Using Neural Networks to Image through Highly Scattering Media

Ben Marrett

29th April 2024

Abstract

Optical imaging through highly scattering media has historically been a challenging problem. Recent developments have shown this scattering to be both linear and deterministic, enabling the control of the propagation of light within such regimes. We utilise a neural network to appropriately characterise the scattering process and thereby image non-invasively behind a highly scattering medium (simulated using a discrete dipole model). The approach brings physical significance to our network, with its two-layer architecture emulating the propagation of light through our system.

1 Introduction

Historically, the propagation of light was only studied within simple scenarios where scattering could be exactly determined [1]. Conventional wisdom suggested that when moving to the multiple scattering regime, all spatial information of the incident light would be lost, thus limiting imaging techniques to shallow depths. In fact, elastic scattering processes are deterministic regardless of complexity, meaning information is scrambled but conserved [2]. Therefore by shaping the phase of the incident field, it is possible to measure and reverse scattering within such complex domains, opening the possibility of imaging within situations previously intractable [3].

For coherent optical light, intensities are too low to introduce impactful non-linear effects and so the scattering process can be considered linear in field [4]. Thus, the propagation of light through a given medium is characterised by its transmission matrix (TM), linking fields at an input and output plane through a system of linear equations [5]. The TM specifies the required input field to produce a given output field, giving spatial control of the field at the output

plane and thus facilitating focusing behind the scattering medium [6]. By exploiting correlations within rows of the TM, one may tilt and shift this focus to image over small areas within a scattering system [7]. The use of scattered, optical light for imaging brings great promise in fields such as medical imaging, where conventional methods using ionising X-ray radiation [8] are often undesirable, and can even be used to improve image resolution [6]. Measurement of the TM is possible if one has access to both the input and output plane [9], however such methods are unproductive when seeking to image non-invasively behind the medium, for instance if the object is located within living tissue. This presents the major challenge of TM measurement without direct access to the object plane.

Recent research has turned to neural networks to characterise scattering using the statistical characteristics of the system learned from large datasets [10] [11]. However, such methods provide little physical insight to the system, instead treating the model as a metaphorical 'black box'. To improve this, we use the framework provided by [12] to construct a two-layer neural network which emulates the physical structure of our system and hence directly reproduces its characteristic TMs as our model's parameters. We apply the network to image a series of fluorescent particles hidden behind a highly scattering medium, simulated using a discrete dipole method [13]. By using training data comprised of intensity measurements taken only at the output plane, the TMs are obtained non-invasively and used to image behind our simulated scattering medium.

2 The Transmission Matrix Formalism

2.1 Complex Media

In free space, ballistic light propagates without interruption, meaning a lens can predictably map light from an object to a detector to form an image. Within complex media, this relation is broken - light is exponentially attenuated with depth and the remaining light is scattered, scrambling its spatial information and thereby degrading the resolution of an object when imaged. In this way, imaging within complex media using ballistic light is restricted at depths L greater than the mean free path l_f [4]. As demonstrated by figure 1, the image is largely resolved at

Figure 1: Water droplets composing fog scatter light, degrading the image quality of a tree at distances $L \approx l_f$ and completely obscuring it for $L > l_f$.

$L < l_f$. However, at $L \approx l_f$, a significant proportion of light is scattered, obscuring the object's features, and at $L > l_f$, the image is reduced to a blur. Therefore, in order to image at depths beyond l_f, one must make use of this scattered light.

Despite its high complexity, scattering within complex media is in fact both linear and deterministic and therefore its propagation can be mathematically characterised [14]. This principle allows for the spatial control of light within such media, and ultimately leads to the ability to image beyond the ballistic regime.

2.2 The Transmission Matrix

For linear, complex media, the propagation of optical light may be characterised by its Green function $G(\mathbf{r}_{in}, \mathbf{r}_{out}, t_{in}, t_{out})$, linking the field at a position $\mathbf{r}_{in}$ and time t_{in} to a position $\mathbf{r}_{out}$ and time t_{out} [15]. For imaging, this may represent the propagation from an input object plane to an output camera plane. If the medium is stationary with time, the time dependence becomes $t = t_{out} - t_{in}$, meaning we can take the Fourier transform to obtain the propagation of a specific frequency ω; our Green function thus becomes $G(\mathbf{r}_{in}, \mathbf{r}_{out}, \omega)$ [4]. This project uses simulations working within a two-dimensional geometry with linearly polarised sources, which motivates the use of a scalar model. Therefore for a surface S^{in} containing all our input sources, the output

optical field of frequency ω can be given:

$$E(\mathbf{r}_{\text{out}}, \omega) = \iint_{S^{\text{in}}} G(\mathbf{r}_{\text{in}}, \mathbf{r}_{\text{out}}, \omega) E(\mathbf{r}_{\text{in}}, \omega) \, d^2\mathbf{r}_{\text{in}} \qquad (1)$$

Now consider our input plane: we may divide S^{in} into N smaller areas S_n^{in}, each of which contains the nth source. We then find the average optical field at the nth input surface as $E_n^{\text{in}} = \int_{S_n^{\text{in}}} E(\mathbf{r}) \, d\mathbf{r}$. Similarly, the output plane is divided into M smaller areas S_m^{out}, each of which contains the mth detector. The average optical field at the mth output surface is thus $E_m^{\text{out}} = \int_{S_m^{\text{out}}} E(\mathbf{r}) \, d\mathbf{r}$ [16]. In this way, we can discretize our Green functions into a matrix of complex coefficients $\mathbf{T}$, each of which linking the field from a given input source E_n^{in} to that at a given output detector E_m^{out} for a specified wavelength:

$$E_m^{\text{out}} = \sum_n^N t_{mn} E_n^{\text{in}} \qquad (2)$$

where t_{mn} is an element of $\mathbf{T}$. This is our transmission matrix and its power lies in its simplicity, linking inputs and outputs via a single complex coefficient regardless of scattering complexity.

For incoherent light, this linear system is broken since there is no constant phase relationship. At detection, incoherent processes are linear in intensity [17] and so our TM will be a real-valued, positive matrix linking input and output intensities. For instance, when imaging fluorescent objects, their response to an incident field is incoherent and so should be described by an intensity TM [12].

2.3   Controlling Propagation

By inverting the matrix equation $\mathbf{E}^{\text{out}} = \mathbf{T}\mathbf{E}^{\text{in}}$, one may calculate the required input field to produce a given output field, thereby gaining spatial control of the light at the output plane. For instance, this allows the formation of a focus at the output plane, which may be raster scanned throughout to produce an image. Without noise, the direct inverse may be taken or a pseudo-inverse for systems where $N \neq M$, such as Moore–Penrose or Tikhonov regularisation [14]. In this case, the reconstruction operator is given $\mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$, where I is the identity matrix, which
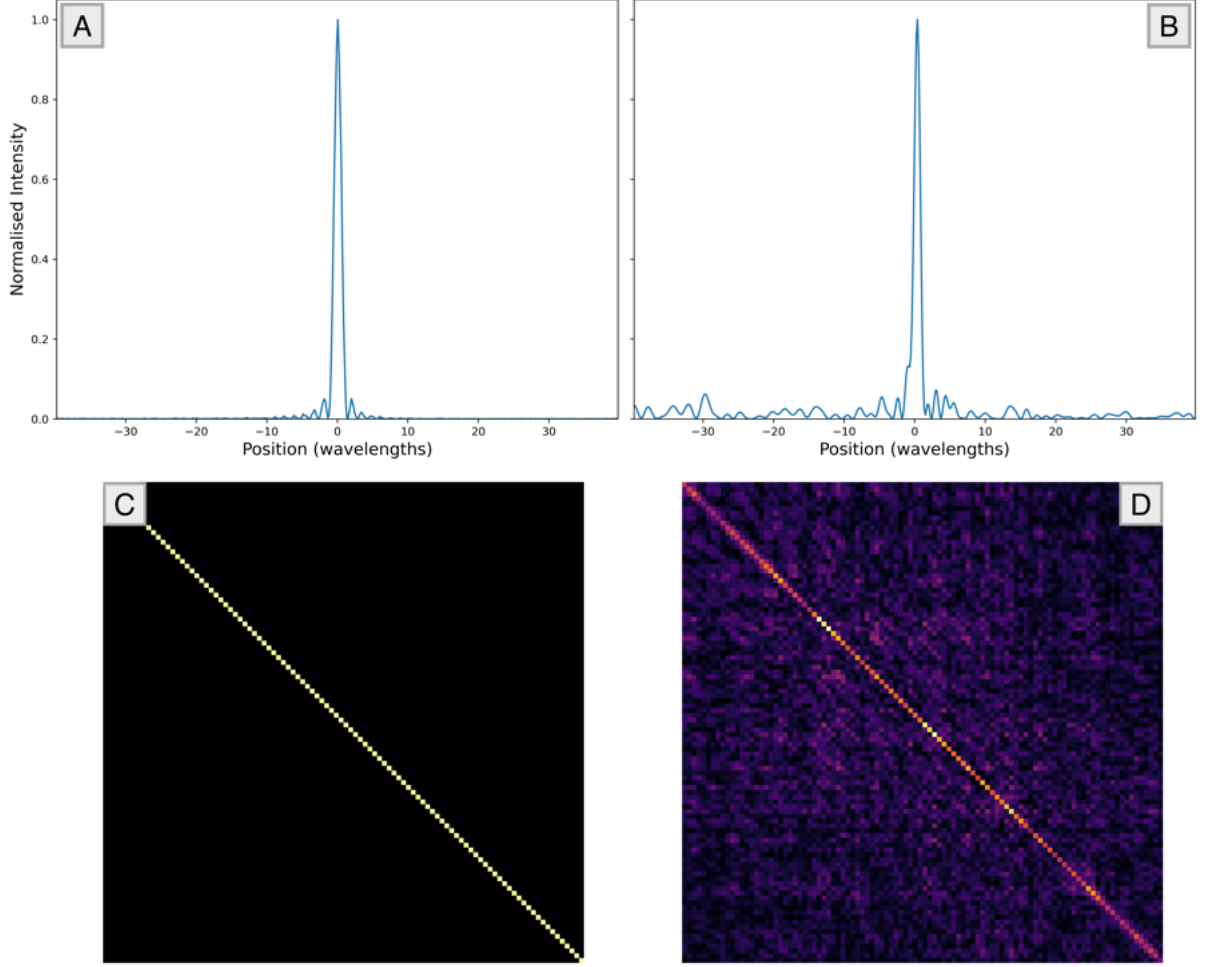
Figure 2: **Calculated intensity for a focus** constructed via (A) a direct inverse and (B) phase conjugation, with their respective **reconstruction operators** (C) $\mathbf{T}^{-1}\mathbf{T}$ and (D) $\mathbf{T}^{\dagger}\mathbf{T}$. A pronounced diagonal represents high reconstruction fidelity. The system is simulated via a two-dimensional discrete dipole method (section 3) with $N = M = 100$.

indicates that a specified field may be perfectly transmitted to the output plane (figure 2C). For focusing, this represents a tight focus at the specified detector and a high signal to noise ratio (SNR) (figure 2A). Focus contrast may be improved by increasing the number of input sources, although this is limited by the wavelength when source spacing reaches $\lambda/2$.

A direct inversion is highly noise sensitive and so other techniques are often used for practical purposes. Phase conjugation employs the conjugate transpose (denoted by $\dagger$) to improve operator stability at the cost of reconstruction fidelity; $\mathbf{T}^{\dagger}\mathbf{T}$ shows a strong diagonal but with noticeable fluctuations elsewhere (figure 2D). The approach produces a diffraction limited focus with a lower SNR but a higher resilience to experimental noise and perturbations (figure 2B).

5

Consequently, a focus may be tilted and shifted over small distances, while one from direct inversion breaks down entirely. One may also notice the use of the conjugate transpose implies that each TM row's conjugate represents the input field to focus on a given output point - this becomes conceptually significant when understanding memory effects (section 5).

3 Modelling the System

3.1 Coupled Equations

Our scattering medium is simulated using a discrete dipole (DD) method [18], built from first principles within python. A scattering medium of thickness L and width βL is approximated as a series identical, non-absorbing point dipoles (figure 3). To achieve a large optical thickness within computational restrictions, we use the two-dimensional (2D) scalar approximation where the electric field and dipole polarizations are orientated directly perpendicular to the plane. Each dipole interacts with both the incident field E_0 and all other dipoles via the free space Green function G_0, and hence for coherent light of frequency ω, the field at the jth dipole, is found as the following linear sum:

$$E_j(\omega) = E_0(\mathbf{r}_j, \omega) + k_0{}^2\alpha(\omega) \sum_{\substack{k=1 \\ k \neq j}}^{K} G_0(\mathbf{r}_j - \mathbf{r}_k, \omega)E_k(\omega) \tag{3}$$

where we sum over the incident field $E_0(\mathbf{r}_j, \omega)$ and the field at all other dipoles $E_k(\omega)$, radiated from $\mathbf{r}_j$ to $\mathbf{r}_k$ via our Green function $G_0(\mathbf{r}_j - \mathbf{r}_k, \omega)$. We denote k_0 as the wavevector in a vacuum and $\alpha(\omega)$ as the dipole polarizability. For our 2D problem, the scalar free-space Green function, connecting the field from a point source $\mathbf{r}_0$ to its radiated field at a point $\mathbf{r}$, is given by:

$$G_0(\mathbf{r} - \mathbf{r}_0, \omega) = \frac{i}{4}H_0^{(1)}(k_0|\mathbf{r} - \mathbf{r}_0|) \tag{4}$$

where $H_0^{(1)}$ is the Hankel function of the first order and zeroth kind [19]. Eq. 3 forms a linear system of equations which may be solved by matrix inversion to obtain the field at all K scattering dipoles. Once known, one may similarly sum over all dipoles to obtain the field at an arbitrary point $\mathbf{r}$:
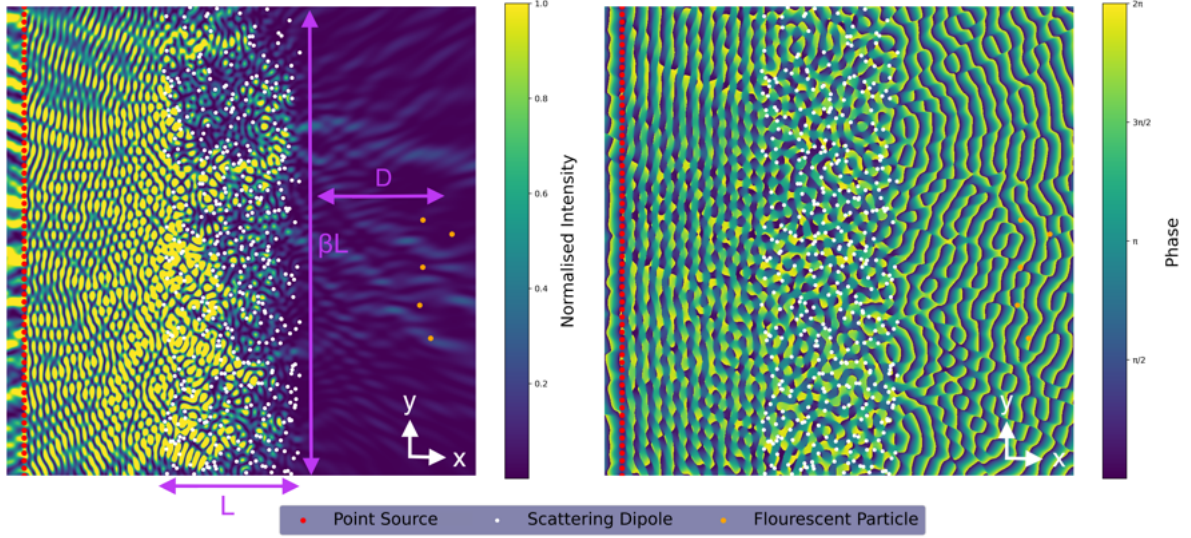
Figure 3: **Intensity and phase for a plane wave input** within a scattering medium of dimension $L \times \beta L$ consisting of point-like, non-absorbing scattering dipoles. Several fluorescent particles are placed behind the medium to be imaged. Intensities are capped to resolve the scattered field.

$$E(\mathbf{r}, \omega) = E_0(\mathbf{r}, \omega) + k_0{}^2 \alpha(\omega) \sum_{k=1}^{K} G_0(\mathbf{r} - \mathbf{r}_k, \omega) E_k(\omega) \tag{5}$$

Figure 3 demonstrates how this technique may be employed to simulate a highly scattering medium. The plotted phase shows the flat wavefront of an incident plane wave distorted at the medium's exit, scrambling spatial information.

3.2 Energy Conservation

Under elastic scattering conditions, no light is absorbed by the medium since all energy is elastically re-scattered. The polarizability of a given dipole for low intensities is given by [20]:

$$\alpha(\omega) = -\alpha_0 \frac{\Gamma/2}{\omega - \omega_0 + i\Gamma/2} \tag{6}$$

where Γ is the spectral linewidth, ω_0 the resonant frequency and $\alpha_0 = 4\pi/k_0{}^2$ for the 2D scalar approximation. From this, we may compute their scattering (σ_s) and extinction (σ_e) cross-sections:

$$\sigma_s = \frac{k_0{}^3}{4} |\alpha(\omega)|^2 \quad : \quad \sigma_e = k_0 \mathrm{Im}[\alpha(\omega)] \tag{7}$$
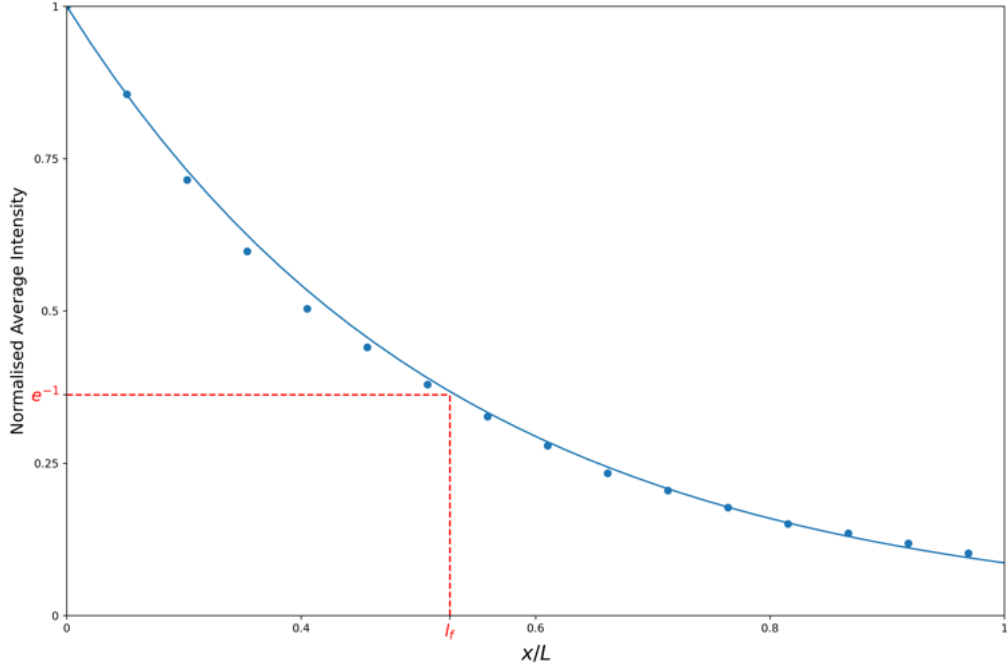
Figure 4: **Exponential decay in intensity with depth** x into a medium of thickness L. Intensities are calculated along a 1D slice for 10,000 random scatterer configurations, then averaged and finally normalised by the incident intensity.

Their ratio gives the fraction of energy re-scattered by the dipole, hence for elastic conditions $\sigma_s = \sigma_e$. Therefore, energy conservation limits the polarizability to:

$$\alpha_{\max} = \frac{4i}{k_0{}^2} \tag{8}$$

which occurs at resonance [21], found by using $\omega = \omega_0$ within Eq. 6. For simplicity and to save computation time, we use the maximum polarizability $\alpha_{\max}$ which conserves energy for all scattering dipoles.

While our scatterers are non-absorbing, intensity is still exponentially attenuated with depth x through our medium [19] (figure 4). Dipole emission is spherically symmetric, meaning a given scatterer radiates energy further through the medium and back to the input in equal parts. Therefore, scatterers at greater depth experience an exciting field of lower amplitude than the preceding and the intensity is gradually reduced. The analytical exponential relation in figure 4 is not exactly produced by simulation, likely because the simulated medium is of finite width
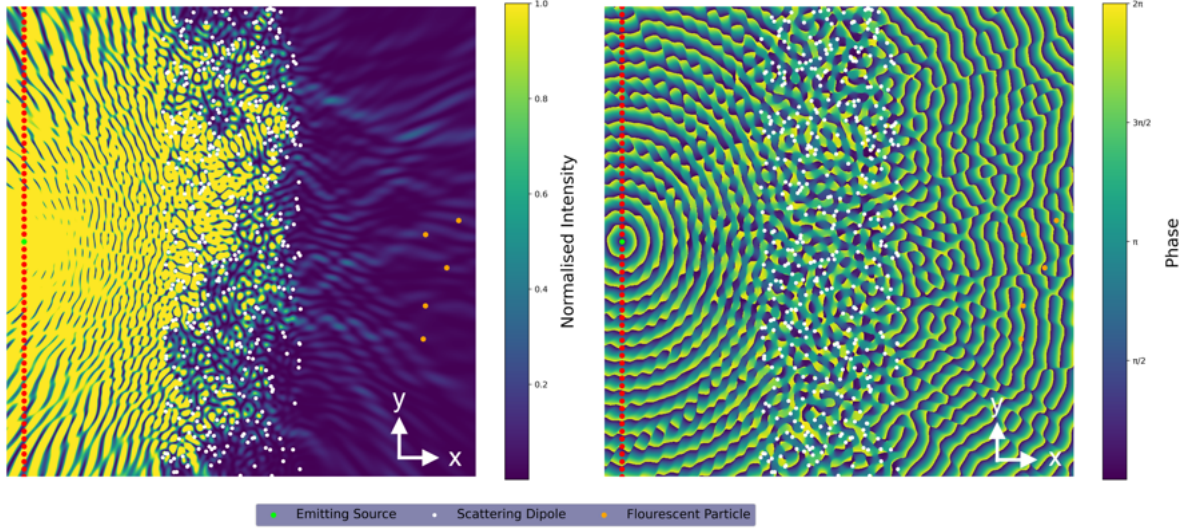
8

Figure 5: **Intensity and phase for a single source input**. The calculated field at the fluorescent particles composes one column of the forward propagation TM. Intensities are capped to resolve the scattered field.

and therefore energy spills and diffracts at edges (suggested by testing varying β). Figure 3 also shows a reduced intensity at the output, along with a distorted plane wave phase at the input from the back-scattered light.

3.3 Calculating the Transmission Matrix

To gain spatial control of the output field, it is necessary to implement a form of computational wavefront shaping, as described by section 2.3. The following method aims to emulate experimental wavefront shaping techniques, such as use of spatial light modulators, which control either the input amplitudes or phases for an array of discrete pixels [22]. To achieve this, the simulated input comprises an array of N non-interacting dipoles (figure 5), each of which may be excited only by its specified input field $\mathbf{E}^{\text{in}}$. Each excited dipole radiates this energy identically to the scatterers by way of the free-space Green function (Eq. 4), and so the incident field E_0 at a point $\mathbf{r}$ may be found by summing over all input contributions:

$$E_0(\mathbf{r}, \omega) = k_0{}^2 \alpha_{\text{max}} \sum_{n=1}^{N} G_0(\mathbf{r} - \mathbf{r}_n, \omega) E_n^{\text{in}}(\omega) \tag{9}$$

With control over individual input components, one may obtain the transmission matrix
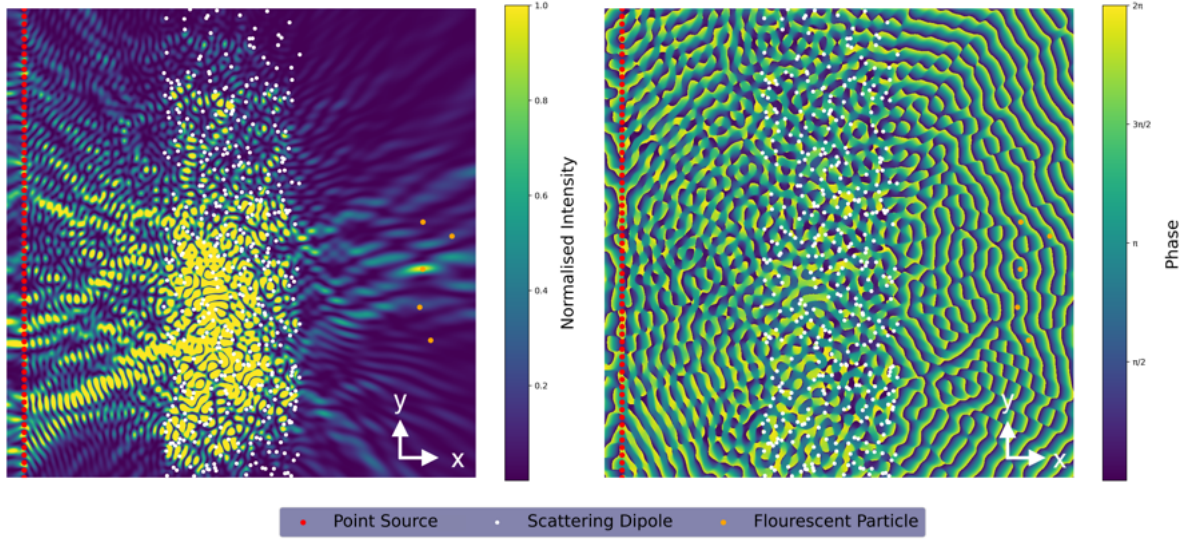
Figure 6: **Intensity and phase for a shaped input field**, focusing on a fluorescent particle located behind the scattering medium. Intensities are capped at the peak focus intensity in order to resolve the focus within this image.

characterising the forward propagation from the input sources to a series of specified output points within the simulation. For the purposes of imaging, this will be an array of fluorescent particles located behind the scattering medium. To produce a given TM column, a single input dipole is excited by an input field of unit amplitude and the output field is calculated at the location of all fluorescent particles (figure 5). Repeating for all input sources produces the full forward propagation TM, denoted $\mathbf{T}_1$. Phase conjugation of the calculated $\mathbf{T}_1$ provides control over the field at the output plane within our simulation, as described within section 2.3. For instance, to focus on the mth fluorescent target, E_m^{out} is set to unity while all target outputs are null. Using the phase conjugated $\mathbf{T}_1$, one may then obtain $\mathbf{E}^{\text{in}}$, which forms the exciting input field (figure 6).

The incoherent, intensity response from a given fluorescent target is characterised by a real-valued positive TM, denoted $\mathbf{T}_2$, linking the excitation intensities at each target to an output measurement plane (section 2.2). For our system, intensity measurements are taken at the location of the input dipoles. $\mathbf{T}_2$ is calculated similarly to $\mathbf{T}_1$, using coherent emitters at the location of the fluorescent particles; taking the absolute square for each element then converts to an intensity TM. In this way we obtain two TMs, $\mathbf{T}_1$ and $\mathbf{T}_2$, characterising the forward and

backward propagation through the system respectively.

4 Neural Networks

4.1 Machine Learning

The meteoric rise of machine learning in recent years has seen a huge surge in its application to a multitude of areas, from large language models [23] to predicting the 3D structure of proteins [24]. Its power stems from its flexibility; rather than explicitly defining a procedure to complete a certain task, machine learning models use a flexible structure of tunable parameters to make a prediction for a given input. Using large datasets which prescribe how the model should behave, parameters are gradually adjusted towards the configuration which best predicts the output for the widest range of inputs. Fundamentally, machine learning is an optimization problem and indeed the approach prescribed within section 4.3 treats it as such.

Perhaps the simplest form of machine learning would be linear regression, where two tunable parameters, the gradient and y-intercept, are optimized to best fit the data. In practice, machine learning models for more complex tasks comprise of millions of parameters, and so it is desirable to structure these in a manner which both lends itself to optimization and minimises model over-fitting. Neural networks are a particular structure with a particular optimisation mechanism which has been shown to scale well [25]. The model's parameters are structured in layers and hence an input is transformed by a series of matrix multiplications to produce an output (figure 7). More complex networks build from this base structure to achieve specific tasks, but the most simple structure described above, referred to as a multi-layer perceptron [26], is desirable to physically characterise light within our scattering system.

4.2 The Mathematics of Learning

The following mathematics is adapted from [27]. A standard neural network consists of layers of neurons, where a_j^l denotes the jth neuron in the lth layer (figure 7). Each neuron is linked to all those in the previous layer by a set of weights, therefore we can say the value of the jth neuron in the lth layer is determined by the weighted sum $a_j^l = \sum_k w_{jk}^l a_k^{l-1}$, where we sum over all
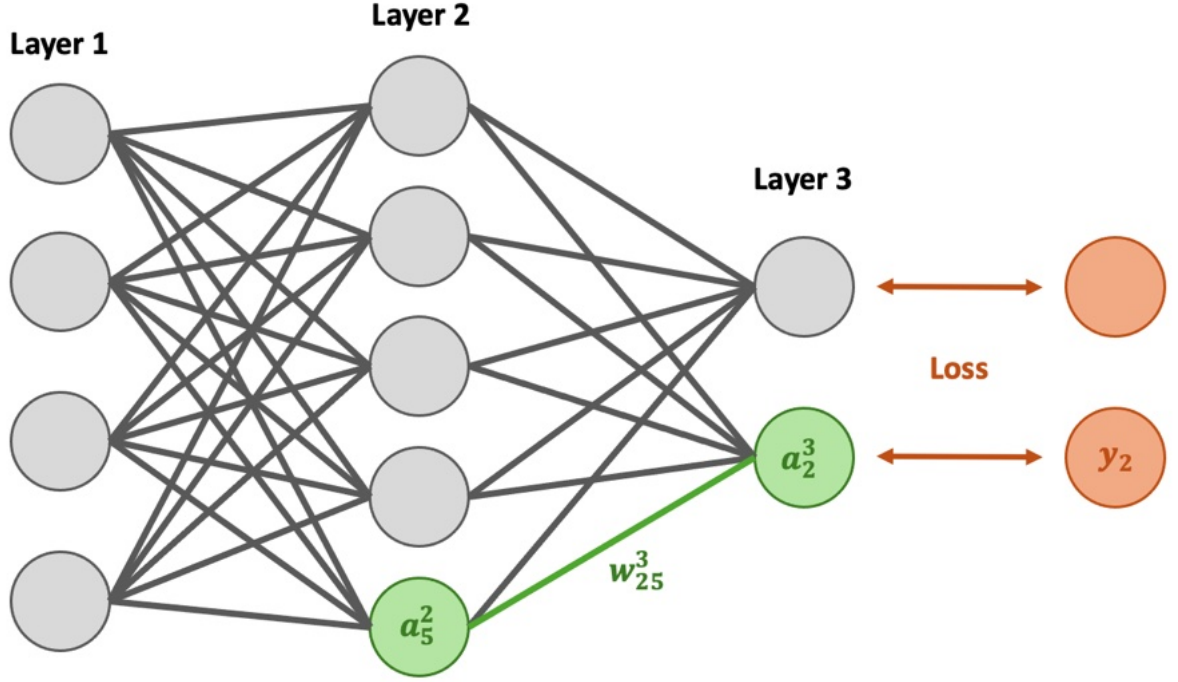
Figure 7: **Basic 3 layer neural network** with neurons linked together through a series of weights and biases. Each set of weights represents a matrix multiplication between layers. The loss is computed at the final layer, assessing the network's accuracy. A notation example is highlighted.

neurons in the previous layer a_k^{l-1}. Most networks also incorporate a bias for inactivity b and an activation function σ, which conceptually ensure a neuron only has a meaningful output when its value reaches a certain threshold, but in the case of this project, they are used to introduce physical significance to the network (section 4.3). Our weighted sum therefore becomes:

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \tag{10}$$

If we now group each layer of neurons into vectors, our weighted sum between individual neurons becomes a matrix product between layers:

$$\mathbf{a}^l = \sigma\left(\mathbf{w}^l \mathbf{a}^{l-1} + \mathbf{b}^l\right) \tag{11}$$

In this way, an input vector is transformed to the output through a sequence of matrix multiplications. More generally, an input could be a tensor of any dimension, but such networks usually employ alternative techniques to intelligently segment and analyse the data, such as

feature extracting kernels within convolutional neural networks [28].

To effectively train a network, one needs some metric to assess its current performance. This is done via a loss function C, which compares the predicted output to the ground truth for a given input. While complex loss functions are used for specified tasks, such as cross-entropy loss for classification [29], the simplest and most widely applicable is the mean squared error (MSE):

$$C = \frac{1}{N} \sum_{\mathbf{x}}^{N} |\mathbf{y}(\mathbf{x}) - \mathbf{a}^L(\mathbf{x})|^2 \tag{12}$$

where for a given input $\mathbf{x}$, the predicted output $\mathbf{a}^L(\mathbf{x})$ is compared to the ground truth $\mathbf{y}(\mathbf{x})$ (L denotes the final layer). Averaging the loss over a set of N training inputs computes the networks accuracy for a wide range of scenarios; indeed, a trained network's predictive power is constrained to scenarios within the bounds of its training data.

In this way, training a network becomes an optimisation problem, with the aim to determine the parameter values which minimise the network's loss. For neural networks, the minimisation technique is termed gradient descent. Conceptually, optimisation can be visualised as a walk down the metaphorical valley of the loss function, and so a step of gradient descent must determine in which direction is the minimum. Compiling the network's current parameters into the vector $\mathbf{W}_{\text{old}}$, one may calculate the current loss as $C(\mathbf{W}_{\text{old}})$ and take its derivative with respect to all parameters to find the direction of the minimum. The current parameters are thus adjusted by following the direction of the calculated gradient:

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \eta \nabla C(\mathbf{W}_{\text{old}}) \tag{13}$$

where η is the learning rate, giving the size of the optimisation step. In practice, this is reduced throughout training as the loss converges on the minimum. Some networks also make use of momentum [30], incorporating previously calculated gradients into the current calculation:

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} - \eta \nabla C(\mathbf{W}_{\text{old}}) - \gamma \Delta \mathbf{W}_{\text{previous}} \tag{14}$$

where γ is the momentum and $\Delta \mathbf{W}_{\text{previous}}$ gives the update to the weights for the previous

step of gradient descent. The use of momentum can help escape local minima in the loss when training, leading to a faster and smoother convergence to the global minimum.

However, a neural network is structured in layers meaning the value of the later neurons is determined by the preceding. Therefore, adjusting the weights in the early layers must account for how the later weights have changed; mathematically, the partial gradient of the loss with respect to parameters at a given layer uses the chain rule containing all partial gradients at the following layers. To do this, the gradient is taken layer by layer, beginning at the last layer and moving backwards. This process is named backpropagation and it provides high computational efficiency by reusing partial gradient calculations for the earlier layers. To further speed computation, the gradient is often estimated using a batch of randomly selected training data in a process known as stochastic gradient descent (SGD). By taking smaller steps of optimisation using a subset of data, the network converges quicker than by taking fewer steps, with each computed using the full training data. Within this project, SGD is treated as a powerful optimizer to statistically obtain the transmission matrices using a large dataset.

4.3 Our Structure

Most machine learning approaches to non-invasive imaging utilise complex neural network structures from which it is difficult to extract any physical meaning. The approach described by [12] aims to improve upon this by constructing a two-layer network architecture directly emulating the physical structure of the system, reproducing the forward and backward propagating $\mathbf{T}_1$ and $\mathbf{T}_2$ as the weights of the model (figure 8A). The network takes the incident field $\mathbf{E}^{\text{in}}$ as its input and predicts the corresponding intensity response from a series of fluorescent targets as the measured intensity $\mathbf{I}^{\text{out}}$ at the output plane.

The weights of the first layer represent the forward propagating $\mathbf{T}_1$, a complex-valued matrix characterising the propagation of the incident field $\mathbf{E}^{\text{in}}$ through the scattering medium to the positions of the fluorescent objects, where the field can be given as $\mathbf{E}^{\text{ex}} = \mathbf{T}_1\mathbf{E}^{\text{in}}$ (figure 8B). The number of hidden layer neurons thus represents the number of fluorescent targets. Since the target's fluorescent response is proportional to the excitation intensity, the activation function at the hidden layer σ_1 will convert to an intensity $\mathbf{I}^{\text{ex}} = \sigma_1(\mathbf{E}^{\text{ex}}) = |\mathbf{E}^{\text{ex}}|^2$; additionally, no bias for
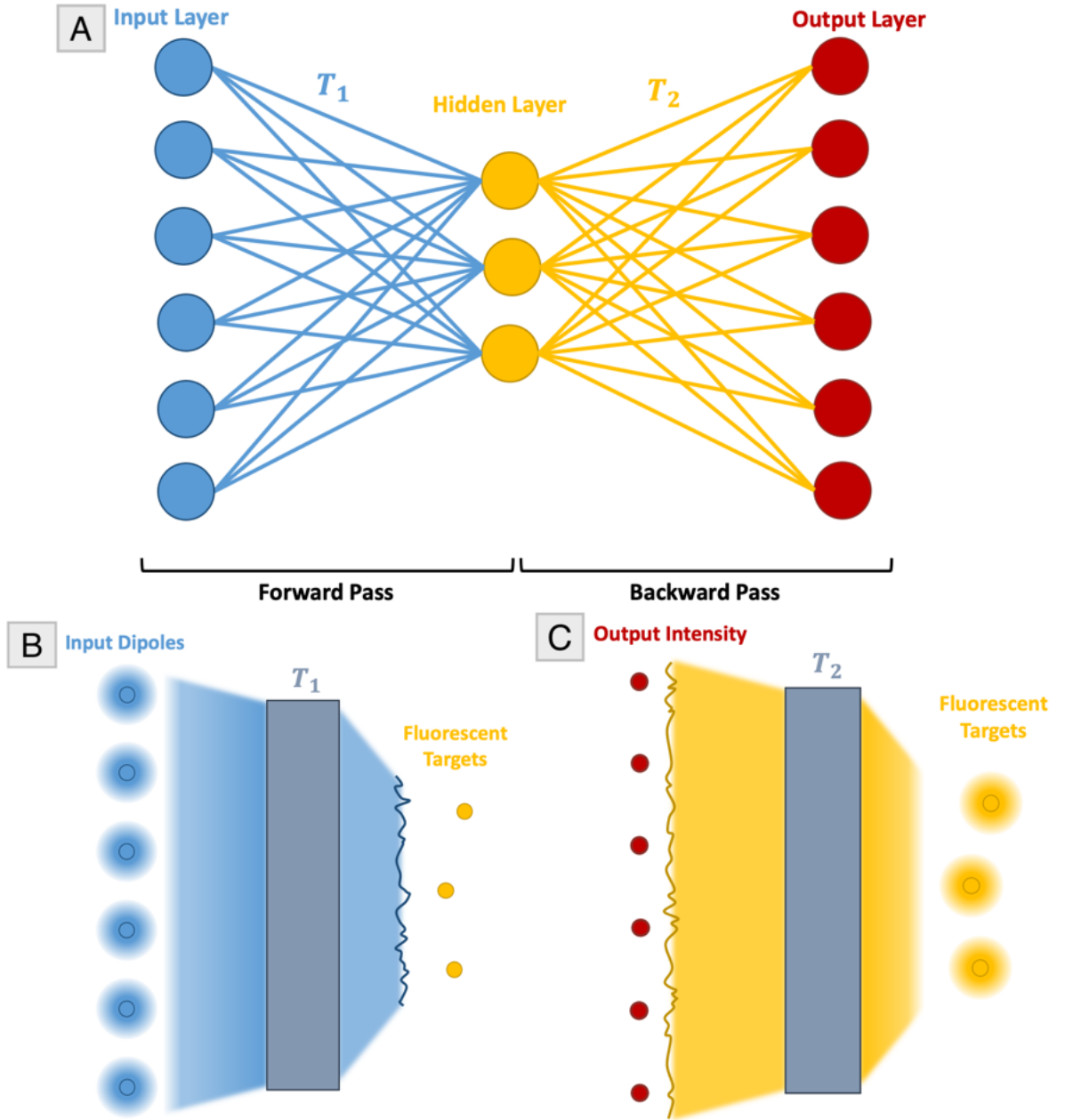
Figure 8: (A) **Physics inspired neural network** which predicts the intensity response from a series of fluorescent targets due to a given input field. Its weights represent the forward and backward propagation transmission matrices, $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively. (B) **Forward pass** through scattering medium by an input field, exciting the fluorescent targets. (C) **Backward pass** of excited target's intensity response through the medium, measured at the output plane.

inactivity is used (ie: $b^1 = 0$). The network's second layer of weights represent the backward propagating $\mathbf{T}_2$, a real valued matrix characterising the incoherent transmission of $\mathbf{I}^{\text{ex}}$ through the medium and measured as $\mathbf{I}^{\text{out}} = \mathbf{T}_2\mathbf{I}^{\text{ex}}$ at the output plane (figure 8C). A linear activation function $\sigma_2(\mathbf{I}^{\text{out}}) = \mathbf{I}^{\text{out}}$ along with no bias for inactivity at the final layer, ensure the network conforms to the physical system. The full transformation for an input $\mathbf{E}^{\text{in}}$ is thus:

$$\mathbf{I}^{\text{out}} = \mathbf{T}_2|\mathbf{T}_1\mathbf{E}^{\text{in}}|^2 \tag{15}$$

To train the network, the ground truth TMs were obtained directly from the DD simulation, as described within section 3.3, and used within Eq. 15 to produce the measured intensities from a set of randomly generated input fields - this comprised the training data. The network's weights were randomly initialised and then updated through a sequence of SGD steps so as to minimise the MSE loss (Eq. 12). Steps of gradient descent were taken using randomly generated mini-batches of training data until the full set had been used once - this represented one epoch of training. For each optimisation step, the correlation was taken between the ground truths, $\mathbf{T}_1$ and $\mathbf{T}_2$, and the approximate matrices given by the network's weights, denoted $\tilde{\mathbf{T}}_1$ and $\tilde{\mathbf{T}}_2$. Training was halted either when both calculated correlations converged to unity or a maximum number of epochs was reached.

5   Constructing an Image

5.1   Memory Effect

Despite the complexity of scattering within complex domains, a scattered field will retain some degree of correlation. For the purpose of imaging, a powerful example is the optical memory effect (ME) [31]. A small tilt or shift in the incident field produces a shifted output intensity pattern, lower in intensity but maintaining its structure. For instance, figure 9 shows a focus produced by a shaped input field; by tilting this field, the focus is translated over small distances. For a given scattering medium of thickness L and distance D from its exit to the output plane, an intensity pattern maintains its structure for translations within its characteristic ME range:
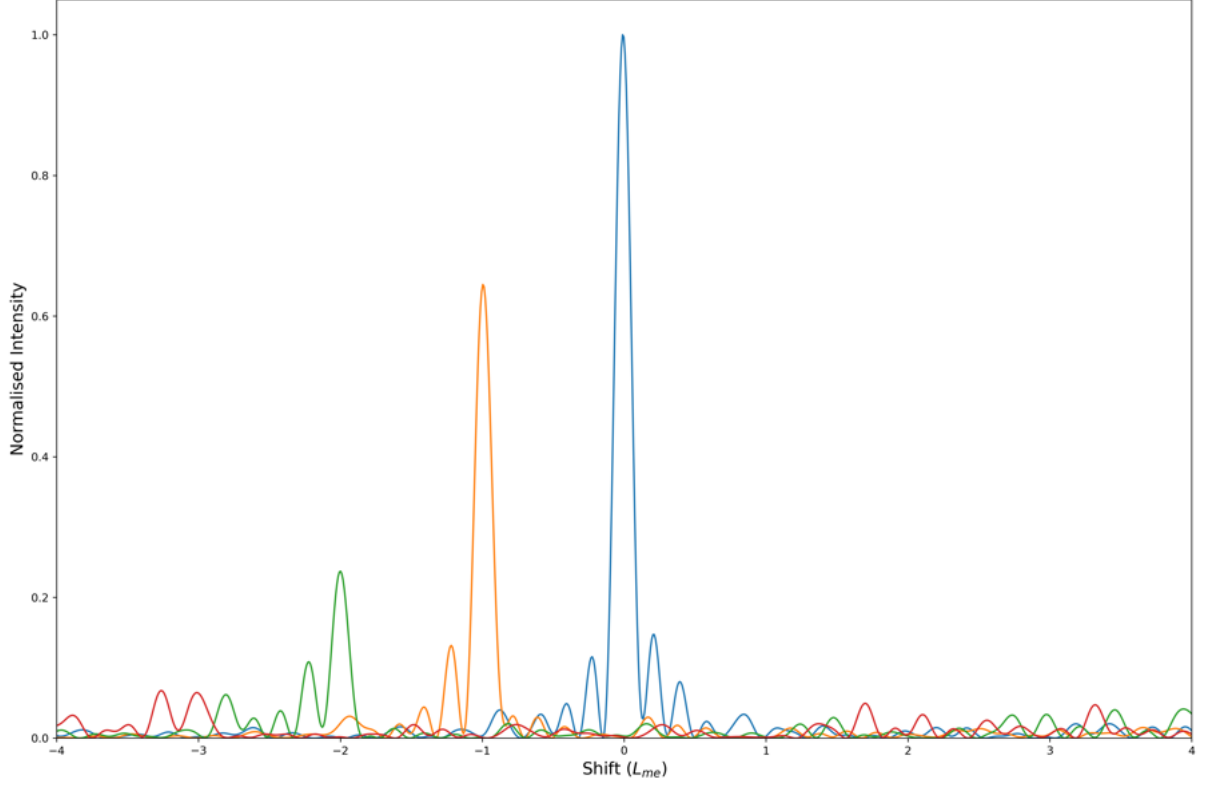
Figure 9: **Series of translated focuses** produced by tilting the incident field. Those shifted within the constraints of the memory effect (shift $\leq L_{me}$) maintain spatial structure, whilst those exceeding this loose spatial structure and eventually collapse into a random speckle pattern. Simulated using discrete dipole method.

$$L_{me} = \frac{D}{kL} \tag{16}$$

where k is the wavevector [16]. The size of a focus is limited to the speckle size:

$$L_{sp} = \frac{2\pi D}{kW} \tag{17}$$

where W is the width of the illumination area. Knowledge of such quantities indicates system geometries within which the memory effect may be employed for imaging (note D is defined as the average distance for volumetric imaging). If $W \gg L$, then $L_{me} \gg L_{sp}$ and so a focus may be scanned a distance greater than its width. Additionally, a higher D increases both quantities such that their ratio remains constant, meaning that a larger object may be imaged but to a lower resolution.

Importantly for non-invasive imaging, memory effects bring about correlations between el-

17

ements of the TM. As suggested within section 2.3, the conjugate for each row of the forward propagating TM represents a shaped input field to focus on a given target. Therefore, rows corresponding to targets separated by $< L_{me}$ will be correlated since they approximately produce two translated foci, with the strength of the correlation inversely related to the target separation.

5.2 Tilting and Translating a Focus

ME assisted imaging requires the ability to translate a focus through space; when using fixed input sources, this can be achieved through use of the appropriate phase masks. For a 2D system, two translations are possible: a tilt perpendicular to the propagation direction and a parallel shift (directions y and x respectively as defined in figure 3). The perpendicular tilt is analogous to physically tilting the input plane, which for fixed inputs corresponds to a linear phase ramp $e^{i\delta\phi_t(a)}$ [32], where $\delta\phi_t(a)$ gives the required phase delay for a source at position a on the input plane to tilt an angle θ_t. Consulting figure 10A, a tilted wavefront of angle θ_t constructs a right angle triangle with the input plane, with the side x corresponds to a phase delay of $\delta\phi_t = kx$, where k is the wavevector. Therefore, for an input source located at a distance a from the array edge, the corresponding phase delay is given by:

$$\delta\phi_t(a) = ka \sin \theta_t \tag{18}$$

Simple trigonometry gives the corresponding horizontal shift $s = D \tan \theta_t$ at the output plane, located a distance D from the medium exit.

To derive the phase mask for a parallel shift, one must consider the phase of individual rays at the point of focus. A focus can be thought of as a point of constructive interference, where rays from all input sources meet with an identical phase. For a ray emitted directly opposite the focus ($a = 0$), its phase at the focus will be given by $\phi_0 = f$, where f is the focal length. For $a \neq 0$, ray's path forms a right angle triangle with the focal distance f (figure 10B), and hence its length may be calculated as $\sqrt{f^2 + a^2}$, corresponding to a phase of $\phi_0 + \delta\phi = k \sqrt{f^2 + a^2}$. To shift a focus a distance d, the appropriate phase mask for a source at a is found by taking the difference in phases for focal lengths f and $f + d$:
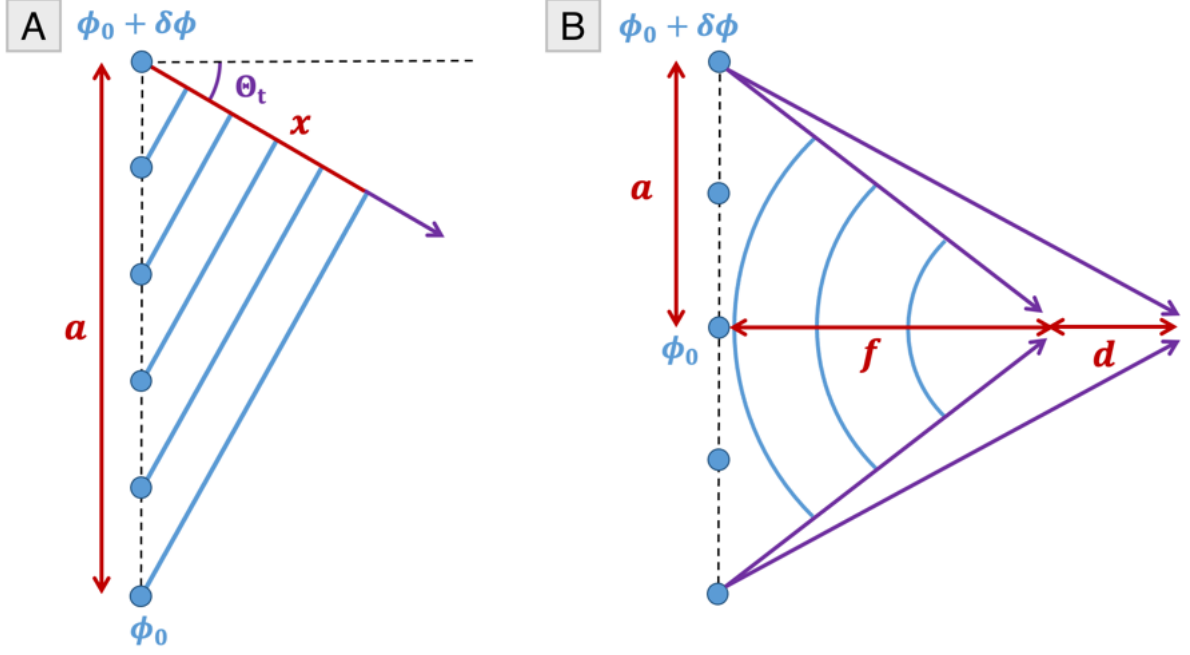
Figure 10: (A) **Wavefronts for a tilted input field** where a linear phase ramp $\delta\phi$ has been applied at input dipoles, at a position a on the input plane, to tilt a plane wave by an angle θ_t. (B) **Wavefronts for a shifted input field**. Taking the difference between input phases with focal lengths f and $f + d$ produces the appropriate phase mask to shift the focus a distance d.

$$\delta\phi_s(a) = k\left(\sqrt{a^2 + (f + d)^2} - \sqrt{a^2 + f^2} \right) \tag{19}$$

Using $e^{i\delta\phi_t(a)}$ thus produces the appropriate shift phase mask for a source a from the input plane centre.

By applying the appropriate shifts and tilts, one may translate a focus through 2D space which can be used to obtain inter-target distances. Since $\mathbf{T}_1$ rows facilitate a focus on a given target (section 2.3), one may apply a series of translations to each row to emulate a focus raster scan. For each translation, a correlation may be taken between the shifted row and all other rows; for targets within the range of the memory effect, this correlation will peak at the translation corresponding to their inter-target distance. In this way, one may construct two matrices $\mathbf{D}_x$ and $\mathbf{D}_y$ containing the inter-target distances along x and y respectively, and the corresponding correlation matrix $\mathbf{C}$ containing the calculated correlation at each translation. These matrices may then be used to construct a map of target position, (section 5.3). While other methods obtain distances through the cross correlation of the speckle pattern (columns of
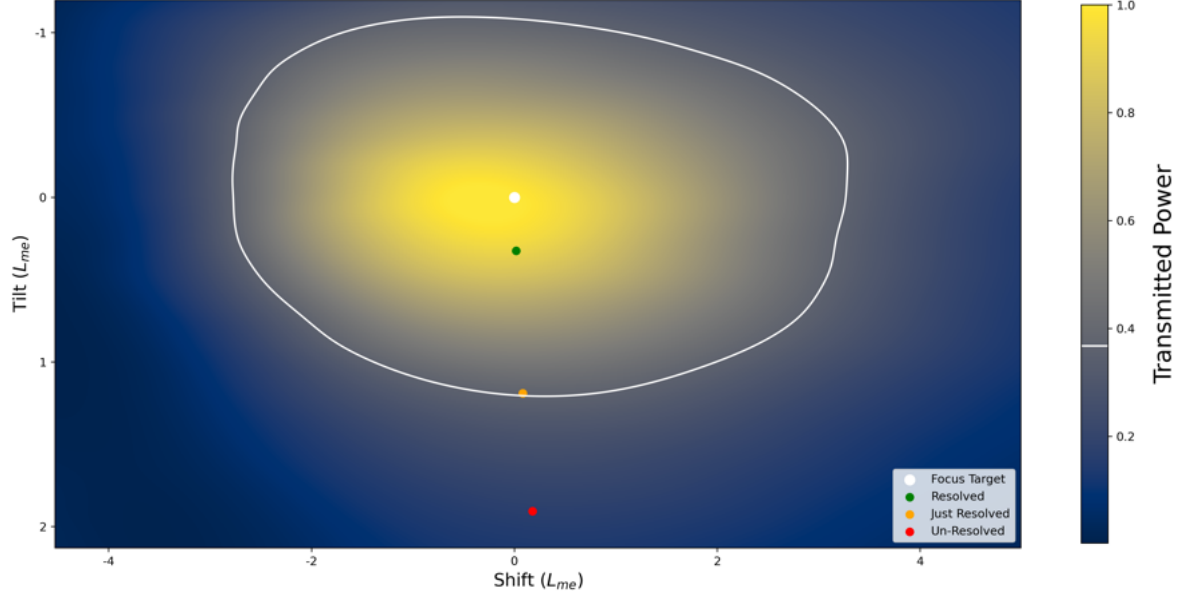
Figure 11: **The isoplanatic patch for a given target**, calculated within a discrete dipole simulation as the transmitted power for a focus translated around the target. The distance to targets located within the $1/e$ contour may be easily calculated, whereas those outside are not resolved. The extent of the patch is greater for shifts than for tilts.

$\mathbf{T}_2$) [33], most implementations do not give access to volumetric information and those that do require specific experimental setups [34].

Shift and tilt memory effects may be utilised for targets located within each other's isoplanatic patch [32]. The size of the isoplanatic patch is given by the area over which a focus for a given target may be translated before it becomes decorrelated. Within figure 11, this is represented using the transmitted power, which will be defined as the ratio between the peak translated focus intensity and the peak focus intensity at the original target. For targets within the isoplanatic patch, the translated focus retains its spatial structure, meaning TM rows will be strongly correlated and a distance may be calculated. Targets located outside the isoplanatic patch will still produce a distance measurement but with a low correlation, hence such distances are given little weight during image construction. Notice the shape of the isoplanatic patch (figure 11) suggests the shift memory effect is more powerful than the tilt by several times, meaning much higher inter-target distances may be reliably calculated along x than along y. However, consulting figure 6 indicates the focus is stretched along y, suggesting while objects may be imaged over a greater extent, this is at the cost of a lower resolution (ie: a bigger L_{sp}).

5.3 Multidimensional Scaling

To construct an image using the distance matrices $\mathbf{D}_x$ and $\mathbf{D}_y$, one must convert the set of distances into a 2D position map. One option is to construct the image patch by patch, although this approach only uses the distance measurements for the nearest neighbour, corresponding to the strongest correlation. The technique described by [35] employs multidimensional scaling (MDS) to construct a distance map using the full distance matrix through the minimisation of a stress function S. For a given coordinate, the stress function quantifies how well the current configuration $\mathbf{x}$ preserves the inter-target distances from $\mathbf{D}$:

$$S(\mathbf{x}) = \left(\sum_{\substack{i=1 \\ i \neq j}}^{N_{\text{target}}} C_{ij}{}^{\alpha} \left(D_{ij} - |x_i - x_j| \right)^2 \right)^{1/2} \tag{20}$$

where C_{ij} and D_{ij} are elements of the full distance and correlation matrices respectively, and α is a constant used to give greater weight to distances with higher correlation ($\alpha = 6$ is used here). The optimisation is run twice, along x and y, to produce a 2D target configuration that may be plotted as an image.

6 Results and Discussion

6.1 Obtaining the Transmission Matrices

We task our network to estimate the forward and backward propagating TMs for a simulated scattering medium with input sources $N_{\text{in}} = 100$, fluorescent targets $N_{\text{target}} = 5$ and output intensity detectors $N_{\text{out}} = 100$, corresponding to the input, hidden and output layers of our physics based network (figure 8). The system is of thickness $L = 5\lambda$ and medium to target propagation distance $D = 80\lambda$, where all distances are given in terms of the wavelength λ. Ground truth TMs $\mathbf{T}_1$, $\mathbf{T}_2$ are generated directly from the DD simulation and the estimated TMs $\tilde{\mathbf{T}}_1$, $\tilde{\mathbf{T}}_2$ are randomly initialised following a standard normal distribution. The network is optimised using training data of size $N_{\text{train}} = 80000$ over 10 epochs of SGD using a momentum value $\gamma = 0.9$.

Figure 12 shows a successful minimisation of the network's loss and the convergence of cor-

relations between the estimated and ground truth TMs to near unity. Figure 12A shows a rapid initial convergence for both transmission matrices, responsible for the sharp initial decrease in the loss, although while $\tilde{\mathbf{T}}_2$ continues to unity, improvements in $\tilde{\mathbf{T}}_1$ are far slower after the first epoch. Since $\mathbf{T}_1$ is used for imaging, this slow convergence restricts the method's use and motivates a further investigation into imaging using fluorescent patterns from the faster obtained $\mathbf{T}_2$. Interestingly, it was found that correlations in the TM rows (due to the memory effect), led to faster convergence than the randomly generated TMs explored by [12]. To further accelerate convergence, a key area to be considered should be the initialisation of the network's weights, preferably using physical priors. Focus should be given to the initialisation of $\mathbf{T}_1$, which begins optimisation at a far lower correlation (figure 12B).

By making use of PyTorch's complex functionality, our network gives a performance gain of several times when compared to previous work [12], converging in tens of seconds with training times increasing with network dimensionality (for instance higher target numbers). When considering real world applications, such as biological imaging, these training times are agreeable with the stability times for ex-vivo biological tissues [36], although the necessity for large training datasets requiring lengthy data collection may affect this. Our network also provides greater flexibility in the training procedure, allowing for techniques such as batch training to further improve speed. Use of PyTorch's native linear scheduler also provided a noticeable gain ($\approx 50\%$) through a linear reduction is the learning rate η throughout training.

While this physics-based network brings benefits to the physical interpretation of the network's parameters, their scale gives rise to difficulties in the choice of hyperparameters, most importantly the learning rate η. Since $\mathbf{T}_2$ is an intensity TM, its elements are the square of $\mathbf{T}_1$ in magnitude and therefore a constant learning rate will only be appropriate for one layer, meaning network optimisation fails. Consider again the optimisation analogy of walking down a valley: the optimisation valley for $\mathbf{T}_1$ may be far larger than $\mathbf{T}_2$, and so an appropriate learning rate for $\mathbf{T}_1$ may miss the $\mathbf{T}_2$ minimum entirely, stepping over the metaphorical valley. To tackle this, different learning rates were used for each layer, although their selection proved difficult and hugely time consuming. In general, a network's learning rate is manually tuned
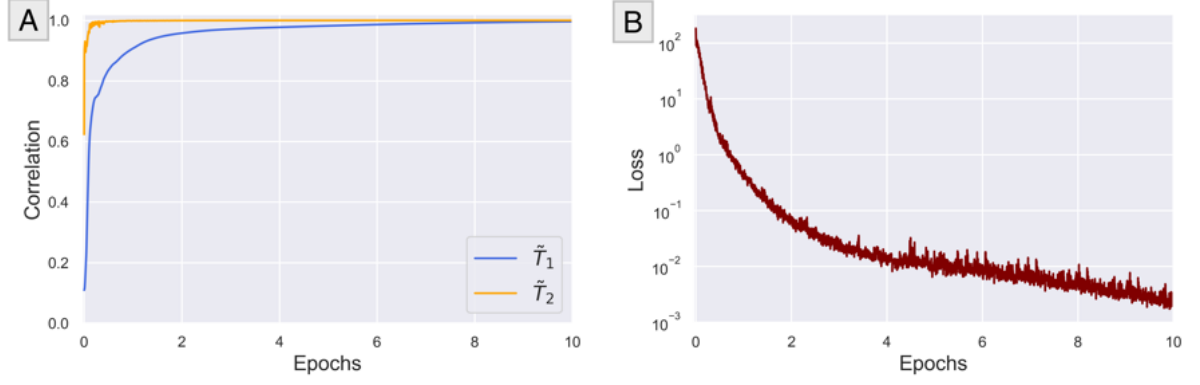
Figure 12: (A) **Correlations between estimated and ground truth transmission matrices**, denoted $\tilde{\mathbf{T}}_1$, $\tilde{\mathbf{T}}_2$ and $\mathbf{T}_1$, $\mathbf{T}_2$ respectively, over 10 epochs of gradient descent optimisation. (B) **Minimisation of the loss** quantifying the network's accuracy in predicting the measured intensity at the output plane.

to the value which best minimises the loss, but for our physically significant network, different systems may have different scaled TMs depending on factors such as medium's transmission. A-priori determination of hyperparameters would require prior knowledge of factors such as the medium's transmission properties, perhaps through measurement of the power transmitted to the output plane. Such knowledge would also prove invaluable in the initialisation of the network's weights at the correct scale. Alternatively, a dynamic learning rate selection could approach the problem such as [37], gradually increasing the learning rate during a preparation optimisation and taking the value for which the loss is minimised the quickest. While such an approach would certainly increase total training time, one may use the fact that the learning rate for $\mathbf{T}_2$ will be the square of $\mathbf{T}_1$, in similar fashion to the TM elements, and so only one learning rate need be manually tuned. A second approach was also trialled, where the training data was scaled so that a constant learning rate may be used for both layers. While this approach successfully retrieved the TMs, these estimates were scaled with respect to the ground truths; however, scaled TMs may still be used for the purpose of imaging and spatial control of light within the medium.

A similar issue also arises for the choice of hidden layer dimension, which must coincide with the number of fluorescent targets for optimal TM retrieval. For an unknown target number, choice of hidden layer is limited to trial and error, with the value for the hidden layer taken which provides the best minimisation of the loss. While this method is effective, the additional

training procedure likely results in a total training time longer than the stability times discussed above. For a continuous object, the hidden layer dimension would instead be liked to the image resolution, rather than the number of targets.

Occasionally during testing, the network failed to converge upon the ground truth TMs. Whilst for some of these failures the origin was unclear, a common fault was the appearance of negative terms within the real, positive $\mathbf{T}_2$. This proved difficult to tackle, with counter-measures seeming to detrimentally affect network convergence. All approaches trialled directly enforced a positive $\mathbf{T}_2$ at each step of optimisation, for instance through taking the absolute value of parameters or by setting all negative parameters to zero. A perhaps less intrusive method to encourage physicality within the network, rather than enforce, would be to incorporate a large penalty to the loss for negative elements of $\mathbf{T}_2$.

The two-layer neural network was also investigated in the presence of random noise. This was added to training data as:

$$\mathbf{I}^{\text{out}} = (1 + \rho \mathbf{N}) \, \mathbf{T}_2 |\mathbf{T}_1 \mathbf{E}^{\text{in}}|^2 \tag{21}$$

where $\mathbf{N}$ is the Gaussian distributed noise matrix and ρ is the noise level (1 gives a noise level equal to the average measured intensity). Figure 13A shows the neural network is robust to noise even for $\rho > 1$. While higher noise levels appear to ruin convergence, one may notice all training examples experience a sharp initial increase in correlation, often followed by a slow decline, possibly as the network over-fits to the changing noise level. This means for higher ρ, peak correlation is often reached before the end of training, and hence the optimal $\tilde{\mathbf{T}}_1$ may be taken from an earlier epoch. The peak correlation is plotted within figure 13B, and shows a power law relation x^n with $n \approx 2.03$. Later analysis (section 6.2) suggest a $\tilde{\mathbf{T}}_1$ correlation level of ≈ 0.95 is required for acceptable target localisation, indicating that imaging capabilities should be robust for roughly $\rho < 5$. To elevate analysis, improvements may be made to the modelling of noise by considering its origin. The most prominent source of noise is likely background fluorescence [35] and so improvements could begin by incorporating background targets into the model or perhaps by considering the interaction between adjacent targets.
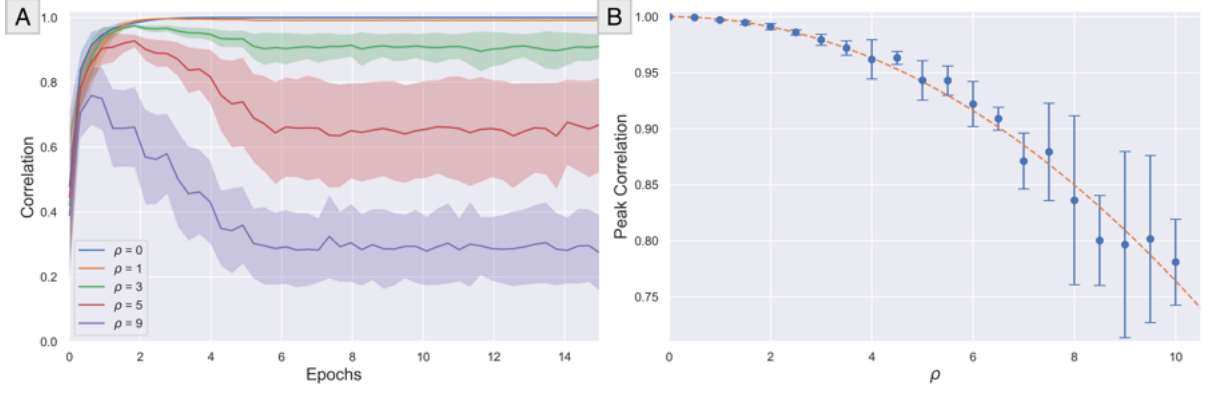
Figure 13: **Influence of noise on TM retrieval** (A) Moving average for $\mathbf{T}_1$ convergence for different Gaussian noise levels ρ, where $\rho = 0$ is no noise and $\rho = 1$ is noise equal to the average output intensity. 4 configurations were used with $N_{\text{target}} = 3$ and their average is plotted along with their standard deviation as the shaded area. Only $\mathbf{T}_1$ correlation is measured due to its importance to imaging (B) Peak correlation during training with noise ρ, along with the standard deviation over the 4 configurations. Dotted line shows a power law fit x^n with $n \approx 2.03$

6.2   Target Localisation

The estimated TMs are employed through ME imaging to produce the inter-target distances, which are then used by an MDS optimisation to construct the map of estimated positions (figure 14). Reconstruction using the shift ME may be achieved over a larger range than the tilt (x and y respectively); this can be understood through a visual inspection of the isoplanatic patch (figure 11), which extends roughly twice as far along x. However, shift ME measurements are to a lower precision due to the larger speckle size, with $\Delta y \approx 0.10\lambda$ while $\Delta x \approx 0.23\lambda$. One can also notice the MDS construction is more powerful for targets with several nearby neighbours, since this provides multiple strongly correlated distance measurements for reconstruction. For instance, targets in the centre of figure 14 are localised to a precision beyond L_{sp}.

Importantly, the imaging technique is limited by the target's sparsity and not their extent, meaning there is no theoretical limit to the size of the image plane so long as adjacent targets are within L_{me}. However, a strong memory effect requires a large medium to target distance D, which limits the application of the method. For instance, embedding two targets within the medium was tested but their positions could not be recovered with this technique. Additionally, isolated targets cannot be accurately located, due to a lack of highly correlated distance measurements.
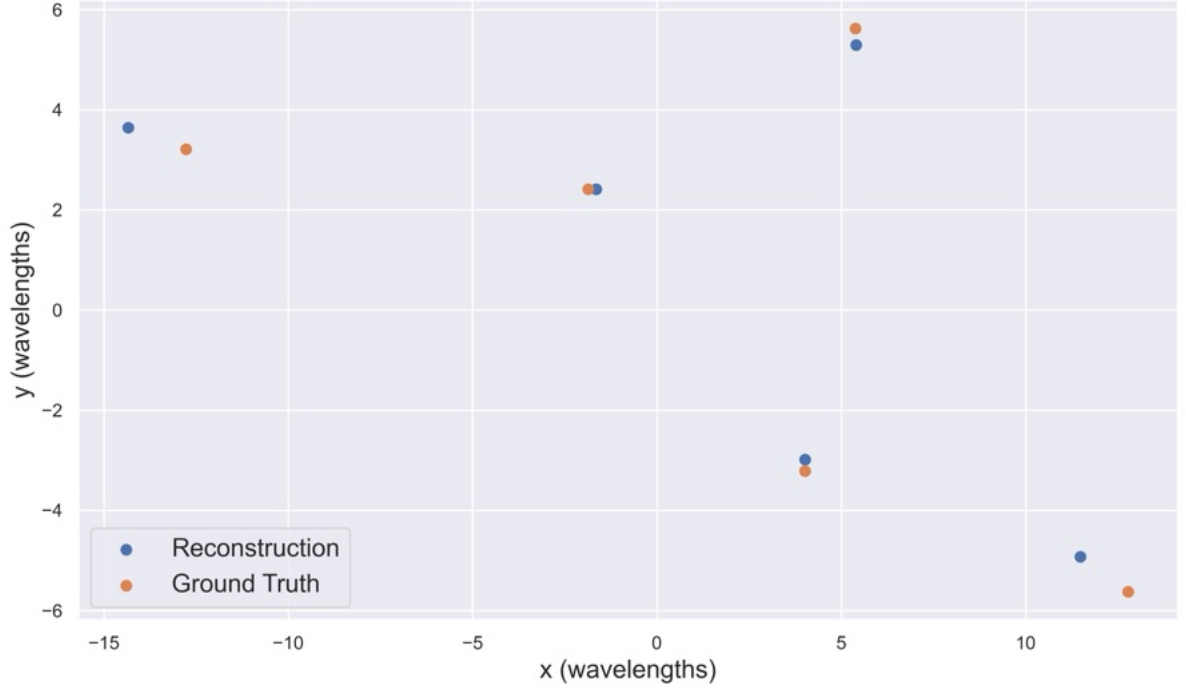
Figure 14: **Target localisation for 5 fluorescent targets using estimated TMs** through memory effect imaging and an MDS reconstruction. Reconstruction along x may be achieved over a larger range than y but to a lower precision, with the error in both reconstructions $\Delta y \approx 0.10\lambda$ and $\Delta x \approx 0.23\lambda$. Targets are located behind a scattering medium of $L_{\mathbf{me}} \approx 8\lambda$.

When training the network, TM correlations do not always reach unity, and so it is helpful to understand what level of convergence is sufficient to produce a strong reconstruction. Figure 15A shows a reduction in the localisation error for more highly correlated TMs along with a lower standard deviation, indicating reconstruction is not only more precise but also more consistent. When inspecting the same data plotted using a logscale (figure 15B), it is clear that the biggest improvement to the reconstruction ability occurs past correlations ≈ 0.95, with a sharp drop in error over several orders of magnitude. This can be viewed in two different ways: as shown in figure 12, optimisation of $\mathbf{T}_1$ for correlations roughly beyond 0.9 is the most difficult and time consuming, meaning the crucial improvements in reconstruction are also the hardest to obtain. The alternate perspective emphasises that correlations greater than 0.95 all produce reconstructions to a precision beyond the speckle ($L_{\mathrm{sp}} \approx 1.5$ for the figure 15 scattering media) and so correlations not converging to unity is not detrimental to imaging capability.

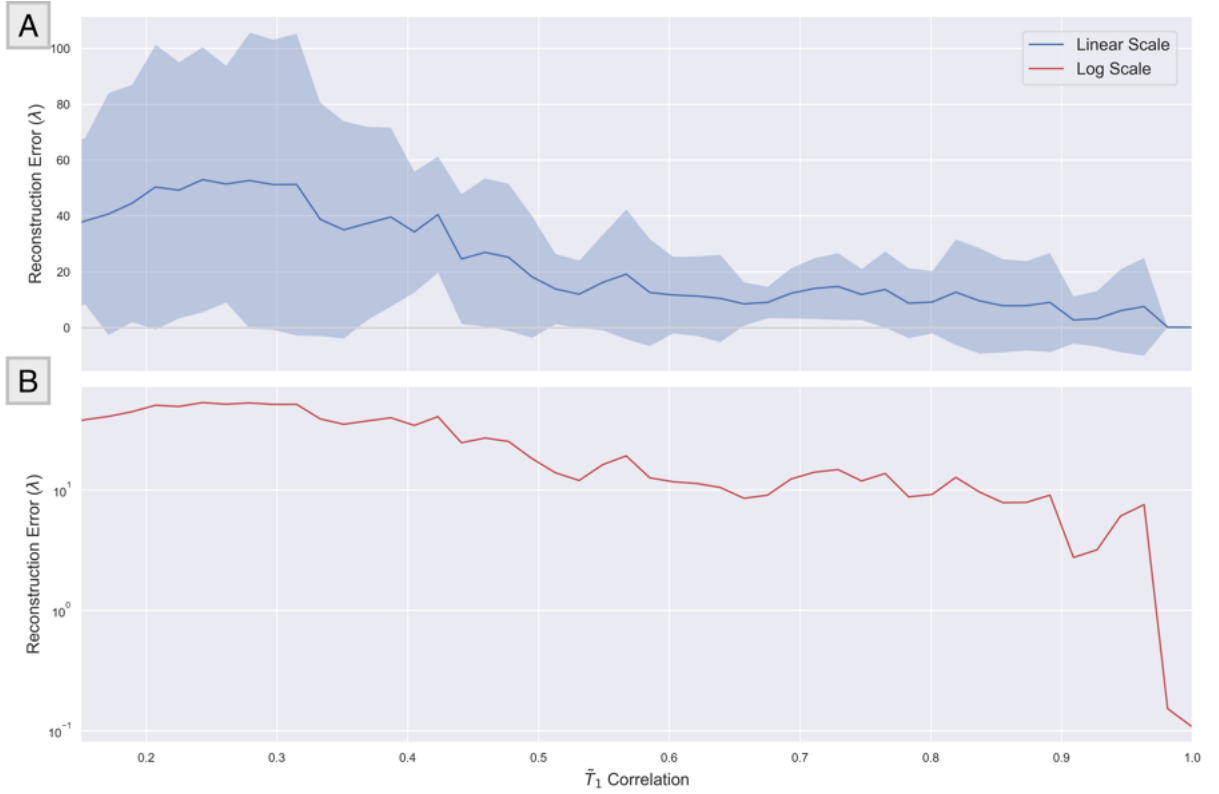The process also gives limited access to localisation of targets directly behind one another.

Figure 15: **The effect of TM convergence upon target reconstruction error**. (A) The error in predicted target positions along *y* (given in wavelengths) is plotted against TM convergence for reconstruction as a moving average, with its standard deviation given by the shaded region. 33000 data points were obtained over optimisation of 12 different configurations. (B) The same data is plotted in logscale, demonstrating the greatest gains to reconstruction precision are made past TM correlations of 0.95.

If one considers a system consisting of two image planes (figure 16A) a phase-conjugated TM, with each target as an output point, may produce a focus upon a target in the second plane without inducing an intensity response from one in the first, since the output field at the first is specified to 0 by the TM. This focus would only be restricted by the speckle size of the system L_{sp} along *x*. So long as the right hidden layer dimensionality is chosen, the network similarly retrieves the TMs and an image may be accurately constructed within the same limitations of L_{me} and L_{sp} (figure 16B). However, such behaviour may be indicative of limitations to the modelling approach. Analysis could be improved through a more rigorous modelling of the target's fluorescent responses; the model should account for the interaction of the target's incoherent emission, perhaps through a similar system of equations to the coherent DD approximation. Additionally, the TM only specifies the field to be zero at a specific point, and so a target of
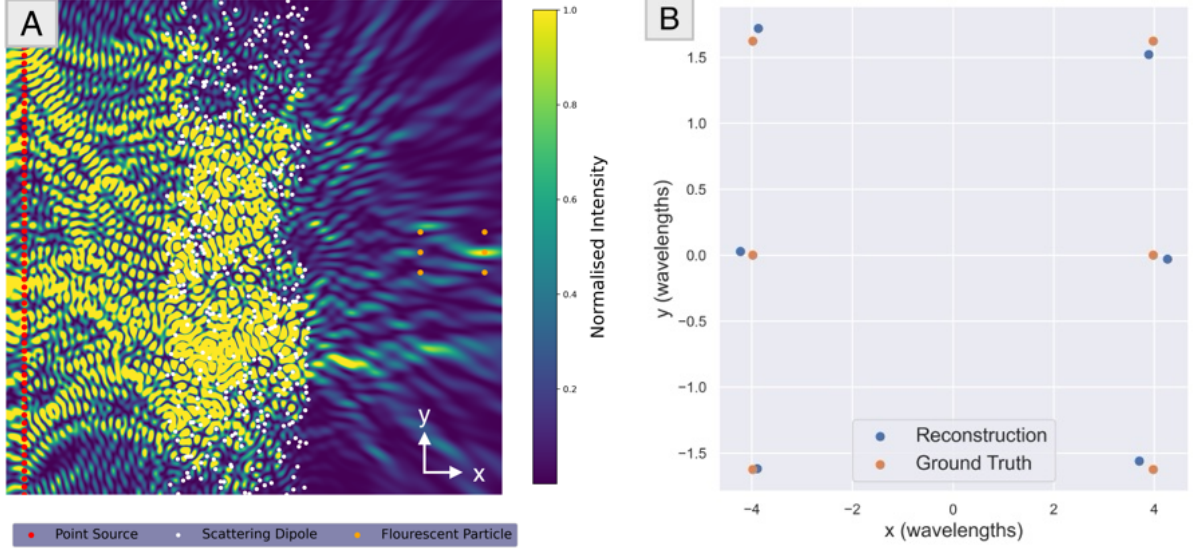
Figure 16: (A) **Focus at target located at back of two plane configuration**. The lower intensity incident at other targets should provoke a limited fluorescent response and therefore minimal interference (B) **Reconstruction of two plane configuration** where one layer of targets is located directly behind the other.

finite volume may fluoresce even if zero intensity is specified at its centre. Current modelling only uses point particles as targets, and so again more rigorous modelling should include volumetric targets. Further investigation may extend this to imaging a large continuous volumetric object by discretising the object into a series of focus points comprising the hidden layer of the network.

Conclusions

In summary, we have presented a two layer physics-based neural network to effectively characterise the propagation of light within a scattering system through the estimation of its transmission matrices. The network is robust to noise and provides a significant performance gain over previous work [12]. The network's main limitations lies with the necessity for a large training dataset along with the difficulties in hyperparameter choice; further investigation should consider an initialisation process grounded in easy to obtain physical priors. The retrieved matrices were used to focus at depth and inherent correlations in the fields (a memory effect) were employed to non-invasively image a series of fluorescent targets at varying distances behind the medium, without the use of phase information. The use of multi-dimensional scaling for im-

age construction has provided localisation precision beyond the characteristic speckle size. It has been shown that a strong image construction does not require perfectly retrieved transmission matrices, although the biggest precision gains are made at the latter stages of training. A more rigorous modelling of the fluorescent targets should be considered to further validate the approach, for instance through accurate modelling of background fluorescence or inter-target interference.

References

[1] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company publishers, 2005.

[2] S. Skipetrov, "Information transfer through disordered media by diffuse waves," *Physical Review E*, vol. 67, no. 3, p. 036621, 2003.

[3] A. P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink, "Controlling waves in space and time for imaging and focusing in complex media," *Nature photonics*, vol. 6, no. 5, pp. 283–292, 2012.

[4] J. Bertolotti and O. Katz, "Imaging in complex media," *Nature Physics*, vol. 18, no. 9, pp. 1008–1017, 2022.

[5] I. M. Vellekoop and A. Mosk, "Focusing coherent light through opaque strongly scattering media," *Optics letters*, vol. 32, no. 16, pp. 2309–2311, 2007.

[6] I. M. Vellekoop, A. Lagendijk, and A. Mosk, "Exploiting disorder for perfect focusing," *Nature photonics*, vol. 4, no. 5, pp. 320–322, 2010.

[7] B. Judkewitz, R. Horstmeyer, I. M. Vellekoop, I. N. Papadopoulos, and C. Yang, "Translation correlations in anisotropically scattering media," *Nature physics*, vol. 11, no. 8, pp. 684–689, 2015.

[8] H. Chen, M. M. Rogalski, and J. N. Anker, "Advances in functional x-ray imaging techniques and contrast agents," *Physical Chemistry Chemical Physics*, vol. 14, no. 39, pp. 13469–13486, 2012.

[9] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, "Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media," *Physical review letters*, vol. 104, no. 10, p. 100601, 2010.

[10] Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica*, vol. 5, no. 10, pp. 1181–1190, 2018.

[11] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," *Light: science & applications*, vol. 7, no. 1, p. 69, 2018.

[12] A. d'Arco, F. Xia, A. Boniface, J. Dong, and S. Gigan, "Physics-based neural network for non-invasive control of coherent light in scattering media," *Optics Express*, vol. 30, no. 17, pp. 30845–30856, 2022.

[13] S. Sukhov, D. Haefner, and A. Dogariu, "Coupled dipole method for modeling optical properties of large-scale random media," *Physical Review E*, vol. 77, no. 6, p. 066709, 2008.

[14] S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," *Nature communications*, vol. 1, no. 1, p. 81, 2010.

[15] C. H. Papas, *Theory of electromagnetic wave propagation*. Courier Corporation, 2014.

[16] S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Controlling light through optical disordered media: transmission matrix approach," *New Journal of Physics*, vol. 13, no. 12, p. 123021, 2011.

[17] J. Perina, *Coherence of light*. Springer Science & Business Media, 1985.

[18] S. Sukhov, D. Haefner, and A. Dogariu, "Coupled dipole method for modeling optical properties of large-scale random media," *Physical Review E*, vol. 77, no. 6, p. 066709, 2008.

[19] R. Pierrat, R. Carminati, and J.-L. Le Gouët, "Photon echoes in strongly scattering media: A diagrammatic approach," *Physical Review A*, vol. 97, no. 6, p. 063816, 2018.

[20] O. Leseur, R. Pierrat, and R. Carminati, "Spatial correlations of the spontaneous decay rate as a probe of dense and correlated disordered materials," *The European Physical Journal Special Topics*, vol. 226, pp. 1423–1432, 2017.

[21] C. C. Kwong, D. Wilkowski, D. Delande, and R. Pierrat, "Coherent light propagation through cold atomic clouds beyond the independent scattering approximation," *Physical Review A*, vol. 99, no. 4, p. 043806, 2019.

[22] S. Rotter and S. Gigan, "Light fields in complex media: Mesoscopic scattering meets wave control," *Reviews of Modern Physics*, vol. 89, no. 1, p. 015005, 2017.

[23] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[24] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasu-vunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[25] I. H. Sarker, "Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective," *SN Computer Science*, vol. 2, no. 3, p. 154, 2021.

[26] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer percep-tron and neural networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009.

[27] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, USA, 2015.

[28] H. Liu, B. Lang, M. Liu, and H. Yan, "Cnn and rnn based payload classification methods for attack detection," *Knowledge-Based Systems*, vol. 163, pp. 332–341, 2019.

[29] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE access*, vol. 8, pp. 4806–4813, 2019.

[30] N. O. Attoh-Okine, "Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance," *Advances in engineering software*, vol. 30, no. 4, pp. 291–302, 1999.

[31] G. Osnabrugge, R. Horstmeyer, I. N. Papadopoulos, B. Judkewitz, and I. M. Vellekoop, "Generalized optical memory effect," *Optica*, vol. 4, no. 8, pp. 886–892, 2017.

[32] S. Li, S. A. Horsley, T. Tyc, T. Čižmár, and D. B. Phillips, "Memory effect assisted imaging through multimode optical fibres," *Nature Communications*, vol. 12, no. 1, p. 3751, 2021.

[33] A. Boniface, B. Blochet, J. Dong, and S. Gigan, "Noninvasive light focusing in scattering media using speckle variance optimization," *Optica*, vol. 6, no. 11, pp. 1381–1385, 2019.

[34] Y. Okamoto, R. Horisaki, and J. Tanida, "Noninvasive three-dimensional imaging through scattering media by three-dimensional speckle correlation," *Optics letters*, vol. 44, no. 10, pp. 2526–2529, 2019.

[35] A. Boniface, J. Dong, and S. Gigan, "Non-invasive focusing and imaging in scattering media with a fluorescence-based transmission matrix," *Nature communications*, vol. 11, no. 1, p. 6154, 2020.

[36] M. Jang, H. Ruan, I. M. Vellekoop, B. Judkewitz, E. Chung, and C. Yang, "Relation between speckle decorrelation and optical phase conjugation (opc)-based turbidity suppression through dynamic scattering media: a study on in vivo mouse skin," *Biomedical optics express*, vol. 6, no. 1, pp. 72–85, 2015.

[37] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE winter conference on applications of computer vision (WACV)*, pp. 464–472, IEEE, 2017.