nature photonics

Article

https://doi.org/10.1038/s41566-024-01578-w

Tunnelling escape of waves

Received: 13 November 2023

Accepted: 21 October 2024

Published online: 3 December 2024

Check for updates

David A. B. Miller [1] ✉, Zeyu Kuang [2] & Owen D. Miller [2]

Applications of waves in communications, information processing and sensing need a clear understanding of how many strongly coupled channels or degrees of freedom exist in and out of volumes of space and how the coupling falls off for larger numbers. Numerical results are possible, and some heuristics exist, but there has been no simple physical picture and explanation for arbitrary volumes. By considering waves from a bounding spherical volume, we show a clear onset of a tunnelling escape of waves that both defines a limiting number of well-coupled channels for any volume and explains the subsequent rapid fall-off of coupling strengths. The approach works over all size scales, from nanophotonics and small radiofrequency antennas up to imaging optics. It gives a unified view from the multipole expansions common for antennas and small objects to the limiting plane and evanescent waves of large optics, showing that all such waves can escape to propagation to some degree, by tunnelling if necessary, and gives a precise diffraction limit.

With emerging nanophotonics, increasingly we can design[1,2] and fabricate sophisticated objects down to wavelength sizes or below. Growing bandwidth demands in radiofrequency wireless communications require we exploit spatial channels more effectively. The resulting antenna systems are growing to increasingly sophisticated structures many wavelengths in size[3–5]. Information processing, such as for artificial intelligence, requires ever increasing numbers of channels, which optics could provide, whether for improved inference in neural networks[6] or more generally in optical interconnects[7] and processing[8]. The numbers of strong channels can determine how much and what kind of information we can measure in sensing applications such as microscopy or imaging or how many designable elements or basis functions we need in design.

We can quite rigorously define such independent channels, mathematically and physically, as 'communication modes'[9,10]. A core question, both for design and applications, is whether we can understand how many different (that is, orthogonal) waves or communication mode channels can propagate in and out of objects or volumes. Such counting is increasingly relevant in optics now that we can control and detect light mode by mode[11–13]. This question has arguably never had a simple answer, especially as we move through objects on scales of a few wavelengths. Although we may believe that diffraction limits lie behind such counting, and we can certainly calculate the orthogonal channels and their coupling strengths for any specific case[10], we lack any simple general model and intuition for important key behaviours (see the examples in ref. 10 and in many other analyses[9,14–18]). In particular, why do the coupling strengths tend to fall off rapidly past some number of well-coupled channels? And, indeed, just what defines that number?

We show here that there is a useful and unified approach with simple results and physical intuition. This approach spans continuously from sub-wavelength objects to large optical scales. A key concept is the idea of tunnelling escape of spherical waves. We find that the onset of this tunnelling can be precisely defined, allowing a clear counting of strongly coupled channels that propagate without tunnelling. Past this number, the tunnelling behaviour explains the fall-off and just how rapid it must be. The complementary problem of waves focusing into a volume similarly obeys a required onset of tunnelling that explains the difficulty of focusing past diffraction limits. The final step in our argument is that, for any wave from some finite volume, we can use a mathematical bounding sphere and the waves emerging from it to describe any wave that would leave (or enter) that bounding sphere; the sets of spherical waves are complete for describing all waves in free space or a uniform medium. Because we can establish the number of spherical waves that can propagate out of this bounding sphere without tunnelling, we have a simple estimate of an upper bound to the number of strongly coupled channels out of any finite volume, and we can

[1]Ginzton Laboratory, Stanford University, Stanford, CA, USA. [2]Department of Applied Physics and Energy Sciences Institute, Yale University, New Haven, CT, USA. ✉e-mail: dabm@stanford.edu

https://doi.org/10.1038/s41566-024-01578-w

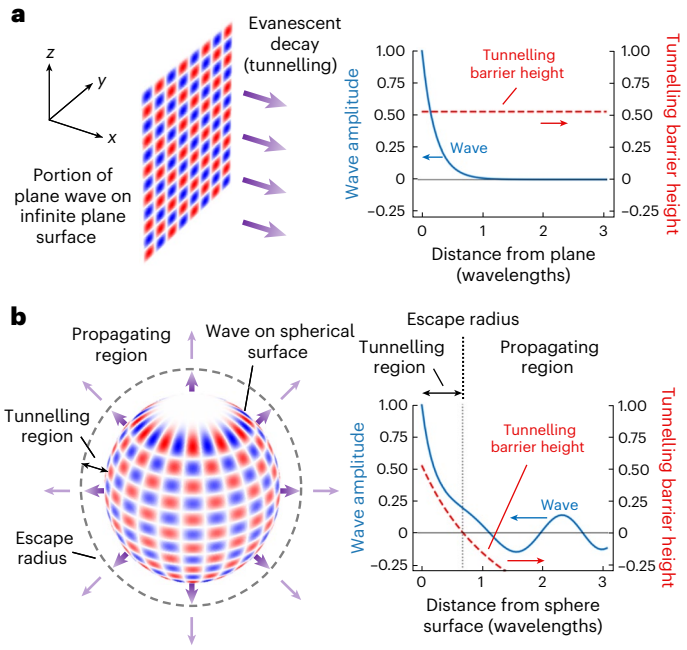

Fig. 1 | Comparison of an evanescent plane wave and an initially tunnelling spherical wave. a, A plane wave emerging from an infinite surface in the y–z plane, showing evanescent (tunnelling) decay in the positive x direction. **b**, A spherical wave emerging from a spherical surface—the initial tunnelling behaviour changes to propagating behaviour past the escape radius. The wave in each case is the amplitude of an oscillating wave as a snapshot in time. The wave in **a** is the simple evanescent, exponentially decaying plane wave. The wave in **b** is the amplitude as a function of radius with the underlying 1/radius fall-off of a spherically expanding wave removed, technically plotted for the phase where it corresponds to the C_n Riccati–Bessel function. Also shown is the effective tunnelling barrier height for the waves. For the plane wave, this is a positive constant. For the spherical wave, it falls off as 1/(radius)^2, becoming zero at the escape radius. When this barrier height is zero or negative, the wave is 'above' the tunnelling barrier, so it becomes propagating for all larger radii. (There is no significance to the point at which the wave and the tunnelling barrier height cross in **a**; these two are different graphs.) Both waves start with unit overall amplitude, at the plane surface for **a** and at the sphere surface for **b**, and with the same initial decay in distance. The plane wave in **a** has a transverse pattern corresponding to $n_z = 1.034$ and $n_y = 0.674$ periods per wavelength in the z and y directions, respectively, equivalent to $k_z = 2\pi n_z$ and $k_y = 2\pi n_y$ radians per wavelength. The spherical wave in **b** has a spherical harmonic angular form with $n = 22$ and $m = 12$. The spherical surface has radius $r_o = 2.9$ wavelengths. The escape radius is ~3.58 wavelengths, so ~0.68 wavelengths larger than r_o. The tunnelling barrier height starts at 0.524 in both cases, being $n_z^2 + n_y^2 - 1 \equiv [(k_z^2 + k_y^2)/k^2] - 1$ for plane waves and $[n(n+1)/(kr)^2] - 1$ for the spherical waves, for radial distance r.

understand that the coupling strength falls off quasi-exponentially past some number because the waves then have to tunnel to escape (or enter) the bounding volume. This onset of tunnelling also gives a precise definition of a diffraction limit for volumes.

Spherical waves and tunnelling

To set up our larger argument, we first describe spherical waves and their behaviour. Just as plane waves describe waves from (infinite) plane surfaces, so spherical waves based on spherical harmonic and spherical Bessel functions usefully describe waves from spherical surfaces or finite objects. Their mathematics is well understood, for example, in relation to scattering from spherical objects[19–23] and from multipole expansions of fields[24,25] (details are provided in Supplementary Texts 1 and 2).

Figure 1 illustrates the argument. For infinite plane waves (Fig. 1a), if the wave varies too rapidly in the y and z directions, we have evanescent decay in the x direction. Such a truly evanescent (literally 'vanishing') wave never escapes to propagate in the x direction. Figure 1b shows a spherical wave that, at a radial distance r_o from the centre, has an amplitude in angle given by a spherical harmonic function of order n. Such spherical waves retain their angular shape as they expand radially.

Suppose this spherical wave initially varies too rapidly (in transverse directions, on the sphere surface) compared to a wavelength. This wave still expands spherically, but must tunnel radially until it reaches some escape radius r_{escn} for this n. Such 'tunnelling' is characterized here by quasi-exponential decay, initially similar to the evanescent decay of the plane-wave case above. After the escape radius, the wave's transverse variation is slow enough that it can start to propagate, in a quasi-oscillatory spherical wave similar to the spatially oscillatory behaviour of a propagating plane wave.

Although the wave amplitude remaining after tunnelling may be quite small, that wave will continue expanding, settling to an ~1/radius decay of its amplitude, as in a spherically expanding propagating wave. So, waves from finite bodies, even if they start out somewhat evanescently, do not remain so. At least to some degree, they escape to propagation.

Scalar waves

To justify this argument, we analyse spherical waves, starting with the scalar case, with a Helmholtz wave equation:

$$\nabla^2 U(\mathbf{r}) + k^2 U(\mathbf{r}) = 0 \tag{1}$$

for a wave $U(\mathbf{r})$ of frequency f_o in a uniform medium such as vacuum, air or an isotropic dielectric. (For some wave velocity v, the wavevector magnitude $k = 2\pi/\lambda_o = \omega/v$, where the wavelength and angular frequency are $\lambda_o = v/f_o$ and $\omega = 2\pi f_o$, respectively.) Using complex waves with time dependence $\exp(-i\omega t)$, first we note that a plane wave solution has the form $U(\mathbf{r}) \propto \exp(i\mathbf{k}\cdot\mathbf{r})$ in space, with wavevector $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}$, where k_x, k_y and k_z are the wavevector components in the x, y and z directions (with corresponding unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$). Such a solution $\exp(i\mathbf{k}\cdot\mathbf{r}) \equiv \exp(ik_x x)\exp(ik_y y)\exp(ik_z z)$, of equation (1), implicitly in infinite space, is separable as $U(\mathbf{r}) = X(x)Y(y)Z(z)$, with $k_x^2 + k_y^2 + k_z^2 = k^2$. If $k_y^2 + k_z^2 > k^2$, then $X(x) \propto \exp(-\kappa x)$, where $\kappa = \sqrt{k_y^2 + k_z^2 - k^2}$; presuming any sources are on the 'left', in the region with $x < 0$, we take the positive square root. This is the classic evanescent wave, decaying exponentially for all $x > 0$. Note that it never becomes a propagating wave for any positive x, and the change from propagating to evanescent is totally abrupt as soon as $k_y^2 + k_z^2 > k^2$.

We can write the separated equation for $X(x)$ (Supplementary Text 1.1) as

$$-\frac{\mathrm{d}^2\xi(\rho)}{\mathrm{d}\rho^2} + V(\rho)\xi(\rho) = E\xi(\rho) \tag{2}$$

where $\rho = k_x x$ and $X(x) \equiv \xi(\rho)$, which is in the form of a one-dimensional Schrödinger equation[26], with a 'potential' energy $V(\rho) = (k_y^2 + k_z^2)/k^2$ (here actually independent of ρ), and an 'eigenenergy' $E = 1$. Quite generally, an equation like equation (2) can be trivially rewritten as

$$\frac{\mathrm{d}^2\xi(\rho)}{\mathrm{d}\rho^2} = [V(\rho) - E\,]\xi(\rho) \tag{3}$$

So, if $V(\rho) > E$ (a positive tunnelling barrier height), the second derivative has the same sign as the function $\xi(\rho)$. That means, for positive values of the function, the function is curving up for increasing ρ (or, for negative function values it is curving down), which is a characteristic we see in exponential curves, and we can call this case 'tunnelling' or (quasi) exponential. Conversely, if $V(\rho) < E$ (a negative tunnelling barrier height), the second derivative has the opposite sign from the function. Then, if the function is positive, it is curving down, forcing it to cross the axis to become negative with increasing ρ. Once it is

negative, it then starts curving up, forcing it to cross the axis again to become positive, and so on. This is a characteristic seen in oscillatory functions, like sine waves, and we call this case 'propagating' or (quasi) oscillatory. (Other recent work[27] has explored an abstract quantum tunnelling approach with finite plane surfaces, though our spherical approach here is different.)

The scalar Helmholtz equation (1) can also be solved in spherical polar coordinates, r, θ and ϕ (for example, ref. 25; Supplementary Text 1.2). Specifically, one can look for separable solutions $U(\mathbf{r}) = R(r)Y(\theta, \phi)$. The angular solutions $Y(\theta, \phi)$ are the spherical harmonics $Y_{nm}(\theta, \phi)$, with $n = 0, 1, 2, \ldots$ and, using the complex forms, $-n \le m \le n$ (for example, ref. 26). The radial solutions $R(r)$ are the spherical Bessel functions $z_n(kr)$, also for $n = 0, 1, 2, \ldots$. Using a dimensionless radial variable $\rho = kr$, these radial functions satisfy the differential equation

$$\rho^2 \frac{\mathrm{d}^2 z_n(\rho)}{\mathrm{d}\rho^2} + 2\rho \frac{\mathrm{d}z_n(\rho)}{\mathrm{d}\rho} + (\rho^2 - n(n+1))z_n(\rho) = 0 \tag{4}$$

Spherical Bessel functions of the first and second kinds, respectively $j_n(\rho)$ and $y_n(\rho)$, give two (real) independent solutions of equation (4). For a given n, linear combinations of these are also solutions. In particular, the spherical Hankel function of the first kind

$$h_n^{(1)}(\rho) = j_n(\rho) + \mathrm{i}y_n(\rho) \tag{5}$$

corresponds to outward-propagating waves at large distances. (The notation $z_n(\rho)$ stands in for any of $j_n(\rho)$, $y_n(\rho)$ or $h_n^{(1)}(\rho)$ in equation (4).) All these solutions have an underlying $1/r$ or $1/\rho$ dependence at large r or ρ, which corresponds to them falling off ultimately as spherically expanding waves. Indeed, specifically,

$$h_n^{(1)}(\rho) \to \mathrm{i}\frac{\exp(\mathrm{i}\rho)}{\rho} \equiv \frac{\mathrm{i}}{k}\frac{\exp(\mathrm{i}kr)}{r} \tag{6}$$

is a spherically expanding propagating wave as $r \to \infty$.

We can multiply by radius to take out this underlying $1/r$ or $1/\rho$ dependence, giving radial solutions then expressed using what are known as Riccati–Bessel functions, $S_n(\rho) = \rho j_n(\rho)$ and $C_n(\rho) = -\rho y_n(\rho)$, and in particular, the 'outgoing' Riccati–Bessel function

$$\xi_n(\rho) = \rho h_n^{(1)}(\rho) \equiv S_n(\rho) - \mathrm{i}C_n(\rho) \tag{7}$$

Such functions can be convenient for viewing the wave more as a function of angle rather than transverse position at any radius.

There is, however, one other very important consequence of using a Riccati–Bessel functions like $\xi_n(\rho)$: they obey the Riccati–Bessel differential equation, which can be written

$$-\frac{\mathrm{d}^2\xi_n(\rho)}{\mathrm{d}\rho^2} + \frac{n(n+1)}{\rho^2}\xi_n(\rho) = \xi_n(\rho) \tag{8}$$

(Note that substituting from equation (7) in equation (8) will recover equation (4).) This equation is now in exactly the Schrödinger form as in equation (2). 'Eigenenergy' E is equal to 1 as before, but now the effective radial potential is

$$V(\rho) = \frac{n(n+1)}{\rho^2} \tag{9}$$

which falls off (as $1/\rho^2$) with radius ρ or r ($= \rho/k$). We can usefully now define the 'escape radius' as

$$\begin{aligned} \rho_{\mathrm{esc}n} &= \sqrt{n(n+1)}\ \text{or, equivalently,} \\ r_{\mathrm{esc}n} &= \frac{\sqrt{n(n+1)}}{k} \equiv \frac{\lambda_\mathrm{o}}{2\pi}\sqrt{n(n+1)} \end{aligned} \tag{10}$$

which marks the boundary between tunnelling and propagating behaviour for the wave, that is, the point r or ρ at which $V(kr) = E(=1)$. So, we have the following outgoing wave solutions:

$$\begin{aligned} U_{nm}(\mathbf{r}) &= h_n^{(1)}(kr)Y_{nm}(\theta, \phi) \\ &\equiv \frac{\xi_n(kr)}{kr}Y_{nm}(\theta, \phi)\ \ n = 0, 1, 2, \ldots, -n \le m \le n \end{aligned} \tag{11}$$

If an outward wave has an angular form $Y_{nm}(\theta, \phi)$, then for $r < r_{\mathrm{esc}n}$, this wave (in the Riccati–Bessel form $\xi_n(kr)$) is tunnelling outwards, but once it passes the escape radius $r_{\mathrm{esc}n}$, it becomes propagating, escaping at least to some degree. This contrasts with (infinite) plane waves of the form $U(\mathbf{r}) \propto \exp(i\mathbf{k}\cdot\mathbf{r})$. If such waves start out as evanescent (so with $k_y^2 + k_z^2 > k^2$), they remain evanescent at all x, eventually vanishing completely.

The wave plotted in Fig. 1b is such a Riccati–Bessel function, showing tunnelling-like behaviour up to the escape radius, and propagating behaviour for larger radii. (See Supplementary Fig. 5 in the Supplementary Text for the behaviour as a function of time.)

Although this transition between tunnelling and propagating behaviour at the escape radius is clear from differential equation (8), it is not at all obvious from the usual algebraic expressions for spherical Bessel or Riccati–Bessel functions, which involve series of inverse powers of the radius together with sine and cosine functions (for example, ref. 25, p. 426); this may be why this tunnelling behaviour is not already better known.

Incidentally, the scalar spherical waves found in this way are also the waves associated with the communication modes[9,10] between spherical surfaces (Supplementary Text 3).

Vector electromagnetic waves

There are three different forms of vector wave solutions to the vector Helmholtz equation (Supplementary Text 2), with one 'longitudinal' and two 'transverse' polarizations. The angular aspects are describable based on three vector spherical harmonic functions. Importantly, the tunnelling in the radial behaviour $R(r)$ described above for scalar waves persists into the vector cases.

Although the longitudinal wave exists for sound and elastic waves, electromagnetic waves have just two, transverse forms. Each of those is separable into radial and angular parts, with the radial parts obeying the same equation as the function $R(r)$ above, but with the angular part being a vector spherical harmonic function. Explicitly, for outgoing waves, we have a set of transverse electric (TE) waves, with the electric field (from equation (134) in the Supplementary Text) given by

$$\begin{aligned} \mathbf{E}_{nm}^{(\mathrm{TE})}(r, \theta, \phi) &= \mathrm{i}\sqrt{\frac{\mu}{\varepsilon}}\, h_n^{(1)}(kr)\mathbf{C}_{mn}(\theta, \phi) \\ &\equiv \mathrm{i}\sqrt{\frac{\mu}{\varepsilon}}\,\frac{\xi_n(kr)}{kr}\mathbf{C}_{mn}(\theta, \phi)\ \ n = 1, 2, \ldots, -n \le m \le n \end{aligned} \tag{12}$$

and transverse magnetic (TM) waves, with the magnetic field (from equation (138) in the Supplementary Text) given by

$$\begin{aligned} \mathbf{H}_{nm}^{(\mathrm{TM})}(r, \theta, \phi) &= \mathrm{i}\, h_n^{(1)}(kr)\mathbf{C}_{mn}(\theta, \phi) \\ &\equiv \mathrm{i}\,\frac{\xi_n(kr)}{kr}\mathbf{C}_{mn}(\theta, \phi)\ \ n = 1, 2, \ldots, -n \le m \le n \end{aligned} \tag{13}$$

Here, $\mathbf{C}_{mn}$ is the vector spherical harmonic function

$$\begin{aligned} \mathbf{C}_{mn}(\theta, \phi) &= \nabla \times [\mathbf{r}Y_{nm}(\theta, \phi)] \\ &\equiv \nabla Y_{nm}(\theta, \phi) \times \mathbf{r}\ n = 1, 2, \ldots,\ \ -n \le m \le n \end{aligned} \tag{14}$$

https://doi.org/10.1038/s41566-024-01578-w

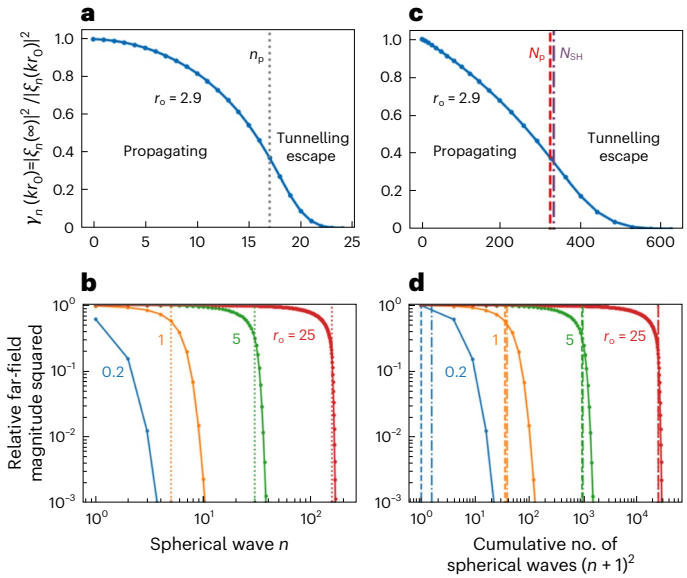

Fig. 2 | Far-field spherical wave strength with increasing order *n*. a–**d**, Relative far-field strength $\gamma_n(kr_o) = |\xi_n(\infty)|^2/|\xi_n(kr_o)|^2$ of the outgoing Riccati–Bessel function of order n for various values of the starting radius r_o in wavelengths, plotted in **a** and **b** against n, and in **c** and **d** parametrically against the total number $(n+1)^2$ of scalar spherical waves up to and including waves of order n. Solid lines are to guide the eye. n_p (vertical dotted lines on **a** and **b**) is the largest n for which all waves start out as propagating from the given radius r_o. $N_p = n_p(n_p+2)$ (vertical dashed lines in **c** and **d**) is the total number of such waves. $N_{SH} = (kr_o)^2$ (vertical dash-dotted lines on **c** and **d**) is the spherical heuristic number. **a** and **c** use linear scales, with $r_o = 2.9$ wavelengths as in Fig. 1. (The $n = 0$ point, which does not exist for electromagnetic waves, is shown in **a** and **c**.) For $r_o = 2.9$ wavelengths, $n_p = 17$; $N_p = 323$; $N_{SH} \simeq 332.01$. **b** and **d** show results for r_o of 0.2, 1, 5 and 25 wavelengths on log scales. The values of n_p, N_p and N_{SH} for $r_o = 0.2, 1, 5, 25$ are, respectively, $n_p = 0, 5, 30, 156$; $N_p = 1, 35, 960, 24{,}648$; $N_{SH} \simeq 1.58, 39.48, 986.96, 24{,}674.01$.

Note explicitly that radial behaviour is given by functions $h_n^{(1)}(kr)$ or $\xi_n(kr)$, just as in the scalar case. The main differences are as follows:

- We have two waves for each choice of n and m, which we can view as being TE- and TM-polarized waves, respectively.
- Their angular form, which is vectorial, is based on the gradient ∇Y_{nm} of the spherical harmonic rather than just the (scalar) spherical harmonic Y_{nm} directly. Because $Y_{nm}(\theta, \phi)$ only depends on angles, not radius, ∇Y_{nm} only has vector components in the $\hat{\theta}$ and $\hat{\phi}$ directions on the sphere surface, so it and the function $\mathbf{C}_{mn}(\theta, \phi)$ contain only transverse components—vectors that lie in the sphere surface.
- In contrast to the scalar solutions (equation (11)), there are no electromagnetic waves for $n = 0$.

The electric field $\mathbf{E}_{nm}^{(TM)}$ corresponding to $\mathbf{H}_{nm}^{(TM)}$ and the magnetic field $\mathbf{H}_{nm}^{(TE)}$ corresponding to $\mathbf{E}_{nm}^{(TE)}$ are given by Supplementary equations (137) and (133), respectively. The TE and TM waves for a given n and m describe perpendicularly polarized waves, with $\mathbf{E}_{nm}^{(TM)}$ perpendicular to $\mathbf{E}_{nm}^{(TE)}$, and similarly for the magnetic fields $\mathbf{H}_{nm}^{(TE)}$ and $\mathbf{H}_{nm}^{(TM)}$. The polarizations of both $\mathbf{E}_{nm}^{(TE)}$ and $\mathbf{H}_{nm}^{(TM)}$ are always transverse (perpendicular to the radius vector $\mathbf{r}$), and the corresponding fields $\mathbf{E}_{nm}^{(TM)}$ and $\mathbf{H}_{nm}^{(TE)}$ are transverse in the far field, though they have radial components in the near field.

This approach corresponds exactly to the multipole expansion of outward-propagating electromagnetic fields. We write fields in forms equivalent to Jackson's definitions[25], and derivations of equations (12) and (13) can be found there[25] (Supplementary Text 2 provides general vector wave derivations). Our tunnelling analysis conceptually connects multipole and quasi-evanescent behaviours in one formalism. These vector spherical waves are the waves associated with the (vector) communication modes between spherical surfaces or volumes[14].

Note that our analysis also tells us that, to observe strong radiation from high-n multipoles from some object, to avoid having to tunnel to escape the object would have to be essentially of the scale of twice the escape radius (so the diameter of the bounding sphere) for that n.

Counting waves from spherical surfaces

The escape radius allows a useful counting of waves associated with a spherical surface of radius r_o. There is a maximum value n_p of n for which all the associated spherical harmonic waves propagate without tunnelling to escape, which requires $r_o > r_{escn}$ for a given n. From equation (10), solving the quadratic equation $n(n+1) = (kr_o)^2$, the largest n for which $kr_o > \sqrt{n(n+1)}$ is

$$n_p(r_o) = \text{floor}\left[\sqrt{(1/4) + N_{SH}} - (1/2)\right] \tag{15}$$

(floor(u) is the largest integer $\le u$). Here, N_{SH} is the 'spherical heuristic number':

$$N_{SH} = (kr_o)^2 \equiv \frac{4\pi r_o^2}{(\lambda_o^2/\pi)} = \frac{A_S}{(\lambda_o^2/\pi)} \tag{16}$$

where $A_S = 4\pi r_o^2$ is the area of the spherical surface. (The concept of N_{SH} was introduced empirically in ref. 10; this algebraic definition of N_{SH} also emerged in analytic work on spherical waves[14]. Note too that the rule of thumb of using n up to ~kr_o in spherical wave scattering expansions (ref. 28, p. 126) is nearly the same as these results.) Note that, for $kr_o < \sqrt{2}$, or, equivalently, $N_{SH} < 2$ or $r_o < r_{esc1}$, where

$$r_{esc1} = \lambda_o/\left(\sqrt{2}\pi\right) \simeq 0.225\lambda_o \tag{17}$$

only the $n = 0$ wave propagates without requiring tunnelling escape. Because there are no $n = 0$ electromagnetic waves, all electromagnetic waves from (bounding) volumes smaller than this radius r_{esc1} must tunnel to escape. This difficulty that electromagnetic waves have in getting out of small volumes is consistent with the well-known Chu antenna limit[29], which states that the antenna Q-factor must

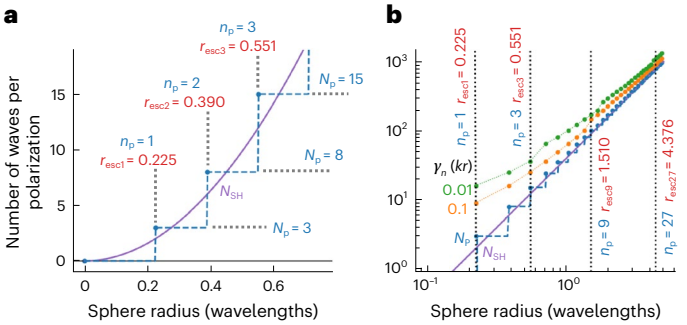

Fig. 3 | Number of waves (per polarization for electromagnetic waves) that start out by propagating, as a function of the sphere radius. This number is shown as the stepped dashed line. The solid line is the spherical heuristic number N_{SH}, showing it is a good approximation, even down to small radii. Also shown are corresponding values of n_p, N_p and the escape radius r_{escn} (in wavelengths). **a**, Results on a linear scale. **b**, Results on a log scale, with example $\gamma_n(kr_o) = |\xi_n(\infty)|^2/|\xi_n(kr_o)|^2$ values as the relative far-field magnitude squared of the spherical wave, with numbers of waves shown for $\gamma_n(kr_o) > 0.1$ and >0.01. Dotted lines connecting points are just to guide the eye. Also indicated are several example n_p values (1, 3, 9 and 27) and their corresponding escape radii r_{escn}, in wavelengths.

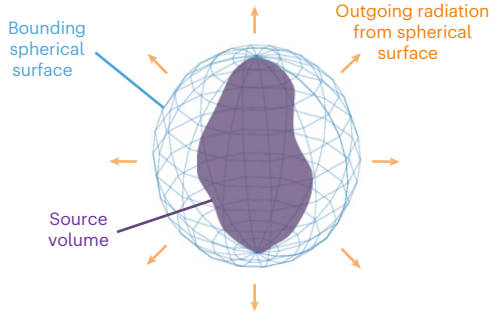

Fig. 4 | A spherical bounding surface that just encloses some source volume or object. Any waves from the source volume that reach the spherical surface can be described in terms of spherical waves on that surface.

increase for small antennas. Indeed, we expect that the energy storage in the quasi-exponential 'tunnelling' part of these waves should form part of the reactive field around such small antennas. Note again that there is no corresponding requirement for scalar waves, such as normal sound waves, which is consistent with small microphones and loudspeakers not having a corresponding limit and high Q-factor requirement, and hence remaining quite effective even for deeply sub-wavelength sizes.

We can now calculate the total number N_{ps} of (scalar) spherical harmonic waves with values of n up to (and including) n_p. Because there are $2n+1$ different m values (and hence spherical harmonics Y_{nm}) for each n, it is the case that

$$N_{ps} = 1 + 3 + 5 + \cdots (2n_p + 1) \equiv \sum_{q=0}^{n_p} (2q+1) = (n_p+1)^2 \qquad (18)$$

For electromagnetic waves, with no $n = 0$ waves, we remove that one wave, giving

$$N_p = N_{ps} - 1 = n_p(n_p + 2) \qquad (19)$$

which is the number of such electromagnetic waves per polarization.

Because the Riccati–Bessel $\xi_n(kr)$ functions take out the underlying $1/r$ decay of spherical waves, a simple metric for how effectively a spherical wave of index n propagates outwards from a radius r_o is the relative far-field magnitude squared:

$$\gamma_n(kr_o) = |\xi_n(\infty)|^2 / |\xi_n(kr_o)|^2 \qquad (20)$$

Incidentally, for a given n, this coupling γ_n is the same for every m value; this might seem surprising given the different associated shapes of Y_{nm}, although analogous behaviours can be seen in other spherical problems, such as the independence of the hydrogen-atom energies from the corresponding m quantum number[26]. Note, too, that n is always the total number of nodal circles in a spherical harmonic, independent of the m value.

In Fig. 2a,c, we plot $\gamma_n(kr_o)$ against n, and in Fig. 2b,d parametrically against the cumulative number $(n+1)^2$ of scalar waves (equation (18)). (Supplementary Fig. 6 also presents curves like Fig. 2a,b for different radii of spherical surfaces.) We see several interesting and useful behaviours in Fig. 2.

- For small volumes—for example, a few wavelengths or smaller in radius—relatively quite a few waves can escape by tunnelling with usefully large (for example, $>10^{-2}$) propagating amplitudes.
- The spherical heuristic number N_{SH} is a good approximation to the total number N_p of waves (per polarization) that start out as propagating.
- As the radius of the spherical volume increases, a smaller fraction of the waves can usefully escape by tunnelling, compared to those that start out as propagating. So, the transition to waves that must tunnel to increasing weak escape is increasingly relatively abrupt for larger volumes. So, in practice, with increasing size, we tend towards a simpler categorization of waves being either propagating or (approximately) evanescent.

The actual number of waves that propagate strictly without tunnelling is a series of steps of integer height as a function of the radius of the spherical surface (Fig. 3). The number N_{SH} is a continuous smooth function that passes through those steps (Fig. 3). Hence, it can, overall, be a good simple estimate for this number N_p.

Note that N_{SH} in equation (16) is also the area $A_S = 4\pi r_o^2$ of the spherical surface divided by an area λ_o^2/π. So N_{SH} corresponds to one such wave for every λ_o^2/π of area on the sphere, connecting this spherical wave behaviour to heuristic diffraction limits in conventional optical systems with more planar surfaces, which limit focal spots to approximately a (square) half wavelength because of diffraction. Note, too, that at moderate to large n, we have $n_p \simeq \sqrt{N_{SH}} = 2\pi r_o/\lambda_o$, which is the circumference of the sphere in wavelengths.

Figure 3 also shows the number of waves that couple to the far field with coupling γ_n (equation (20)) greater than specific values of 0.1 and 0.01. (These would correspond to horizontal lines at 10^{-1} and 10^{-2} on Fig. 2d.) Some such waves will be tunnelling to escape, but might still be practically useful, for example, for communications or sensing. We see that there are quite substantial numbers of such additional waves for a small sphere radius, though relatively fewer for larger radii.

Although we consider only outgoing waves explicitly here, in practice these same numbers of waves also correspond essentially to the number of incoming waves that can penetrate into an empty spherical (bounding) volume (Supplementary Text 4). Spherical waves beyond this limit will essentially reflect off an empty spherical volume of free space (becoming standing waves), so objects smaller than this volume are essentially invisible to, or 'self-cloaked'[30] from, such waves.

Waves from arbitrary volumes

Now we can formally complete our argument on counting strongly coupled waves from arbitrary volumes. The sets of scalar and electromagnetic spherical waves we have constructed are complete for describing any (outgoing) wave on a spherical surface. Hence, they can also describe any wave emerging from sources in a volume enclosed by that bounding spherical surface (Fig. 4). Because we have established how many orthogonal basis-set waves can emerge from this spherical surface without tunnelling, we have established an upper bound on the maximum number of orthogonal waves that could emerge from the source volume without tunnelling. We could also extend this estimate to allow for waves that could tunnel from the spherical surface to escape to some specified degree.

Heuristic result for restricted solid angles

Once the spherical surface becomes several wavelengths or larger in size, the number of waves (per polarization) that propagate without tunnelling becomes quite large. At radii of ~3, 5 and 10 wavelengths, for example, these numbers are ~350, 1,000 and 4,000, respectively. So, we can divide N_{SH} by 4π steradians to obtain, for a spherical surface of radius r_S, the (approximate) number of propagating waves per unit solid angle:

$$N_\Omega \simeq \frac{\pi r_S^2}{\lambda_o^2} \qquad (21)$$

Imagine, then, some receiving surface of area A_R at some (perpendicular) distance L from the centre of the spherical surface (Fig. 5), so

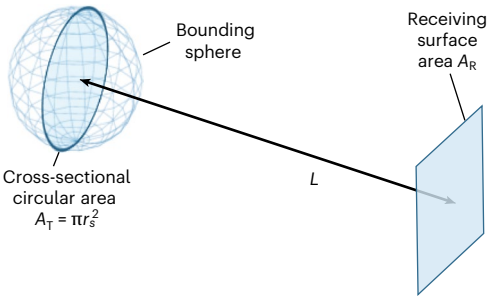

Fig. 5 | Construction for waves to a finite receiving surface. A bounding spherical volume of radius r_S surrounds a source of circular cross-sectional area $A_T = \pi r_S^2$ communicating to a receiving surface of area A_R at a distance L.

subtending a solid angle $\Omega \simeq A_R/L^2$. Then from equation (21) we could estimate a total number of propagating wave channels

$$N_H = \frac{\pi r_S^2}{\lambda_o^2}\frac{A_R}{L^2} = \frac{A_T A_R}{\lambda_o^2 L^2} \tag{22}$$

where $A_T = \pi r_S^2$, which we note is the apparent circular cross-sectional area of the source volume (Fig. 5). We note that equation (22) is exactly the number previously deduced as the 'paraxial heuristic number' of well-coupled channels between planar source and receiver spaces[10] (equation (64) in ref. 10). (See also appendix A in ref. 10 for the many heuristic derivations of the number.) So, if we count only those waves that do not require tunnelling to escape, we can derive previous heuristic results for the 'diffraction-limited' number of channels in paraxial optical systems. (Note, too, that N_Ω, equation (21), is not itself restricted to paraxial cases.) These approaches could also be viewed as asking for the approximate number of these (spherical harmonic) basis functions required to adequately describe the resulting possible waves on the receiving surface from such a source volume.

Discussion and conclusions

We have shown a unified way of thinking about waves in and out of volumes, from the propagating and evanescent fields of large optics to the multipole expansions of antennas and nanophotonics. This is based rigorously on the propagation of spherical waves associated with the spherical bounding surface around some volume or object. For all volumes from approximately a wavelength scale and upwards, a maximum number of well-coupled waves or orthogonal channels is understood as those that do not have to tunnel to escape the spherical surface. This onset of tunnelling corresponds to a 'knee' in coupling strength after which coupling falls rapidly because of the tunnelling. A corresponding escape radius $r_{escn} = (\sqrt{n(n+1)})\lambda_o/2\pi$ characterizes the largest order n of spherical wave of wavelength λ_o that can propagate from such a spherical surface without initially tunnelling. A spherical heuristic number N_{SH}, corresponding to one wave for every λ_o^2/π of area on the spherical surface, approximately but usefully characterizes the number of well-coupled waves (per polarization) and the position of the 'knee' in coupling strength. With increasing radius of the volume, the relative fall-off from tunnelling becomes progressively more abrupt, asymptoting towards the complete abruptness of the onset of truly evanescent behaviour for infinite plane waves. Note, though, that such truly evanescent waves are an artefact of the assumption of infinite plane waves; all corresponding waves from finite bodies eventually escape to some degree by tunnelling. (Note, incidentally, that similar radial tunnelling, propagating and escape behaviour can be derived for circular and cylindrical waves (Supplementary Text 5).)

Finally, this approach allows us to propose 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.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-024-01578-w.

References

1. Jiao, Y., Fan, S. & Miller, D. A. B. Demonstration of systematic photonic crystal device design and optimization by low-rank adjustments: an extremely compact mode separator. *Opt. Lett.* **30**, 141–143 (2005).
2. Molesky, S. et al. Inverse design in nanophotonics. *Nat. Photon.* **12**, 659–670 (2018).
3. Ouyang, C., Liu, Y., Zhang, X. & Hanzo, L. Near-field communications: a degree-of-freedom perspective. Preprint at https://arxiv.org/abs/2308.00362 (2023).
4. Gong, T. et al. Holographic MIMO communications: theoretical foundations, enabling technologies and future directions. *IEEE Commun. Surv. Tutor.* **26**, 196–257 (2024).
5. Wang, Z. et al. Extremely large-scale MIMO: fundamentals, challenges, solutions and future directions. *IEEE Wirel. Commun.* **31**, 117–124 (2023).
6. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
7. Miller, D. A. B. Attojoule optoelectronics for low-energy information processing and communications. *J. Lightwave Technol.* **35**, 346–396 (2017).
8. McMahon, P. L. The physics of optical computing. *Nat. Rev. Phys.* **5**, 717–734 (2023).
9. Miller, D. A. B. Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths. *Appl. Opt.* **39**, 1681–1699 (2000).
10. Miller, D. A. B. Waves, modes, communications and optics: a tutorial. *Adv. Opt. Photon.* **11**, 679–825 (2019).
11. Miller, D. A. B. Self-configuring universal linear optical component. *Photon. Res.* **1**, 1–15 (2013).
12. Fontaine, N. K. et al. Photonic lanterns, 3-D waveguides, multiplane light conversion and other components that enable space-division multiplexing. *Proc. IEEE* **110**, 1821–1834 (2022).
13. Bogaerts, W. et al. Programmable photonic circuits. *Nature* **586**, 207–216 (2020).
14. Kuang, Z., Miller, D. A. B. & Miller, O. D. Bounds on the coupling strengths of communication channels and their information capacities. Preprint at https://doi.org/10.48550/arXiv.2205.05150 (2022).
15. Sekehravani, E. A., Leone, G. & Pierri, R. Evaluation of the number of degrees of freedom of the field scattered by a 3D geometry. *Sensors* **23**, 4056 (2023).
16. Pizzo, A. & Lozano, A. On Landau's eigenvalue theorem for line-of-sight MIMO channels. *IEEE Wirel. Commun. Lett.* **11**, 2565–2569 (2022).
17. Solimene, R., Maisto, M. A., Romeo, G. & Pierri, R. On the singular spectrum of the radiation operator for multiple and extended observation domains. *Int. J. Antennas Propag.* **2013**, e585238 (2013).
18. Migliore, M. D. On the role of the number of degrees of freedom of the field in MIMO channels. *IEEE Trans. Antennas Propag.* **54**, 620–628 (2006).
19. Mie, G. Beiträge zur optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* **330**, 377–445 (1908).

20. Debye, P. Der Lichtdruck auf Kugeln von beliebigem Material. *Ann. Phys.* **335**, 57–136 (1909).
21. Hancu, I. M., Curto, A. G., Castro-López, M., Kuttge, M. & van Hulst, N. F. Multipolar interference for directed light emission. *Nano Lett.* **14**, 166–171 (2014).
22. Tzarouchis, D. & Sihvola, A. Light scattering by a dielectric sphere: perspectives on the Mie resonances. *Appl. Sci.* **8**, 184 (2018).
23. Dorodnyy, A., Smajic, J. & Leuthold, J. Mie scattering for photonic devices. *Laser Photon Rev.* **17**, 2300055 (2023).
24. Hansen, W. W. A new type of expansion in radiation problems. *Phys. Rev.* **47**, 139–143 (1935).
25. Jackson, J. D. *Classical Electrodynamics* (Wiley, 1999).
26. Miller, D. A. B. *Quantum Mechanics for Scientists and Engineers* (Cambridge Univ. Press, 2008).
27. Creagh, S. C. & Gradoni, G. Slepian eigenvalues as tunnelling rates. *Ann. Phys.* **449**, 169204 (2023).
28. Bohren, C. F. & Huffman, D. R. *Absorption and Scattering of Light by Small Particles* (Wiley, 1983).
29. Wheeler, H. Small antennas. *IEEE Trans. Antennas Propag.* **23**, 462–469 (1975).
30. Miller, D. A. B. On perfect cloaking. *Opt. Express* **14**, 12457–12466 (2006).

Data availability
All associated data and materials are available in the paper.

Acknowledgements
D.A.B.M. was supported by AFOSR grant no. FA9550-21-1-0312. Z.K. and O.D.M. were supported by AFOSR grant no. FA9550-22-1-0393 and by the Simons Collaboration on Extreme Wave Phenomena Based on Symmetries (award no. SFI-MPS-EWP-00008530-09).

Author contributions
D.A.B.M. wrote the paper with input and comments from all authors, performed the calculations and generated the graphics. Z.K. and O.D.M. contributed the approach to spherical and cylindrical wave descriptions. D.A.B.M. contributed the tunnelling interpretation. All authors reviewed and edited the paper before submission.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41566-024-01578-w.

Correspondence and requests for materials should be addressed to David A. B. Miller.

Peer review information *Nature Photonics* thanks the anonymous reviewers for their contribution to the peer review of this work.