https://doi.org/10.1038/s41566-023-01191-3

Nonlinear optics from the viewpoint of interaction time

Jacob B. Khurgin

In recent decades, progress in developing better nonlinear materials has not been as rapid as wished. Here I propose that this may be explained by considering the simple view of nonlinear optical phenomena as being determined mostly by the length of interaction time between photons and matter. Tentative routes towards improvements in the efficiency of nonlinear optical phenomena are suggested.

The field of nonlinear optics (NLO) as we know it was born shortly after the first demonstrations of lasers in the early 1960s[1]. Bloembergen's early treatise *Nonlinear Optics* appeared in 1965[2], followed two decades later by a more comprehensive treatment by Shen, *Principles of Nonlinear Optics*[3]. By that time, the fundamental science of NLO had been developed, and most of the second- and third-order nonlinear phenomena were well understood and experimentally characterized. If one takes a look at the first edition of Shen's work (published in 1984), one can see that the ensuing decades have not added much to what was contained in that book. Forty years on, one still reads papers about revolutionary applications such as all-optical switching[4] and ultrafast signal processing[5] in new exotic materials, whereas for the relatively limited list of practical applications (harmonic, optical parametric, continuum and optical frequency-comb generation) one still relies on a handful of well-developed nonlinear crystals ($LiNbO_3$, KTP, BBO, metal chalcogenides), and on optical fibres and silicon waveguides. The situation is strikingly different from many other areas of photonics. For example, take lasers[6,7], where rapid progress has made many otherwise excellent 40-year-old books limited in scope, as they contain no mention of the devices that are the mainstay of photonics today: quantum cascade, Yb-fibre, Ti-sapphire, GaN blue lasers and vertical-cavity surface-emitting lasers, to name a few. Past years have seen an ebb and flow of news about promising nonlinear materials or schemes. The incomplete list includes nonlinear polymers, semiconductor quantum wells and dots, nanotubes, photonic crystals, slow light, plasmonics and metamaterials, graphene and other 2D materials, perovskites, topological materials and many others.

With a few notable exceptions (photonic crystal fibres[8] for continuum generation, and various new materials for saturable absorbers) the research on these topics, while extending our horizons, has yielded relatively modest practical results. The long-foretold all-optical digital computer[9] has not materialized, and the jury is still out on all-optical neural networks[10]. If anything, among the greatest successes of NLO in recent decades have been advances in the areas where NLO effects are deleterious, for example in techniques to mitigate nonlinearities in high-capacity fibre communication links[11] (I am not touching on the different and very successful field of nonlinear spectroscopy[12] here). With six decades of NLO research behind us, in my view it would be worth taking a fresh look at the limits of NLO, trying to develop a unified

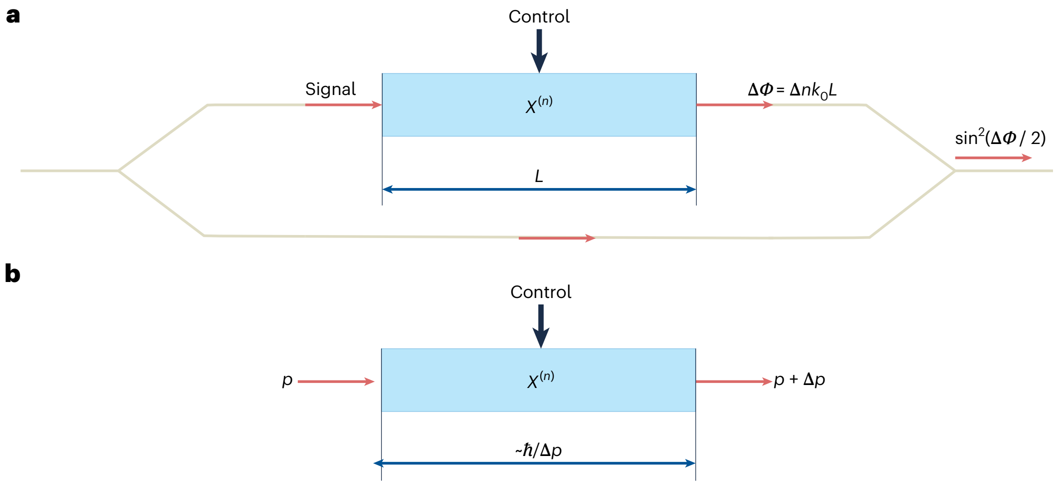

Fig. 1 | Equivalence of switching via refractive index modulation and measuring the change of the photon momentum. a, An example of arrangement for nonlinear switching of signal light by a control light using a Mach–Zehnder interferometer. One can use all types of interferometers or filters with essentially the same requirement for efficient switching, $\Delta\Phi_{\mathrm{nl}} \sim \pi/2$. **b**, Converting nonlinear phase shift into intensity switching is equivalent to measuring the momentum change $\Delta p = \Delta n \hbar k_0$ and requires a minimum interaction length, in accordance with Heisenberg's uncertainty principle.

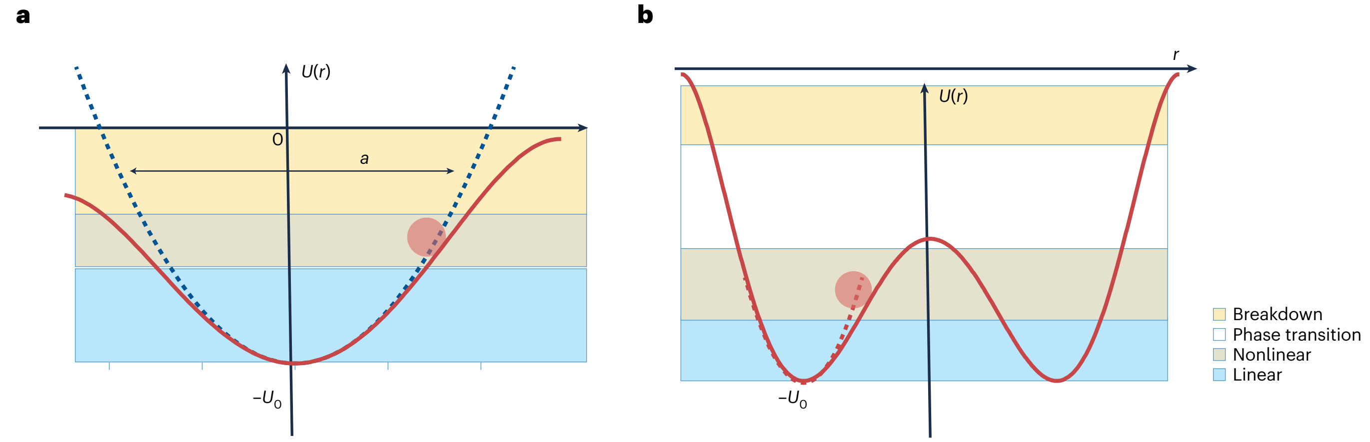

Fig. 2 | Optical nonlinearity is related to optical breakdown. a, Anharmonic potential showing that nonlinearity becomes appreciable only close to optical breakdown. **b**, Anharmonic potential in the vicinity of phase transition.

picture of its diverse phenomena, and providing engineers with intuitive understanding of the boundaries of the possible.

This discourse is neither a comprehensive review nor a tutorial. Nor it is an attempt to divine the future of all-optical processing and other attention-catching applications. Whereas the visionary scientist's role is like that of an architect conceiving inspiring designs, I see my role as that of a structural engineer coldly assessing whether the design will stand on its own and the roof will not leak. What follows is my own subjective commentary formed by four decades of experience, with which some readers may disagree. I focus exclusively on nonlinearities associated with the real part of the refractive index (switching, modulation and frequency conversion), because it is these that tend to rouse the imagination and portend all-optical networks in which the optical signals shuffle smoothly between the components, changing wavelengths and modulation formats with lightning speed. Thus, important (even though less futuristic) and more developed areas of saturable absorption for mode-locking, nonlinear and Raman spectroscopies are left out of the scope. Also left out is terahertz NLO, because this region is closer to the electronic rather than optical domain.

A proper figure of merit (and why we should not 'divide by zero')

The established metric for the NLO properties is nonlinear susceptibility $\chi^{(n)}$, or, in the case of third-order nonlinearity, nonlinear index n_2. Either of these measures describes the permittivity change imposed by unit electric field ($\chi^{(2)}E$) or the square of it ($\chi^{(3)}E^2$). If modulation of the index is time-dependent, it results in frequency conversion. However, the index modulation *per se* means very little unless it is accomplished over sufficient length L. Indeed, the efficiency of frequency conversion increases as $\sin^2(\chi^{(n)}E^{n-1}k_0L/2n)$ where wavevector $k_0 = \omega/c$[3]. Nonlinear index change is $\Delta n = \Delta\varepsilon/2n = \chi^{(n)}E^{n-1}/2n$, so conversion efficiency changes as $\sin^2(\Delta\Phi_{\mathrm{nl}})$, where the nonlinear phase shift is $\Delta\Phi_{\mathrm{nl}} = \Delta nk_0L$. Thus, achieving 100% conversion efficiency in sum- or difference-frequency generation, or in four-wave mixing, just like achieving 100% switching in a nonlinear interferometer or coupler, requires a nonlinear phase shift of the order of at least π (or π/2 for a push–pull arrangement), as shown schematically in Fig. 1a.

Although this fact is known, it is less recognized that this relation is fundamental and follows from the uncertainty principle. Indeed, change of index is tantamount to change of momentum $\Delta p = \Delta n\hbar k_0$, and to measure or register this change, the length L over which the measurement is performed must be such that the product $\Delta pL \geq \hbar/2$, which postulates $\Delta\Phi_{\mathrm{nl}} \geq 1/2$ as shown in Fig. 1b. The equality sign corresponds to Gaussian probability distribution and, for uniform distribution, meaning propagation with constant velocity $\Delta\Phi_{\mathrm{nl}} \sim \pi/2$, appears a reasonable condition. Note that all the presented analysis is of an 'order of magnitude' nature, and the goal is not to provide a 'design guide' but simply to outline the realm of the possible.

Thus, from the practical point of view, the value of nonlinear susceptibility $\chi^{(n)}$ is far less relevant than the maximum nonlinear phase shift attainable for a given intensity of light $\Delta\Phi_{\mathrm{nl}}$, or alternatively the intensity, $I_{\pi/2}$, required to get a 90° nonlinear phase shift. Obviously, the phase shift is limited by the interaction length, which for various nonlinearities is limited by absorption, by fabrication or simply by the desire to limit the size of a device.

This puts into proper perspective the numerous reports of 'giant' nonlinearities in two-dimensional materials[13,14], metal films and various metasurfaces[15]. Taking vanishingly small nonlinear output and dividing by nearly zero (a very small thickness, a few angstroms in the case of 2D materials) may indeed yield values of $\chi^{(n)}$ that seemingly warrant high-impact publications, but the stubborn fact remains that the efficiency is still vanishingly small.

Intuitive understanding

Having settled the issue of boosting nonlinear susceptibility using the magic of division by zero, let us turn attention to the issue of why the nonlinearity is so weak. On the most fundamental level, the answer is simple: photons happen to be bosons that carry no charge and are not subject to the Pauli exclusion principle, so they do not interact with each other directly, but only via intermediary short-lived states of the matter (typically electrons and holes). Or one can say that that photons in the non-vacuum medium are no longer pure photons but polaritons – coupled field–matter excitations (quasiparticles) that do interact with each other, albeit very weakly. The fact that the interactions are so weak is the main reason that photons are just about perfect for unimpeded long-distance transmission of information, but, at the same time, it makes manipulation of information with photons a daunting task.

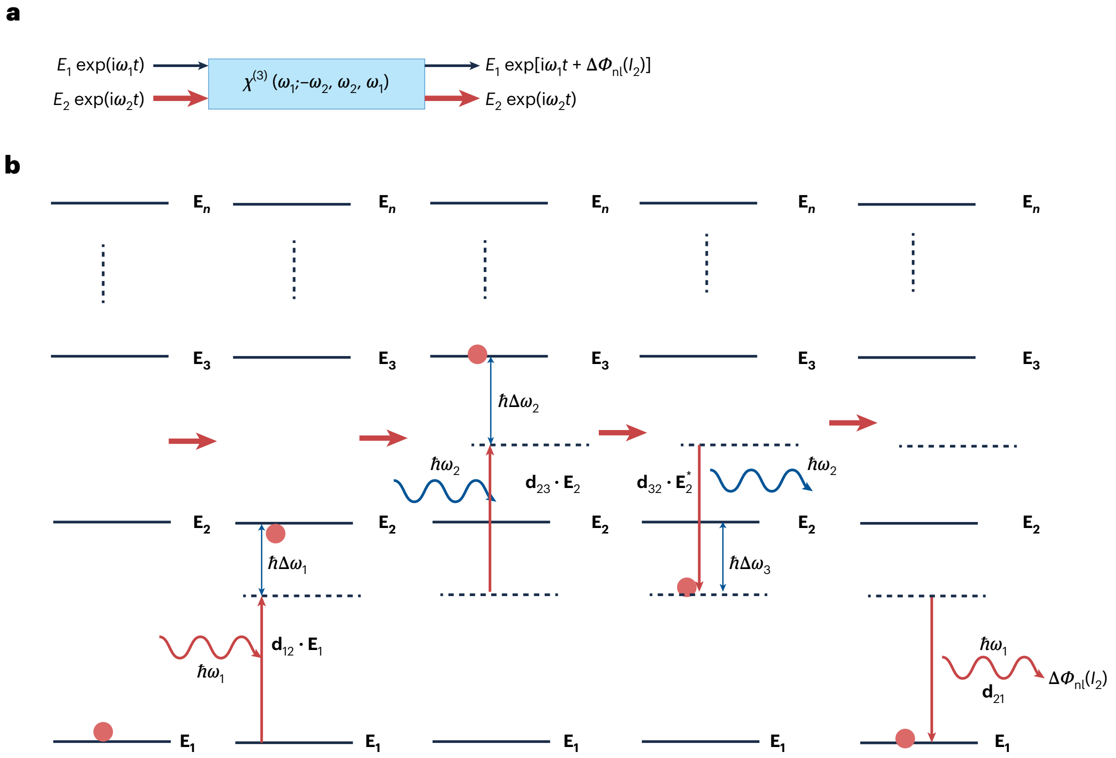

Fig. 3 | Feynman diagram for third-order nonlinearity. **a**, Schematics of a typical third-order nonlinear optical process: generation of nonlinear polarization at frequency ω_1 and nonlinear phase shift $\Delta\Phi_{nl}$ caused by a pump of frequency ω_2. **b**, Feynman diagram of this process.

One way to quantify the weakness of nonlinear interactions is to consider a typical anharmonic potential as shown in Fig. 2a, which may describe the binding potential $U(r)$ of an electron in a chemical bond of characteristic size a (of the order of 1–3 Å for bonds active in the optical range) with binding energy U_0 (a few eV). For small deviations from the equilibrium position, the potential is very well described by a harmonic (parabolic) function, shown as a dashed line. One can loosely define this region as 'linear'. As displacement approaches $a/2$, the electron is no longer strongly bound, and one finds the electron in the 'optical breakdown' region. As one can see, the 'nonlinear' region is squeezed between the linear and breakdown regions, and it is only when one is on the verge of breakdown that nonlinear change of permittivity becomes commensurate with the linear one.

In this most primitive model, polarizability as well as the linear susceptibility $\chi^{(1)} = \varepsilon - 1$ depends on the curvature of the potential, and the curvature changes substantially (factor of ~2) when the electron energy changes by an amount commensurate with U_0. To achieve roughly 100% change in susceptibility, either each valence electron must acquire energy commensurate with U_0, or one could change the curvature by a factor of ~2 by distorting the lattice so that U_0 is changed by that factor. Either way, one can then make a conjecture: if the energy density stored (not necessarily absorbed) in the material is roughly $\Delta u_{st} \sim u_0 = NU_0$, where N is the density of valence electrons, then the electronic part of susceptibility $\chi^{(1)} = \varepsilon - 1$ changes by roughly 100%. It follows that the change of index is proportional to the energy, as in third-order nonlinearity, and it can be roughly estimated as $\Delta n/n \sim \frac{1}{2}\Delta\varepsilon/\varepsilon \sim \Delta u_{st}/u_0$.

The question of what value of energy should be taken as U_0 does not have an unambiguous answer, but it should be noted that in the most simple potential well model, the depth U_0 is commensurate with the split between the lowest and the first excited state. In semiconductors and dielectrics, this split corresponds to the 'mean' splitting of valence and conduction band and is commonly referred to as the Penn gap[16], which is always in the range of a few electron volts for materials that are transparent in the optical/near-infrared range. On the higher end of the estimate, one can consider cohesive energy (per atom)[17] as U_0, and that value is typically closer to 5–10 eV; whereas when it comes to the nonlinearities associated with nonparabolicities in the bands[18], it is roughly the fundamental direct-bandgap energy that should be considered, which is rather lower than the Penn gap. However, these energies (as well as electron affinity and ionization energy) are all on the same 1–10 eV scale. Therefore, the order-of-magnitude assessments performed here are valid no matter what is chosen as U_0. As the density of valence electrons is $N \sim 1-2\times10^{23}\,\mathrm{cm}^{-3}$ and $U_0 \sim 1-10\,\mathrm{eV}$, it seems that the energy density required to change refractive index by 100% is on the scale of $u_0 \sim 10^4 - 3\times10^5\,\mathrm{J\,cm}^{-3}$. To verify the sanity of this analysis, one can consider very diverse mechanisms of changing refractive index using thermal, electro-optic and acousto-optic $\chi^{(2)}$ and $\chi^{(3)}$. Details of analysis with different materials are given in ref. 19, and for all these materials u_0 falls into the range outlined above. Materials transparent in the visible and ultraviolet regions generally have larger binding energy, and thus are characterized by larger u_0 and lower nonlinearity than materials that operate at longer wavelengths.

The meaning of this remarkable result is that the energy density required to change index by a given amount Δn is $\Delta u_{st} \sim (\Delta n/n)u_0$, which is more or less similar in every third-order nonlinear material (for the same wavelength range), but the power density $I = \Delta u_{st}/\tau$ can be changed by orders of magnitude by changing the 'energy storage' time, τ. This time can be the lifetime of an electron on a given energy level or band, thermal diffusion time, photon lifetime in the resonator, or photon propagation time among many others. As we shall see, engineering the storage or interaction time τ and accepting (almost inevitable) bandwidth limitation is precisely what all the conceivable pathways amount to in the end.

Before continuing with impact of interaction time, let us estimate the scale of nonlinear susceptibility the same model of Fig. 2a. If one defines the binding ('atomic') field $E_0 \sim U_0/ea$ (ref. 20), it follows that when an applied optical field approaches E_0, the displacement becomes comparable to a, and nonlinear change of permittivity is on the scale of 100%. Therefore, the scale of nonlinear susceptibility is $\chi^{(n)} \sim (\varepsilon - 1)/E_0^{n-1}$. For a typical solid-state medium, linear susceptibility $\chi^{(1)} = \varepsilon - 1$ is of the order of 2–20, with lower values corresponding to materials transparent all the way into the ultraviolet, and higher ones to materials transparent only in the infrared range. Note that it is assumed that the entire lattice participates in the optical process – if one is looking for the response of a material due to dopants, then nonlinear susceptibility is scaled down by the fraction of 'active' entities.

As the order of magnitude of E_0 is between 10^{10} and 10^{11} V m^{-1} and permittivity ranges between 2 and 20, the typical value of $\chi^{(3)}$is $10^{-18} - 10^{-22}\,\mathrm{m}^2\,\mathrm{V}^{-2}$, corresponding to nonlinear index $n_2 \sim 10^{-13} - 10^{-16}\,\mathrm{cm}^2\,\mathrm{W}^{-1}$. Note that typically materials with wider bandgaps, transparent throughout the visible and near-ultraviolet range, are more strongly bound: that is, the energy U_0 in them is higher and the bond length is shorter, so their nonlinear index is lower than in narrower-bandgap materials that are transparent only in the infrared region. Therefore, when comparing different nonlinear materials, it is important to compare them in the same spectral region. When doing that, one unsurprisingly discovers that for materials with comparable transparency ranges, the off-resonant (that is, not associated with absorption) nonlinear indices are also comparable, with not much room left for engineering 'giant' nonlinearities. For second-order (and other even-order) nonlinearity, the situation is more complex, as lack of inversion symmetry is required, and different materials have vastly different $\chi^{(2)}$. Hence there still exists opportunity to engineer higher $\chi^{(2)}$, but the optimization is still bounded by a hard limit of $\chi^{(2)} \sim (\varepsilon - 1)/E_0 \sim 10^{-11} - 10^{-9}$ m V^{-1}, once again in the absence of resonance.

Overall, the simple picture presented here indicates that no great enhancement of off-resonant nonlinearity is possible unless one can engineer a potential that is more intricate than Fig. 2a. One way is to engineer an optically induced phase transition[21] with a less gradual change of the potential (Fig. 2b). We shall briefly consider this possibility below, but first we return to the role of the interaction time in the nonlinear process.

The Feynman diagram picture

Turning to a quantum model of nonlinearity, we invoke the language of Feynman diagrams[22]. Here nth-order nonlinear process can be represented as a sequence of $n+1$ acts of absorption and emission of photons. For example, a linear or first-order process consists of absorption of a photon and subsequent emission of a photon of the same frequency with a phase delay. A second-order process, such as sum-frequency generation, involves consecutive absorption of two photons with frequencies ω_1 and ω_2, followed by emission of a single photon of sum frequency $\omega_3 = \omega_1 + \omega_2$, and so on. (The order of absorption and emission processes can be permuted, and then summation over all possible permutations or 'pathways' must be performed, but this is not pertinent to the qualitative discussion here.) As an example, consider the cross-phase modulation process, schematically shown in Fig. 2a, where nonlinear polarization at frequency ω_1 is proportional to the intensity of light I_2 at frequency ω_2, which causes nonlinear phase shift $\Delta\Phi(I_2)$. In the Feynman diagram shown in Fig. 2b, photons of frequency ω_1 and ω_2 are first (virtually) absorbed and then emitted, with the net result of intensity at one of the frequencies ($I_2 \sim |E_2|^2$) modulating the phase of the other one. The nonlinear polarization can then be written as

$$P^{(3)}(\omega_1; -\omega_2, \omega_2, \omega_1) \sim N\mathbf{d}_{21} \frac{\mathbf{d}_{32} \cdot \mathbf{E}_2^*}{\hbar\Delta\omega_3} \frac{\mathbf{d}_{23} \cdot \mathbf{E}_2}{\hbar\Delta\omega_2} \frac{\mathbf{d}_{12} \cdot \mathbf{E}_1}{\hbar\Delta\omega_1} \qquad (1)$$

Now, the probability of the mth step, when an electron transitions from energy level E_m to the level E_{m+1}, can be evaluated as $\mathbf{d}_{m,m+1} \cdot \mathbf{E}_m/\hbar\Delta\omega_m$ where $\mathbf{d}_{m,m+1}$ is the matrix element of a dipole, $\mathbf{E}_m$ is electric field of frequency ω_m, and $\hbar\Delta\omega_m = E_{m+1} - E_1 - \sum_{k=1}^{m} \pm\hbar\omega_k$ is the difference between the actual energy acquired by the electron as it resides at an energy level E_{m+1} and the energy gained from absorbed photons (+ sign) or lost to emitted photons (– sign). One can rewrite the numerators in Eq. (1) as $\Omega_m\tau_m$, where $\Omega_m = \mathbf{d}_{m,m+1} \cdot \mathbf{E}_m/\hbar$ is the Rabi frequency, that is, the rate of photon–media interaction, and $\tau_m = 1/\Delta\omega_{m+1}$ as the effective interaction time (often called virtual lifetime) which can be understood as such from the time–energy uncertainty relation. In the off-resonant case, when $\hbar\Delta\omega_m \sim U_0$ one can see that each term $d_{m,m+1}/\hbar\Delta\omega$ in Eq. (1) is just the inverse of the atomic field E_0 as previously defined.

Now, the rate Ω_m depends on just a single material parameter $\mathbf{d}_{m,m+1}$, whose maximum value is essentially limited by the size of the bond a (of the order of 1–3 Å). One can also apply the oscillator sum rule to obtain $d_{m,m+1} \le \hbar\sqrt{1/2m_0U_0}$, which results in essentially the same result of $d \le 2e$Å. And this leaves only one parameter that can be changed within a very wide range, spanning orders of magnitude – the effective interaction time τ at each step. One can think of the importance of interaction time using the analogy of a climber on the wall in a gym, whose odds of successful climb and descent look better if a firm grip is assured at each step. Let us now see how it plays out for a typical example of nonlinear phase modulation, and what trade-offs are involved in the process of enhancing nonlinearity via increase in the interaction time.

Fast and slow nonlinearity

So, it appears that the obvious way to enhance nonlinearity is to increase the interaction time by reducing the detuning $\Delta\omega$, that is, by taking advantage of the intrinsic resonance of matter, as for instance in atomic vapours[23] or rare-earth dopants[24]. But the resonances remain sharp only for as long as the concentration of the active entities stays low. Hence the increase in nonlinearity of each atom or ion is more than cancelled out by the reduction in their number. In a typical solid medium, the resonances are broadened into wide bands, and even the excitonic resonances are broad, so one inevitably gets absorption when attempting to approach a resonance. Absorption is the origin of the 'slow' nonlinearities[18] when the electron makes a 'real' optically induced transition between bands or within the same band. Once the electron is excited, several different processes can take place, such as absorption saturation[25], exciton screening[26], rise of electron temperature with subsequent change of effective mass[27], or simply rise of the lattice temperature[28]. Each of these processes is accompanied by change of the index of refraction and is characterized by its own characteristic time τ, which can be recombination time, intraband scattering time, thermal diffusion time and so on. The range of characteristic times is very wide – from tens of femtoseconds in intraband processes to milliseconds for thermal diffusion times, with recombination rates being anywhere from picoseconds to tens of microseconds. The energy density stored in the medium can be estimated as $\Delta u_{st} \sim \alpha I\tau$, where α is absorption coefficient and I is power density of propagating light (it can be the power of the signal for the self-phase modulation case, or the power of control light I_2 for the cross-phase modulation case of Fig. 3). Then, the order of magnitude of the optically induced index change is $\Delta n_{\text{slow}} \sim \alpha nI\tau/2u_0$, and the maximum phase shift is determined by the propagation length equal to $1/\alpha$, that is, $\Delta\Phi_{\text{slow}} \sim k_0 nI\tau/2NU_0$. Far from resonance, one can estimate linear susceptibility as $\chi^{(1)} = n^2 - 1 \sim Ne^2a^2/\varepsilon_0 U_0$, or $1/N \sim e^2a^2/\varepsilon_0 U_0(n^2-1)$, therefore

$$\Delta\Phi_{\text{slow}} = \frac{1}{2}\frac{n}{n^2-1}\frac{e^2a^2\tau}{\varepsilon_0 U_0^2}k_0 I \qquad (2)$$

Let us now find the phase shift imposed by the 'fast' nonlinearity, far from resonance:

$$\Delta\Phi_{\text{fast}} = \frac{1}{2n}\chi^{(3)}E^2k_0L \sim \frac{n^2-1}{2n^2}\frac{e^2a^2}{U_0^2}\eta_0 Ik_0L, \qquad (3)$$

where $\eta_0 = 377\,\Omega$ is vacuum impedance and L is the propagation length, which in the absence of absorption can be arbitrarily long. The fast nonlinearity occurs on a timescale of $1/\Delta\omega$, far from resonance on the sub-femtosecond scale, and, for all practical purposes, can be considered nearly instant. We can now establish a connection between fast and slow phase shifts:

$$\frac{\Delta\Phi_{\text{fast}}}{\Delta\Phi_{\text{slow}}} = \frac{\left(n^2-1\right)^2}{n^3}\frac{\varepsilon_0\eta_0 L}{\tau} = \frac{\left(n^2-1\right)^2}{n^4}\frac{nL}{c\tau} \sim \frac{\tau_{\text{tr}}}{\tau}, \qquad (4)$$

where $\tau_{\text{tr}} = Ln/c$ is the transit (propagation) time. Therefore, no matter whether the nonlinearity is instant or 'slow', the nonlinear index change can always be written as

$$\Delta\Phi = \frac{1}{2}\frac{n}{n^2-1}\frac{\tau_{\text{eff}}}{\varepsilon_0 E_0^2}k_0 I \qquad (5)$$

where the effective interaction time τ_{eff} is either a propagation time in the case of fast ('instant') nonlinearity or some actual intrinsic relaxation time in the case of slow nonlinearity. The switching intensity that provides the required π/2 shift is then $I_{\pi/2} \sim 2n\varepsilon_0 E_0^2\lambda_0/\tau_{\text{eff}}$.

The key feature here is that τ_{eff} is the only parameter that can be changed by orders of magnitude. One can enhance nonlinear shift (reduce $I_{\pi/2}$), but with inevitable trade-offs that are different for slow and fast nonlinearities. For slow nonlinearity, the actual speed (or bandwidth) of the nonlinear device gets reduced as $1/\tau_{\text{eff}}$, while for instant nonlinearity, increase in transit time only increases the latency, which in many applications is less deleterious than decrease in speed. Another important difference between 'slow' and 'fast' nonlinearities is that the energy of light is dissipated in slow absorption-based nonlinearity but is preserved in the case of fast nonlinearity and can in principle be reused. Overall, one can write the product of switching intensity and delay time (which can be thought of as the inverse of the gain–bandwidth product used in electronics) as

$$I_{\pi/2}\tau_{\text{eff}} \sim n\varepsilon_0 E_0^2\lambda_0 \sim 10^1 - 10^3\,\text{GW ps cm}^{-2} \qquad (6)$$

with lower numbers corresponding to mid-infrared and higher numbers to visible–ultraviolet. Here we need to emphasize that although for slow nonlinearity Eq. (6) can be interpreted as switching energy (per unit area), no such connection can be made for fast nonlinearity in travelling-wave geometry where the switching energy is determined by the bit length $U_{\text{sw}} \sim I_{\pi/2}T_{\text{bit}}$: that is, it is reduced considerably if the bit duration T_{bit} is much shorter than transit time (as is the case in nonlinear fibre optics).

Intrinsic or extrinsic enhancements

When it comes to enhancement of nonlinearity by increase of interaction time, one may distinguish between intrinsic and extrinsic methods. For fast nonlinearity, the intrinsic enhancement consists of using the resonances inside the material. As we have already discussed, approaching resonance and still staying away from absorption requires very narrow resonances, attainable in such media as atomic vapours but not in condensed matter. Even using excitonic resonances[29,30] is not going to enhance the nonlinear phase shift by a lot. This is because excitons typically contain only the states near the centre of the Brillouin zone – only a small fraction of all states – so although the excitonic effects on imaginary part of permittivity (absorption) are indeed significant, the real part of permittivity is not strongly affected. One can take this argument even further and mention that other peculiarities of band structures, such as Dirac and Weyl[31] points, hardly affect the refraction index and real part of nonlinear susceptibility simply because they represent but a tiny fraction of all the states contributing to the n and real $\chi^{(3)}$.

At the same time, in slow nonlinearity engineering, the characteristic relaxation time τ makes perfect sense. When the required switching time is t_{sw}, it makes sense to keep τ equal to a fraction of it, say $\tau \sim 0.2 - 0.5t_{\text{sw}}$, and not any shorter, as shortening will only reduce the nonlinearity. With today's requirements of $t_{\text{sw}} \sim 10$ ps (for 100-Gbps communications) and prospectively going to $t_{\text{sw}} \sim 1$ ps, most of the nonlinearities based on interband transitions are too slow, although one may have defect-rich materials in which recombination times are reduced to within a picosecond range. On the other hand, the intraband relaxation in most materials occurs on the scale of 100s of femtoseconds. For example, hot carriers in transparent conductive oxides (TCOs)[32] have been shown to thermalize in sub-picosecond times. That is part of the reason that TCOs have attracted such interest. Previously, the same relaxation times were observed in III–V semiconductors using spectral-hole-burning measurements[33]. One can also easily engineer the relaxation times for inter-subband transitions in quantum wells[34].

One should note that in terms of their nonlinear indices, 'slow' nonlinearities are always stronger than 'fast' ones, but, as we have emphasized, it is the nonlinear phase shift that matters. Therefore, according to Eq. (5), one can achieve comparable results in a slow medium that is a few micrometres long with relaxation time of 1 ps and in a fast medium a couple of hundreds of micrometres long with 1 ps propagation time. The advantage of the slow medium would, of course, be the smaller size, albeit at the cost of heat dissipation.

A different (and complementary) approach to interaction time enhancement is doing it externally. The most straightforward method is using the 'slow light effect'[35] in a highly dispersive medium, in which the transit time is $\tau_{tr} = Ln_g/c$, where $n_g = c\mathrm{d}k/\mathrm{d}\omega > n$ is a group index. When dispersion is considered, one can attain the same phase shift in a shorter length. A similar effect is achievable with photonic crystals[36] and plasmonic waveguides[37] (plagued by loss). By far the most widely used way of enhancing the interaction time is by using some type of resonant photonic structure. In a simple Fabry–Perot or ring resonator, the effective interaction time is enhanced by a finesse F, which is roughly the number of round trips a photon makes inside the cavity. Alas, this approach has a major shortcoming, as the interaction time increase is accompanied by a reduction in the bandwidth (and not latency).

What is important to mention is that recent years have seen multiple sophisticated resonant schemes designed to achieve 'giant enhanced nonlinearity'. One might mention Fano resonances[38], electromagnetically induced transparency[39], various nanoantennas and metasurfaces[40], photonic bound states in continuum[41], anapoles[42] or topological modes[43], and this list can be continued. No matter which scheme is used, the effective interaction time for nonlinearity is simply $\tau_{\mathrm{eff}} = Q/\omega$. The Q factor in most of the sophisticated schemes does not exceed Q for a simple microresonator[44], and it is mostly determined by the intrinsic absorption. It seems that with $Q \sim 10^3$ one may achieve effective interaction times on the scale of a few picoseconds; any further increase would affect the bandwidth too much. It can be shown that in other sophisticated schemes of enhancing nonlinearity, such as enhancement near exceptional points[45], one can also relate that increase to the interaction time.

It is instructive to emphasize the importance of interaction time using the example of 'cascaded' nonlinearity – that is, using the second-order processes sequentially, rather than a third-order process[46]. The most widely known example is generation of third harmonic by first doubling the fundamental frequency ω and then mixing the second harmonic 2ω with ω to obtain 3ω (ref. 47). This cascaded process, known to be far more efficient than directly generating the third harmonic, is the operating principle of frequency tripling in all commercial laser systems. Cascading can also emulate the entire spectrum of third-order nonlinear processes. As shown in ref. 46, in cascaded nonlinearity the intensity required for a 90° phase shift is $I_{\pi/2}^{(\mathrm{casc})} \sim I_{\pi/2}/\omega\tau_{tr}$: that is, effective nonlinearity is enhanced by the propagation factor $Q = \omega\tau_{tr}$. Obviously, if one uses a resonant cavity, then the enhancement is just the cavity Q. As explained in ref. 46, it is achieved by coupling one of the short-lived 'virtual' excitations in Fig. 3 to the photon at intermediate frequency and thus extending its lifetime.

Conclusions

The importance of the role of what can be broadly defined as an interaction time in the NLO has been emphasized. The strength of photon–matter interaction depends only on the value of the Hamiltonian and the lifetime of the excitation (real or virtual), referred to here as interaction time. The strongest interaction is the dipole transition, and its maximum value is limited by what is typically a bond length, at most a couple of angstroms. That leaves the interaction time as the only flexible parameter capable of enhancing effective nonlinearity. This enhancement can be achieved intrinsically, typically by operating in the absorption region and exciting real carriers with characteristic lifetimes that can vary by orders of magnitude. The inevitable trade-off is reduced switching speed, but even for a 'slow' nonlinearity, speed can be fast enough for some applications. TCO materials capable of operating at less than 1 ps are a prime example of such nonlinearity.

It is far less feasible to intrinsically enhance interaction time for the 'ultra-fast' nonlinearities operating away from absorption (that is, with only virtual carriers being excited), mostly because in practical nonlinear media the resonances are very broad as individual energy levels broaden into the bands.

With both slow and (especially) ultrafast nonlinearities, it is the extrinsic enhancement of interaction time that is bound to yield the best results. The enhancement can be achieved in both propagating (travelling-wave) and localized geometry. The most obvious way to achieve enhancement is simply by using long propagation lengths, such as in fibres, where, despite all the hype surrounding other methods, the best results for all-optical switching, modulation and frequency conversion have been achieved. To reduce the physical length while keeping interaction time long, one may resort to various photonic structures with low group velocity (enabling 'slow light'). Using a travelling-wave geometry allows one to enhance the nonlinearity while keeping bandwidth reasonably wide and only increasing the latency.

Alternatively, various photonic resonant structures may be used for enhancement. Despite their proliferation over the past decade, it is my opinion that a simple microresonator remains the best option for practical enhancement. The degree of the enhancement in localized resonant structures is ultimately limited by the reduction of the bandwidth, hence the ultra-high Q-factors might not be required. In the end, one should always be able to judiciously choose which (slow or fast) nonlinearity to use and how to enhance it, based on the demands of each concrete application, for example whether the emphasis should be placed on speed, size or low loss.

As far as new NLO materials entrants are concerned, there is plenty of room for development of materials with desirable properties of high damage threshold, low insertion loss, low cost, ability to be integrated on a silicon platform and many others. But it is my view that the fundamental laws of physics would not allow enhancements of efficiencies (or reducing power requirements) for a given wavelength and required speed by orders of magnitude, notwithstanding anticipated future claims of giant nonlinearities. Perhaps one way to look at the enhancement of strength rather than interaction time is to exploit phase transitions in which the effective potential (Fig. 1b) becomes anharmonic well below the optical damage threshold, but this transition needs to be electronic rather than ionic in nature for operation in the near-infrared to visible region. Potential candidates might be the optically induced Mott–Andersen localization transition[48] or other collective effects in solids.

To summarize, a view from a less familiar angle has been provided on the state of developments in NLO. Key factors that limit the efficiency of NLO phenomena and prevent NLO from entering new applications, especially in all-optical data processing, are pointed out. I hope that the universal language of 'interaction time engineering' presented here will be used to enhance and customize NLO phenomena. On whether all-optical switching and processing based on NLO will ever materialize, I will withhold my opinion, as unexpected discoveries may occur and new demands may arise in the not-so-distant future.

Jacob B. Khurgin ✉
Johns Hopkins University, Baltimore, MD, USA.
✉e-mail: jakek@jhu.edu

References

1. Terhune, R. W., Maker, P. D. & Savage, C. M. *Phys. Rev. Lett.* **8**, 404–406 (1962).
2. Bloembergen, N. *Nonlinear Optics; A Lecture Note and Reprint Volume* (Benjamin, 1965).
3. Shen, Y. R. *The Principles of Nonlinear Optics* (Wiley, 1984).
4. Nozaki, K. et al. *Nat. Photon.* **4**, 477–483 (2010).
5. Chai, Z. et al. *Adv. Opt. Mater.* **5**, 1600665 (2017).
6. Svelto, O. *Principles of Lasers* (Plenum, 1976).
7. Siegman, A. E. *Lasers* (University Science Books, 1986).
8. Dudley, J. M., Genty, G. & Coen, S. *Rev. Mod. Phys.* **78**, 1135 (2006).
9. Mingaleev, S. & Kivshar, Y. *Opt. Photonics News* **13**, 48–51 (2002).
10. Wetzstein, G. et al. *Nature* **588**, 39–47 (2020).
11. Millar, D. S. et al. *IEEE J. Sel. Top. Quantum Electron.* **16**, 1217–1226 (2010).
12. Levenson, M. *Introduction to Nonlinear Laser Spectroscopy* 2nd edition (Elsevier, 2012).
13. You, J., Bongu, S., Bao, Q. & Panoiu, N. *Nanophotonics* **8**, 63–97 (2019).
14. Hong, X. et al. *Research* https://doi.org/10.34133/2020/9085782 (2020).
15. Lee, J. et al. *Nature* **511**, 65–69 (2014).
16. Ravindra, N., Auluck, S. & Srivastava, V. *Phys. Status Solidi B* **93**, K155–K160 (1979).
17. Harrison, W. A. *Solid State Theory* (Courier, 1980).
18. Khurgin, J. B., Clerici, M. & Kinsey, N. *Laser Photonics Rev* **15**, 2000291 (2021).
19. Khurgin, J. B. Preprint at https://arxiv.org/abs/2207.05569 (2022).
20. Boyd, R. W. *Nonlinear Optics* (Academic, 2020).
21. Rajpurohit, S., Das Pemmaraju, C., Ogitsu, T. & Tan, L. Z. *Phys. Rev. B* **105**, 094307 (2022).
22. Ward, J. *Rev. Mod. Phys.* **37**, 1 (1965).
23. Spillane, S. M. et al. *Phys. Rev. Lett.* **100**, 233602 (2008).
24. Kumari, V., Kumar, V., Malik, B. P., Mehra, R. M. & Mohan, D. *Opt. Commun.* **285**, 2182–2188 (2012).
25. Karabulut, İ. & Baskoutas, S. *J. Appl. Phys.* **103**, 073512 (2008).
26. Schmitt-Rink, S., Chemla, D. & Miller, D. A. *Adv. Phys.* **38**, 89–188 (1989).
27. Secondo, R., Khurgin, J. & Kinsey, N. *Opt. Mater. Express* **10**, 1545–1560 (2020).
28. Minovich, A., Neshev, D. N., Dreischuh, A., Krolikowski, W. & Kivshar, Y. S. *Opt. Lett.* **32**, 1599–1601 (2007).
29. Schmitt-Rink, S., Chemla, D. & Miller, D. A. *Phys. Rev. B* **32**, 6601 (1985).
30. Takagahara, T. *Phys. Rev. B* **36**, 9293 (1987).
31. Khurgin, J. *Appl. Phys. Lett.* **104**, 161116 (2014).
32. Kinsey, N. & Khurgin, J. *Opt. Mater. Express* **9**, 2793–2796 (2019).
33. Shah, J. & Lucent Technologies. *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* 2nd enlarged edition (Springer, 1999).
34. Rabinovich, W., Beadie, G. & Katzer, D. *IEEE J. Quantum Electron.* **34**, 975–981 (1998).
35. Khurgin, J. B. *Adv. Opt. Photonics* **2**, 287–318 (2010).
36. Hamachi, Y., Kubo, S. & Baba, T. *Opt. Lett.* **34**, 1072–1074 (2009).
37. Huang, Y., Min, C., Dastmalchi, P. & Veronis, G. *Opt. Express* **23**, 14922–14936 (2015).
38. Limonov, M. F., Rybin, M. V., Poddubny, A. N. & Kivshar, Y. S. *Nat. Photon.* **11**, 543–554 (2017).
39. Harris, S. E., Field, J. & Imamoğlu, A. *Phys. Rev. Lett.* **64**, 1107 (1990).
40. Sain, B., Meier, C. & Zentgraf, T. *Adv. Photonics* **1**, 024002 (2019).
41. Carletti, L., Koshelev, K., De Angelis, C. & Kivshar, Y. *Phys. Rev. Lett.* **121**, 033903 (2018).
42. Grinblat, G., Li, Y., Nielsen, M. P., Oulton, R. F. & Maier, S. A. *ACS Nano* **11**, 953–960 (2017).
43. Smirnova, D., Leykam, D., Chong, Y. & Kivshar, Y. *Appl. Phys. Rev.* **7**, 021306 (2020).
44. Xuan, Y. et al. *Optica* **3**, 1171–1180 (2016).
45. Miri, M.-A. & Alù, A. *Science* **363**, eaar7709 (2019).
46. Khurgin, J. B., Obeidat, A., Lee, S. & Ding, Y. J. *JOSA B* **14**, 1977–1983 (1997).
47. Kaminskii, A. et al. *Appl. Phys. B* **93**, 865–872 (2008).
48. Li, S., Khurgin, J. B. & Lawandy, N. M. *Opt. Commun.* **115**, 466–470 (1995).

Acknowledgements

The author appreciates support by the DARPA DSO NLM programme, and the encouragement and assistance of his co-workers P. Noir and S. Artois.

Competing interests

The author declares no competing interests.