

Uni-directional optical pulses, temporal propagation, and spatial and temporal dispersion

P. Kinsler*

*Physics Department, Lancaster University, Lancaster LA1 4YB, United Kingdom. and
Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, United Kingdom.*

(Dated: Monday 4th December, 2017)

I derive a temporally propagated uni-directional optical pulse equation valid in the few cycle limit. Temporal propagation is advantageous because it naturally preserves causality, unlike the competing spatially propagated models. The exact coupled bi-directional equations that this approach generates can be efficiently approximated down to a uni-directional form in cases where an optical pulse changes little over one optical cycle. They also permit a direct term-to-term comparison of the exact bi-directional theory with its corresponding approximate uni-directional theory. Notably, temporal propagation handles dispersion in a different way, and this difference serves to highlight existing approximations inherent in spatially propagated treatments of dispersion. Accordingly, I emphasise the need for future work in clarifying the limitations of the dispersion conversion required by these types of approaches; since the only alternative in the few cycle limit may be to resort to the much more computationally intensive full Maxwell equation solvers.

I. INTRODUCTION

The importance of having simple and robust methods for the propagation of optical pulses has attracted increasing attention in recent years. This is due to the multitude of applications [1] in which ever shorter pulses are used to act like a strobe-lamp that takes snapshots of ultrafast processes [2, 3], or sub-wavelength electric field profiles are created [4–6] to achieve detailed control over atomic or molecular responses. Alternatively, strong nonlinearity can be used to construct pulses that are both wide-band and temporally extended, such as (white light) supercontinua [7–9]. Such nonlinearities can even be used to generate sub-structure that is instead temporally confined, as in optical rogue waves [10]; or even the temporally and spatially localized filamentation processes [11–13].

This progress towards achieving ever shorter pulse durations, with their associated larger spectral bandwidths, and higher pulse intensities, has been pushing traditional pulse propagation models to their limits, or breaking them. If we want to avoid the computational expense of always resorting to high resolution Maxwell's equations solvers coupled to detailed material response models, we need to be confident that our simpler, less demanding approaches still work. In particular we need to have a clear idea of the physics that may have been removed, and what the side-effects of those approximations are. In this paper I use a directional approach, whose relatively simple and straightforward derivation allows easy comparison of both the approximate and the exact propagation equations, whilst still resulting in the analytical and numerical convenience of a first-order wave equation.

However, unlike perhaps the most common approach in nonlinear optics, I consider propagation in time rather than along some chosen spatial axis [14, 15]. The most significant distinctions between temporal and spatial propagation approaches are summarized on figs. 1 and 2 respectively. No-

tice that the initial conditions (starting states) are completely different, and that only the temporal propagation model is going to provide causal solutions in a straightforward way. The temporally propagated evolution equations for the spatial wave profiles do not give us access to the frequency spectrum [16], but instead we have the wavevector spectrum, as well as a causally appropriate time evolution [17]. This requires an alternate approach to the temporal response of the propagation medium, which highlights the existing and often somewhat poorly characterised approximations inherent in spatially propagated treatments of dispersion. Of course, full finite element and/or FDTD [18, 19] pulse propagation methods can also be used, but here my aim is to simplify the time propagated approach in a directional approach in line with common spatially propagated methods.

In section II I derive a customized second order wave equation from Maxwell's equations, and reorganize it to define the material properties appropriately and set up the factorization stage. In section III, I use a factorization method [14] that allows us to construct an explicitly bi-directional model, and which is then approximated to the popular uni-directional limit in section IV. Section IV D remarks on commonly used modifications that can be applied to the equations given in sections III and IV, typically in order to clarify their properties, simplify them, or compare them to existing models. Having derived this main result – the propagation equations and their various specializations – I turn in section V to the consequences of temporal propagation and the perspective it provides on the handling of mixed spatial and temporal dispersive effects. This is then followed in section VI by a specific comparison between the ordinary spatially propagated nonlinear Schrödinger (NLS) equation – perhaps the most widely investigated equation in nonlinear optics – and its temporally propagated counterpart. Finally, in section VII, I present my conclusions.

II. SECOND ORDER WAVE EQUATION

My starting point is the standard macroscopic Maxwell's equations, where I aim for a solution based on a uniform and

*<https://orcid.org/0000-0001-5744-8146>;
Dr.Paul.Kinsler@physics.org

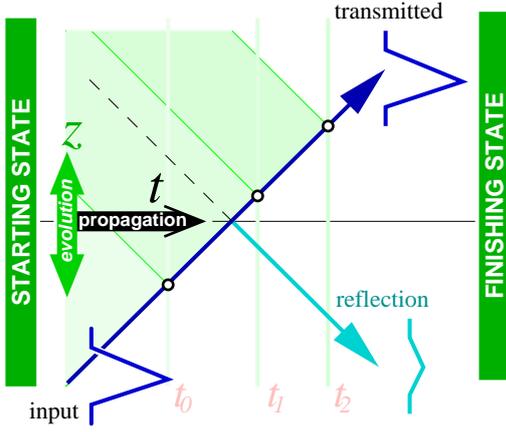


FIG. 1: Temporal propagation of waves, where disturbances (or pulses) evolve either forward or backward in space. At any point in the propagation, we know the spatial behaviour of our wave field at all points in space, as indicated by the pale vertical lines. The pale shaded triangles indicate the past light cones (i.e. the causal past) of a wave element (black circles) at selected times along the path of the disturbance. In a temporally propagated numerical simulation which has a maximum wave speed, the causal past matches the computational past. A notional interface has been added to the diagram to show how a reflection would behave. Figure used with permission from [16].

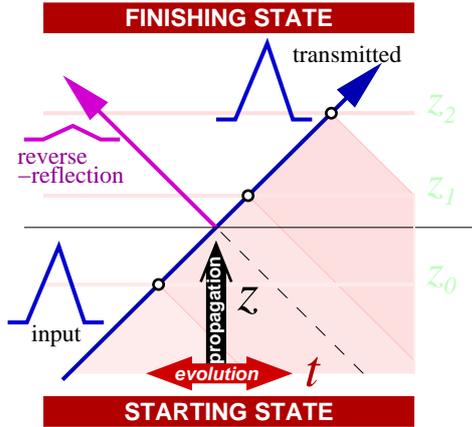


FIG. 2: Spatial propagation of waves, where disturbances (or pulses) evolve either forward or backward in time. At any point in the propagation, we know the full time behaviour of our wave field – both history and future, as indicated by the pale horizontal lines. The light shaded triangles indicates the computational past of a wave element at particular points (black circles) along the path of the disturbance. Note that unlike the temporally propagated case shown in fig. 1, the computational past of a spatially propagated system is not the same as the causal past. A notional interface has been added to the diagram to show how reflections behave – i.e. in an unexpected way [20]. This is because a reflection should have been put in the initial conditions, but was not due to an (assumed) lack of knowledge of that future behaviour. Figure used with permission from [16].

source free dielectric medium, but still intending to allow material properties that are as general as possible. A typical approach to this is to construct a second order wave equation for the electric field \mathbf{E} , as results from the substitution of the $\nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{J}$ Maxwell's equation into $\nabla \times \mathbf{E} = -\partial_t \mathbf{B}$ (see e.g. [21]). Here, however, I want to follow the displacement field \mathbf{D} (not \mathbf{E}) since it naturally occurs in conjunction with a time derivative – likewise, the other important field is \mathbf{B} rather than \mathbf{H} . Thus we rearrange the curl Maxwell's equations to emphasize both their causal properties [17] and their time derivatives, as

$$\mu_0 \partial_t \mathbf{D} = \mu_0 \nabla \times \mathbf{H} - \mu_0 \mathbf{J} = \nabla \times \mathbf{B} - \mu_0 \nabla \times \mathbf{M} - \mu_0 \mathbf{J}, \quad (2.1)$$

$$\epsilon_0 \partial_t \mathbf{B} = -\epsilon_0 \nabla \times \mathbf{E} - \epsilon_0 \mathbf{K} = -\nabla \times \mathbf{D} + \nabla \times \mathbf{P} - \epsilon_0 \mathbf{K}. \quad (2.2)$$

Note that for both of these equations, we will also (as usual) want to specify the material response, whether dielectric or magnetic, in terms of a dielectric polarization $\mathbf{P} = \mathbf{D} - \epsilon_0 \mathbf{E}$ and a magnetization $\mathbf{M} = \mathbf{B} - \mu_0 \mathbf{H}$. Either of these equations given above might be either in the spatial domain (with argument $\mathbf{r} = (r_x, r_y, r_z)$), or the wavevector domain (with argument $\mathbf{k} = (k_x, k_y, k_z)$).

Now we define the total magnetization \mathbf{M} in terms of a component \mathbf{M}_L linearly dependent on \mathbf{B} , and another component \mathbf{M}_B containing the rest. As a result, the excess polarization $\mathbf{M}_\mu \equiv \mathbf{M}_B$, which will typically include any nonlinearity, is also a function of \mathbf{B} and not \mathbf{H} . In traditional \mathbf{H} -based spatially-propagated approaches, the temporal response (“dispersion”) can be incorporated naturally, here it cannot. We have that

$$\mathbf{M}(\mathbf{r}) = \mathbf{M}_L(\mathbf{r}) + \mathbf{M}_B(\mathbf{r}) = \alpha_\mu(\mathbf{r}) \mathbf{B}(\mathbf{r}) + \mathbf{M}_B(\mathbf{r}), \quad (2.3)$$

which in the spatial frequency (wavevector \mathbf{k}) domain is

$$\mathbf{M}(\mathbf{k}) = \mathbf{M}_L(\mathbf{k}) + \mathbf{M}_B(\mathbf{k}) = \alpha_\mu(\mathbf{k}) \star \mathbf{B}(\mathbf{k}) + \mathbf{M}_B(\mathbf{k}), \quad (2.4)$$

since after a Fourier transform from \mathbf{r} into \mathbf{k} , the products become spatial convolutions (“ \star ”); which emphasises that anything wavevector dependent is necessarily a non-local concept. Note that this is the converse of the usual situation where a temporal response is described using time-domain convolutions, and these become products in the frequency domain.

Now we define the total polarization \mathbf{P} in terms of a component \mathbf{P}_L linearly dependent on \mathbf{D} , and another component \mathbf{P}_D containing the rest. As a result, the excess polarization $\mathbf{P}_\epsilon \equiv \mathbf{P}_D$, which will typically include any nonlinearity, is also a function of \mathbf{D} . We have that

$$\mathbf{P}(\mathbf{r}) = \mathbf{P}_L(\mathbf{r}) + \mathbf{P}_D(\mathbf{r}) = \alpha_\epsilon(\mathbf{r}) \mathbf{D}(\mathbf{r}) + \mathbf{P}_D(\mathbf{r}), \quad (2.5)$$

which in the spatial frequency (wavevector \mathbf{k}) domain is

$$\mathbf{P}(\mathbf{k}) = \alpha_\epsilon(\mathbf{k}) \star \mathbf{D}(\mathbf{k}) + \mathbf{P}_D(\mathbf{k}). \quad (2.6)$$

With these definitions for $\partial_t \mathbf{D}$, $\partial_t \mathbf{B}$, and for \mathbf{M} and \mathbf{P} set up, we can now proceed with the derivation of a second order

wave equation for \mathbf{D} –

$$\mu_0 \partial_t \mathbf{D} = \nabla \times \mathbf{B} - \mu_0 \nabla \times \mathbf{M}_L - \mu_0 \nabla \times \mathbf{M}_B - \mu_0 \mathbf{J} \quad (2.7)$$

$$= \nabla \times \mathbf{B} - \nabla \times \alpha_\mu \mathbf{B} - \mu_0 \nabla \times \mathbf{M}_B - \mu_0 \mathbf{J} \quad (2.8)$$

$$= \nabla \times \mathbf{B} - \alpha_\mu \nabla \times \mathbf{B} - (\nabla \alpha_\mu) \times \mathbf{B} - \mu_0 \nabla \times \mathbf{M}_B - \mu_0 \mathbf{J} \quad (2.9)$$

$$= \nabla \times [1 - \alpha_\mu] \mathbf{B} - \mu_0 \nabla \times \mathbf{M}_B - \mu_0 \mathbf{J}. \quad (2.10)$$

Now take the time derivative of this equation, use the fact that α_μ is independent of time, define $a_\mu = 1 - \alpha_\mu$ and $a_\varepsilon = 1 - \alpha_\varepsilon$; and then substitute for $\partial_t \mathbf{B}$, so that

$$\varepsilon_0 \mu_0 \partial_t^2 \mathbf{D} = \varepsilon_0 \nabla \times a_\mu \partial_t \mathbf{B} - \varepsilon_0 \mu_0 \nabla \times \partial_t \mathbf{M}_B - \varepsilon_0 \mu_0 \partial_t \mathbf{J} \quad (2.11)$$

$$c^{-2} \partial_t^2 \mathbf{D} = -\nabla \times a_\mu [\nabla \times \mathbf{D} - \nabla \times \mathbf{P} + \varepsilon_0 \mathbf{K}] - c^{-2} \nabla \times \partial_t \mathbf{M}_B - c^{-2} \partial_t \mathbf{J} \quad (2.12)$$

$$\partial_t^2 \mathbf{D} = -c^2 \nabla \times a_\mu [\nabla \times \mathbf{D} - \nabla \times \alpha_\varepsilon \mathbf{D} - \nabla \times \mathbf{P}_D + \varepsilon_0 \mathbf{K}] - \nabla \times \partial_t \mathbf{M}_B - \partial_t \mathbf{J} \quad (2.13)$$

$$= -c^2 \nabla \times a_\mu [\nabla \times (1 - \alpha_\varepsilon) \mathbf{D} - \nabla \times \mathbf{P}_D + \varepsilon_0 \mathbf{K}] - \nabla \times \partial_t \mathbf{M}_B - \partial_t \mathbf{J} \quad (2.14)$$

$$= -c^2 \nabla \times a_\mu \nabla \times a_\varepsilon \mathbf{D} + c^2 \nabla \times a_\mu \nabla \times \mathbf{P}_D - c^2 \nabla \times a_\mu \varepsilon_0 \mathbf{K} - \nabla \times \partial_t \mathbf{M}_B - \partial_t \mathbf{J} \quad (2.15)$$

Remember that in this picture, all field properties are specified as functions of our chosen primary field \mathbf{D} ; notably, the magnetic field \mathbf{B} which is usually the main argument for \mathbf{M}_B must be determined from the calculated \mathbf{D} using the relevant Maxwell's equation (i.e. eqn. (2.2)).

A. Simple case: isotropic & homogeneous

Since trying to treat all the details of eqn. (2.15) correctly leads to excessively complicated expressions, I will first simplify it down into the case where (a) the material's reference properties a_μ , a_ε do not vary in space, and (b) the effective monopole current is zero ($\mathbf{K} = 0$). Since (magnetic) monopoles do not exist, and all non reference properties can still be encoded in \mathbf{P}_D , \mathbf{M}_B , or \mathbf{J} , this simplification does not (of itself) impose any approximation. However, the first simplification means that our reference properties cannot incorporate any exact knowledge about the spatial structure of the propagation medium¹. This restriction on obtaining the best possible match between the reference behaviour and the exact behaviour of the propagation medium will typically affect later approximations, where we assume deviations from the

¹ This differs from the situation present in the complementary spatially propagated approach [14], where the medium's linear temporal response typically *can* be subsumed into the reference behaviour. Nevertheless, in neither case can the *spatial* properties be incorporated exactly into the reference behaviour; although approximations that assume a particular k dependence can be made.

reference behaviour are small. One additional advantage of selecting constant reference parameters α_μ and α_ε is that we can easily replace them with matrices to allow for e.g. birefringence and cross-polarization couplings.

The simplified second order wave equation for changes in the displacement field $\mathbf{D}(\mathbf{r})$ as it propagates forward in time is

$$\partial_t^2 \mathbf{D} = -c^2 a_\mu a_\varepsilon \nabla \times \nabla \times \mathbf{D} + c^2 a_\mu \nabla \times \nabla \times \mathbf{P}_D - \nabla \times \partial_t \mathbf{M}_B - \partial_t \mathbf{J} \quad (2.16)$$

$$= +c^2 a_\mu a_\varepsilon \nabla^2 \mathbf{D} - c^2 a_\mu a_\varepsilon \nabla \nabla \cdot \mathbf{D} - c^2 a_\mu \nabla^2 \mathbf{P}_D + c^2 a_\mu \nabla \nabla \cdot \mathbf{P}_D - \nabla \times \partial_t \mathbf{M}_B - \partial_t \mathbf{J}. \quad (2.17)$$

Note that the retention of the $\nabla \cdot \mathbf{D}$ and $\nabla \cdot \mathbf{P}_D$ terms allows us to include charge effects, as is needed when discussing high-power nonlinear situations such as filamentation [22].

In the spatial Fourier “wavevector” regime, we replace $\nabla \rightarrow i\mathbf{k}$, so that $k^2 = \mathbf{k} \cdot \mathbf{k}$, and reorganize slightly. The equation for changes in the displacement field $\mathbf{D}(\mathbf{k})$ as it propagates forward in time is

$$\partial_t^2 \mathbf{D} - c^2 a_\mu a_\varepsilon k^2 \mathbf{D} = -c^2 a_\mu k^2 \mathbf{P}_D - \partial_t \mathbf{J} - i\mathbf{k} \times \partial_t \mathbf{M}_B - c^2 a_\mu a_\varepsilon \mathbf{k} \mathbf{k} \cdot \mathbf{D} + c^2 a_\mu \mathbf{k} \mathbf{k} \cdot \mathbf{P}_D \quad (2.18)$$

$$\partial_t^2 \mathbf{D} - \Omega^2 \mathbf{D} = +\Omega^2 a_\varepsilon^{-1} \mathbf{P}_D - \partial_t \mathbf{J} - i\mathbf{k} \times \partial_t \mathbf{M}_B - \Omega^2 k^{-2} \mathbf{k} \mathbf{k} \cdot \mathbf{D} + \Omega^2 k^{-2} a_\varepsilon^{-1} \mathbf{k} \mathbf{k} \cdot \mathbf{P}_D, \quad (2.19)$$

where we have defined a reference frequency Ω for the propagation, based on the reference material properties, i.e.

$$\Omega^2(\mathbf{k}) = c^2 a_\mu a_\varepsilon k^2. \quad (2.20)$$

We could – if we believed we understood the spatial properties of the propagation medium sufficiently well – simply alter this definition by incorporating an alternate \mathbf{k} dependence that mimics an assumed spatial dispersion. However, it needs to be understood that such a step is distinctly *ad hoc*, since it lacks the convolutions over \mathbf{k} necessary for any accurate representation of the material structure. But, that said, such an approach may still have considerable practical utility; and indeed is functionally equivalent to how spatial dispersion can be handled as a non-local (and hence, strictly speaking, a non-causal) process: e.g. those that import a waveguide dispersion calculated from the (spatial) modes given by the guide's transverse cross-section. To allow for this possibility, in the following I will replace the physically rigorous magnitude k argument for Ω with \mathbf{k} , its vector counterpart². This allows

² In some cases, depending on the approximate model chosen, this new $\Omega(\mathbf{k})$ definition may also need to be convolved with other parts of the expression, e.g. $\Omega^2(k) \mathbf{P}_D(\mathbf{k}) \rightarrow \Omega^2(\mathbf{k}) \star \mathbf{P}_D(\mathbf{k})$, although such an additional complication is likely to prevent us obtaining much conceptual or mathematical benefit. Consequently, in what follows I do not allow for such convolutions.

for not only more complicated assumed spatial dispersions, but also additional orientation dependence. However, these benefits (only) arise as a result of approximation. Strictly speaking, both on the grounds covered above, and ones taking stricter view of causality [16], many simplistic notions of spatial dispersion involving adding some convenient k dependence is incompatible with locality³.

Unlike for the spatially propagated scheme in [14], we have not included temporal response of the propagation medium in our definition of the directional fields, since it is incompatible with our temporal propagation. To include this we must model it explicitly as is done in Maxwell solvers such as FDTD [18, 19]; or approximate it as if it were a spatial effect (see e.g. [24], or the discussions in the later section V.

III. FACTORIZATION

I now factorize [14, 25] the second order wave equation for \mathbf{D} . This process neatly avoids some of the approximations necessary in traditional approaches, and takes its name from the fact that the LHS of eqn. (2.17) or (2.19) is a simple difference of squares which might be factorized. Indeed, this was done by Blow and Wood in 1989 [27], albeit in a rather ad hoc – but nevertheless effective – fashion. Since the factors are just of the form $(\partial_t \mp ik)$, taken individually they each look like a simple forward (or backward) directed first order wave equation. The factorization method therefore allows us to define a pair of counter-propagating Greens functions, which divides the original second order wave equation into a pair of oppositely directed first order wave equations that are coupled together. In addition, it allows us to straightforwardly compare the exact bi-directional and approximate uni-directional theories term by term, which is in distinct contrast to other approaches, where the backward or unwanted parts are approximated away piecemeal, and may not be suitable for direct comparison. A short summary of the factorization method, adapted to this new context from earlier work [14], is given in appendix A.

An important point to remember is that the choice of a_ϵ, a_μ and therefore $\Omega(\mathbf{k})$ in eqn. (2.19) defines the specific Greens functions used. This means that it also defines the basis upon which we will then propagate the displacement field \mathbf{D} .

In this section and the next, the derivation will closely follow the procedure, language, and terminology used in earlier work [14]. However, despite the mathematical similarities, it is important not to lose sight of the distinct physical differences. That previous work focussed on propagation of optical fields along a spatial axis, whereas here I instead consider propagation in time. Since my aim is to compare and contrast the two approaches, as discussed at length in sections V and

VI, it is useful to enable comparisons not only at the endpoint of the derivation, but also at each step along the way.

A. Bi-directional wave equations

A pair of bi-directional wave equations suggests similarly bi-directional fields, so I split the electric displacement field into forward (\mathbf{D}^+) and backward (\mathbf{D}^-) directed parts, with $\mathbf{D} = \mathbf{D}^+ + \mathbf{D}^-$. In the following, remember that $\mathbf{P}_D, \mathbf{M}_B$, etc all depend on the full field \mathbf{D} , and *not* only one partial field (e.g. only \mathbf{D}^+). This is an important point since we see that they then drive both the forward and backward evolution equations equally.

The coupled bi-directional first order wave equations for the directed fields \mathbf{D}^\pm are *propagated* forward in time whilst being *evolved* forward and backward in space are

$$\begin{aligned} \partial_t \mathbf{D}^\pm(\mathbf{k}) &= \pm i\Omega(\mathbf{k})\mathbf{D}^\pm(\mathbf{k}) \pm \frac{i\Omega^2(\mathbf{k})a_\epsilon^{-1}}{2\Omega(\mathbf{k})}\mathbf{P}_D(\mathbf{D}, t; \mathbf{k}) \\ &\mp \frac{\mathbf{J}(t; \mathbf{k})}{2\Omega(\mathbf{k})} \mp \frac{i}{2\Omega(\mathbf{k})}\mathbf{k} \times \mathbf{M}_B(\mathbf{D}, t; \mathbf{k}) \\ &\mp \frac{i\Omega^2(\mathbf{k})}{2\Omega(\mathbf{k})k^2}\mathbf{k}\mathbf{k} \cdot \mathbf{D}(\mathbf{k}) \pm \frac{i\Omega^2(\mathbf{k})a_\epsilon^{-1}}{2\Omega(\mathbf{k})k^2}\mathbf{k}\mathbf{k} \cdot \mathbf{P}_D(\mathbf{D}, t; \mathbf{k}) \end{aligned} \quad (3.1)$$

$$\begin{aligned} \partial_t \mathbf{D}^\pm(\mathbf{k}) &= \pm i\Omega(\mathbf{k})\mathbf{D}^\pm(\mathbf{k}) \pm \frac{i\Omega(\mathbf{k})}{2a_\epsilon}\mathbf{P}_D(\mathbf{D}, t; \mathbf{k}) \\ &\mp \frac{\mathbf{J}(t; \mathbf{k})}{2\Omega(\mathbf{k})} \mp \frac{i}{2\Omega(\mathbf{k})}\mathbf{k} \times \mathbf{M}_B(\mathbf{D}, t; \mathbf{k}) \\ &\mp \frac{i\Omega(\mathbf{k})}{2} \frac{\mathbf{k}\mathbf{k}}{k^2} \cdot \mathbf{D}(\mathbf{k}) \pm \frac{i\Omega(\mathbf{k})}{2a_\epsilon} \frac{\mathbf{k}\mathbf{k}}{k^2} \cdot \mathbf{P}_D(\mathbf{D}, t; \mathbf{k}) \end{aligned} \quad (3.2)$$

Here I have replaced the partial time derivatives on the RHS with over-dots, this being to emphasise that if our expression is to remain strictly causal [17], then our models for current \mathbf{J} and magnetization \mathbf{M}_B should return their *time-derivatives* as explicit functions of known quantities (here, typically \mathbf{D} or t). For example, a current model specification such as $\mathbf{J} = \sigma \mathbf{D}$ would mean that $\dot{\mathbf{J}} = \partial_t \mathbf{D}$, giving a $\partial_t \mathbf{D}$ on both sides.

Note that factorized equations can be rebuilt into a single second-order equation by taking the sum and difference, then substituting one into the other with the assistance of a time derivative (see section IV.B of [25]).

A remaining clarification required is to understand the meaning of “forward” and “backward” in the context of three dimensional space – this being less straightforward than in the spatially propagated case, where one spatial axis is selected for propagation⁴, and forward and backward refers to time t . Here, we should understand “forward” and “backward” as being with respect to a given choice of \mathbf{k} .

³ Of course, it is possible to treat spatial dispersion in a properly causal way; a notable example being the hydrodynamic plasmon model [23]. This takes propagation in the electron fluid into account, giving rise to an additional k^2 dependence.

⁴ Typically a cartesian axis, although other choices are possible [15].

B. Propagation, evolution, and directed fields

Since we have chosen to enforce *propagation* towards later times t on our solutions of the wave equations, the fields $\mathbf{D}^\pm(t)$ are *directed* forwards and backwards in space; these fields then *evolve* forwards and/or backwards in space as t increases. I use this terminology (propagated, directed, evolved) throughout this paper to mean these three specific things. This usage is consistent with that used in the alternative spatially propagated approach [14], but in that case the fields $\mathbf{E}^\pm(z)$ are directed in time, and evolve forwards and/or backwards in time as z increases.

The wave equation eqn. (3.2) which evolves the directed fields \mathbf{D}^\pm as they propagate forward in t , has two types of terms on the RHS: and I call these the “reference” and “residual” parts [14, 28].

Reference evolution (or “*underlying evolution*”) is that given by $\pm i\Omega(\mathbf{k})\mathbf{D}^\pm$ term, and is determined by our chosen a_ϵ and a_μ , and any additional ad hoc refinements. By itself, it would describe an ordinary oscillatory evolution where the field oscillations in $\mathbf{D}^\pm(\mathbf{r})$ would move forward (+) or backwards (−) in space. This is analogous to the choice of reference when constructing directional fields [29], but here is done for temporal rather than spatial propagation [20].

Residual evolution accounts for the discrepancy between the true evolution and the underlying/reference evolution. It consists of all parts of the material response *not* included in a_ϵ, a_μ (and hence $\Omega(\mathbf{k})$); i.e. it contains all of the rest of the terms on the RHS of eqn. (3.2). These will usually consist of any non-linear polarization, or spatially or orientationally dependent linear terms; they are analogous to the correction terms used in methods based on directional fields. Alternatively, such contributions are called “source” terms by Ferrando et al. [25]. Generally we will hope they only provide a weak perturbation, so that we might make the (desirable) uni-directional approximation discussed later; but the factorization procedure itself is valid *regardless* of their strength.

C. Reference evolution: choice of Ω and the resulting D^\pm

I now consider how our choice of Ω will affect the relative sizes of the forward and backward directed \mathbf{D}^+ and \mathbf{D}^- . I therefore now consider the example of a simple medium in which the field is known to propagate with frequency ω ; but instead we choose a reference evolution determined by a frequency Ω that is different from ω . E.g., for a linear isotropic medium we could exactly define $\omega^2 = \Omega^2 + \Delta^2$; but in general we would just have some residual (source) term Ω . As a consequence, the definitions of forward and backward *directed* fields do not correspond exactly to what the wave equation actually will *evolve* forward and backward, as we *propagate* towards later times.

The second order wave equation is $(\partial_t^2 + \Omega^2)\mathbf{D} = -\Omega$, which in the linear case has $\Omega = \Delta^2\mathbf{D}$, so that $(\partial_t^2 + \omega^2)\mathbf{D} = 0$.

The factorization in terms of Ω is then

$$\partial_t \mathbf{D}^\pm = \pm i\Omega \mathbf{D}^\pm \pm \frac{i\Omega}{2\Omega}. \quad (3.3)$$

Now if we choose the situation where our field \mathbf{D} only evolves forward, we know that $\mathbf{D} = \mathbf{D}_0 \exp[i\omega t]$. Consequently \mathbf{D}^\pm must have matching oscillations: i.e. $\mathbf{D}^\pm = \mathbf{D}_0^\pm \exp[i\omega t]$, even though \mathbf{D}^- is directed backwards. Substituting these into eqn. (3.3) gives

$$\mathbf{D}_0^- = \frac{\Omega - \omega}{\Omega + \omega} \mathbf{D}_0^+. \quad (3.4)$$

This tells us how much \mathbf{D}^- we need to combine with \mathbf{D}^+ so that our pulse evolves forward; the \mathbf{D}^- is strongly coupled to \mathbf{D}^+ , and so will be dragged forward against its usual preference. This interdependence of \mathbf{D}^\pm is generic – no matter what the origin of the discrepancy between Ω and the true evolution of the field (i.e. the residual/source terms such as mismatched linear behaviour, nonlinearity, etc): some non-zero backward directed field \mathbf{D}^- must exist but still evolve forwards with \mathbf{D}^+ . Comparable behaviour is also seen in the directional fields approach of Kinsler et al. [29].

Typically we will hope that this residual \mathbf{D}^- contribution is small enough so as to be negligible. If we assume $\mathbf{D}^- \simeq 0$, then we find that $\omega \simeq \Omega + \Delta^2/2\Omega$, which is just the expansion of $\omega = (\Omega^2 + \Delta^2)^{1/2}$ to first order in Δ^2/Ω^2 . Following this, we find that eqn. (3.4) then says that $\mathbf{D}_0^- \simeq (\Delta^2/4\Omega^2)\mathbf{D}_0^+$, which now provides us with the scale on which \mathbf{D}^- can be ignored. Apart from the simple (linear) case where we know Δ^2 , the true frequency ω might be difficult to determine, and in nonlinear propagation any local estimate of the frequency can change during propagation.

Further, if we choose $\Omega = \Omega(k)$ with a wave-vector dependence, then we see that the source-like terms inherit that dispersion; and this is almost inevitable, since typically $\Omega \sim ck$. This means, for example, that even if we start with model of nonlinear polarization without any k dependence, our factorized equations will have a k dependence on the nonlinear terms. This is the complementary process to what happens in the spatially propagated approaches of Kinsler et al. [14, 29], where choosing a dispersive reference has consequences for the behaviour of the correction terms.

IV. UNI-DIRECTIONAL WAVE EQUATIONS

We can now make a single, well defined class of approximation, and simplify the exact coupled bi-directional evolution of \mathbf{D}^\pm down to just one uni-directional first order wave equation. As with the complementary spatially propagated theory [14], this does not require a moving frame, a smooth envelope, or to assume inconvenient second order derivatives are somehow negligible, as was necessary in traditional treatments. The approximation assumes that the residual terms are weak in comparison to the (reference) $\pm i\Omega\mathbf{D}$ term – e.g. weak nonlinearity or orientation dependence. This means that I can assert that if we start with $\mathbf{D}^- = 0$, then \mathbf{D}^- will stay negligible. In this context, “weak” means that no significant change

in the backward field is generated in a time shorter than one period (“slow evolution”); and that small effects do not build up gradually over propagation times of many periods (“no accumulation”).

Slow evolution is where the size of the residual terms is much smaller than that of the underlying linear evolution – i.e. smaller than $\Omega\mathbf{D}$. This allows us to write down straightforward inequalities which need to be satisfied. It is important to note the close relationship between these and a good choice of Ω , as discussed in subsection III C. If Ω is not a good enough match, there always be significant contributions from both forward and backward directed fields; and even if nothing ends up *evolving* backwards, an ignored backward directed field will result in (e.g.) miscalculated nonlinear effects, since the total field $\mathbf{D} = \mathbf{D}^+ + \mathbf{D}^-$ will be wrong.

No accumulation occurs when the evolution of any backward directed field \mathbf{D}^- is dominated by its coupling via the residual terms to the forward directed field \mathbf{D}^+ ; and not by its preferred underlying backward evolution. No accumulation means that forward evolving field components do not couple to field components that evolve backwards; this the typical behaviour since the phase mismatch between forward evolving and backward evolving components is $\sim 2\Omega$; in essence it is comparable to the common rotating wave approximation (RWA). This rapid relative oscillation means that backward evolving components never accumulate, as each new addition will be out of phase with the previous one; it is not quite a “no reflection” approximation, but one that asserts that the many micro-reflections do not combine to produce something significant. An estimate of the terms required to break this approximation are given in appendix B; generally speaking this is a much more robust approximation than the slow evolution one. Of course, a periodic temporal modulation of the medium will give periodic residual terms, and these can be engineered to force phase matching. In spatially propagated analyses, where we talk of spatial (wavevector) phase matching and not this temporal (frequency) phase matching [30], this would be a periodicity based on a relatively small phase mismatch (see e.g. quasi phase matching in Boyd [31]); but might even go as far as matching the backward wave (see e.g. [32]).

Note however that small forward perturbations from the residual terms *can* accumulate on the forward evolving field components, as indeed can backward perturbations accumulate on the backward evolving field components. Despite the fact that the residual terms acting on the forward and backward field evolutions are the same size, forward evolving components of the residuals can accumulate on the forward evolving field because they are phase matched; whereas backward residuals are not, and rapidly average to zero.

A. Residual terms

The handling and criteria for different types of residual terms has been discussed already [14] for the spatially propagated case. While it would be straightforward to repeat that level of detail here, the basic methodology is virtually identi-

cal except for the different roles played by the t and \mathbf{r} arguments. Accordingly I will consider only the most commonly considered residuals – the dielectric scalar and vector terms as well as currents. Those interested in magnetic or magneto-electric effects can adapt the process to their own systems in a similar manner.

The relevant total polarization \mathbf{P}_D can then be decomposed into pieces before seeing how each might satisfy the slow evolution criteria. Thus we write

$$\begin{aligned} \mathbf{P}_D(\mathbf{D}, \mathbf{r}, t) &= a_\varepsilon(\mathbf{D}, \mathbf{r}, t)\mathbf{D}(\mathbf{r}, t) + \mathbf{U}_D(\mathbf{D}, \mathbf{r}, t) \\ &= a_L(\mathbf{r}, t) * \mathbf{D}(\mathbf{r}, t) + a_N(\mathbf{D}, \mathbf{r}, t) * \mathbf{D}(\mathbf{r}, t) \\ &\quad + \mathbf{U}_L(\mathbf{r}, t) * \mathbf{D}(\mathbf{r}, t) + \mathbf{U}_N(\mathbf{D}, \mathbf{r}, t). \end{aligned} \quad (4.1)$$

The scalar part is represented by a_ε , and might contain (non-reference) linear parts and time response (a_L), perhaps a convolution (“*”) over the past, or angle dependence; it might also include nonlinear contributions a_N such as a third order nonlinearity with $a_N\mathbf{D} \propto (\mathbf{D} \cdot \mathbf{D})\mathbf{D}$. The vector part \mathbf{U}_D , if present, could be due to something like a second order nonlinearity, which couples the ordinary and extra-ordinary field polarizations. This description of the material parameters is not restrictive, might e.g. include any order of nonlinearity.

B. Slow evolution?

Having categorized the relevant residual terms, it now remains to treat each one individually and determine the conditions under which the oppositely directed field can nevertheless remain negligible: i.e., for \mathbf{D}^\pm , we have that $\mathbf{D}^\mp \simeq 0$, where the scalar a_L contains the linear response of the material that is both isotropic and lossless (or gain-less); since here it is a time-response function, it is convolved with the electric displacement field \mathbf{D} . Note that the field vector \mathbf{D} , and indeed the material parameter a_L are all functions of time t and space $\mathbf{r} = (x, y, z)$; the polarization \mathbf{P}_D and its components a_ε , \mathbf{U}_ε are a functions of time t , space \mathbf{r} , and the field \mathbf{D} .

Below, the field vector \mathbf{D} will be split into components D_i and D_i^\pm , with $\mathbf{D}^+ + \mathbf{D}^- = \mathbf{D} \equiv (D_x, D_y, D_z)$ and $i \in \{x, y, z\}$. Like we also have wavevector components k_i from $\mathbf{k} = (k_x, k_y, k_z)$, where $k_\perp^2 = k_x^2 + k_y^2$; and further, k_j is also used as a substitute symbol to represent any one of k_x, k_y , or k_z .

Any corrections due to imperfect choice of Ω will appear in the \mathbf{P}_D polarization correction term. However, unlike in the spatially propagated case, there is no diffraction-like term – here, that is quite naturally part of the reference behaviour.

Firstly, we consider the scalar polarization terms a , which could be either linear ($a_L(k)$) or nonlinear ($a_N(\mathbf{D}, k)$). These might represent e.g. the time-response of the medium (and hence its dispersion), or be some nonlinear response such as the third-order Kerr nonlinearity. Since these are locally correlated in space, in the wavevector picture they include convolutions over k . The criterion therefore is

$$\frac{i\Omega a * |D_i^+ + D_i^-|/2}{i\Omega |D_i^\pm|} \simeq \frac{a * |D_i^\pm|/2}{|D_i^\pm|} \simeq \frac{a}{2} \ll 1. \quad (4.2)$$

In the important linear case, $a \equiv a_L$ is independent of \mathbf{D} , which means that the pulse properties play no role, and only the material parameters are constrained. In the nonlinear case, there is a further constraint, which is on the peak intensity of the pulse, but still there are no smoothness assumptions or bandwidth restrictions.

Secondly, we consider the linear and nonlinear vector terms from \mathbf{U} . This criterion is similar to the scalar case in eqn. (4.2), but with \mathbf{U} replacing $\phi\mathbf{D}$. Thus for $i \in \{x, y, z\}$, we can write down constraints for each component U_i , which are

$$i\Omega|U_i|/2 \ll i\Omega|D_i^\pm| \implies |U_i| \ll 2|D_i^\pm|. \quad (4.3)$$

In the linear case, $\mathbf{U} \equiv \mathbf{U}_L$, and the linear relationship between \mathbf{U}_L and \mathbf{D} again means that the criterion only constrains the material parameters contained in \mathbf{U}_L , not the field profile or amplitude. The nonlinear case is similar to the scalar one, and the peak pulse intensity is restricted; e.g. for a $\chi^{(2)}$ medium, $|\mathbf{U}_N| \sim \chi^{(2)}|\mathbf{D}|$.

An additional complication that occurs in this vector case is that a field consisting of only one field polarization component (e.g. D_x^+) may induce a driving in the orthogonal (and initially zero) components (e.g. D_y^\pm). This means that both D_y^\pm fields will be driven with the same strength, so that it is far from obvious that we can set D_y^- to zero, but still keep the D_y^+ without being inconsistent. However, it has already been noted that phase matching ensures that forward residuals accumulate, whilst the non-matched backward residuals are subject to the RWA, and become negligible: hence we can still rely on eqn. (4.3), albeit under caution.

Thirdly, we have the residual terms indicated by the presence of time-varying currents. These are

$$|J_i| \ll 2\Omega^2|D_i^\pm|. \quad (4.4)$$

Fourthly, we have the divergence terms $\mathbf{k}\mathbf{k} \cdot \mathbf{D}$, $\mathbf{k}\mathbf{k} \cdot \mathbf{P}_D$, which amount to charge density contributions either from \mathbf{D} or \mathbf{P}_D , with $\sigma_P(k) = a_\epsilon^{-1}\mathbf{k} \cdot \mathbf{P}_D$ and $\sigma_D(k) = \mathbf{k} \cdot \mathbf{D}$. Thus if the wavevector is oriented along a unit vector \mathbf{u} , then

$$\begin{aligned} u_i\Omega|\sigma|/2k &\ll i\Omega|D_i^\pm| \\ \sigma u_i &\ll 2k|D_i^\pm|. \end{aligned} \quad (4.5)$$

is the relevant criterion.

To summarize, these various criteria assert that modulations away from the reference evolution must be weak. Notably, weak nonlinearity is invariably guaranteed by material damage thresholds [33]. However, and in contrast to the usually favoured (but strictly speaking less causal) spatially propagated methods, the lack of any rigorous way to incorporate dispersion into the reference frequency $\Omega(k)$ makes it harder to satisfy the criteria for the ordinary linear response.

C. Uni-directional equation for \mathbf{D}^+

In a situation in which all of the above slow-evolution criteria hold, we can be sure that the backward directed field \mathbf{D}^- is only driven by a negligible amount, and if the no-accumulation condition also holds, then neither will there be any build up of backward evolving contributions to the field. Consequently, we can be sure that an initially negligible \mathbf{D}^- remains so, and eqn.(3.2) becomes

$$\begin{aligned} \partial_t \mathbf{D}^+(\mathbf{k}) = & +i\Omega(\mathbf{k})\mathbf{D}^+(\mathbf{k}) + \frac{i\Omega(\mathbf{k})}{2a_\epsilon} \mathbf{P}_D(\mathbf{D}^+, t; \mathbf{k}) \\ & - \frac{\mathbf{J}(t; \mathbf{k})}{2\Omega(\mathbf{k})} - \frac{i}{2\Omega(\mathbf{k})} \mathbf{k} \times \mathbf{M}_B(\mathbf{D}^+, t; \mathbf{k}) \\ & - \frac{i\Omega(\mathbf{k})}{2} \frac{\mathbf{k}}{k^2} \mathbf{k} \cdot \mathbf{D}^+(\mathbf{k}) + \frac{i\Omega(\mathbf{k})}{2a_\epsilon} \frac{\mathbf{k}}{k^2} \mathbf{k} \cdot \mathbf{P}_D(\mathbf{D}^+, t; \mathbf{k}). \end{aligned} \quad (4.6)$$

where now the polarization \mathbf{P}_D , magnetization, and divergence are solely dependent on the forward directed field \mathbf{D}^+ .

D. Modifications

Just as for the comparable spatially propagated derivation [14] we might also apply some of the strategies used in other approaches to get a more tractable and/or simpler evolution equation. Unlike in ‘‘traditional’’ derivations [34–37], none of these are required, but they nevertheless may be useful.

Briefly, a *co-moving frame* might be added, using $z' = z + tv_f$. This is a simple linear process that causes no extra complications; the leading RHS icD^+ term is replaced by $i(c \mp v)D^\pm$. Note that setting $c = v$ will freeze the phase velocity of the pulse, not the group velocity. Also, a *carrier-envelope* separation could be implemented using $D(z) = A(z) \exp[i(\omega_1 t - k_1 z)] + A^*(z) \exp[-i(\omega_1 t - k_1 z)]$ defining the envelope $A(z)$ with respect to wavevector k_1 and carrier frequency ω_1 . Multiple envelopes centred at different wavevectors and carrier frequencies (k_i, ω_i) could also be used to split the field up. This is typically done when the field principally comprises a number of distinct narrowband components (e.g. [31, 38, 39]). The wave equation can then be separated into one equation for each piece, coupled by the appropriate wavevector-matched polarization terms (c.f. [30]).

Typically, such modifications not only make the model easier to understand, but can also reduce the computational requirements when propagating the fields. This is in addition to the advantage given by the unidirectional model having only *one* first order differential equation for the field, in comparison to two in the case of FDTD. However, these advantages for field propagation may still be affected by numerical issues of stability and accuracy, and must be placed in the context of other auxiliary equations which may also be being solved, such as for temporal dispersion, an electron density, or other dynamic effects.

V. DISPERSION

The distinct contrast between the physical meaning of this temporally propagated wave equation and spatially propagated ones [14, 27, 31] now gives us an opportunity to consider the respective roles of temporal and spatial dispersion. This is because the temporally propagated picture holds all spatial information at any given moment in time, and so is a natural arena in which to treat spatial dispersion; whereas the spatially propagated picture, as we know, holds all temporal information at any given point in space, and so is a natural arena in which to treat temporal dispersion. However, the comparison is not as straightforward as we would like, so in what follows I restrict the discussion to the context of the simple slab waveguide.

Nevertheless, we should always keep in mind that the only way to do temporal dispersion correctly in a temporally propagated model is to explicitly model the time-response of the medium, just as you would do in (e.g.) FDTD methods [19].

A dielectric slab waveguide is a high index planar sheet (the core) of high index material clad on both sides by semi-infinite volumes of lower index material. Here the waveguide is taken to have thickness d in the perpendicular x direction, propagation in the z direction, and with core and cladding permittivities ϵ_1 and ϵ_0 . In the continuous wave (CW) limit, the spatial structure of the waveguide gives rise to frequency dependent properties that are usually called ‘‘spatial dispersion’’. Even for this simple slab design, the matching of the boundary conditions between core and cladding regions gives the dispersion relation a non trivial form; in this case it is given by the solution to a transcendental equation [40]. Typically, this is written in a way implying we want to calculate k_x and $\beta = k_z$ from ω , i.e. for the transverse electric (TE) field modes we have

$$T(k_x d) = k_x^{-1} \sqrt{\omega^2 \mu_0 (\epsilon_1 - \epsilon_2) - k_x^2}, \quad (5.1)$$

$$\text{where } \beta^2(\omega) = \omega^2 \mu_0 \epsilon_1 - k_x^2, \quad (5.2)$$

and $T(k_x d)$ is either $\tan(k_x d)$ or $-\cot(k_x d)$.

However, we can now notice that this expression looks like one where we know ω , and from that have to calculate the matching wavevector properties – i.e. that this is an implicitly spatially propagated representation. Therefore, for the temporally propagated case that is the focus of this paper, where \mathbf{k} is the known quantity, eqn. (5.2) should be rewritten. Accordingly we use k_z for β (since wavevector is no longer a propagation reference), and Ω for ω (since frequency now is a propagation reference); thus we have

$$c^2 \Omega^2(k_x) = k_x^2 \frac{\epsilon_1}{(\epsilon_1 - \epsilon_2)} [T^2(k_x d) + 1]. \quad (5.3)$$

$$\text{where } k_z^2 = c^2 \Omega^2(k_x) - k_x^2. \quad (5.4)$$

Here we see that the natural (reference) propagation frequency is defined solely by the transverse wavevector k_x , as is the matching evolution wavevector k_z . While apparently simpler than the traditional form, since Ω is directly given by k_x , we usually want to neglect any direct knowledge of the transverse

properties. This means that what we really want is not $\Omega(k_x)$, but $\Omega(k_z)$, the inverse of the traditionally preferred $\beta(\omega)$.

Despite these complications, which are in any case rather similar to those we see in the traditional spatially propagated picture, we can still find the same waveguide modes and their dispersions, but this time given as functions of k_z , not ω . To simplify the presentation, I now assume that for some n -th mode of the waveguide, we can expand about some central evolution wavelength k_{zn} , which corresponds to a reference frequency $\Omega_n = \Omega(k_{zn})$. This means we can write

$$\Omega(k_z) = \Omega_n + \sum_m \gamma_m (k_z - k_{zn})^m, \quad (5.5)$$

where $\gamma_1 = c + \bar{\gamma}_1$ has two contributions – that due to the vacuum (just c), and the estimated modifications due to spatial effects ($\bar{\gamma}_1$).

Now we arrive at the crucial point in the argument, where we also decide to add the effects of whatever *temporal* response the material might have to our propagation calculation; but also decide that an approximate model where we modify the expression above in eqn. (5.5) is sufficient. This is the converse of the traditional procedure, where the waveguide dispersion is added to whatever $\beta(\omega)$ is appropriate for the medium’s time response. Consequently, to proceed further we only need to make very similar assumptions: (a) that the two sources of dispersion are weak and so can simply be added, (b) that the temporal response is assumed to have negligible effect on the waveguide’s propagating modes, and (c) that the temporal dispersion is the same at all points in the waveguide.

The temporal response of the waveguide is quite naturally written in terms of a wavevector dependent on a specified frequency, and here is taken to have been simplified to a quadratic. Following the recommendations of [28], I also exclude imaginary components (e.g. loss or gain terms) in either the wavevector reference β or the frequency, intending to incorporate such effects as necessary at some later stage. Consequently, for the n -th mode, we have

$$\beta(\omega) = \beta_n + \kappa_1 (\omega - \omega_n) + \kappa_2 (\omega - \omega_n)^2, \quad (5.6)$$

where $\kappa_1 = c^{-1} + \bar{\kappa}_1$ has two contributions – that due to the vacuum (just $1/c$), and the modification due to the temporal response ($\bar{\kappa}_1$).

In order to map this into the temporally propagated picture, with a reference frequency being dependent on a wavevector; we replace ω with Ω , and β with k_z . Since we have chosen a quadratic form, the resulting expression can then be solved⁵ to find Ω . With $\Omega_n = ck_{zn}$, this gives

$$\Omega(k_z) = \Omega_n - \frac{\kappa_1}{2\kappa_2} \left[1 \mp \sqrt{1 + 4(k_z - k_{zn}) \kappa_2 / \kappa_1^2} \right], \quad (5.7)$$

which we could expand into a power series for small $k_{zn} - k_z$. Since κ_1 includes the vacuum, and dispersion κ_2 is assumed

⁵ In general – if there are more terms in the expansion, or β has some specific and non-trivial analytic form – this step is more complicated.

to be small, we should make the top sign choice since it gives the physically relevant small frequency corrections. The linear term in the series has the coefficient $\gamma_{1T} = 1/\kappa_1$ and the quadratic term $\gamma_{2T} = \kappa_2/\kappa_1^3$.

We now combine the two dispersive effects – the spatial/geometric/waveguide ones based on parameters γ_i , and converted temporal ones based on κ_i , to give a total reference frequency based on k_z of

$$\Omega'(k_z) = \Omega_n + \bar{\gamma}_1 (k_z - k_{zn}) + \sum_{m>1} \gamma_m [(k_z - k_{zn})]^m, \\ - \frac{\kappa_1}{2\kappa_2} \left[1 - \sqrt{1 + 4(k_z - k_{zn}) \kappa_2 / \kappa_1^2} \right] \quad (5.8)$$

$$\simeq \Omega_n + \left(\bar{\gamma}_1 + \frac{1}{\kappa_1} \right) (k_z - k_{zn}) + \left(\gamma_2 - \frac{\kappa_2}{\kappa_1^3} \right) (k_z - k_{zn})^2. \quad (5.9)$$

Note that we have to avoid incorporating the vacuum contribution twice when summing the dispersive effects, which means that $\bar{\gamma}_1$ appears instead of γ_1 .

In the complementary case, useful for spatial propagation, we can follow the converse – and indeed the *usual* procedure – to convert the CW effect of the waveguide's spatial structure $\Omega(k_z)$ into a frequency domain form [40, 41]. This then can be combined with temporal response of $\beta(\omega)$ to get

$$\beta'(\omega) = \beta_0 + \left(\bar{\kappa}_1 + \frac{1}{\gamma_1} \right) (\omega - \omega_0) + \left(\kappa_2 - \frac{\gamma_2}{\gamma_1^3} \right) (\omega - \omega_0)^2. \quad (5.10)$$

Discussion

We might have hoped to find a less approximate comparison between the spatially propagated and the temporally propagated approaches to dispersion what has just been discussed. Further, although qualitative comparisons can also be made, ideally we would like a specific and quantitative comparison as done for e.g. testing the unidirectional approximation [33].

That this is not as simple as it might sound can be demonstrated as follows: write two mathematically identical but physically and notationally distinct propagation equations, i.e.

$$\partial_z E(\omega) = +i\kappa(\omega)E(\omega) \quad (5.11)$$

$$\partial_t D(z) = +i\Omega(z)D(z). \quad (5.12)$$

For simplicity, this comparison is done in the 1+1D limit and for unidirectional equations; the residual ω or z behaviour has been merged with the reference behaviour to give single propagation parameters $\kappa(\omega)$ or $\Omega(z)$.

Superficially, the comparison between eqn. (5.11) and (5.12) looks promising. However, the vacuum part of $\kappa(\omega)$ is ω/c , but the vacuum part of $\Omega(z)$ is in contrast a spatial derivative, i.e. $c\partial_z$. Even if it were possible to match up the functional forms of the non-vacuum temporal response of $\kappa(\omega) - \omega/c$ with a spatial structure giving $\Omega(z) - c\partial_z$, the vacuum parts would remain unmatched. Alternatively, we

could compare the z -propagated ω -domain to the t -propagated k -domain. However, although in this case the reference behaviours match, but the k -domain model now has a convolution that does not appear in the ω -domain one.

As a result, it is unclear how well approximating temporal dispersion as spatial properties should work; but then it is equally unclear how well approximating spatial properties as temporal dispersion should work either. Nevertheless, this latter process is so ubiquitous in optics as to be effectively invisible, and – more importantly – does not seem to give rise to significant problems. Removing either of the mismatches discussed above – either in reference behaviour or convolution – requires a narrowband limit, a process which necessarily obscures the parameter regime where interesting tests and comparisons can (and should) be made – i.e. the short pulse, large bandwidth limit. Indeed, the stringent nature of the approximation would defeat the entire point of the comparison.

Nevertheless, it may be possible to design numerical simulations containing structures and temporal responses whose scales are carefully graduated so that the mundane bandwidth issues can be kept separate from the interesting propagation ones. Such a design would, as desired, allow the propagation approximations (only) to be compared and contrasted by numerical simulation. However, this is a non-trivial task, and one which I leave for later work.

VI. PULSE PROPAGATION

As an example, I will now consider pulse propagation in a simple optical fibre model [21, 26, 40, 41], where the dominant interesting effect is the third order nonlinearity present in silica. Here I compare 1+1D propagation such a waveguide, in order to elucidate the differences between the standard spatially propagated picture (as discussed in [14]) and the temporally propagated picture derived here. This comparison is made much clearer by not only the use of the factorization method in both cases, but also because it enabled both derivations to closely follow the same steps and approximations.

The assumptions made are those of transverse fields, weak dispersive corrections, and weakly nonlinear response; these all allow us to decouple the forward and backward wave equations. This decoupling means that without using any extra approximations, we can simplify our description and treat forward only pulse propagation. The specific example chosen here is for an instantaneous cubic nonlinearity, but it is easily generalized to non-instantaneous cases or even other scalar nonlinearities.

A. Spatially propagated NLSE

The commonly used spatially propagated NLS for $E_x^+(\omega)$ [31, 34, 42] based on a directional factorization [14] can be

written

$$\partial_z E_x^+ = +i\beta E_x^+ + \sum_m \kappa_m \cdot (\omega - \omega_0)^m E_x^+ + \frac{ik_0^2}{2\beta} \mathcal{F} \left[\chi^{(3)} E_x^2(t) E_x^+(t) \right], \quad (6.1)$$

where $\mathcal{F}[\dots]$ is the Fourier transform that converts the nonlinear polarization term into its frequency domain form. Here we have chosen a fixed, non-dispersive reference wavevector β , the temporal response of the propagation medium is encoded in the coefficients κ_j . For the temporal (time response) of the material, this is an in-principle exact re-representation of that response as a Taylor series expansion; although in practise the series is usually truncated after a few terms and the propagating fields restricted to within a bandwidth where only those terms are significant.

The important point to remember is that in this spatially propagated picture we do know the full time history, at least in a computational sense – i.e. as we compute a solution to the propagation equation. This “computational past” [16, 43], comprising all previously computed values of the field properties and material state means that a full frequency spectrum and response is known. As a result, the time response of the material *can* be directly encoded in an accurate and useful dispersion relation, and then be approximated as a Taylor series in ω . A depiction of this spatial propagation scheme was shown in fig. 2

Since this propagation is assumed to be taking place in a waveguide, we also have its geometric – wavelength sensitive – properties to consider. However, as discussed in the previous section we typically just re-represent those *as if they were instead generated by a temporal response* (see e.g. [40, 41]) Thus the coefficients κ_j usually combine both the effect of time response *and* geometric properties into a single total dispersion; this is the very basis of dispersion compensation in optical fibres, since the material (temporal) dispersion is offset by the waveguide design (spatial dispersion) [40].

B. Temporally propagated NLSE

Using the approach derived and presented in this paper, we can develop a time propagated counterpart to eqn. (6.1). In a purely mathematical sense, it is, apart from notation, identical to eqn. (6.1). However, in physical terms, the change is not trivial, since the physical meaning and boundary conditions differ significantly. Starting from eqn. (4.6) we can get a propagation equation in wavevector space for $D_x^+(k)$ which reads

$$\partial_t D_x^+ = +i\Omega D_x^+ + \sum_m \gamma_m \cdot (k_z - k_{z0})^m D_x^+ + \frac{i\Omega}{2a_e} \mathcal{G} \left[\chi^{(3)} D_x^2(z) D_x^+(z) \right], \quad (6.2)$$

where $\mathcal{G}[\dots]$ is the Fourier transform that converts the spatially localized nonlinear polarization into its wavevector domain form. *Here*, all dispersive behaviour has been combined into

the coefficients γ_j , which mimic the common notion of spatial dispersion. Such terms can arise directly from a Taylor expansion in k of a relevant CW dispersion relation about some suitable reference wavevector; however, although we can use the expansion to represent the spatial structure of the propagation medium (e.g. a waveguide), the conversion from spatial structure to k expansion is only relatively straightforward in the CW limit.

If the propagation medium also has a temporal response (i.e. temporal dispersion), and if we do not want to model it explicitly, then we can use the approach described in the previous section, where the temporal response of the material was (approximately) represented as if it were instead due to spatial properties.

Finally, note that in ordinary temporally propagated simulations, the computational past and the physical (temporal) past are identical; this is not the case for spatially propagated simulations, as mentioned above. A depiction of the temporal propagation scheme was shown in fig. 1

C. Discussion

Now that we have seen and compared the two complementary versions of unidirectional NLS equations, it is worthwhile to emphasize some of their features. In a spatial propagation model, the back-evolving wave is distinctly problematic, because it is travelling backwards in time. If our boundary conditions are strictly *initial* (starting) conditions as in Fig. 2, and we disallow sources of reverse-propagating waves at the endpoint of our propagation – then the back-evolving wave is non-causal. This means we can argue it needs to be removed (or guaranteed negligible) on purely physical grounds. In contrast, in temporal propagation, the back-evolving wave is simply evolving in the opposite (spatial) direction to our pulse, and there is no argument for removing it on purely physical grounds. Nevertheless, it is convenient to do so, as long as the criteria discussed in subsection IV B and appendix B hold, in order to derive the simple (6.2).

Note that if the third order nonlinearity is preserved accurately (as for the instantaneous $\chi^{(3)}$ in (6.1) and (6.2)), the dispersion is correctly accounted for, and the necessary constraints hold, pulses of any bandwidth can be evolved correctly by either form (see e.g. [26] for the spatial case). However, if we introduce a carrier-envelope picture to get a simpler NLS equation, we can drop the third harmonic generation contribution and retain only the self-phase modulation (SPM) nonlinear term. In spatial propagation, such an approximation restricts the allowed pulse bandwidth to one third of the carrier frequency, but in temporal propagation this is instead a restriction on the *wavevector* bandwidth. This constraint also helps ensure that the spectrum does not extend onto the negative frequency or wavevector range.

VII. CONCLUSION

I have derived a general first order wave equation for unidirectional pulse propagation in time. The scope was kept as general as possible, allowing for arbitrary dielectric polarization, magnetization, diffraction, and free electric charge and currents. This propagation equation has a utility all of its own, allowing efficient unidirectional propagation in a strictly causal manner (i.e. time directed). However, it requires comparable approximations to those used in deriving spatially propagated unidirectional propagation equations, although the specific tradeoffs are different.

The propagation equation was derived by first factorizing the second order wave equation for \mathbf{D} into an exact bi-directional model. Next, I applied the same type of well defined type of approximation to all non-trivial effects (e.g. non-linearity, diffraction), and so reduced the bi-directional propagation equations down to a simpler first order uni-directional wave equation. One feature of this factorization method is that it provides an easy way to do a term-to-term comparison of the exact bi-directional theory and approximate uni-directional descriptions.

In addition to its use where the causal simulation of a propagation problem is of strict concern, it also illuminates the

process of handling material response and spatial structure in waveguides by constructing a single combined dispersion model. As discussed in section V, we can now see that this simple addition of temporal and spatial dispersions in an ad hoc manner stands on rather poor foundations, particularly in the few-cycle pulse limit. Further, it is hard to see any straightforward way of testing the limitations of this common procedure, which poses an interesting challenge for the future. That said, in most applications, we expect that a combined dispersion will remain a perfectly adequate approximation, since the pulses are rarely so short, and it seems reasonable to assume that the character of the pulse evolution should remain similar, even if details might differ. Lastly, the lessons learnt from this investigation could also be applied to other types of wave propagation, most notably acoustic systems (see e.g. [20]).

Acknowledgments

The great majority of the work here was done whilst at Imperial College London, supported by EPSRC (grant number EP/K003305/1); but finished when at Lancaster University, again supported by EPSRC (the Alpha-X project EP/N028694/1).

-
- [1] T. Brabec and F. Krausz, *Rev. Mod. Phys.* **72**, 545 (2000), doi:10.1103/RevModPhys.72.545.
 - [2] P. B. Corkum, *Phys. Rev. Lett.* **71**, 1994 (1993), doi:10.1103/PhysRevLett.71.1994.
 - [3] K. J. Schafer, B. Yang, L. F. DiMauro, and K. C. Kulander, *Phys. Rev. Lett.* **70**, 1599 (1993), doi:10.1103/PhysRevLett.70.1599.
 - [4] T. Fuji, J. Rauschenberger, C. Gohle, A. Apolonski, T. Udem, V. S. Yakovlev, G. Tempea, T. W. Hansch, and F. Krausz, *New J. Phys.* **7**, 116 (2005), doi:10.1088/1367-2630/7/1/116.
 - [5] S. B. P. Radnor, L. E. Chipperfield, P. Kinsler, G. H. C. New, *Phys. Rev. A* **77**, 033806 (2008), arXiv:0803.3597, doi:10.1103/PhysRevA.77.033806.
 - [6] J. Solanpää, J. A. Budagosky, N. I. Shvetsov-Shilovski, A. Castro, A. Rubio, and E. Räsänen, *Phys. Rev. A* **90**, 053402 (2014), doi:10.1103/PhysRevA.90.053402.
 - [7] R. R. Alfano and S. L. Shapiro, *Phys. Rev. Lett.* **24**, 584 (1970), doi:10.1103/PhysRevLett.24.584.
 - [8] R. R. Alfano and S. L. Shapiro, *Phys. Rev. Lett.* **24**, 592 (1970), doi:10.1103/PhysRevLett.24.592.
 - [9] J. M. Dudley, G. Genty, and S. Coen, *Rev. Mod. Phys.* **78**, 1135 (2006), doi:10.1103/RevModPhys.78.1135.
 - [10] D. R. Solli, C. Ropers, P. Koonath, and B. Jalali, *Nature* **450**, 1054 (2007), doi:10.1038/nature06402.
 - [11] A. Braun, G. Korn, X. Liu, D. Du, J. Squier, and G. Mourou, *Opt. Lett.* **20**, 73 (1995), doi:10.1364/OL.20.000073.
 - [12] S. L. Chin, S. Petit, F. Borne, and K. Miyazaki, *Jpn. J. Appl. Phys.* **38**, L126 (1999), doi:10.1143/JJAP.38.L126.
 - [13] H. M. Milchberg, Y. H. Chen, Y. H. Cheng, N. Jhajj, J. P. Palastro, E. W. Rosenthal, S. Varma, J. K. Wahlstrand, S. Zahedpour, *Phys. Plasmas* **21**, 100901 (2014), doi:10.1063/1.4896722.
 - [14] P. Kinsler, *Phys. Rev. A* **81**, 013819 (2010), arXiv:0810.5689, doi:10.1103/PhysRevA.81.013819.
 - [15] P. Kinsler (2012), “Directional pulse propagation in beam, rod, pipe, and disk geometries”, arXiv:1210.6794.
 - [16] P. Kinsler (2014), “Negative Frequency Waves? Or: What I talk about when I talk about propagation”, arXiv:1408.0128.
 - [17] P. Kinsler, *Eur. J. Phys.* **32**, 1687 (2011), the arXiv version has additional appendices, arXiv:1106.1792, doi:10.1088/0143-0807/32/6/022.
 - [18] K. S. Yee, *IEEE Trans. Antennas Propagat.* **14**, 302 (1966), doi:10.1109/TAP.1966.1138693.
 - [19] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, *Comput. Phys. Commun.* **181**, 687 (2010),

- doi:10.1016/j.cpc.2009.11.008.
- [20] P. Kinsler (2012),
“Acoustic waves: should they be propagated forward in time, or forward in space?”,
arXiv:1202.0714.
- [21] G. P. Agrawal,
Nonlinear Fiber Optics
(Academic Press Inc., Boston, 2007), 4th ed.,
ISBN 978-0-12-369516-1.
- [22] L. Berge and S. Skupin,
Discrete Continuous Dyn. Syst. (DCDS-A) **23**, 1099 (2009),
doi:10.3934/dcds.2009.23.1099.
- [23] C. Ciraci, J. B. Pendry, and D. R. Smith,
ChemPhysChem **14**, 1109 (2013),
doi:10.1002/cphc.201200992.
- [24] S. J. Carter,
Phys. Rev. A **51**, 3274 (1995),
doi:10.1103/PhysRevA.51.3274.
- [25] A. Ferrando, M. Zia, P. F. de Cordoba, D. Binosi, and A. Montero,
Phys. Rev. E **71**, 016601 (2005),
doi:10.1103/PhysRevE.71.016601.
- [26] G. Genty, P. Kinsler, B. Kibler, and J. M. Dudley,
Opt. Express **15**, 5382 (2007),
doi:10.1364/OE.15.005382.
- [27] K. J. Blow and D. Wood,
IEEE J. Quantum Electronics **25**, 2665 (1989),
doi:10.1109/3.40655.
- [28] P. Kinsler,
Phys. Rev. A **79**, 023839 (2009),
arXiv:0901.2466,
doi:10.1103/PhysRevA.79.023839.
- [29] P. Kinsler, S. B. P. Radnor, and G. H. C. New,
Phys. Rev. A **72**, 063807 (2005),
arXiv:physics/0611215,
doi:10.1103/PhysRevE.75.066603.
- [30] P. Kinsler, G. H. C. New, and J. C. A. Tyrrell (2006),
arXiv:physics/0611213.
- [31] R. W. Boyd,
Nonlinear Optics
(Academic Press Inc., New York, 2008), 3rd ed.,
ISBN 978-0-12-369470-6.
- [32] S. E. Harris,
Appl. Phys. Lett. **9**, 114 (1966),
doi:10.1063/1.1754668.
- [33] P. Kinsler,
J. Opt. Soc. Am. B **24**, 2363 (2007),
arXiv:0707.0986,
doi:10.1364/JOSAB.24.002363.
- [34] T. Brabec and F. Krausz,
Phys. Rev. Lett. **78**, 3282 (1997),
doi:10.1103/PhysRevLett.78.3282.
- [35] P. Kinsler and G. H. C. New,
Phys. Rev. A **67**, 023813 (2003),
arXiv:physics/0212016,
doi:10.1103/PhysRevA.67.023813.
- [36] M. Geissler, G. Tempea, A. Scrinzi, M. Schnürer, F. Krausz,
and T. Brabec,
Phys. Rev. Lett. **83**, 2930 (1999),
doi:10.1103/PhysRevLett.83.2930.
- [37] M. Scalora and M. E. Crenshaw,
Opt. Comm. **108**, 191 (1994),
doi:10.1016/0030-4018(94)90647-5.
- [38] P. Kinsler and G. H. C. New,
Phys. Rev. A **69**, 013805 (2004),
arXiv:physics/0411120v1,
doi:10.1103/PhysRevA.69.013805.
- [39] P. Kinsler and G. H. C. New,
Phys. Rev. A **72**, 033804 (2005),
arXiv:physics/0606111,
doi:10.1103/PhysRevA.72.033804.
- [40] K. S. Thyagarajan and A. Ghatak,
Fiber Optic Essentials
(Wiley-IEEE Press, 2007),
ISBN 978-0-470-09742-7,
doi:10.1002/9780470152560.
- [41] J. K. Shaw,
Mathematical Principles of Optical Fiber Communications,
CBMS-NSF Regional Conference Series in Applied Mathematics
(SIAM, 2004),
ISBN 978-0-89871-708-2,
doi:10.1137/1.9780898717082.
- [42] J. Laegsgaard,
Opt. Express **15**, 16110 (2007),
doi:10.1364/OE.15.016110.
- [43] M. A. Ajaib (2013),
arXiv:1302.5601.

Appendix A: Factorizing

Here I present a simple overview of the mathematics of the factorization procedure, since full details can be found in Ferrando et al [25]. In the calculations below, I alter the physics but not the mathematical process (of [25]) to instead transform into frequency space, where the t -derivative ∂_t is converted to $i\omega$. This presentation closely matches that by Kinsler [14], but altered away from the spatial propagation scheme used there so as to be appropriate for temporal propagation used here. Also, we have that $\Omega^2 = c^2 k^2 / n^2$, and the unspecified residual term is denoted Q . The second order wave equation can then be written

$$[\partial_t^2 + \Omega^2] D = -Q \quad (\text{A.0.1})$$

$$[-\omega^2 + \Omega^2] D = -Q \quad (\text{A.0.2})$$

$$D = \frac{1}{\omega^2 - \Omega^2} Q = \frac{1}{(\omega - \Omega)(\omega + \Omega)} \quad (\text{A.0.3})$$

$$= \frac{-1}{2\Omega} \left[\frac{1}{\omega + \Omega} - \frac{1}{\omega - \Omega} \right] Q. \quad (\text{A.0.4})$$

Now $(\omega - \Omega)^{-1}$ is a forward-like (Green's function) propagator for the field, but note that in my terminology, it *evolves* the field. The complementary backward-like propagator is $(\omega + \Omega)^{-1}$. As already described in the main text, we now

write $D = D^+ + D^-$, and split the two sides up to get

$$D^+ + D^- = \frac{-1}{2\beta} \left[\frac{1}{\omega + \Omega} - \frac{1}{\omega - \Omega} \right] Q \quad (\text{A.0.5})$$

$$D^\pm = \frac{\pm 1}{2\Omega} \frac{1}{\omega \mp \Omega} Q \quad (\text{A.0.6})$$

$$[\omega \mp \Omega] D^\pm = \pm \frac{1}{2\Omega} \frac{1}{\omega \mp \Omega} Q \quad (\text{A.0.7})$$

$$i\omega D^\pm = \pm i\Omega D^\pm \pm \frac{i}{2\Omega} Q. \quad (\text{A.0.8})$$

Finally, we transform the frequency space $i\omega$ terms back into normal time to give t derivatives, resulting in the final form

$$\partial_t D^\pm = \pm i\Omega D^\pm \pm \frac{i}{2\Omega} Q. \quad (\text{A.0.9})$$

Appendix B: The no accumulation approximation

In the main text, I describe the no accumulation approximation in spectral terms as a RWA approximation. However, it is hard to set a clear, accurate criterion for the RWA approximation to be satisfied in the general case, since it requires knowledge of the entire propagation before it can be justified. In this appendix, I take a different approach to determine the conditions under which the approximation will be satisfied. This presentation closely matches that for the earlier spatial propagation version by Kinsler [14], but has been changed to match the temporal propagation used here.

First, consider a forward evolving field so $D = D_0 \exp(i\omega t)$, and therefore

$$D_0^- = \frac{\omega - \Omega}{\omega + \Omega} D_0^+ = \xi D_0^+, \quad (\text{B.0.1})$$

where as noted ω can be difficult to determine, and may even change dynamically; here we can assume it corresponds to the propagation wavevector that would be seen at if all the conditions holding at a chosen position also held everywhere

else. On this basis, we can (might) even (try to) define $\omega = \omega(t)$, where by analogy to the linear case we might assert that $\omega^2(t) = \Omega^2 + \Omega(t)/D(t)$, so that for small Ω , we have $\omega D \simeq \Omega(D + \Omega/2\Omega^2)$.

Let us start by assuming our field is propagating and evolving forwards (only), with perfectly matched D^\pm fields; so that $D^- = \xi D^+$. but then it happens that Ω changes by $\delta\Omega$ over a small interval δt , likewise ξ changes by $\delta\xi$. The D^\pm will no longer be matched, and now the total field splits into two parts that evolve in opposite directions. The part that continues to evolve forward has D^+ nearly unchanged, but the forward evolving D^- has changed size (and is now $\propto (\xi - \delta\xi)$) to stay perfectly matched according to the new Ω . The rest of the old D^- ($\propto \delta\xi$) now propagates backwards, taking with it a tiny fraction of the original D^+ (and is $\propto \xi \delta\xi$).

Comparing the two backward evolving D^- components at t and $t + \delta t$, and taking the limit $\delta t \rightarrow 0$ enables us to estimate that the backward evolving D^- field changes according to

$$\partial_t D_{0,backward}^- = \frac{2\Omega}{(\omega + \Omega)^2} [\partial_t \omega] D_{0,forward}^+. \quad (\text{B.0.2})$$

Now, using the small- Ω approximation for ω , we can write

$$\partial_t D_{0,backward}^- = \frac{1}{(\omega + \Omega)^2} [\partial_t \Omega] e^{-i\omega t}. \quad (\text{B.0.3})$$

where the exponential part removes any oscillations due to the linear part of Ω ; i.e. if $\Omega = \chi D$ then

$$\partial_t D_{0,backward}^- = \frac{1}{(\omega + \Omega)^2} [\partial_t \chi]. \quad (\text{B.0.4})$$

So here we see that backward evolving fields are only generated from forward evolving fields due to changes in the underlying conditions (i.e. either material response or pulse properties), but that for the reflection to be strong those changes will have to be significant on the order of a period, or be periodic so that phase matching of the the backward wave could occur.