Understanding and exploiting advanced optics with communication modes: fundamentals, applications and limits

stanford.io/4oZy7bf

David Miller
Stanford University

Questions

Three basic questions for optics and waves for information

- What are the "best" channels for communicating with waves?
  e.g., out of one volume or surface to another
  These should be independent of one another
  so "orthogonal" in some mathematical and physical sense
  "**communication modes**"
- How many of those channels are there?
- How do these channels affect what optics we can design and make?

Answering these questions will

- let us understand "diffraction" limits beyond classical optics and for arbitrary structures
  including nanostructures and metamaterials
- give us limits to size and thickness
  even when using the best metasurface designs
  "**overlapping nonlocality**"
- give us some quite fundamental limits

Questions

There is a good way of thinking about this

and coming to definite and quite useful answers

but it is not like many of the previous ways of thinking about optics and waves

Note in particular that the answers generally are *not* any of the "standard" sets of "modes"

plane waves, Laguerre- or Hermite-Gaussians, orbital angular momentum beams, Bessel beams …

and those approaches can lead us to paradoxes and mistakes

but the answers are well defined mathematically

being quite straightforward to calculate

and they have real physical meaning

A different way of thinking about modes and waves

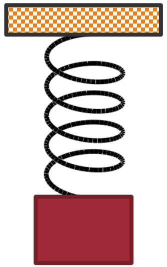
We are used to modes for

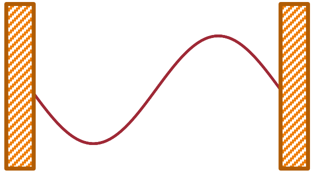
- resonators
- propagating modes in waveguides

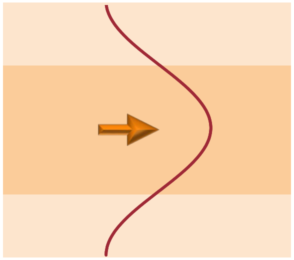
We like "modes" because they are economical

- We can use a few mode amplitudes
  - not fields at every point

We can often "count" modes meaningfully







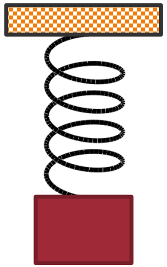
A different way of thinking about modes and waves

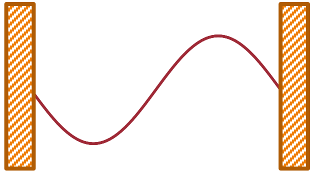
Modes have very useful mathematical properties, e.g.,

orthogonality

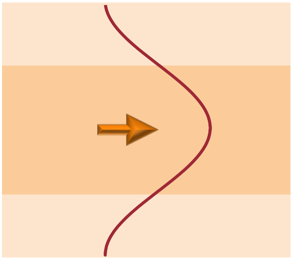
completeness

We can give a definition of a mode





A mode is an eigenfunction of an eigen problem describing a physical system



A different way of thinking about modes and waves

When we look generally at

communications with waves

or scatterers, optical devices, or nanostructures

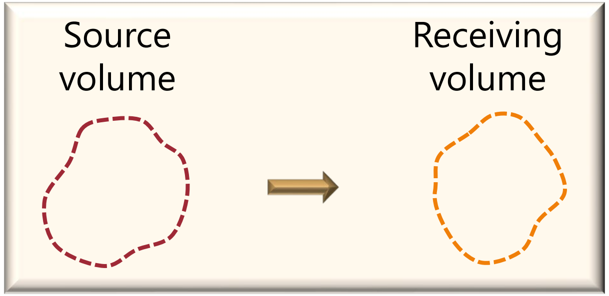
we can think in terms of

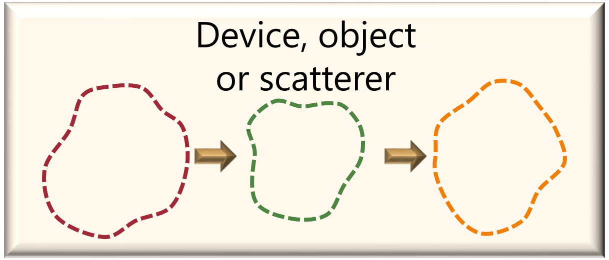
- a "source" or input space
- and a "receiving" or output space

We can ask first

what is the best choice of source function that

leads to the strongest wave in the receiving space





"Waves, modes, communications, and optics: a tutorial," Adv. Opt. Photon. **11**, 679-825 (2019)

A different way of thinking about modes and waves

We see immediately a difference compared to previous “beams”

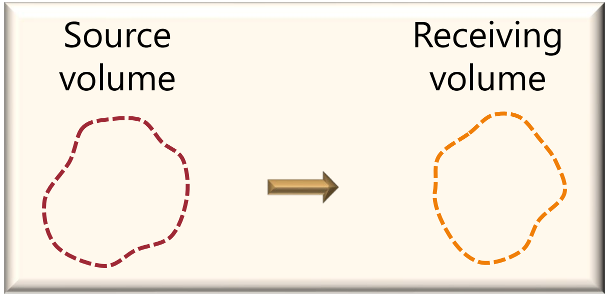
- Our answer will involve **two** functions
  - one in the source space
    - and one in the receiving space

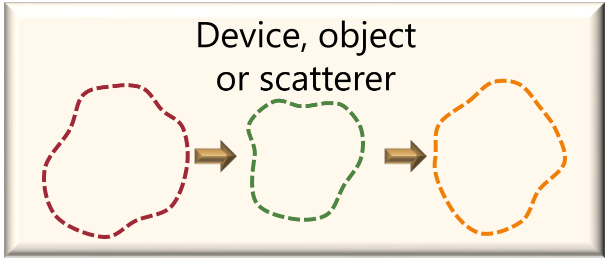
If we want to use the word “mode”

- these are “modes” in **two** spaces
  - not one space

The answer is **not** the “beam” between the spaces

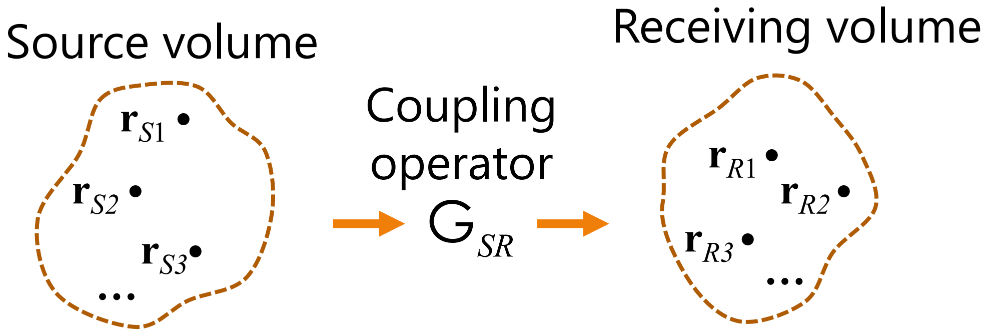
- though we can calculate that afterwards if we want





"Waves, modes, communications, and optics: a tutorial," Adv. Opt. Photon. **11**, 679-825 (2019)

Constructing examples with point sources and receivers



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

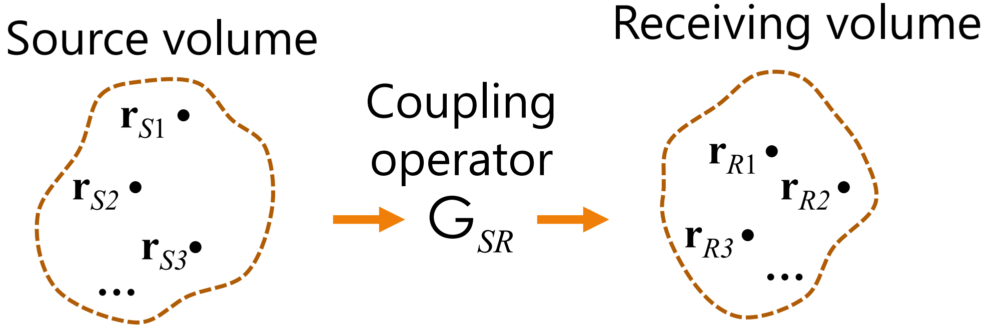
We can see how this works first for a finite number of point sources and receivers

e.g., "loudspeakers" at positions $\mathbf{r}_{S1}$, $\mathbf{r}_{S2}$, $\mathbf{r}_{S3}$, etc., in the source volume

and "microphones" at positions $\mathbf{r}_{R1}$, $\mathbf{r}_{R2}$, $\mathbf{r}_{R3}$, etc., in the receiving volume

There will be some "coupling operator" or Green's function G_{SR} that tells us the wave from any point source

Constructing examples with point sources and receivers



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

For a simple scalar wave, like a monochromatic sound wave in air

the Green's function could be written as

$$G_{\omega}\left(\mathbf{r}_R;\mathbf{r}_S\right) = -\frac{1}{4\pi}\frac{\exp\left(ik\left|\mathbf{r}_R-\mathbf{r}_S\right|\right)}{\left|\mathbf{r}_R-\mathbf{r}_S\right|}$$

This is simply saying that a "unit" point source will generate a spherically expanding wave like this

3 sources and receivers

For these source and receiving points

using the Green's function

$$G_\omega\left(\mathbf{r}_R;\mathbf{r}_S\right) = -\frac{1}{4\pi}\frac{\exp\left(ik\left|\mathbf{r}_R-\mathbf{r}_S\right|\right)}{\left|\mathbf{r}_R-\mathbf{r}_S\right|}$$

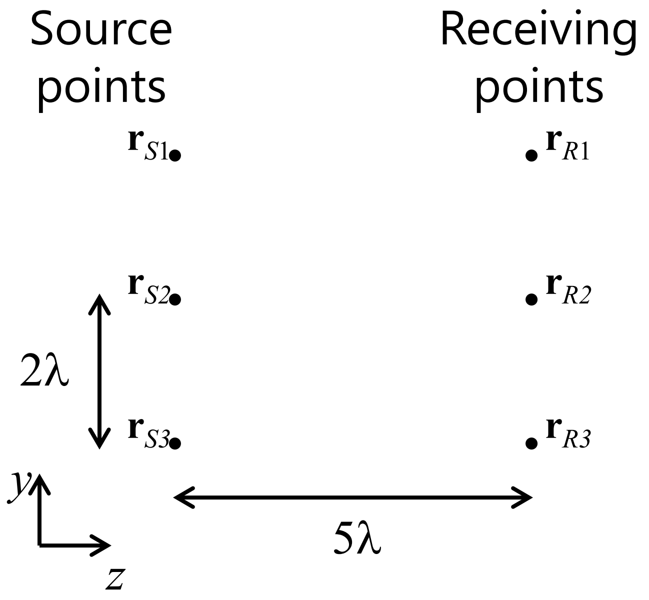
gives a matrix of connections

(for unit wavelength λ)

from each source point

to each receiving point

$$\mathrm{G}_{SR} \cong \frac{-1}{62.83}\begin{bmatrix} 1 & -0.7+0.6i & -0.64+0.45i \\ -0.7+0.6i & 1 & -0.7+0.6i \\ -0.64+0.45i & -0.7+0.6i & 1 \end{bmatrix}$$



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

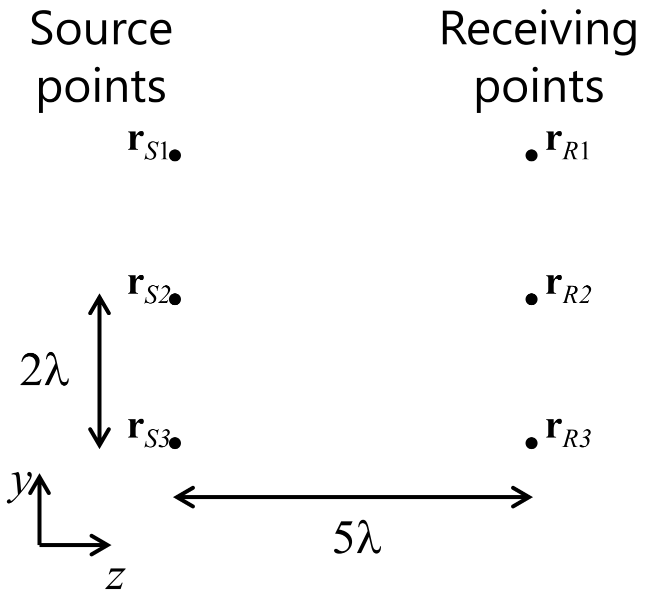
Suppose the source amplitudes at the points $\mathbf{r}_{S1}$, $\mathbf{r}_{S2}$, $\mathbf{r}_{S3}$ were, respectively, a_{S1}, a_{S2}, a_{S3}

We could write those as the column vector

$$|\psi\rangle = \begin{bmatrix} a_{S1} \\ a_{S2} \\ a_{S3} \end{bmatrix}$$

where we have introduced the "Dirac" "bra-ket" notation, here with the "ket" $|\psi\rangle$

as a simple way of writing a column vector that we are labelling as ψ



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

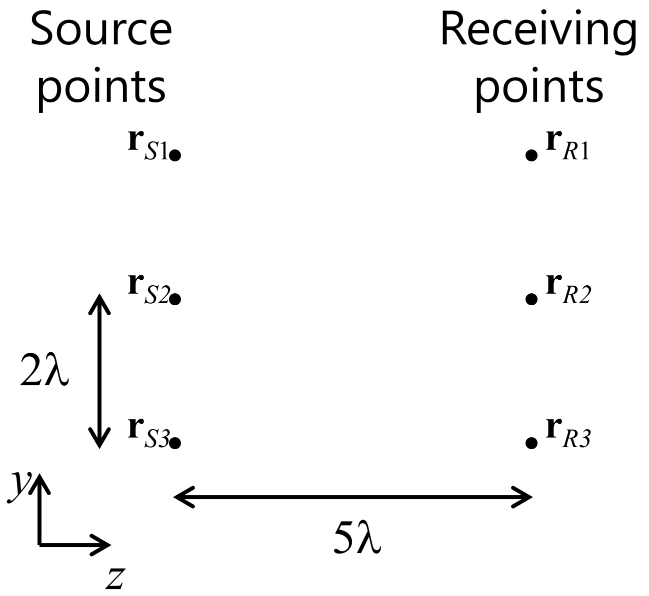
Then the resulting vector of amplitudes

$$|\phi\rangle = \begin{bmatrix} a_{R1} \\ a_{R2} \\ a_{R3} \end{bmatrix}$$

at the receiving points $\mathbf{r}_{R1}$, $\mathbf{r}_{R2}$, $\mathbf{r}_{R3}$

would be given by the matrix-vector product

$$|\phi\rangle = G_{SR} |\psi\rangle$$



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

If we think that a good measure of the arriving power at each point $\mathbf{r}_{R1}$, $\mathbf{r}_{R2}$, $\mathbf{r}_{R3}$ is given, respectively, by $|a_{R1}|^2, |a_{R2}|^2, |a_{R3}|^2$ then the total received power would be

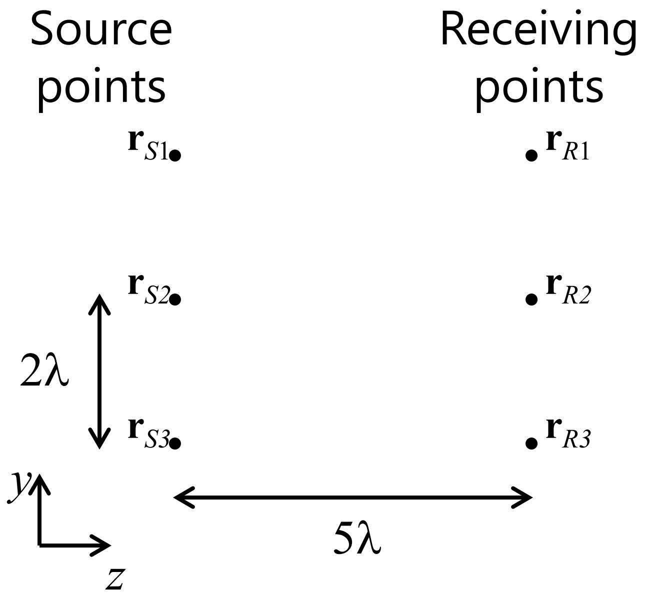
$$P = |a_{R1}|^2 + |a_{R2}|^2 + |a_{R3}|^2 \equiv \begin{bmatrix} a_{R1}^* & a_{R2}^* & a_{R3}^* \end{bmatrix} \begin{bmatrix} a_{R1} \\ a_{R2} \\ a_{R3} \end{bmatrix}$$

Using the Dirac "bra-ket" notation

$$\langle \phi | \equiv \begin{bmatrix} a_{R1}^* & a_{R2}^* & a_{R3}^* \end{bmatrix}$$

where the "bra" $\langle \phi |$ is the Hermitian adjoint (complex conjugate of the transpose) of $| \phi \rangle$

then $P = \langle \phi || \phi \rangle \; (\equiv \langle \phi | \phi \rangle)$



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

But $|\phi\rangle = \mathrm{G}_{SR}|\psi\rangle$

and, by the normal rules of matrix algebra

$$\langle\phi| = \langle\psi|\mathrm{G}_{SR}^{\dagger}$$

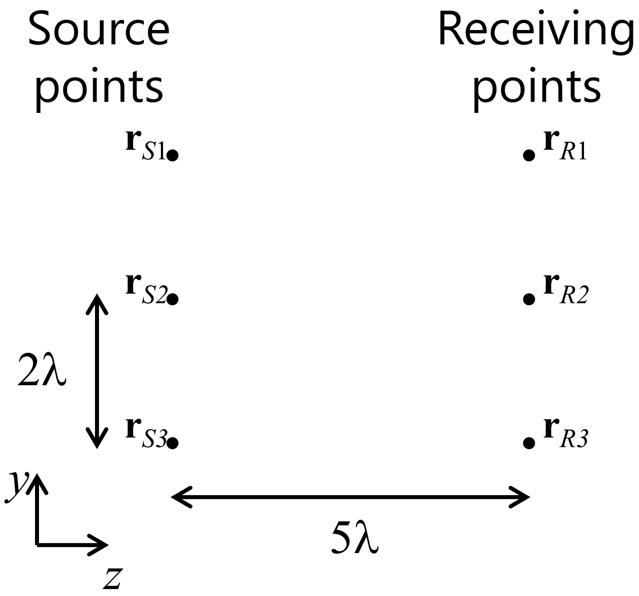
where the superscript "dagger" means we are taking the Hermitian adjoint (conjugate transpose) of the matrix

Hence the received power is

$$P = \langle\phi|\phi\rangle \equiv \langle\phi||\phi\rangle = \langle\psi|\mathrm{G}_{SR}^{\dagger}\mathrm{G}_{SR}|\psi\rangle$$

We want to find the (normalized) source

to give the largest received power



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

Now, the matrix $G_{SR}^{\dagger} G_{SR}$

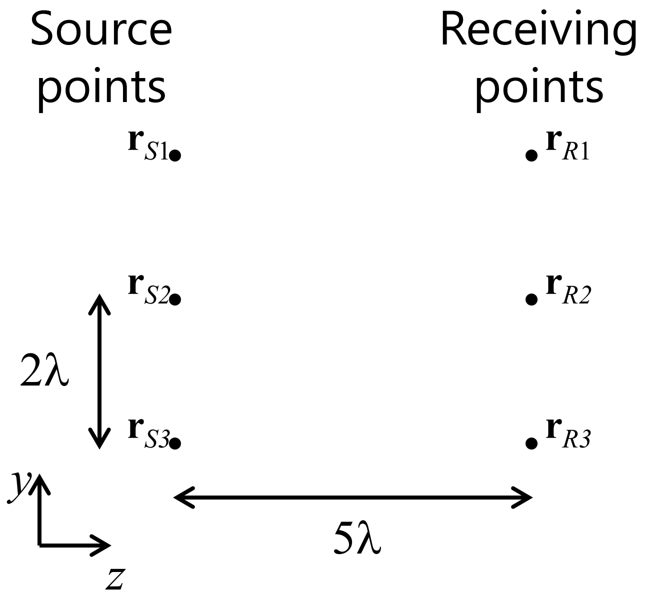
is square, positive, and Hermitian, so has

- positive real eigenvalues, written as $|s_j|^2$
- orthogonal eigenvectors

and the "best" choice of source function

is then simply the first eigenfunction

corresponding to the largest eigenvalue



Solving the eigen problem $G_{SR}^{\dagger} G_{SR} |\psi_{Sj}\rangle = |s_j|^2 |\psi_{Sj}\rangle$

finds a set of orthogonal source functions $|\psi_{Sj}\rangle$

that, when ordered by their eigenvalues, from largest to smallest, give

the set of "best" choices for source functions

in order from best downwards

"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

From these orthogonal source functions

the corresponding resulting wave is

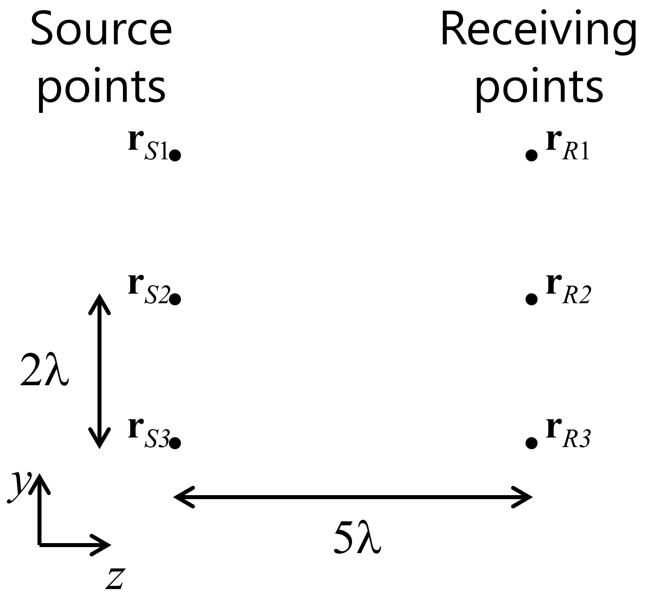
$$\mathrm{G}_{SR}\left|\psi_{Sj}\right\rangle = s_j \left|\phi_{Rj}\right\rangle$$

All these $\left|\phi_{Rj}\right\rangle$ are easily shown to be orthogonal also

and are also the solutions to the complementary eigenproblem

$$\mathrm{G}_{SR}\mathrm{G}_{SR}^{\dagger}\left|\phi_{Rj}\right\rangle = \left|s_j\right|^2 \left|\phi_{Rj}\right\rangle$$

Note these two problems have the same eigenvalues $\left|s_j\right|^2$



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

This process is the **singular value decomposition** of the matrix G_{SR}

The result

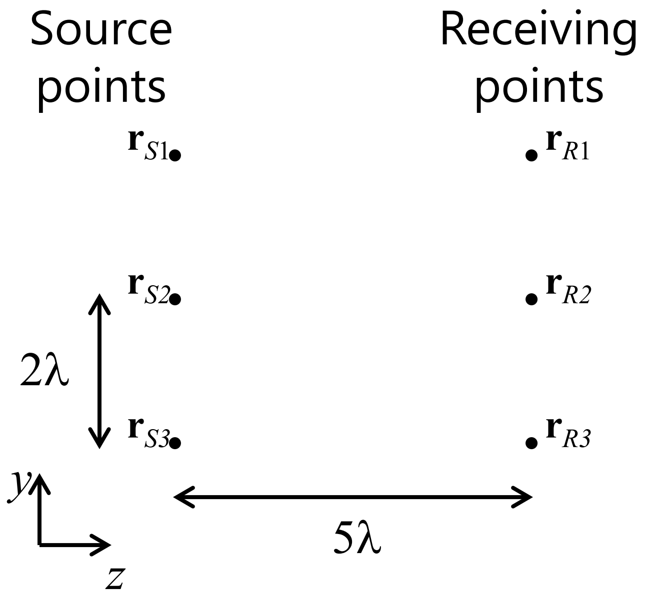
$$G_{SR}\left|\psi_{Sj}\right\rangle = s_j\left|\phi_{Rj}\right\rangle$$

is saying that we have a set of orthogonal source functions

that couple one by one

to a set of orthogonal resulting waves

The (amplitude) coupling strength s_j is the **singular value**



We have established the **communication mode pairs** of functions

"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

3 sources and receivers

Returning to our example problem with three source and receiver points

with

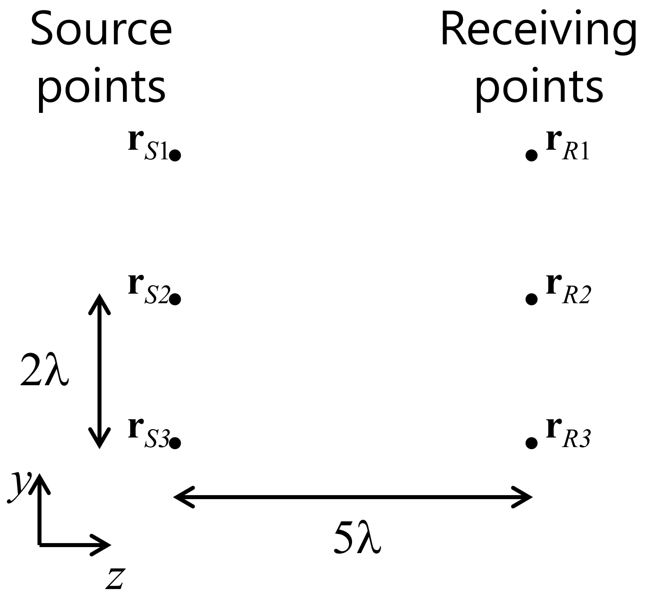
$$G_{SR} \cong \frac{-1}{62.83}\begin{bmatrix} 1 & -0.7+0.6i & -0.64+0.45i \\ -0.7+0.6i & 1 & -0.7+0.6i \\ -0.64+0.45i & -0.7+0.6i & 1 \end{bmatrix}$$

we can now formally find these various functions

Note we can add up the modulus squared of all the matrix elements, giving

$$S = 7.67/(62.83)^2$$

We will return to this number later



Communications modes for 3 sources and receivers

With this matrix, the orthogonal eigenvectors of $G_{SR}^{\dagger}G_{SR}$ are

$$|\psi_{S1}\rangle = \begin{bmatrix} 0.41 \\ -0.81+0.1i \\ 0.41 \end{bmatrix} \qquad |\psi_{S2}\rangle = \begin{bmatrix} -0.71 \\ 0 \\ 0.71 \end{bmatrix} \qquad |\psi_{S3}\rangle = \begin{bmatrix} 0.58 \\ 0.57-0.07i \\ 0.58 \end{bmatrix}$$

and the corresponding eigenvectors of $G_{SR}G_{SR}^{\dagger}$ are

$$|\phi_{R1}\rangle \cong \begin{bmatrix} 0.41 \\ -0.81-0.1i \\ 0.41 \end{bmatrix} \qquad |\phi_{R2}\rangle \cong \begin{bmatrix} -0.71 \\ 0 \\ 0.71 \end{bmatrix} \qquad |\phi_{R3}\rangle \cong \begin{bmatrix} 0.58 \\ 0.57+0.07i \\ 0.58 \end{bmatrix}$$

which in this symmetric problem are the complex conjugates of the source vectors

though that is not generally the case

Communications modes for 3 sources and receivers

$$|\psi_{S1}\rangle = \begin{bmatrix} 0.41 \\ -0.81+0.1i \\ 0.41 \end{bmatrix} \quad |\psi_{S2}\rangle = \begin{bmatrix} -0.71 \\ 0 \\ 0.71 \end{bmatrix} \quad |\psi_{S3}\rangle = \begin{bmatrix} 0.58 \\ 0.57-0.07i \\ 0.58 \end{bmatrix}$$

$$|\phi_{R1}\rangle \cong \begin{bmatrix} 0.41 \\ -0.81-0.1i \\ 0.41 \end{bmatrix} \quad |\phi_{R2}\rangle \cong \begin{bmatrix} -0.71 \\ 0 \\ 0.71 \end{bmatrix} \quad |\phi_{R3}\rangle \cong \begin{bmatrix} 0.58 \\ 0.57+0.07i \\ 0.58 \end{bmatrix}$$

These solutions are essentially unique

There is only one set of such orthogonal channels

The modes are these (complex) drive and receive vectors

not the "beam" in between the sources and receivers

Using a communications mode

A given source vector gives

the relative amplitudes and phases

to drive the three "loudspeakers"

to drive a communications mode channel

$$|\psi_{S1}\rangle = \begin{bmatrix} 0.41 \\ -0.81+0.1i \\ 0.41 \end{bmatrix}$$

And a given receiving vector gives

the relative amplitudes and phases

for adding up the signals from the "microphones"

to receive a given communications mode channel

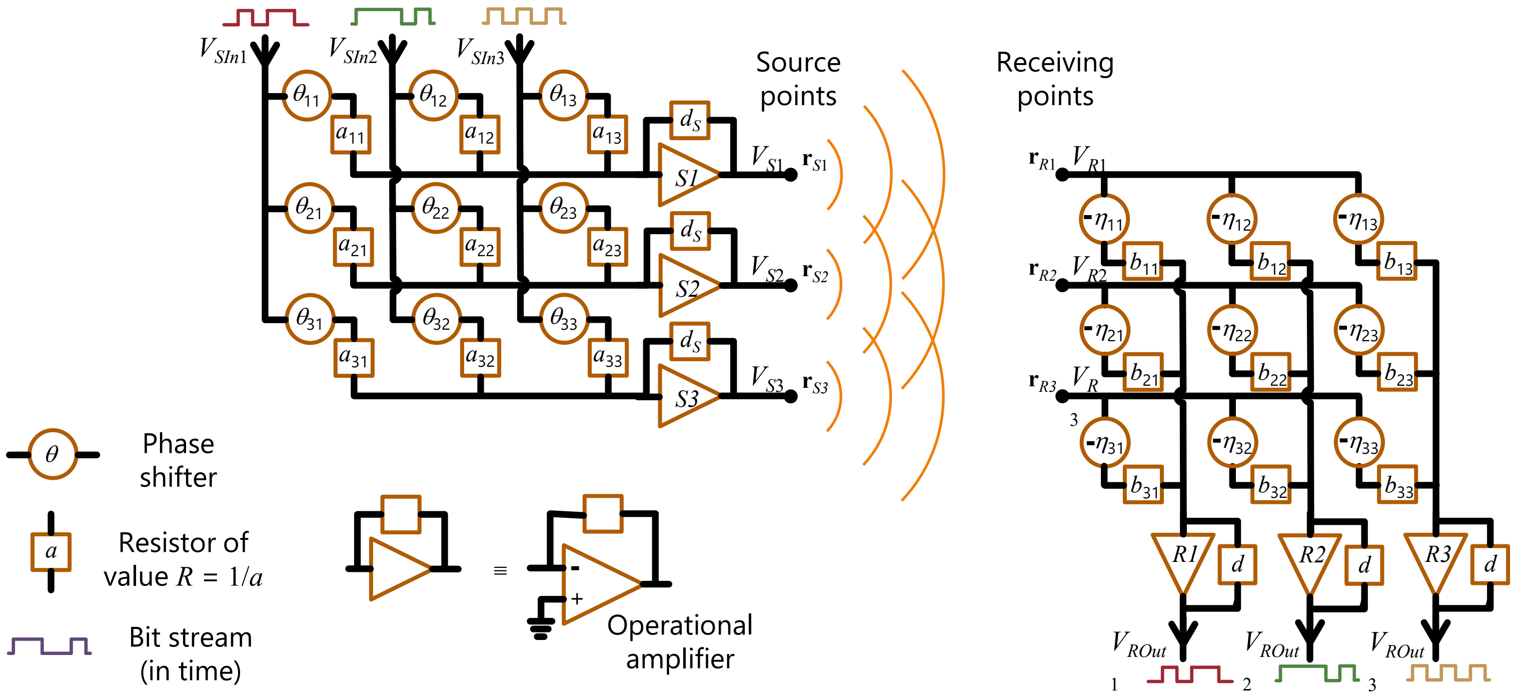
$$|\phi_{R1}\rangle \cong \begin{bmatrix} 0.41 \\ -0.81-0.1i \\ 0.41 \end{bmatrix}$$

Note that, though we use all three sources and all three receiver points for each communication mode or channel

so all source functions and all receiver functions are fully overlapping in their respective spaces

they can all be used simultaneously without cross-talk

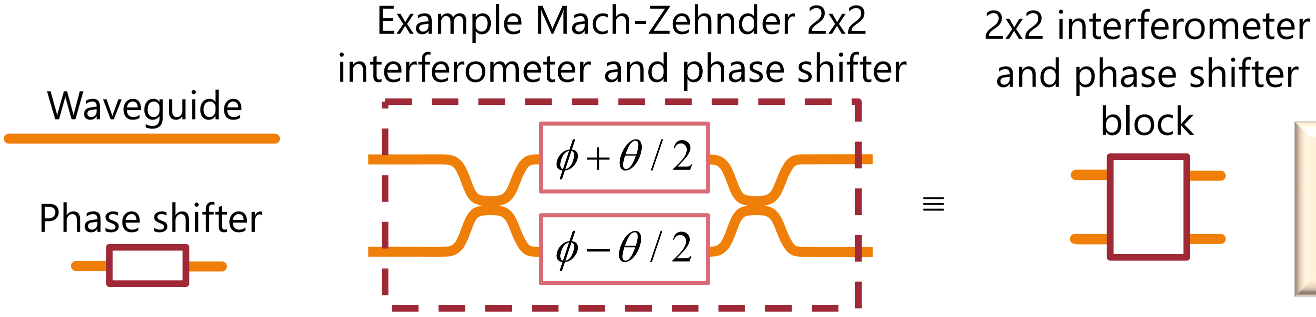
Apparatus to use the 3 channels in acoustics or r.f.



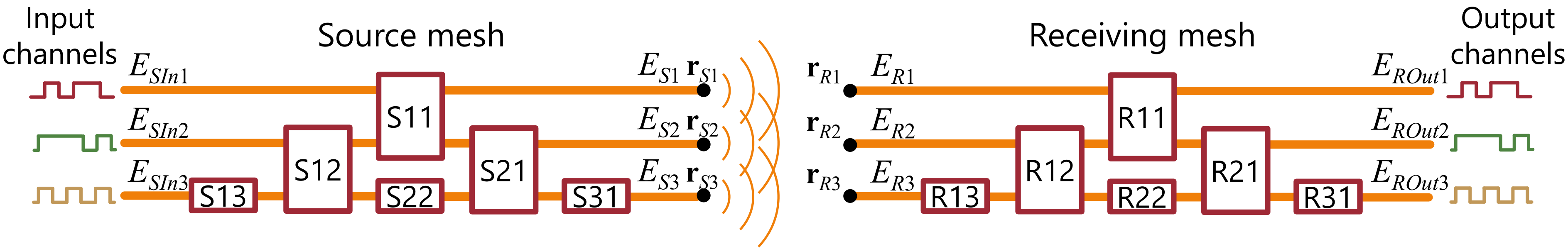
The source network generates the vectors of source amplitudes for each channel

The receiving network separates the channels out again

Apparatus to use the three channels in optics



DM, Optica 7, 794 (2020)
DM, Adv. Opt. Photon. 11, 679 (2019)
DM, Photon. Res. 1, 1 (2013)
DM, J. Lightwave Tech. 31, 3987 (2013)



The source network generates the vectors of source amplitudes for each channel

The receiving network separates the channels out again

Singular value decomposition (SVD)

The mathematical process of singular value decomposition

is more commonly thought of just as a factorization of a matrix

For any linear operator D

at least as long as it is bounded, i.e., finite output for finite input

we can perform the singular value decomposition

$$\mathrm{D} = \mathrm{V}\mathrm{D}_{diag}\mathrm{U}^{\dagger} \quad \text{or equivalently} \quad \mathrm{D} = \sum_m s_m |\phi_m\rangle\langle\psi_m|$$

U and V are unitary operators ($\mathrm{U}^{\dagger}$ is automatically also unitary)

D_{diag} is a diagonal operator with elements s_m

which are called the singular values

$|\psi_m\rangle$ are the columns of U (and $\langle\psi_m|$ are the rows of $\mathrm{U}^{\dagger}$)

$|\phi_m\rangle$ are the columns of V

A sum rule

Note that, for the matrix elements g_{ij} of D

evaluated on any orthonormal basis sets

the sum of the $|g_{ij}|^2$ is the same as the sum of the $|s_q|^2$

and we can usefully write this as the **sum rule** S

$$S = \sum_q |s_q|^2 = \sum_i \sum_j |g_{ij}|^2$$

This sum rule is important below for many reasons

It can be evaluated without solving the problem

and it gives a limit on the number and strength of connections

Singular values and the sum rule

The modulus squared of the singular values are the "power" coupling strengths

For our 3 sources and receivers are

$$|s_1|^2 \simeq \frac{3.41}{62.83^2} \quad |s_2|^2 \simeq \frac{2.89}{62.83^2} \quad |s_3|^2 \simeq \frac{1.37}{62.83^2}$$

so the channels are not all equally strongly coupled

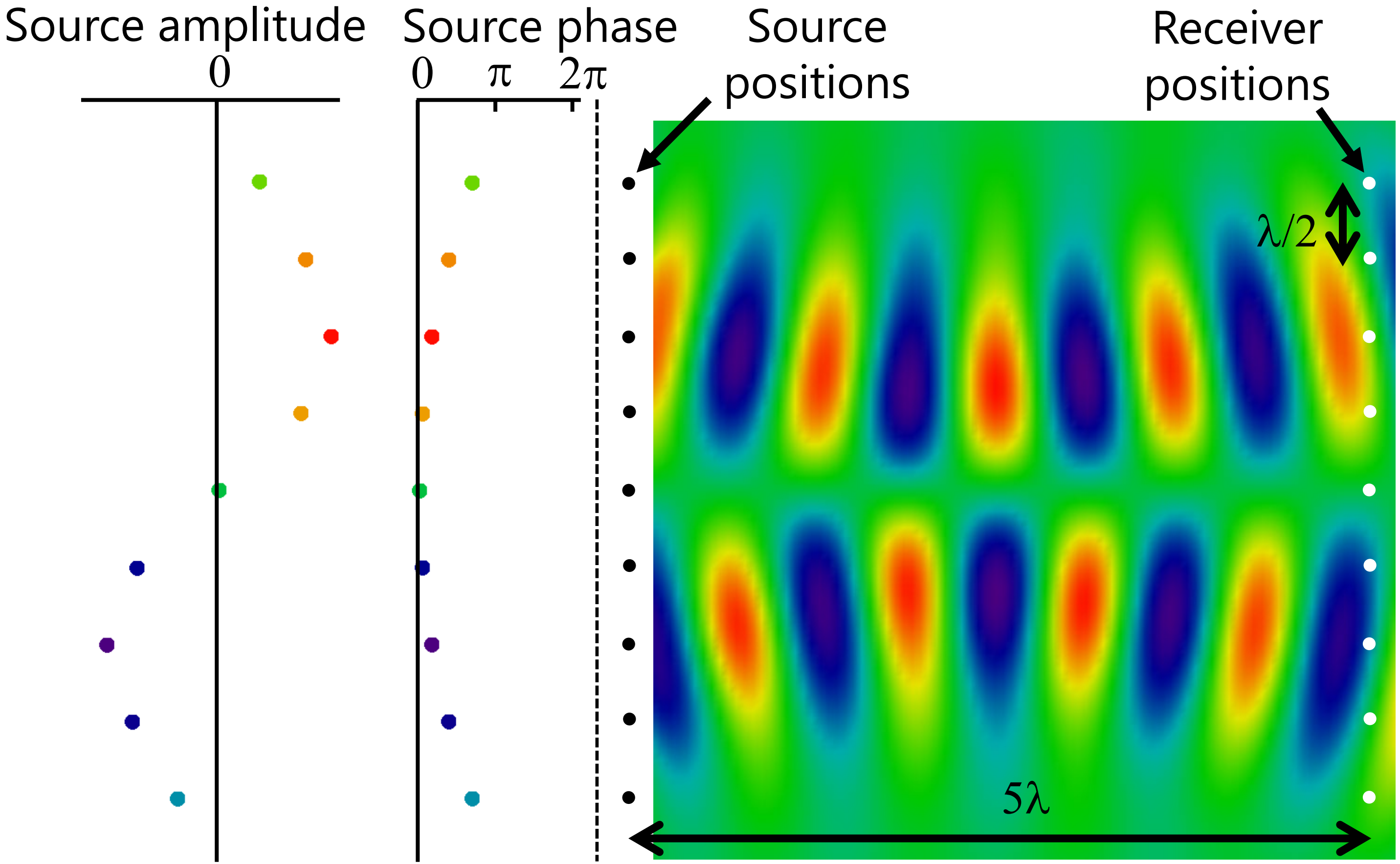
Note that

$$|s_1|^2 + |s_2|^2 + |s_3|^2 \cong \frac{3.41}{62.83^2} + \frac{2.89}{62.83^2} + \frac{1.37}{62.83^2} = \frac{7.67}{62.83^2} = S$$

where S is also the sum of the squares of the matrix elements

We already know the sum of the squares of these coupling strengths before finding the channels

Communication modes with 9 source and receiver points



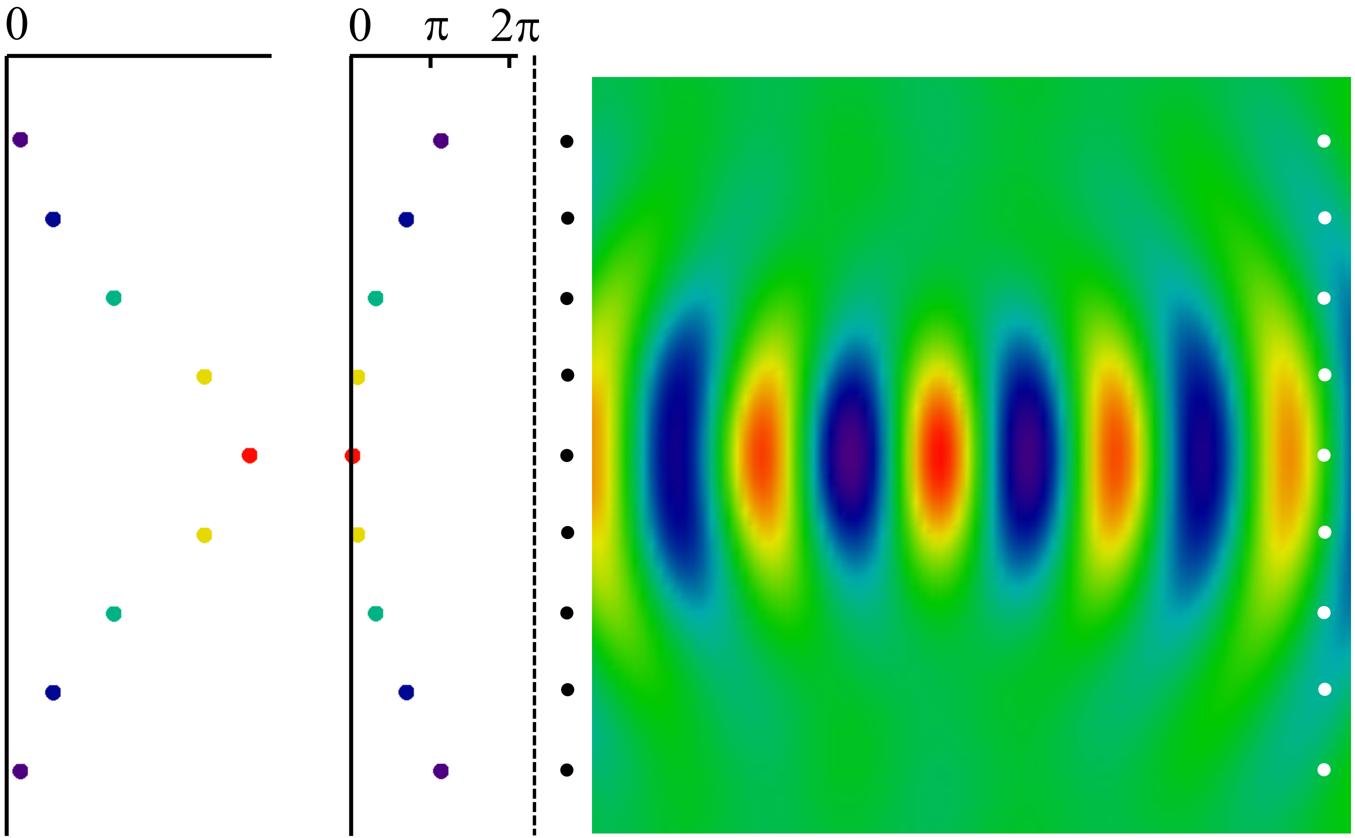
Mode 1

$$|s_1|^2 \simeq 28.54\% \text{ of } S$$

Note the source phases are "curved" to give some focusing

This curvature is "found" by the SVD math

Communication modes with 9 source and receiver points



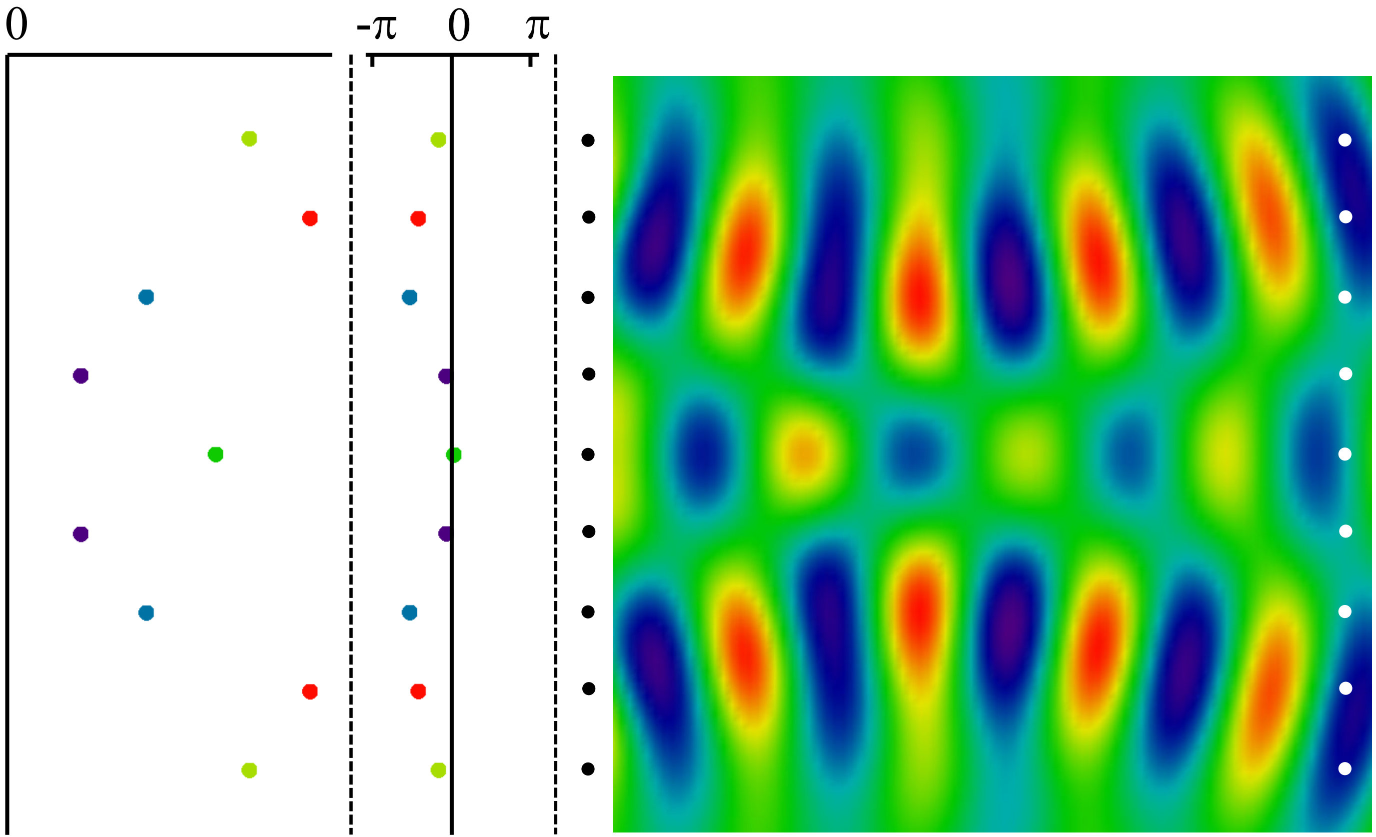
Mode 2

$|s_2|^2 \simeq 28.07\%$ of S

Note the source phases are "curved" to give some focusing

This curvature is "found" by the SVD math

Communication modes with 9 source and receiver points



Mode 3

$|s_3|^2 \simeq 26.28\%$ of S

Note we can begin to see this wave "missing" the receivers

consistent with its slightly lower coupling strength

Communication modes with 9 source and receiver points

Increasing the number of sources and receivers to 9

gives 9 orthogonal channels

but not 9 good communications channels

We have 3 good channels, 2 weaker though usable ones

and 4 so weak as to be essentially useless

The mode coupling strengths are obeying the sum rule

We have "run out" of sum rule by ~ mode 6 or 7

Increasing the number ("dimensionality") or sources and/or receivers

does not necessarily correspondingly increase the number of usable channels

This sum rule is one reason why

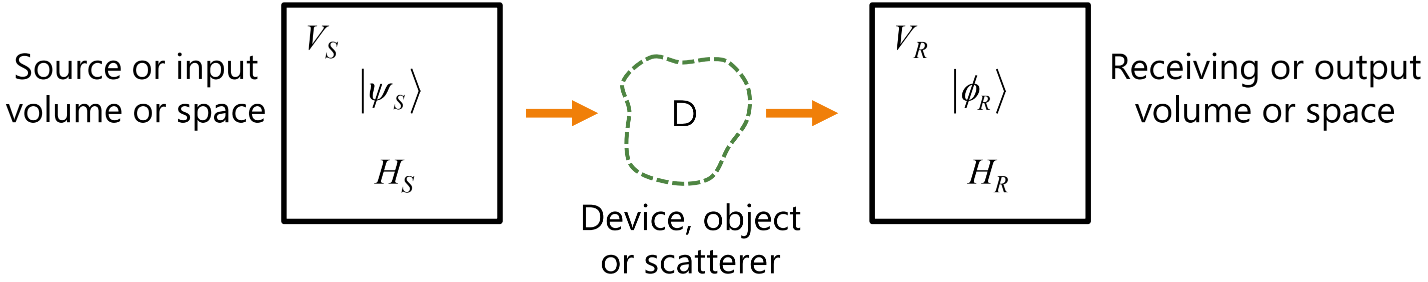
we never have infinite numbers of usable channels in communicating with waves

Mode number, j	% of S	Cum. % of S
1	28.54	28.54
2	28.07	56.61
3	26.28	82.89
4	14.34	97.23
5	2.62	99.84
6	0.16	~100
7	0.0038	~100
8	0.000037	~100
9	0.000000089	~100

Key points about the SVD approach

- This approach is easily extended to full vector fields
  - the underlying mathematics is the same
- We can always think about these problems just using a sufficiently dense set of points in the source and receiving volumes
  - Though it takes some mathematics to prove it
    - requiring functional analysis
  - such approaches do converge to the results for continuous functions
- Note that SVD can be performed for any matrix
  - So, communication modes
    - orthogonal channels
  - can be found for *any* optical system from sources to receivers
    - including ones involving complicated scattering
  - and for any shapes (surfaces, volumes) for sources and receivers

Mode-converter basis sets



One immediate consequence is that

because we can perform the SVD of any linear operator D

For any linear optical system, we have what we can call

the **mode-converter basis sets** of functions

a set of orthogonal source functions

that lead, one by one

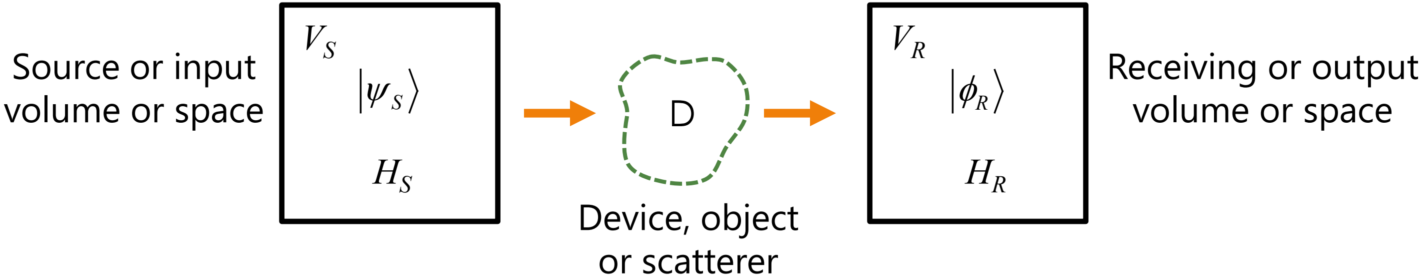
to a set of corresponding orthogonal received waves

We can generalize to consider the source and receiver spaces as

Hilbert spaces, H_S and H_R, of functions

"All linear optical devices are mode converters," Opt. Express **20**, 23985 (2012)

Mode-converter basis sets



In turn, that means that

there is a set of orthogonal channels through any linear scatterer

which are given by these mode-converter input and output function pairs

Decomposing optical systems

We can always perform the singular value decomposition of an optical component or system

So any linear optical system can be described as a mode-converter

"All linear optical devices are mode converters," Opt. Express **20**, 23985 (2012)

These sets of modes turn out to have basic physical significance

"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679-825 (2019)

Waves, modes, communications and optics

“Waves, modes, communications and optics,”
Adv. Opt. Photon. 11, 679-825 (2019)

For any linear optical system

singular value decomposition gives

an optimal, orthogonal set of “input” functions that map, one-by-one, to an optimal orthogonal set of “output” functions

These allow

- A rigorous “communications mode” counting of communications channels
  - including the conclusion that there is always a finite number of usable channels
    - including specific new limits for various optical systems
- A general form of diffraction theory, valid for all sizes and shapes of objects
- The most economical “mode-converter basis” description of any linear optics
- New versions of Kirchhoff’s radiation laws, valid for all objects
  - including nanophotonics and non-reciprocal systems …
- A new, “mode by mode” version of Einstein’s A & B coefficient argument
- A new quantization of the radiation field in any volume
- An understanding of why optics needs thickness, and how much it needs
- The real reason why we can only get so many orthogonal waves in or out of any volume

DM, L. Zhu, and S. Fan, “Universal modal radiation laws for all thermal emitters,” PNAS **114**, 4336 (2017)

Radiation laws from modal optics

This universal "modal" way of looking at optics

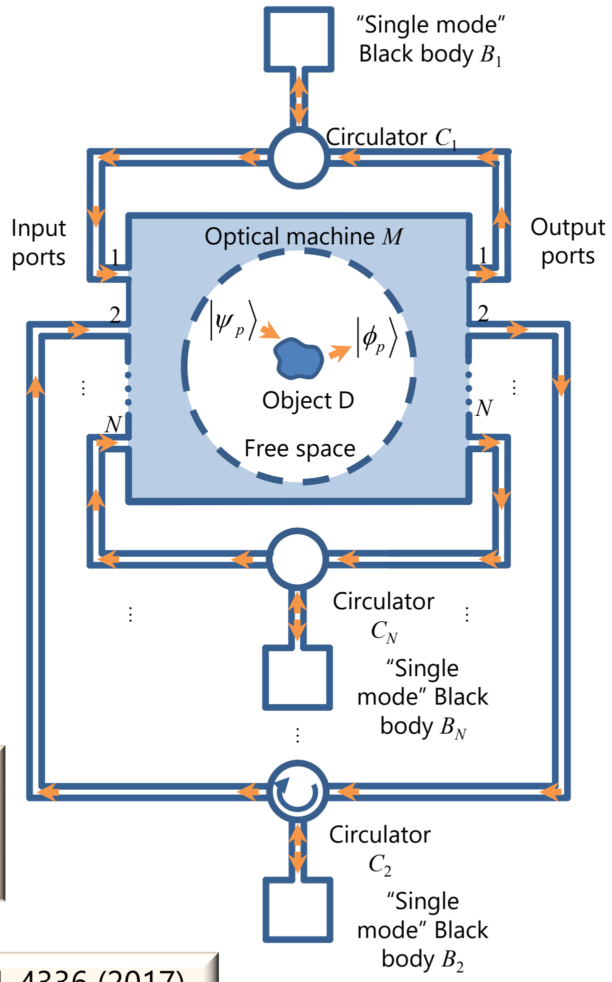
- also allows basic wave results
  - the correct (and modal) Kirchhoff radiation laws for thermal emission

E.g., the absorptivity of an input mode-converter basis function

- equals the emissivity into the corresponding output mode-converter basis function
  - even for non-reciprocal objects

Shows that these mode-converter basis functions

- have basic physical meaning and significance



A "though experiment" machine that leads to modal radiation laws for arbitrary objects, including non-reciprocal ones.

"Universal modal radiation laws for all thermal emitters," PNAS **114**, 4336 (2017)

Structuring light waves in 3D volumes with high precision using communication mode optics

stanford.io/4oZy7bf

Vinicius S. de Angelis, Ahmed H. Dorrah,
Leonardo A. Ambrosio,
David A. B. Miller, and Federico Capasso

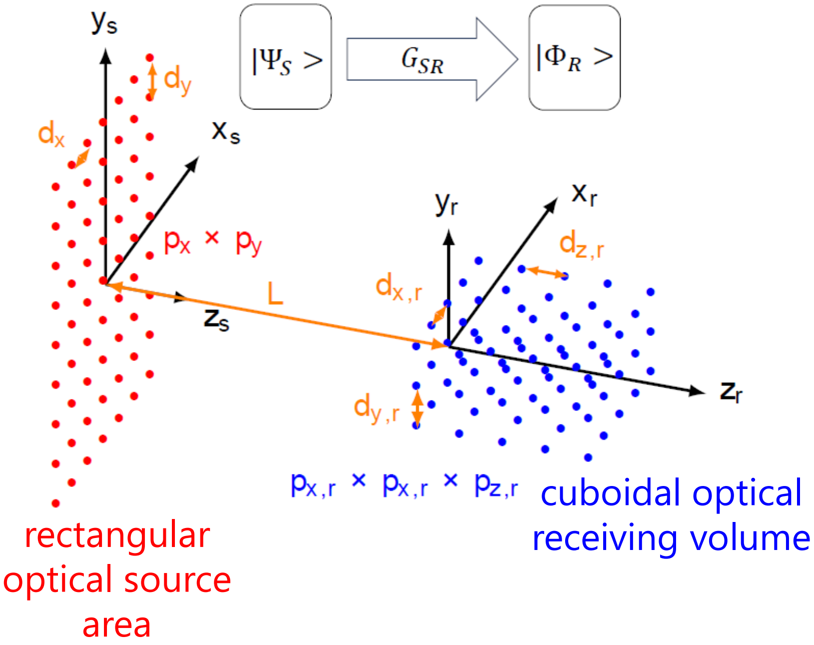
Generating volume fields using communication modes

Note that quite generally

- we can calculate the best "source" output, e.g.,
  - interferometer mesh layer settings
  - spatial light modulator settings
  - metasurface layer
- to generate any volume light field

The "best" way to do this

- is to start by calculating the "communication modes"
  - between the source volume or surface
    - and the "receiving" volume where we want to generate the fields
      - and use those to calculate the necessary source for a given desired volume field



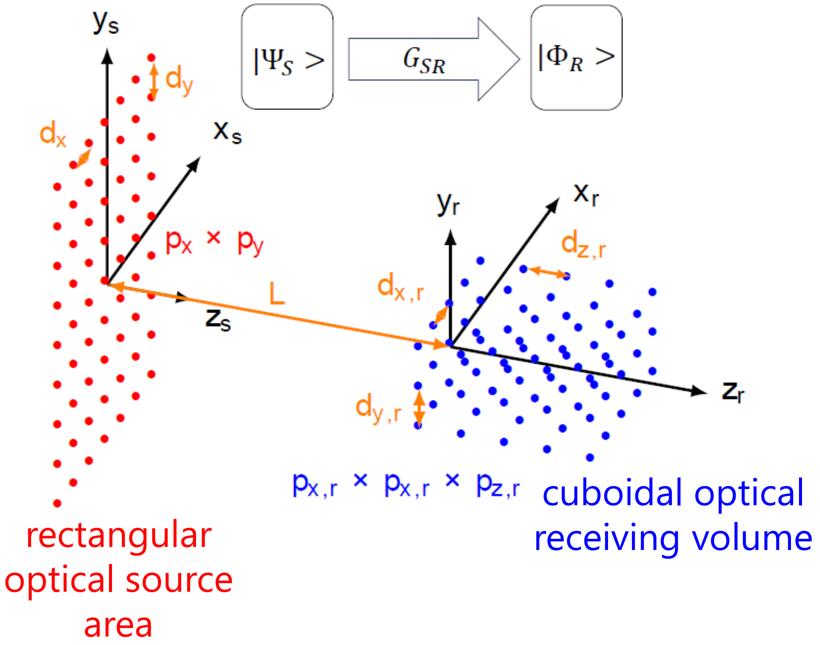
"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

de Angelis et al., "Optimal structured light waves generation in 3D volumes using communication mode optics," arXiv:2411.10865. Optica (to be published)

Generating volume fields using communication modes

This can be done by first approximating both source and receiving spaces by a dense set of point sources and receiver points

- calculating the connection amplitudes
  - between each such source and receiver point pair
    - using the (free-space) Green's function
      - to construct a matrix version of the Green's function



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

de Angelis et al., "Optimal structured light waves generation in 3D volumes using communication mode optics," arXiv:2411.10865. Optica (to be published)

Generating volume fields using communication modes

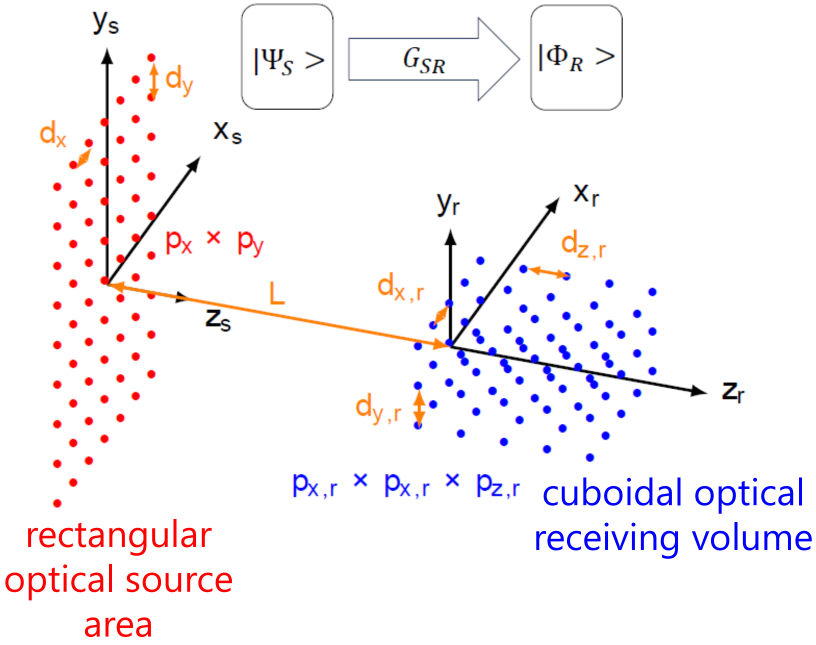
Then perform the singular value decomposition (SVD) of the matrix to give

- the best (and only) set of orthogonal source functions $\left|\psi_j\right\rangle$
  - that connect, one by one, to the best (and only) set of corresponding orthogonal received wave functions $\left|\phi_j\right\rangle$

These are the "communication modes"

- each being a pair of functions, one in each space
  - with a known coupling amplitude s_j between each pair
    - which is the "singular value" that comes out of the SVD calculation

Note we only do this SVD calculation once



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

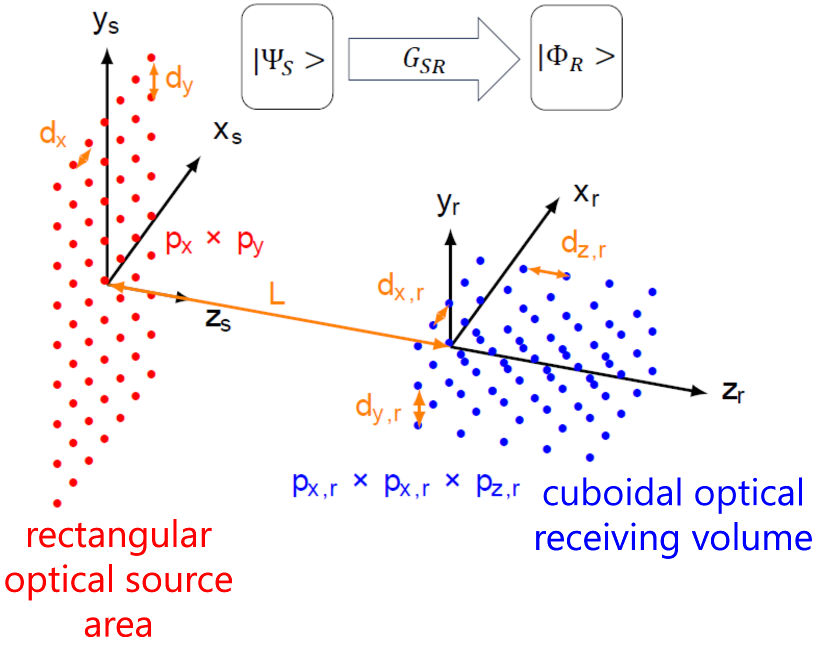
Generating a volume field

For example, we can choose dense sets of points in

- a planar rectangular optical source "volume"
  - which might be a spatial light modulator or a metasurface
- a cuboidal optical receiving volume
  - where we want to construct a field

We can use the free-space Green's function to construct the corresponding matrix G_{SR}

- and perform the SVD of it
  - to get the source and receiving
  - communication mode vectors
    - and coupling strengths (singular values)



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

de Angelis et al., "Optimal structured light waves generation in 3D volumes using communication mode optics," arXiv:2411.10865. Optica (to be published)

Mathematics for generating arbitrary waves with communication modes

Suppose we want a specific wave $|\phi_{Ro}\rangle$ in the receiving space

We expand it in the "receiving" communication modes as $|\phi_{Ro}\rangle = \sum_j a_j |\phi_{Rj}\rangle$

where $a_j = \langle \phi_{Rj} | \phi_{Ro} \rangle$ is the "inner product" or "overlap integral"

Since $G_{SR} |\psi_{Sj}\rangle = s_j |\phi_{Rj}\rangle$

this is the SVD "pairing" between source and received waves

"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

to generate any specific component $a_q |\phi_{Rq}\rangle$ for this expansion

we need an amplitude a_q / s_q of the source function $|\psi_{Sq}\rangle$

So, the required source function $|\psi_{So}\rangle$ to generate $|\phi_{Ro}\rangle$ is

$$|\psi_{So}\rangle = \sum_j \frac{a_j}{s_j} |\psi_{Sj}\rangle \equiv \sum_j \frac{1}{s_j} \langle \phi_{Rj} | \phi_{Ro} \rangle |\psi_{Sj}\rangle$$

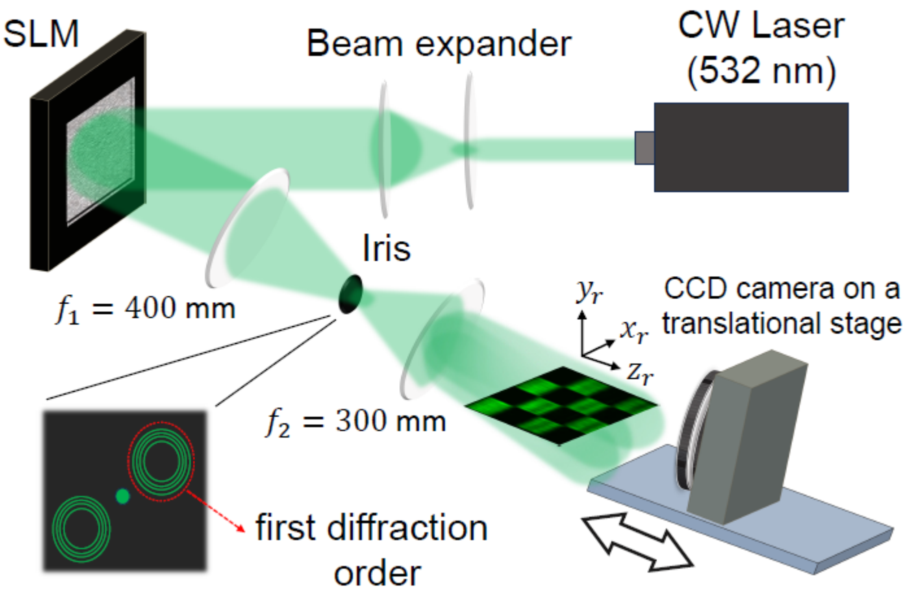
This lets us generate any desired wave in the receiving space

even if the coupling strengths s_j are not the same for every communication mode

Creating a 3-D optical field

Using this approach, we can generate an arbitrary desired 3D field

- calculating the necessary amplitudes for the pixel "point sources"
  - on a spatial light modulator
    - to generate the field of interest in the 3D volume



de Angelis et al., "Optimal structured light waves generation in 3D volumes using communication mode optics," arXiv:2411.10865.
Optica (to be published)

Creating a 3-D optical field

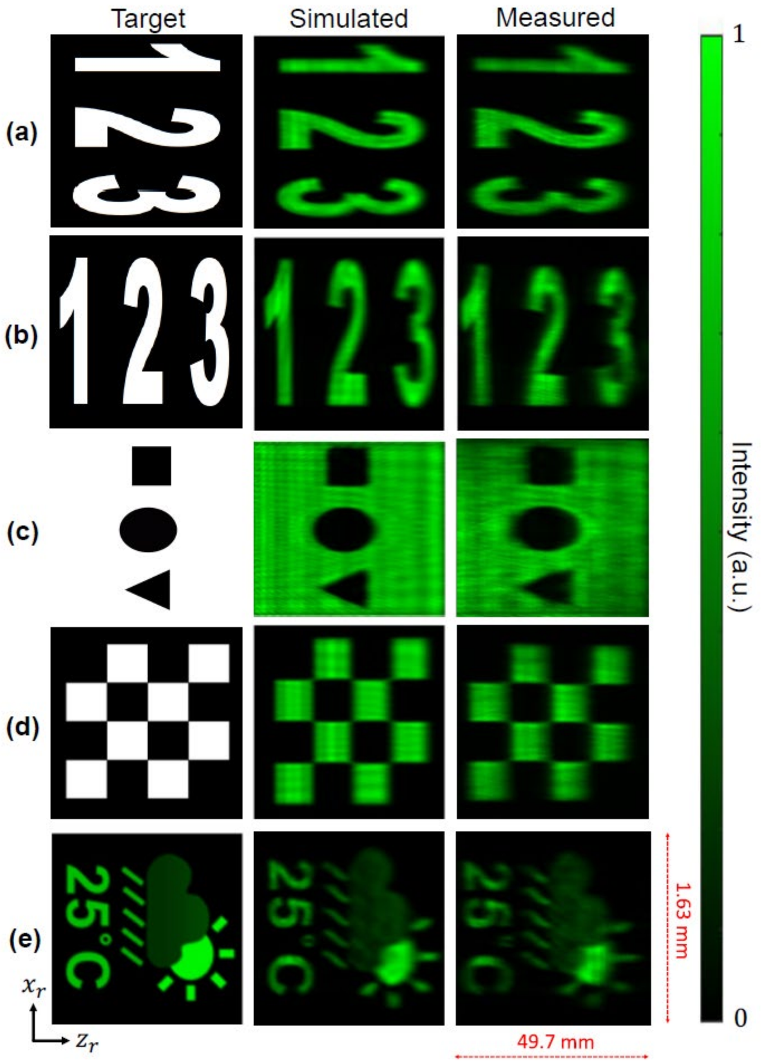
Using this approach, we can generate an arbitrary desired 3D field

calculating the necessary amplitudes for the pixel "point sources"

on a spatial light modulator

to generate the field of interest in the 3D volume

de Angelis et al., "Optimal structured light waves generation in 3D volumes using communication mode optics," arXiv:2411.10865. Optica (to be published)

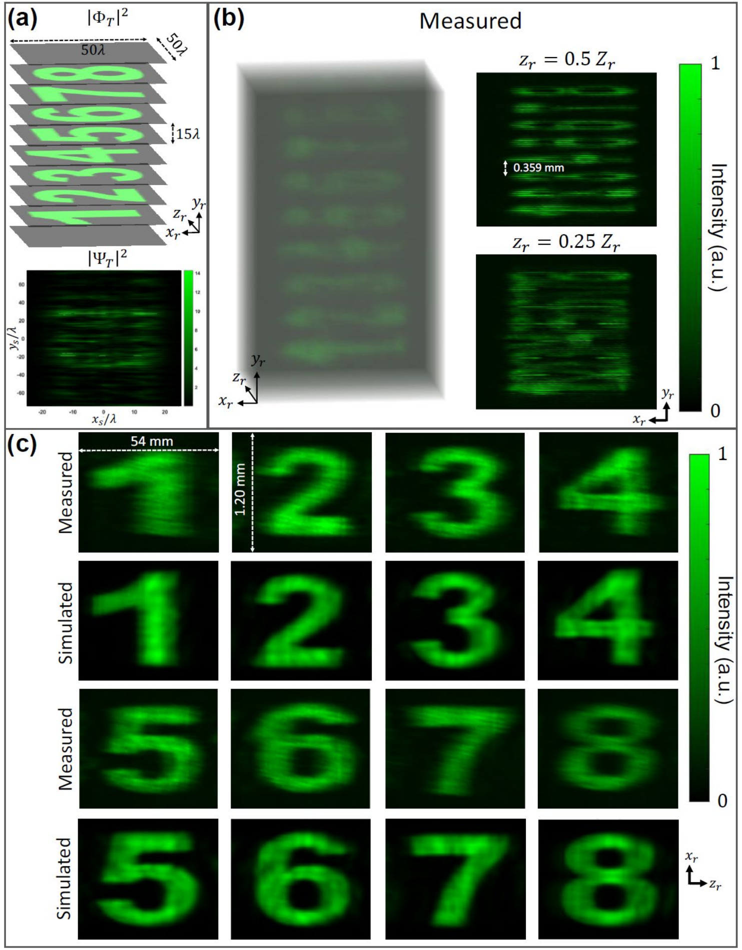


Creating a 3-D optical field

Using this approach, we can generate an arbitrary desired 3D field

- calculating the necessary amplitudes for the pixel "point sources"
  - on a spatial light modulator
    - to generate the field of interest in the 3D volume

de Angelis et al., "Optimal structured light waves generation in 3D volumes using communication mode optics," arXiv:2411.10865.
Optica (to be published)

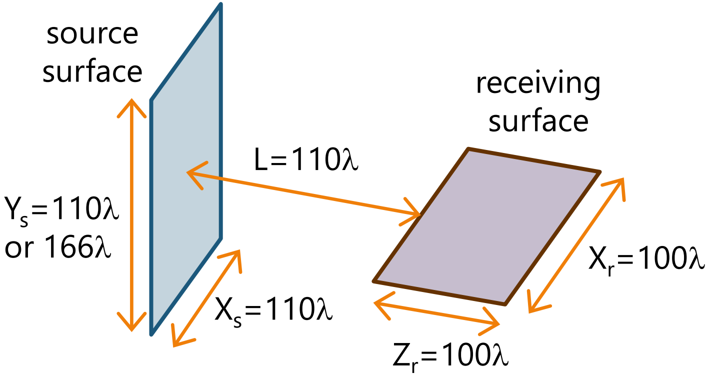


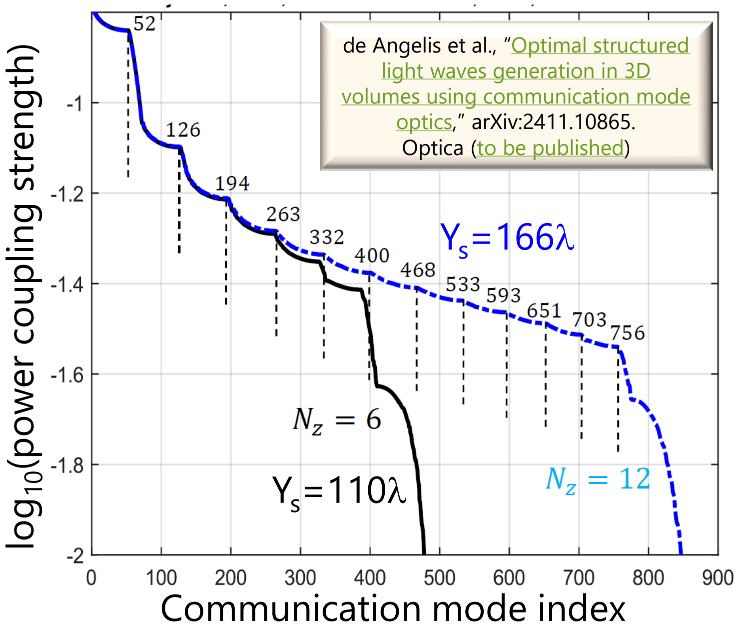
Number of usable modes

We can also understand just what light fields we can create
- because we can only create those that use strongly enough coupled communication modes
  - Beyond that, other fields are impossible

We can estimate how many modes we can have
- based on simple heuristics
- supported by the deeper understanding
  - that waves beyond a given complexity
    - have to tunnel to escape a source
      - Hence the rapid fall-off in coupling strengths (singular values) beyond some number of communication modes

D. A. B. Miller, Z. Kuang, and O. D. Miller, "Tunneling escape of waves," Nat. Phot., Dec. 3, 2024





Waves, modes, and minimum thicknesses for optics

stanford.io/4oZy7bf

DM, "Why optics needs thickness," Science 379, 41 (2023)

Why optics needs thickness

For metasurfaces and metastructures

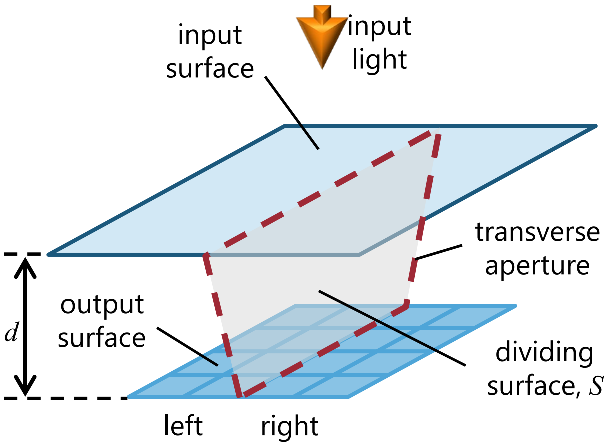
and for compact optics generally

we need to understand whether they need thickness

Can we make a given optical device in just one “layer”, for example?

Generally, no.

But why?



David Miller, "Why optics needs thickness," Science 379, 41 (2023)

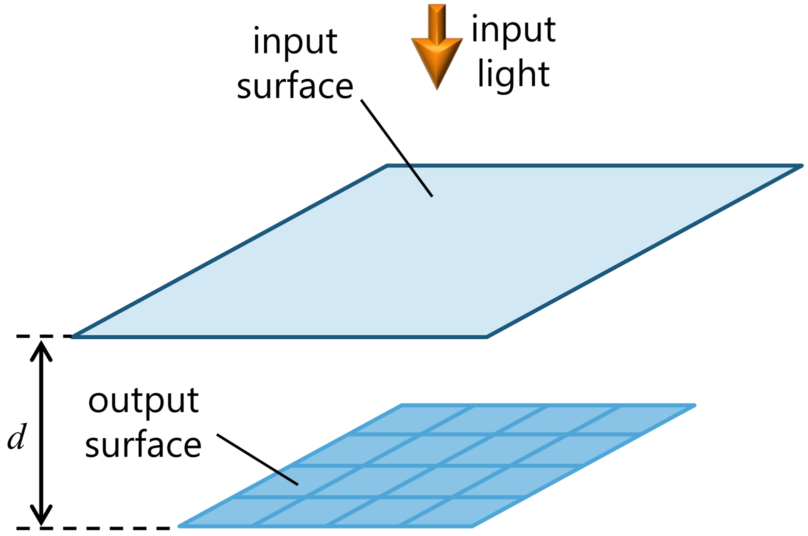
Why optics needs thickness

Think of an optical system with

- an input surface
  - such as a lens surface or metasurface
- an output surface
  - such as an image sensor plane
- with a distance *d* between them

Note we are not yet specifying what is between these two surfaces

- and we will not need to do so



"Why optics needs thickness," Science 379, 41 (2023)

The key idea – channels through a transverse aperture

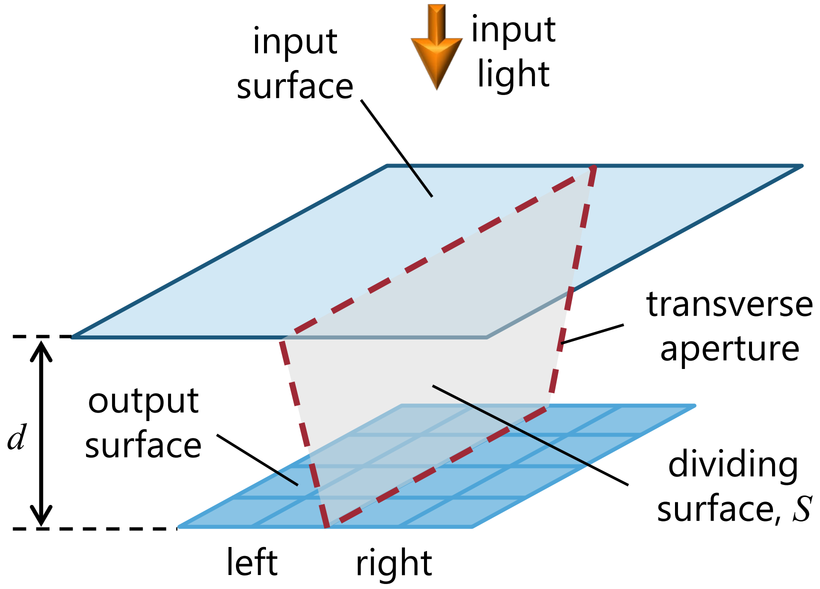
Now imagine we divide each surface in two parts
- left and right
  - by passing an imaginary mathematical dividing surface *S* through them

This defines a "**transverse aperture**"

Because of what we want the system to do
- some number *C* of channels must pass
  - from right to left (or left to right)
    - through this aperture

We call *C* the "**overlapping nonlocality**"

The transverse aperture must be large enough
- for these channels to propagate through it
  - which requires minimum area and/or thickness



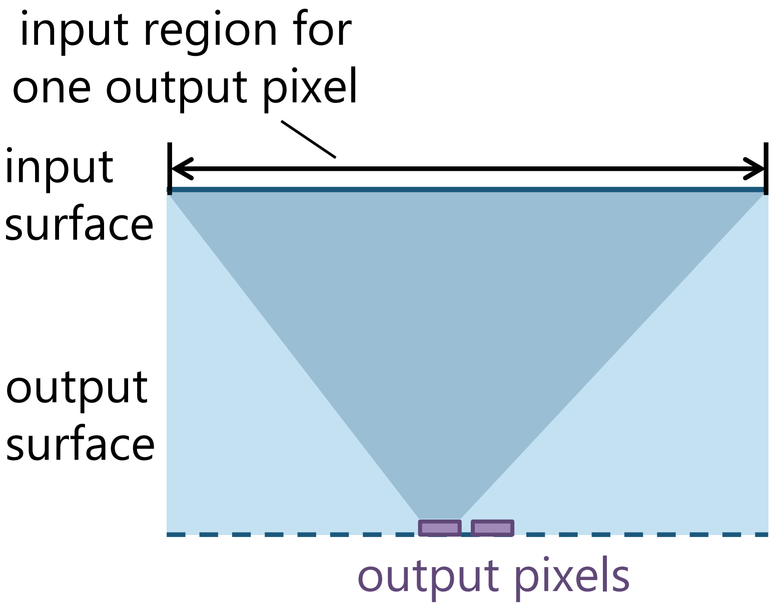
"Why optics needs thickness," Science 379, 41 (2023)

Nonlocality in optics

nonlocality

the output at one point depends on
the input at many points

Imager example



Nonlocality in optics

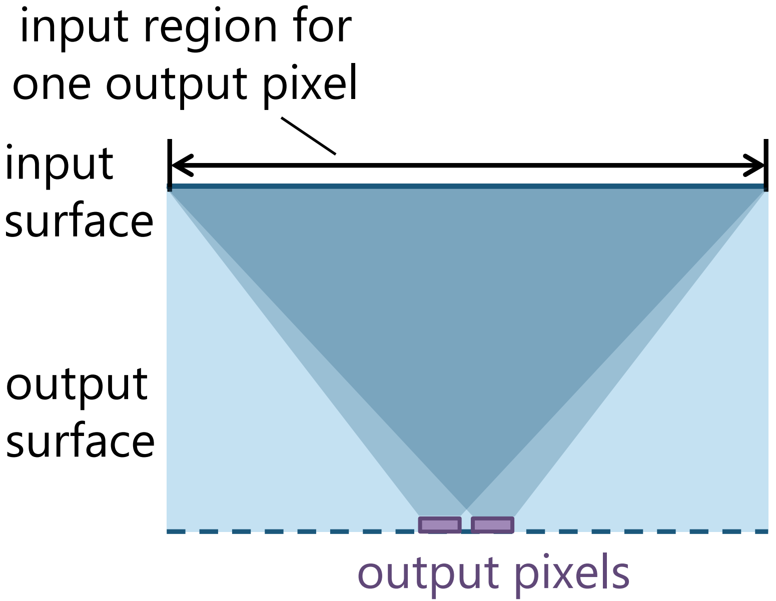
nonlocality

- the output at one point depends on the input at many points

overlapping nonlocality

- the input regions for different output points overlap with one another

Imager example



Nonlocality in optics

nonlocality

- the output at one point depends on the input at many points

overlapping nonlocality

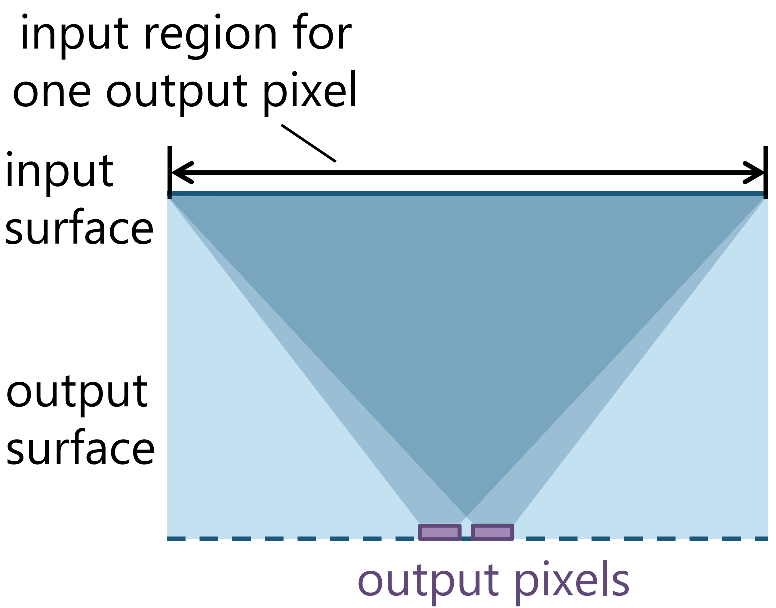
- the input regions for different output points overlap with one another

overlapping nonlocality C

- loosely, the number of such overlapping "channels"

For an imager, C ends up being half the number of pixels

Imager example



Nonlocality in optics

nonlocality

- the output at one point depends on the input at many points

overlapping nonlocality

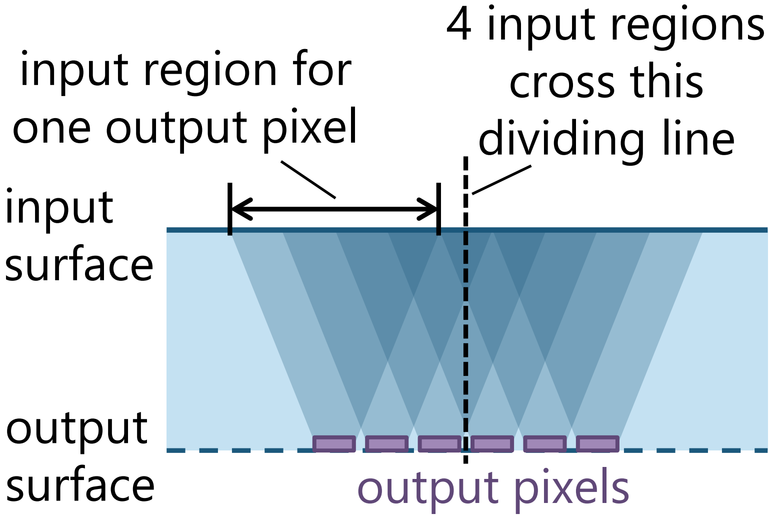
- the input regions for different output points overlap with one another

overlapping nonlocality C

- loosely, the number of such overlapping "channels"

For this example, C is 4

Space-invariant example
e.g., image differentiator



$C = 4$

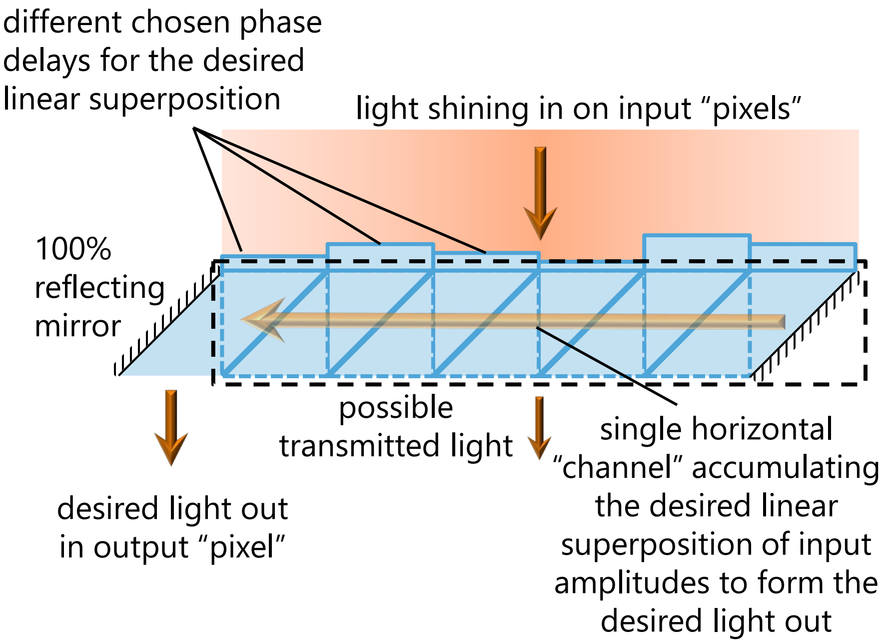
Nonlocality in optics

A system of beamsplitters collects possibly all the light from 6 different input regions

- so, with a “nonlocality” of 6

to only one output “pixel”

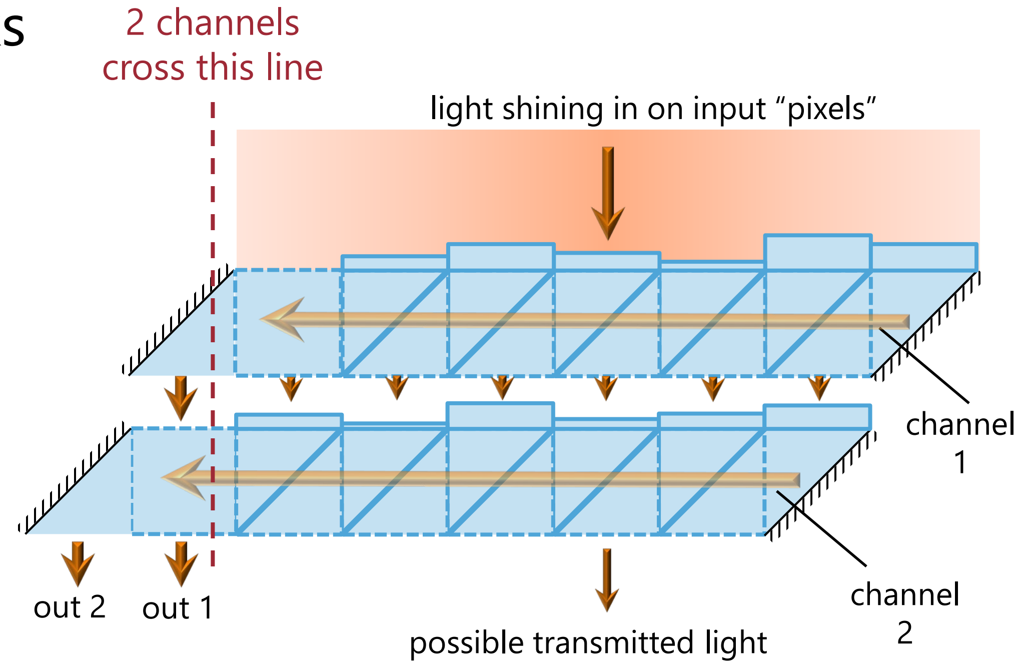
- at the extreme left
  - so, with no overlap in the nonlocality
  - i.e., $C = 1$ “channels”



Nonlocality in optics

Two rows of beamsplitter blocks

- collect two orthogonal 6-element light beams
  - into two separate outputs
    - with an **overlapping nonlocality** $C = 2$



How big a transverse aperture for a given C?

For a 1D system with free-space wavelength λ_o and maximum refractive index n_{max}

we presume we need a thickness

$$\Delta d \geq \lambda_o / 2\alpha n_{max}$$

for each channel

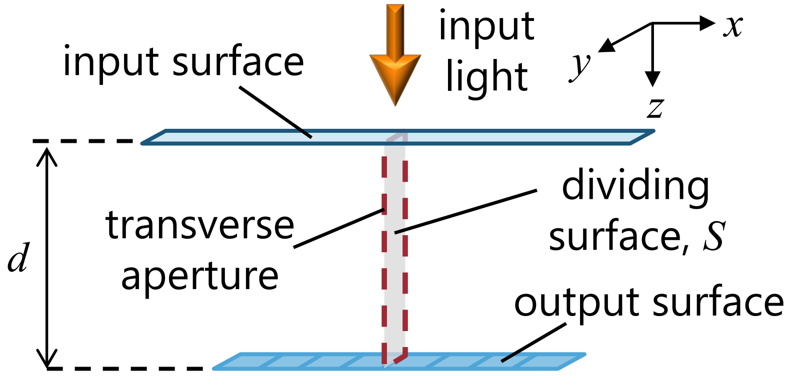
where we allow for some practical factor $\alpha < 1$

which comes from some practical restriction on the range of

usable angles

or usable k-space

inside the device



Thickness of a one-dimensional imager

For some value of C_x in a one-dimensional device

with $\Delta d \geq \lambda_o / 2\alpha n_{max}$ of thickness per channel

then $d \geq C_x \Delta d$

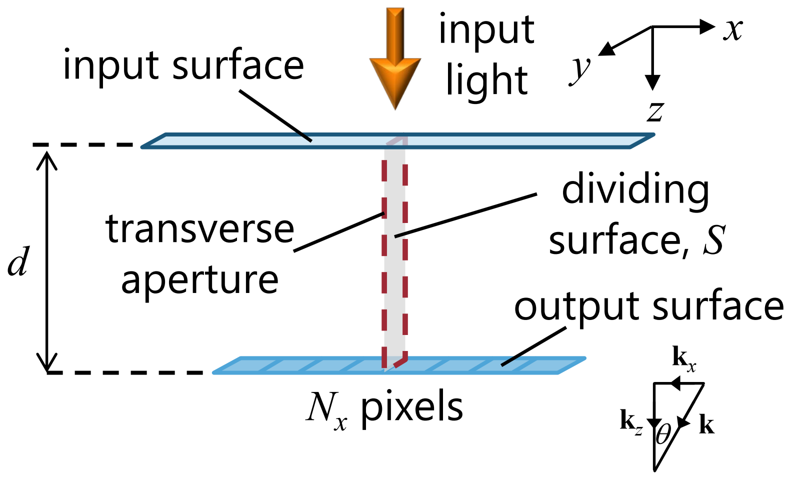
so $d \geq C_x \lambda_o / 2\alpha n_{max}$

For our one-dimensional imager with $C_x = N_x / 2$

$$d \geq N_x \lambda_o / 4\alpha n_{max}$$

For $\lambda_o = 700$ nm, $N_x = 4000$ (one "line" of a 12 MP smartphone camera), $n_{max} = 1.5$ and no "rays" past 45° angle, $d \geq 1.6$ mm

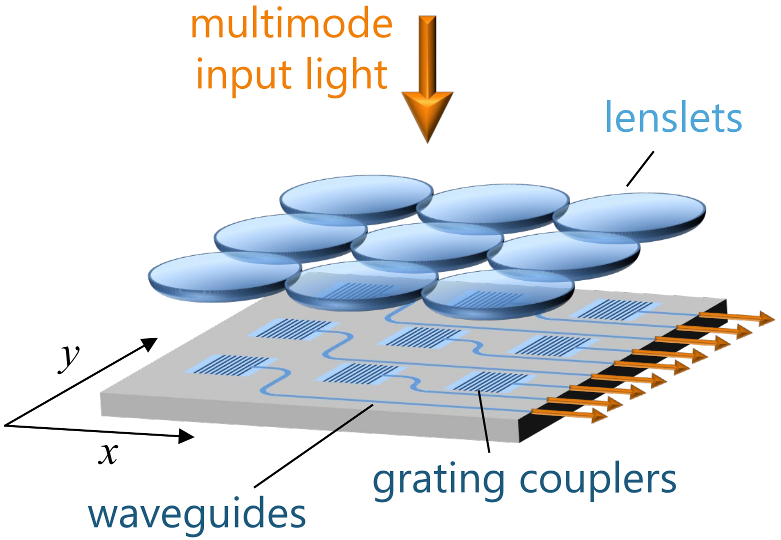
Note: this is the limit for a 2D imager using conventional lenses because of dimensional interleaving restrictions



See V. Blahnik, O. Schindelbeck, Advanced Optical Technologies 10, 145 (2021) for general discussion of modern smartphone cameras

Dimensional interleaving

Can we just "interleave" the channels
taking degrees of freedom that were in x
and interleave them into y?
In principle, yes – the "supercoupler" does this



"supercoupler"
converts 2D input modes
to output modes in a 1D line
e.g., in waveguides

Dimensional interleaving

Can we just "interleave" the channels

taking degrees of freedom that were in x

and interleave them into y?

In practice, this "dimensional interleaving" is much harder

None of the following appear to support dimensional interleaving

- free-space propagation
- conventional imaging systems
- simple dielectric stack structures
- 2-D photonic crystals

In such cases, the thickness of these 2-D systems

may end up as the 1-D limit

with C_x as overlapping nonlocality in the longer, x direction

$$d \geq \frac{C_x \lambda_o}{2\alpha n_{max}}$$

Why optics needs thickness

For the imager example

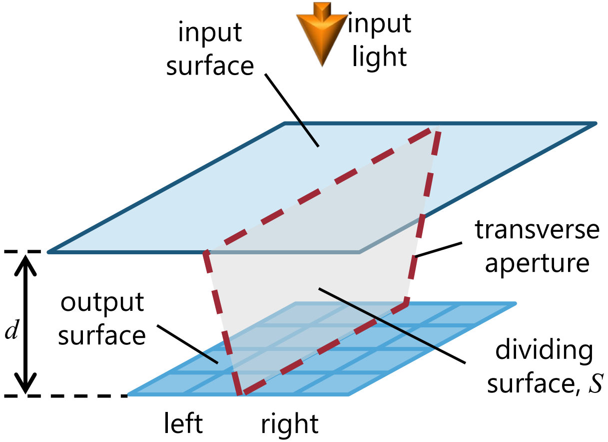
- we used a somewhat heuristic way of counting the necessary channels

The formal way to deduce the necessary thickness

- is to write down the matrix that couples input "pixels" to output "pixels"
  - then perform the singular value decomposition of the coupling between
    - the left input side to the right output side and
    - the right input side to the left output side
      - retaining and counting only the channels that need significant coupling

Once we choose the desired function (matrix)

- the necessary overlapping nonlocality follows
  - regardless of how we implement the device

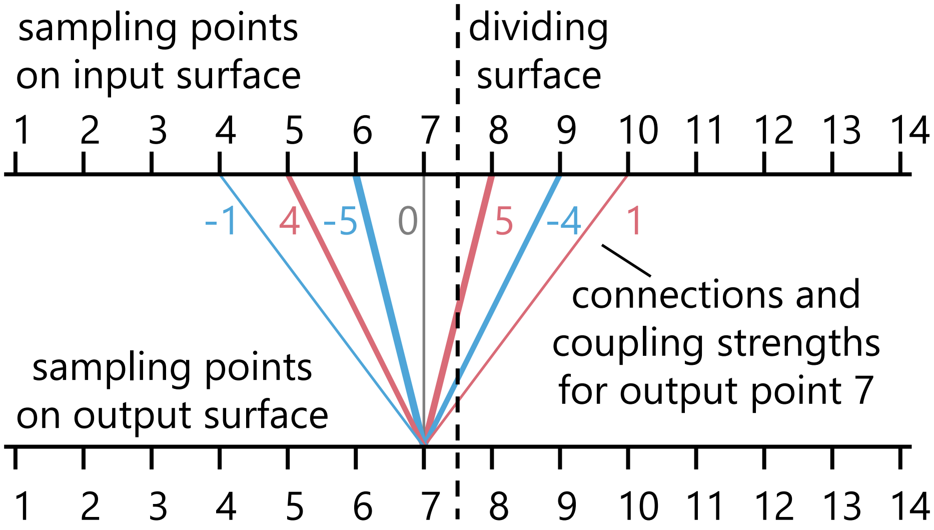


"Why optics needs thickness," Science 379, 41 (2023)

A pixelated differentiator

Consider a 5th order finite difference derivative kernel
- formed from a
  - -1, 4, -5, 0, 5, -4, 1
- weighting of adjacent input points

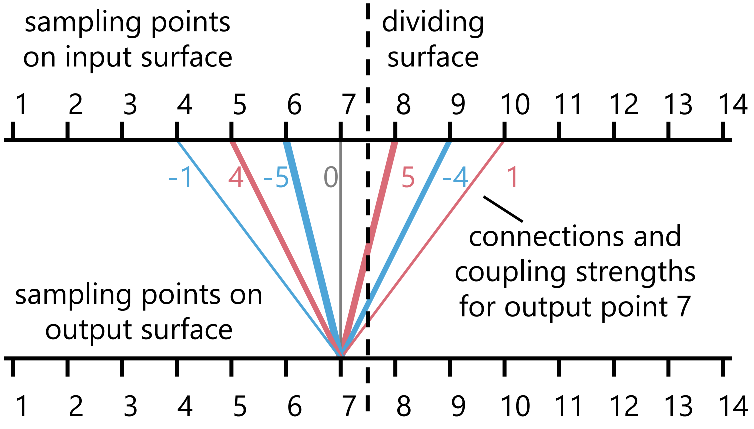
In this case, we can set up a matrix D
- which gives all the connection strengths between inputs and outputs
- for the full "space-invariant" kernel

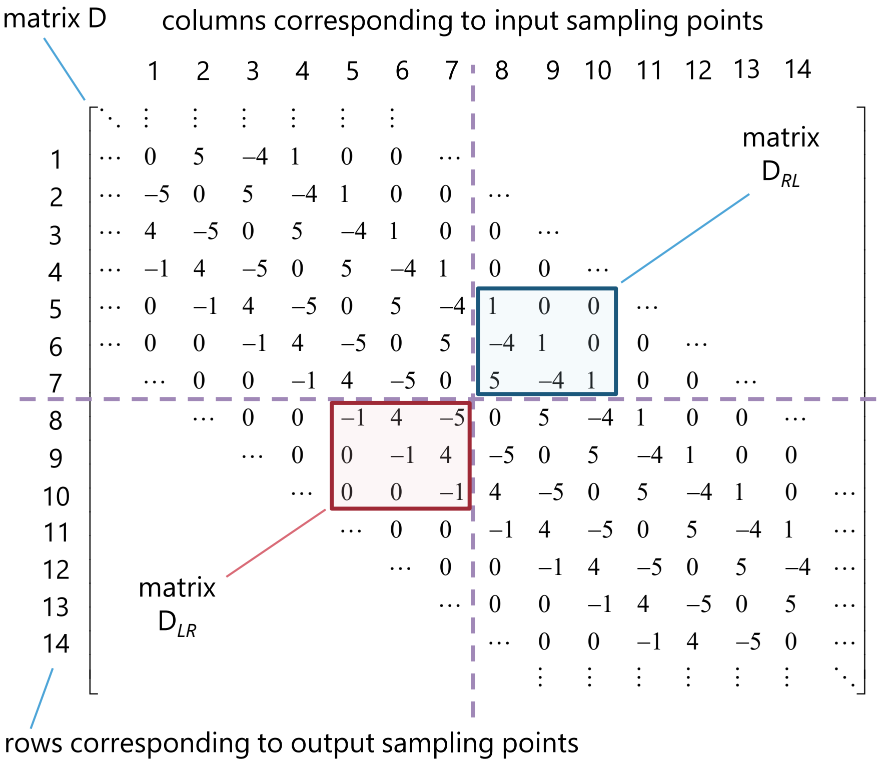


A pixelated differentiator

We can construct the full matrix D of the full "space-invariant" kernel

- arbitrarily choosing one vertical position for the dividing surface
  - between pixels 7 and 8

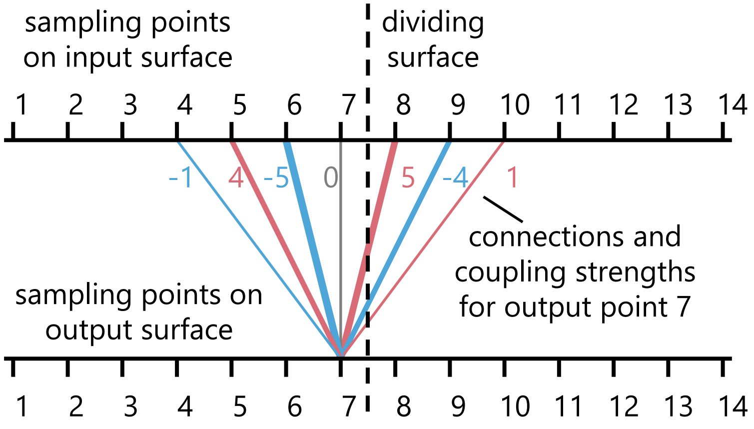


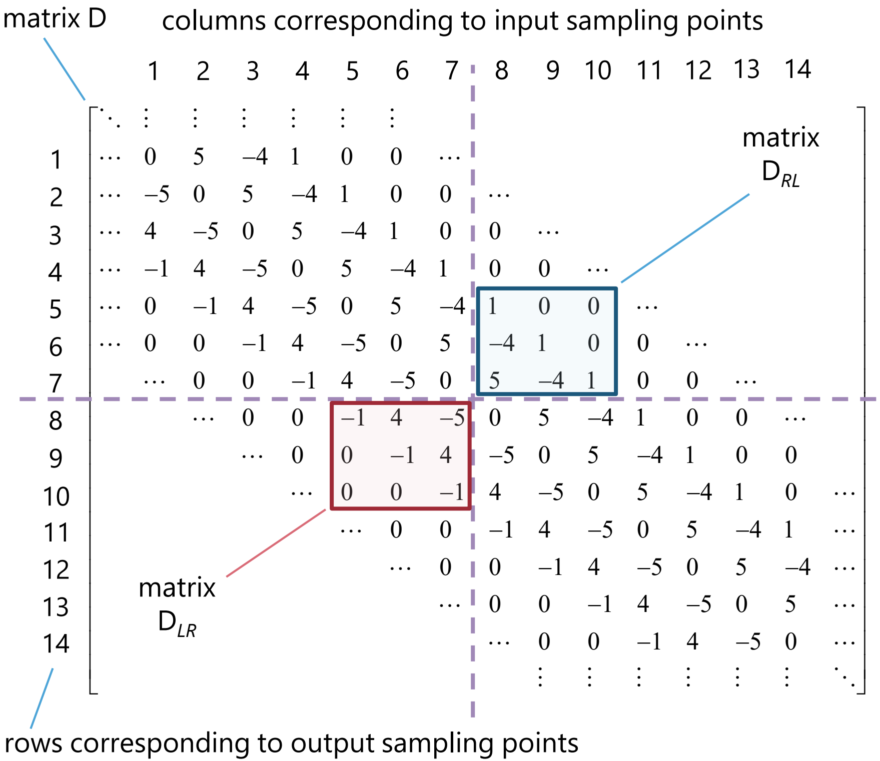


A pixelated differentiator

Sub-matrix D_{RL} gives all the connections
  from the right inputs to the left outputs

Sub-matrix D_{LR} gives all the connections
  from the left inputs to the right outputs





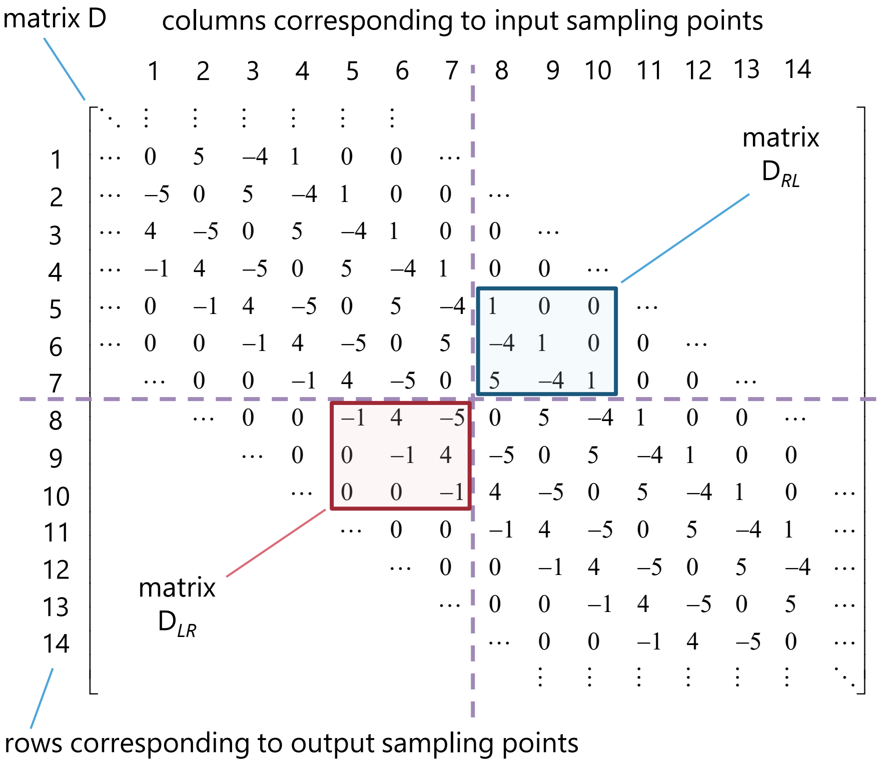
Singular-value decomposition approach

We can count directly as before
- deducing $C = 6$

But with these matrices
- we can take another formal approach -
  - singular value decomposition (SVD) of the matrices D_{RL} and D_{LR}

which gives C_{RL} and C_{LR} as the numbers of singular values of these matrices

Though we don't need this approach here
- we can use this approach for other problems where counting is not so clear



See "Waves, modes, communications, and optics: a tutorial," Adv. Opt. Photon. **11**, 679 (2019) for the SVD approach to optics

Example – metastructure for smoothed derivative

Wang et al. designed a "thick" 2D photonic crystal to perform a smoothed ("Gaussian") derivative with kernel

$$D(u;x) = \frac{(x-u)}{\beta}\exp\left(-\frac{(x-u)^2}{\beta^2\Delta_t^2}\right)$$

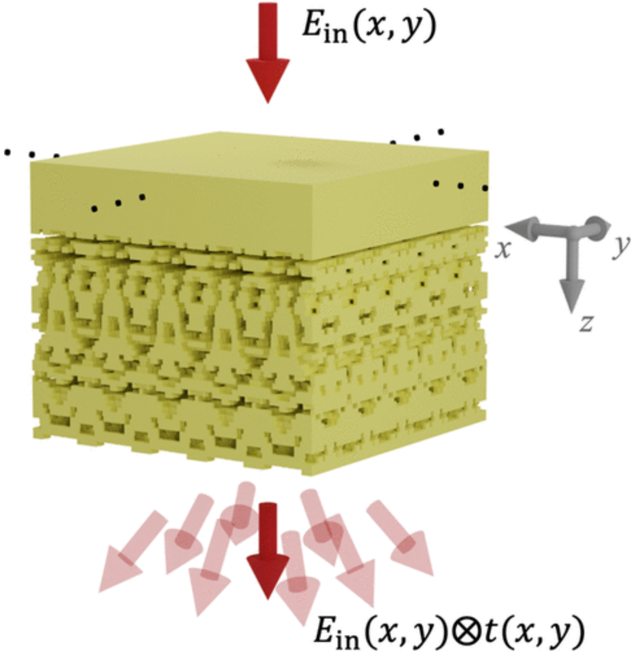
The "divided" kernel has ~ 6 significant singular values

so we should need ~ 6 physical channels through the "transverse aperture"

The thickness of the actual designed structure is ~ 6 wavelengths thick

so more than thick enough at half a wavelength thickness per channel

obeying the proposed (1D) limit here



H. Wang, W. Jin, C. Guo, N. Zhao, S. P. Rodrigues, S. Fan, ACS Photonics 9, 1358–1365 (2022)

Why optics needs thickness

This explains the necessary thickness of, e.g.,

- smart phone cameras, which are within a factor of 3 of this limit
- “space plates” intended to make imagers thinner
- metasurface/metastructure devices as the “kernel” becomes more nonlocal

  e.g., as in image differentiation

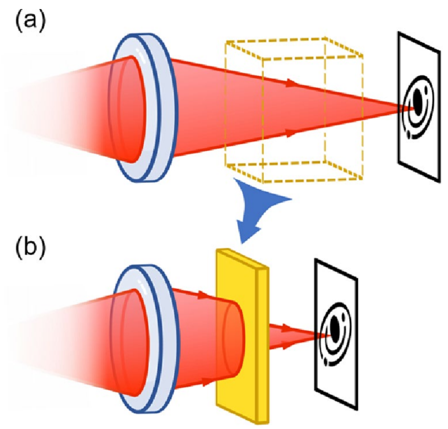
More generally

it clarifies just why optics needs thickness

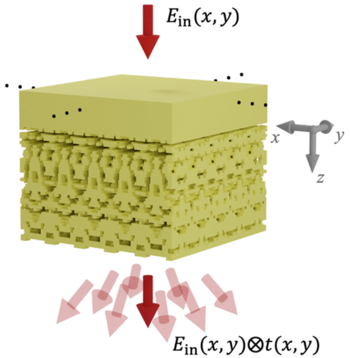
and how much it needs

based formally on an SVD approach

"Why optics needs thickness," Science 379, 41 (2023)



Guo et al., Optica 7, 1133 (2020)



H. Wang et al., ACS Photonics 9, 1358 (2022)

Additional related topics not covered in this talk

- Conservation of “modal étendue”, the number of modes in loss-less optics
- Discussion of non-reciprocal systems
- Extended discussion of dimensional interleaving
- Limits on “supercoupler” dimensions and flat optical systems
- Comparison with other “space-plate” designs
- Effect of displacing the output compared to the input
- “Sampling theory” approach to devices described in k-space
- Other example kernels
  - space-invariant – e.g., Daubechies wavelets, finite impulse response filters
  - space-variant – e.g., Fourier transform optics, interconnection networks
  - extension to complex kernels

"Why optics needs thickness," Science 379, 41 (2023)

Conclusions

The communication mode / mode-converter basis set approach

- lets us rigorously define the channels that describe optical systems
- with many uses in fundamental optics and applications
  - including properly counting channels
  - understanding what fields we can actually generate
  - and what fields can get in and out of arbitrary volumes

Adding in the idea of overlapping nonlocality

- lets us understand broad classes of optical systems
  - including basic requirements on thickness

Meshes of Mach-Zehnder interferometers as universal as self-configuring optics

stanford.io/4oZy7bf

David Miller
Stanford University

Mach-Zehnder interferometer meshes

Mach-Zehnders meshes as universal optical systems

- completely programmable for linear operations
- useful for low to moderate complexity problems
- specific topologies of meshes support
  - self-configuration
    - with simple set-up algorithms
- useful also for thought experiments
  - since they show how universal linear operations can be performed with optics
- fundamentally, they illustrate that
  - any linear operation at a given wavelength can be factorized
    - into successive two-beam interferences

Mach-Zehnder interferometer meshes

Many potential applications

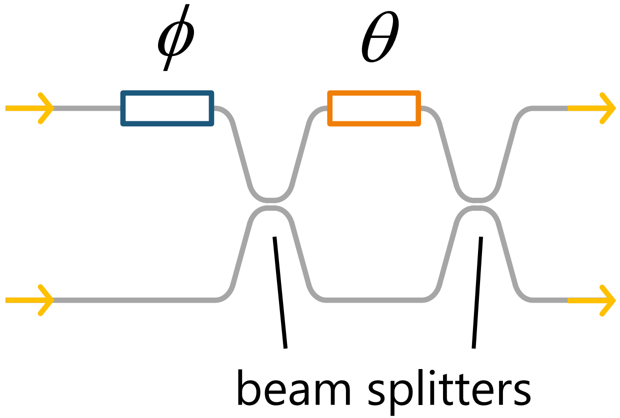
- linear algebra operations like matrix multiplication and inversion
- optical applications
  - self-aligning beam couplers
  - separating overlapping orthogonal beams
  - modal spatial filtering
  - measuring amplitude and phase of optical fields
  - finding the best channels through an optical system
  - analyzing partially coherent light
  - programmable spectral filters

  …

Nulling a Mach-Zehnder output

Consider a waveguide Mach-Zehnder interferometer (MZI)

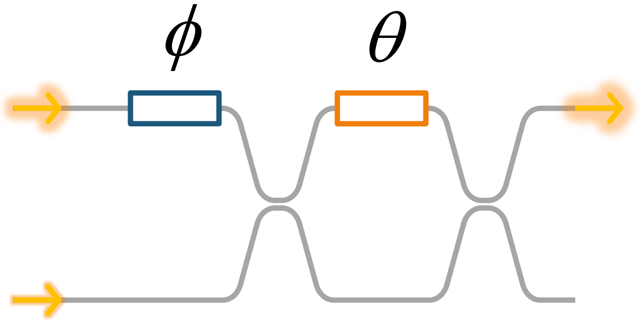
- formed from two "50:50" beam splitters
- and at least two phase shifters
  - one, ϕ, to control the relative phase of the two inputs
  - a second, θ, to control the relative phase on the interferometer "arms"



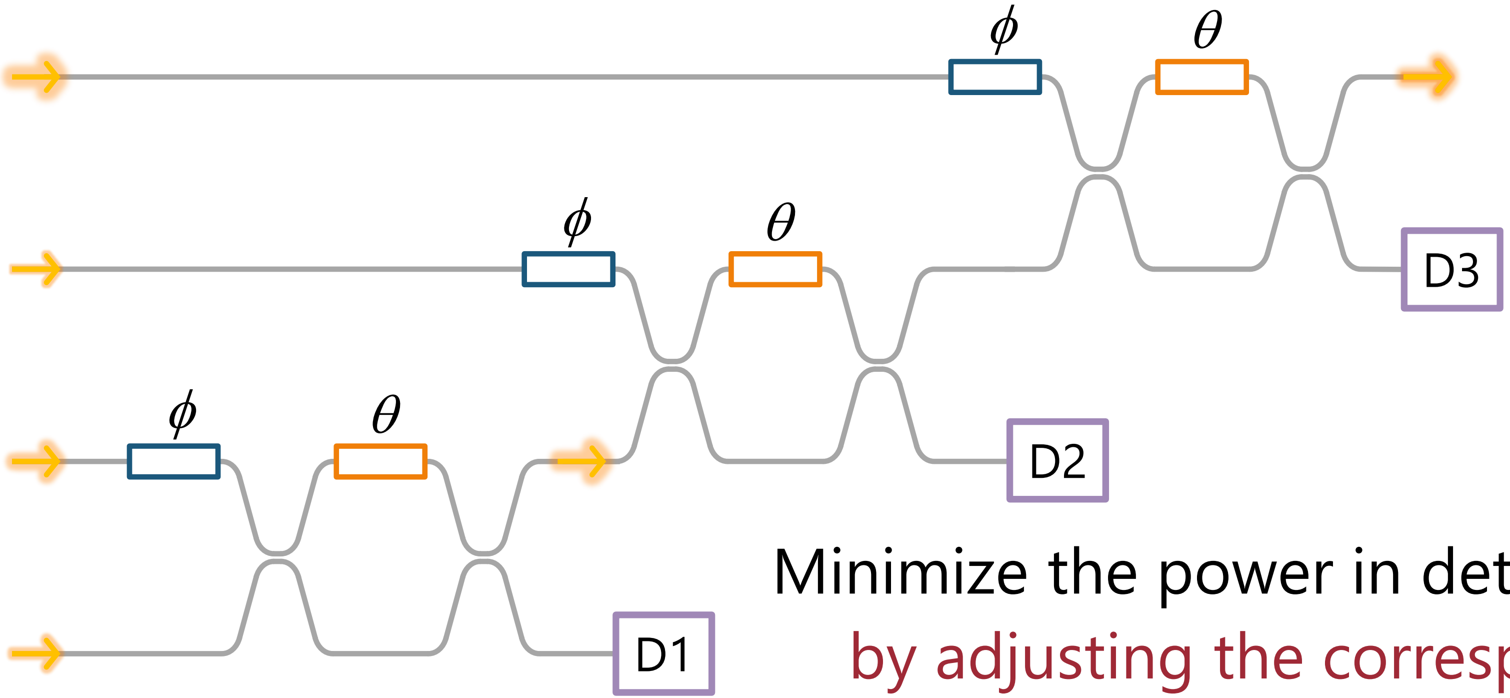
Nulling a Mach-Zehnder output

In such an MZI with 50:50 beamsplitters

- for any relative input amplitudes and phases
  - we can “null” out the power at the bottom output
    - by two successive single-parameter power minimizations
      - first, using ϕ
      - second, using θ



"Diagonal line" self-aligning coupler



Minimize the power in detector D1

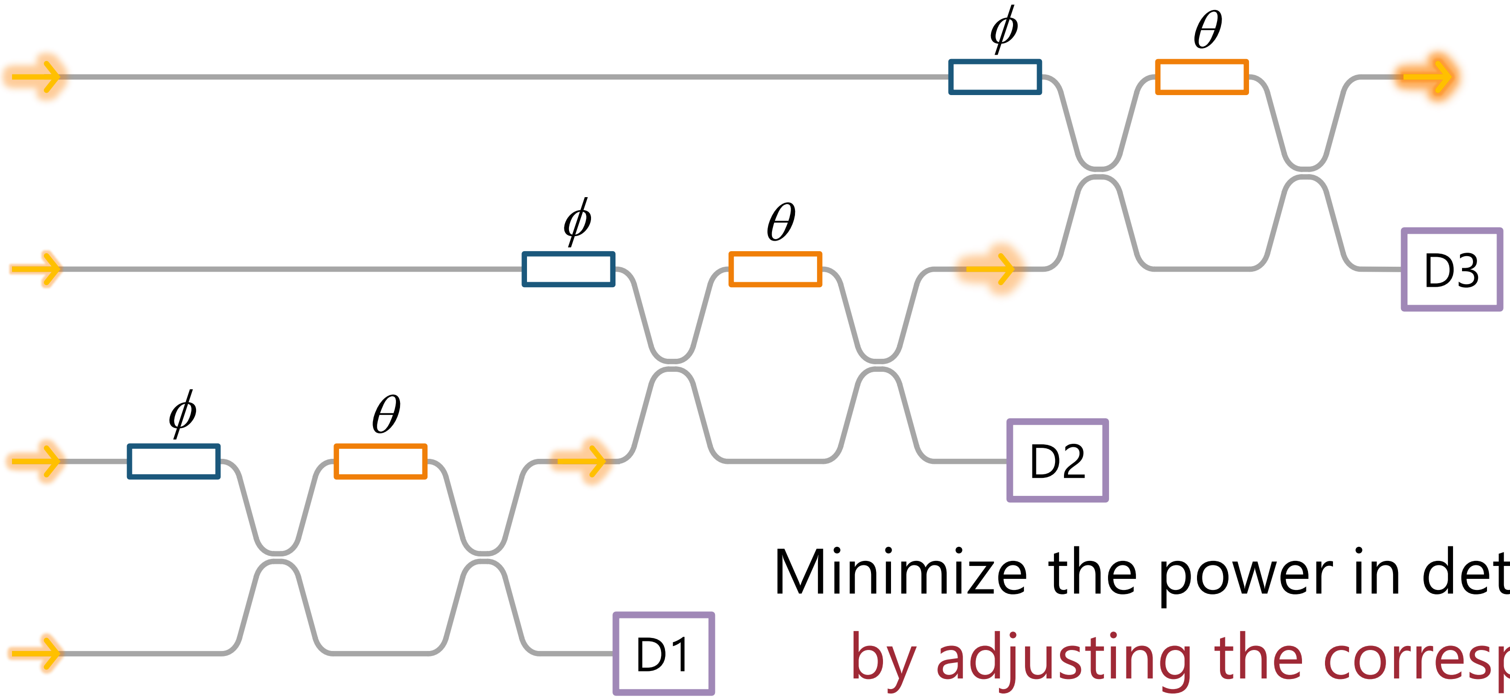
by adjusting the corresponding ϕ

and then θ

putting all power in the upper output

"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

"Diagonal line" self-aligning coupler



Minimize the power in detector D2

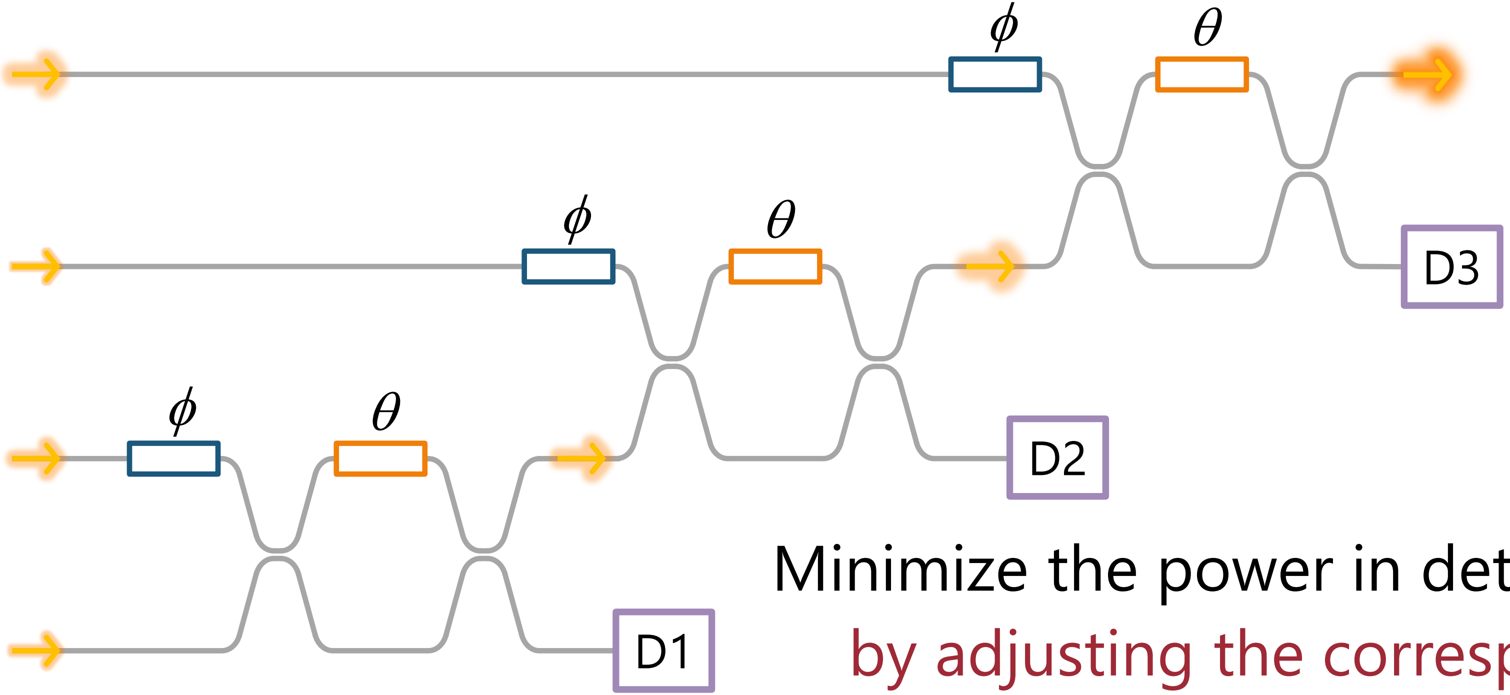
by adjusting the corresponding ϕ

and then θ

putting all power in the upper output

"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

"Diagonal line" self-aligning coupler



"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

Minimize the power in detector D3

by adjusting the corresponding ϕ

and then θ

putting all power in the upper output

Self-aligning beam coupler

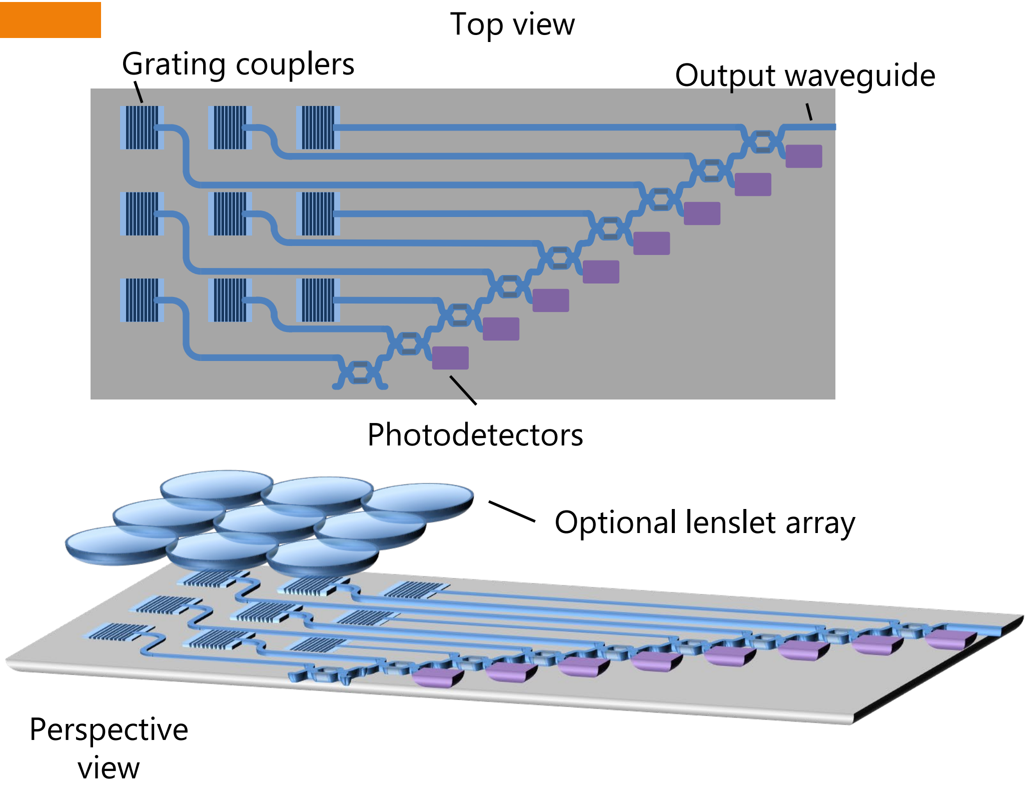
Grating couplers could couple a free-space beam to a set of waveguides

Then

we could automatically couple all the power to the one output guide

This could run continuously

tracking changes in the beam

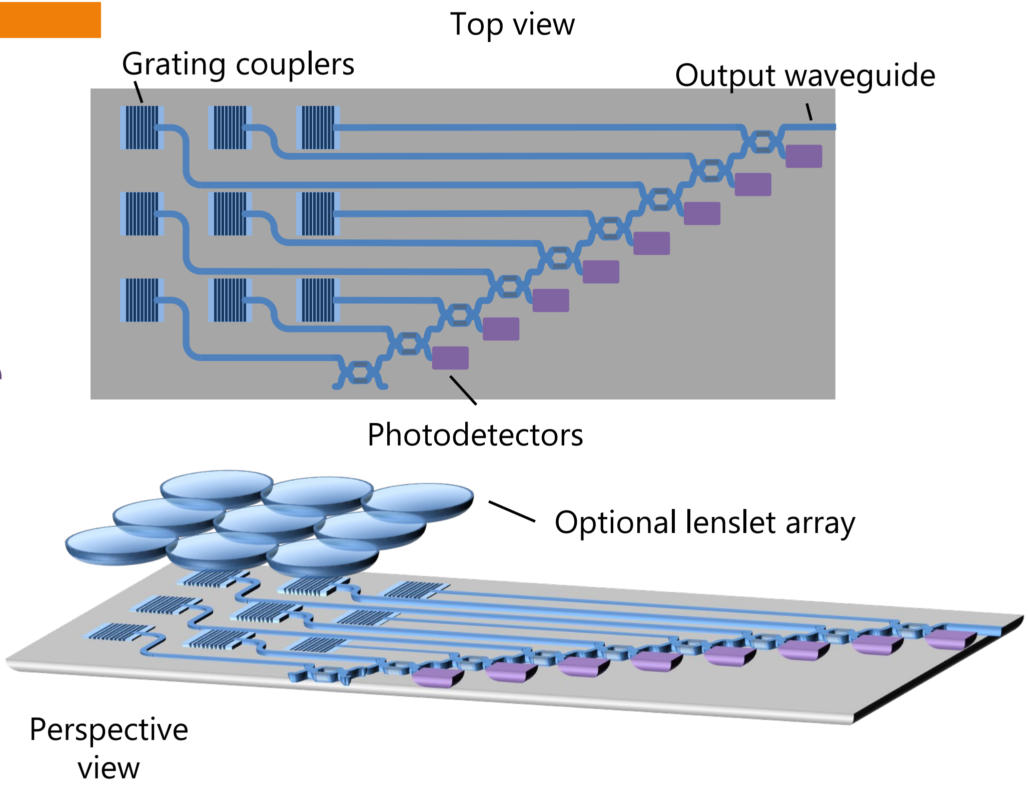


"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

Self-aligning beam coupler

This has several different uses

- tracking an input source
  both in angle and focusing
- correcting for aberrations
- analyzing amplitude and phase of the components of a beam
  - …

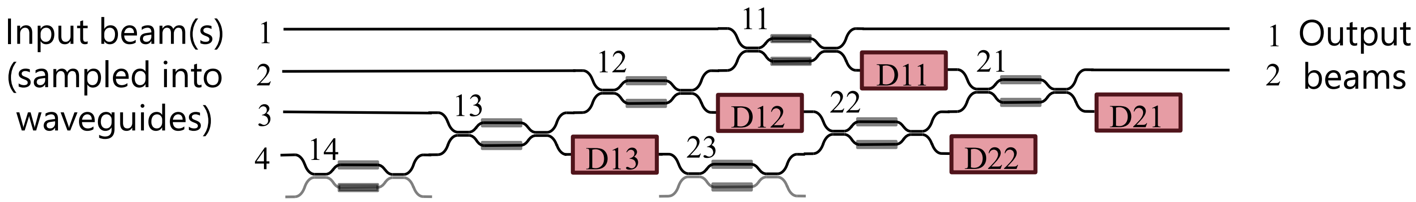


"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

Separating beams with interferometer meshes

stanford.io/4oZy7bf

Separating multiple orthogonal beams



"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

Once we have aligned beam 1 to output 1 using detectors D11 – D13

an orthogonal input beam 2 would pass entirely into the detectors D11 – D13

If we make these detectors mostly transparent

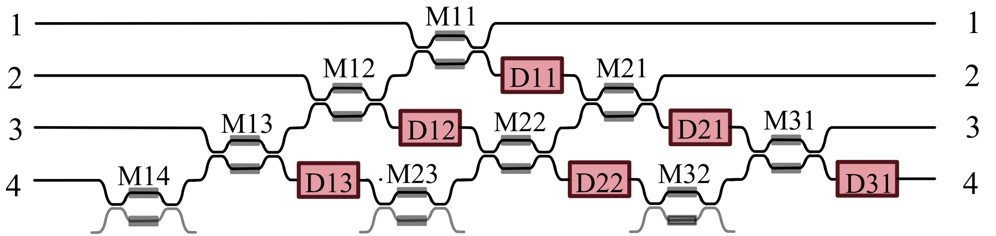
this second beam would pass into the second diagonal "row"

where we self-align it to output 2 using detectors D21 – D22

separating two overlapping orthogonal beams to separate outputs

Separating multiple orthogonal beams

Input beam(s) (sampled into waveguides)



Output beams

"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

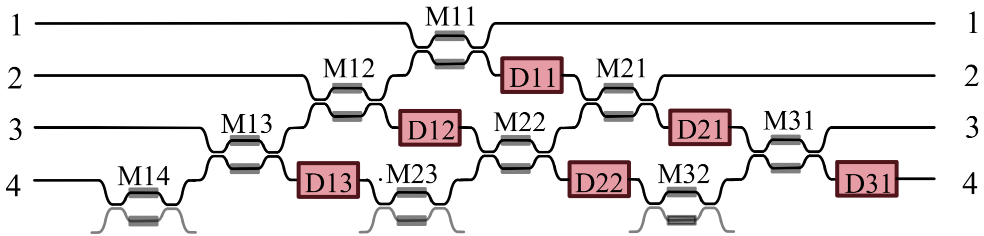
Adding more rows and self-alignments

separates a number of orthogonal beams

equal to the number of beam “segments”, here, 4

Separating multiple orthogonal beams

Input beam(s) (sampled into waveguides)



Output beams

"Self-aligning universal beam coupler," Opt. Express **21**, 6360 (2013)

If we put identifying "tones" on each orthogonal input "beam"

and have the corresponding diagonal row of detectors look for that tone

then the mesh can continually adapt to the orthogonal inputs

even when they are all present at the same time

and even if they change

Self-configuring beam separator

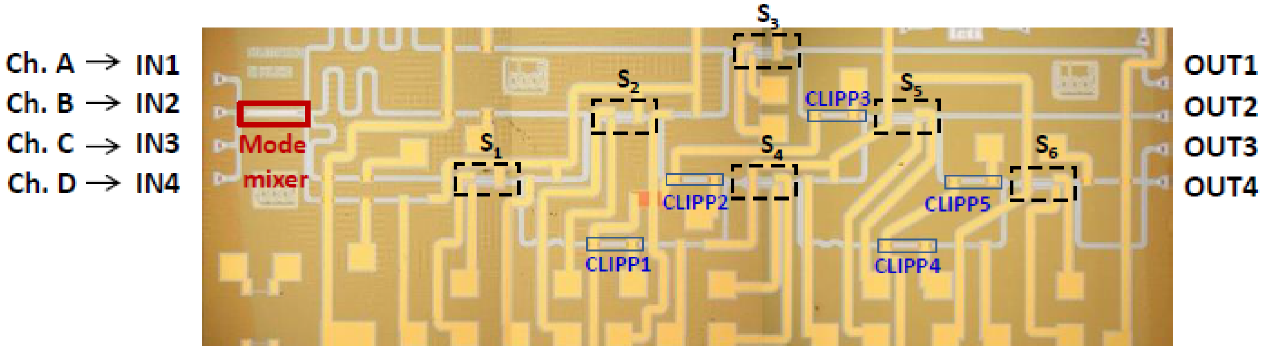
Light from four input fibers

deliberately mixed in a mode mixer

are automatically separated out again by a mesh of interferometers

by sequential power maximizations

without calculations



A. Annoni et al., "Unscrambling light – automatically undoing strong mixing between modes," Light Science & Applications 6, e17110 (2017)

See, e.g., review W. Bogaerts et al., "Programmable photonic circuits," Nature 586, 207 (2020)

Optical and mathematical linear operations with meshes

stanford.io/4oZy7bf

Universal matrix multiplier chip

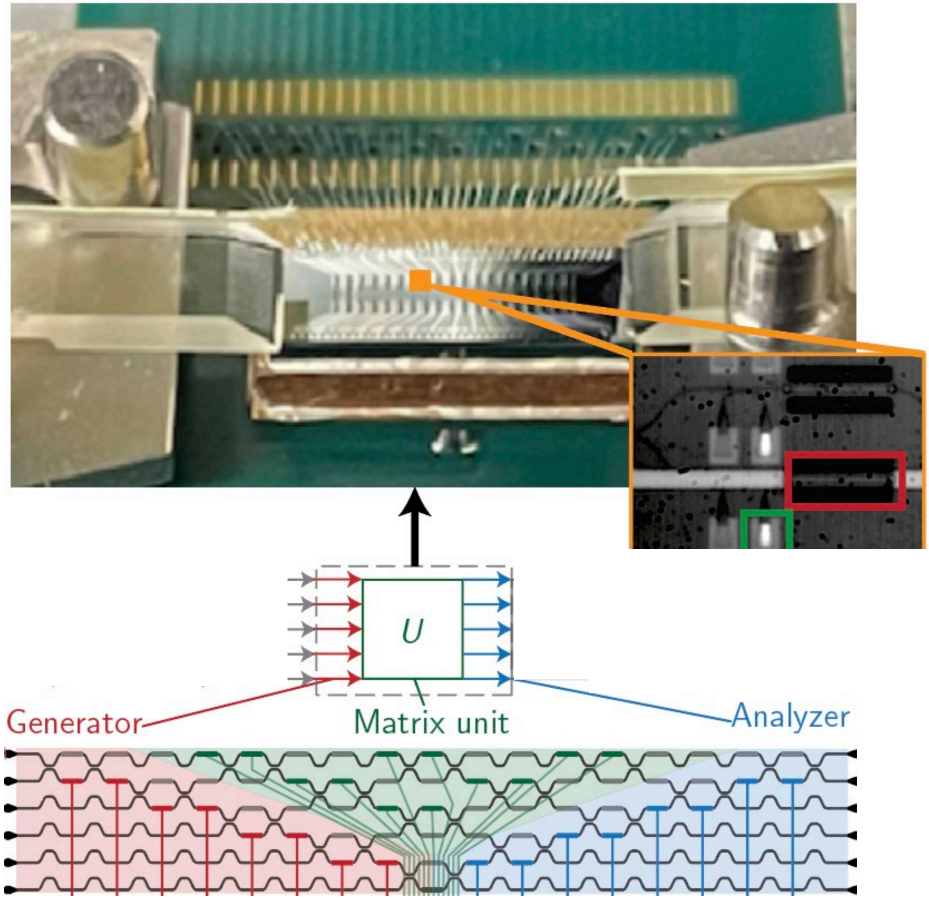
Universal matrix multiplying chip

"4x4" unitary Mach-Zehnder mesh with

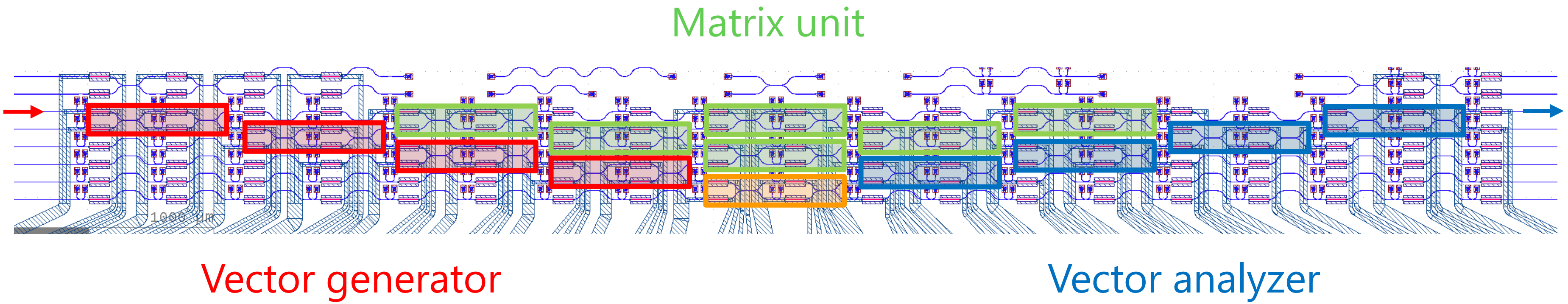
- a "generator" to create any complex input vector
- an "analyzer" to measure the complex output vector

This can be programmed to implement

any "unitary" (loss-less) transformation from the inputs to the outputs



Mask layout and block diagram



Universal matrix multiplier chip

Full complex matrix multiplication

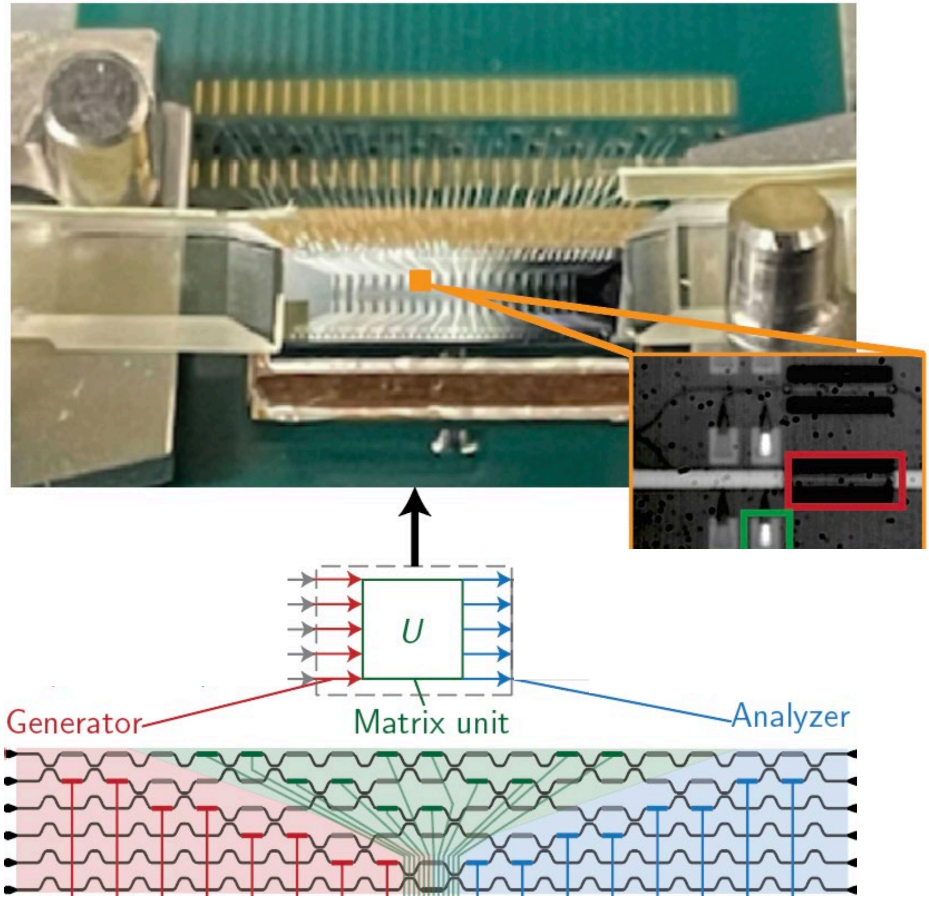
with vector generation and vector analysis

Photonic back-propagation neural net training

S. Pai, Z. Sun, T. W. Hughes, T. Park, B. Bartlett, I. A. D. Williamson, M. Minkov, M. Milanizadeh, N. Abebe, F. Morichetti, A Melloni, S. Fan, O. Solgaard, D. A. B. Miller, "Experimentally realized in situ backpropagation for deep learning in photonic neural networks," **Science** 380, 398-404 (2023)

Digital matrix multiplication for cryptography

S. Pai, T. Park, M. Ball, B. Penkovsky, M. Dubrovsky, N. Abebe, M. Milanizadeh, F. Morichetti, A. Melloni, S. Fan, O. Solgaard, and D. A. B. Miller, "Experimental evaluation of digitally verifiable photonic computing for blockchain and cryptocurrency," **Optica** 10, 552-560 (2023)



“Flipping round” the SVD

Now, we know that we can construct any unitary linear operator in optics

- using a mesh of interferometers

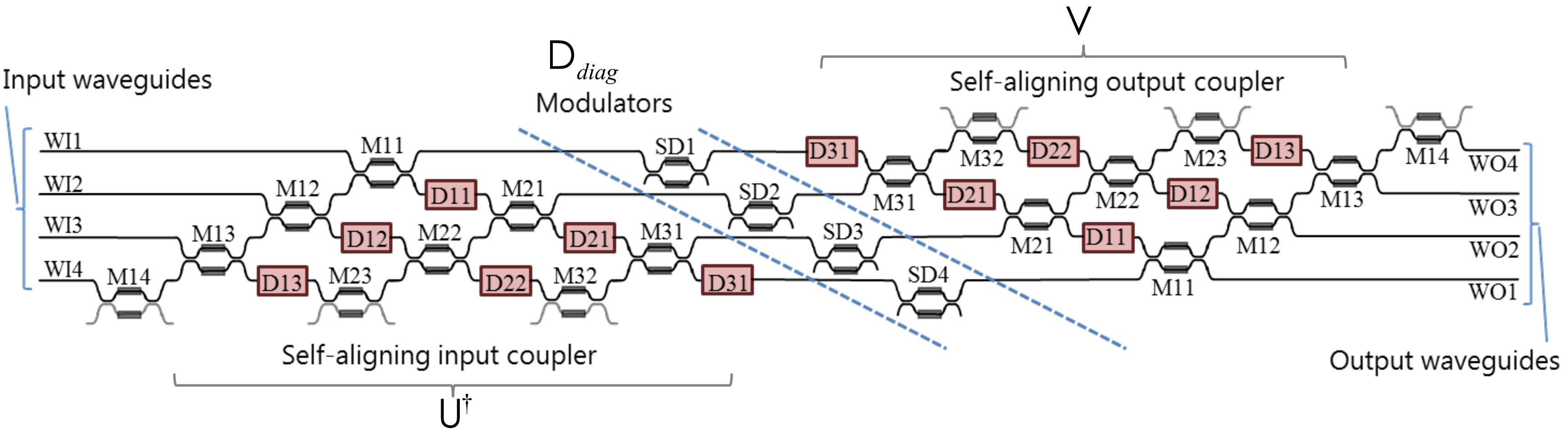
And we now know we can perform the SVD of any linear optical system

- which decomposes it mathematically into a product of three operators
  - a unitary, a diagonal and a unitary

Can we take one more step

- and emulate *any* linear operator with interferometer meshes?

General multiple mode converter



So, the optical "units" in the mesh implement the singular value decomposition $D = VD_{diag}U^{\dagger}$

So, for an optical system of a given dimensionality

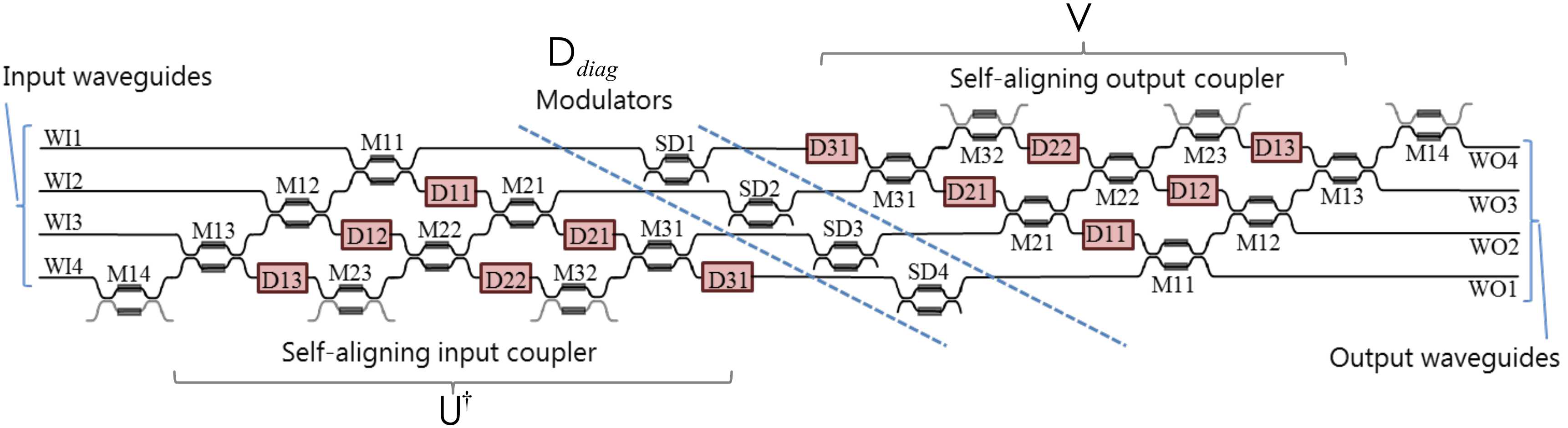
we can emulate any linear optical system

Note we are implementing an arbitrary linear optical component

by constructing it using its mode converter basis sets

"Self-configuring universal linear optical component," Photon. Res. **1**, 1-15 (2013).

General multiple mode converter



The input mode converter basis functions are the ones that
are converted to light in single waveguides in the middle

The output mode converter basis functions are the ones
generated by light in a single waveguide in the middle

"Self-configuring universal linear optical component," Photon. Res. **1**, 1-15 (2013).

The coupling strengths from input to output mode-converter modes
are the singular values implemented by the modulators in the middle

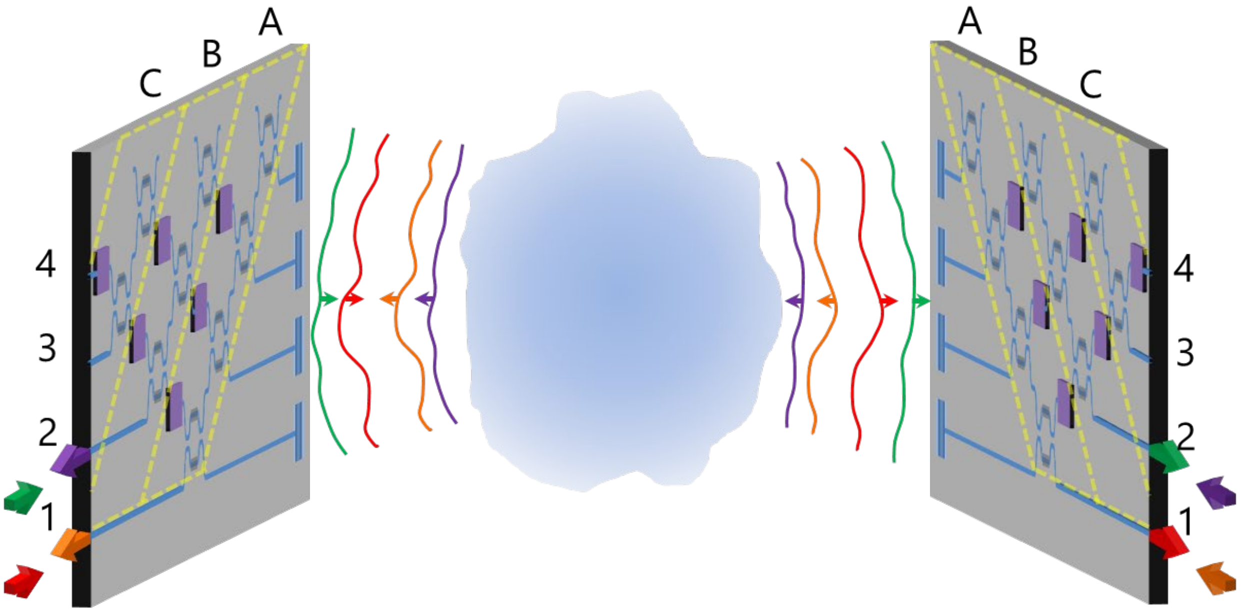
Establishing optimum orthogonal channels

In this architecture, using meshes on both sides

we proposed we could find optimal orthogonal channels through a scatterer

between waveguides on the left and waveguides on the right

by iterating back and forward between the two sides



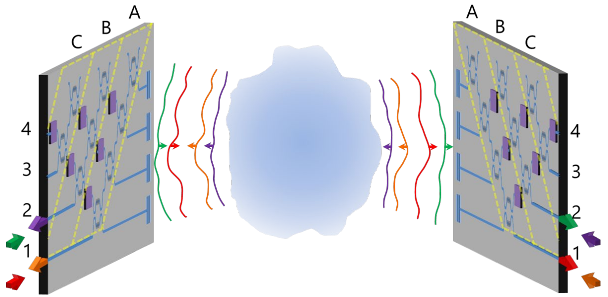
"Establishing optimal wave communication channels automatically," J. Lightwave Technol. 31, 3987 (2013)

Using optics to *perform* linear algebra

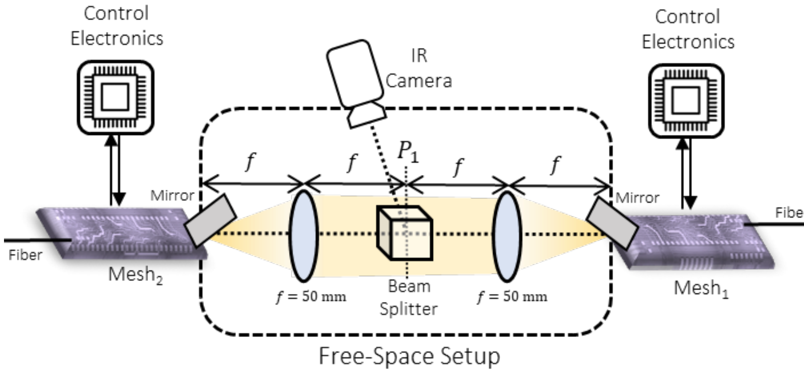
By power maximizing on rows of the mesh at both sides

- this circuit can automatically find the best orthogonal channels between the two sides
  - physically performing the singular-value decomposition of the optical system

This is a true optical computer!



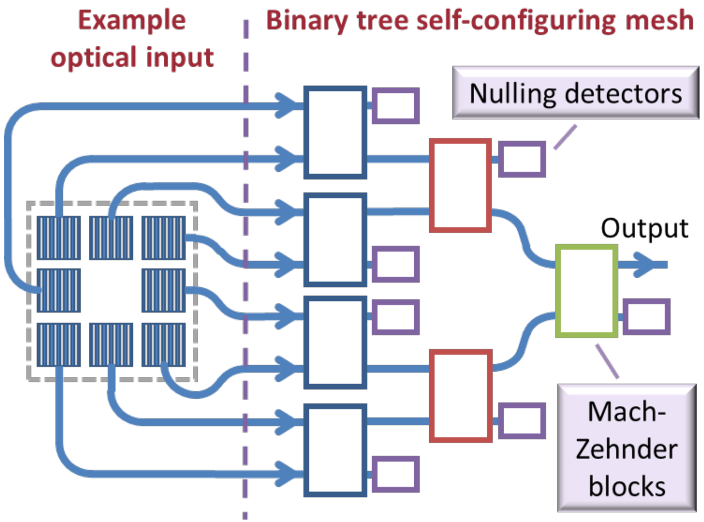
"Establishing optimal wave communication channels automatically," J. Lightwave Technol. 31, 3987 (2013)



S. SeyedinNavadeh et al., "Determining the optimal communication channels of arbitrary optical systems using integrated photonic processors," Nat. Photon. **18**, 149-155 (2024)

Measuring and generating arbitrary beams

- Self-configuring this “binary tree” layer to route all power to the output
  - automatically measures the relative amplitudes and phases of the input light
    - with the values deduced from the resulting mesh settings.
- Run backwards, it can generate any beam emerging from the “inputs”
  - generation of arbitrary beams
  - reference-free measurement of arbitrary beams



“Analyzing and generating multimode optical fields using self-configuring networks,” Optica 7, 794 (2020)

See also J. Bütow et al. "Spatially resolving amplitude and phase of light with a reconfigurable photonic integrated circuit," Optica 9, 939 (2022)

Optically separating exoplanets

Finding exoplanets around distant stars is optically very challenging

the star may be 10^{10} times brighter than the planet

and the planet may lie in the weak wings of the star's diffraction pattern in the telescope

Interferometer meshes may allow

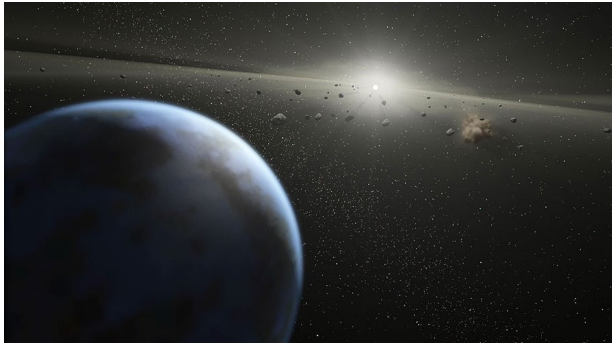
optimized modal filtering

to suppress the star "modes"

to improve the rejection of the star light

Preliminary experiments with meshes are already showing ~ 90dB rejection

Dan Sirbu et al., "AstroPIC: near-infrared photonic integrated circuit coronagraph architecture for the Habitable Worlds Observatory," Proc. SPIE 13092, 130921T (2024)



Use a programmable photonic mesh to provide optimal modal filtering to reject star light and pass possible exoplanet light

Separating partially coherent light

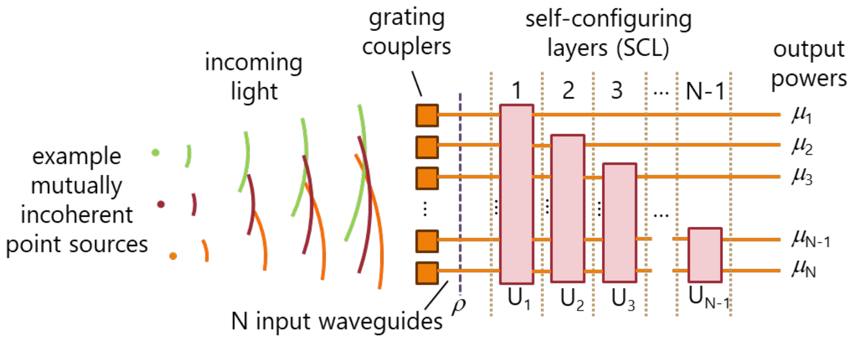
With partially coherent input light

by power maximizing on the successive self-configuring layers

this circuit can measure the coherency matrix of that light

simultaneously separating it into its mutually incoherent and mutually orthogonal components

No other known apparatus can apparently perform this separation



Roques-Carmes et al., "Measuring, processing, and generating partially coherent light ..." LSA **13**, 260 (2024)

Programmable and self-configuring filters

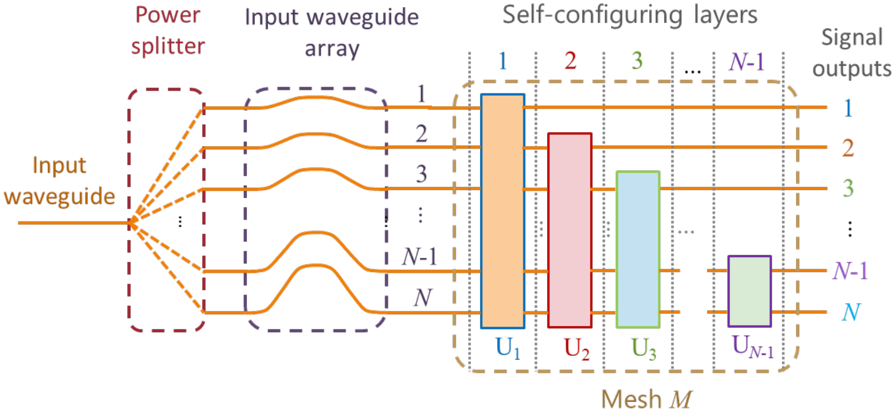
This proposed circuit can function like an arrayed waveguide grating filter

- but has a spectral response that is fully programmable
  - so it can implement any linear combination of such filter functions
  - and allows multiple different simultaneous filter functions

It can also

- self-configure to specific wavelengths
- reject N-1 arbitrary wavelengths
- measure and separate temporally partially coherent light
  - the Karhunen-Loève decomposition

D. A. B. Miller, C. Roques-Carmes, C. G. Valdez, A. R. Kroo, M. Vlk, Shanhui Fan, and O. Solgaard, "Universal programmable and self-configuring optical filter," Optica **12**, 1417-1426 (2025)



Conclusions

stanford.io/4oZy7bf

- Interferometer meshes allow many different and fully programmable optical and mathematical functions
  - with self-configuration to adapt to the problem of interest
  - and to stabilize what is otherwise a very complex interferometric circuit
- Applications include
  - mathematical operations
  - reference-free measurement of arbitrary optical fields
  - generation of arbitrary beams
  - automatically finding best channels
  - modal filtering to separate and suppress arbitrary beams
  - separation and measurement of partially coherent light
  - extensions to similar concepts in the frequency domain
- Conceptual realizations from this work include
  - understanding that any linear optical operation can be reduced to two-beam interferences
  - thought experiments utilizing arbitrary linear optical transformations

Tunneling escape of waves

stanford.io/4oZy7bf

David Miller, *Stanford University*
Zeyu Kuang, Owen Miller, *Yale University*

Why the abrupt fall-off past some number of channels

Why do we *always* see

regardless of the shape of the source and receiving volumes or surfaces

some number of "well coupled" channels

followed by an abrupt, quasi-exponential fall-off in couplings past this number

and just what gives this number?

Is there some underlying piece of physics we are missing?

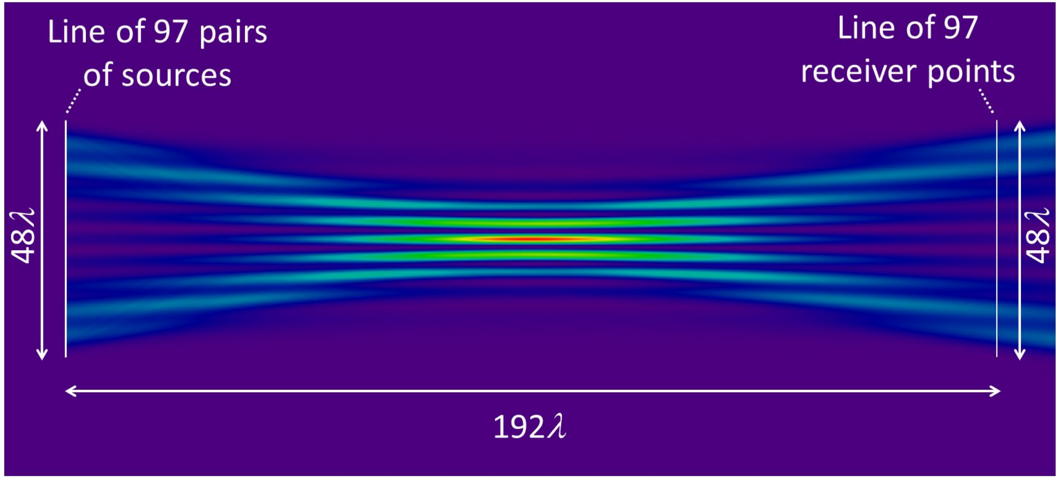
We might argue it is "just" diffraction limits

but that still does not explain the rapid fall-off

Communication modes for a large paraxial example

Now we consider a large line of sources and a line of receiver points

with an approximately "paraxial" set of dimensions

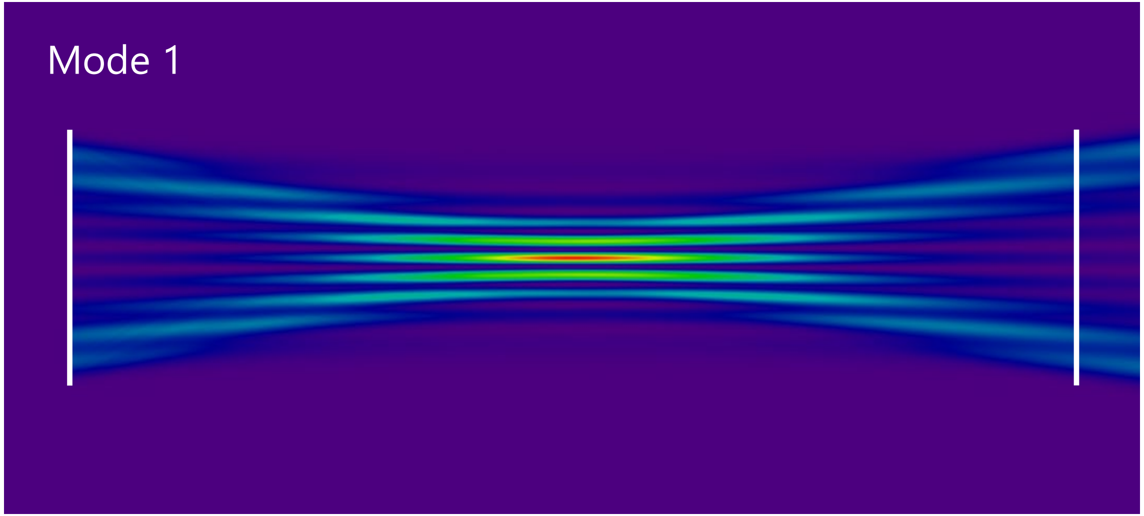


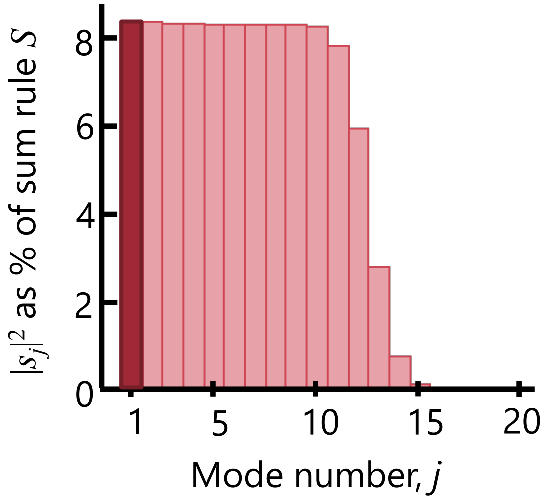
"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

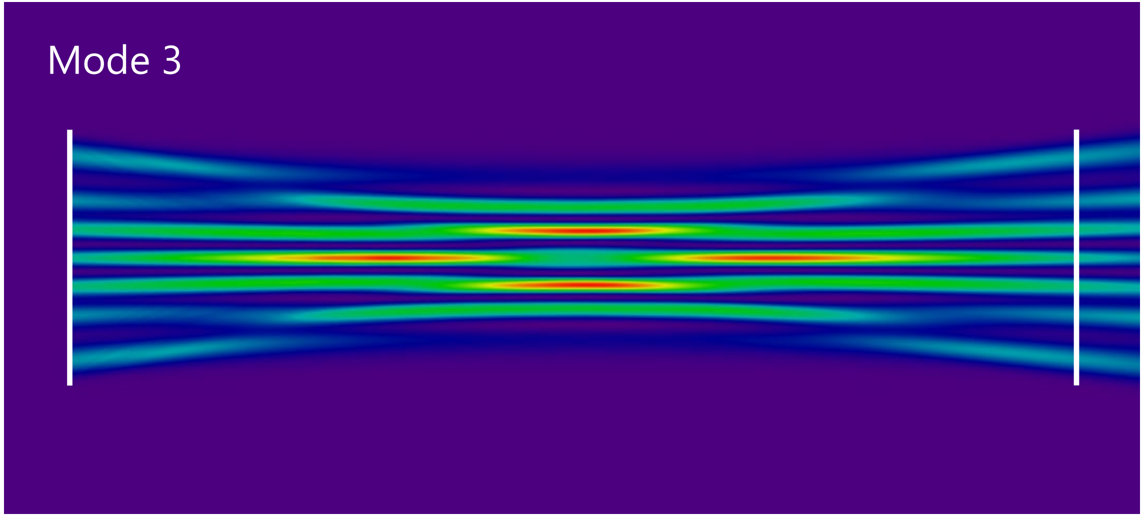
and we establish the communication modes between them

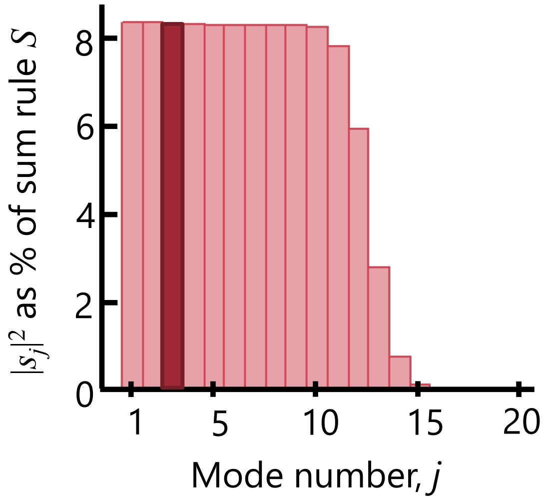
The picture shows the cross-section of the intensity in the plane

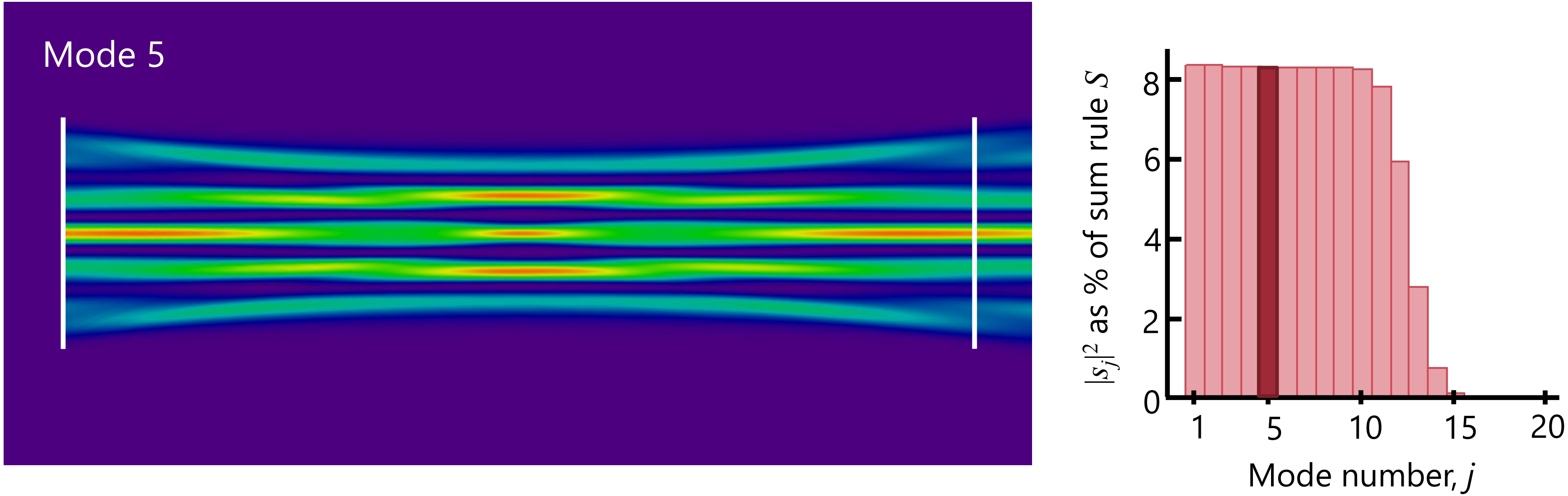
here for the most strongly coupled mode


Mode 1

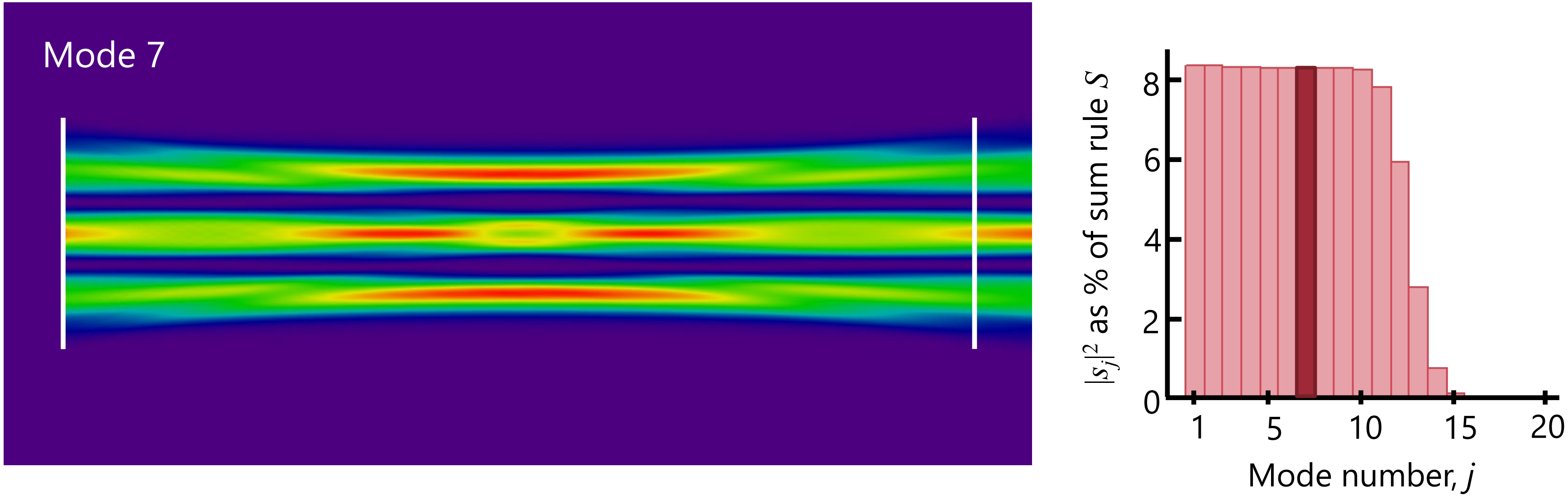

$|s_j|^2$ as % of sum rule S
8
6
4
2
0
1
5
10
15
20
Mode number, j


Mode 3

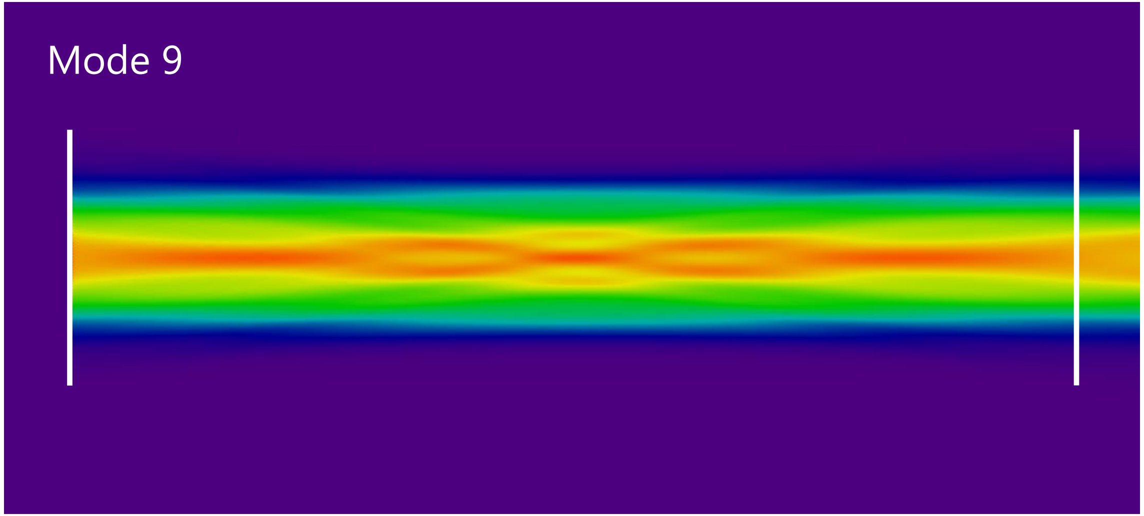

$|s_j|^2$ as % of sum rule S
8
6
4
2
0
1
5
10
15
20
Mode number, j

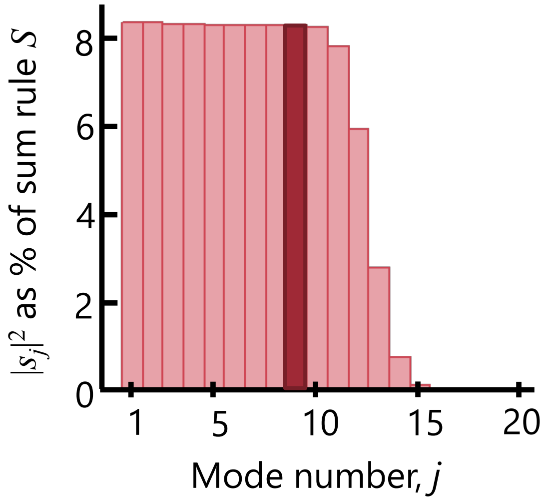


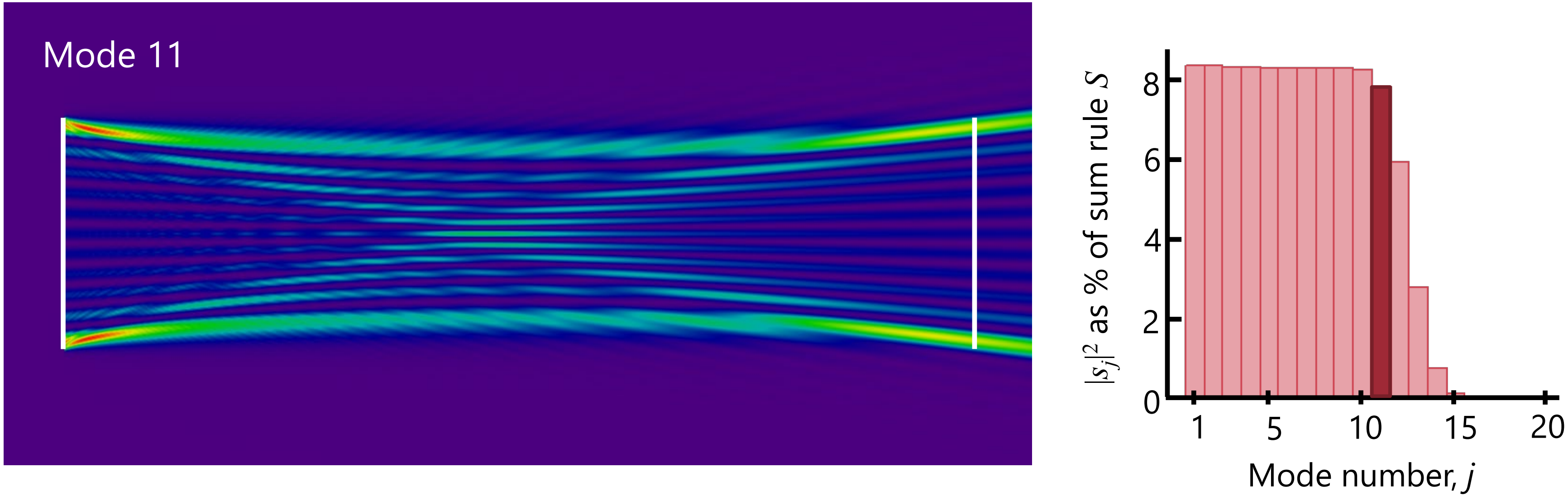
Mode 5
$|s_j|^2$ as % of sum rule S
0
2
4
6
8
1
5
10
15
20
Mode number, j



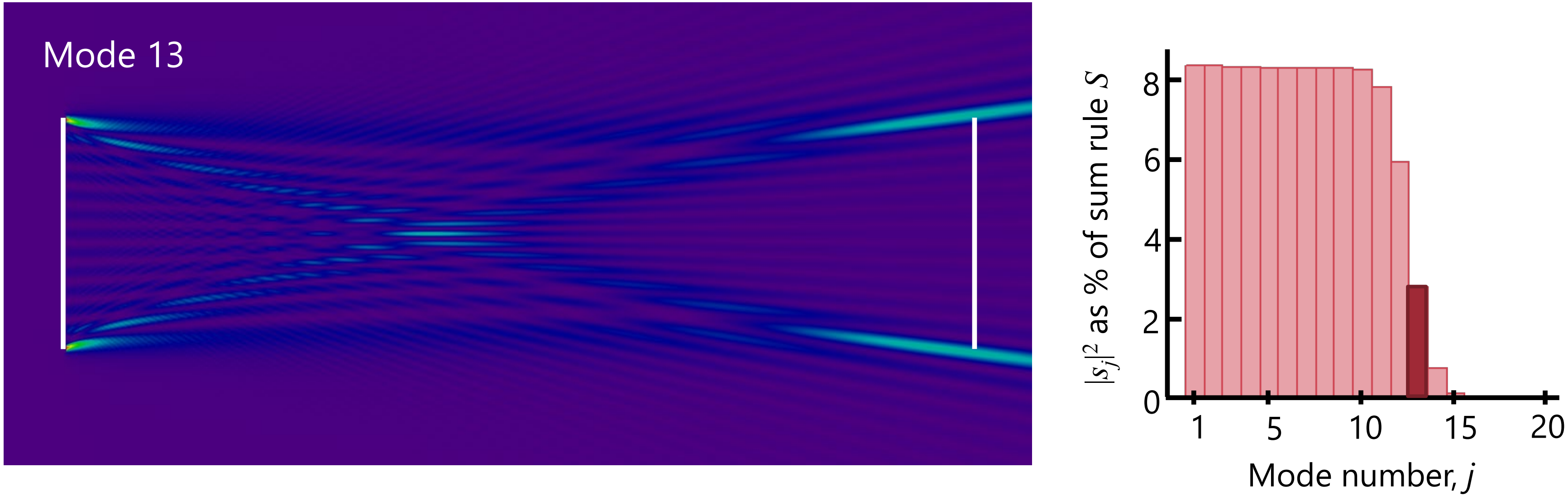
Mode 7
$|s_j|^2$ as % of sum rule S
0
2
4
6
8
1
5
10
15
20
Mode number, j


Mode 9

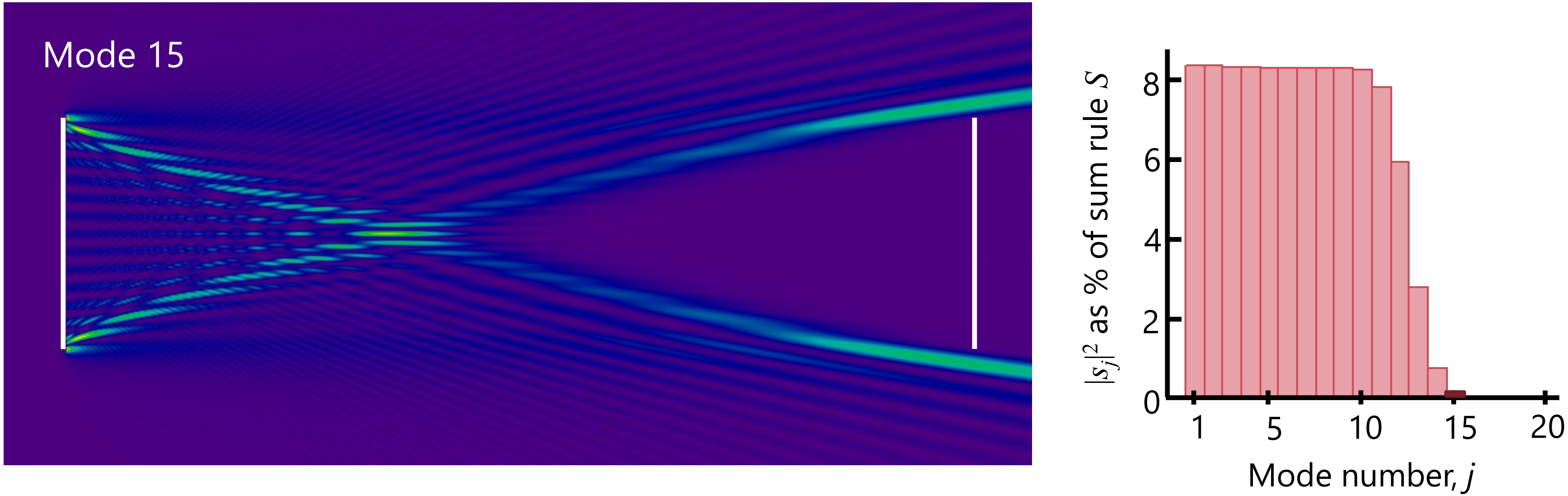

$|s_j|^2$ as % of sum rule S
0
2
4
6
8
1
5
10
15
20
Mode number, j



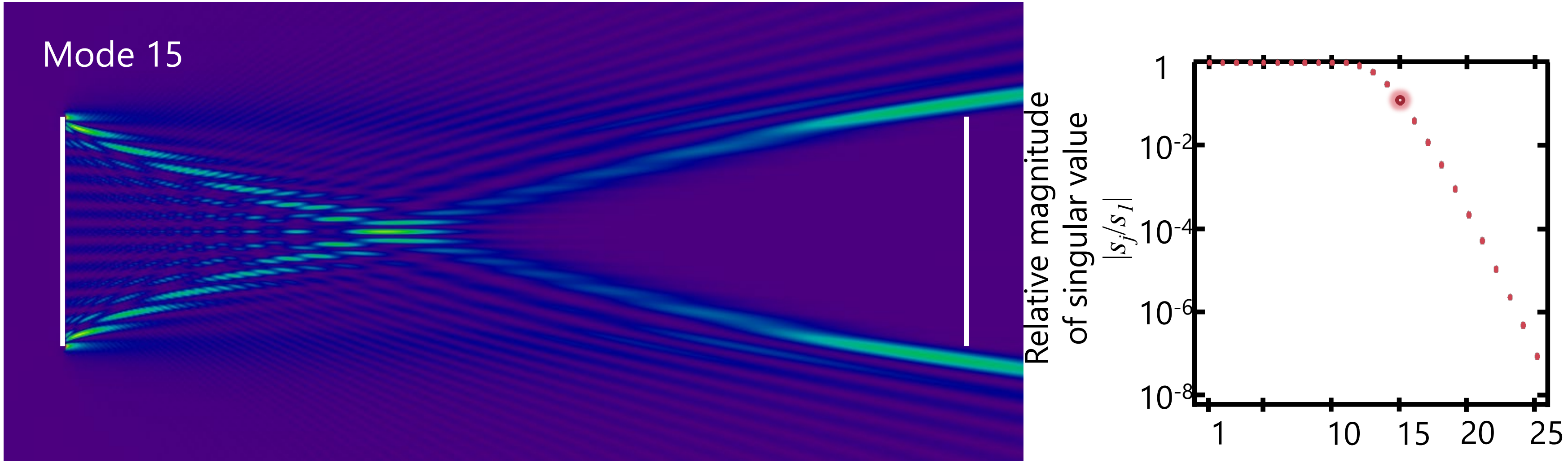
Mode 11
$|s_j|^2$ as % of sum rule S
0
2
4
6
8
1
5
10
15
20
Mode number, j



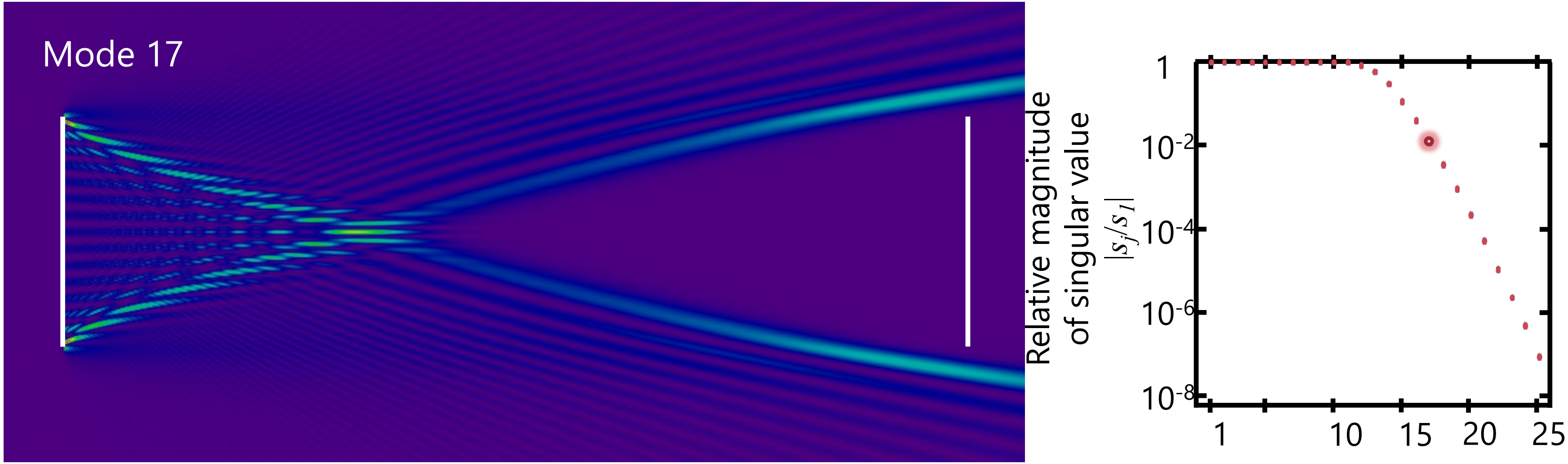
Mode 13
$|s_j|^2$ as % of sum rule S
0
2
4
6
8
1
5
10
15
20
Mode number, j



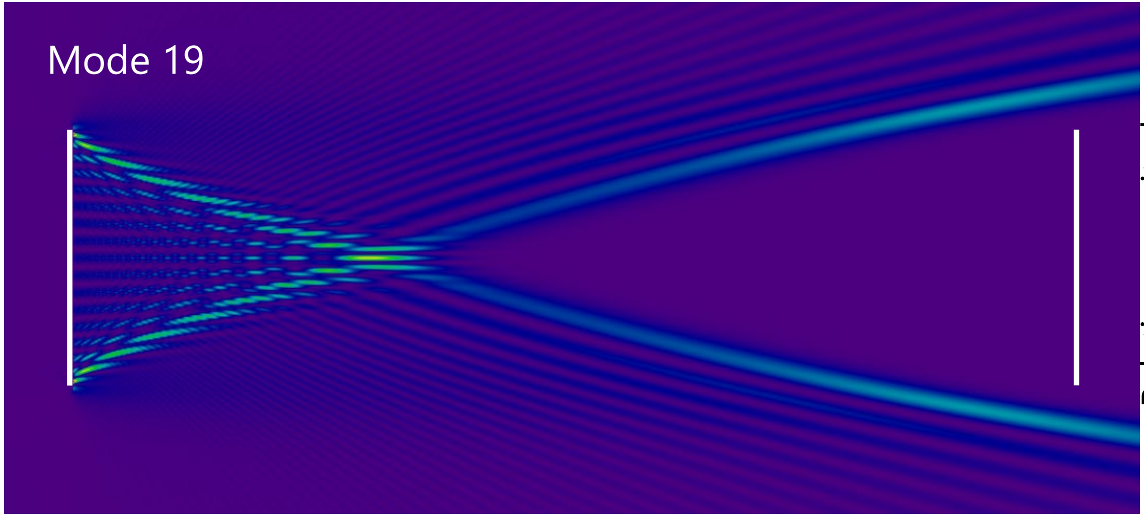
Mode 15
$|s_j|^2$ as % of sum rule S
0
2
4
6
8
1
5
10
15
20
Mode number, j

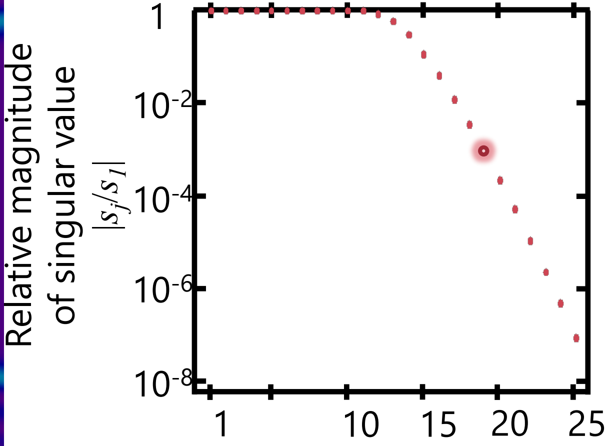


Mode 15
Relative magnitude of singular value
$|s_j/s_1|$
1
10^{-2}
10^{-4}
10^{-6}
10^{-8}
1
10
15
20
25

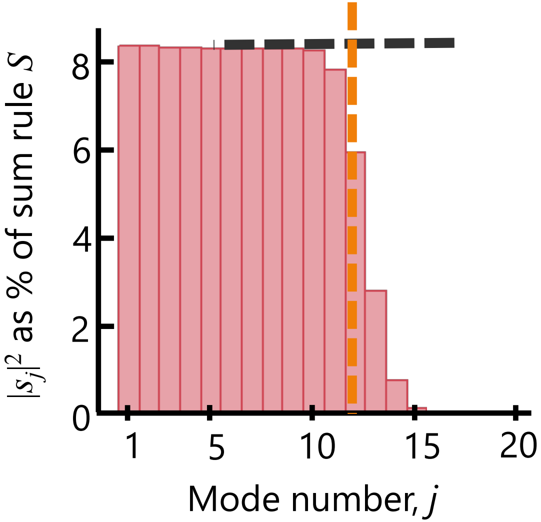


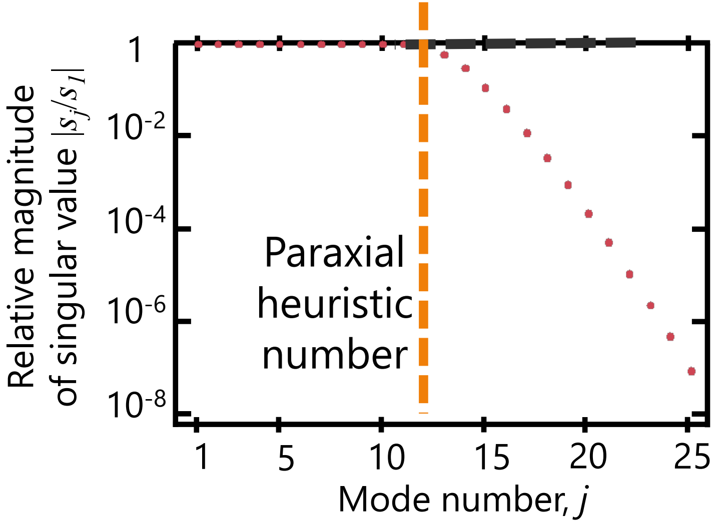
Mode 17
Relative magnitude of singular value
|s_j/s_1|
1
10^-2
10^-4
10^-6
10^-8
1
10
15
20
25


Mode 19


Relative magnitude
of singular value
|s_j/s_1|
1
10^-2
10^-4
10^-6
10^-8
1
10
15
20
25

Paraxial heuristic number and paraxial degeneracy





"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

Paraxial heuristic number

$N_H \sim W_S W_R / \lambda L$

for source and receiver widths W_S, W_R

separation L

wavelength λ

Once we pass the number we expect from conventional "diffraction limits"

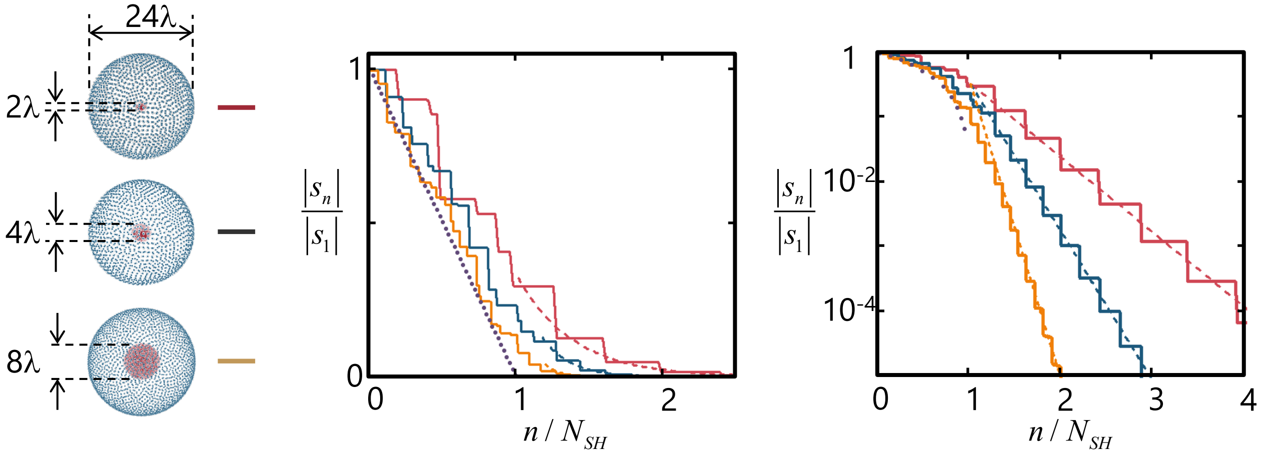
coupling strengths for further communication modes

fall off drastically and somewhat exponentially

We might think this is because the waves "miss" the receiving space

but that is not the general explanation

3D examples – concentric spherical shells



"Waves, modes, communications and optics," Adv. Opt. Photon. 11, 679 (2019)

Z. Kuang, D. A. B. Miller, and O. D. Miller, "Bounds on the Coupling Strengths of Communication Channels and their Information Capacities," IEEE Trans. Antennas and Propagation, **73**, 3974 (2025)

Concentric spherical shell source and receiver spaces

- are not easily analyzed by conventional "diffraction limit" theories
  - and do not show "paraxial degeneracy"
    - and the waves from the source space cannot "miss" the receiving space
      - but we still get some characteristic number of well-coupled communication modes
        - and a quasi-exponential fall-off of coupling beyond that

Waves from arbitrary volumes

How can we count the maximum number of well coupled waves (at a given frequency)

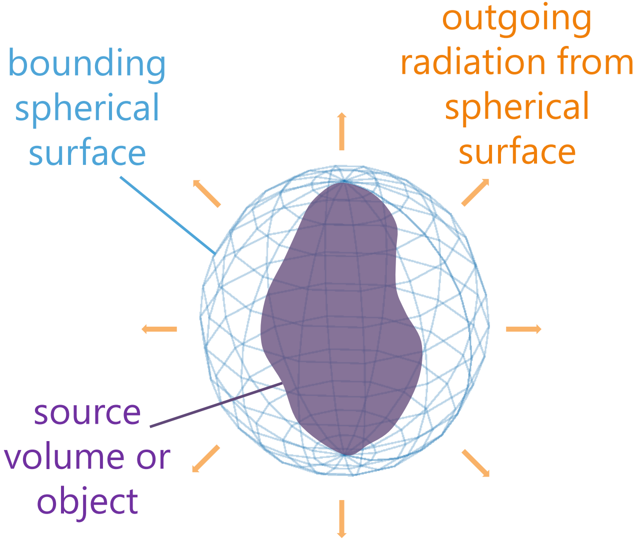
from some finite volume?

Our approach

Surround the volume with a mathematical "bounding" spherical surface

Count the number of well-coupled waves possible from this spherical surface

which then becomes the upper bound for waves from the source volume



D. A. B. Miller, Z. Kuang, O. D. Miller, "Tunneling escape of waves," Nat. Photon. **19**, 284–290 (2025)

Waves from arbitrary volumes

We show that, for spherical waves

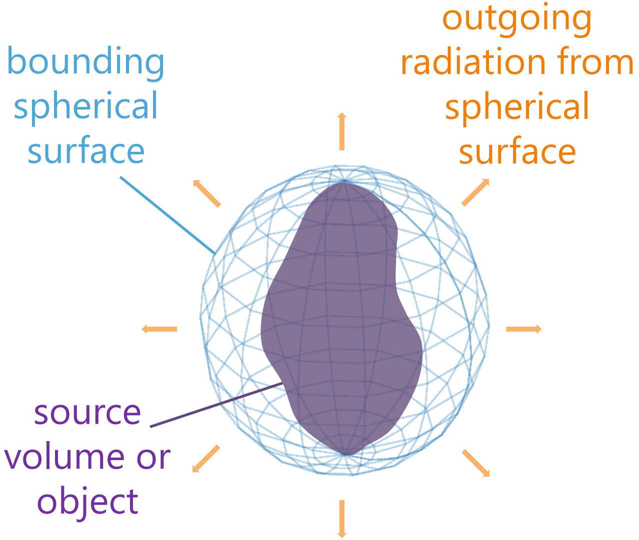
- with one key mathematical trick
  - there is a very simple and physical result

Beyond a certain simple threshold of "complexity" of spherical waves

- they must "tunnel" to escape

Because the fall-off from tunneling is generally so rapid

- this threshold effectively tells us the maximum number of well-coupled waves
  - and explains the quasi-exponential fall-off



D. A. B. Miller, Z. Kuang, O. D. Miller, "Tunneling escape of waves," Nat. Photon. **19**, 284–290 (2025)

Waves in spherical coordinates

In spherical coordinates r, θ, and ϕ

the solution to the wave equation separates to

$$U_{nm}(\mathbf{r}) = z_n(kr)Y_{nm}(\theta,\phi)$$

where $z_n(kr)$ is one of the spherical Bessel functions of order n, and

$Y_{nm}(\theta,\phi)$ is a spherical harmonic

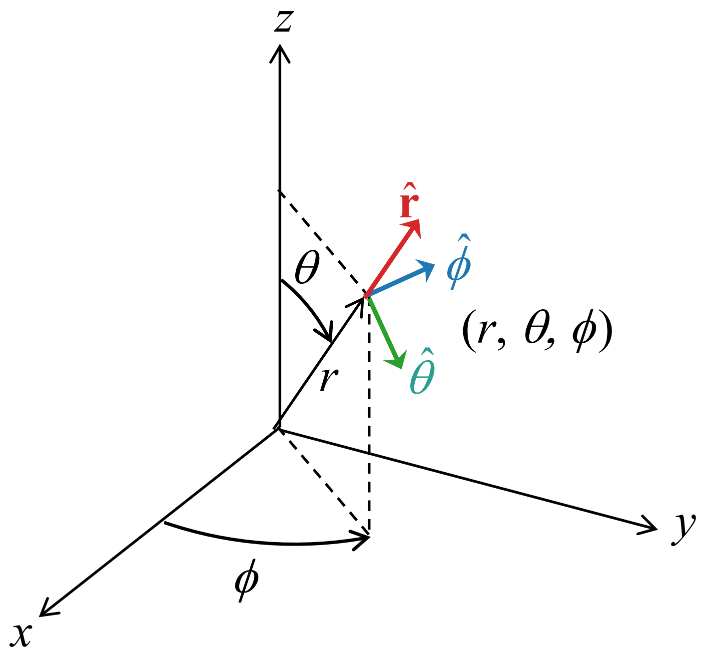
Here m and n are integers with

$n = 0,1,2,\ldots$ and $-n \le m \le n$

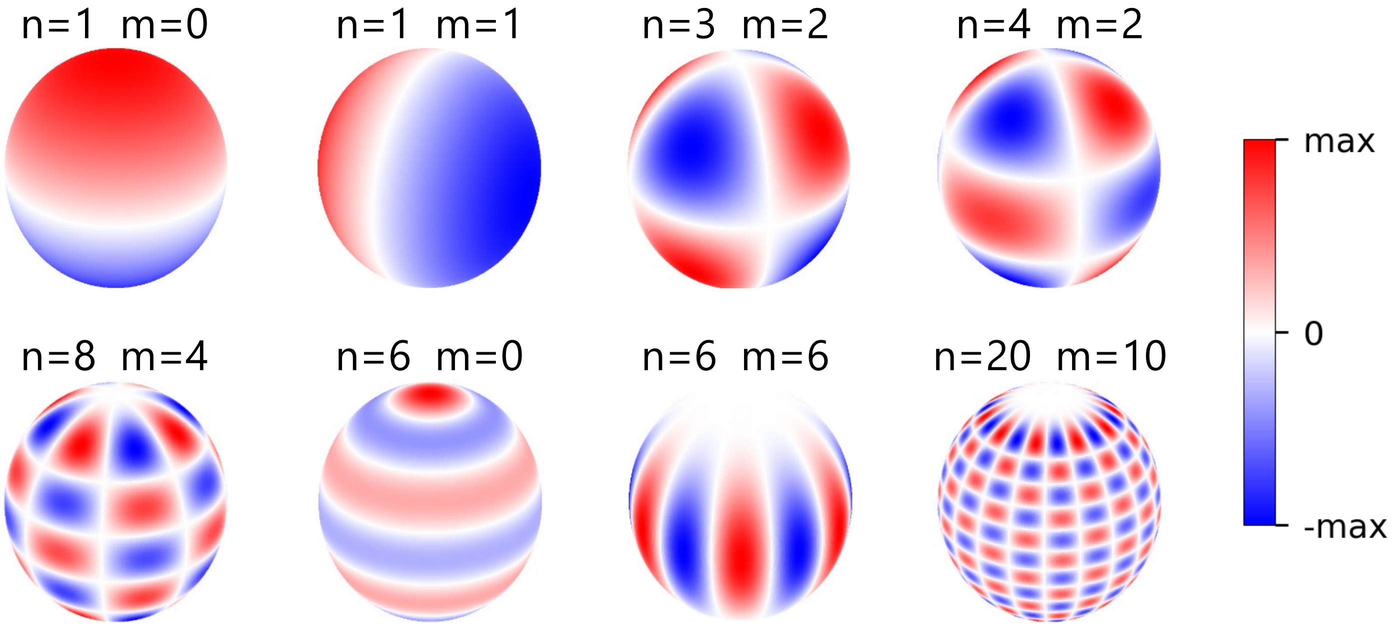
So, if we know the largest n for waves to propagate without tunneling

we can easily add up the total number of waves up to and including that n

$2n+1$ for each n



Spherical harmonics



Spherical harmonics are functions of angle only, and can be plotted on a spherical surface

They have *n* nodal circles altogether, with |*m*| through the poles (in their real form)

Escape radius

Specifically, for a given "order" n of spherical wave

there is an "escape radius"

$$r_{escn} = \frac{\sqrt{n(n+1)}}{k} \equiv \frac{\lambda_o}{2\pi}\sqrt{n(n+1)}$$

So, if the radius r_o of the spherical surface of interest

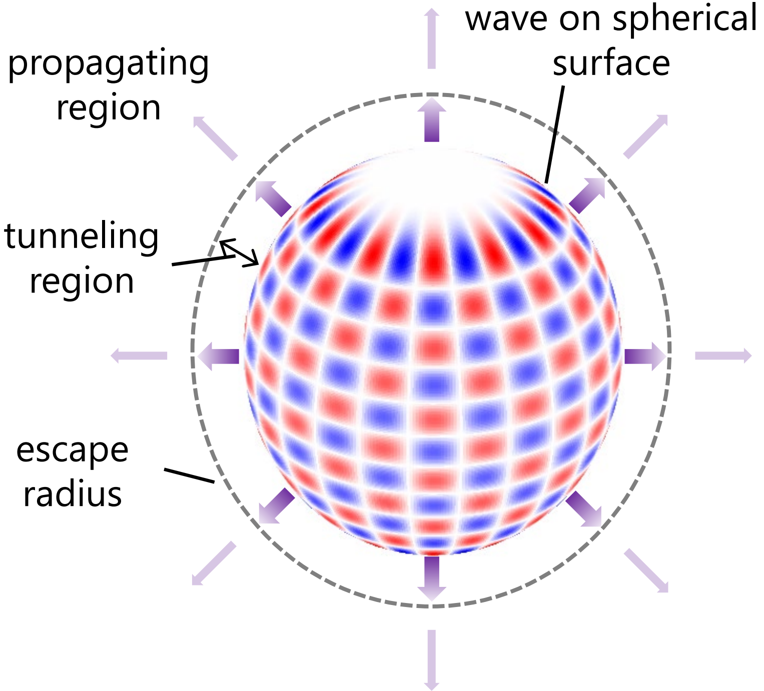
is smaller than the escape radius

for some order n of spherical wave

a wave with this n must tunnel

until it reaches the escape radius

after which it can propagate



D. A. B. Miller, Z. Kuang, O. D. Miller, "Tunneling escape of waves," Nat. Photon. **19**, 284–290 (2025)

Spherical Bessel functions and equation

Spherical Bessel functions obey

$$\rho^2 \frac{d^2 z_n(\rho)}{d\rho^2} + 2\rho \frac{dz_n(\rho)}{d\rho} + \left(\rho^2 - n(n+1)\right) z_n(\rho) = 0$$

Classic radial standing wave solutions are

j_n which grows quasi-exponentially for small radii

and is quasi-oscillatory for larger radii

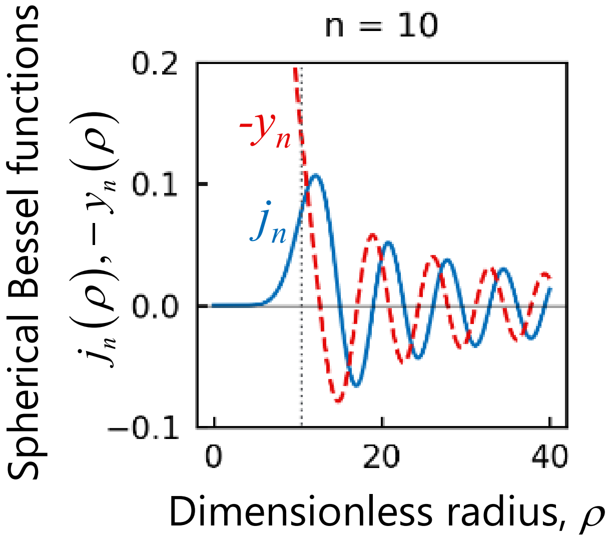
y_n which is singular at the origin

decaying quasi-exponentially for small radii

becoming quasi-oscillatory at large radii

Physically, ρ here is the dimensionless radius

$$\rho = kr \equiv 2\pi \frac{r}{\lambda}$$



Taking out the 1/radius dependence

Since the spherical Bessel functions have an underlying 1/radius dependence at large radius

- as appropriate for what are ultimately spherically expanding waves

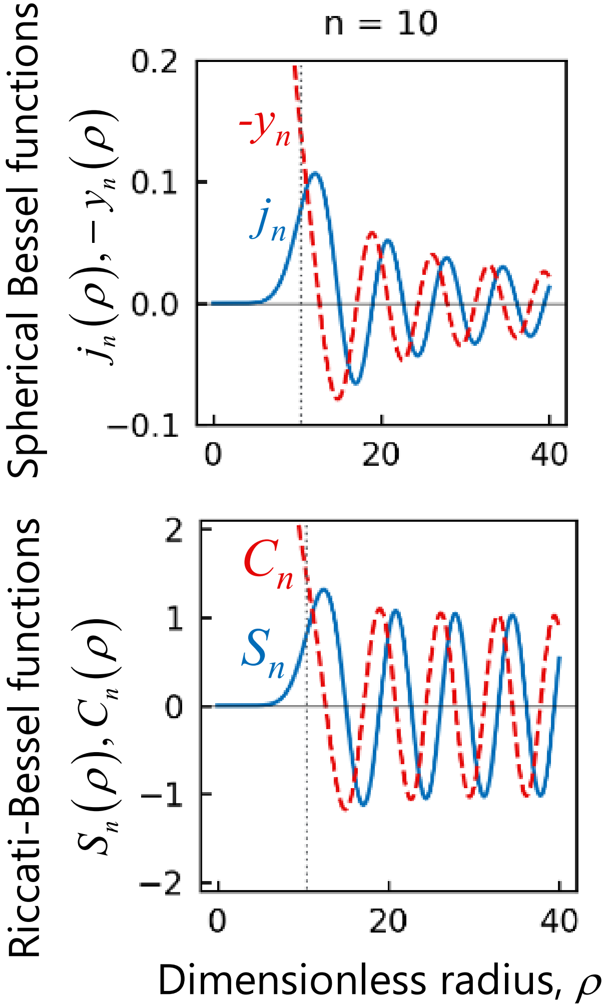
- it could be useful to remove that dependence
  - multiplying by radius
  - which gives functions corresponding to power per unit solid angle
  - rather than power per unit area

So, we recast in terms of such functions

- known as Riccati-Bessel functions

$$S_n(\rho) = \rho j_n(\rho) \qquad C_n(\rho) = -\rho y_n(\rho)$$

$$\xi_n(\rho) = \rho h_n^{(1)}(\rho) \equiv S_n(\rho) - iC_n(\rho)$$



Riccati-Bessel equation

Given that the spherical Bessel functions satisfy

$$\rho^2 \frac{d^2 z_n(\rho)}{d\rho^2} + 2\rho \frac{dz_n(\rho)}{d\rho} + \left(\rho^2 - n(n+1)\right) z_n(\rho) = 0$$

then we can easily check that all the Riccati-Bessel functions satisfy

$$\rho^2 \frac{d^2 \zeta_n}{d\rho^2} + \left(\rho^2 - n(n+1)\right)\zeta = 0$$

We can rearrange that to

$$-\frac{d^2 \zeta_n}{d\rho^2} + \frac{n(n+1)}{\rho^2}\zeta_n = \zeta_n$$

Riccati-Bessel "Schrödinger" equation

But wait!!!!!!

$$-\frac{d^2\zeta_n}{d\rho^2}+\frac{n(n+1)}{\rho^2}\zeta_n=\zeta_n$$

is in the form of a Schrödinger equation

$$-\frac{d^2\zeta_n}{d\rho^2}+V(\rho)\zeta_n=E_n\zeta_n$$

with effective radial potential

$$V(\rho)=\frac{n(n+1)}{\rho^2}$$

and the same "eigenenergy" $E_n=1$ for all n

Tunneling escape and escape radius

With the equation $$-\frac{d^2\zeta_n}{d\rho^2}+\frac{n(n+1)}{\rho^2}\zeta_n=\zeta_n$$

if the "potential energy" exceeds the "total energy", i.e., if

$\frac{n(n+1)}{\rho^2}>1$ or, equivalently $n(n+1)>\rho^2$

the wave will be tunneling rather than propagating

So, for each n, there is an "escape radius" $$\rho_{escn}=\sqrt{n(n+1)}$$

or, equivalently, in dimensioned form $$r_{escn}=\frac{\sqrt{n(n+1)}}{k}\equiv\frac{\lambda_o}{2\pi}\sqrt{n(n+1)}$$

Evanescent and spherical escaping waves

Both plane and spherical waves start with the same tunneling barrier height

and hence the same initial decay

but the falling barrier height for the spherical wave

means it eventually escapes

to being a propagating wave

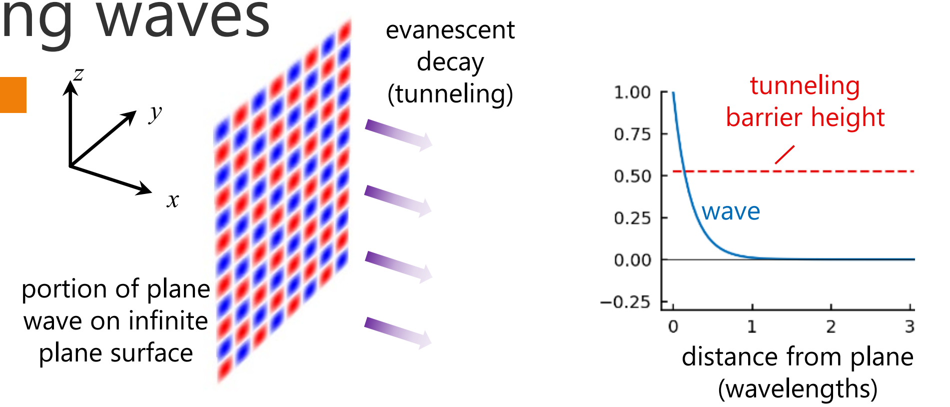
Note *all* such spherical waves

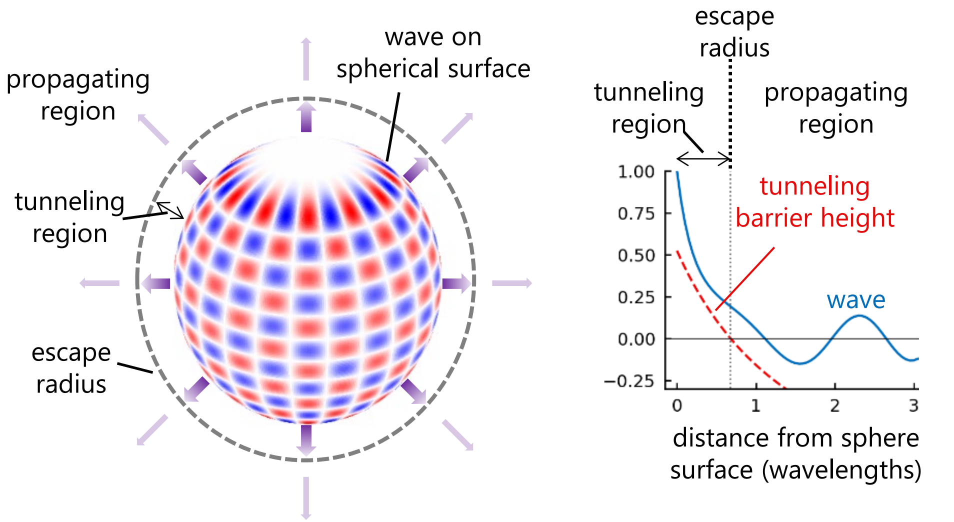
eventually escape to some degree

though the evanescent plane wave

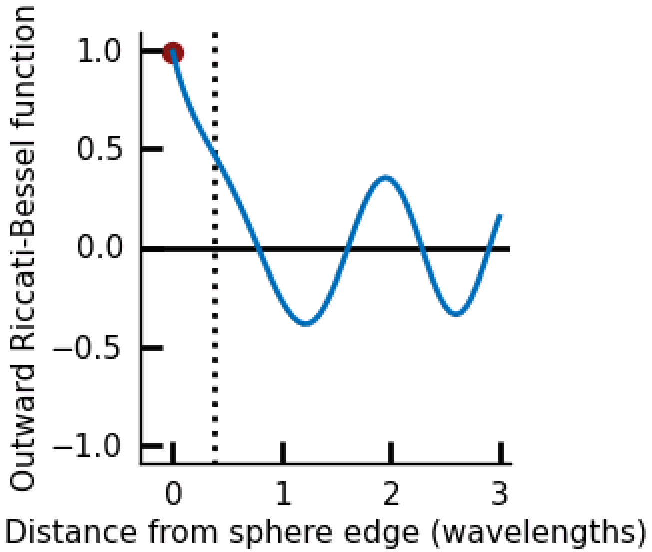
never does

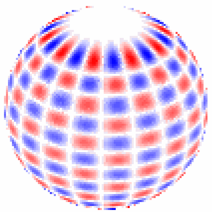
This is an artifact of the "infinite" extent of the plane wave





Snapshot in time of a spherical wave





Real part of outward (Riccati- Bessel) spherical wave

Starting spherical surface radius 2.9 wavelengths

Wave with n = 20, m = 10, escape radius 3.26 wavelengths

Note the angular shape is constant as the wave expands

Outward wave propagation

As time progresses

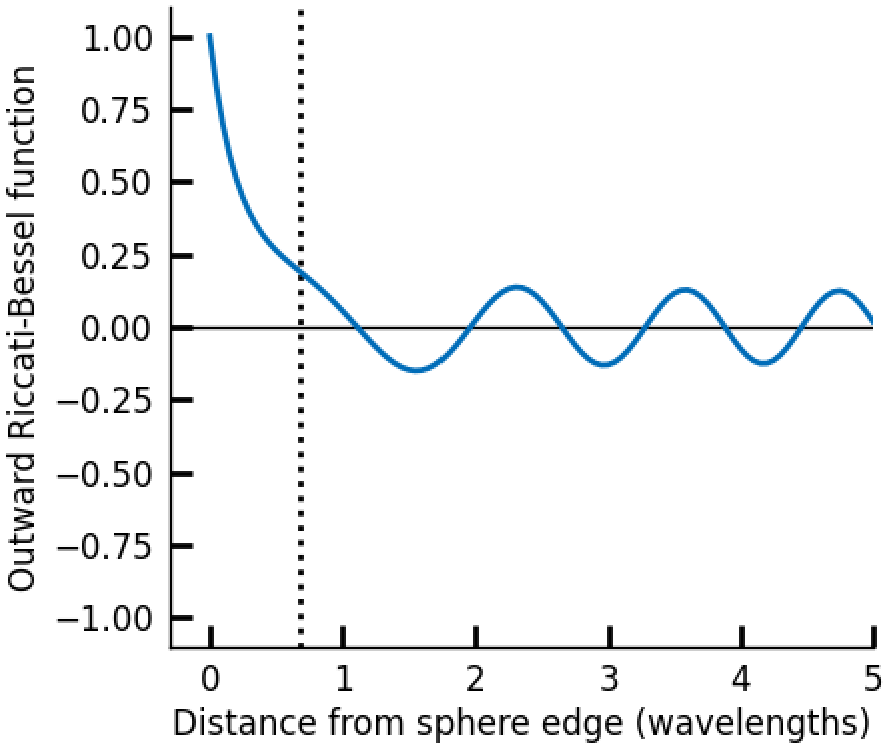
- the wave beyond the escape radius
  - propagates outwards

We plot the outward Riccati-Bessel wave

- as a function of time
  - technically the real part of

$$\xi_n(2\pi r)\exp(-i\omega t)$$

    - normalized to unit amplitude at the sphere edge
- for a sphere of radius 2.9 wavelengths
  - with $n = 22$
    - which has an escape radius of
      - 3.58 wavelengths



$$r_o = 2.9 \quad n = 22 \quad r_{escn} = 3.58$$

Outward wave propagation

As time progresses

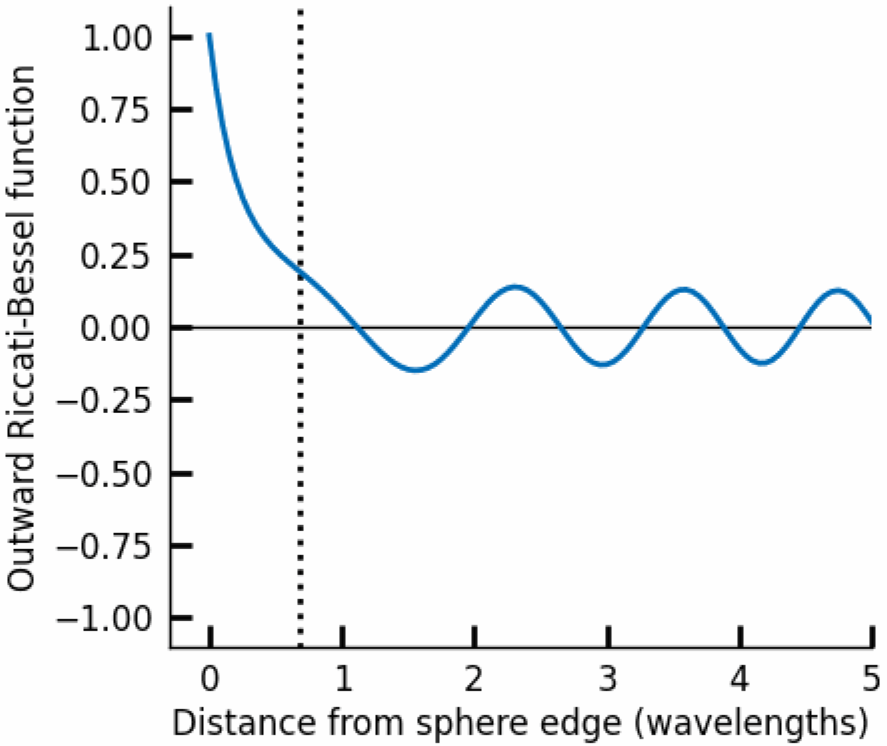
- the wave beyond the escape radius
  - propagates outwards

We plot the outward Riccati-Bessel wave

- as a function of time
  - technically the real part of

$$\xi_n(2\pi r)\exp(-i\omega t)$$

  - normalized to unit amplitude at the sphere edge
- for a sphere of radius 2.9 wavelengths
  - with $n = 22$
    - which has an escape radius of
      - 3.58 wavelengths



$r_o = 2.9 \quad n = 22 \quad r_{escn} = 3.58$

Spherical heuristic number

The threshold for tunneling is easy to characterize

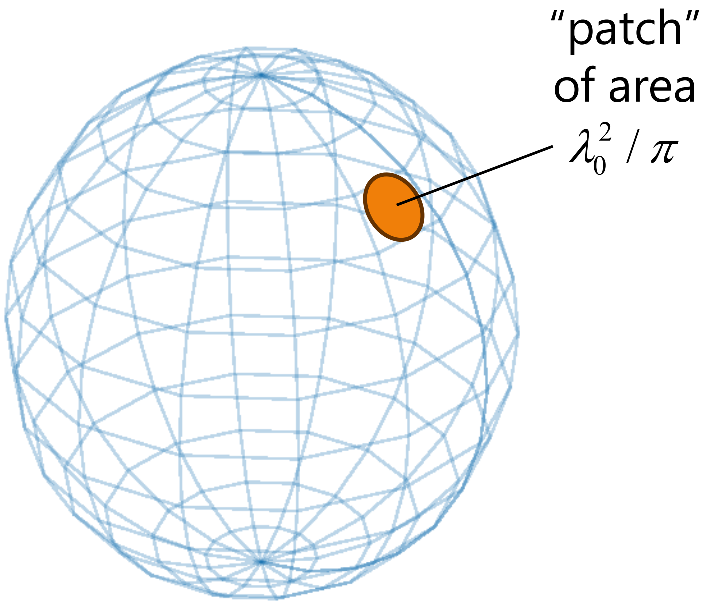
and gives a simple answer for the number of waves that do not need to tunnel

This is well approximated by the spherical heuristic number

$$N_{SH} = (kr_o)^2 \equiv \left(\frac{2\pi r_o}{\lambda_o}\right)^2 \equiv \frac{4\pi r_o^2}{(\lambda_o^2/\pi)} \equiv \frac{A_S}{(\lambda_o^2/\pi)}$$

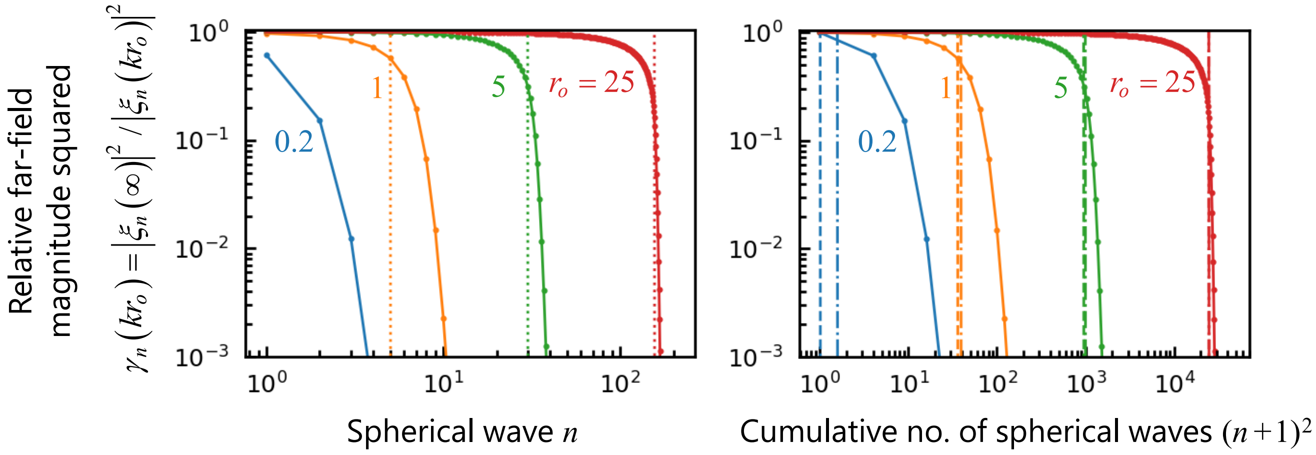
where A_S is the sphere area

so one "propagating" wave for every λ_o^2/π of surface area



D. A. B. Miller, Z. Kuang, O. D. Miller, "Tunneling escape of waves," Nat. Photon. **19**, 284–290 (2025)

Relative far-field magnitude squared



dotted lines – n_p
dashed lines – N_{SH}
dot-dashed lines - N_p
(the exact counting result)

As the size of the spherical surface increases

the cut-off becomes increasingly relatively abrupt

tending towards the "absolutely abrupt" cut-off of evanescent waves

Note the spherical heuristic number N_{SH} is a good approximation to the total exact number N_p of "propagating" waves even down to ~ 1 wavelength of radius

Defining the diffraction limit

We can now construct a precise definition of
the "diffraction limit"

For a wave interacting with a volume
the wave passes the diffraction limit
if any spherical component of the wave must tunnel to enter or leave the bounding spherical surface enclosing the volume

No n=0 electromagnetic outgoing waves

Electromagnetic waves behave mostly similarly to scalar (e.g., acoustic) waves

but, unlike for acoustic waves, there is no electromagnetic wave that is uniform in angle

Mathematically, there is no $n = 0$ wave in electromagnetism

If the first outgoing electromagnetic waves (so, for $n = 1$)

are not to require tunneling to escape

the bounding spherical volume must be at least

$$r_{esc1} = \lambda_o / \left(\sqrt{2}\,\pi\right) \simeq 0.225\lambda_o$$

in radius or, equivalently, in diameter

$$d = \sqrt{2}\lambda_o / \pi \simeq 0.45\lambda_o$$

(consistent with the well-known Chu limit on antenna Q)

Note: The escape radius for $n = 0$ acoustic waves is, however, zero

so, there is always one acoustic wave that can escape without tunneling

no matter how small the emitter or microphone

Perfect cloaking - An optical "white hole"?

- In this "white hole", incoming light
  - appears to be mostly "sucked" into the "white hole" in the middle
    - The phase fronts all "fall" rapidly into the "white hole"
  - and then the light is regenerated
    - The phase fronts rapidly re-emerge from the "white hole"
- How do we make this optical "white hole"?
  - Note: it may be simpler than you think

Perfect cloaking - An optical "white hole"?

So, what does it take to build this cloak?

Absolutely nothing

at least for this wave

If the wave is too complicated

i.e., if it is trying to violate the "diffraction limit"

it can't even effectively get into the volume

and it "reflects off free space"

This is the "inward wave" version of the tunneling escape

with the wave trying to tunnel to get in

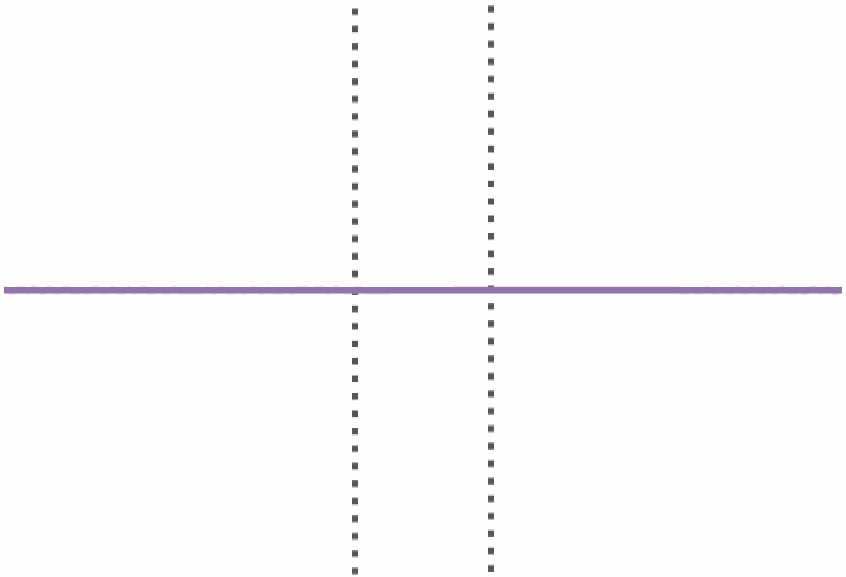
Interestingly, the pulse

looks as if it propagates right through!

See also Z Jacob, L V Alekseyev, and E Narimanov, Opt. Express 14, 8247 (2006)

Perfect cloaking?

Watch the blue dot, which propagates at the usual "phase velocity" of the wave



It appears to move right through the volume at a constant speed

Conclusions

There is a unified way of thinking about waves

based on waves from a spherical surface

from the propagating and evanescent fields of large optics

to the multipole expansions of antennas and nanophotonics

This approach gives a clear intuition

based on the onset of spherical wave tunneling

that

- explains how many waves can easily get in or out of a volume

  and why the fall-off is so abrupt past this number
- gives a rigorous and precise diffraction limit definition
- can also derive previous heuristic results

D. A. B. Miller, Z. Kuang, O. D. Miller, "Tunneling escape of waves," Nat. Photon. **19**, 284–290 (2025)

stanford.io/4oZy7bf

Funding from AFOSR grant FA9550-21-1-0312