Mutual Information between Reflected and Transmitted Speckle Images

N. Fayard, A. Goetschy, R. Pierrat, and R. Carminati[*]

ESPCI Paris, PSL Research University, CNRS, Institut Langevin, 1 rue Jussieu, F-75005 Paris, France

We study theoretically the mutual information between reflected and transmitted speckle patterns produced by wave scattering from disordered media. The mutual information between the two speckle images recorded on an array of N detection points (pixels) takes the form of long-range intensity correlation loops that we evaluate explicitly as a function of the disorder strength and the Thouless number g. Our analysis, supported by extensive numerical simulations, reveals a competing effect of cross-sample and surface spatial correlations. An optimal distance between pixels is proven to exist that enhances the mutual information by a factor Ng compared to the single-pixel scenario.

When waves propagate in complex environments, their information content is spread out in space and encoded into complicated speckle patterns, eventually recorded as two-dimensional images at the output of the medium. A central issue is the quantification of the information content in speckle patterns, and its use for imaging, power deposit, or information transfer [1–3]. Much effort has been made to take advantage of the existence of spatial correlations in speckles measured in transmission. Various schemes based on the memory effect of short-range correlations (termed C_1^{TT} hereafter) have been developed to image an object through an opaque screen [4,5], while long-range correlations (C_2^{TT}), which capture nonlocal information, have been demonstrated to be useful for increasing energy delivery through turbid media [6,7].

Very recently, the existence of cross-correlations between speckle patterns measured in reflection and transmission has been demonstrated, and the shape of the intensity correlation function has been characterized in regimes ranging from quasiballistic to diffusive transport [8,9]. These correlations suggest the possibility to acquire information about a transmitted speckle from a measurement restricted to the reflection half-space. This is of crucial importance for sensing, imaging, and communicating through turbid media, and for the control of wave transmission through disordered scattering environments by wave front shaping techniques [1,2,10]. In this Letter, we quantify the amount of mutual information (MI) between transmitted and reflected speckles, and analyze the dependance of the MI on the disorder strength and the geometrical parameters characterizing the detection process (number of detectors and their interdistance). Our theory is formulated in terms of the transport mean free path (independently of the details of the microstructure), and applies for an arbitrary space dimension as long as the wave propagation remains diffusive.

The scheme of the gedanken experiment is represented in Fig. 1(a). A slab of a disordered medium is illuminated by a plane wave, and the speckle intensity profile is recorded with a CCD camera placed at the input side. The transmitted speckle, potentially recorded with another camera, is assumed to be unknown. Let $I_i^R = |E_i|^2$ be the reflected intensity measured on pixel i (or detector i) and $x_i = I_i^R/\langle I_i^R \rangle$ be the normalized intensity, the brackets $\langle \cdots \rangle$ denoting an ensemble average over statistical realizations of the disordered medium. The reflected speckle image is represented by the vector $\mathbf{x}$ of size N equal to the number of pixels of the camera. Similarly, the transmitted unknown image is wrapped up into a vector $\mathbf{y}$. In a statistical description of the disordered medium, configurations of disorder are generated by a stochastic process, and $\mathbf{x}$ and $\mathbf{y}$ are random variables. A quantitative estimate of the statistical dependence between $\mathbf{x}$ and $\mathbf{y}$, or equivalently between the two speckle images, is given by their MI, defined as the difference between the entropy of $\mathbf{x}$ and $\mathbf{y}$ considered separately and the entropy of the pair $\{\mathbf{x}, \mathbf{y}\}$ [11]:

$$\mathcal{I} = \iint d\mathbf{x}d\mathbf{y}\, p(\mathbf{x}, \mathbf{y})\log_2\left(\frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})}\right). \qquad (1)$$

Here, $p(\mathbf{x})$, $p(\mathbf{y})$, and $p(\mathbf{x}, \mathbf{y})$ are joint probability density functions (PDFs). The MI is sensitive to all types of statistical dependence, beyond that captured by correlations. In particular, a vanishing MI is strictly equivalent to statistical independence. Moreover, it provides a direct connection with Shannon entropy and information theory concepts. Note the important difference with the study of channel capacity for multiple-input multiple-output protocols, in which the MI between input and output signals is evaluated [3,12–14]. Contrary to these protocols, here the input signal is not random, there is no external noise, and $\mathbf{x}$ is not the injected signal but the output signal in reflection.
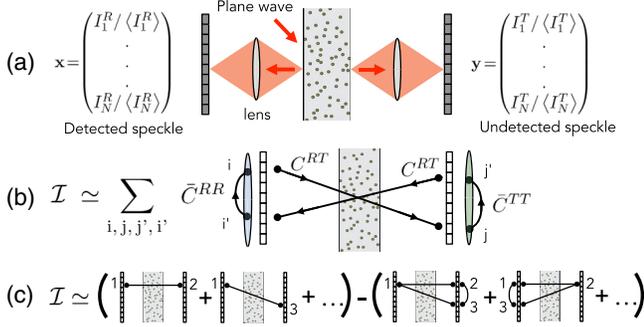
FIG. 1. (a) Schematic view of a disordered slab illuminated by a plane wave. The reflected speckle **x** produced at the sample surface is registered on a CCD camera with N pixels, and its mutual information $\mathcal{I}$ with the transmitted speckle **y** is evaluated. (b) Diagrammatic representation of the trace formula (3) as a sum of correlation loops. (c) Diagrammatic representation of the approximation (6).

One difficulty in evaluating the MI lies in the fact that the PDFs $p(\mathbf{x})$, $p(\mathbf{y})$, and $p(\mathbf{x},\mathbf{y})$ are theoretically unknown. Only marginal distributions, such as $p(x_1)$ and $p(y_1)$, as well as two-point correlations functions ($C_{ii'}^{RR} = \langle \delta x_i \delta x_{i'} \rangle$, $C_{jj'}^{TT} = \langle \delta y_j \delta y_{j'} \rangle$, and $C_{ij}^{RT} = \langle \delta x_i \delta y_j \rangle$ with $\delta x = x - 1$ and $\delta y = y - 1$) have been calculated for disordered media [9,15,16]. In the limit of small pairwise correlations, however, we will show that $\mathcal{I}$ can be expressed as a combination of the previous correlators only, even if the field amplitudes E_i cannot be modeled as complex Gaussian random variables.

First, we express $p(\mathbf{x},\mathbf{y})$ in terms of $p(\mathbf{x})$ and $p(\mathbf{y})$. This joint PDF is entirely characterized by the set of correlators $\langle x^{\{n\}} y^{\{m\}} \rangle = \langle x_1^{n_1} \cdots x_N^{n_N} y_1^{m_1} \cdots y_N^{m_N} \rangle$. Since **x** and **y** are weakly correlated in the multiple scattering regime [8,9], we search for leading corrections to the independent variable result $\langle x^{\{n\}} y^{\{m\}} \rangle = \langle x^{\{n\}} \rangle \langle y^{\{m\}} \rangle$. To proceed, we represent the field E_i as a sum of propagators along all possible scattering trajectories $\mathcal{S}$ inside the medium, $E_i = \sum_{\mathcal{S}} E_i^{\mathcal{S}}$. Hence, each term $x_i^{n_i} \propto |E_i|^{2n_i}$ contains n_i replica of complex propagators $E_i^{\mathcal{S}}$ and $E_i^{\mathcal{S}*}$. The leading correction to the independent result involves all combinations of correlations between two reflection propagators and two transmission propagators. Counting these combinations yields $\langle x^{\{n\}} y^{\{m\}} \rangle = \langle x^{\{n\}} \rangle \langle y^{\{m\}} \rangle + \sum_{i,j} n_i^2 n_j^2 \langle \delta x_i \delta y_j \rangle \langle x^{\{n\}-1_i} \rangle \langle y^{\{m\}-1_j} \rangle$, where the notation $x^{\{n\}-1_i} = x_1^{n_1} \cdots x_i^{n_i-1} \cdots x_N^{n_N}$ is used. This expression gives the moments of $p(\mathbf{x},\mathbf{y})$ in terms of the moments of $p(\mathbf{x})$ and $p(\mathbf{y})$. Then, standard algebra, detailed in the Supplemental Material [17], allows us to cast the joint PDF in the form $p(\mathbf{x},\mathbf{y}) = p(\mathbf{x}) p(\mathbf{y}) [1 + \sum_{i,j} u_{ij}(\mathbf{x},\mathbf{y})]$ with

$$u_{ij}(\mathbf{x},\mathbf{y}) = \langle \delta x_i \delta y_j \rangle \frac{\partial_{x_i}[x_i \partial_{x_i} p(\mathbf{x})]}{p(\mathbf{x})} \frac{\partial_{y_j}[y_j \partial_{y_j} p(\mathbf{y})]}{p(\mathbf{y})}. \quad (2)$$

Second, we insert the previous decomposition into Eq. (1), and express the logarithm as a power series of the correlation function $\sum_{ij} u_{ij}(\mathbf{x},\mathbf{y})$. By keeping the first nonzero term in the power expansion, we obtain the following trace formula [17]:

$$\mathcal{I} \simeq \frac{1}{2 \ln 2} \text{Tr}[C^{RT} \bar{C}^{TT} C^{RT} \bar{C}^{RR}]. \quad (3)$$

In this expression we have introduced three $N \times N$ matrices, with elements defined as $C_{ij}^{RT} = \langle \delta x_i \delta y_j \rangle$,

$$\bar{C}_{jj'}^{TT} = \int d\mathbf{y} \frac{\partial_{y_j}[y_j \partial_{y_j} p(\mathbf{y})] \partial_{y_{j'}}[y_{j'} \partial_{y_{j'}} p(\mathbf{y})]}{p(\mathbf{y})}, \quad (4)$$

and $\bar{C}_{ii'}^{RR}$ in which $p(\mathbf{y})$ is replaced by $p(\mathbf{x})$. Equation (3) has a clear interpretation: the MI between the reflected and transmitted speckle patterns is the sum of all correlation loops ($i \to j \to j' \to i' \to i$), as illustrated in Fig. 1(b). In each loop, the correlation between pixels in different images is carried by pairwise cross-sample long-range coupling (C_{ij}^{RT} and $C_{j'i'}^{RT}$), whereas the correlations within each image ($\bar{C}_{i'i}^{RR}$ and $\bar{C}_{jj'}^{TT}$) are more complicated since they are nonlocal, involving the full distributions $p(\mathbf{x})$ and $p(\mathbf{y})$.

To make the interpretation of the trace formula (3) even more transparent, we further assume that the distance between pixels in each image is larger than the free-space wavelength λ, so that correlations within each image remain small. As detailed in the Supplemental Material [17], this allows us to approximate the transmission PDF as $p(\mathbf{y}) = \prod_k p(y_k)[1 + \sum_{j<j'} u_{jj'}(y_j, y_{j'})]$, where $p(y_k) = e^{-y_k}[1 + C_2^{TT}(y^2/4 - y + 1/2)]$. Here, $C_2^{TT} \simeq \langle (\delta y)^2 \rangle - 1$ is the leading non-Gaussian local correction to Rayleigh statistics [25,26]. The reflection side PDF $p(\mathbf{x})$ takes the same functional form, with C_2^{RR} replacing C_2^{TT}. With this simplification, the matrix elements (4) reduce to

$$\bar{C}_{jj'}^{TT} = (1 - C_2^{TT}) \delta_{jj'} - (1 - 2C_2^{TT}) \langle \delta y_j \delta y_{j'} \rangle (1 - \delta_{jj'}). \quad (5)$$

This result is a first order expansion in C_2^{TT} that can be generalized to higher order if needed, as discussed in the Supplemental Material [17]. However, if we operate in a regime where local corrections C_2^{TT} and C_2^{RR} are much smaller than unity, we simply get $\bar{C}^{TT} = \mathbb{1} - C^{TT}$ and $\bar{C}^{RR} = \mathbb{1} - C^{RR}$, where the diagonal elements of the matrices C^{TT} and C^{RR} are zero. In this case, the trace formula (3) becomes

$$\mathcal{I} \simeq \frac{1}{2 \ln 2} \text{Tr}[(\mathbb{1} - C^{TT} - C^{RR})(C^{RT})^2]. \quad (6)$$

Hence, the existence of pairwise long-range correlations inside each image tends to reduce the MI between the two images compared to the result without surface correlation, $\text{Tr}[(C^{RT})^2]/2 \ln 2$ [see Fig. 1(c) for an illustration]. This means that long-range cross-sample correlations

and long-range surface correlations compete with each other, suggesting that a balance can be found that maximizes the MI for certain geometrical configurations of detectors. This effect is analyzed at the end of this Letter.

In order to validate the theoretical prediction (3) or its approximation (6), we have performed numerical simulations of wave propagation in two-dimensional (2D) disordered slabs with various thicknesses L and scattering mean free paths ℓ. Subwavelength dipole scatterers were placed at random positions inside the slab, and the scalar wave equation was solved numerically using the coupled-dipole method [17]. For each set of parameters, $M = 10^8$ disorder realizations were typically generated numerically, and $\mathbf{x}$ and $\mathbf{y}$ were calculated at the sample input and output surfaces, for various numbers N of detectors and interdistances a between detectors. Then, from the sets of data $\{\mathbf{x}_\alpha, \mathbf{y}_\alpha\}_{\alpha=1,...,M}$, an estimator of the MI was built, based on entropy estimates from nearest neighbor distances [17,27]. Such an estimator is expected to be more accurate than binning estimators—which consist in partitioning the support of $\mathbf{x}$ and $\mathbf{y}$ into bins—for which the bias potentially grows exponentially with the dimension N of $\mathbf{x}$ and $\mathbf{y}$ [28].

Let us first analyze the simplest situation where a single pair of detection points is considered ($N = 1$). The approximation (6) takes the simple form $\mathcal{I} \simeq C^{RT}(\Delta r)^2/2\ln 2$, where Δr is the transverse distance between the detection points placed on both sides of the sample. As shown in Fig. 2(a), this prediction agrees well with the direct estimate of Eq. (1), proving that the MI between $\mathbf{x}$ and $\mathbf{y}$ essentially boils down to the square of their correlation function C^{RT} for $N = 1$. In the multiple scattering regime ($kL \gg k\ell \gg 1$ with $k = 2\pi/\lambda$) C^{RT} is transported along diffusive paths exploring a transverse distance $\sim L$ [9], and the MI is vanishingly small for $\Delta r \gtrsim L$.

For a larger number of detectors ($N > 1$), the behavior of the MI becomes more complex. Let us analyze its

dependence on the interdistance a between detectors. Results corresponding to samples with two different thicknesses L are presented in Fig. 2(b). Here, also we obtain very good agreement between numerical estimates and the trace formula (3) completed by Eq. (5), in which the values of correlators have been obtained from simulations. This confirms that the MI in multiple scattering environments can be computed from the combination of pairwise correlators only. We distinguish three regimes in Fig. 2(b) that can be interpreted by means of the approximation (6). For detector spacing a larger than the extent L of $C^{RT}(\Delta r)$, the MI is driven by detectors placed in front of each other only. Thus, it is independent of a and N times larger than the MI obtained with a single pair of detectors with $\Delta r = 0$ [see Fig. 2(a)]. When a is progressively reduced, the MI starts to increase since more and more pairwise cross-sample correlations get activated. In the absence of correlations between the various components of $\mathbf{x}$ or $\mathbf{y}$, this increase would hold for arbitrary small spacing a. However, we observe that the MI reaches a maximum for a certain critical spacing below which it falls down, thereby revealing the effect of surface correlations. The latter contain both short-range and long-range contributions [15]. Short-range contributions, responsible for the size λ of speckle spots, explain the convergence of the MI towards its $N = 1$ limit when $a \ll \lambda$. Indeed, the MI cannot be increased by adding detectors located in the same speckle spot. Nevertheless, a qualitative analysis of Eq. (6) only does not allow us to infer which contribution triggers the value of the critical distance, and to explain why the MI is globally reduced when the thickness of the medium increases.

To clarify these observations, we studied the dependence of the correlators $C^{RT}(\Delta r)$, $C^{TT}(\Delta r)$, and $C^{RR}(\Delta r)$ on L and ℓ (in the regime $kL \gg k\ell \gg 1$). Simulation results for plane wave illumination and various sets of parameters
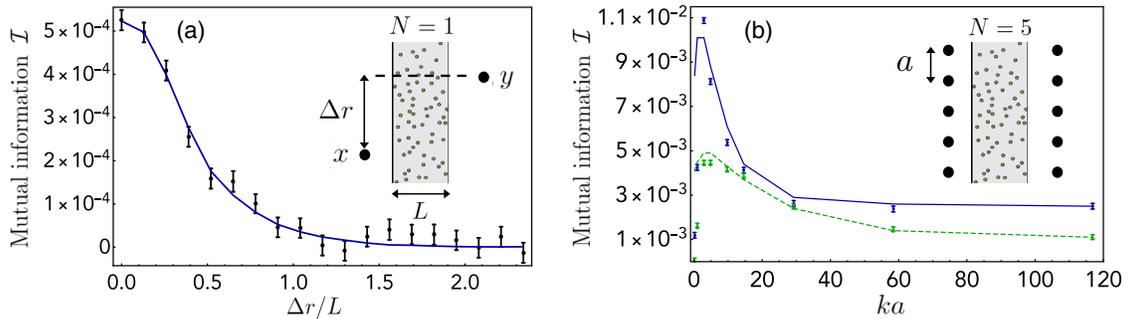

FIG. 2. MI as a function of distances between detectors. The theoretical prediction (3) (lines) is compared to the numerical estimation (dots) for various numbers of detectors N, thicknesses L, and mean free paths ℓ. (a) MI between the intensities measured in reflection ($x = I^R/\langle I^R \rangle$) and transmission ($y = I^T/\langle I^T \rangle$) versus transverse distance Δr. Parameters in the simulation: $kL = 30$, $k\ell = 10$. (b) MI between two sets of $N = 5$ detectors versus detector spacing a, for two thicknesses $kL = 30$ (solid line) and $kL = 80$ (dashed line), and fixed $k\ell = 10$. The constant residual biases in the numerical estimates (dots) have been removed, according to the procedure detailed in the Supplemental Material [17]. Note that the agreement between simulations and theory for $ka \lesssim 1$ can only be qualitative since the hypothesis of weak surface correlations is not fully satisfied.
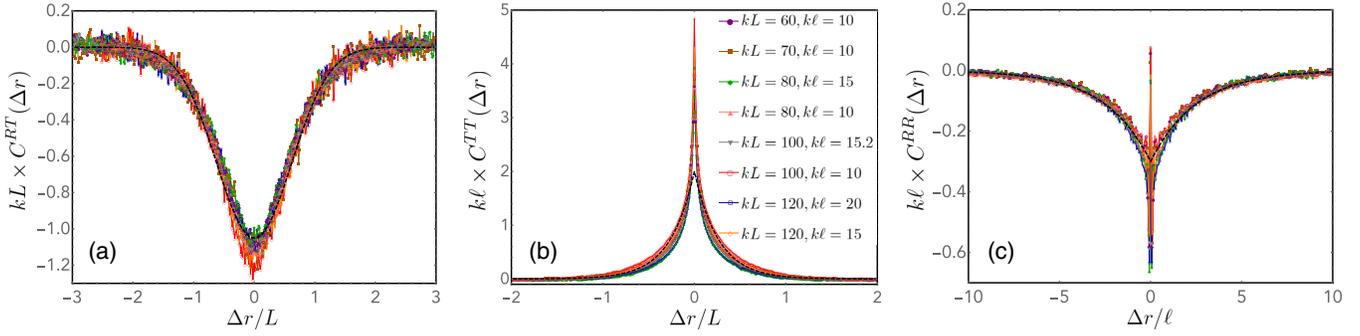
FIG. 3. Scaling of the three correlation functions that are building blocks of the MI: (a) $C^{RT}(\Delta r)$, (b) $C^{TT}(\Delta r)$, and (c) $C^{RR}(\Delta r)$. Numerical results (dots) were obtained by solving the wave equation in 2D, for different values of L and ℓ (see inset). Gaussian contributions (C_1^{TT} and C_1^{RR}), which are short range, have been removed for clarity. Dashed lines are simple fitting functions for $\Delta r \gtrsim \ell$ [Gaussian in (a) and exponentials in (b) and (c)].

$\{kL, k\ell\}$ are shown in Fig. 3. When properly normalized, data points collapse on single curves, suggesting the following scalings: $C^{RT}(\Delta r) = -f_1(\Delta r/L)/(kL)^{d-1}$ for all Δr, and $C^{TT}(\Delta r) = (L/\ell)f_2(\Delta r/L)/(kL)^{d-1}$ and $C^{RR}(\Delta r) = -f_3(\Delta r/\ell)/(k\ell)^{d-1}$ for $\Delta r \gtrsim \ell$. Here, f_1, f_2, and f_3 are three positive decaying functions of range and amplitude close to unity, and d is the space dimension. Theoretical justifications for these scalings are given in the Supplemental Material [17]. Note that, contrary to the well-established behavior of the long-range component of C^{TT}, C^{RT} and C^{RR} do not scale as $\sim 1/g$, where $g = k\ell(kL)^{d-2}$ is the Thouless number of a box of size L [29–31]. In particular, C^{RT} is independent of ℓ. This means that our initial assumption of weak reflection-transmission correlation is a priori robust against mechanisms affecting the transport mean free path, such as structural correlation of the medium [32–34], or near-field coupling between scatterers in dense materials [35–37]. We also point out that the long-range component of C^{RR} is negative, extending over a few mean free paths because waves explore such a distance in the transverse direction before being reflected [38,39]. Finally, for practical calculations, it is instructive to note that the functions f_1, f_2, and f_3 are reasonably well fitted by a Gaussian and two exponentials (see Fig. 3).

The simple scaling forms of the three correlators allow us to push forward the analytic calculation of the trace formula (3), in particular in the interesting limit of a large number of detectors ($N \gg 1$). As C^{RT}, C^{TT}, and C^{RR} are Toeplitz-type matrices, we may use an extension of Szegö's theorem to evaluate the trace for arbitrary spacing a [40]. To simplify the discussion, we focus on the situation where detectors (pixels) are equally spaced in all directions on the surface, in the regime $\ell \lesssim a \ll L$ (see the Supplemental Material [17] for a study in the general case). In this regime, the contribution of C^{RR} is negligible, and the remaining sums over indices in the development of the trace of the matrix product can be replaced by space integrals on the surface. The approximation (6) becomes [17]

$$\mathcal{I} \simeq \frac{N}{2 \ln 2 [(kL)(ka)]^{d-1}} \left(c_{RT} - \frac{c_{TT}}{(ka)^{d-1}} \frac{L}{\ell} \right), \quad (7)$$

where $c_{RT} = \int d\mathbf{r} f_1(r)^2$ and $c_{TT} = \iint d\mathbf{r} d\mathbf{r}' f_1(r) f_2(r') \times f_1(|\mathbf{r} + \mathbf{r}'|)$ are two numerical constants of order unity. The result in Eq. (7) supports previous qualitative observations: the MI scales linearly with the number of detectors, and decreases when the sample thickness increases because the cross-sample correlation C^{RT} itself is reduced. Interestingly, when we normalize Eq. (7) by the MI measured for a single detector $\mathcal{I}_1 = C^{RT}(0)^2/2 \ln 2$, we obtain $\mathcal{I}/\mathcal{I}_1 \propto N(c_{RT}u - c_{TT}u^2/g)$, where $u = (L/a)^{d-1}$. This shows that $\mathcal{I}$ exhibits a maximum triggered by the long-range component of C^{TT}, of the form $\mathcal{I}^{\max} \sim Ng\mathcal{I}_1$, for a critical interdistance a^* much larger than the wavelength and potentially larger than ℓ [$(L/a^*)^{d-1} = (c_{RT}/2c_{TT})g$ so that $a^* \sim \lambda(L/\ell)^{1/(d-1)}$]. Hence, the MI for an array of N detectors with optimized interdistance is enhanced by a factor $Ng \gg 1$ compared to the MI for a single detector. These considerations are confirmed in Fig. 4 by the good agreement between the numerical
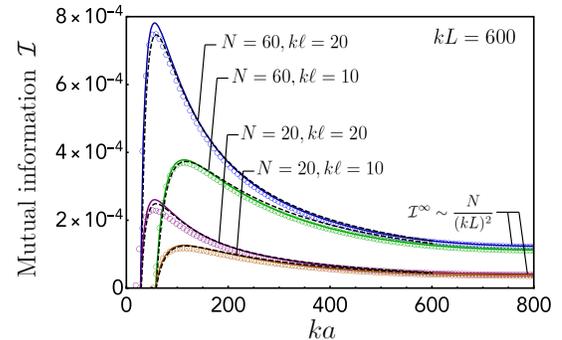

FIG. 4. MI computed from the trace formula (6). Open circles are numerical calculations of the trace, using the scaling functions for C^{RT}, C^{TT}, and C^{RR} identified in Fig. 3. Solid lines stand for the analytic result presented in the Supplemental Material [17], and dashed lines for its approximation (7) valid for $\ell \lesssim a \lesssim L$.

evaluation of the trace (6) and the full analytic prediction detailed in the Supplemental Material [17] and its approximation (7). In particular, denoting by $\mathcal{I}^\infty = N\mathcal{I}_1$ the MI obtained in the large spacing regime $a \gtrsim L$ where only front side correlations contribute, we clearly observe the enhancement factor $\mathcal{I}^{\max}/\mathcal{I}^\infty \sim g$.

In summary, we have presented a quantitative treatment of the MI between two speckle images produced on opposite sides of a multiple scattering medium. The dependence of the MI on length scales characterizing the medium and on the detection geometry highlights the entangled and competitive contributions of long-range intensity correlations. In particular, using an array of N detectors with interdistance a to record the speckle image, the MI can be increased by a factor of Ng compared to the single detector case for a critical value of $a \gg \lambda$. This enhancement factor could guide the development of experimental protocols to measure the MI, in configurations such as that in Ref. [9] for which Ng can be made very large. Although our approach does not give the recipe to recover from $\mathbf{x}$ the information contained in $\mathbf{y}$, or vice versa, it provides quantitative estimates of the MI, and conditions for its optimization, that should help the design of new setups dedicated to information recovery or transfer in complex media.

---

*remi.carminati@espci.fr

[1] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, Nat. Photonics **6**, 283 (2012).

[2] S. Rotter and S. Gigan, Rev. Mod. Phys. **89**, 015005 (2017).

[3] A. Moustakas, H. Baranger, L. Balents, A. Sengupta, and S. Simon, Science **287**, 287 (2000).

[4] J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, Nature (London) **491**, 232 (2012).

[5] O. Katz, P. Heidmann, M. Fink, and S. Gigan, Nat. Photonics **8**, 784 (2014).

[6] S. M. Popoff, A. Goetschy, S. F. Liew, A. D. Stone, and H. Cao, Phys. Rev. Lett. **112**, 133903 (2014).

[7] C. W. Hsu, S. F. Liew, A. Goetschy, H. Cao, and A. D. Stone, Nat. Phys. **13**, 497 (2017).

[8] N. Fayard, A. Cazé, R. Pierrat, and R. Carminati, Phys. Rev. A **92**, 033827 (2015).

[9] I. Starshynov, A. M. Paniagua-Diaz, N. Fayard, A. Goetschy, R. Pierrat, R. Carminati, and J. Bertolotti, arXiv:1707.03622.

[10] S. H. Simon, A. L. Moustakas, M. Stoychev, and H. Safar, Phys. Today **54**, No. 9 38 (2001).

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[12] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications* (Now Publishers, Delft, 2004).

[13] S. E. Skipetrov, Phys. Rev. E **67**, 036621 (2003).

[14] A. Goetschy and A. D. Stone, Phys. Rev. Lett. **111**, 063901 (2013).

[15] E. Akkermans and G. Montambeaux, *Mesoscopic Physics of Electrons and Photons* (Cambridge University Press, Cambridge, England, 2007).

[16] M. C. W. van Rossum and T. M. Nieuwenhuizen, Rev. Mod. Phys. **71**, 313 (1999).

[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.120.073901, which includes Refs. [18–24], that contains (1) the proof of the trace formula; (2) details about wave propagation simulation; (3) an analysis of the mutual information estimator and its bias for various N; (4) a justification of the scaling form of long-range correlation functions; (5) the analytic calculation of the trace formula based on properties of Toeplitz matrices.

[18] E. Kogan and M. Kaveh, Phys. Rev. B **52**, R3813 (1995).

[19] T. M. Nieuwenhuizen and M. C. W. van Rossum, Phys. Rev. Lett. **74**, 2674 (1995).

[20] M. Lax, Phys. Rev. **85**, 621 (1952).

[21] C. W. J. Beenakker, Rev. Mod. Phys. **69**, 731 (1997).

[22] D. B. Rogozkin and M. Y. Cherkasov, Phys. Lett. A **214**, 292 (1996).

[23] P. A. Mello, J. Phys. A **23**, 4061 (1990).

[24] P. A. Mello, E. Akkermans, and B. Shapiro, Phys. Rev. Lett. **61**, 459 (1988).

[25] N. Shnerb and M. Kaveh, Phys. Rev. B **43**, 1279 (1991).

[26] E. Kogan, M. Kaveh, R. Baumgartner, and R. Berkovits, Phys. Rev. B **48**, 9404 (1993).

[27] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).

[28] R. Moddemeijer, Signal Processing **16**, 233 (1989).

[29] M. J. Stephen and G. Cwilich, Phys. Rev. Lett. **59**, 285 (1987).

[30] S. Feng, C. Kane, P. A. Lee, and A. D. Stone, Phys. Rev. Lett. **61**, 834 (1988).

[31] R. Pnini and B. Shapiro, Phys. Rev. B **39**, 6986 (1989).

[32] S. Fraden and G. Maret, Phys. Rev. Lett. **65**, 512 (1990).

[33] L. F. Rojas-Ochoa, J. M. Mendez-Alcaraz, J. J. Sáenz, P. Schurtenberger, and F. Scheffold, Phys. Rev. Lett. **93**, 073903 (2004).

[34] O. Leseur, R. Pierrat, and R. Carminati, Optica **3**, 763 (2016).

[35] T. Nieuwenhuizen, A. Burin, Y. Kagan, and G. Shlyapnikov, Phys. Lett. A **184**, 360 (1994).

[36] A. Dogariu and R. Carminati, Phys. Rep. **559**, 1 (2015).

[37] R. Rezvani Naraghi, S. Sukhov, J. J. Sáenz, and A. Dogariu, Phys. Rev. Lett. **115**, 203903 (2015).

[38] D. B. Rogozkin and M. Y. Cherkasov, Phys. Rev. B **51**, 12256 (1995).

[39] L. Froufe-Pérez, A. Garcia-Martin, G. Cwilich, and J. Saenz, Physica (Amsterdam) **386A**, 625 (2007).

[40] R. M. Gray, Found. Trends Commun. Inf. Theory **2**, 155 (2006).

Supplementary Information
Mutual information between reflected and transmitted speckle images

N. Fayard, A. Goetschy, R. Pierrat, and R. Carminati,

ESPCI Paris, PSL Research University, CNRS, Institut Langevin, 1 rue Jussieu, 75005 Paris, France

I. PROOF OF THE TRACE FORMULA

Our goal here is to prove Eqs. (3), (5), and (6) of the main text (MT), and provide a generalization of Eq. (5) that includes second-order corrections in C_2^{TT}.

First we search for an expression of the jpdf $p(\mathbf{x}, \mathbf{y})$ in terms of $p(\mathbf{x})$ and $p(\mathbf{y})$. This distribution is characterized by the set of correlators

$$\langle x^{\{n\}} y^{\{m\}} \rangle = \langle x_1^{n_1} ... x_N^{n_N} y_1^{m_1} ... y_N^{m_N} \rangle. \tag{1}$$

As explained in the MT, we are interested in leading corrections to the independent result $\langle x^{\{n\}} y^{\{m\}} \rangle = \langle x^{\{n\}} \rangle \langle y^{\{m\}} \rangle$. We adopt a path-integral-type representation of each field E_i ($E_i = \sum_{\mathcal{S}} E_i^{\mathcal{S}}$, where $\mathcal{S}$ is a scattering trajectory), such that each term $x_i^{n_i} \propto |E_i^R|^{2n_i}$ contains n_i replica of propagators $E_i^{R,\mathcal{S}}$, $E_i^{R,\mathcal{S}*}$. The same representation is used for $y_j^{n_j} \propto |E_j^T|^{2n_j}$. We insert this decomposition into Eq. (1) and retain trajectories that provide non-zero contributions after averaging over scatterer positions. Leading corrections to the independent result are due to correlations that involve propagator quadruplets, such as $\{E_i^{R,\mathcal{S}_1}, E_i^{R,\mathcal{S}_2 *}, E_j^{T,\mathcal{S}_3}, E_j^{T,\mathcal{S}_4 *}\}$. The number of these quadruplets formed from replica is $n_i^2 n_j^2$. In addition, their weight is $\langle \delta x_i \delta y_j \rangle = \langle x_i y_j \rangle - \langle x_i \rangle \langle y_j \rangle$. The correlator (1) becomes:

$$\langle x^{\{n\}} y^{\{m\}} \rangle \simeq \langle x^{\{n\}} \rangle \langle y^{\{m\}} \rangle$$
$$+ \sum_{i,j} n_i^2 n_j^2 \langle \delta x_i \delta y_j \rangle \langle x^{\{n\}-1_i} \rangle \langle y^{\{m\}-1_j} \rangle, \tag{2}$$

where we use the notation $x^{\{n\}-1_i} = x_1^{n_1} \ldots x_i^{n_i-1} \ldots x_N^{n_N}$. Then, we compute the characteristic function

$$g(\mathbf{z}, \mathbf{z}') = \langle e^{\mathbf{z} \cdot \mathbf{x} + \mathbf{z}' \cdot \mathbf{y}} \rangle. \tag{3}$$

By inserting the result (2) into the series expansion of $g(\mathbf{z}, \mathbf{z}')$ and using the property

$$\sum_{n_i=0}^{\infty} \frac{n_i z_i^{n_i}}{(n_i - 1)!} x^{\{n\}-1_i} = z_i \, \partial_{z_i} \left(z_i \sum_{n_i=0}^{\infty} \frac{z_i^{n_i}}{n_i!} x^{\{n\}} \right), \tag{4}$$

we get

$$g(\mathbf{z}, \mathbf{z}') = g(\mathbf{z}) g(\mathbf{z}')$$
$$+ \sum_{i,j} \langle \delta x_i \delta y_j \rangle z_i \, \partial_{z_i} [z_i g(\mathbf{z})] z_j' \, \partial_{z_j'} [z_j' g(\mathbf{z}')]. \tag{5}$$

The distribution $p(\mathbf{x}, \mathbf{y})$ follows by taking the inverse Laplace transform of $g(\mathbf{z}, \mathbf{z}')$. It can be written in the following convenient form:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}) p(\mathbf{y})[1 + u(\mathbf{x}, \mathbf{y})], \tag{6}$$

where the correction term $u(\mathbf{x}, \mathbf{y})$ involves all possible pairwise correlations between the reflected and the transmitted speckles:

$$u(\mathbf{x}, \mathbf{y}) = \sum_{i,j} \langle \delta x_i \delta y_j \rangle v_i(\mathbf{x}) v_j(\mathbf{y}), \tag{7}$$

$$v_i(\mathbf{x}) = \frac{\partial_{x_i}[x_i \partial_{x_i} p(\mathbf{x})]}{p(\mathbf{x})}. \tag{8}$$

We are now in position to compute explicitly the mutual information (MI). With the representation (6), it reads

$$\mathcal{I} = \frac{1}{\ln 2} \iint d\mathbf{x} d\mathbf{y} \, p(\mathbf{x}) p(\mathbf{y})[1 + u(\mathbf{x}, \mathbf{y})] \ln [1 + u(\mathbf{x}, \mathbf{y})]. \tag{9}$$

We then expand the integrand as

$$(1 + u) \ln(1 + u) = u + \sum_{n=2}^{\infty} \frac{(-1)^n}{n(n-1)} u^n, \tag{10}$$

and note that the first term of this series does not contribute to (9). Indeed, using the notation $\langle \ldots \rangle_0 = \iint d\mathbf{x} d\mathbf{y} \, p(\mathbf{x}) p(\mathbf{y})(\ldots)$, we obtain

$$\langle u(\mathbf{x}, \mathbf{y}) \rangle_0 = \sum_{i,j} \langle \delta x_i \delta y_j \rangle \langle v_i(\mathbf{x}) \rangle_0 \langle v_j(\mathbf{y}) \rangle_0$$
$$= 0, \tag{11}$$

since $\int d\mathbf{x} \, \partial_{x_i}[x_i \partial_{x_i} p(\mathbf{x})] = 0$. This result holds whatever the distributions $p(\mathbf{x})$ and $p(\mathbf{y})$. Therefore, MI reduces to

$$\mathcal{I} = \sum_{n=2}^{\infty} \frac{(-1)^n}{n(n-1)} \langle u(\mathbf{x}, \mathbf{y})^n \rangle_0. \tag{12}$$

Inasmuch as the images $\mathbf{x}$ and $\mathbf{y}$ are weakly correlated, we retain in the expansion (12) the first term only. The term $\langle u(\mathbf{x}, \mathbf{y})^2 \rangle_0$ gives rise to the trace formula:

$$\mathcal{I} \simeq \frac{1}{2\ln 2} \text{Tr} \left[C^{RT} \bar{C}^{TT} C^{RT} \bar{C}^{RR} \right], \tag{13}$$

where we introduced three $N \times N$ matrices defined by their elements, $C_{ij}^{RT} = \langle \delta x_i \delta y_j \rangle$,

$$\bar{C}_{jj'}^{TT} = \langle v_j(\mathbf{y}) v_{j'}(\mathbf{y}) \rangle_0 \qquad (14)$$
$$= \int d\mathbf{y} \, \frac{\partial_{y_j}[y_j \partial_{y_j} p(\mathbf{y})] \partial_{y_{j'}}[y_{j'} \partial_{y_{j'}} p(\mathbf{y})]}{p(\mathbf{y})},$$
$$\bar{C}_{i'i}^{RR} = \langle v_{i'}(\mathbf{x}) v_i(\mathbf{x}) \rangle_0 \qquad (15)$$
$$= \int d\mathbf{x} \, \frac{\partial_{x_{i'}}[x_{i'} \partial_{x_{i'}} p(\mathbf{x})] \partial_{x_i}[x_i \partial_{x_i} p(\mathbf{x})]}{p(\mathbf{x})}.$$

Next, we assume that the distance a between consecutive pixels or detectors used to record the images is larger than the wavelength λ. In this case, only weak but long-range parts of pairwise correlations inside each image contribute to the jpdf $p(\mathbf{y})$ and $p(\mathbf{x})$. In order to find explicit forms for these distributions, we proceed in the same way as for the moments (1), *i.e.* we consider in the expansion of $\langle y^{\{n\}} \rangle$ (or $\langle x^{\{n\}} \rangle$) all corrections due to the correlation of field quadruplets. We get an expansion similar to Eq. (2):

$$\langle y^{\{n\}} \rangle = \prod_j \langle y^{n_j} \rangle + \sum_{j<j'} n_j^2 n_{j'}^2 \langle \delta y_j \delta y_{j'} \rangle \langle y^{n_j-1} \rangle \langle y^{n_{j'}-1} \rangle$$
$$\times \prod_{k \neq j, k' \neq j'} \langle y_k^{n_k} \rangle \langle y_{k'}^{n_{k'}} \rangle. \qquad (16)$$

Note that the sum runs over indices $j < j'$ instead of i, j in Eq. (2) to avoid redundant counting of pair correlations in the same image. The corresponding distribution, obtained by computing the inverse Laplace transform of the characteristic function, has the same structure as Eq. (6):

$$p(\mathbf{y}) = \prod_k p(y_k) \left[1 + \sum_{j<j'} \langle \delta y_j \delta y_{j'} \rangle v_j(y_j) v_{j'}(y_{j'}) \right], \quad (17)$$

where the definiton of $v_j(y_j)$ follows from Eq. (8):

$$v_j(y_j) = \frac{\partial_{y_j}[y_j \partial_{y_i} p(y_j)]}{p(y_j)}. \qquad (18)$$

The remaining unkown quantity in Eq. (17) is the distribution of intensity recorded on a single detector, $p(y)$. This distribution has been calculated exactly in the nineties using random matrix theory as well as microscopic diagrammatic approaches [1, 2]. Here, we are interested in a tractable approximation of this distribution, that includes second-order corrections in $C_2^{TT} \simeq \langle (\delta y)^2 \rangle - 1$. For this purpose, we follow the approach of Ref. 3, which is similar in spirit to the counting procedure used previously to evaluate the moments. The moment

$\langle y^n \rangle$ are given by

$$\langle y^n \rangle \simeq \sum_{k=0}^{E(n/2)} \left(\frac{C_2^{TT}}{2} \right)^k N_{n,k}, \qquad (19)$$

$$= (n!)^2 \sum_{k=0}^{E(n/2)} \frac{\left(C_2^{TT} \right)^k}{4^k k! (n-2k)!}. \qquad (20)$$

Here C_2^{TT} is the weight of non-Gaussian corrections to the Rayleigh statistics, due to correlations of field quadruplets in the form of Hikami boxes. In addition, $N_{n,k}$ is the number of field combinations that contain k Hikami boxes and $n - 2k$ pairs of fields that form diffusons. In the following we keep terms in Eq. (20) up to the second-order ($k \leq 2$), so that the characteristic function $g(z) = \langle e^{zy} \rangle$ reads

$$g(z) = \frac{1}{1-z} + \frac{z^2}{2(1-z)^3} C_2^{TT} + \frac{3z^4}{4(1-z)^5} (C_2^{TT})^2, \quad (21)$$

and the distribution $p(y)$ becomes

$$p(y) = e^{-y} \left[1 + h_1(y) C_2^{TT} + h_2(y) (C_2^{TT})^2 \right], \qquad (22)$$

with $h_1(y) = (y^2 - 4y + 2)/4$ and $h_2(y) = (y^4 - 16y^3 + 72y^2 - 96y + 24)/32$. Hence, the coefficient given by Eq. (18) and appearing in Eq. (17) is explicitly given by

$$v_j(y_j) = \delta y_j + h_3(y_j) C_2^{TT} + h_4(y_j) (C_2^{TT})^2, \qquad (23)$$

with $\delta y_j = y_j - 1$, $h_3(y) = -y^2 + 3y - 1$ and $h_4(y) = (7y^3 - 39y^2 + 50y - 10)/4$. The jpdf $p(\mathbf{x})$ has a functional form identical to Eq. (17) with C_2^{RR} replacing C_2^{TT} in Eq. (22).

Finally, we compute the matrix elements (14) and (15), using Eqs. (17), (22) and (23). This is a tedious but straightfoward calculation. For example, the first-order expansion in C_2^{TT} and $\langle \delta y_j \delta y_\alpha \rangle$ of the coefficient given by Eq. (8) that enters into the definitions (14) and (15) reads

$$v_j(\mathbf{y}) = \delta y_j + (1 - 2y_j) \sum_{\alpha \neq j} \langle \delta y_j \delta y_\alpha \rangle \delta y_\alpha + h_3(y_j) C_2^{TT}$$
$$+ \sum_{\alpha \neq j} \langle \delta y_j \delta y_\alpha \rangle [h_5(y_j) \delta y_\alpha + (1 - 2y_j) h_3(y_\alpha)] C_2^{TT}, \quad (24)$$

with $h_5(y) = 5y^2 - 12y + 3$. The final result, obtained after integration over $\mathbf{y}$, looks quite simple:

$$\bar{C}_{jj'}^{TT} = \left[1 - C_2^{TT} + 5 \left(C_2^{TT} \right)^2 \right] \delta_{jj'}$$
$$- (1 - 2C_2^{TT}) \langle \delta y_j \delta y_{j'} \rangle (1 - \delta_{jj'}). \qquad (25)$$

This completes the proof of Eq. (5) of the MT, including second-order correction in C_2^{TT}. We note that second-order correction shows up in the diagonal matrix element only. We also stress that the result (25) is a first order expansion in $\langle \delta y_j \delta y_{j'} \rangle$ ($j \neq j'$), as the distribution (17) itself. Numerical simulations confirm that this approximation is sufficient for $a \gtrsim \lambda$.

II. WAVE EQUATION SIMULATIONS WITH THE COUPLE DIPOLE METHODS

The numerical simulations were performed using the coupled-dipole method [4]. This is an exact method providing that the scatterers can be described by dipoles of subwavelength dimension. Here we considered cylindrical scatterers of subwavelength cross-sections illuminated by an incident field polarized along their longitudinal axis, so that the wave equation is effectively two dimensional. We first generate a configuration of the disorder by placing randomly N_c cylinders in a box, the longitudinal size of which is given by the thickness L. The other dimension is chosen to be large enough to mimic a slab geometry (10 times the thickness L). To avoid scatterers overlaps, a minimum distance is forced between them. This distance is small enough not to introduce disorder correlations. Moreover, to lower the number of scatterers inside the system and save computational time, the polarizability α of each scatterer has been chosen such that it maximizes the scattering cross-section $\sigma_s = k^3|\alpha|^2/4$ while verifying energy conservation [i.e. the scattering cross-section σ_s should be equal to the extinction cross-section $\sigma_e = k\,\mathrm{Im}(\alpha)$]. This leads to $\alpha = 4i/k^2$. By adjusting the number density of scatterers ρ, we can simulate systems with different scattering mean-free paths $\ell = 1/(\rho\sigma_s)$. The multiple interactions between the scatterers are described by a set of N_c linear equations which reads

$$E_j = E_0(\mathbf{r}_j) + k^2\alpha \sum_{n=1, n\neq j}^{N_c} G_0(\mathbf{r}_j - \mathbf{r}_n)E_n \qquad (26)$$

where E_j is the exciting electric field on scatterer j lying at position $\mathbf{r}_j$. E_0 is the incident field (plane-wave at normal incidence) and $G_0(\mathbf{r}-\mathbf{r}_0)$ is the Green function in vacuum which gives the electric field produced at position $\mathbf{r}$ by a source dipole lying at position $\mathbf{r}_0$. Its expression reads

$$G_0(\mathbf{r}-\mathbf{r}_0) = \frac{\mathrm{i}}{4}H_0(k|\mathbf{r}-\mathbf{r}_0|), \qquad (27)$$

where H_0 is the Hankel function of the first kind. Once this system is solved and the exciting fields are known, a similar equation is used to compute the field at any position at the sample surfaces:

$$E(\mathbf{r}) = E_0(\mathbf{r}) + k^2\alpha \sum_{n=1}^{N_c} G_0(\mathbf{r}-\mathbf{r}_n)E_n. \qquad (28)$$

Finally, disorder averages are performed to compute the speckle correlation function $C^{RT}(\Delta\mathbf{r})$, $C^{TT}(\Delta\mathbf{r})$, and $C^{RR}(\Delta\mathbf{r})$. These correlation function are defined as

$$C^{XY}(\Delta\mathbf{r}) = \frac{\langle I^X(\mathbf{r})I^Y(\mathbf{r}')\rangle}{\langle I^X(\mathbf{r})\rangle\langle I^Y(\mathbf{r}')\rangle} - 1, \qquad (29)$$
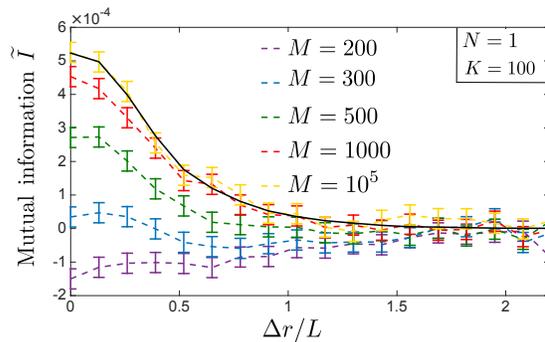

Figure 1. MI between two detectors as function of their transverse distance Δr. Numerical estimates (dashed lines) are computed from Eq. (33) and compared to the theoretical result $\mathcal{I} = C^{RT}(\Delta r)^2/2\ln 2$ (solid line).

where $\mathbf{r}' = \mathbf{r} + \Delta\mathbf{r}$, the superscripts X and Y stand for R or T, and the intensities are defined as $I^X(\mathbf{r}) = |\delta E^X(\mathbf{r})|^2$, with $\delta E^X(\mathbf{r}) = E^X(\mathbf{r}) - \langle E^X(\mathbf{r})\rangle$. As an example, for disorder strength $k\ell = 10$ and sample thickness $kL = 100$, we used $N_c = 1750$ dipoles and 6.4 millions of configurations.

III. MUTUAL INFORMATION ESTIMATION

Most intuitive estimators of MI between $\mathbf{x}$ and $\mathbf{y}$ are binning estimators, that consist in partitioning the supports of $\mathbf{x}$ and $\mathbf{y}$ into bins, representing jpdf by histograms built from M realizations, and approximating MI by a finite sum. Generally, these estimators suffer from bias due to finite M and finite bin size [5]. In dimension $N = 1$, it is possible to find a bin size that minimizes the bias, but no such strategy is available for the case $N > 1$. In particular, bias remain non-zero in the limit $M \to \infty$ and it grows drastically with the dimension N.

In order to limit the previous bias issue, we used an estimator of MI based on entropy estimates built from nearest neighbor distances, measured in the space spanned by $\mathbf{x}$ and $\mathbf{y}$ [6]. In the following, we briefly summarize the approach of Ref. [6], with the aim of clarifying the bias dependence on the system parameters. First we interpret the entropy $\mathcal{H}(\mathbf{x}, \mathbf{y})$ as the average of $\log p(\mathbf{z})$, with $\mathbf{z} = (\mathbf{x}, \mathbf{y})$. Its unbiased estimator, built from the data set $\{\mathbf{z}_\alpha\}_{\alpha=1\dots M}$, is $\widetilde{\mathcal{H}}(\mathbf{x}, \mathbf{y}) = -\overline{\log p(\mathbf{z}_\alpha)}$, where we use the notation $\overline{(\dots)_\alpha} = M^{-1}\sum_{\alpha=1}^M (\dots)_\alpha$. Then, we construct an estimate of $\log p(\mathbf{z}_\alpha)$ by considering the ball centered in $\mathbf{z}_\alpha$ that contains its K nearest neighbors. Let us denote by ϵ_α the diameter of this ball. If it is small enough, we approximately have $p(\mathbf{z}_\alpha) \propto p_\alpha(\epsilon_\alpha)/\epsilon_\alpha^{2N}$, where $p_\alpha(\epsilon)$ is the probability to get a realization in the ball of diameter ϵ centered at $\mathbf{z}_\alpha$. Up to an irrelevant constant, we
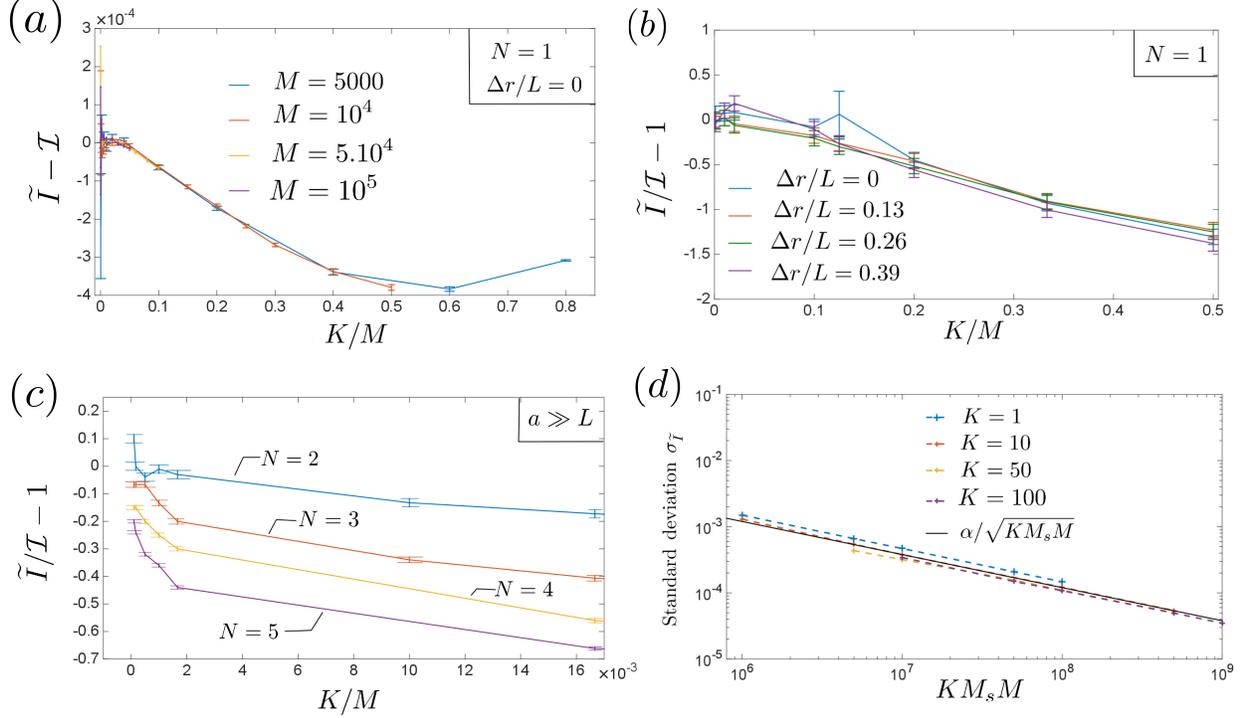
Figure 2. Dependence of the bias $\tilde{\mathcal{I}} - \mathcal{I}$ on the number of realizations M and neighbors K (a), on the MI $\mathcal{I}$ that decreases with Δr (b), and on the numbers of detectors N (c). Scaling of the standard deviation of the estimator given in Eq. (33) (d). In all panels, the estimator $\tilde{\mathcal{I}}$ has been averaged over M_s sets of M realizations to reduce the standard deviation. Parameters of the wave propagation simulation are $kL = 30$ and $k\ell = 10$.

obtain

$$\widetilde{\mathcal{H}}(\mathbf{x}, \mathbf{y}) \simeq -\overline{\log p_\alpha(\epsilon_\alpha)} + 2N \overline{\log \epsilon_\alpha}. \qquad (30)$$

This representation is useful because we can replace the first term by its average $\int d\epsilon P_K(\epsilon) \log p(\epsilon)$, where $P_K(\epsilon)$ is the probability to find a ball of diameter ϵ containing K realizations. By expressing $P_K(\epsilon)$ in terms of $p(\epsilon)$, we easily compute the previous integral to obtain

$$\widetilde{\mathcal{H}}(\mathbf{x}, \mathbf{y}) \simeq -\psi(K) + \psi(M) + 2N \overline{\log \epsilon_\alpha}, \qquad (31)$$

where ψ is the digamma function. Formula (31) is true, in principle, for any value of K. However, ϵ_α increases with K, so that previous approximations may break down at large K, resulting in a large bias (see below for illustration).

The same procedure can be adopted to construct an estimator of $\mathcal{H}(\mathbf{x})$. The only subtle point is that K has been defined in the space of dimension $2N$ spanned by $\mathbf{z}$, so that the effective number of neighbors in the marginal space of dimension N is different from K. In particular, this number depends on the choice of the norm. In the following, we use the maximum norm $|\mathbf{z}| = \max(|\mathbf{x}|, |\mathbf{y}|)$, so that $\epsilon = \max(\epsilon^{\mathbf{x}}, \epsilon^{\mathbf{y}})$. In that case, the number of neighbors in the marginal space is approximately $K^{\mathbf{x}} \simeq n^{\mathbf{x}} + 1$, where $n^{\mathbf{x}}$ is the number of elements contained

in the ball (defined in the marginal space) of diameter ϵ. Hence, we have

$$\widetilde{\mathcal{H}}(\mathbf{x}) \simeq -\overline{\psi(n_\alpha^{\mathbf{x}} + 1)} + \psi(M) + N \overline{\log \epsilon_\alpha}. \qquad (32)$$

By combining Eq. (31) and Eq. (32), we finally obtain an estimator of $\mathcal{I} = \mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{x}, \mathbf{y})$:

$$\widetilde{\mathcal{I}} = \psi(K) + \psi(M) - \frac{1}{M} \sum_{\alpha=1}^{M} [\psi(n_\alpha^{\mathbf{x}} + 1) + \psi(n_\alpha^{\mathbf{y}} + 1)]. \qquad (33)$$

An important property of this estimator is that its bias remains moderate even at large N and it tends to zero when M tends to infinity.

In practice, we transform the data $\{\mathbf{z}_\alpha\}_{\alpha=1\ldots M}$ to make them almost uniformly distributed, in view of minimizing the bias. This is possible because MI is invariant under homeomorfic tranformation of the variables $\mathbf{x}$ and $\mathbf{y}$. Using the fact that the marginal distributions of the components of $\mathbf{x}$ and $\mathbf{y}$ are close to the Rayleigh distribution, we apply the transformation $x_i \to x_i' = e^{-x_i}$ and $y_j \to y_j' = e^{-y_j}$ for $i, j = 1 \ldots N$. Then, we choose a value of K (according to a strategy discussed below) and, for each $\mathbf{z}_\alpha'$, we search for ϵ_α which is twice the distance to the Kth neighbor of $\mathbf{z}_\alpha'$ in the sense of the maximal norm. Finally, we evaluate $n_\alpha^{\mathbf{x}'}$ and $n_\alpha^{\mathbf{y}'}$ in the two marginal spaces and compute (33).
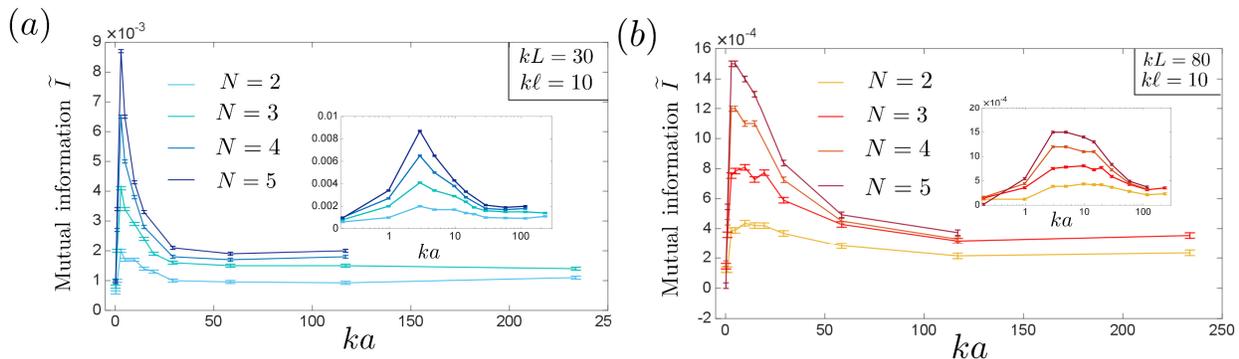
Figure 3. MI between two sets of N detectors vs detector spacing a, for two thicknesses $kL = 30$ (a) and $kL = 80$ (b), and fixed $k\ell = 10$. Insets show the same data in log scale, in order to emphasize that the position and the amplitude of the maximum are in agreement with theoretical predictions: $ka^* \propto L/\ell$ and $\mathcal{I}^{\max}/\mathcal{I}^\infty \simeq k\ell/4 \simeq 2.5$ ($c_{RT} \simeq c_{TT} \simeq 1.2$ in 2D, see Sec. V for details).

In order to establish a reliable procedure for minimizing errors in estimates of MI, we studied the dependence of the the bias $\mathcal{B} = \langle \tilde{\mathcal{I}} \rangle - \mathcal{I}$ and standard deviation $\sigma_{\tilde{\mathcal{I}}}$ of the estimator (33) on the different parameters of the problem: number of realizations M, number of nearest neighbors K, value of the true MI $\mathcal{I}$, and number of detectors N. An illustration of the strong dependence of the bias on M and $\mathcal{I}$ is given in Fig. 1, where we represent the same content as Fig. 2(a) of the MT — i.e. the MI between two detectors as a function of their transverse distance — for different values of M. The comparison with the theoretical result $\mathcal{I} = C^{RT}(\Delta r)^2/2\ln2$ (solid lines) shows that the bias decreases at large M and gets smaller when we reduce $\mathcal{I}$ by increasing Δr. More precisely, the analysis of the bias for fixed values of N and $\mathcal{I}$ reveals that the latter depends on the parameter K/M only. This scaling dependence is shown in Fig. 2(a) and agrees with similar analysis performed in Ref. 6. In addition, the bias appears to be proportional to $\mathcal{I}$ [see Fig. 2(b)], so that it can be expressed in the form

$$\mathcal{B} \simeq \mathcal{I}\, f_N\left(\frac{K}{M}\right). \tag{34}$$

The function f_N is represented in Fig. 2(c) for various N.

We also studied the behavior of the standard deviation $\sigma_{\tilde{\mathcal{I}}}$ and found that it is independent of $\mathcal{I}$ and N (data not shown), while it depends on M and K as $\sim (KM)^{-1/2}$. As expected from the construction of the estimator (33), $\sigma_{\tilde{\mathcal{I}}}$ decreases at large K whereas the bias increases with K. We can further reduce $\sigma_{\tilde{\mathcal{I}}}$ without affecting significantly the computation time of the estimator by averaging Eq. (33) over M_s sets of M realizations. Hence, the standard deviation takes the form:

$$\sigma_{\tilde{\mathcal{I}}} \simeq \frac{\alpha}{\sqrt{KM_sM}}, \tag{35}$$

where $\alpha \simeq 1.3$, according to Fig. 2(d).

In our simulations, we used the scaling forms (34) and (35) to infer the values of K and M that are required to minimize errors in MI estimates. For a standard deviation satisfying $\sigma_{\tilde{\mathcal{I}}}/\mathcal{I} \lesssim \epsilon$ where ϵ is some small number, we get $K \gtrsim \alpha^2/M'\mathcal{I}^2\epsilon^2 \simeq 10^{-8}/\mathcal{I}^2\epsilon^2$, since the product $M' = M_sM$ is kept fixed in our study ($M' = 10^8$). This gives us the value of K required to estimate $\mathcal{I}$. Furthermore, a constraint on the relative bias in the form $\mathcal{B}/\mathcal{I} \lesssim \epsilon$ translates into $f_N(K/M) \lesssim \epsilon$, from which we find the constraint on M using Fig. 2(c). Following this analysis, we chose the following parameters in the MT: $K = 10^2$, $M = 10^5$ in Fig. 2(a) of the MT; $K = 10$, $M = 10^5$ for $kL = 30$ and $K = 10^2$, $M = 10^6$ for $kL = 80$ in Fig. 2(b) of the MT.

Finally, we provide more details about Fig. 2(b) of the MT which focuses on the dependence of MI on the distance a between detectors. We have represented in Fig. (3) the MI computed from Eq. (33) for various number N of detectors. We observe that the estimate is not exactly linear with N, which is the signature of residual bias in the data. In order to get rid of them, we take advantage of Eq. (34) to express the true MI in the form

$$\mathcal{I} = \beta\tilde{\mathcal{I}}, \tag{36}$$

$$\beta = \left[1 + f_N(K/M)\right]^{-1}. \tag{37}$$

The proportionality coefficient β depends on N only. In particular, it is independent of the distance a, so that we can evaluate it in the large spacing regime $a \gg L$ where only frontside correlators contribute to MI:

$$\beta = \frac{N\mathcal{I}_1}{\tilde{\mathcal{I}}^\infty}. \tag{38}$$

Figure 2(b) of the MT was obtained from Fig. (3) by applying the rescaling given by Eqs. (36) and (38), with $\mathcal{I}_1$ deduced from Eqs. (13) and (25).

IV. SCALING FORMS FOR THE LONG-RANGE COMPONENTS OF C^{RT}, C^{TT}, AND C^{RR}

The correlations functions C^{RT}, C^{TT}, and C^{RR} are defined according to Eq. (29), where the intensities are square modulus of the fluctuating parts of the fields. In this way, we remove spurious interferences between mean fields and scattered fields, that vanish in the limit $L \gg \ell \gg \lambda$. The Gaussian contribution to Eq. (29), denoted C_1, is obtained by pairing fields to form averages of complex conjugate pairs. On the other hand, non-Gaussian contributions necessarily involve scattering paths that connect four fields, since $\langle \delta E_X(\mathbf{r}) \rangle = 0$. By noting $\langle \dots \rangle_c$ the non-Gaussian contributions and omitting the spatial dependence in the notation, we obtain

$$C^{XY}(\Delta\mathbf{r}) = \frac{|\langle \delta E^X \delta E^{Y*} \rangle|^2}{\langle I^X \rangle \langle I^Y \rangle} + \frac{\langle \delta E^X \delta E^{X*} \delta E^Y \delta E^{Y*} \rangle_c}{\langle I^X \rangle \langle I^Y \rangle}. \tag{39}$$

The first term of Eq. (39) is the C_1 contribution. As far as C^{TT} and C^{RR} are concerned, this contribution is of the order of unity at short distances $\Delta\mathbf{r} \lesssim \lambda$. On the other hand, it is negligible in C^{RT} for all distances as long as $L \gg \lambda$ [7]. In the following, we disregard this well-known contribution and focus on the non-Gaussian term in Eq. (39). In the weak scattering regime $k\ell \gg 1$, it is dominated by four field correlations made of one Hikami box, which are termed C_2. C_2^{TT} and C_2^{RR} contain both short-range and long-range components, while C_2^{RT} does not contain any short-range feature. In what follows, we provide scaling forms for long-range components only, which we will label $C_{2,\text{long}}$.

Two frameworks are available to compute $C_{2,\text{long}}^{XY}(\Delta\mathbf{r})$: the microscopic diagrammatic approach and random matrix theory (RMT) [8, 9]. The diagrammatic formalism is in principle more powerful since it allows to compute all short-range and long-range features of the correlation functions. However, great care must be taken to properly account for leading contributions, in particular when reflected fields are involved [10, 11]. To avoid such complications, we will mostly rely on the RMT approach which accounts, by construction, for all contributions ensuring flux conservation [12]. As we will show below, one major drawback is that it is based on an isotropy assumption that captures, in an open slab geometry, the integrated correlation $\int C_{2,\text{long}}^{XY}(\Delta\mathbf{r}) \mathrm{d}\Delta\mathbf{r}$ only. This turns out to be sufficient to infer the scalings we are interested in.

With the help of the path-integral-type representation already used in Section I, we can show that far-field correlation functions, $C^{XY}(\mathbf{k}_b, \mathbf{k}_{b'})$, are simply related to speckle correlation functions recorded at the sample surface, $C^{XY}(\Delta\mathbf{r})$. In particular if we compute the correlators in the observation directions $\mathbf{k}_b = \mathbf{k}_{b'}$, C_2^{XY} satisfy

the following relations:

$$C_2^{TT}(\mathbf{k}_b, \mathbf{k}_b) = 2 \int \frac{\mathrm{d}\Delta\mathbf{r}}{\mathcal{A}} C_{2,\text{long}}^{TT}(\Delta\mathbf{r}), \tag{40}$$

$$C_2^{RR}(\mathbf{k}_b, \mathbf{k}_b) = 2 \int \frac{\mathrm{d}\Delta\mathbf{r}}{\mathcal{A}} C_{2,\text{long}}^{RR}(\Delta\mathbf{r}), \tag{41}$$

$$C_2^{RT}(\mathbf{k}_b, \mathbf{k}_b) = \int \frac{\mathrm{d}\Delta\mathbf{r}}{\mathcal{A}} C_2^{RT}(\Delta\mathbf{r}), \tag{42}$$

where $\mathcal{A} = W^{d-1}$ is the transverse area covered by the input illumination. Note the difference in the prefactors: there are two possibilities to form field pairing between the surface and the far-field for C_2^{TT} and C_2^{RR}, while there is only one for C_2^{RT} (the remaining pairing is negligible because it is made of mean fields crossing the full sample).

Next, we evaluate the left hand-side of previous equations using RMT. Let us remind briefly the approach [13, 14]. First we express the intensities $I^X(\mathbf{k}_b)$ as elements of transmission and reflection matrices of the slab, t and r: $I^T(\mathbf{k}_b) \propto |t_{ba}|^2$ and $I^R(\mathbf{k}_b) \propto |r_{ba}|^2$, where the subscript a stands for the input plane wave $\mathbf{k}_a$. Second we use singular value decomposition of t and r, taking into account constraints imposed by flux conservation and time-reversal symmetry:

$$t = U\sqrt{\tau}V^\dagger, \tag{43}$$

$$r = -V^*\sqrt{1-\tau}V^\dagger. \tag{44}$$

The RMT approach consists in assuming that U, V, and τ are three independent random matrices, with U and V uniformly distributed in the unitary group, and τ a diagonal matrix whose elements are the so-called transmission eigenvalues. The size of the matrices is equal to the number of propagating channels inside a waveguide of transverse section $\mathcal{A}$, $N_1 = (kW/\pi)^{d-1}$. Using well-known statistical properties of random unitary matrices, we first compute the leading contributions to the correlators $\langle I^X(\mathbf{k}_b) I^Y(\mathbf{k}_{b'}) \rangle$ in the limit $N_1 \gg 1$:

$$\langle I^T(\mathbf{k}_b) I^T(\mathbf{k}_{b'}) \rangle \propto \frac{1 + \delta_{b,b'}}{N_1^4} \left[\left(1 - \frac{1}{N_1}\right) \langle \mathrm{Tr}^2(\tau) \rangle \right. $$
$$\left. + \langle \mathrm{Tr}(\tau^2) \rangle \right], \tag{45}$$

$$\langle I^R(\mathbf{k}_b) I^R(\mathbf{k}_{b'}) \rangle \propto \frac{1 + \delta_{b,b'}}{N_1^4} \left[\left(1 - \frac{2}{N_1}\right) \langle \mathrm{Tr}^2(1-\tau) \rangle \right. $$
$$\left. + \langle \mathrm{Tr}(1-\tau)^2 \rangle \right], \tag{46}$$

$$\langle I^R(\mathbf{k}_b) I^T(\mathbf{k}_{b'}) \rangle \propto \frac{1 + \delta_{b,b'}}{N_1^4} \left[\left(1 - \frac{1}{N_1}\right) \langle \mathrm{Tr}(\tau) \mathrm{Tr}(1-\tau) \rangle \right. $$
$$\left. + \langle \mathrm{Tr}(\tau(1-\tau)) \rangle \right]. \tag{47}$$

We stress that the RMT approach does not capture properly the dependence of these correlators in $\mathbf{k}_b - \mathbf{k}_{b'}$,
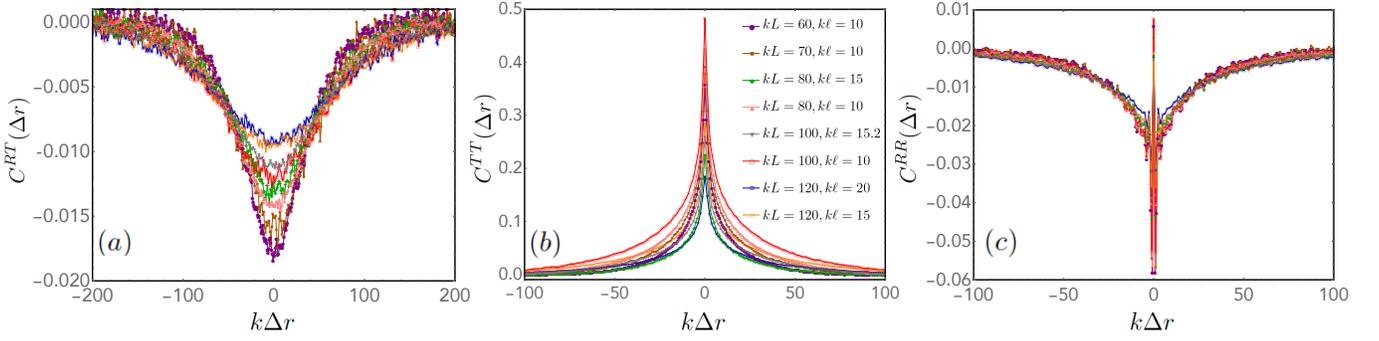
Figure 4. Spatial dependence of the non-Gaussian part of the three correlation functions that build up the MI: (a) $C^{RT}(\Delta r)$ (b) $C^{TT}(\Delta r)$ (c) $C^{RR}(\Delta r)$. Numerical results were obtained by solving the wave equation in 2D, for different values of L and ℓ (see inset). The gaussian contribution to the correlation [first term of Eq. (39)], which is short-range, has been removed for clarity.

because it assumes a quasi-one-dimensional geometry ($W \ll L$), in which the density of states filters out wave-vector components $\mathbf{k}_b \neq \mathbf{k}_{b'}$. However, in the case $\mathbf{k}_b = \mathbf{k}_{b'}$, there is no filtering so that predictions from RMT are valid in the slab geometry as well ($W \gg L$).

From Eqs. (45), (46), and (47), we can evaluate the correlation functions $C^{XY}(\mathbf{k}_b, \mathbf{k}_{b'})$ explicitly, using $\langle \mathrm{Tr}(\tau^2) \rangle = 2N_1 \bar{\tau}/3$, where $\bar{\tau} \propto \ell/L$ is the mean transmission eigenvalue. After having removed the C_1 contribution, we finally obtain

$$C_2^{TT}(\mathbf{k}_b, \mathbf{k}_b) = \frac{4}{3} \frac{1}{N_1 \bar{\tau}}, \tag{48}$$

$$C_2^{RR}(\mathbf{k}_b, \mathbf{k}_b) = \frac{2}{N_1} \left(-1 + \frac{2\bar{\tau}}{3} \right), \tag{49}$$

$$C_2^{RT}(\mathbf{k}_b, \mathbf{k}_b) = \frac{1}{N_1} \frac{-2/3 + \bar{\tau}}{1 - \bar{\tau}}. \tag{50}$$

We now have all the building blocks to evaluate the scaling forms of $C_{2,\mathrm{long}}^{XY}(\Delta r)$. Combining previous formula with Eqs. (40), (41), and (42), in the regime $\bar{\tau} \ll 1$, we find

$$\left(\frac{k}{\pi} \right)^{d-1} \int \mathrm{d}\Delta r \, C_{2,\mathrm{long}}^{TT}(\Delta r) = \frac{2}{3\bar{\tau}}, \tag{51}$$

$$\left(\frac{k}{\pi} \right)^{d-1} \int \mathrm{d}\Delta r \, C_{2,\mathrm{long}}^{RR}(\Delta r) = -2, \tag{52}$$

$$\left(\frac{k}{\pi} \right)^{d-1} \int \mathrm{d}\Delta r \, C_2^{RT}(\Delta r) = -\frac{2}{3}. \tag{53}$$

Typical transmitted waves, which cross the sample through a diffusion process, explore a transverse distance $\sim L$. On the other hand, typical reflected waves explore a transverse distance $\sim \ell$. For these reasons, we look for scaling functions in the form $C^{RT}(\Delta r) = \alpha_{RT} f_1(\Delta r/L)$, $C_{2,\mathrm{long}}^{TT}(\Delta r) = \alpha_{TT} f_2(\Delta r/L)$, and $C_{2,\mathrm{long}}^{RR}(\Delta r) = \alpha_{RR} f_3(\Delta r/\ell)$, where f_1, f_2, and f_3 are three positive decaying function of range and amplitude close to unity.

Inserting these trials functions into Eqs. (51), (52), and (53), we find

$$\alpha_{RT} = -\frac{1}{(kL)^{d-1}}, \tag{54}$$

$$\alpha_{TT} = \frac{1}{(kL)^{d-1}} \frac{L}{\ell}, \tag{55}$$

$$\alpha_{RR} = -\frac{1}{(k\ell)^{d-1}}, \tag{56}$$

where numerical prefactors have been absorbed in the definition of f_1, f_2, and f_3. Hence, we recover the scaling $C_{2,\mathrm{long}}^{TT} \sim 1/g$, with $g = k\ell(kL)^{d-2}$, which has been popularized in the eighties, as well as the scaling of C^{RT} established recently with the help of the microscopic diagrammatic approach [7, 15–17].

We confirmed in Fig. 3 of the MT the validity of previous scaling functions by showing that correlation functions, computed numerically from wave equation simulation, collapse on single curves when properly normalized. For completeness, we report in Fig. 4 the same data without normalization.

V. ANALYTIC CALCULATION OF THE TRACE FORMULA

We provide here an analytic evaluation of the trace formula (13), in the form of the approximation (6) of the MT. We reproduce it for convenience:

$$\mathcal{I} \simeq \frac{1}{2\ln 2} \mathrm{Tr} \left[(\mathbb{1} - C^{TT} - C^{RR})(C^{RT})^2 \right]. \tag{57}$$

The elements of the three $N \times N$ matrices under the trace are assumed to be of the following form:

$$C_{ij}^{RT} = -\frac{1}{(kL)^{d-1}} f_1\left(|\mathbf{r}_i - \mathbf{r}_j|/L\right), \qquad (58)$$

$$C_{jj'}^{TT} = \frac{1}{(kL)^{d-1}} \frac{L}{\ell} f_2\left(|\mathbf{r}_j - \mathbf{r}_{j'}|/L\right), \qquad (59)$$

$$C_{ii'}^{RR} = -\frac{1}{(k\ell)^{d-1}} f_3\left(|\mathbf{r}_i - \mathbf{r}_{i'}|/\ell\right), \qquad (60)$$

where d is the space dimension and the functions f_1, f_2, and f_3 are three positive decaying functions of amplitude and range close to unity (see Sec. IV for justification of these scalings). In the following, we focus for simplicity on the situation where detectors are separated by a distance a in all directions of the surface.

Let us start our analysis with the regime where the density of detectors is sufficiently large to approximate sums overs indices appearing in the development of Eq. (57) by integration over surface positions. According to the range of f_1, f_2, and f_3, this regime requires $a \lesssim L$ for C^{RT} and C^{TT}, and $a \lesssim \ell$ for C^{RR}. In fact, this last condition may be relaxed since the contribution of C^{RR} will turn out to be negligible for all values of $\lambda \lesssim a \lesssim L$ (see the discussion below). In this continous limit, we make the approximations:

$$\sum_{i,j} (C_{ij}^{RT})^2 \simeq N \int \frac{d\boldsymbol{\Delta r}}{a^{d-1}} C^{RT}(\Delta r)^2, \qquad (61)$$

$$\sum_{i,\alpha,\beta} C_{i\alpha}^{RT} C_{\alpha\beta}^{XX} C_{\beta i}^{RT} \simeq N \iint \frac{d\boldsymbol{\Delta r}\, d\boldsymbol{\Delta r}'}{a^{2(d-1)}} C^{RT}(\Delta r)$$
$$\times C^{XX}(\Delta r') C^{RT}(|\boldsymbol{\Delta r} + \boldsymbol{\Delta r}'|), \quad (62)$$

where the superscript X stands for R or T. Let us note $\mathcal{A}(W)$ the area of each image, parametrized by a typical transverse dimension W: $\mathcal{A}(W) \sim W^{d-1}$. With the scaling expressions (58), (59), and (60), the mutual information (57) becomes

$$\mathcal{I} \simeq \frac{N}{2\ln 2[(kL)(ka)]^{d-1}} \left[c_{RT} - \frac{c_{TT}(L/\ell) - c_{RR}}{(ka)^{d-1}} \right], \quad (63)$$

where we introduced the coefficients

$$c_{RT} = \int_{\mathcal{A}(W/L)} d\mathbf{r} f_1(r)^2, \qquad (64)$$

$$c_{TT} = \iint_{\mathcal{A}(W/L)} d\mathbf{r} d\mathbf{r}' f_1(r) f_2(r') f_1(|\mathbf{r} + \mathbf{r}'|), \qquad (65)$$

$$c_{RR} = \int_{\mathcal{A}(W/L)} d\mathbf{r} \int_{\mathcal{A}(W/\ell)} d\mathbf{r}' f_1(r) f_3(r') f_1\left(|\mathbf{r} + (\ell/L)\mathbf{r}'|\right). \qquad (66)$$

In the limit $W \gg L \gg \ell$, these coefficients become constants of the order of unity and the result (63) reduces to Eq. (7) of the MT. For example, if we use the fitting functions $f_1(r) = \alpha_1 e^{-\beta_1 r^2}$ and $f_2(r) = \alpha_2 e^{-\beta_2 r}$ as shown in Fig. 3 of MT, we get, in dimension $d = 2$,

$c_{RT} = \alpha_1^2 \sqrt{\pi/2\beta_1} \simeq 1.2$ and $c_{TT} = \pi\alpha_1^2\alpha_2 e^{\beta_2^2/2\beta_1}[1 - \mathrm{erf}(\beta_2/\sqrt{2\beta_1})]/\beta_1 \simeq 1.2$ ($\alpha_1 \simeq 1.05$, $\beta_1 \simeq 1.3$, $\alpha_2 \simeq 2$, and $\beta_2 \simeq 3.8$ were used to fit the data).

We now turn to the analytic evaluation of the trace (57) without assuming any condition on a. This is possible in the limit of large number of detectors ($N \gg 1$), where we can use powerful theorems for spectral properties of Toeplitz matrices [18]. Furthermore, we restrict the analysis to the dimension $d = 2$, where numerical simulations have been performed (see Sec. II). In this case, the matrices C^{RT}, C^{TT}, and C^{RR} are standard Toeplitz matrices (C_{ij} depends on $i - j$ only), while for $d = 3$ we would have to deal with block Toeplitz matrices. Extension of the Szegö's theorem for products of Toeplitz matrices allows us to express the trace (57) as

$$\mathcal{I} = \frac{N}{2\ln 2} \int_0^{2\pi} \frac{d\mu}{2\pi} \left[1 - \tilde{C}^{TT}(\mu) - \tilde{C}^{RR}(\mu) \right] \tilde{C}^{RT}(\mu)^2, \qquad (67)$$

where

$$\tilde{C}^{XY}(\mu) = \sum_{n=-\infty}^{\infty} C^{XY}(a|n|) e^{in\mu}. \qquad (68)$$

Using Eqs. (58), (59), and (60) for $d = 2$, we obtain

$$\mathcal{I} = \frac{N}{2\ln 2(kL)^2} \left(S^{RT} - \frac{S^{TT} - S^{RR}}{k\ell} \right), \qquad (69)$$

where S^{RT}, S^{TT}, and S^{RR} are three series defined as

$$S^{RT} = \sum_n f_1\left(\frac{|n|a}{L}\right)^2, \qquad (70)$$

$$S^{TT} = \sum_{n,p} f_1\left(\frac{|n|a}{L}\right) f_1\left(\frac{|p|a}{L}\right) f_2\left(\frac{|n+p|a}{L}\right), \quad (71)$$

$$S^{RR} = \sum_{n,p} f_1\left(\frac{|n|a}{L}\right) f_1\left(\frac{|p|a}{L}\right) f_3\left(\frac{|n+p|a}{\ell}\right). \quad (72)$$

As discussed in the MT, f_1 is well approximated by a Gaussian function, $f_1(r) = \alpha_1 e^{-\beta_1 r^2}$, so that S^{RT} reduces to a Jacobi theta function:

$$S^{RT} = \alpha_1^2 \theta\left[2\beta_1\left(\frac{a}{L}\right)^2 \right]. \qquad (73)$$

Here we use the notation $\theta(x) \equiv \theta_3(0, e^{-2x})$, where $\theta_3(u, q) = 1 + 2\sum_{n=1}^{\infty} q^{n^2} \cos(2nu)$. If we also use Gaussian models for f_2 and f_3 [$f_2(r) = \alpha_2 e^{-\beta_2 r^2}$ and $f_3(r) = \alpha_3 e^{-\beta_3 r^2}$], we find

$$S^{TT} = \alpha_1^2\alpha_2 \theta\left[(\beta_1 + \beta_2)\left(\frac{a}{L}\right)^2 \right]$$
$$\times \theta\left[\beta_1 \frac{\beta_1 + 2\beta_2}{\beta_1 + \beta_2}\left(\frac{a}{L}\right)^2 \right]. \qquad (74)$$

and

$$S^{RR} = \alpha_1^2 \alpha_3\, \theta\left[\left(\beta_3 + \beta_1 \frac{\ell^2}{L^2}\right)\left(\frac{a}{\ell}\right)^2\right]$$
$$\times\, \theta\left[\beta_1 \frac{2\beta_3 + \beta_1 \ell^2/L^2}{\beta_3 + \beta_1 \ell^2/L^2}\left(\frac{a}{L}\right)^2\right]. \quad (75)$$

Simple approximations of these functions can be obtained using $\theta(x) \simeq \sqrt{\pi/x}$ for $x \ll 1$, and $\theta(x) \simeq 1$ for $x \gg 1$. In this way, we can recover the result (63). We also find that $S^{TT} \sim (L/a)^2$ for $a \ll L$, $S^{RR} \sim L/a$ for $\ell \ll a \ll L$, and $S^{RR} \sim (\ell/a)(L/a)$ for $a \ll \ell$. Hence, $S^{RR} \ll S^{TT}$ for $a \ll L$ in the multiple scattering regime $L \gg \ell$. This is the reason why we neglected the contribution of reflection correlations to mutual information in Eq. (7) of the MT. In the opposite limit $a \gg L$, we find $S^{RT} \sim 1$, $S^{TT} \sim 1$, $S^{RR} \sim 1$, so that surface correlations become negligible and $\mathcal{I}^\infty \sim N/(kL)^2$.

Although Gaussian fits for $f_2(r)$ and $f_3(r)$ in the regime $r \gtrsim \ell$ are satisfactory, we used in the MT exponential fits $[f_2(r) = \alpha_2 e^{-\beta_2 r}$ and $f_3(r) = \alpha_3 e^{-\beta_3 r}]$ that give better agreement with correlation functions computed from wave propagation simulations. The solid line in Fig. 4 of the MT represents the result (69) with S^{RT} given by Eq. (73), and S^{TT} and S^{RR} by Eqs. (71) and (72) with exponential forms for f_2 and f_3.

[1] E. Kogan and M. Kaveh, Phys. Rev. B **52**, R3813 (1995).
[2] T. M. Nieuwenhuizen and M. C. W. van Rossum, Phys. Rev. Lett. **74**, 2674 (1995).
[3] E. Kogan, M. Kaveh, R. Baumgartner, and R. Berkovits, Phys. Rev. B **48**, 9404 (1993).
[4] M. Lax, Phys. Rev. **85**, 621 (1952).
[5] R. Moddemeijer, Signal Process. **16**, 233 (1989).
[6] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[7] I. Starshynov, A. M. Paniagua-Diaz, N. Fayard, A. Goetschy, R. Pierrat, R. Carminati, and J. Bertolotti, (2017), arXiv:1707.03622.
[8] C. W. J. Beenakker, Rev. Mod. Phys. **69**, 731 (1997).
[9] M. C. W. van Rossum and T. M. Nieuwenhuizen, Rev. Mod. Phys. **71**, 313 (1999).
[10] D. B. Rogozkin and M. Y. Cherkasov, Phys. Rev. B **51**, 12256 (1995).
[11] D. B. Rogozkin and M. Y. Cherkasov, Phys. Lett. A **214**, 292 (1996).
[12] P. A. Mello, J. Phys. A **23**, 4061 (1990).
[13] P. A. Mello, E. Akkermans, and B. Shapiro, Phys. Rev. Lett. **61**, 459 (1988).
[14] L. Froufe-Pérez, A. Garcia-Martin, G. Cwilich, and J. Saenz, Physica A **386**, 625 (2007).
[15] M. J. Stephen and G. Cwilich, Phys. Rev. Lett. **59**, 285 (1987).
[16] S. Feng, C. Kane, P. A. Lee, and A. D. Stone, Phys. Rev. Lett. **61**, 834 (1988).
[17] R. Pnini and B. Shapiro, Phys. Rev. B **39**, 6986 (1989).
[18] R. M. Gray, Found. Trends Commun. Inf. Theory **2**, 155 (2006).