UNIVERSITY
OF OSLO

IIR filterdesign

Sverre Holm

INF3470

Digital signalbehandling

Filterdesign

1. Spesifikasjon
   - Kjenne anvendelsen
   - Kjenne designmetoder (hva som er mulig, FIR/IIR)
2. Approksimasjon
   - Fokus her
3. Analyse
   - Filtre er som regel spesifisert i frekvensdomenet
   - Også analysere i tid (fase, forsinkelse, ...)
4. Realisering
   - DSP, FPGA, PC: Matlab, C, Java ...

Sources

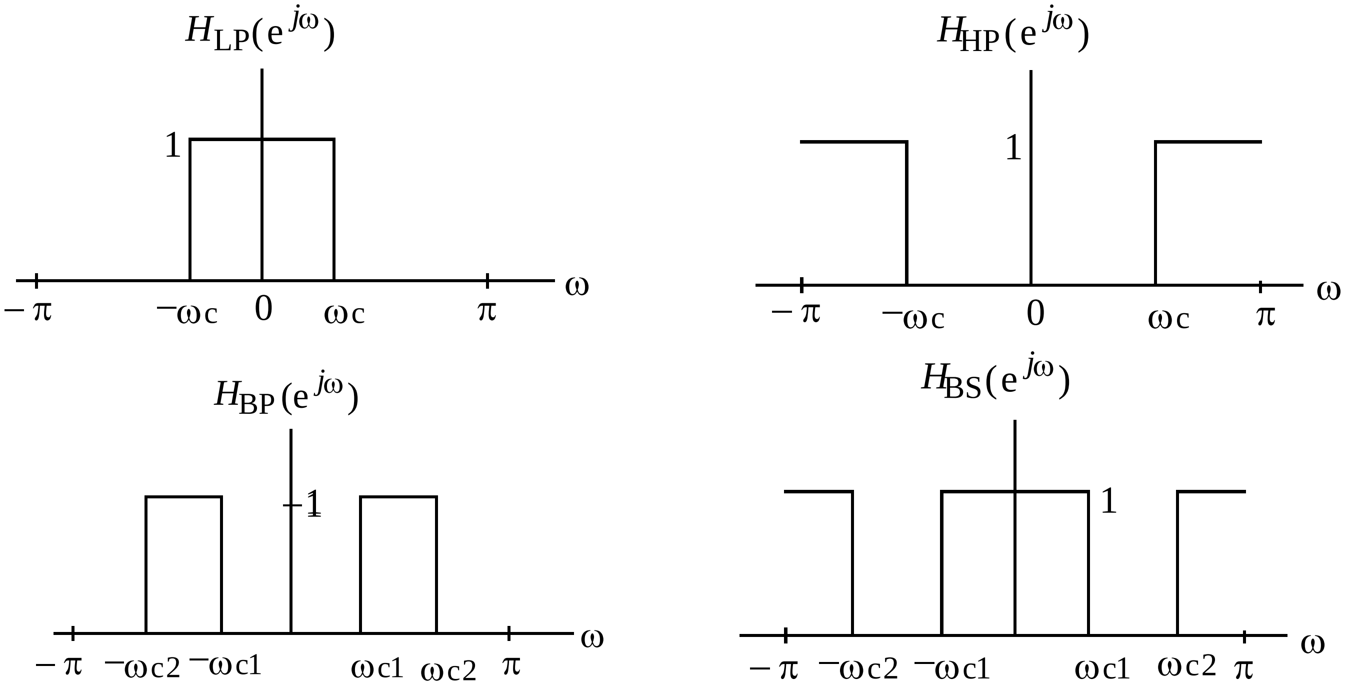
- The slides about Digital Filter Specifications have been adapted from slides by S. Mitra, 2001
- Butterworth, Chebychev, etc filters are based on Wikipedia
- Builds on Oppenheim & Schafer with Buck: Discrete-Time Signal Processing, 1999.

IIR kontra FIR

- IIR filtre er mer effektive enn FIR – færre koeffisienter for samme magnitude-spesifikasjon
- Men bare FIR kan gi eksakt lineær fase
  - Lineær fase $\Leftrightarrow$ symmetrisk h[n]
    $\Rightarrow$ Nullpunkter symmetrisk om |z|=1
  - Lineær fase IIR? $\Rightarrow$ Poler utenfor enhetssirkelen
    $\Rightarrow$ ustabilt
- IIR kan også bli ustabile pga avrunding i aritmetikken, det kan ikke FIR

Ideal filters

- Lavpass, høypass, båndpass, båndstopp



Prototype low-pass filter

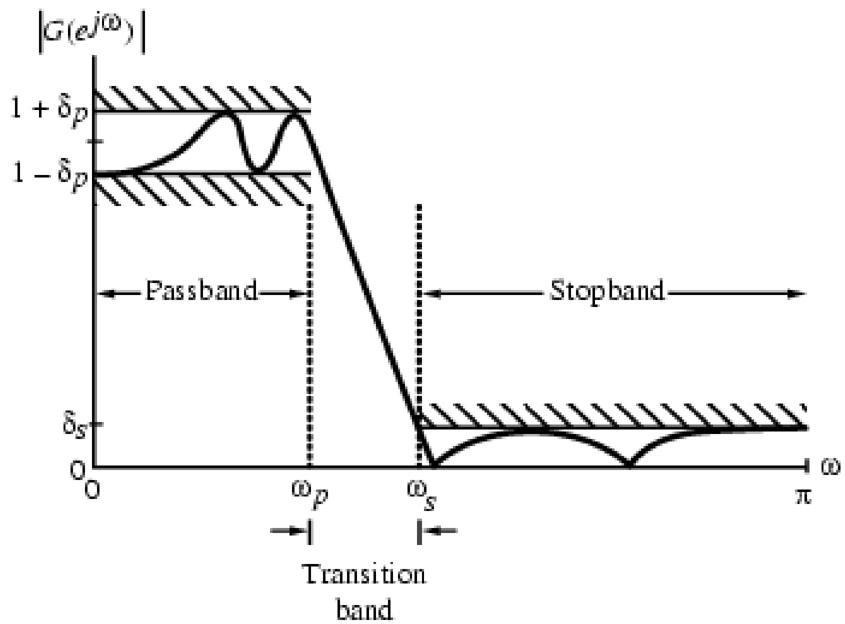
- All filter design methods are specificed for low-pass only
- It can be transformed into a high-pass filter
- Or it can be placed in series with others to form band-pass and band-stop filters

Digital Filter Specifications

- As the impulse response corresponding to each of these ideal filters is noncausal and of infinite length, these filters are **not realizable**
- **In practice**, the magnitude response specifications of a digital filter in the passband and in the stopband are given with some acceptable tolerances
- In addition, a transition band is specified between the passband and stopband

Digital Filter Specifications

- The magnitude response $|G(e^{j\omega})|$ of a digital lowpass filter may be specified as:



Digital Filter Specifications

- **Passband**: $0 \leq \omega \leq \omega_p$
  - We require that $|G(e^{j\omega})| \approx 1$ with an error $\pm\delta_p$, i.e.,

$$1-\delta_p \leq \left|G(e^{j\omega})\right| \leq 1+\delta_p, \quad |\omega| \leq \omega_p$$

- **Stopband**: $\omega_s \leq \omega \leq \pi$
  - We require that $|G(e^{j\omega})| \approx 0$ with an error δ_s, i.e.,

$$\left|G(e^{j\omega})\right| \leq \delta_s, \quad \omega_s \leq |\omega| \leq \pi$$

Digital Filter Specifications

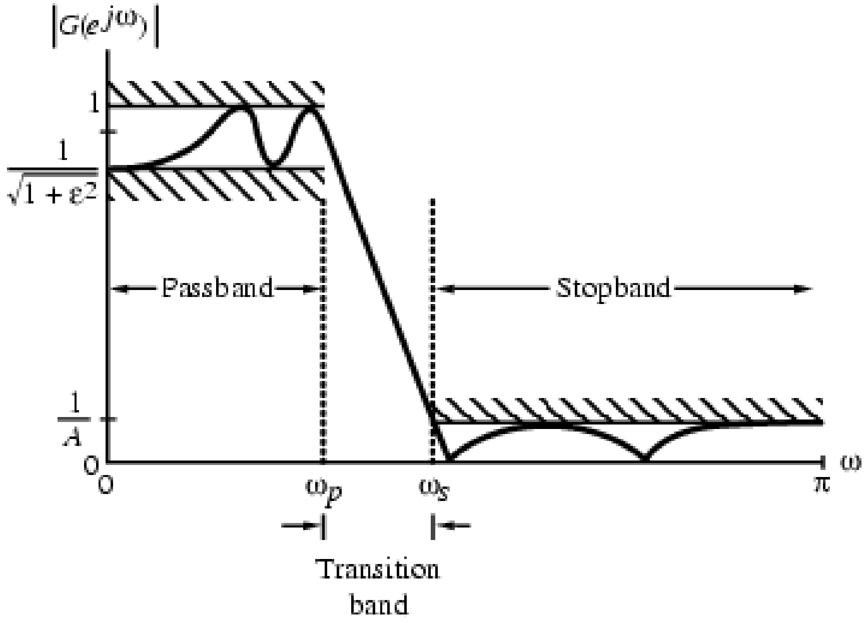
- ω_p - passband edge frequency
- ω_s - stopband edge frequency
- δ_p - peak ripple value in the passband
- δ_s - peak ripple value in the stopband
- Properties:
  - $G(e^{j\omega})$ is a periodic function of ω
  - $|G(e^{j\omega})|$ of a real-coefficient digital filter is an even function of ω
- Consequence: Filter specifications are given only for $0 \leq |\omega| \leq \pi$

Digital Filter Specifications

- Specifications are often given in terms of loss function:
  $G(\omega)=-20\ \log_{10}|G(e^{j\omega)}|$ in dB
- Peak passband ripple
  $\alpha_p = -20\log_{10}(1-\delta_p)$ dB
- Minimum stopband attenuation
  $\alpha_p = -20\log_{10}(\delta_s)$ dB

Digital Filter Specifications

- Magnitude specifications may alternately be given in a normalized form as indicated below



Normalized frequencies

- Real values
  - Real frequencies: f_p, f_s, f_{sample}
  - Angular frequencies: $\omega=2\pi f$
- Normalized values
  - Angular frequencies: $\omega=0...2\pi$ where $f_{sample} \Leftrightarrow 2\pi$
  - Normalized frequencies f=0 ... 2 where 2 is the sampling frequency (0...1 is the useful range): MATLAB filter design

FIR and IIR Digital Filter

- Difference equation

$$\sum_{k=0}^{k=N} a_k y[n-k] = \sum_{k=0}^{k=M} b_k x[n-k]$$

- Transfer function

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}}$$

← FIR-filtre med bare nullpunkter har ingen analog ekvivalent

← Analoge filter har bare poler ⇔ IIR uten FIR-del

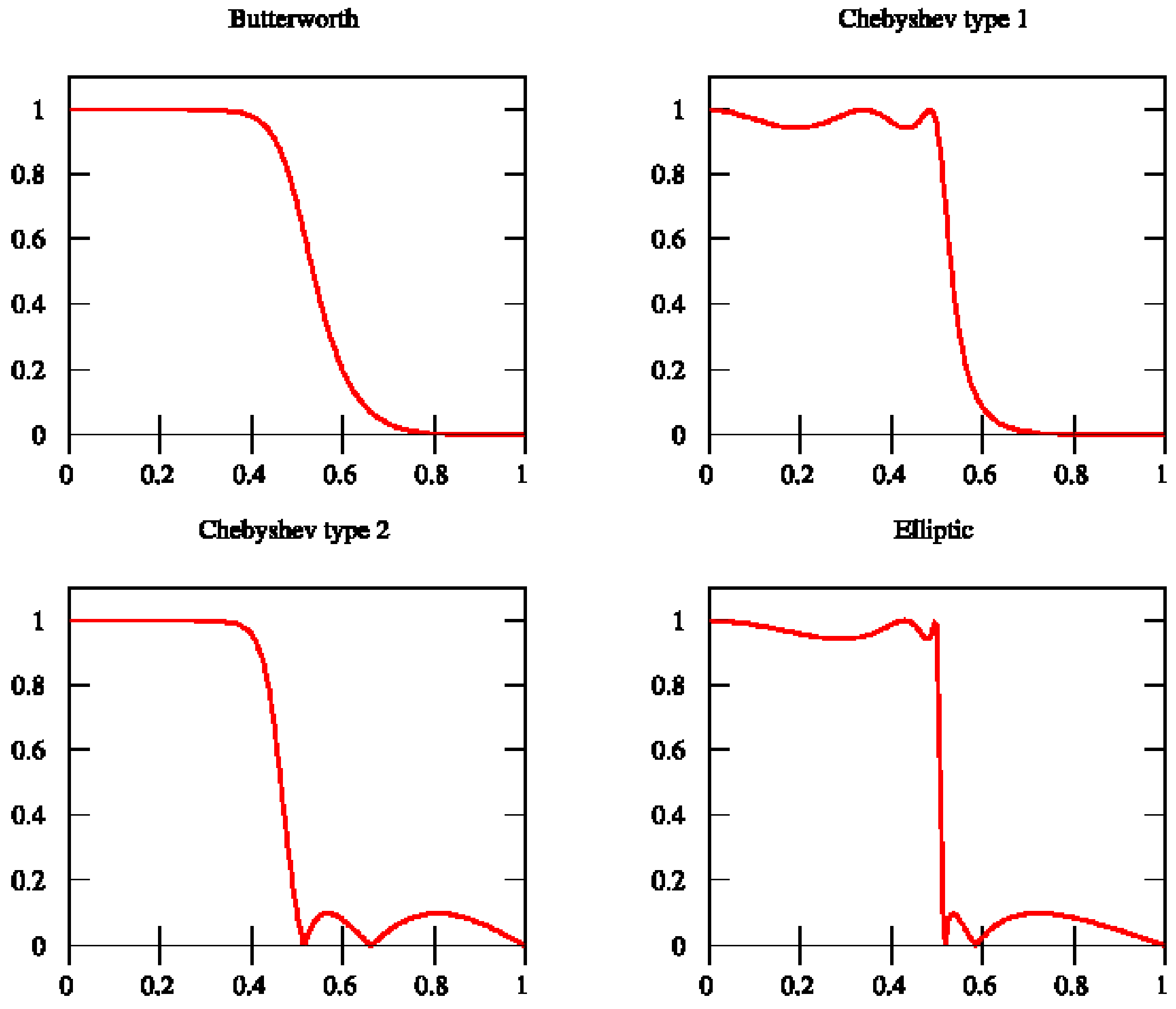
  - General: IIR - Infinite Impulse Response
  - FIR - Finite Impulse Response
    - » N=0, no feedback, always stable

Pol- og nullpunktsplassering

- Poler innenfor enhetssirkel ⇔ Stabil og kausal
- Poler og nullpunkter i kompleks konjugerte par ⇔ Reell impulsrespons
- Alle nullpunkter finnes speilet om enhetssirkelen ⇔ Lineær fase.
- Viktig! Lineær fase og reelle koeffisienter ⇔ Nullpunkter finnes i grupper av fire.

- Alle nullpunkter er speilbildet av en pol ⇔ Allpass system
- Alle nullpunkter innenfor enhetssirkel ⇔ Minimum fase system og inversfilter eksisterer
- Alle nullpunkter utenfor enhetssirkel ⇔ Kausal maksimum fase

Standard Analog Filter functions (magnitude)



Standard Analog Filter Functions

- Butterworth filter
  - no gain ripple in pass band and stop band, slow cutoff
- Chebyshev filter (Type I)
  - no gain ripple in stop band, moderate cutoff
- Chebyshev filter (Type II)
  - no gain ripple in pass band, moderate cutoff
- Elliptic filter
  - gain ripple in pass and stop band, fast cutoff

- Bessel filter
  - no group delay ripple, no gain ripple in both bands, slow gain cutoff

- Linkwitz-Riley filter
  - Used for crossover filters for loudspeakers

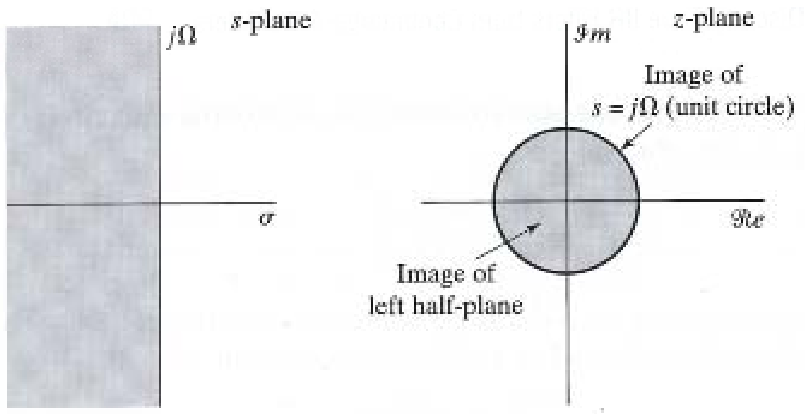
Laplace vs z-transform

- Laplace, s=σ+jΩ:

$$H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}dt$$

- Z-transform, z=e$^{j\omega}$:

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n}$$



Avbildning mellom s- og z-plan
ved bilineær transform

Filter Slopes

- **Decade**: 10 x frequency, e.g. 100 Hz – 1 kHz is one decade
- **Octave**: 2 x frequency (octave=8 for the white piano keys), e.g. 100 Hz -> 200 Hz
  - 1. order filter: rolls off at −6 dB per octave (−20 dB per decade)
  - 2. order filter: the response decreases at −12 dB per octave (-40 dB per decade)
  - 3. order at −18 dB, and so on.

Group Delay

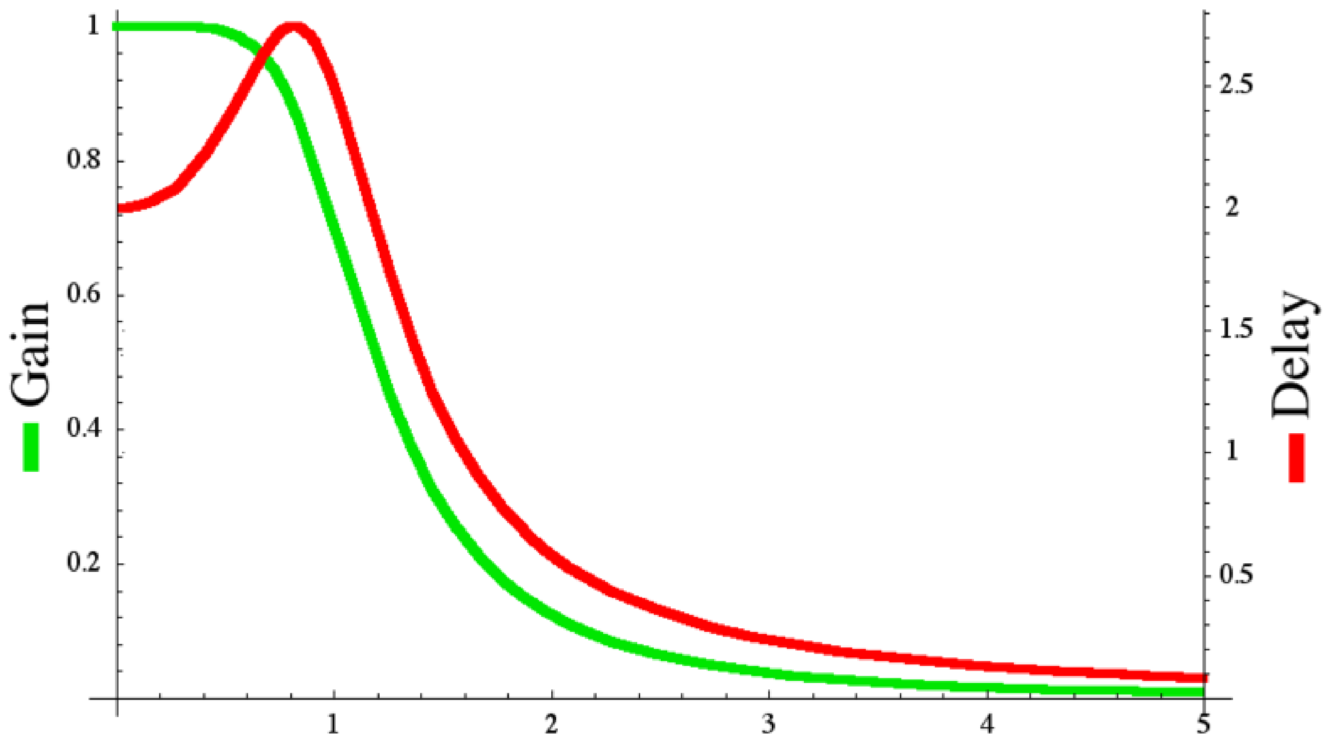
- The group delay is the derivative of the phase with respect to angular frequency
- It is a measure of the distortion in the signal introduced by phase differences for different frequencies.

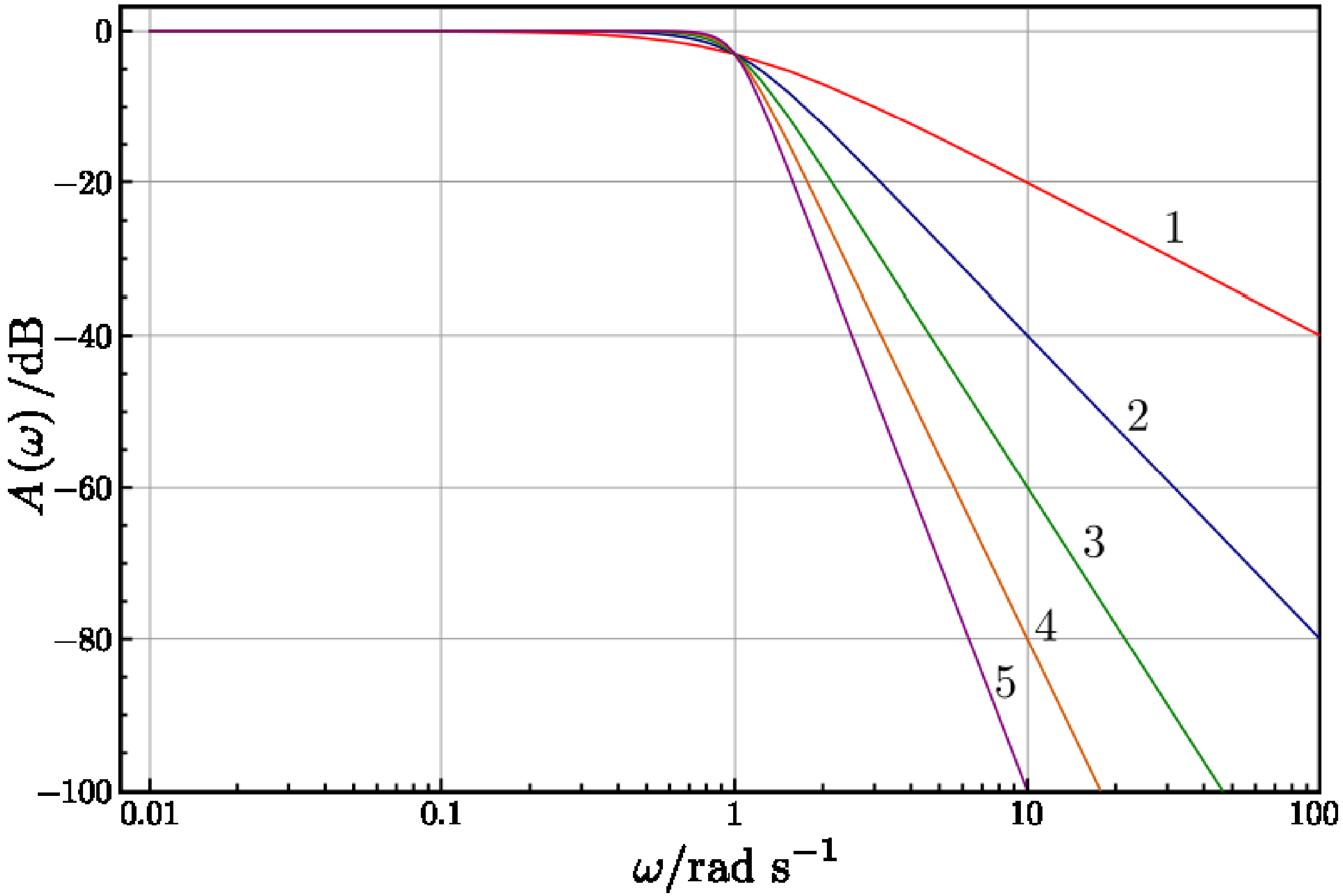
$$\tau_g = -d[\arg(H(e^{j\omega}))] / d\omega$$

Butterworth Filter

- Maximally flat (has no ripples) in the passband, and rolls off towards zero in the stopband.
- When viewed on a logarithmic plot, the response slopes off linearly towards negative infinity.
- Butterworth filters have a monotonically changing magnitude function with ω.
- First described by British engineer Stephen Butterworth in "On the Theory of Filter Amplifiers", Wireless Engineer, vol. 7, 1930, pp. 536-541.

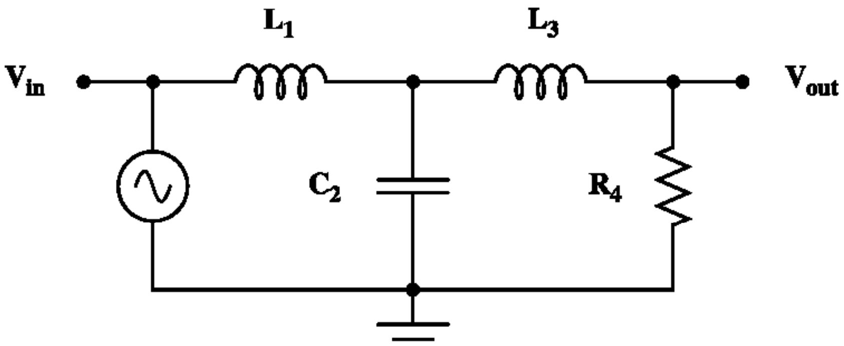
3. Order Butterworth with ω_p=1



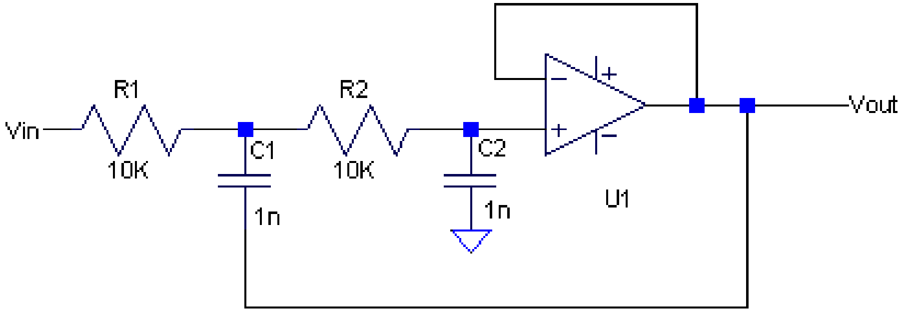


Plot of the gain of Butterworth low-pass filters of orders 1 through 5.
Note that the slope is 20n dB/decade where n is the filter order.

Analog Butterworth



3. order passive low pass filter (Cauer topology).



2. order active filter (Sallen-Key topology)

Butterworth

- Frequency response:

$$G^2(\omega) = |H(j\omega)|^2 = \frac{1}{1 + (\frac{\omega}{\omega_0})^{2n}}, s = \sigma + j\omega$$

- Transfer function

$$H(s) = \frac{1}{B_n(s)}$$

- Butterworth polynominals (n even, n odd)

$$B_n(s) = \prod_{k=1}^{\frac{n}{2}} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\pi\right) + 1 \right] \quad B_n(s) = (s+1)\prod_{k=1}^{\frac{n-1}{2}} \left[s^2 - 2s\cos\left(\frac{2k+n-1}{2n}\pi\right) + 1 \right]$$

Butterworth

n	Factors of Polynomial $B_n(s)$
1	$(s + 1)$
2	$s^2 + 1.4142s + 1$
3	$(s + 1)(s^2 + s + 1)$
4	$(s^2 + 0.7654s + 1)(s^2 + 1.8478s + 1)$
5	$(s + 1)(s^2 + 0.6180s + 1)(s^2 + 1.6180s + 1)$

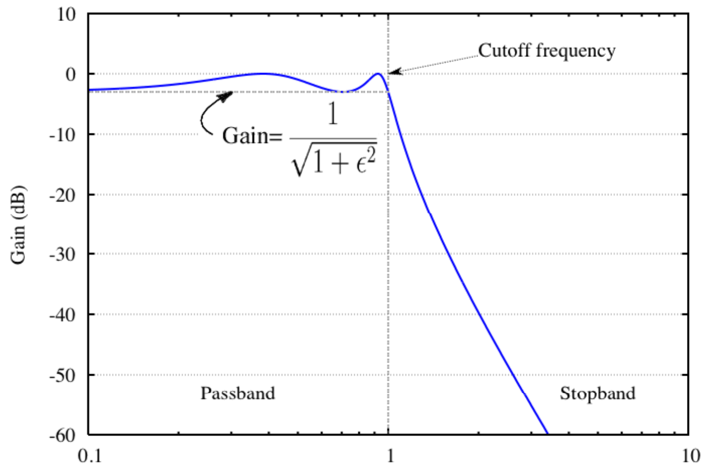
Butterworth Filter

- N'th order filter: all derivatives of the gain up to and including the 2N-1'th derivative are zero at ω=0, resulting in "maximal flatness".
- In decibels, the high-frequency roll-off is 20n dB/decade, or 6n dB/octave (not only Butterworth – all filters)

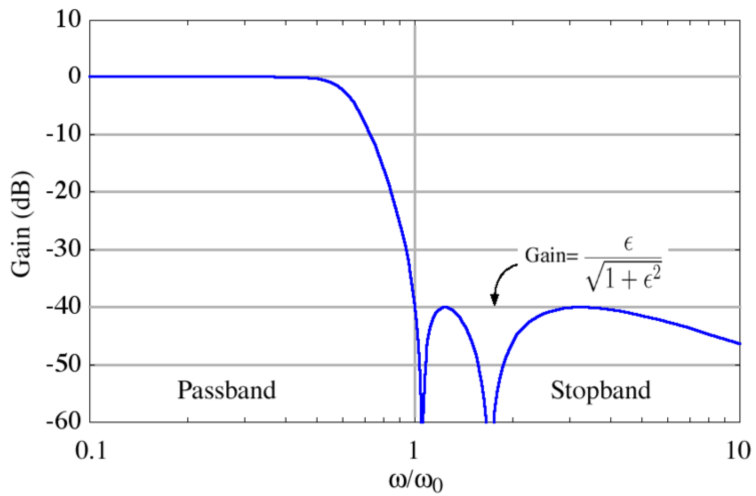
Chebyshev filter

- Norsk: Tsjebysjeff
- Steeper roll-off and more passband ripple (type I) or stopband ripple (type II) than Butterworth filters.
- Minimize the error between the idealized filter characteristic and the actual, but with ripples in the passband.
- Named after Pafnuty Chebyshev (1821-1894) Пафну́тий Льво́вич Чебышёв, because they are defined in terms of Chebyshev polynomials.

Chebyshev type I and type II

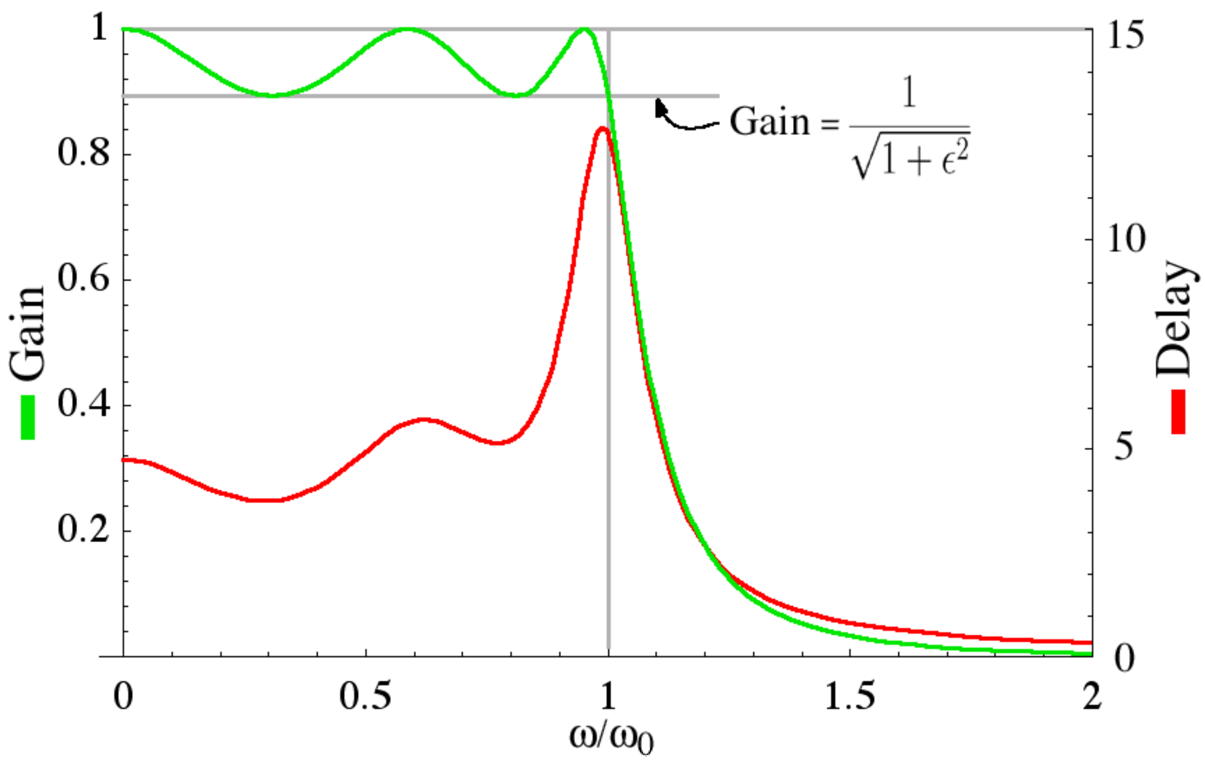


- The frequency response of a fourth-order type I Chebyshev low-pass filter with ε = 1



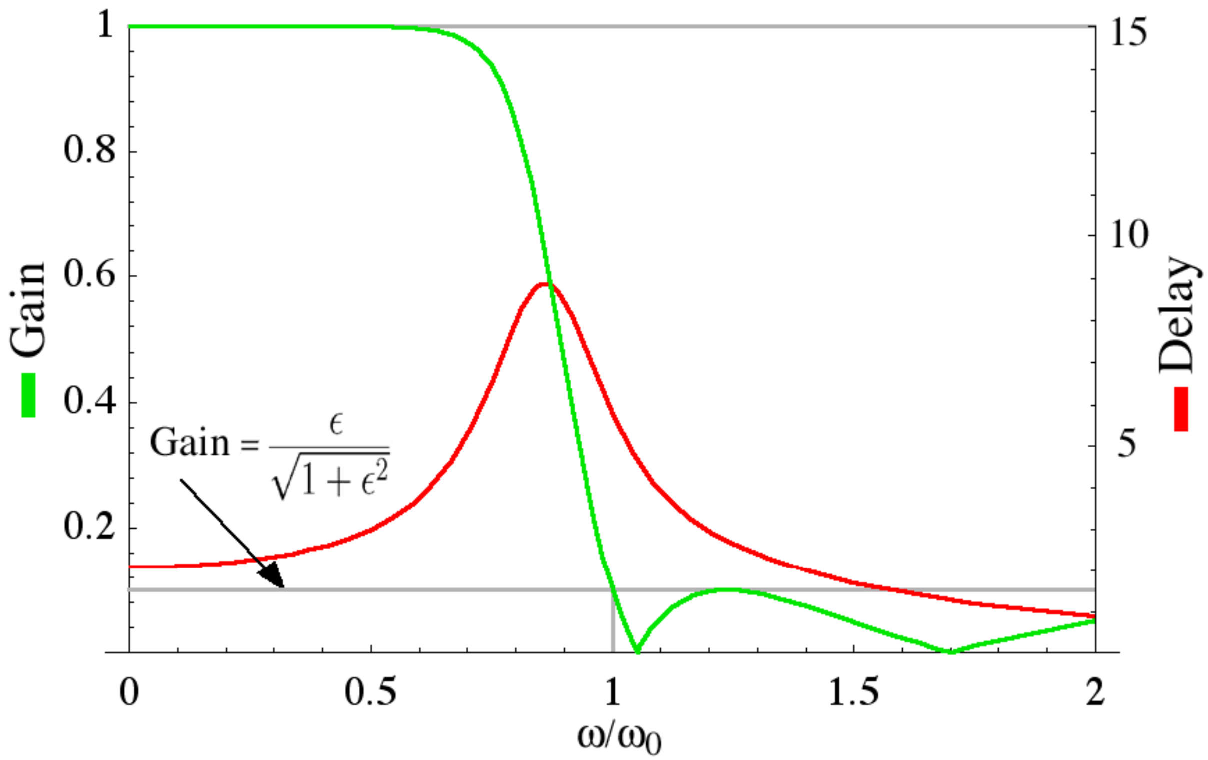
- The frequency response of a fifth-order type II Chebyshev low-pass filter with ε = 0.01

5. Order Chebyshev type I (ε=0.5)



There are ripples in the gain and the group delay in the passband but not in the stop band.

5. Order Chebyshev type II (ε=0.1)



There are ripples in the gain in the stop band but not in the pass band.

Chebyshev

- Frequency response (type I, type II)

$$G_n(\omega) = |H_n(j\omega)| = \frac{1}{\sqrt{1+\epsilon^2 T_n^2(\frac{\omega}{\omega_0})}}, or \frac{1}{\sqrt{1+\frac{1}{\epsilon^2 T_n^2(\frac{\omega}{\omega_0})}}}$$

- ε is the ripple factor, ω_0 is the cutoff frequency and $T_n()$ is a nth order Chebyshev polynomial.

$$T_n(x) = \cos(n \cos^{-1} x)$$

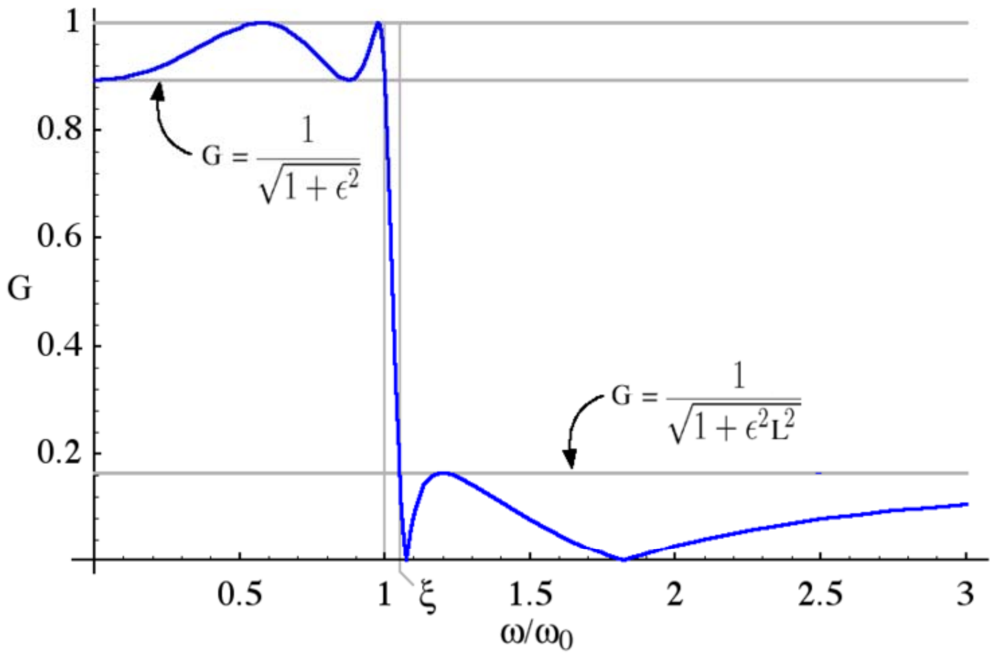
Elliptic Filters

- An elliptic filter (also known as a Cauer filter) has equalized ripple (equiripple) behavior in both the passband and the stopband.
- The amount of ripple in each band is independently adjustable.
- No other filter of equal order can have a faster transition between the passband and the stopband, for the given values of ripple (whether the ripple is equalized or not).

Cauer

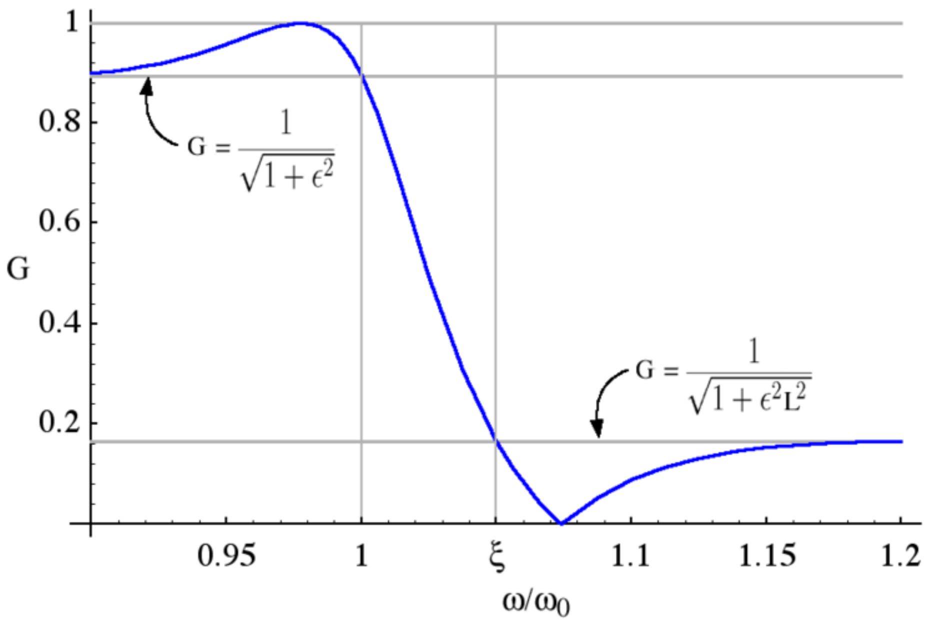
- Wilhelm Cauer (June 24, 1900 – April 22, 1945) was a German mathematician and scientist.
- He is most noted for his work on the analysis and synthesis of electronic filters and his work marked the beginning of the field of network synthesis.
- Prior to his work, electronic filter design was an art, requiring specialized knowledge and intuition. Cauer placed the field on a firm mathematical footing, providing a theoretical basis for the rational design of electronic filters.

4. Order Elliptic Filter



- The frequency response of a fourth-order elliptic low-pass filter with ε=0.5 and ξ=1.05.
- Also shown are the minimum gain in the passband and the maximum gain in the stopband, and the transition region between normalized frequency 1 and ξ

4. Order Elliptic Filter



- A closeup of the transition region of the previous plot

Elliptic filter

- Frequency response:

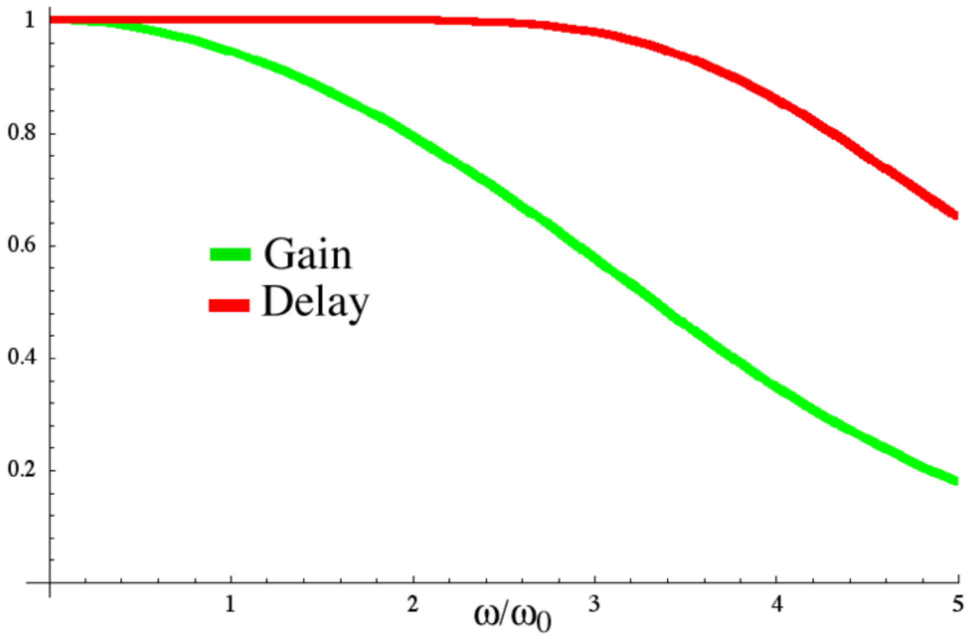
$$G_n(\omega) = |H_n(j\omega)| = \frac{1}{\sqrt{1 + \epsilon^2 R_n^2(\xi, \frac{\omega}{\omega_0})}}$$

- where $R_n()$ is the nth-order Jacobian elliptic rational function and ω_0 is the cutoff frequency, ε is the ripple factor, ξ is the selectivity factor

Bessel Filter

- A Bessel filter is a filter with a maximally flat group delay ($\approx$ linear phase response).
- Analog Bessel filters are characterized by almost constant group delay across the entire passband.
- The filter that best preserves the wave shape of filtered signals in the passband.
- Named after Friedrich Bessel (1784–1846) as the filter polynomial is expressed with Bessel functions

4. Order Bessel Filter



- A plot of the gain and group delay for a fourth-order low pass Bessel filter.
- Note that the transition from the pass band to the stop band is much slower than for other filters, but the group delay is practically constant in the passband.
- The Bessel filter maximizes the flatness of the group delay curve at zero frequency.

Bessel filter

- Transfer function

$$H(s) = \frac{\theta_n(0)}{\theta_n(s/\omega_0)}$$

- where $\theta_n(s)$ is a reverse Bessel polynomial, ω_0 is the cut-off frequency

- No Matlab function

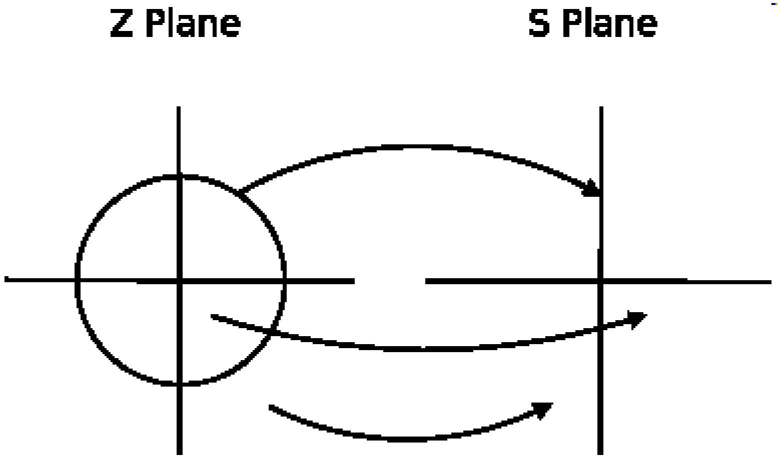
Comparison

- Butterworth: maximally flat amplitude response
- Bessel: maximally flat group delay
- Compared with a Chebyshev Type I/Type II filter or an elliptic filter, the Butterworth filter has a slower roll-off, and thus will require a higher order to implement a particular stopband specification.
- However, Butterworth filter will have a more linear phase response in the passband than the Chebyshev Type I/Type II and elliptic filters.
- Chebyshev filters are sharper than the Butterworth filter; they are not as sharp as the elliptic one, but they show fewer ripples over the bandwidth.
- Elliptic filters are sharper than all other filters, but they show ripples on the whole bandwidth.

IIR filters

Transform analog prototype to digital domain

- s-plane to z-domain: $s=\sigma+j\Omega \Leftrightarrow z=e^{j\omega}$
- Frequency axis: $s=j\Omega$ maps to $|z|=1$
- Stability, causality maintained
- $\Omega=0 \Leftrightarrow \omega=0$ always: LP $\Leftrightarrow$ LP



http://en.wikibooks.org/wiki/Digital_Signal_Processing/Bilinear_Transform

Transform from s-plane to z-plane

1. Impulse invariance
   - Impulse response is a sampled version of the analog one
   - Aliasing as $\Omega=\Omega_s \Leftrightarrow \omega=\pi$
   - We are not fond of dealing with aliasing, avoid it if we can
2. Bilinear transform
   - Let $\Omega=\infty$ be mapped to $\omega=\pi$
   - Nonlinear transform to go from H(s) to H(z)

$$s = \frac{2}{T_d}\frac{1-z^{-1}}{1+z^{-1}}$$

IIR Design: Transform mellom analog og digital

- Rett fram: sampling av impulsresponsen, h(t) til $h_s[n]$ ⇔ Impulsinvarians-metoden

- Konsekvens: aliasing for alle deler av frekvensresponsen som er over F=S/2

Impulsinvarians

- $S=1/t_s \Rightarrow h_s[n]=t_s\, h(nt_s) \Rightarrow$

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} H_a(j\frac{\omega}{T} + j\frac{k2\pi}{T})$$

- Hvis H_a er båndbegrenset, så $Ha(j\omega)=0$ for $|\omega|>\pi/T$:

$$H(e^{j\omega}) = H_a(j\frac{\omega}{T})$$

Impulsinvarians

- $z=e^{st} = e^{\sigma t_s} e^{j\Omega t_s}$

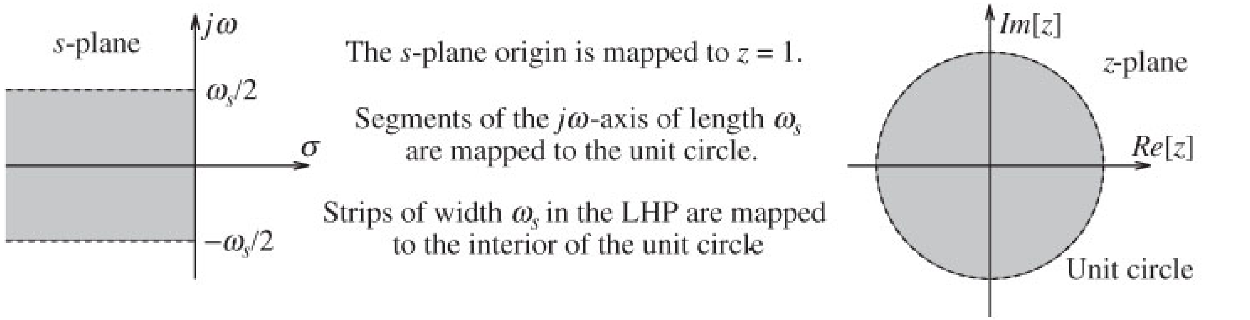


FIGURE 9.2 Characteristics of the mapping $z \Rightarrow \exp(st_s)$. Each strip of width ω_s in the left half of the s-plane is mapped to the interior of the unit circle in the z-plane. Each segment of the $j\omega$-axis in the s-plane of length ω_s maps to the unit circle itself. Clearly, the mapping is not unique

- Frekvensaksen $s=j\Omega$ transformeres til $|z|=1$
- Stabilitet, kausalitet beholdes
- $\Omega=0 \Leftrightarrow \omega=0$ alltid: LP $\Leftrightarrow$ LP
- Aliasing for alle frekvenser over S/2

Bedre: Bilineær transform (kap 9.6)

- La $\Omega = 0$ bli transformert til $\omega=0$
- La $\Omega=\infty$ bli transformert til $\omega=\pi$
- Ikke-lineær transform fra H(s) til H(z):

$$s = \frac{2}{T_d}\frac{1 - z^{-1}}{1 + z^{-1}}$$

  - Sjekk verdier for z=1, -1, $\pm$ j
- Ekvivalent transform mellom frekvenser:

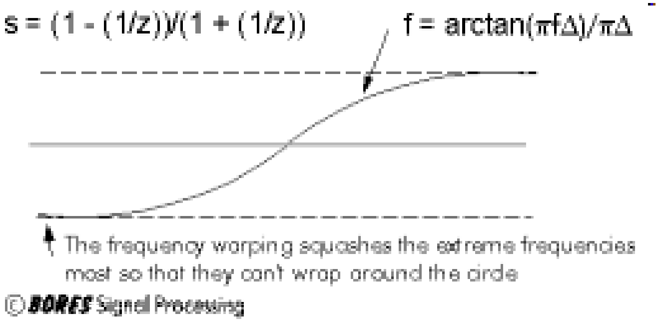
$$\Omega = \frac{2}{T_d}\tan(\omega/2)$$

  - Vises ved å sette inn $z=e^{j\omega}$

Bilineær transform

- Ingen aliasing, beholder stabilitet og kausalitet, men forvrenger frekvensaksen:

$$\Omega = \frac{2}{T_d}\tan(\omega/2)$$



- Best for lavpass, minst endring av passbånd
  - For små ω: $\Omega \approx \omega/T_d$
- Derfor designes først propotyp lavpassfiltre
  - For HP, BP etc: Start med analog LP => digital LP => digital HP etc

Matlab IIR Digital Filter Design

- Order Estimation
- For IIR filter design using bilinear transformation, MATLAB statements to determine the order and bandedge are:

```
[N, Wn] = buttord(Wp, Ws, Rp, Rs);
[N, Wn] = cheb1ord(Wp, Ws, Rp, Rs);
[N, Wn] = cheb2ord(Wp, Ws, Rp, Rs);
[N, Wn] = ellipord(Wp, Ws, Rp, Rs);
```

Matlab IIR Digital Filter Design

- Filter Design
- For IIR filter design using bilinear transformation, MATLAB statements to use are:

```
[b, a] = butter(Nb, Wn)
[b, a] = cheby1(Nc, Rp, Wn)
[b, a] = cheby2(Nc, Rs, Wn)
[b, a] = ellip(Ne, Rp, Rs, Wn)
```

- No need to think about bilinear transform
- Transfer function can be computed using freqz(b, a, w) where w is a set of angular frequencies

Matlab IIR Digital Filter Design

- Design an elliptic IIR lowpass filter with F_p=0.8 kHz, F_s=1 kHz, F_{sample}=4 kHz, α_p = 0.5 dB, α_s=40 dB
- Code fragments used are:

```
[N,Wn] = ellipord(0.8/(4/2), 1/(4/2), 0.5, 40);
```

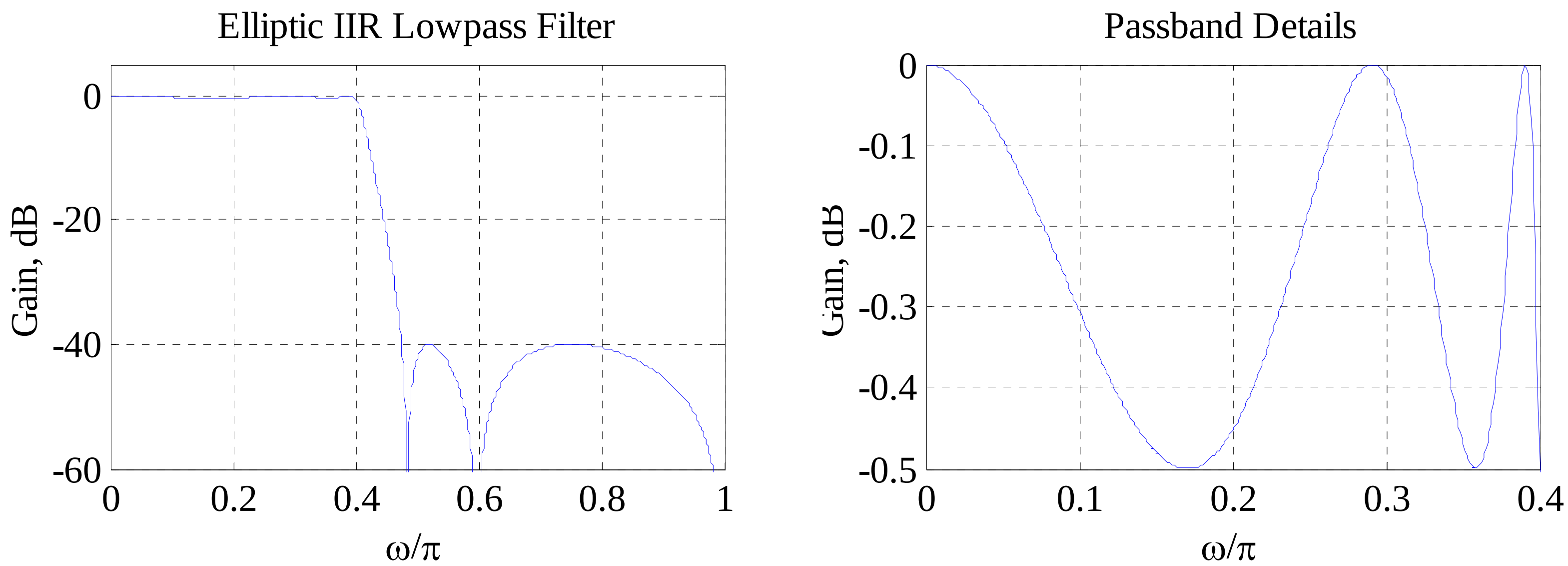
  – Result: N=5. order, W_n=0.4

    (compare with N_b=18, N_c=8)

```
[b, a] = ellip(N, 0.5, 40, Wn);
```

- Full example: `IIRdesign.m`

Matlab IIR Digital Filter Design



9.8 Effekter av endelig ordlengde

- Koeffisienter blir avkortet
  - Som regel blir det en liten endring av frekvensrespons
  - Katastrofal feil: en pol innenfor |z|=1 kan havne utenfor (bare IIR)
- Aritmetikken foregår med endelig presisjon
  - Filteret blir et ikke-lineært system
  - Kvantiseringsstøy og avrundingsstøy
  - Feil pga overstyring
- Limit cycles: Oscillasjoner på utgangen uten inngang
  - Bare i IIR da det trenger tilbakekobling
  - Fullskala oscillasjoner hvis overstyring folder rundt ($x>x_{max} \Rightarrow -x_{max}$) i stedet for metning ($x>x_{max} \Rightarrow x_{max}$)
  - Små oscillasjoner hvis avrunding istedet for avkorting
  - http://cnx.org/content/m11928/latest/