OPTICA

Universal programmable and self-configuring optical filter: supplement

DAVID A. B. MILLER,[1,*] **CHARLES ROQUES-CARMES,**[1] **CARSON G. VALDEZ,**[1] **ANNE R. KROO,**[1] **MAREK VLK,**[1,2] **SHANHUI FAN,**[1] **AND OLAV SOLGAARD**[1]

[1]*Ginzton Laboratory, Stanford University, 348 Via Pueblo Mall, Stanford, California 94305, USA*
[2]*Department of Physics and Technology, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway*
*dabm@stanford.edu

Universal programmable and self-configuring optical filter: Supplemental document

David A. B. Miller[*,1], Charles Roques-Carmes[1], Carson G. Valdez[1], Anne R. Kroo[1], Marek Vlk[1,2], Shanhui Fan[1] and Olav Solgaard[1]

[1]*Stanford University, Ginzton Laboratory, 348 Via Pueblo Mall, Stanford CA 94305*
[2]*Department of Physics and Technology, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway*
**dabm@stanford.edu*

We give extended mathematical analysis and discussion of the approach to filtering discussed in the main text and an expanded discussion of some extensions of device usage and architectures. S1 gives a detailed formalism for analyzing the filter behaviors. S2 gives a complete analysis of a simple single-layer filter. S3 gives extended discussion of the classes and range of possible filter functions, especially for multilayer filters, and of approaches for programming the necessary matrices onto the interferometer mesh. S4 gives a detailed discussion of the analysis of temporal coherence and its measurement by the system. S5 discusses some other operating modes and extensions in more detail.

S1. Formalism for analyzing filter behavior

To understand this filter approach in detail, we can set up some physical and mathematical formalism. Much of this is standard for analyzing devices such as arrayed waveguide grating (AWG) routers, but we give this in relatively complete form because we need to extend it, especially for waveguide lengths beyond those used in conventional AWGs and for analysis of multiple-layer interferometer meshes, which allow many novel filters.

General description

Consider the system as illustrated schematically in Fig. S1, which is a generalized version of Fig. 1(a) of the main text.
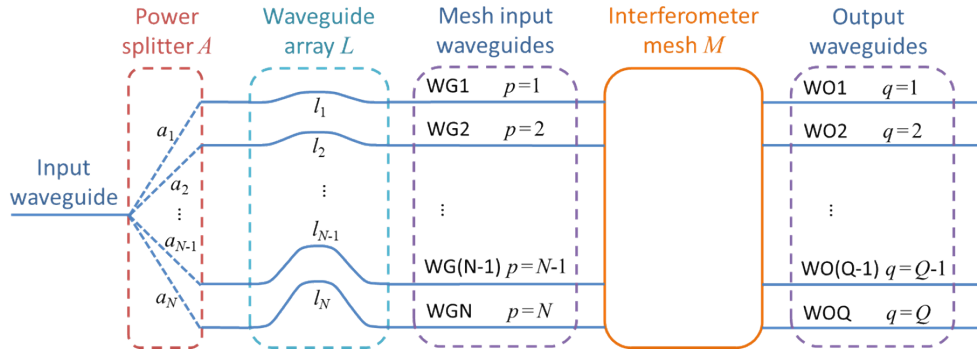

Fig. S1. General block diagram of the device.

All waveguides are presumed to be single-spatial-mode guides. We presume a single input waveguide. In this analysis, we consider one wavelength or frequency at a time, so, some angular frequency ω, which would correspond to a conventional frequency $f = \omega / 2\pi$, and a free-space wavelength $\lambda = c / f = 2\pi c / \omega$ where c is the velocity of light in free space. Since this is a linear, time-invariant system (at least over the time-scales relevant for optical frequencies), different frequencies are never mixed to generate new frequencies within the system, so the response of the system to a sum of light at different frequencies is just the sum of the responses to the individual frequencies.

Power splitting

An input field at this angular frequency ω and of amplitude x in the single input waveguide is split by this power splitter A among the waveguides in the array L with amplitude coefficients a_1 to a_N. In general, this split need not be equal, but if it were, for a lossless splitter these amplitudes would all be

$$a_p = 1/\sqrt{N} \tag{S1}$$

in magnitude; then the relative powers, presumed proportional to the modulus squared of the amplitudes, would add to 1. Whether or not the split is equal, quite generally, for a lossless system by power conservation we write

$$\sum_{p=1}^{N} |a_p|^2 = 1 \tag{S2}$$

We have written the modulus squared here (i.e., $|a_p|^2$) in each case, though in fact, without loss of generality, we could take these coefficients a_p to be real because we will have other points at which we can add arbitrary phases mathematically if we need to. (It will not matter in the final operation if there are different fixed phases associated with these splitting amplitudes; such phases can readily be calibrated out of the system [1].) If there is some common loss in the waveguides or the power splitter, we can simply presume some constant loss factor for the entire system and proceed otherwise as if the system is lossless.

Waveguide lengths

The waveguides in the array L will be fabricated to a set of lengths l_1 to l_N, which we could take as the waveguide lengths from the output of the power splitter to the input to the mesh. In the analysis that follows, the total length of the guides is of little or no interest; what matters is the difference in the lengths of the guides. (Lengths that are common to all paths just correspond to the same phase factor over all paths for a given frequency, and such a common phase factor is of no particular importance here.) For convenience, we can define such differences relative to some "center" length l_c. This could be the length of some "center" waveguide (if there is one), or some average length; in practice, we can just choose it to be some length that is convenient for our algebra. (It can also be the length of the first waveguide if that is more convenient for analysis.) Then we work with the relative length difference of waveguide p compared to l_c. So, we use the quantities

$$\delta l_p = l_p - l_c \tag{S3}$$

Various parameters in the analysis, such as power-splitting coefficients, beamsplitter ratios, and controlled phase shifts, could all depend on frequency, in which case we would regard these parameter values as being the ones at the angular frequency ω of interest. As a first approximation and for simplicity of analysis, we presume these parameters do not depend on frequency or wavelength, at least for narrow bandwidths. We do retain frequency dependence of the effective refractive index of the waveguide mode, however, since this is important in the system behavior through group velocity effects, and we discuss this next.

Waveguide phase delays and effective refractive indexes

In waveguides, some care is required with appropriate refractive index values, both for the effective refractive index at some frequency and with the effects of frequency-dependence or dispersion of that index. Especially in single-mode waveguides with small cross-sections, first, the effective (phase) refractive index for light in the guide at a given frequency, n_r, can differ significantly from the underlying refractive index n_o of the core material itself [2], in part

because some of the field amplitude is in the lower-refractive index waveguide cladding material. Second, the group index n_g may be quite different again, especially because of strong modal dispersion effects in small guides.

To understand refractive index effects and phase delays in single-mode guides, first we can formally define the phase refractive index for the waveguide mode

$$n_r = \frac{c\beta(\omega)}{\omega} \tag{S4}$$

where $\beta(\omega)$ is the (phase) propagation constant of the mode at (angular) frequency ω. Now we choose some "center" or reference (angular) frequency ω_c (corresponding to a conventional frequency $f_c = \omega_c / 2\pi$). This need not be in the actual center of some frequency band of interest, though conveniently it can be. For future use, we can define the effective wavelength inside the waveguide at the center frequency as

$$\lambda_c = \frac{c}{n_r f_c} \equiv \frac{2\pi c}{n_r \omega_c} \tag{S5}$$

We can usefully approximate n_r by a first-order expansion of its behavior with respect to frequency as

$$n_r(\delta\omega) = n_c + \delta\omega \frac{dn_r}{d\omega} \tag{S6}$$

where n_c is the (phase) index at the center frequency, $dn_r / d\omega$ is the rate of change of this refractive index n_r with respect to the angular frequency ω, and

$$\delta\omega = \omega - \omega_c \tag{S7}$$

is the (angular) frequency separation from the center (angular) frequency.

The relative phase delay in propagating through waveguide p can then be written

$$
\begin{aligned}
\delta\phi_p(\delta\omega) &= 2\pi n_r(\delta\omega)\frac{\delta l_p}{\lambda} = n_r(\delta\omega)\frac{\omega}{c}\delta l_p \\
&= \left(n_c + \delta\omega\frac{dn_r}{d\omega}\right)\frac{(\omega_c + \delta\omega)}{c}\delta l_p
\end{aligned}
\tag{S8}
$$

Dropping terms in $(\delta\omega)^2$, consistent with taking the first-order expansion as in Eq. (S6), Eq. (S8) leads to

$$\delta\phi_p(\delta\omega) = \left(n_c\omega_c + n_g\delta\omega\right)\frac{\delta l_p}{c} \tag{S9}$$

where

$$n_g = n_c + \omega_c\frac{dn_r}{d\omega} \tag{S10}$$

which is typically referred to as the group index. (Note that, if n_r did not depend on frequency – so we would then have $n_r = n_c$ – then Eq. (S9) would become the simple expression $\delta\phi_p(\delta\omega) = (\omega / c)n_r\delta l_p$.)

In the design of AWGs, it will often be the case that the waveguide lengths are chosen so that, at a center frequency,

$$n_c \omega_c \delta l_p / c = 2u_p \pi \tag{S11}$$

for some integers u_p, which is equivalent to saying that the lengths of the guides (compared to the "center" waveguide length) are integer numbers u_p of λ_c [the wavelength in the waveguide at the center frequency, Eq. (S5)], i.e.,

$$\delta l_p = u_p \lambda_c \tag{S12}$$

Such an approach leads to simpler analysis, and we take this approach here. Then Eq. (S9) can be rewritten in the simpler form

$$\delta \phi_p \left(\delta \omega \right) = n_g \delta \omega \delta l_p / c \tag{S13}$$

where we have dropped the integer multiples of 2π as in Eq. (S11) from Eq. (S9) since they make no difference when in the argument of any sine or cosine trigonometric function. In that case, only the group index would then be involved in the frequency response of the system [3].

In practice, it will be useful to work with length increments between guides that are themselves a specific integer number m_o of λ_c. Such a choice will mean that the free-spectral range of the system (essentially, the useful spectral range – see below for a detailed discussion) will be a correspondingly smaller fraction of f_c. (For example, the width of the telecommunication C-band is $\sim 1/44$ of its center frequency, which would suggest using $m_o = 44$ to get a filter of an appropriate free-spectral range.) So, we can choose some underlying length increment for design

$$\delta l_o = m_o \lambda_c = \frac{2m_o \pi}{n_c \omega_c} \tag{S14}$$

We can now usefully write

$$\delta l_p = m_p \delta l_o \tag{S15}$$

where $m_p m_o = u_p$.

We can now also usefully introduce the effective time delay δt_o for the wave at frequency $\omega_c + \delta \omega$ passing through that same waveguide length increment δl_o; specifically,

$$\delta t_o = n_g \delta l_o / c \tag{S16}$$

in which case Eq. (S13) becomes

$$\delta \phi_p \left(\delta \omega \right) = m_p \delta \omega \delta t_o \tag{S17}$$

For example, and for our explicit calculations, as mentioned in the main text, we presume a typical silicon photonics waveguide design that is a silicon strip waveguide 500 nm wide by 220 nm tall, with silicon dioxide cladding. We use material refractive indices of 1.444 [4] for the silicon dioxide cladding and $n_o = 3.4757$ [5] for the silicon core, as appropriate for 1550 nm free-space wavelength. Using an eigenmode solver (EMOpt [6]) with these parameters, at a "center" wavelength of 1550 nm we deduce a (phase) refractive index of $n_c = 2.45932$ and a group index of $n_g = 4.05647$ for the TE polarized mode in such a guide. The low phase index is in part because of penetration of the waveguide mode into the lower index cladding material surrounding that silicon core. The group index is strongly influenced by modal dispersion. (In calculations by others for other thin silicon guides [7], the group index for TE polarized modes can be as high as $n_g \simeq 4.25$ near 1550 nm (free-space) wavelengths.) So, in our analysis below, as in other analyses of AWG routers (see, e.g., [3]), we will include such dispersive effects in

the waveguides, at least to lowest order. The group index also varies slightly over the wavelength range of interest (such as the telecommunications C-band), but for simplicity in our calculations we use the above group index value for all wavelengths in our calculations.

Modeling the combination of splitter, waveguide array and interferometer mesh

We use the convention here that a forward propagating wave at angular frequency ω has the form $\exp[i(\omega t - \beta(\omega)z)]$, where t is time and we take z to be the appropriate distance along the waveguide. (If necessary, we can take the real part of any result to give a real field or amplitude.) The interferometer mesh in Fig. S1 can be represented by some matrix M with matrix elements M_{qp}. This mesh and matrix will have N input waveguides and Q output waveguides. Here p indexes the mesh input waveguides WG1 to WGN and q indexes the output waveguides WO1 to WOQ. N and Q can be equal, but they need not be; Q can also be greater than or less than N. A simple filter need only have one signal output waveguide (so we can have $Q=1$).

For an amplitude at (angular) frequency $\omega = \omega_c + \delta\omega$ (as defined in Eq. (S7)) of value x in the input waveguide, by definition the output field in the waveguide of index q is given by

$$y_q(\delta\omega) = \left[\sum_{p=1}^{N} a_p M_{qp} \exp(-i\, m_p \delta\omega \delta t_o)\right] x(\delta\omega) \tag{S18}$$

where we have used the phase delay from Eq. (S17) in each waveguide.

In the language of frequency filters [8], we can rewrite Eq. (S18) as (Eq. (2) of the main text)

$$y_q(\delta\omega) = H_q(\delta\omega) x(\delta\omega) \tag{S19}$$

where (Eq. (3) of the main text)

$$H_q(\delta\omega) = \sum_{p=1}^{N} a_p M_{qp} \exp(-i\, m_p \delta\omega \delta t_o) \tag{S20}$$

is the frequency response of the filter as seen in output waveguide q. For analysis of some simple systems, we may presume equal power splitting $a_p = 1/\sqrt{N}$ as in Eq. (S1). Then for simplicity of notation, we rewrite the matrix elements as

$$\tilde{M}_{qp} = a_q M_{qp} = \frac{M_{qp}}{\sqrt{N}} \tag{S21}$$

(this will allow the matrix $\tilde{M}$ to be an actual unitary matrix in practice), giving the simpler form

$$H_q(\delta\omega) = \sum_{p=1}^{N} \tilde{M}_{qp} \exp(-i\, m_p \delta\omega \delta t_o) \tag{S22}$$

For the case where the m_p are successive integers, the series in Eq. (S20) or Eq. (S22) essentially corresponds to a Fourier series, and N is then the "order" of the filter [8]. Altogether, the system is implementing Q different filter frequency responses H_q, with a possibly different such frequency response seen in each different (signal) output waveguide q.

Note in Eq. (S18), or equivalently Eq. (S20), the response at frequency $\delta\omega$ (relative to the "center" frequency ω_c) is the same as the response at frequency $\delta\omega + 2\pi n / \delta t_o$ for any integer n; because m_p is an integer,

$$\exp\left(-im_p\left[\delta\omega+\frac{2\pi n}{\delta t_o}\right]\delta t_o\right)=\exp\left(-i\left[m_p\delta\omega\delta t_o-2\pi m_p n\right]\right)$$
$$=\exp\left(-im_p\delta\omega\delta t_o\right)\exp\left(-i2\pi m_p n\right)=\exp\left(-im_p\delta\omega\delta t_o\right)\times 1=\exp\left(-im_p\delta\omega\delta t_o\right) \qquad \text{(S23)}$$

So, each term in the sum is the same if we add $2\pi n/\delta t_o$, which makes the response periodic in frequency with a "free-spectral range" of

$$\omega_{FSR}=2\pi/\delta t_o \qquad \text{(S24)}$$

or, in conventional frequency terms

$$f_{FSR}=1/\delta t_o \qquad \text{(S25)}$$

or, explicitly, using Eq. (S16)

$$f_{FSR}=c/n_g\delta l_o \qquad \text{(S26)}$$

So, in understanding what filter functions we can generate, we note that only one such free-spectral range matters (with the response in every other free-spectral range being identical). So, we restrict our analysis to the "zeroth" free-spectral range, which we can choose to regard as being from (angular) frequency $\delta\omega=-\pi/\delta t_o$ to $\delta\omega=\pi/\delta t_o$.

Filter functions, their representations, inner products, and Hilbert space

Much discussion of linear filter behavior is based on approaches that consider functions that may be continuous over all times and all frequencies or may be based on infinite sums of discrete and equally spaced frequencies or times. Approaches that work in such cases include Fourier transforms and z-transforms. Our case is somewhat different. First, we generally only have finite numbers of underlying physical functions we are considering, with the finite number being related to the number of waveguides in our arrays. Second, we may not even have uniform spacings of time delays between our different waveguides. As a result, it can be clearer to consider the filter functions of interest as being more generally in Hilbert spaces that can incorporate these restrictions. (Hilbert spaces are often implicit in the common approaches to analyzing and designing filters, but discussions of Fourier and z-transforms typically do not formally need to emphasize or point out this underlying mathematics.) Our use of Hilbert spaces enables us to be quite explicit about what functions we can consider (they are all the ones that can exist in our Hilbert spaces), retaining underlying mathematics of orthogonality and functional analysis without presuming that full Fourier or z-transforms must be used or be applicable.

With this background, to start, consider two possible filter functions formed as in Eq. (S20)

$$H_a\left(\delta\omega\right)=\sum_{p=1}^{N}\mu_{ap}\exp\left(-i\,m_p\delta\omega\delta t_o\right)\equiv\left|H_a\right\rangle$$
$$H_b\left(\delta\omega\right)=\sum_{p=1}^{N}\mu_{bp}\exp\left(-i\,m_p\delta\omega\delta t_o\right)\equiv\left|H_b\right\rangle \qquad \text{(S27)}$$

where we can consider that we are replacing the product of coefficients like $M_{qp}a_p$ with a single coefficient like μ_{ap} or μ_{bp} in writing these expansions. (We will clarify the meaning and use of Dirac notations like $\left|H\right\rangle$ below.) Generally, we can view such functions as existing mathematically in some "filter-function" function (or vector) space. With the formal mathematical definition of an inner product, which we can choose as

$$\langle H_a | H_b \rangle \equiv \frac{\delta t_o}{2\pi} \int_{-\pi/\delta t_o}^{\pi/\delta t_o} \left[\sum_{p=1}^{N} \mu_{ap}^* \exp\left(im_p \delta\omega\delta t_o\right) \right] \left[\sum_{j=1}^{N} \mu_{bj} \exp\left(-im_j \delta\omega\delta t_o\right) \right] d\left(\delta\omega\right) \quad \text{(S28)}$$

such a function or vector space essentially becomes a Hilbert space, mathematically.

Since quite generally

$$\int_{-\pi/\delta t_o}^{\pi/\delta t_o} \exp\left(iu\delta\omega\delta t_o\right) \exp\left(-iv\delta\omega\delta t_o\right) d\left(\delta\omega\right) = \frac{2\pi}{\delta t_o} \delta_{uv} \quad \text{(S29)}$$

for any integers u and v and with the Kronecker delta δ_{uv}, then the inner product reduces to

$$\langle H_a | H_b \rangle = \sum_{p=1}^{N} \sum_{j=1}^{N} \mu_{ap}^* \delta_{pj} = \sum_{p=1}^{N} \mu_{ap}^* \mu_{bj} \quad \text{(S30)}$$

Equivalently, we can formally define the set of "Fourier-like" basis functions

$$h_p\left(\delta\omega\right) = \sqrt{\frac{\delta t_o}{2\pi}} \exp\left(-im_p \delta\omega\delta t_o\right) \equiv |h_p\rangle \quad \text{(S31)}$$

each of which is also normalized over a free-spectral range (that is, the inner product of the functions with themselves is 1, i.e., $\langle h_p | h_p \rangle = 1$). We call this a "Fourier-like" basis rather than just a Fourier basis because, first, we only have a finite number N of orthogonal basis functions, rather than having p run from $-\infty$ to $+\infty$ as in a conventional Fourier series expansion, and, second, in the general case, the m_p need not be successive integers; for example, they might be given by a non-redundant array set such as a Golomb ruler, as discussed in the main text.

Then, any filter functions in our filter-function Hilbert space can be represented on this basis, Eq. (S31), by the vectors

$$|H_a\rangle \equiv \begin{bmatrix} \mu_{a1} \\ \vdots \\ \mu_{aN} \end{bmatrix}, \quad |H_b\rangle \equiv \begin{bmatrix} \mu_{b1} \\ \vdots \\ \mu_{bN} \end{bmatrix} \quad \text{(S32)}$$

When using a Dirac notation like this in which $|\alpha\rangle$ represents a column vector, then $\langle\alpha|$ represents a corresponding row vector with complex conjugated elements; formally $\langle\alpha| \equiv (|\alpha\rangle)^\dagger$, where quite generally in Dirac notation the superscript "†" denotes the Hermitian adjoint or conjugate transpose of a matrix or vector. In this Dirac notation, we see the vector product $\langle H_a \| H_b \rangle$ [conventionally shortened to the notation $\langle H_a | H_b \rangle$ as in Eq. (S30)] gives the same answer as the definition in Eq. (S28). Whether we use the representations of Eq. (S27) with inner product definition Eq. (S28) or of Eq. (S32) with inner product definition as the vector product $\langle H_a \| H_b \rangle \equiv \langle H_a | H_b \rangle$, the inner product $\langle H_a | H_b \rangle$ lets us define that two non-zero frequency response functions H_a and H_b are orthogonal if and only if $\langle H_a | H_b \rangle = 0$.

With the orthogonal basis functions as in Eq. (S31), we can also rewrite the filter functions of Eq. (S27) in the form

$$H_a(\delta\omega) \equiv \sqrt{\frac{2\pi}{\delta t_o}} \sum_{p=1}^{N} \mu_{ap} |h_p\rangle \equiv |H_a\rangle$$

$$H_b(\delta\omega) \equiv \sqrt{\frac{2\pi}{\delta t_o}} \sum_{p=1}^{N} \mu_{bp} |h_p\rangle \equiv |H_b\rangle$$

(S33)

Incidentally, a major reason for using Dirac notation, in addition to its algebraic simplicity and convenience, is that it allows us to refer to a function in the Hilbert function space in a way that is independent of what representation (basis set) we are using to specify the function. The representation in Eq. (S27) is in terms of the (infinite) set of amplitudes corresponding to the values of the continuous variable $\delta\omega$ within a free spectral range, whereas the representation in Eq. (S33) is in terms of an expansion over a discrete and countable basis set $|h_p\rangle$. Whichever way we represent it, we are referring to the same function, e.g., $|H_a\rangle$ or $|H_b\rangle$, in the mathematical Hilbert space of frequency response functions.

Because we can represent any frequency response function in the space of possible frequency response functions for this apparatus as a (weighted) sum over the N orthogonal functions $h_p(\delta\omega) \equiv |h_p\rangle$, the Hilbert space of possible frequency response functions is N-dimensional. The Fourier-like basis as in Eq. (S31) is one possible convenient basis, but of course we can choose any other N-dimensional orthogonal basis formed by a unitary transform from this basis. With such basis sets we can represent any filter function that can be formed from this set of N waveguides with their specified lengths δl_p as in Eq. (S15). One other such useful basis could be the "sinc-like" functions we would obtain as the frequency responses of AWG outputs (see the discussion below for the analysis of a simple filter. Another other particularly interesting basis would be the set of orthogonal functions that are the eigenfunctions of the (temporal) coherency matrix, as we discuss below in relation to partially coherent light.

Design of matrix elements to implement a filter function

If we know what filter function $|H\rangle$ we want, then we can find the best implementation of it, on the basis given by the functions as in Eq. (S31), by formally projecting it onto this basis to establish the coefficients μ_p. Explicitly

$$\mu_p = \sqrt{\frac{\delta t_o}{2\pi}} \langle h_p | H \rangle \equiv \sqrt{\frac{\delta t_o}{2\pi}} \int_{-\pi/\delta t_o}^{\pi/\delta t_o} \exp(im_p \delta\omega \delta t_o) H(\delta\omega) d(\delta\omega)$$

(S34)

Note that, as mentioned above, the integers m_p could be just consecutive integers for each successive integer p (leading to successive waveguide lengths δl_p that are be equally spaced, as in a conventional AWG), but they need not be. In the case of the non-redundant array filters in the main text, these m_p can be quite widely and relatively arbitrarily spaced integers. Nonetheless, the process described here, with the coefficients as in Eq. (S34) still gives us the best possible filter design (in a "least squares" sense) on the resulting basis. In general, linear combinations of basis functions of the form in Eq. (S31) tell us what filter functions are possible and hence, formally, in what filter-function Hilbert space we are working.

Time-domain description

Though we will mostly use the frequency-domain description as discussed above, we can briefly note the time-domain description. We use a time-domain discussion below when considering the (temporal) mutual coherence function and (temporal) coherency matrix.

8

It is straightforward to write the impulse response of the filter directly. If we presume a (Dirac) δ-function input at time $t = 0$, then the output of one row q of the filter in time is simply

$$\tilde{H}_q(t) = \sum_{p=1}^{N} a_p M_{qp}\, \delta\left(t - m_p \delta t_o\right) \tag{S35}$$

(Here we are neglecting some overall propagation time t_{prop} through the system for mathematical simplicity, and because it likely will not matter to us in applications; we could incorporate that by subtracting it inside the argument of each δ-function, delaying the output accordingly.) All we are saying here is that, if the input to the device is a short impulse (written mathematically as a δ-function), then the output is a string of N δ-functions, delayed by times $m_p \delta t_o$, and with corresponding amplitudes $a_p M_{qp}$. This time-domain response, Eq. (S35), with appropriate normalization, is the Fourier transform of the frequency-domain response in Eq. (S20) (Eq. (3) of the main text).

S2. Analysis of a simple filter

One filter construction, common in AWGs [9–14], is where the power is split equally between the waveguides as in Eq. (S1) and where the successive waveguides increase in length by equal amounts δl_o. For the moment, we are interested in only one "row" q of the matrix, which is equivalent to considering only one (signal) waveguide output at a time; for example, with just a single "layer" of MZIs in the device, as in Fig. 1(b) or (c) of the main text, the matrix would anyway have only one row (so, obviously, then $q = 1$). We call this a simple filter.

We will formally "center" these length differences to get an average guide length difference of zero for mathematical convenience. (The guides themselves will, of course, all have positive lengths; we are essentially just subtracting off the length of the "center" or average guide.) So, we write

$$\delta l_p = \left[p - 1 - \frac{(N-1)}{2} \right] \delta l_o \equiv \left[p - \frac{(N+1)}{2} \right] \delta l_o \tag{S36}$$

(Note that formally, for even N, this would mean that δl_p could be a half integer number of the length increments δl_o, rather than an integer number as formally presumed in Eq. (S15). However, the lengths δl_p will still be spaced by integer numbers of length increments δl_o since p is an integer; so, this possible additional $\delta l_o / 2$ just corresponds to an overall phase factor common to all elements in the sums over the waveguide responses, and such overall phase factors are generally of no particular significance in the operation of the device. This does, however, lead to a minor complicating detail in discussing free-spectral ranges below.)

From Eq. (S17), therefore, we have relative phase delays in the guides of

$$\delta\phi_p(\delta\omega) = \left(p - \frac{(N+1)}{2} \right) \delta\omega \delta t_o$$

As mentioned in the main text, we can usefully add a controllable "tilt" to the input phase shifters of the mesh (e.g., PS1 to PS4 in Fig. 1 of the main text), adding a phase increment γ between each successive waveguide, where γ is a real constant that can be chosen to be positive, negative or zero. This will allow us to tune the "center" frequency of the filter after we set it up. Here, for algebraic convenience (but with no change in the underlying physics), we center the phase tilts around the middle of the set of input waveguides. So, instead of the form $\delta_p = m_p \gamma$ for this "tilted" phase delay as in Eq. (7) of the main text, with our simple choice of $m_p = p$ and this centering, instead we use a "tilted" phase delay

$$\delta_p = \left(p - \frac{(N+1)}{2} \right) \gamma$$

We choose uniform matrix element magnitudes for this filter. With our presumed equal power splitting, and using the simplified notation of Eq. (S22) where we incorporate the corresponding $\sqrt{1/N}$ splitting amplitude into the matrix elements, those elements in our single row q of interest are then

$$\tilde{M}_{qp} = \frac{1}{\sqrt{N}} \exp\left(-i \left(p - \frac{(N+1)}{2} \right) \gamma \right) \tag{S37}$$

Substituting these matrix elements in Eq. (S22) gives

$$\begin{aligned}
y_q\left(\delta\omega\right) &= \frac{1}{N}\left[\sum_{p=1}^{N} \exp\left(-i\left[p - \frac{(N+1)}{2} \right]\gamma \right)\exp\left(-i\left(p - \frac{(N+1)}{2} \right)\delta\omega\delta t_o \right) \right] x\left(\delta\omega\right) \\
&= \frac{1}{N}\left[\sum_{p=1}^{N} \exp\left(-i\left[p - \frac{(N+1)}{2} \right]\alpha \right) \right] x\left(\delta\omega\right) \\
&= \frac{1}{N}\left[\sum_{s=0}^{N-1} \exp\left(-i\left[s - \frac{(N-1)}{2} \right]\alpha \right) \right] x\left(\delta\omega\right) \\
&= \frac{1}{N}\exp\left(i\frac{(N-1)}{2}\alpha \right)\left[\sum_{s=0}^{N-1} \exp\left(-is\alpha \right) \right] x\left(\delta\omega\right)
\end{aligned} \tag{S38}$$

where $s = p - 1$ and for convenience we have defined $\alpha = \gamma + \delta\omega\delta t_o$.

Now, the summation in Eq. (S38) is a geometric series of the form

$$\sum_{s=0}^{N-1} z^s = \begin{cases} \dfrac{1-z^N}{1-z}, & \text{for } z \neq 1 \\ N, & \text{for } z = 1 \end{cases} \tag{S39}$$

where $z \equiv \exp\left(-i\alpha\right)$. Hence, for $\alpha \neq 2m\pi$, where m is any positive or negative integer or zero

$$\begin{aligned}
y_q\left(\delta\omega\right) &= \frac{1}{N}\exp\left(i\frac{(N-1)}{2}\alpha \right)\left[\frac{1-\exp(-iN\alpha)}{1-\exp(-i\alpha)} \right] x\left(\delta\omega\right) \\
&= \frac{1}{N}\frac{\exp(iN\alpha/2)}{\exp(i\alpha/2)}\left[\frac{1-\exp(-iN\alpha)}{1-\exp(-i\alpha)} \right] x\left(\delta\omega\right) \\
&= \frac{1}{N}\left[\frac{\exp(iN\alpha/2)-\exp(-iN\alpha/2)}{\exp(i\alpha/2)-\exp(-i\alpha/2)} \right] x\left(\delta\omega\right) \\
&= \frac{1}{N}\left[\frac{2i\sin(N\alpha/2)}{2i\sin(\alpha/2)} \right] x\left(\delta\omega\right) = \frac{1}{N}\frac{\sin(N\alpha/2)}{\sin(\alpha/2)} x\left(\delta\omega\right)
\end{aligned} \tag{S40}$$

We can also rewrite this result as

$$y_q(\delta\omega) = \frac{1}{N}\frac{\alpha/2}{\alpha/2}\left[\frac{\sin(N\alpha/2)}{\sin(\alpha/2)}\right]x(\delta\omega) = \frac{\alpha/2}{\sin(\alpha/2)}\frac{\sin(N\alpha/2)}{N\alpha/2}x(\delta\omega)$$
$$= \frac{\text{sinc}(N\alpha/2)}{\text{sinc}(\alpha/2)}x(\delta\omega)$$

(S41)

where we take the definition

$$\text{sinc}\,x \equiv \begin{cases} \dfrac{\sin x}{x}, & x \neq 0 \\ 1, & x = 0 \end{cases}$$

(S42)

or explicitly now for $\alpha \neq 2m\pi$

$$y_q(\delta\omega) = \frac{1}{N}\frac{\sin\left(N[\delta\phi_c + \gamma + \delta\omega\delta t_o]/2\right)}{\sin\left([\delta\phi_c + \gamma + \delta\omega\delta t_o]/2\right)}x(\delta\omega)$$

(S43)

or equivalently

$$y_q(\delta\omega) = \frac{\text{sinc}\left(N[\delta\phi_c + \gamma + \delta\omega\delta t_o]/2\right)}{\text{sinc}\left([\delta\phi_c + \gamma + \delta\omega\delta t_o]/2\right)}x(\delta\omega)$$

(S44)

For $\alpha = 2m\pi$, since

$$\exp\left(i\frac{(N-1)2m\pi}{2}\right) = \exp\left(i(N-1)m\pi\right) = (-1)^{m(N-1)}$$

(S45)

then

$$y_q(\delta\omega) = (-1)^{m(N-1)}x(\delta\omega)$$

(S46)

The periodicity of the behavior of this filter in frequency space is easier to see from the "sine" form, Eq.(S43). If we change the frequency of the input from $\omega_a = \omega_c + \delta\omega_a$ to $\omega_b = \omega_c + \delta\omega_b$ where $\delta\omega_b = \delta\omega_a + 4m\pi/\delta t_o$ for any integer m, then no matter what is the integer value of N, the relation between the input $x(\delta\omega)$ and the output $y_q(\delta\omega)$ is unchanged because we have simply added an multiple of 2π to the argument of each sine function ($2Nm\pi$ on the sine in the numerator, and $2m\pi$ on the sine in the denominator). That would give a periodicity in angular frequency with a period or frequency increment $4\pi/\delta t_o$. While this is strictly correct, more commonly in filter response we may be interested in the periodicity of the "power" response of the filter, which would depend on the modulus squared of the output of the filter. In that case, we have a relation of the form

$$\left|y_q(\delta\omega)\right|^2 = \frac{1}{N^2}\frac{\sin^2\left(N[\delta\phi_c + \gamma + \delta\omega\delta t_o]/2\right)}{\sin^2\left([\delta\phi_c + \gamma + \delta\omega\delta t_o]/2\right)}\left|x(\delta\omega)\right|^2$$
$$= \frac{1}{N^2}\frac{1-\cos\left(N[\delta\phi_c + \gamma + \delta\omega\delta t_o]\right)}{1-\cos\left(\delta\phi_c + \gamma + \delta\omega\delta t_o\right)}\left|x(\delta\omega)\right|^2$$

(S47)

which is periodic with an angular frequency increment $2\pi/\delta t_o$. The apparent difference between the field and power frequency periodicities is because of the possible sign change as evident by the expression Eq. (S46). If N is even, the sign of the output changes as we change angular frequency by $2\pi/\delta t_o$, but the intensity response does not care about such a sign change. (This minor point on the apparent doubling of the free-spectral range is algebraically

because we formally "centered" the waveguide lengths about the average, and hence for even values of N we have introduced a $\delta l_o / 2$ shift in the waveguide lengths compared to those of Eq. (S15).) With this understanding, we can formally state that the free-spectral range of this filter is as defined above in Eqs. (S24) - (S26).

Within the free-spectral range, in addition to one frequency at which the input and output magnitudes are the same, there are $N-1$ frequencies at which the response is identically zero. If we presume that the frequency $\omega_0 = \omega_c + \delta\omega_0$ is such that

$$\delta\phi_c + \gamma + \delta\omega_0 \delta t_o = 2m\pi \tag{S48}$$

i.e., for some integer m

$$\delta\omega_0 = \frac{2m\pi - \delta\phi_c - \gamma}{\delta t_o} \tag{S49}$$

this frequency corresponds to a maximum output from the filter (both sinc functions in Eq. (S44) are then 1). Then the set of frequencies

$$\omega_s = \omega_c + \delta\omega_0 + s\frac{2\pi}{N\delta t_o}, \quad s = 1, 2, \ldots, N-1 \tag{S50}$$

will all correspond to zero output, so these frequencies, which are separated by

$$\Delta\omega = 2\pi / N\delta t_o \tag{S51}$$

or a conventional frequency separation

$$\Delta f = 1 / N\delta t_o \tag{S52}$$

are all perfectly rejected by the filter (at least for the output corresponding to this matrix row).

We can think of these (angular) frequencies $\omega_0, \omega_1, \ldots, \omega_{N-1}$ (or the corresponding conventional frequencies $f_0, f_1, \ldots, f_{N-1}$ obtained by dividing by 2π) and their shifted replicas in the other free spectral ranges as the reference comb for the system. (See, e.g., [15] for discussion of comb sources.) Note that, if we choose $\gamma_q = 2m\pi - \delta\phi_c$ for some integer m, we are effectively able to "center" the response of the filter round the chosen frequency ω_c just by imposing this "tilt" on the phase shifts of the mesh. This completes the formal analysis of the simple filter in the main text, with frequency responses as in Fig. 3(a) of the main text.

S3. Classes and range of filter operations

Meshes and matrices

With interferometer meshes, we can usefully separate the architectures and operations into those that implement unitary matrices (at least within some constant, such as an overall loss of the system) and those that implement non-unitary matrices. We can make mesh architectures in which we add loss, with controllable absorbers or by deliberately "dumping" power to external waveguides, which would generally make the corresponding matrix M represented by them be non-unitary. Mesh architectures more typically start by having no added internal loss or external "dump" ports, in which case they are represented by unitary matrices, typically written as U or V or their Hermitian adjoints (conjugate transposes) $U^\dagger$ or $V^\dagger$ (which mathematically are also necessarily unitary).

One particularly useful way to represent or construct any arbitrary non-unitary matrix is to use the singular-value decomposition (SVD) (see, e.g., [16]), in which we can write any matrix M as the matrix product

$$M = VD_{diag}U^\dagger \tag{S53}$$

where U and V are unitary matrices and D_{diag} is a diagonal matrix with a set of (generally complex) numbers s_j, known as the singular values, as the diagonal elements.

Equivalently, we can think of the matrix as mapping each of a set of orthogonal "input" vectors $|\psi_j\rangle$ to corresponding orthogonal "output" vectors $|\phi_j\rangle$ with coupling amplitudes s_j. The functions $|\psi_j\rangle$ are the eigenfunctions of the operator $M^\dagger M$, and the functions $|\phi_j\rangle$ are the eigenfunctions of the operator $MM^\dagger$, with positive real eigenvalues $|s_j|^2$, in both cases.

These sets are unique for any given matrix (at least within any arbitrary linear combinations of degenerate eigenvectors) – there is exactly one set of orthogonal "input" functions or vectors that maps, one by one, to exactly one set of corresponding "output" functions or vectors. These pairs of eigenfunctions, which can be called the "mode-converter basis sets" [16,17] for matrices representing optical systems, generally have fundamental physical properties and meaning in optical systems, including physical laws that apply only to them (see, e.g., [16,18]). The existence of these pairs of "input" and "output" orthogonal functions also automatically implies that any linear optical system has a set of orthogonal channels "through" it.
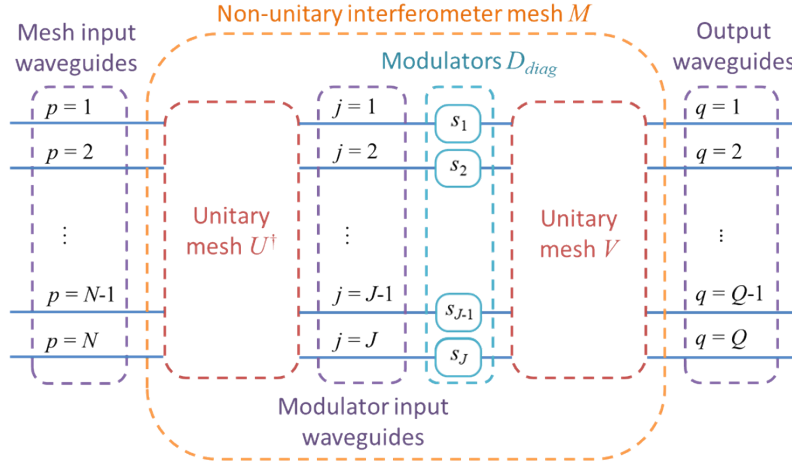

Fig. S2. Block diagram of a non-unitary matrix M implemented by singular-value decomposition [19] onto a succession of a unitary matrix $U^\dagger$, implemented by a corresponding "input" unitary mesh, a diagonal matrix D_{diag}, implemented by a line of modulators, and a unitary matrix V, implemented by an "output" unitary mesh. Such a matrix M need not be square; Q can be equal to N, or can be greater than or less than N. The number of modulators J need never exceed the smaller of Q or N.

In this notation, the $|\psi_j\rangle$ are the corresponding columns of U (and the corresponding $\langle\psi_j|$ are the rows of $U^\dagger$), and the $|\phi_j\rangle$ are the corresponding columns of V. In this notation, Eq. (S53) can be rewritten equivalently as

$$M = \sum_j s_j |\phi_j\rangle\langle\psi_j| \tag{S54}$$

Generally, then, any linear optical system between an "input" and an "output" space can be completely described by these coupled pairs of orthogonal functions $|\psi_j\rangle$ and $|\phi_j\rangle$ and their coupling strengths s_j, hence the term "mode-converter basis sets" [16,17]. If viewed as the set of communication channels between the input or "source" space and the output or "receiving"

space, these same function pairs and coupling strengths can be viewed equivalently as the "communication modes" [16,20,21] of the system.

This SVD decomposition gives a general way to implement arbitrary matrices with interferometer meshes [19], as illustrated in Fig. S2. Each of the unitary matrices $U^\dagger$ and V can be implemented with a unitary interferometer mesh and the matrix D_{diag} can be implemented with a line of modulators between these two unitary meshes. The outputs of the $U^\dagger$ unitary mesh feed the inputs of the modulators, and the outputs of the modulators feed the inputs of the V unitary mesh. The complex field transmission of a given modulator in waveguide j is just chosen to be the singular value s_j.

The SVD architecture for implementing arbitrary matrices was introduced in Ref. [19], and has been employed by others subsequently (e.g., Ref. [22]). This method is only as complicated as it needs to be [23]; the number of adjustable parameters in the system overall (e.g., phase shifters) is essentially equal to the number of real numbers required to specify an arbitrary $Q \times N$ complex matrix.

Range of possible filter functions

Given the ability to implement arbitrary unitary or non-unitary matrices with interferometer meshes, a wide range of filter functions can be implemented. Unlike some previous transversal or finite-impulse-response filters [24–26], we can contemplate a mesh with multiple outputs, each corresponding to a row of the matrix M, so the filter here can have multiple such frequency filtering behaviors simultaneously, one for each such mesh output. Unlike the fixed behavior of the outputs of a conventional AWG filter, we can have separate control over the filter functions for different outputs. This is an unusual capability for a physical filtering system, so we should clarify just what filters can be generated, both individually and in simultaneous sets.

As noted above, we can usefully view the filter frequency-response functions we can generate as formally being functions in a Hilbert space of possible functions. For N waveguides in the array, we can define this Hilbert space quite precisely as having the N dimensional basis set as in Eq. (S31) and the inner product as in Eq. (S28) or, more compactly, Eq. (S30), with frequency response functions that can be written as in Eq. (S27) or more compactly Eq. (S33).

In the discussion that follows, just as we did for the analysis of the simple filter above, for simplicity we presume equal power splitting, so all of the $a_p = 1/\sqrt{N}$ as in Eq. (S1), and we will fold these a_p into the rescaled matrix elements $\tilde{M}_{qp}$ for simplicity of notation, as in Eq. (S22). We will also presume waveguide length differences based on integer multiples of an underlying length δl_o, which will give all such filters a free-spectral range as in Eq. (S51), $\Delta\omega = 2\pi / N\delta t_o$, or equivalently as in Eq. (S52), $\Delta f = 1/ N\delta t_o$.

Just what different filter functions we can generate depends on whether the mesh is a unitary mesh or a more general non-unitary mesh (e.g., as in Fig. S2). Quite generally, though, for some arbitrary desired filter function $H(\delta\omega) \equiv |H\rangle$, the "best" version of that filter we can generate (in a "least squares" sense), will be the one with the matrix elements

$$\tilde{M}_{qp} = \frac{1}{\sqrt{N}}\langle h_p | H \rangle \equiv \sqrt{\frac{2\pi}{N\delta t_o}} \int_{-\pi/\delta t_o}^{\pi/\delta t_o} \exp(ip\delta\omega\delta t_o) H(\delta\omega) d(\delta\omega) \tag{S55}$$

This approach of finding the best representation of a function on a given basis by using the inner products with the basis functions as the expansion coefficients is standard in such linear algebra and functional analysis [27].

Unitary filters and filter sets

With M implemented as a unitary mesh, because of the construction of the multilayer unitary filter, the multiple filter functions it implements are physically guaranteed to be orthogonal to one another (at least if the mesh is "perfect" – that is, with 50:50 beam beamsplitters and the same loss on all paths through the mesh); each layer effectively implements a different row of a unitary matrix, and such rows are necessarily orthogonal. Any mutually orthogonal set of filter functions on the basis as in Eq. (S31) can be implemented. Because of the unitarity, the mesh overall will be loss-less; all the input power will appear somewhere in the N outputs of the mesh itself (whether we use all N of them or not).

If we express the filter functions as (normalized) vectors of amplitudes as in Eq. (S32), for example,

$$\left| H_q \right\rangle \equiv \begin{bmatrix} \tilde{M}_{q1} \\ \vdots \\ \tilde{M}_{qN} \end{bmatrix} \tag{S56}$$

any $Q\ (\leq N)$ such orthogonal vectors can be implemented in the outputs of a mesh with Q "layers" (such as the self-configuring layers as in Fig. 2 of the main text).

For example, instead of working directly on the Fourier-like basis of Eq. (S31), we could unitarily transform to and work with what we could call the "AWG" basis, which is functions each like those in Eqs. (S43) or (S44), but shifted (by effective mathematical phase tilts) so that the peak of a given filter function lines up with the zeros of every other such filter function. The actual filters that can be implemented do not depend on which such mathematical basis we choose, however.

Non-unitary filters and filter sets

Though unitary meshes are restricted to orthogonal filter functions, for non-unitary meshes, as in Fig. S2, in principle we can program any desired set of filter functions on the same basis. For N waveguides in the array, we can program up to N different filter functions, and these need not be orthogonal; those filter functions, on the Fourier-like basis Eq. (S31), are then simply the rows of the (generally non-unitary) matrix M. We do, however, need to consider power conservation in the final set of filters that can be generated. Though it would be possible to add gain instead of (lossy) modulation in generating a non-unitary mesh, adding such gain can be technologically challenging, so we presume we will not do so. Given that constraint, a procedure for setting up the best mesh can be as follows:

1. Write down the initially desired matrix M_{init} with the expansion coefficients of the desired filter functions as the rows.

2. Mathematically perform the SVD of M_{init}.

3. Note the required singular values, and in particular note the largest one, which, without loss of generality, we can call s_1.

4. Divide the matrix by s_1 (or, at least, its magnitude) to obtain $M_{actual} = M_{init} / |s_1|$. This guarantees that the largest singular value in M_{actual} is no larger than 1 in magnitude, which means physically that we do not need to introduce gain in the system. (The SVD of this new matrix M_{actual} is the same as that of M_{init} except that all the singular values are divided by $|s_1|$). Such new singular values can be implemented by the line of modulators as in Fig. S2.

With this approach, we can implement the desired set of non-orthogonal filter functions, subject to their overall magnitudes being reduced, if necessary, by this factor $1/|s_1|$. This ability to generate multiple non-orthogonal filter functions simultaneously and physically in optics, and with independent programmability, is an unusual one.

Mesh architectures and topologies

As we consider the full range of filters that could be made with this approach, we can usefully consider the topologies of architectures made from 2×2 blocks such as the MZIs in Fig. 1 of the main text [28]. Quite generally, the architectures we are considering are all "forward-only"; there are no reflections or loops in the network. In filter terminology, because of the absence of loops or reflections, these filters are all "finite impulse response" or (non-autoregressive) "moving average" filters [8] (as opposed to "infinite impulse response" or "autoregressive" filters). In network topology nomenclature, these are "directed acyclic graphs". The 2×2 (e.g., MZI) blocks can be viewed as the "nodes" in the graph and the waveguides between these nodes are the graph "edges".
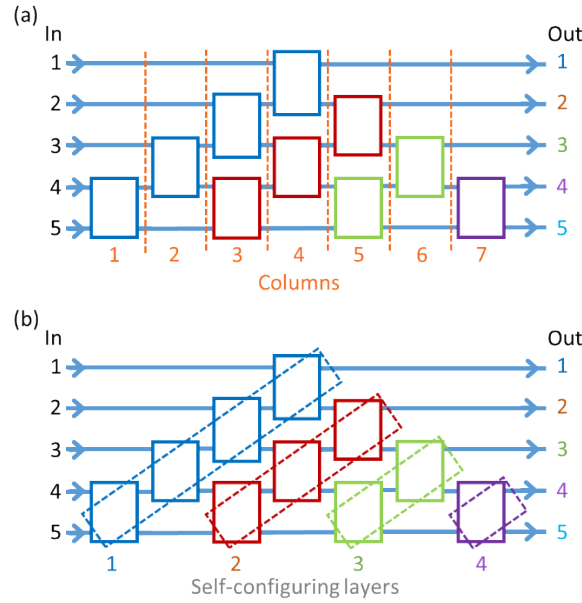

Fig. S3. Two important topologies for 2×2 blocks (such as MZIs) in forward-only (directed acyclic graph) mesh networks, illustrated for a triangular mesh. (a) Column topology, in which all blocks in a column can be configured or calibrated in parallel; all forward-only networks can be arranged into this form. (The triangular network is unusual in that no waveguide crossings are required in such an arrangement.) (b) Self-configuring layer topology; in such a layer, each layer input connects to the "signal" output by only one path through the blocks in the layer. Not all forward-only networks have such topologies, though any network formed from cascades of successive self-configuring layers automatically does.

Any such directed acyclic graph can be arranged into "columns" without changing the topology. (As discussed in Ref. [28], this arrangement corresponds to Dijkstra's algorithm [29].) The blocks or nodes in a column do not affect the inputs to any other blocks or nodes in the same column, and so they can be calibrated or set in parallel. With rectangles representing the 2×2 blocks, Fig. S3(a) shows the corresponding "column" topology in a "triangular" mesh [30] architecture.

Another important topology of some directed acyclic graphs or forward-only mesh architectures is the self-configuring layer topology [1] (Fig. S3(b)). Topologically, a self-configuring layer is one in which each input port is connected by only one path through the blocks to a single "signal output" port (or waveguide). The diagonal line and (symmetric) binary tree architectures [31] have this property, as do hybrids of them [1]. Conveniently, cascades of multiple such layers can be made. If we start with one such layer with N inputs, we can essentially "extract" this signal output from that layer as a useful output. Then the other $N-1$ outputs (which we could call "drop-port" outputs) are passed as inputs to a succeeding self-configuring layer (which has its own "signal output" port). If we wish, we can continue this for a total of up to $N-1$ layers to form a full cascade.

Such a full cascade implements a fully programmable $N \times N$ unitary matrix. Fig. S3(b) shows self-configuring layers in a full cascade in a triangular mesh architecture, which we see can also be viewed as being formed from successive cascaded "diagonal line" self-configuring layer mesh architectures (as also in Fig. 1(c))). Not all mesh architectures can be factored into such successive self-configuring layers; notably, the rectangular architecture [30] cannot be factored this way.

Detectors can be added between columns or self-configuring layers to sample the amount of power in one or both outputs of a 2×2 block, which can be convenient for calibrating or setting up the network. Such detectors can be "mostly transparent" detectors [32] in the waveguides, or can use waveguide "taps" to sample a small amount of power from the guide to a conventional detector [33]. Other techniques based on two output detectors [34] can also be used to deduce the power passing through any phase shifter in a forward-only mesh. With such monitoring of internal powers in the mesh, simple progressive techniques [28,35] allow mesh calibration by power maximization based on shining in specific vectors of amplitudes. Any given self-configuring layer can also be set up progressively just with a detector at its signal output; to configure the first 2×2 block, we set up the other blocks to route this power to the signal output of that layer, and can proceed similarly through the subsequent 2×2 blocks in the layer [31]; meshes based on cascaded self-configuring layers can also be set up progressively just with detectors at the outputs [1,19,31] in an extension of this approach.

Programming matrices on meshes

Generally, the matrices corresponding to collections of 2×2 blocks can be set up as matrix products of the 2×2 matrices corresponding to each block. With a calibrated mesh, and with calculation of the factorization of the overall matrix into such 2×2 blocks, arbitrary matrices can be set up this way. See, for example, Ref. [28] for a general discussion of this approach. Ref. [35] shows how to calibrate any forward-only mesh.

Programming self-configuring layers

Architectures made from successive self-configuring layers are particularly easy to set up and to calibrate [1,19,31,35]. Explicit procedures for diagonal line and symmetric binary tree mesh layers are given in Ref. [1] and its supplementary material, including both simple calibration procedures and explicit formulas for mesh settings for a given vector. Self-configuring layers only require one input vector for each self-configuring layer to be calibrated or programmed. Shining in a vector $|\psi_1\rangle$ and optimizing progressively for minimum powers at the drop ports of each successive 2×2 block configures the first such layer to have a matrix row of $\langle\psi_1|$. For each successive self-configuring layer j in a unitary mesh, shining in a vector $|\psi_j\rangle$ that is orthogonal to all the preceding such vectors $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_{j-1}\rangle$ can similarly configure the succeeding matrix row to $\langle\psi_j|$, giving a simple way of programming any unitary matrix to the

set of rows $\langle \psi_j |$ just by successively shining in the vectors $| \psi_j \rangle$ and configuring these successive layers.

It is also particularly straightforward to calculate the required settings of the self-configuring layer blocks to implement a desired matrix row. Conceptually, we can do this by imagining we are running the system backwards, shining light backwards into the signal output to generate a specific vector coming backwards out of the inputs to that mesh layer. Quite generally with unitary mesh networks (Ref. [1] supplementary section S5, Ref. [35]) if a given input vector $| \psi \rangle$ gives some output vector $| \phi \rangle$, then if instead we shine a vector of amplitudes $(| \phi \rangle)^*$ backwards into the "outputs", the vector that emerges from backwards from the inputs will be of the form $(| \psi \rangle)^*$. (When we say "of the form", we mean that the relative amplitudes and phase delays are the same as those of the vector $(| \psi \rangle)^*$. If there is any overall common loss in the forward direction, there will be the same loss in the backwards direction, so the overall amplitude of the backwards vector will be reduced accordingly.) This is a general property of reciprocal loss-less optical systems, and relates to phase conjugation (see Ref. [1] supplementary material for an explicit discussion of these phase conjugating properties in mesh mathematics); $(| \phi \rangle)^*$, which is the column vector whose elements are the complex conjugates of those in the vector $| \phi \rangle$, can be called the phase conjugate of $| \phi \rangle$.

Now, the phase conjugate of a beam going "forwards" in a single-mode guide is just the same mode going "backwards". If we supposed that the layer is set up to take all of the power in vector $| \psi \rangle$ and put it in the signal output port of that layer, then just shining light backwards into that signal output port (so "phase conjugating" that simple output) would therefore generate the vector $(| \psi \rangle)^*$ coming backwards out of the mesh "inputs". So, to design the settings of the self-configuring layer to take all of the power in input vector $| \psi \rangle$ and put it at its signal output, we simply imagine that we are shining light back into the signal output and calculate how to generate the vector $(| \psi \rangle)^*$ emerging backwards from the "inputs". For the diagonal line mesh, this is particularly obvious.

Using the diagonal mesh of Fig. 1(c), we set up the θ phase shifter of MZI3 in Fig. 1(c), to split out the desired fraction of power into WG1, and then set the ϕ phase shifter of MZI3 to give the desired phase. Then we set up the θ phase shifter of MZI2 to split out the desired fraction of the remaining power into WG2 and set its phase with the ϕ of MZI2, and so on. For other self-configuring layer meshes such as the symmetric binary tree, the calculations are slightly different but are still similarly progressive and straightforward.

In calculating how to set up successive self-configuring layers, we simply note the unitary matrices we have set up in earlier layers and compensate for those. So, for example, having set up the first layer in the mesh to correspond to the first matrix row $\langle \psi_1 |$, which will correspond to a unitary matrix U_1 as in Fig. 2 of the main text, then to calculate how to set up the second row so it implements the row $\langle \psi_2 |$, we set it up to send all the power in the vector $U_1 | \psi_2 \rangle$ to its signal output, and so on for each successive layer, multiplying all the unitary matrices corresponding to all previous layers as necessary. This algebra is explicitly given in [19].

Calculating how to set up the general non-unitary matrix in the SVD configuration is also straightforward with unitary meshes made from successive self-configuring layers; each of the two unitary meshes can be set up (or even self-configured) in the same way.

Simultaneous self-configuration with identifying tones

In an input signal comprised of multiple different optical frequencies, by putting different identifying tones on different specific optical frequencies in that input signal, such as some small amplitude modulation at a given tone frequency, layers in the network can be self-configured to extract such specific optical frequencies to specific signal outputs. For example, the detection process for maximizing or minimizing power, as required for self-configuration in a given layer, can also incorporate electronic filters that have the system look only for such specific tones and the amplitude of the corresponding modulation in the detected signals. (The use of such modulation tones or frequencies for setting up self-configuring layers is discussed in [31] and used, for example, in Ref. [32].) Such a process then can optimize that layer to extract, to its signal output, signals on this specific optical frequency. (This extraction of multiple frequencies by multiple layers can only be perfect if the corresponding filter functions are orthogonal, of course.)

S4. Analyzing temporal partial coherence with self-configuring filters

Here we give a formal discussion of the use of the device concept to analyze temporal partial coherence. We use a modal description of partial temporal coherence, using approaches from references such as Wolf [36] and others [37]. We consider the amplitude of the field mode in the single-mode input waveguide to be a single time-varying scalar field function $x(t)$. We formally presume that this function of time is one of a statistical ensemble $x_j(t)$ of such possible functions, where j indexes all the possible functions in the ensemble. Quite generally, the (first-order) coherence (or coherence function) of such an ensemble between the values of the function at two times t_1 and t_2 can be written as

$$\Gamma(t_1, t_2) = \left\langle x^*(t_1) x(t_2) \right\rangle_j \tag{S57}$$

where by $\langle \cdot \rangle_j$ we mean the average over this statistical ensemble. In our case, we choose to write

$$t_1 = t + u\delta t_o, \quad t_2 = t + v\delta t_o \tag{S58}$$

where t is time, δt_o is our time delay element characterizing the waveguide array (as in Eq. (S16)), and for the moment u and v are real numbers (though later those will be integers characterizing waveguide relative lengths).

We take these $x_j(t)$ to be ergodic random variables, by which we mean that an average over this ensemble of possibilities at some given time is the same as a time average. This in turn requires we presume the process is also "stationary", with probabilities independent of the chosen time origin. This choice means we presume that (average) power measurements we might make on these fields will be independent of time. With these ergodic and stationary assumptions, the average over time will be the same as the ensemble average. We can also freely shift the time origin for the averaging because of the stationarity property, and any such coherence function is then only a function of time differences. As a result, we can define the mutual coherence function $\Gamma(\tau)$ of the field as [36] (Eq. (8) of the main text)

$$\Gamma(\tau) = \left\langle x^*(t) x(t+\tau) \right\rangle_t \tag{S59}$$

where $\langle \cdot \rangle_t$ denotes time averaging, and where

$$\tau \equiv (v - u)\delta t_o \tag{S60}$$

This time averaging can be expressed as the integral

19

$$\Gamma(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^*(t) x(t+\tau) dt \qquad (S61)$$

Note, incidentally, because this is presumed to be a stationary process, shifting the time origin again by τ, makes no difference to the averaging. Hence

$$\Gamma(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^*(t-\tau) x(t) dt = \left[\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^*(t) x(t-\tau) dt \right]^* = \Gamma^*(-\tau) \qquad (S62)$$

Consider now the spectrometer system from Fig. 2 of the main text, with the "AWG-like" set of waveguides with lengths spaced progressively by increments δl_o. In this case, it is more useful to think of the waveguides as giving a set of time-delayed versions of the input signal. So, we have progressive relative time delays in waveguide p of $p\delta t_o$, with $\delta t_o = n_g \delta l_o / c$ (Eq. (S16), also Eq. (1) of the main text), and with equal power splitting, so a factor $a_p = 1/\sqrt{N}$ as in Eq. (S1) above. After passing through the power splitter and waveguide array, the input to the mesh M is a set of time-delayed replicas of the signal $x(t)$ of $(1/\sqrt{N})x(t-p\delta t_o)$ in the pth input waveguide. As a result, the output field in waveguide q is

$$y_q(t) = \frac{1}{\sqrt{N}} \sum_{p=1}^{N} M_{qp} x(t-p\delta t_o) \qquad (S63)$$

(We could also view this formally as the result of convolution of the input signal $x(t)$ with the impulse response of the system in waveguide q, as in Eq. (S35).)

We can formally write

$$y_q^*(t) = \frac{1}{\sqrt{N}} \sum_{s=1}^{N} M_{qs}^* x^*(t-s\delta t_o) \equiv \frac{1}{\sqrt{N}} \sum_{s=1}^{N} M_{sq}^\dagger x^*(t-s\delta t_o) \qquad (S64)$$

So, the time-averaged power in waveguide q is

$$\begin{aligned}
\left\langle \left| y_q(t) \right|^2 \right\rangle_t &\equiv \left\langle y_q(t) y_q^*(t) \right\rangle_t = \left\langle \frac{1}{N} \sum_{p=1}^{N} M_{qp} x(t-p\delta t_o) \sum_{s=1}^{N} M_{sq}^\dagger x^*(t-s\delta t_o) \right\rangle_t \\
&= \left\langle \frac{1}{N} \sum_{p=1}^{N} \sum_{s=1}^{N} M_{qp} x^*(t-s\delta t_o) x(t-p\delta t_o) M_{sq}^\dagger \right\rangle_t \\
&= \frac{1}{N} \sum_{p=1}^{N} \sum_{s=1}^{N} M_{qp} \left\langle x^*(t-s\delta t_o) x(t-p\delta t_o) \right\rangle_t M_{sq}^\dagger \\
&= \frac{1}{N} \sum_{p=1}^{N} \sum_{s=1}^{N} M_{qp} \left\langle x^*(t) x(t+(s-p)\delta t_o) \right\rangle_t M_{sq}^\dagger \\
&= \frac{1}{N} \sum_{p=1}^{N} \sum_{s=1}^{N} M_{qp} \Gamma((s-p)\delta t_o) M_{sq}^\dagger \equiv \frac{1}{N} \sum_{p=1}^{N} \sum_{s=1}^{N} M_{qp} \Gamma_{ps} M_{sq}^\dagger \\
&\equiv \frac{1}{N} \left(M \Gamma M^\dagger \right)_{qq}
\end{aligned} \qquad (S65)$$

where what we will now call the temporal coherency matrix Γ in the last line has the matrix elements (Eq. (9) of the main text)

$$\Gamma_{ps} \equiv \Gamma((s-p)\delta t_o) \qquad (S66)$$

These matrix elements therefore represent a discretized version of the mutual coherence function as in Eq. (S59). Note that, by Eq. (S62),

$$\Gamma_{ps} = \Gamma_{sp}^*$$ (S67)

So, this matrix Γ is Hermitian, which means it has real eigenvalues and orthogonal eigenvectors. We can therefore choose the rows of the unitary matrix M to correspond to the eigenvectors of Γ. In that case, the matrix $M\Gamma M^\dagger$ is diagonal, and, for unit input power, the output powers as given by Eq. (S65) are the eigenvalues of Γ. Because the left hand side of Eq. (S65) is the average of a positive number, we can see directly that all the eigenvalues of Γ are necessarily non-negative.

Now, in the presence of the input field, if we adjust the first layer of the mesh to maximize its output power, then we are variationally setting the first row of the matrix to correspond to the eigenvector with the largest eigenvalue. Leaving those settings, we can then similarly maximize the power from the second mesh layer, which will set the combination of the first and second layers to correspond to the second eigenvector and so on. In this way, we can establish the eigenvectors and eigenvalues of Γ, hence measuring this discretized version of the coherency matrix or coherence function. (This is the same sequential mesh optimization procedure introduced by some of us in the context of measuring the spatial coherency matrix [38].)

Note too that the input field has then been separated by the mesh into its eigen components, which in this case correspond to the "natural" mode or Karhunen-Loève expansion, providing the separate field at the signal outputs of the mesh. These output powers will also be mutually completely incoherent. We are not aware of another approach that accomplishes both this measurement and this non-destructive physical separation, which is simultaneous for all these components and is without any loss (other than background losses in the system).

Incidentally, having set up the matrix M in this way by physically separating into its natural mode components, it is now straightforward to understand exactly what set of filter functions we have constructed. We know the matrix elements M_{qp}, and so we can simply deduce the filter functions from Eq. (S20) (Eq. (3) of the main text).

Above we have presumed for simplicity in the derivation that the waveguides in the array have lengths $p\delta t_o$ where p is the summation index, so p takes on sequential integer values. We note that, as a result, several matrix elements of Γ_{ps} are identical (e.g., $\Gamma_{12} = \Gamma_{23}$), so we are essentially measuring the same value of the coherence function multiple times. If instead we chose sets of relative lengths $m_p\delta t_o$ such that all the differences between the various m_p values corresponded to different integers, then we could avoid this redundancy, and obtain measured values of $\Gamma(\tau)$ at a larger number of different time-delay values τ for the same number of physical measurements in the apparatus. Of course, this simply corresponds to making the set of integers m_p form a non-redundant set, as in a Golomb ruler, and as discussed for the non-redundant array filters in the main text. So, using such a non-redundant array approach for the waveguides gives us a more efficient method of measuring the coherence function for a larger number of different time (delay) values.

S5. Other operating modes and extensions

Here we give an extended discussion of some additional device operating modes and extensions to the device architecture.

Running the device backwards

Though we have so far considered the device working forwards from the input waveguide to one or more output waveguides, as in the architectures of Fig. 1 or 2, there are several possibly useful ways of running the device backwards.

One example would be wavelength multiplexing. We could, for example, shine light of a first wavelength back into the signal output 1 of Fig. 2. We could then adjust the settings of that first mesh layer to maximize the power coming back out of the input waveguide of the entire device. This would be equivalent to lining up the peak of the response as in Fig. 3 to coincide with that wavelength. Then, leaving those settings, we could now shine light of a second wavelength back into signal output 2 of Fig. 2, and adjust the settings of the second layer to again maximize the power coming back out of the input waveguide. If that second wavelength happened to line up with one of zeros of the filter response we have effectively set on the first layer, then we should be able to couple all of this second wavelength backwards out of the input waveguide. To the extent that the second wavelength does not line up perfectly with such a zero, there will be some loss in multiplexing this second wavelength (which would appear as loss in the power splitter, scattering light out of it). Nonetheless, even with some loss from such imperfect spacing of the wavelengths to be combined, we should still be able to couple much of the power from this second wavelength back out of the input guide. We can continue to extend this for different wavelengths backwards into the outputs of other layers in Fig. 2. A feature of this multiplexer is that it will still multiplex wavelengths together (with some loss) even if they are not quite equally spaced, and without knowing in advance just what the wavelengths are.

A second example of running the device backwards would be to operate the device as a wavelength-dependent mirror by adding an actual mirror or reflector at the signal output of a layer. Such a configuration used as a mirror in a laser cavity could allow tuning of the laser, for example over the free-spectral range of the device, with the laser running at or near the wavelength corresponding to the peak power transmission to that signal output.

A third example would be to use the device as a pulse generator. If the various outputs of a mesh as in Fig. 2 were each fed backwards by different mutually coherent elements of a frequency comb, the mesh could combine them controllably to generate pulses at a repletion rate given by the difference frequency between adjacent frequencies in the comb. Generally, we can imagine various architectures related to those of the schemes as reviewed in Weiner et al. [39] for pulse generation and shaping, with architectures such as those of Fig. 2 replacing grating dispersive elements, for example, and allowing quite arbitrary pulse shape control.

Other modalities exploiting partial coherence

Separating high-coherence sources or spectral lines

If the input spectrum contains some narrow lines amid a broad, incoherent background, the approach above proposed for measuring the coherence function will tend to separate out such coherent lines preferentially in the first layer or layers of the mesh because those will tend to be the strongest components that are coherent with themselves and mutually incoherent with respect to the rest of the spectrum. Note that this will happen without prior knowledge of the frequency of any such spectral lines.

Configuring for absorption lines rather than spectral emission lines

The discussion so far has largely concerned the detection and separation of specific frequencies or wavelengths of light to different outputs. This is also appropriate if we are performing spectroscopy where we are looking for different spectral lines in the emission from gases of atoms or molecules, for example. Often, however, spectroscopy is concerned with looking for corresponding absorption lines, again, for example, in gases of atoms or molecules. In such a

case, the material to be investigated, such as a gas, may be illuminated with a broad, incoherent spectrum of light, such as from the sun or some light bulb. To look for absorption lines, one can then configure or tune the first mesh layer to look for minimum rather than the maximum power in its signal output. Configuring a second layer to look for a maximum output power would tend to mean that the second layer was effectively measuring the background light power in regions with minimal absorption. The amount of absorption from the spectral line of interest can then be deduced by comparing the output of this first layer to the output of this second layer. Alternatively, if we want to compare the relative strength of two such absorption lines, so we can tell the relative abundance of two species, for example, we can compare the minimum power from two layers, one tuned to the first spectral line and the second tuned to the second such line.

Separating channels without knowing the wavelengths

Signals from different lasers can generally be considered mutually incoherent. So, for example, if we have several telecommunications signals from different lasers incident at the same time in the input, a mesh such as that in Fig. 2 could at least partially separate them automatically just based on their mutual incoherence as discussed above. In such a case, we might decide to optimize each layer based not on maximum output power, but instead on, say, the degree of opening of an eye diagram or a minimum in a bit error rate (see also the discussion below on optimization on other parameters). If the signals from the different lasers are of somewhat different powers and if their wavelengths are separated by more than the basic spectral resolution of the system (as given by the widths of the "peaks" in Fig. 3, for example), then we can expect that these signals can be usefully separated, at least approximately. Note that such a separation does not require advance knowledge of the specific wavelengths of the signals.

Extensions to the device architecture

NxQ meshes

Working with a large number N of input waveguides together with a smaller number Q of self-configuring layers can allow a relatively compact mesh (in terms of the number of MZIs required) that nonetheless has good spectral resolution and discrimination between different signals or spectra. Such an architecture could, for example, be programmed to pick out just a few (Q) spectral lines in a larger spectrum, presenting their powers at the Q output waveguides. It could also be a good mesh for separating a few wavelength channels when the input wavelengths are not known, as discussed above, because, with a large N, it more easily discriminates closely spaced channels.

Adding a controllable power splitter

Instead of the fixed power splitter of Figs. 1 and 2, we could instead use a controllable splitter, similar to the diagonal line or binary tree networks as shown in Fig. 1 but run backwards (i.e., with the input light shining "backwards" into the signal "output" of the layer). Such a device could allow us to pre-compensate for different losses in different waveguides so that the powers leaving the waveguides as they couple into the mesh are all approximately equal; this might be particularly useful for the non-redundant array architectures where waveguide lengths can be very different. In such a device, only controllable power splitting would be necessary (so only the θ phase shifters in the MZIs) because phase differences between guides are already handled by the mesh M.

Successive layers or filters for better rejection

With ideal components, a given layer can completely reject a given wavelength. Of course, imperfections exist; rejection will not generally be perfect, and some of that wavelength will also be passed to the next layer. For better rejection, one simple strategy is then to configure

the second layer also to reject the same wavelength. If there is no other leakage of light in the system, we can expect this second layer to lead to a doubling of the overall rejection (in decibels or some other logarithmic scale).

Of course, as with any filtering system, we can cascade complete filters, here also including another waveguide array, to improve rejection. Such a technique could be particularly effective with non-redundant array filters because the second set of waveguides could use a different set of lengths so their spectral "leakage" would not line up with the leakage in the first filter.

Optimization on other parameters

Instead of optimizing based on maximizing or minimizing power out of the signal port or ports, the device could be run to optimize other measurable parameters. For example, it could optimize based on "eye-opening" in telecommunications eye-diagrams or on minimum bit error rate, which essentially is a technique to correct for pulse dispersion and/or crosstalk with other channels. A related concept is to maximize the measured signal-to-noise ratio, which tends to create matched filters [40,41]. A single-layer filter could compensate and extract the single best channel, and multilayer filters could extract or compensate additional channels.

Handling imperfections and perfect MZIs

One of the possible imperfections in MZIs is that the beamsplitter split ratio will not in general be 50:50. That imperfection can limit rejection, for example, especially in "diagonal line" architectures. (Given approximately equal desired power splitting overall in a diagonal line, the overall power splitting of the individual MZIs in a long line tends to vary from nearly unity at one end of the line to nearly zero at the other. Symmetric binary trees are less susceptible to this issue since the corresponding power splitting of all MZIs tends to be near 0.5 for each MZI [42].)

One solution to this issue is to use "double" or "perfect" MZIs in which the actual beamsplitters are themselves implemented by "beamsplitter" MZIs. Such a "beamsplitter" MZI can behave as a 50:50 beamsplitter even for fabricated split ratios as bad as 85:15 [43]. There is also an algorithm that allows these MZIs to be set to 50:50 effective split ratio based only on successive power minimizations or maximizations [43], and high rejection has been reported in such configurations [44]. All such "perfect" MZIs inside a larger mesh can be set up with a sequence of such power minimization or maximization algorithms [43].

Dithering for derivative spectra

Derivative techniques can be generally useful in spectroscopy; they can identify line centers precisely, for example, as those correspond to zeros in the first derivative, and overall background light can be easily rejected. One way to perform such derivative spectroscopy here would be to add some small modulation of the γ tuning parameter, at some frequency ν and look for modulation at that frequency for first-derivative spectra (or at 2ν for second-derivative spectra) at the signal output of a given layer. This could be achieved by adding small modulations of amplitude $\propto m_p$ (as in Eq. (7) of the main text) to the "input" phase shifters (e.g., PS1 to PS4 in Fig. 1). A particularly simple way of doing that would be to modulate the temperature of the entire waveguide array at such a frequency ν since that would add phase shift appropriately proportional to the waveguide length (i.e., $\propto m_p$) in every waveguide in the array.

Spatio-spectral device

The spectral mesh concept discussed here could also be combined with spatial meshes. For example, if we had, say, W different input waveguides, with their inputs coming from different grating couplers or other spatial mode couplers, we could split each of those into its own array

of N waveguides of different lengths. We could then feed all those waveguides into a mesh with $W \times N$ inputs. That mesh could then implement simultaneous spatial and spectral operations. For example, it could possibly extract different spatial modes from a multimode fiber and perform different dispersion-compensating spectral filtering on different modes. It could also operate simultaneously on spatial and spectral partial coherence. Such a concept could also be extended to using non-unitary meshes (e.g., in the SVD architecture [19]).

References

1. D. A. B. Miller, "Analyzing and generating multimode optical fields using self-configuring networks," Optica **7**, 794–801 (2020).
2. X. Luo, Y. Cao, J. Song, X. Hu, Y. Cheng, C. Li, C. Liu, T.-Y. Liow, M. Yu, H. Wang, Q. J. Wang, and P. G.-Q. Lo, "High-Throughput Multiple Dies-to-Wafer Bonding Technology and III/V-on-Si Hybrid Lasers for Heterogeneous Integration of Optoelectronic Integrated Circuits," Front. Mater. **2**, Art. 28 (2015).
3. M. K. Smit and C. Van Dam, "PHASAR-based WDM-devices: Principles, design and applications," IEEE J. Sel. Top. Quantum Electron. **2**, 236–250 (1996).
4. I. H. Malitson, "Interspecimen Comparison of the Refractive Index of Fused Silica," J. Opt. Soc. Am. **55**, 1205–1209 (1965).
5. H. H. Li, "Refractive index of silicon and germanium and its wavelength and temperature derivatives," J. Phys. Chem. Ref. Data **9**, 561–658 (1980).
6. A. Michaels, "EMOpt," https://emopt.readthedocs.io/en/latest/ (2019).
7. E. Dulkeith, F. Xia, L. Schares, W. M. J. Green, and Y. A. Vlasov, "Group index and group velocity dispersion in silicon-on-insulator photonic wires," Opt. Express, OE **14**, 3853–3863 (2006).
8. C. K. Madsen and J. H. Zhao, *Optical Filter Design and Analysis: A Signal Processing Approach* (Wiley, 1999).
9. M. K. Smit, "New focusing and dispersive planar component based on an optical phased array," Electron. Lett. **24**, 385–386 (1988).
10. H. Takahashi, S. Suzuki, K. Kato, and I. Nishi, "Arrayed-waveguide grating for wavelength division multi/demultiplexer with nanometre resolution," Electron. Lett. **26**, 87–88 (1990).
11. C. Dragone, "An N*N optical multiplexer using a planar arrangement of two star couplers," IEEE Photon. Technol. Lett. **3**, 812–815 (1991).
12. K. Okamoto, K. Takiguchi, and Y. Ohmori, "16-channel optical add/drop multiplexer using silica-based arrayed-waveguide gratings," Electron. Lett. **31**, 723–724 (1995).
13. E. J. Stanton, N. Volet, and J. E. Bowers, "Low-loss demonstration and refined characterization of silicon arrayed waveguide gratings in the near-infrared," Opt. Express, OE **25**, 30651–30663 (2017).
14. F. Kish, V. Lal, P. Evans, S. W. Corzine, M. Ziari, T. Butrie, M. Reffle, H.-S. Tsai, A. Dentai, J. Pleumeekers, M. Missey, M. Fisher, S. Murthy, R. Salvatore, P. Samra, S. Demars, N. Kim, A. James, A. Hosseini, P. Studenkov, M. Lauermann, R. Going, M. Lu, J. Zhang, J. Tang, J. Bostak, T. Vallaitis, M. Kuntz, D. Pavinski, A. Karanicolas, B. Behnia, D. Engel, O. Khayam, N. Modi, M. R. Chitgarha, P. Mertz, W. Ko, R. Maher, J. Osenbach, J. T. Rahn, H. Sun, K.-T. Wu, M. Mitchell, and D. Welch, "System-on-Chip Photonic Integrated Circuits," IEEE J. Sel. Top. Quantum Electron. **24**, 1–20 (2018).
15. A. Pasquazi, M. Peccianti, L. Razzari, D. J. Moss, S. Coen, M. Erkintalo, Y. K. Chembo, T. Hansson, S. Wabnitz, P. Del'Haye, X. Xue, A. M. Weiner, and R. Morandotti, "Micro-combs: A novel generation of optical sources," Physics Reports **729**, 1–81 (2018).
16. D. A. B. Miller, "Waves, modes, communications, and optics: a tutorial," Adv. Opt. Photon. **11**, 679–825 (2019).
17. D. A. B. Miller, "All linear optical devices are mode converters," Opt. Express **20**, 23985–23993 (2012).
18. D. A. B. Miller, L. Zhu, and S. Fan, "Universal modal radiation laws for all thermal emitters," PNAS **114**, 4336–4341 (2017).
19. D. A. B. Miller, "Self-configuring universal linear optical component," Photon. Res. **1**, 1–15 (2013).
20. D. A. B. Miller, "Spatial channels for communicating with waves between volumes," Opt. Lett. **23**, 1645–1647 (1998).

21. D. A. B. Miller, "Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths," Appl. Opt. **39**, 1681–1699 (2000).
22. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," Nat. Photon. **11**, 441–446 (2017).
23. D. A. B. Miller, "How complicated must an optical component be?," J. Opt. Soc. Am. A **30**, 238–251 (2013).
24. K. Sasayama, M. Okuno, and K. Habara, "Coherent optical transversal filter using silica-based waveguides for high-speed signal processing," J. Lightwave Technol. **9**, 1225–1230 (1991).
25. X. Xu, G. Ren, T. Feleppa, X. Liu, A. Boes, A. Mitchell, and A. J. Lowery, "Self-calibrating programmable photonic integrated circuits," Nat. Photon. **16**, 595–602 (2022).
26. X. Liu, G. Ren, X. Xu, A. Dubey, T. Feleppa, A. Boes, A. Mitchell, and A. Lowery, "'Dial up' Photonic Integrated Circuit Filter," J. Lightwave Technol. **41**, 1775–1783 (2023).
27. D. A. B. Miller, "An introduction to functional analysis for science and engineering," http://arxiv.org/abs/1904.02539 (2019).
28. S. Pai, I. A. D. Williamson, T. W. Hughes, M. Minkov, O. Solgaard, S. Fan, and D. A. B. Miller, "Parallel Programming of an Arbitrary Feedforward Photonic Network," IEEE J. Sel. Top. Quantum Electron. **26**, 6100813 (2020).
29. E. W. Dijkstra, "A note on two problems in connexion with graphs," Numer. Math. **1**, 269–271 (1959).
30. W. Bogaerts, D. Pérez, J. Capmany, D. A. B. Miller, J. Poon, D. Englund, F. Morichetti, and A. Melloni, "Programmable photonic circuits," Nature **586**, 207–216 (2020).
31. D. A. B. Miller, "Self-aligning universal beam coupler," Opt. Express **21**, 6360–6370 (2013).
32. A. Annoni, E. Guglielmi, M. Carminati, G. Ferrari, M. Sampietro, D. A. B. Miller, A. Melloni, and F. Morichetti, "Unscrambling light—automatically undoing strong mixing between modes," Light Sci. Appl. **6**, e17110–e17110 (2017).
33. A. Ribeiro, A. Ruocco, L. Vanacker, and W. Bogaerts, "Demonstration of a 4 × 4-port universal linear circuit," Optica **3**, 1348–1357 (2016).
34. S. Pai, C. Valdez, T. Park, M. Milanizadeh, F. Morichetti, A. Melloni, S. Fan, O. Solgaard, and D. A. B. Miller, "Power monitoring in a feedforward photonic network using two output detectors," Nanophotonics https://doi.org/10.1515/nanoph-2022-0527 (2023).
35. D. A. B. Miller, "Setting up meshes of interferometers; reversed local light interference method," Opt. Express **25**, 29233–29248 (2017).
36. E. Wolf, "New theory of partial coherence in the space–frequency domain. Part I: spectra and cross spectra of steady-state sources," J. Opt. Soc. Am. **72**, 343–351 (1982).
37. J. W. Goodman, *Statistical Optics*, 2nd edition (Wiley, 2015).
38. C. Roques-Carmes, S. Fan, and D. A. B. Miller, "Measuring, processing, and generating partially coherent light with self-configuring optics," Light Sci. Appl. **13**, 260 (2024).
39. A. M. Weiner, "Ultrafast optical pulse shaping: A tutorial review," Optics Comm. **284**, 3669–3692 (2011).
40. G. Turin, "An introduction to matched filters," IRE Trans. Inform. Theory **6**, 311–329 (1960).
41. E. Marom, "Optical delay line matched filters," IEEE Trans. Circ. Syst. **25**, 360–364 (1978).
42. S. Pai, O. Solgaard, S. Fan, and D. A. B. Miller, "Scalable and self-correcting photonic computation using balanced photonic binary tree cascades," http://arxiv.org/abs/2210.16935 (2022).
43. D. A. B. Miller, "Perfect optics with imperfect components," Optica **2**, 747–750 (2015).
44. C. M. Wilkes, X. Qiang, J. Wang, R. Santagati, S. Paesani, X. Zhou, D. A. B. Miller, G. D. Marshall, M. G. Thompson, and J. L. O'Brien, "60 dB high-extinction auto-configured Mach–Zehnder interferometer," Opt. Lett. **41**, 5318–5321 (2016).