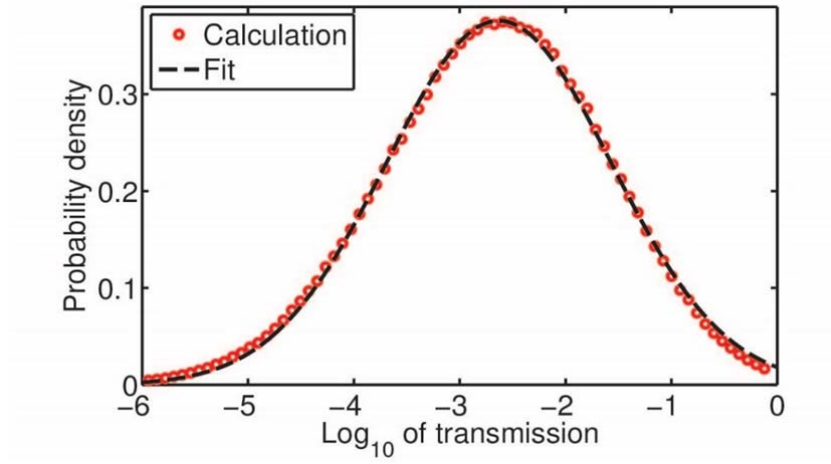




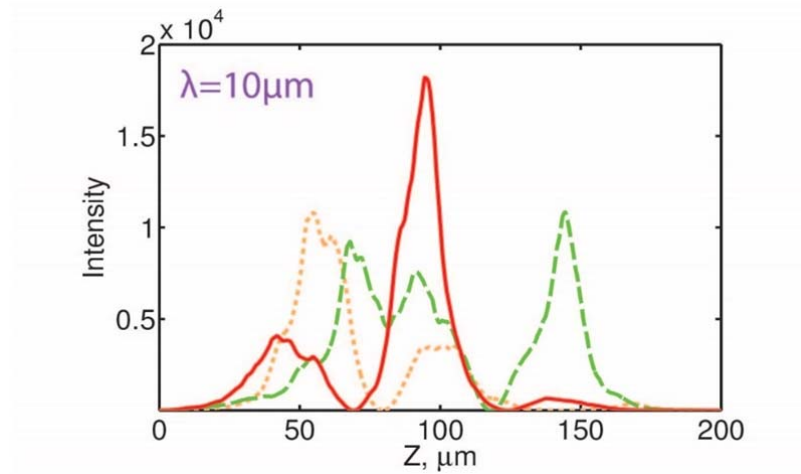
2

Supplementary Figure 1. Probability density of the log of transmission through an $N=300$ structure at $\theta = 60^\circ$ and $\lambda = 1\mu m$.

3

4

5



6

Supplementary Figure 2. Three localized modes for $\theta = 60^\circ$ and $\lambda = 10\mu m$. These modes are found by examining various realizations and finding the three "strongest" ones which have the highest transmission (out of an ensemble of 10,000 realizations). We show the calculated intensity distribution of those realizations, with the intensity normalized relative to incident intensity.

7

8

9

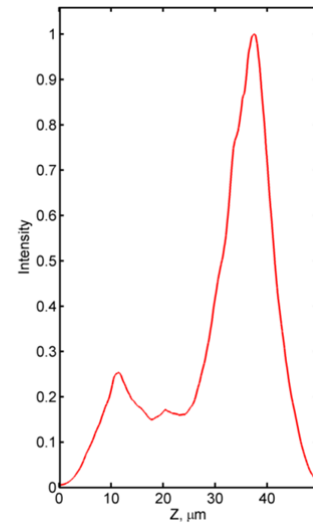
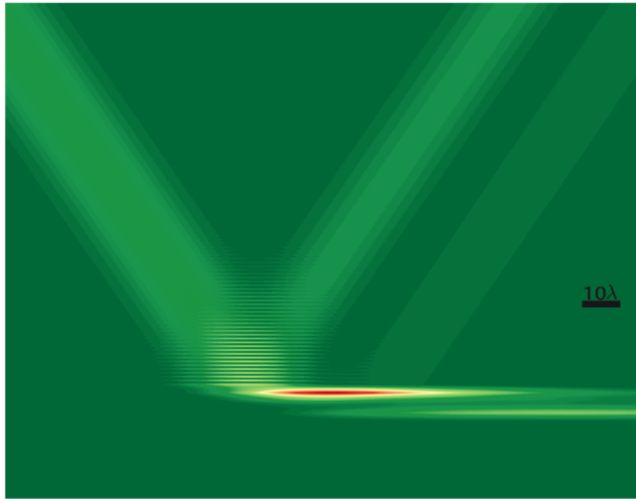
10

11

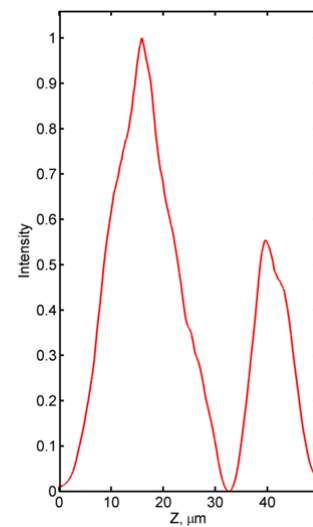
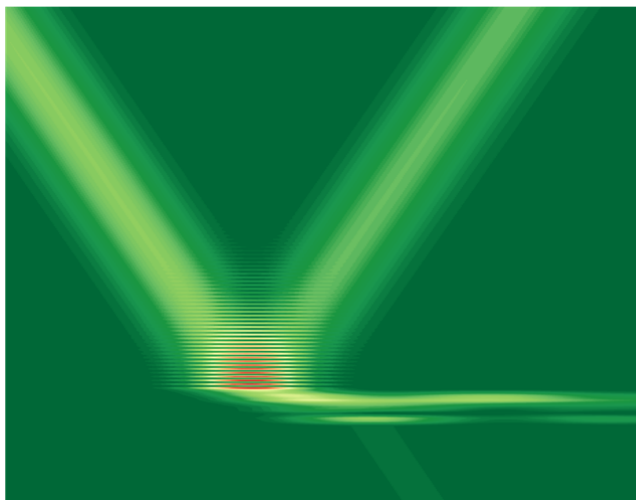
12

Supplementary Figure 3. **(a-j)** In all of the subfigures below, the left hand side shows the intensity profile of the light for an incoming Gaussian beam incident at the critical angle, while the right hand side shows a longitudinal cross-section of the intensity of the trapped component inside the multilayer normalized to the peak intensity. Red represents the maximum intensity and green the minimum. The boundaries of this structure and the scale are the same as in Figs. 1b,c of the main text. The scale-bar in **a** applies to all figures below.

a.



b.

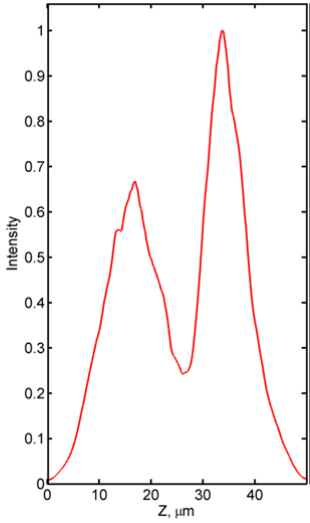
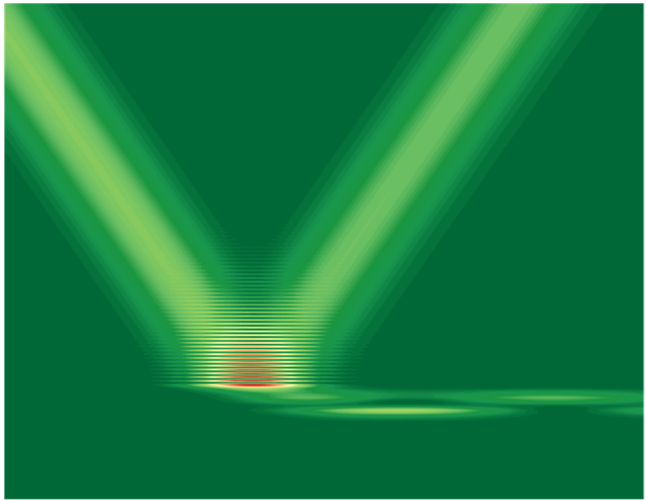


26

27

c.

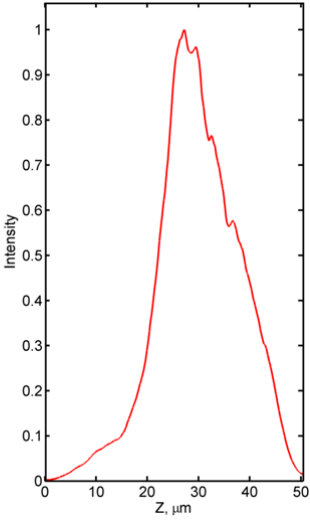
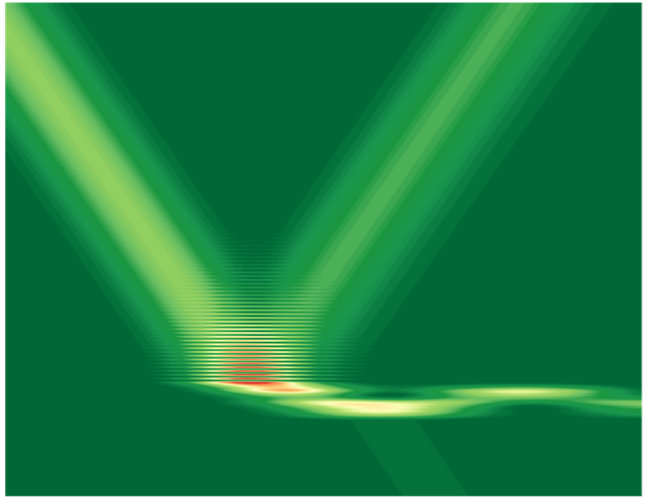
28

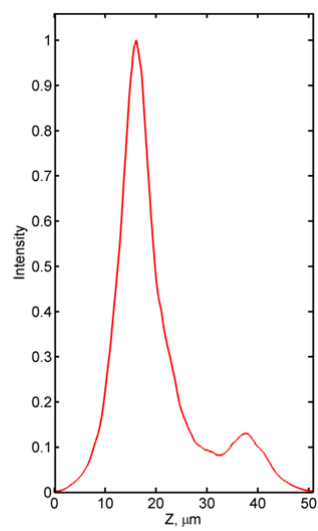
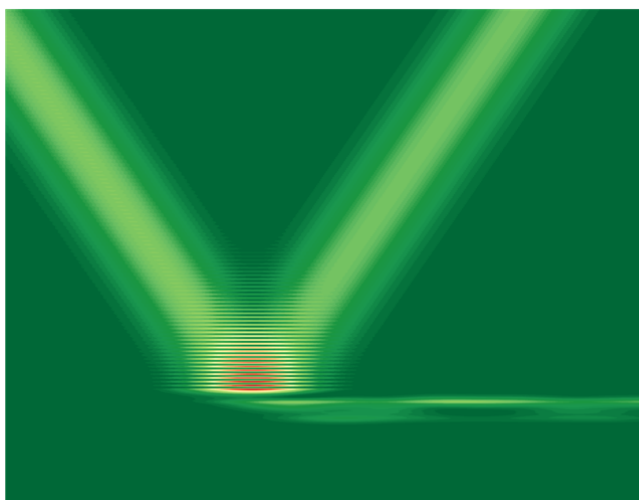


29

d.

30



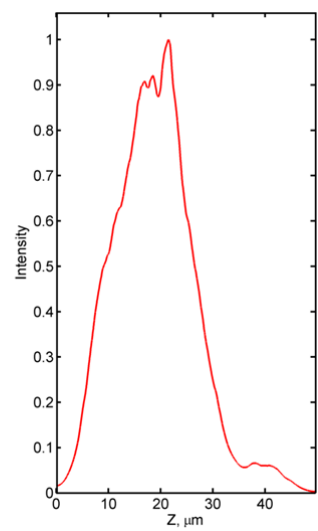
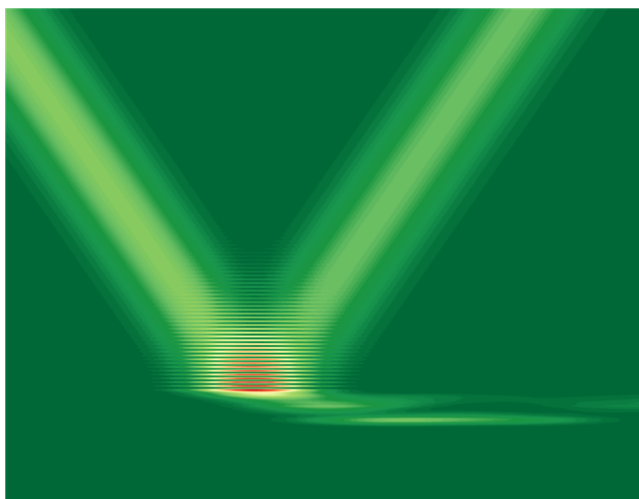


e.

31

f.

32



g.

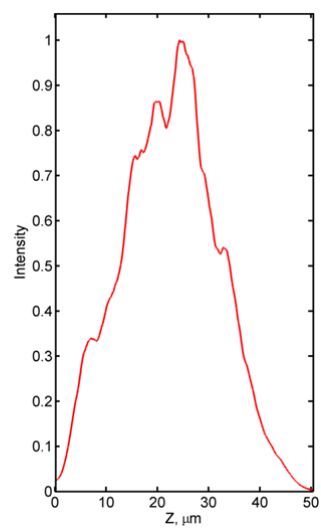
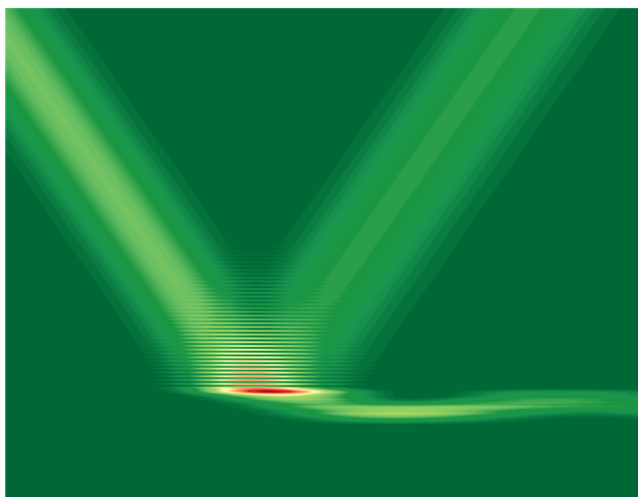
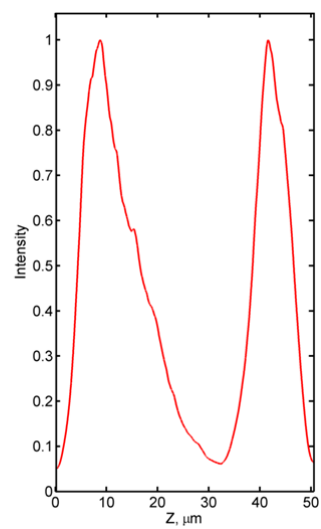
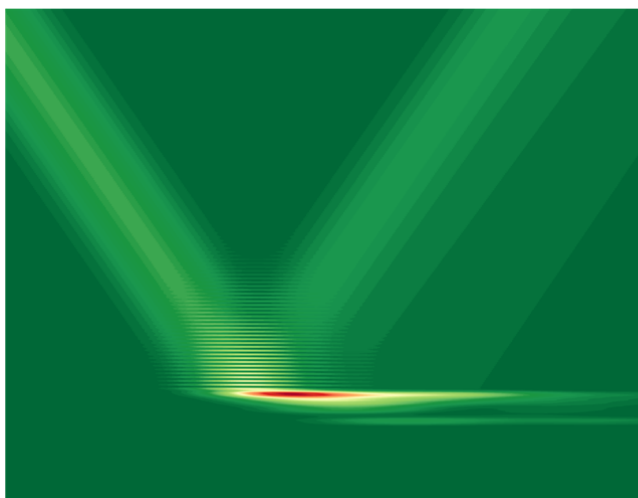
33

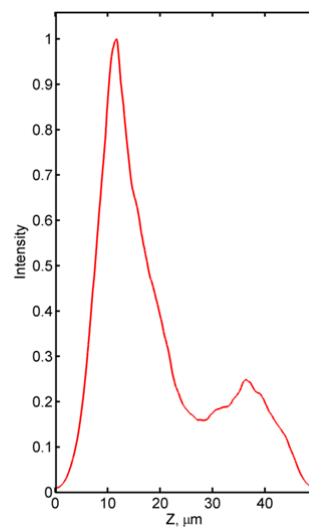
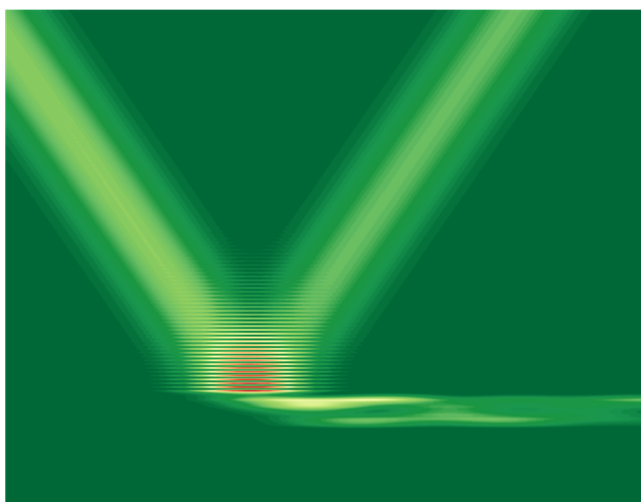
h.

34

i.

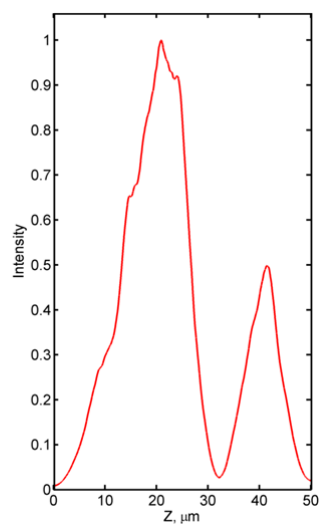
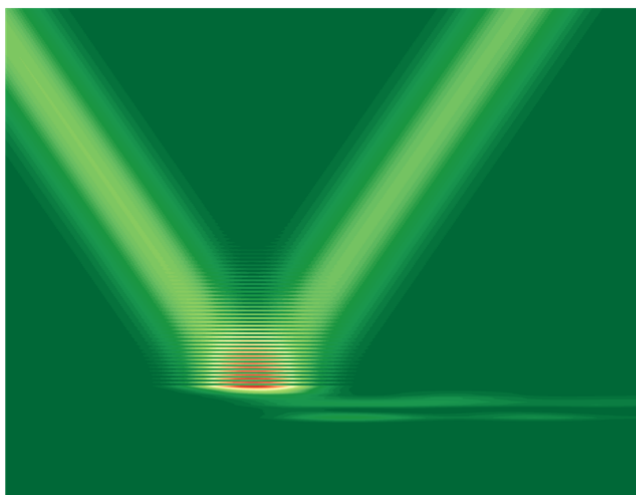
35



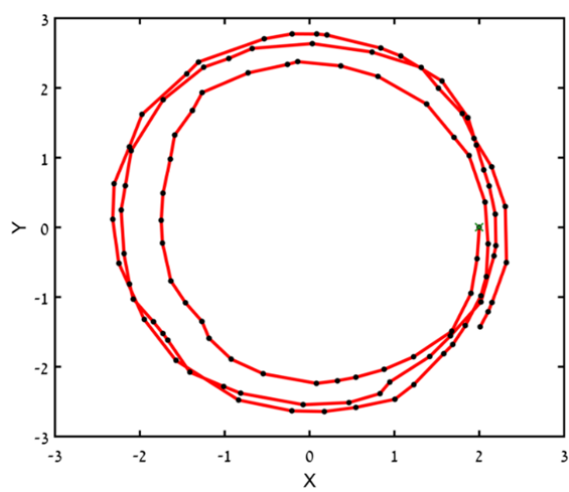


j.

36

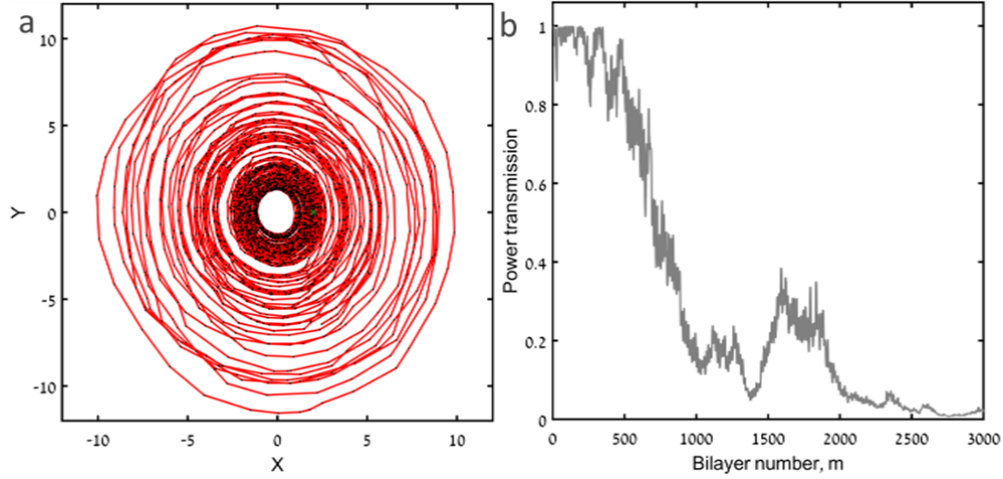


37

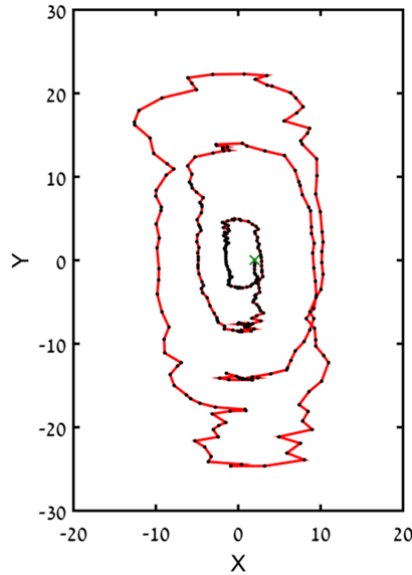


38

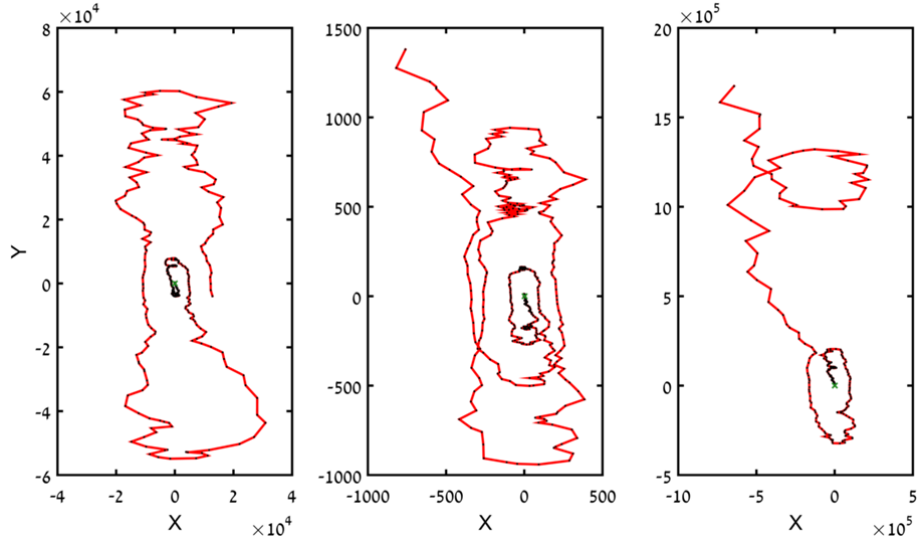
Supplementary Figure 4. Trajectory in the $X - Y$ plane for a disorder sample with $N = 100$ layer pairs, at $\theta = 20^\circ$, $\lambda = 1 \mu m$, $d = 10 nm$ (and all of the other parameters set as in the main paper). The black points are the values of X_m, Y_m , the green marker shows X_1, Y_1 and the red line guides the eye and connects subsequent dots.



Supplementary Figure 5. **(a.)** Trajectory in the $X - Y$ plane for $N = 3000$ layer pairs and $\theta = 20^\circ$. **(b.)** Power transmission as a function of the number of layers in the multilayer for the same disordered realization.

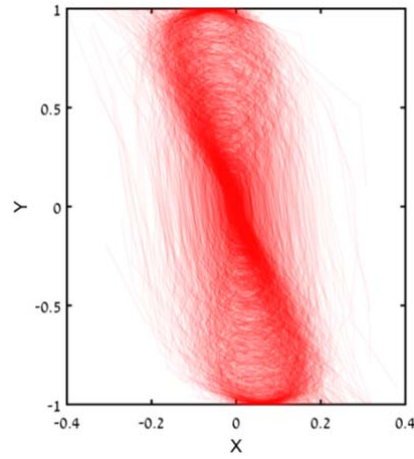


Supplementary Figure 6. A typical trajectory, for $N = 300$ and $\theta = 55^\circ$.



52

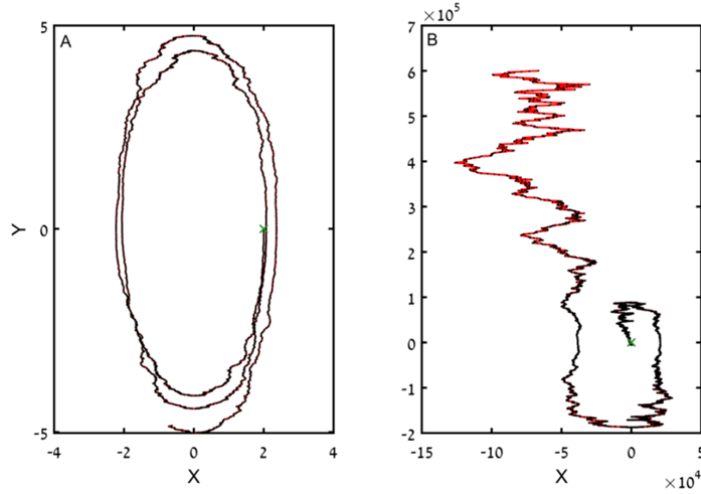
Supplementary Figure 7. Three typical trajectories, for $N = 600$ and $\theta = \theta_c = 60^\circ$. 53
 Note that the degree of localization (value of the maximum X and Y) differs in the 54
 different realizations, but also that the ratio of the X and Y maxima is similar. 55



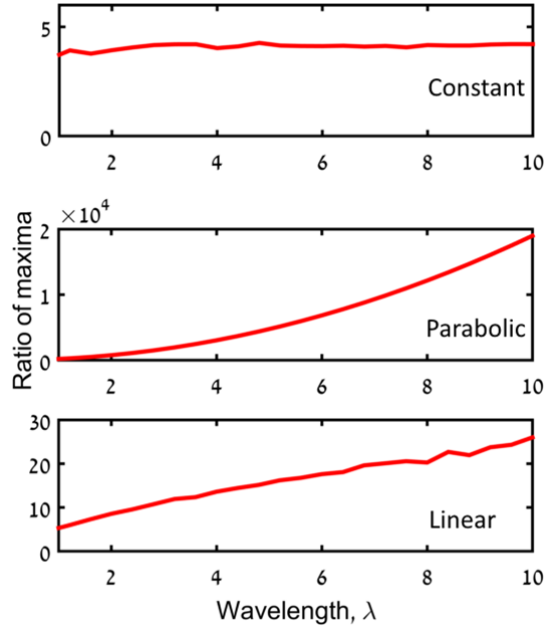
56

Supplementary Figure 8. The trajectories in 1000 realizations of disorder plotted over 57
 one another for $N = 3000$ and $\theta = \theta_c = 60^\circ$. 58

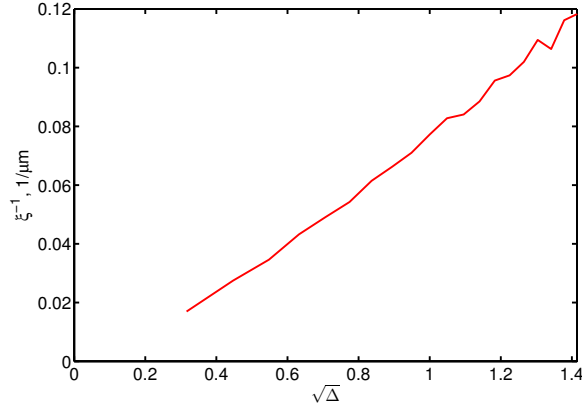
59



Supplementary Figure 9. **(a)** Typical trajectory in the $X-Y$ plane, for $N = 2000$, $\theta = 55^\circ$ and $\lambda = 10\mu m$. **(b)** Trajectory for $N = 10^4$ and $\theta = 60^\circ$.



Supplementary Figure 10. The variation of the average ratio of maxima of X and Y with the wavelength in an ensemble of 1000 realizations. The ratio is calculated after propagation through a large number of layers, large enough that $Nd \gg \xi(\lambda)$. **(a)** for near critical incidence, $\theta = 55^\circ$, **(b)** for critical incidence (at $\theta_c = 60^\circ$) in a periodic disorder-free structure, **(c)** for critical incidence to a disordered structure.



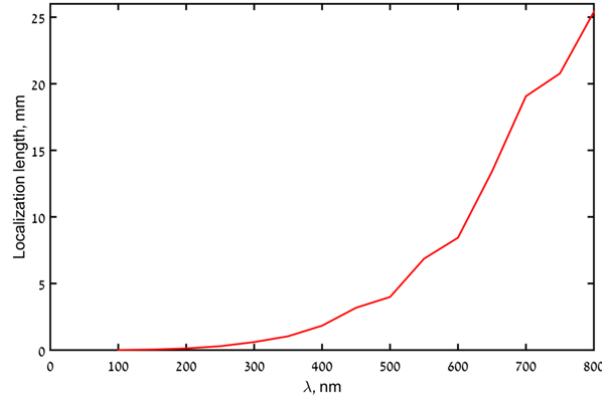
71

Supplementary Figure 11. Inverse localization length ($1/\xi$) as a function of $\sqrt{\Delta}$,
 showing a roughly linear dependence. In this figure we use the same parameters as
 Fig. 2a of the main text and $\lambda = 1\mu m$.

73

74

75



76

Supplementary Figure 12. Localization length as a function of wavelength, for
 correlated disorder with $d_{2n} = d_{2n+1} = random$. Note that the localization length scale
 is in mm and becomes so long that it takes prohibitively long times to continue the
 calculation to longer wavelengths. For the same reason, each point is the average over
 only 30 realizations. These calculations display a good fit to a λ^4 dependence.

77

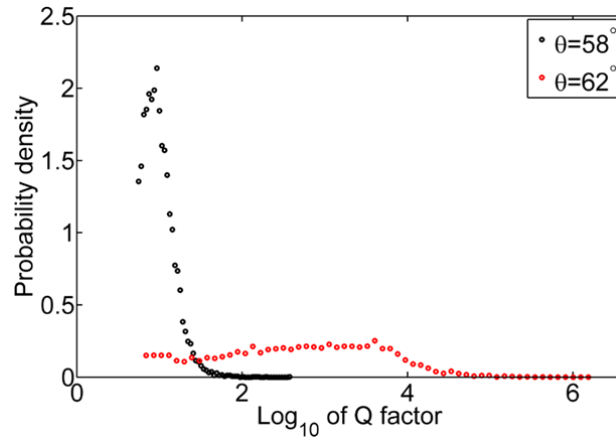
78

79

80

81

82



83

Supplementary Figure 13. Probability density of the 1st mode's Q-factor for two
 angles of incidence $\theta = 58^\circ$ and $\theta = 62^\circ$ below and above critical incidence. Results
 were extracted by taking distribution of the maximum intensity in an ensemble of 10^4
 realizations of disorder and neglecting harder-to-identify diffusive modes with a Q-
 factor below 5.

84

85

86

87

88

89

Supplementary Note 1. The Abbe Transfer matrix formalism 90

The main tool in our calculations is the Abbe¹ transfer matrix formalism. This formalism is described in many places, but for completeness, we describe the Abbe formalism here and in the notations used throughout the article. 91
92
93

We consider a multilayer structure composed of alternating dielectric layers, each layer with uniform permeability and a permittivity. The dielectric constant just beyond the input and output of the structure have arbitrary dielectric constants 94
95
96

$$\begin{aligned}\mathcal{E}(z < 0) &= \mathcal{E}_{in} \\ \mathcal{E}\left(\sum_1^{n-1} d_n < z < \sum_1^n d_n\right) &= \mathcal{E}_n, [1] \\ \mathcal{E}\left(\sum_1^N d_n < z\right) &= \mathcal{E}_{out}\end{aligned} \quad 97$$

for $n=0,1,2,\dots, 2N-1$ where $\mathcal{E}_n = \mathcal{E}_a$ for even values of n and $\mathcal{E}_n = \mathcal{E}_b$ for odd values of n . Consider now the decomposition of the (TE polarized) electromagnetic field into forward and backward propagating waves $E_{tot}(z) = (E_+ + E_-)\hat{y}$ where $E_+(nd < z < (n+1)d) = U_n e^{ik_{z,n}(z-nd)-i\omega t}$, $E_-(nd < z < (n+1)d) = V_n e^{-ik_{z,n}(z-nd)-i\omega t}$ and $k_{z,n} = k_0 \sqrt{\mathcal{E}_n - \mathcal{E}_{in} \sin^2(\theta)}$, where θ is the angle defined in the text (see Fig. 1 of the main text). U_n, V_n are complex constants indicating the amplitude and phase of the wave immediately after the interface. Outside the structure, we assume the wave propagates (is not evanescent), so that the propagation constants $k_{z,0}$ and $k_{z,n}$ are real. 98
99
100
101
102
103
104
105
106

The transfer matrix relating the field in two adjacent layers is^{2,3}: 107

$$\begin{pmatrix} U_n \\ V_n \end{pmatrix} = T_{n,n+1} \begin{pmatrix} U_{n+1} \\ V_{n+1} \end{pmatrix} = \frac{1}{t_{n,n+1}} \begin{pmatrix} \exp(-ik_{z,n}d) & r_{n,n+1} \exp(-ik_{z,n}d) \\ r_{n,n+1} \exp(+ik_{z,n}d) & \exp(+ik_{z,n}d) \end{pmatrix} \begin{pmatrix} U_{n+1} \\ V_{n+1} \end{pmatrix} \quad [2] \quad 108$$

The transfer matrix for a stack of such layers, extending from the n^{th} to the m^{th} surface (with $n < m$), is therefore: 109
110

$$\begin{pmatrix} U_n \\ V_n \end{pmatrix} = T_{n,m} \begin{pmatrix} U_m \\ V_m \end{pmatrix} = \begin{pmatrix} A_{m,n} & B_{m,n} \\ C_{m,n} & D_{m,n} \end{pmatrix} \begin{pmatrix} U_m \\ V_m \end{pmatrix} = \prod_{k=n}^m T_{k,k+1} \begin{pmatrix} U_m \\ V_m \end{pmatrix} \quad [3] . \quad 111$$

After calculating the relevant transfer matrices, the transmission coefficient for a 112

wave incident on the n^{th} layer is $1/A_{m,n}$ and the reflection coefficient is $B_{m,n}/A_{m,n}$. 113

Given the set of transfer matrices and assuming the illumination is incident from the 114

left ($U_0 = 1, V_N = 0$), we find $\begin{pmatrix} U_n \\ V_n \end{pmatrix}$ inside a structure with N layers ($n < N$): 115

$$\begin{pmatrix} U_n \\ V_n \end{pmatrix} = \frac{1}{A_{0,n}A_{n,N} + B_{0,n}C_{n,N}} \begin{pmatrix} A_{n,N} \\ C_{n,N} \end{pmatrix} [4] . \quad 116$$

Repeating this process for various angles of incidence to give the different plane wave 117

components of an incident Gaussian wave, we find the field distribution for an 118

incident beam. 119

This is how all the results presented in this article are calculated, including all the 120

figures. The results of these calculations are also corroborated by scattering matrix 121

calculations. While both scattering and transfer formalisms are analytically identical, 122

the scattering matrix calculations are known to be more stable numerically in systems 123

with evanescent waves. However, for the case at hand, we find practically identical 124

results in both formalisms. 125

126

Supplementary Note 2. Characterization of the GHl regime 127

Both evanescence and localization lead to exponential decay of the fields transmitted 128
through a multilayer stack. How then can localization be definitively demonstrated in 129
a medium in which the wave is evanescent in some parts of the sample? This can be 130
done by showing unique signatures of localized waves are present in an ensemble of 131
disordered samples. Here we consider three such characteristics: (1) the variation of 132
the scaling of transmission upon the strength of disorder, (2) the equality of the 133
variance of the logarithm of transmission and the absolute value of the average of the 134
logarithm of transmission, as required by the single parameter scaling hypothesis, ad 135
(3) the presence of spatially localized modes. These characteristics have been 136
demonstrated in the main text. Here we specifically contrast these properties to those 137
of typical systems supporting evanescent waves. 138

1. Scaling of transmission. In the Goos-Hanchen localization regime (see 139
definition of GHl in the main text), we find that, without disorder, i.e, in a 140
perfectly periodic multilayer, the transmission and intensity inside the structure is 141
a periodic function of the thickness of the overall structure, which resembles the 142
transmission through a Fabry-Perot cavity. Therefore, full transmission can be 143
obtained even through very thick structures⁴. The exponential decay in the 144
(average) transmission of a random ensemble, as seen in Figs. 2a,b of the main 145
text, only appears when disorder is added. Moreover, the exponential decay of 146
transmission depends strongly on the degree of randomness of the layer 147
thicknesses. For example, drawing the layer thicknesses from a rectangular 148
distribution from the 8 to 12 nm range (instead of the 2 to 18 nm range used all 149
throughout the article), leads to a substantial increase in the localization length ξ 150
and weakens all reported localization effects. 151

2. Statistics of transmission. As shown in Supplementary Figure 1a, below, the 152
statistical distribution of the transmission through the multilayer sample is log- 153
normal, which is the typical distribution for localized waves in 1D disordered 154
systems^{5,6}. In addition, the variance of the log of transmission is essentially equal 155
to the magnitude of the average of the log of transmission. This feature is in 156
accord with the single-parameter scaling (SPS) hypothesis, which is a known 157

characteristic of Anderson localization in 1D^{6,7}. Specifically, we find that
 $\text{var}(\ln T) \approx 1.05L / \xi$ or better throughout the GHL regime.

3. Modal makeup of field. The field inside the sample does not decay monotonically, as occurs for evanescent waves. Instead, the intensity profile of the EM field can exhibit one or more peaks inside the sample, with intensities much higher than the input intensity. In particular, for realizations associated with high transmission, such as those shown in Supplementary Figure 1b, relatively high Q-factors are seen (even for $d \sim \lambda/1000$). The intensity profiles in Supplementary Figure 2 are the modes of the system. The frequency dependence of these modes is found to be Lorentzian, and the typical length scale of the modes is comparable to the value of ξ found in the paper (for example, in Fig. 2a). These features are all characteristic of Anderson localized systems (for example, see ^{8,9}). The spacing between the modes is generally greater than the width of these modes (see for example Fig 5b and the top of Fig 5a), which complies with the Thouless criterion for localization^{9,10}.

Note that here and in the main text, we refer to quasi-normal modes as 'modes'. Quasi-normal modes are the resonances of an open system and exhibit a Lorentzian linewidth. When referring in the main text to the “first mode”, we mean the quasi-normal mode with the lowest frequency possible for a specific angle of incidence. For a potential of finite height and support, such a first mode is always expected to exist. It is essential distinguish between quasi-normal transmission modes and the transverse modes of a waveguide, which are a completely different entity, related to the waveguide profile.

Supplementary Note 3. Localization of beams: additional examples

In any physical experiment, the incident beam is a wave packet of plane waves which carries finite power, unlike idealized plane waves. Supplementary Figure 3, shows 10 examples of a 2D Gaussian beam incident at the critical angle for different random realizations of disorder in the structure, all with the same parameters as the structure as in Fig. 1c.

Due to random nature of localization, there is a significant difference between the manifestation of localization in different disordered realizations. But notably, all of these examples have much in common. Mainly, they all show a very strong decay in transmission. The exact amount of transmitted light, however, changes drastically, where some examples (sub figure **i**) show no transmission and some (subfigure **b** and **d**) show a small amount of light can get through. In addition, in most of the figures there is a visible trapping effect, induced by a localized mode (as in subfigure **a** and **f**), but some figures (as in subfigure **i**) show almost no transverse propagation of trapped energy.

Supplementary Note 4. Analytic derivation of the scaling 201

relation of the localization length 202

Introduction and definitions 203

Our goal here is to show that, in the vicinity of the critical angle for TIR, the 204
localization length scales with the wavelength as $\xi \sim \lambda$ for asymptotically long 205
wavelengths and therefore that localization remains wavelength scaled in deep 206
subwavelength structures. The technique relies on the Hamiltonian mapping 207
method¹¹, where we map the problem of transmission through a stack of layers of 208
random thicknesses to a trajectory in the plane of $X = E_y$, the transverse electric field, 209
and $Y = -ik_z^{-1}B_z$, the longitudinal component of the magnetic field. Without disorder, 210
these trajectories are known to be simple ellipses. Under the influence of weak 211
disorder, the elliptical trajectories are perturbed and acquire multiple "kicks" from the 212
disorder. Each kick can change the radius of the trajectory in a random manner, either 213
increasing it or decreasing it, but on average the effect of disorder accumulates, 214
manifesting itself in the exponential decay of the field, indicating that the wave is 215
localized due to the disorder. 216

In the analytic treatment which follows, the matrix formalism we use differs from the 217
Abbe formalism described above. In the formalism used here, we use the X, Y basis 218
instead of the U, V basis (of forward and backward waves). Otherwise, we use the 219
same parameters and notation as elsewhere in the text. Specifically, we assume that 220
 $d = d_a = d_b$, that the range of variations in d is comparable to d , and that $d \ll \lambda$. 221

Since the field in the entire bilayer in deep subwavelength structures is almost 222
constant, it is sufficient to take only the field value at a single point at the beginning 223
of the high-permittivity layer in each bilayer X_m, Y_m . This point is arbitrarily chosen 224
to be the beginning (leftmost point) of the high-permittivity layer. We further assume 225
that the odd layer in the m^{th} bilayer, where $n = 2m - 1$, are high permittivity layers, in 226
which the waves propagate with $k_H \in \mathbb{R}$ and that the even layer in the each bilayer (227
 $n = 2m - 2$) has low permittivity with $k_L \in i\mathbb{R}$. We only consider waves incident 228
exactly at the critical angle, $\theta = \theta_c$ and TE polarization. From the definition of θ_c , we 229

$$\text{have } k_{z,\text{eff}} = \sqrt{\bar{\epsilon}k_0^2 - k_x^2} = 0 \quad \text{and} \quad \text{therefore} \quad k_{z,H} = \sqrt{\epsilon_H k_0^2 - k_x^2} = \sqrt{(\epsilon_H - \bar{\epsilon})k_0^2} \quad 230$$

$$= -i\sqrt{(\epsilon_L - \bar{\epsilon})k_0^2} = -ik_{z,L}. \text{ A discussion of the behavior for other angles of incidence} \quad 231$$

$$\text{appears in the main paper. As a result, the ratio of the layer impedances is } Z_a Z_b^{-1} = i. \quad 232$$

$$\text{We then have, } \varphi = \varphi_a = -i\varphi_b = k_{z,H}d. \text{ We emphasize that the matrix solution is an} \quad 233$$

$$\text{exact solution of Maxwell's equations.} \quad 234$$

$$\text{Following the conventions of the transfer matrix formalism, we do not calculate the} \quad 235$$

$$\text{field transmitted through the structure, but rather the field amplitude required at the} \quad 236$$

$$\text{input of the structure to obtain a field with unit amplitude at the output. Therefore,} \quad 237$$

$$\text{larger radii of a trajectory imply that the field decays significantly before reaching a} \quad 238$$

$$\text{(lower) value at the input. Thus, proving that the wave is localized entails showing} \quad 239$$

$$\text{that the trajectories diverge. Proving that this localization is due to disorder and not} \quad 240$$

$$\text{simple evanescence is addressed above in Supplementary Note 2. This convention has} \quad 241$$

$$\text{a significant added benefit of simplifying the interpretation of visualizations of} \quad 242$$

$$\text{calculated trajectories.} \quad 243$$

$$\text{Using this notation, the field amplitudes in the } n^{\text{th}} \text{ bilayer are related to the amplitude} \quad 244$$

$$\text{of the adjacent cell by} \quad 245$$

$$X_{m+1} = \tilde{A}_m X_m + \tilde{B}_m Y_m \quad Y_{m+1} = -\tilde{C}_m X_m + \tilde{D}_m Y_m \quad [5]. \quad 246$$

$$\text{The factors } \tilde{A}_m, \tilde{B}_m, \tilde{C}_m, \tilde{D}_m \text{ are the transfer matrix coefficients for the bilayer} \quad 247$$

$$\text{structure}^{38}. \text{ By taking the lowest orders in } \frac{d}{\lambda} \text{ of these coefficients, we obtain} \quad 248$$

$$\begin{aligned} \tilde{A}_m &= 1 + \varphi^2 - \frac{1}{2} \eta_H^2(m) - \frac{1}{2} \eta_L^2(m) + i \eta_H(m) \eta_L(m) \\ \tilde{B}_m &= -\frac{1}{3} \varphi^2 + \eta_H(m) - i \eta_L(m) \\ \tilde{C}_m &= 2\varphi + \eta_H(m) + i \eta_L(m) \\ \tilde{D}_m &= 1 - \varphi^2 - \frac{1}{2} \eta_H^2(m) - \frac{1}{2} \eta_L^2(m) - i \eta_H(m) \eta_L(m), \end{aligned} \quad [6] \quad 249$$

$$\text{where } \eta_j(m) = k_j(d_{jm} - d), \text{ with } d_{Hm}, d_{Lm} \text{ are widths of the high and low permittivity} \quad 250$$

$$\text{layers at period of index } m. \quad 251$$

252

Unperturbed periodic system

253

As a constructive example, we begin with the disorder-free case, where we perform the following canonical transformation:

254

255

$$\begin{aligned} X_m &= v^{-1} Q_m \cos \tau - v P_m \sin \tau & Q_m &= v X_m \cos \tau + v Y_m \sin \tau \\ Y_m &= v^{-1} Q_m \sin \tau + v P_m \cos \tau & P_m &= -v^{-1} X_m \cos \tau + v^{-1} Y_m \sin \tau. \end{aligned} \quad [7] \quad 256$$

The new coordinates, Q_m, P_m , occupy a circular trajectory, instead of the elliptical trajectory of X_m, Y_m . We require that Q, P rotate at a fixed angular frequency

257

258

$\gamma = \frac{1}{2} \varphi^2$. This value can be found by substituting in the expression for $\tilde{A}_m$,

259

$\tilde{B}_m, \tilde{C}_m, \tilde{D}_m$ and following the resulting algebra, as in³⁸. Note that for calculating γ , higher order terms in φ must be considered, but not elsewhere in the derivation. More simply, we can use the following dispersion relation, whose validity for describing the bulk behavior in our case is shown in⁴

260

261

262

263

$$\cos \gamma = \cos \varphi_a \cos \varphi_b - \frac{1}{2} (Z_a Z_b^{-1} + Z_b Z_a^{-1}) \sin \varphi_a \sin \varphi_b \quad [8]. \quad 264$$

With Z_a (Z_b) the wave impedance in the a (b) layer. Here, γ , the phase accumulation across a single pair of layers, is related to the effective medium wavenumber via $k_{z,\text{eff}} = \gamma / d$. Usually, we expect to that $\gamma \sim \varphi$, but at critical incidence, the linear term

265

266

267

exactly cancels since $Z_a Z_b^{-1} = -Z_b Z_a^{-1}$ and we find $\gamma = \frac{1}{2} \varphi^2$ for $\lambda \ll d$.

268

To solve for Q_m, P_m and find v, τ , we note that the factors A_m, B_m, C_m, D_m are greatly simplified for the disorder-free case, with $\eta_L(m) = \eta_H(m) = 0$:

269

270

$$\begin{aligned} A_m &= 1 + \varphi^2 & B_m &= -\frac{1}{3} \varphi^2 \\ C_m &= 2\varphi & D_m &= 1 - \varphi^2. \end{aligned} \quad [9] \quad 271$$

272

To find the values of v, τ , we calculate

$$Q_{m+1} = (A_m \cos^2 \tau + B_m \cos \tau \sin \tau - C_m \cos \tau \sin \tau + D_m \sin^2 \tau) Q_m + v^2 (-A_m \cos \tau \sin \tau + B_m \cos^2 \tau + C_m \sin^2 \tau + D_m \cos \tau \sin \tau) P_m. \quad [10]$$

In order to obtain rotation with an angular frequency γ , we require

$$\begin{aligned} A_m \cos^2 \tau + B_m \cos \tau \sin \tau - C_m \cos \tau \sin \tau + D_m \sin^2 \tau &= \cos \gamma \\ -A_m \cos \tau \sin \tau + B_m \cos^2 \tau + C_m \sin^2 \tau + D_m \cos \tau \sin \tau &= v^{-2} \sin \gamma, \end{aligned} \quad [11]$$

giving

$$2\phi \sin^2 \tau = v^2 \sin \gamma \quad [12].$$

Similarly,

$$\begin{aligned} P_{m+1} &= (-A_m \cos \tau \sin \tau + B_m \sin^2 \tau - C_m \cos^2 \tau + D_m \cos \tau \sin \tau) Q_m \\ &+ v^{-2} (A_m \sin^2 \tau - B_m \cos \tau \sin \tau - C_m \cos \tau \sin \tau - D_m \cos^2 \tau) P_m, \end{aligned} \quad [13]$$

giving

$$2\phi \cos^2 \tau = v^2 \sin \gamma \quad [14].$$

Thus,

$$\tau = \frac{\phi^2}{4}, \quad v^2 = \frac{2}{\phi} \quad [15].$$

Having found the parameters of the unperturbed trajectory, we would proceed to calculate the effect of disorder, in a manner similar to ¹¹. However, this calculation yields the unphysical result of a localization length which is independent of the wavelength for asymptotically large wavelengths, because the disorder-coupled magnetic field is much larger than the magnetic field existing in the absence of disorder. The calculation above shows that in the absence of disorder B_z is of order ϕ^2 , which is a manifestation of the wavelength dependence of the effective medium

permittivity, which converges as $\Delta(\varphi) = \varepsilon_{\text{eff}} - \bar{\varepsilon} \sim \varphi^2$ (see discussion in ⁴). On the 295
 other hand, the disorder provides coupling of the electric field to the magnetic field 296
 which is of order φ (through the $v^2 \tilde{B}_m \cos^2 \tau$ term). Therefore, to properly treat the 297
 magnetic field in the presence of disorder we suggest to change the unperturbed state 298
 and take as an ansatz a different basic state for which B_z is of order φ . In what 299
 follows we justify this ansatz further, by numerical evidence. 300

301

Numerical investigation of the trajectories

303

The trajectories in the $X - Y$ plane can be calculated numerically, providing a vivid image of the system's behavior. We start with the case of near-normal incidence, where we expect a perturbed-elliptical trajectory which spirals away due to disorder. Note that in all the figures below, we draw only the real part of the fields. For the parameters here, each circuit is nearly circular. The same simulation implemented on a larger multilayer shows the trajectory spiraling out considerably. This corresponds to the decay of transmission and to localization.

310

For $\theta = 55^\circ$, closer to critical incidence, the trajectory becomes less regular and more stretched. The impact of disorder in this case is seen to be stronger, inducing larger random jumps in the trajectory.

313

At the critical angle, the effect of disorder becomes dominant and the trajectories are more complex and random. The jumps in the trajectory are larger and the trajectories include loops or wild zig zags or may rapidly fly away. Nevertheless, for sufficiently large N , the trajectory always revolves around the center of the pattern and eventually forms a highly distorted and elongated spiral pattern. The ratio of the maxima of X and Y in all trajectories of sufficient length is, on average, roughly the same.

320

Unlike the cases of lower angles of incidence, the effect of disorder here is dramatic and different realizations of disorder produce different and highly distorted trajectories. The resemblance to an elliptical trajectory is also much weaker. However, normalizing all all of trajectories in an ensemble to their maximum radius and drawing them one on top of the other produces a picture that is easier to interpret visually.

326

When the wavelength is increased asymptotically, one might expect the trajectory to become closer and closer to the unperturbed elliptical trajectory. Indeed, for most angles of incidence, this is precisely what happens. However at the critical angle, the effect of disorder is not diminished as the wavelength increases. For $\lambda = 10^4 d$, the trajectories remain as distorted in exceedingly long wavelengths. The number of

331

bilayers, N , required for the trajectory to evolve does increase, but it does not
converge to the unperturbed trajectory.

For example, Supplementary Figure 9b shows a trajectory that is qualitatively similar
to the ones in Supplementary Figure 7, except for the larger *ratio* of the maxima of X
and Y . To determine the rate of elongation in the disordered system, we can calculate
the ensemble average of this ratio as a function of the wavelength.

Away from the critical angle, the ratio of the maxima of X and Y is almost
independent of the wavelength (Supplementary Figure 10a). At critical incidence, this
ratio is wavelength-dependent, but as our earlier analysis suggested - this dependence
is not parabolic as in periodic structures (Supplementary Figure 10b). In
Supplementary Figure 10c, this ratio still increases in a disordered system at critical
incidence, but at a much slower, roughly linear, rate. Such a linear dependence in the
wavelength complies with our intuition that the size of B_z should be dominated by
the disorder-induced coupling coefficient, which was shown earlier to be linearly
related to the wavelength.

347

Scaling of the localization length

348

In light of the numerical analysis above, we suggest the following ansatz – the basic trajectory appears to be an ellipse with $v^4 \sim \varphi^{-1}$ (instead of $v^4 \sim \varphi^{-2}$ found for the unperturbed trajectory). There is a strong intuitive argument for this assumption: the strong coupling of Y to X inflates the trajectory and effectively prevents the magnitude of B_z from being as low as predicted by effective medium theory. At the same time, it should also be stressed that this assumption is only expected to produce the dependence in φ and therefore the scaling of the localization length, but not the exact dependence or other features of the system.

349

350

351

352

353

354

355

356

With this assumption, we can now demonstrate that the localization length scales linearly with the wavelength, as we have found numerically in the main text. To do so, we find recurrence relations for the radius, defined as, $R_m^2 = Q_m^2 + P_m^2$. Keeping only the lowest order terms in φ , and using $v^2 \sim \sqrt{\varphi}$, as per our assumption, we obtain

357

358

359

360

361

$$Q_{m+1} = Q_m + 2\varphi^{-\frac{1}{2}}(\eta_H(m) + i\eta_L(m))P_m [16],$$

362

$$P_{m+1} = P_m - 2\varphi^{\frac{1}{2}}(2\varphi + \eta_H(m) - i\eta_L(m)) [17].$$

363

Thus,

364

$$R_{m+1}^2 = R_m^2 + 4\varphi^{-\frac{1}{2}}(\eta_H(m) + i\eta_L(m))Q_mP_m + 4\varphi^{-1}(\eta_H(m) + i\eta_L(m))^2P_m^2 [18].$$

365

We denote $\sigma_H^2 = \langle \eta_H^2(m)_m \rangle$, the variance of d_n in the high permittivity layers and

366

$\sigma_L^2 = \langle \eta_L^2(m)_m \rangle$ for the low permittivity layers. We assume that no correlation exist,

367

so that $\langle \eta_H(m)\eta_L(m) \rangle = 0$ as stated earlier. We then average over n , using

368

$\langle \eta_H(m) \rangle = \langle \eta_L(m) \rangle = 0$ to get

369

$$\frac{R_{m+1}^2}{R_m^2} = \varphi 4\sigma^2 \cos^2 \theta_m [19],$$

370

where $\tan \theta_m = Q_m P_m^{-1}$ and $\sigma^2 = \sigma_H^2 + \sigma_L^2$. Here, we neglect correlation terms such as $\eta_H(n) \sin \theta_m$, since, for long wavelengths, θ_m changes slowly (i.e., over a large number of layers), whereas η_H changes rapidly and averages to zero in the absence of long-range correlation in d_n . Since $\cos^2 \theta_n$ is generally a factor of order unity, we neglect it for the purposes of scaling and are left with the relation

$$\xi = \frac{2d}{\ln \frac{R_{m+1}^2}{R_m^2}} \sim \frac{1}{k_H \sigma^2} \sim \frac{\lambda}{\sqrt{\Delta}} [20],$$

where $\Delta = \epsilon_H - \epsilon_{\text{eff}} = \frac{1}{2}(\epsilon_H - \epsilon_L)$.

The linear scaling of ξ with the wavelength, which is what we find here, is one of the main findings of this work. It corresponds to the localization length remaining short, typically a few wavelengths, even for exceedingly long wavelengths, as demonstrated numerically in Fig. 2a in of the main text. In a similar vein, the scaling with Δ can also be verified numerically and goes as indicated in Supplementary Equation 11. It is interesting to note that it differs from the $\Delta^{-\frac{3}{2}}$ dependence predicted theoretically in²¹ for a related scenario.

Supplementary Note 5. Effect of interbilayer correlations on 386
wavelength scaling 387

In the analysis above, we assumed that η_H, η_L are uncorrelated random variables. 388
However, considering the derivation above, we see that if $|\eta_H| = |\eta_L|$ at every bilayer 389
then the v^1 term of $\tilde{B}_n$ cancels. Accordingly, in such a case, the scaling we find no 390
longer holds and we find instead a $\xi \sim (\lambda/d)^4$ dependence, as demonstrated the 391
figure below. We do not derive this alternative scaling relation explicitly here – the 392
derivation and its result for this case are very similar to those in ³⁸. However, the 393
contrast between the correlated case and the uncorrelated cases is exceedingly large. 394
For example, two structures which differ only on a deep subwavelength (say, 395
nanometric) level correlations can have an entirely different optical response, either 396
reflecting or transmitting light completely, with localization lengths that differ by 397
many orders of magnitude This is another example to the sensitivity of this regime to 398
subwavelength details, where certain correlated random stacks exhibit complete 399
delocalization and other correlations exhibit complete localization. 400
401

Supplementary Note 6. Q-factor distribution for the first mode 402

In Fig. 6c of the main text, we show the wavelength at which the first mode of a 403
structure is obtained. We can also find the Q-factor distribution of these modes for 404
two angles of incidence, one below and one above the critical angle. These Q-factors 405
are important for potential applications, and can also help us to better understand the 406
transverse dynamics in Figs. 1b,c, since they indicate the transverse localization 407
length. 408

These Q-factors are important for the analysis in the main text for several reasons. 409
First, because they demonstrate the strength of the localization relative to normal 410
incidence localization. Q-factors reach 10^6 which is remarkably high for Anderson 411
localization, especially for thin samples. Moreover, the presence of high-Q modes, 412
facilitates the enhanced transmission effects we describe in the main text, as well as 413
the wide spread in the transverse direction of a finite beam. A better understanding the 414
Q-factor distribution is required to understand these effects. 415

416

References	417
	418
[1] F. Abeles, La théorie générale des couches minces, <i>J. Phys. Radium</i> , 11 , 307–310 (1950).	419 420
[2] Katsidis, C. C. & Siapkas, D. I. General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference. <i>Appl. Opt.</i> 41 , 3978-3987 (2002).	421 422 423
[3] Heavens, O.S. <i>Optical Properties of Thin Solid Films</i> . (Dover Publications, New York, USA, 1991).	424 425
[4] Herzig Sheinfux, H., Kaminer, I, Plotnik, Y., Bartal, G. and Segev, M. Subwavelength multilayer dielectrics: ultrasensitive transmission and breakdown of effective-medium theory. <i>Phys. Rev. Lett.</i> 113 , 243901 (2014).	426 427 428
[5] Lagendijk, A., van Tiggelen, B. and Wiersma, D. S. Fifty years of Anderson localization. <i>Phys. Today</i> 62 , 24-29 (2009).	429 430
[6] Anderson, P. W., Thouless, D., Abrahams, E. and Fisher, D. New method for a scaling theory of localization. <i>Phys. Rev. B</i> 22 , 3519-3526 (1980).	431 432
[7] Deych, L., Lisyansky, A., and Altshuler, B. L. Single-parameter scaling in one-dimensional Anderson localization: exact analytical solution. <i>Phys. Rev. B</i> 64 , 224202 (2001).	433 434 435
[8] Bertolotti, J., Gottardo, S., Wiersma, D., Ghulinyan, M. and Pavesi, L. Optical necklace states in Anderson localized 1D systems. <i>Phys. Rev. Lett.</i> 94 , 113903 (2005).	436 437
[9] J. Wang, A. Z. Genack, Transport through modes in random media. <i>Nature</i> . 471 , 345–8 (2011).	438 439
[10] Sebbah, P., Hu, B., Klosner, J., and Genack, A. Z. Extended quasimodes within nominally localized random waveguides. <i>Phys. Rev. Lett.</i> 96 , 183902 (2006).	440 441
[11] Izrailev, F. & Makarov, N. Localization in correlated bilayer structures: from photonic crystals to metamaterials and semiconductor superlattices. <i>Phys. Rev. Lett.</i> 102 , 203901–1 (2009).	442 443 444
	445
	446
	447
	448