MATERIALS AND STRUCTURES FOR NONLINEAR PHOTONICS

Xin Gai, Duk-Yong Choi, Steve Madden and Barry Luther-Davies
Centre for Ultra-high-bandwidth Devices of Optical Systems, Laser Physics Centre, RSPE, The Australian National University, Canberra ACT 2600, Australia

1. INTRODUCTION

Research into all-optical processing dates back to the 1980s when the so-called “electronic bottleneck”, the limited speed of electronics, was seen as the major barrier that would limit the information capacity of fiber-optic networks. This was the era pre-dating the development of wavelength division multiplexing (WDM) and, hence, research was exclusively focused on time division multiplexed (TDM) signals on a single wavelength channel [1]. In this “first era” of optical signal processing the main idea was to use the ultra-fast third order nonlinearity of Kerr-like materials to create fast switches so that an interleaved signal could be de-multiplexed efficiently to different output ports of a device. Although there was a lot of work on self-switched devices, where routing was controlled by varying the signal power itself, the most practical implementation involved switching the state of the device using separate control pulses.

A canonical optical switch consisted of a nonlinear Michelson interferometer in which a third order nonlinear element existed in only one of the arms. At the input port the beam was split in two and one half propagated through the linear arm and the other through the nonlinear arm to the output. At the output the signals interfered at a second beam splitter and the power was divided between the two output ports depending on the relative phase of the beams. At low power if the phase shift in the two arms was identical, the two beams interfered constructively to create an output at the “through” port. However, if the power was increased (or a control pulse was present) such that the nonlinear phase change that occurred in the nonlinear arm was \( \Delta \phi_{nl} = k_n I_2 L = \pi \) (\( L \) is the length of the nonlinear element; \( n_2 \) its nonlinear refractive index, \( I_2 \) the light intensity and \( k_n \) the free space wavenumber), the power would be switched to the “cross” port. This type of nonlinear optical switch, therefore, relied on a nonlinear phase shift caused by the real part of the third order nonlinearity.

Of course, in this description, losses are ignored and no account is taken of the time dependence of the pulse that can lead to partial switching and pulse distortion. This can only be eliminated for the special case where the propagating pulses are temporal solitons [2] or when the control pulse is flat-topped and long compared with the signal pulse. Losses are, however, invariably present and can have a number of important consequences. Firstly, linear (and nonlinear) losses lead to power dissipation and this causes heating of the device. Via the thermo-optic effect this creates a slow phase change that can imbalance the interferometer. Thermal effects can be particularly serious in high average power applications, such as processing of telecommunications signals, or when micro-resonators with very small thermal mass are used as a switch [3]. Secondly, losses linear or nonlinear, reduce the effective path length in the nonlinear medium and this can limit the available nonlinear phase change because the power decays in the propagation direction. Finally, in a case of a simple Michelson interferometer, loss can imbalance the arms reducing the contrast of the switch. These matters were considered extensively in the early 1990s and a number of figures of merit were proposed to compare different switching devices and materials [1]. Whilst there is now little research into these true optical switches, except in the case of photonic crystal cavities, these remain relevant today.

One of the most widely quoted figures of merit (FOM) is associated with two-photon absorption [4]. The nonlinear refractive index, \( n_2 \), and the two-photon absorption (2PA) coefficient, \( B_{2PA} \), reflect the real and imaginary components of the third order susceptibility. They are related via the Kramers-Kronig relation just as is the case of the linear refractive index and the linear absorption coefficient. As a consequence, two-photon absorption can be present in any third order nonlinear material operated at photon energies close to the bandgap. Figures of merit (FOM) define the nonlinear phase change that can be achieved over an absorption length. For two-photon absorption, the nonlinear phase change and the nonlinear absorption coefficient are both linearly dependent on intensity and hence the resulting FOM_{2PA} depends only on material parameters and is written \( \text{FOM}_{2PA} = n_2 \beta_{2PA} \). For efficient all-optical processing we ideally require \( \text{FOM}_{2PA} \gg 1 \). A similar FOM can be written for linear absorption as reads \( \text{FOM}_{1PA} = n_1 \alpha \) [1] where \( n_1 \) is the maximum intensity that the material can sustain (e.g. limited by optical damage, or by the maximum power available for the device) and \( \alpha \) is the linear absorption coefficient. Taking this further, in materials where three photon absorption may exist (3PA) we obtain
Absorption can be shown that a strong nonlinear absorption can exist at wavelengths well into the mid infrared for indirect gap materials. One of the more useful results was published by Sheik-Bahae et al [5-7], and was based on a simple two-band model for direct-gap semiconductors. Their analysis resulted in expressions for the dependence of $n_2$ and $\beta_{2PA}$ on the ratio of the photon to bandgap energies, and had the form $n_2 = A/(n_0^2E_g^3)\tilde{g}_2(\hbar\omega/E_g)$ and $\beta_{2PA} = B/(n_0^2E_g^3)\tilde{F}_2(\hbar\omega/E_g)$ where $A$ and $B$ are constants; $n_0$ is the linear refractive index; and $E_g$ the bandgap energy. The functions $\tilde{F}_2$ and $\tilde{g}_2$ are reproduced in Fig. 1(a). These expressions have been shown to provide a reasonable fit to experimental measurements for a wide range of materials [5, 6]. For the case of indirect-gap semiconductors the situation is somewhat more complex but was analyzed by Dinu [8] whose results are shown in Fig 1(b). In both these cases the dispersion of the nonlinear coefficients contain similar features. Firstly, as the frequency increases, the third order nonlinearity becomes resonantly enhanced above its value at DC frequency as the two-photon resonance close to half the band gap of the material is approached. Once this resonance is passed, two photon absorption ($\tilde{F}_2, \tilde{\beta}(\omega)$) grows and maintains a high value all the way up to the band edge of the material. The influence of the single photon resonance has a marked effect on the real part of the nonlinearity and, in general, results in a reversal in the sign of the nonlinearity from positive to negative. In the case of the direct gap semiconductors this is predicted to occur at $\approx 0.7E_g$ whilst for indirect materials it is located within the bandgap of the material at around $1.25E_g$. In both these models $\beta_{2PA} = 0$ when $\hbar\omega > E_g/2$, thus in this regime the $FOM_{2PA}$ becomes very large. Whilst this is true, one must remember that the nonlinear polarization of the material contains higher order terms. For example, the imaginary part of $\chi_2$ gives rise to 3PA and this is not necessarily small as, for example, has been demonstrated recently for silicon [9]. Thus, beyond half its band-gap silicon has the intensity dependent $FOM_{2PA}$ mentioned above. In fact as the frequency is progressively reduced we expect that higher and higher orders of nonlinear absorption can appear. For the case of the direct gap materials, the dispersion of the high order multi-photon absorption was predicted by Wherrett [10] and consists of a series of similar curves to that shown for 2PA in Fig 1(a), but displaced progressively lower frequencies with a cut-off frequencies corresponding to $\hbar\omega = E_g/3$, $E_g/4$, etc. To our knowledge no similar analysis exists for indirect gap materials. Of course it is generally assumed that such higher order absorption can be neglected but this may not always be the case as has been illustrated by recent results in silicon, which shows that strong nonlinear absorption can exist at wavelengths well into the mid infrared [9].

**Figure 1:**

1(a) Dispersion functions for the nonlinear refraction ($\tilde{g}_2$) and two photon absorption ($\tilde{F}_2$) for direct gap materials (from[7]); 1(b) Similar relationships for $n_2(\omega)$ ($=n_2$) and $\beta(\omega)$ ($=\beta_{2PA}$) for indirect gap materials (from [8]).

So far we have concentrated on the ultrafast Kerr nonlinearity, but intensity-dependent changes in the phase of the light can also be due to other mechanisms. Apart from the thermal effects, mentioned above, the most common of these is due to the creation of free carriers that lead to a change of index as described by the Drude model. Free carrier effects in semiconductors can be strong and have proved very useful for all-optical switching. However, in passive materials such as silicon, free carriers created by single or multi-photon absorption lead to strong free carrier absorption (FCA) and this is generally detrimental. However, in active devices based on III-V semiconductors, the losses can be compensated using laser action and, hence, the semiconductor optical amplifier (SOA) has proved to be a very effective structure for nonlinear optics. As outlined in more detail below, SOAs main limitation arise because of the finite lifetime of the free carriers which leads to a speed limitation.

Finally in the late 1990s, the so-called “cascaded” second order nonlinearity was (re-)discovered and showed that a nonlinear phase shift could be generated using second order nonlinear materials [11].
Implemented in PPLN this has led to very efficient all-optical processors that are effective even at the single photon level. Their general characteristics are described below.

In what follows we summarize the properties of materials and devices using Kerr, cascaded and free carrier nonlinearities outlining the material and device properties and some of their capabilities and limitations for all-optical signal processing.

2. ALL-OPTICAL PROCESSING USING $\chi^2$ NONLINEARITIES

In the late 1990s it was realized that an intensity-dependent phase shift that was used for all-optical switching could also be obtained by cascading two second-order ($\chi^2$) frequency conversion processes, creating an effective third order nonlinearity without nonlinear loss [11]. In this early work, cascading involved imperfectly phase matched second harmonic generation and degenerate parametric down conversion operating simultaneously. Because of the phase mismatch, power flowed from the low frequency pump to its second harmonic and back again. In this process the down-converted signal originating from the second harmonic accumulated a small phase shift (positive or negative depending on the sign of the phase mismatch) relative to the unconverted pump so when these combined coherently, the phase of the pump was slightly modified. If coherence length of the SHG/DFG processes was small compared with the device length (the case of large phase mismatch) the phase shift increased linearly with device length, and also linearly with pump intensity mimicking the behavior of a $\chi^3$ nonlinear material. Various schemes for all-optical switching using cascading were reported [12].

The advent of WDM significantly changed the emphasis of the research into all-optical signal processing away from switching devices. For example, the challenge in a WDM network was not, at least until recently, to de-multiplex signals at speeds beyond the limits of electronics, but to transfer data from one to another wavelength or to multi-cast data on many channels simultaneously. Conventionally such processes used direct detection and electronic modulation of a wavelength shifted laser or multiple lasers but this can become power-hungry and complex as channel counts rise. An alternative and potentially better solution is to use nonlinear optics to directly convert light at one wavelength to another using nonlinear mixing processes. Thus, optical signal processing using $\chi_2$ nonlinear optics became mostly based on four-wave mixing (4WM) whilst in the case of $\chi_3$ devices three-wave mixing (3WM) was employed. Many of the wave mixing devices are astigmatic to the data format and can readily preserve the amplitude and phase of the incoming signal.

![Diagram](Image)

Fig 2: (a): Schematic of 4WM in which two pump photons ($\omega_p$) mix with a signal photon ($\omega_s$) to create an idler ($\omega_i$) such that $2\omega_p=\omega_s+\omega_i$; (b) schematic of 3WM in which difference frequency mixing between a single high frequency pump photon ($\omega_p$) and a signal ($\omega_s$) creates an idler ($\omega_i$) symmetrically positioned around $\omega_p/2$ such that $\omega_i=\omega_s+\omega_p$; (c) cascaded 3WM where second harmonic generation converts a pump at $\omega_p$ to $2\omega_p$ which then undergoes 3WM with the signal to create an idler.

4WM and 3WM are shown schematically in Fig 2. In the case of 4WM two pump photons at frequency $\omega_p$ mix with a signal photon at $\omega_s$ to create an idler photon at frequency $\omega_i$ for which energy conservation requires $2\omega_p=\omega_s+\omega_i$. In this case $\omega_i=\omega_s, \pm \Delta \omega$ where $\Delta \omega < \omega_p$ which means that all three frequencies lie in the telecommunications band where there are readily available sources. Since two photons at the pump are destroyed and new photons created at $\omega_s, \omega_i$, the signal and idlers experience gain.

The case of 3WM is quite similar, however, here a single pump photon decays into signal and idler photons so that the energy conservation condition now reads $\omega_i=\omega_p+\omega_s$. Again $\omega_i=\omega_s, \pm \Delta \omega$ but now $\Delta \omega < \omega_p$, and hence a high frequency pump is required that lies well outside the telecommunications band where sources are generally less compatible with telecommunications technology. This difference
between the pump and signal/idler frequencies in 3WM would be a large disadvantage were it not possible to eliminate it using cascading. To achieve this a strong pump at frequency $\omega_0=\omega_{id}$ in the telecommunications band is frequency doubled in the $\chi^2$ material to create a new second harmonic pump at $2\omega_0$. Difference frequency generation then allows this second harmonic pump to mix with the signal to create a new idler spaced symmetrically around the original low frequency pump. For this case the pump, signal and idler all exist in the telecommunications band although it must be remembered that the process inevitably generates a new high frequency pump at $2\omega_0$ which needs to optimally couple with the signal and idler waves to achieve high overall conversion efficiency. Just like in the 4WM process, two pump photons are thus destroyed and their energy transferred to the signal and idler exactly mimicking 4WM in a $\chi^2$ material.

To achieve low operating powers, compatible with telecommunications systems, 3WM must take place in a non-centrosymmetric crystalline waveguide exhibiting very strong $\chi^2$ nonlinearity with perfect phase matching. The most common material of choice has been periodically-poled lithium niobate (PPLN) for which there are several well-established routes for waveguide fabrication. In addition, the technology for electric field poling of lithium niobate is well developed and complex grating patterns can be engineered, for example, to increase the bandwidth for phase matching. Of the approaches to waveguide fabrication that have been developed, the most widely used is the so-called reverse proton exchange process (RPE) [13, 14]. This involves, firstly, the preparation of a substrate using electric field poling to periodically invert the ferroelectric domains. This is then followed by proton exchange and annealing steps that define the waveguide channel creating as annealed proton exchange (APE) waveguides [15, 16]. The losses, nonlinearity and field overlap between second harmonic pump and signal and idler fields are not optimum in APE structures and so a second exchange process with a lithium-rich melt is used to replace proton by lithium near the channel surface thus creating a RPE waveguide. These waveguides have lower losses because the interacting fields are pushed further from the top surface, and better overlap between the interacting modes which also enhances the nonlinearity. RPE waveguides typically have losses are 0.1-0.2dB/cm and modes well-matched to standard SMF-28 optical fiber which reduces coupling loss [17]. Whilst RPE waveguides have been used in many demonstrations of all-optical signal processing, they do suffer from some issues with photorefractive damage. This can be alleviated by operation at elevated temperatures or by using alternative materials and fabrication methods such as Ti in-diffusion. A fuller description of the various fabrication approaches can be found in the references [17].

An important parameter for any $\chi^2$ waveguide is its normalized efficiency $\eta_{out}$. For the cascaded process the output power at the idler can be written in the non-depleted pump approximation as $P_{out} = (\eta_{out} L^2 P_p/2)^2 P_s$, where $P_p$ and $P_s$ are the pump and signal powers respectively. Typically $\eta_{out} \approx 1\text{W}^{-1}\text{cm}^2$ and this implies that if the losses are negligible, 100% conversion can be obtained for a pump power of only 80mW in a 5cm long device at low signal levels. In fact one of the remarkable features of the devices based on cascaded nonlinearities is their high dynamic range and ability to up-convert single signal photons to access high detector sensitivity [18].

The second important parameter is the phase matching bandwidth for the interaction. For DFG the conversion scales with $\text{sinc}^2(\Delta k L/2)$, where $L$ is the device length and $\Delta k = k_p - k_s - k_i$ is the phase mismatch with $k_{p,s,i}$ are the $k$-vectors of the pump, signal and idler at frequencies $\omega_p$, $\omega_s$ and $\omega_i$, respectively, and $K_g = 2\pi/\Gamma_g$ with $\Gamma_g$ being the grating period. Similarly for the SHG process $\Delta k = k_p - 2k_{id} - K_g$. In either case the output fall to half it maximum when $\Delta k = 0.89\pi/L$. The DFG bandwidth substantially exceeds that for SHG and is typically 70nm FWHM for a 50mm long device, however, for SHG the bandwidth is only $\approx 0.2\text{mm}$ for a similar device length. This rather narrow pump bandwidth can be overcome either by engineering multiple gratings into the crystal to allow several different pumps to be used simultaneously [19] or by using two tunable pump beams [17].

PPLN based all-optical processors have proven highly capable in applications such as wavelength conversion [20]; dispersion compensation via phase conjugation [21]; digital signal processing including header recognition, time-slot interchange [22], etc. For an extensive review of their capabilities the reader is referred to recent reviews by Langrock et al [17] and Willner et al [23]. In terms of limitation there are three of consequence. The first arises from group velocity dispersion (GVD) because of the need to convert the pump in the telecomm band near 1550nm to its harmonic around 775nm. This large frequency difference results in GVD of $\approx 0.3ps/mm$ which means that 50mm long devices are restricted to operate with pulses $\approx 15ps$ in duration. Whilst techniques have been proposed to circumvent this issue, such as the introduction of delays lines to retime the fundamental and second harmonic waves [24], the
large GVD makes it more difficult to process signals at very high rates. Nevertheless optical processing at 640Gb/s has been reported [25]. A second limitation arises from the single polarization nature of the interaction, and generally requires the use of polarization diversity techniques. Finally, in a photonic world increasingly dominated by silicon waveguide devices, PPLN remains a material that is difficult to integrate to create highly functional circuits and because of the interactions are phase matched require temperature stabilization. Countering these are the ability of PPLN devices to operate with low noise; minimal cross talk, high bandwidth and high efficiency and no chirp making them a powerful platform for all-optical signal processing.

3. ALL-OPTICAL PROCESSING USING $\chi^3$ NONLINEARITIES

3WM can achieve efficient all-optical processing but the use of the $\chi^2$ nonlinearity requiring a non-centrosymmetric material limits its applicability to only a few materials and these are generally incompatible with current CMOS processing technology. As a result 4WM using the $\chi^3$ nonlinearity that exists in all materials has started to dominate all-optical processing. In this case, the pump signal and idler all lie in the telecommunications band, group velocity dispersion can be small, and phase matching can be achieved by engineering the structure of the waveguide so that the dispersion is anomalous. A large number of materials can be used for FWM and combined with the simply device geometry and, in most cases, compatibility with CMOS processing, allow all-optical devices to be integrated at low cost. In this section we describe some of the materials dependent factors that influence the efficiency and bandwidth for FWM.

For waveguides that utilize the $\chi^3$ nonlinearity, the nonlinear phase change is generally written in the form $\Delta \phi_{nl} = \gamma P L$, where $P$ is the laser power; $L$ is the propagation length; and $\gamma$ is the waveguide nonlinear parameter $\gamma = k_0 n_2 / A_{eff}$, with $k_0 = 2\pi / \lambda$ the free space wavenumber; and $A_{eff}$ the effective area of the waveguide mode. In a lossless waveguide, the 4WM conversion efficiency from signal to idler in the non-depleted pump approximation, can be written $\eta_{4WM} = (\Delta \phi_{nl})^2 = \gamma^2 P^2 L^2$, and is directly determined by nonlinear phase change. It is worth noting here that compared with switching devices where $\Delta \phi_{nl}$ needs to be large (>π), FWM can be quite efficient for small values of $\Delta \phi_{nl} = 0.3$. The nonlinear parameter $\gamma$ is important for any $\chi^3$ waveguide because a high $\gamma$ value minimizes $P$ and $L$, and small, low power devices are essential in most applications. In practice $P$ is often limited by energy constraints in the system, and $L$ has a maximum value determined by the loss of the waveguide. In a lossy waveguide, $L$ is replaced by an effective length written $L_{eff} = (1 - \exp(-\alpha L)) / \alpha$ where $\alpha$ is the linear loss coefficient. Because of these constraints, as high a value of $\gamma$ as possible is always preferable and this motivates the use of materials with high $n_2$ and waveguides with small $A_{eff}$.

The phase matching condition for 4WM is generally written $-\Delta P \Delta k = \beta_2 \Delta \omega^2 < 0$, where $\Delta k = 2k_p - k_s - k_i$ is the phase mismatch with $k_p, k_s, k_i$ are the k-vectors of the pump, signal and idler at frequencies $\omega_{p,s,i}$ respectively, $\Delta \omega = [\omega_{p} - \omega_{s} - \omega_{i}]$ is the frequency difference between pump and signal (or idler), and $\beta_2$ is the second order dispersion of waveguide. The dispersion of waveguide $\beta(\omega) = k(\omega) / 2\pi n_{eff}$ can be expanded as a Taylor series as $\beta(\omega + \Delta \omega) = \beta_0 + \beta_1 \Delta \omega + \beta_2 \Delta \omega^2 / 2 + \beta_3 \Delta \omega^3 / 3 + \cdots$, where $\beta_0 = d^2 \beta / d\omega^2 (\omega_0, 2, \ldots)$ at $\omega_0$, $n_{eff}$ is the waveguide effective index. The first order of the dispersion $\beta_1 = 1 / n_{g} = n_{g} / c_0$ where $n_g$ is the group velocity, $n_{g}$ is the group index and $c_0$ is the light speed in free space. The second orders of the dispersion $\beta_2 = d^2 \beta / d\omega^2 = -\lambda^2 D / 2\pi n_{g}$, where $D$ is the group velocity dispersion (GVD) of waveguide. $\beta_1$ and $\beta_2$ are the most critical parameters for all-optical processing because the pulse envelope propagates at $v_g$ and pulse broadening is relating to GVD.

![Figure 3: The group velocity dispersion (GVD) of Ge$_{11.9}$As$_{52.5}$Se$_{36.5}$ nanowire as a function of film thickness and wavelength. The waveguide width is fixed at 630nm.][26]
In order to achieve phase matching for FWM, the waveguide dispersion must be anomalous, that is $\beta_2<0$ and when this is achieved, power in the signal and idler waves grows exponentially at the expense of the pump. However, most materials exhibit normal dispersion, $\beta_2>0$, at telecommunications frequencies and it is, therefore, necessary to engineer the waveguide structure to produce sufficient anomalous waveguide dispersion so that the normal material dispersion is compensated. This means that for any particular material, only specific waveguide designs can be used and these determine the minimum mode area, $A_{eff}$, and hence the $\gamma$ of the waveguide. Fig. 3 is an example of such dispersion engineering for a 630nm wide chalcogenide nanowire and shows the effect of changing the waveguide height on the TM mode dispersion.[26] In this example, the waveguide has to be $>$450nm thick in order to achieve anomalous dispersion at 1550nm. In other materials, like crystalline silicon, anomalous dispersion occurs in smaller structures potentially leading to higher $\gamma$. Thus material dispersion becomes the dominant parameter that ultimately determines the achievable value of $\gamma$ at a particular wavelength.

In the phase matching relation, $\Delta \omega$ represents the bandwidth of 4WM and it is determined by $\beta_2$ and $\gamma P$. $\beta_2$ close to zero and a large $\gamma P$ will support 4WM over a very wide bandwidth. This is important for many devices required for telecommunications and allows 4WM to span the whole S- C- and L-bands. As a result, the ‘zero-dispersion’ is a target for dispersion engineering as well as a large $\gamma$ value.

![Figure 4](image1.png)

**Figure 4.** (a) The imaginary part of Fourier transform of the Raman response function of Ge$_{11.5}$As$_{24}$Se$_{64.5}$, As$_2$S$_3$ and SiO$_2$.[27] (b) The real part of Fourier transform of the Raman response function of Ge$_{11.5}$As$_{24}$Se$_{64.5}$, As$_2$S$_3$ and SiO$_2$.[27] (c) The experimental result of Raman modulated 4WM in a Ge$_{11.5}$As$_{24}$Se$_{64.5}$ chalcogenide nanowire.[28]

The phase-matching condition of 4WM, in fact, demonstrates the relation between the nonlinear phase change and the dispersive phase change. However, this conventional phase matching relation ignores the influence of other nonlinearities notably Raman scattering. In fact Raman scattering can cause an additional phase change that originates from the real part of the Fourier transform of the Raman response function $Re[hR(\omega)]$.[27, 29] Thus, in materials where the bandwidth for FWM extends as far as the Raman bands, the phase matching condition must be modified to include this additional phase term and thus reads $-4\gamma P \approx |Re[hR(\omega)]|<\Delta k=\beta_2 \Delta \omega^2<0$, where $R$ is the fractional Raman factor which normally range from 0.1 to 0.2 depending on material. When the $Re[hR(\omega)]>1$, the inequality becomes $-4\gamma P<\Delta k=\beta_2 \Delta \omega^2<0$, and Raman scattering does not contribute to the phase change. When $Re[hR(\omega)]<1$, it cause a positive phase change which broadens the phase-matching condition and the gain for 4WM. On the other hand, if $Re[hR(\omega)]<1$, then the Raman phase change is negative, the phase-matching condition narrows and the 4WM gain is reduced.

Figs. 4(a) and (b) shows the imaginary and real parts of $hR(\omega)$ for SiO$_2$, As$_2$S$_3$ and Ge$_{11.5}$As$_{24}$Se$_{64.5}$ glasses.[27] Typically $Re[hR(\omega)]$ remains $=1$ for small detuning, and then exhibits a resonant behavior to increase up several times right after the Raman peak of $Im[hR(\omega)]$ before dropping sharply. At large detuning, $Re[hR(\omega)] \rightarrow 0$, but never returns to its original value of unity. As a result, the bandwidth of 4WM is restricted by the spectrum of Raman scattering even if $\beta_2$ is near zero and $\gamma$ is large. 4(c) shows the effect of $Re[hR(\omega)]$ on 4WM for a Ge$_{11.5}$As$_{24}$Se$_{64.5}$ chalcogenide nanowire.[28] When the $Re[hR(\omega)]$ recovers to zero at large detuning, the gain is less than 70% of that $Re[hR(\omega)]=1$. The maximum bandwidth for high gain 4WM can be predicted using the Raman detuning frequency for the material, and is 130nm for Ge$_{11.5}$As$_{24}$Se$_{64.5}$; 180nm for As$_2$S$_3$; 190nm for SiO$_2$; 220nm for amorphous Si; and over 250nm for crystalline Si assuming a pump at 1550nm.

**Properties of $\chi_3$: Materials and Devices**
Very many materials and device structures have been considered for $\chi^3$ nonlinear optics although only a few of these have actually been used in demonstrations of optical signal processing. One of the most successful has been silica-based highly nonlinear optical fiber (HLNF) that has been used for parametric signal processing for more than two decades [30]. Despite its inherent compatibility with fiber optic systems, the small material nonlinearity implies large device lengths and additionally the fiber geometry makes integration of multiple structures for signal processing impractical. Several alternative media with much larger material nonlinearity have emerged including bismuth oxide glass; crystalline and amorphous silicon; silicon nitride; AlGaAs; chalcogenide glasses; and high-index doped silica-based glasses. In this section we summarize the characteristics of a few of these materials and discuss device properties relevant to a few on-chip platforms that use the nonlinear Kerr effect and 4WM based all-optical processing. Because of the large amount of work based on the following materials we focus on crystalline silicon (c-Si); hydrogenated amorphous silicon (a-Si:H); and silicon nitride that are all compatible to complementary metal oxide semiconductor (CMOS) processes and the somewhat less CMOS-compatible chalcogenide glasses.

**Crystalline silicon:**
In the past decade a new application for silicon has emerged as a material for photonics. The main driver for this has been the interconnect bottleneck now facing electronic processors. The major advantage of silicon photonic devices is that they can leverage the precision of advanced CMOS processing technology. However, in order to create a small-scale communication network, various functional photonic components must be integrated onto the silicon platform. This has resulted in some remarkable developments including germanium light emitters [31] and germanium photo-detector [32] integrated onto silicon chips as well as high-speed optical modulators [33]. There has also been substantial interest in all-optical signal processing utilizing the $\chi^3$ nonlinearity of Si.

Si is transparent from 1100 nm to 7000 nm and linear absorption can, therefore, be low across the whole optical communications bands between 1260 and 1675 nm, but two-photon absorption is large. This results in a relatively low $\text{FOM}_{2\text{PA}}$ of -0.4 at 1550nm corresponding to a Kerr nonlinear coefficient $n_2=4.4\times10^{-14} \text{ cm}^2/\text{W}$ and $\beta_{2\text{PA}}=8.4\times10^{-10} \text{ cm/W}$ [34]. This means that Si is, perhaps, not the best material for nonlinear photonics but since it can be used to produce complex circuits at low-cost in high-volume there has been a large incentive to explore its use in all-optical signal processing. Si also has excellent thermal conductivity and high damage resistance and because of its high refractive index can confine light into a very small mode and this can lead to exceptionally large values of $\chi$. In addition, waveguide designs that lead to near optimal confinement also correspond to those which achieve anomalous dispersion.

The silicon-on-insulator (SOI) platform has become the foundation of silicon photonics and comprises of 220nm thick top silicon layer sitting on thermally oxidized (1–2 μm SiO$_2$ layer) silicon substrate. Figure 1a shows a schematic of a typical silicon photonic nanowire, and the optical field distribution in the fundamental TE mode. The large refractive index contrast between Si ($n = 3.45$) and SiO$_2$ ($n = 1.45$) or air ($n = 1$) leads to a strong light confinement which makes it possible to scale down the size of the waveguides mode to approximately 0.1 μm$^2$. Combining high $n_2$ and small $A_{eff}$ yields an extremely large nonlinear parameter $\gamma$ of 300 W$^{-1}$m$^{-1}$ [35]. On the other hand this rectangular geometry makes much lower effective index in TM so that the mode area is a few times larger than that of TE.

Silicon photonic wire waveguides are typically fabricated by plasma etching with electron-beam or deep-ultraviolet (DUV) lithography used to define the sub-micron pattern. Nowadays, propagation losses of 1–2 dB/cm [36] are routinely achieved mostly limited by light scattering from the sidewalls but this is low enough since the devices are short (a few cm or less). A more difficult issue has been light coupling into the photonic wire due to a large mode mismatch between nanoscale silicon waveguides and fibers. Two coupling schemes are commonly employed to overcome this issue are to use surface grating couplers [36] and Si inverse tapers [37] for which the coupling efficiency reaches around 40-70%.
In spite of its many advantages over other media, silicon-based all-optical signal processing inevitably suffers from nonlinear loss and patterning effects resulting from large TPA and free carrier absorption (FCA). The lifetime of free carriers generated by two-photon absorption is in the range of several hundred ps to several hundred ns [38] and slow carrier dynamics can limit the speed of signal processing. Several approaches have been proposed to mitigate the issue as shown in Figure 5-(b), (c), and (d). The main idea is that direct light-silicon interaction can be avoided by engineering the Si waveguide structure. The first is to employ a thin strip waveguide and cover it with a material possessing both a large nonlinear Kerr coefficient ($n_2$) and a small nonlinear loss ($\beta_{\text{2PA}}$) [39]. In this case the structure makes TM mode extend far into the cladding material so that most of the nonlinear effect can be due to the upper cladding. Figure 5-(c) shows a so-called “slotted waveguide” where trenches several tens to hundreds nanometer wide are formed along the middle of Si waveguide, and this slot is filled with another nonlinear medium. With an appropriate design the light intensity in the slot is enhanced a few times relative to the unstructured waveguide [39]. Photonic crystal waveguides have also been demonstrated to further increase nonlinearity by slowing-down the light speed [40]. This slow-light effect can boost the light-matter interaction proportional to the group index squared $(c/v_g)^2$.

To conclude, it is worthwhile noting that in spite of the limitations of silicon, two recent reports clearly show that these can be overcome with careful attention to device design. Forster et al. [41] achieved peak conversion efficiencies $>10\%$ and conversion bandwidths $>150$ nm in carefully dispersion-engineered Si nanowire waveguides. For this they used rather fat and thick Si core (300 nm tall and 500 nm wide) and operated with the TM mode. The work of Denmark Technical University group is also remarkable in that ultra-high bit rate signals were processed in a nominal Si nanowire without incurring TPA and associated FCA problems [42]. They demonstrated the wavelength conversion of 640 Gbit/s signal by low-power 4WM in a 3.6-mm long silicon waveguide with a switching energy of $\sim 110$ fJ/bit, which is low enough to reduce nonlinear absorption.

**Hydrogenated amorphous silicon (a-Si:H):**
Over the years a-Si:H has been extensively deployed in applications from photovoltaics to flat-panel displays but its applications in lightweight technology is just evolving. As a nonlinear medium for all-optical signal processing, amorphous silicon outperforms crystalline silicon in some respects. First of all, this material retains CMOS-compatibility so that low-cost mass manufacturing is available. Another attractive feature of a-Si:H comes from the fact that it can be deposited at low temperature (below 400°C) on almost any substrate including glass, metal and even plastic, whilst crystalline silicon (SOI) does not offer this flexibility. Moreover, the low temperature deposition allows back-end integration of a-Si:H photonic components on pre-processed CMOS electronic chips without any damage to the underlying metal routing wires [43]. In addition it also proves to be a highly promising nonlinear material for power efficient photonic devices due to its ultra-high optical nonlinearity and low nonlinear loss relative to crystalline Si.
Plasma-enhanced chemical vapour deposition (PECVD) is a well-established technique for depositing high quality amorphous silicon-hydrogen alloy (~10at% H) at low temperature. Silane (SiH₄) is the dominant precursor gas, mixed with hydrogen, helium, or argon as a diluent. If amorphous silicon is deposited by sputtering it has mid-bandgap absorption in the near infrared due to dangling bonds. These Si dangling bonds, however, can be saturated by incorporating hydrogen during PECVD, resulting in the low optical attenuation below 1 dB/cm [44]. The fabrication of a-Si:H photonic nanowaveguides is identical to that of SOI; i.e. optical or electron-beam lithography followed by plasma etching and SiO₂ PECVD for cladding. a-Si:H nanowires have similar or slightly smaller dimension compared with crystalline Si because of its higher index. The linear propagation losses achieved in a-Si:H waveguides lies in the same range as c-Si; that is a few dB/cm, and even ~1 dB/cm has been achieved by adopting a delicate thickness control technique [45].

The nonlinear characteristics of the material are favourable with a Kerr nonlinearity (nₑ) reported to be several or up to 10 times [46] higher than that of c-Si. It is debatable as to whether TPA in a-Si:H is a still the dominant nonlinear loss mechanism in the telecommunication band, since TPA should vanish since the photon energy is less than the half the bandgap (E₂/2~0.85 eV) of the medium. Nevertheless, it is generally considered to have similar level of nonlinear absorption to c-Si. Because of the high nonlinearity, γ can reach 3,000 W⁻¹m⁻¹ in a 500 nm wide, 200 nm high nanowire [47]. Utilizing this extremely high nonlinear parameter in a-Si:H, Wang et al.[47] demonstrated wavelength conversion through nonlinear parametric processes with maximum conversion efficiency of -13 dB using mere 15 mW of pump peak power. In addition on-chip parametric amplification with 26.5 dB gain was attained in a-Si:H photonic wire at telecommunication wavelengths [48].

A key concern of hydrogenated amorphous silicon devices has been its lack of stability. Kuyken et al. [49] observed a decrease of parametric amplification with the time of exposure to pump light. They asserted that this results from material degradation and has the same origin to the Staebler-Wronski effect [49], which is well known to the a-Si solar cell community. Most recent results, however, demonstrated that deterioration in amorphous silicon may not be an issue. The large scatter in nₑ, β₂PA (or β₃PA), FOM₃PA, γ, and free carrier lifetime in the literature is another nagging problem. For instance, the measured β₂PA spans from 41×10⁻¹⁰ to 2.5×10⁻¹⁰ cm/W. The origin of this large variation remains an open question, but it is a general thought to be due to differences in the film deposition process. Even so, a-Si:H is quite a promising platform for ultra-compact all-optical signal processing chips operating at low power.

Silicon nitride (SiN): This is another CMOS-compatible material that has recently emerged as a candidate for nonlinear photonics. Since SiN can be deposited from gas precursors in a standard silicon processing environment, it benefits from flexibility in deposition parameters which produce different film characteristics. Despite nₑ of SiN (nₑ≈2.5×10⁻¹⁵ cm²/W at 1.55μm) being more than an order of magnitude lower than c-Si, the most significant advantage of SiN over crystalline and amorphous Si comes from the absence of 2PA or multi-photon absorptions and associated FCA due to its wide bandgap (E₂~5eV) [50].
Low-pressure chemical vapour deposition (LPCVD) at high temperature (~800 °C) is the preferred method to grow low loss SiN layer and leads to minimal hydrogen contamination in the film. Using this method, propagation losses of 0.5dB/cm loss was obtained in a SiN waveguide with 1 µm$^2$ cross section [51]. In order to facilitate back-end integration with pre-existing electronic circuits, however, PECVD below 400 °C was also demonstrated to produce SiN waveguides with only ~ 1dB/cm propagation losses. Here, the absorption in the near IR was controlled by reducing the number of hydrogen bonds (Si-H and N-H) in the grown SiN by employing techniques such as adding helium into the plasma [52] and replacing NH$_3$ with a N$_2$ precursor [53].

The group at Cornell University [51] demonstrated on-chip multiple wavelength source exploiting 4FM optical parametric oscillation in a SiN micro-ring. (Figure 6) They used a single pump laser tuned to the resonance of the ring, then ~1000 times higher light power circulating in the ring which leads to cascaded 4WM, and allowing narrow multi-wavelength lines.

Diaz et al. [54] reported a systematic comparison of eight nonlinear media for parametric all-optical signal processing using degenerate 4WM wavelength conversion. Figure 7 represents the minimum pump power required to attain a given conversion efficiency (CE). Here CE is the ratio of the idler power at the waveguide output to the signal power at its input. Crystalline and amorphous silicon can operate at low pump power because of their very high $\gamma$, but their CE is limited to around 10% due to TPA and FCA. Whilst the performance of SiN is inferior to c-Si and a-Si because of its smaller nonlinear parameter, it can handle much higher pump power without suffering nonlinear losses. An additional difficulty, however, is that the large mode area needed for dispersion engineering which is around 1µm$^2$ results in the footprint of SiN being two orders higher than the equivalent devices made in c-Si and a-Si:H and this makes it difficult to achieve compact devices.
Figure 7 Minimum pump power required to achieve a given CE for the different materials [54]

**Chalcogenide glass:** Chalcogenide glasses have attracted a lot of interest for all-optical signal processing. These are amorphous, highly nonlinear materials containing the chalcogen elements S, Se and Te covalently bonded with glass forming materials such as Ge, As, P, Sb or Si. Chalcogenide glasses provide low optical loss in the telecommunication bands (1310 nm and 1550 nm) and transmission out to beyond 8µm in the infrared. These glasses have high linear refractive indices (2.0-3.0 at 1550 nm) which leads to small mode volumes in dispersion-engineered nanowires and a Kerr nonlinear index similar to c-Si (≈ a few hundred × silica) [55, 56]. Most importantly they generally have negligible 2PA which leads to FOM\textsubscript{2PA} > 100 [57-59]. In most chalcogenide glasses, there are no free carrier effects and this provides a distinct advantage compared with the semiconductors.

As\textsubscript{2}S\textsubscript{3} is one of the most well known chalcogenide glasses with a linear index of 2.43 at 1550nm, a nonlinear index over 120 times silica and FOM\textsubscript{2PA} over 320 and a glass transition temperature over 170°C. The first reports demonstrating the efficacy of As\textsubscript{2}S\textsubscript{3} fiber for all-optical processing were produced by Asobe [60]. More recently a large number of ultra-fast nonlinear devices have been demonstrated using rib waveguides made from this glass including a signal regenerator [61]; RF spectrum analyser [62, 63]; dispersion compensator using mid-span spectral inversion [64]; TDM demultiplexers [65]; wavelength convertor [66]; parametric amplifier [29, 67]; as well as a supercontinuum source [68].

Although, As\textsubscript{2}S\textsubscript{3} glass has, therefore, exhibited some good properties for all-optical processing, even better chalcogenide materials would be of advantage particularly if higher nonlinearity and an absence of photosensitivity can be obtained [69]. High nonlinear index and low TPA has been demonstrated in Ge-As-Se, Ge-As-S-Se and As-S-Se compositions [55, 56, 58]. Germanium rich compositions have reported nonlinearities between 200-900 times that of fused silica. It was found that the optical properties of the glasses can be improved by fine tuning the composition in the Ge-As-Se system.

The Ge\textsubscript{11.5}As\textsubscript{24}Se\textsubscript{64.5} glass is one of a small family of Ge-As-Se glass compositions that show high nonlinearity, particularly high thermal and optical stability when exposed to near band edge light and good film forming properties. It is known that the physical properties of ternary chalcogenides vary significantly as a function of their chemical composition and mean coordination number (MCN = the sum of the products of the valency times the atomic abundance of the constituent atoms) and that MCN can be used to categorize the basic properties of the glass network. In the case of Ge-As-Se glasses, it has been found that their linear and nonlinear refractive indices, optical losses; elastic properties, etc, all vary strongly with MCN. Ge\textsubscript{11.5}As\textsubscript{24}Se\textsubscript{64.5}, for which MCN=2.47, lies in the so-called “intermediate” phase (IP) which lies between the “floppy” and “stressed-rigid” glass networks. Of particular significance is that films produced by thermal evaporation in this region have similar properties to those of bulk glasses – an unusual behaviour for films made from ternary chalcogenides.
Compared with As$_2$S$_3$ glass, which has been the workhorse for all-optical devices to date, Ge$_{11.5}$As$_{24}$Se$_{64.5}$ has a higher linear refractive index (2.65 compared with 2.43) which leads to better mode confinement and, as would be expected from Miller’s rule, a higher nonlinear index, $n_2$ (8.6x10$^{-14}$cm$^2$/W c.f. 3x10$^{-14}$cm$^2$/W for As$_2$S$_3$ at 1500nm) [70]. No TPA could be detected during z-scan measurements made on bulk Ge$_{11.5}$As$_{24}$Se$_{64.5}$ samples, however, the FOM has been found to be ≈ 60 from measurements of the power-dependence of the transmission for Ge$_{11.5}$As$_{24}$Se$_{64.5}$ rib waveguides, as shown in Fig. 8 (a) [71]. Ge$_{11.5}$As$_{24}$Se$_{64.5}$ nanowires with $\gamma$=135 W$^{-4}$m$^{-1}$ have been reported with losses around 1.5-2.5dB/cm [26, 28]. 4WM and SC were demonstrated in such nanowires as well as polarization independent properties using the structure shown in Fig. 8 (b) [28]. These nanowires were also used to demonstrate correlated quantum photon pair generation [27, 72]. High Q (>700,000) photonic crystal cavities are also made from this material for optical switching application, as shown in Fig. 9 [73].

4. ALL-OPTICAL PROCESSING IN SEMICONDUCTOR OPTICAL AMPLIFIERS

Semiconductor optical amplifiers (SOAs) have a long history in all-optical processing with their popularity stemming from the range and strength of nonlinear effects available; from the early commercial availability of fibre pigtailed devices; from the capacity for on chip integration of multiple functions/devices; and from the potential for low power operation with net conversion gain. Such is the diversity of architectures that it is beyond the scope of this summary to cover these in their entirety and the reader is instead referred to some of the excellent recent review articles and books on the topic for the wider view, e.g. [74-81]. Here the types of nonlinear effects will be briefly reviewed, the limitations facing SOAs examined along with some of the proposed remedies and the potential limits, and the best and latest achievements briefly summarized.

4.1 TYPES AND ORIGINS OF NONLINEAR EFFECTS IN SEMICONDUCTOR OPTICAL AMPLIFIERS

As electrically pumped devices, SOAs are dominated by carrier density driven nonlinearities. These result in effects operating on the amplitude, phase and polarization of optical signals. The first and most obvious effect is due to the signal power dependence of the SOA gain which saturates as a result of carrier depletion as the input power is raised. Thus, cross gain modulation (XGM) occurs where an intense pulse modulates a co-propagating CW beam at a different wavelength. This process is commonly used in wavelength conversion. Gain saturation is also used in optical limiting where the SOA output power becomes essentially independent of the input power, thereby reducing amplitude noise on high signal levels. This is particularly useful for non-return to zero (NRZ) modulation.

The Kramers-Kronig relation means gain or absorption changes are always accompanied by changes in refractive index and also therefore phase change in the propagating signals. Refractive index power dependence of course manifests as a $\chi^3$ effect. Ascribing a precise value to the nonlinearity is troublesome.
due to the differing contributing parts [82], but to give some comparative indication of the effective size of $n_2$, values from $2 \times 10^{-12}$ up to $\sim 1 \times 10^{-9}$ cm$^2$/W have been reported, e.g. [83-85]. However the response time of the electron plasma is not as fast as the response of bound electron states in glasses, for example, leading to speed limitations as will be discussed shortly. Cross phase modulation (XPM) between a strong pulsed pump and a weaker CW probe can also be exploited in SOAs for wavelength conversion by placing the SOA inside a Mach-Zehnder or other interferometer.

Nonlinear polarisation rotation (NPR) in SOAs has also been increasingly harnessed for high speed nonlinear processing. This results from a combination of the different gain saturation behaviours of the TE and TM modes, and also from the intrinsic waveguide birefringence in the SOA (with both geometrical and stain components) which can also be modulated by changes in the carrier density. Thus a signal launched into both polarization states sees a “waveplate” with an effective input power dependent retardation, and this can be used for all optical processing.

4.2 IMPAIRMENTS IN SOA DEVICES

4.2.1 CARRIER INDUCED SWITCHING SPEED LIMITATIONS

The strengths of SOAs for nonlinear processing stem from the strong carrier-based nonlinearity in the devices, but this also leads to their main shortcoming. A major driver for all-optical signal processing is speed, but the relaxation times of carrier-based nonlinearities are long compared with those due to bound electron (Kerr) nonlinearities in dielectrics. This is the most significant impairment for SOA based devices and so will be explored in some detail.

The dynamics of the gain compression and its associated phase shift has been studied by quite a number of groups, e.g. [83, 86-101] with good summaries of the varying contributing processes and their distinct timescales being available e.g. [75, 76, 102]. It should also be noted that the SOA gain region structure (bulk vs Quantum well (QW) vs Quantum dot (QD) [103]) also strongly influences the relative contributions of these components. The recovery time is also affected by the input power, the polarization, the level of bias current applied to the SOA, the dimensions and design of the active region, and the pump wavelength, e.g. [92, 94, 97-101]. This makes it hard to compare recovery times in any absolute sense, but nonetheless some broad generalizations of structural impacts, and mitigating approaches is in order. It is also important to note that the phase recovery has different characteristics to the gain recovery.

Hall et al. [83], undertook one of the most comprehensive earlier studies at 1550nm in bulk and strained multi-quantum well devices, showing that carrier injection and carrier heating effects dominated the dynamic recovery of the gain and phase. Under the conditions studied in their pump-probe experiment, the QW device exhibited faster gain recovery than the bulk albeit with a small but long lived tail, but broadly similar phase recovery to the bulk device, however at about one third the amplitude. Responses from the devices studied are reproduced from their paper below in Fig. 10.

![Fig. 10 - Bulk gain left and phase underneath, MQW to the right from [83]](image)
QD SOAs exhibit a different response again with very rapid gain recovery but rather slower and smaller phase response as shown in Fig. 11. The reason for the more rapid gain recovery in QD devices is considered to be the action of the wetting layer used to grow the dots as a charge reservoir that feeds the dots locally and therefore quickly. The wetting layer itself can also deplete but this has to be refilled via normal interband mechanisms and so displays the characteristic long time constant (visible as a very small and long exponential recovery tail in Fig. 11).

![Fig. 11 — Gain and phase recovery of typical QD SOA (from [104])](image)

Taking the published results as a whole, the obvious question is how short can the recovery times go? As noted earlier operating parameters influence the recovery times as was recently graphically illustrated by Cleary et al. [75, 105]. By optimizing the pump and probe wavelengths relative to the SOA gain peak, they achieved 10–90% gain recovery times from 3–5ps in an MQW SOA, up to 7x improvement over some of the unoptimized conditions. The phase response was improved by factors up to 2x. QD SOAs have also demonstrated similar gain recovery times [104, 106]. These results suggest bit rates of 3–500GB/s should be possible at least using XGM effects.

There are other effects that can also be harnessed in the quest for ultimate speed. To improve pure carrier recovery these include doping to reduce the carrier lifetime in the bulk/feeding zones, exploiting tunnelling effects in QW and QD structures [91], the use of pulsed or CW optical holding/reset beams[86, 107, 108], and increasing the energy gap of QDs to the wetting layer [97] amongst others. Exactly how far the intrinsic recovery time can be pushed is not yet clear, but modelling has shown that speeds of 1TB/s appear possible in some devices[93].

Sub-picosecond phase response times without long lived tails have also been reported both experimentally & theoretically in MQW SOAs and have been used in NPR based wavelength conversion experiments under carefully tailored low current injection regimes where two photon absorption dominates [109-111]. However the major issue with this mode of operation is the large pulse energy required for switching (~10μJ meaning an average power at the SOA input of 5W for a 1TB/s PRBS RZ stream!) and the observation that the output pulses were broadened considerably (nonetheless at 500fs these are still the shortest pulses ever wavelength translated in an SOA).

In addition to directly attacking the carrier recovery, architectural design can also aid in improving the overall system response time. To cite a single example from many, in XGM based wavelength conversion the nature of the phase response can also be exploited as it causes the long recovery tail to be red shifted on the probe and so a blue shifted bandpass filter can remove it, e.g. [112]. This effect was used to demonstrate XGM based wavelength conversion at 640GB/s SOAs with a carrier recovery time of ~60ps [113]. A wide variety of other architectures have also been developed to enhance overall switching speed as are discussed elsewhere, e.g. [74-81].

The discussion above relates to nonlinear devices utilising the gain or XPM effects, but the Kerr effect in SOAs is large enough to enable Four Wave Mixing (FWM) based processing. There are two distinct scenarios for FWM operation, one where intense CW pumping is used and the pulsed input signal is relatively weak (often for optical phase conjugation), and the other where the input pulsed signal is strong and the CW probe used to generate the wavelength shifted idler is weak. In the latter case all the recovery processes & timescales discussed above clearly apply and this scenario has also been analysed by several authors for interactions between pulses, e.g. [88, 114, 115]. In the former case, the recovery processes are strongly suppressed due to the strong CW pump leading to the question of how fast a signal can be accurately phase conjugated or wavelength translated by this process?

Surprisingly there seem to be no analyses of the ultimate speeds achievable, though Kikcuhi et al. [116] Summerfield et al. [117] and Kim et al. [118] recognized that the ratio of CW pump power to signal...
power is critical and that a trade off exists with conversion efficiency (due to carrier lifetime effects) and SNR for bulk SOAs. In QD SOAs different and less stringent restrictions apply to the conversion efficiency eg. [88, 103] but the trade off still exists. Experimentally, Kikuchi et al. [119] demonstrated phase conjugation of a 1.6ps pulse train in an SOA cascade and successful transmission through 40km of standard single mode fibre limited only by third order fibre dispersion differences resulting from the wavelength translation. Taken together this suggests that the potential for TB/s CW pumped FWM wavelength translation and OPC in SOAs are good.

4.2.2 OTHER LIMITATIONS AND IMPAIRMENTS OF SOA NONLINEAR ELEMENTS

Speed aside, there are other potential limitations in using SOAs as nonlinear elements. First and foremost all the nonlinear effects in an SOA except the Kerr effect require amplitude modulation. Phase, subcarrier, and polarization encoding are increasingly important in today’s communication systems and amplitude modulation is no longer implicit. Where modulation formats such as NRZ DQPSK are employed, the signal has no amplitude modulation and XGM, XPM, and NPR cannot be harnessed for nonlinear signal processing. FWM however is a format transparent effect and remains highly useful for wavelength translation and phase conjugation.

Another obvious issue in using a device with gain is the linear addition of amplified spontaneous emission noise and its effect on optical signal to noise ratio, but also the less well studied nonlinear interactions between noise and the signal. OSNR is well understood and can be engineered and optimized with well proven approaches. Nonlinear noise based interactions in SOAs have been studied by numerous groups but most recently and with the least approximation by Connelly et al. [120] for phase encoded signals in the context of line amplifiers operating in the saturated regime (i.e. as nonlinear elements are). However further study is needed to clarify whether the effects of the various nonlinear noise mixing processes in operation present a limiting factor to high speed data processing.

A further limitation often considered to limit the applicability of SOAs is multichannel crosstalk. This relates to the imposition of the modulation of other channels onto the channel under consideration by XGM, XPM, and NPR as a channel ensemble passes through an SOA. In general this is an issue for devices relying on these effects as the modulated signal is usually the pump and so it is rather difficult to suppress the crosstalk as the nonlinear effect being exploited relies upon it. However it can sometimes be managed enabling multichannel operation with acceptable performance, e.g [121].

There is however one potential exception to this which exploits the properties of QD SOAs. QD SOAs have a range of dot sizes, each size amplifying a reasonably narrow portion of the optical spectrum through the homogenous broadening of the theoretically atom like DOS. Thus if the signals are sufficiently separated in wavelength that they each address different QD cohorts, then they will not impact on each other’s gain to first order [103] (at a more detailed level they are all fed by the same wetting layer which may deplete based on the total instantaneous power, this however as seen previously is a lower level effect in QD SOAs with appropriate design and operating conditions).

Multichannel Crosstalk is also an issue in devices employing FWM. Here all the same XGM mediated effects can apply if the powers are high enough, but in addition, there are also multiple interchannel FWM products that can fall inside the desired signal bandwidth [117]. Whilst certain optimisations can minimise this, the effects cannot be eliminated completely. Multichannel devices usually operate with a strong CW pump that suppresses the gain compression based effects, and early work demonstrated that even for purely intensity modulated data that the effects were acceptable for four channel conversion [122]. With NRZ phase modulated data the gain compression effects are absent leaving only the intermixing effects, and four channel NRZ DQPSK modulated signals were wavelength converted by FWM in a QD SOA with < 4dB penalties from intermixing [123]. Much further work however remains to optimise the systems architecture to attain higher channel counts.

4.3 CURRENT STATE OF THE ART AND FUTURE PROSPECTS FOR SOA NONLINEAR PROCESSING

Whilst the range of applications and therefore relevant results is too broad to summarise meaningfully here, there are a number of key achievements for SOA based nonlinear devices related to speed, conversion range, and conversion efficiency which will now be briefly discussed. Taking speed first, Table I gives an overview of some of the fastest demonstrations using both pulses and full data streams.
Table I. – Details of fastest and most recent nonlinear switching demonstrations in SOAs

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Nonlinear effect</th>
<th>Data Rate (GB/s)</th>
<th>Data format</th>
<th>Pulse width (ps)</th>
<th>Operation</th>
<th>Conversion efficiency (dB)</th>
<th>Pump Power</th>
<th>Detection penalty (dB @10^-9 BER)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[113]</td>
<td>XGM, blue shifted bandpass filter</td>
<td>640</td>
<td>RZ-OOK</td>
<td>0.8</td>
<td>TD M demux, 640 to 40 GB/s</td>
<td>~22</td>
<td>8dBm Ave.</td>
<td>13.5</td>
</tr>
<tr>
<td>[124]</td>
<td>XPM in symmetric Mach-Zehnder</td>
<td>640</td>
<td>RZ-DPSK</td>
<td>0.38</td>
<td>TD M demux, 640 to 40 GB/s</td>
<td>?</td>
<td>17dBm Ave.</td>
<td>4</td>
</tr>
<tr>
<td>[125]</td>
<td>XGM, blue shifted bandpass filter</td>
<td>320</td>
<td>RZ-OOK</td>
<td>1.2</td>
<td>TD M demux, 320 to 40 GB/s</td>
<td>~19</td>
<td>8dBm Ave.</td>
<td>8</td>
</tr>
<tr>
<td>[126]</td>
<td>XGM, blue shifted bandpass filter</td>
<td>320</td>
<td>RZ-OOK</td>
<td>1.2</td>
<td>Full C band wavelength shift</td>
<td>-13</td>
<td>7dBm Ave.</td>
<td>5</td>
</tr>
<tr>
<td>[127]</td>
<td>XGM, blue shifted bandpass filter</td>
<td>320</td>
<td>RZ-OOK</td>
<td>1.0</td>
<td>Wavelength shift</td>
<td>~17</td>
<td>5.4dBm Ave.</td>
<td>~10</td>
</tr>
<tr>
<td>[128]</td>
<td>FWM</td>
<td>320</td>
<td>RZ-OOK</td>
<td>1.2</td>
<td>Wavelength shift &amp; phase conjugation</td>
<td>-18</td>
<td>9dBm CW</td>
<td>4</td>
</tr>
<tr>
<td>[119]</td>
<td>FWM</td>
<td>-</td>
<td>Pulses</td>
<td>1.6</td>
<td>Wavelength shift &amp; phase conjugation</td>
<td>-5</td>
<td>7.3dBm CW</td>
<td>-</td>
</tr>
<tr>
<td>[110]</td>
<td>NPR</td>
<td>-</td>
<td>pulses</td>
<td>0.12</td>
<td>Wavelength shift</td>
<td>12</td>
<td>10pJ per pulse</td>
<td>Output pulse 0.5ps</td>
</tr>
</tbody>
</table>

There are a number of pertinent points to consider from this summary. High bit rates, up to 640 GBit/s, have been achieved with error free transmission and under special circumstances that even shorter pulses have been converted. However this has not been achieved to date will all nonlinear effects, nor with very low detection penalties. The detection penalties at best were 4dB, this is not a small number compared to the essentially penalty free results achieved in for example passive waveguide demultiplexers using FWM at even 1.2TBit/s [65], although such schemes are much less power efficient than SOAs. The origins of the penalties seem to be connected in these two cases of 640GBit/s performance to the filtering required to speed up the SOA response, particularly for the XGM based method when used for demultiplexing.

Conversion efficiencies for the system experiments were also relatively low though the input powers were much lower than used in passive nonlinear waveguide devices. In particular, [126] shows performance that appears close to usable in a real system. However there was only a single demonstration at high speed of a device that could be format independent using FWM [128]. This would of course allow for higher bit rates through multilevel coding schemes, and the use of phase only modulation would also reduce XGM effects considered to be partly responsible for the detection penalty (the output OSNR of ~23dB was though to make up the remnant part).

With further research and perhaps utilizing some of the techniques discussed earlier the detection penalties might be decreased in the future, conversion efficiencies increased and possibly operation at 1.2TBit/s demonstrated as some modelling has predicted [93].

Where speed is a lesser concern (e.g. for 40GBit/s RZ-OOK or 100GBit/s coherent schemes), then some impressive results have been achieved in terms of conversion efficiency, range, and regenerative performance. Table II summarises a few of the more recent and impressive examples.

The work of Contestabile et al. [129, 130] has conclusively demonstrated that ultra-broadband wavelength conversion with gain is possible in a single device with minimal signal degradation, and that it can be format and bitrate independent at least to symbol rates <100GBaud/s. Additionally the phase conjugation function can also be obtained if desired for distortion compensation. Perhaps the only shortcoming of their experiment was that the device was not polarization insensitive, but this can be addressed with dual pump schemes [131, 132] and an appropriately optimized QD SOA, or by using a polarization diversity approach to obtain the best performance [133].

Filion et al. [134] recently demonstrated wideband conversion with >10% efficiency of even high order multiplexed signals for the first time. They also demonstrated data rates up to 320GBit/s with low penalties including the phase conjugation function.
Given the de facto nature of DWDM technology, the conversion of multiple channels in a single device is clearly also a priority. Matsuura et al. [123] demonstrated that acceptable penalties can be achieved for wavelength translation/OPC in a single SOA even with coherent transmission. The penalties resulted from intra-channel mixing products, meaning that to process a complete DWDM comb four or less channel banding would be required, as partly demonstrated by Andriolli et al. [136].

Taking this body of work, it is clear that the prognosis for system deployment of format transparent highly efficient wideband wavelength translation devices using in SOAs is very good. All optical regeneration of coherent signals in an SOA is becoming competitive. Unlike On-Off Keying where an electronic regenerator can be comparatively simple, coherent versions are relatively complex and expensive. Saturation effects in SOAs can also be used to limit amplitude noise without introducing significant phase noise (or even reducing it in some cases) to improve the signal Q factor or reduce the detection penalty. Porzi et al. [135] and Wu et al. [121] demonstrated significant improvements in the signal quality of coherently coded data this way, and in the case of Wu et al. did this with 8 channels of polarisation and wavelength division multiplexed data. In both cases these operations were performed in a simple integratable single SOA based device.

Finally there have also been developments to enable effects based on FWM to be mimicked in SOAs. Optical phase conjugation has recently been accomplished for the first time based on non saturated FWM processes in SOAs. In Ref. [129] DM comb four or less channels were demonstrated using 1D photonic crystal nanocavity,” Opt E

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Nonlinear effect</th>
<th>Operation</th>
<th>Conversion range (nm @ X dB eff.)</th>
<th>Pump Power</th>
<th>Detection penalty</th>
<th>Bit rate</th>
<th>Data format</th>
<th># channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>[130]</td>
<td>Single pump FWM</td>
<td>Wavelength shift /OPC 11.2 @-&gt;0 28 @&gt;-10</td>
<td>6dBm CW</td>
<td>1</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>QPSK 8 PSK 16 QAM</td>
</tr>
<tr>
<td>[129]</td>
<td>Dual pump FWM</td>
<td>Wavelength shift /OPC 100 @ &gt;0 140 @ &gt;-10</td>
<td>3dBm CW per pump</td>
<td>? error free max error vector mag. 14%</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>QPSK 8 PSK 16 QAM 0FD</td>
</tr>
<tr>
<td>[134]</td>
<td>Single pump FWM</td>
<td>Wavelength shift /OPC 32 @ -&gt;-10</td>
<td>5dBm CW per pump</td>
<td>~1.5</td>
<td>~2</td>
<td>256</td>
<td>320</td>
<td>16 QAM 64 QAM</td>
</tr>
<tr>
<td>[123]</td>
<td>Single pump FWM</td>
<td>Multichannel OPC</td>
<td>-</td>
<td>10dBm CW</td>
<td>4</td>
<td>50</td>
<td>N-RZ-DPSK</td>
<td>4</td>
</tr>
<tr>
<td>[135]</td>
<td>Single pump FWM</td>
<td>Regeneration via saturation of FWM eff.</td>
<td>10 @ &gt;-20</td>
<td>13.5dBm CW Q raised from 6 to &gt;12.5</td>
<td>10</td>
<td>N-RZ-DPSK</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>[121]</td>
<td>XPM plus filter</td>
<td>Regeneration</td>
<td>N/A</td>
<td>1dBm Ave. /channel</td>
<td>-2</td>
<td>20</td>
<td>PM-RZ-PSK</td>
<td>8</td>
</tr>
</tbody>
</table>

Table II – High conversion efficiency and multiple format NLO processing demonstrations in SOAs

REFERENCES
26. X. Gai, S. Madden, D.Y. Choi, D. Bulla, & B. Luther-Davies, "Dispersion engineered Ge(11.5)As(24)Se(64.5) nanowires with a nonlinear parameter of 136W(-1)m(-1) at 1550nm," Opt Express 18, 18866-18874 (2010).


71. A. Prasad, "Ge-As-Se chalcogenide glasses for all-optical signal processing," in *Laser Physics Center*(Australian National University, 2010).
73. X. Gai, B. Luther-Davies, & T.P. White, "Photonic crystal nanocavities fabricated from chalcogenide glass fully embedded in an index-matched cladding with a high Q-factor (> 750,000)," Opt Express **20.** 15503-15515 (2012).