: Grønn ut	7: Grønn retur (jord)	12: Skjerm ID 1 eller data fra skjerm
3: Blå ut	8: Blå retur (jord)	13: Horisontal shift
4: Ikke i bruk	9: Ikke i bruk	14: Vertikal shift
5: Jord	10: Shift retur (jord)	15: Skjerm ID 3 eller klokke

Digitale skjermkontakter



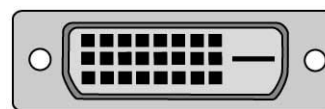
Single Link DVI-I



Dual Link DVI-I



Single Link DVI-D



Dual Link DVI-D

Figur 28; Her er en oversikt over den analoge tilkoblingen, VGA, og de forskjellige digitale tilkoblingene, DVI-typerne.

Uansett hva slags type signal som føres kommer det fra datamaskinens skjermkort. Skjermkortet er, dersom det ikke er integrert på selve hovedkortet, en egen separat komponent. Det har som oppgave å tolke den digitale informasjonen som kommer fra hovedkortet, og dermed fra prosessoren og minnet, og lage bilder av det. Skjermkortet består derfor av en inngang for strømforsyning og data fra hovedkortet, en prosessor, GPU, som bestemmer hva som skal gjøres med hver enkelt piksel på skjermen, et minne, VRAM, for å huske informasjon om hver enkelt piksel, i noen tilfeller hele bilder. Og selvfølgelig en utgang mot skjermen slik at vi kan se resultatet. Denne utgangen kan da være digital og/eller analog gjennom en omformer.

Prosessoren på skjermkortet, "Graphic Processing Unit" eller GPU, ligner på datamaskinens prosessor, "Central Processing Unit", men er spesielt konstruert for å utføre komplekse matematiske og geometriske operasjoner som er nødvendig for å lage bilder. Noen av de raskeste GPU'ene har faktisk flere transistorer enn en gjennomsnittlig CPU. Å forvandle data til bilder er en av datamaskinens tyngste oppgaver og skjermkortet trekker derfor mye strøm og lager mye varme. Skjermkortets minne, VRAM, må være veldig raskt og er derfor nesten alltid "dual ported", som vil si at det kan skrives til og leses fra samtidig.

Bilderørskjermer – CRT

”Cathode Ray Tube”, CRT, er fortsatt den klart vanligste skjermtypen og benytter den samme teknologien som i gamle TV’er, figur 29. Slike skjermer blir ofte kalt bilderør- eller katoderørskjermer. Selv om LCD-skjermer blir stadig mer populært, sverger en del fortsatt til CRT på grunn av dens overlegne egenskaper, enn så lenge. CRT har nemlig et mer lyssterkt skjermbilde, større innsynsvinkel, skjermen oppdateres oftere og kan ha større oppløsning.

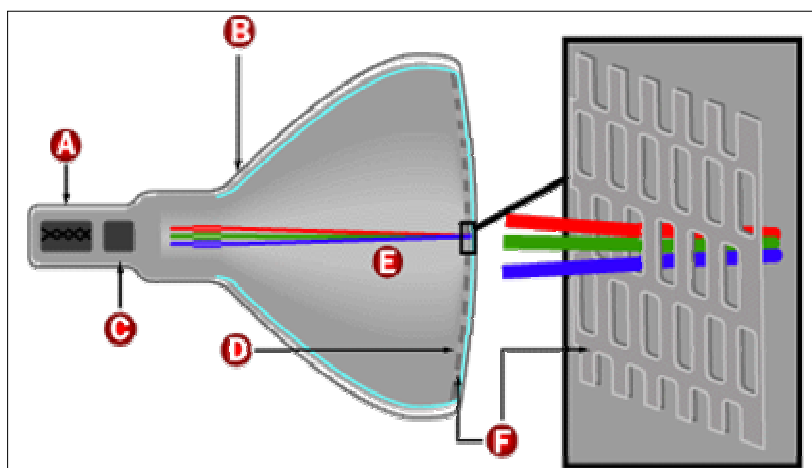


Figur 29; Bilde av en vanlig CRT-skjerm.

Innvendig består en CRT-skjerm av en glasskjegle, se figur 30, og bak denne sitter en elektronkanon bestående av en katode, A, og to eller flere anoder, C. Katoden er et filament, ikke ulikt glødetråden i en lyspære. Når denne blir varmet opp i vakuum sender den ut elektroner som trekkes i riktig retning av anodene. Den første kalles fokuserende anode og har, som navnet tilsier, som funksjon å samle alle elektronstrålene til én. Den andre er en akselererende anode som gir elektronene farten de trenger for å nå frem til skjermen.

Skjermen er litt konkav, noe som skyldes at avstanden fra den siste anoden skal være tilnærmet lik til alle punkter på skjermen for å unngå forsinkelser. Innsiden av skjermen er dekket av et tynt lag med fosfor (D) som lyser når det blir truffet av elektroner.

I svart/hvitt skjermer er det én type fosfor som sender ut hvitt lys når det blir truffet av elektroner, mens det i fargeskjermer er tre forskjellige typer fosfor ordnet som striper eller punkter i forhold til hverandre. Disse sender ut henholdsvis rødt, grønt og blått lys. I fargeskjermer består derfor elektronkanonen i realiteten av tre elektronkanoner som hver sender ut en elektronstråle, E. I hver av disse kan intensiteten reguleres separat og på denne måten produsere alle mulige farger. Forskjellige typer fosfor kan ha variasjoner i det emitterte lysets bølgelengde og hvor lenge det emitterer etter å ha blitt truffet av elektronet. På innsiden av fosforet er det i fargeskjermer også et tynt metallgitter kalt skyggegitte, F. Hullene i gitteret er tilpasset stripene eller punktene i fosforen. Dette er et hjelpemiddel for å skille mellom pikslene slik at bildet blir skarpt.



Figur 30; Her er en skjematisk oversikt over glasskjeglen, katodene og anodene i en CRT-skjerm. De store enkeltbokstavene i teksten refererer til bokstavene figuren.

Innsiden av glasskjeglen er ellers dekket med en elektrisk leder, B, som fanger opp elektronene som kastes tilbake fra skjermen. Den tykke, svarte ledningen på bildet under er koblet til denne.

Slik som beskrevet til nå vil elektronstrålen, E, gå rett frem i vakuumet og lyse opp ett punkt akkurat i midten av skjermen. For å lage et bilde er vi avhengige av å styre denne elektronstrålen. Det gjøres ved hjelp av to spoler av kobbertråd rundt glassrøret, en for å styre strålen i vertikal retning og en for horisontal. Disse kan sees i figur 31. På denne måten kan elektronstrålen styres til hvilket som helst punkt på skjermen ved å variere spenningen over disse.



Figur 31; Innmaten i en vanlig CRT-skjerm. Man kan skimte de to spolene av kobbertråd.

Visningen av bilder

CRT-skjermen viser bilder ved å styre elektronstrålen i horisontale linjer fra venstre til høyre på samme måte som når vi leser en tekst. Strålen skrur så av og starter til venstre igjen på neste linje. Slik fortsetter den til alle linjene er fylt og hopper så, som når øynene våre "leser" en tekst, øverst til venstre igjen. Disse forflytningene kalles "shift" eller "retrace", og det er viktig å merke seg at elektronstrålen slås av under "shift'et". Vi har med andre ord horisontale og vertikale "shift".

En vanlig skjerm har rundt 400 slike linjer og hele skjermen oppdateres gjerne 60 ganger i sekundet. TV-skjermer, og noen PC-skjermer, bruker en teknikk kalt "interlacing" der bare annenhver linje oppdateres hver gang. Elektronstrålen har da fortsatt 60 vertikale "shift" i sekundet, men hele bildet oppdateres i realiteten bare 30 ganger i sekundet. Med 400 linjer 30 ganger per sekund blir det 12 000 linjer i sekundet, en frekvens som for noen er hørbar som en veldig lys tone. De fleste PC-skjermer oppdaterer riktig nok hele skjermen 60 ganger i sekundet fordi det reduserer flimringen.

Dersom du har prøvd å filme en CRT-skjerm med et videokamera har du sannsynligvis sett at det gir et dårlig resultat. Du vil enten se intens flimring og/eller en sort stripe som ruller nedover skjermen. Flimringen skyldes enten at kameraet du bruker tar bilder med en annen frekvens enn den skjermen din har og den sorte stripa kommer av forskjellen i hvordan kameralinsen og øyet ditt ser fosforpunktene. Et vanlig videokamera er nemlig langt fra så følsomt for lys som det menneskelige øyet, og vil derfor ikke se de opplyste punktene like lenge etter at de har blitt truffet av elektroner. Begge disse fenomenene kan unngås ved å forandre skjermens frekvens eller ved rett å slett å filme en LCD-skjerm, som ikke fungerer på samme måte som en CRT-skjerm.

Flatskjermer – LCD

”Liquid crystal display”, LCD, er den vanligste typen flatskjerm i dag og den blir stadig mer populær, figur 32. Teknologien med flytende krystaller har vært i bruk i lang tid og eksempler finnes overalt rundt oss i digitale klokker, cd-spillere, mikrobølgeovner, bilstereo osv. De aller fleste slike små display benytter LCD-teknologi. At LCD skjermer til datamaskiner er blitt så populært skyldes hovedsakelig prisen og størrelsen. LCD-teknologien er nemlig langt billigere å produsere enn f.eks plasmaskjermer, og mindre og lettere enn CRT. I tillegg bruker de langt mindre strøm og de er derfor standard i alle bærbare datamaskiner.



Figur 32; Her kan man se en vanlig LCD-skjerm. Slike skjermer tar veldig lite fysisk plass, sammenlignet med CRT-skjermer.

Flytende krystaller

De fleste har lært at materie forekommer i tre forskjellige former, nemlig fast form, væske og gass. Derfor høres det veldig rart ut med flytende krystaller. I krystaller har nemlig molekylene en fast struktur og posisjon i forhold til hverandre, mens det i væsker er helt motsatt. Her kan molekylene bevege seg nærmest fritt i forhold til hverandre.

Så i hvilken bås setter vi flytende krystaller? Svaret er ingen. De kan bare defineres som et sted i mellom fast og flytende materie og derfor har de fått sitt noe motsigende navn. Molekylene holder som regel en fast posisjon i forhold til hverandre, men kan også forandre posisjoner. Det viser seg likevel at flytende krystaller har flere av egenskapene til en væske enn fast stoff. For å lage flytende krystaller av et fast stoff trengs det en god del varme, deretter trengs det kun en liten temperaturforandring for å forandre den til væske. Dette forklarer blant annet hvorfor laptop'er og andre LCD-skjermer kan fungere dårlig på kalde eller veldig varme dager.

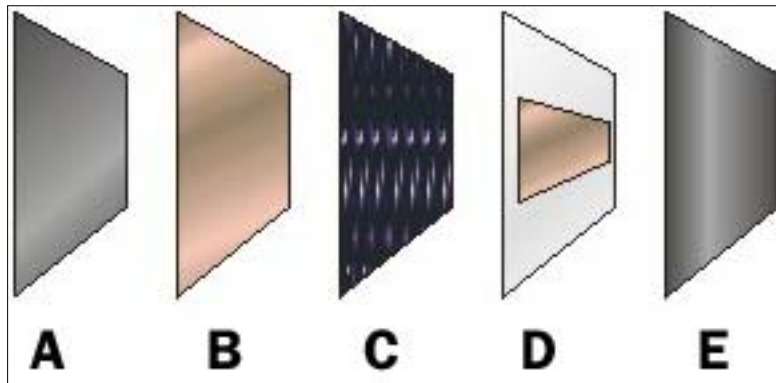
Det finnes mange forskjellige typer flytende krystaller. De som brukes i skjermer kalles ”twisted nematics”, TN. De ønsker til enhver tid å organisere seg i en spiralform og reagerer på strøm på en slik måte at spiralen retter seg ut. Hvor mye spiralen rettes ut avhenger av strømstyrken som går igjennom den. Denne typen flytende krystaller brukes i LCD-skjermer fordi de reagerer forutsigbart på strøm og fordi de på denne måten kan slippe igjennom en bestemt mengde lys.

LCD-skjermens sammensetning

Teknikken bak en LCD-skjerm hviler på fire grunnsteiner:

- Lys kan polariseres.
- Flytende krystaller kan reflektere lyset slik at polarisasjonsretningen forandres.
- Flytende krystaller kan forandre form når de blir påvirket av strøm.
- Det finnes gjennomsiktige materialer som kan lede strøm.

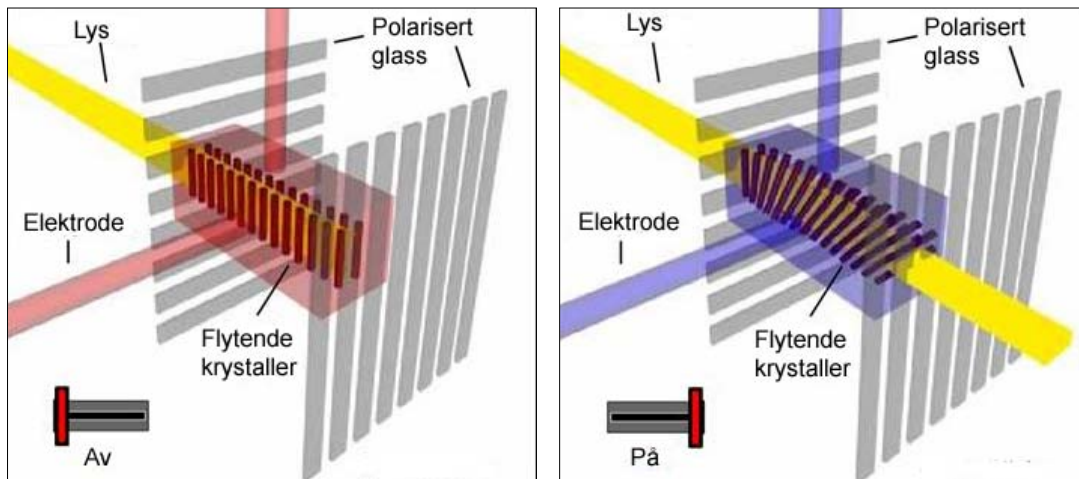
For å lage en skjerm trenger du to glassplater, figur 33; A og F, som er polarisert på utsiden og har en gjennomsiktig elektrode på innsiden, B og D. Et tynt lag med polymer strykes utover den ene glassplaten, på den siden som ikke har den polariserende filmen. Denne har som funksjon å innrette de flytende krystallene i samme retning som polariseringen på glassplaten. Så legges flere lag med TN-krystaller, C, og til slutt den siste glassplaten. Denne må ha polarisering vinkelrett på polariseringen til den første glassplaten. Nå vil det nederste laget med TN-krystaller ha samme retning som den første polarisasjonen og de påfølgende lagene vil være dreid litt om sin egen akse i forhold til den forrige helt til det siste laget som har samme retning som den siste polarisasjonen. Spiralen har altså en kvart omdreining.



Figur 33; de forskjellige lagene med glass, plast og elektroder.

Når lyset nå treffer den første glassplaten blir det polarisert. Deretter vil hvert ledd i spiralen sende lyset videre til det neste, men med en liten dreining av polarisasjonsretningen for hver gang. Når lyset når enden av spiralen vil det dermed være polarisert i samme vinkel som det siste leddet. Det vil derfor være vinkelen til det siste leddet i forhold til det første som bestemmer hvor mye lys som skal slippe igjennom, med andre ord lysintensiteten. Denne dreiningsvinkelen til spiralen kan vi som tidligere nevnt styre ved hjelp av strøm.

Når lyset nå treffer den første glassplaten blir det polarisert. Deretter vil hvert ledd i spiralen sende lyset videre til det neste, men med en liten dreining av polarisasjonsretningen for hver gang. Når lyset når enden av spiralen vil det dermed være polarisert i samme vinkel som det siste leddet. Det vil derfor være vinkelen til det siste leddet i forhold til det første som bestemmer hvor mye lys som skal slippe igjennom, med andre ord lysintensiteten. Denne dreiningsvinkelen til spiralen kan vi som tidligere nevnt styre ved hjelp av strøm.



Figur 34; Hvordan hver spiral av flytende krystall fungerer.

Passive rutenett

Det finnes to forskjellige måter å organisere enkeltpunktene/pikslene på slik at man kan styre hver enkelt separat. Den enkleste kalles på engelsk "Passive Matrix". Her aktiveres pikslene av et enkelt rutenett av et gjennomsiktig strømførende materiale, oftest indium-tinn oksid. På innsiden den ene glassplaten legges vertikale ledere og på den andre horisontale. Ved å sende en ladning ned en av de vertikale lederne og legge jord på en horisontal vil dette aktivere

pikslen der de to møtes. Dette kan virke som et elegant system, men det har sine negative sider. For det første tar det lang tid når dette skal gjøres med alle pikslene. Vi får dermed lav responstid og oppdateringsfrekvens som kan sees ved fenomenet "ghost" som for eksempel halen av piler du ser når du beveger musa raskt over skjermen eller etter raske bevegelser i filmer. I tillegg medfører passive rutenett dårlig spenningskontroll, det vil si at ikke bare den ønskede pikslen rettes ut, men også til en viss grad pikslene rundt. Dette fører igjen til dårlig kontrast på skjermen.

Aktive rutenett

Det andre systemet, kalt "Active Matrix", baserer seg på tynn film transistor, TFT-teknologi. TFT er i realiteten ørsmå transistorer og kondensatorer ordnet i et rutenett på den ene glassplaten. For å aktivere en piksel skrur man på den raden og sender så en ladning vertikalt nedover den riktige kolonnen. Siden alle de andre radene ikke er aktivert treffer ladningen kun den ønskede kondensatoren, som holder på denne helt til neste oppdateringssyklus. Denne teknologien er både raskere og mer nøyaktig, men koster også mer å produsere.

Lys

Det er viktig å merke seg at LCD-skjermer ikke lyser selv, slik som CRT-skjermene. Enkle LCD'er som klokker osv. er avhengige av å reflektere lys fra eksterne lyskilder for at du skal kunne se noe. De fleste LCD'er, deriblant dataskjermer, har baklys slik at de også kan benyttes i liten belysning og mørke. I dataskjermer er det lysrør øverst, på sidene og kanskje bak skjermen som skaper lyset. En reflekterende overflate bak hele skjermen fordeler dette lyset jevnt over hele skjermen. Som nevnt tidligere har LCD ulempen at de ikke er like lyssterke som CRT. Det skyldes at nærmere 50 % av lyset aldri kommer ut av gjennom skjermen.

Farger

For at en LCD skal kunne vise farger må hver piksel ha tre underpikslar. En rød, en grønn og en blå, akkurat som for CRT. Dagens skjermer har ved finjustering av transistorer og kondensatorer fått til 256 grader av gjennomlysning i spiralene, fra hvitt til svart. Dette kombinert med de tre fargepikslene gjør at dagens skjermer kan vise hele 16,8 millioner forskjellige farger (256 grader av rød x 256 grader av grønn x 256 grader av blå)! For å klare dette med et aktivt rutenett kreves et svimlende antall transistorer. Ta for eksempel en oppløsning på 1280x960 som er vanlig for en bærbar datamaskin: til dette trengs 1280 x 960 x 3 underpikslar, eller totalt 3 686 400 transistorer spredt utover glassflaten.

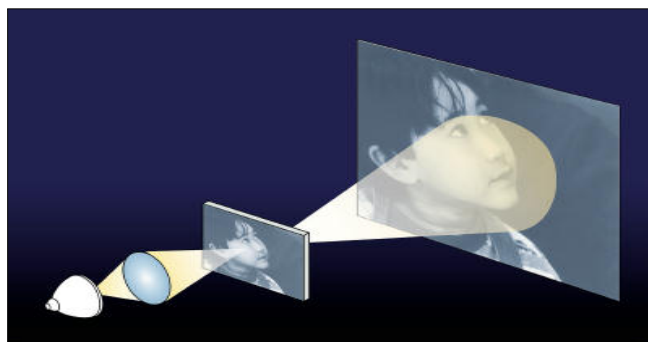
Dersom det under produksjonen eller seinere oppstår en feil med en av disse transistorene vil det komme til syne som en defekt piksel. Dette er for øvrig grunnen til at store LCD-skjermer er veldig kostbare. Jo større de er, jo større er sjansene for at de inneholder en defekt piksel og må kasseres. Produsentene av virkelig store LCD'er må kassere rundt 40 % av alle skjermene de lager, salget av de resterende må derfor dekke kostnadene for disse også.

Andre typer monitorer og projektorer

En PC kan også kobles opp mot en rekke andre monitorer, for eksempel TV-skjermer. CRT og LCD-skjermer vil fungere veldig likt som PC-versjonene, mens plasmaskjermer fungerer på en helt annen måte. Plasmaskjermer har det til felles med CRT at de selv sender ut lys og ligner

LCD'er ved at de også har passive rutenett. I plasma er det derimot en liten "lampe" i hver underpiksel, fargepiksel, som aktiveres av ladningen som sendes. Man lager så små "lamper" ved hjelp av bittesmå lommer mellom glassplatene som fylles med en gass av frie ioner. En spenning over en slik lomme vil trekke elektronene mot den ene platen og protonene mot den andre, og kollisjoner mellom disse eksiterer fotoner, altså lys. Med god lysstyrke, synsvinkel og liten tykkelse ser plasmaskjermen ut til å ha tatt de beste egenskapene til CRT og LCD, grunnen til at de ikke brukes til datmaskiner i svært liten grad er at de er veldig dyre å produsere.

På mange arbeidsplasser er PC'er også i bruk sammen med projektorer, figur 35. Dette brukes gjerne for å kunne kommunisere med flere mennesker, for eksempel under et foredrag eller en demonstrasjon. En projektor sender bildet i form av lys mot en reflektiv plate eller lerret og er derfor avhengig av lite eller ikke noe lys i lokalet. Inne i projektoren sitter det i mange tilfeller faktisk en liten CRT eller LCD skjerm. Denne kan enten gjennomlyses eller brukes som speil for å sende bildet ut. I begge tilfeller må signalet først speilvendes av projektoren.



Figur 35; her viser man prinsippet bak en projektor på en veldig enkel måte.

Datamaskinens fremtid [120-123]

De færreste betviler at datamaskinen har en fremtid, men hvordan den blir og hvilken retning utviklingen tar kan ingen vite sikkert. Den nærmeste fremtiden vet vi likevel litt om. Som nevnt tidligere er det å være et jag om å lage mindre, lettere, billigere og bedre komponenter. Dette vil ganske sikkert føre til mindre og lettere maskiner som likevel yter bedre enn dagens. Gapet mellom bærbare datamaskiner og skrivebordsmodeller vil bli mindre, og kanskje vil den bærbare innta den tradisjonelle PC'ens posisjon.

Transistorer er de komponentene som forminskes fortest og om noen få år kommer vi til et interessant veiskille. Det finns nemlig en fysisk grense for hvor små man kan lage en transistor, den kan ikke være mindre enn at et elektron kan gå igjennom det. Når man kommer ned på en så liten skala i størrelsesorden, vil ikke lenger klassisk fysikk gjelde for elektronet. Man må i så fall bruke kvantefysikk til å forklare bevegelser og tilstander til en slik partikkel, men det er ingen som vet hvordan enda, selv om forskere har noen gode ideer som å utnytte "spinnnet" på et elektron.

Kommunikasjonen mellom datamaskiner vil trolig også gjennomgå en stor utvikling. Dette ser vi i dag spesielt på områdene sikkerhet, nett-betaling, beskyttede databaser, e-signatur, og informasjonshåndtering som for eksempel søkemotorer. Hastigheten på en internetforbindelse vil også vokse meget fort, og man kan utnytte denne båndbredden for eksempel ved å overføre store mengder digital informasjon i form av video/lyd, og kvaliteten på disse signalene vil være

utrolig bra. Digitale skjermer kan utnytte de digitale signalene og vise bilder som vil se "ekte" ut.

Det er veldig sannsynlig at den personlige datamaskinen vil bli håndholdt og vil også muligens ha mange andre komponenter integrert, som telefon, LAN, satellittmottager og kamera. Disse finns allerede, men de er ikke i nærheten av å konkurrere med bærbare datamaskiner og stasjonære. Integrerte kretser, som RFID-brikker, vil bli billigere å produsere og kan sys fast i klær, opereres under huden eller plasseres i andre gjenstander. Dette er en radiobrikke som kan sende informasjon ut rundt seg i en viss radius. Denne brikken kan for eksempel inneholde identiteten, blodtypen og DNA på den personen som har den under huden.

Noen ser for seg fremtidens PC'er som et litt tykt ark som du kan rulle eller brette sammen og ta med deg. Dette er tenkt styrt ved hjelp av trykk på skjermen, en såkalt "touch-screen". Det er generelt en tendens i utviklingen av nye skjermer at de lages bredere da dette passer bedre med det menneskelige synsfeltet. Skjermer vil kunne bli større, men først og fremst vil de bli klarere og mer behagelige å se på. Bildet viser en visjon om at PC-er i fremtiden har projektor istedenfor skjerm, og at det over alt vil finnes parabolignende skjermer der man kan jobbe med behagelige og virkelighetsnære bilder, figur 36

Lagringskapasiteten i en datamaskin, om det er primærminnet eller sekundær minnet, vil også vokse meget fort. Noen forestiller seg "holografisk lagring", at man lagrer datamengder i 3 dimensjoner, som i en terning. Dette kan øke lagringsmengden tusenvis av ganger i forhold til dagens løsninger som harddisker eller CD og DVD. I løpet av 2006 vil også en ny type DVD-plater komme ut. Disse kalles for HD-DVD og Blu-ray-DVD og kan lagre mellom 30 til 100 Gigabytes per plate.

Vi lever i en spennende tidsalder der teknologiene overalt rundt oss utvikler seg svært fort. Og det blir spesielt spennende å se hvordan teknologien i datamaskiner vil se ut om noen få år!



Figur 36; noen ser for seg at fremtidens skjerm bokstavelig talt vil delvis eller helt omslutte oss!

Referanser

Tallene i klammer, [], i teksten viser til referansene her:

- 20: home.online.no/~rkaste/dhindex.htm
- 21: <http://heim.ifi.uio.no/~knuthe/db-historie/datahistorie.html>
- 22: www.pbs.org/nerds/timeline/index.html
- 23: http://wikipedia.org/wiki/category:history_of_computing
- 24: www.stud.ntnu.no/~runevilfiles/tdt4160.pdf
- 25: <http://public.web.cern.ch>
- 26: "Den forunderlige reisen gjennom datahistorien", Arild Haraldsen, Tano Aschehoug, 1999
- 27: "Fra kuleramme til PC - Datamaskinens historie og betydning", Nils Kr. Rossing, Arne, Asphjell, Einar J. Aas, Vitensenteret Trondheim, 1998

- 30: <http://www.pcworld.no/index.cfm?fuseaction=artikkel&id=0225E623-0550-E4BF-154551AE1C13DA64#tekst>
- 31: <http://www.webopedia.com/TERM/m/motherboard.html>
- 32: <http://en.wikipedia.org/wiki/Motherboard>
- 33: <http://computer.howstuffworks.com/motherboard.htm>

- 40: www.vuc-ribeamt.dk/underv/jja/tastaturets_historie.htm
- 41: www.mit.edu/~jcb/dvorak/history.html
- 42: <http://computer.howstuffworks.com/keyboard.htm>
- 43: "Physics for Scientists and Engineers with modern physics", Serway og Beichner, Harcourt, 2000

- 50: <http://club.fnatic.com/technology/articles/7/>
- 51: www.museskade.dk/artikler.asp?t=5&id=11109
- 52: <http://computer.howstuffworks.com/mouse.htm>

- 59: "Electronic Circuits and Applications", Stephen D. Senturia og Bruce D. Wedlock, Wiley, 1975
- 60: <http://www.lucent.com/minds/transistor/tech.html>
- 61: <http://en.wikipedia.org/wiki/Transistor>
- 62: <http://bmrc.berkeley.edu/courseware/ICMfg92/>
- 63: <http://www.pbs.org/transistor/science/info/chip.html>
- 64: <http://electronics.howstuffworks.com/diode.htm/printable>
- 65: http://whatis.techtarget.com/definition/0,,sid9_gci213216,00.html
- 66: <http://cerebellum.no/flasher/PS/periodiskesystem.html>
- 67: <http://www.leb.e-technik.uni-erlangen.de/lehre/mm/html/start.htm>
- 68: <http://www.pbs.org/transistor/science/info/chip.html>
- 69: <http://library.thinkquest.org/12666/junction.html>
- 70: <http://www.pbs.org/transistor/science/info/transmodern.html>
- 71: http://www.st-andrews.ac.uk/~www_pa/Scots_Guide/first11/part7/page1.html
- 72: <http://computer.howstuffworks.com/microprocessor.htm/printable>
- 73: <http://computer.howstuffworks.com/boolean.htm/printable>
- 74: <http://computer.howstuffworks.com/bytes.htm/printable>
- 75: <http://www.lucent.com/minds/transistor/uses.html>
- 76: <http://www.webopedia.com/TERM/m/microprocessor.html>

- 77: http://whatis.techtarget.com/definition/0,,sid9_gci213216,00.html
- 80: <http://computer.howstuffworks.com/ram.htm>
- 81: <http://www.webopedia.com/TERM/R/RAM.html>
- 82: http://en.wikipedia.org/wiki/Random_access_memory
- 83: http://en.wikipedia.org/wiki/Read-only_memory
- 84: <http://computer.howstuffworks.com/rom.htm>
- 85: <http://computer.howstuffworks.com/computer-memory.htm>
- 86: <http://computer.howstuffworks.com/cache.htm>
- 90: <http://computer.howstuffworks.com/hard-disk.htm>
- 91: http://en.wikipedia.org/wiki/Hard_drive
- 92: <http://electronics.howstuffworks.com/cd.htm>
- 93: <http://electronics.howstuffworks.com/dvd.htm>
- 94: http://en.wikipedia.org/wiki/Compact_disc
- 95: <http://computer.howstuffworks.com/removable-storage.htm>
- 100: <http://computer.howstuffworks.com/monitor.htm/printable>
- 101: <http://monitorworld.com/faq.html>
- 102: <http://hitmill.com/html/screen.html>
- 103: <http://computer.howstuffworks.com/graphics-card.htm/printable>
- 104: <http://www.mpeg.org/MPEG/index.html>
- 105: <http://electronics.howstuffworks.com/tv.htm/printable>
- 106: <http://www.viewsonic.com/monitoruniversity/lcdvcr.htm>
- 107: <http://electronics.howstuffworks.com/question336.htm>
- 108: <http://electronics.howstuffworks.com/lcd.htm/printable>
- 109: [http://computer.howstuffworks.com/laptop.htm/printable:](http://computer.howstuffworks.com/laptop.htm/printable)
- 110: <http://electronics.howstuffworks.com/projection-tv.htm>
- 111: <http://electronics.howstuffworks.com/home-theater.htm>
- 112: <http://electronics.howstuffworks.com/plasma-display.htm>
- 113: <http://www.beareview.com/links/plasma.html>
- 120: <http://www.itavisen.no/showArticle.php?articleId=1306174>
- 121: <http://www.itavisen.no/showArticle.php?articleId=1299088>
- 122: http://www.hardware.no/nyheter/harddisk/terabyte-disker_i_lopet_av_2007/16080
- 123: http://www.hardware.no/nyheter/lagring/1_tb_hvd_snart_underveis/20312