Optical digital filtering, splitting and combining

The theory about digital filters has been introduced. The digital filter designed is needed as a first step to the optical design of the desired filter, to be able to get the coefficients that will be mapped into the optical filter parameters. But some differences exist between both designs:

- In the digital procedure, the filter is designed using a low pass filter, which means that we start to work with a baseband signal. However, the optical filter is designed to work in a high frequency band, with a repeated spectrum every FSR (Free Spectral Range).
- In the digital procedure, the coefficients that we can obtain are real, whereas in the optical design the coefficients will be complex.
- In digital filters, we have introduced the complexity factor as a number of filter stages (or degree of the polynomials). In optical designs the complexity is not only the number of stages but the number of tuning elements (heaters) also (combination of both), which is explained later

Optical Filter principles:

The Filter functions arise from the interference of two or more waves that are delayed relative to each other. The incoming signal is split into multiple paths by a division of the wave front or amplitude. And after travelling along different paths (in length) the fields are combined. For this combination it is needed that the waves have the same polarization, frequency and be temporally coherent over the longest delay length. And when the combination arises their relatives phases determine if the interference is constructive or destructive. The phase φ for each path is the product of the distance traveled L and the propagation constant **β (β = 2πn/λ)**, where n is the effective index, so the phase for each path is then expressed as a multiple of β L, $\boldsymbol{\varphi_N = N \beta L}$ where N is an integer. The key to analyzing optical filters using Z-transforms is that each delay be an integer multiple of a unit delay length L. The total electrical field is at the end the sum of the all optical paths given by

$$\mathbf{E = E_0 e^{-j\varphi 0} + E_1\ e^{-j\varphi 1} + E_2 e^{-j\varphi 2} + \ldots}$$

where the complex mode amplitude is denoted by E. To obtain the Z-transform, the phase must be expressed as a multiple of the unit delay T, where T is giving by

$$\mathbf{T = nLc}$$

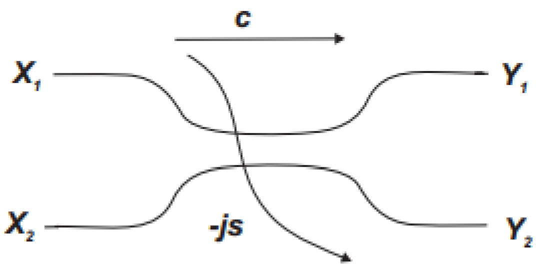
where c is the speed of the light. Rewriting β L in terms of Ω yields **β L = Ω[ne(Ω)L/c] = ΩT(Ω)** For a dispersion less delay line, T is a constant and we obtain is

$$\mathbf{E = E0 + E1z{-}1 + E2z{-}2 + \ldots}$$

by substituting z−1 for$e^{-j\Omega T}$. As it was said in Section 3.1, the optical frequency response is periodic with FSR of 1/T. The center frequency **fc = c/λc** is defined so that L is equal to an integer number of wavelengths, **mλc = nL**. At λc, the contributions from each path of length N L differente by 2π and thus add constructively in the absence of any other source of relative phase difference.

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} c & \text{-j}s \\ \text{-j}s & c \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

where c =$\sqrt{1-k}$ and -js = -j$\sqrt{k}$, X1 and X2 are the inputs of the coupler and Y1 and Y2 are the outputs.



Directional coupler and its parameters.

Digital filter complexity

At this point, an important factor to keep in mind can be introduced: the complexity of the filter. In digital design, the complexity of the filter is given by the number of the stages, in other words, the degree of the polynomials in numerator and denominator in . This number of stages depends mainly on the slope of the filter response: the steeper the response is (a narrower transition band) the higher the order of the filter will be. And it is an important issue if we want to fabricate the filter in a limited space (in our case, integrated on a chip).

Poles and Zeros

The filter input and output are related by a weighted sum of inputs and previous outputs. The relationship for a discrete linear system with a discrete input signal is as follows:

$$y(n) = b_0x(n) + b_1x(n-1) + \ldots + b_Mx(n-M) - a_1y(n-1) - \ldots - a_Ny(n-N)$$

The weights are given by the a and b coefficients. The Z-transform of this discrete sequence results in:

$$y(n) = b_0 + b_1z^{-1} + b_2z^{-2} \ldots + b_Mz^{n-M} - a_1z^{-1} - \ldots - a_Nz^{n-N}$$

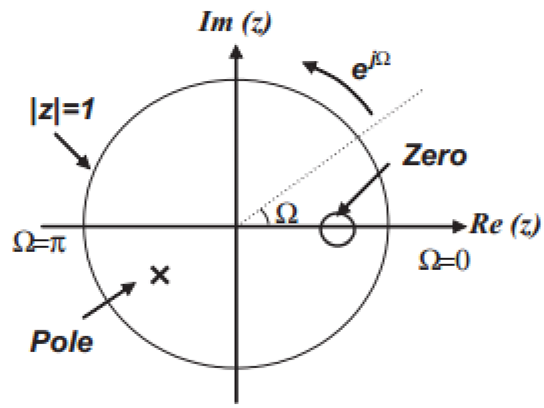
and the transfer function can be obtained, that is a ratio of polynomials:

$$H(z) = \frac{\sum_{m=0}^{M} b_m z^{-m}}{1 + \sum_{n=1}^{N} a_n z^{-n}} = \frac{B(z)}{A(z)}$$

where A(z) and B(z) are M th and N th order polynomials in the denominator and nominator of H(z), respectively. The expression for H(z) can also be written in term of the roots of the polynomials as follows:

$$H(z) = \frac{\Gamma z^{N-M} \Pi_{m=1}^{M} (z - z_m)}{\Pi_{n=1}^{N} (z - p_n)}$$

The roots are complex numbers and are given different names depending on whether they are from the numerator or denominator polynomials. The zeros of the numerator, also called the zeros of H(z), are reprinted by zm. Zeros that occur on the denominator polynomial, designated by pn, are called poles. The gain is set by Γ. For passive filters, the transfer function can never be greater than 1, so Γ has a maximum value determined by max**{| H(z) |z=ejΩ } = 1**. A pole-zero diagram is a convenient way to reveal the locations of the poles and zeros of a filter in the complex plane. The unit circle is also depicted in the diagram and one trip around it corresponds to 2×π (one FSR in optical filters). An example pole-zero diagram is shown in Figure

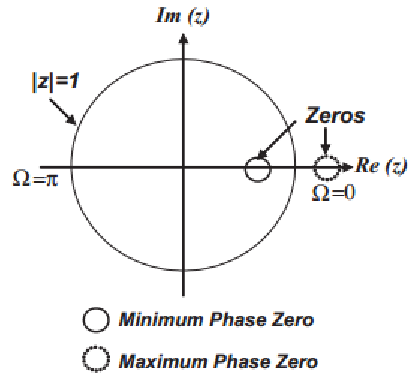


Poles and Zeros diagram.

Digital filters are classified by the polynomials. A Moving Average (MA) filter has only zeros and is also referred to as a Finite Impulse Response (FIR) filter. It consist only of feed-forward paths. An Autoregressive (AR) filter has only poles and contains one or more feedback paths. A pole produces an impulse response with an infinite number of terms in contrast to the finite number of terms of MA filters. Filters with both poles and zeros are referred to as an Autoregressive Moving Average (ARMA) filters. An Infinite Impulse Response (IIR) filter must contain at least one pole. Since the IIR designation is ambiguous with respect to whether a filter has a zero or not, the MA/AR/ARMA terminology is used.

Linear Phase Filters

An important class of filters are those having linear-phase. Those filters have a constant group delay and thus are dispersion-less. No phase distortion is induced on to the signal. A distortion-less filter has a magnitude response that is flat across the frequency band of the input signal and the phase response in the passband region is a linear function of frequency. Linear phase filters are important in applications where no phase distortion is allowed. A Moving Average filter or FIR filter can be designed to have a linear phase. Following with the previous subsection, if all the zeros are inside the unit circle, a minimum-phase response results. Conversely, if all the zeros are outside the unit circle, a maximum-phase response results. The minimum-phase system implies a minimum



Maximum and Minimum phased zeros representation.

group delay function whereas the maximum group delay occurs in a maximum-phase system. As an aside, all stable AR filters are minimum-phase. Linear-phase filters can have very sharp transitions in the magnitude response, although many stages are required, without introducing dispersion.

Digital filter design procedure

So far the theory about filter description has be introduced. Now a short explanation is made about the procedure itself. Actually several approaches exists for the design of digital filters. some of them can be summarized in the following subsections, because they will be useful in, where the optical design procedure (together with the digital one) is detailed.

Classical design methods for Autoregressive Moving Average filters The classical methods for filter design can only be used when the filter to be designed has the same number of poles and zeros in its response, which means the same polynomial order both in the numerator and the denominator. Then, the steps to design a digital filter can be summarized as follow:

1. The first step is to choose the desired spectral range as well as the ripple conditions in the pass band and the attenuation.
2. Starting with the procedure, normalize the frequency range is needed, between 0 and 2π, where the 2π is the Frequency Spectral Range (FSR). And afterwards, normalize the cut-off and stop band frequencies.
3. The next step is just calculating the coefficients in the Laplace domain (analog filters), with the formulas we have from each filter type (see next subsections).
4. After that, it is necessary to change to the Z-domain in order to work with delays (to make the filter digital), using the bilinear transform, and get the poles and zeros. This bilinear transformation arises from the relation between the frequency response of the analog filter H(s) and the digital filter H(z).
   We can make the equality $H(z = e^{j\omega t} = H(j_{\Omega}(\omega))$ where $\Omega(\omega)$ is a nonlinear function called frequency warping. Then the bilinear transform can be expressed as

$$z = \frac{1+s}{1-s} \Rightarrow s = \frac{z-1}{z+1}$$

and the frequency warping is obtained by setting z = $e^{j\omega t}$

$$j\Omega(\omega) = \frac{\exp^{j\omega t} - 1}{\exp^{j\omega t} + 1} = j\frac{\sin(\omega t/2)}{\cos(\omega t/2)}$$

5. With these poles and zeros, we have the transfer function of the filter (the polynomials), therefore, the filter is designed.

The process to design the filter is started using a prototype analog low-pass filter H(s), with cut-off frequency $\omega c = 1$. For each class of analog filter, there are equations which constrains the possible values of ε, δ, ω1 and ω2. These equations also involve the filter order n.

Butterworth method

Butterworth filters have a magnitude response which is maximally flat near the center frequency, and declines monotonically for $\omega > 0$. The order n of a Butterworth filter is characterized by

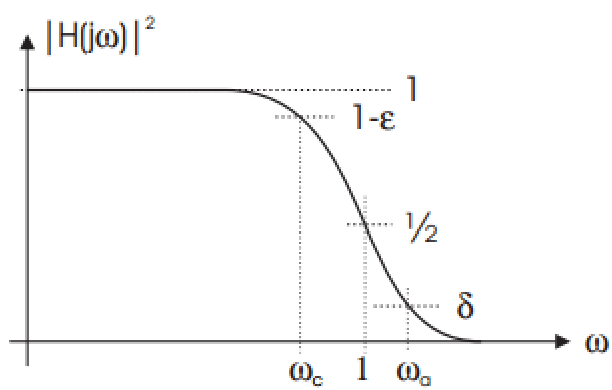
$$\|H(j\omega)\|^2 = \frac{1}{1 + \omega^{2n}}$$

where n is the order of the filter. The Butterworth squared magnitude response is monotonely decreasing for $\omega > 0$, is equal to 1 for $\omega = 0$ and to 1/2 at the cutoff frequency $\omega = 1$, and approaches 0 as ω approaches infinity. It can be seen in Figure below. The poles must satisfy the equation

$$1 + (js)^{2n} = 0$$

obtained by setting $\omega = s/j = -js$. The roots of this polynomial lie on the unit circle and are given by

$$\lambda_k = e^{j\Omega_k}, \theta_k = \frac{\pi}{2}(1 + \frac{2k - 1}{n})$$



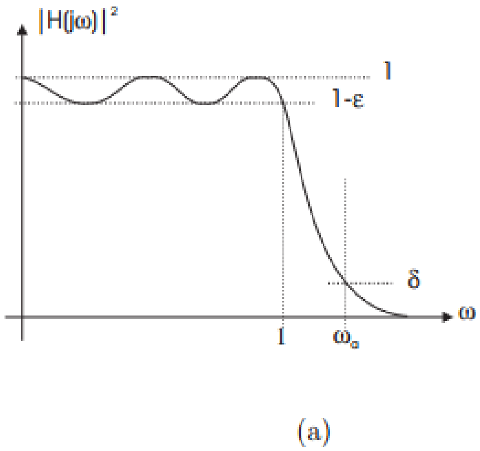
Transfer function of a Butterworth filter.

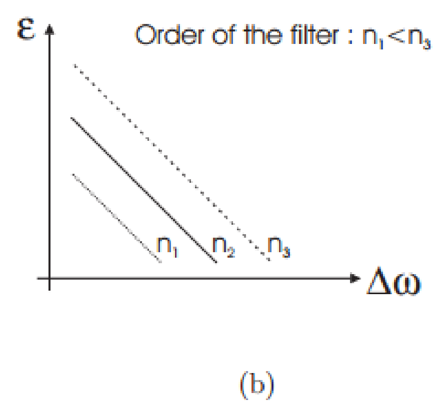
Chebyschev method

Chebyschev filters have a magnitude response which exhibits equal ripple for frequencies between 0 and 1, and declines monotonically for ω >1. The order n Chebyschev filter is characterized by

$$\|H(j\omega)\|^2 = \frac{1}{1 + \frac{\varepsilon}{1+\varepsilon} T_n^2(\omega)}$$

where Tn is the Chebychev polynomial, that can be easily founded in filter design literature. There is a tradeoff between ripple magnitude ε and transition bandwidth ωu − 1. It can be seen in Figure below





(a)Transfer function of a Chebyschev filter, (b) The tradeoff between the ripple magnitude and the transition band in the filter.

The poles of the transfer function satisfy the polynomial equation

$$1 + \frac{\varepsilon}{1-\varepsilon} T_n^2(-\mathrm{j}s) = 0$$

The poles may be shown to be

$$\lambda_k = \sinh(\mu)\cos(\theta_k) + \mathrm{j}\cosh(\mu)\sin(\theta_k)$$

Where

$$\theta_k = \frac{\pi}{2}(1 + \frac{2k-1}{n})$$

And

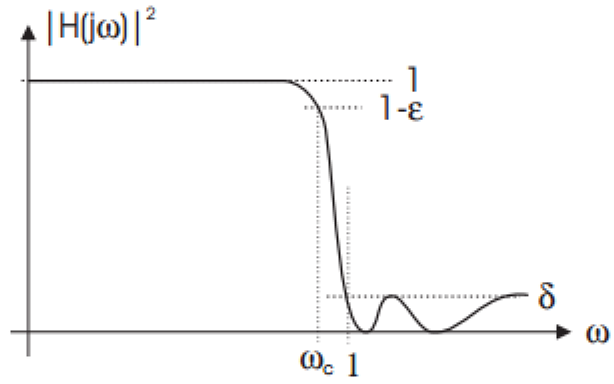
$$\mu = \frac{1}{n}\sinh^{-1}(\sqrt{\frac{1-\varepsilon}{\varepsilon}})$$

Inverse Chebyschev method

Inverse Chebyschev filters have a magnitude response which is monotone decreasing for frequencies between 0 and ωc and exhibits equal ripple for frequencies ω >1. The order n inverse Chebyschev filter is characterized by

$$\|H(j\omega)\|^2 = \frac{\frac{\delta}{1-\delta}T_n(\omega^{-1})^2}{1+\frac{\delta}{1-\delta}T_n(\omega^{-1})^2}$$

The transfer function can be seen in Figure below



Transfer function of a Inverse Chebyschev filter.

The zeros of the transfer function satisfy the polynomial equation

$$T_n(\mathrm{j}/s) = 0$$

The zeros may be shown to be

$$\omega_k = \frac{1}{\cos(\frac{(2k-1)\pi}{2n})}$$

The poles satisfy

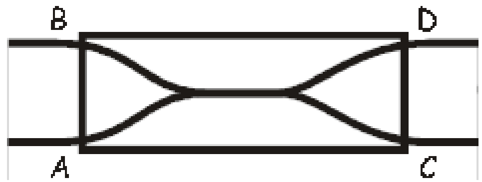
$$1 + \frac{\delta}{1-\delta}T_n(\frac{\mathrm{j}}{2})^2$$

Splitting and Combining

Electrically connecting two uninsulated copper wires together is as simple as bringing them into contact at any one point along their length. This allows electrical signals traveling along the wire to be readily divided and sent to tow or more destinations. However the splitting of light signals traveling a long optical fiber is not so simple. Optical fiber is designed to contain the light within the core (using total internal reflection) and guide it to the other end. This means that very little of the light that enters the fiber can escape along its length. Clearly then, splitting light signals cannot be achieved by simply “ tapping ” into the fiber at one point in the same way as for copper conductors.

One method of achieving optical splitting relies on the fact that, in practice, a small amount of light does escape from small cored glass fiber. That being the case, it’s possible to transfer some of the light from one fiber to another by placing them sufficiently close together over a sufficient length. An obvious variation on this idea involves increasing the closeness of the coupling (and thereby reducing the length over which the coupling must be done) by fusing the fibers’ cores

together. The optical device designed to split light in this way is called a fused fiber coupler and Emona FOTEx has tow of them.
The construction of fused fiber couplers is reflected in their schematic symbol as shown in figure below.



Finally the fused-fiber coupler can be used to combine signals instead of splitting them. To explain, consider the example of a signal connected to A, form the discussion so far we know that the signal appears on both parts D (strong) and c (weak). Now , if another signal is connected to port B at the same time, that signal must also appear on both ports D (weak) and C (strong). Clearly, both of the output ports consist of light from both of the sources and have combined. This is a very handy feature if these devices that is used in later experiments.