Multi-dimensional Photonic Processing, A Discrete-Domain Approach: Part I – Design Techniques

Le Nguyen Binh
Department of Electrical and Computer Systems Engineering,
Monash University, Clayton Victoria 3168 Australia.

Abstract
The information transport infrastructure is currently based and directionally developed towards super-dense wavelength division multiplexing with provisional routing and switching in spatial (e.g. routing to different sub-network elements) and spectral (e.g. multiplexing/demultiplexing, filtering, adding/dropping of wavelength channels) domain. The processing of photonic signals is becoming very important in these diverse domains. This paper is the Part I of three parts which has sought to integrate the fields of discrete signal processing and fiber-optic signal processing, integrated photonics and/or possibly nano-photonics to establish a methodology based on which physical systems can be implemented. Because fiber-optics is essentially one-dimensional planar medium, the methodology has been proposed in order to implement 2-D signal processing using 1-D sources and processors. A number of 2-D filter design algorithms are implemented. These algorithms are applicable to photonic filters that perform 2-D processing. The developed 2-D filter design methods are generic allowing the proposals of several photonic signal processing (PSP) architectures in Part II and part II to enable efficient coherent lightwave signal processing.
Key words:
Multi-dimensional, Photonic, Design

1. Introduction to Photonic Signal Processing

In recent years, there has been a notable increase in the number of applications that require an extremely fast signal processing speed that cannot be met by current all-electronic technology. Photonic signal processing (PSP) opens the possibilities for meeting the demands of such high-speed processing by exploiting the ultra high bandwidth capability of lightwave signals with specific applications in the field of photonic communications and fiber optic sensor networks.

1.1 Photonic Signal Processing: A brief overview

The field of signal processing is concerned with the conditioning of a signal to fit certain required characteristics such as bandwidth, amplitude, and phase. Conventional techniques of signal processing make direct or indirect use of electronics. For example, frequency filtering, a most important signal processing procedure, can be performed through direct electronic means such as tunable IC filters or indirectly by digitizing the input for subsequent processing by computers or special purpose digital signal processing chips. Although a high performance can be obtained using either of the techniques, electronic methods suffer from physical limitations that govern the maximum processing speed. The demands for high performance beyond that achievable by electronic means have been increasing recently due to the increase in computationally demanding real-time processing applications.
Using lightwaves instead of electronic signals as the information carrier in signal processing is an appealing concept.
The full potential of the technology has been accelerated in recent years due to the invention and discovery of photonic crystals. Several important advances have been made in utilizing light as the information carrier including real-time spatial-light modulators and electro-optic devices, micro-ring resonators, photonic crystal fibers, guided wave crystal photonics, super-prisms. Another incentive for using light as the information carrier is the superiority of fiber-optic communication systems which offer the wide bandwidth properties of photonic fiber medium. To fully exploit the capability of photonic systems, PSP is very essential.
The field of photonic signal processing (PSP) can be divided into two distinctive approaches which are outlined in the following sections.

1.2 Spatial and Temporal Approach

The first use of lightwaves for signal processing applications was developed as early as in 1968 when an “integrated photonic correlator” {1(a)] and laser inscription device [1(b)] consisting of spatial light modulators and lenses in a planar waveguide was suggested. Further developments along this line were made and several experimental devices including acousto-optic spectrum analyzer, a time-integrating acousto-optic correlator, a hybrid electro-optic/acousto-optic vector multiplier, a high-speed electro-optic analog-to-digital

Manuscript received May 5, 2009
Manuscript revised May 20, 2009

converter, and several fiber delay-line processors [2] were demonstrated.

The advantage of the spatial and temporal approach over the conventional electronic approach can be seen in light of the fact that lenses which are 2-D devices, have Fourier transform properties and can therefore act as a massively parallel Fourier transform processor. Taking advantage of the massive parallelism can mean the removal of the Von-Neumann bottleneck of present-day digital computers. Although an all-photonic computer does not seem feasible in the near future, a hybrid photonic-electronic computer offering ultra-high speed processing capability that could be realized by combining photonic information processing for some specific functions and electronics for general operation [2].

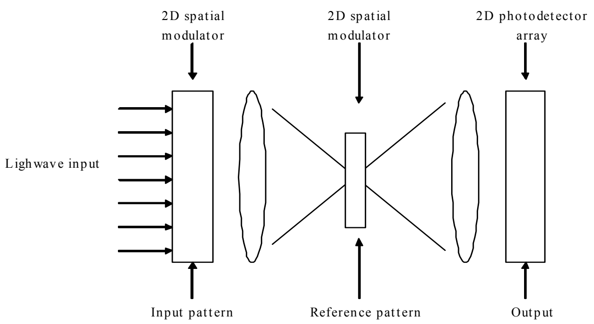


Figure 1: Spatial Fourier optical signal processor. The photonic active components are the acousto-optic diffractors.

The drawback with spatial and temporal approach is the fact that the signal processing is performed in analogue manner. As shown in *Figure 1*, lightwaves carrying different signals must travel through different media therefore suffering acoustic diffraction resulting in crosstalk [2]. It is interesting to note that using holographic techniques, several layers of neural nets can be implemented with each layer in parallel format making spatial and temporal approach a suitable technique for neural network implementation. Although this technique may be useful in implementation of opto-electronic computer, the approach is not suitable for signals that have been transmitted through photonic fiber communication networks. Such signals are sequentially linear and to be processed by a spatial and temporal processor, a conversion into a suitable 2-D format using demultiplexing devices and laser arrays will be required. The following section introduces a technique which is ideal for lightwave signals from guided media such as photonic fibers and photonic crystals.

The spatial structures can be translated in fiber and integrated photonic forms using planar lightwave circuit (PLC) using silica-on-silicon technology, for example the array waveguide filters acting as wavelength muxes and demuxes and spatial separators.

1.3 Fiber-Optic Delay Line Approach

Guided-wave photonics and fiber optics provide alternative architectures for PSP to the classic spatial or time integrating architecture introduced in *Section 1.2*. The main advantage of guided-wave systems over spatial and temporal system is the wide bandwidth property available with photonic fiber transmission medium. For example, a silica fiber with a nominal 5μs delay can store 1 GHz bandwidth signals for time periods less than one millisecond [2]. Another advantage of guided wave optics can be stated as the elimination of acoustic diffraction. However, since photonic fiber is essentially a 1-D medium (signal propagates along one axis - that of the fiber), this architecture sacrifices the 2-D nature of light that is utilized in time and space integrating architectures. In effect, in guided-wave systems, the advantage of massive 2-D parallel processing capability of light is sacrificed for the wide bandwidth of guided wave optics which enables high speed processing. Despite this limitation which confines the use of fiber-optic technology to signals from guided lightwave transmission medium, the simple fact that the current major usage of photonic systems is in communication systems makes the technology useful as it presents the possibilities of removing the bottleneck caused by opto-electronic conversion and therefore ensuring full utilization of fiber bandwidth. So far various uses have been found for fiber-optic signal processors as frequency filters, matched filters, correlators, and waveform and sequence generators [3-10].

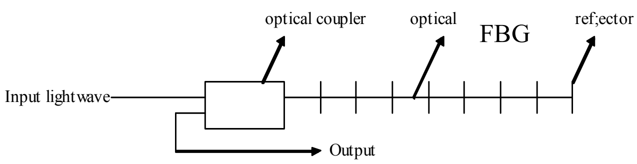


Figure 2: Fiber-optic delay line processor. The coupler can be replaced by a3-port optical circulator. The reflector can be a fiber Bragg grating (FBG).

Figure 2 shows one possible configuration of fiber-optic processor. Although filter coefficients were realized using reflectors in *Figure 2, o*ther in-line components such as photonic attenuator/amplifier can also be used for implementing filter coefficients. It is evident that the operation of fiber-optic delay line filters is similar to that of digital filters. In fact, the correct term to describe the fiber-optic signal processing would be 'discrete-time PSP' rather than digital signal processing as the range of the input or output signal is not digital at all. In any case, the discrete-time property makes it possible to apply the well developed z-transform techniques to filter design. In Section 2, the application of the z-transform techniques for analysis and design of fiber-optic systems is discussed in detail.

1.4 Motivation

The demand for multi-dimensional photonic signal processing (M-ary PSP) can be attributed to various factors due the growing feasibility of high-capacity digital transmission networks capable of transmitting ultra-high bit rate and time division multiplexing up to 160 Gb/s as well as fiber optical sensor networks.

A problem with the implementation of such systems is the lack of devices that are capable of processing an enormous amount of data associated with multi-dimensional signals. With photonic transmission networks becoming the transport infrastructure, PSP technique has become increasingly more desirable compared to O/E and E/O conversion techniques. As discussed in Section 1.3, fiber-optic signal processing systems are ideal for such processing demands for several reasons: all-optical (or photonic) processing of photonic information of optical communication systems are possible using fiber-optic signal processing; 2-D signals usually require much higher bandwidth than 1-D signals and therefore must be processed by a high bandwidth system to allow real-time performance; it is likely that future telecommunication networks would be all fiber-optic.

	Spatial and Temporal	Fiber-optic
Principle operating mode	unguided	guided
Components used	lenses, light modulators, mirrors, masks, LED or laser arrays, slits	lasers or LED's, optical fibers, optical amplifiers (OA), attenuators, reflectors
Time mode	continuous-time	discrete-time
Flexibility	hard to change configuration once developed	easy to adjust the function using different tab values
Analysis method	difficult (some Fourier transforms)	well known z-transform method
Accuracy	low	high
Cross-talk	yes	no
Major use	Photonic computing	Communication signal processing
Parallel processing capability	massive parallel processing	limited parallel processing

Table 1-1: Outline of the two different approaches to PSP

2. Multidimensional Signal Processing

Multidimensional signal processing enables processing of signals that depend on more than one co-ordinate. Although many concepts of multidimensional signal processing are straightforward extensions of 1-D signal processing theory, there are also significant differences that need to be clarified, particularly when referred to photonics. Discussions of multidimensional signal processing in this paper is limited to 2-D signal processing applicable to photonics that is by far the most important class of multidimensional signal processing.

2.1 Multi-dimensional Signal

One may define multidimensional signals as signals whose values at a certain instance of time, space, or other coordinates depend on more than one variable. In 2-D signal processing, each of the properties depends on both x and y direction and therefore the concepts of spatial signal, and therefore spatial frequency must be introduced. Spatial frequency does not depend on time, but rather depends on the spatial variations of the 2-D signal. There are two distinct spatial frequencies, one in x direction and one in y direction. 2-D signals form the most important class of multidimensional signals and methods developed for 2-D signal can be generalized to signals of larger dimensions. This paper concentrates on developing filter design methods for two-dimensional signals.

2.2 Discrete Domain Signals

A signal domain can be either continuous or discrete. For digital signal processing purposes however, it is convenient to 'sample' continuous domain signals at a discrete interval so that in effect it has a discrete domain. In 1D, signal to be processed or stored in a sequential manner can be sampled at discrete intervals of time or direction. Put into an equation form, 1D signal can be represented by a train of scaled impulses as

$$a(t) = \lim_{\Delta \to 0} \sum_{k=-\infty}^{\infty} a(k\Delta)\delta(t - k\Delta)\Delta \qquad (1)$$

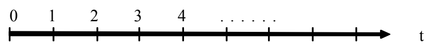
where Δ is the sampling period and n denotes the sequence number. The sampled signal can be infinite in extent and reflects this accordingly. If Δ is infinitely short, then above expression reduces to the representation of a continuous signal as expected.

In 2-D, a natural extension to (1) can be made as s

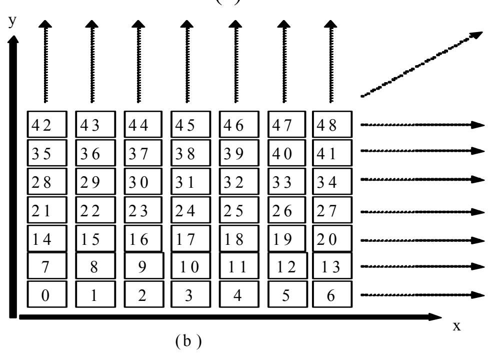
$$a(x,y)=\lim_{\Delta_1\to 0}\lim_{\Delta_2\to 0}\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} a(k_1\Delta_1,k_2\Delta_2)\,\delta(x-k_1\Delta_1,y-k_2\Delta_2)\Delta_1\Delta_2 \quad (2)$$

There is an important difference between sampling of 1D signals and 2-D signals in practice. Assuming there is only one sampling device, 1D signal such as the one shown in *Figure 3(a)* can be sampled by taking values at discrete intervals. If the signal duration is infinite in extent, no truncation is needed as the transfer function defines the limit even if the signal is not periodic. For 2-D signals with infinite duration, this is not the case. As it can be seen in Figure 3(b*)*, if the 2-D signal was sampled infinitely in one dimension, the part of the 2-D signal which extends in the other dimension will never be sampled. For 2-D signals, there is always a predefined limit on how many samples are taken in each dimension. After reaching the limit in one dimension, the coordinate on the other dimension is incremented by one sampling period and the sampling process continues until the number of samples in the first dimension again reaches the limit. The process is repeated until the pre-defined number of samples in the second dimension is reached. The consequence is that the process gives a train of sampled 2-D signals stretched out in 1D as shown in *Figure 3*(c). For 1-D discrete time processing of 2-D signals, the signal must be sampled in this way so that the processor can implement the delays z_1^{-1} and z_2^{-1} using only one dimensional delay photonic element. The limiting of sample space is similar to windowing or truncation performed on 1-D signals for some signal processing operations such as discrete Fourier transform (DFT).
Discrete space form of 2-D signal with predefined limits can be expressed by

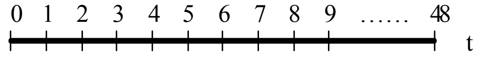
$$a[n_1,n_2]=\sum_{k_1=0}^{n_1}\sum_{k_2=0}^{n_2} a[k_1,k_2]\delta[n_1-k_1,n_2-k_2] \quad (3)$$



(a)



(b)



(c)

Figure 3 (a) Infinite extent 1-D signal.(b) 2-D signal with finite predefined limit of 7×7(each index refer to the crossing at the bottom-left corner of the grid it belongs to) and (c) The signal in (b) fed into 1-D signal processor

Using the above form of coding 2-D signals in a linear sequence, 2-D signal processing using 1-D medium such as optical fiber can be made possible.

2.3 Multi-dimensional Discrete Signal Processing

Having made a reasonable compromise in the size of the predefined limit, i.e. truncation window size, the Nyquist rate can be applied to 2-D signals to determine the sampling rate. In 1D, the Nyquist rate is twice the highest frequency component of the sampled signal and defines the sampling rate necessary to preserve the entire bandwidth of the signal.
In 2-D, the direction in which the Nyquist rate is applied must be made clear as sampling in one dimension at the Nyquist rate may not guarantee the preservation of the 2-D signal if the signal varies faster with respect to the other dimension. To preserve the entire 2-D signal bandwidth, sampling must be performed at twice the highest spatial frequency component of the 2-D signal in any direction in the sampled space. For example, consider a signal which has a 20 GSamples/s component in n_1-axis but has a 60 GSamples/s component at 70° from n_1-axis. In this case, the sampling rate of 40 GSamples/s in both dimensions is not adequate as the signal has a frequency component of 60 GSamples/s ×sin(70°)=56 GSamples/s along n_2-axis. Since the sampling rates in the both dimensions are usually kept the same, sampling rate of 56×2= 112 GSamples/s in both dimensions will preserve the entire 2-D signal bandwidth.
In discrete-time signal processing, the term *normalized frequency* is used to describe a frequency independent of the system sampling frequency. The concept is applied in 2-D processing with a straightforward extension to spatial frequency.

2.4 Separability of 2-D Signals

A 2-D sequence is *separable* if it can be represented by a product of two 1-D sequences as shown in (2.4.1). Separable sequences form an important and special, but limited class of 2-D sequences. Many results in 1-D theory have a simple extension for separable 2-D sequences whereas for non-separable sequences such extensions often do not exist. If a 2-D sequence is separable, the separability can be exploited to reduce the processing requirements resulting in considerably less amount of computation. Unfortunately, most 2-D sequences are not separable.

$$s[n_1, n_2] = f[n_1]g[n_2] \tag{4}$$

An example of a separable sequence is the unit sample sequence $\delta(n_1, n_2)$ shown in (5)*(a)* [10]. Other examples of separable sequences include the unit step sequence $u(n_1, n_2)$ and the example in (5)*(b).*

$$\delta[n_1, n_2] = \delta[n_1]\delta[n_2]............(a)$$

$$a^{n_1}b^{n_2} + b^{n_1+n_2} \;=\; \left(a^{n_1} + b^{n_1}\right)b^{n_2}.............(b) \tag{5}$$

2.5 Separability of 2-D Signal Processing Operations

Similar to 2-D sequences, a 2-D signal processing operation can be classified as separable or non-separable. The consequence of an operation being separable is that the operation yields the correct answer when it is performed in two independent cascade stages with each stage performing the operation with respect to only one of the independent variables. The situation is illustrated in *Figure 4*.

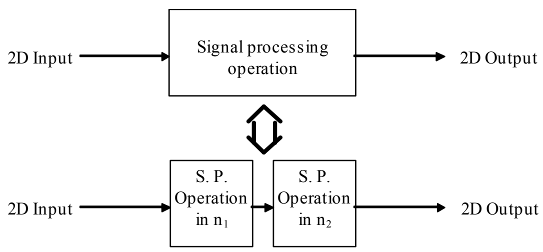


Figure 4: A separable 2-D signal processing operation

An example of a 2-D separable signal processing operation is double integration. A double integration procedure can be expressed as

$$F(n_1, n_2) = \int_{n_2=-\infty}^{\infty} \int_{n_1=-\infty}^{\infty} f(n_1, n_2)dn_1dn_2 \tag{6}$$

The 2-D sequence $f(n_1, n_2)$ is integrated with respect to n_1 first, and then with respect to n_2. The two procedures can be put in cascade and thus double integration operation is classified as a *separable operation*. Note that the separability of the 2-D input sequence $f(n_1,n_2)$ is not a pre-requisite for the success of the operation. In addition, separable signal processing operations have separable impulse responses. The 2-D signal processing can be performed by convoluting the 2-D input with the 1-D filter impulse response in one dimension, and the operation can be completed by convoluting the result of the first convolution with the filter impulse response in the other dimension. It is therefore clear that the operations can be performed using a cascade stage of two filters. As with the case of separable sequences, separable operations form a special class of 2-D signal processing operations. Most signal processing operations are not separable.

In discrete domain, separable operations can be expressed in terms of a product of two z-transform transfer functions. For example, is double integration using Simpson's rule for digital integration [11].

$$H(z_1, z_2) = \frac{T_1}{2}\left[\frac{1+z_1^{-1}}{1-z_1^{-1}}\right] \cdot \frac{T_2}{2}\left[\frac{1+z_2^{-1}}{1-z_2^{-1}}\right] \tag{7}$$

It is clear that $H(z_1,z_2)$ is separable. For other functions, the separability is often not the case. A circularly symmetric 2-D digital low pass filter cannot be separated into a product of two functions each dealing with only one kind of delays (z_1 *or* z_2). In such cases, a way of dealing with non-separability must be found as cascade stages will no longer work. It is the non-separability of most 2-D signal processing functions that makes implementation of 2-D filters a difficult task.

$$\begin{aligned} H(z_1, z_2) = \;& 1 + 0.9z_1^{-1} + 0.9z_2^{-1} + 0.8z_1^{-1}z_2^{-1} \\ & -0.1z_1^{-2} - 0.05z_1^{-1}z_2^{-2} - 0.05z_1^{-2}z_2^{-1} \\ & -0.1z_2^{-2} + 0.1z_1^{-3} + 0.1z_1^{-3}z_2^{-1} \\ & +0.07z_1^{-3}z_2^{-2} + 0.07z_1^{-2}z_2^{-3} - 0.05z_1^{-3}z_2^{-3} \end{aligned} \tag{8}$$

Another advantage of having a separable implementation is the issue of stability. The stability analysis of non-separable filters is very difficult and there are no known simple methods of checking the stability of 2-D filters directly from the transfer function or from pole-zero plots as is the case with 1-D systems[1]. However, with separable filters if 1-D sub-sections are stable, then the overall stability is guaranteed. Stability of 1-D filters can be guaranteed by having all system poles inside the unit circle.

3. Filter Design Methods for 2-D PSP

In Section 2, the concepts of 2-D signal processing have been introduced. Out of many possible mathematical

[1] For more detailed discussion on stability checking using position of poles, refer to Section 4, [25].

models for 2-D systems, the model best suited to fiber-optic signal processing must be found. In this section, two different mathematical models of representing 2-D systems are presented and a brief introduction to 2-D filter design methods is given.

3.1 2-D Filter Specifications

To specify a filter, two approaches can be adopted. One approach specifies a filter in mathematical form by specifying the transfer function or the state-space equations of the filter. This method can specify the exact behavior of the filter. The other approach specifies a filter by its transfer characteristics of magnitude and phase response or impulse response of the filter. This later approach is more intuitive than the former because it is easy to see how the filter would behave in practical implementation. However the accuracy of the filter then depends on the accuracy of the specification therefore can sometimes be inadequate. In any case, the later approach must go through the mathematical description before implementation. Developing a method for designing and implementing a filter from its dynamic characteristics therefore encompasses the mathematical description.

The method developed in this section assumes the spatial frequency responses of the filter to be specified. The 2-D photonic filter design process can be as follows: Specification of magnitude or impulse response the desired 2-D filter; Development of transfer function or state-space description of the 2-D filter; Development of signal flow diagram of the 2-D filter and Development of photonic implementation of the 2-D filter.

To specify a filter using its frequency response, both magnitude and the phase responses need to be supplied. However, designing a filter with a certain phase response is a very difficult task. In many cases of interest, a condition of linear phase is all that is required and in this paper, the condition is adhered to. The reason for requiring a linear phase can be explained by the Fourier transform of a 1-D linear phase filter as

$$V(f)e^{-j\varphi\omega} \Leftrightarrow v(t-\varphi) \tag{9}$$

In (9), the phase $\phi\omega$ is proportional to frequency. The Fourier transform (FT) of linear phase on the right of shows that a linear phase corresponds to pure time delay. The result is extendible to 2-D simply by substituting frequency by spatial frequency and time delay by spatial delay. A non-linear phase response leads to non-uniform delays and thus inter-symbol interference (ISI).

3.2 Mathematical Model of 2-D Discrete Photonic Systems

3.2.1 Transfer Function Description

2-D transfer function description of the filter can also be explained by using a 2-D difference equation. As in 1-D, 2-D transfer functions can readily be turned into 2-D difference equations

$$H(z_1,z_2) = \frac{T_1}{2}\left[\frac{1+z_1^{-1}}{1-z_1^{-1}}\right]\cdot\frac{T_2}{2}\left[\frac{1+z_2^{-1}}{1-z_2^{-1}}\right] \tag{10}$$

The equivalent difference equation is given by

$$\begin{aligned} y(n_1,n_2) = \frac{T_1T_2}{4}&[x(n_1,n_2)+x(n_1-1,n_2)+x(n_1,n_2-1) \\ &+x(n_1-1,n_2-1)]+y(n_1-1,n_2) \\ &+y(n_1,n_2-1)-y(n_1-1,n_2-1) \end{aligned}$$

Figure 5 illustrates the sample points which are summed in $y(n_1,n_2)$ of (10). Because we are dealing with spatial delay and not time delay, the actual implementation of delay depends on the signal transmission format. If all points of the 2-D signal are transmitted in parallel, then the delays need not be time delay raising the possibility of parallel processing similar to that of spatial and temporal architecture.

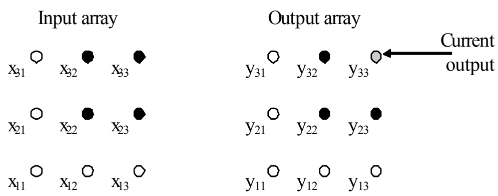


Figure 5: Illustration of the difference eqn. (10)

A 1-D integrator transfer function amounts to just the first half or the second half of $H(z_1,z_2)$ in (10) and can readily be turned into a signal flow graph (SFG) as shown in Figure 6.

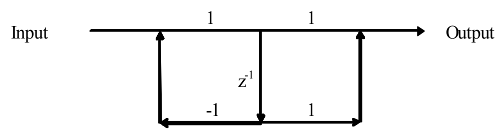


Figure 6: 1-D trapezoidal integrator signal flow diagram

The SFG of 2-D version of the trapezoidal integrator transfer function H is shown in Figure 5. A notable difference is the presence of two different delay elements. In spatial terms, one represents the vertical delay and the other represents the horizontal delay.

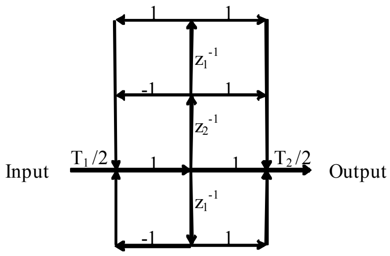


Figure 7: 2-D trapezoidal integrator SFG diagram

Whether a SFG can readily be turned into photonic domain depends on its structure. A major obstacle that prevents a direct translation of SFGs into photonic circuits is the number of complex interconnections. Too many complex interconnections can result in the loss of modularity making the photonic implementation of the transfer function difficult.

Transfer functions can be manipulated into potentially useful forms for photonic implementation. One such manipulation technique is called *continued fraction expansion realization (CFER)* as described below. Although the method results in reduced number of filter components, not all transfer function can be expanded easily. A method of designing a filter transfer function that can be expanded using continued fraction expansion is therefore required. Such a design method does not exist currently and as a consequence the use of CFER is confined to only transfer functions that are expandable.

Continued Fraction Expansion

Given a transfer function, the numerator of the transfer function is recursively long-divided by the denominator until the remainder is only a simple fraction. A possible form for a fraction which has been expanded using continued fraction expansion as

$$H(z_1,z_2) = C_1 + \cfrac{1}{A_1 z_1 + \cfrac{1}{C_2 + \cfrac{1}{B_1 z_2 + \ddots}}} \tag{11}$$

It is intuitively obvious that such expansions do not exist for all polynomial fractions. A method of checking the existence of such expansion is given in [12-16].

3.2.2 State-Space Equation Description

State-space description can be seen as an alternative description method to the transfer function description. The advantages offered by state-space description include the notion of observability and controllability. Although such concepts are useful in 2-D dynamic control system, the applications of the concepts are not obvious in 2-D signal processing. At best, the main advantage of using state-space approach can be stated as the established techniques of 1-D state-space theory such as algorithms to manipulate state-space matrices to obtain a reduced order system.

An Algorithm for Conversion of a 2-D Transfer Function into 2-D State-space Equation

In [14], a 2-D state-space description is formulated from a 2-D FIR transfer function description using the following method. Ref.[12] has given a more generalized case of transfer functions of 2-D IIR filters.

Formulation of state-space equations from transfer function [14]

A 2-D FIR transfer function can be expressed by

$$H(z_1,z_2) = \sum_{n_1=0}^{N_1}\sum_{n_2=0}^{N_2} h(n_1,n_2) z_1^{-n_1} z_2^{-n_2} \tag{12}$$

State-space form can be expressed as

$$\begin{bmatrix} x^h(i+1,j) \\ x^v(i,j+1) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix} \begin{bmatrix} x^h(i,j) \\ x^v(i,j) \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} u(i,j) \equiv \mathbf{Ax} + \mathbf{b} \quad \text{(a)}$$

(13)

$$y(i,j) = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}^h(i,j) \\ \mathbf{x}^v(i,j) \end{bmatrix} + du(i,j) \equiv \mathbf{cx} + du \quad \text{(b)}$$

$$\begin{aligned}
&\mathbf{A}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{N_1-1} \\ 0 & \mathbf{0} \end{bmatrix}, \qquad \mathbf{A}_2 = \begin{bmatrix} h_{1N_2} & \cdots & h_{11} \\ \vdots & \ddots & \vdots \\ h_{N_1N_2} & \cdots & h_{N_11} \end{bmatrix} \\
&\mathbf{A}_3 = \mathbf{0}, \qquad \mathbf{A}_4 = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{N_2-1} \\ 0 & \mathbf{0} \end{bmatrix} \\
&\mathbf{b} = \begin{bmatrix} h_{10} & \cdots & h_{N_10} \mid 0 & \cdots & 0 & 1 \end{bmatrix}^T \\
&\mathbf{c} = \begin{bmatrix} 1 & 0 & \cdots & 0 \mid h_{0N_2} & \cdots & h_{01} \end{bmatrix} \\
&d = h_{00}
\end{aligned} \tag{14}$$

An Algorithm to Convert a 2-D State-space equations into a 2-D Transfer Function

For state-space approach to be useful, there must be a method of converting 2-D state-space Eqs into 2-D transfer function form. An algorithm to perform such a task could not be found in standard text books of digital signal processing, and therefore it had to be devised independently. The algorithm given in this section performs conversion from a 2-D state-space equations specified in (15) to a 2-D transfer function description.

By rearranging (12), we can obtain the transfer function form of the same system with input denoted by x and output denoted by y as

$$H(z_1, z_2) = \mathbf{C}\left[\begin{pmatrix} z_1\mathbf{I}_m & \\ & z_2\mathbf{I}_n \end{pmatrix} - \mathbf{A}\right]^{-1}\mathbf{B} \tag{15}$$

What makes the implementation of above equation difficult is the presence of matrix inverse. Because the matrix inside the bracket in (15). (15) describes a 2-D system, the determinant of the matrix contains two independent variables z_1 and z_2 and cannot therefore be solved by simply obtaining the eigenvalues of the matrix and cross-multiplying to get coefficients of variables as in 1-D determinant calculation.

An Algorithm to obtain the characteristic polynomial of a matrix describing a 2-D system

1. Let A be the matrix of size m×n describing a 2-D system and Z be a zero matrix of size m×n.
2. Let $z_1 = 0$, $z_2 = 0$.
3. Let A' be a matrix formed by eliminating m-z_1 rows and n-z_2 columns from matrix A.
4. Let $Z_{z1,z2} = Z_{z1,z2} + \Delta(A')$. Δ is 1-D determinant operator.
5. Repeat step 3 and 4 until all combinations of z_1 *rows and* z_2 *columns* are tried.
6. Repeat step 3 to 5 with different value of z_1 *and* z_2 until all coefficients of the characteristic polynomial Z are found.
7. Reverse the signs of elements of Z whose indices sum to an odd number.

The resulting matrix Z contains the coefficients of 2-D characteristic polynomial in format as.

$$H(z_1, z_2) = \sum_{n_1=0}^{N_1}\sum_{n_2=0}^{N_2} h(n_1, n_2) z_1^{-n_1} z_2^{-n_2}$$

is expressed in matrix form as (16)

$$Z = \begin{bmatrix} h_{00} & h_{01} & \cdots & h_{0n} \\ h_{10} & h_{11} & \cdots & h_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m0} & h_{m1} & \cdots & h_{mn} \end{bmatrix}$$

For the 2-D transfer function description, denominator and the numerator can be calculated by

$$\begin{aligned} \mathbf{h}_a &= \det(\mathbf{A}) \\ \mathbf{h}_b &= \det(\mathbf{A} - \mathbf{B}\times\mathbf{C}) - \det(\mathbf{A}) \times (D-1) \end{aligned} \tag{17}$$

3.3 Filter Design Methods

3.3.1 Direct Design Methods

Direct design methods include *window method*, *frequency sampling method*, *transformation method* for FIR filter implementations, and *impulse response method* for IIR filters. All of the design methods listed are similar to the 1-D methods of the same name involving some extensions of the concepts into 2-D and they all result in non-separable transfer functions.

The general format of non-separable filter transfer functions generated by FIR filter design methods is given in Section 2 and the most general form of 2-D FIR filter signal flow diagram is shown in Figure 8. In *Section 4*, the details of the algorithms and the implementations of 2-D direct design FIR filters are discussed. Direct design methods for IIR filters are also discussed.

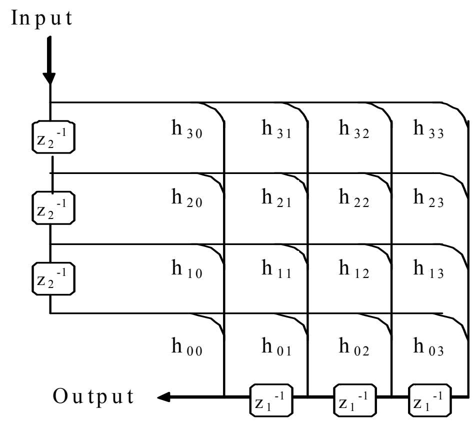


Figure 8: General form of FIR filter signal flow diagram [15]

3.3.2 Use of Matrix Decomposition

In [16], Mitra introduces a method where matrix decomposition is used to separate a non-separable function into a cascade of two separable filter stages each one involving only one set of delay elements. A drawback with Mitra's method is that there is no general structure for filter implementation. The filter structure is therefore heavily dependent on the transfer function and this lack of generality of filter design limits the usefulness of the method in photonic filter implementation where the final product corresponds closely to the SFG representation of the filter transfer function. Figure 9 shows a part of the filter implementation.

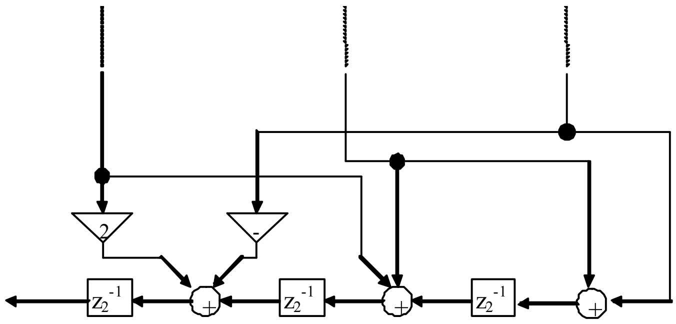


Figure 9: A subsection of filter in [16].

Another approach which uses matrix decomposition is by decomposition of 2-D magnitude specification into the sum of products of 1-D magnitude specifications. Because 2-D magnitude specification becomes a set of two 1-D magnitude specifications, design process of a 2-D filter reduces to a set of 1-D filter designs for which established design methods are aplenty. In addition, the resulting 2-D filter is separable. Section 4 discusses a number of 2-D filter design algorithms based on matrix decomposition of 2-D magnitude specification.

4. Direct 2-D Filter Design Methods

As the first of two streams of 2-D filter design methods, direct 2-D filter design methods are introduced. This class of 2-D filter design methods does not use matrix decomposition and instead uses 2-D magnitude specifications directly to produce non-separable designs.

4.1 FIR and IIR Structures in 2-D Signal Processing

A digital filter can be divided into two broad classes, FIR (Finite Impulse Response) and IIR (Infinite Impulse Response). FIR filters only use feed forward structure and therefore are non-recursive, whereas IIR filters use feedback as well as feed forward structure and therefore are recursive. The impulse response of an IIR filter is infinite in duration, therefore the name 'infinite impulse response filter'.

Given a 2-D frequency response specification, one can either try to formulate a FIR filter transfer function or an IIR filter transfer function. There are several factors which must considered when deciding which structure to implement for a given frequency response specification. (i) *Linearity of the phase response of the filter; (ii) Stability of the filter; and (iii) Order of the implemented filter.*

Linearity of the designed filter's phase response is very important as explained previously in *Section 3.1*. Linear phase FIR filters are very easy to design as the condition for the linear phase is simply a symmetric impulse response which in turn is guaranteed if the magnitude response of the 2-D filter is symmetric about the two axis. For 2-D IIR filters, phase linearity is much more difficult to guarantee. Often IIR filters are specified only with magnitude characteristics and the phase response is generally accepted for what it is(which is non-linear). The lack of control over phase response of IIR filters limits its usefulness in many applications [10].

Stability is a very important issue in designing of any dynamic system which requires no explanations. The advantage of 2-D FIR filter over 2-D IIR filter regarding the issue of stability is that for 2-D FIR filters, stability is inherent in its definition. Since the impulse response of FIR filter is finite in duration, bounded input results in bounded output and the filter is therefore always stable. Although 2-D IIR filters can be designed to be stable, as mentioned in *Section 2.5* there is no simple algorithm for checking the condition for stability of a 2-D IIR filter. A mathematical theory involving complex cepstrum to check for the stability condition of 2-D filters is quite involved and most algorithms for checking the 2-D stability simply repeat 1-D stability condition over the 2-D space many times over which can be computationally very inefficient [10]. As a consequence of the lack of usable algorithms or simple method for stability testing, there is no known method of designing stable 2-D IIR filters [10]. In practice, 2-D FIR filters are therefore much more preferred to 2-D IIR filters.

Order of the filter refers to the number of delay elements in the numerator (or the denominator if the filter is IIR). Order of the filter has a direct consequence in the final implementation of the filter as the determining factor of number of processing elements. Higher order filters require more processing elements than lower order filters which makes low order filters more desirable. One distinctive advantage of IIR filters over FIR filters is that the order of the filter required for a given magnitude specification is smaller. FIR filters can sometimes require an excessive order(the definition of 'excessive' depends on the implementation medium - for example in software implementation of a digital filter, a 1000th order filter might be quite acceptable but with fiber-optic delay line filters, the maximum order is well below 50). For example an integrator, which has an infinite impulse response by nature can be described using a first order filter within 12.5% error [11] whereas achieving the same error with 1st order FIR filters would be impossible. In addition, 2-D filters usually require much higher order filters than 1-D filters of similar transition band requirements and it is therefore important that the 2-D filter design methods keep the order of the filter to an acceptable level.

4.2 Frequency Sampling Method

As the first of the direct 2-D filter design methods, 2-D frequency sampling method produces an FIR filter with the minimum of fuss. Using the fact that the transfer function of a FIR filter is same as the impulse response of the filter, 2-D frequency sampling method takes discrete Fourier transform of even-spaced samples of 2-D frequency response and uses the result as the coefficients for 2-D transfer function. It is noted that the procedure is identical to that of 1-D frequency sampling method.

It is observed in [10] that the filter designed using frequency sampling method is not optimal as far as number of delay elements is concerned. Also, the frequency response of the filter is controlled only by the sampling rate of the frequency sampling and the nature of the frequency samples. For example, increasing the frequency sampling rate will increase the number of discrete sampling points of the impulse response and hence will result in a filter with a better frequency response but with a larger order. It is also found that frequency response can be improved considerably especially around the transition band if the ideal frequency response takes account of the transition frequency values.

A filter design example using frequency sampling method

Design Aim: A low pass filter with normalized cut off frequency of 0.5 in both dimensions.

Method: Frequency sampling method

Program used: FIR2-DFS.m

Result: The magnitude response of the designed filter is shown below. Filter is of 33×33 rd order and the error is 3.63%.

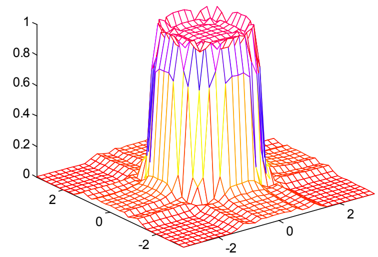


Figure 10: Magnitude response of a filter designed using frequency sampling method

The filter error is calculated using *Eq. 4-1* where $2N_1+1$ and $2N_2+1$ are order of the filter in n_1 and n_2 dimension, respectively. H_d is the ideal frequency response, H_f is the actual filter response, and Ω_1 and Ω_2 are the frequency sampling rates.

$$e = \frac{\sum_{n_1=-N_1}^{N_1} \sum_{n_2=-N_2}^{N_2} \left| H_d[n_1\Omega_1, n_2\Omega_2] - H_f[n_1\Omega_1, n_2\Omega_2] \right|}{(2N_1-1)(2N_2-1)} \quad (18)$$

The magnitude response shown below is obtained by incorporating the transition band values into the ideal frequency response parameter. It can be seen that the ripple in the transition band has disappeared and the error is found to be only 2.74% which is nearly 1% less than that obtained without any transition band consideration.

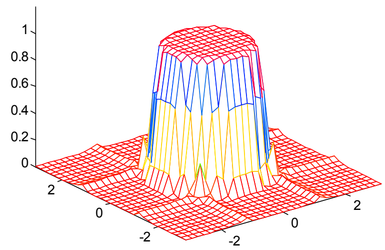


Figure 11: Magnitude response of the filter designed with transition band consideration

The response shown below is obtained with a filter of order 20×20. The error is found to be 5.85% which compares unfavorably with 3.63% obtained with the first design. Clearly lower order filters result in considerably worse performance.

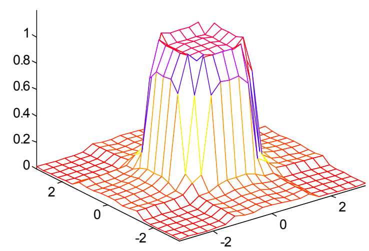


Figure 12: Magnitude response of filter of order 20×20

4.3 Windowing Method

The window method for 2-D filters use a 2-D window instead of 1-D window to achieve a finite impulse response sequence in 2-D. As with 1-D windowing method, the 2-D windowing method begins by performing a Fourier transform on the desired frequency response expression. The impulse response of the filter in 2-D is then multiplied by the expression for window. The purpose of multiplying by a windowing function is to reduce the effect of sharp transitions in the transition band and also to make the impulse response finite through truncation. The windowing function is chosen so that the frequency response is least affected and the impulse response is as short as possible. It is noted in [10] that the performances of window method and the frequency sampling method are similar and an example is therefore omitted.

4.4 McClellan Transformation Method

McClellan transformation method takes an entirely different approach to the design process of 2-D filters. The idea is to transform a 1-D FIR filter into a 2-D filter of the desired characteristic. The 1-D filter can be designed using any 1-D filter design method so that its frequency response is a cross-section of the desired 2-D filter response (see Figure 13).

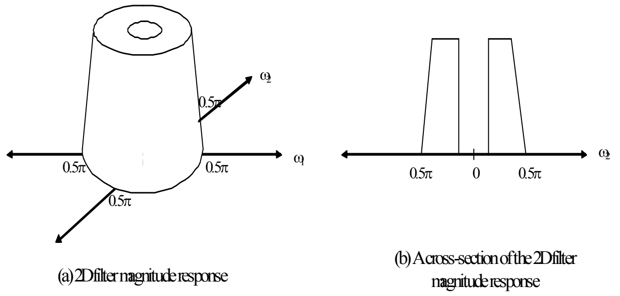


Figure 13: Desired 2-D filter magnitude specification and the required 1-D filter specification

Given the transfer function of 1-D FIR filter, each coefficient is multiplied by the transformation function T which is a function of ω_1 *and* ω_2. The resulting transfer function is also a function of ω_1 *and* ω_2 and describes a 2-D filter with the desired magnitude response.

$$H(\omega_1,\omega_2) = h(0) + \sum_{n=1}^{N} 2h(n)\cdot[T(\omega_1,\omega_2)]^n$$
$$\text{with } T(\omega_1,\omega_2) = \sum_{(n_1,n_2)\in R_T}\sum t(n_1,n_2)\cdot e^{-j\omega_1 n_1}\cdot e^{-j\omega_2 n_2} \tag{19}$$

In (19), R_T is the region of support of $t(n_1, n_2)$ which describes the transformation function. It should be noted that using different transformation functions, many 2-D filters can be designed from a single 1-D filter. It is also important to note that as long as the 1-D filter is a linear phase filter, the transformed 2-D filter is also a linear phase filter as long as phase of the transformation function *T* is linear (i.e. the transfer function is symmetric about the zero delay point since multiplying a linear phase function by another linear phase function does not affect the linearity of the phase of the resultant function.

2-D filter design using transformation method

Design Aim: 2-D low pass filter with frequency cut off at 0.5 in both dimensions

Method: McClellan transformation method. The filter order is 13×13 and the transformation function used is given by

$$\begin{aligned} T(\omega_1,\omega_2) = &-\frac{1}{2}+\frac{1}{4}e^{-j\omega_1}+\frac{1}{4}e^{-j\omega_2}+\frac{1}{4}e^{j\omega_1}+\frac{1}{4}e^{j\omega_2} \\ &+\frac{1}{8}e^{-j\omega_1}\cdot e^{-j\omega_2}+\frac{1}{8}e^{j\omega_1}\cdot e^{j\omega_2}+\frac{1}{8}e^{j\omega_1}\cdot e^{-j\omega_2}+\frac{1}{8}e^{-j\omega_1}\cdot e^{j\omega_2} \end{aligned} \tag{20}$$

Graphically, above transformation function can be expressed as in Figure 14. Figure 15 shows the frequency response of the designed filter. It is clear that the performance of the filter designed using the transformation method is somewhat worse than that designed using frequency sampling method. A factor which should be taken into consideration is the order of the filter. The order of the filter in is only 13×13. The reason for such large difference lies with the coefficients of the resulting filter of transformation method. In Figure 15(b), the coefficients of the filter as originally designed by the transformation method is shown as the form of impulse response of the filter. Clearly, it is a 33×33 order filter, however the surrounding 10 rows and columns do not contribute to the filter response at all and therefore can be removed without affecting the response of the filter and the remaining coefficients constitute a 13×13 order filter.

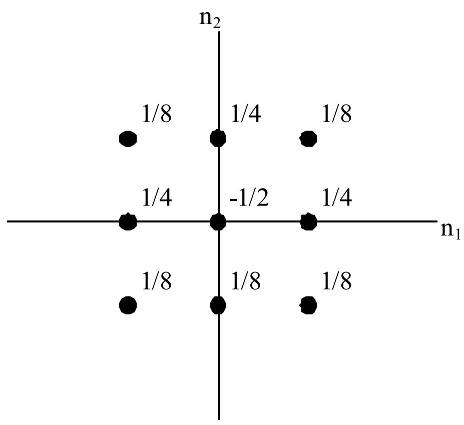
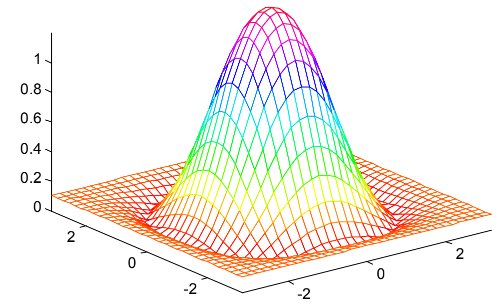


Figure 14: Transformation sequence used [10]

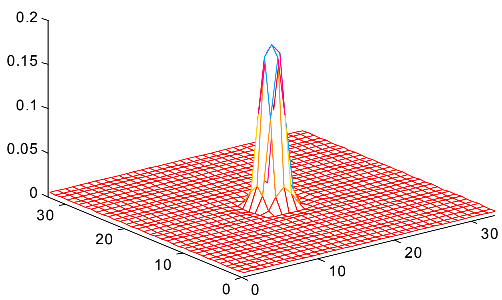
Matlab program used: FIR2-DTF.m

Result: The error calculated = 13.71%.

The transform function has a large bearing on the eventual filter transfer function. The filter shown in Figure 16 is of the same order as the filter in)



(a)



(a)

Figure 15: (a) Frequency response of a filter of order 13×13 designed using transformation method (b) Impulse response(filter coefficients) of the filter designed by transformation method.

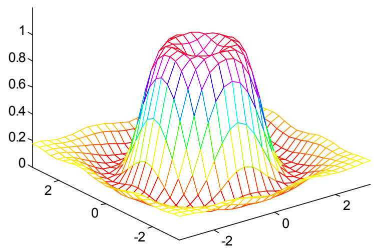


Figure 16: 2-D filter designed using a different transform function

Figure *15*(a) and 1-D prototype function is identical. However, the transformation function is now a 7×7 function[2]. It can be seen that the transition band is much more distinct whereas the stopband is not as well attenuated as the filter response in)

Figure *15*(a*)*. The overall error of the filter whose frequency response is shown in Figure 16 is found to be 12.84%. It is clear that the good performance in passband and the transition band is offset by the poor stopband attenuation. Finally, it is noted that the resulting transfer function is a non-separable FIR transfer function.

5. Concluding Remarks: Part I

The objective of the research this series of three parts is to explore possible ways of realizing a 2-D signal processing system using fiber-optic signal processing architecture. This part I describes a general technique for designing 2-D filters. Numerous examples of utilization of the technique are given. Although the discussion is focused on fiber-optic systems, the design procedure for 2-D filters are just as applicable to any other signal processing architectures. For example, the 2-D filter order reduction method given in Section 4 can be used to simplify 2-D lightwaves systems which may or may not be fiber-optic systems.

The design of 2-D filters is classified into two different classes. One class used matrix decomposition to reduce the design of 2-D filters into a set of 1-D filter design procedures. The other class uses direct extensions of 1-D filter design methods. It is found that neither has a distinctive superiority over another and that the designer has to choose what is the best for the particular application, most likely by designing both and comparing the performances. All of the design procedures are implemented using the MATLAB™ programming language.

Part II will describe the techniques of matrix decomposition methods.ong these techniques, the multiple stage singular value decomposition method performs the best whereas for direct methods, frequency sampling method produced filters with smallest errors. A 2-D Filter order reduction method is applied to make fiber- and integrated optic signal processing more feasible. The technology allows the filter designer to produce filters of orders that are implementable in practice without sacrifices in performance. Part III will deal with different possible filter structures are proposed and illustrated for photonic implementation of 2-D filters. Filter structures for FIR and IIR filters are also shown and examples are given in Section 9.

---

[2] [13] discusses methods of designing such transformation functions in great detail

References

[1] (a) Ngo N.Q. and L.N. Binh, *"Programmable incoherent Newton-Cotes Optical Integrator"*, *Optics Communications*, vol. 119, 1995, pp. 390-402; (b) Robert R. Thomson, Ajoy K. Kar1, and Jeremy Allington-Smith, "Ultrafast laser inscription: an enabling technology for astrophotonics", Optics Express, 2 February 2009 / Vol. 17, No. 3, pp. 1964-2969.

[2] H. F. Taylor, "*Application of Guided-Wave Optics in Signal Processing and Sensing,*" Proc. IEEE, vol. 75, No 11, pp1524-1535, November 1987.

[3] See for example B. Moslehi, J. W. Goodman, M. Tur, and H. J. Shaw, "Fiber-Optic Lattice Signal Processing," Proc. IEEE, Vol. 72, No 7, July 1984;

[4] C.K. Marsden and J.H. Zhao "*Optical filter analysis and design: A digital signal processing approach*", J Wiley, NY 1999.

[5] J. Capmany, and J. Cascon, "Optical programmable transversal filters using fiber amplifiers," Electronic Letters, vol. 28, pp 1245-1246, 1992

[6] B. Moslehi, "*Fiber-optic filters employing optical amplifiers to provide design flexibility*," Electronic Letters, Vol. 28, pp 226-228, 1992

[7] E. Heyde and R. A. Minasian, "*Photonic signal processing of microwave signals using an active fiber Bragg grating pair*" IEEE Trans. Microwave Th.Tech , pp 1463 -1466, 1997.

[8] F.T.S. Yu and I.C. Khoo, *Principles of Optical Engineering*, Section 9.

[9] L. N. Binh, N. Q. Ngo, and S. F. Luk, "*Graphical representation and analysis of the Z-shaped double-coupler optical resonator*," IEEE J. Lightwave Technol., vol. 11, pp 1782-1792, 1993

[10] J. S. Lim, *Two-Dimensional Signal and Image Processing*, Englewood Cliffs, NJ, Prentice Hall, 1990

[11] Binh L.N., "*Generalized analysis of a double-coupler doubly-resonant optical circuit*", *IEE Proceedings Optoelectronics*, vol. 142, Dec.1995, pp. 296-304.

[12] S. Kung, B. C. Levy, M. Morf, T. Kailath, "*New results in 2-D systems theory, Part II: 2-D state-space models-realization and the notions of controllability, observability, and minimality*," Proc. IEEE, Vol. 65, No. 6, June 1977, pp 945-961

[13] R. Mersereau, W. F. G. Mecklenbrauker, and T.F. Quatieri, Jr., "*McClellan Transformations for Two-Dimensional Digital Filtering: I-Design*", IEEE Trans. Circuits and Systems., VOL. CAS-23, No. 7, July 1976, pp 405-422

[14] A. Antoniou and W Lu, "Design of two-dimensional digital filters by using the singular value decomposition," IEEE Trans. Circ. Systems, Vol. CAS-34, No. 10, Oct 1987, pp1191-1198. http://www.ds.eng.monash.edu.au/techrep/reports/post2003index.html.

[15] Z-J Mou, "*Efficient 2-D and Multi-D Symmetric FIR Filter Structures*", IEEE Trans. Signal Processing, Vol. 42, No. 3, March 1994

[16] S. Mitra, A. Sagar, and N Pendergrass, "*Realizations of two-dimensional recursive digital filters,*" IEEE Trans. Circuits Syst. , vol. Cas-22(3), 1975, pp177-184.

Le Nguyen Binh received the degrees of B.E. (Hons) and PhD in Electronic Engineering and Integrated Photonics in 1975 and 1980 respectively, both from the University of Western Australia, Nedlands, Western Australia. In 1981 he joined in the Department of Electrical Engineering of Monash University after a three year period with CSIRO Australia as a research scientist developing parallel micro-computing systems. He was appointed to reader of Monash University in 1995. He has worked for and lead major research and development programs in the Department of Optical Communications of Siemens AG Central Research Laboratories in Munich and the Advanced Technology Centre of Nortel Networks in Harlow, U.K. He was a visiting professor in 2007-2008 and is currently an adjunct professor of the Faculty of Engineering of Christian Albretchs University of Kiel, Germany. Dr. Binh has published more than 200 papers in leading journals and international conferences and three books, one on Photonic Signal Processing (2007) and the other on Digital Optical Communications (2008) both by the publishing house CRC Press (Florida, USA). His third book scheduled for publication in 2009 is "Optical Fiber Communications Systems: Principles, Practices and MATLAB Simulink Models" also by the same publisher. He is an associate editor of the International Journal of Optics and the Research Letters in Optics, Series Editor of the Series "Optics and Photonics" for CRC Press Pub.

His current research interests are Tb/s optical transmission systems and networks, especially advanced modulation formats and associate electronic equalization methods, photonic signal processing, integrated photonics and multi-bound soliton fiber lasers.